The influence of blood-group-related antigens on the intestinal microbiome

Dissertation

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Christian-Albrechts-Universität zu Kiel

vorgelegt von

Philipp Rausch

Kiel, 2015

Erster Gutachter: **Prof. Dr. John F. Baines**
Zweiter Gutachter: **Prof. Dr. Hinrich Schulenburg**

Tag der mündlichen Prüfung: **Kiel, 23.07.2015**
Zum Druck genehmigt: **Kiel,**

Der Dekan,

Für meine Familie, für meine Lieben.

Table of Contents

Zusammenfassung

Das AB0-Blutgruppensystem wurde vor mehr als hundert Jahren beschrieben, allerdings ist ihre biologische und evolutionäre Bedeutung noch immer nicht vollständig geklärt. Blutgruppenantigene, speziell Zuckerketten (Glykane), dienen vielen Bakterien als Nahrungsgrundlage, aber auch als wichtige Bindungsstelle mit der Darmschleimhaut. Unterschiede in diesen Antigenen haben somit Auswirkungen auf die Zusammensetzung und Funktionalität dieser Artengemeinschaften. Auch Umwelteinflüsse, wie das Nahrungsspektrum des Wirtes, oder die ihn umgebenden Bakteriengemeinschaften, sind relevante Faktoren die wirtsassoziierte Bakteriengemeinschaften beeinflussen und Auswirkung auf deren Entwicklung und die Entstehung Krankheiten haben. In dieser Arbeit untersuche ich den Einfluss von verschiedenen Umweltfaktoren und Blutgruppenantigenen auf die mikrobiellen Gemeinschaften des Darmes und ihre mögliche Rolle bei der Entstehung von Krankheiten.

Im ersten Kapitel dieser Doktorarbeit untersuche ich die Verteilung aktiver und inaktiver Darmbakterien zwischen verschiedenen menschlichen Populationen (Deutschland, Litauen, Indien), im Kontext entzündlicher Darmerkrankungen (Morbus Crohn, Colitis Ulcerosa). Die Analysen zeigen universelle- und populationsspezifische bakterielle Krankheits-charakteristika. Diese Unterschiede treten besonders deutlich unter den aktiven Bakterien zutage. Meine Resultate zeigen mögliche biogeographische Unterschiede der Darmflora zwischen verschiedenen Wirtspopulationen und populationsspezifische Krankheitscharakteristika auf. Darüber hinaus können die Unterschiede zwischen den erkrankten Probandenpopulationen Informationen für angepasste Behandlungsstrategien liefern.

Kapitel II und III befassen sich mit dem Einfluss des Morbus Crohn Risikogens *FUT2* (*α-1,2-Fucosyltransferase*) auf die Darmflora. Um die Auswirkung von Unterschieden in der Glykan Zusammensetzung zu untersuchen, nutzte ich humane Proben (Colonbiopsien, Galle) und Mausmodelle. Mit diesem Ansatz war es mir möglich die Veränderungen der Bakteriengemeinschaften durch genetische- und Umwelteinflüsse, wie auch während pathologischer Prozesse, näher zu untersuchen. Das *FUT2* Gen ist im Menschen sehr polymorph und spielt eine zentrale Rolle in der Synthese von AB0 Blutgruppenantigen in allen exokrinen Geweben. Für dieses Gen wurden viele konservierte Mutationen beschrieben, die die Synthese der Blutgruppenantigene unterbinden und im homozygoten Fall zur Abwesenheit von AB0-Antigenen in Geweben und Ausscheidungen führen („nonsecretor“ Mutationen). Diese Mutationen wurden mit Resistenzen gegen eine Vielzahl von Krankheitserregern in Verbindung gebracht, aber auch mit einer erhöhten Anfälligkeit für entzündliche Darmerkrankungen. In unseren Experimenten konnte ich nachweisen, dass

Expressionsunterschiede von Blutgruppenantigenen, wie auch Entzündungsprozesse, Auswirkungen auf die Zusammensetzung und die Diversität der Darmflora haben. Diese Unterschiede könnten mitunter die Anfälligkeit gegenüber chronischen entzündlichen Darmerkrankungen beeinflussen (Kapitel II). Auch ökologische Prozesse, wie zum Beispiel primär Sukzession und deren Interaktion mit Wirtscharakteristika (z.B. *FUT2*), sind von entscheidender Bedeutung für das Verständnis der Beziehung zwischen Wirt und seiner Bakteriengemeinschaft. Diese Interaktionen beeinflussen die Entwicklung, wie auch die Stabilität, von Bakteriengemeinschaften und sind somit bedeutend für deren Homöostase und Funktionalität (Kapitel III).

In Kapitel IV untersuche ich die Interaktion zwischen dem Blutgruppengen *B4galnt2* (*β-1,4-N-Acetylgalactosaminyltransferase 2*) und der Darmflora während einer Infektion mit *Salmonella enterica* ssp. ser. Typhimurium. In Mäusen führen Mutationen einer *cis*-regulatorischen Sequenz zu einer Veränderung des Expressionsmusters und einer Verlagerung der Expression vom Darmepithel zum vaskulären Endothel, was eine Blutungsdiathese zur Folge hat. Die Erhaltung eines solchen Allels über evolutionäre Zeitspannen legt einen Konflikt verschiedener Selektionsdrücke nahe. So wäre ein Kompromiss zwischen der Resistenz gegenüber enteropathogenen Erregern (Verlust von Antigenen im Darmepithel) und einer Beeinträchtigung der Blut-Homöostase ein mögliches Szenario. Ich konnte feststellen, dass Unterschiede in der Bakterienzusammensetzung zwischen den verschiedenen Genotypen zu einer höheren Stabilität der Bakteriengemeinschaft und niedrigereren Infektionsanfälligkeit in Mäusen ohne *B4galnt2* Expression im Darmepithel führen (Kapitel IV).

Zusammengenommen verdeutlichen diese Ergebnisse den Einfluss von Blutgruppenantigenen und Umweltfaktoren auf die Zusammensetzung und das Verhalten von mikrobiellen Gemeinschaften, sowie deren möglichen Einfluss auf die Fitness des jeweiligen Wirtes. Die Balance zwischen der Anpassung an lokale Bakteriengemeinschaften, Resistenz gegenüber Pathogenen und dem Verlust vorteilhafter symbiotischer Beziehungen und Funktionen, könnte somit eine Erklärung für die Variabilität von Genen wie *FUT2* und *B4galnt2* liefern. Diese Ergebnisse eröffnen auch Perspektiven für populations- und genotyp spezifische Behandlungs- und Präventivmaßnahmen bei entzündlichen Darmerkrankungen im Menschen.

Abstract

Blood groups were discovered more than one hundred years ago, but their evolutionary- and biological role is still not fully understood. Blood-group-related glycan structures can serve as attachment sites and nutrient sources for intestinal bacteria, thus, changes in the glycan repertoire may result in changes in bacterial community structure and functionality. However, environmental differences such as diet, hygiene, or the environment from which the bacterial community is recruited also exert strong influences on community structure. This variation, introduced by genetic or environmental factors, may also influence the development- and susceptibility to diseases. This thesis is dedicated to the analysis of microbial community characteristics with respect to (i) environmental variables (*i.e.* geography) and (ii) variation in blood-group-related antigens in the gastrointestinal tract and their possible contribution to disease development.

In the first chapter of this thesis I investigate the differences of the active and stagnant microbial communities between human populations (Germany, Lithuania, and India) in healthy- and diseased subjects (Crohn Disease, Ulcerative Colitis). This allowed me to identify universal- and population-specific patterns of microbial communities and dysbiosis, specifically among the active microbial community members. My results indicate biogeographic patterns in mucosa associated microbial communities and population-specific disease signatures, which may entail specific treatment strategies in the future.

Chapters II and III aim to explore the impact of variation in the Crohn Disease risk locus *FUT2* (*α-1,2-fucosyltransferase*) and inflammatory diseases on the microbial communities. To investigate the influence of host-specific glycan composition on the microbial community I took different approaches, using human sample material (colonic biopsies, bile samples) and mouse models. This allowed me to investigate changes in the ecology of the host associated microbial communities, through genetic, environmental, and pathological influences. The *FUT2* gene shows widespread sequence variation in human populations and is required for expression of ABO blood group antigens in all bodily secretions and mucosal surfaces. Several highly conserved mutations have been identified in different human populations, which eliminate blood group expression in excretory tissues (so called “nonsecretor” mutations). These mutations have been associated with decreased susceptibility to several infectious agents, but increased susceptibility to inflammatory bowel diseases. I could show how this variation in blood group expression translates into changes of the composition and diversity of microbial communities, potentially influencing susceptibility to chronic inflammation in the gastrointestinal tract (Chapter II). Furthermore, ecological processes like initial colonization, succession, and their intricate interaction with host genetic traits are of crucial importance to understand host-microbiome interactions and

represent potential drivers in the development of community imbalances or disease (Chapter III).

In chapter IV I investigate the role of interactions between the blood-group-related gene *B4galnt2* (*β-1,4-N-Acetylgalactosaminyltransferase 2*) and the intestinal microbiota in susceptibility to the enteric pathogen *Salmonella enterica* ssp. ser. Typhimurium. Naturally occurring *cis*-regulatory variation at *B4galnt2* leads to a tissue-specific switch from gut epithelial expression to expression in the vascular endothelium and results in a bleeding diathesis. The maintenance of such variation could be related to a trade-off scenario between resistance to enteric pathogens (absence of certain glycans in the gut) and prolonged bleeding after injury. I identified microbial community factors present in mice without gut epithelial *B4galnt2* expression, which are responsible for a higher community resilience and lower susceptibility to *Salmonella* infection.

In summary, these findings provide strong evidence for the influence of blood-group-related antigens and the environment on the microbial communities with potential fitness consequences for the host. The balance between adaptation to local microbial communities, pathogen resistance, and loss of potentially beneficial bacterial symbionts and functions might thus contribute to the patterns of long-term-balancing selection in genes like *FUT2* and *B4galnt2*. These results may lead to future population- and genotype-specific measures for treatment and prevention of inflammatory bowel diseases.

Introduction

Microbial communities-structure, function, and distribution

With the advent of high throughput sequencing [1-3] we gained great insight not only into the molecular makeup and inner workings of the human genome, but also into the diversity of our “second genome” encoded in the vast microbial communities within and around the human body [4-6]. Bacteria and archaea have shared their environments for three billion years and even sparked the rise of eukaryotic life through several instances of endosymbiosis (Cyanobacteria-Chloroplast; α-Proteobacteria-Mitochondria) [7, 8] and shaped major transitions in the biotic and abiotic environment [9-11]. Besides the intimate endosymbiotic relationship of eukaryotic organisms with prokaryotes, single- or multi-cellular organisms interacted and continue to interact in a wide variety of symbiotic and antagonistic ways that shaped the evolution of both sides [12]. Not only biotic, but also abiotic habitats are constantly colonized and influenced by prokaryotic life, ranging from common habitats (*e.g.* benthic- and marine water, soil, stratosphere) to the most extreme environments (*e.g.* hydrothermal vents, acid mine drainage, deep sea sediments, deep lithosphere; [13-16]) and manmade habitats (*e.g.* apartments, aquifers, water treatment plants) [17-19]. This incredible diversity of habitats, the long evolutionary time, and huge population sizes left their traces and re-enforced the diversification of prokaryotic species in conjunction with their genomic and metabolic capacities. This constant competition for resources, space, and even genetic building blocks [20] promoted the occupation of new emerging niches and speciation within and around them. Bacteria are therefore not only the basis of almost all food chains, directly as primary producers or indirectly as essential parts of primary producers (*i.e.* chloroplasts, nodula symbionts, mitochondria), but also can influence each intermediate step of them (*e.g.* anaerobic respiration, interspecies hydrogen transfer, methanogenesis, acetogenesis). Thus, they are able to extend host functions such as pathogen resistance [21, 22] or alter ecosystem functions and stability on a macro-ecological scale [23, 24].

Microbial communities range in their complexity from the very well characterized, low diversity systems like the acid mine biofilm [13, 25], to highly heterogeneous and phylogenetically rich assemblages found *e.g.* in the soil [26, 27]. Members of those assemblages appear highly adapted to their environment and to other bacteria sharing their respective habitats. The observation of highly stable and intricate metabolic cycles in cubic centimeters of soil [28-30] to the most abundant carbon fixing bacteria travelling through the seas (*e.g. Pelagibacter spec.*) are examples for the range from locally restricted to global interaction [31-33]. Together with their high abundance and metabolic versatility, bacterial communities are major keystones in most local and global matter and energy cycles [34]. All of these communities are influenced by diverse ecological forces which can even be

generalized and extended to population genetic concepts like drift (*i.e.* ecological drift), selection (*i.e.* environmental selection, species sorting), dispersal, and mutation/speciation at each stage of community development [35, 36].

An interesting feature of microbial communities is their distribution over the globe. The dominant theory of microbial biogeography states a continuing colonization of randomly dispersing bacteria over a global scale, which is followed by environmental filtering of those bacterial assemblies by biotic or abiotic (*i.e*. "everything is everywhere, but the environment selects") [37]. Active dispersal of microbes, however, is limited to a relatively small scale (~40 µm/s), due to cell size and mode of motility (flagellar propelling [38], twitching motility [39], gliding [40], or sliding [41]). Passive dispersal by the means of phoresis [42] or by environmental currents (*e.g*. air, water), seems possible to occur on a global scale [43, 44]. The passive dispersal in conjunction with small cell size and high population densities, enables microbes to disperse in high numbers even on a global scale if they can tolerate the physiological stress [45]. This implies that through the immense number of cells and high reproduction individual microbial species probably do not go extinct, disperse continuously and passively. Environmental studies observed striking ubiquitous bacterial distributions among habitat types [44, 46], which can even show different or opposing trends to macro biological systems (*e.g.* lack of elevation gradients) [47, 48]. Moreover, whether succession leads to an increase in microbial community diversity [49, 50], saturation and stabilization of climax communities is debatable [51].

Microbial communities also show similarity distance-decay relationships [52], which represents a hallmark of neutral, distance- and regional pool- dependent dispersal. Furthermore, different but adjacent habitats often show very different microbial communities, which makes pure random dispersal and settlement unlikely to be the main effect driving community differentiation [53]. However, microbial communities do not seem to disperse and settle completely at random in the environment as their composition is influenced by latitude, depth, and other abiotic factors [54, 55]. These differences seem to be an amalgam of environmental selection, neutral, and even historical effects. So far it is almost impossible to distinguish between historical effects and the regionally acting contemporary environment, despite by analyses of endemicity of bacterial groups [56], or reconstruction of habitat characteristics by historical samples [57]. Endemism of bacterial taxa would be a strong sign of ancestral/historical overlapping communities [56, 58].To definitively prove the endemicity of bacteria is an impossible task of explicitly showing the absence of a bacterial group conclusively through shear sampling limitations and scale, while presence alone also does not mean the bacterium is an autochthonous and active member of the community. However, some examples of potential endemicity exist [59, 60], even though most investigations so far

revealed broad habitat distributions of bacteria [44, 45], but little overlap among soil-, freshwater-, marine- and host-associated habitat types [12]. Furthermore, these observations are also highly influenced by the timescale under examination as seasonality, incidental disturbances, and succession alters community patterns, especially in environments with high dispersal [52, 61, 62]. Microbial communities can change at daily [63], seasonal [62, 64, 65], or annual scales [28, 62, 66]. Therefore, determining factors that are important to shape microbial communities are not only dependent on the spatial-, but also on the temporal scale, specifically as size and generation time of many bacteria are small and environmental influences ubiquitous.

The heterogeneity of biogeographic concepts in microbial community ecology is illustrated in the already mentioned distance-decay relationships, where similarity of species assemblages decreases with physical distance among communities. Several studies showed this effect among microbial communities [35], but it is most readily detected in the extreme cases (*e.g.* distance <7 km [52, 54], continental distance [67]) due to a balance between ecological drift and colonization. These observations are further influenced by time itself, as seasonality and succession alter community patterns at a daily [63], seasonal [62, 64, 65], or annual scale [28, 62, 66]. Also undetected environmental factors, which are auto-correlated with the spatiotemporal distance between communities, can overwrite or induce false interpretations of microbial community relationships.

Patchy species distributions, which occur through dispersal limitation and the action of ecological drift, are often observed in host associated communities. These communities are contained within a host and cannot be easily transferred without direct contact (*i.e.* dispersal limitation [68, 69]), while individuality of hosts in terms of diet [12, 70, 71], resident community [22], genetics [72-74], and random community fluctuations enforce differentiation among host communities, increase individuality and patchiness (*i.e.* species sorting and ecological drift [75]). Leaf associated microbial communities of *Pinus ponderosa* apparently adapted to the host species instead of being differentiated by geographic or environmental distance on a transcontinental scale [53], as does the root associated microbial community of *Zea mais* varieties [76]. This may be a result of selective recruitment of specific facultative bacterial symbionts from the environment as observed in nodula (*i.e.* dispersal limitation, see paragraph "*Host associated microbial communities*"), or other undetected environmental factors correlated to the specific habitats and not considered in these analyses. However, host associated communities that offer highly individualized niches (genetics, diet, etc.) with potentially strong dispersal limitations are another promising venue to investigate not only endemism and bacterial biogeography, but also how co-evolutionary- and community ecological patterns shape those processes. In Chapter I we investigate the influences of host

population and potentially associated life style, and environmental differences on the gut microbiota in inflammatory bowel disease patients and healthy controls.

Host associated microbial communities

Niche specialization has led to the early colonization of biotic habitats first available with the development of metazoan life forms. So are the surfaces of nearly all plant [53, 76-79] and animal [12] life inhabited by a high number of prokaryotic cells, potentially by more than their actual bodies consist of. For example, the colon of an adult human is home to a staggering number of microorganisms of more than 100 trillion individuals, outnumbering all other microbial communities associated with other body surfaces [80-82]. The taxonomic variation of community members in the vertebrate bowel and over body surfaces is very high, lacking a clearly defined shared core flora (at high taxonomic resolution) already within a single host species and thus display a potential high degree endemicity [83]. This variation, however, is strongest on high taxonomic levels (species, genera diversity) and decreases at lower taxonomic resolution (phylum diversity). However variable the observed communities are, there is a certain consistency in the functional spectrum of the communities within and among host populations and host associated microbial populations [84-86]. This highlights the functional redundancy in those assemblages, and the dependence on a minimal set of core functions to sustain those communities and the interaction with the host [71, 85, 87]. The overlapping requirements and sheer number of bacteria usually associated to hosts drives the abundance of those highly redundant functions and genes, but also facilitates the occurrence of rare, unique functions which fine-tune and connect functional groups/compartments of the microbiome [85, 87]. These microbial functions supplement functions of the host, such as polysaccharide digestion [88] and the synthesis of essential nutrients [89, 90], in addition to conferring resistance against pathogens and parasites [22, 91-93]. Even though the functional variation is smaller than taxonomic variation among hosts, different community compositions can lead to functional and metabolic repertoires with direct consequences for the host. As seen in aphids, the restrictive vertical transmission and the associated genetic and taxonomic bottleneck of the symbionts (transmission bottleneck) leaves its footprints in the bacterial genomes and reinforces the association of a specific microbial community to its host-species and vice versa [94, 95], with clear signs of co-evolution with their host species.

Eukaryotic organisms can therefore be described as “*holobionts*”, an aggregate of interacting eukaryotic and prokaryotic organisms, genes, and gene products, lending features the host could not evolve alone, and have selection acting on both sides of the

symbiosis [96, 97]. This explains to some extent the relative low diversity and specificity of phyla in host associated bacterial communities, which consist mainly of Bacteroidetes and Firmicutes, and to a lesser extent Proteobacteria, Actinobacteria, Verrrumicrobia [86]. However, even though there is little diversity in the large phylogenetic groups, species and genera show immense diversity within hosts and have strong associations to certain groups such that community composition appears to be relatively correlated to genetic divergence of hosts [71, 98, 99], even though differences in diet can outweigh these patterns by overwriting the selective nutrient environment offered by the host [71]. Furthermore, only a small number of taxa have been shown to be shared across environmental and host associated communities [12, 69, 100], which indicates strong species sorting by the host environment on the potential symbionts [69, 101-103]. Recruitment and environmental pressures further drive specialization of the functional and taxonomic spectrum of the holobiont, specifically diet [12, 70, 71, 100, 104]. However, depending on the source communities (biotic, abiotic, species specific, species nonspecific) and already residing bacterial populations, the succession process can have very different dynamics, but appears to favor bacterial communities pre-adapted to the respective host environment [69]. The coevolution of symbiotic communities and its host can lead to the loss of essential genes in either host or symbiont and go so far as to facilitate speciation of hosts while preventing hybridization with sister species not carrying compatible microbes [105, 106]. How those selective forces are acting on the "*holobiont*" and potentially reinforce speciation is still under debate [107, 108].

Microbial communities of multicellular organisms appear to be recruited by several mechanisms, being it gradients and interplay of certain small effector molecules (flavonoids, nod factor, [109]) or by the induction of trapping mechanisms or suitable attachment sites in plants and animals (*e.g*. cilia, mucus [110]). These weak initial interactions, mostly facilitated through an interplay of glycan residues with binding proteins (*i.e.* pili, lectins, adhesins, LPS binding proteins) can later solidify [111]. These recruitment mechanisms date back to the origin of metazoans, with the development of C-type lectins (glycan binding protein), which are important for bactivory and initiation of colony formation, and might have been one of the first steps to multicellularity [112, 113]. As observed in invertebrates, colonization in vertebrates appears to be quite specific as well [69], and could even happen before the first direct contact to the external environment via passage through the placenta [114, 115]. The initial transmission is by definition associated with a taxonomic and genetic bottleneck of the communities, which influences the dynamics of community and host fitness over a long period [116-118]. Like most macro-ecological systems, the gut microbial community is characterized by processes like succession, and appears to be very deterministic. Infant bacterial communities show common transitions [117, 119], with a transition of early generalists to adapted specialists [117, 119], or the shift from facultative anaerobes to strict

anaerobes [120]. Early communities are thus relatively similar to each other, but distinct from an adult microbiome. These patterns are influenced by the starting conditions and early community priming, through *e.g.* delivery mode, early nutrition, or medication, but also by the surrounding source environment [69, 121, 122]. The host innate- and adaptive immune system, or other colonization barriers like the gastric passage, are other obstacles for potential colonists that are in many ways unique to the respective individual. In addition, the intrinsic adaptive capability of the already present microbiota exerts a strong selective force on potentially colonizing bacteria [69], which also acts as a colonization barrier for pathogens via competition or bactericidal activity [22, 69, 91-93, 123]. Together environmental-, physiological-, neutral-, and community intrinsic forces shape a taxonomically highly individualized microbial community. Selection on the recruitment of a certain microbial community by the host seems therefore plausible [96, 105], as some bacteria are heritable and have a direct influence on the host metabolism [124]. Further, breast milk and its specific glycan spectrum attract and nourish a variety of specific bacteria, and thus set the stage for the deterministic succession sequence [125, 126]. These processes appear to enforce coevolution and heredity of microbial community members through the transfer of bacteria from mother to newborn via delivery and nourishment, but also by enforcing the colonization of specific and potentially beneficial bacteria through milk [121, 124, 125].

Genetic elements that alter community composition and/or favor the propagation of probiotic bacteria (or viruses) could rise in frequency, together with their associated community members. This may therefore explain the relative similarity of the human associated microbiome in comparison to environmental bacterial communities [12], but also the dependence of hosts and certain bacteria on each other [91, 127-136]. Specifically in early stages of human development, infections of the intestinal tract can result in a high mortality [137], making a resistance mechanism like a resistant/resilient microbial community highly beneficial and obtainable via the recruitment and interference with bacteria through breast milk. Depending on the underlying tissue characteristics and set of bacteria able to colonize, the communities become different along the physicochemical gradients of the host body. As such, microbial communities differ within the gastrointestinal tract [138], but also among exposed body surfaces such as skin [139], and other mucous environments like the vaginal tract [140] or the oral cavity [86, 141]. These differences between body sites and locations within a body site correlate strongly with the underlying attributes of the respective site, such as nutrient availability, pH and moisture, and can be changed through external disturbances (*e.g.* antibiotics [142], washing [143]). Recolonization from neighboring body sites and environmental sources (*e.g.* diet) [69, 144, 145] can therefore continuously replenish the communities and reduce the impact of stochastic species loss by ecological drift [145, 146], if bacteria are able to cope with those conditions.

Hosts can be pictured as uninhabited (newborn host, gnotobiotic animal) or already inhabited islands (colonized host) that are subject to colonization by source communities (other already colonized hosts, environment) [69, 75, 147, 148]. Physical proximity thereby dictates the likelihood of colonization from the local species pool, while environmental factors and the resident community alter the probability of establishment and extinction, comparable to succession on islands [69, 149, 150]. Another interesting parallel between island biogeography and host associated microbial communities is the high level of endemism at the low taxonomic levels [151]. This may be the result of individual selective pressures on the microbial communities, such as rapid occupation of niches, immune system evasion, and adaptive radiations (including horizontal gene transfer) within the relatively isolated communities of the gastrointestinal system [152]. Further, as niche space [146] among conspecific hosts is comparable but colonization is different due to a multitude of factors, the taxonomic composition may differ substantially, while community functions are conserved. Another possibility is that the functional repertoire of the community converges due to similar resources, while taxonomically the communities substantially differ [86, 153]. Microbial communities further show a tendency to associate with subpopulations of hosts, which drives the emergence of the recently described “enterotypes” or “enterogradients” [154] in several host species [155-157] and host tissues [140, 154, 156]. These “community types” appear independent of ethnicity and physical distance among hosts, while some individual host factors (*e.g.* diet, gender) might play a role [104, 158, 159] as do intrinsic community dynamics like stochastic species loss [146, 160]. These patterns may ultimately not be “enterotypes”, but “enterogradients”, and represent local optima dominated by specific driver microbes [154, 160].

These community differences, facilitated by environmental differences, differences in host genetics and life style, may however change the host response to environmental influences and even contribute to disease susceptibility. As part of this thesis I investigated disease associated changes in microbiome composition across different human populations and disease cohorts (Germany, Lithuania, and India). This led to the discovery of population specific and inter-population responses of the microbiome to inflammatory bowel diseases (see Chapter I).

The role and evolution of host glycans

In recent years, molecular biology mainly focused on the biological entities of DNA, RNA, and Proteins. Oligosaccharides are so far relatively undervalued for their role in evolution and development, despite playing important roles in protein folding and cell trafficking/regulation [161-163]. Glycans further developed a variety of other functions since they became part of the eukaryotic cell, *i.e.* developmental cues, structural components (chondroitin, hyaluronic acid), facilitators of cell-cell, and cell-matrix interactions, barriers, and protective layers [164]. Almost all cells are covered with different sugar chains ("glycocalyx"); either attached to lipids, proteins or excreted, and are thus the direct interface to the environment. Glycan profiles, the spectrum of distinct carbohydrate chains, show high inter- and intraspecific variability (*i.e.* hyaluronic acids restricted to vertebrates, N-glycolylneuraminic acid lost in humans [165]), but also conserved themes (N-glycosylation [166]) and strong conservation of glycosyltransferase functions among eukaryotes, bacteria and archaea despite high sequence divergence [167, 168], or even independent origin (convergence, [168-170]). These glycans are synthesized by a huge variety of glycosyltransferases into linear and branched structures by specific elongation at reactive sites of lipids, proteins, or saccharide precursor molecules (*e.g*. UDP-Glucose, GDP-Fucose, CMP-Sialic acid) during their passage through the Endoplasmic Reticulum and Golgi apparatus, and are thus major energy commitments for cells. During the building process, variation in glycan chains can arise, so called micro-heterogeneity, which either emerges through the speed of synthesis, different rates of glycosyltransferases, and/or compartmentalization of the synthesis. Together with the high influence of bond orientation and bond location, glycans can produce a stunning variability of structures even between different glycosylation sites of the same protein [171]. Thus, glycan structures are not hard-coded in the genome like protein- or RNA sequences, but depend on the pace of synthesis, available substrate pools and state of the cell, despite their high substrate specificities [172-174]. Small environmental differences and cell location can therefore lead to changes in the composition of cells and their products, leading to short term diversity- and variation generating machinery, which does not necessarily lead to long lasting, detrimental, and heritable effects in cell homeostasis as changes in the genome do. Glycan profiles, specifically the terminal branches, are able to change quickly in response to external and internal stressors. This allows the organism/cell to adapt rapidly to new environmental characteristics, which is important for example in embryonic development [175-178]. However, the array of glycans tends to be relatively conserved within the same species and is met by the variation of glycan binding proteins called lectins [179]. Furthermore, several glycans usually interact to facilitate a specific or unspecific binding, which thus allows certain variation without the loss functionality or cell recognition [164, 180]. A complete and

unbalanced loss of certain glycans, however, has developmental and even lethal effects, so-called congenital disorders of glycosylation [181-184].

A well-known example for intraspecific glycan polymorphism is the ABO blood group system in humans, which has parallels in other primate species [185] and long term conserved genetic polymorphisms [186, 187], hallmarks of balancing selection [187]. These genes have several homologs in a wide variety of other species and thus a long evolutionary history, in which single genes merged, were regained, duplicated, or lost [188]. Until recently the leading hypothesis for the co-occurrence of ABO blood group polymorphism in primates was due to the action of convergent evolution. In this scenario the ancestral A allele duplicated and independently evolved into B, which was lost and regained at least six times in the course of evolution [189, 190] and appears to show expression differences between new- and old world monkeys (presence on erythrocytes or tissue, respectively) [191]. However, evidence for long-term balancing selection conserving this polymorphism for at least 20 million years has regained support [185, 186, 192]. This makes the ABO polymorphism one of the rare validated trans-species polymorphisms known today (*e.g.* MHC/HLA locus). Already on the level of humans, the polymorphism of this locus on chromosome 9 is astonishing, as for each antigen class (A, B, and O/H) numerous alleles (358 alleles in the blood group antigen mutation database, June 2015) and high nucleotide diversity have been identified [187]. Since the discovery of the ABO blood groups [186, 193-195] and their molecular and genetic underpinnings [196-199], scientists hypothesized about their ancestral function and evolution [200]. The preferred hypothesis so far suggests an ancestral role in cell-cell adhesion and development [201], which over the course of gene duplication and neofunctionalization developed into the blood group system, which plays a role in host-pathogen coevolution [202, 203]. Carriers of different blood- and histo-blood groups accordingly display differential susceptibilities to pathogens [204] and other maladies [205-210], but do not show any other disorders. Blood-group antigens usually occur with a specific non-self antibody (allograft recognition), with strong binding affinity and the ability to recruit the complement system that leads to opsonization and lysis of cells [211]. This feature and the polymorphic nature of the potential attachment sites, can prevent or reduce the vertical spread of encapsulated virus particles or other pathogens through a population differing in its cell surface antigens [202]. Bacteria carry highly diverged glycosyltransferases [188], allowing them to express ABO surface glycans [188, 212-214], which can serve as targets of the respective host antibodies [212, 213, 215, 216]. The antibody response to non-self ABO blood groups needs to be primed first by an interaction of the developing immune system with non-self antigens. The microbial community and other sources of environmental antigens seem to be play an important role for priming the immune system for their respective ABO blood group early in life [213, 215-217]. This early immune priming through

enteric- or environmental bacteria even extends to the production of antibodies targeting other pathogens such as malaria [218].

Other ways to counter the adaptation to-, and exploitation of glycan variation by potential pathogens is the abandonment of glycosylation pathways [219-221] or the loss of glycan binding/signaling molecules that act as an entry point for pathogens (*e.g. SIGLEC's* [222, 223]). This of course also comes at a cost, such as mismatches of zygotes [224], failing zygote implantation [183] and developmental defects if certain glycan structures cannot be built properly [181-184], or if they are exploited by metastatic cancer cells [225, 226]. This variation even leads to glycan mediated reproductive isolation, as incompatibilities at the level of reproductive cells, implantation tissue, or during ontogeny evolve and could fix under selective pressure from pathogens or other linked benefits (*e.g.* increased reproduction) and facilitate genetic and glycan differentiation among populations [219, 227-229]. Furthermore, glycosylated proteins or vesicles also act as decoys that exploit the binding specificity of an invading pathogen and prevent/reduce direct tissue attachment [230-232].

ABO antigens are not only present on the surface of erythrocytes, but in a high percentage of individuals in exocrine tissues and excretions, the so called histo-blood groups [233]. Many studies revealed fucosylated glycans, specifically ABO related antigens, in high abundance in the gastrointestinal mucosa [234-236] and breast milk [237, 238]. The genes responsible for the initial steps in glycan synthesis of the H/O-antigen are paralogous to their hematopoietic counterparts (*FUT1*-H/O blood group gene, *FUT2*-secreted H/O blood group/*Se* gene, *Sec1*-pseudogenized secreted blood group gene in humans, Chromosome 19q13.3), but overlap in several glycan modifying down-stream pathways even though they differ in their expression profile (*FUT1*: endothelium, bone marrow; *FUT2, FUT3*: mucosal surfaces and exocrine glands) [163, 239-241]. The *α-1,2-fucosyltransferases* encoded by *FUT1* and *FUT2* use similar precursor molecules (Galβ1-3/4GlcNAc-R), which are then fucosylated at the galactose residue (type 1 precursor for *FUT2:* Galβ1-3GlcNAc-R; type 2 precursor for *FUT1*: Galβ1-4GlcNAc-R). This H/O-antigen serves as the substrate for the A- and B glycosyltransferases (*α-1,3-GalNAc transferase* and *α-1,3-Gal transferase*), which add N-acetyl-D-galactosamine or a Galactose to the Galactose residue of the H/O-antigen and represent a significant component of the mucosal lining in association with glycoproteins like mucins [234-236]. The unfucosylated- and fucosylated type 1 glycans can also be modified by the *Le* gene cluster (*FUT5-FUT3-FUT6*) via *α-1,3-fucosyltransferase* or *α-1,4-fucosyltransferase* activity, giving rise to Le^a or Le^b (for type 2 precursors Le^x / Le^y, respectively). These products then reside on lipids or proteins, either bound to the cell membrane (in the case of the *FUT1* pathway), excreted (in the case of *FUT2*), or passively taken up by erythrocytes (*FUT2* H-antigen, *FUT3* Lewis antigens).

Many other fucosyltransferase genes have been identified in humans with differing degrees of conservation between homologs and paralogs [169]. These enzymes can be divided into two families based on conserved peptide motifs (group one- *α-1,2-fucosyltransferases*: *FUT1*, *FUT2*, *Sec1*; *α-1,6-fucosyltransferases*: *FUT8*; *protein-O-fucosyltransferases*: *POFUT1*, *POFUT2*) [242]. The second group comprises of *α-1,3-fucosyltransferases* (*FUT4-FUT7*, *FUT9-FUT11*), and the *α-1,3/4-fucosyltransferases* (*FUT3*, *FUT5*) [243]. These fucosyltransferase families seem to originate from successive duplication events followed by neofunctionalization [169]. Several fucosyltransferases share similar acceptor molecules and reactions, but differ in their substrate affinity, location of expression and developmental timing [244]. For example, *FUT3* is expressed in the epidermal tissue (also secreted) encoding a *α-1,3/4-fucosyltransferase*, with broad binding spectrum but a higher affinity to type 1 acceptors (Le^a/Le^b) than for type 2 precursors molecules (Le^x/Le^y) [245]. *FUT6*, the "plasma-type" fucosyltransferase, shows a very high sequence similarity and a small chromosomal distance to *FUT3*, also with its main product Le^x. However, the acceptor molecule, glycosylation targets, and tissue expression are very different from the *FUT3* enzyme (reviewed in [246]).

For several of those genes polymorphisms are known (*FUT1* [247], *FUT2* [239, 248], *FUT3* [249, 250], *FUT5* [251], *FUT6* [252, 253], *FUT7* [254, 255], *FUT8* [256], *FUT9* [257], POFUT1 [258]), however at very different frequencies as some are redundant and others essential for developmental processes (*i.e. POFUT1*, *POFUT2*) [259]. Especially mutations of the *FUT2* gene are very frequent and diverse [260], with a center of diversity in Africa and several independent population-specific haplotypes [220, 239, 248, 261-263]. The most common loss-of-function mutation in Caucasians *W143X* (*G428→A*, *rs601338* [239, 248]) reaches an allele frequency of almost 0.45 and renders homozygote carriers to "nonsecretors", unable to express ABO-, and several Lewis antigens in bodily secretions (*i.e.* mucosal secretions, saliva, milk) [239, 248, 264]. This variability at the phenotype- and molecular level, appears to be caused by long-term-balancing selection occurring in several human populations [187, 265] and potentially also other vertebrate species [266]. This variation was estimated to be conserved for 2.6-5.3 million years in the hominid lineage, five times longer than comparable neutral loci [248].

Variation at genes involved in glycan synthesis is also described in mice. Wild mice carry a common *cis*-regulatory mutation of the blood group glycosyltransferase *β-1,4-N-Acetylgalactosaminyltransferase 2* (*B4galnt2),* which shows strong signs of long-term-balancing selection [267] including trans-species polymorphism [268]. This glycosyltransferase directs the biosynthesis of the Sd(a) antigen [269] and is conserved across all vertebrates [270]. This *cis*-regulatory mutation confers a tissue specific switch in

B4galnt2 expression from the gut epithelium to the blood vessel endothelium [271]. The consequence of vascular expression is aberrant glycosylation of the blood coagulation factor *von Willebrand factor* (*VWF*), which accelerates *VWF* clearance from the blood stream and prolongs bleeding after injury [272]. This allele, termed "*Modifier of von Willebrand Factor-1*" (*Mvwf1*) [273], was first described in RIIIS/J mice, as this strain serves as model system for the bleeding disorder *von Willebrand disease* [273]. This variation in *B4galnt2* expression has been maintained in the mouse lineage for several million years, despite its potentially detrimental effects on blood homeostasis [268]. Previous studies found an influence of *B4galnt2*-glycans on the intestinal microbial community [274], and point towards the maintenance of both expression types through the action of host-pathogen interactions, either conferred by the presence/absence of the antigen in the intestinal mucosa, or mediated via their influence on the intestinal microbial communities [267, 274].

These patterns imply a variable trade-off scenario, in which different alleles have a fluctuating context- and time dependent benefit compared to the others. However, also frequency-dependent selection, as facilitated by the tracking of the abundant alleles through pathogens can maintain variability of a trait or gene over long timescales. For example secretor and nonsecretors have been found to have differential susceptibility to a wide array of bacterial, viral and fungal pathogens [230, 275-279]. The presence of "dispensable" or variable glycan structures in tissues and mucosal surfaces may thus provide an additional opportunity to evade certain agents by varying this trait in the host [230, 275-279]. As a result, carbohydrate blood-group associated genes show strong signals of selection [187, 265, 267, 280], and long-term conservation of alleles [186, 248], surpassed in the case of *FUT2* only by other prominent alloimmune loci like the MHC/HLA [265, 280].

Glycans and their role in host-microbe interaction

The intestinal epithelium is the largest environmental interface of the human body and is covered with a dense mucus layer, the site of interaction with the gastrointestinal microbial community. There are two different types of highly glycosylated molecules called mucins, gel-forming mucins (MUC2, MUC5AC, MUC5B, MUC6) which form extremely large polymers and transmembrane mucins (MUC1, MUC3, MUC4, MUC12, MUC13, MUC16, MUC17), which cover the surface of enterocytes and other epithelial and form a part of the "glycocalyx". The expression of mucins and their glycan repertoire [235, 236] is highly heterogeneous, which also influences the macroscopic structure of the mucosal lining (thickness, density [281]). Along with these characteristics come changes in the permeability and thus integration into the somatic immune system and metabolism, but also conserved

gradients of glycan composition over the intestinal tract [235, 236]. The mucus of the small intestine is characterized by its thick aqueous composition, its high permeability for bacteria and nutrients, the lack of sulphated Lewis antigens, and high abundance of fucosylated residues [235]. The mucosa of the distal gastrointestinal tract, on the other hand, shows a clear structural difference with a lower impermeable viscous mucus layer [282], and an aqueous rapidly dissolving mucus generated by host- and bacterial endopeptidases [281, 283]. These differences in mucosa structure serve the higher demand of nutrient resorption in the small intestine with a permeable mucosa, and the high bacterial burden within the large intestine, with a rigid bilayer to avoid invasion of the adjacent dense microbial community [281, 282, 284]. Thus, the interplay of attaching commensals, mutualists, or pathogens is altered by the mucus composition, which is regulated on a multigene level, including genes that determine the expression of blood-group-related antigens, which can also be manipulated by the community itself. The microbial community also changes along the gastrointestinal tract, which might be promoted through these described physico-chemical and nutrient gradients [236] and is reflected in the gene content of the respective microbial community [285]. It has been shown that luminal and mucosal communities are fairly different and that communities change drastically along the gastrointestinal tract [138].

The human genome possesses only a small number of carbohydrate active enzymes, which restricts the capabilities of complete glycan breakdown to starch, lactose and sucrose. Bacteria, on the other hand, can encode many of those multigene enzyme machineries (*e.g. starch-utilization-system*/*Sus* in *B. thetaiotamicron*), which enables them to digest and ferment even the highly intertwined and complex glycans of the plant cell walls (cellulose, hemicellulose, pectin) their hosts take up [88]. However, host glycans are an important staple of the bacterial community due to their relatively low complexity and stable supply, while complex dietary glycans are provided in pulses to the community. This may represent a reasonable selective pressure for bacteria associated with the mucosal surface to acquire functions to degrade the ubiquitously abundant host-derived mucins and glycans [101].

Microbes dependent in many different ways on host glycans, not only in a time of low nutrient supply through the host's diet. They are a nutrient source, but also attachment sites and a signaling medium. Many bacteria therefore show genomic adaptations to the specific hosts and their respective tissue and glycan characteristics [101, 286]. Specifically, the proximity of mucus-dwelling bacteria to the underlying tissue positions them to be of great importance and influence for homeostasis, but also makes those communities more dependent on the underlying mucus properties [284, 287].

Many prokaryotes [288, 289], protozoans [290] and even macroparasites [291] digest the mucosal layer by enzymes secreted into their environment to degrade and uncover more

sufficient attachment molecules on the cells from which other community members may also benefit or increase the likelihood of tissue invasion [292-296]. Several pathogens have the ability to modify the glycan profile to match their own binding capabilities (adhesins/lectins, LPS) via glycosidases (*e.g. Streptococcus pneumonia*-sialidases/neuramidases [297]), *Pseudomonas aeruginosa*-neuramidases [298])). The spectrum of carbohydrate active genes (CAZymes) of the residing microbial community also changes over the intestinal tract, as does the glycan composition and physical characteristics of the excreted mucus membrane [235, 282, 285]. Glycan liberation by bacteria is however not only a factor driving virulence, it also acts as a community service at the base of many cross feeding relationships among intestinal bacteria [299, 300]. Highly glycosidic bacteria with a wide repertoire of glycosidases, like *Bacteroides thetaiotamicron* and other highly specialized glycan digesters [301], are keystones in those relationships that establish niches for other resident microbes and can influence host metabolism, infection, and development, especially in the context of degradation of fucosylated glycans [300, 302-304]. Thus, it is no surprise that a highly successful early bacterial colonizer with a large repertoire of carbohydrate active enzymes was observed in a competitive colonization experiment, considering the interdependence of bacteria and host glycans [69]. The dependence of bacteria on host glycans is further exemplified by the induction of several glycosyltransferase genes, specifically fucosyltransferases, after cessation of fucosylated milk oligosaccharides during weaning [305-309].

The digestion and fermentation of glycans and mucin proteins by bacteria leads to the production of short-chain fatty acids (acetate, butyrate, propionate; SCFAs) through putrification and fermentation [310], which are integral for intestinal homeostasis. Butyrate is the main nutrient source of enterocytes/colonocytes and directly influences the somatic immune system via activation of regulatory cell populations [311-313] and appears to repress the development of colonic tumors [314]. Propionate and acetate are also absorbed by the host and incorporated into lipids [315]. The generation of these SCFAs has not only influence on the host itself, but can further influence the community structure as it represents a new pool of organic electron acceptors for butyrate fermentation [316, 317] or repression of pathogen colonization via acetate production [318]. Through the action of desulfating bacteria on sulfated glycans and the fermentation of the mucin proteins, new niches can open up for sulfate reducing bacteria, or for glycan digesting bacteria by unmasking the glycan core structures [294]. Desulfating bacteria and glycan fermenting bacteria further produce high amounts of hydrogen and carbon dioxide, which are substrates for keystone members like methanogens and actetogens [102, 319]. These well-connected food chains and substrate pools make these community processes highly efficient [102, 319] and strengthen the association among community members and their hosts [286].

Supplementation of animals with fucosyllactose (also present in milk) results in proliferation of *Barnesiella* and provides higher resistance against chemically induced colitis [125, 320]. Thus, breast milk, with its high fucose glycan content, might directly interact with the infant microbial community and nurture specific community members which increase immune tolerance. Between secretor- and nonsecretor mothers, these interactions may however differ due to the different glycan content of their breast milk [264], which change successional order and long-term development of the community [126, 230, 238, 321], but also its resistance against pathogen invasion [126, 230, 238, 321]. Fucose excretion by the host itself has also been described in response to inflammation, and might serve as a stabilization mechanism for the resident microbiota [322].

Molecular mimicry, the decoration of cells by host derived or related surface markers is widely distributed among pathogens (*e.g. Plasmodium sp.*, Influenza virus, *Neisseria gonorrhoeae*) and mutualists alike (*e.g. Bacteroides thetaiotamicron* [323], *Campylobacter jejuni* [324]), but also act as a solidifying agent for biofilms (*e.g.* levans, dextrans [325]). These traits increase colonization success (colony formation, predator/immune protection) and help to avert or even directly manipulate and exploit host immune related pathways by this molecular masquerade [167, 223, 326]. Mimicry can even result in cross kingdom interaction through the exploitation of glycans by bacteria and viruses alike, which can be mutualistic [304, 327, 328] or antagonistic [329]. Also *Vibrio fisheri* and *Salmonella Typhimurium* are dependent on certain glycan structures to invade and attach to the host epithelial cells, while *Salmonella* transforms the LPS layer after invasion to avoid host recognition and binds fucosylated glycans to contact the mucosal lining of its respective host [330, 331]. Commensals and pathogens can further establish their own niche, direct host manipulation through initiation of specific glycosyltransferases [305-308] and induction of mucin expression [309], or by excretion of their own “glycan environment” to establish a biofilm [325] or by altering the chemical and physical properties of the surrounding mucus [332]. Bacterial glycan products like polysaccharide A from *Bacteroides fragilis* may also have direct benefits for the host, as it interferes with the immune system by lowering the T-cell response and increasing immunologic tolerance, and interference with other gut pathogens like *Helicobacter hepaticus* [136, 333]. This and other glycans produced by this bacterium are of great importance for successful colonization within the competitive gut microbiome, also by immune evasion and tolerigenic immune modulation [136, 333-335].

Bacterial glycan liberation might also be an additional trigger to stimulate tissue- and immune system development. Especially fucose may play a central role as a signaling molecule that stimulates a baseline immune response [336]. This feature might give additional support to the role of probiotic and highly glycosidic bacteria in immune

homeostasis as described for *Bacteroides thetaiotamicron* [323, 337, 338] or *Lactobacilli* [339]. Bacterial exploitation of the intestinal environment, together with adaptation to the resources and the immune system may have led to a reduced virulence and increased tolerance of the host [340, 341], leading to dependence of the immune system on the bacterial interaction [91, 127-136]. Bacterial products, like short chain fatty acids or glycans, have a tolerigenic and immune modulatory effect on the immune system [136, 311, 313, 333-335].

Factors altering the interplay of microbial communities at the mucosal barrier in a beneficial way, being it bacterial recruitment, nutrient provision, or colonization resistance, may thus be under constant selection. Pathogens might track the altered environmental factors (*i.e.* glycan profile) and community profiles, which in turn changes the adaptive value of those genes together with its aided microbial community and glycan profile. Thus, a variable environment might enforce variation in such a trait, depending on the context and the potential trade-offs, *i.e.* differential pathogen susceptibility. Accordingly, the development of the mucosa and its variable glycan spectrum can be seen primarily as a defense strategy [284, 287]. The mucosa thus acts for one as a medium for antimicrobial peptides (*e.g.* REGIIIγ), secretory IgA, physical barrier and a residence for bacteria, but also as the medium for host-bacteria interaction and communication [342]. This development goes so far as the induction of several genes [305-309] and a normal host immune response depend on the priming by bacterial communities [91, 129-136, 340]. Further, basic interaction of eukaryotes and bacteria, via glycan binding and recognition (C-type lectins), appears to be the basis for the self-nonself recognition and basic multicellularity in the early metazoans. These mechanisms are still important for defense against bacteria (bactericidal lectins, *e.g.* REGIIIγ), but also serve in immune regulation and inflammation in vertebrates (selectin mediated leukocyte recruitment) [112, 113], and may still play a role for the establishment and maintenance of symbiosis [94, 343].

Inflammatory bowel disease and the microbiome

As already discussed, microbial communities influence the host on numerous levels and pathways. This brought the “neglected organ“ into the focus of medical life sciences. Especially in the context of complex diseases like inflammatory diseases, microbial dysbiosis in interaction with genetic susceptibilities have become a main pillar for the explanation of disease development [344, 345].

The rise of autoimmune diseases has been linked to the decrease of infectious diseases and a modern highly sanitary life style (decreased antigen exposure) [346-348].

Also the current food regimen has influenced microbial communities and their host in a detrimental way [349]. For example, processed-, and high caloric foods showed in recent studies their high inflammatory and glycemic potential through the influence of ubiquitously used artificial sweeteners and emulsifiers by influencing the gut mucosa and microbial communities alike [350, 351]. This may further explain the unprecedented increase in IBD cases in countries which recently adopted a Western, industrialized lifestyle [352, 353] and the rise in mortality (>30.000 deaths/year in 2010) and 2.9 million disability adjusted life years caused by this disease complex [354, 355]. However, genetic predisposition cannot be neglected, as strong genetic determinants exist which play essential roles in bacterial recognition, immune homeostasis and host-community interaction (*i.e. Nod2*/*Card15*) [135, 356, 357]. However, the "purely" genetic effect for IBD (as estimated by genome-wide association studies) explains only a relatively small variation of its incidence in human populations (Crohn Disease- 13.6% (30 loci); Ulcerative Colitis- 7.5% (23 loci); IBD- 110 loci) [344]. This is another indicator for the prevailing theory of genetic- and environmental interactions driving the pathology of chronic inflammatory diseases and speaking in favor of a "multiple-hits" scenario in which genetically susceptible hosts face one or several environmental and immunological imbalances, which can tip the balance towards exacerbated immune responses [358].

Many other genes with lower penetrance have been associated to inflammatory diseases, which play important roles in immune function and regulation against bacterial infections (innate and adaptive), autophagy and Endoplasmic Reticulum stress, goblet cell function, healing, generation of reactive oxygen species, but also for barrier integrity and glycosylation [344, 359]. Consistent associations with IBD have been identified in the glycosylation of antibodies and other immune related factors (*e.g. BACH2*, *IL6ST*), which are of great importance for antibody-, and general immune function [360, 361].

Of special interest in this thesis is the main nonsecretor mutation in Caucasians, which was recently established as a prominent risk factor for Crohn Disease [209, 362, 363]. Older studies already identified associations to other inflammatory diseases like necrotizing colitis [208] and rheumatic fever [364]. Studies showed the abundance of fucosylated glycans, specifically ABO related antigens, make up a major component of the mucosal lining [234-236] together with one of its main glycosylation targets MUC2, which is an essential protein component of the mucosal barrier [282, 365]. Thus, changes in the glycan profile and viscosity through the absence of essential glycan components (ABO antigens) increase the permeability of the mucosal lining. This could lead to a higher concentration of bacteria and their effector molecules close to the tissue, enabling tissue contact and invasion, and finally inflammation. Normally, glycosylated mucins and antibacterial proteins are continuously

produced and secreted, triggered by a broad array of pro-inflammatory cytokines and counteract invading bacteria. However, pathogenic bacteria and other environmental factors (*i.e.* emulsifiers) can interfere with the mucosal surfaces and may tip the balance towards acute and chronic inflammation [350]. A thin mucus layer, decreased numbers of goblet cells, and a reduced mucin secretion in the mucosal lining are a hallmark of human IBD and a compromised mucosal barrier. In mice, a reduction of goblet cells, a lack of *Muc2* expression, or aberrant core 1 O-glycosylation leads to spontaneous inflammation due to a less stable mucosal barrier [282, 365, 366].

Changes in glycosylation have also been observed in serum during the course of IBD [367] as a result of acute-phase protein glycosylation [368, 369], and changes in MUC2 glycosylation in Ulcerative Colitis have been observed (absence of sulphated/sialylated O-glycan structures) [370]. Whether those responses are a consequence or a prerequisite for intestinal inflammation is however not yet clear. Glycosylation dramatically changes in response to infection and inflammation and might increase immune cell recruitment at the location of inflammation, but may also repress further pathogen binding and propagation within the host body or its gastrointestinal tract. Specifically in the case of *Fut2*, targeted excretion of fucosylated glycans may serve as a counter measure to buffer pathogen or inflammation induced disturbances of the microbiota [322]. Fucose glycans can induce proliferation of a potentially probiotic bacterium which reduces chemical induced colitis [320].

FUT2 nonsecretor mutations have been associated to lower susceptibility to a number of pathogens such as *Helicobacter pylori, Norovirus,* or *Campylobacter jejuni* [230, 275, 276]. Nonsecretor status was on the other hand associated to increased risk for vaginal candidiasis [371], *Streptococcus* and *Neisseria* infections [372, 373], *Haemophilus* infections [374]. Several of these infectious agents still account for many childhood deaths (*i.e. Campylobacter jejuni*, *Neisseria meningitidis*) all over the world and may thus represent a stronger selective force than exerted by associated late-onset diseases. Further, not only does this gene influence the abundance of single bacteria like *Helicobacter pylori* [375], but seems to alter the composition and structure of the bacterial communities as a whole [376-378]. This extends the range of potential trade-off scenarios with bacteria, up to the point of altering the accumulated community services and bacterial interactions in response to the glycan environment, with a considerable impact on host health and fitness [300, 304, 322, 379]. Furthermore, the absence of fucosylated glycans may alter bacterial digestion- and fermentation products and thus the spectrum of short-chain fatty acids [380]. These differences may arise from communities adapted to the respective glycan spectrum present in either secretor or nonsecretor hosts and might carry down to differences in coloncyte homeostasis and potentially disease.

Chapter I: Geographic patterns of the standing and active human gut microbiome in health and inflammatory bowel disease

The intestinal microbial communities are an important biological entity influencing metabolism [158, 381], immune status [382], and are themselves a barrier to fight of pathogens [22]. These communities are also considered to play a fundamental part in the establishment of inflammatory bowel diseases (*i.e.* Cohn's Disease, Ulcerative Colitis) and other autoimmune diseases (*e.g.* Psoriasis). With the increasing frequency of inflammatory diseases in industrialized societies, together with the increase of processed food consumption, antibiotic use, and cleanliness, it is assumed that bacterial communities disturbed by those factors might fall into a dysbiotic state, which on the one hand can make them more susceptible to invasion by opportunistic pathogens, but also renders the community to an agent of local or systemic inflammation in its host. However, little is known if the observed patterns, mostly obtained from single cohort studies, are also transferrable to other unrelated host populations and if one can identify universal patterns of dysbiosis in inflammatory bowel diseases. As microbial communities are highly variable and strongly influenced by food consumption, healthcare and the external environment from which it colonizes the respective hosts, populations of different societal and geographic origin may show strong population specific community compositions and dynamics. These differences could influence the developmental routes and final states of the dysbiosis that might be the result or the trigger of inflammation in their hosts. So far, most studies investigating the microbial composition of biotic- and abiotic environments focused on the abundance bacteria measured by the abundance of the rRNA-gene (DNA) and rarely on the expressed rRNA (RNA). The analysis of the transcribed sequences allows one to infer which bacteria are allochthonous (passive) and autochthonous (active) community members, and might help to differentiate between the stagnant seed bank of spore forming bacteria (*e.g.* diapause), environmentally acquired transient members, and actively transcribing bacteria.

To explore the differences in healthy and diseased microbial communities among different populations, we investigated the mucosal attached microbiome in colonic biopsy samples of healthy individuals and inflammatory bowel disease patients (Cohn's Disease, Ulcerative Colitis) from Europe (Germany, Lithuania) and India. We profiled communities not only based on the genomic rRNA gene, but also on the expressed RNA, enabling us to approximate the activity of community members.

Community analyzes based on DNA and RNA derived community profiles reveal strong population differences among cohorts, especially between European and Indian populations. These can be the result of an interaction of similarity distance decay

relationships between communities, as well as genetic, environmental, and cultural differences correlated to the geographic distances. Furthermore, we observe mainly population-specific disease clusters, and thus different community dynamics and structures as a result of inflammatory bowel diseases in different host populations. However, when the geographic effects are factored out, a weak but universal footprint of disease becomes apparent, which may reflect some commonalities of the community trajectories between health conditions. Disease effects are more strongly pronounced in the active community regarding a decreased community complexity and community differentiation in the diseased cohorts. The active communities in general show a much lower diversity compared to the stagnant DNA derived profiles, which may be explained by the smaller autochthonous actively contributing microbial community (measured via RNA), in contrast to the community profile considering also allochthonous transient members (DNA based).

In summary, only a limited universal microbial association of inflammatory bowel diseases is observable on an international scale. Future studies, however, may interrogate the functionality of the communities, which may show a stronger universal theme of community changes according to disease, as functional redundancy of the microbiome might obscure these effects in the taxonomic approach presented here.

Publication:

Rehman, A., **P. Rausch**, J. Wang, J. Skieceviciene, G. Kiudelis, K. Bhagalia, D. Amarapurkar, L. Kupcinskas, S. Schreiber, P. Rosenstiel, J. F. Baines and S. Ott (2015). "Geographical patterns of the standing and active human gut microbiome in health and IBD." Gut.

Gut Online First, published on January 7, 2015 as 10.1136/gutjnl-2014-308341

ORIGINAL ARTICLE

Geographical patterns of the standing and active human gut microbiome in health and IBD

Ateequr Rehman,[1] Philipp Rausch,[2,3] Jun Wang,[2,3] Jurgita Skieceviciene,[1,4] Gediminas Kiudelis,[5] Ketan Bhagalia,[6] Deepak Amarapurkar,[6] Limas Kupcinskas,[4,5] Stefan Schreiber,[1,7] Philip Rosenstiel,[1] John F Baines,[2,3] Stephan Ott[7]

▸ Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/gutjnl-2014-308341).

For numbered affiliations see end of article.

Correspondence to
Professor Philip Rosenstiel; Schittenhelmstr. 12, Institut für Klinische Molekular Biologie Kiel D-24105, Germany; p.rosenstiel@mucosa.de
Professor John F Baines; Arnold-Heller-Str. 3, Haus 17, Institut für Experimentelle Medizin, Kiel D- 24105, Germany; j.baines@iem.uni-kiel.de
Dr. Stephan Ott; Arnold-Heller-Str. 3, Haus 6, Klinik für Innere Medizin I Kiel D-24105, Germany; s.ott@mucosa.de

AR and PR contributed equally.

Received 25 September 2014
Revised 19 November 2014
Accepted 30 November 2014

ABSTRACT

Objective A global increase of IBD has been reported, especially in countries that previously had low incidence rates. Also, the knowledge of the human gut microbiome is steadily increasing, however, limited information regarding its variation on a global scale is available. In the light of the microbial involvement in IBDs, we aimed to (1) identify shared and distinct IBD-associated mucosal microbiota patterns from different geographical regions including Europe (Germany, Lithuania) and South Asia (India) and (2) determine whether profiling based on 16S rRNA transcripts provides additional resolution, both of which may hold important clinical relevance.

Design In this study, we analyse a set of 89 mucosal biopsies sampled from individuals of German, Lithuanian and Indian origins, using bacterial community profiling of a roughly equal number of healthy controls, patients with Crohn's disease and UC from each location, and analyse 16S rDNA and rRNA as proxies for standing and active microbial community structure, respectively.

Results We find pronounced population-specific as well as general disease patterns in the major phyla and patterns of diversity, which differ between the standing and active communities. The geographical origin of samples dominates the patterns of β diversity with locally restricted disease clusters and more pronounced effects in the active microbial communities. However, two genera belonging to the *Clostridium leptum* subgroup, *Faecalibacteria* and *Papillibacter*, display consistent patterns with respect to disease status and may thus serve as reliable 'microbiomarkers'.

Conclusions These analyses reveal important interactions of patients' geographical origin and disease in the interpretation of disease-associated changes in microbial communities and highlight the added value of analysing communities on both the 16S rRNA gene (DNA) and transcript (RNA) level.

Significance of this study

What is already known on this subject?
- IBD impacts microbial community structure in the gut.
- Microbial communities differ between human populations, potentially driven by variation in genetic polymorphism, lifestyle and environmental conditions.
- Diversity within and between bacterial communities change over a lifetime.

What are the new findings?
- Pathological community patterns observed in IBD are influenced by local, population-specific factors, but also show shared elements between the different cohorts.
- The active bacterial component (rRNA) shows lower diversity and intercohort variation than the standing diversity (rDNA).
- The bacterial communities investigated on the RNA level reveal stronger disease-associated patterns.

How might it impact on clinical practice in the foreseeable future?
- Variation of mucosal microbial communities among human populations constitutes an important factor when considering the microbiome as a target for treatment of IBD.
- The identification of disease-associated microbial patterns shared between distinct geographical regions might be specifically useful to tailor a set of 'microbiomarkers' for a molecular assessment of IBD.

INTRODUCTION

While a steady increase of IBD has been observed over the past decades in North America and Europe, a dramatic rise in incidence rates is observed in countries that have recently adopted a Western industrialised lifestyle, for example, East and South Asia, or states of the former Soviet Union.[1 2] Suspected environmental factors include increased levels of hygiene (decrease in antigen contacts), changes in nutritional habits and the industrialisation of food production and preservation.[3–6] It can be assumed that all these environmental cues specifically act on the acquisition, composition and stability of the intestinal microbiome. Only recently have international studies begun to explore microbial communities in human populations at different body sites on broader geographical scales.[7 8] A common concept in microbial biogeography proposes a world-wide, passive dispersal of bacteria, followed by environmental filtering of bacterial assemblies (ie, 'everything is everywhere, but the environment selects').[9] Many

To cite: Rehman A, Rausch P, Wang J, *et al*. *Gut* Published Online First: [*please include* Day Month Year] doi:10.1136/gutjnl-2014-308341

studies addressed this hypothesis in an environmental context,[10–12] but how those globally distinct environmental microbial assemblages are translated into the stable adult intestinal microbiome is still poorly understood. However, it seems obvious that physical distance and variables that correlate with this (eg, temperature), the genetic makeup of the host, diet and other sociocultural habits will play decisive roles. Whatever the exact reasons for geographical differences in human-associated microbial communities, they are likely important factors to be considered for understanding disease aetiologies among human populations. Qin *et al*[13] were among the first to identify a common faecal microbial gene catalogue and to discern communities according to an underlying condition of IBD not restricted to a single human population. More recent in-depth analyses of faecal communities included divergent populations,[8] but only recently focused on disease states.[14]

It can be hypothesised that the microbial communities tightly associated with the intestinal mucosa might be under stronger host control[15] and less subject to transient perturbations compared with the luminal microbiota. Although they might have a greater impact on homeostasis, mucosal communities are comparatively understudied. In this study, we investigate the impact of the two major forms of IBD, Crohn's disease (CD) and UC, on the bacterial communities associated with the colonic mucosa in a geographical context. Colonic biopsies were obtained from patients with IBD and controls originating from Germany, Lithuania and India. While most previous studies focused on the 16S rDNA level, we employ bacterial community profiling on the levels of both 16S rDNA and rRNA. This comprehensive approach distinguishes between standing and active microbial communities, which together enables us to explore variation among geographically distinct microbiomes and also to relate these differences to the patterns observed in IBD.

MATERIAL AND METHODS

Human samples

Colonic biopsies were taken from the sigmoid region of healthy subjects and patients in clinical remission. The diagnoses of UC and CD were based on standard clinical, endoscopical, radiological and histological criteria. All samples and phenotype information were pseudonymised before the procedure. All individuals agreed to participate by giving informed consent at least 24 h before sampling. Details on age, sex, disease status and medication are provided in table 1. Due to mean differences in age between population cohorts, a normalisation within populations was performed to account for this in interpopulation comparisons by subtracting the minimum age within each population.

Nucleic acids extraction and 16S rRNA pyrosequencing

DNA and RNA were extracted using the Qiagen Allprep DNA/RNA as previously described (see online supplementary material[16]). RNA was reverse transcribed to cDNA using random hexamers (Qiagen, Hilden, Germany). Nucleic acid extraction and reverse transcription of Indian samples were performed on-site in India. Reverse transcribed cDNA and genomic DNA were freeze dried and transported on dry ice to Germany for further processing. Frozen biopsies sampled in Lithuania were transported on dry ice to be processed in Kiel, Germany. The 16S rRNA gene (RNA and DNA) was amplified with the 27F-338R primer pair and sequenced as described before.[17] Sequences were processed using Mothur V.1.15.0,[18] and filtered using stringent quality criteria (see online supplementary methods).

Statistical analysis

α Diversity and β diversity indices (Jaccard and Bray–Curtis (square root transformed)) were calculated in R.[19–21] FASTUniFrac was used to calculate the unweighted and normalised weighted UniFrac metrics.[22] Statistical analysis of community distances was performed with non-parametric distance-based analysis of variance (ANOVA) using '*adonis*', Mantel correlation, Procrustes analysis and fitting of centroids were implemented in the 'vegan' package for R and tested with 10^5 permutations to assess significance.[23 24] Redundancy Analysis (RDA) was carried out on Hellinger-transformed Operational Taxonomic Unit (OTU) tables and tested using a permutative ANOVA approach.[25] Comparisons of means (ie, phyla abundances, α diversity) followed a linear model framework using standard model selection procedures (minimising AIC values without a significant loss of fit) requiring normally distributed residuals. Indicator species analysis was implemented via the R package 'indicspecies' with 10^5 permutations.[26] The activity of genera and species was estimated through rRNA/rDNA ratio, while divisions of and by zero were set to zero. Differentially active bacteria were detected by Kruskal–Wallis tests. p Values of the genera/OTU associations (rDNA, rRNA, activity) were adjusted using the Benjamini and Hochberg procedure.

RESULTS

Phylum abundances are influenced by disease status and sampling population

To investigate the influence of IBD on the mucosa-associated microbiota in a broader geographical context, sigmoidal biopsies were obtained from ~10 each of healthy controls, patients with CD and UC, residing in Germany, Lithuania and India, totalling 89 samples (cohort details in table 1). Pyrosequencing of the V1–V2 region of the 16S rRNA gene was performed on the level of both 16S rDNA and rRNA (reverse transcribed to cDNA, see the Materials and methods section). Normalisation (~1000 sequences per individual) yielded 88000 rDNA and 86974 rRNA sequences. A single control rDNA sample from Germany and single control and CD rRNA samples from Lithuania (ie, in total three samples) were not included in further analysis due to low sequencing coverage. Species-level OTUs (97% identity OTUs) were clustered using the combined rDNA and rRNA-level datasets, and split accordingly. This resulted in a community coverage of 83.45±5.08% and 89.22±5.54% of species for rDNA and rRNA, respectively (Good's coverage, see online supplementary figure S1).

We first analysed phylum abundances in a global manner (ie, across all three populations), whereby complex differences between standing (rDNA) and active (rRNA) communities were observed for most phyla (see online supplementary figure S2 and S3). Overall, Bacteroidetes and Proteobacteria show inverse effects among the active and standing microbial communities and are negatively associated with each other (rDNA: $r=-0.527$, $p=1.33\times10^{-7}$, rRNA: $r=-0.254$, $p=0.0176$). Bacteroidetes show a significant increase with age in the rDNA samples (figure 1A), whereas the rRNA samples further reveal influences of disease status on the abundance of active Bacteroidetes, mainly by a higher abundance in UC samples across populations (figure 1B, see online supplementary table S1). Proteobacteria abundance decreases with age in the rDNA-based samples (figure 1E), inversely with Bacteroidetes. The abundance of active Proteobacteria in contrast does not decrease with age, but displays a decrease in patients with UC compared with patients with CD and healthy samples, which is

Table 1 Patient information for each population

Metadata	Germany	Lithuania	India
DNA			
Control	9	10	11
Crohn's Disease	10	9	9
UC	10	10	10
Male/female	14/15	10/18	21/9
# NA gender	0	1	0
Mean age (original)	35.966±12.974 SD	45.893±18.089 SD	37.172±13.782 SD
Mean age (normalised)	19.966±12.974 SD	26.893±18.089 SD	20.172±13.782 SD
Age range (original)	16–63	19–81	17–67
Age range (normalised)	0–47	0–62	0–50
# NA age	0	1	1
Medication (yes/no)	10/19	13/16	25/5
Antibiotics	2	2	16
Probiotics	4	0	7
5-ASA	2	11	17
TNF-block	3	0	1
Azathioprin	4	0	4
MTX	1	0	0
Corticoids	5	1	9
RNA			
Control	10	9	11
Crohn's Disease	10	8	9
UC	10	10	10
Male/female	14/16	10/17	21/9
# NA gender	0	0	0
Mean age (original)	35.667±12.853 SD	46.370±18.253 SD	37.172±13.782 SD
Mean age (normalised)	19.667±12.853 SD	27.370±18.253 SD	20.172±13.782 SD
Age range (original)	16–63	19–81	17–67
Age range (normalised)	0–47	0–62	0–50
# NA age	0	0	1
Medication (yes/no)	10/20	13/14	25/5
Antibiotics	2	2	16
Probiotics	5	0	7
5-ASA	2	11	17
TNF-block	3	0	1
Azathioprin	4	0	4
MTX	1	0	0
Corticoids	5	1	9

ASA, 5-aminosalicylic acid; MTX, methotrexate; NA, not available; TNF, tumour necrosis factor.

also influenced by the subject's gender (figure 1F, see online supplementary table S1). The Firmicutes abundances based on rDNA mainly display differences between healthy controls and patients with UC across populations, especially among German and Lithuanian samples (figure 1C), which is confirmed in separate analyses for each population (see below, see online supplementary table S2). The Firmicutes abundances based on rRNA show significant differences between European (Germany, Lithuania) and Indian samples, as well as between pathologies within and among the sampling cohorts (figure 1D, see online supplementary table S1).

Second, we analysed each single population separately. This reveals pronounced differences especially in Firmicutes between disease groups within each population, based on both rRNA and rDNA, although the relative phylum-level patterns between investigated groups are not consistent among populations (see online supplementary table S2). In particular, Firmicutes abundance is the lowest in healthy German samples compared with diseased individuals, while Lithuanian and Indian patients with CD show the lowest Firmicutes abundances. Bacteroidetes, on the other hand, show common patterns of age and pathology in Lithuanian and Indian patients but not in Germans. Bacteroidetes show also a population-independent increase in abundance in the standing and active bacteria among healthy and UC subjects. Proteobacteria also display an increased abundance in CD among Lithuanians and Indians, while no apparent effects were present in German samples (see online supplementary table S2).

In summary, we revealed interesting age-related patterns for both Bacteroidetes and Proteobacteria, while population-specific disease-related patterns are present among the Firmicutes. Furthermore, basing analyses on 16S rRNA in general provided greater resolution in detecting disease and population-specific effects.

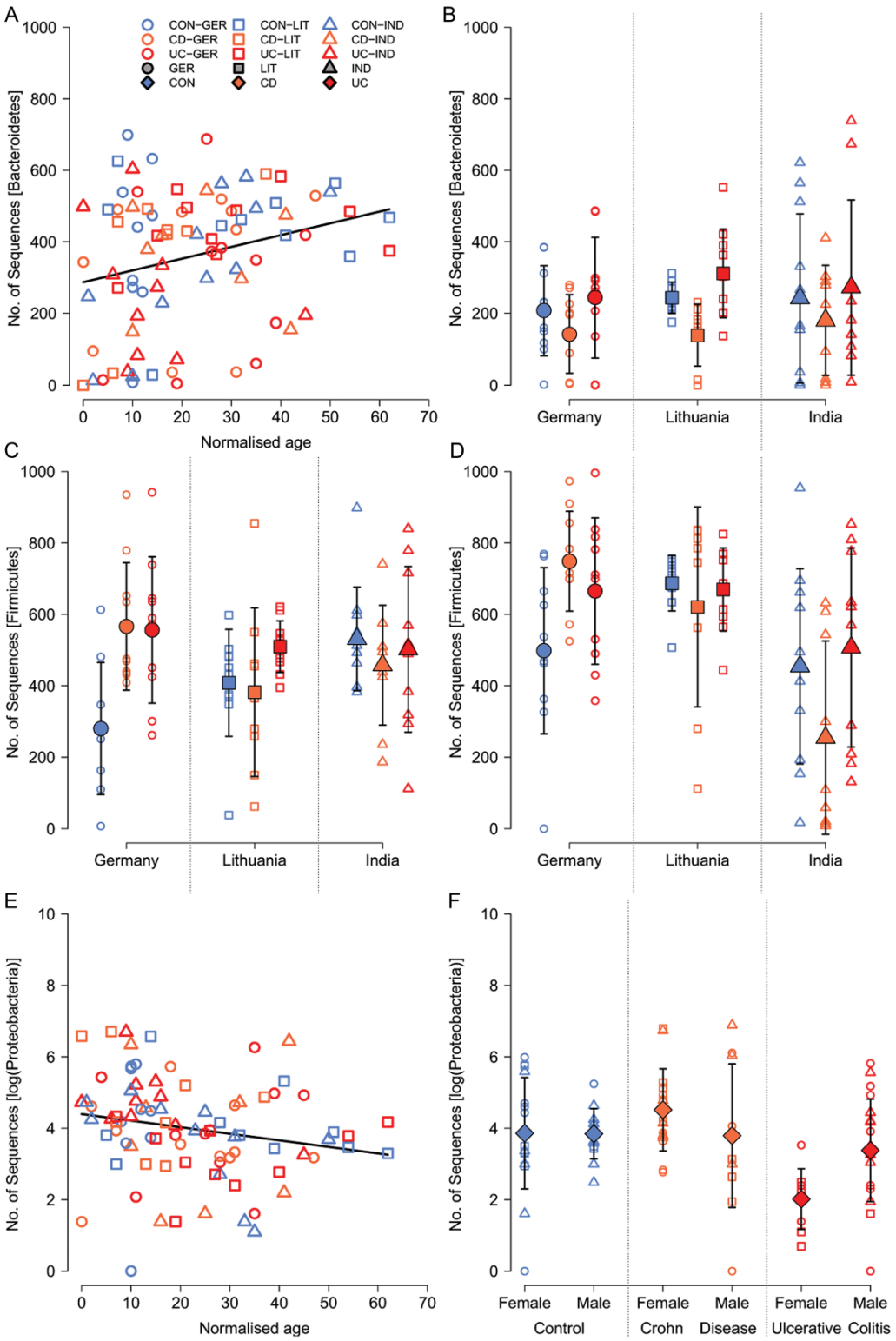


Figure 1 Comparative analysis of mucosa-attached bacterial communities at the phylum level. Plots of phyla abundances based on 16S rDNA (A, C, E) and rRNA (B, D, F) visualise the effects of the best statistical model (Firmicutes, Bacteroidetes, Proteobacteria, error bars represent SD). CD, Crohn's Disease; CON, control; GER, Germany; IND, India; LIT, Lithuania.

Patterns of bacterial diversity within and between individuals is influenced by age, population-specific effects and disease

α Diversity

We focused our analysis on a panel of diversity measures which provide information about the approximate species number,[27] entropy and evenness of the community,[19] as well as its phylogenetic diversity.[20] Interestingly, we find significantly higher diversity in rDNA-based samples and a moderate correlation between the species diversities of the standing and active communities (figure 2A–C).

First, we analysed the panel of α diversity indices globally among all samples. Investigating species richness (using Chao1 index), we observe increases of species number with age in the standing microbial community (rDNA), while species richness in the active communities (rRNA) increases with age and shows significantly lower diversity among patients with CD (figure 2D, E, table 2). By applying Shannon entropy,[19] which represents the distribution of species in a sample, we also find an increase in diversity with age in the rDNA-based and rRNA-based communities (see online supplementary figure S4A, table 2). Phylogenetic diversity of the standing community is also correlated with a subject's age, but increases only in patients with CD (see online supplementary figure S4C, table 2). The rRNA-based samples display differences between CD and healthy controls, between CD and UC, as well as between European and Indian samples (see online supplementary figures S4B,D, S5 and table S3; also see table 2), with the highest level of species and phylogenetic diversity among healthy individuals.

The increase of community diversity with age can be a sign of community succession, that is, a change in community structure over time, or a lack of colonisation resistance. To further investigate potential confounding effects of disease on those succession patterns, we analysed each disease state and population cohort separately. Interestingly, the strongest signal of succession is present in the active and standing bacterial communities of patients with CD, while in healthy individuals and patients with UC, diversity does not consistently increase with age (see online supplementary figures S5, S6 and supplementary table S3). Thus, in summary, the rRNA-based samples display reduced diversity compared with rDNA-based samples, and at the same time provide more resolution to detect influences of sampling region and disease compared with rDNA. The age-related patterns of increasing species diversity appear to be largely limited to CD and may point towards a reduced colonisation resistance of the disturbed microbial communities in IBD. Further,

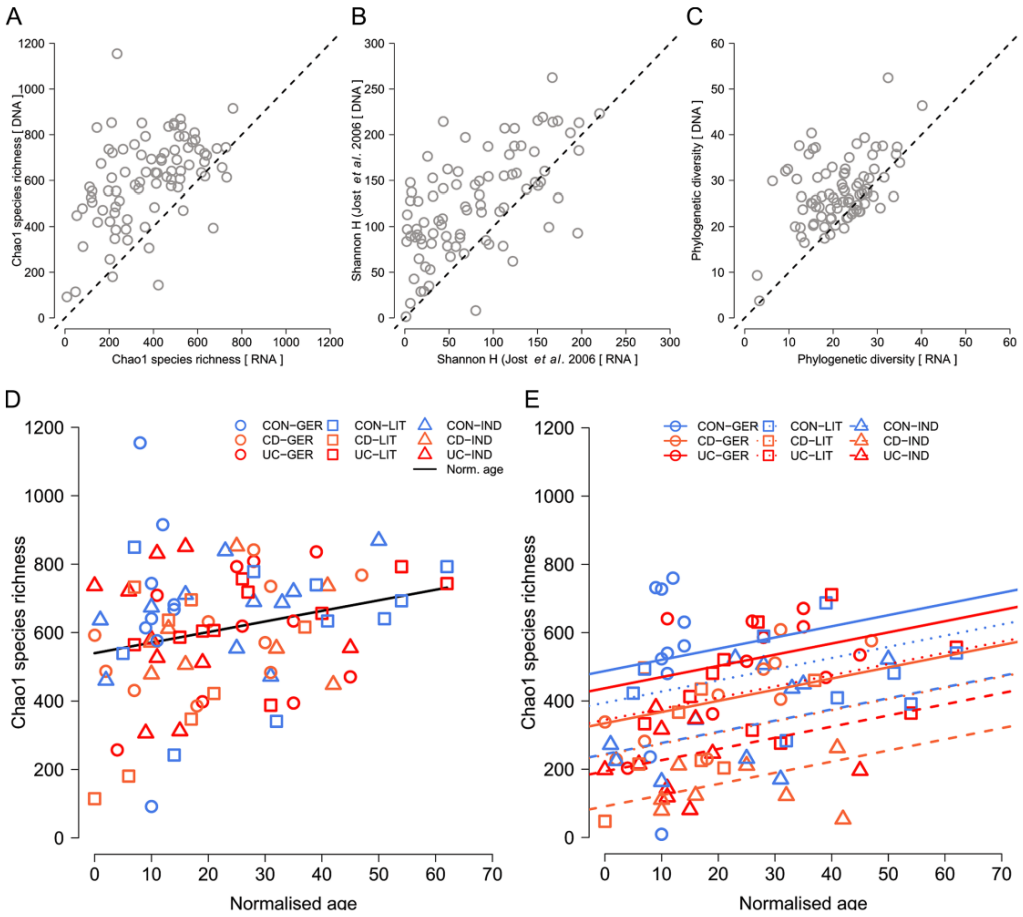


Figure 2 Analysis of mucosa-attached bacterial communities identifies a common increase of bacterial diversity with age regardless of diagnosis and geographical origin. Correlation of α diversity metrics based on rDNA and rRNA (Chao1 species richness: r=0.493, $p=1.97\times10^{-6}$ (A); Shannon H (Jost): r=0.609, $p=9.223\times10^{-11}$ (B); phylogenetic diversity: r=0.355, p=0.001 (C)). Species richness according to the best statistical model in rDNA (D) and rRNA (E) derived communities (table 2; for details on Shannon H and phylogenetic diversity, see figure S4). CD, Crohn's Disease; CON, control; GER, Germany; IND, India; LIT, Lithuania.

Table 2 Statistical analyses of α diversity based on species distribution (Shannon H), richness (Chao1) and phylogenetic diversity in DNA-based and RNA-based samples

α Diversity metric	Model factors	DF	F	P	adj. R^2	Posthoc	p Value*
DNA							
Shannon H	Normalised age	1	10.532	0.002	0.101	–	–
Chao1 species richness	Normalised age	1	5.656	0.020	0.052	–	–
Phylogenetic diversity	Disease	2	2.258	0.111	0.062	CON—CD	0.007
						CON—UC	0.304
						UC—CD	0.245
	Normalised age	1	0.203	0.654		–	–
	Disease : Normalised age	2	2.935	0.059		CON—CD	0.028
						CON—UC	0.495
						UC—CD	0.048
RNA							
Shannon H	Disease	2	5.006	0.009	0.358	CON—CD	0.020
						CON—UC	0.926
						UC—CD	0.130
	Population	2	14.726	<0.001		IND—GER	<0.001
						LIT—GER	0.712
						LIT—IND	0.003
	Normalised age	1	12.895	0.001		–	–
Chao1 species richness	Disease	2	8.356	0.001	0.451	CON—CD	0.001
						CON—UC	0.560
						UC—CD	0.039
	Population	2	23.849	<0.001		IND—GER	<0.001
						LIT—GER	0.079
						LIT—IND	0.001
	Normalised age	1	10.425	0.002		–	–
Phylogenetic Diversity	Disease	2	5.208	0.007	0.198	CON—CD	0.0164
						CON—UC	0.0920
						UC—CD	0.9282
	Population	2	4.739	0.011		IND—GER	0.0177
						LIT—GER	0.5718
						LIT—IND	0.4897
	Normalised age	1	6.078	0.016		–	–

*Single step corrected family wise p value for posthoc comparisons.
CD, Crohn's Disease; CON, control; GER, Germany; IND, India; LIT, Lithuania.

although systematic differences are present between geographical locations, a consistent pattern with respect to disease status nested *within* each location, based on rRNA, is that the diversity decreases from healthy individuals, followed by patients with UC and is the lowest in patients with CD.

β Diversity

To further evaluate the contribution of geographical origin, disease status and their interactions, we performed analyses based on the phylogenetic β diversity measure UniFrac (weighted and unweighted), as well as on metrics considering the shared presence (Jaccard) or abundance (Bray–Curtis) of species level OTUs. First, we applied a model including each factor on all β diversity metrics using non-parametric multivariate ANOVA ('adonis', see the Materials and methods section). Second, we complemented these analyses with individual pairwise comparisons with respect to only population of origin and disease status (table 3). The analyses show that population is the most influential factor, displaying significant differences especially between European and Indian samples, for all four β diversity measures in rDNA-based and rRNA-based samples (table 3, see online supplementary figure S7). The influence of disease status alone is less apparent, with relatively small differences among the rDNA-based samples (table 3, see online supplementary figure S7A, C, E, G). By contrast, rRNA-based samples reveal significant influences of disease for all β diversity measures, and individual pairwise comparisons between health conditions also uncovered differences between the two pathologies (table 3, see online supplementary figure S7B, D, F, H). Interactions between population and disease were more pronounced than disease alone, displaying significant regional disease-associated communities based on presence/absence and abundance of bacteria in the rDNA-based and rRNA-based samples (see online supplementary figure S7A–D). Interestingly, in addition to the greater number of significant influences detected, consistently more variation is explained among the rRNA-based samples (table 3, also see online supplementary figure S7 and table S4), which are correlated with the standing community profiles (rDNA), but differ in the abundance of several genera (see online supplementary figure S8). Changes in community composition with respect to age were also observed, especially in the phylogenetic profile of rDNA-based and rRNA-based samples (table 3, see online supplementary figure S7).

To directly assess the influences of population origin and disease condition, we applied RDA to model these effects on bacterial communities using individual bacterial distributions

Table 3 β diversity analyses via non-parametric distance-based analysis of variance (*adonis*) using population, disease condition, their interaction, age and the pairwise comparisons among countries and disease conditions (bold face highlights significant comparisons)

		DNA			RNA		
Metric	**Model**	**F**	**R^2**	**p Value**	**F**	**R^2**	**p Value**
UniFrac (unweighted)	Population	2.457	0.054	**0.0001**	3.589	0.077	**0.0001**
	Disease	1.135	0.025	0.1164	1.482	0.032	**0.0034**
	Disease: population	1.261	0.055	**0.0061**	1.146	0.049	0.0568
	Normalised age	2.050	0.024	**0.0002**	1.956	0.023	**0.0012**
	GER—LIT	1.532	0.027	**0.0092**	1.883	0.033	**0.0002**
	GER—IND	2.440	0.041	**0.0001**	4.096	0.066	**0.0001**
	LIT—IND	3.279	0.054	**0.0001**	4.504	0.076	**0.0001**
	CD—UC	1.062	0.019	0.2520	1.297	0.023	0.0597
	CON—UC	1.151	0.019	0.1348	1.423	0.024	**0.0189**
	CON—CD	1.051	0.018	0.2855	1.397	0.025	**0.0387**
UniFrac (weighted)	Population	2.881	0.062	**0.0083**	5.206	0.106	**0.0001**
	Disease	0.971	0.021	0.4192	2.344	0.048	**0.0068**
	Disease: population	1.503	0.065	0.0914	1.253	0.051	0.1646
	Normalised age	5.017	0.056	**0.0032**	2.787	0.032	**0.0121**
	GER—LIT	1.434	0.025	0.1813	2.040	0.036	**0.0499**
	GER—IND	2.619	0.044	**0.0285**	5.646	0.089	**0.0001**
	LIT—IND	4.526	0.074	**0.0054**	6.963	0.112	**0.0002**
	CD—UC	0.855	0.015	0.4439	3.532	0.060	**0.0038**
	CON—UC	0.999	0.017	0.3563	1.576	0.026	0.1152
	CON—CD	0.840	0.015	0.4633	1.303	0.023	0.2150
Jaccard	Population	2.161	0.048	**0.0001**	3.552	0.076	**0.0001**
	Disease	1.131	0.025	**0.0336**	1.502	0.032	**0.0001**
	Disease: population	1.128	0.050	**0.0064**	1.283	0.055	**0.0001**
	Normalised age	1.483	0.017	**0.0005**	1.397	0.016	**0.0056**
	GER—LIT	1.651	0.029	**0.0001**	2.530	0.044	**0.0001**
	GER—IND	2.282	0.038	**0.0001**	4.111	0.066	**0.0001**
	LIT—IND	2.483	0.042	**0.0001**	3.689	0.063	**0.0001**
	CD—UC	1.100	0.019	0.1231	1.439	0.025	**0.0073**
	CON—UC	1.125	0.019	0.0694	1.293	0.022	**0.0191**
	CON—CD	1.074	0.019	0.1679	1.422	0.025	**0.0065**
Bray–Curtis	Population	2.895	0.063	**0.0001**	4.783	0.100	**0.0001**
	Disease	1.217	0.027	**0.0388**	1.688	0.035	**0.0003**
	Disease: population	1.179	0.051	**0.0244**	1.264	0.053	**0.0150**
	Normalised age	1.654	0.019	**0.0029**	1.647	0.019	**0.0095**
	GER—LIT	2.009	0.035	**0.0005**	3.090	0.053	**0.0001**
	GER—IND	3.061	0.051	**0.0001**	5.431	0.086	**0.0001**
	LIT—IND	3.469	0.057	**0.0001**	5.338	0.088	**0.0001**
	CD—UC	1.195	0.021	0.1105	1.839	0.032	**0.0017**
	CON—UC	1.172	0.020	0.1260	1.389	0.023	**0.0414**
	CON—CD	1.105	0.019	0.2187	1.345	0.024	0.0729

CD, Crohn's Disease; CON, control; GER, Germany; IND, India; LIT, Lithuania.

among those environmental factors. To test whether local disease patterns are present, we included the interaction between disease and population in the RDA model, which indicates the presence of significant local disease effects in the standing and active microbial communities (figure 3A, C). The variation explained by these models is relatively small (rDNA: adjusted R^2=0.042; rRNA: adjusted R^2=0.102), which stresses the high interpopulation and intrapopulation variability of the microbiome. However, an interesting observation is the dominating influence of population and that common disease effects are observed only in higher, less important dimensions of the ordinations (see online supplementary figures S9 and S10). To explore whether general influences of IBD can be observed, we cancelled out the influence of host population beforehand using partial Redundancy Analysis (pRDA). This revealed significant disease clusters in both datasets (rDNA: $F_{2.83}$=1.183, p=0.040, R^2=0.026; rRNA: $F_{2.82}$=1.574, p=0.001, R^2=0.033). Thus, these results highlight the importance of geographically restricted environmental factors driving microbial community differentiation, leaving a weak but universal disease imprint after correction for sampling population (figure 3B, D). Further analyses on the single populations and disease subsets support the existence of regional disease microbiomes, as similar pathologies differ in their microbial communities between sampling regions, which appear stronger in the active communities (see online supplementary table S4). As medication, in particular antibiotic use, can influence microbial communities and confound these analyses, we further investigated these variables, but

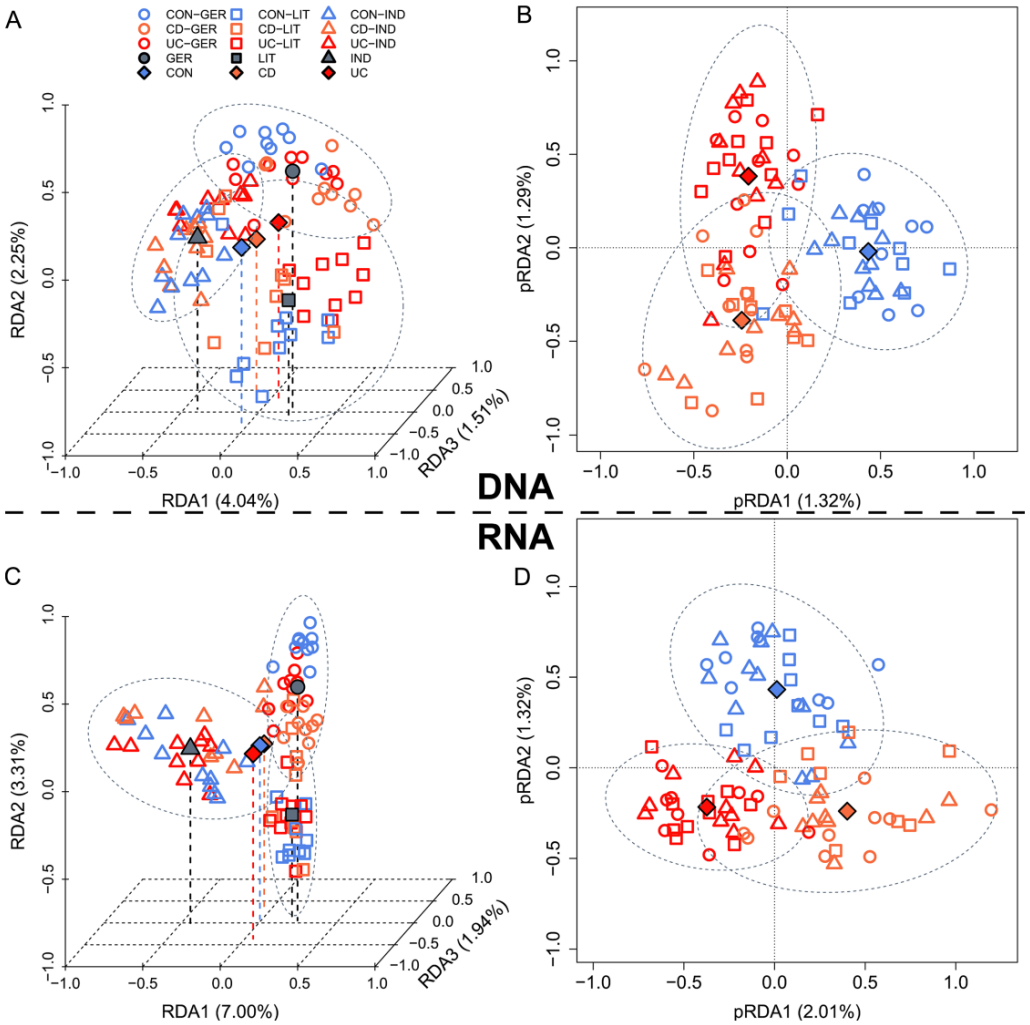


Figure 3 Identification of population of origin and disease effects on microbial community structures: influence of active (16S rRNA) and standing (16S rDNA) bacteria. Redundancy analysis (RDA) (A and C) and partial RDA (B and D) of DNA-based and RNA-based datasets. An RDA of DNA-based microbial communities reveals strong influence of sampling population and local disease regimes (population-$F_{2.79}$=2.650, R^2=0.058, p=0.001; disease-$F_{2.79}$=1.194, R^2=0.026, p=0.027; disease by population-$F_{4.79}$=1.163, R^2=0.051, p=0.018) and RDA on the active community with higher explanatory power and an increased influence of disease condition (C: population-$F_{2.78}$=4.747, R^2=0.099, p=0.001; disease-$F_{2.78}$=1.587, R^2=0.033, p=0.001; disease by population-$F_{4.78}$=1.269, R^2=0.053, p=0.005). The axes shown in the pRDA are the main axes of variation (rDNA: pRDA 1-$F_{1.83}$=1.198, p=0.073, pRDA 2-$F_{1.83}$=1.167, p=0.083; rRNA: pRDA 1-$F_{1.82}$=1.896, p=0.001; pRDA 2-$F_{1.82}$=1.251, p=0.059). For additional significant dimensions of A and C, see online supplementary figures S9 and 10. CD, Crohn's Disease; CON, control; GER, Germany; IND, India; LIT, Lithuania.

identified only minor effects within and among the different populations (see online supplementary analyses).

Indicator bacteria analysis shows strong disease by population associations

To identify bacteria that are more frequently present and abundant with respect to population, disease or their interaction, we used indicator species analysis.[26] This analysis was performed at the level of consensus genera (classification-based) and species-level OTUs for both the rDNA-based and rRNA-based samples, in addition to a proxy for bacterial activity (rRNA/rDNA). The indicator genera based on rDNA (N=28, see online supplementary table S5) are associated mostly with the sample population and, accordingly, those associated with disease are mostly predictive for pathologies within a certain population (see online supplementary figure S11 and table S5). The same patterns are also present in the active communities (n=24; see online supplementary figure S12 and table S6), with differentiation of some highly abundant genera between Europe and India (ie, *Bacteroides*, *Prevotella*, respectively). Active *Bacteroides* also appear to be significantly more abundant in the European control and UC subjects. As with rDNA, most genera display associations with disease within single populations, while *Papillibacter* shows preferential occurrence in healthy controls across all populations, representing a potentially active universal indicator of health status (see online supplementary table S6). An analysis of species-level OTUs shows qualitatively similar results (rDNA: n=24, see online supplementary figure S13; rRNA: N=122, online supplementary figures S14 and S15), but offers a more detailed view on certain low abundant

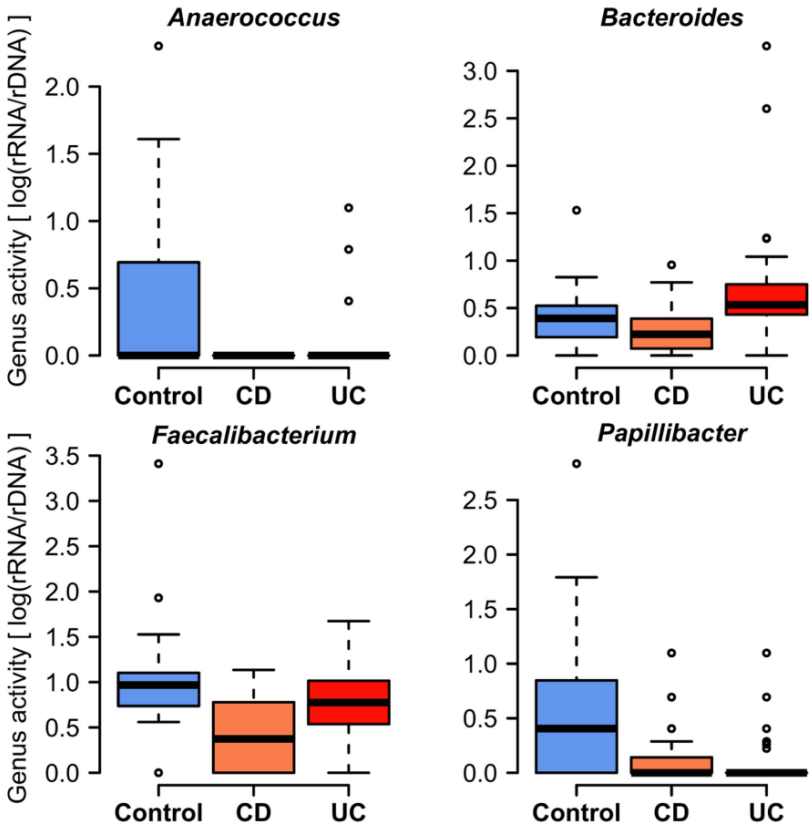


Figure 4 Analysis of bacterial genera that are differentially active between diseases. The plots show the mean ratio of rRNA/rDNA as a proxy of metabolic activity of *Anaerococcus* (p=0.014; rRNA/rDNA=0.900±0.348 SEM), *Papillibacter* (p=0.014; rRNA/rDNA=1.513±0.579 SEM), *Bacteroides* (p=0.014; rRNA/rDNA=2.033±0.893 SEM) and *Faecalibacterium* (p=0.014; rRNA/rDNA=0.606±0.125 SEM; p Values adjusted by false discovery rate (FDR); see also online supplementary table S9).CD, Crohn's Disease.

taxa, such as the association of *Chloroplasts/Cyanobacteria* to Indian samples, possibly originating from higher plant intake in this region or higher colonisation with *Cyanobacteria* (see online supplementary tables S7 and S8). Again, the association of individual bacterial species to disease conditions is rare, and even absent in the rDNA-based analyses (see online supplementary figure S14 and table S7).

Further analysis of bacterial activity (rDNA/rRNA ratio) reveals several taxa with different mean activities (ie, dormant vs active) between pathologies (see online supplementary table S9) and among populations (see online supplementary tables S9, S10 and figure S16). An interesting generalisable finding is that the genera *Papillibacter* and *Anaerococcus* appear more active in healthy controls, which further strengthens the role of *Papillibacter* as a general health indicator. *Bacteroides*, on the other hand, is more active in patients with UC and *Faecalibacterium* lies almost dormant in patients with CD (figure 4).

DISCUSSION

There is increasing evidence that diet and socioeconomic conditions largely influence the composition of the intestinal microbiome, yet relatively little is known about exact differences among human populations. This is especially true for studies linking bacterial dysbioses to immune-mediated diseases, or more specifically to IBD.[28] Most investigations were conducted on individual European or American focus populations,[7 13 29] although these diseases are increasing worldwide.[1] The patterns of microbial biogeography in IBD, depicting shared and private dysbiotic events between human populations, may help to understand the role of the microbiome in disease aetiology, and are of high relevance for any diagnostic and/or therapeutic approach targeting the microbiome.

Irrespective of community profiling on the 16S rDNA or rRNA level, the most influential variable throughout the analyses is the population origin of the sample. These differences in composition according to population may arise due to a number of factors. Present-day factors would include the differences in the surrounding source environments between, for example, Europe and the Indian subcontinent, as well as the accompanying cultural (eg, diet) and genetic differences between human hosts. These abiotic and biotic factors, amplified by differences in bacterial dispersal and transmission, are likely to affect diversity on multiple levels and increase the differences between microbial communities between host populations.[30 31] While the relative importance of those factors may be difficult to disentangle in this study, a recent fine-scale survey of the intestinal microbiota in wild mouse populations indicated a predominant influence of geographical distance, which in this case was stronger than the underlying genetic distance.[15]

The results of our phylum abundance analysis show congruences with other studies. The increase of Firmicutes in IBD in European samples was previously observed,[17 32] while others failed to find this pattern or even found the opposite.[28 33 34] The Indian samples, in particular, showed either no or only weak signs of differential phylum abundance and diversity among disease conditions, by contrast with former lower coverage experiments in this ethnic group.[35] These conflicting results might originate from the different sampling areas (here exclusively sigmoid colon) and types of samples (biopsies vs stool) in the respective studies. An interesting perspective emerged from the relationship of Bacteroidetes and Proteobacteria, which

show age-dependent patterns and a negative correlation across multiple populations, which could be the result of competitive exclusion of Proteobacteria by Bacteroidetes. The exclusion of Proteobacteria was suggested by a study on infant microbiomes[36] and seems to be supported in our adult cohort. The capability of Bacteroidetes to digest complex sugars, their ancient symbiotic relationship with their hosts[37 38] and their central position in the gut microbiome[39] might help to sustain their abundance over time. Further, Proteobacteria seem to belong to the early colonisers of the mammalian gut and may, therefore, be less competitive than well-adapted late colonisers over the course of succession.[36 40] Succession processes are also apparent through the increasing diversity within (α-diversity measures) and between (β-diversity analyses) subjects with age.[41] We observed an average decrease in active diversity in Indian patients and in patients with CD (figure 2E, also see online supplementary figure S4B, D), which might indicate less stable communities, although the connection between stability and diversity is not yet fully understood.[42] Furthermore, the strongest differences between the sampled populations lies among the diseased individuals, which together with disease-specific increases in α diversity with age, point towards higher community turnover and decreased colonisation resistance among diseased subjects.[43] The standing microbial communities (rDNA-based) show a general increase in species number and evenness with age (figures 2D, also see online supplementary figure S4A), while phylogenetic diversity remains relatively constant over time (figure 4C), which may be a result of the replacement of Proteobacteria by Bacteroidetes, two deep-branching bacterial groups. Likewise, the correlation of β diversity with age could be a product of succession with community turnover over time.[44]

Contrasting patterns between the active and standing community members are present at the level of phylum abundances as well as α and β diversity. The reduced species richness among the active samples and the stronger differences between populations and pathologies might be the result of limited sequencing coverage, as rRNA of rare or less active members may be outcompeted during sequencing by more active bacteria, thereby reducing the number of observed species. Nevertheless, communities obtained from either active or standing bacteria are correlated, but appear to emphasise different processes and patterns. These differences may be of particular importance in the context of IBD, as active bacteria and their products play a more significant role in inflammation than dormant bacteria.[45] Dormant bacteria, on the other hand, can balance community disturbances and maintain diversity in the microbiome as a 'seed bank'.[46]

With respect to the understanding and interpretation of disease-associated microbial patterns, the influence of the study population is of great importance as it overshadows that of disease condition. This, in part, may be one explanation for the often inconsistent findings among studies of patients with IBD.[47] However, we identified two interesting exceptions displaying consistency across all populations. The first concerns the activity of *Faecalibacteria,* which is specifically reduced among patients with CD (figure 4). This adds to a growing list of examples demonstrating a reduction of *Faecalibacteria* in the context of CD,[48 49] indicating it may be a true hallmark of CD-associated communities. A second interesting and not yet described association is the increased prevalence and activity of *Papillibacter* in healthy subjects compared with both CD and UC across all three populations (figure 4, see online supplementary figure S8). *Papillibacter* is a relative of *Faecalibacteria*, and both belong to the *Clostridium leptum* subgroup,[50] which are common butyrate producers. This finding further emphasises the importance of short-chain fatty acid producers for enterocyte homeostasis.[51] A recent targeted case–control study of the *Clostridium leptum* subgroup in an independent Indian cohort revealed similar results in the context of IBD,[52] and further supports the use of this group as 'biomarkers'.

By contrast with the low number of 'universal' disease indicators, we identified a greater number of taxa displaying disease-by-population associations (see online supplementary tables S5–S8). Another interesting observation, the increased activity, and to some extent abundance of *Bacteroides* in patients with UC, suggests a high adaptability and exploitation of the disturbed mucosa in IBD by this genus.[53] Also, in the light of recent findings in a large, early onset biopsy cohort for CD, we found several bacteria negatively associated with CD in common (eg, *Bacteroides*, *Blautia*, *Ruminococcus*, *Roseburia*, *Coprococcus*, *Lachnospiraceae*, *Faecalibacteria*).[54] Associations of those bacteria were mainly restricted to European samples, again stressing population-specific differences in microbiome composition and the need for broad sampling. The only shared genus positively associated with CD is *Prevotella*, which in our study associates with Indian samples and has been reported to associate with non-Western microbiomes.[41] A possible concern with these results may lie with differences in diagnostic criteria between study cohorts, which could contribute to the heterogeneity in disease patterns. However, no over-representation of, for example, known pathogenic genera is identified among the taxa specific to IBD in any given location. An exception may be the higher abundance of the genus, *Campylobacter*, among the German controls, but this does not argue in favour of differences in diagnosis due to, for example, failure to identify pathogens.

In summary, our study provides several important findings that advance our understanding of the forces shaping diversity of the intestinal microbiota and their relationship with the disease. These include the influence of age and host population on numerous aspects of community composition and structure. We identify both shared and private IBD-related signatures regarding bacterial abundances, activity and community diversity in the investigated cohorts. It is important to note that our observations were made, in part, at the level of actively transcribing community members and highlights the merits of additional 16S rRNA profiling as a promising approach to identify disease-relevant bacterial-derived biomarkers in future studies.

Author affiliations
[1]Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany
[2]Max Planck Institute for Evolutionary Biology, Plön, Germany
[3]Institute for Experimental Medicine, Christian-Albrechts-University of Kiel, Kiel, Germany
[4]Institute for Digestive Research, Medical Academy, Lithuanian University of Health Sciences, Kaunas, Lithuania
[5]Department of Gastroenterology, Medical Academy, Lithuanian University of Health Sciences, Kaunas, Lithuania
[6]Bombay Hospital and Medical Research Center, Mumbai, India
[7]Department of General Internal Medicine, Christian-Albrechts-University of Kiel, University Hospital Schleswig-Holstein, Kiel, Germany

Acknowledgements We thank all study participants and Manuela Kramp and Dorina Ölsner for excellent technical assistance.

Contributors AR, SS, and SO designed the research; AR, PRa., JS, GK, KB, DA, LK, SO and PRo performed the research; PRa, JW and JFB analysed the data; PRa, PRo and JFB wrote the paper.

Funding This work was supported by the Deutsche Forschungsgemeinschaft ExC 306 Excellence Cluster 'Inflammation at Interfaces' (CL Nucleotide lab, CL CCIM and RA Envirome) and the Broad Medical Research Program (IBD-0248R).

Competing interests None.

Patient consent Obtained.

Ethics approval University Hospital Schleswig-Holstein Ethics Committee (B231/98 and A154/06); Kaunas Regional Biomedical Research Ethics Committee (P2-84/2003); and Bombay Hospital (Mumbai, Maharashtra State) and Research Center (dated 8th July 2009).

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Raw sequence data and related metadata can be accessed at the European Nucleotide Archive (ENA) under the accession number PRJEB6172.

REFERENCES

1 Thia KT, Loftus JEV, Sandborn WJ, *et al*. An update on the epidemiology of inflammatory bowel disease in Asia. *Am J Gastroenterol* 2008;103:3167–82.
2 Burisch J, Pedersen N, Čuković-Čavka S, *et al*. East–West gradient in the incidence of inflammatory bowel disease in Europe: the ECCO-EpiCom inception cohort. *Gut* 2014;63:588–97.
3 Hampe J, Heymann K, Krawczak M, *et al*. Association of inflammatory bowel disease with indicators for childhood antigen and infection exposure. *Int J Colorectal Dis* 2003;18:413–7.
4 Gent AE, Coggon D, Swarbrick ET, *et al*. Inflammatory bowel-disease and domestic hygiene in childhood. *Gastroenterology* 1994;106:A686.
5 Guarner F, Bourdet-Sicard R, Brandtzaeg P, *et al*. Mechanisms of disease: the hygiene hypothesis revisited. *Nat Clin Pract Gastroenterol Hepatol* 2006;3:275–84.
6 Hugot JP, Alberti C, Berrebi D, *et al*. Crohn's disease: the cold chain hypothesis. *Lancet* 2003;362:2012–5.
7 Consortium THMP. Structure, function and diversity of the healthy human microbiome. *Nature* 2012;486:207–14.
8 Yatsunenko T, Rey FE, Manary MJ, *et al*. Human gut microbiome viewed across age and geography. *Nature* 2012;486:222–7.
9 Baas-Becking LGM. Geobiologie of Inleiding Tot de Milieukunde. 1934.
10 Fierer N, Jackson RB. The diversity and biogeography of soil bacterial communities. *Proc Natl Acad Sci* 2006;103:626–31.
11 Nemergut DR, Costello EK, Hamady M, *et al*. Global patterns in the biogeography of bacterial taxa. *Environ Microbiol* 2011;13:135–44.
12 Lindström ES, Langenheder S. Local and regional factors influencing bacterial community assembly. *Environ Microbiol Rep* 2012;4:1–9.
13 Qin J, Li R, Raes J, *et al*. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 2010;464:59–65.
14 Karlsson FH, Tremaroli V, Nookaew I, *et al*. Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature* 2013;498:99–103.
15 Linnenbrink M, Wang J, Hardouin EA, *et al*. The role of biogeography in shaping diversity of the intestinal microbiota in house mice. *Mol Ecol* 2013;22:1904–16.
16 Rehman A, Lepage P, Nolte A, *et al*. Transcriptional activity of the dominant gut mucosal microbiota in chronic inflammatory bowel disease patients. *J Med Microbiol* 2010;59:1114–22.
17 Rausch P, Rehman A, Künzel S, *et al*. Colonic mucosa-associated microbiota is influenced by an interaction of Crohn disease and FUT2 (Secretor) genotype. *Proc Natl Acad Sci* 2011;108:19030–5.
18 Schloss PD, Westcott SL, Ryabin T, *et al*. Introducing mothur: open source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 2009;75:7537–41.
19 Jost L. Partitioning diversity into independent alpha and beta components. *Ecology* 2007;88:2427–39.
20 Faith DP. Conservation evaluation and phylogenetic diversity. *Biol Conserv* 1992;61:1–10.
21 Kembel SW, Cowan PD, Helmus MR, *et al*. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* 2010;26:1463–4.
22 Hamady M, Lozupone C, Knight R. Fast UniFrac: facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data. *ISME J* 2010;4:17–27.
23 Oksanen J, Blanchet FG, Kindt R, *et al*. vegan: Community Ecology Package. 2011. http://CRAN.R-project.org
24 Anderson MJ. A new method for non-parametric multivariate analysis of variance. *Austral Ecology* 2001;26:32–46.
25 Legendre P, Legendre L. *Numerical ecology*. 2nd English edn. Developments in Environmental Modelling, 1998;20:i–xv, 1–853.
26 De Cáceres M, Legendre P, *et al*. Improving indicator species analysis by combining groups of sites. *Oikos* 2010;119:1674–84.
27 Chao A. Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* 1987;43:783–91.
28 Frank DN, St. Amand AL, Feldman RA, *et al*. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc Natl Acad Sci* 2007;104:13780–5.
29 Costello EK, Lauber CL, Hamady M, *et al*. Bacterial community variation in human body habitats across space and time. *Science* 2009;326:1694–7.
30 Whitaker RJ, Grogan DW, Taylor JW. Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science* 2003;301:976–8.
31 Falush D, Wirth T, Linz B, *et al*. Traces of human migrations in helicobacter pylori populations. *Science* 2003;299:1582–5.
32 Lepage P, Häsler R, Spehlmann ME, *et al*. Twin study indicates loss of interaction between microbiota and mucosa of patients with ulcerative colitis. *Gastroenterology* 2011;141:227–36.
33 Gophna U, Sommerfeld K, Gophna S, *et al*. Differences between tissue-associated intestinal microfloras of patients with Crohn's disease and ulcerative dolitis. *J Clin Microbiol* 2006;44:4136–41.
34 Manichanh C, Rigottier-Gois L, Bonnaud E, *et al*. Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* 2006;55:205–11.
35 Verma R, Verma AK, Ahuja V, *et al*. Real-time analysis of mucosal flora in patients with inflammatory bowel disease in India. *J Clin Microbiol* 2010;48:4279–82.
36 Trosvik P, Stenseth NC, Rudi K. Convergent temporal dynamics of the human infant gut microbiota. *ISME J* 2009;4:151–8.
37 Backhed F, Ley RE, Sonnenburg JL, *et al*. Host-bacterial mutualism in the human intestine. *Science* 2005;307:1915–20.
38 Xu J, Mahowald MA, Ley RE, *et al*. Evolution of symbiotic bacteria in the distal human intestine. *PLoS Biol* 2007;5:e156.
39 Xu J, Bjursell MK, Himrod J, *et al*. A genomic view of the human-bacteroides thetaiotaomicron symbiosis. *Science* 2003;299:2074–6.
40 Bazzaz FA. Physiological ecology of plant succession. *Annu Rev Ecol Syst* 1979;10:351–71.
41 Lozupone C, Stombaugh J, Gonzalez A, *et al*. Meta-analyses of studies of the human microbiota. *Genome Res* 2013;23:1704–14.
42 May RM. Will a large complex system be stable? *Nature* 1972;238:413–4.
43 Stecher B, Hardt W-D. Mechanisms controlling pathogen colonization of the gut. *Curr Opin Microbiol* 2011;14:82–91.
44 Faith JJ, Guruge JL, Charbonneau M, *et al*. The long-term stability of the human gut microbiota. *Science* 2013;341:1237439.
45 Bron PA, van Baarlen P, Kleerebezem M. Emerging molecular insights into the interaction between probiotics and the host intestinal mucosa. *Nat Rev Micro* 2012;10:66–78.
46 Jones SE, Lennon JT. Dormancy contributes to the maintenance of microbial diversity. *Proc Natl Acad Sci* 2010;107:5881–6.
47 Sokol H, Lay C, Seksik P, *et al*. Analysis of bacterial bowel communities of IBD patients: What has it revealed? *Inflamm Bowel Dis* 2008;14:858–67.
48 Sokol H, Pigneur B, Watterlot L, *et al*. Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc Natl Acad Sci* 2008;105:16731–6.
49 Fujimoto T, Imaeda H, Takahashi K, *et al*. Decreased abundance of Faecalibacterium prausnitzii in the gut microbiota of Crohn's disease. *J Gastroenterol Hepatol* 2013;28:613–19.
50 Shen J, Zhang B, Wei G, *et al*. Molecular profiling of the clostridium leptum subgroup in human fecal microflora by PCR-denaturing gradient gel electrophoresis and clone library analysis. *Appl Environm Microbiol* 2006;72:5232–8.
51 Segain J-P, de la Blétière DR, Bourreille A, *et al*. Butyrate inhibits inflammatory responses through NFκB inhibition: implications for Crohn's disease. *Gut* 2000;47:397–403.
52 Kabeerdoss J, Sankaran V, Pugazhendhi S, *et al*. *Clostridium leptum* group bacteria abundance and diversity in the fecal microbiota of patients with inflammatory bowel disease: a case-control study in India. *Bmc Gastroenterol* 2013;13:20.
53 Wexler HM. Bacteroides: the good, the bad, and the Nitty-Gritty. *Clin Microbiol Rev* 2007;20:593–621.
54 Gevers D, Kugathasan S, Denson Lee A, *et al*. The treatment-naive microbiome in new-onset Crohn's Disease. *Cell Host Microbe* 2014;15:382–92.

Supplemental Information

Geographic patterns of the standing and active human gut microbiome in health and inflammatory bowel disease

Ateequr Rehman[1*], Philipp Rausch[2,3*], Jun Wang[2,3], Jurgita Skieceviciene[1,7], Gediminas Kiudelis[4], Ketan Bhagalia[5], Deepak Amarapurkar[5], Limas Kupcinskas[4,7], Stefan Schreiber[1,6], Philip Rosenstiel[1,#], John F. Baines[2,3,#], Stephan Ott[6,#]

*Contributed equally to this study; [#]**corresponding authors**

[1]Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, D-24105 Kiel, Germany

[2]Max Planck Institute for Evolutionary Biology, D-24306 Plön, Germany

[3]Institute for Experimental Medicine, Christian-Albrechts-University of Kiel, D-24105 Kiel, Germany

[4]Department of Gastroenterology, Medical Academy, Lithuanian University of Health Sciences, Kaunas, Lithuania

[5]Bombay Hospital and Medical Research Center, Mumbai 400020, India

[6]Department of General Internal Medicine, Christian-Albrechts-University of Kiel, University Hospital Schleswig-Holstein, Campus Kiel, D-24105 Kiel, Germany

[7]Institute for Digestive Research, Medical Academy, Lithuanian University of Health Sciences, Kaunas, Lithuania

1. Supplemental Materials and Methods

DNA/RNA extraction and reverse transcription

In brief, biopsies frozen directly in liquid nitrogen were homogenized using stainless steel pestles cooled with liquid nitrogen, followed by immediate addition of 600 µl of RLT plus buffer. Mechanical homogenization was next performed in Lysing matrix E tubes (MP Biomedicals Germany) using the FASTPrep FP 120 instrument (Thermo Scientific). This Lysate was incubated at room temperature to enhance host and bacterial cell lysis. The homogenate was centrifuged through a QIAshredder column (Qiagen, Hilden Germany). The remainder of the procedure followed the manufacturer's instructions as previously described [1] and RNA was reverse transcribed with random hexamers (Qiagen, Hilden, Germany). 10 µl of RNA was reverse transcribed to cDNA in a final reaction of 20 µl using the High

Capacity cDNA Reverse Transcription Kit (Life Technologies, Darmstadt, Germany). Each reaction mixture consisted of 2 µl of 10 x RT bufffer, 0.8 µl of 25 x dNTP Mix (100 mM), 2 µl of 10 x RT Random hexamer Primers, 1 µl of MultiScribe™ Reverse Transcriptase, 1 µl of RNase inhibitors, and the reaction volume was adjusted to 10 µl. Reaction mixtures were incubated for 10 min at 25°C, 120 minutes at 37°C, followed by 85°C for 5 minutes.

16S rRNA/rRNA gene amplification and pyrosequencing

The 16S rRNA and rRNA genes were subsequently amplified and sequenced as described in Rausch *et al.* (2011) [2] using forward (27F 5′-***CTATGCGCCTTGCCAGCCCGC*****TCAG**TCAGAGTTTGATCCTGGCTCAG -3′) and reverse (338R 5′-***CGTATCGCCTCCCTCGCGCCA***TCAGXXXXXXXXXXCATGCTGCCTCCCGTAGGAGT -3′) fusion primers. These fusion primer pairs are able to amplify bacteria specific 16S rRNA hypervariable region V1 and V2. The 454 Life Sciences primer B (forward) and A (reverse) adapter sequences are denoted as bold and italics while the underlined sequences represent the broadly conserved bacterial primers 27F and 338R. A two-base linker sequence (TC/CA) and four-base key (TCAG) were added as recommended by Roche (454). A unique 10 base multiplex identifier (MID; designated as XXXXXXXXXX) was added to the reverse primer to tag each PCR product. 100 ng of template DNA/cDNA was added to 25 µl PCR reaction. PCR was performed using Phusion® Hot Start DNA Polymerase (Finnzymes, Espoo, Finland) with following cycling conditions: initial denaturation for 30 sec at 98°C; 30 cycles of 9 sec at 98°C, 30 sec at 55°C, and 30 sec at 72°C; final extension for 10 min at 72°C. All reactions were performed in duplicate and combined after PCR. Amplicons were run on 2.0% agarose gel for size control, subsequently PCR products were excised and purified with the Qiagen MiniElute Gel Extraction Kit (Qiagen, Hilden, Germany). Purified amplicons were quantified with the Quant-iT™ dsDNA BR Assay Kit on a Qubit® fluorometer (Invitrogen, Darmstadt, Germany). Equimolar amounts of purified PCR product were pooled prior to emulsion PCR and subsequently sequenced on 454 GS-FLX using Titanium sequencing chemistry.

Sequence processing

Raw sequences were filtered by mothur v.1.15.0 [3]. To reduce sequencing errors we required no ambiguous bases, a mean quality score of ≥20 and a minimum length of 200 nucleotides. Chimeric sequences were determined using ChimeraSlayer and Usearch 4.1 (UCHIME algorithm) with a curated seed database based on the SILVA gold sequence collection [4, 5]. Sequences were confirmed as bacterial using the RDP classifier with a 60% bootstrap threshold [6]. For all downstream analyses of diversity and habitat association, we

took a random subset of 1000 sequences per sample to normalize the read distribution. The remaining sequences were aligned to the curated Greengenes database (release 2010) [7] using the k-mer alignment procedure PyNAST [8] and subsequent OTU binning via Usearch 4.1 [5]. Phylogenetic tree construction was carried out using FastTree 2.1 with a CAT substitution model and gamma correction [9].

2. Supplemental Analyses

Analysis of beta diversity with respect to pharmaceutical treatments. As medication may influence microbial communities and confound analyses, we investigated these variables (listed in Table S1) and corrected for biases in the frequency of medication among populations and disease conditions. Specifically, we analyzed the use of antibiotics, probiotics, immunosuppressants (AZA, ASA, TNF blockers) and corticosteroids.

No effect of any pharmaceutical treatment was discovered when partial redundancy analysis was applied with treatment as explanatory- and population and disease as conditioning variables (rDNA: antibiotics- P=0.113, probiotics- P=0.101, immune suppressants- P=0.796, corticosteroids- P=0.202; rRNA: antibiotics- P=0.201, probiotics- P=0.211, immune suppressants- P=0.913, corticosteroids- P=0.700). Thus, after controlling for the effects of population and disease, no influence of pharmaceutical treatment was detected. Analysing the effect of pharmaceutical treatments within populations (*adonis*, stratified by disease condition) revealed a weak, non-significant influence of corticosteroid treatment on the abundance of stagnant bacteria in the German samples ($F_{1,27}$=1.259, R^2=0.045, P=0.079) and immune suppressants influence the abundance of phylogenetic branches (*i.e.* weighted UniFrac; $F_{1,27}$=2.343, R^2=0.080, P=0.051). No significant impact of medication could be found in the Lithuanian samples based on DNA or RNA. In the Indian samples only a weak influence of probiotics on the standing community (Jaccard: $F_{1,28}$=1.171, R^2=0.040, P=0.064; Bray-Curtis: $F_{1,28}$=1.382, R^2=0.037, P=0.033; weighted UniFrac: $F_{1,28}$=2.383, R^2=0.078, P=0.056) and a weak effect on the compositon of the active microbial community (Jaccard: $F_{1,28}$=1.174, R^2=0.040, P=0.098) is present. Thus, by and large these results indicate that prior medication introduces no major biases in the global analysis, which is still dominated by the effects of population and disease.

3. Supplemental Figures

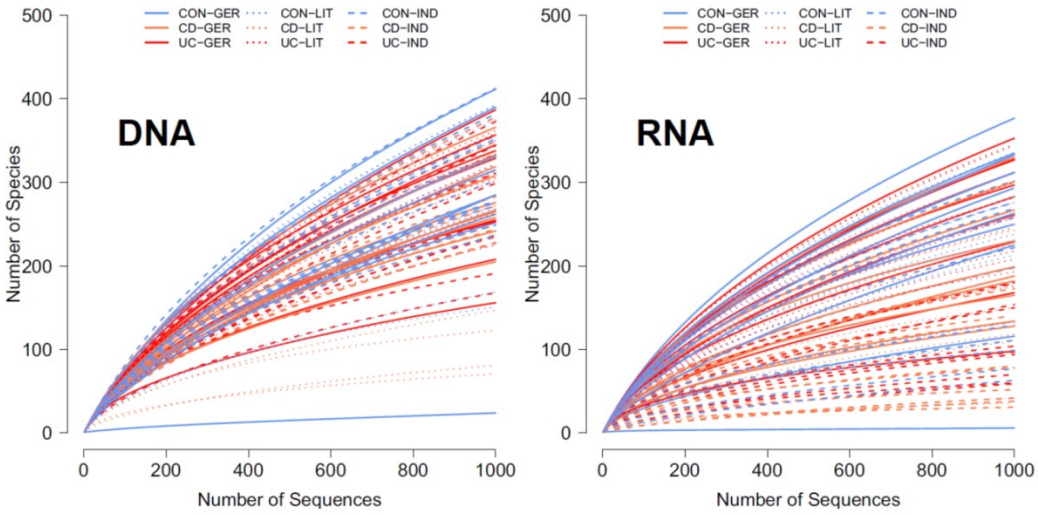


Figure S1: Individual rarefaction curves for rDNA- and rRNA-derived samples, assessed by stepwise random sampling in increments of 10 sequences.

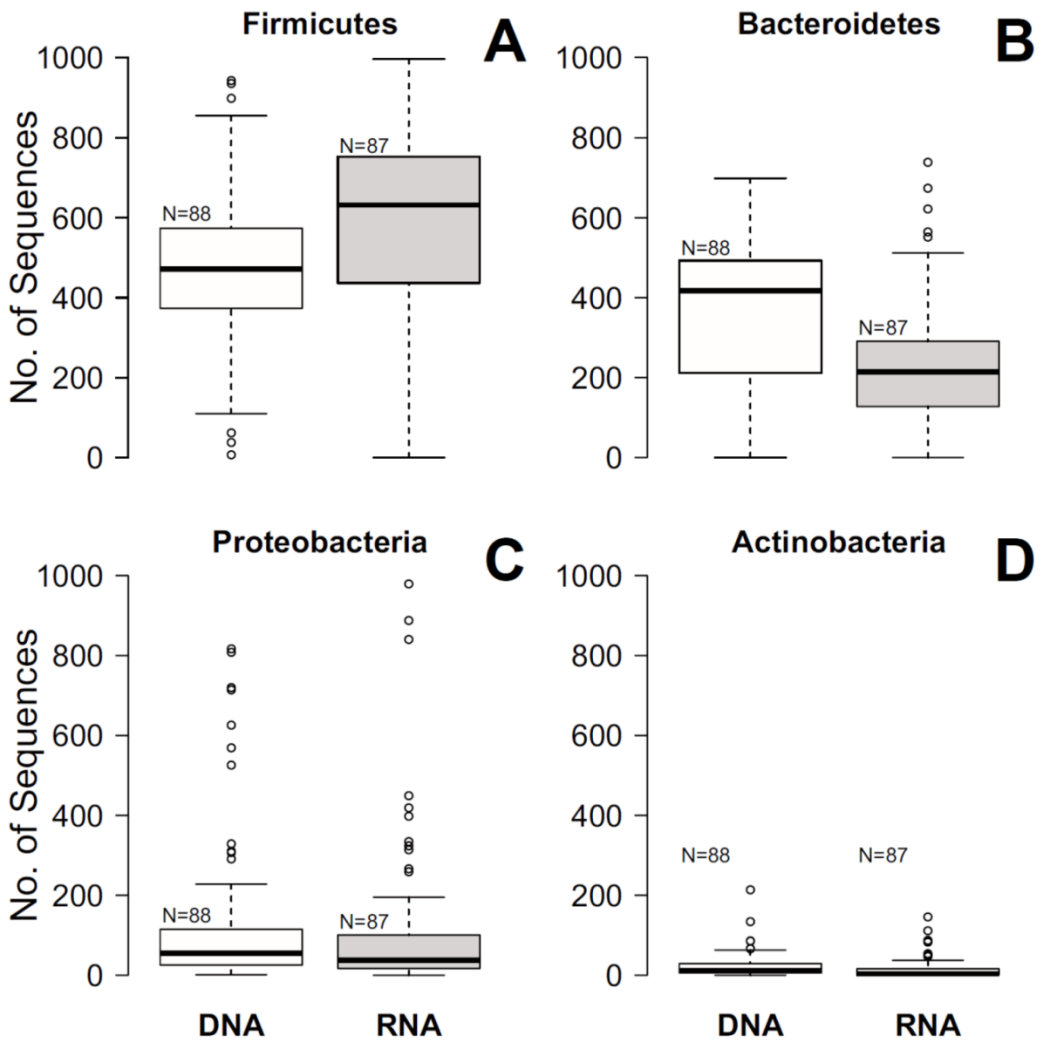


Figure S2: Comparison of phyla abundances between rDNA and rRNA (Firmicutes: V=946.5, P=6.99×10^{-5}; Bacteroidetes: V=3240.5, P=6.042×10^{-10}; Proteobacteria: V=2044, P=0.344; Actinobacteria: V=2483.5, P=0.0001; $N_{DNA,RNA}$=86, paired Wilcoxon test).

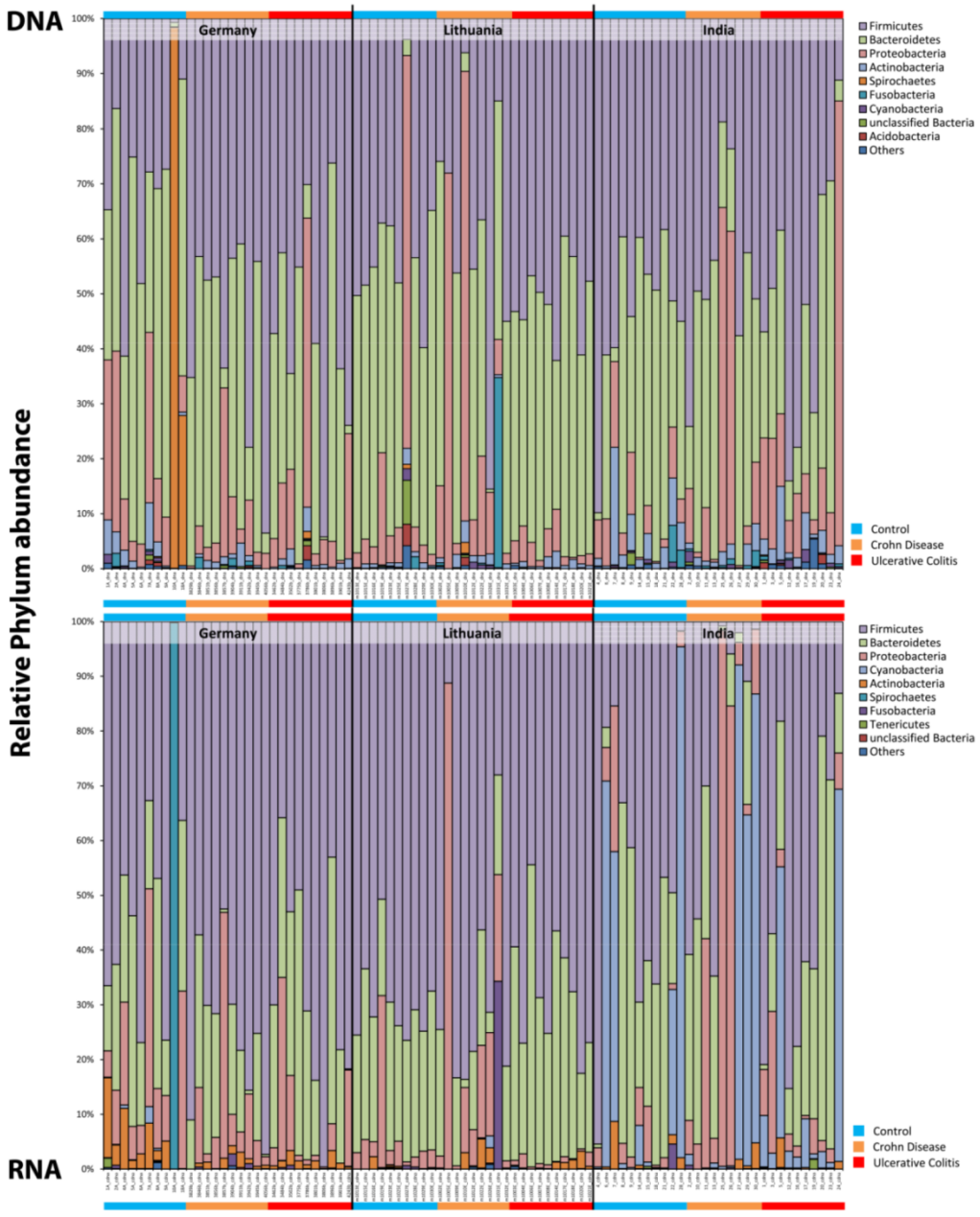


Figure S3: Individual phylum abundances in the rDNA- and rRNA-derived communities.

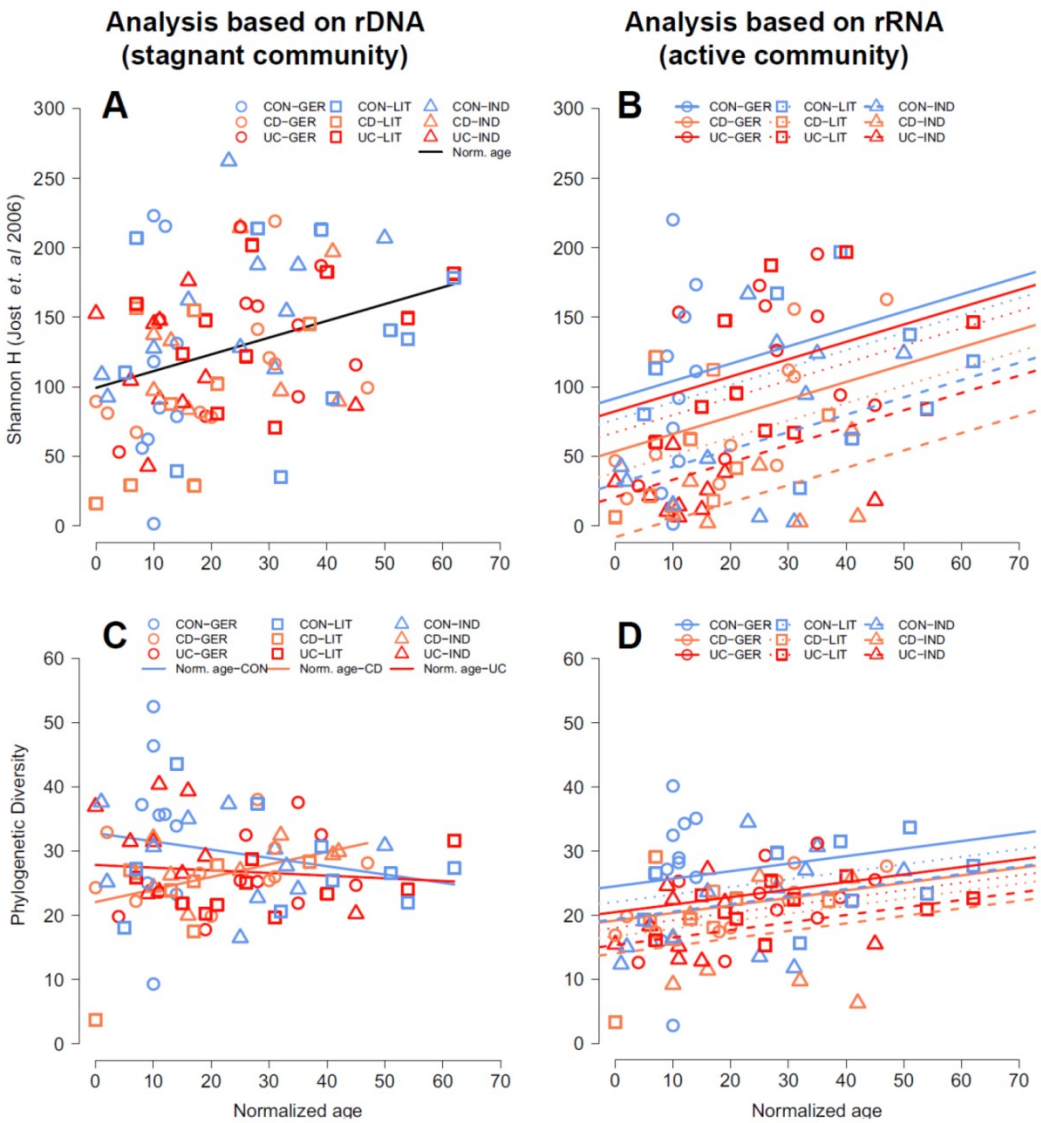


Figure S4: Distribution of alpha diversity among samples within rDNA [A, C] - and rRNA-derived [B, D] communities focusing on the distribution of species [Shannon H (Jost), A, B] and phylogenetic diversity [C, D] (see Table 1, Figure 2B, 2C).

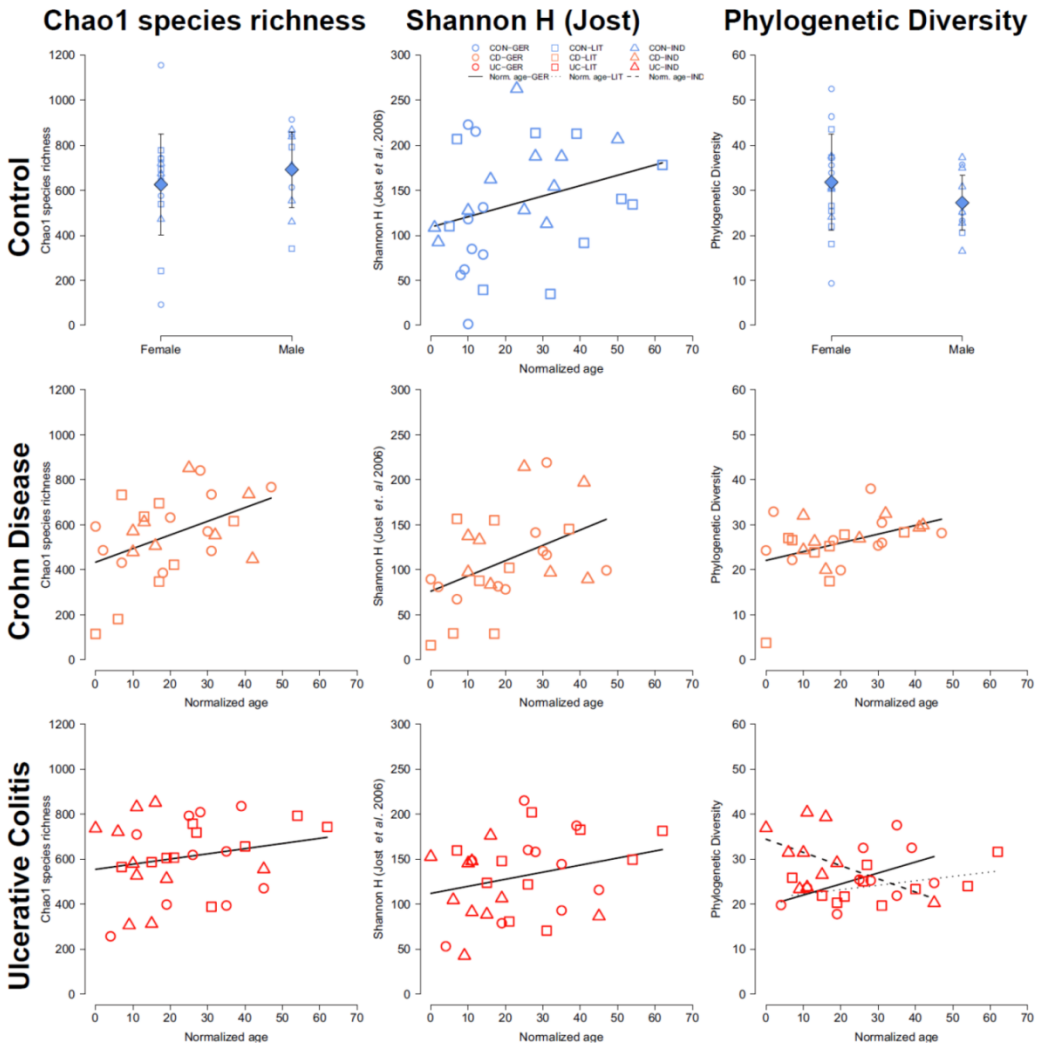


Figure S5: Analyses of alpha diversity of the rDNA-based bacterial species OTUs (97% identity) within healthy and diseased patient subsets (error bars represent SD, see also Table S3).

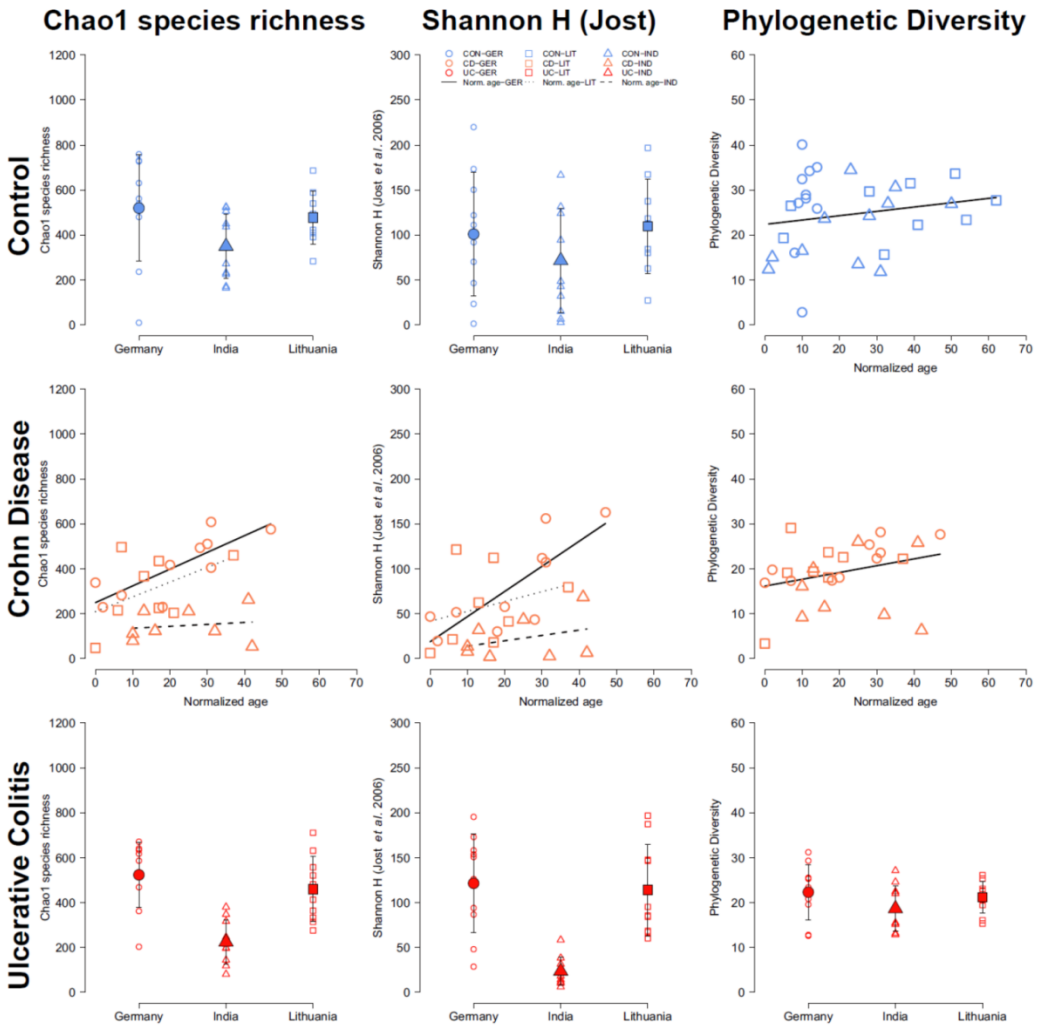


Figure S6: Analyses of alpha diversity of the rRNA-based bacterial species OTUs (97% identity) within healthy and diseased patient subsets (error bars represent SD, see also Table S3).

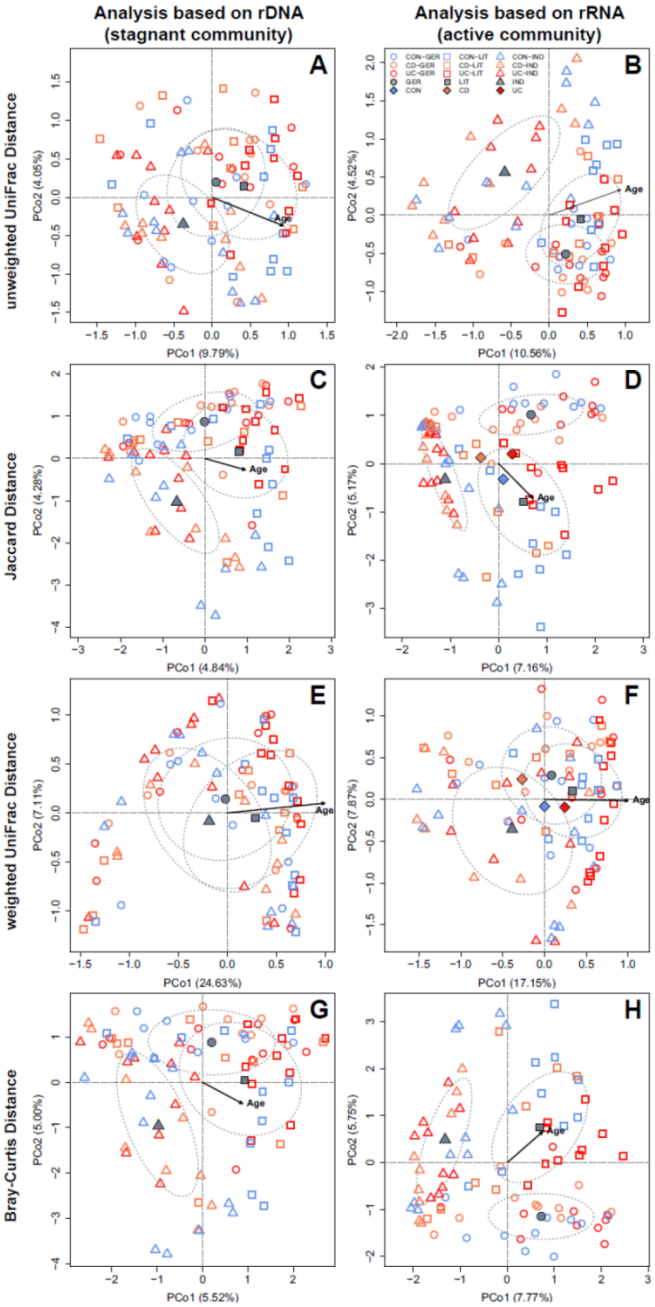


Figure S7: Principle coordinate analyzes of unweighted (A, B) UniFrac distances with centroids for the most significant factors (rDNA: population: R^2=0.145, P<0.0001, age: R^2=0.182, P=0.0002; rRNA: population: R^2=0.306, P<0.0001, age: R^2=0.122, P=0.004). C and D visualize the Jaccard distance based on OTU presence/absence using PCoA (rDNA: population: R^2=0.296, P=0.0001, age: R^2=0.199, P<0.0001; rRNA: population: R^2=0.412, P<0.0001, disease: R^2=0.055, P=0.026, age: R^2=0.199, P=0.0001). Panels E and F show the effects influencing the abundance of phylogenetic clades using the weighted UniFrac metric

(rDNA: population: R^2=0.051, P=0.068; age: R^2=0.117, P=0.006; rRNA: population: R^2=0.169, P=0.0001; age: R^2=0.093, P=0.018; disease: R^2=0.068, P=0.018). The effects seen with the weighted UniFrac metric might be obscured through the presence of an “arch-effect” in the ordination. The abundance considering OTU based metric Bray-Curtis (G, H) emphazises the effects of age and population especially among rRNA derived samples (rDNA: population: R^2=0.290, P=0.0001; age: R^2=0.183, P=0.0001; rRNA: population: R^2=0.457, P<0.0001; age: R^2=0.154, P=0.001).

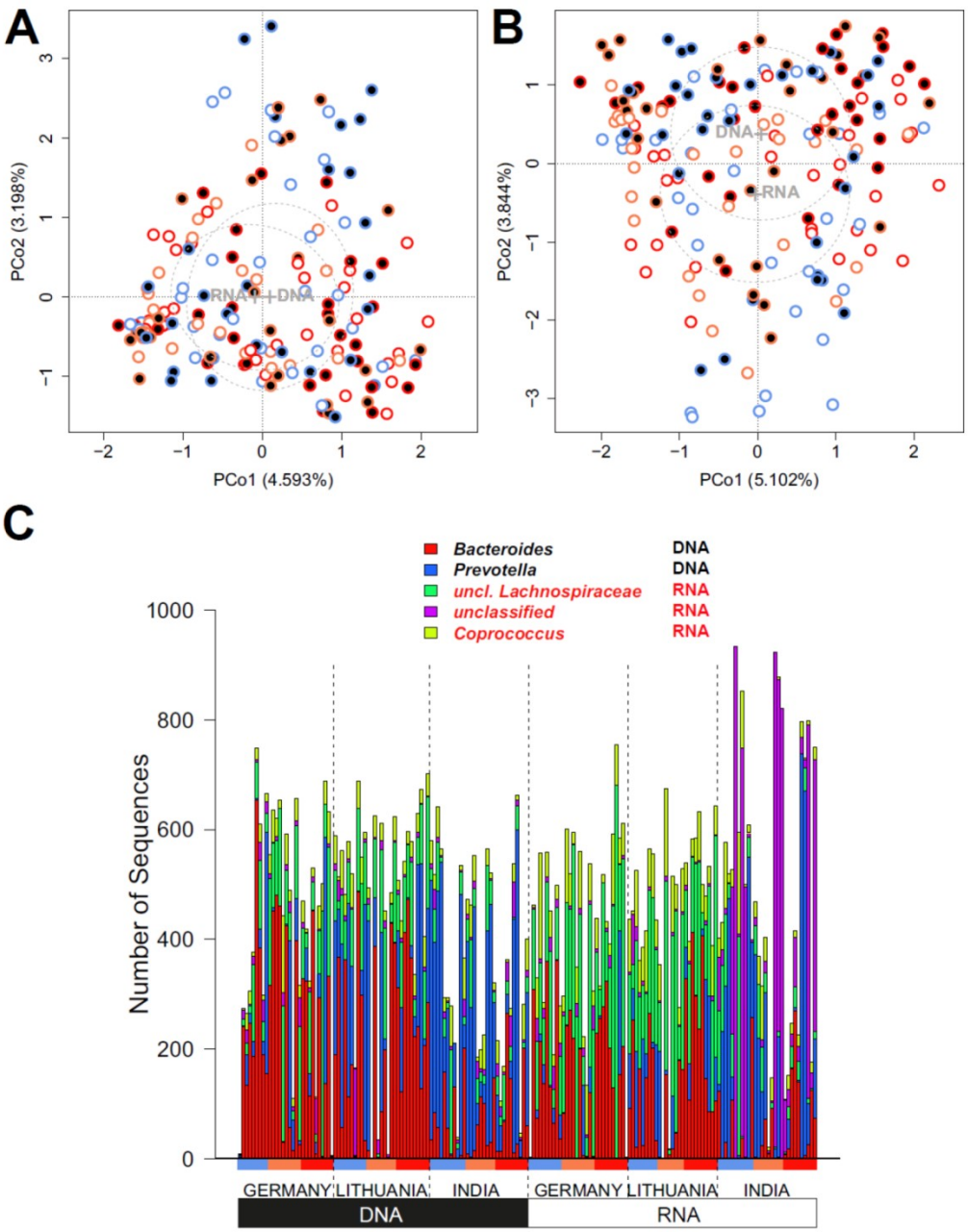


Figure S8: Clustering (A- Jaccard, B- Bray-Curtis) and correlation of rDNA (filled circle) and rRNA (open circle) derived community distances (Jaccard: Mantel- ρ=0.6951, P=0.0002; Procrustes- $m12$=0.3214, P=0.0002; Clustering: R^2=0.004, P=0.4986; Bray-Curtis: Mantel- ρ=0.7008, P=0.0002; Procrustes- $m12$=0.3013, P=0.0002, Clustering: R^2=0.0545, P<0.0001). Panel C displays significantly overrepresented genera in rDNA (black label) and rRNA (red label) with a minimum abundance of at least 1% (color bar: blue- healthy, orange- CD, red- UC).

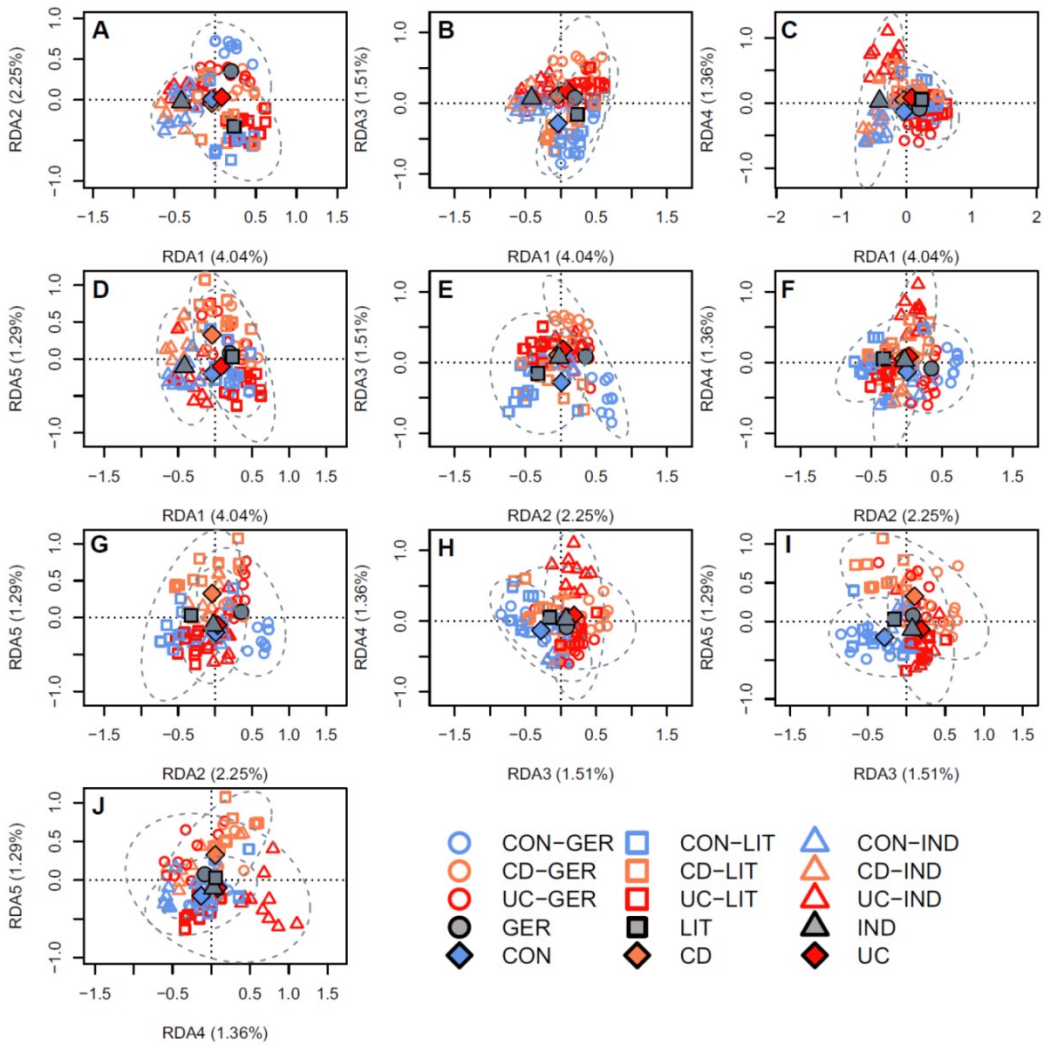


Figure S9: Panel of all significant dimensions of RDA of the rDNA-based dataset (RDA1: $F_{1,79}$=3.693, P=0.001; RDA2: $F_{1,79}$=2.054, P=0.001; RDA3: $F_{1,79}$=1.376, P=0.011 ; RDA4: $F_{1,79}$=1.239, P=0.043 ; RDA5: $F_{1,79}$=1.177, P=0.086, see also Figure 3A). Ellipses were drawn according to a *post hoc* correlation of variables and show 95% confidence interval of sample distribution, best explaining the differences among communities (A-G population, H-J disease condition).

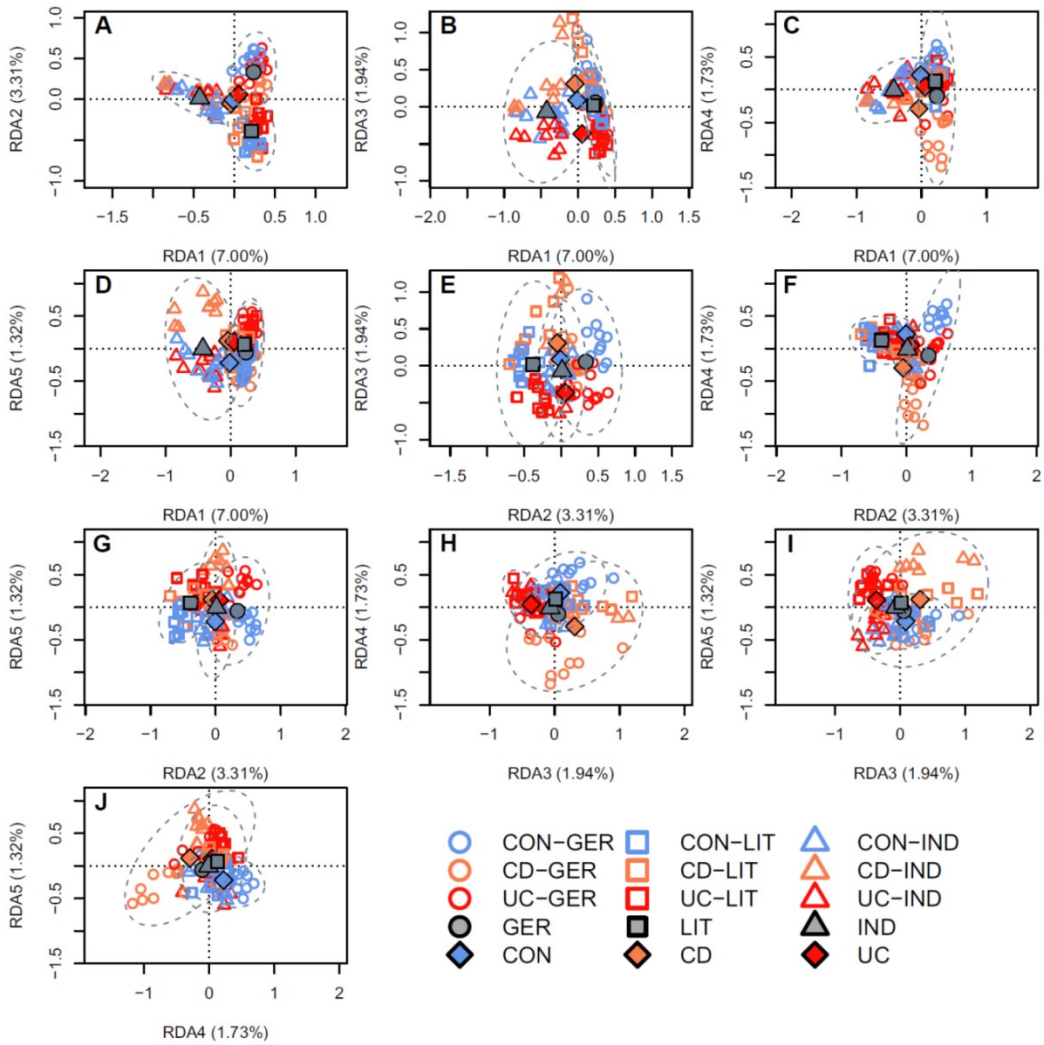


Figure S10: Panel of all significant dimensions of RDA of the rRNA-based dataset (RDA1: $F_{1,78}$=6.698, P=0.001; RDA2: $F_{1,78}$=3.172, P=0.001; RDA3: $F_{1,78}$=1.861, P=0.001; RDA4: $F_{1,78}$=1.658, P=0.005; RDA5: $F_{1,78}$=1.260, P=0.050, see also Figure 3C). Ellipses were drawn according to a *post hoc* correlation of variables and show 95% confidence interval of sample distribution, best explaining the differences among communities (A-G population, H-J disease condition).

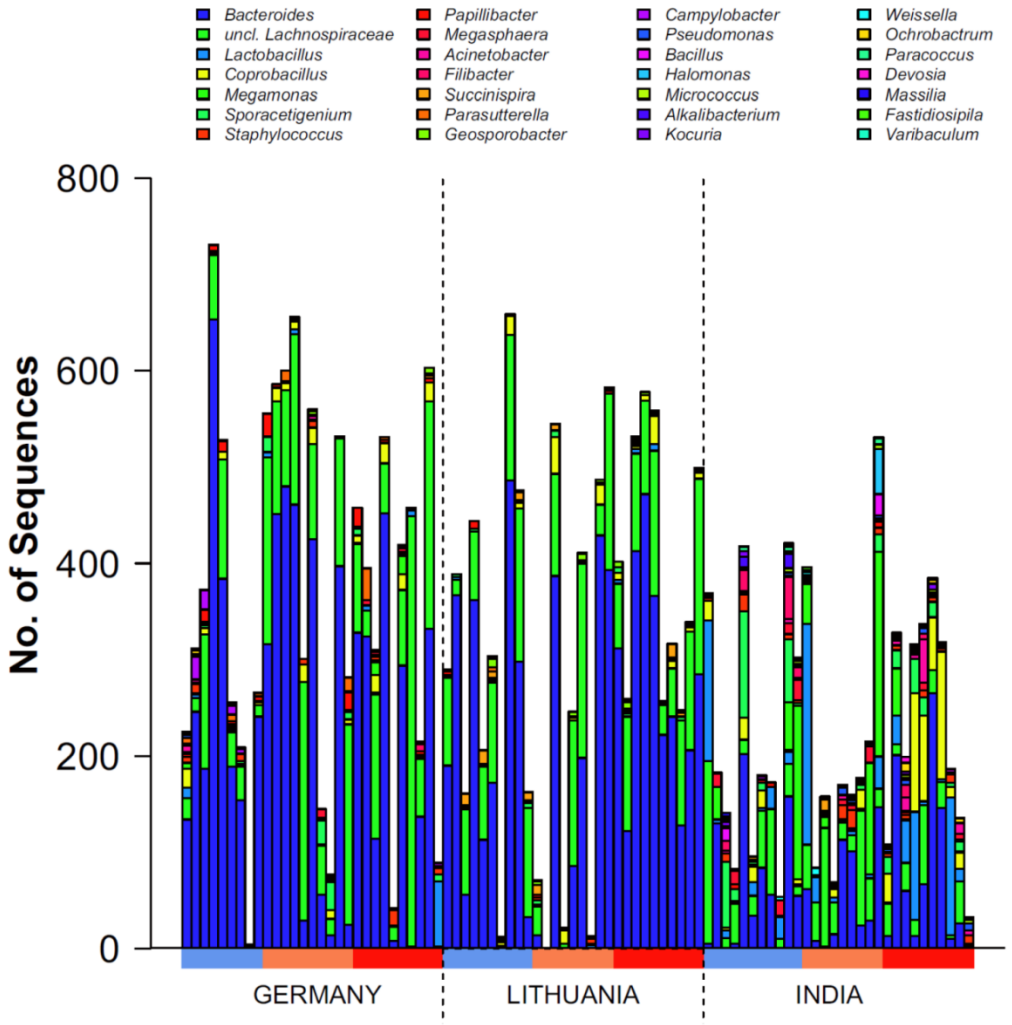


Figure S11: Distribution of indicator consensus genera in samples based on rDNA (for further details see Table S5; color bar: blue- healthy, orange- CD, red- UC).

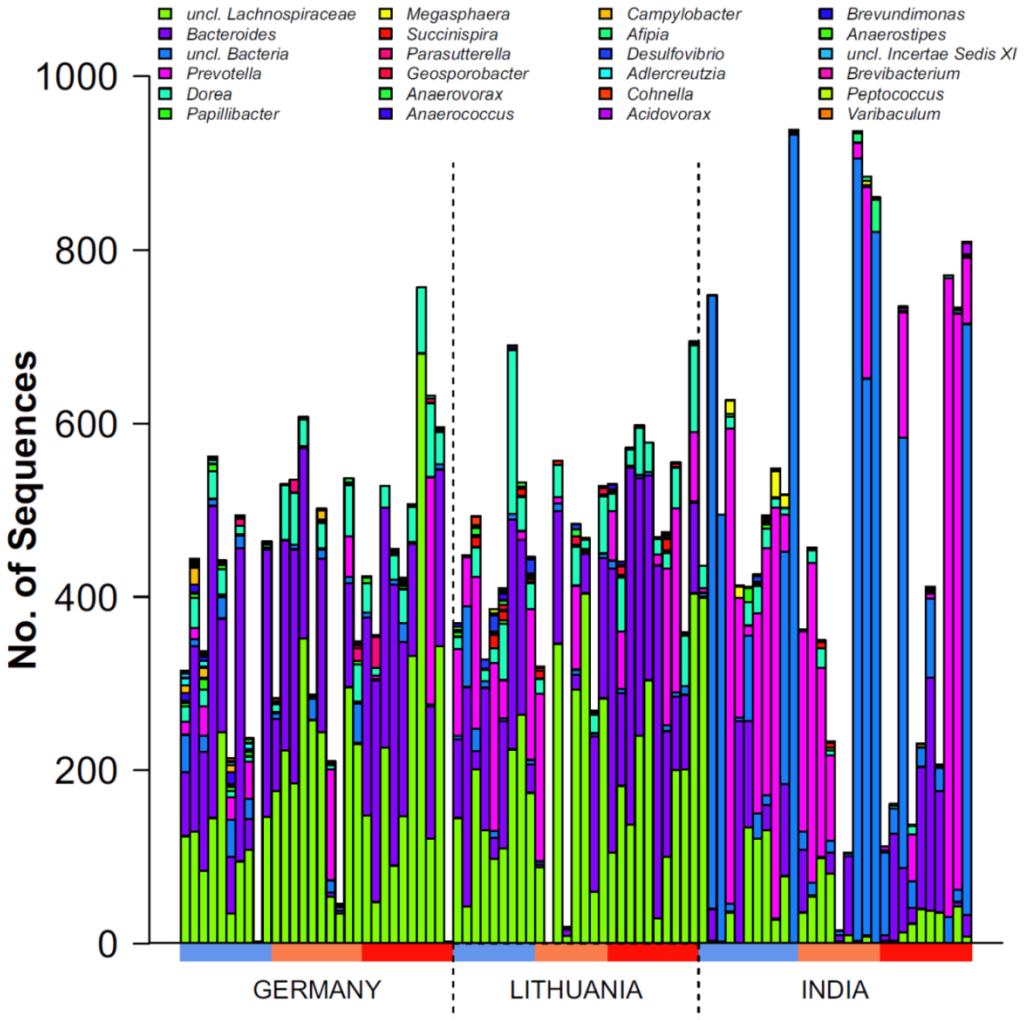


Figure S12: Distribution of indicator consensus genera in samples based on rRNA (for further details see Table S6; color bar: blue- healthy, orange- CD, red- UC).

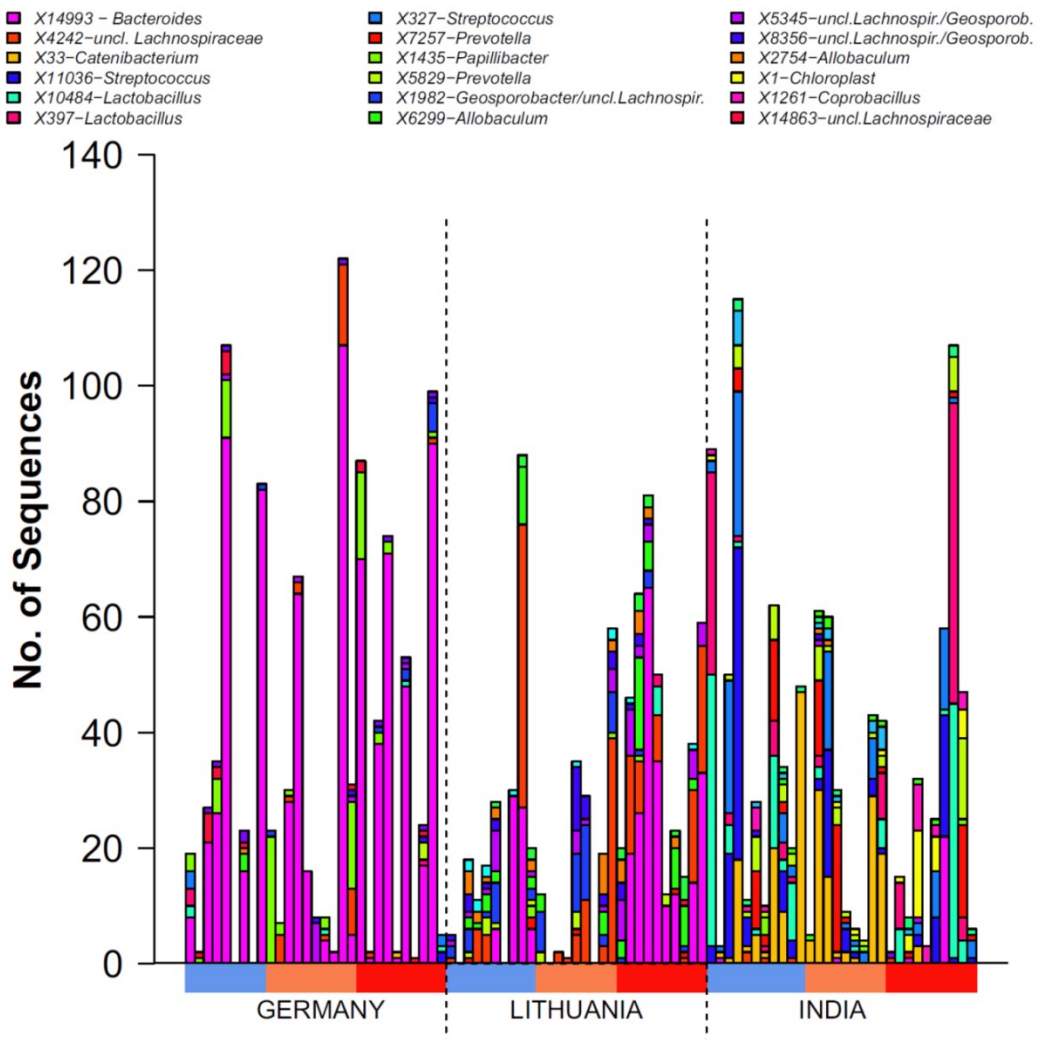


Figure S13: Distribution of indicator species OTUs in samples based on rDNA (for further details see Table S7; color bar: blue- healthy, orange- CD, red- UC).

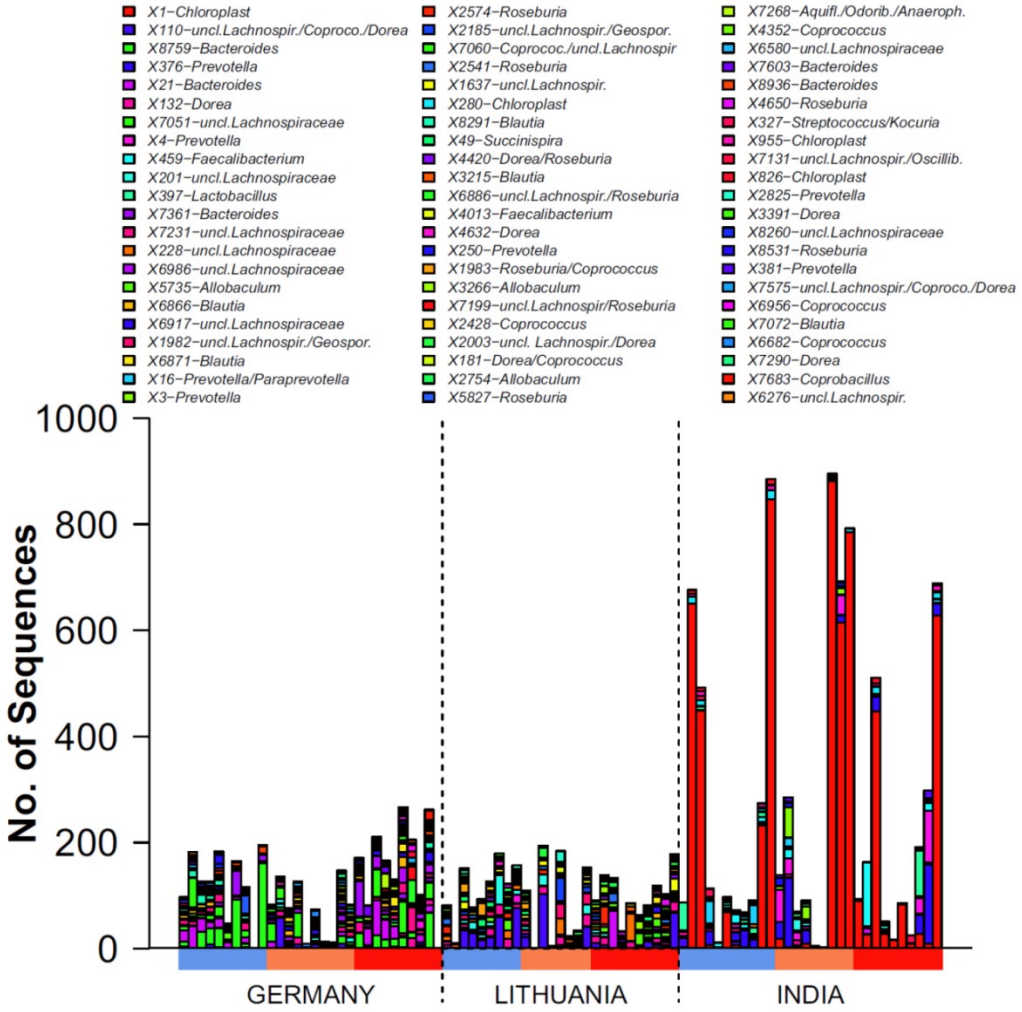


Figure S14: Distribution of indicator species OTUs in samples based on rRNA (part 1; for further details see Table S8; color bar: blue- healthy, orange- CD, red- UC).

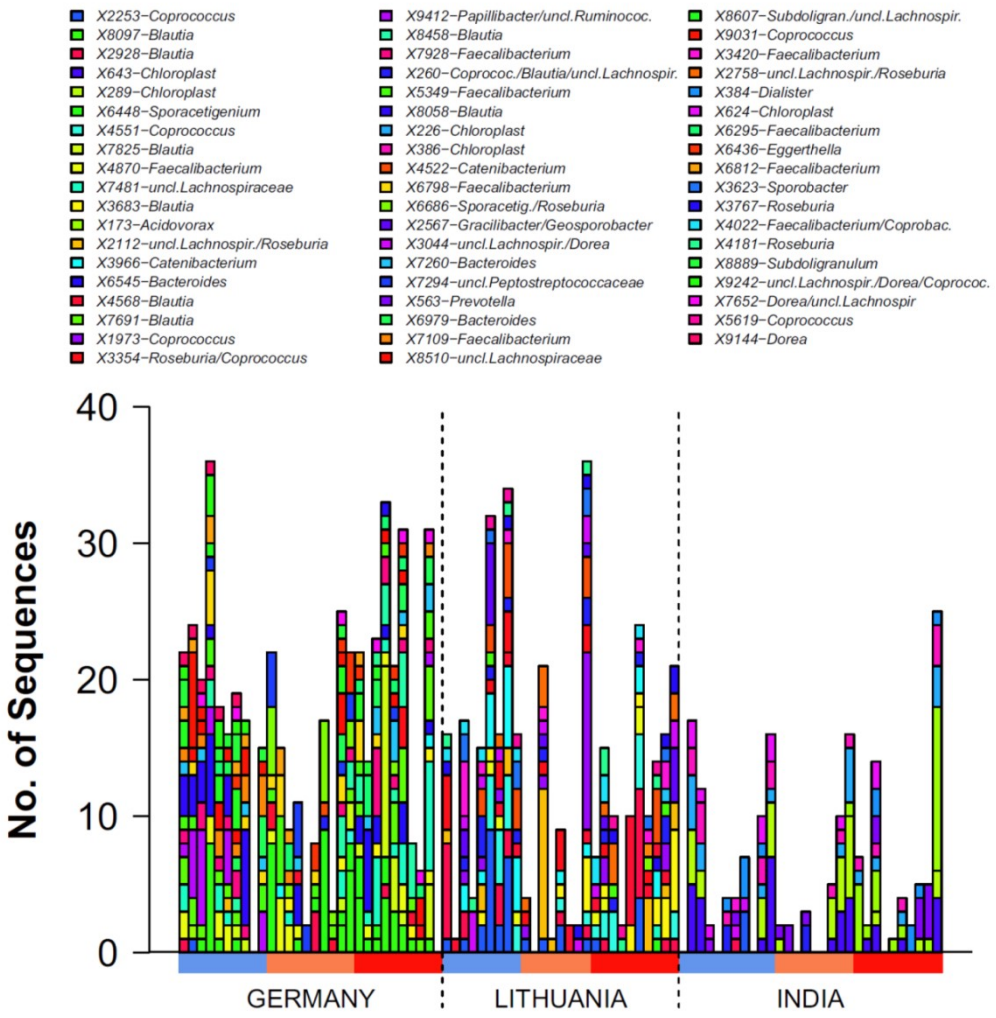


Figure S15: Distribution of indicator species OTUs in samples based on rRNA (part 2; for further details see Table S8; color bar: blue- healthy, orange- CD, red- UC).

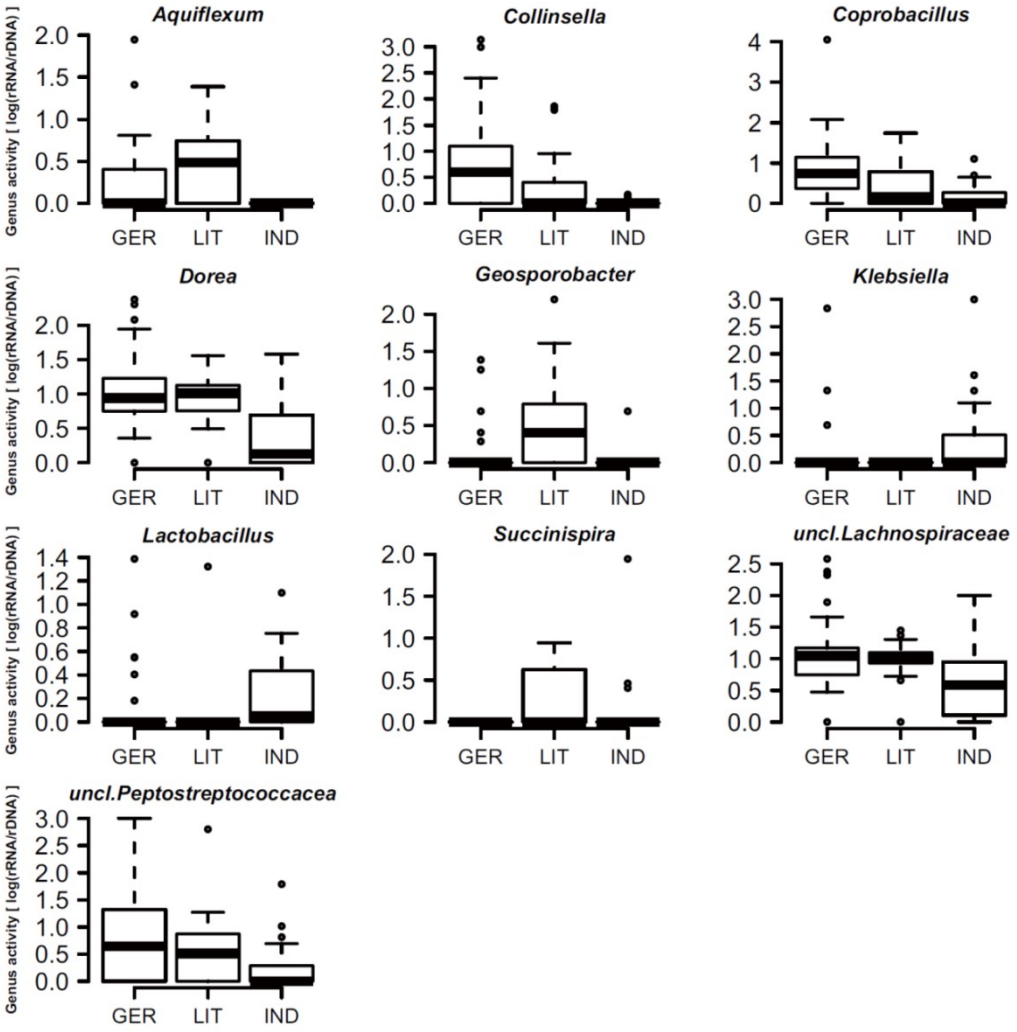


Figure S16: The plot shows the activity (rRNA/rDNA) of significant differentially active genera among the populations (Aquiflexum *P*=0.001; Collinsella *P*=0.001; *Coprobacilli P*=0.002; *Dorea P*=0.001; *Geosporobacter P*=0.001; *Klebsiella P*=0.038; *Lactobacillus P*=0.005; *Succinispira P*=0.030; unclassified *Lachnospiraceae P*=0.038; unclassified *Peptostreptococcaceae P*=0.046; *P*-values adjusted by FDR; for further details see Table S9).

4. Supplemental Tables

Table S1: Analyses of the major phyla abundances in DNA- and RNA-based samples (bold face highlights significant comparisons; * single step corrected family wise *P*-Value; ** transformation (log(X+1)); *** transformation (X^2)).

	Phylum	Model Factors	*df*	*F*	*P*	adj. R^2	Post Hoc	*P* *
DNA	Bacteroidetes	normalized age	1	6.360	**0.014**	0.059	-	-
	Firmicutes	Population	2	1.357	0.264		-	-
		Disease	2	3.229	**0.045**		CON - CD	0.766
							CON - UC	**0.041**
							UC - CD	0.216
		Population:Disease	4	4.525	**0.002**		LIT (CD) - GER (CD)	0.077
							GER (CON) - GER (CD)	**0.012**
							IND (CON) - GER (CON)	**0.038**
							GER (UC) - GER (CON)	**0.018**
							LIT (UC) - GER (CON)	0.093
	Proteobacteria **	normalized age	1	3.612	0.061	0.030	-	-
RNA	Bacteroidetes	Disease		4.739	**0.011**		CON - CD	0.126
							CON - UC	0.503
							UC - CD	**0.008**
	Firmicutes ***	Population	2	8.780	**< 0.001**		IND - GER	**0.002**
							LIT - GER	0.976
							LIT - IND	**0.001**
		Disease	2	0.796	0.455		-	-
		Population:Disease	4	2.894	**0.027**		IND (CD) - GER (CD)	**0.001**
							IND (CON) - GER (CD)	0.061
							LIT (CD) - IND (CD)	0.053
							LIT (CON) - IND (CD)	**0.022**
							GER (UC) - IND (CD)	**0.016**
							LIT (UC) - IND (CD)	**0.028**
	Proteobacteria **	Disease	2	6.352	**0.003**		CON - CD	0.574
							CON - UC	**0.038**
							UC - CD	**0.003**
		Sex	1	0.374	0.542		-	-

	Sex:Disease	2	3.829	**0.026**		F (UC) - F (CD) F (UC) - F (CON) M (CD) - F (UC) M (CON) - F (UC)	**0.001** **0.019** 0.055 **0.033**

Table S2: Analyses of Phylum abundances within population and disease subsets (bold face highlights significant comparisons; * transformation (log(X+1)) ; ** transformation (X^2) ; *** transformation ($X^{1/2}$)).

Subset	Dataset	Phylum	Best model	*DF*	*F*	*P*
Control	DNA	Firmicutes	Population	2	8.217	**0.00191**
			Normalized Age	1	0.005	0.94692
			Population:Normalized Age	2	6.033	**0.00754**
		Bacteroidetes	Normalized Age	1	3.554	0.06984
		Proteobacteria *	Normalized Age	1	2.318	0.13910
	RNA	Firmicutes **	Population	2	3.052	0.06594
			Normalized Age	1	0.164	0.68944
			Population:Normalized Age	2	2.993	0.06911
		Bacteroidetes	1		NA	NA
		Proteobacteria *	1		NA	NA
Crohn Disease	DNA	Firmicutes	Population	2	2.135	0.13930
		Bacteroidetes	Normalized Age	1	2.716	0.11240
		Proteobacteria *	Normalized Age	1	0.503	0.48583
			Sex	1	0.149	0.70323
			Normalized Age:Sex	1	7.306	**0.01299**
	RNA	Firmicutes **	Population	2	10.050	**0.00068**
		Bacteroidetes	Normalized Age	1	1.952	0.17510
		Proteobacteria *	Normalized Age	1	0.016	0.90173
			Sex	1	1.824	0.19060
			Normalized Age:Sex	1	9.250	**0.00599**
Ulcerative Colitis	DNA	Firmicutes	Population	2	0.274	0.76240
			Normalized Age	1	0.013	0.91100
			Population:Normalized Age	2	2.538	0.10010
		Bacteroidetes	Population	2	2.902	0.07215
		Proteobacteria *	Population	2	4.613	**0.01890**
	RNA	Firmicutes	1		NA	NA
		Bacteroidetes	1		NA	NA
		Proteobacteria *	Population	2	4.014	**0.02977**
Germany	DNA	Firmicutes	Disease	2	6.792	**0.00424**
		Bacteroidetes	1		NA	NA
		Proteobacteria *	1		NA	NA
	RNA	Firmicutes	Disease	2	4.219	**0.02544**
		Bacteroidetes	1		NA	NA
		Proteobacteria ***	1		NA	NA
Lithuania	DNA	Firmicutes	Disease	2	4.449	**0.02223**
		Bacteroidetes **	Normalized Age	1	2.573	0.12181
			Sex	1	1.629	0.21406
			Normalized Age:Sex	1	6.165	**0.02042**
		Proteobacteria *	Diagnosis	2	4.979	**0.01645**
			Normalized Age	1	0.474	0.49836
			Sex	1	3.047	0.09485
			Normalized Age:Sex	1	7.371	**0.01265**
	RNA	Firmicutes **	Sex	1	3.060	0.09251
		Bacteroidetes	Disease	2	7.684	**0.00264**
		Proteobacteria *	Disease	2	16.505	**0.00005**
			Sex	1	3.040	0.09586
			Disease:Sex	2	8.374	**0.00212**
India	DNA	Firmicutes	Disease	2	1.065	0.36117

		Normalized Age	1	0.138	0.71347
		Disease:Normalized Age	2	3.912	**0.03449**
	Bacteroidetes	Disease	2	1.120	0.34331
		Normalized Age	1	2.307	0.14244
		Disease:Normalized Age	2	4.798	**0.01813**
	Proteobacteria *	Normalized Age	1	5.093	**0.03232**
RNA	Firmicutes	Disease	2	2.215	0.12860
	Bacteroidetes ***	Disease	2	2.489	0.12630
	Proteobacteria *	Disease	2	3.133	0.05978

Table S3: Analyses of alpha diversities within population and disease subsets (bold face highlights significant comparisons; * transformation (X^2); ** transformation ($X^{1/2}$); *** transformation ($X^{1/4}$)).

Subset	Dataset	Alpha diversity	Best model	*DF*	*F*	*P*
Control	DNA	Shannon H	Normalized Age	1	2.744	0.10880
		Chao1	Sex	1	0.789	0.38190
		PD	Sex	1	1.905	0.17850
	RNA	Shannon H	Population	2	1.121	0.34070
		Chao1	Population	1	2.957	0.06963
			Sex	1	3.362	0.07818
		PD	Normalized Age	1	1.053	0.31360
Crohn Disease	DNA	Shannon H	Normalized Age	1	5.817	**0.02389**
		Chao1	Normalized Age	1	6.094	**0.02108**
		PD *	Normalized Age	1	4.245	0.05037
	RNA	Shannon H **	Population	2	6.997	**0.00445**
			Normalized Age	1	8.457	**0.00815**
		Chao1	Population	2	12.762	**0.00021**
			Normalized Age	1	9.691	**0.00507**
		PD	Normalized Age	1	2.398	0.13460
Ulcerative Colitis	DNA	Shannon H	Normalized Age	1	2.197	0.14940
		Chao1	Normalized Age	1	1.253	0.27250
		PD **	Population	2	3.285	0.05482
			Normalized Age	1	0.412	0.52697
			Population:Normalized Age	2	4.036	**0.03082**
	RNA	Shannon H	Population	2	15.031	**0.00004**
		Chao1	Population	2	14.296	**0.00006**
		PD	Population	2	3.979	0.05588
Germany	DNA	Shannon H	Normalized Age	1	3.679	0.06573
		Chao1	Normalized Age	1	0.512	0.48040
		PD	Sex	1	2.227	0.14720
	RNA	Shannon H	Normalized Age	1	5.687	**0.02410**
		Chao1	Disease	2	1.590	0.22304
			Normalized Age	1	6.538	**0.01674**
		PD	Disease	2	1.739	0.19546
			Normalized Age	1	4.626	**0.04096**
Lithuania	DNA	Shannon H	Normalized Age	1	4.245	**0.04950**
		Chao1 *	Normalized Age	1	6.513	**0.01693**

		PD	Normalized Age	1	1.600	0.21720
	RNA	Shannon H	Disease	2	3.348	0.05220
		Chao1	Disease	2	3.779	**0.03743**
		PD	Normalized Age	1	5.910	**0.02257**
India	DNA	Shannon H	Normalized Age	1	2.898	0.10020
		Chao1	Normalized Age	1	0.833	0.36940
		PD	Normalized Age	1	1.743	0.19790
	RNA	Shannon H ***	Disease	2	3.782	**0.03564**
		Chao1 **	Disease	2	8.815	**0.00113**
		PD	Disease	2	1.891	0.17050

Table S4: Analyses of beta diversities within populations and disease subsets (bold face highlights significant comparisons).

			DNA			RNA		
Metric	Subset	Factor	*F*	R^2	*P*	*F*	R^2	*P*
UniFrac (unweighted)	Control	Population	1.6475	0.1088	**0.0003**	1.9924	0.1286	**0.0001**
	CD		1.4702	0.1052	**0.0014**	1.7771	0.1290	**0.0002**
	UC		1.8737	0.1219	**0.0001**	2.1313	0.1364	**0.0001**
	Germany	Disease	1.2393	0.0870	**0.0217**	1.3582	0.0914	**0.0019**
	Lithuania		1.2458	0.0875	**0.0387**	1.2575	0.0949	**0.0098**
	India		1.1714	0.0798	0.0695	1.1731	0.0800	0.1197
UniFrac (weighted)	Control	Population	1.8783	0.1221	0.0585	2.8899	0.1763	**0.0011**
	CD		1.1853	0.0866	0.2822	2.1255	0.1505	**0.0257**
	UC		3.0631	0.1849	**0.0068**	2.7172	0.1676	**0.0008**
	Germany	Disease	1.0029	0.0716	0.4032	1.3119	0.0886	0.1781
	Lithuania		2.1866	0.1440	**0.0413**	3.0611	0.2033	**0.0015**
	India		0.9692	0.0670	0.4235	1.0698	0.0734	0.3584
Jaccard	Control	Population	1.5208	0.1013	**0.0001**	2.0434	0.1315	**0.0001**
	CD		1.4019	0.1008	**0.0002**	1.8786	0.1354	**0.0001**
	UC		1.4877	0.0993	**0.0001**	2.2119	0.1408	**0.0001**
	Germany	Disease	1.1828	0.0834	**0.0063**	1.3337	0.0899	**0.0001**
	Lithuania		1.1366	0.0804	**0.0294**	1.3132	0.0986	**0.0008**
	India		1.0674	0.0733	0.1632	1.4237	0.0954	**0.0022**
Bray-Curtis	Control	Population	1.9711	0.1274	**0.0001**	2.7267	0.1680	**0.0001**
	CD		1.4835	0.1061	**0.0016**	2.0374	0.1451	**0.0003**
	UC		1.8089	0.1182	**0.0001**	2.5762	0.1603	**0.0001**
	Germany	Disease	1.2405	0.0871	**0.0452**	1.5652	0.1039	**0.0008**
	Lithuania		1.2287	0.0864	0.0534	1.4127	0.1053	**0.0040**
	India		1.1118	0.0761	0.1759	1.2533	0.0850	0.1409
RDA	Control	Population	1.8443	0.1202	**0.0001**	2.6036	0.1617	**0.0001**
	CD		1.4641	0.1048	**0.0005**	2.1251	0.1504	**0.0001**
	UC		1.6726	0.1102	**0.0001**	2.5655	0.1597	**0.0001**

Germany	Disease	1.2347	0.0867	**0.0221**	1.5545	0.1033	**0.0001**
Lithuania		1.1698	0.0826	0.0625	1.3272	0.0996	**0.0049**
India		1.1157	0.0763	0.1423	1.2523	0.0849	0.1148

Table S5: DNA-based indicator consensus genera.

Association	Classification	Abundance	*r*	*P*	Adj. *P*
GER	*Papillibacter*	183	0.375	0.0006	0.0269
GER	*Parasutterella*	100	0.292	0.0024	0.0575
IND	*Acinetobacter*	142	0.292	0.0003	0.0169
IND	*Alkalibacterium*	34	0.261	0.0011	0.0379
IND	*Bacillus*	56	0.282	0.0002	0.0117
IND	*Devosia*	14	0.281	0.0034	0.0759
IND	*Fastidiosipila*	11	0.357	0.0016	0.0481
IND	*Filibacter*	119	0.3	0.0022	0.0556
IND	*Kocuria*	30	0.423	0.0000	0.0038
IND	*Lactobacillus*	1032	0.375	0.0001	0.0061
IND	*Massilia*	14	0.341	0.0025	0.0575
IND	*Megamonas*	739	0.332	0.0005	0.0223
IND	*Megasphaera*	151	0.421	0.0002	0.0122
IND	*Micrococcus*	41	0.481	0.0000	0.0038
IND	*Paracoccus*	27	0.407	0.0000	0.0038
IND	*Pseudomonas*	58	0.362	0.0011	0.0379
IND	*Sporacetigenium*	568	0.335	0.0016	0.0481
IND	*Weissella*	29	0.339	0.0009	0.0370
LIT	*Geosporobacter*	79	0.434	0.0001	0.0102
LIT	*Succinispira*	115	0.41	0.0002	0.0137
GER+IND	*Staphylococcus*	234	0.351	0.0021	0.0556
GER+LIT	*Bacteroides*	15226	0.466	0.0000	0.0038
GER+LIT	*uncl.Lachnospiraceae*	6795	0.352	0.0017	0.0481
IND (UC)	*Acinetobacter*	142	0.505	0.0000	0.0205
GER (CON)	*Campylobacter*	71	0.542	0.0002	0.0341
IND (UC)	*Coprobacillus*	967	0.563	0.0003	0.0384
IND (CD)	*Halomonas*	51	0.336	0.0007	0.0606
IND (UC)	*Ochrobactrum*	28	0.525	0.0009	0.0666
GER (CON)	*Varibaculum*	5	0.607	0.0002	0.0341
IND (CD)	*Weissella*	29	0.499	0.0007	0.0606

Table S6: RNA-based indicator consensus genera.

Association	Classification	Abundance	*r*	*P*	Adj. *P*
CON	*Papillibacter*	132	0.395	0.0002	0.0895
GER	*Adlercreutzia*	42	0.424	0.0001	0.0043
GER	*Anaerostipes*	15	0.335	0.0013	0.0436
GER	*Parasutterella*	87	0.285	0.0019	0.0555
IND	*Megasphaera*	92	0.335	0.0004	0.0159
IND	*Prevotella*	6827	0.372	0.0011	0.0408
IND	*unclassified*	7084	0.474	0.0000	0.0012
LIT	*Anaerovorax*	77	0.432	0.0000	0.0027
LIT	*Geosporobacter*	82	0.475	0.0000	0.0012
LIT	*Succinispira*	91	0.392	0.0004	0.0159
GER+LIT	*Bacteroides*	10160	0.402	0.0003	0.0159
GER+LIT	*Dorea*	2076	0.42	0.0001	0.0048
GER+LIT	*uncl.Lachnospiraceae*	11507	0.483	0.0000	0.0012
GER (CON)	*Adlercreutzia*	42	0.622	0.0000	0.0143
GER (CON)	*Brevibacterium*	10	0.475	0.0057	0.0931
GER (CON)	*Brevundimonas*	19	0.452	0.0053	0.0931
GER (CON)	*Campylobacter*	66	0.511	0.0019	0.0741
GER (CON)	*Papillibacter*	132	0.449	0.0056	0.0931
GER (CON)	*uncl.Incertae Sedis XI*	13	0.511	0.0019	0.0741
GER (CON)	*Varibaculum*	9	0.554	0.0004	0.0394
LIT (CON)	*Anaerovorax*	77	0.471	0.0032	0.0790
LIT (CON)	*Cohnella*	40	0.443	0.0038	0.0790
LIT (CON)	*Desulfovibrio*	62	0.497	0.0035	0.0790
LIT (CON)	*Peptococcus*	9	0.453	0.0039	0.0790
LIT (CON)	*Succinispira*	91	0.485	0.0026	0.0779
IND (CON)	*Megasphaera*	92	0.545	0.0008	0.0580
IND (CD)	*Afipia*	62	0.424	0.0024	0.0779
IND (UC)	*Acidovorax*	31	0.501	0.0003	0.0394
GER (CON)+LIT (CON)	*Anaerococcus*	76	0.461	0.0055	0.0931
IND (CON. CD. UC)	*unclassified*	7084	0.474	0.0038	0.0790
GER (CON)+GER (UC) +LIT (CON)+LIT (UC)	*Bacteroides*	10160	0.468	0.0013	0.0741

Table S7: DNA-based indicator species level OTUs.

Association	OTU-ID	Classification	Abundance	*r*	*P*	Adj. *P*
GER	X15998	*Bacteroides*	12	0.454	0.0002	0.0531
GER	X14993	*Bacteroides*	1113	0.438	0.0002	0.0531
GER	X14863	*uncl.Lachnospiraceae*	19	0.372	0.0001	0.0425
GER	X1435	*Papillibacter*	91	0.368	0.0002	0.0531
IND	X12146	*Prevotella*	12	0.442	0.0001	0.0425
IND	X33	*Catenibacterium*	208	0.439	0.0001	0.0425

IND	X84	*Dialister*	9	0.422	0.0001	0.0425
IND	X327	*Streptococcus*	119	0.407	0.0001	0.0425
IND	X5829	*Prevotella*	76	0.404	0.0001	0.0425
IND	X7257	*Prevotella*	94	0.374	0.0001	0.0425
IND	X2423	*Prevotella*	19	0.363	0.0001	0.0425
IND	X11036	*Streptococcus*	158	0.361	0.0001	0.0425
IND	X1261	*Coprobacillus*	23	0.344	0.0002	0.0531
IND	X10484	*Lactobacillus*	155	0.315	0.0001	0.0425
IND	X1	*Chloroplast*	34	0.299	0.0002	0.0531
IND	X397	*Lactobacillus*	129	0.284	0.0002	0.0531
LIT	X6299	*Allobaculum*	71	0.445	0.0001	0.0425
LIT	X2754	*Allobaculum*	35	0.44	0.0001	0.0425
LIT	X13027	*Coprococcus*	12	0.425	0.0002	0.0531
LIT	X14248	*Allobaculum*	17	0.409	0.0002	0.0531
LIT	X1982	*Geosporobacter/ uncl.Lachnospiraceae*	71	0.409	0.0001	0.0425
LIT	X5345	*uncl. Lachnospiraceae/ Geosporobacter*	63	0.407	0.0001	0.0425
LIT	X4242	*uncl.Lachnospiraceae*	233	0.378	0.0002	0.0531
LIT	X8356	*uncl.Lachnospiraceae/ Geosporobacter*	40	0.363	0.0001	0.0425

Table S8: RNA-based indicator species level OTUs.

Association	OTU-ID	Classification	Abundance	*r*	*P*	Adj. *P*
GER	X6866	*Blautia*	133	0.526	0.0001	0.0102
GER	X7691	*Blautia*	24	0.498	0.0001	0.0102
GER	X7051	*uncl. Lachnospiraceae*	366	0.495	0.0001	0.0102
GER	X7928	*Faecalibacterium*	21	0.482	0.0001	0.0102
GER	X8260	*uncl. Lachnospiraceae*	43	0.477	0.0001	0.0102
GER	X6580	*uncl. Lachnospiraceae*	53	0.458	0.0001	0.0102
GER	X6986	*uncl. Lachnospiraceae*	155	0.451	0.0001	0.0102
GER	X8291	*Blautia*	89	0.445	0.0001	0.0102
GER	X6798	*Faecalibacterium*	19	0.442	0.0001	0.0102
GER	X6917	*uncl. Lachnospiraceae*	122	0.442	0.0002	0.0161
GER	X7231	*uncl. Lachnospiraceae*	170	0.430	0.0001	0.0102
GER	X8458	*Blautia*	22	0.428	0.0001	0.0102
GER	X7575	*Uncl. Lachnospiraceae/ Dorea/Coprococcus*	40	0.428	0.0002	0.0161
GER	X8097	*Blautia*	34	0.427	0.0001	0.0102
GER	X7072	*Blautia*	39	0.424	0.0003	0.0221
GER	X7260	*Bacteroides*	15	0.422	0.0004	0.0269
GER	X8510	*uncl. Lachnospiraceae*	13	0.422	0.0001	0.0102
GER	X7290	*Dorea*	38	0.417	0.0001	0.0102
GER	X6812	*Faecalibacterium*	10	0.415	0.0003	0.0221

GER	X8759	*Bacteroides*	566	0.415	0.0001	0.0102
GER	X7652	*Dorea/ uncl. Lachnospiraceae*	7	0.411	0.0012	0.0499
GER	X7131	*uncl. Lachnospiraceae/ Oscillibacter*	44	0.408	0.0001	0.0102
GER	X6871	*Blautia*	116	0.406	0.0001	0.0102
GER	X6276	*uncl. Lachnospiraceae*	37	0.392	0.0004	0.0269
GER	X6295	*Faecalibacterium*	11	0.390	0.0006	0.0343
GER	X6979	*Bacteroides*	13	0.387	0.0011	0.0479
GER	X8607	*Subdoligranulum/ uncl. Lachnospiraceae*	13	0.387	0.0008	0.0402
GER	X8889	*Subdoligranulum*	9	0.386	0.0009	0.0420
GER	X7481	*uncl. Lachnospiraceae*	28	0.381	0.0001	0.0102
GER	X4870	*Faecalibacterium*	28	0.377	0.0013	0.0517
GER	X8936	*Bacteroides*	50	0.376	0.0002	0.0161
GER	X21	*Bacteroides*	481	0.372	0.0009	0.0420
GER	X7109	*Faecalibacterium*	13	0.371	0.0006	0.0343
GER	X6436	*Eggerthella*	10	0.363	0.0008	0.0402
GER	X6545	*Bacteroides*	25	0.360	0.0009	0.0420
GER	X5349	*Faecalibacterium*	20	0.354	0.0007	0.0368
GER	X7294	*uncl. Peptostreptococcaceae*	15	0.351	0.0012	0.0499
GER	X6682	*Coprococcus*	38	0.349	0.0001	0.0102
GER	X6448	*Sporacetigenium*	31	0.349	0.0001	0.0102
GER	X9412	*Papillibacter/ uncl. Ruminococcaceae*	23	0.344	0.0007	0.0368
GER	X8531	*Roseburia*	43	0.331	0.0002	0.0161
GER	X6956	*Coprococcus*	39	0.326	0.0007	0.0368
GER	X7361	*Bacteroides*	186	0.324	0.0001	0.0102
GER	X7603	*Bacteroides*	52	0.308	0.0007	0.0368
GER	X7683	*Coprobacillus*	38	0.295	0.0001	0.0102
GER	X7825	*Blautia*	29	0.287	0.0005	0.0309
GER	X7199	*uncl. Lachnospiraceae/ Roseburia*	61	0.282	0.0002	0.0161
GER	X7268	*Aquiflexum/Odoribacter/ Anaerophaga*	56	0.269	0.0014	0.0552
GER	X7060	*Coprococcus/ uncl. Lachnospiraceae*	106	0.264	0.0004	0.0269
GER	X6886	*uncl. Lachnospiraceae/ Roseburia*	76	0.237	0.0004	0.0269
GER	X5827	*Roseburia*	56	0.217	0.0007	0.0368
IND	X386	*Chloroplast*	19	0.493	0.0001	0.0102
IND	X1	*Chloroplast*	5952	0.482	0.0001	0.0102
IND	X289	*Chloroplast*	31	0.457	0.0001	0.0102
IND	X643	*Chloroplast*	33	0.437	0.0001	0.0102
IND	X280	*Chloroplast*	89	0.430	0.0001	0.0102
IND	X226	*Chloroplast*	19	0.413	0.0001	0.0102

IND	X250	*Prevotella*	66	0.406	0.0002	0.0161
IND	X624	*Chloroplast*	11	0.390	0.0011	0.0479
IND	X376	*Prevotella*	565	0.381	0.0001	0.0102
IND	X384	*Dialister*	11	0.370	0.0007	0.0368
IND	X955	*Chloroplast*	47	0.367	0.0015	0.0587
IND	X826	*Chloroplast*	44	0.360	0.0002	0.0161
IND	X563	*Prevotella*	14	0.359	0.0011	0.0479
IND	X4	*Prevotella*	331	0.328	0.0013	0.0517
IND	X381	*Prevotella*	40	0.306	0.0008	0.0402
IND	X16	*Prevotella/Paraprevotella*	114	0.295	0.0009	0.0420
IND	X327	*Streptococcus/Kocuria*	47	0.294	0.001	0.0453
IND	X397	*Lactobacillus*	195	0.277	0.0001	0.0102
IND	X173	*Acidovorax*	26	0.275	0.0013	0.0517
IND	X3	*Prevotella*	110	0.259	0.0007	0.0368
IND	X459	*Faecalibacterium*	265	0.243	0.0006	0.0343
LIT	X4420	*Dorea/Roseburia*	82	0.552	0.0001	0.0102
LIT	X2541	*Roseburia*	102	0.531	0.0001	0.0102
LIT	X4551	*Coprococcus*	29	0.488	0.0001	0.0102
LIT	X3391	*Dorea*	43	0.469	0.0001	0.0102
LIT	X2754	*Allobaculum*	56	0.456	0.0001	0.0102
LIT	X132	*Dorea*	445	0.439	0.0002	0.0161
LIT	X4352	*Coprococcus*	55	0.428	0.0001	0.0102
LIT	X1637	*uncl. Lachnospiraceae*	100	0.426	0.0001	0.0102
LIT	X110	*uncl. Lachnospiraceae/ Coprococcus/Dorea*	673	0.425	0.0001	0.0102
LIT	X181	*Dorea/Coprococcus*	56	0.422	0.0005	0.0309
LIT	X5735	*Allobaculum*	134	0.414	0.0002	0.0161
LIT	X3215	*Blautia*	79	0.413	0.0003	0.0221
LIT	X1982	*uncl. Lachnospiraceae/ Geosporobacter*	117	0.412	0.0001	0.0102
LIT	X4022	*Faecalibacterium/Coprobacillus*	9	0.408	0.0004	0.0269
LIT	X3044	*uncl. Lachnospiraceae/Dorea*	15	0.404	0.0003	0.0221
LIT	X5619	*Coprococcus*	6	0.400	0.0006	0.0343
LIT	X3266	*Allobaculum*	63	0.396	0.0001	0.0102
LIT	X3966	*Catenibacterium*	25	0.395	0.0002	0.0161
LIT	X4632	*Dorea*	68	0.390	0.0003	0.0221
LIT	X49	*Succinispira*	87	0.390	0.0005	0.0309
LIT	X3354	*Roseburia/Coprococcus*	23	0.383	0.001	0.0453
LIT	X2428	*Coprococcus*	59	0.379	0.0009	0.0420
LIT	X2253	*Coprococcus*	36	0.378	0.0001	0.0102
LIT	X3623	*Sporobacter*	9	0.378	0.0006	0.0343
LIT	X3767	*Roseburia*	9	0.377	0.0008	0.0402
LIT	X4181	*Roseburia*	9	0.377	0.0012	0.0499
LIT	X260	*Coprococcus/Blautia/ uncl. Lachnospiraceae*	20	0.376	0.0009	0.0420
LIT	X4650	*Roseburia*	48	0.374	0.0001	0.0102
LIT	X4522	*Catenibacterium*	19	0.374	0.0013	0.0517

LIT	X3683	*Blautia*	27	0.371	0.0005	0.0309
LIT	X2574	*Roseburia*	107	0.368	0.0001	0.0102
LIT	X2758	*uncl. Lachnospiraceae/ Roseburia*	11	0.357	0.001	0.0453
LIT	X2003	*uncl. Lachnospiraceae/ Dorea*	57	0.346	0.0001	0.0102
LIT	X4013	*Faecalibacterium*	68	0.346	0.0012	0.0499
LIT	X3420	*Faecalibacterium*	12	0.337	0.0009	0.0420
LIT	X228	*uncl. Lachnospiraceae*	160	0.337	0.0002	0.0161
LIT	X2928	*Blautia*	33	0.331	0.0012	0.0499
LIT	X201	*uncl. Lachnospiraceae*	262	0.322	0.0013	0.0517
LIT	X1983	*Roseburia/Coprococcus*	64	0.318	0.0004	0.0269
LIT	X2567	*Gracilibacter/Geosporobacter*	15	0.314	0.0006	0.0343
LIT	X2185	*uncl. Lachnospiraceae/ Geosporobacter*	106	0.310	0.0001	0.0102
LIT	X2112	*uncl. Lachnospiraceae/ Roseburia*	25	0.307	0.0002	0.0161
LIT	X2825	*Prevotella*	43	0.302	0.0005	0.0309
LIT	X4568	*Blautia*	24	0.301	0.0011	0.0479
LIT	X1973	*Coprococcus*	23	0.255	0.0005	0.0309
GER (CD)	X6686	*Sporacetigenium/Roseburia*	17	0.542	0.0001	0.0416
GER (CON)	X9144	*Dorea*	6	0.756	0.0001	0.0416
GER (CON)	X8058	*Blautia*	20	0.660	0.0001	0.0416
GER (CON)	X9412	*Papillibacter/ uncl.Ruminococcaceae*	23	0.655	0.0002	0.0572
GER (CON)	X8260	*uncl.Lachnospiraceae*	43	0.655	0.0001	0.0416
GER (CON)	X9242	*uncl.Lachnospiraceae/ Dorea/Coprococcus*	8	0.610	0.0002	0.0572
GER (CON)	X8291	*Blautia*	89	0.609	0.0002	0.0572
GER (CON)	X9031	*Coprococcus*	13	0.477	0.0001	0.0416
GER (CON)+GER (UC)	X7691	*Blautia*	24	0.588	0.0001	0.0416
GER (CON)+ GER (UC)	X7928	*Faecalibacterium*	21	0.560	0.0002	0.0572
GER (UC)	X6986	*uncl.Lachnospiraceae*	155	0.607	0.0001	0.0416
IND (CD)	X250	*Prevotella*	66	0.616	0.0001	0.0416
IND (CD)	X3	*Prevotella*	110	0.514	0.0001	0.0416
IND (UC)	X173	*Acidovorax*	26	0.533	0.0001	0.0416
LIT (CON)+LIT (UC)	X4420	*Dorea/Roseburia*	82	0.640	0.0001	0.0416
LIT (CON)+LIT (UC)	X4551	*Coprococcus*	29	0.569	0.0002	0.0572

Table S9: Activity of indicator consensus genera for disease condition and host population (rRNA/rDNA).

Disease	Genus (RDP classified)	Mean activity (CON)	Mean activity (CD)	Mean activity (UC)	χ^2	*P*	Adj. *P*
CON	*Anaerococcus*	0.900	0.000	0.123	16.170	0.000308	0.014
UC	*Bacteroides*	0.585	0.349	2.033	16.664	0.000241	0.014
CON+UC	*Faecalibacterium*	2.618	0.606	1.286	18.487	0.000097	0.014
CON	*Papillibacter*	1.513	0.203	0.179	16.260	0.000295	0.014
Population	**Genus (RDP classified)**	**Mean activity (Germany)**	**Mean activity (Lithuania)**	**Mean activity (India)**	χ^2	*P*	Adj. *P*
GER+LIT	*Aquiflexum*	0.500	0.773	0.000	25.429	0.000003	0.001
GER+LIT	*Collinsella*	2.930	0.609	0.018	17.969	0.000125	0.005
GER	*Coprobacillus*	3.599	0.751	0.260	20.652	0.000033	0.002
GER+LIT	*Dorea*	2.449	1.654	0.591	21.509	0.000021	0.001
LIT	*Geosporobacter*	0.253	1.037	0.033	24.261	0.000005	0.001
GER+IND	*Klebsiella*	0.716	0.000	1.117	12.861	0.001612	0.037
IND	*Lactobacillus*	0.230	0.102	0.328	17.581	0.000152	0.005
LIT	*Succinispira*	0.000	0.323	0.236	13.552	0.001141	0.030
IND	*unclassified*	0.807	1.347	92.160	12.383	0.002046	0.038
GER+LIT	*uncl.Lachnospiraceae*	2.693	1.782	1.167	12.547	0.001886	0.038
GER+LIT	*uncl.Peptostreptococcaceae*	2.328	1.265	0.387	11.808	0.002728	0.046

Table S10: Activity of indicator species level OTUs (rRNA/rDNA).

Association	OTU-ID	Genus (RDP classified)	Mean activity (Germany)	Mean activity (Lithuania)	Mean activity (India)	χ^2	*P*	Adj. *P*
GER	X1435	*uncl.Ruminococcaceae/Papillibacter*	0.086	0.000	0.000	14.769	0.0006209	0.033

GER	X629	*Staphylococcus*	0.310	0.000	0.000	17.088	0.0001947	0.017
GER	X6545	*Bacteroides*	0.248	0.000	0.000	14.769	0.0006207	0.033
GER	X6580	*uncl.Lachnospiraceae*	0.616	0.074	0.000	15.811	0.0003687	0.024
GER	X6791	*Dorea*	0.899	0.193	0.000	16.350	0.0002816	0.023
GER	X6798	*Faecalibacterium*	0.267	0.000	0.000	14.771	0.0006201	0.033
GER	X6866	*Blautia*	0.793	0.142	0.000	24.931	0.0000039	0.002
GER	X6871	*Blautia*	0.713	0.081	0.000	18.093	0.0001178	0.013
GER	X7051	*uncl.Lachnospiraceae*	1.712	0.066	0.000	37.346	0.0000000	0.000
GER	X7060	*Coprococcus*	1.230	0.000	0.000	14.769	0.0006207	0.033
GER	X7231	*uncl.Lachnospiraceae*	1.297	0.037	0.000	21.173	0.0000252	0.005
GER	X7290	*Dorea*	0.635	0.062	0.000	13.954	0.0009331	0.044
GER	X7361	*Bacteroides*	0.647	0.031	0.000	19.041	0.0000734	0.010
GER	X8097	*Blautia*	0.437	0.000	0.000	19.453	0.0000597	0.008
GER	X8291	*Blautia*	0.931	0.000	0.000	19.452	0.0000597	0.008
GER	X8531	*Roseburia*	0.515	0.000	0.000	14.770	0.0006205	0.033
GER	X8759	*Bacteroides*	1.104	0.000	0.000	32.148	0.0000001	0.000
GER+LIT	X1167	*Blautia*	0.603	0.975	0.078	14.282	0.0007918	0.039
GER+LIT	X5960	*Coprococcus*	0.506	0.454	0.000	14.204	0.0008234	0.040
IND	X1	*Streptophyta*	0.000	0.000	33.122	18.472	0.0000975	0.011
IND	X33	*Catenibacterium*	0.000	0.000	0.266	16.223	0.0003000	0.023
IND	X397	*Lactobacillus*	0.000	0.000	0.869	18.472	0.0000975	0.011
LIT	X110	*uncl.Lachnospiraceae*	0.739	2.757	0.353	17.970	0.0001253	0.013
LIT	X1196	*Roseburia*	0.051	1.153	0.367	21.606	0.0000203	0.005
LIT	X1982	*uncl.Lachnospiraceae*	0.038	0.918	0.000	21.121	0.0000259	0.005
LIT	X201	*uncl.Lachnospiraceae*	0.106	1.579	0.383	15.183	0.0005047	0.031
LIT	X2013	*uncl.Bacteroidetes*	0.668	1.492	0.000	16.033	0.0003299	0.023
LIT	X203	*unclassified_Lachnospiraceae*	0.364	1.014	0.033	22.478	0.0000132	0.004
LIT	X2245	*Faecalibacterium*	0.000	0.579	0.433	14.455	0.0007265	0.037
LIT	X2428	*Coprococcus*	0.017	0.460	0.000	15.891	0.0003543	0.024

LIT	X2541	*Roseburia*	0.100	1.127	0.000	27.341	0.0000012	0.001
LIT	X2574	*Roseburia*	0.207	1.270	0.000	20.156	0.0000420	0.007
LIT	X2754	*uncl.Firmicutes*	0.069	0.426	0.000	17.520	0.0001569	0.014
LIT	X30	*unclassified_Lachnospiraceae*	0.525	1.676	0.351	16.231	0.0002989	0.023
LIT	X4242	*unclassified_Lachnospiraceae*	0.498	0.789	0.000	22.275	0.0000146	0.004
LIT	X4352	*Coprococcus*	0.069	0.596	0.000	17.653	0.0001468	0.014
LIT	X4632	*Dorea*	0.138	0.889	0.000	15.570	0.0004159	0.026
LIT	X4650	*Roseburia*	0.026	0.573	0.000	16.020	0.0003322	0.023
LIT	X5735	*uncl.Firmicutes*	0.224	0.374	0.000	22.448	0.0000133	0.004
LIT	X87	*Roseburia*	0.497	1.764	0.877	13.759	0.0010287	0.047

Supplementary References:

1 Rehman A, Lepage P, Nolte A, Hellmig S, Schreiber S, Ott SJ. Transcriptional activity of the dominant gut mucosal microbiota in chronic inflammatory bowel disease patients. Journal of Medical Microbiology 2010;**59**:1114-22.

2 Rausch P, Rehman A, Künzel S, Häsler R, Ott SJ, Schreiber S*, et al.* Colonic mucosa-associated microbiota is influenced by an interaction of Crohn disease and FUT2 (Secretor) genotype. Proceedings of the National Academy of Sciences 2011;**108**:19030-5.

3 Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB*, et al.* Introducing mothur: Open Source, Platform-independent, Community-supported Software for Describing and Comparing Microbial Communities. Appl Environ Microbiol 2009;**75**:7537-41.

4 Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J*, et al.* SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucl Acids Res 2007;**35**:7188-96.

5 Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 2010.

6 Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. Appl Environ Microbiol 2007;**73**:5261-7.

7 DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K*, et al.* Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. Appl Environ Microbiol 2006;**72**:5069-72.

8 Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics 2010;**26**:266-7.

9 Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLoS One 2010;**5**:e9490.

Chapter II: The influence of *FUT2* in shaping the human microbial community in health and disease

In the post antibiotics era of medicine, the primary contributor to human mortality and morbidity in the western world is no longer infectious diseases, but complex, usually late-onset chronic diseases have taken the lead [346]. These modern-day maladies have been extensively studied using population-wide genetic and environmental screens and only relatively small effect sizes were ascertained so far for single genetic [344, 362, 363, 383] or environmental traits [346, 384]. Thus, the leading hypotheses to explain the development and spread of those diseases propose multiple disturbances on a genetically susceptible host background that trigger the onset of symptoms. The complex microbial communities inhabiting almost every part of the human body have recently been recognized as potential mediators for environmental influences, genetic susceptibilities and their interaction [385]. The microbiome was subsequently associated to a number of maladies including inflammatory diseases [345, 386], complex metabolic diseases [130, 387], and even cancer development [388, 389].

The human *α-1,2-fucosyltransferase* gene (*FUT2*) codes for an enzyme transferring a fucose residue to short ABO blood group precursor molecules in excretory tissues and shows widespread functional polymorphisms in humans. These loss-of-function and activity reducing mutations appear to be maintained for a long time and evolved independently in geographically isolated populations [220, 248, 263]. These patterns signify long-term balancing selection, which conserved the functional and genetic variation of *FUT2* among human populations. Being a nonsecretor, *i.e*. homozygous for loss-of-function mutations in the *FUT2* gene and therefore unable to excrete ABO blood group antigens, has been associated with lower susceptibility to a number of pathogens such as *Helicobacter pylori, Norovirus,* or *Campylobacter jejuni* [230, 275, 276]*,* but also to marked increases in the risk for necrotizing colitis [208], decreased vitamin titers [390], and inflammatory bowel disease [209]. These benefits of being a nonsecretor suggest a pathogen driven selection on those ABO subverting mutations, which come at the cost of other susceptibilities and a higher risk of long lasting chronic inflammatory diseases.

Here, we focus on the influence of *FUT2* secretor status and genotype on the microbial community in two inflammatory diseases with a strong overlap in coincidence and genetic predispositions. Crohn's Disease a main form of inflammatory bowel disease and primary sclerosing cholangitis, an inflammation and fibrosis of the bile duct, have several common genetic associations [391], including the G428→A mutation in *FUT2*, the main nonsecretor generating mutation in Caucasians [239]. It is a valid assumption that there

exists a gut-liver immune axis, as another established risk gene for inflammatory bowel disease (IL10 [392]) gives strong indication for such a link and indicates a prominent role for the microbial community facilitating inflammation [393]. The liver is also the main location of microbial acetate (short chain fatty acid) recycling to build lipids [315] essential to the host, such as butyrate [311-313], which is highly dependent on nutrient access of the community, while obesity associated microbial community changes can lead to hepatic cancer [388]. The dependence of bacteria on intestinal glycans, not only under limitation of other carbohydrate sources [302, 394], is well established [395]. Recent studies showing the direct initiation of host glycosylation programs by the intestinal bacteria [303, 305-308] further underline the ability of microbes to establish their own environment/niche and to persist in it [323, 394], but also their direct dependence on glycan provision by the host itself. Thus, by changing the carbohydrate spectrum in the mucosa, specifically the depletion of fucosylated glycans, can attract different bacterial species, alter community. These community wide changes may thus increase community instabilities and set the stage for dysbioses and connected health risks through a more inflammatory active or invadable bacterial assemblage in the gastrointestinal tract [22, 396].

We found significant differences in community composition between secretor status and even *FUT2* genotype in healthy and diseased mucosa associated bacteria in the colon and bile duct. Interestingly, the community complexity shows comparable patterns among genotypes, health conditions, and disease types. Furthermore, we identified single bacterial species and genera with potential pathogenic capabilities in nonsecretor individuals. This indicates a potential for instable communities and lower colonization resistance and thus a higher risk for the development of inflammation in nonsecretors.

Publications:

Rausch, P., A. Rehman, S. Künzel, R. Häsler, S. J. Ott, S. Schreiber, P. Rosenstiel, A. Franke and J. F. Baines (2011). "Colonic mucosa-associated microbiota is influenced by an interaction of Crohn disease and FUT2 (Secretor) genotype." Proceedings of the National Academy of Sciences 108(47): 19030-19035.

Folseraas, T., E. Melum, **P. Rausch**, B. D. Juran, E. Ellinghaus, A. Shiryaev, J. K. Laerdahl, D. Ellinghaus, C. Schramm, T. J. Weismüller, D. N. Gotthardt, J. R. Hov, O. P. Clausen, R. K. Weersma, M. Janse, K. M. Boberg, E. Björnsson, H.-U. Marschall, I. Cleynen, P. Rosenstiel, K. Holm, A. Teufel, C. Rust, C. Gieger, H. E. Wichmann, A. Bergquist, E. Ryu, C. Y. Ponsioen, H. Runz, M. Sterneck, S. Vermeire, U. Beuers, C. Wijmenga, E. Schrumpf, M. P. Manns, K. N. Lazaridis, S. Schreiber, J. F. Baines, A. Franke and T. H. Karlsen (2012). "Extended analysis of a genome-wide association study in primary sclerosing cholangitis detects multiple novel risk loci." Journal of Hepatology 57(0): 366-375.

Colonic mucosa-associated microbiota is influenced by an interaction of Crohn disease and *FUT2* (*Secretor*) genotype

Philipp Rausch[a,b], Ateequr Rehman[c], Sven Künzel[b], Robert Häsler[c], Stephan J. Ott[d], Stefan Schreiber[c,d], Philip Rosenstiel[c], Andre Franke[c], and John F. Baines[a,b,1]

[a]Institute for Experimental Medicine, Christian-Albrechts-University of Kiel, D-24105 Kiel, Germany; [b]Max Planck Institute for Evolutionary Biology, D-24306 Plön, Germany; [c]Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, D-24105 Kiel, Germany; and [d]Department of General Internal Medicine, Christian-Albrechts-University of Kiel, University Hospital Schleswig-Holstein, Campus Kiel, D-24105 Kiel, Germany

Edited by Norman R. Pace, University of Colorado, Boulder, CO, and approved October 11, 2011 (received for review April 29, 2011)

The *FUT2* (*Secretor*) gene is responsible for the presence of ABO histo-blood group antigens on the gastrointestinal mucosa and in bodily secretions. Individuals lacking a functional copy of *FUT2* are known as "nonsecretors" and display an array of differences in susceptibility to infection and disease, including Crohn disease. To determine whether variation in resident microbial communities with respect to *FUT2* genotype is a potential factor contributing to susceptibility, we performed 454-based community profiling of the intestinal microbiota in a panel of healthy subjects and Crohn disease patients and determined their genotype for the primary nonsecretor allele in Caucasian populations, *W143X* (G428A). Consistent with previous studies, we observe significant deviations in the microbial communities of individuals with Crohn disease. Furthermore, the *FUT2* genotype explains substantial differences in community composition, diversity, and structure, and we identified several bacterial species displaying disease-by-genotype associations. These findings indicate that alterations in resident microbial communities may in part explain the variety of host susceptibilities surrounding nonsecretor status and that *FUT2* is an important genetic factor influencing host–microbial diversity.

Crohn disease (CD) is one of the two major forms of inflammatory bowel disease (IBD) characterized by chronic relapsing/remitting inflammation of the gastrointestinal (GI) tract (1, 2). It is posited that an abnormal immune response to microbes inhabiting the GI tract contributes to CD etiopathogenesis, which may arise from perturbed gene-by-environment interactions and include host genetic and immune factors, environmental triggers, and GI microbes. The importance of the GI microbiota is emphasized by numerous observations of alterations or imbalances of the GI microbiota in IBD patients (3–7) as well as by examples of animal models failing to manifest disease after germ-free rederivation (8). Although considerable advances in identifying genetic susceptibility loci in CD have been made, with more than 71 loci known to date (9), still very little is known regarding the details of interactions between individual susceptibility variants, environmental factors, and the GI microbiota (10, 11).

The *FUT2* (*Secretor*) gene encodes an α-1,2-fucosyltransferase responsible for the expression of ABO histo-blood group antigens on the GI mucosa and in bodily secretions. Individuals bearing at least one functional allele are known as "secretors," whereas those homozygous for loss-of-function mutations display a "nonsecretor" phenotype. It was recently shown that nonsecretor status is associated with CD susceptibility (12), and the wealth of information on the population genetics of *FUT2* and its role in numerous host–microbe interactions make it an ideal candidate for describing the possible interactions between a genetic susceptibility variant and the endogenous microbiota.

Several mutations leading to the nonsecretor phenotype exist in human populations and display evidence of being maintained by strong selective pressure (13–16). A large body of evidence suggests that this maintenance may be because of numerous tradeoffs surrounding host–microbe interactions. For example, nonsecretors are resistant to infection with the Norwalk (Noro) (17) and respiratory viruses (18) but are more susceptible to duodenal ulcers (19), rheumatic fever (20), and cholera (21). Furthermore, the breast milk of secreting mothers provides protection against *Campylobacter jejuni* to their offspring by exploiting the binding affinity of the bacterium to fucosyloligosaccharides (22).

Despite variability at *FUT2* mediating susceptibility to numerous pathogens, its overall conservation in mammals (23) indicates an important functional role. Experiments in mice have shown that the resident microbiota induce expression of *FUT2* in the GI tract upon weaning, suggesting a role in the maintenance of homeostasis between mammalian hosts and their microbiota (24, 25). However, to date, no study of the effect of *FUT2* on the overall composition and structure of the adult intestinal microbiota has been published. In this study, we have performed a survey of the colonic mucosa-attached microbiota in a panel of CD patients and controls and analyzed microbial community composition and structure with respect to disease and *FUT2* genotype. As previously reported, we observe significant deviations in the mucosal communities of CD individuals. In addition, we demonstrate significant disease-by-genotype influences with respect to microbial community composition, diversity, and structure.

Results

To determine the influence of *FUT2* expression on the colonic mucosa-associated microbial communities of healthy and CD individuals, we genotyped 47 individuals (29 CD, 18 controls) for the primary nonsecretor allele in Caucasian populations, *W143X* (G428A; rs601338) (13, 15), and generated sequence libraries of the bacterial 16S rRNA gene by using a multiplex barcoded pyrosequencing approach. Individuals homozygous for the functional allele "G" are denoted as *SeSe* (8 CD, 7 controls), those homozygous for the loss-of-function allele "A" are given as *sese* (6 CD, 3 controls), and heterozygotes are represented by *Sese* (15 CD, 8 controls) (Table S1). The allele most often associated with the nonsecretor phenotype in Asian populations (A385T; rs1047781) is also present in European populations at low frequency (0.4%) (15), but genotyping revealed all individuals to be homozygous for the functional variant at this position.

Author contributions: A.R., S.S., and J.F.B. designed research; P. Rausch, A.R., S.K., R.H., and A.F. performed research; P. Rausch, A.R., S.J.O., P. Rosenstiel, and J.F.B. analyzed data; and P. Rausch, P. Rosenstiel, A.F., and J.F.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequences reported in this paper have been deposited in the European Nucleotide Archive, www.ebi.ac.uk/ena/data/view/ERP000888 (accession no. ERP000888).

[1]To whom correspondence should be addressed. E-mail: j.baines@iem.uni-kiel.de.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1106408108/-/DCSupplemental.

A total of 46,990 16S rRNA gene sequence reads spanning variable regions 1 and 2 (V1 and V2) were analyzed after quality filtering and normalization of read number to 1,000 randomly chosen sequences per individual, on which all subsequent analyses are based (*Materials and Methods*).

To provide an initial overview of the sequence reads associated with *FUT2* genotype and disease status, we classified sequences at the phyla level by using Ribosomal Database Project (RDP) classifier (26, 27). A single *Sese* control individual displayed sequences belonging almost entirely (91%) to the phylum Spirochaetes, whereas this phylum was absent from all other individuals (Fig. S1). This individual was most likely displaying asymptomatic human intestinal spirochaetosis at the time of sampling and was removed because of its unclear clinical significance (28) and severe effect on bacterial community structure. All other individuals displayed communities dominated by the Firmicutes, Bacteroidetes, and Proteobacteria along with other less abundant phyla (Fig. 1). Overall, a significant increase of Firmicutes (ANOVA, $P = 0.034$) and corresponding decrease in Proteobacteria (ANOVA, $P = 0.02$) and Actinobacteria (ANOVA, $P = 0.002$) is apparent among CD individuals compared with controls (*Statistical Analysis* in *Materials and Methods* and Table S2). Differences were also observed with respect to secretor status, with a significant increase in Bacteroidetes among nonsecretors (ANOVA, $P = 0.036$). The abundance of Proteobacteria differed with respect to not only disease status but also genotype, with a significant increase in *Sese* compared with *sese* individuals, regardless of disease status (ANOVA, $P = 0.047$). Interestingly, several comparisons between control and CD individuals at the phyla level revealed differences only among secretors. Proteobacteria and Actinobacteria were significantly lower in abundance among CD individuals (post hoc Mann–Whitney test, $P = 0.012$ and 0.006, respectively), whereas Firmicutes were significantly higher in abundance (post hoc Mann–Whitney test, $P = 0.005$). These differences, however, were only observed among secretors, and no significant difference was present among nonsecretors (Table S2), indicating that, at the level of the relative abundance of the major phyla, control nonsecretors display more similarity to their diseased counterparts than do secretors.

Bacterial Diversity Within and Between Individuals. To evaluate aspects of bacterial diversity that may be influenced by genotype and disease status, we first applied measures of alpha diversity, which describe species composition in one specific habitat of interest and can be informative of community functioning (29). Because microbial communities are highly diverse and are often poorly amenable to the diversity measures commonly used in community ecology (30), we used several different measures focusing on different aspects of community assembly, including species richness, evenness, and abundance based on operational taxonomic units (OTUs) at a 97% sequence similarity (species level) threshold, in addition to phylogenetic distance (Fig. S2 *A–D*). Phylogenetic diversity is a measure of alpha diversity that takes phylogenetic divergence into account (31). Compared with controls, a significant reduction in phylogenetic diversity was apparent among CD individuals (ANOVA, $P = 0.0009$; *Statistical Analysis* in *Materials and Methods*) (Fig. S2*A*). However, the same model also revealed a significant influence of *FUT2* genotype (ANOVA, $P = 0.024$). The Chao1 index of estimated species richness displayed no difference with respect to disease status, but marginal differences were apparent between genotypes (ANOVA, $P = 0.076$, Fig. S2*B*). Interestingly, heterozygotes (*Sese*) displayed a general reduction in diversity compared with both homozygous genotypes, although only comparisons with *sese* genotypes were significant (post hoc Mann–Whitney test, phylogenetic diversity, $P = 0.046$; Chao1, $P = 0.05$). The modified Shannon entropy and evenness by Jost (32) displayed a similar pattern, but no differences were significant (Fig. S2 *C* and *D*). These results indicate that a reduction in bacterial diversity within CD individuals compared with controls was observable when taking the phylogenetic relationship of the present species into account, whereas measures based solely on observed species number and abundance failed to reveal differences with respect to disease status. Thus, a similar number of species is present in CD individuals compared with controls, but they are on average more closely related. Because of the significant influence of *FUT2* genotype, we focused the remainder of the analysis on genotype rather than on secretor status.

Although alpha diversity measures describe aspects of community structure within a given individual, they do not reveal the similarities or differences in communities between individuals (i.e., beta diversity). Thus, to reveal the relationships of the microbial communities hosted by individuals differing by *FUT2* genotype and disease status, we performed several analyses describing the differences in bacterial community composition and structure between individuals, also taking age and sex as potential confounding factors into account. UniFrac is a phylogenetic-based beta diversity measure that represents the genetic distance between communities by comparing the shared versus unique branch lengths underlying different communities (33, 34). To analyze the distance between the communities, we used multivariate analysis of variance [analysis of dissimilarity or "*adonis*" (*Materials and Methods*) (35)] of the unweighted UniFrac metric, which ignores taxon abundance and thus reduces it to binary presence/absence data. This analysis revealed a highly significant distinction between the microbial communities of control and CD subjects (*adonis*: $R^2 = 0.053$, $P < 0.0001$). For further analysis, we analyzed UniFrac distances via principal coordinate analysis (PCoA) (Fig. S3 *A–C*). Disease status and the interaction between disease status and genotype were strongly correlated with the ordination of the unweighted UniFrac values on all three axes (goodness of fit for disease status $R^2 = 0.186$, $P < 0.0001$; genotype–disease interaction $R^2 = 0.266$, $P = 0.0005$), indicating a significant contribution of these factors to variation in community composition between individuals. This finding indicates that a significant amount of the variation in bacterial community composition between individuals can be explained not only by disease status but also by disease-by-genotype interactions. We also observed a significant correlation of age (goodness of fit $R^2 = 0.257$, $P = 0.006$). Additional analysis of the axis scores with a linear model framework revealed a strong correlation of axes one and two with disease status (PCo1: $P = 0.013$; PCo2: $P < 0.0001$) and a gradient among the genotypes on axis three ($P = 0.053$) (Table S3).

In addition to phylogenetic-based measures of beta diversity, we used OTUs (97% threshold) to analyze the similarity in species distributions between communities. To measure community similarity, we used the Jaccard and Bray-Curtis indices, which are classical ecological beta diversity measures based on the ratio of shared and unique species relative to the total number of

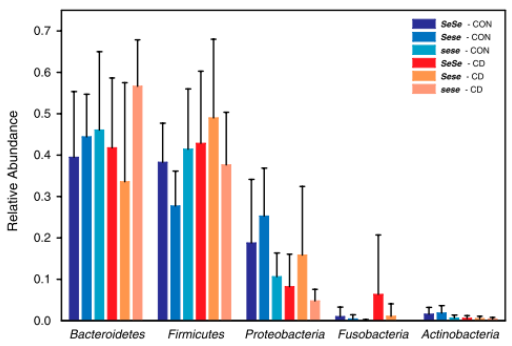


Fig. 1. Distribution of the major phyla with respect to disease status and genotype (error bars indicate SD). The corresponding results of the statistical analyses are presented in Table S2. CON, control.

species present between two communities, respectively. This analysis revealed a significant impact of disease status (*adonis*: Bray-Curtis, $R^2 = 0.039$, $P = 0.004$; Jaccard, $R^2 = 0.032$, $P = 0.004$). Furthermore, a significantly higher proportion of variance in community structure could be explained by adding the interaction between disease and genotype to this model (*adonis*: Bray-Curtis, disease status $R^2 = 0.039$, $P = 0.003$; genotype–disease interaction $R^2 = 0.103$, $P = 0.037$; Jaccard, disease status $R^2 = 0.032$, $P = 0.003$; genotype–disease interaction $R^2 = 0.097$, $P = 0.04$). Differences in bacterial community structure are also apparent in PCoA analyses of these beta diversity indices that reveal strong differences with respect to disease status, sex, and genotype (Fig. S3 *D–I* and Table S3). To compare the ordinations of beta diversity measures based on OTUs and those that consider phylogeny, we conducted a Procrustes analysis (36) using the Procrustean randomization test (PROTEST). Jaccard and Bray-Curtis displayed a strong overlap with each other (PROTEST: $M = 0.345, r = 0.809, P < 0.001$) as well as to unweighted UniFrac (PROTEST: Jaccard, $M = 0.649$, $r = 0.592$, $P < 0.001$; Bray-Curtis, $M = 0.390, r = 0.781, P < 0.001$). Thus, despite measuring different aspects of community structure, metrics based on shared abundance or phylogenetic distance appear to converge to related solutions.

To validate the factors responsible for community clustering and gain deeper insight into the actual determinants explaining this community clustering, it is advisable to use ordination methods such as canonical correspondence analysis (CCA) because these methods use the full range of OTU data and not merely the relationships among the individuals, allowing specific hypothesis testing. CCA models assume a unimodal distribution of species around an optimum (niche) to calculate the relationships among individuals via an iterative process. To avoid a potential confounding influence of age and sex, we performed partial CCA with these variables as conditioning factors. With this approach, we identified significant clusters of individuals by using artificial gradients based on disease status and the interaction with genotype (Fig. 2). Model optimization for CCA was carried out for maximizing the explained variance using a minimum of variables as well as a strong contribution of these variables to the model (total inertia = 17.739, constrained inertia = 2.195, conditioned inertia = 0.828, unconstrained inertia = 14.716 (82.96%), and explained variance by constraints and conditions = 17.04%). By this procedure, we obtained two highly significant axes explaining 5.99% of the variance in the data. Disease status ($F_{1,38} = 1.202$, $\chi^2 = 0.466$, $P = 0.008$), genotype ($F_{2,38} = 1.120$, $\chi^2 = 0.867$, $P = 0.023$), and genotype–disease status interaction ($F_{2,38} = 1.113$, $\chi^2 = 0.862$, $P = 0.06$) contributed to the fit of the ordination as assessed via permutation tests (10^5 permutations). An additional related unconstrained method [detrended correspondence analysis (DCA)] also revealed a strong influence of genotype and disease status (Table S3 and Fig. S3 *J–L*), thus verifying that disease status and *FUT2* genotype significantly contribute to the differences in community structure observed between individuals. Although these analyses describe factors that significantly contribute to differences in community structure, they explain a relatively small proportion of the total interindividual variability. Other factors such as diet, lifestyle, additional genetic factors, and differences in disease manifestation likely contribute to the remaining unexplained variation in intestinal communities. However, additional analysis of different disease subphenotypes (i.e., colitis, ileitis, and ileocolitis) did not reveal any significant patterns in Dataset S1.

Indicator Species. To identify individual bacterial taxa that contribute to the patterns identified by multivariate analysis, we analyzed OTUs at both the species (≥97% identity) and genus (≥95% identity) levels by using a common ecological measure of species habitat association, i.e., "indicator species" analysis (37). We defined our "habitats" by disease status, genotype, and genotype within the disease status. Using this analysis, we identified numerous species-level OTUs distributed across the three major

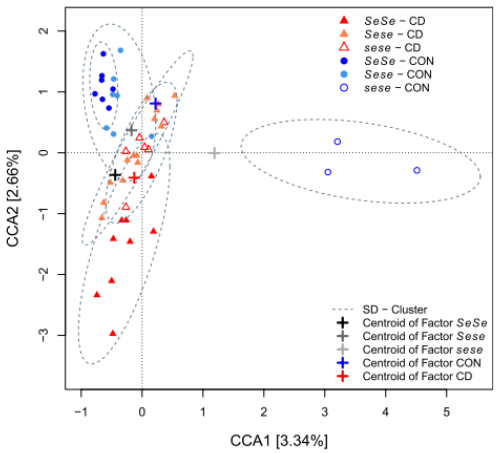


Fig. 2. Ordination of individuals using CCA of disease status and its interaction with genotype, conditioned by the factors age and sex. The ordination includes only nonredundant factors that were validated by variance inflation scores ranging between 1.1 and 3.6. Only axes that significantly contribute to the explained variance (expl. var.) are displayed (CCA1: expl. var. = 3.34%, $P = 0.007$; CCA2: expl. var. = 2.66%, $P = 0.03$; CCA3: expl. var. = 2.44%, $P = 0.171$; CCA4: expl. var. = 2.19%, $P = 0.476$; CCA5: expl. var. = 1.76%, $P = 0.944$).

phyla that were more frequently present and/or abundant in a given habitat(s) (Table S4). The analysis with respect to disease status identified 10 species-level OTUs restricted to healthy individuals belonging to *Prevotella*, *Lactobacillus*, *Coprobacillus*, *Clostridium*, *Faecalibacterium*, and *Stenotrophomonas*. The analysis with respect to genotype revealed only two OTUs belonging to *Coprococcus* and unclassified Lachnospiraceae. However, extending the analysis to genotypes within disease status revealed numerous interesting associations (Fig. 3). An OTU belonging to *Lactobacillus* and two OTUs belonging to *Stenotrophomonas* were associated with healthy secretor genotypes (*SeSe* and *Sese*). Likewise, five OTUs were identified among healthy *sese* individuals and belonged to *Prevotella*, *Brevundimonas*, unclassified Lachnospiraceae, *Sutterella*, and *Faecalibacterium*. Interestingly, three OTUs belonging to *Alistipes*, unclassified Lachnospiraceae, and *Coprococcus* were found to be associated with *sese* individuals with CD, suggesting that they may contribute to a bacterial community structure that is a subphenotype among CD patients and depends on *FUT2* genotype. Performing this analysis at the genus level identified a single, but strong, association of an unclassified genus belonging to the family Lachnospiraceae that is more abundant among CD individuals and an additional 15 genera associated with healthy individuals (Fig. S4 and Table S4).

Discussion

In this study, we sought to determine the influence of the nonsense SNP *W143X* at *FUT2*, a major susceptibility variant for CD, on the composition and structure of the intestinal mucosa-associated microbiota in both healthy and CD individuals. By using high-throughput sequencing of bacterial 16S rRNA gene sequences at a single location along the GI tract, we revealed several unique aspects of bacterial communities with respect to both disease status and *W143X* genotype. Our analysis of bacterial diversity both within and between individuals revealed significant influences of disease and disease-by-genotype interactions. Furthermore, through the application of an ecological measure of species habitat association to disease and genotype habitats, we identified important candidate bacterial species and genera contributing to these patterns.

The distribution of reads according to bacterial phyla was concordant with previously published analyses of the intestinal

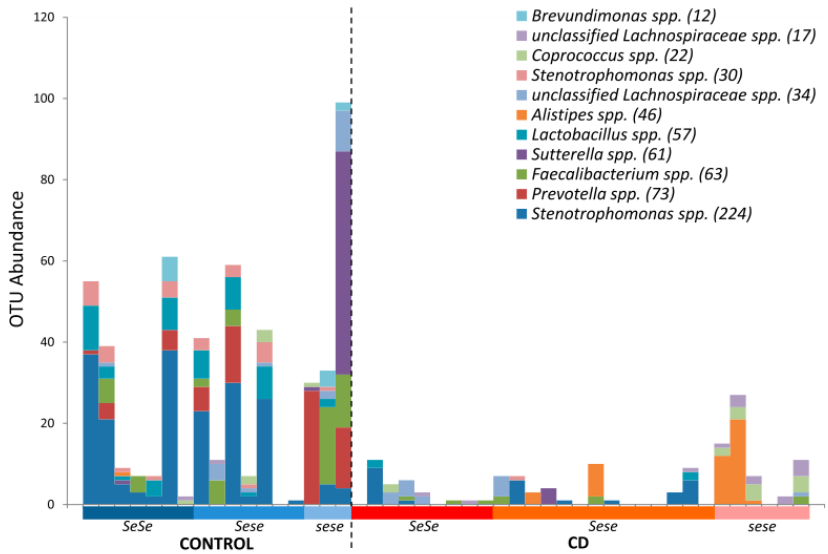


Fig. 3. Abundance of species-level OTUs identified by indicator species analysis with respect to genotype within disease status. Numbers in parentheses indicate the total read number in the normalized dataset (1,000 reads per individual).

microbiota (38) but deviated somewhat from previous investigations of CD patients (e.g., a decrease in the relative abundance of Proteobacteria was found as opposed to an increase) (5, 39). This difference may be attributable to the fact that samples in these studies were taken from acutely inflamed tissue or from multiple locations throughout the gut. Cloning bias and/or overlooked diversity because of different sequencing technologies also could contribute to this discrepancy (40) as well as account for the general lack of decreased alpha diversity in CD patients observed in our study. Another explanation for the latter could be a loss of specific taxa and corresponding replacement by other members of the bowel community. This scenario may maintain the number of observed species but result in a different phylogenetic relationship among them, thus impacting phylogenetic- rather than OTU-based diversity indices, which was observed in Dataset S1. Our observation of a significant increase in prevalence and abundance of a single bacterial genus in the context of CD compared with 15 such genera in healthy individuals (Fig. S4 and Table S4) also supports this explanation.

The contribution of *W143X* genotype to variation in bacterial communities between individuals is, in some respect, overshadowed by the effects of CD. In CD, the architecture of the intestinal mucosal lining undergoes a drastic remodeling process. Apart from the increased influx of inflammatory cells, even in remission phases without obvious inflammation, studies have shown an increased intestinal permeability in CD patients and defects in the amount and structure of the dense carbohydrate-rich layer of mucus covering the intestinal epithelium (41, 42). However, combining the effects of CD and the underlying factor of *W143X* genotype explained significantly more variation between microbial communities, supporting the hypothesis of an interaction of these three elements. Interestingly, some differences in community composition and structure, i.e., alpha diversity and the relative abundance of the major phyla, were less apparent between control and CD nonsecretors than in the other genotypes. Given the overall conservation of *FUT2* among mammals (23), these observations associated with a lack of functional enzyme activity in the GI tract could be indicative of an altered mucosal constitution that may contribute to the manifestation and progression of CD.

The intestinal mucosal barrier has a high regenerative capacity and builds up a pivotal barrier not only for mechanical or chemical stressors but also for those induced by the microbiota. The immunological interface of the intestines is composed of a multilayered structure from the luminal-secreted mucus, the intestinal epithelial cells, and the underlying mucosa-associated migratory immune cells. Bacteria are mostly restricted to the outer layer of the mucus and interact directly with the underlying epithelia only in pathological conditions (43). These adjacent bacteria are able to trigger the secretion of specific glycans, such as blood groups antigens, which in return can serve as nutrients or attachment sites for the bacteria, highlighting the close interaction of resident bacteria with the gut (24, 25, 44). Thus, the clustering we observed according to *W143X* genotype might be explained by the common presence/absence of preferred attachment site(s), which are, in several cases, fucosylated glycoconjugates (45). We hypothesize that the glycoconjugate profile of the mucosa of heterozygotes (*Sese*) may differ from each respective homozygote, for example, via competition for substrates with other glycosyltransferases expressed in the GI tract, resulting in an influence on the availability of bacterial attachment sites or nutritional resources (46). Variation in α-1,2-fucosyltransferase activity was observed in fucosylated milk proteins, and it is hypothesized that variation not only at the level of secretor status but also at the level of genotype plays an important role in determining which substrates are glycosylated by the same set of enzymes (46). Experimental evidence also points toward a dose effect of FUT2, with significant differences among *SeSe* and *Sese* individuals (47).

To shed light on the groups contributing to differences in beta diversity at higher taxonomic levels, we performed an indicator species analysis at both the genus and species levels. At the genus level, we identified among CD patients only a single taxa belonging to the family Lachnospiraceae. The family Lachnospiraceae has been recovered in numerous metagenomic surveys and belongs to the core gut microbiota (48) but was also previously demonstrated to be increased in ileal CD (49) and in mice with dextran sodium sulfate (DSS)-induced colitis (50).

At the species level, most OTUs displayed a strong association with control individuals, stressing the compositional difference with respect to disease status. Importantly, several of these OTUs are considered to be probiotic [e.g., *Lactobacillus* (51) and *Faecalibacterium* (52)] and/or were previously observed to be reduced in the context of CD (e.g., *Faecalibacterium prausnitzii*). Our observation that OTUs belonging to the *Lactobacillus* genus are particularly associated with healthy secretors is also consistent with the fact that they possess adhesins specifically targeting ABO blood group antigens in the mucosa (53), which may contribute to the probiotic function of *Lactobacillus* by blocking the attachment of potential pathogens to mucosal surfaces (54).

Interestingly, although *Stenotrophomonas* is becoming increasingly important as a nosocomial pathogen—associated with airway infections, weak bacteremia in immunocompromised subjects, resistance to antibiotics, and even life-threatening chronic enteritis (55, 56)—several species' OTUs belonging to this genus were identified among our controls. Notable features of this genus are its weak invasiveness but its variety of colonization mechanisms and strong ability to form biofilms, making it a successful colonizer of various hosts (56). This genus was more prevalent among *SeSe* and *Sese* controls, which might indicate a preference for mucosal substrates containing blood group antigens.

Finally, we identified several species significantly associated with *sese* individuals, in both control and CD groups. Unclassified species belonging to the family Lachnospiraceae were identified as indicator species in both healthy and CD nonsecretors. As mentioned above, this group becomes more abundant upon experimental induction of colitis in mice by the chemical irritant DSS (50). Because DSS treatment changes the characteristics of the inner and outer mucosal barriers, facilitating bacterial penetration by reduced thickness and permeability (42), it is possible that this group may also display increased invasiveness in undisturbed mucosa lacking protective ABO antigens. Furthermore, flagellins of this bacterial group serve as elicitors of inflammation in CD (57). Another noteworthy observation is an OTU belonging to *Prevotella* that is associated with *sese* controls. Because these bacteria are able to digest mucins and are of increasing clinical relevance for chronic infections, it is possible that they contribute to mucosa impairment (58, 59).

Our results offer important insight into both the host genetic basis of diversity in the intestinal microbiota and the potential means through which alternative *FUT2* alleles contribute to disease susceptibility. Although several individual interactions between microbes and *FUT2*-dependent antigens in the GI tract were known before this study, our results indicate that differences in the composition and structure of bacterial communities according to *FUT2* genotype may contribute to CD susceptibility. Because of their long-term maintenance and repeated evolutionary origin (14, 15), loss-of-function mutations at *FUT2* are extremely common among human populations. Given the association of genetic variants at *FUT2* with multiple immune phenotypes, further understanding of the role of *FUT2* in maintaining homeostasis between mammalian hosts and their complex associated microbial communities may considerably contribute to future improvements in preventative and therapeutic patient care in acute infectious and chronic inflammatory diseases.

Materials and Methods

Human Samples. The biopsy bank of the outpatient clinic of the Department of General Internal Medicine of University Hospital Schleswig-Holstein was screened for individuals with a diagnosis of CD and for healthy controls of Caucasian (northern European) ancestry. Symptoms were in remission at the time of sampling, and all biopsies were taken from noninflamed tissue (for details, see *SI Materials and Methods*). All procedures related to patients and healthy subjects were approved by the University Hospital Schleswig-Holstein ethics committee (B231/98 and A154/06) and follow the guidelines of the Declaration of Helsinki.

Genotyping. Functionally tested TaqMan SNP Genotyping Assays (Applied Biosystems) were used to genotype the primary nonsecretor allele in Caucasian populations, *W143X* (G428A; rs601338) (13, 15), and A385T (rs1047781) on an automated platform (60). All process data were written to and administered by a previously described database-driven laboratory information management system (LIMS) (61).

DNA Extraction and 16S rRNA Gene Pyrosequencing. DNA from sigmoid colonic biopsies was extracted with the AllPrep DNA/RNA Mini Kit (Qiagen) following the manufacturer's instructions with the addition of a bead-beating step after the addition of the RLT buffer to enhance cell lysis. The 27F–338R region of the 16S rRNA gene was amplified and sequenced on the 454 Life Sciences GS-FLX platform using Titanium sequencing chemistry (*SI Materials and Methods*).

Sequence Processing and Quality Control. Raw sequences were trimmed by a Perl script using a Smith–Waterman alignment algorithm to identify multiplex identifier primer sequences. A mean quality score of ≥20 and a minimum length of 200 nt for the coupled V1–V2 region was required. Sequences were then aligned to the highly curated seed database from SILVA (62) using a *k*-mer alignment procedure as implemented in mothur 1.12 (63). Sequences that did not match the defined core region of the seed alignment were manually removed. Chimeric sequences were removed by using the UCHIME function in UCLUST 3.0 (64) with the SILVA gold database as a reference. For all subsequent analyses, we used a random subset of 1,000 sequences per individual to normalize the read distribution (a single individual was included with 990 sequences), which was previously suggested as a good balance between sample number and coverage (65). This sampling depth corresponded to an average Good's coverage value of 0.86 ± 0.07 SD. Sequences were confirmed as bacterial by using the Ribosomal Database Project (RDP) classifier with an 80% bootstrap threshold (26, 27). Aligned sequences were used to compute a distance matrix and group-related sequences into OTUs using mothur (OTU abundances are provided as Dataset S1). Phylogenetic tree construction was carried out by using FastTree 2.0 with a generalized time-reversible (GTR) substitution model (66).

Statistical Analysis. Alpha diversity indices were calculated in R (67). Phylogenetic diversity was calculated according to Faith et al. (31). Analyses of phyla abundances and alpha diversity were performed with an ANOVA framework. Alternative models were tested sequentially by using all possible combinations of predictor variables (i.e., genotype, disease status, age, and sex), and the best model was chosen according to the Akaike information criterion. FAST UniFrac was used to calculate the unweighted UniFrac metric (33, 34). For statistical analysis of beta diversity indices, we performed nonparametric matrix-based analysis of variance by using *adonis* implemented in the vegan package for R (35, 68). PCoAs were performed in R without constraints. Goodness of fit was assessed with 10^5 permutations on all three axes. A correlation of each single axis with specific factors (i.e., genotype, disease status, age, and sex) was assessed by using linear models with model selection procedures as described above. For the comparison of ordinations, we used Procrustes analysis with 10^5 permutations (36). CCA was carried out with the least number of variables to assess the highest explained variation (69).

Indicator species analysis, as described by De Cáceres et al. (37), was implemented in the R package indicspecies and based on 10^5 permutations. To reduce the number of candidate OTUs, we applied the sample discrimination (SIMPER) method implemented in PRIMER 6 (70). We limited the analysis to OTUs with a minimum contribution of 0.5% to the overall similarity within each cluster as assessed by the Bray-Curtis distance. The thresholds for *P* values of OTU association were adjusted for multiple testing by the method of Benjamini and Hochberg (71). We set the threshold for significance after correction at the 5% level and that for trends in the data at 10%.

ACKNOWLEDGMENTS. We thank all study participants and Katja Cloppenborg-Schmidt, Tanja Wesse, and Manuela Kramp for excellent technical assistance. This work was supported by the Deutsche Forschungsgemeinschaft Excellence Cluster "Inflammation at Interfaces," the National Genome Research Network (NGFN) "Systematic Genomics of Chronic Inflammatory Barrier Diseases" (Subprojects GP1 and 10), and the Max Planck Society.

1. Schreiber S, Rosenstiel P, Albrecht M, Hampe J, Krawczak M (2005) Genetics of Crohn disease, an archetypal inflammatory barrier disease. *Nat Rev Genet* 6:376–388.
2. Xavier RJ, Podolsky DK (2007) Unravelling the pathogenesis of inflammatory bowel disease. *Nature* 448:427–434.
3. Ott SJ, et al. (2004) Reduction in diversity of the colonic mucosa associated bacterial microflora in patients with active inflammatory bowel disease. *Gut* 53:685–693.
4. Lepage P, et al. (2005) Biodiversity of the mucosa-associated microbiota is stable along the distal digestive tract in healthy individuals and patients with IBD. *Inflamm Bowel Dis* 11:473–480.
5. Frank DN, et al. (2007) Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc Natl Acad Sci USA* 104:13780–13785.
6. Sokol H, Lay C, Seksik P, Tannock GW (2008) Analysis of bacterial bowel communities of IBD patients: What has it revealed? *Inflamm Bowel Dis* 14:858–867.
7. Qin J, et al.; MetaHIT Consortium (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464:59–65.
8. Taurog JD, et al. (1994) The germfree state prevents development of gut and joint inflammatory disease in HLA-B27 transgenic rats. *J Exp Med* 180: 2359–2364.
9. Franke A, et al. (2010) Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet* 42:1118–1125.
10. Frank DN, et al. (2011) Disease phenotype and genotype are associated with shifts in intestinal-associated microbiota in inflammatory bowel diseases. *Inflamm Bowel Dis* 17:179–184.

11. Rehman A, et al. (2011) Nod2 is essential for temporal development of intestinal microbial communities. *Gut* 60:1354–1362.
12. McGovern DPB, et al.; International IBD Genetics Consortium (2010) Fucosyltransferase 2 (*FUT2*) non-secretor status is associated with Crohn's disease. *Hum Mol Genet* 19:3468–3476.
13. Kelly RJ, Rouquier S, Giorgi D, Lennon GG, Lowe JB (1995) Sequence and expression of a candidate for the human *Secretor* blood group α(1,2)fucosyltransferase gene (*FUT2*). Homozygosity for an enzyme-inactivating nonsense mutation commonly correlates with the non-secretor phenotype. *J Biol Chem* 270:4640–4649.
14. Koda Y, Tachida H, Soejima M, Takenaka O, Kimura H (2000) Ancient origin of the null allele se^{428} of the human ABO-secretor locus (*FUT2*). *J Mol Evol* 50:243–248.
15. Ferrer-Admetlla A, et al. (2009) A natural history of FUT2 polymorphism in humans. *Mol Biol Evol* 26:1993–2003.
16. Andrés AM, et al. (2009) Targets of balancing selection in the human genome. *Mol Biol Evol* 26:2755–2764.
17. Lindesmith L, et al. (2003) Human susceptibility and resistance to Norwalk virus infection. *Nat Med* 9:548–553.
18. Raza MW, et al. (1991) Association between secretor status and respiratory viral illness. *BMJ* 303:815–818.
19. Evans DA, Horwich L, McConnell RB, Bullen MF (1968) Influence of the ABO blood groups and secretor status on bleeding and on perforation of duodenal ulcer. *Gut* 9: 319–322.
20. Haverkorn MJ, Goslings WR (1969) Streptococci, ABO blood groups, and secretor status. *Am J Hum Genet* 21:360–375.
21. Chaudhuri A, DasAdhikary CR (1978) Possible role of blood-group secretory substances in the aetiology of cholera. *Trans R Soc Trop Med Hyg* 72:664–665.
22. Ruiz-Palacios GM, Cervantes LE, Ramos P, Chavez-Munguia B, Newburg DS (2003) *Campylobacter jejuni* binds intestinal H(O) antigen (Fucα1, 2Galβ1, 4GlcNAc), and fucosyloligosaccharides of human milk inhibit its binding and infection. *J Biol Chem* 278:14112–14120.
23. Abrantes J, Posada D, Guillon P, Esteves PJ, Le Pendu J (2009) Widespread gene conversion of α-2-fucosyltransferase genes in mammals. *J Mol Evol* 69:22–31.
24. Bry L, Falk PG, Midtvedt T, Gordon JI (1996) A model of host-microbial interactions in an open mammalian ecosystem. *Science* 273:1380–1383.
25. Meng D, et al. (2007) Bacterial symbionts induce a FUT2-dependent fucosylated niche on colonic epithelium via ERK and JNK signaling. *Am J Physiol Gastrointest Liver Physiol* 293:G780–G787.
26. Cole JR, et al.; Ribosomal Database Project (2003) The Ribosomal Database Project (RDP-II): Previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. *Nucleic Acids Res* 31:442–443.
27. Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 73:5261–5267.
28. Schmiedel D, et al. (2009) Rapid and accurate diagnosis of human intestinal spirochetosis by fluorescence in situ hybridization. *J Clin Microbiol* 47:1393–1401.
29. Peter H, et al. (2011) Function-specific response to depletion of microbial diversity. *ISME J* 5:351–361.
30. Hughes JB, Hellmann JJ, Ricketts TH, Bohannan BJM (2001) Counting the uncountable: Statistical approaches to estimating microbial diversity. *Appl Environ Microbiol* 67:4399–4406.
31. Faith DP (1992) Conservation evaluation and phylogenetic diversity. *Biol Conserv* 61: 1–10.
32. Jost L (2007) Partitioning diversity into independent α and β components. *Ecology* 88: 2427–2439.
33. Lozupone C, Knight R (2005) UniFrac: A new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* 71:8228–8235.
34. Hamady M, Lozupone C, Knight R (2010) Fast UniFrac: Facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data. *ISME J* 4:17–27.
35. Anderson MJ (2001) A new method for non-parametric multivariate analysis of variance. *Austral Ecol* 26:32–46.
36. Peres-Neto P, Jackson D (2001) How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. *Oecologia* 129:169–178.
37. De Cáceres M, Legendre P, Moretti M (2010) Improving indicator species analysis by combining groups of sites. *Oikos* 119:1674–1684.
38. Eckburg PB, et al. (2005) Diversity of the human intestinal microbial flora. *Science* 308: 1635–1638.
39. Gophna U, Sommerfeld K, Gophna S, Doolittle WF, Veldhuyzen van Zanten SJO (2006) Differences between tissue-associated intestinal microfloras of patients with Crohn's disease and. *J Clin Microbiol* 44:4136–4141.
40. Temperton B, et al. (2009) Bias in assessments of marine microbial biodiversity in fosmid libraries as evaluated by pyrosequencing. *ISME J* 3:792–796.
41. Pullan RD, et al. (1994) Thickness of adherent mucus gel on colonic mucosa in humans and its relevance to colitis. *Gut* 35:353–359.
42. Johansson MEV, et al. (2010) Bacteria penetrate the inner mucus layer before inflammation in the dextran sulfate colitis model. *PLoS ONE* 5:e12238.
43. Johansson MEV, et al. (2008) The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria. *Proc Natl Acad Sci USA* 105:15064–15069.
44. Hoskins LC, Boulding ET (1976) Degradation of blood group antigens in human colon ecosystems. I. In vitro production of ABH blood group-degrading enzymes by enteric bacteria. *J Clin Invest* 57:63–73.
45. Magalhães A, et al. (2009) Fut2-null mice display an altered glycosylation profile and impaired BabA-mediated *Helicobacter pylori* adhesion to gastric mucosa. *Glycobiology* 19:1525–1536.
46. Erney RM, et al. (2000) Variability of human milk neutral oligosaccharides in a diverse population. *J Pediatr Gastroenterol Nutr* 30:181–192.
47. Marionneau S, Airaud F, Bovin NV, Le Pendu J, Ruvoën-Clouet N (2005) Influence of the combined ABO, FUT2, and FUT3 polymorphism on susceptibility to Norwalk virus attachment. *J Infect Dis* 192:1071–1077.
48. Sekelja M, Berget I, Næs T, Rudi K (2011) Unveiling an abundant core microbiota in the human adult colon by a phylogroup-independent searching approach. *ISME J* 5: 519–531.
49. Willing BP, et al. (2010) A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. *Gastroenterology* 139:1844–1854, e1.
50. Nagalingam NA, Kao JY, Young VB (2011) Microbial ecology of the murine gut associated with the development of dextran sodium sulfate-induced colitis. *Inflamm Bowel Dis* 17:917–926.
51. Lebeer S, Vanderleyden J, De Keersmaecker SCJ (2008) Genes and molecules of lactobacilli supporting probiotic action. *Microbiol Mol Biol Rev* 72:728–764.
52. Sokol H, et al. (2008) *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc Natl Acad Sci USA* 105:16731–16736.
53. Watanabe M, et al. (2010) Identification of a new adhesin-like protein from *Lactobacillus mucosae* ME-340 with specific affinity to the human blood group A and B antigens. *J Appl Microbiol* 109:927–935.
54. Chan RC, Reid G, Irvin RT, Bruce AW, Costerton JW (1985) Competitive exclusion of uropathogens from human uroepithelial cells by *Lactobacillus* whole cells and cell wall fragments. *Infect Immun* 47:84–89.
55. Hellmig S, et al. (2005) Life-threatening chronic enteritis due to colonization of the small bowel with *Stenotrophomonas maltophilia*. *Gastroenterology* 129:706–712.
56. Ryan RP, et al. (2009) The versatility and adaptation of bacteria from the genus *Stenotrophomonas*. *Nat Rev Microbiol* 7:514–525.
57. Duck LW, et al. (2007) Isolation of flagellated bacteria implicated in Crohn's disease. *Inflamm Bowel Dis* 13:1191–1201.
58. Rho JH, et al. (2005) A novel mechanism for desulfation of mucin: Identification and cloning of a mucin-desulfating glycosidase (sulfoglycosidase) from *Prevotella* strain RS2. *J Bacteriol* 187:1543–1551.
59. Lucke K, Miehlke S, Jacobs E, Schuppler M (2006) Prevalence of *Bacteroides* and *Prevotella* spp. in ulcerative colitis. *J Med Microbiol* 55:617–624.
60. Hampe J, et al. (2001) An integrated system for high throughput TaqMan based SNP genotyping. *Bioinformatics* 17:654–655.
61. Teuber M, et al. (2005) Improving quality control and workflow management in high-throughput single-nucleotide polymorphism genotyping environments. *J Assoc Lab Autom* 10:43–47.
62. Pruesse E, et al. (2007) SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* 35: 7188–7196.
63. Schloss PD, et al. (2009) Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75:7537–7541.
64. Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–2461.
65. Hamady M, Knight R (2009) Microbial community profiling for human microbiome projects: Tools, techniques, and challenges. *Genome Res* 19:1141–1152.
66. Price MN, Dehal PS, Arkin AP (2010) FastTree 2—Approximately maximum-likelihood trees for large alignments. *PLoS ONE* 5:e9490.
67. R Development Core Team (2009) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna).
68. Oksanen J, et al. (2011) *vegan: Community Ecology Package* (Comprehensive R Archive Network), R package Version 1.17-6, http://cran.r-project.org/web/packages/vegan/index.html.
69. ter Braak CJF (1986) Canonical correspondence analysis: A new eigenvector technique for multivariate direct gradient analysis. *Ecology* 67:1167–1179.
70. Clarke K, Gorley R (2006) *PRIMER v6: User Manual/Tutorial* (PRIMER-E Ltd, Plymouth, UK).
71. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc, B* 57:289–300.

Supporting Information

Rausch et al. 10.1073/pnas.1106408108

SI Materials and Methods

Human Samples. The biopsy bank of the outpatient clinic of the Department of General Internal Medicine of University Hospital Schleswig-Holstein was screened for individuals with a diagnosis of Crohn disease (CD) ($n = 29$) and for healthy controls ($n = 18$) of Caucasian (northern European) ancestry. CD samples were selected based on the following criteria: (*i*) macro- and microscopically noninflamed sigmoid region biopsies available in biopsy bank and (*ii*) clinical remission with Crohn disease Activity Index (CDAI) score of <150 at time of sampling. The inflammatory activity was independently scored by two investigators. Further patient characteristics are given in Table S1.

All diagnoses were based on standard criteria via radiological and endoscopic examinations. Indications for colonoscopy were monitoring of therapy response and cancer surveillance in CD patients and participation in a volunteer study for the healthy subjects (exclusion of intestinal pathologies before nutritional intervention). All samples (pairs of matching DNA/biopsy) and phenotype information were pseudonymized before the procedure. All procedures related to patients and healthy subjects were approved by the University Hospital Schleswig-Holstein ethics committee (B231/98 and A154/06) and follow the guidelines of the Declaration of Helsinki. All individuals agreed to participation by giving informed consent at least 24 h before the study.

16S rRNA Gene Pyrosequencing. The 16S rRNA gene was amplified by using forward (5′-**CTATGCGCCTTGCCAGCCCGCTCAG***T-C*AGAGTTTGATCCTGGCTCAG-3′) and reverse (5′-**CGTATCGCCTCCCTCGCGCCATCAG**XXXXXXXXXX*CA*TGCTGCCTCCCGTAGGAGT-3′) primers flanking the V1 and V2 hypervariable regions. The 454 Life Sciences primer B (forward) and A (reverse) adapter sequences are denoted in boldface, and the underlined sequences represent the broadly conserved bacterial primers 27F and 338R. A 2-base linker sequence (TC/CA; shown in italics) was added as recommended by Roche (454). A unique 10-base multiplex identifier (designated as XXXXXXXXXX) was added to the reverse primer to tag each PCR product. Template DNA (100 ng) was added to 25-μL PCR reactions performed with Phusion Hot Start DNA Polymerase (Finnzymes). The cycling conditions were as follows: initial denaturation for 30 s at 98 °C; 30 cycles of 9 s at 98 °C, 30 s at 55 °C, and 30 s at 72 °C; and final extension for 10 min at 72 °C. All reactions were performed in duplicate and combined after PCR. PCR products were extracted with the Qiagen MiniElute Gel Extraction Kit and quantified with the Quant-iT dsDNA Broad-Range Assay Kit on a NanoDrop 3300 fluorometer. Equimolar amounts of purified PCR product were pooled and further purified with AMPure beads (Agencourt). A sample of each library was run on an Agilent Bioanalyzer before emulsion PCR and sequencing as recommended by Roche. Amplicon libraries were subsequently sequenced on a 454 Life Sciences GS-FLX using Titanium sequencing chemistry.

Influence of CD Subphenotypes. Because different subphenotypes of CD (i.e., ileal, colonic, ileocolonic) were present in our patient sample, we investigated their influence on the assembly of microbial communities (Table S1). We detected no significant differences in alpha diversity measures between subphenotypes (Shannon H, $F_{2,26} = 0.129$, $P = 0.88$; Shannon evenness (1), $F_{2,26} = 0.132$, $P = 0.877$; Chao1, $F_{2,26} = 0.135$, $P = 0.874$; phylogenetic diversity, $F_{2,26} = 0.041$, $P = 0.960$) nor any significant differences between communities by using measures of beta diversity (*adonis*: Jaccard, $R^2 = 0.065$, $P = 0.889$; Bray-Curtis, $R^2 = 0.06$, $P = 0.913$; unweighted UniFrac, $R^2 = 0.058$, $P = 0.981$).

1. Jost L (2007) Partitioning diversity into independent α and β components. *Ecology* 88: 2427–2439.

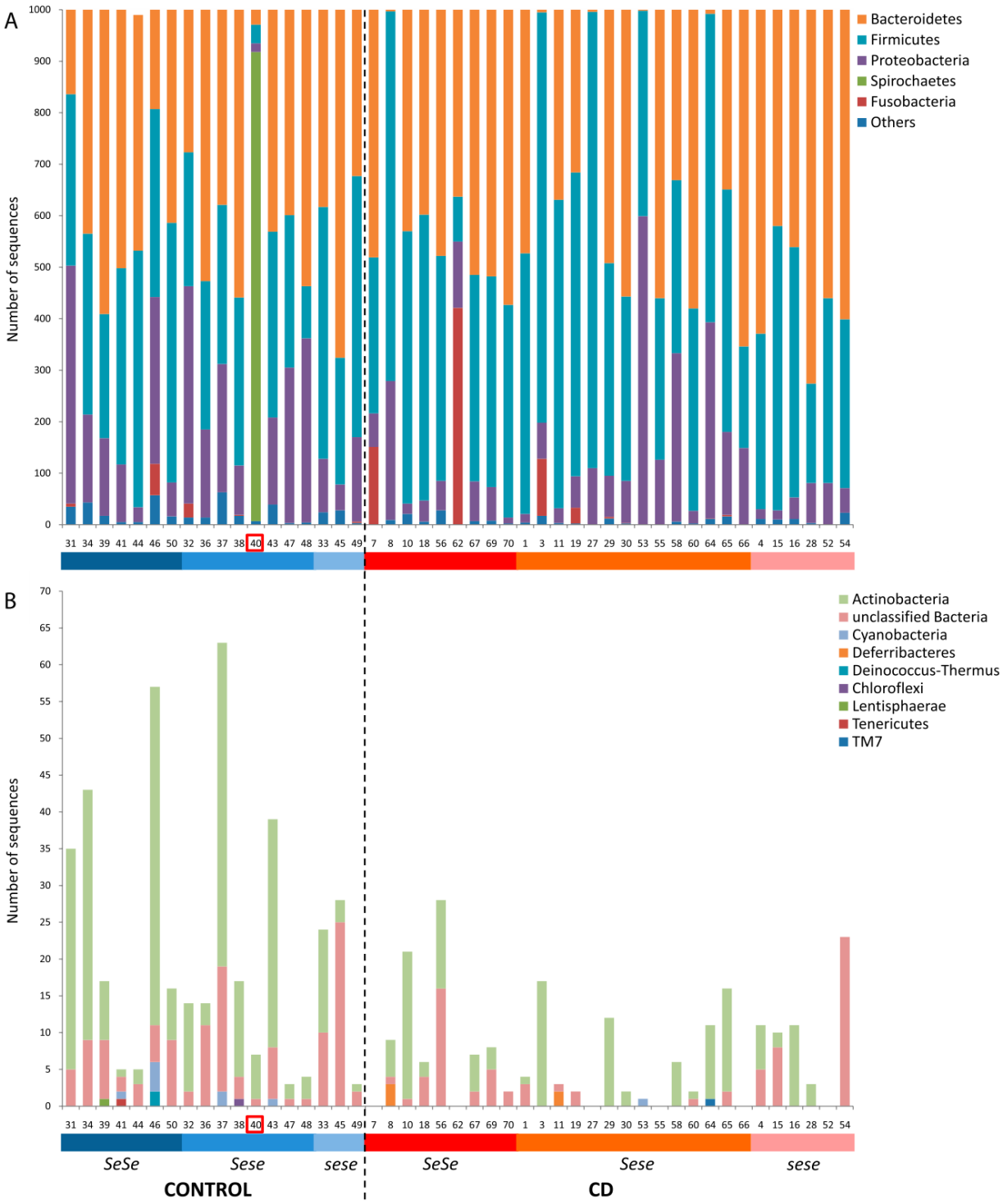


Fig. S1. (*A*) Distribution of read number among the major phyla in sampled individuals. Individual #40 (boxed in red) was excluded from further analysis because of suspected human intestinal spirochaetosis. (*B*) Read number among the rare phyla and unclassified bacteria comprising the "Others" category in *A*.

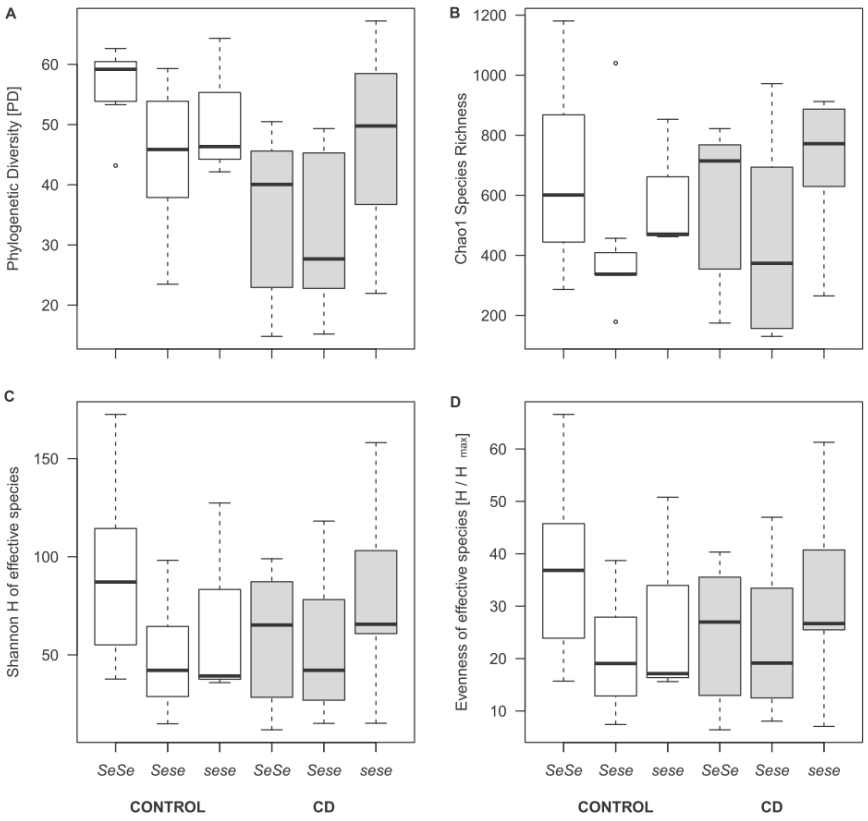


Fig. S2. Mean alpha diversity measures based on species-level operational taxonomic units (OTUs). (*A*) Faith's phylogenetic diversity (1). (*B*) Chao1 species richness metric. (*C*) Shannon H of effective species numbers. (*D*) Shannon evenness of effective species numbers. Error bars indicate SD.

1. Faith DP (1992) Conservation evaluation and phylogenetic diversity. *Biol Conserv* 61:1–10.

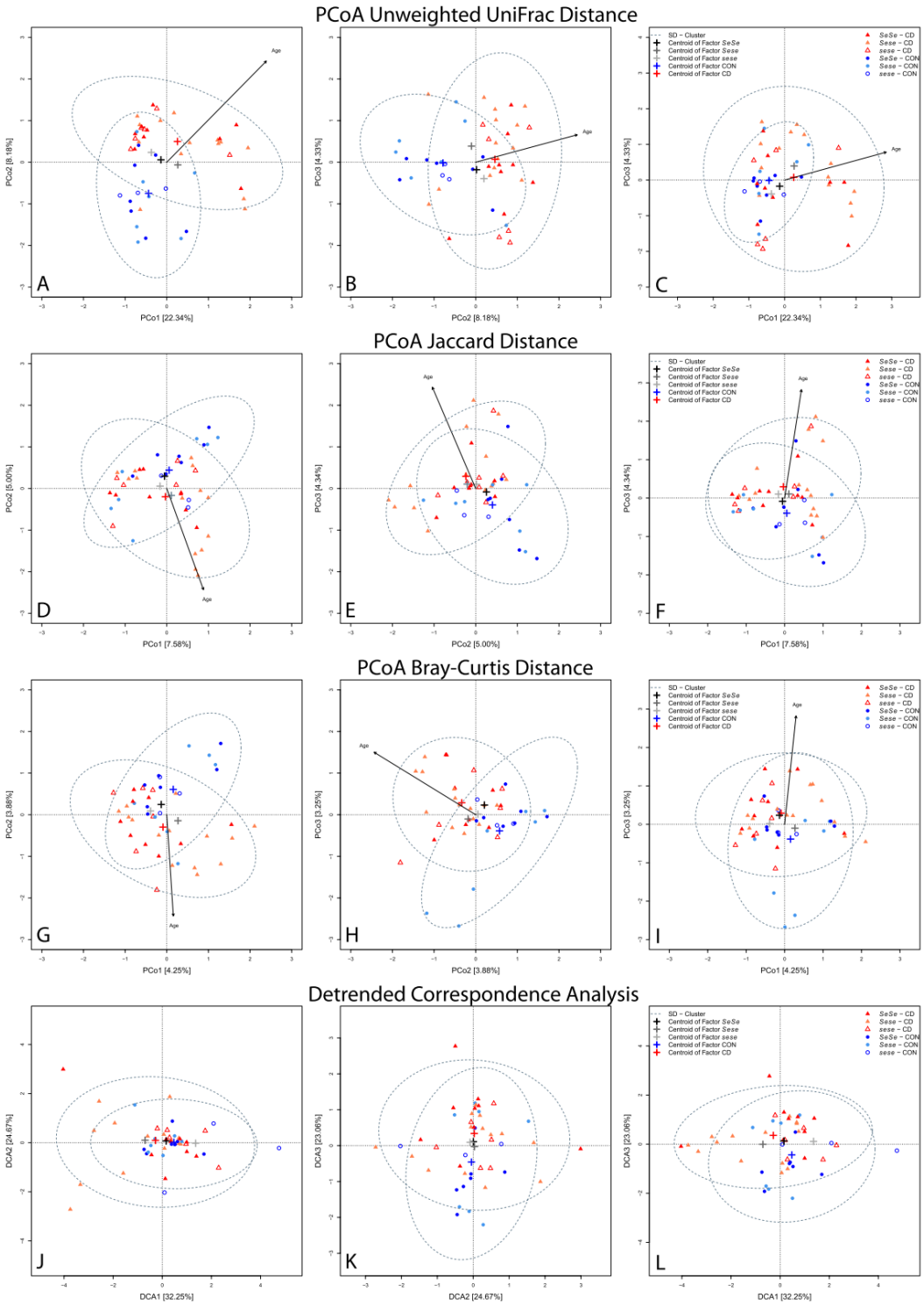


Fig. S3. (*A–C*) Ordination of the unweighted UniFrac distance by principal coordinate analysis (PCoA) in three dimensions (see main text for statistical analysis). (*D–F*) PCoA of the Jaccard index in three dimensions. Centroids of disease status (goodness of fit $R^2 = 0.1048$, $P = 0.002$), genotype within disease status ($R^2 = 0.183$, $P = 0.039$), and age (goodness of fit $R^2 = 0.263$, $P = 0.005$) are correlated with all three axes. Analysis by linear models revealed a correlation of sex with the first axis ($P = 0.047$), whereas the second and third axes show the gradients between disease status ($P = 0.009$) and age ($P = 0.001$), respectively (Table S3). (*G–I*) PCoA of the Bray-Curtis index (data Wisconsin-transformed) in three dimensions. The centroids of disease status (goodness of fit $R^2 = 0.154$, $P < 0.0001$), genotype within disease status (goodness of fit $R^2 = 0.26$, $P = 0.0008$), and age of the subjects (goodness of fit $R^2 = 0.213$, $P = 0.017$) are correlated with

Legend continued on following page

all three axes. Linear models of the site scores revealed a correlation of sex with the first axis ($P = 0.079$), whereas the second and third axes display significant gradients in disease status ($P = 0.0001$) and genotype within disease status ($P = 0.008$), respectively (Table S3). (*J–L*) Detrended correspondence analysis (DCA) visualized in three dimensions. The centroids of disease status (goodness of fit $R^2 = 0.064$, $P = 0.038$), genotype (goodness of fit $R^2 = 0.137$, $P = 0.006$), and genotype–disease status interaction ($R^2 = 0.238$, $P = 0.006$) are highly correlated with all three axes. Analysis by linear models revealed a correlation of genotype with the first axis ($P = 0.003$), whereas the second axis shows no correlation with known factors. The third dimension is significantly influenced by disease status ($P = 0.009$) (Table S3).

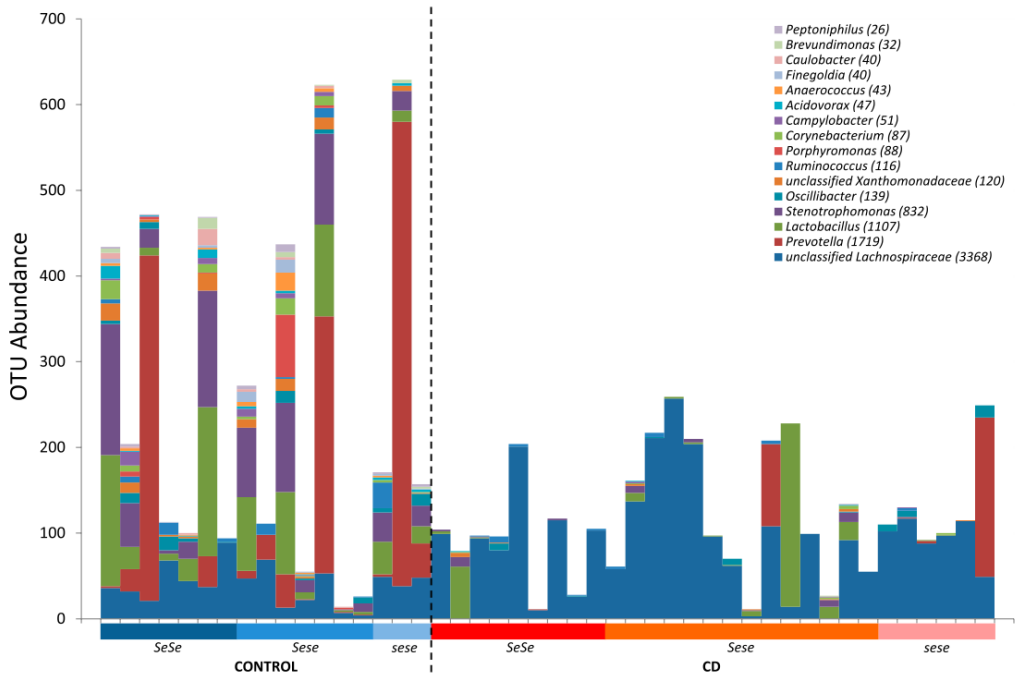


Fig. S4. Abundance of genus-level OTUs identified by indicator species analysis with respect to health status. Numbers in parentheses indicate the total read number in the normalized dataset (1,000 reads per individual).

Table S1. Patient characteristics

	CD				Control			
Parameter	All	*SeSe*	*Sese*	*sese*	All	*SeSe*	*Sese*	*sese*
n	29	8	15	6	18	7	8	3
Sex ratio, female/male	20/9	4/4	13/2	3/3	14/4	6/1	6/2	2/1
Median age, y	39	35.5	43	43	26	28	26.5	25
Age range, y	27–55	29–48	27–55	36–48	23–30	25–30	23–30	23–26
Medication					N/A	N/A	N/A	N/A
Cortisone	5/29	1	3	1				
Azathioprine	1/29	1	0	0				
5-Aminosalicylic acid (5-ASA)	2/29	1	0	1				
Anti-TNFα	2/29	1	0	1				
Disease subtype					N/A	N/A	N/A	N/A
Ileal	8/29	3	4	1				
Ileocolonic	16/29	3	9	4				
Colonic	4/29	2	1	1				
Undetermined	1/29	0	1	0				

N/A, not applicable; *SeSe*, homozygous secretor; *Sese*, heterozygous secretor; *sese*, nonsecretor.

Table S2. Statistical analysis of phyla abundances

		Phyla			
Test	Categories tested	Actinobacteria	Firmicutes	Bacteroidetes	Proteobacteria
ANOVA	Disease status	**0.002**	**0.034**	N/A	**0.020**
	Secretor status	N/A	N/A	**0.036**	N/A
	Genotype	N/A	N/A	N/A	**0.047**
Post hoc Mann–Whitney					
Subsets					
Secretor	Disease status	**0.006***	**0.005***	N/A	**0.012**
Nonsecretor	Disease status	0.603	0.714	N/A	0.095
Genotype	*SeSe:sese*	N/A	N/A	0.121	0.295
	SeSe:Sese	N/A	N/A	0.728	0.154
	Sese:sese	N/A	N/A	**0.044**	**0.017**
Control	*SeSe:sese*	N/A	N/A	N/A	0.5167
	SeSe:Sese	N/A	N/A	N/A	0.3374
	Sese:sese	N/A	N/A	N/A	0.067
CD	*SeSe:sese*	N/A	N/A	N/A	0.596
	SeSe:Sese	N/A	N/A	N/A	0.156
	Sese:sese	N/A	N/A	N/A	0.0622

Significant *P* values are indicated in boldface. N/A, no test applied.
*Significant after sequential Bonferroni correction.

Table S3. Linear model analysis of PCoA and DCA axis scores

Ordination	Transformation	Axis	Model	Akaike information criterion	Factors	Adjusted R^2	*P*
PCoA	None	1, 2, 3	Null	127.205	1		
Unweighted UniFrac	None	1	Best	122.732	Disease status	0.112	0.0134
	None	2	Best	103.740	Disease status	0.412	<0.0001
	None	3	Best	124.938	Genotype	0.087	0.0534
PCoA Jaccard	None	1, 2, 3	Null	115.314	1		
	None	1	Best	113.150	Sex	0.066	0.0471
	None	2	Best	110.158	Disease status	0.125	0.0093
	None	3	Best	106.015	Age	0.200	0.0011
PCoA Bray-Curtis	Wisconsin	1, 2, 3	Null	116,312	1		
	Wisconsin	1	Best	115.060	Sex	0.047	0.0795
	Wisconsin	2	Best	102.922	Disease status	0.268	0.0001
	Wisconsin	3	Best	108.887	Disease status × genotype	0.230	0.0078
DCA	None	1	Null	175.828	1		
	None		Best	167.067	Genotype	0.207	0.0026
	None	2	Null	130.402	1		
	None		Best	> 130.402	N/A	N/A	N/A
	None	3	Null	135.010	1		
	None		Best	129.755	Disease status	0.127	0.0088

N/A, no test applied.

Table S4. List of indicator species and genera and their properties

Taxonomic level	Association	Direction	R^2	*P* (Benjamini and Hochberg-adjusted)*	OTU ID	OTU classification to genus level (no. of reads in normalized dataset)†	Comments
Species	Disease status	Control	0.435	**0.00001 (0.00025)**	1117	*Lactobacillus* spp. (173)	Normal member of the colon microbiome. probiotic characteristics (1, 2)
		Control	0.446	**0.00007 (0.0007)**	2270	*Lactobacillus* spp. (554)	Normal member of the colon microbiome, probiotic characteristics (1, 2)
		Control	0.495	**0.00005 (0.00063)**	2420	*Lactobacillus* spp. (57)	Normal member of the colon microbiome, probiotic characteristics (1, 2)
		Control	0.373	**0.0004 (0.00286)**	2267	*Prevotella* spp. (73)	Mucosa-associated anaerobe, found in the upper and lower gastrointestinal (GI) tract, pathogenic with increasing relevance and resistance (3, 4)
		Control	0.266	0.0146 (0.073)	2165	*Coprobacillus* spp. (377)	Altered abundance in inflammatory bowel disease (IBD) cases and belonging to the core microbiome of the gut (5, 6)
		Control	0.482	**0.00001 (0.00025)**	2460	*Stenotrophomonas* spp. (519)	Ubiquitous, nosocomial pathogen (7)
		Control	0.479	**0.00014 (0.00117)**	2562	*Stenotrophomonas* spp. (224)	Ubiquitous, nosocomial pathogen (7)
		Control	0.514	**0.00004 (0.000625)**	3631	*Stenotrophomonas* spp. (30)	Ubiquitous, nosocomial pathogen (7)
		Control	0.358	**0.00194 (0.0108)**	3844	*Faecalibacterium* spp. (63)	Normal member of the colon microbiome, depletion with potential role in IBD and suspected as probiotic with potential to reduce inflammatory responses (8–11)
		Control	0.33	**0.00157 (0.0098)**	2551	Unclassified Clostridiales spp. (80)	
	Genotype	*sese*	0.545	0.00082 (0.057)	442	Unclassified Lachnospiraceae spp. (17)	Increased in dextran sodium sulfate (DSS) mouse models (12)
		sese	0.504	0.00211 (0.073)	3774	*Coprococcus* spp. (22)	Saccharolytic/cellulolytic/amylolytic, also fucose (13, 14), associated with CD (15)
	Genotype–disease status	*SeSe* and *Sese*-Control	0.576	0.0172 (0.0909)	2420	*Lactobacillus* spp. (57)	Normal member of the colon microbiome, probiotic characteristics (1, 2)

Table S4. Cont.

Taxonomic level	Association	Direction	R^2	*P* (Benjamini and Hochberg-adjusted)*	OTU ID	OTU classification to genus level (no. of reads in normalized dataset)†	Comments
		SeSe and *Sese*-Control	0.557	0.0207 (0.097)	2562	*Stenotrophomonas* spp. (224)	Ubiquitous, nosocomial pathogen (7)
		SeSe and *Sese*-Control	0.598	0.011 (0.065)	3631	*Stenotrophomonas* spp. (30)	Ubiquitous, nosocomial pathogen (7)
		sese-Control	0.69	**0.00417 (0.0339)**	2267	*Prevotella* spp. (73)	Mucosa-associated anaerobe, found in the upper and lower GI tract, pathogenic with increasing relevance and resistance (3, 4)
		sese-Control	0.516	0.01463 (0.0816)	2488	*Brevundimonas* spp. (12)	Pathogen in immunocompromised subjects and distributed in environment (16–18)
		sese-Control	0.536	0.0211 (0.097)	3689	Unclassified Lachnospiraceae spp. (34)	Increased in DSS mouse models (12)
		sese-Control	0.551	**0.00449 (0.0339)**	2193	*Sutterella* spp. (61)	Member of the colon microbiome and associated with GI infections (3, 10, 19, 20)
		sese-Control	0.717	**0.003 (0.0302)**	3844	*Faecalibacterium* spp. (63)	Normal member of the colon microbiome, depletion with potential role in IBD (8–10)
		sese-CD	0.511	0.02053 (0.0972)	4018	*Alistipes* spp. (46)	Normal member of the GI microbiome (6, 21)
		sese-CD	0.759	**0.00073 (0.0302)**	442	Unclassified Lachnospiraceae spp. (17)	Increased in DSS mouse models (12)
		sese-CD	0.608	0.00906 (0.0565)	3774	*Coprococcus* spp. (22)	Saccharolytic/cellulolytic/amylolytic, also fucose (13, 14), associated with CD (15)
Genus	Disease status	Control	0.498	**0.00003 (0.00036)**		Unclassified Xanthomonadaceae (120)	Associated with gut and oral microbiome but majorly environmental organisms with some pathogenicity, especially in plants (7, 22–25)
		Control	0.409	**0.00001 (0.00024)**		*Acidovorax* (47)	Skin microbiome, associated with earthworm development and gut symbiosis (26–28)
		Control	0.540	**0.00001 (0.00024)**		*Stenotrophomonas* (832)	Ubiquitous, nosocomial pathogen (7)
		Control	0.342	**0.00823 (0.03039)**		*Ruminococcus* (116)	Core gut and fecal microbiome, increased in healthy subjects (6, 29, 30)

Table S4. Cont.

Taxonomic level	Association	Direction	R^2	*P* (Benjamini and Hochberg-adjusted)*	OTU ID	OTU classification to genus level (no. of reads in normalized dataset)†	Comments
		Control	0.306	0.01572 (0.0503)		*Prevotella* (1,719)	Mucosa-associated pathogen with increasing relevance and resistance (3)
		Control	0.210	**0.00136 (0.00593)**		*Porphyromonas* (88)	Saccharolytic, mucosa-associated, cause of gingivitis and soft tissue infections (3, 14)
		Control	0.406	**0.00025 (0.001714)**		*Peptoniphilus* (26)	Chronic soft tissue infections, diabetic ulcers, periodontitis (31–33)
		Control	0.386	**0.00975 (0.0334)**		*Oscillibacter* (139)	Anaerobic *Oscillibacter* was identified in clamp intestines, unknown ecology but found in healthy human guts (34–36)
		Control	0.331	0.020603 (0.07809)		*Lactobacillus* (1,107)	Normal member of the colon microbiome, probiotic characteristics (1, 2)
		Control	0.360	**0.00012 (0.00096)**		*Finegoldia* (40)	Opportunistic pathogen on mucosa (37), associated with chronic infections (31, 32)
		Control	0.396	**0.00061 (0.00336)**		*Corynebacterium* (87)	Highly diverse group, pathogens and saprophytes (38–40)
		Control	0.343	**0.00009 (0.000864)**		*Caulobacter* (40)	Stomach flora (41)
		Control	0.390	**0.00072 (0.003456)**		*Campylobacter* (51)	Cause of enteritis and increased IBD risk (42)
		Control	0.350	**0.00246 (0.00984)**		*Brevundimonas* (32)	Pathogen in immunocompromised subjects and distributed in environment (16–18)
		Control	0.344	**0.00002 (0.00032)**		*Anaerococcus* (43)	Chronic soft tissue infections, diabetic ulcers, periodontitis (31–33)
		CD	0.486	**0.00063 (0.00336)**		Unclassified Lachnospiraceae (3,368)	Increased in DSS mouse models (12)

*Values in boldface indicate significance after Benjamini and Hochberg adjustment (43).
†Classification obtained on the genus level by Ribosomal Database Project (RDP) classifier at the 80% bootstrap threshold (44, 45).

1. Lebeer S, Vanderleyden J, De Keersmaecker SCJ (2008) Genes and molecules of lactobacilli supporting probiotic action. *Microbiol Mol Biol Rev* 72:728–764.
2. Zoetendal EG, et al. (2002) Mucosa-associated bacteria in the human gastrointestinal tract are uniformly distributed along the colon and differ from the community recovered from feces. *Appl Environ Microbiol* 68:3401–3407.
3. Falagas ME, Siakavellas E (2000) *Bacteroides, Prevotella,* and *Porphyromonas* species: A review of antibiotic resistance and therapeutic options. *Int J Antimicrob Agents* 15:1–9.
4. Eckburg PB, et al. (2005) Diversity of the human intestinal microbial flora. *Science* 308:1635–1638.
5. Kassinen A, et al. (2007) The fecal microbiota of irritable bowel syndrome patients differs significantly from that of healthy subjects. *Gastroenterology* 133:24–33.
6. Tap J, et al. (2009) Towards the human intestinal microbiota phylogenetic core. *Environ Microbiol* 11:2574–2584.
7. Hauben L, Vauterin L, Moore ERB, Hoste B, Swings J (1999) Genomic diversity of the genus *Stenotrophomonas*. *Int J Syst Bacteriol* 49:1749–1760.
8. Sokol H, et al. (2009) Low counts of *Faecalibacterium prausnitzii* in colitis microbiota. *Inflamm Bowel Dis* 15:1183–1189.
9. Marteau P (2009) Bacterial flora in inflammatory bowel disease. *Dig Dis* 27(Suppl 1):99–103.
10. Dethlefsen L, Relman DA (2010) Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proc Natl Acad Sci USA* 108 (Suppl 1):4554–4561.
11. Sokol H, et al. (2008) *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc Natl Acad Sci USA* 105:16731–16736.
12. Nagalingam NA, Kao JY, Young VB (2011) Microbial ecology of the murine gut associated with the development of dextran sodium sulfate-induced colitis. *Inflamm Bowel Dis* 17: 917–926.
13. Palmonari A, Stevenson DM, Mertens DR, Cruywagen CW, Weimer PJ (2010) pH dynamics and bacterial community composition in the rumen of lactating dairy cows. *J Dairy Sci* 93: 279–287.
14. Duncan SH, Louis P, Flint HJ (2007) Cultivable bacterial diversity from the human colon. *Lett Appl Microbiol* 44:343–350.
15. Wensinck F, van de Merwe JP, Mayberry JF (1983) An international study of agglutinins to *Eubacterium, Peptostreptococcus* and *Coprococcus* species in Crohn's disease, ulcerative colitis and control subjects. *Digestion* 27:63–69.
16. Han XY, Andrade RA (2005) *Brevundimonas diminuta* infections and its resistance to fluoroquinolones. *J Antimicrob Chemother* 55:853–859.
17. Egert M, Schmidt I, Bussey K, Breves R (2010) A glimpse under the rim—The composition of microbial biofilm communities in domestic toilets. *J Appl Microbiol* 108:1167–1174.
18. Grossart H-P, Dziallas C, Leunert F, Tang KW (2010) Bacteria dispersal by hitchhiking on zooplankton. *Proc Natl Acad Sci USA* 107:11959–11964.
19. Wexler HM, et al. (1996) *Sutterella wadsworthensis* gen. nov., sp. nov., bile-resistant microaerophilic *Campylobacter gracilis*-like clinical isolates. *Int J Syst Bacteriol* 46:252–258.
20. Sakon H, Nagai F, Morotomi M, Tanaka R (2008) *Sutterella parvirubra* sp. nov. and *Megamonas funiformis* sp. nov., isolated from human faeces. *Int J Syst Evol Microbiol* 58:970–975.
21. Hoyles L, McCartney AL (2009) What do we mean when we refer to Bacteroidetes populations in the human gastrointestinal microbiota? *FEMS Microbiol Lett* 299:175–183.
22. Conti S, dos Santos SS, Koga-Ito CY, Jorge AO (2009) Enterobacteriaceae and pseudomonadaceae on the dorsum of the human tongue. *J Appl Oral Sci* 17:375–380.
23. Rajilić-Stojanović M, Smidt H, de Vos WM (2007) Diversity of the human gastrointestinal tract microbiota revisited. *Environ Microbiol* 9:2125–2136.
24. Dewhirst FE, et al. (2010) The human oral microbiome. *J Bacteriol* 192:5002–5017.
25. da Silva ACR, et al. (2002) Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature* 417:459–463.
26. Thakuria D, Schmidt O, Finan D, Egan D, Doohan FM (2010) Gut wall bacteria of earthworms: A natural selection process. *ISME J* 4:357–366.
27. Davidson SK, Stahl DA (2008) Selective recruitment of bacteria during embryogenesis of an earthworm. *ISME J* 2:510–518.
28. Price LB, et al. (2010) The effects of circumcision on the penis microbiome. *PLoS ONE* 5:e8422.
29. Joossens M, et al. (2011) Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. *Gut* 60:631–637.
30. Liu C, Finegold SM, Song Y, Lawson PA (2008) Reclassification of *Clostridium coccoides, Ruminococcus hansenii, Ruminococcus hydrogenotrophicus, Ruminococcus luti, Ruminococcus productus* and *Ruminococcus schinkii* as *Blautia coccoides* gen. nov., comb. nov., *Blautia hansenii* comb. nov., *Blautia hydrogenotrophica* comb. nov., *Blautia luti* comb. nov., *Blautia producta* comb. nov., *Blautia schinkii* comb. nov. and description of *Blautia wexlerae* sp. nov., isolated from human faeces. *Int J Syst Evol Microbiol* 58:1896–1902.
31. Price LB, et al. (2009) Community analysis of chronic wound bacteria using 16S rRNA gene-based pyrosequencing: Impact of diabetes and antibiotics on chronic wound microbiota. *PLoS ONE* 4:e6462.
32. Dowd SE, et al. (2008) Polymicrobial nature of chronic diabetic foot ulcer biofilm infections determined using bacterial tag encoded FLX amplicon pyrosequencing (bTEFAP). *PLoS ONE* 3:e3326.
33. Kumar PS, Griffen AL, Moeschberger ML, Leys EJ (2005) Identification of candidate periodontal pathogens and beneficial species by quantitative 16S clonal analysis. *J Clin Microbiol* 43:3944–3955.
34. Iino T, Mori K, Tanaka K, Suzuki K-i, Harayama S (2007) *Oscillibacter valericigenes* gen. nov., sp. nov., a valerate-producing anaerobic bacterium isolated from the alimentary canal of a Japanese corbicula clam. *Int J Syst Evol Microbiol* 57:1840–1845.
35. Walker AW, et al. (2011) Dominant and diet-responsive groups of bacteria within the human colonic microbiota. *ISME J* 5:220–230.
36. Mondot S, et al. (2011) Highlighting new phylogenetic specificities of Crohn's disease microbiota. *Inflamm Bowel Dis* 17:185–192.
37. Goto T, et al. (2008) Complete genome sequence of *Finegoldia magna*, an anaerobic opportunistic pathogen. *DNA Res* 15:39–47.
38. Gao Z, Tseng CH, Pei Z, Blaser MJ (2007) Molecular analysis of human forearm superficial skin bacterial biota. *Proc Natl Acad Sci USA* 104:2927–2932.
39. Ventura M, et al. (2007) Genomics of Actinobacteria: Tracing the evolutionary history of an ancient phylum. *Microbiol Mol Biol Rev* 71:495–548.
40. Madigan MT, Martinko JM (2006) *Brock Biology of Microorganisms* (Pearson Prentice Hall, Upper Saddle River, NJ), 11th Ed.
41. Bik EM, et al. (2006) Molecular analysis of the bacterial microbiota in the human stomach. *Proc Natl Acad Sci USA* 103:732–737.
42. Nell S, Suerbaum S, Josenhans C (2010) The impact of the microbiota on the pathogenesis of IBD: Lessons from mouse infection models. *Nat Rev Microbiol* 8:564–577.
43. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc, B* 57:289–300.
44. Cole JR, et al.; Ribosomal Database Project (2003) The Ribosomal Database Project (RDP-II): Previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. *Nucleic Acids Res* 31:442–443.
45. Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 73: 5261–5267.

Dataset S1. OTU abundances per individual at 97% sequence similarity threshold

Dataset S1

ARTICLE IN PRESS

Research Article

EASL EUROPEAN ASSOCIATION FOR THE STUDY OF THE LIVER | JOURNAL OF HEPATOLOGY

Extended analysis of a genome-wide association study in primary sclerosing cholangitis detects multiple novel risk loci

Trine Folseraas[1,2,3,†], Espen Melum[1,2,3,†], Philipp Rausch[4,5], Brian D. Juran[6], Eva Ellinghaus[7], Alexey Shiryaev[1,2,3], Jon K. Laerdahl[8,9], David Ellinghaus[7], Christoph Schramm[10], Tobias J. Weismüller[11,12], Daniel Nils Gotthardt[13], Johannes Roksund Hov[1,2,3], Ole Petter Clausen[3,14], Rinse K. Weersma[15], Marcel Janse[15], Kirsten Muri Boberg[1], Einar Björnsson[16], Hanns-Ulrich Marschall[16], Isabelle Cleynen[17], Philip Rosenstiel[7], Kristian Holm[1], Andreas Teufel[18], Christian Rust[19], Christian Gieger[20], H-Erich Wichmann[21,22,23], Annika Bergquist[24], Euijung Ryu[25], Cyriel Y. Ponsioen[26], Heiko Runz[27], Martina Sterneck[28], Severine Vermeire[17], Ulrich Beuers[26], Cisca Wijmenga[29], Erik Schrumpf[1,3], Michael P. Manns[11,12], Konstantinos N. Lazaridis[6], Stefan Schreiber[7,30], John F. Baines[4,5], Andre Franke[7], Tom H. Karlsen[1,2,31,*]

[1]*Norwegian PSC Research Center, Department of Transplantation Medicine, Oslo University Hospital, Rikshospitalet, Oslo, Norway;* [2]*Research Institute for Internal Medicine, Oslo University Hospital, Rikshospitalet, Oslo, Norway;* [3]*Faculty of Medicine, University of Oslo, Oslo, Norway;* [4]*Institute for Experimental Medicine, Christian-Albrechts-University, Kiel, Germany;* [5]*Max Planck Institute for Evolutionary Biology, Plön, Germany;* [6]*Center for Basic Research in Digestive Diseases, Division of Gastroenterology and Hepatology, Mayo Clinic, College of Medicine, Rochester, Minnesota, United States;* [7]*Institute of Clinical Molecular Biology, Christian-Albrechts-University, Kiel, Germany;* [8]*Centre for Molecular Biology and Neuroscience (CMBN) and Department of Microbiology, Oslo University Hospital, Rikshospitalet, Oslo, Norway;* [9]*Bioinformatics Core Facility, Department of Informatics, University of Oslo, Oslo, Norway;* [10]*1st Department of Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany;* [11]*Department of Gastroenterology, Hepatology and Endocrinology, Hannover Medical School, Hannover, Germany;* [12]*Integrated Research and Treatment Center-Transplantation (IFB-tx), Hannover Medical School, Hannover, Germany;* [13]*Department of Medicine, University Hospital of Heidelberg, Heidelberg, Germany;* [14]*Division of Pathology, Oslo University Hospital Rikshospitalet, Oslo, Norway;* [15]*Department of Gastroenterology and Hepatology, University Medical Center Groningen and University of Groningen, The Netherlands;* [16]*Department of Internal Medicine, Institute of Medicine, Sahlgrenska Academy and University Hospital, Gothenburg, Sweden;* [17]*Department of Gastroenterology, University Hospital Gasthuisberg, Leuven, Belgium;* [18]*1st Department of Medicine, University of Mainz, Mainz, Germany;* [19]*Department of Medicine 2, Grosshadern, University of Munich, Munich, Germany;* [20]*Institute of Genetic Epidemiology, Helmholtz Center Munich, German Research Center for Environmental Health, Neuherberg, Germany;* [21]*Institute of Epidemiology I, Helmholtz Center Munich, German Research Center for Environmental Health, Neuherberg, Germany;* [22]*Institute of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany;* [23]*Klinikum Grosshadern, Munich, Germany;* [24]*Department of Gastroenterology and Hepatology, Karolinska University Hospital Huddinge, Stockholm, Sweden;* [25]*Division of Biomedical Statistics and Informatics, Mayo Clinic College of Medicine, Rochester, Minnesota, United States;* [26]*Department of Gastroenterology and Hepatology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands;* [27]*Department of Human Genetics, University Hospital of Heidelberg, Heidelberg, Germany;* [28]*Department of Hepatobiliary Surgery and Transplantation, University Medical Center Hamburg-Eppendorf, Hamburg, Germany;* [29]*Department of Genetics, University Medical Center Groningen and University of Groningen, Groningen, The Netherlands;* [30]*Department for General Internal Medicine, Christian-Albrechts-University, Kiel, Germany;* [31]*Division of Gastroenterology, Institute of Medicine, University of Bergen, Bergen, Norway*

Keywords: Primary sclerosing cholangitis; Genome-wide association study; Single nucleotide polymorphism; Immunogenetics.
Received 6 October 2011; received in revised form 7 March 2012; accepted 26 March 2012

* Corresponding author. Address: Norwegian PSC Research Center, Department of Transplantation Medicine, Oslo University Hospital Rikshospitalet, Postboks 4950 Nydalen, N-0424 Oslo, Norway. Tel.: +47 23 07 2469; fax: +47 23 07 4869.
E-mail address: t.h.karlsen@medisin.uio.no (T.H. Karlsen).
† These authors contributed equally to this work.

Abbreviations: PSC, primary sclerosing cholangitis; SNP, single nucleotide polymorphism; GWAS, genome-wide association study; IBD, inflammatory bowel disease; CEPH, Centre d'Etude du Polymorphisme Humain; HWE, Hardy–Weinberg-Equilibrium; LD, linkage disequilibrium; CMH, Cochran–Mantel–Haenszel; BD, Breslow-Day; HET, heterogeneity; PBC, primary biliary cirrhosis; Chr, chromosome; Al, alleles; OR, odds ratio; CI, confidence interval; GRAIL, Gene Relationships Across Implicated Loci.

Please cite this article in press as: Folseraas T et al. Extended analysis of a genome-wide association study in primary sclerosing cholangitis detects multiple novel risk loci. J Hepatol (2012), http://dx.doi.org/10.1016/j.jhep.2012.03.031

Research Article

Background & Aims: A limited number of genetic risk factors have been reported in primary sclerosing cholangitis (PSC). To discover further genetic susceptibility factors for PSC, we followed up on a second tier of single nucleotide polymorphisms (SNPs) from a genome-wide association study (GWAS).
Methods: We analyzed 45 SNPs in 1221 PSC cases and 3508 controls. The association results from the replication analysis and the original GWAS (715 PSC cases and 2962 controls) were combined in a meta-analysis comprising 1936 PSC cases and 6470 controls. We performed an analysis of bile microbial community composition in 39 PSC patients by 16S rRNA sequencing.
Results: Seventeen SNPs representing 12 distinct genetic loci achieved nominal significance ($p_{replication}$ <0.05) in the replication. The most robust novel association was detected at chromosome 1p36 (rs3748816; $p_{combined} = 2.1 \times 10^{-8}$) where the *MMEL1* and *TNFRSF14* genes represent potential disease genes. Eight additional novel loci showed suggestive evidence of association (p_{repl} <0.05). *FUT2* at chromosome 19q13 (rs602662; $p_{comb} = 1.9 \times 10^{-6}$, rs281377; $p_{comb} = 2.1 \times 10^{-6}$ and rs601338; $p_{comb} = 2.7 \times 10^{-6}$) is notable due to its implication in altered susceptibility to infectious agents. We found that *FUT2* secretor status and genotype defined by rs601338 significantly influence biliary microbial community composition in PSC patients.
Conclusions: We identify multiple new PSC risk loci by extended analysis of a PSC GWAS. *FUT2* genotype needs to be taken into account when assessing the influence of microbiota on biliary pathology in PSC.

Introduction

Primary sclerosing cholangitis (PSC) is a chronic cholestatic liver disease of unknown etiology, characterized by progressive inflammation and fibrosis of the bile ducts, leading to liver cirrhosis in many cases [1]. An important clinical feature of PSC is the frequent autoimmune manifestations in other organ systems, most commonly inflammatory bowel disease (IBD), which is reported in 62–83% of PSC patients of Northern European descent [1]. In addition, approximately 25% of PSC patients are affected with at least one autoimmune disease outside the liver and colon [2], most commonly, type 1 diabetes, thyroid disease, rheumatoid arthritis and psoriasis. Shared genetic susceptibility could potentially explain the frequent occurrence of these immune-related co-morbidities in PSC.

The importance of genetic risk factors in PSC is demonstrated by heritability studies estimating siblings of PSC patients to be 9–39 times more likely to develop PSC than the general population [1]. The contribution of genetic variants in the HLA complex on chromosome 6p21 to the risk of PSC is well established [1]. Previously, strong evidence for associated risk factors outside the HLA region has been reported at chromosome 3p21 in *MST1* and chromosome 2q13 near *BCL2L11* [3]. In addition, suggestive PSC associations have been reported at six additional loci harboring the likely susceptibility genes *GPBAR1*, *IL2RA*, *GPC5*/*GPC6*, *IL2*/*IL21*, *CARD9*, and *REL* [3–6]. In the most recent PSC GWAS [3], only the top 23 associated regions were selected for replication genotyping.

Replication attempts of promising, but lower ranked markers from GWAS have yielded valuable findings in several other diseases, and have been particularly fruitful when combined with a semi-hypothesis-driven approach taking into consideration gene content and potential biological relevance [7]. In an attempt to identify novel susceptibility loci in PSC, we selected a second tier of promising associated markers from an available PSC GWAS [3] for replication in an independent cohort.

Materials and methods

Study subjects

The discovery panel included a total of 715 PSC cases and 2962 healthy controls, with 332 PSC cases and 262 controls from Scandinavia, and 383 PSC cases and 2700 controls from Germany [3]. The replication panel consisted of a total of 1221 PSC cases and 3508 controls, with 289 PSC cases and 820 controls from Scandinavia, 561 PSC patients and 2063 controls from Central Europe, and 371 PSC cases and 625 controls from the United States (US).

The recruitment of study subjects is described in detail in the Supplementary Methods section of the Supplementary Material.

Written informed consent was obtained from all study participants. The study conforms to the ethical guidelines of the 1975 Declaration of Helsinki and was approved by the South-Eastern Norwegian Regional Ethics Committee (S-93178 and S-08872b).

Association analysis of genome-wide data

Genome-wide genotyping, imputation and quality control in the discovery panel are described in the Supplementary Methods. Association analysis of the genotyped and imputed SNPs in the discovery panel was performed using a logistic regression procedure implemented in the R statistical package version 2.9.1 (http://www.r-project.org). The six first principal components generated by the EIGENSTRAT software [8] after removal of population outliers were included as co-variates in the regression to correct for residual population structure.

Selection of SNPs for extended replication

A total of 2,466,182 SNPs were included in the association analysis of the GWAS dataset. All nominally associated SNPs (p value <0.05) outside the HLA complex (defined as position 25–35 million base pairs on chromosome 6p21), which had not been previously subjected to replication genotyping, were considered potential candidates for replication. We aimed at integrating a priori knowledge on disease mechanisms in PSC and genetic susceptibility factors for related conditions and traits by implementing two SNP prioritization strategies for replication genotyping.

Strategy I
All non-HLA SNPs with a GWAS p value (p_{GWAS}) <0.05 (n = 134,466) were aligned with SNPs reported to be associated in immune-mediated and chronic inflammatory diseases and relevant biochemical parameters as listed in the Catalog of Genome-Wide Association Studies (http://www.genome.gov/26525384) (accessed 23.04.2010). Supplementary Table 2 lists the phenotypes taken into consideration.

Strategy II
Genetic loci harboring SNPs with robust statistical evidence of association in the GWAS ($p_{GWAS} < 1.0 \times 10^{-4}$) underwent literature-based assessment taking potential pathogenetic mechanisms in PSC into account [1]. Before this assessment, the number of candidate SNPs with a $p_{GWAS} < 1.0 \times 10^{-4}$ was reduced by applying a linkage disequilibrium (LD) clumping threshold of $r^2 = 0.8$ and distance = 50 kb in PLINK version 1.06 [9] using LD measurements from the HapMap project.

To further avoid redundant genotyping, the total number of SNPs selected with strategies I and II was clumped using a LD clumping threshold of $r^2 = 0.8$ and distance = 2000 kb in PLINK version 1.06 [9]. In *FUT2*, two redundant coding SNPs (rs601338 and rs281377) were purposely retained. In addition, a tag (rs4143332, $r^2 = 1.0$, $D' = 1.0$) for the most strongly associated SNP (rs3134792) in *HLA-B* was included. If assay design was not possible for the index SNPs from the clumping, SNPs identified to be in LD with the SNP in the clumping procedure were used instead.

Please cite this article in press as: Folseraas T et al. Extended analysis of a genome-wide association study in primary sclerosing cholangitis detects multiple novel risk loci. J Hepatol (2012), http://dx.doi.org/10.1016/j.jhep.2012.03.031

JOURNAL OF HEPATOLOGY

Table 1. Clinical characteristics of the PSC patients included in the study.

	Genome-wide analysis		Replication analysis		
	Scandinavian	German	Scandinavian	Central Europe	United States
Number of PSC patients	332	383	289	561	371
Median age at diagnosis (range)	34 (10-76)	32 (9-76)	37 (12-74)	35 (2-81)	41 (8-77)
Male patients (%)	241 (73%)	259 (69%)	199 (70%)	354 (64%)	239 (64%)
IBD data available for (n) patients	331	364	249	544	329
IBD (%)	260 (79%)	245 (67%)	196 (79%)	347 (64%)	262 (80%)
IBD subtype					
Ulcerative colitis (%)	64%	51%	66%	47%	66%
Crohn's disease (%)	10%	10%	8%	13%	8%
IBD unclassified (%)	4%	7%	5%	4%	6%
CCA data available for (n) patients	330	375	243	461	329
CCA diagnosed (%)	35 (11%)	10 (3%)	27 (11%)	28 (6%)	27 (8%)

Clinical characteristics of the patients included in the study stratified according to study panel in the genome-wide analysis or the replication analysis.
IBD, inflammatory bowel disease; CCA, cholangiocarcinoma.

Association analysis of replication data and meta-analysis

Replication genotyping and quality control are described in the Supplementary Methods. Association analysis in the three replication panels was performed by the Cochran–Mantel–Haenszel (CMH) test as implemented in the PLINK version 1.06 [9]. Along with the CMH test, a Breslow-Day (BD) test for heterogeneity of odds ratios was performed. Meta-analysis of the summary statistics in the discovery and replication panels was performed using the Meta-Analysis Tool for genome-wide association scans, METAL (http://www.sph.umich.edu/csg/abecasis/Metal) (version released 2010-08-01) (see Supplementary Methods for details).

To adjust for multiple testing in the replication, a strict Bonferroni correction was applied assuming independence between the 45 SNPs analyzed (Bonferroni-adjusted p value threshold at 0.05/45 = 0.0011).

Association signal plots, linkage disequilibrium calculations, and pathway analysis

Regional association plots were generated from the GWAS data using the LocusZoom software (Supplementary Methods) [10]. For all the other LD calculations, data from the HapMap project and PLINK version 1.06 [9] were used. To examine the functional relationship among genomic PSC risk regions, we performed a Gene Relationships Across Implicated Loci (GRAIL) pathway analysis (http://www.broadinstitute.org/mpg/grail/) (Supplementary Methods).

Biliary FUT2 phenotyping, FUT2 sequence alignment, and profiling of biliary microbiota composition

To demonstrate presence of FUT2 in the bile duct epithelium, immunostaining for the α (1,2) fucose-specific lectin *Ulex europaeus* agglutinin-I was performed. The biliary FUT2 phenotyping is described in detail in Supplementary Methods. Sequence alignment of human *FUT2* with a range of different species was performed to evaluate evolutionary conservation (see figure legend of Supplementary Fig. 2). A fragment of the 16S rRNA gene spanning the V1 and V2 hypervariable regions was amplified from processed bile samples of PSC patients and sequenced using 454 GS-FLX Titanium sequencing chemistry as described [11] (Supplementary Methods).

Results

SNP selection and genotyping results

Seven hundred and fifteen PSC patients and 2962 controls were successfully genotyped and analyzed in the GWAS (Table 1). Among the 2,466,182 genotyped and imputed SNPs analyzed, a subset of 59 SNPs were selected for follow-up; 37 SNPs based on SNP selection strategy I and 22 SNPs based on SNP selection strategy II. Following quality pruning of the replication dataset, a total of 45 SNPs were included in the association analysis.

Association results

A total of 17 SNPs achieved nominal significance in the replication (p_{repl} <0.05) and demonstrated effect sizes in the same direction as observed in the discovery cohort (Table 2). Results for the additional 28 SNPs studied are provided in Supplementary Table 1. The associated SNPs implicate 12 genetic loci, nine of which had not been previously reported to be associated with PSC. The association results for SNPs with nominally significant replication with the same direction of effect were homogenous ($p_{HETEROGENEITY\ (HET)}$ >0.05) when the GWAS discovery and replication cohorts were compared, except for one marker (rs11936230). Within the three panels constituting the replication cohort, the significant association results were also largely homogenous, however, two markers (rs11682163 and rs281377) had a significant Breslow-Day test, implying heterogeneity of odds ratios in between the three panels for these SNPs. Four SNPs at *MMEL1*/*TNFRSF14*, *IL2*/*IL21*, and *CARD9* demonstrated association results robust to correction for multiple testing using Bonferroni's method (p <0.0011) in the replication analysis (Table 2).

The strongest finding in the replication panel was the missense SNP rs3748816 (M[A**T**G] → T[A**C**G]) located in *MMEL1*, which achieved genome-wide significance [12] in the combined analysis ($p_{comb} = 2.1 \times 10^{-8}$) (Table 2). The association signal at this locus (1p36) encompasses *MMEL1* and extends into the tumor necrosis factor receptor superfamily member 14 (*TNFRSF14*) gene (Fig. 1).

FUT2 associations and bile microbial community composition

Eight novel loci demonstrated suggestive evidence for association in the replication panel (p_{repl} <0.05), but were not robust to

Please cite this article in press as: Folseraas T et al. Extended analysis of a genome-wide association study in primary sclerosing cholangitis detects multiple novel risk loci. J Hepatol (2012), http://dx.doi.org/10.1016/j.jhep.2012.03.031

Table 2. Allele frequencies and results from the association analyses for SNPs with association results reaching nominal significance with consistent effect sizes in the replication.

					Genome-wide analysis				Replication analysis							
					Allele frequencies (Cases/Controls)				Allele frequencies (Cases/Controls)							
Chr	SNP	Position	Locus	AI	Scandinavia (332/262)	Germany (383/2700)	*p* value*	OR (95% CI)*	Scandinavia (289/820)	Central Europe (561/2063)	United States (371/625)	*p* value†	OR (95% CI)†	BD *p* value	Combined *p* value	Het *p* value	Selection‡
1	rs3748816	2,516,606	*TNFRSF14/MMEL1*	G/A	0.29/0.34	0.28/0.33	6.5×10^{-4}	0.79 (0.69-0.90)	0.31/0.35	0.30/0.34	0.27/0.35	8.7×10^{-6}	0.79 (0.71-0.88)	0.45	2.1×10^{-8}	0.92	I
4	rs13132245	123,269,042	*KIAA1109/IL2/IL21*	G/A	0.16/0.21	0.12/0.17	7.3×10^{-5}	0.69 (0.57-0.83)	0.13/0.18	0.12/0.14	0.12/0.15	3.0×10^{-4}	0.77 (0.67-0.89)	0.65	1.2×10^{-7}	0.40	II
4	rs13119723	123,437,763	*KIAA1109/IL2/IL21*	G/A	0.13/0.18	0.09/0.14	4.1×10^{-5}	0.65 (0.53-0.80)	0.15/0.19	0.13/0.15	0.12/0.16	4.0×10^{-4}	0.78 (0.68-0.90)	0.87	9.9×10^{-8}	0.32	II
9	rs4077515	138,386,317	*CARD9*	T/C	0.48/0.45	0.43/0.40	0.045	1.14 (1.00-1.31)	0.49/0.45	0.47/0.42	0.45/0.42	5.3×10^{-4}	1.18 (1.08-1.30)	0.95	7.6×10^{-5}	0.55	I
16	rs2903692	11,146,284	*CLEC16A*	A/G	0.28/0.33	0.31/0.36	0.0014	0.80 (0.92-0.70)	0.27/0.33	0.32/0.35	0.32/0.36	0.0012	0.85 (0.76-0.94)	0.43	6.4×10^{-6}	0.64	I
2	rs13017599	61,017,835	*REL*	A/G	0.40/0.35	0.40/0.37	0.027	1.16 (1.32-1.02)	0.40/0.36	0.42/0.38	0.38/0.37	0.0026	1.16 (1.05-1.28)	0.42	1.9×10^{-4}	0.88	I
21	rs11203203	42,709,255	*UBASH3A*	A/G	0.39/0.35	0.42/0.37	0.0025	1.22 (1.39-1.07)	0.39/0.34	0.39/0.36	0.40/0.36	0.0033	1.16 (1.05-1.28)	0.68	2.9×10^{-5}	0.60	I
19	rs602662	53,898,797	*FUT2*	A/G	0.56/0.50	0.53/0.46	7.7×10^{-5}	1.30 (1.49-1.14)	0.53/0.47	0.48/0.45	0.51/0.50	0.0034	1.15 (1.05-1.27)	0.29	1.9×10^{-6}	0.21	II
21	rs9976767	42,709,459	*UBASH3A*	G/A	0.48/0.46	0.48/0.44	0.028	1.15 (1.02-1.30)	0.45/0.43	0.47/0.44	0.49/0.43	0.0055	1.14 (1.04-1.26)	0.44	4.0×10^{-4}	0.99	I
19	rs601338	53,898,486	*FUT2*	A/G	0.50/0.44	0.47/0.40	4.5×10^{-5}	1.33 (1.52-1.16)	0.51/0.44	0.46/0.42	0.48/0.48	0.0061	1.14 (1.04-1.26)	0.062	2.7×10^{-6}	0.14	II
2	rs11682163	3,727,421	*ALLC*	C/T	0.49/0.39	0.49/0.44	7.9×10^{-5}	1.29 (1.47-1.14)	0.50/0.45	0.49/0.44	0.43/0.45	0.010	1.13 (1.03-1.24)	0.048	7.6×10^{-6}	0.14	II
19	rs281377	53,898,415	*FUT2*	T/C	0.43/0.49	0.46/0.52	9.3×10^{-6}	0.71 (0.61-0.83)	0.39/0.47	0.47/0.49	0.44/0.44	0.011	0.88 (0.80-0.97)	0.025	2.1×10^{-6}	0.059	II
22	rs5771069	48,777,607	*IL17REL*	A/G	0.48/0.54	0.49/0.53	0.0057	0.82 (0.95-0.72)	0.45/0.49	0.47/0.50	0.48/0.50	0.028	0.90 (0.82-0.99)	0.89	5.8×10^{-4}	0.43	I
4	rs11936230	123,236,205	*KIAA1109/IL2/IL21*	T/C	0.38/0.30	0.41/0.35	1.9×10^{-5}	1.33 (1.17-1.52)	0.38/0.34	0.38/0.36	0.38/0.36	0.039	1.11 (1.01-1.22)	0.79	1.9×10^{-5}	0.040	II
12	rs4149056	21,222,816	*SLCO1B1*	C/T	0.13/0.17	0.16/0.17	0.044	0.84 (1.00-0.70)	0.14/0.17	0.15/0.16	0.13/0.15	0.043	0.87 (0.76-1.00)	0.60	0.0045	0.76	I
22	rs4820599	23,320,213	*GGT1*	G/A	0.29/0.24	0.31/0.28	0.035	1.16 (1.01-1.33)	0.29/0.28	0.29/0.27	0.31/0.28	0.043	1.11 (1.00-1.23)	0.77	0.0038	0.69	I
12	rs3184504	110,368,991	*SH2B3*	T/C	0.54/0.52	0.57/0.51	0.0016	1.23 (1.08-1.40)	0.47/0.45	0.51/0.49	0.51/0.47	0.050	1.10 (1.00-1.21)	0.67	4.6×10^{-4}	0.21	I

Complete association results for SNPs reaching nominal significance with consistent effect sizes in the replication. For the genome-wide analysis, the allele frequencies were calculated based on allele dosages and are listed separately for the German and Scandinavian discovery panels. For the replication analysis, allele frequencies are given for all three panels making up the combined replication panel. The combined *p* values for the genome-wide analysis and the replication analysis were generated using the METAL software [34]. The genomic positions refer to NCBI's build 36. SNP selection strategies I and II indicate SNPs selected based on the two SNP selection strategies. The HLA association was confirmed for rs4143332 ($p_{repl} = 4.1 \times 10^{-77}$, $p_{comb} = 4.4 \times 10^{-123}$) (not listed in the table).

Chr, chromosome; AI, alleles; OR, odds ratio; CI, confidence interval; BD, Breslow-Day; Het, heterogeneity.

*Odds ratios and *p* values derived from logistic regressions of allele dosages including the six first principal components from the principal components analysis as covariates.

†*p* values generated by using the Cochrane–Mantel–Haenszel test.

‡SNP selection strategy.

Please cite this article in press as: Folseraas T et al. Extended analysis of a genome-wide association study in primary sclerosing cholangitis detects multiple novel risk loci. J Hepatol (2012), http://dx.doi.org/10.1016/j.jhep.2012.03.031

JOURNAL OF HEPATOLOGY

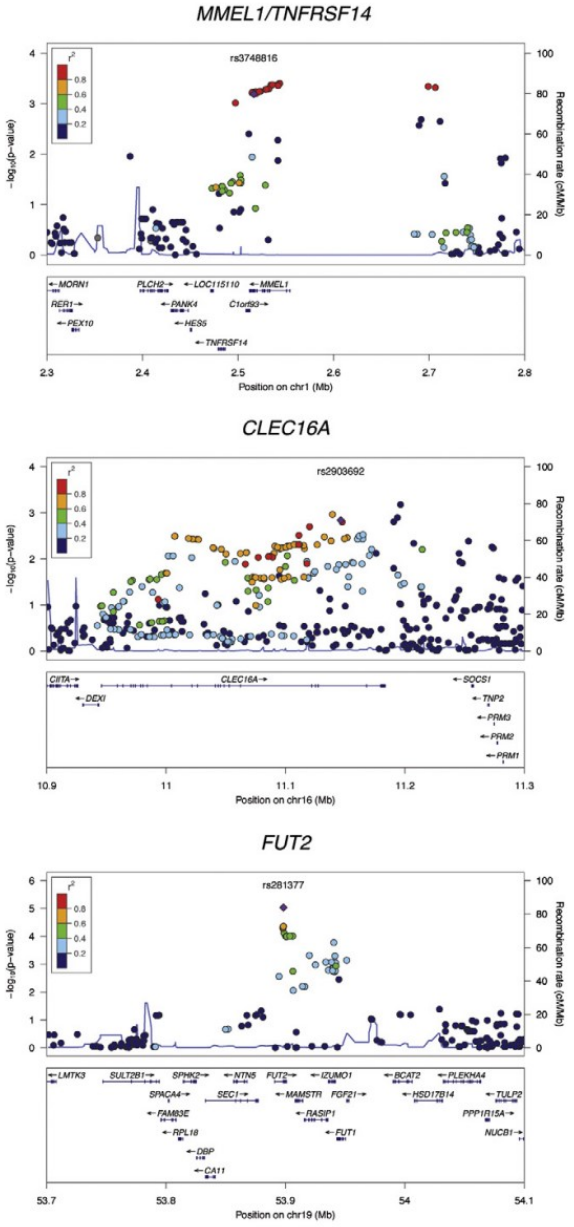


Fig. 1. Regional association plots for *MMEL1/TNFRSF14*, *CLEC16A* and *FUT2*. The association results for both the genotyped and imputed SNPs are represented by the $-\log_{10}$ *p* value plotted against the genomic position. The index SNP is indicated by a purple diamond while the colors of the remaining SNPs indicate the linkage disequilibrium with the index SNP. The recombination rates were derived from the HapMap project and are represented by the thin blue lines. The plots were generated using the LocusZoom software [10].

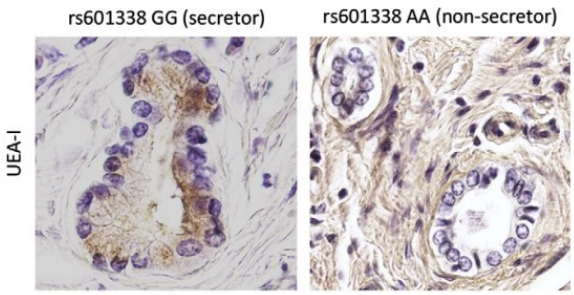


Fig. 2. Biliary FUT2 phenotyping. Lectin staining of the hilar liver biopsy specimens from PSC patients. Paraffin-embedded sections from individuals with AA (non-secretor) and GG (secretor) variants of the *FUT2* rs601338 SNP were used to evaluate the expression of α(1,2) fucosylated glycans in the bile duct epithelium. H antigen, detected with the α(1,2) fucose-specific lectin *Ulex europaeus* agglutinin-I (UEA-I) (brown staining), is expressed on the apical surface of the biliary epithelial layer of the secretor variant, while it is absent on non-secretor epithelia (Original magnification 400x).

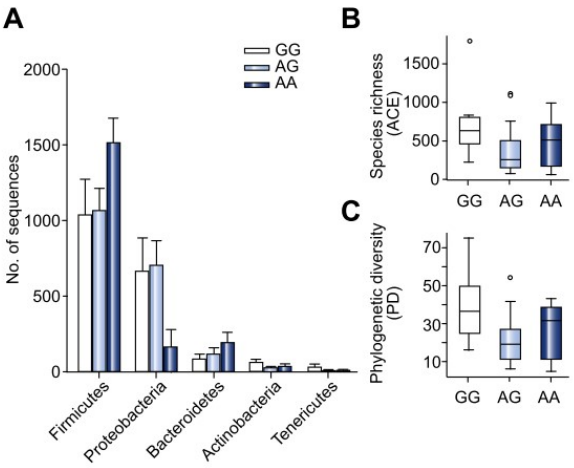


Fig. 3. Influence of *FUT2* genotype on phyla abundances and alpha diversity. (A) Mean abundances (±SE) of the major phyla with respect to genotype at the non-sense SNP rs601338 (W (**TGG**) → ∗(**TAG**). (B) The Abundance based Coverage Estimator (ACE) as a measure of the approximated species richness [35]. (C) Alpha diversity measured as the sum of the total branch length in each sample (i.e. Phylogenetic Diversity) [36]. All values are based on the normalized dataset (2000 reads per individual).

correction for multiple testing (Table 2). Of these, three SNPs in the Fucosyltransferase 2 gene (*FUT2*, Table 2 and Fig. 1) at 19q13 (a synonymous SNP, rs281377 ($p_{comb} = 2.1 \times 10^{-6}$), a non-sense SNP, rs601338 (W (**TGG**) → ∗(**TAG**)) ($p_{comb} = 2.7 \times 10^{-6}$) and a missense SNP, rs602662 (G (**GGT**) → S (**AGT**)) ($p_{comb} = 1.9 \times 10^{-6}$) are of particular interest, given the associations between *FUT2* genotype and several infectious diseases. Strong LD between these SNPs ($r^2_{rs601338\ vs.\ rs281377} = 0.84$, $r^2_{rs601338\ vs.\ rs602662} = 0.76$ and $r^2_{rs281377\ vs.\ rs602662} = 0.57$) implies that the putatively functional allele(s) cannot be directly determined by the association analysis. The rs601338 polymorphism generates a premature stop codon (W143X) and a truncated, dysfunctional FUT2 protein [13]. By immunostaining for the α(1,2) fucose-specific lectin *Ulex europaeus* agglutinin-I (Fig. 2), we show that this stop codon appears to have a profound effect on expression of α(1,2) fucosylated glycans in bile duct epithelium. The rs602662 polymorphism results in an amino acid substitution (G247S) which may be detrimental to

Please cite this article in press as: Folseraas T et al. Extended analysis of a genome-wide association study in primary sclerosing cholangitis detects multiple novel risk loci. J Hepatol (2012), http://dx.doi.org/10.1016/j.jhep.2012.03.031

Research Article

FUT2 function, as shown by strong evolutionary conservation of this residue (Supplementary Fig. 2).

To evaluate a potential functional role of FUT2 secretor status and genotype as defined by rs601338 in the pathophysiology of PSC, we analyzed the bacterial community composition of bile samples from 39 PSC patients (8 homozygous for the functional allele "G", 21 heterozygous and 10 homozygous for the loss-of-function allele "A" (W (T**G**G) → ∗(T**A**G)), since we previously observed genotype-dependent changes in community structure of the colon in the context of Crohn's disease [11]. Intriguingly, the abundances of several phyla displayed significant differences with respect to *FUT2* genotype and secretor status similar to those observed in the colon. In particular, the Firmicutes are significantly elevated and the Proteobacteria significantly decreased among non-secretors, in addition to differences observed in the Bacteroidetes, Actinobacteria, and Tenericutes (Fig. 3A and Supplementary Table 3). Similarly, alpha diversity measures display significant differences with respect to genotype similarly to what observed in the colon, i.e. a decrease of diversity in the heterozygous state compared to both homozygous genotypes (Fig. 3B and C and Supplementary Table 4). Finally, differences in inter-individual variability, *i.e.* beta diversity, are also apparent (*adonis*: unweighted UniFrac-$r^2 = 0.074$, $p = 0.055$ (genotype); normalized weighted UniFrac- $r^2 = 0.057$, $p = 0.049$ (secretor status)) (Supplementary Fig. 4A and B). Using the full species assemblage to test for differentiation with respect to *FUT2* genotype also revealed marginal results (Redundancy Analysis: $F = 1.140$, $p = 0.085$ (Supplementary Fig. 4C).

Additional suggestive associations

Among the other associated SNPs, rs2903692 ($p_{comb} = 6.4 \times 10^{-6}$) is located in the gene C-type lectin domain family 16, member A (*CLEC16A*, Table 2 and Fig. 1). Two SNPs, rs11203203 ($p_{comb} = 2.9 \times 10^{-5}$) and rs9976767 ($p_{comb} = 4.0 \times 10^{-4}$), localize in an intron of ubiquitin associated and SH3 domain containing A gene (*UBASH3A*, Table 2) and are in LD ($r^2 = 0.63$). The rs11682163 SNP ($p_{comb} = 7.6 \times 10^{-6}$) is localized in allantoicase (*ALLC*, Table 2) at 2p25. At 22q13, rs5771069 ($p_{comb} = 5.8 \times 10^{-4}$) is a missense SNP (L(C**T**T) → P (C**C**T)) in interleukin 17 receptor E-like (*IL17REL*, see Table 2). At 12p12, the rs4149056 SNP ($P_{comb} = 0.0045$) is a missense SNP (V(G**T**G)→ A (G**C**G) in an exon of the solute carrier organic anion transporter family gene (*SLCO1B1*). The rs4820599 SNP ($p_{comb} = 0.0038$) is localized in an intron of the gamma-glutamyltransferase 1 gene (GGT1) at 22q11. The rs3184504 SNP ($p_{comb} = 4.6 \times 10^{-4}$) represents a missense mutation (W (**T**GG) → R (**C**GG) in an exon of the SH2B adaptor protein 3 gene (*SH2B3*, Table 2) at 12q24.

The associated SNPs at 4q27 (rs13132245; $p_{comb} = 1.2 \times 10^{-7}$ and rs13119723; $p_{comb} = 9.9 \times 10^{-8}$) are localized 42 kb upstream of and in an intron of *KIAA1109*, respectively. Both rs13119723 and another genetic variant residing in this region, rs6822844, which is in LD with the replicated SNPs at rs13132245 ($r^2 = 0.85$) and rs13119723 ($r^2 = 0.66$), have been reported nominally associated with PSC in two candidate gene studies [5,6]. The replicated SNPs at 9q34 (rs4077515; $p_{comb} = 7.6 \times 10^{-5}$) in the caspase-recruitment domain family, member 9 gene (*CARD9*) and at 2p16 (rs13017599; $p_{comb} = 1.9 \times 10^{-4}$) near v-rel reticuloendotheliosis viral oncogene (*REL*) (Table 2 and Supplementary Fig. 1) also belong to loci which previously have been reported nominally associated with PSC [5]. A GRAIL analysis of previously published and hereby detected PSC risk loci demonstrated that several of the loci are functionally related (Supplementary Fig. 3A and B).

Discussion

In the largest PSC cohort presented, we performed replication genotyping and a combined analysis of 45 SNPs not followed up in a previous GWAS in PSC [3]. We identified one novel PSC risk locus with association results below the threshold for genome-wide significance along with suggestive evidence for replication of 8 novel additional loci. The PSC-associated *FUT2* variant was shown to significantly influence the bile microbial community composition in PSC patients.

The replicated SNP at 1p36 (*MMEL1-TNFRSF14*) represents the first genetic overlap demonstrated between PSC and primary biliary cirrhosis (PBC) [14]. Interestingly, the SNP demonstrates opposite effect sizes in PSC and PBC [14]. Considering the peak association signal of the observed non-synonymous SNP at *MMEL1* (Fig. 1), the suggestive presence of MMEL1 protein expression in bile duct cells and glandular cells of the gall bladder (http://www.proteinatlas.org) and the fundamental role of other membrane metallo-endopeptidase family members in processes of metabolism, *MMEL1* has a relevant candidate role at this locus. For PSC, *TNFRSF14* is also an intriguing candidate. The TNFRSF14 protein is expressed on T lymphocytes and the mucosal epithelium and acts as a receptor for the proinflammatory cytokine LIGHT [15]. TNFRSF14 signaling has also inhibitory effects on B and T cells, and this dual role is thought to regulate immune tolerance [16]. Mice lacking Btla ($Btla^{-/-}$mice), the mediator responsible for the inhibitory actions of TFNRSF14, spontaneously develop an autoimmune hepatitis-like disease accompanied by severe bile duct epithelium inflammation with nuclear pleomorphism and irregularity of duct outlines [17], paralleling features of histopathological liver biopsy specimens in human PSC [18].

FUT2 encodes an enzyme (Galactoside 2-alpha-L-fucosyltransferase 2) involved in protein glycosylation, including the ABH blood-antigen synthesis pathway [13]. Genetic variants giving rise to a non-functional, truncated FUT2 enzyme result in an inability to synthesize ABH antigens on mucosal surfaces and in salivary glands, which is referred to as non-secretor status [13]. *FUT2*-determined secretor status has been observed to affect susceptibility to a number of infectious agents [19], possibly by altering the recognition and binding of pathogen adhesins to their preferred carbohydrate receptors on mucosal surfaces [20]. Homozygosity for the PSC-associated non-sense variant rs601338 has been reported to account for >95% of the prevalence of the non-secretor status in Caucasians [21]. The non-sense SNP rs601338 is also a risk factor for Crohn's disease [19]. Although the bile duct is an environment distinct from the portions of the GI tract most frequently afflicted by Crohn's disease, we observe strikingly similar changes with respect to the abundance of major bacterial phyla and the level and pattern of bacterial diversity compared to previous observations in the colon [11]. Thus, these general patterns appear to be consistent between different habitats throughout the GI tract, which may underlie the overlapping association of *FUT2* to different chronic inflammatory disorders. Importantly, based on the present data, further enquiries into this topic need to take into account both biliary *FUT2* expression and the biliary microbiome in the contaminated bile

Please cite this article in press as: Folseraas T et al. Extended analysis of a genome-wide association study in primary sclerosing cholangitis detects multiple novel risk loci. J Hepatol (2012), http://dx.doi.org/10.1016/j.jhep.2012.03.031

JOURNAL OF HEPATOLOGY

Table 3. Associations in other autoimmune diseases for PSC loci with robust or suggestive association in the current study.

Locus	Candidate gene(s) and SNP(s) associated in the current study	Autoimmune phenotypes with a reported association at the locus	Associated SNP	References
1p36	*MMEL1-TNFRSF14*	Celiac disease	**rs3748816**	Dubois *et al.*, *Nat Genet* 42, 295 (2010)
	rs3748816	Multiple sclerosis	rs6684864	Blanco-Kelly *et al.*, *Genes Immun* 12, 145 (2011)
			rs3748816	Ban *et al.*, *Genes Immun* 11, 660 (2010)
		Primary biliary cirrhosis	**rs3748816**	Hirschfield *et al.*, *Nat Genet* 42, 655 (2010)
		Rheumatoid arthritis	rs3890745	Raychaudhuri *et al.*, *Nat Genet* 40, 1216 (2008)
		Ulcerative colitis	rs734999	Anderson *et al.*, *Nat Genet* 43, 246 (2011)
2p15	*REL*	Celiac disease	rs13003464	Dubois *et al.*, *Nat Genet* 42, 295 (2010)
	rs13017599	Crohn's disease	rs10181042	Franke *et al.*, *Nat Genet* 42, 1118 (2010)
		Psoriasis	rs702873	Strange *et al.*, *Nat Genet* 42, 985 (2010)
		Psoriatic arthritis	**rs13017599**	Ellinghaus *et al.*, *J Invest Dermatol* 132, 1133 (2012)
		Rheumatoid arthritis	**rs13017599**	Gregersen *et al.*, *Nat Genet* 41, 820 (2009)
			rs13031237	Stahl *et al.*, *Nat Genet* 42, 508 (2010)
		Ulcerative colitis	rs13003464	McGovern *et al.*, *Nat Genet* 42, 332 (2010)
2p25	*ALLC*	n.a.		
	ra11682163			
4q27	*IL2-IL21*	Celiac disease	**rs13119723**	van Heel *et al.*, *Nat Genet* 39, 827 (2007)
	rs13132245		rs6822844	Hunt *et al.*, *Nat Genet* 40, 395 (2008)
	rs13119723		rs13151961	Dubois *et al.*, *Nat Genet* 42, 295 (2010)
	rs11936230	Graves' disease	rs17388568	Todd *et al.*, *Nat Genet* 39, 857 (2007)
		Psoriatic arthritis	**rs13119723**	Liu *et al.*, *PLoS Genet* 4, e1000041 (2008)
			rs13151961	Liu *et al.*, *PLoS Genet* 4, e1000041 (2008)
		Rheumatoid arthritis	**rs13119723**	Stahl *et al.*, *Nat Genet* 42, 508 (2010)
			rs6822844	Maiti *et al.*, *Arthritis Rheum* 62, 323 (2010)
		Systemic lupus erythematosus	rs6822844	Maiti *et al.*, *Arthritis Rheum* 62, 323 (2010)
		Sjögren's syndrome	rs6822845	Maiti *et al.*, *Arthritis Rheum* 62, 323 (2010)
		Type 1 diabetes	rs4505848	Barrett *et al.*, *Nat Genet* 41, 703 (2009)
			rs6534347	WTCCC, *Nature* 447, 661 (2007)
		Ulcerative colitis	rs17388568	Anderson *et al.*, *Nat Genet* 43, 246 (2011)
9q35	*CARD9*	Ankylosing spondylitis	**rs4077515**	Burton *et al.*, *Nat Genet* 39, 1329 (2007)
	rs4077515		**rs4077515**	Pointon *et al.*, *Genes Immun* 11, 490 (2010)
			rs10781500	Evans *et al.*, *Nat Genet* 43, 761 (2011)
		Crohn's disease	**rs4077515**	Franke *et al.*, *Nat Genet* 42, 1118 (2010)
			rs10870077	Zhernakova *et al.*, *Am J Hum Genet* 82, 1202 (2008)
		Ulcerative colitis	**rs4077515**	McGovern *et al.*, *Nat Genet* 42, 332 (2010)
			rs10781500	Barrett *et al.*, *Nat Genet* 41, 1330 (2009)
			rs10781499	Anderson *et al.*, *Nat Genet* 43, 246 (2011)
			rs10870077	Zhernakova *et al.*, *Am J Hum Genet* 82, 1202 (2008)
12p12	*SLCO1B1*	n.a.		
	rs4149056			

(***continued on next page***)

Please cite this article in press as: Folseraas T et al. Extended analysis of a genome-wide association study in primary sclerosing cholangitis detects multiple novel risk loci. J Hepatol (2012), http://dx.doi.org/10.1016/j.jhep.2012.03.031

Research Article

Table 3 (*continued*)

Locus	Candidate gene(s) and SNP(s) associated in the current study	Autoimmune phenotypes with a reported association at the locus	Associated SNP	References
12q31	*SH2B3*	Celiac disease	rs653178	Hunt *et al.*, *Nat Genet* 40, 395 (2008)
	rs3184504		rs653178	Dubois *et al.*, *Nat Genet* 42, 295 (2010)
			rs653178	Zhernakova *et al.*, *PLoS Genet* 7, e1002004 (2011)
		Multiple sclerosis	**rs3184504**	Alcina *et al.*, *Genes Immun* 11, 439 (2010)
		Rheumatoid arthritis	**rs3184504**	Stahl *et al.*, *Nat Genet* 42, 508 (2010)
			rs653178	Zhernakova *et al.*, *PLoS Genet* 7, e1002004 (2011)
		Systemic lupus erythematosus	rs17696736	Gateva *et al.*, *Nat Genet* 41, 1228 (2009)
		Type 1 diabetes	**rs3184504**	Barrett *et al.*, *Nat Genet* 41, 703 (2009)
			rs17696736	WTCCC, *Nature* 7, 661 (2007)
16p14	*CLEC16A*	Celiac disease	rs12928822	Dubois *et al.*, *Nat Genet* 42, 295 (2010)
	rs2903692	Multiple sclerosis	rs11865121	De Jager *et al.*, *Nat Genet* 41, 776 (2009)
			rs2903692	Martinez *et al . Ann Rheum Dis* 69, 309 (2010)
		Primary biliary cirrhosis	rs12924729	Mells *et al.*, *Nat Genet* 43, 329 (2011)
		Rheumatoid arthritis	rs6498169	Martinez *et al.*, *Ann Rheum Dis* 69, 309 (2010)
		Systemic lupus erythematosus	rs12708716	Gateva *et al.*, *Nat Genet* 41, 1228 (2009)
		Type 1 diabetes	rs12708716	Cooper *et al.*, *Nat Genet* 40, 1399 (2008)
			rs2903692	Martinez *et al.*, *Ann Rheum Dis* 69, 309 (2010)
19q13	*FUT2*	Crohn's disease	rs504963	McGovern *et al.*, *Hum Mol Genet* 19, 3468 (2010)
	rs602662		**rs602662**	McGovern *et al.*, *Hum Mol Genet* 19, 3468 (2010)
	rs601338		**rs601338**	McGovern *et al.*, *Hum Mol Genet* 19, 3468 (2010)
	rs281377		rs281379	Franke *et al.*, *Nat Genet* 42, 1118 (2010)
22q11	*GGT1*	n.a.		
	rs4820599			
21q22	*UBASH3A*	Celiac disease	**rs11203203**	Zhernakova *et al.*, *PLoS Genet* 7, e1002004 (2011)
	rs11203203	Rheumatoid arthritis	**rs11203203**	Stahl *et al.*, *Nat Genet* 42, 508 (2010)
	rs9976767		**rs11203203**	Zhernakova *et al.*, *PLoS Genet* 7, e1002004 (2011)
		Type 1 diabetes	**rs9976767**	Grant *et al.*, *Diabetes* 58, 290 (2009)
			rs11203203	Barrett *et al.*, *Nat Genet* 41, 703 (2009)
		Vitiligo	**rs11203203**	Jin *et al.*, *N Engl J Med* 362, 1686 (2010)
22q13	*IL17REL*	Ulcerative colitis	**rs5771069**	Anderson *et al.*, *Nat Genet* 43, 246 (2011)
	rs5771069		**rs5771069**	Franke *et al.*, *Nat Genet* 42, 292 (2010)

All currently reported associations in autoimmune diseases for the nominally replicated candidate genes studied in the present paper are listed in the table. The index SNP from the current study is indicated in bold. In studies where multiple SNPs at the locus are reported associated to the respective disease, the SNP with the lowest *p* value is presented in the table. Significantly associated index SNPs from the current study are additionally presented in the table.
WTCCC, Wellcome Trust Case Control Consortium; n.a, no associations to autoimmune diseases reported for genetic variants in these genes.

ducts of these patients, as well as the potential effects of rs602662 genotype on FUT2 function [22,23].

Among the additional novel suggestive loci, the *CLEC16A* association was most robust. CLEC16A has a predicted C-type lectin domain structure [24] and C-type lectins serve as central players in the immune response by modulating toll receptor function [25]. A role for toll-like receptor signaling in PSC pathogenesis has been suggested [26], but whether CLEC16A is implicated can only be speculated. Like for *CLEC16A*, the suggestive associations observed at *SLCO1B1* and *GGT* might also involve immunological mechanisms since SLCO1B1 and GGT are involved in the hepatic uptake and interconversion of the inflammatory mediator leukotriene C4, respectively [27,28]. The associations at *SLCO1B1* and *GGT* can also reflect other roles of these genes as the PSC associated variants have been shown to alter blood levels of bilirubin and gamma glutamyltransferase, respectively, in previous GWAS [29,30].

The potential roles of the PSC associations at *IL2/IL21* and *REL/PUS10* that are strengthened by the current study have previously been extensively reviewed [31]. In a disease like PSC, in which no effective medical therapy is available, the *IL2* and *IL2RA* genetic findings could prove to be of unique translational value (e.g. an IL-2 monoclonal antibody (Daclizumab/Zenapax) is under clinical trials in other conditions). Like for *FUT2*, the associations observed at *CARD9* point towards the interaction of host genetics and microbiota. When stimulated by intracellular pathogens,

Please cite this article in press as: Folseraas T et al. Extended analysis of a genome-wide association study in primary sclerosing cholangitis detects multiple novel risk loci. J Hepatol (2012), http://dx.doi.org/10.1016/j.jhep.2012.03.031

JOURNAL OF HEPATOLOGY

CARD9 interacts with NOD2 and induces cytokine production or activates the NF-KB pathway [32].

It has been established that susceptibility loci in related conditions are typically implicated in several disorders [33]. What is increasingly evident, and strengthened by our study is that PSC belongs to the family of autoimmune diseases, and that the overlapping genes are not restricted to IBD susceptibility genes (Table 3). For a rare disease, a clear understanding of the overlap with other immune mediated diseases may provide therapeutic opportunities otherwise not within scope. There is also a strong need to expand on the genetic susceptibility to PSC in study populations recruited outside Northern Europe and the United States.

In conclusion, we identify multiple PSC risk loci involved in the regulation of the immune response and the present study demonstrates a considerable overlap in the genetic susceptibility with other autoimmune diseases not restricted to IBD. The *FUT2* association potentially represents other components of the pathogenesis, and the significant influence of *FUT2* genotype on biliary microbial composition demonstrated by our study should be taken into account when further investigating the influence of microbiota on biliary pathology.

Financial support

The study was supported by The Norwegian PSC Research Center (http://ous-research.no/nopsc/) and the German Ministry of Education and Research (BMBF) through the National Genome Research Network (NGFN) and the Integrated Research and Treatment Center – Transplantation (reference number: 01EO0802) and the PopGen biobank (http://www.popgen.de). The US part of the study was supported by the NIH (DK 84960). The project received infrastructure support through the Norwegian Functional Genomics Programme (FUGE) via the "CIGENE" platform (http://www.cigene.no/), the Research Computing Services at the University of Oslo and the DFG excellence cluster "Inflammation at Interfaces" (http://www.inflammation-at-interfaces.de/).

Conflict of interest

The authors who have taken part in this study declared that they do not have anything to disclose regarding funding or conflict of interest with respect to this manuscript.

Acknowledgements

The authors are indebted to all PSC patients and healthy controls for their participation. We thank Katja Cloppenborg-Schmidt, Ilona Urbach, Irene Pauselis, Tanja Wesse, Tanja Henke, Rainer Vogler, Hege Dahlen Sollid, Bente Woldseth, and Liv Wenche Torbjørnsen for expert technical help. Paul R. Berg and Silje Karoliussen are specially thanked for providing expert assistance in the genotyping process. We are grateful to Arthur Kaser and Michael Nothnagel for helpful discussions. Benedicte A. Lie and The Norwegian Bone Marrow Donor Registry at Oslo University Hospital, Rikshospitalet in Oslo are acknowledged for sharing the healthy Norwegian controls. Dr. Felix Braun, Dr. Wolfgang Kreisel, Dr. Thomas Berg and Dr. Rainer Günther are acknowledged for contributing German PSC patients.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jhep.2012.03.031.

References

[1] Karlsen TH, Schrumpf E, Boberg KM. Update on primary sclerosing cholangitis. Dig Liver Dis 2010;42:390–400.
[2] Saarinen S, Olerup O, Broome U. Increased frequency of autoimmune diseases in patients with primary sclerosing cholangitis. Am J Gastroenterol 2000;95:3195–3199.
[3] Melum E, Franke A, Schramm C, Weismuller TJ, Gotthardt DN, Offner FA, et al. Genome-wide association analysis in primary sclerosing cholangitis identifies two non-HLA susceptibility loci. Nat Genet 2011;43:17–19.
[4] Karlsen TH, Franke A, Melum E, Kaser A, Hov JR, Balschun T, et al. Genome-wide association analysis in primary sclerosing cholangitis. Gastroenterology 2010;138:1102–1111.
[5] Janse M, Lamberts LE, Franke L, Raychaudhuri S, Ellinghaus E, Muri Boberg K, et al. Three ulcerative colitis susceptibility loci are associated with primary sclerosing cholangitis and indicate a role for IL2, REL, and CARD9. Hepatology 2011;53:1977–1985.
[6] Stallhofer J, Denk GU, Glas J, Laubender RP, Goke B, Rust C, et al. Analysis of IL2/IL21 Gene Variants in Cholestatic Liver Diseases Reveals an Association with Primary Sclerosing Cholangitis. Digestion 2011;84:29–35.
[7] Trynka G, Zhernakova A, Romanos J, Franke L, Hunt KA, Turner G, et al. Coeliac disease-associated risk variants in TNFAIP3 and REL implicate altered NF-kappaB signalling. Gut 2009;58:1078–1083.
[8] Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38:904–909.
[9] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559–575.
[10] Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 2010;26:2336–2337.
[11] Rausch P, Rehman A, Kunzel S, Hasler R, Ott SJ, Schreiber S, et al. Colonic mucosa-associated microbiota is influenced by an interaction of Crohn disease and FUT2 (Secretor) genotype. Proc Natl Acad Sci U S A 2011;108:19030–19035.
[12] Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 2005;6:95–108.
[13] Kelly RJ, Rouquier S, Giorgi D, Lennon GG, Lowe JB. Sequence and expression of a candidate for the human Secretor blood group alpha(1,2)fucosyltransferase gene (FUT2). Homozygosity for an enzyme-inactivating nonsense mutation commonly correlates with the non-secretor phenotype. J Biol Chem 1995;270:4640–4649.
[14] Hirschfield GM, Liu X, Han Y, Gorlov IP, Lu Y, Xu C, et al. Variants at IRF5-TNPO3, 17q12-21 and MMEL1 are associated with primary biliary cirrhosis. Nat Genet 2010;42:655–657.
[15] Steinberg MW, Shui JW, Ware CF, Kronenberg M. Regulating the mucosal immune system: the contrasting roles of LIGHT, HVEM, and their various partners. Semin Immunopathol 2009;31:207–221.
[16] Sedy JR, Gavrieli M, Potter KG, Hurchla MA, Lindsley RC, Hildner K, et al. B and T lymphocyte attenuator regulates T cell activation through interaction with herpesvirus entry mediator. Nat Immunol 2005;6:90–98.
[17] Oya Y, Watanabe N, Owada T, Oki M, Hirose K, Suto A, et al. Development of autoimmune hepatitis-like disease and production of autoantibodies to nuclear antigens in mice lacking B and T lymphocyte attenuator. Arthritis Rheum 2008;58:2498–2510.
[18] Chapman RW, Arborgh BA, Rhodes JM, Summerfield JA, Dick R, Scheuer PJ, et al. Primary sclerosing cholangitis: a review of its clinical features, cholangiography, and hepatic histology. Gut 1980;21:870–877.
[19] McGovern DP, Jones MR, Taylor KD, Marciante K, Yan X, Dubinsky M, et al. Fucosyltransferase 2 (FUT2) non-secretor status is associated with Crohn's disease. Hum Mol Genet 2010;19:3468–3476.
[20] Mentis A, Blackwell CC, Weir DM, Spiliadis C, Dailianas A, Skandalis N. ABO blood group, secretor status and detection of Helicobacter pylori among patients with gastric or duodenal ulcers. Epidemiol Infect 1991;106:221–229.

Please cite this article in press as: Folseraas T et al. Extended analysis of a genome-wide association study in primary sclerosing cholangitis detects multiple novel risk loci. J Hepatol (2012), http://dx.doi.org/10.1016/j.jhep.2012.03.031

Research Article

[21] Svensson L, Petersson A, Henry SM. Secretor genotyping for A385T, G428A, C571T, C628T, 685delTGG, G849A, and other mutations from a single PCR. Transfusion (Paris) 2000;40:856–860.

[22] Serpa J, Mendes N, Reis CA, Santos Silva LF, Almeida R, Le Pendu J, et al. Two new FUT2 (fucosyltransferase 2 gene) missense polymorphisms, 739G– > A and 839T– > C, are partly responsible for non-secretor status in a Caucasian population from Northern Portugal. Biochem J 2004;383:469–474.

[23] Silva LM, Carvalho AS, Guillon P, Seixas S, Azevedo M, Almeida R, et al. Infection-associated FUT2 (Fucosyltransferase 2) genetic variation and impact on functionality assessed by in vivo studies. Glycoconj J 2010;27:61–68.

[24] Hakonarson H, Grant SF, Bradfield JP, Marchand L, Kim CE, Glessner JT, et al. A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. Nature 2007;448:591–594.

[25] Svajger U, Anderluh M, Jeras M, Obermajer N. C-type lectin DC-SIGN: an adhesion, signalling and antigen-uptake molecule that guides dendritic cells in immunity. Cell Signal 2010;22:1397–1405.

[26] Karrar A, Broome U, Sodergren T, Jaksch M, Bergquist A, Bjornstedt M, et al. Biliary epithelial cell antibodies link adaptive and innate immune responses in primary sclerosing cholangitis. Gastroenterology 2007;132:1504–1514.

[27] Jedlitschky G, Keppler D. Transport of leukotriene C4 and structurally related conjugates. Vitam Horm 2002;64:153–184.

[28] Mistry D, Stockley RA. Gamma-glutamyl transferase: the silent partner? Copd 2010;7:285–290.

[29] Johnson AD, Kavousi M, Smith AV, Chen MH, Dehghan A, Aspelund T, et al. Genome-wide association meta-analysis for total serum bilirubin levels. Hum Mol Genet 2009;18:2700–2710.

[30] Yuan X, Waterworth D, Perry JR, Lim N, Song K, Chambers JC, et al. Population-based genome-wide association studies reveal six loci influencing plasma levels of liver enzymes. Am J Hum Genet 2008;83:520–528.

[31] Karlsen TH, Kaser A. Deciphering the genetic predisposition to primary sclerosing cholangitis. Semin Liver Dis 2011;31:188–207.

[32] Hara H, Saito T. CARD9 versus CARMA1 in innate and adaptive immunity. Trends Immunol 2009;30:234–242.

[33] Zhernakova A, van Diemen CC, Wijmenga C. Detecting shared pathogenesis from the shared genetics of immune-related diseases. Nat Rev Genet 2009;10:43–55.

[34] Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26:2190–2191.

[35] Chao A. Estimating the population size for capture-recapture data with unequal catchability. Biometrics 1987;43:783–791.

[36] Faith DP. Conservation evaluation and phylogenetic diversity. Biol Conserv 1992;61:1–10.

Please cite this article in press as: Folseraas T et al. Extended analysis of a genome-wide association study in primary sclerosing cholangitis detects multiple novel risk loci. J Hepatol (2012), http://dx.doi.org/10.1016/j.jhep.2012.03.031

Supplementary material

Supplementary Methods

Recruitment of Study Subjects

The diagnosis of PSC was based on standard clinical, biochemical, cholangiographic and histological criteria [1] with exclusion of secondary causes of sclerosing cholangitis. The diagnoses of IBD and cholangiocarcinoma used for descriptive purposes were as far as possible based on endoscopic and histological examinations. The PSC patients in the Scandinavian discovery and replication panels were recruited on admission to Oslo University Hospital, Rikshospitalet (Oslo, Norway), Karolinska University Hospital Huddinge (Stockholm, Sweden) and Sahlgrenska University Hospital, (Gothenburg, Sweden). The German PSC patients in the discovery and replication panels were recruited on admission to the Grosshadern University Clinic (Munich, Germany) and the University Hospital of Heidelberg (Heidelberg Germany), or through the Northern German biobank popgen (http://www.popgen.de) for patients recruited at the University Medical Center Hamburg-Eppendorf (Hamburg, Germany), the Hannover Medical School (Hannover,Germany), the University Hospital of Mainz (Mainz, Germany), the Christian-Albrechts-University Hospital Kiel (Kiel, Germany), the University Hospital Freiburg (Freiburg, Germany) and the Charité University Hospital Berlin (Berlin, Germany). The PSC patients from Belgium and the Netherlands in the Central European replication panel were recruited via the University Hospital Leuven (Leuven, Belgium), the Academic Medical Center (Amsterdam, the Netherlands) and the University Medical Center Groningen (Groningen, the Netherlands). The PSC patients and controls in the US replication panel have been recruited via the PSC Resource of Genetic Risk, Environment and Synergy Studies (P.R.O.G.R.E.S.S.) (http://mayoresearch.mayo.edu/lazaridis_lab/genomics_of_psc.cfm)

Healthy controls for the Scandinavian discovery and replication panels were randomly selected from the Norwegian Bone Marrow Donor Registry. Healthy controls for the German discovery panel and the Central European replication panel were selected among blood donors recruited via the Northern German biobank popgen (http://www.popgen.de) and the Southern German population-based study KORA F4 [2]. In addition, volunteers recruited via the University Hospital Leuven (Leuven, Belgium) contributed to the healthy controls in the Central European replication panel.

Genome-Wide Genotyping, Imputation and Quality Control

Genome-wide single nucleotide polymorphism (SNP) genotyping in the discovery panel was performed using the Affymetrix® Genome-Wide Human SNP Array 6.0 (Affymetrix, Santa Clara, CA, USA). Extensive quality control of the genome-wide dataset has previously been described [3]. In brief, samples were excluded based on the following criteria; samples failing array specific quality measures, samples with a genotyping call-rate <95%, samples with a mismatch between gender recorded and gender calculated based on genotype data and samples representing duplicates, related individuals, heterozygosity outliers or ethnic outliers (identified by the EIGENSTRAT software [4]). SNPs were excluded if they had a call rate <95%, a minor allele frequency <1%, or had a genotype distribution deviating from Hardy-Weinberg-Equilibrium (HWE) in the healthy controls.

Imputation of non-genotyped SNPs was performed using MACH version 1.0.16 software [5] and the HapMap Release 22 (http://hapmap.ncbi.nlm.nih.gov) CEU (Centre d´Etude du Polymorphisme Humain (CEPH)) as the reference dataset. All imputed markers underwent the same stringent quality control procedures as genotyped SNPs along with the requirement of good imputation quality (defined as an r2 > 0.3).

Replication Genotyping and Quality Control

Replication genotyping was performed using the Sequenom MassARRAY® iPLEX® Gold system at the Centre for Integrative Genetics (Norwegian University of Life Sciences, Ås, Norway). The iPLEX® Gold method for allele assignment is based on detection of allele specific primer extension products of different masses dependent on the sequence analyzed, by matrix-assisted laser desorption ionization-time of flight mass spectrometry [6].

SNPs with a genotyping call-rate < 0.90 in the iPLEX® genotyping (n = 6; rs2117032, rs11218714, rs4077515, rs608418, rs11252885, rs1000113) were re-genotyped with TaqMan™ SNP Genotyping Assays (Applied Biosystems, Foster City, CA, USA) and analyzed with SDS 2.3 software (Applied Biosystems). In addition, SNPs with a suboptimal call-rate (defined as 0.90-0.95) in the iPLEX® genotyping showing nominally significant association in a preliminary analysis (P Cochran-Mantel-Haenszel (CMH) < 0.05) (n=3; rs5771069, rs3184504, rs601338) were re-genotyped with TaqMan™ to exclude genotyping artifacts. The remaining SNPs with a suboptimal call-rate 0.90-0.95 not showing nominally significant association in this preliminary association analysis (PCMH > 0.05), were not re-genotyped by TaqMan and excluded from the final association analysis (n=12; rs11574637,

rs6498169, rs684134, rs10516487, rs762421, rs2205960, rs11865121, rs12422102, rs7915365, rs7765733, rs10252204 and rs1054611).

In the final replication dataset, samples with a genotyping call-rate < 0.50 (n = 14) and SNPs with a genotyping call-rate < 0.95 (n = 1), minor allele frequency < 0.01 (n = 0) or significant deviation from HWE analyzed separately in the different panels in the healthy controls (PHWE <0.001) (n=1) were excluded from further analyses.

Meta-analysis

Meta-analysis of the summary statistics in the discovery and replication panels was performed using the Meta-Analysis Tool for genome-wide association scans, METAL (http://www.sph.umich.edu/csg/abecasis/Metal) (version released 2010-08-01).

The METAL algorithm calculates a *Z*-score for each marker reflecting the magnitude and direction of the reference allele along with a test for heterogeneity [7]. A Z-score and P-value is then calculated from the weighted average of the *Z*-statistics, where the weight is proportional to the square root of the sample size of the GWA and replication panels. To take account of differences in case-control ratios in the panels the effective sample size was used for weighting [7].

Association Signal Plots and Linkage Disequilibrium Calculations

Regional association plots were generated from the GWAS data using the LocusZoom software [8]. The LocusZoom software calculates LD information based on the HapMap CEU population, and integrates this with custom association results and data on recombination rates from the HapMap project. The different colors for the SNPs plotted indicate the LD with the index SNP. For regions with more than one SNP genotyped in the replication, the SNP with the lowest *P*-value in the discovery panel was used as the index SNP except for the IL2/IL21 region where rs13119723 was used since this SNP is more commonly reported in other conditions. For all other LD calculations data from the HapMap project and PLINK version 1.06 [9] were used.

Biliary FUT2 Phenotyping

To identify the presence of FUT2 in bile duct epithelium, formalin-fixed, paraffin-embedded hilar liver sections harvested at the time of transplantation in PSC patients were selected according to FUT2 genotype status (rs601338 *GG*, n=3, rs601338 *AA* n=1).

Lectin staining was used to identify the presence of *FUT2* determined α-1,2-fucosylation of the H antigen. Following deparaffination, rehydration and blocking of endogenous peroxidase activity with 3% H202 in methanol, sections were incubated with 10% BSA (bovine serum albumin) in PBS (phosphate-buffered saline) for 30 min at room temperature and incubated with the biotinylated lectin UEA-I (*Ulex europaeus*, Sigma Aldrich, St. Louis, MO, USA) for 1 hour. Subsequently, the specimens were washed in PBS and incubated with avidin-biotin complex (Vectastain Elite ABC kit, Vector Laboratories, Burlingame, CA, USA) for 30 min according to the manufacturer's recommendations. Staining was performed with DAB peroxidase substrate kit, 3,3´-diaminobenzidine (Vector Laboratories), followed by counterstaining with Harris haemotoxylin (Cell Path, Newton, United Kingdom). Negative control staining was performed by replacing lectin with PBS [10].

Bacterial DNA extraction and 16S rRNA gene pyrosequencing

Bile samples were collected during ERCP (n=74) or liver transplantation (N=14) from 76 PSC patients recruited at Oslo University Hospital, Rikshospitalet (Oslo, Norway). After stringent quality control of 16S rRNA sequence data, 37 ERCP and 2 transplantation samples were included in bacterial community analysis (see below). The *FUT2* genotypes of these patients were determined using Sequenom MassARRAY® iPLEX® Gold as described in the section "Replication Genotyping and Quality Control". To extract bacterial DNA from ERCP and transplant samples, approximately 200 µL of bile was centrifuged at 13300 for 100 minutes, the supernatant was removed, and the resulting pellet was extracted using the MolYsis Complete5 kit (Molzym©, Bremen, Germany) following the manufacturer's instructions. Samples were eluted with 35 µl PCR grade water. Negative controls mimicking the extraction procedure on PCR grade water were included to account for any potential environmental influence. The 16S rRNA gene was amplified and sequenced using 454 GS-FLX Titanium sequencing chemistry as described [11].

Sequence processing and quality control of 454 sequences

Raw sequences were trimmed and filtered considering only perfect matches to primer and MID sequences. Additional filtering was carried out via a 50 bp sliding window mean quality threshold of 35. To further reduce sequencing errors we required no ambiguous bases and a maximum homopolymer length of 10 bp. Further noise reduction was carried out by pre-clustering the sequences according to a threshold of 2 bp differences over the whole sequence length following the recommendations of Schloss et al. [12, 13]. Chimeric sequences were detected using the UCHIME algorithm (USEARCH 4.2.66) with the Greengenes gold database as a reference [14, 15]. The remaining sequences were aligned to the curated Greengenes database (release 2010) [16] using the k-mer alignment procedure PyNAST [17, 18]. OTU binning was performed via the UCLUST algorithm [14] with 97% and 99% identity used for further analysis. The resulting OTU table was screened by the “SourceTracker” algorithm, whereby negative extraction controls (N=4) were used as a training set [19]. Only samples displaying an estimated proportion of 10% contamination or less and a final minimal processed read number of 2000 (with the exception of a single individual with 1949) were included in further analyses. For analyses of bacterial diversity, random subsets of 2000 sequences per sample were generated to normalize the read distribution. The aligned sequences were used to cluster sequences into Operational Taxonomic Units (OTUs) using USEARCH 4.2.66 with extraction of representative sequences via average distance clustering for each OTU. Phylogenetic tree construction was carried out using FastTree v2.1 with a CAT substitution model and gamma correction on the selected sequences of each OTU [20]. Classification of sequences was performed using the Naïve Bayesian Classifier implemented in mothur (v1.23) with a 60% bootstrap threshold and the updated Greengenes reference taxonomy (10000 permutations) [12, 21].

Statistical analysis of microbiota differences

Comparisons of phyla abundances were performed using Kruskal-Wallis and post hoc Mann-Whitney U tests. For analysis of alpha diversity, the abundance based coverage estimator (ACE) and Phylogenetic Diversity (PD) were calculated in R [22-24]. In order to use linear models to assess the influence of genotype and secretor status, alpha diversity indices were square root transformed to meet the requirement of a normal distribution and models were selected by the Akaike Information Criterion (AIC). As a phylogenetic measure of beta diversity, FASTUniFrac was used to calculate the unweighted and normalized weighted UniFrac metric [25, 26]. Analysis of community composition based on beta diversity metrics was performed using the nonparametric matrix based analysis of variance “adonis” (Analysis

of dissimilarity) implemented in the “vegan” package for R with 105 permutations to assess significance [27, 28]. UniFrac distances were analyzed by Principal Coordinate Analysis [28]. Goodness of fit of the proposed clusters and their respective centroids was assessed with 105 permutations. Further testing of species clustering with respect to genotype was carried out by Redundancy Analysis (RDA) with a prior Hellinger transformation to linearize the OTU distribution [29, 30].

Gene Relationships Across Implicated Loci (GRAIL) Pathway Analysis

To examine the functional relationship among genomic PSC risk regions, we performed a GRAIL pathway analysis (http://www.broadinstitute.org/mpg/grail/) (Supplementary Figure 3A and 3B). The GRAIL software is a statistical tool that utilizes text mining of published abstracts in the PubMed database to identify and quantify functional similarity among genes within disease associated regions [31]. The GRAIL analysis gives a P_{text} score to each disease region, which is a statistical significance score that reflects the degree of relatedness among genes at different loci. GRAIL corrects its significance score for multiple testing by adjusting for the number of genes in the region.

In the GRAIL analysis we included six SNPs that have shown suggestive or robust association in previous studies in PSC (rs3134792; HLA-B, rs6720394; BCL2L11, rs12612347; GPBAR1, rs3197999; MST1, rs9524260, GPC5/6 and rs12412095; IL2RA), for the IL2RA SNP a tag (r2=1) was used, since the lead SNP rs10905718 was not in HapMap release 22 [3, 32]. Furthermore we added the 12 lead SNPs from the respective 12 suggestive or robust susceptibility loci identified in the present study (Table 2). The GRAIL analyses were performed with the following settings: HapMap release 22/ HG18, HapMap population CEU, default settings for SNP rs Number submission and all SNPs as query and seed.

We performed two analyses, one based on PubMed abstracts published up to April 2011 (Supplementary Figure 3B) and one based on only those PubMed abstracts published prior to December 2006 (Supplementary Figure 3A). The first approach reflects the current state of literature information on biological mechanisms underlying genetic associations, but will potentially also be confounded the high number of GWAS data published in PubMed after 2006. The second approach avoids the majority of GWAS publications and the evaluation of gene relationships based on listing of multiple genes in regions identified as associated to the phenotypes under study in these GWAS. To visualize the results from the GRAIL analysis the VIZ-GRAIL software (http://www.broadinstitute.org/mpg/grail/vizgrail.html) [33] with default parameters was used.

Supplementary Table 1. Allele frequencies and results from the association analyses for SNPs with association results not reaching nominal significance with consistent effect sizes in the replication.

					Genome-wide analysis				Replication analysis						
					Allele frequencies (Cases/Controls)				Allele frequencies (Cases/Controls)						
Chr	**SNP**	**Position**	**Locus**	**Alleles**	**Scandinavia (332/262)**	**Germany (383/2700)**	***P*-value***	**OR (95% CI)***	**Scandinavia (289/820)**	**Central Europe (561/2063)**	**United States (371/625)**	***P*-value†**	**OR (95% CI)†**	**BD P-value**	**SNP selection strategy**
1	rs2377570	31,071,356	*SDC3/SNORD85*	A/G	0.26/0.19	0.27/0.22	6.3E-05	1.34 (1.55-1.16)	0.23/0.20	0.20/0.21	0.22/0.22	0.82	1.01 (0.90-1.14)	0.29	II
1	rs12144426	164,418,658	*FAM78B*	T/A	0.24/0.31	0.21/0.25	3.5E-05	0.72 (0.62-0.84)	0.25/0.27	0.24/0.24	0.26/0.24	0.94	1.00 (0.89-1.11)	0.48	II
2	rs1990760	162,832,297	*IFIH1*	C/T	0.39/0.38	0.44/0.39	0.0070	1.21 (1.39-1.05)	0.37/0.37	0.40/0.38	0.41/0.41	0.72	1.02 (0.92-1.12)	0.76	I
3	rs7638558	60,020,841	*FHIT*	C/T	0.02/0.00	0.04/0.02	3.6E-05	2.35 (3.52-1.57)	0.01/0.00	0.02/0.01	0.01/0.02	0.83	1.04 (0.69-1.57)	0.24	II
3	rs983513	79,277,847	*ROBO1*	G/A	0.45/0.53	0.39/0.45	2.5E-05	0.75 (0.65-0.86)	0.48/0.49	0.48/0.46	0.49/0.48	0.30	1.05 (0.96-1.16)	0.55	II
4	rs17005387	123,156,648	*KIAA1109/IL2/IL21*	A/G	0.02/0.01	0.03/0.02	2.6E-05	2.41 (3.63-1.60)	0.01/0.02	0.02/0.02	0.03/0.02	0.34	1.19 (0.84-1.67)	0.26	II
4	rs993704	125,931,510	*ANKRD50*	A/G	0.30/0.24	0.36/0.30	2.3E-05	1.35 (1.54-1.17)	0.26/0.26	0.32/0.30	0.27/0.27	0.29	1.06 (0.95-1.18)	0.78	II
4	rs10857102	125,937,433	*ANKRD50*	A/G	0.26/0.20	0.30/0.25	3.2E-05	1.36 (1.57-1.18)	0.25/0.25	0.31/0.29	0.26/0.26	0.23	1.07 (0.96-1.19)	0.78	II
5	rs1000113	150,220,26	*IRGM*	T/C	0.07/0.08	0.05/0.07	0.015	0.72 (0.56-0.94)	0.07/0.07	0.07/0.07	0.08/0.08	0.74	1.03 (0.86-	0.86	I

		9											1.24)		
5	rs11747270	150,239,060	*IRGM*	G/A	0.07/0.08	0.05/0.08	0.021	0.74 (0.58-0.96)	0.07/0.07	0.07/0.07	0.08/0.08	0.81	1.02 (0.85-1.23)	0.88	I
6	rs394683	5,043,412	*LYRM4*	C/T	0.26/0.31	0.24/0.30	2.2E-05	0.73 (0.85-0.63)	0.26/0.29	0.30/0.28	0.26/0.30	0.37	0.95 (0.86-1.06)	0.036	II
6	rs4713859	35,514,131	*PPARD/FANCE*	C/T	0.02/0.03	0.02/0.05	5.3E-05	0.43 (0.65-0.29)	0.03/0.03	0.05/0.05	0.05/0.06	0.81	1.03 (0.82-1.29)	0.67	II
7	rs590099	18,302,863	*HDAC9*	A/G	0.16/0.13	0.21/0.16	2.6E-05	1.42 (1.67-1.20)	0.13/0.17	0.16/0.17	0.17/0.16	0.13	0.90 (0.79-1.03)	0.16	II
7	rs17806432	76,798,820	*PION/FGL2*	T/C	0.14/0.11	0.15/0.11	1.2E-05	1.59 (1.29-1.95)	0.13/0.10	0.10/0.11	0.11/0.08	0.27	1.09 (0.94-1.27)	0.11	II
7	rs6979188	76,821,451	*PION/FGL2*	T/C	0.10/0.06	0.10/0.06	2.1E-05	1.64 (1.30-2.05)	0.07/0.07	0.07/0.07	0.08/0.06	0.49	1.07 (0.89-1.28)	0.68	II
8	rs10156297	127,621,008	*FAM84B*	C/A	0.16/0.13	0.18/0.12	6.8E-05	1.42 (1.20-1.70)	0.13/0.12	0.12/0.13	0.13/0.12	0.86	0.99 (0.86-1.14)	0.23	II
11	rs11218714	122,000,841	*UBASH3B*	G/A	0.30/0.35	0.27/0.34	3.2E-05	0.75 (0.65-0.86)	0.32/0.31	0.31/0.31	0.33/0.31	0.77	1.02 (0.92-1.12)	0.61	II
11	rs722449	132,656,146	*OPCML*	A/G	0.06/0.03	0.08/0.04	8.8E-06	1.82 (2.36-1.40)	0.04/0.05	0.05/0.04	0.04/0.05	0.74	1.04 (0.83-1.30)	0.45	II
12	rs3764021	9,724,895	*CLEC2D*	T/C	0.41/0.45	0.45/0.47	0.032	0.87 (0.77-0.99)	0.45/0.46	0.47/0.47	0.44/0.48	0.36	0.96 (0.87-1.05)	0.38	I
12	rs608418	10,023,132	*CLEC12A*	T/C	0.57/0.47	0.55/0.49	1.4E-05	1.32 (1.17-1.50)	0.52/0.50	0.49/0.48	0.50/0.50	0.49	1.03 (0.94-1.14)	0.90	II
12	rs2117032	20,965,389	*SLCO1B3*	C/T	0.35/0.31	0.38/0.35	0.045	1.15 (1.31-1.00)	0.35/0.35	0.34/0.36	0.39/0.38	0.73	0.98 (0.89-1.08)	0.54	I
13	rs9576711	38,521,621	*STOML3*	T/G	0.11/0.15	0.06/0.10	2.3E-05	0.60 (0.48-0.76)	0.11/0.13	0.12/0.08	0.09/0.11	0.39	1.07 (0.92-1.24)	2.6E-04	II
13	rs1413040	38,605,939	*STOML3*	A/G	0.26/0.34	0.22/0.26	9.8E-05	0.74 (0.86-0.64)	0.30/0.30	0.33/0.27	0.28/0.30	0.041	1.11 (1.00-	0.027	II

													1.23)		
15	rs289404	83,367,321	*PDE8A*	G/T	0.15/0.21	0.14/0.18	6.5E-05	0.69 (0.83-0.57)	0.19/0.20	0.21/0.23	0.24/0.25	0.078	0.90 (0.81-1.01)	0.70	II
16	rs9888739	31,220,754	*ITGAM*	T/C	0.11/0.13	0.10/0.12	0.026	0.78 (0.63-0.97)	0.09/0.10	0.10/0.11	0.12/0.12	0.61	0.96 (0.82-1.12)	0.88	I
16	rs7190071	71,742,579	*ZFHX3*	T/C	0.39/0.30	0.35/0.30	6.9E-06	1.39 (1.20-1.60)	0.34/0.37	0.33/0.36	0.34/0.35	0.026	0.89 (0.81-0.99)	0.62	II
18	rs4310957	69,756,844	*FBXO15*	G/T	0.32/0.24	0.31/0.27	9.0E-06	1.42 (1.65-1.21)	0.29/0.32	0.28/0.28	0.31/0.30	0.81	0.99 (0.89-1.10)	0.56	II
20	rs6080774	17,600,784	*RRBP1*	A/G	0.04/0.03	0.06/0.04	6.1E-05	2.20 (3.24-1.50)	0.02/0.02	0.03/0.02	0.02/0.02	0.077	1.32 (0.97-1.81)	0.70	II

Complete association results for SNPs not reaching nominal significance with consistent effect sizes in the replication. For the genome-wide analysis, the allele frequencies were calculated based on allele dosages and are listed separately for the German and Scandinavian discovery panels. For the replication analysis, allele frequencies are given for all three panels making up the combined replication panel. The genomic positions refer to NCBI's build 36. SNP selection strategy I and II indicate SNPs selected based on alignment of SNPs with a $P_{GWAS} < 0.05$ (n=134,466) with SNPs reported associated to immune-mediated and chronic inflammatory diseases and traits as listed in the Catalog of Genome-Wide Association Studies (http://www.genome.gov/26525384) (accessed 23.04.2010) [34] and strong evidence of association in the GWAS, as indicated by a GWAS *P*- value of (P_{GWAS}) < 1.0×10^{-4}, in loci considered to be relevant to potential pathogenetic mechanisms in PSC, respectively.

Chr, chromosome; OR, Odds Ratio; CI, Confidence Interval; BD, Breslow-Day

*Odds ratios and P-values derived from logistic regressions of allele dosages including the six first principal components from the principal components analysis as covariates [4].

†*P*-values generated by using the Cochrane-Mantel-Haenszel test [35].

Supplementary Table 2. Genome-wide association studies on immune-mediated and chronic inflammatory traits taken into consideration in SNP prioritization strategy 1.

Disease/Trait	First Author	Journal	PubMed ID
AIDS	Le Clerc	*J Infect Dis*	19754311
AIDS (progression)	Limou	*J Infect Dis*	19115949
Ankylosing spondylitis	The Australo-Anglo-American Spondyloarthritis Consortium (TASC)	*Nat Genet*	20062062
Anti-cyclic Citrullinated Peptide Antibody	Cui	*Mol Med*	19287509
Arthritis (juvenile idiopathic)	Hinks	*Arthritis Rheum*	19116933
	Behrens	*Arthritis Rheum*	18576341
Asthma	Sleiman	*N Engl J Med*	20032318
	Mathias	*J Allergy Clin Immunol*	19910028
	Himes	*Am J Hum Genet*	19426955
	Moffatt	*Nature*	17611496
	Li	*J Allergy Clin Immunol*	20159242
Asthma (childhood onset)	Hancock	*PLoS Genet*	19714205
Asthma (toluene diisocyanate-induced)	Kim	*Clin Exp Allergy*	19187332
Atopic dermatitis	Esparza-Gordillo	*Nat Genet*	19349984
Atopy	Castro-Giner	*BMC Med Genet*	19961619
Behcet's disease	Fei	*Arthritis Res Ther*	19442274
Bilirubin levels	Sanna	*Hum Mol Genet*	19419973
Biochemical measures	Zemunik	*Croat Med J*	19260141
Celiac disease	Hunt	*Nat Genet*	18311140
	van Heel	*Nat Genet*	17558408
	Dubois	*Nat Genet*	20190752
Chronic Hepatitis C infection	Rauch	*Gastroenterology*	20060832
Chronic Obstructive Pulmonary Disease	Pillai	*PLoS Genet*	19300482
	Cho	*Nat Genet*	20173748

C-reactive protein	Elliott	*JAMA*	19567438
	Reiner	*Am J Hum Genet*	18439552
	Ridker	*Am J Hum Genet*	18439548
Crohn's disease	Barrett	*Nat Genet*	18587394
	Raelson	*Proc Natl Acad Sci USA*	17804789
	Franke	*PLoS ONE*	17684544
	WTCCC	*Nature*	17554300
	Parkes	*Nat Genet*	17554261
	Libioulle	*PLoS Genet*	17447842
Crohn's disease and Sarcoidosis (combined)	Franke	*Gastroenterology*	18723019
Cystic fibrosis severity	Gu	*Nature*	19242412
Diabetic nephropathy	Pezzolesi	*Diabetes*	19252134
Drug-induced liver injury (flucloxacillin)	Daly	*Nat Genet*	19483685
Eosinophilic esophagitis (pediatric)	Rothenberg	*Nat Genet*	20208534
Gallstones	Buch	*Nat Genet*	17632509
Hematological and biochemical traits	Kamatani	*Nat Genet*	20139978
Hepatitis B	Kamatani	*Nat Genet*	19349983
HIV-1 control	Fellay	*PLoS Genet*	20041166
HIV-1 viral setpoint	Fellay	*Science*	17641165
Idiopathic pulmonary fibrosis	Mushiroda	*J Med Genet*	18835860
Inflammatory bowel disease	Kugathasan	*Nat Genet*	18758464
	Duerr	*Science*	17068223
Inflammatory bowel disease (early onset)	Imielinski	*Nat Genet*	19915574
Kawasaki disease	Burgner	*PLoS Genet*	19132087
Knee osteoarthritis	Nakajima	*PLoS One*	20305777
Leprosy	Zhang	*N Engl J Med*	20018961
Lupus	Cervino	*Ann NY Acad Sci*	17911428
Malaria	Jallow	*Nat Genet*	19465909
Multiple Sclerosis	Bahlo	*Nat Genet*	19525955

	De Jager	*Nat Genet*	19525953
	Baranzini	*Hum Mol Genet*	19010793
	Aulchenko	*Nat Genet*	18997785
	Comabella	*PLoS ONE*	18941528
	Hafler	*N Engl J Med*	17660530
	Jakkula	*Am J Hum Genet*	20159113
Multiple Sclerosis (age of onset)	Baranzini	*Hum Mol Genet*	19010793
Multiple Sclerosis (severity)	Baranzini	*Hum Mol Genet*	19010793
Neuromyeltis optica	Kim	*Neurobiol Dis*	19850125
Neutrophil count	Okada	*Hum Mol Genet*	20172861
Osteoarthritis	Zhai	*J Med Genet*	19508968
	Kerkhof	*Arthritis Rheum*	20112360
Periodontitis	Schaefer	*Hum Mol Genet*	19897590
Plasma levels of liver enzymes	Yuan	*Am J Hum Genet*	18940312
Primary biliary chirrosis	Hirschfield	*N Engl J Med*	19458352
Psoriasis	Nair	*Nat Genet*	19169254
	Zhang	*Nat Genet*	19169255
	Liu	*PLoS Genet*	18369459
	Capon	*Hum Mol Genet*	18364390
Rheumatoid arthritis	Gregersen	*Nat Genet*	19503088
	Raychaudhuri	*Nat Genet*	18794853
	Julia	*Arthritis Rheum*	18668548
	Plenge	*Nat Genet*	17982456
	Plenge	*N Engl J Med*	17804836
	WTCCC	*Nature*	17554300
Sarcoidosis	Hofmann	*Nat Genet*	18690218
Serum bilirubin levels	Johnson	*Hum Mol Genet*	19414484
Serum soluble E-selectin	Paterson	*Arterioscler Thromb Vasc Biol*	19729612
Soluble leptin receptor levels	Sun	*Hum Mol Genet*	20167575
Soluble levels of adhesion molecules	Barbalic	*Hum Mol Genet*	20167578
Systemic lupus erythematosus	Han	*Nat Genet*	19838193

	Graham	*Nat Genet*	18677312
	Harley	*Nat Genet*	18204446
	Hom	*N Engl J Med*	18204098
	Kozyrev	*Nat Genet*	18204447
	Yang	*PLoS Genet*	20169177
Systemic sclerosis	Zhou	*Arthritis Rheum*	19950302
Type 1 diabetes	Wallace	*Nat Genet*	19966805
	Barrett	*Nat Genet*	19430480
	Cooper	*Nat Genet*	18978792
	Grant	*Diabetes*	18840781
	Hakonarson	*Diabetes*	18198356
	Hakonarson	*Nature*	17632545
	Todd	*Nat Genet*	17554260
Ulcerative colitis	Asano	*Nat Genet*	19915573
	Barrett	*Nat Genet*	19915572
	Silverberg	*Nat Genet*	19122664
	Franke A	*Nat Genet*	20228798
	McGovern	*Nat Genet*	20228799
	Franke	*Nat Genet*	18836448
Vitiligo	Birlea	*J Invest Dermatol*	19890347

The table lists all the genome-wide association studies on immune-mediated and chronic inflammatory traits that were taken into consideration in SNP prioritization strategy 1. The studies were identified using the Catalog of Genome-Wide Association Studies (http://www.genome.gov/26525384) (accessed 23.04.2010) [34].

Supplementary Table 3. Evaluation of abundance patterns in the major phyla via Kruskal-Wallis and *post hoc* Mann-Whitney U tests.

					post hoc test (MWU)		
Phylum	Factor	*df*	X^2	*P*-Value	Factor*	*W*	*P*-Value
Firmicutes	Secretor status	1	3,725	0.054	NA[†]		
Proteobacteria	Secretor status	1	7,303	0.007	NA		
Proteobacteria	Genotype	2	7,323	0.026	AA - AG	49	0.019
					GG - AA	12	0.015
					GG - AG	85	0.981
Bacteroidetes	Secretor status	1	0,648	0.421	NA		
Actinobacteria	Genotype	2	6,344	0.042	AA - AG	120	0.539
					GG - AA	25	0.197
					GG - AG	30	0.009
Tenericutes	Genotype	2	7,023	0.030	AA - AG	143.5	0.092
					GG - AA	25	0.193
					GG - AG	38.5	0.021

*Genotype at the SNP rs601338 with G being the functional allele.

[†]Test not applied.

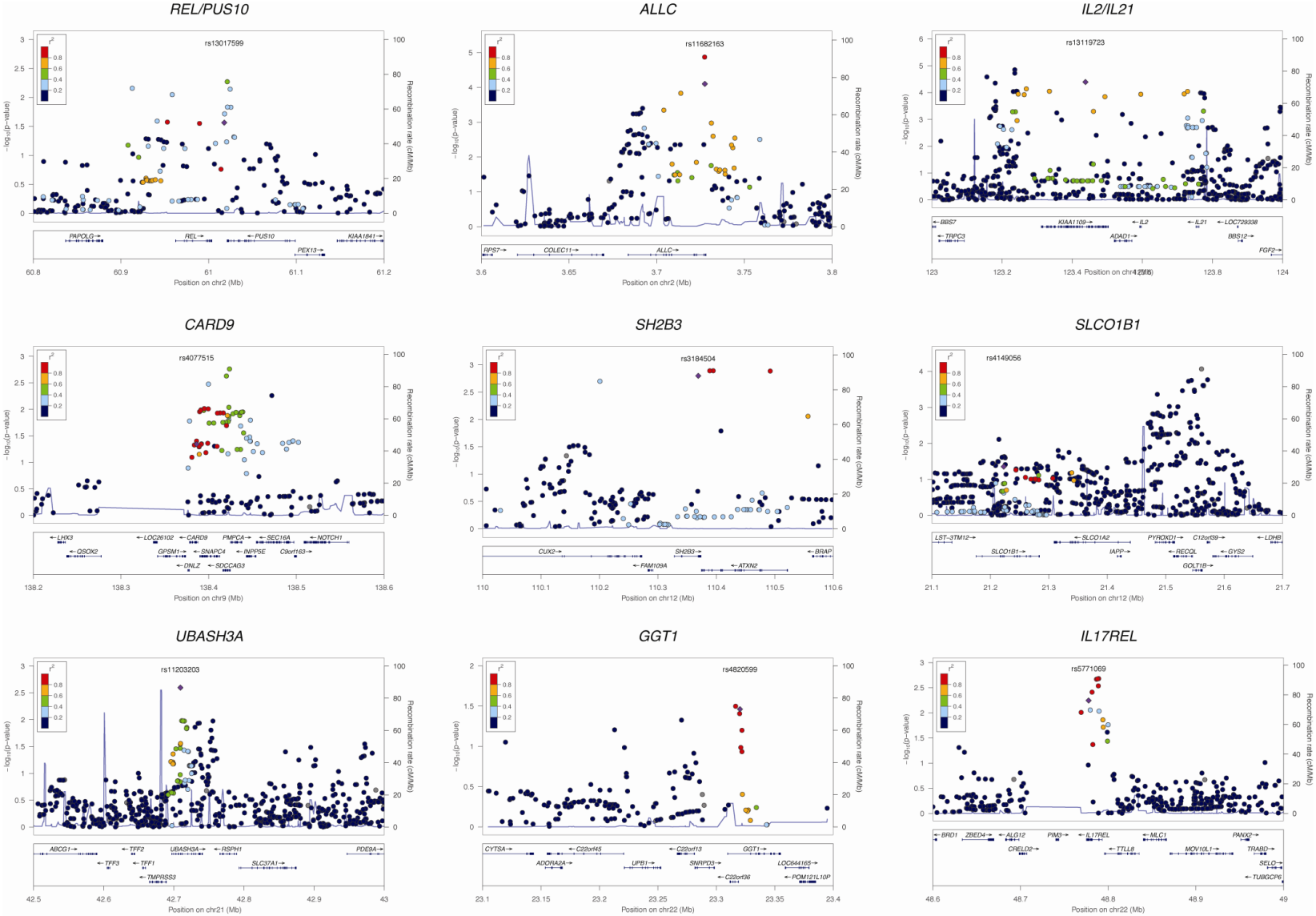
MWU, Mann-Whitney U

Supplementary Table 4. Analysis of alpha diversity via linear modeling.

					post hoc Tukey-HSD	
Alpha diversity	Factor	*df*	*F*-Value	*P*-Value	Factor*	*P*-Value
ACE	Genotype	2	3.015	0.062	AA - AG	0.657
					GG - AA	0.334
					GG - AG	0.049
Phylogenetic Diversity	Genotype	2	3.930	0.029	AA - AG	0.553
					GG - AA	0.259
					GG - AG	0.022

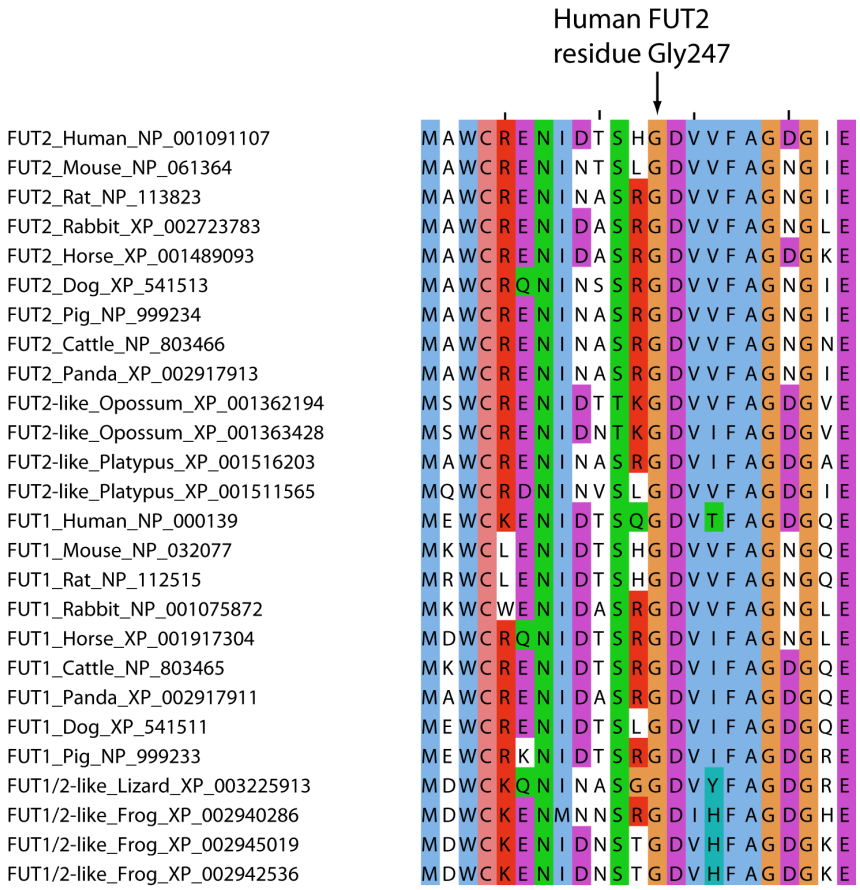
*Genotype at the SNP rs601338 with G being the functional allele.

Supplementary Figure 1. Regional association plots for additional nominally replicated regions



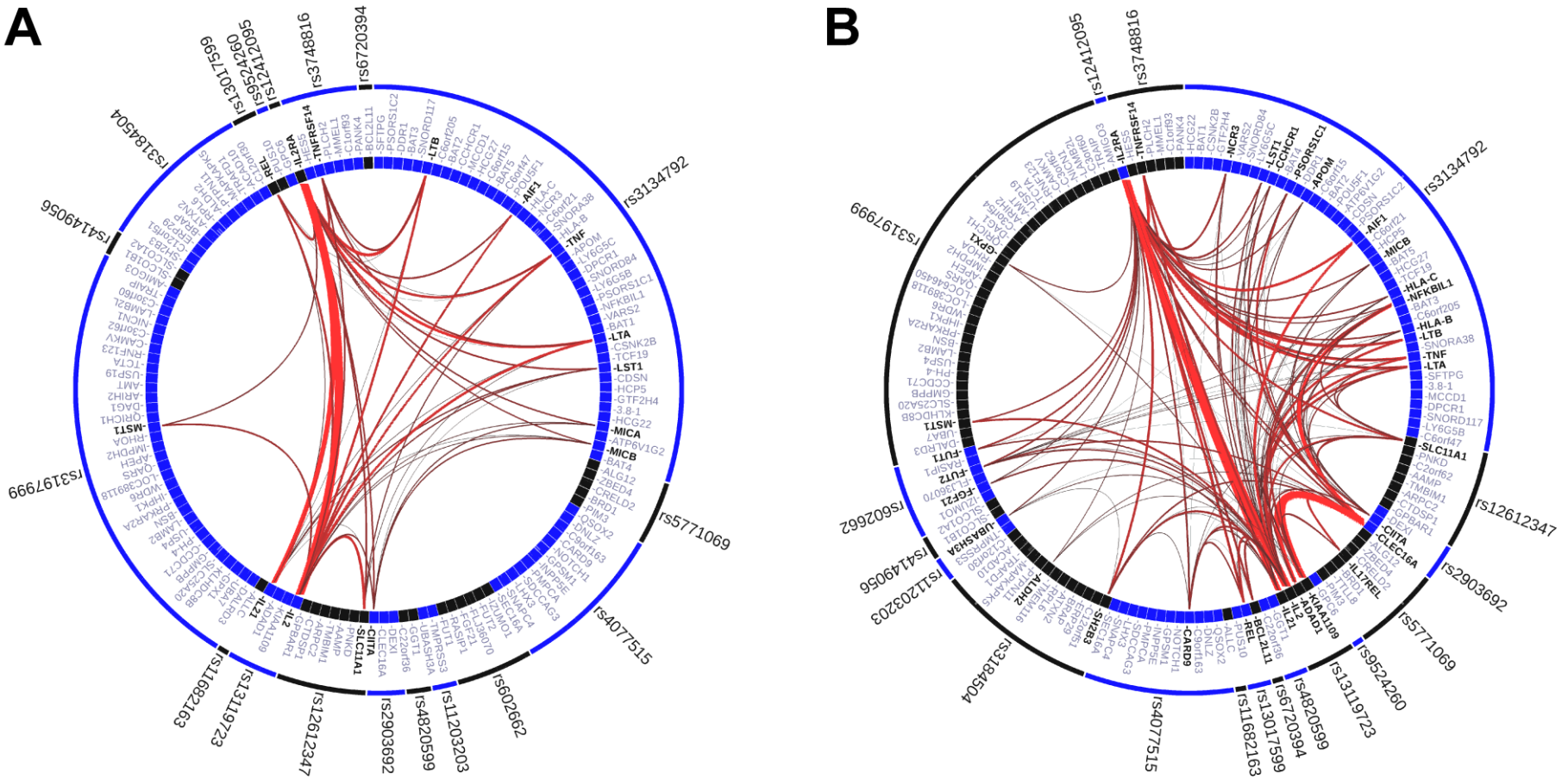
Association results from the genetic regions with nominally replicated SNPs not shown in Figure 1 in the main manuscript. The plots were generated using the LocusZoom software [8]. The association results for both the genotyped and imputed SNPs are represented by the –log10 *P*-value plotted against the genomic position. The index SNP is marked out with a purple diamond while the colors of the remaining SNPs indicated the linkage disequilibrium (LD) with the index SNP. The recombination rates were derived from the HapMap project and are represented by the thin blue lines.

Supplementary Figure 2. Sequence analysis for a segment of human *FUT2* containing Gly247.



A multiple sequence alignment of residues 235-257 of human *FUT2* and homologous sequences from tetrapods shows Gly247 to be evolutionary conserved in both *FUT2* and *FUT1* in mammals, lizard and frogs, strongly suggesting functional importance. The sequences were obtained from the RefSeq protein sequence database [36]. Fold recognition modeling with Phyre [37] indicates that the Bradyrhizobium *NodZ* fucosyltransferase [38] is the closest homolog of *FUT2* with a known 3D structure and that *FUT2* Gly247 is localized in the active site of the enzyme, in the second loop of the conserved β-α-β glycogen phosphorylase/glycosyltransferase (*GPGTF*) motif described by Wrabl and Grishin [39].

Supplementary Figure 3. GRAIL pathway analysis



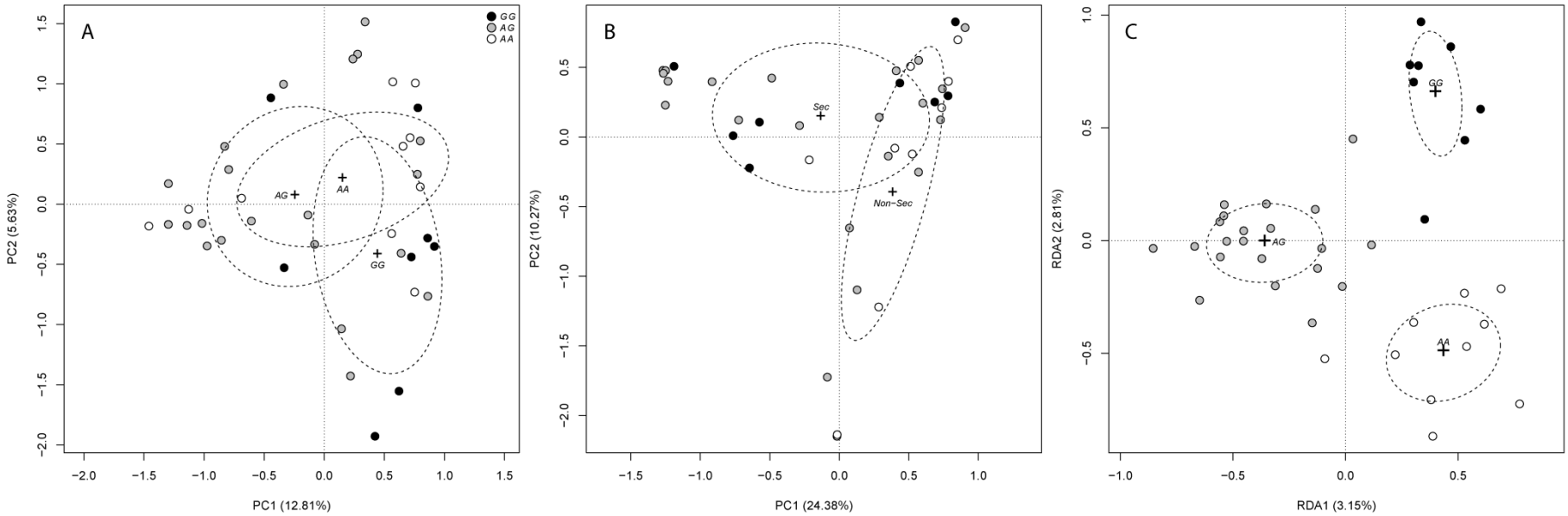
The figures demonstrate GRAIL [31] output results visualized with the VIZ-GRAIL software [33]. Outer circle boxes represent lead SNPs previously or currently identified as suggestive or robust PSC susceptibility loci used as input for the GRAIL analysis (see Supplementary Methods). Inner circle represents genes and genomic regions defined by the outer circle lead SNPs that were identified in the GRAIL analysis based on LD

characteristics, genes scored to a P_{text} < 0.05 in the GRAIL analysis are indicated in bold The lines between the inner circle genes represent functionally related genes within different loci, the thickness of lines is proportional to the relative similarity of the genes connected by the lines and inversely proportional to the number of genes within the loci that the genes are derived from.

(A) GRAIL pathway analysis based on PubMed abstracts published prior to December 2006.

(B) GRAIL pathway analysis based on PubMed abstracts published prior to April 2011.

Supplementary Figure 4. Influence of *FUT2* genotype on beta diversity



(A) PCoAs of the unweighted UniFrac metric based on presence/absence of phylogenetic branches and (B) normalized weighted UniFrac incorporating the abundances of phylogenetic branches. Centroids were positioned and evaluated using an iterative approach and clusters are denoted by the standard deviation of the weighted averages (dashed ellipses) around the centroids (unweighted UniFrac: r^2=0.110, P=0.074; normalized weighted UniFrac: r^2=0.104, P=0.016). (C) Community relationship (99% OTUs) in an environment spanned only by FUT2 genotypes explains 5.96% of the total variation in the bacterial species distribution (Redundancy Analysis: F=1.140, P=0.085; RDA1: F=1.206, P=0.093; RDA2: F=1.075, P=0.266).

References

[1] Chapman RW, Arborgh BA, Rhodes JM, Summerfield JA, Dick R, Scheuer PJ, et al. Primary sclerosing cholangitis: a review of its clinical features, cholangiography, and hepatic histology. Gut 1980;21:870-877.

[2] Wichmann HE, Gieger C, Illig T. KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. Gesundheitswesen 2005;67 Suppl 1:S26-30.

[3] Melum E, Franke A, Schramm C, Weismuller TJ, Gotthardt DN, Offner FA, et al. Genome-wide association analysis in primary sclerosing cholangitis identifies two non-HLA susceptibility loci. Nat Genet 2011;43:17-19.

[4] Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38:904-909.

[5] Li Y, Abecasis GR. Mach: 1.0: rapid haplotype reconstruction and missing genotype inference. Am J Hum Genet 2006;S79:2290.

[6] Storm N, Darnhofer-Patel B, van den Boom D, Rodi CP. MALDI-TOF mass spectrometry-based SNP genotyping. Methods Mol Biol 2003;212:241-262.

[7] Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26:2190-2191.

[8] Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 2010;26:2336-2337.

[9] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559-575.

[10] Fujitani N, Liu Y, Okamura T, Kimura H. Distribution of H type 1-4 chains of the ABO(H) system in different cell types of human respiratory epithelium. J Histochem Cytochem 2000;48:1649-1656.

[11] Rausch P, Rehman A, Kunzel S, Hasler R, Ott SJ, Schreiber S, et al. Colonic mucosa-associated microbiota is influenced by an interaction of Crohn disease and FUT2 (Secretor) genotype. Proc Natl Acad Sci U S A 2011;108:19030-19035.

[12] Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 2009;75:7537-7541.

[13] Schloss PD, Gevers D, Westcott SL. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. PloS one 2011;6:e27310.

[14] Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 2010;26:2460-2461.

[15] Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 2011;27:2194-2200.

[16] DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 2006;72:5069-5072.

[17] DeSantis TZ, Jr., Hugenholtz P, Keller K, Brodie EL, Larsen N, Piceno YM, et al. NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. Nucleic Acids Res 2006;34:W394-399.

[18] Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics 2010;26:266-267.

[19] Knights D, Kuczynski J, Charlson ES, Zaneveld J, Mozer MC, Collman RG, et al. Bayesian community-wide culture-independent microbial source tracking. Nat Methods 2011;8:761-763.

[20] Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for large alignments. PloS one 2010;5:e9490.

[21] Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol 2007;73:5261-5267.

[22] Chao A. Estimating the population size for capture-recapture data with unequal catchability. Biometrics 1987;43:783-791.

[23] Faith DP. Conservation Evaluation and Phylogenetic Diversity. Biol Conserv 1992;61:1-10.

[24] Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, et al. Picante: R tools for integrating phylogenies and ecology. Bioinformatics 2010;26:1463-1464.

[25] Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol 2005;71:8228-8235.

[26] Hamady M, Lozupone C, Knight R. Fast UniFrac: facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data. ISME J 2010;4:17-27.

[27] Anderson MJ. A new method for non-parametric multivariate analysis of variance. Austral Ecol 2001;26:32-46.

[28] Oksanen J, Blanchet FG, Kindt R, Legendre P, O'Hara RB, Simpson GL. Community Ecology Package. 1.17-6 ed: http://CRAN.R-project.org. 2011.

[29] Legendre P, Gallagher ED. Ecologically meaningful transformations for ordination of species data. Oecologia 2001;129:271-280.

[30] Legendre P. Studying beta diversity: ecological variation partitioning by multiple regression and canonical analysis. J Plant Ecol-Uk 2008;1:3-8.

[31] Raychaudhuri S, Plenge RM, Rossin EJ, Ng AC, Purcell SM, Sklar P, et al. Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. PLoS Genet 2009;5:e1000534.

[32] Karlsen TH, Franke A, Melum E, Kaser A, Hov JR, Balschun T, et al. Genome-wide association analysis in primary sclerosing cholangitis. Gastroenterology 2010;138:1102-1111.

[33] Raychaudhuri S. VIZ-GRAIL: visualizing functional connections across disease loci. Bioinformatics 2011;27:1589-1590.

[34] Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A 2009;106:9362-9367.

[35] Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 1959;22:719-748.

[36] Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 2011;39:D38-51.

[37] Kelley LA, Sternberg MJ. Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc 2009;4:363-371.

[38] Brzezinski K, Stepkowski T, Panjikar S, Bujacz G, Jaskolski M. High-resolution structure of NodZ fucosyltransferase involved in the biosynthesis of the nodulation factor. Acta Biochim Pol 2007;54:537-549.

[39] Wrabl JO, Grishin NV. Homology between O-linked GlcNAc transferases and proteins of the glycogen phosphorylase superfamily. J Mol Biol 2001;314:365-374.

Chapter III: Dependence of microbial community development on the host and maternal *α-1,2-fucosyltransferase* gene

An important host factor influencing the structure and diversity of microbial communities is the glycan composition on the mucosal surfaces [274, 376, 397]. These sugar chains represent a major source of nutrients and attachment sites for resident bacteria [395], but also a target for numerous pathogen species [230, 275-278] and commensals [337, 338]. The importance of those molecules for host-microbe interaction is further stressed by the direct initiation of glycosyltransferase expression through resident microbial communities [308, 398].

As ecological communities are under constant change and highly dependent on initial colonization conditions, their dynamics over space and time must be considered [399]. Species colonizing habitats early are often characterized by a high colonization capability and fast reproduction, but relatively low competitiveness against following specialists [400]. Thus, over the course of succession (community assembly), the composition and functional repertoire of a community changes/converges on a high degree of specialization and stability [66]. The primary inoculation by the mother via passage through the urogenital tract and contact to other surfaces in the earliest life stages represents the first and potentially very important factor determining the long-term composition and functionality of these microbial communities [401], which may also be influenced by the parental and juvenile genotype. The importance of parental transmission has been shown to some extent by the community similarity of siblings compared to unrelated individuals in the study by Turnbaugh *et al* 2006 [158]. Noteworthy is that mono- and dizygotic human twins do not show a strong difference between their communities, which emphasizes the importance of shared environmental factors during early upbringing and a lesser role of genetic similarity [104, 158]. The long lasting effects of differences in early microbial colonization are also well represented in the long lasting differences in community profiles between children born vaginally and through caesarian section [121].

The well-known *α-1,2-fucosyltransferase* encoded by the gene *FUT2* is responsible for the presence of ABH blood group antigens in bodily secretion and shows widespread variation in human populations [220, 261-263]. Several loss-of-function-mutations have been identified in the human population, which either subvert glycan fucosylation entirely (so called “nonsecretor”) [262] or drastically decrease enzyme effectivity [402]. These changes in the mucosal glycan repertoire, *i.e.* lack of ABH blood group antigens, have been linked to variation in susceptibility to several infectious diseases [230, 275-278] and changes in the microbial communities [376, 378, 397] and being a recently established risk factor for

inflammatory bowel diseases [209, 397]. Furthermore, trans-generational effects on Campylobacter and Norovirus susceptibility have been described for *FUT2* in humans [230, 238]. But how ecological forces in the microbial communities, such as initial colonization, succession, and host genetics interact, has yet to be established. Current efforts in understanding the etiology of chronic inflammatory disorders focus on gene-by-environment interactions, whereby disease manifests in genetically susceptible individuals after environmental disturbance, which may be either mediated and/or amplified through the intestinal microbial community [403]. Here, we focus on the influence of *Fut2* genotype and initial microbial colonization on the microbiome over time to identify how important the legacy of bacterial communities on their development and complexity is.

Publications:

Rausch P, Künzel S, Rosenstiel P, Baines J (unsubmitted manuscript). "Dependence of microbial community development on the host and maternal *α-1,2-fucosyltransferase* gene".

Dependence of microbial community development on the host and maternal *α-1,2-fucosyltransferase* gene

Philipp Rausch[1,2], Sven Künzel[3], Philip Rosenstiel[4], John F. Baines[1,2,*]

[1] Department of Evolutionary Genomics, Max Planck Institute for Evolutionary Biology, Plön, Germany

[2] Institute for Experimental Medicine, Christian-Albrechts-University of Kiel, Kiel, Germany

[3] Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Biology, Plön, Germany

[4] Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany

* Corresponding author (baines@evolbio.mpg.de)

Abstract

The *FUT2* gene encodes an *α-1,2-fucosyltransferase* responsible for the expression of ABO histo-blood group antigens on mucosal surfaces and bodily secretions. Individuals bearing at least one functional allele are known as “secretors”, whereas those homozygous for loss-of-function mutations, which seem to be maintained by strong selective pressures, display a “nonsecretor” phenotype. A large body of evidence suggests this polymorphism is maintained by numerous trade-offs surrounding host-microbe interactions. Further, nonsecretor individuals are more susceptible to Crohn Disease, which may be mediated by alterations in the microbiome. Here, we investigated the dynamics of microbial communities with respect to genotype using a *Fut2*-deficient mouse model, taking initial colonization and community assembly into account. We found strong differences in community assembly, diversity, and composition of microbial communities over time depending on the *Fut2* genotype of individual mice and their parents/grandparents. Communities were differentiated by *Fut2* genotype early in host development, although these differences faded over time. In contrast, the influence of community complexity appears to increase with time, with the highest diversity in *Fut2* $^{+/+}$ mice. Thus, during the process of colonization, we identified patterns of community specialization and stabilization that are influenced by host genotype.

Introduction

Microbial communities represent a complex phenotype composed of a multitude of bacterial taxa and functions with a strong influence on host fitness. Bacteria provide basic functions for the host such as increased reabsorption, breakdown and generation of nutrients [1, 2], immune regulation [3-6], pathogen resistance [7] and developmental cues for the gut [8]. Changes in community composition have been linked to adverse health effects such as obesity [9, 10], diabetes [11, 12], and Crohn Disease [13, 14], making it a potential target for their treatment and prevention. How the host's genetic makeup influences bacterial assemblages and the functions they provide is a subject of intensive research [15-19].

As ecological communities are under constant change and highly dependent on initial conditions, their dynamic changes over space and time must be considered [20]. Species colonizing habitats early are often characterized by a high colonization capability and fast reproduction, but relatively low competitiveness against subsequent specialists [21]. Thus, over the course of succession (community assembly), the composition and functional repertoire of a community changes dramatically and converges on a high degree of specialization and stability [22-24]. The primary inoculation by the mother via passage through the birth canal in the earliest life stages represents the first and potentially very important colonization step that determines long-term composition and functionality of host-associated microbial communities [25, 26]. The importance of parental transmission has been shown to some extent by the community similarity of siblings (monozygotic/dizygotic) compared to unrelated individuals in the study by Turnbaugh et al. 2006 [10] and recently more extensively by Goodrich et al. [19]. Noteworthy is that human twin pairs do not show a strong difference between their communities, which emphasizes the importance of shared environmental factors during early upbringing, but also the definite role of broad genetic similarity as seen in monozygotic twins [10, 19, 27].

An important host factor that influences the structure and diversity of microbial communities is the glycan composition on the mucosal surfaces [14, 28, 29]. These sugar chains represent a major source of nutrients and attachment sites for resident bacteria [30], but also a target for numerous pathogen species [31-35] and commensals [36, 37]. The initiation of glycosyltransferase expression by the resident microbial community therefore only further emphasizes the importance of those molecules for host-microbe interaction [38, 39].

The well-known *α-1,2-fucosyltransferase* encoded by the gene *FUT2* is responsible for the presence of ABH blood group antigens in bodily secretion and shows widespread variation in human populations [40-43]. Several highly conserved loss-of-function mutations

have been identified in the human population, which either subvert glycan fucosylation entirely (so called “nonsecretor”) [41] or drastically decrease enzyme effectivity [44]. These changes in the mucosal glycan repertoire, *i.e.* the lack of ABH blood group antigens, have been linked to variation in disease susceptibility [31-35], changes in the microbial communities [14, 29, 45], and a recently established risk factor for inflammatory bowel diseases [29, 46]. Furthermore, transgenerational effects on *Campylobacter* and *Norovirus* susceptibility have been described for *FUT2* in humans and mice, mainly described as interference of glycosylated milk components with the infective agents [33, 47, 48]. It also appears that fucose and mucin expression via *Il-22* might also play a direct or indirect role in the clearance of helminth infections [49] and bacterial infections [50-52].

How ecological forces in the microbial communities such as initial colonization, succession, and host genetics interact, has yet to be established. Current efforts in understanding the etiology of chronic inflammatory disorders focus on gene-by-environment interactions, whereby disease manifests in genetically susceptible individuals after environmental disturbance. This effect may be either mediated or amplified through changes in the intestinal microbial community, by changing the composition and function further escalating these imbalances [53, 54]. Recent studies indicate a role of *Fut2* expression in the maintenance and restoration of the microbial community and tissue integrity via fucose supplementation upon immune activation [50-52, 55, 56]. Here, we focus on the influence of *Fut2* genotype on the microbiome in the context of microbial colonization during development.

Results

Throughout the study, analyses performed with respect to genotype status are noted accordingly: *Fut2* $^{+/+}$, *Fut2* $^{+/-}$, *Fut2* $^{-/-}$; which account for differences in gene dose. Analyses performed with respect to the presence/absence of *Fut2* glycans are indicated as “secretor” (*Fut2* $^{+/+}$ & *Fut2* $^{+/-}$) versus “nonsecretor” (*Fut2* $^{-/-}$). To generate littermates of all three *Fut2* genotypes, C57BL6/J (*Fut2* $^{+/+}$) mice were initially crossed to *Fut2* $^{-/-}$ mice to create *Fut2* $^{+/-}$ mice, which were subsequently intercrossed. Thus, in addition to *Fut2* genotype and secretor status, we also accounted for the direction of the initial cross, *i.e. Fut2* $^{+/+}$ females × *Fut2* $^{-/-}$ males versus *Fut2* $^{-/-}$ females × *Fut2* $^{+/+}$ males, which we refer to as “secretor dam” versus “nonsecretor dam”, respectively. Fecal samples were taken at four time points: one-, three-, five- and eleven weeks post-weaning. Mucosal tissue samples (*i.e.* jejunum, ileum, cecum, and colon) were taken after sacrificing animals at eleven weeks post-weaning.

Phylum and indicator analyses: To explore the microbial communities at a basic level, we first investigated the dynamics of the major microbial taxonomic groups over time (feces) and location within the gastrointestinal tract (intestinal tissue sampled after sacrifice). This reveals changes in the major bacterial groups Bacteroidetes and Firmicutes with respect to secretor status and initial breeding direction over the time course. The abundance of Proteobacteria changes over time mainly with respect to breeding direction (see Figure 1, Table 1). Bacteroidetes decrease over time in secretor animals, which appears more pronounced in those whose grand dam was also a secretor. In nonsecretor mice Bacteroidetes increase over time, but again show a steeper increase in animals originating from a mouse lineage inoculated by a secretor grand dam (Figure 1A; Table 1). For the phylum Firmicutes we find effects mirroring those of the Bacteroidetes, being more abundant in the lineage inoculated by secretor dams, but showing an overall increase in secretors and a decrease in nonsecretors over time (see Figure 1, Table 1). Proteobacteria on the other hand do not appear to be influenced by host *Fut2* genotype, but primarily by initial breeding direction. Specifically, Proteobacteria decrease only in the lineage initially founded by nonsecretor dams (Figure 1C).

When examining phylum abundances within mucosal communities, a weak decrease of Firmicutes and increase of Bacteroidetes is present in the ileal mucosa of *Fut2* $^{+/-}$ mice. Proteobacteria abundance shows differences with respect to secretor status in the ileal mucosa, with a higher abundance in secretors stemming from secretor dams and the opposite effect in mice originating from nonsecretor dams. Proteobacteria are the only bacterial group in the jejunum which shows an effect of *Fut2* genotype, particularly an increase in abundance with the number of functioning *Fut2* alleles (*i.e.* an additional dose effect between *Fut2* $^{+/-}$ and *Fut2* $^{+/+}$ secretors).

In the more proximal regions of the gastrointestinal (GI) tract (cecum, colon), the effects of *Fut2* genotype/secretor status vanish among the larger microbial groups. Especially in the cecum, Bacteroidetes display no dependence on *Fut2* genotype, but a decreased abundance in the mucosa of animals stemming from secretor dams. In the cecum only Proteobacteria are influenced by *Fut2* genotype, where the abundance of Proteobacteria is reduced in *Fut2* $^{-/-}$ mice. In the colonic mucosa on the other hand, no effect of breeding direction or *Fut2* genotype was identified (Table S1).

We next investigated the microbial communities on the level of single genera and species that are characteristic of a specific *Fut2* genotype, secretor status, or breeding direction using indicator species analysis. One week after weaning no genus was characteristic for any combination of *Fut2* alleles (Table 2), while unclassified Alphaproteobacteria, *Anaerotruncus* (Firmicutes) and the genus *Escherichia-Shigella* (β-

Proteobacteria) are characteristic of the breeding lineage founded by a nonsecretor dam, and *Paraprevotella* (Bacteroidetes) occurs preferentially in the secretor dam lineage. In later time points, several bacteria are detected with characteristic occurrence in secretor or nonsecretor genotypes. Most notable is the consistent signal of *Ruminococcus* (Firmicutes), which is an indicator of nonsecretor individuals over several weeks in the fecal time course (Table 2). In the mucosal communities we find no overlap of indicator bacteria between the single parts of the gastrointestinal tract or the fecal community, which reflects the differentiation between regions of the gastrointestinal tract. Of interest are the genera associated to secretor genotypes, such as *Staphylococcus* (*Fut2* $^{+/+}$ in the entire GI tract, jejunum), *Streptococcus* (*Fut2* $^{+/+}$ in the jejunum), and *Lactobacillus* (*Fut2* $^{+/+}$ in the cecum), which are known to bind fucosylated blood group antigens [57-60]. Members of the TM7 candidate group also show an association to the lack of blood group related antigens in the colonic mucosa, either directly by the host genotype (*Fut2* $^{-/-}$), or indirectly via the breeding direction (*Fut2* $^{-/-}$ dam, Table S2). Further, we investigated the preference of genera for one of the breeding directions (*Fut2* $^{+/+}$ or *Fut2* $^{-/-}$ dam). *Clostridia* were highly associated to fecal and mucosal samples of the nonsecretor mouse line. Interestingly, in animals derived from nonsecretor dams, we find *Staphylococci* and *Propionibacteria*, which are facultatively anaerobic opportunistic bacteria commonly associated to the body surface (Table 2). The described patterns are based on consensus genera, but are consistent on the level of species level OTUs (Table S3-S5).

Alpha diversity: A hallmark for understanding the resilience and productivity of ecological communities is to investigate their complexity. The complexity of such a community can be measured in different ways by incorporating the number of observed species (species richness), their distribution (Shannon entropy), or phylogenetic relatedness (Net Relatedness Index, Nearest Taxon Index) [61, 62].

Our dataset allows us to analyze the turnover of microbial communities not only over time, but also to compare the influences of ecological factors including the direction of microbial transmission, host genotype and location within the gastrointestinal tract. Over the time course we identified effects of maternal transmission in the fecal microbial communities on species richness, distribution and phylogenetic diversity, showing mainly comparable starting diversities, strong deviations at subsequent time points and a final convergence of diversity at the end of the time course (Table 3, Figure 2). Species richness decreases after weaning among animals with a nonsecretor dam, but equilibrates towards the end of the time course (TP11; Figure S1, Table S6). The evenness of the species abundance distributions, as described by Shannon entropy, shows different trajectories over time according to breeding direction and *Fut2* genotype, resulting in significant differences in diversity between

Fut2 genotypes at the last time point (*Fut2* $^{+/+}$ > *Fut2* $^{+/-}$ > *Fut2* $^{-/-}$; Figure S1, Table S6). The differences in community trajectories are even more evident between the *Fut2* $^{+/+}$ mice with respect to the secretor status of the dam.

Considering the phylogenetic structure of the bacterial communities we found stronger phylogenetic clustering in secretor- than in nonsecretor mice. This clustering deteriorates over time until the communities become phylogenetically unstructured- (secretors, NRI ~ 0), to overdispersed in nonsecretors (NRI < 0). Secretor status thus determines the trajectory of community assembly on a broad phylogenetic scale. This shows that succession of bacterial communities results in a phylogenetically less restricted community (phylogenetic dispersion) and thus in a potentially more resilient community.

The phylogenetic relatedness among closely related species (NTI), on the other hand, is increased (restricted) in mice derived from secretor dams. The trajectory of secretor dam communities cluster quickly and remain so, while the bacterial communities derived from nonsecretor dams first diversify (decrease of NTI), and then later become more restricted (increase of NTI). Overall, the microbial communities passed from either breeding direction converge towards the end of the time course, but take different trajectories during this process.

The results of these analyses speak for strong habitat filtering outweighing the subtle differences in the starting communities, as seen in the richness and evenness analyses. However, the broad diversification of phylogenetic groups over time (NRI: time span- $F_{1,101}$=21.018, P<0.0001, $R^2_{marginal}$=0.067) is potentially driven by the colonization and establishment of distantly related groups (*e.g.* phyla, classes). Within the larger phylogenetic groups, the phylogenetic distance among species and genera decreases (NTI: time points- $F_{1,101}$=23.153, P<0.0001, $R^2_{marginal}$=0.250) depending on the breeding direction. Mucosal microbial communities at different locations along the gastrointestinal tract do not differ according to breeding line or genotype. Only in the ileum are species richness and entropy significantly lower in the breeding lineage founded by secretor dams (Table S7).

Beta diversity: To investigate the differences between microbial communities we employed beta diversity analyses measuring differential presence, abundance, or phylogenetic relatedness of microbial species (Jaccard, Bray-Curtis, unweighted UniFrac) among samples. Formal tests of community differentiation (Redundancy Analysis, distance based Redundancy Analysis [63-65]) according to *Fut2* genotype or secretor status alone revealed no significant differences between communities. Neither in tissue samples nor in fecal communities was a differentiation of microbial communities by genotype/secretor status alone observable, even though the communities change over time (Jaccard: $F_{3,132}$=1.365,

P<0.0001, R^2=0.030; Bray-Curtis: $F_{3,132}$=1.748, P<0.0001, R^2=0.038; UniFrac: $F_{3,132}$=1.375, P=0.0017, R^2=0.030). However, a consistent factor influencing microbial communities of the mucosa and feces is the direction of the initial breeding crosses (see Table S8). Comparable, long-lasting and potentially fitness relevant siring effects have been recently observed in mice overwriting effects of immune relevant genes [66]. Interestingly, when we consider the interaction of *Fut2* genotype and the direction of the initial breeding crosses we find strong differences in the fecal bacterial communities on the levels of phylogenetic distance, shared presence and abundance until five weeks after weaning. This effect decreases over time, while breeding direction remains significant even in the mucosa-associated microbial communities (Table S8, Table S9). When we investigate the effect of *Fut2* on the basis of secretor status no effects are observable, implying that the dose of *Fut2* expression is a determinant of bacterial community composition. However, the analysis of all time points together reveals that breeding direction and time influence fecal communities, while *Fut2* genotype nested within breeding direction shapes community structure and composition (Table S10, Figure S3, Figure S4). A combined analysis of all sampled mucosa attached microbial communities reveals strong genotype effects in addition to their interactions with breeding direction, revealing an overall influence on the mucosal community along the whole gastrointestinal tract (Table 5). Single anatomical sites, however, show consistent effects only with respect to breeding direction (Table S9).

Next, we investigated community distances within a genotype class (community variability), which reveals a time-dependent decrease of overall community variability and thus a stabilization/homogenization of bacterial community composition. However, in contrast to the differences of community composition between genotypes, which diminish over time, the differences in community variability increase between *Fut2* genotypes over time (TP11-secretor status: Jaccard: P=0.0344; Bray-Curtis: P=0.0342; UniFrac: P=0.3002; Euclidean: P=0.0435). Differences between homozygous secretors (*Fut2* $^{+/+}$) and nonsecretors (*Fut2* $^{-/-}$) are most apparent (Jaccard: P=0.0189; Bray-Curtis: P=0.0217; UniFrac: P=0.1548; Euclidean: P=0.0756; Figure S5). This reflects a more homogenous or restrictive microbial colonization in *Fut2* $^{-/-}$ animals over time. Breeding direction, on the other hand, had no observable effect on the variability of fecal communities. Only the mucosal communities of the jejunum and all anatomical sites combined display higher variability in the breeding line founded by nonsecretor dams (jejunum- UniFrac: P=0.012; all GIT locations- Jaccard: P=0.008, Bray-Curtis: P=0.004, UniFrac: P=0.002).

Network analysis: To infer potential relationships between bacteria we used co-occurrence/correlation networks of bacterial genera. To construct the networks for each time point, we split the abundance table accordingly and calculated measures of determination on

the same set of bacteria and animals for each stratum using a compositionality considering correlation procedure [67]. We measured different aspects of bacterial network positions to approximate the structural importance of single members within those communities. To assess this we used four different measures of the importance/centrality of network components, including the number of connections/interactions of each bacterium in the network (node degree), the Google™ PageRank™ index [68], the well-known "betweenness centrality" [69], and closeness centrality [69]. In simple terms, PageRank™ assigns a high importance to genera that interact with other important genera, while betweenness centrality measures the number of interactions in which the respective genus is a mediator (on the shortest path between any genus pair). Closeness centrality, on the other hand, measures how distant a node is from all other nodes in the network. When we compare the networks between single time points, patterns that best match the processes of community succession and stabilization emerge. The number of interactions between bacterial genera decreases over time (node degree: ρ=-0.497, $P<1.00 \times 10^{-15}$). Further, not only the number of connections, but also the direction of those interactions changes from a higher number of negative interactions to an almost even distribution of positive and negative interactions (positive/negative ratio; TP1: 0.292, TP3: 0.492, TP5: 0.791, TP11: 0.776), while their overall strength increases (mean real value weights; TP1: 0.171, TP3: 0.221, TP5: 0.228, TP11: 0.234; Figure S6). Furthermore, we identified an increase of the overall importance of single genera within those assemblages. The PageRankTM as a generalized importance index reveals an increase with time (ρ=0.148, $P=2.21 \times 10^{-6}$) as well as the mediating role of single bacteria (betweeness: ρ=0.0879, P=0.0876), while the community members become more central and closer connected to all other members (ρ=−0.7643, $P<2.20 \times 10^{-16}$; see Figure 4B). This hints towards a stabilization of a core set of strongly interacting bacteria over time, with increasing mutualism and less competition between single genera, which widens the network and increases its modularity (diameter; TP1: 1.381, TP3: 2.037, TP5: 1.973, TP11: 3.197; modularity/number of modules; TP1: 0.311/7, TP3: 0.488/8, TP5: 0.580/11, TP11: 0.562/11; Figure S7) [70]. Thus, importance of each respective genus in the community increases, which is also reflected in the increased strength of correlations between bacteria over time (ρ=0.120, P=0.0001).

Interestingly, the association strength (indicator value) of bacteria to secretor status reveals a strong negative relationship with the node degree of bacteria, which increases over time. With community development strong indicator genera for either secretor- or nonsecretor status appear to become more central in the network, a pattern more pronounced for indicator genera of secretor status (see Tab.S11). This implies that bacteria preferring fucosylated sugars in their environment also interact more, or depend more on interactions with other bacteria. Thus, the loss of ABO antigens in the gut could destabilize

the microbial communities by restricting potential syntrophic and stabilizing relationships. We further tested the relationship of network importance with the association of bacteria to the respective breeding direction. Over time this relationship becomes stronger, especially for the bacteria associated to the mouse line founded by secretor dams (see Table S11). Overall, the associations appear strongest in animals directly- or indirectly influenced by the presence of ABH antigens, being it genotype or founding genotype, making bacteria associated to the presence of fucosylated glycans more important for the microbial community.

To test the robustness of the microbial communities against different disturbance regimes, we next simulated random loss of community members or targeted depletion of important keystone members (*i.e.* network hubs- based on the number of interactions) via random and targeted sequential removal of network nodes [71]. For each respective co-occurrence network we measured the resulting changes in the network characteristics such as network fragmentation (number and size of connected components), size (diameter) and clustering (transitivity, closeness). As expected for ecological communities, which are mainly described as scale free networks, we found high resistance against random failures, but fast deterioration of communities after removing small fractions of important nodes (see Figure 4C, Figure S8, attack vs. random failure). The highest resistance against community collapse was observed shortly after weaning, where many weak interactions seem to prevent the community from collapsing/disintegration (Figure 4, Figure S8). Over time many interactions appear to solidify while others are lost during succession, which increases the importance of single associations and therefore the vulnerability of the system. Thus, removal of important mediators (targeted attack) increases community disturbance, while random removal has far less impact. The networks consistently break up into more fragments during targeted attack compared to random network failure (see Figure 4C, Figure S8), as does the size of the biggest connected component. The transitivity of the networks deteriorates much quicker compared to random removal throughout the time course, similar to the network diameter (spiking) and closeness centrality. These tests illuminate community resilience in early phases of community development, but also the susceptibility of climax communities to the removal of key players. By comparing the behavior of the empirical networks to several simulated networks of similar size and different topologies, we revealed that under random error- and targeted attack regime the empirical networks behave similar to simulated networks of the same degree distribution (Erdös-Renyi, degree sequence, power law (power=4), small world, Figure S9, Table S12). Specifically, communities from the last time point appear more resilient than any simulated topology.

To investigate the role of the *Fut2* genotype or secretor status on the robustness of microbial communities, we sequentially removed the 25% of bacteria from the network that were most strongly associated to either secretor- or nonsecretor status. This revealed several interesting characteristics, such as a higher vulnerability of the communities to the removal of bacteria with a certain environmental preference (attack) than random removal of bacteria (failure). Especially the removal of bacteria associated to secretors influences the networks earlier in the removal sequence and more strongly than the removal of nonsecretor-associated bacteria, as signified by network fragmentation (*i.e.* number of connected components; Figure 4C; Figure S10, S11). Especially in early community development (TP1, TP3) the average number of resulting subnetworks is higher than after removal of nonsecretor-associated bacteria. The size of the biggest community component (core community), on the other hand, does not differ dramatically from the patterns during random removal of bacteria (Figure S10, S11). Similar patterns are observed when the analyses are based on indicators for *Fut2* genotype, where especially the removal of bacteria associated to the *Fut2* $^{+/+}$ genotype show the strongest effects on network characteristics. Thus, microbes that associate with blood group related antigens appear to be more central and interactive in the microbial communities, making them crucial to balance community disturbances (Figure S12, S13). Removal of bacteria associated to the breeding direction founded by a secretor dam disturbed the potential interactions earlier than the removal of bacteria associated to the nonsecretor founded mouse line (Figure S14, S15). The disturbances introduced by removal of secretor dam-associated bacteria are on average higher than random removal of network members.

Discussion

Microbial communities can be viewed as a plastic and complex phenotype of multicellular hosts, which is influenced by numerous factors like host genetics, initial founding members (*e.g.* mode of delivery), community disturbances (antibiotics), or diet. Over the last years glycosyltransferases have received attention from evolutionary and microbiome researchers [14, 28, 72-74]. Glycans represent a major part of the mucosa serving as attachment sites for as well as nutrient sources for microbes [37, 75, 76]. These genes appear essential for host-bacteria homeostasis, as their expression is directly triggered by the resident microbial community [38, 39].

Bacteria selectively take up and incorporate glycans to evade host immune recognition, increase their colonization success or use them directly as a nutrient source [8, 36, 75]. In concordance with this, we show examples of bacterial interactions with ABH

antigens including *Staphylococcus* [58] and *Lactobacillus* [57, 59, 60] in the mucosa of secretor mice, which are known to bind fucosylated blood group antigens. The strong variation in *Fut2* expression over time/gut development [77, 78] may thus also explain the late stabilization and relatively high variability of indicator taxa (*i.e. Ruminococcus*). We also observed a decrease of *Fut2* effects on the mucosa-associated microbial communities along the gastrointestinal tract. This may be attributed to the decrease of glycosyltransferase expression and α-1,2-fucosylated glycans/H antigens along the proximodistal gradient and by local expression patterns in the gastrointestinal tract and differences in the mucosal structure [78, 79]. Major bacterial groups such as Bacteroidetes and Firmicutes show differences with respect to secretor status in interaction with the breeding direction, as does the phylogenetic distribution of species. Glycan liberation through the microbial community influences the colonization success of commensals and pathogens [80]. Liberated fucose has been linked to increased resistance of intestinal cells via preemptive cytokine production [56]. Furthermore, fucosylation appears as an important factor induced by the resident community in type 3 innate lymphoid cells [52]. This could represent a response of intestinal epithelial cells to buffer disturbances in microbial communities during infection, through the modulation of quorum sensing- and virulence- mechanisms, which modulate the microbial community directly through the immune system and indirectly via microbial interactions [50-52, 55].

Stabilization of the microbial communities, especially among the nonsecretors, develops from a pattern of relative species clustering (co-occurring bacteria closer related than expected by chance) to a pattern of phylogenetic overdispersion (co-occurring bacteria more distantly related than expected by chance). Early in development the microbial communities are phylogenetically clustered, potentially due to the colonization bottleneck and the common origin from secretor parents (*Fut2* $^{+/-}$), which are early environmental filters. With time, communities in secretor individuals develop into neutrally-assembled communities, or in the case of nonsecretor individuals into phylogenetically overdispersed communities. This can be interpreted as a sign of competitive exclusion of moderately related bacteria with comparable metabolic dependencies [81, 82], or as a sign of community facilitation with mutualistic interactions between distantly related species [82, 83]. The previously common resource fucose, as provided via the milk of the secretor mothers, becomes scarce and later absent in nonsecretor mice, which increases the competition for other available glycans or forces communities to compensate by recruiting more/different functions from distantly related bacteria (facilitation, phenotype differences). In this case the type of interactions that drive the phylogenetic overdispersion (phenotype matching, phenotype displacement) can be important for community resistance against potential invaders like pathogens [84]. Most bacteria in the investigated communities show weak negative correlations among each other, and implies weak competitive interactions. Together, the competitive nature of communities

and the overall increase in pairwise phylogenetic distance (decrease of NRI) may also increase biotic resistance against invaders [84]. The abundance of weak competitive links in communities seems to have stabilizing effects on ecological communities, as reactions of taxa are asynchronous, which balances the reduction of one species by the complementary increase of other community members [85-87]. This so called “portfolio”- or “insurance” effect can decrease the effect of environmental disturbance by a release of competition [88-90].

Thus, aside from the direct effects of *Fut2* genotype or secretor status, we found strong indirect trans-generational effects, which highlight the importance of initial microbial colonization interacting with the underlying genotype in the development of microbial communities. This becomes evident in the dominant effect of initial inoculation on the diversity within and between bacterial communities. The differentiation among breeding lineages carries down from the level of bacterial phyla to single genera and species. Communities founded by a nonsecretor dam are characterized by an increased abundance of aero tolerant and opportunistic pathogens such as *Streptococcus*, *Staphylococcus* and *Propionibacteria*. These bacteria have been shown to be the dominating bacteria in infants delivered via caesarian section, whereby they represent the mother’s skin microbiota as the first inoculum [26]. The relevance of initial transmission is also shown in the high abundance of members of the *Clostridium leptum* group (*Cluster IV*) in the mucosa of *Fut2* $^{+/+}$ inoculated mice. This bacterial group is well known for its butyrate production, providing essential nutrients for enterocytes and fostering intestinal homeostasis. Community diversity also shows surprisingly strong responses to breeding direction, especially on the trajectories of community development over time, while the starting and final diversity do not differ between breeding directions regarding number, distribution, and relative relatedness of species. The importance of breeding and cohousing effects has also been shown in *TLR* knockout mice, whereby the signals of these immune related genes on the luminal and mucosal microbial communities were overshadowed by effects of cohousing and legacy [66].

However, these patterns may represent not only an effect of the initial bacterial inoculum, which is in our case influenced by initial dams genotype (*Fut2* $^{-/-}$/*Fut2* $^{+/+}$), but also indirectly by the absence of fucosylated oligosaccharides and antibodies in the mothers’ milk [91, 92]. These molecules have been shown to exert strong influences on microbes [33, 47, 93, 94] and immune development [48, 92]. Furthermore in addition to vertical community transfer and modulation (“legacy” effect) being important factors for microbiome development, the offspring’s genotype itself influences phyla, genera and species abundance in the fecal microbial community. The interaction of the *Fut2* genotype/secretor status and initial microbial inoculation has been reported in the context of *Campylobacter jejuni*, enterotoxigenic *Escherichia coli*, and *Calicivirus* infections [33, 47]. Mothers with a

functioning *FUT2* gene appear to reduce pathogen binding to the offspring's' intestinal mucosa through the provision of fucosylated milk oligosaccharides as decoys that compete for binding with the pathogens, other taxa and the offspring's mucosa [33]. The intrinsic expression of *Fut2* is only initiated after weaning, mainly by signals of the microbial communities, which further underscores the importance of *Fut2* for host-microbial homeostasis [37-39, 75, 76]. The genital tract, the first bacterial encounter for the newborn, is also a mucosal site of active *Fut2* expression [95, 96] and home to a variety of bacteria important for the initial colonization process of newborns and disease resistance [26, 96, 97]. This species poor community may be altered by the absence of *Fut2* expression as well, especially as the main taxa of the vaginal microbiome (*e.g. Lactobacilli*) are known to bind ABH blood group antigens [57, 60]. These early differences of inoculation influence the development of the gut microbial communities most likely by changing the order of species succession [98-100], potentially imposing different deterministic dynamics on the trajectories of community development and composition. The interaction of bacteria and the immune system and neutral dynamics may further enhance early compositional differences [98-101]. Early microbial exposure and differences in the community complexity can thus have long lasting effects on the development of the immune system, even priming the host for potential inflammatory and autoimmune disease [102-104].

The aforementioned differences in trajectories are well reflected by the development of terminal phylogenetic clustering and species diversity. Species diversity and clustering quickly plateau in the breeding lineage founded by a secretor dam, which could be the result of competition among closely related taxa leading to phylogenetic clustering [82]. The opposite effect is present in the lineage founded by a nonsecretor dam, which does not diversify and only later begins to show a similar phylogenetic clustering. This could be interpreted as colonization by a wide range of bacteria after weaning, which only later specialize to the same level as observed in the other breeding direction. In general, NTI increases over time (ρ=0.2967, P=0.0005), while NRI decreases (ρ=-0.3298, P=8.80 × 10^{-5}). Closely related taxa appear to exclude each other, while distantly related taxa can be more diverse through reduced niche overlap correlated with phylogenetic distance. Thus, secretor status mainly influences broad phylogenetic patterns over time, while the breeding direction influences the number of species and occurrence of closely related species.

Community resilience is another cornerstone of host-microbial homeostasis and is suspected to play a major role in the development of dysbioses. A pattern we observed among the microbial communities shows changes in community resistance over time depending on the underlying *Fut2* genotype, secretor status or breeding direction. Complex systems like the host associated microbial communities can dampen disturbances through

functional redundancy and decentralized structure, even though community composition over time is not completely stable [23]. We tested the resistance of communities to random removal of single bacteria and targeted attacks on central community members *in silico* in order to simulate natural community fluctuation (random removal, random failure) and the introduction of a highly competitive pathobiont (*e.g. C. difficile*) or a narrow-spectrum antibiotic (targeted removal), respectively. These strategies were previously employed in order to investigate the stability of complex networks such as the internet, biological networks (protein interaction, metabolic networks) and ecological networks [71, 105-108]. Modular, differentiated networks tend to be more robust to random fluctuations; as such disruptions are kept compartmentalized within the modular structure and do not spread throughout the network [71, 106]. Modularity has also been described for human microbial communities, and certain functional modules have been associated to inflammatory bowel disease risk [109]. High modularity, as observed in our microbial co-occurrence networks, has been shown to allow for a higher total abundance of community members and their diversity, which implies a higher productivity of structured communities by reducing interspecific competition [110, 111]. This reduction of competitive interactions and increased modularity is also present over the successional gradient of the co-occurrence networks in our study (see Network analysis). These analyses further revealed characteristics of error tolerance and attack susceptibility in the empirical networks comparable to aspects of small world and exponential networks [71]. However, the microbial communities appear astoundingly robust against random attacks/failure, specifically when the potential climax community is reached, outperforming all simulated topologies (Figure S9).

We extended this concept to the removal of bacteria with a preference for occurring in a specific *Fut2* genotype, secretor status, and breeding direction, which disturbed the communities more than expected by chance. These heuristics revealed a higher importance of bacteria associated to secretor genotypes for community homeostasis. This effect may be a product of cross feeding relationships with bacteria essential to liberate the ABH blood group antigens from the mucosal surfaces, or to cope with the absence of this nutrient source or attachment site [36, 80, 112]. Furthermore, the bacteria most characteristic for the secretor status are also exert greater influences on other bacteria and thus the abundance of many other bacteria in the community, as glycan liberation through the microbial community influences the colonization of commensals and pathogens [80]. Liberated fucose may represent a link between the microbial community and the immune system, further stressing the importance of *Fut2* associated microbes [50-52, 55, 56]. These results illustrate the importance of specialized key members in microbial community function and may open up a different analysis strategy to evaluate community dynamics and characteristics.

These results, however, are approximations and the real food-web structure remains elusive in such complex multispecies assemblages. The integration of metabolic traits, their trophic role in community functioning [113], cell size [114], and growth dynamics [115] in such models might further enhance our understanding of microbial community vulnerability to disturbance as an origin of dysbiosis and disease, which represent significant future challenges in microbiome research.

Material and Methods

Animal Husbandry: We used the B6.129X1-*Fut2* ^{tm1Sdo}/J mouse model [116] kept in independently ventilated cages under specific pathogen free conditions to investigate changes in the microbial community. Animals were bred by mating a *Fut2* $^{-/-}$ male or female with a respective wild type C57Bl/6J (*Fut2* $^{+/+}$) mouse, depending on "breeding direction". The resulting heterozygote offspring were mated to obtain experimental animals different by *Fut2* genotype. Feces were sampled at most 5 days after weaning, and then every 7 days for 11 weeks and dissected after the final sampling. Fecal samples were collected on ice and immediately transferred to -80°C until being processed. To obtain the mucosa associated microbial communities after dissection (TP11), the tissue was washed in 4 ml RNAlater® (Ambion®; Carlsbad, CA, USA) and stored separately from the luminal content in a fresh tube containing 1.5 ml RNAlater®. Samples preserved in RNAlater® were left over night at 4°C, were spun down, and supernatant was removed before storage at -80°C. To avoid cross-contamination, instruments were rinsed and cleaned with 70% ethanol between anatomic sites. All animal procedures were approved by the Research Animal Ethics Committee of Schleswig-Holstein.

DNA extraction and 16S rRNA gene pyrosequencing: DNA was extracted from fecal and RNAlater® washed mucosa samples (stored at -80°C) of each animal according to the manufacturer's instruction with the Qiagen Stool DNA Isolation Kit (Qiagen, Hilden, Germany). The 16S rRNA gene was amplified using forward (5´-CTATGCGCCTTGCCAGCCCGCTCAGTCAGAGTTTGATCCTGGCTCAG-3´) and reverse (5´-CGTATCGCCTCCCTCGCGCCATCAGXXXXXXXXXXCATGCTGCCTCCCGTAGGAGT-3´) primers flanking the V1 and V2 hypervariable regions (27F-338R) and were sequenced following the methods described in Rausch *et al.* 2011.

Sequence processing and quality control: Raw sequences were trimmed using mothur 1.23.1 requiring no ambiguous bases, a mean quality score of ≥ 35 and a minimum length of 200 nucleotides for the coupled V1-V2 region [117] and removal of chimeric sequences detected by Usearch 4.25 (overlap of *denovo* and database centered) [118]. Sequences

were subsampled to 1500 reads per sample, classified via RDP classifier with ≥ 60% bootstrap threshold [119, 120] and aligned to the SILVA database [121]. OTUs were binned in mothur using the average distance clustering and phylogenetic tree construction on representative OTU sequences was carried out using FastTree 2.1 with a gamma corrected CAT substitution model [122]. We achieved a relatively high coverage of 88.49 ± 6.15% at the species level over all samples, for fecal samples 87.70 ± 4.02% at TP1, 83.81 ± 7.32% at TP3, 86.13 ± 6.96% at TP5, and 84.26 ± 4.26% at TP11, respectively. Mucosa associated microbial communities were sequenced at a level of 87.31 ± 3.81% for cecal tissue, 92.17 ± 3.13% for colonic tissue, 94.11 ± 2.45% for the jejunum and 94.98 ± 2.39 for ileal tissue.

Statistical analysis: Species diversity indices (species richness, Shannon-Weaver index) as well as phylogenetic diversity were calculated in R [123, 124]. Phylogenetic measures were derived using species occurrences, resulting in unweighted NRI (Net Relatedness Index) and NTI (Nearest Taxon Index). These measures represent phylogenetic effect sizes inferred by contrasting the observed relatedness patterns and a null model, by maintaining species occurrence frequency and sample species richness over 999 iterations [125]. The phylogenetic measures of beta diversity, unweighted/weighted UniFrac, were calculated in mothur and provide insight into differentially present or abundant phylogenetic lineages between communities [126]. Solely species based beta diversity metrics based on shared OTU presence (Jaccard distance), or shared abundance (Bray-Curtis distance) were calculated in the vegan package for R [124]. Statistical analyses of community composition based on different beta diversity metrics was performed with nonparametric matrix based analysis of variance using “adonis” implemented in the vegan package for R with 105 permutations [124, 127]. Adonis models were reduced until only significant factors remained. Distances were ordinated via Principal Coordinate Analysis (PCoA) and fit of clusters was assessed via an iterative process (105 permutations). For constrained ordination (Redundancy Analysis, RDA) the OTU table was Hellinger transformed and RDA was carried out following Legendre and Legendre [65]. Significance of factors and axes in RDA and distance based RDA (Jaccard, UniFrac, Bray-Curtis) was determined using a permutative ANOVA approach (5000 permutations, stratified by time point or GIT location in global analyses). For univariate analyses of repeated measurements (*i.e.* time course), linear mixed models with mouse ID as a random variable and a cage dependent variance structure were applied and reduced by model selection using the conditional AIC criterion and its weights [128]. For analyses within a certain time point or GIT location, cage was treated as a random variable for LMM analysis, with normality of model residuals after refitting of the final model under REML as a requirement [129]. Indicator species analysis was based on 105 permutations using the generalized indicator value (IndVal.g) to assess the predictive value

of a taxon for each respective host phenotype/category (taxon frequency ≥ 0.1) [130]. P-values of the genera and OTU associations were adjusted by the Benjamini-Hochberg procedure [131]. To assess the predictability of the microbial abundances between time points, we calculated the real value of the non-parametric spearman correlation coefficients for each present bacterium between each respective time point and compared those using Wilcoxon rank tests. Genera networks were generated using the SparCC algorithm (100 iterations, 105 permutations) as implemented in mothur to avoid spurious correlations induced by compositionality [67]. The networks were constructed from the correlation matrices and weighted by the correlation coefficient in igraph for R using only associations with $P \leq 0.005$ [132]. Calculation of centrality scores and network manipulation were also carried out in igraph, whereby weights were transformed to real values for the derivation of centralities and modularity. Network robustness tests were performed by random attacks on networks by sequentially removing 25% of network nodes randomly over 1000 iterations, and mean values of the network characteristics for each fraction were used for further analyzes. Targeted attacks were performed by sequentially removing the highest connected bacteria (highest node degree) from the network. We also simulated network attacks based on the association strength of bacteria to a host characteristic as measured by its indicator value (IndVal.g). The top 25% genera were used for sequential removal, and permuted 1000 times to exclude directional effects. Artificial random networks of similar sizes were constructed (100 iterations) based on a similar degree distribution [133] or by random and evenly distributed associations (Erdös-Renyi) [134], and partially rewired random networks (rewiring probability k; k=0.6, k=0.8) which represent networks with small world characteristics. The scale-free networks of similar sizes were constructed with a power law degree distribution (p=1, p=2, p=4) [135], as were the exponential networks (p=4, p=6, p=8). These networks were subjected to random and targeted attack as described earlier (100 iterations in each constructed network). To assure comparability of results, all network characteristics were related to the respective simulated network before disturbance.

Figures

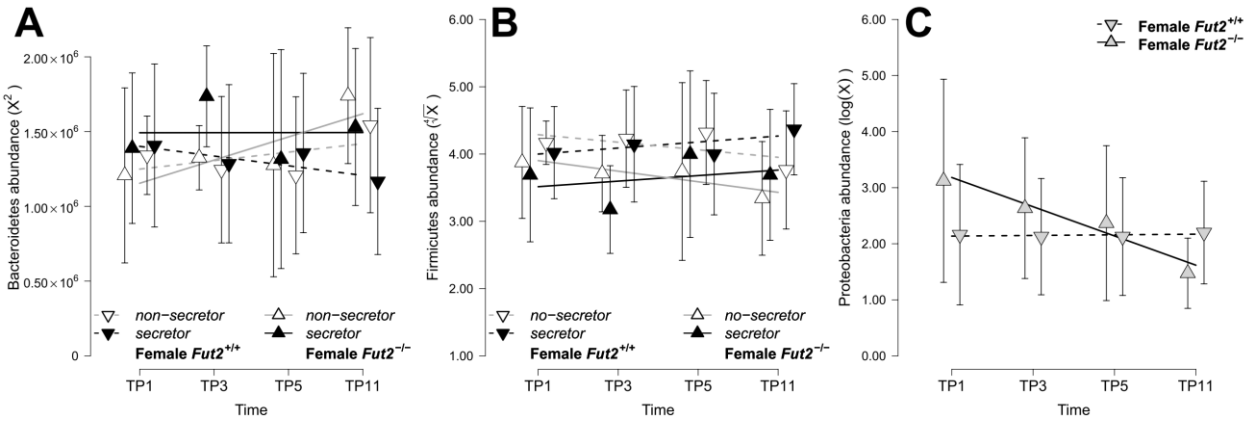


Figure 1: Analysis of the major bacterial phyla over the fecal time course incorporating *Fut2* genotype/secretor status or the mouse breeding lineage (founded by *Fut2* $^{-/-}$ or *Fut2* $^{+/+}$ dam).

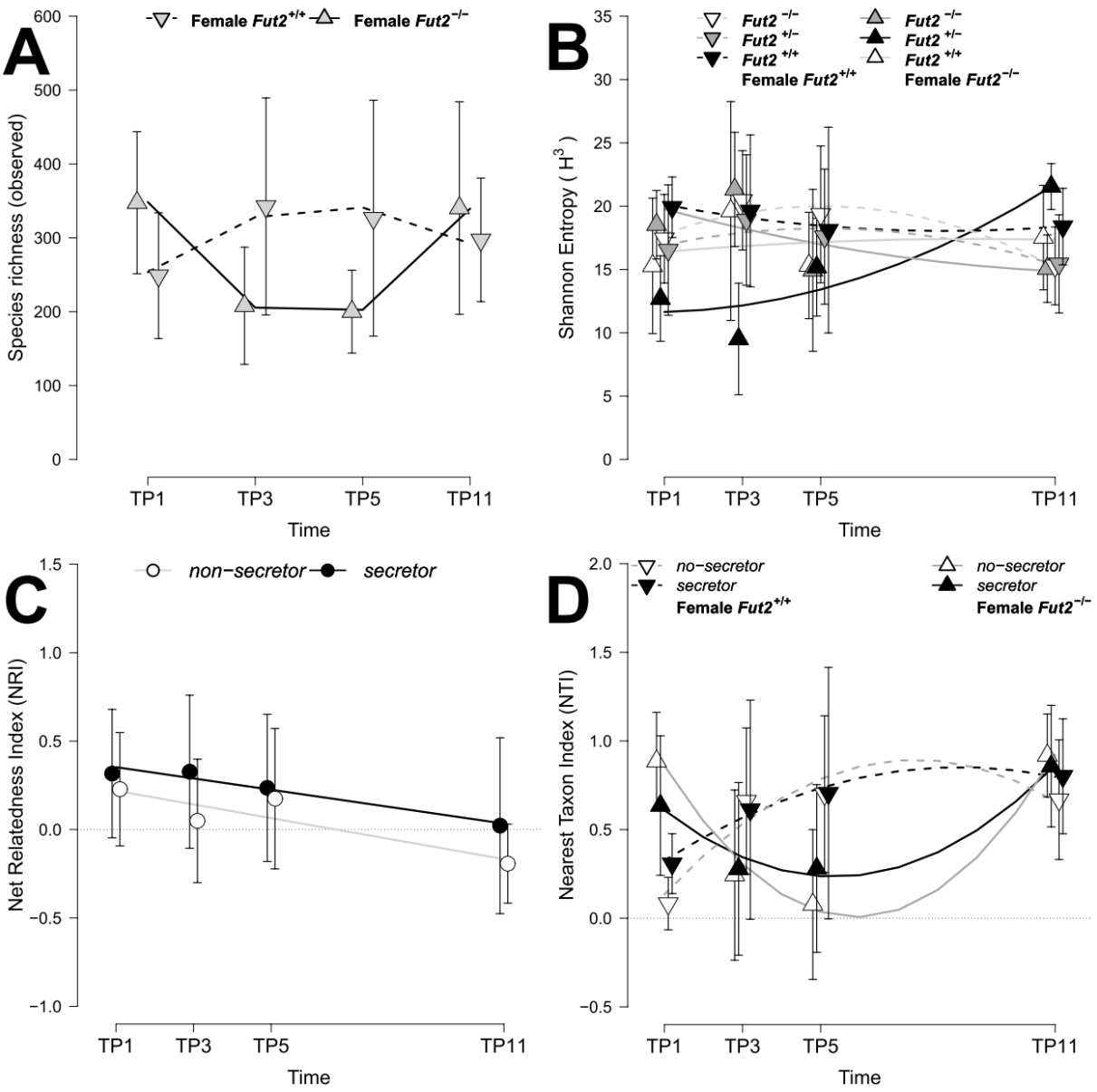


Figure 2: Analyses of species richness (A), community complexity (B; Shannon Entropy), general phylogenetic clustering (C) and terminal phylogenetic clustering (D) of the communities over time. The best statistical model for each diversity metric is plotted (see Table 1).

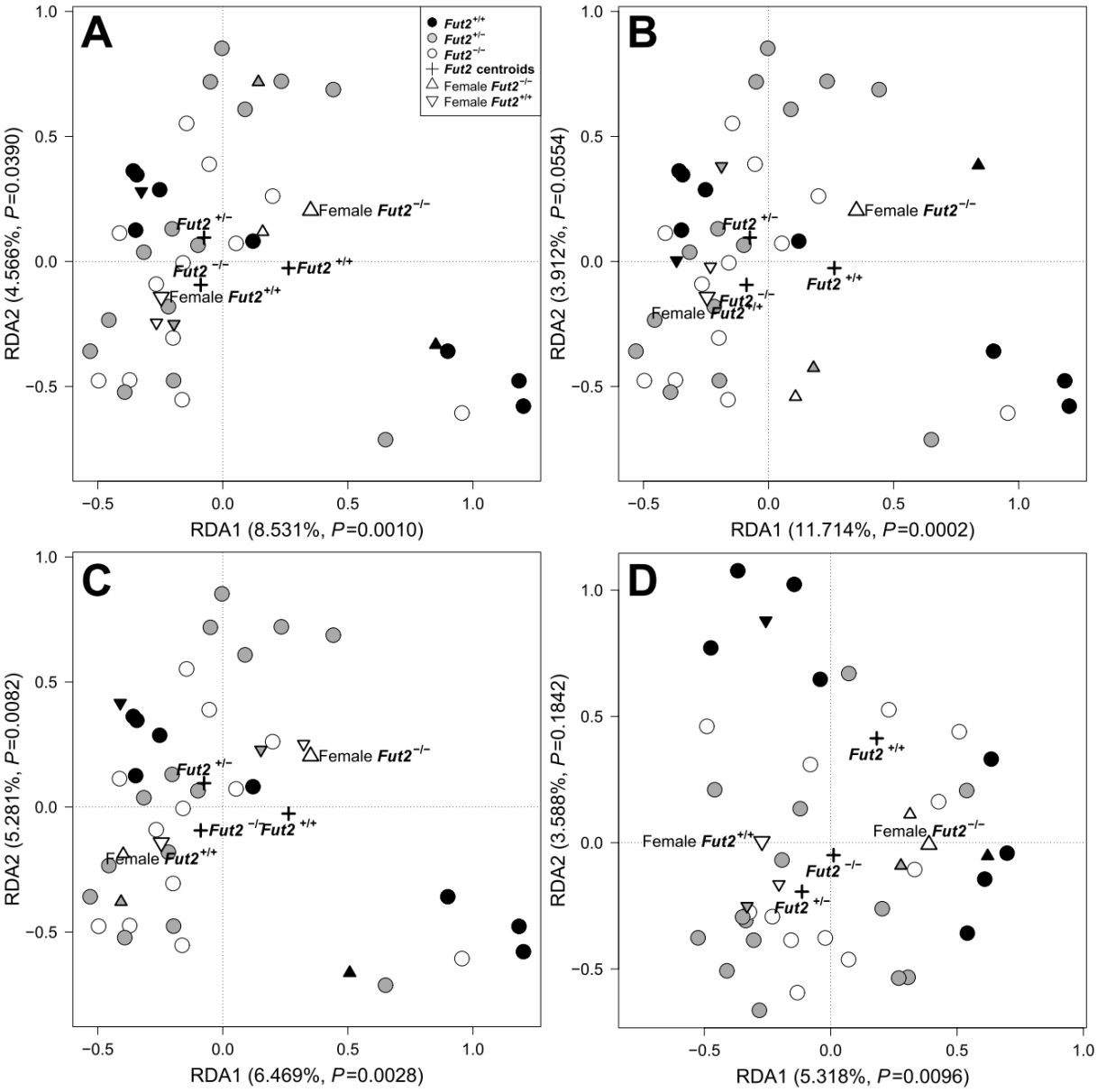


Figure 3: Redundancy analysis of the microbial communities for each individual fecal time point (see Tab.4).

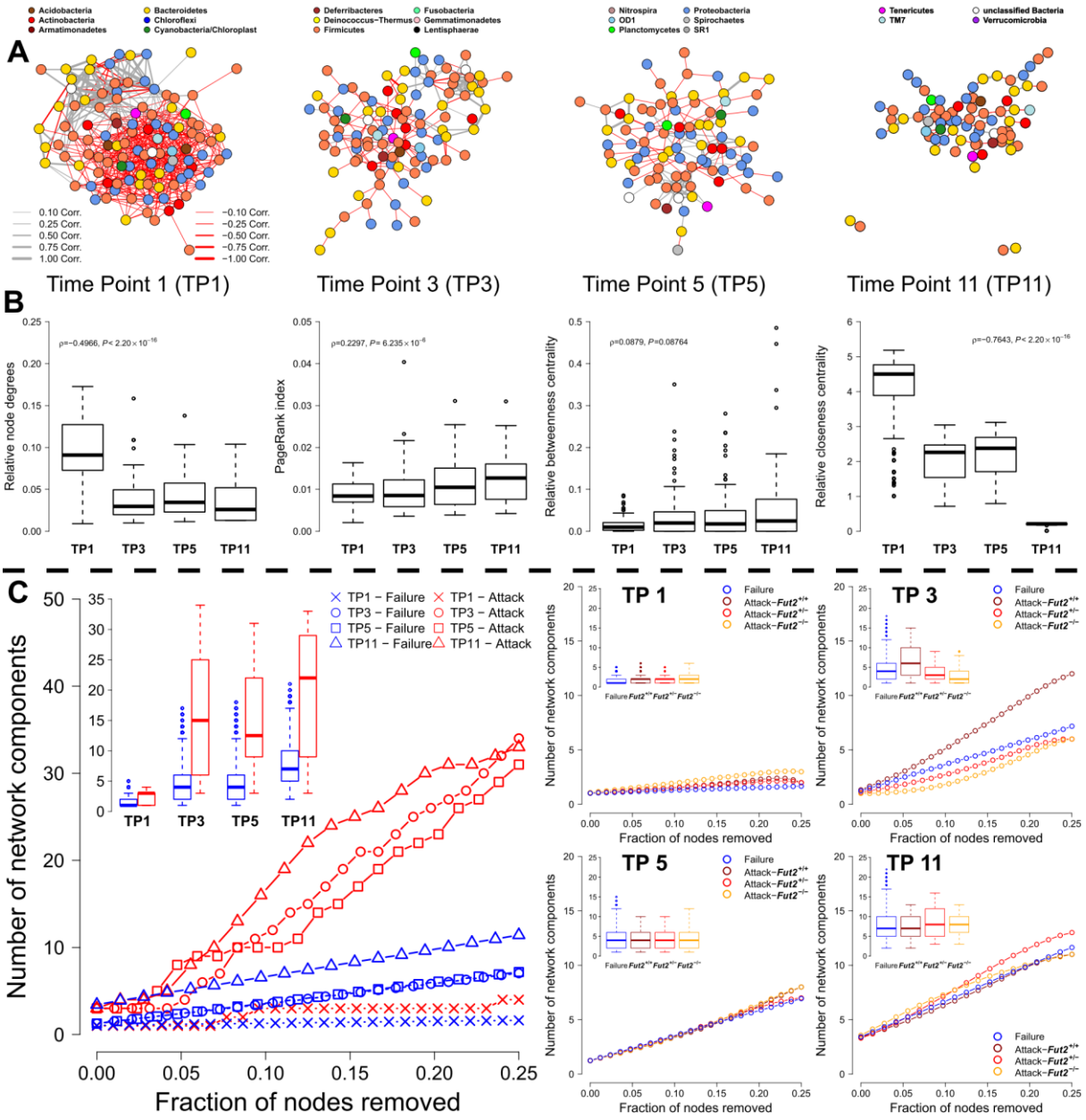


Figure 4: (A) Correlation networks of time point 1 to 11 reveal general network density to be decreasing (TP1=0.0948, TP3=0.0359, TP5=0.0402, TP11=0.0343), as well as decreasing centralization (TP1=33.196, TP3=18.080, TP5=11.801, TP11=11.101, Hubscore), whereas the diameter is increasing (TP1=1.381, TP3=2.037, TP5=1.974, TP11=3.197). (B) Analyses of node characteristics among the interaction networks between time points focusing on the number of connections of single genera (node degree), their importance based on the quality of its connectedness (PageRank^{TM}), and the importance of single bacteria as mediators between assemblages (betweeness). (C) Analysis of network robustness based on sequential random removal (network failure) and targeted attack on the most integrated genera (highest number of connections), measured as the decay of the networks into smaller connected components. The robustness of bacterial interaction networks was further tested for each time point by random removal treatment (blue, failure) and targeted removal of the

top 25% of bacteria associated to the *Fut2* $^{-/-}$ (darkred), *Fut2* $^{+/-}$ (red) or *Fut2* $^{-/-}$ associated bacteria (orange) based on 1000 iterations.

Tables

Table 1: Abundances of the three most abundant phyla within and across all fecal time points.

	Phylum	Factor	*DF*	*F*-Value	*P*-Value	marg.R^2
All	Bacteroidetes*	*Intercept*	1,99	1770.732	< 0.0001	0.053
		Secretor	1,31	0.077	0.7829	
		Timepoint**	1,99	2.159	0.1449	
		Direction	1,31	3.725	**0.0628**	
		Secretor:Timepoint	1,99	8.326	**0.0048**	
		Timepoint:Direction	1,99	7.573	**0.0070**	
	Firmicutes#	*Intercept*	1,99	2257.109	< 0.0001	0.153
		Secretor	1,31	0.061	0.8066	
		Timepoint	1,99	1.772	0.1862	
		Direction	1,31	11.468	**0.0019**	
		Secretor:Timepoint	1,99	9.937	**0.0021**	
		Timepoint:Direction	1,99	7.548	**0.0071**	
	Proteobacteria†	*Intercept*	1,100	235.528	< 0.0001	0.243
		Direction	1,32	0.650	0.4260	
		Timepoint	1,100	4.739	**0.0318**	
		Direction:Timepoint	1,100	24.502	**< 0.0001**	
TP1	Firmicutes‡	*Intercept*	1,24	128.956	< 0.0001	
TP3	Firmicutes‡	*Intercept*	1,24	148.341	< 0.0001	
		Direction	1,8	5.020	**0.0554**	
TP5	Firmicutes‡	*Intercept*	1,24	156.233	< 0.0001	
TP11	Firmicutes‡	*Intercept*	1,23	66.117	< 0.0001	
		Secretor	1,23	4.993	**0.0355**	
TP1	Bacteroidetes*	*Intercept*	1,24	199.426	< 0.0001	
TP3	Bacteroidetes	*Intercept*	1,24	900.370	< 0.0001	
		Direction	1,8	3.328	0.1056	
TP5	Bacteroidetes*	*Intercept*	1,24	163.012	< 0.0001	
TP11	Bacteroidetes***	*Intercept*	1,23	48.595	< 0.0001	
		Secretor	1,23	6.823	**0.0156**	
TP1	Proteobacteria†	*Intercept*	1,24	57.255	< 0.0001	
TP3	Proteobacteria†	*Intercept*	1,24	89.235	< 0.0001	
TP5	Proteobacteria†	*Intercept*	1,24	57.559	< 0.0001	
TP11	Proteobacteria†	*Intercept*	1,24	115.108	< 0.0001	
		Direction	1,8	4.587	**0.0646**	

* X^2 transformed; # $X^{1/4}$ transformed; † log(X+1) transformed; ‡ $X^{1/2}$ transformed; *** X^3 transformed; ** Timepoint- time coded as order of observation (0-3)

Table 2: Indicator genera for *Fut2* genotype, secretor status and breeding direction (gray shading highlights taxa with multiple associations to the same factor among time points).

Time point	Genus	Association	*IndVal*	*P*-Value	P_{FDR}
TP 1	*uncl. Bacteria*	nosec	0.8235	0.0105989	0.5087
TP 3	*uncl. Bacteria*	nosec	0.8130	0.0374963	0.4218
	Erysipelotrichaceae incertae sedis	nosec	0.6678	0.0035996	0.1620
	Robinsoniella	nosec	0.7378	0.0269973	0.4218
	Butyricicoccus	sec	0.6030	0.0294971	0.4218
TP 5	*Ruminococcus*	nosec	0.4880	0.0383962	1.0000
TP 11	*uncl. Porphyromonadaceae*	nosec	0.7798	0.0370963	0.4822
	Rikenella	nosec	0.5311	0.0467953	0.4822
	Ruminococcus	nosec	0.5622	0.0295970	0.4822
TP 1-11	*uncl. Bacteria*	nosec	0.7669	0.0083992	0.2100
	Ruminococcus	nosec	0.4882	0.0012999	0.0650
TP 1	*uncl. Bacteria*	*Fut2* $^{-/-}$	0.7292	0.0074993	0.3600
	uncl. Burkholderiales	*Fut2* $^{+/+}$	0.6005	0.0483952	0.9569
TP 3	*uncl. Bacteria*	*Fut2* $^{-/-}$	0.7112	0.0464954	0.6974
	Erysipelotrichaceae incertae sedis	*Fut2* $^{-/-}$	0.6481	0.0066993	0.3015
	Oscillibacter	*Fut2* $^{+/-}$	0.7268	0.0368963	0.6974
TP 5	*Marvinbryantia*	*Fut2* $^{+/+}$	0.6862	0.0178982	0.8770
TP 11	*uncl. Porphyromonadaceae*	*Fut2* $^{-/-}$	0.6614	0.0469953	0.7703
	Odoribacter	*Fut2* $^{+/+}$	0.8117	0.0004000	0.0180
TP 1-11	*uncl. Bacteria*	*Fut2* $^{-/-}$	0.6511	0.0133987	0.3166
	Ruminococcus	*Fut2* $^{-/-}$	0.4617	0.0022998	0.1150
	Oscillibacter	*Fut2* $^{+/-}$	0.6209	0.0384962	0.4780
	Butyrivibrio	*Fut2* $^{+/+}$	0.3660	0.0477952	0.4780
	uncl. Proteobacteria	*Fut2* $^{+/+}$	0.4804	0.0189981	0.3166
TP 1	*uncl. Alphaproteobacteria*	*Fut2* $^{-/-}$ dam	0.6631	0.0306969	0.1842
	Anaerotruncus	*Fut2* $^{-/-}$ dam	0.7080	0.0110989	0.1316
	uncl. Bacteroidetes	*Fut2* $^{-/-}$ dam	0.8130	0.0158984	0.1316
	Escherichia/Shigella	*Fut2* $^{-/-}$ dam	0.7002	0.0191981	0.1316
	Parabacteroides	*Fut2* $^{-/-}$ dam	0.8665	0.0189981	0.1316
	uncl. Bacteroidales	*Fut2* $^{+/+}$ dam	0.7714	0.0034997	0.1316
	Barnesiella	*Fut2* $^{+/+}$ dam	0.7813	0.0172983	0.1316
	Paraprevotella	*Fut2* $^{+/+}$ dam	0.7226	0.0384962	0.2053
	uncl. Porphyromonadaceae	*Fut2* $^{+/+}$ dam	0.8260	0.0126987	0.1316
TP 3	*uncl. Bacteroidetes*	*Fut2* $^{-/-}$ dam	0.8222	0.0038996	0.0585
	Parasutterella	*Fut2* $^{-/-}$ dam	0.8826	0.0019998	0.0450
	uncl. Prevotellaceae	*Fut2* $^{-/-}$ dam	0.8237	0.0197980	0.1485
	Alistipes	*Fut2* $^{+/+}$ dam	0.7930	0.0436956	0.2553
	Bacteroides	*Fut2* $^{+/+}$ dam	0.8673	0.0103990	0.0936
	Helicobacter	*Fut2* $^{+/+}$ dam	0.7900	0.0066993	0.0754
	uncl. Lachnospiraceae	*Fut2* $^{+/+}$ dam	0.8689	0.0008999	0.0405
TP 5	*uncl. Erysipelotrichaceae*	*Fut2* $^{-/-}$ dam	0.5058	0.0486951	0.3871
	Prevotella	*Fut2* $^{-/-}$ dam	0.8834	0.0001000	0.0049
	uncl. Prevotellaceae	*Fut2* $^{-/-}$ dam	0.8484	0.0036996	0.0906
	Robinsoniella	*Fut2* $^{-/-}$ dam	0.7265	0.0263974	0.3368
	Alistipes	*Fut2* $^{+/+}$ dam	0.8302	0.0369963	0.3626
	uncl. Rikenellaceae	*Fut2* $^{+/+}$ dam	0.7377	0.0274973	0.3368
TP 11	*uncl. Prevotellaceae*	*Fut2* $^{-/-}$ dam	0.7790	0.0236976	0.5332
	uncl. Alphaproteobacteria	*Fut2* $^{+/+}$ dam	0.7065	0.0207979	0.5332
TP 1-11	*uncl. Bacteroidetes*	*Fut2* $^{-/-}$ dam	0.7712	0.0004000	0.0067
	Escherichia/Shigella	*Fut2* $^{-/-}$ dam	0.4772	0.0016998	0.0212

Prevotella	*Fut2* $^{-/-}$ dam	0.7026	0.0002000	0.0050
uncl. Prevotellaceae	*Fut2* $^{-/-}$ dam	0.7805	0.0001000	0.0050
Bacteroides	*Fut2* $^{+/+}$ dam	0.8236	0.0026997	0.0270
Barnesiella	*Fut2* $^{+/+}$ dam	0.6493	0.0089991	0.0633
Clostridium Cluster IV	*Fut2* $^{+/+}$ dam	0.4669	0.0102990	0.0633
uncl. Firmicutes	*Fut2* $^{+/+}$ dam	0.7209	0.0455954	0.2229
uncl. Lachnospiraceae	*Fut2* $^{+/+}$ dam	0.7868	0.0037996	0.0317
Marvinbryantia	*Fut2* $^{+/+}$ dam	0.6362	0.0113989	0.0633

Table 3: Alpha diversity across all fecal sampling time points based on species richness (observed number of species), abundance distribution (Shannon H) and phylogenetic community structure (NRI/NTI).

Alpha diversity	Model Factors	*DF*	*F*-Value	*P*-Value	marg.R^2
Species Richness observed	*Intercept*	1,98	509.246	< 0.0001	0.180
	Direction	1,32	1.798	0.1894	
	Timepoint (poly) * #	2,98	0.862	0.4257	
	Direction:Timepoint (poly)	2,98	24.137	**< 0.0001**	
Shannon Entropy (X^2 transformed)	*Intercept*	1,94	813.694	< 0.0001	0.146
	Fut2	2,3	0.049	0.9522	
	Timespan (poly) †	2,94	21.978	**< 0.0001**	
	Direction	1,3	1.748	0.1961	
	Fut2:Timespan (poly)	4,94	34.422	**< 0.0001**	
	Timespan (poly):Direction	2,94	7.169	**0.0013**	
Net Relatedness Index (NRI)	*Intercept*	1,101	25.602	< 0.0001	0.097
	Secretor	1,32	7.615	**0.0095**	
	Timespan	1,101	22.348	**< 0.0001**	
Nearest Taxon Index (NTI)	*Intercept*	1,98	568.972	< 0.0001	0.760
	Secretor	1,31	4.919	**0.0340**	
	Direction	1,31	8.963	**0.0054**	
	Timespan (poly)	2,98	29.053	**< 0.0001**	
	Direction:Timespan (poly)	2,98	30.749	**< 0.0001**	

* Timepoint- time coded as order of observation (0-3); † Timespan coded as time in weeks (0, 2, 4, 10); # fitted as second order polynomial

Table 4: Community differentiation according to *Fut2* genotype and breeding direction among all fecal time points combined, based on shared abundance (Bray-Curtis), shared presence (Jaccard), phylogenetic relatedness (unweighted UniFrac) and distribution of species (Euclidean/Redundancy Analysis).

Dataset	Distance	Factors	DF	*F*-Value	*P*-Value	R^2	adj. R^2
Fecal time points (TP 1-11)	Bray-Curtis	*Fut2*	2,130	1.0072	0.3198	0.0484	0.0118
		Direction	1,130	2.1329	0.0002		
		Fut2:Direction	2,130	1.2335	0.0002		
	Jaccard	*Fut2*	2,130	1.0036	0.349	0.0477	0.0111
		Direction	1,130	2.0928	0.0002		
		Fut2:Direction	2,130	1.2071	0.0002		
	UniFrac (unweighted)	*Fut2*	2,130	1.0976	0.1394	0.0578	0.0216
		Direction	1,130	3.1389	0.0002		
		Fut2:Direction	2,130	1.3213	0.0102		
	Redundancy Analysis	*Fut2*	2,130	0.9681	0.4208	0.0757	0.0402
		Direction	1,130	4.5829	0.0002		
		Fut2:Direction	2,130	2.066	0.0002		
Gastrointestinal tract (Jejunum, Ileum, Cecum, Colon)	Bray-Curtis	*Fut2*	2,136	1.2161	0.0008	0.0508	0.0159
		Direction	1,136	2.5043	0.0002		
		Fut2:Direction	2,136	1.1730	0.001		
	Jaccard	*Fut2*	2,136	1.1832	0.0012	0.0511	0.0162
		Direction	1,136	2.6433	0.0002		
		Fut2:Direction	2,136	1.1570	0.0034		
	UniFrac (unweighted)	*Fut2*	2,136	1.105	0.0322	0.0551	0.0204
		Direction	1,136	3.1875	0.0002		
		Fut2:Direction	2,136	1.2662	0.0018		
	Redundancy Analysis	*Fut2*	2,136	1.7277	0.0004	0.070	0.0358
		Direction	1,136	4.0415	0.0002		
		Fut2:Direction	2,136	1.3690	0.0106		

References

1. Walker AW, Duncan SH, McWilliam Leitch EC, Child MW, Flint HJ. pH and Peptide Supply Can Radically Alter Bacterial Populations and Short-Chain Fatty Acid Ratios within Microbial Communities from the Human Colon. Appl Environ Microbiol. 2005;71(7):3692-700. doi: 10.1128/aem.71.7.3692-3700.2005.

2. Tasse L, Bercovici J, Pizzut-Serin S, Robe P, Tap J, Klopp C, et al. Functional metagenomics to mine the human gut microbiome for dietary fiber catabolic enzymes. Genome Res. 2010;20(11):1605-12. doi: 10.1101/gr.108332.110.

3. Abt Michael C, Osborne Lisa C, Monticelli Laurel A, Doering Travis A, Alenghat T, Sonnenberg Gregory F, et al. Commensal Bacteria Calibrate the Activation Threshold of Innate Antiviral Immunity. Immunity. 2012;37(1):158-70. doi: http://dx.doi.org/10.1016/j.immuni.2012.04.011.

4. Hill DA, Siracusa MC, Abt MC, Kim BS, Kobuley D, Kubo M, et al. Commensal bacteria-derived signals regulate basophil hematopoiesis and allergic inflammation. Nat Med. 2012;18(4):538-46. doi: http://www.nature.com/nm/journal/v18/n4/abs/nm.2657.html#supplementary-information.

5. Hill DA, Artis D. The influence of commensal bacteria-derived signals on basophil-associated allergic inflammation. Gut Microbes. 2013;4(1):76-83.

6. Sonnenberg GF, Monticelli LA, Alenghat T, Fung TC, Hutnick NA, Kunisawa J, et al. Innate Lymphoid Cells Promote Anatomical Containment of Lymphoid-Resident Commensal Bacteria. Science. 2012;336(6086):1321-5. doi: 10.1126/science.1222551.

7. Endt K, Stecher B, Chaffron S, Slack E, Tchitchek N, Benecke A, et al. The Microbiota Mediates Pathogen Clearance from the Gut Lumen after Non-Typhoidal Salmonella Diarrhea. PLoS Pathog. 2010;6(9):e1001097. doi: 10.1371/journal.ppat.1001097.

8. Hooper LV. Bacterial contributions to mammalian gut development. Trends in Microbiology. 2004;12(3):129-34. PubMed PMID: WOS:000220460900007.

9. Ley RE, Turnbaugh PJ, Klein S, Gordon JI. Microbial ecology - Human gut microbes associated with obesity. Nature. 2006;444(7122):1022-3. doi: 10.1038/nature4441022a. PubMed PMID: ISI:000242971100042.

10. Turnbaugh P, Ley R, Mahowald M, Magrini V, Mardis E, Gordon J. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature. 2006;444(7122):1027 - 31. PubMed PMID: doi:10.1038/nature05414.

11. Wen L, Ley RE, Volchkov PY, Stranges PB, Avanesyan L, Stonebraker AC, et al. Innate immunity and intestinal microbiota in the development of Type 1 diabetes. Nature. 2008;455(7216):1109-13. doi: http://www.nature.com/nature/journal/v455/n7216/suppinfo/nature07336_S1.html.

12. Vijay-Kumar M, Aitken JD, Carvalho FA, Cullender TC, Mwangi S, Srinivasan S, et al. Metabolic Syndrome and Altered Gut Microbiota in Mice Lacking Toll-Like Receptor 5. Science. 2010;328(5975):228-31. doi: 10.1126/science.1179721.

13. Frank DN, St. Amand AL, Feldman RA, Boedeker EC, Harpaz N, Pace NR. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. Proceedings of the National Academy of Sciences. 2007;104(34):13780-5. Epub 2007/08/19. doi: 10.1073/pnas.0706625104. PubMed PMID: 17699621; PubMed Central PMCID: PMC1959459.

14. Rausch P, Rehman A, Künzel S, Häsler R, Ott SJ, Schreiber S, et al. Colonic mucosa-associated microbiota is influenced by an interaction of Crohn disease and *FUT2* (Secretor) genotype. Proceedings of the National Academy of Sciences. 2011;108(47):19030-5. Epub 2011/11/10. doi: 10.1073/pnas.1106408108. PubMed PMID: 22068912; PubMed Central PMCID: PMC3223430.

15. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, et al. A core gut microbiome in obese and lean twins. Nature. 2009;457(7228):480-U7. doi: 10.1038/nature07540. PubMed PMID: WOS:000262519200047.

16. Benson AK, Kelly SA, Legge R, Ma F, Low SJ, Kim J, et al. 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. 2010;107(44):18933-8. doi: 10.1073/pnas.1007028107.

17. McKnite AM, Perez-Munoz ME, Lu L, Williams EG, Brewer S, Andreux PA, et al. Murine Gut Microbiota Is Defined by Host Genetics and Modulates Variation of Metabolic Traits. PLoS One. 2012;7(6). doi: e39191

10.1371/journal.pone.0039191. PubMed PMID: WOS:000305583300114.

18. Manichanh C, Borruel N, Casellas F, Guarner F. The gut microbiota in IBD. Nat Rev Gastroenterol Hepatol. 2012;9(10):599-608.

19. Goodrich Julia K, Waters Jillian L, Poole Angela C, Sutter Jessica L, Koren O, Blekhman R, et al. Human Genetics Shape the Gut Microbiome. Cell. 2014;159(4):789-99. doi: http://dx.doi.org/10.1016/j.cell.2014.09.053.

20. Knight R, Jansson J, Field D, Fierer N, Desai N, Fuhrman JA, et al. Unlocking the potential of metagenomics through replicated experimental design. Nat Biotech. 2012;30(6):513-20.

21. Bazzaz FA. Physiological ecology of plant succession. Annu Rev Ecol Syst. 1979;10:351-71. doi: 10.1146/annurev.es.10.110179.002031. PubMed PMID: WOS:A1979HU59600014.

22. Caporaso JG, Lauber C, Costello E, Berg-Lyons D, Gonzalez A, Stombaugh J, et al. Moving pictures of the human microbiome. Genome Biology. 2011;12(5):R50. PubMed PMID: doi:10.1186/gb-2011-12-5-r50.

23. Faith JJ, Guruge JL, Charbonneau M, Subramanian S, Seedorf H, Goodman AL, et al. The Long-Term Stability of the Human Gut Microbiota. Science. 2013;341(6141):1237439. Epub 2013/07/06. doi: 10.1126/science.1237439. PubMed PMID: 23828941; PubMed Central PMCID: PMC3791589.

24. La Rosa PS, Warner BB, Zhou Y, Weinstock GM, Sodergren E, Hall-Moore CM, et al. Patterned progression of bacterial populations in the premature infant gut. Proceedings of the National Academy of Sciences. 2014;111(34):12522-7. doi: 10.1073/pnas.1409497111.

25. Trosvik P, Stenseth NC, Rudi K. Convergent temporal dynamics of the human infant gut microbiota. ISME J. 2009;4(2):151-8. Epub 2009/08/28. doi: http://www.nature.com/ismej/journal/v4/n2/suppinfo/ismej200996s1.html. PubMed PMID: 19710708.

26. Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, Fierer N, et al. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. Proceedings of the National Academy of Sciences. 2010;107(26):11971-5. doi: 10.1073/pnas.1002601107.

27. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. Human gut microbiome viewed across age and geography. Nature. 2012;486(7402):222-7. Epub 2012/06/16. doi: http://www.nature.com/nature/journal/vaop/ncurrent/abs/nature11053.html#supplementary-information. PubMed PMID: 22699611; PubMed Central PMCID: PMC3376388.

28. Staubach F, Künzel S, Baines AC, Yee A, McGee BM, Bäckhed F, et al. Expression of the blood-group-related glycosyltransferase B4galnt2 influences the intestinal microbiota in mice. ISME J. 2012;6(7):1345-55. Epub 2012/01/27. doi:

http://www.nature.com/ismej/journal/vaop/ncurrent/suppinfo/ismej2011204s1.html. PubMed PMID: 22278669; PubMed Central PMCID: PMC3379640.

29. Folseraas T, Melum E, Rausch P, Juran BD, Ellinghaus E, Shiryaev A, et al. Extended analysis of a genome-wide association study in primary sclerosing cholangitis detects multiple novel risk loci. Journal of Hepatology. 2012;57(0):366-75. Epub 2012/04/24. doi: 10.1016/j.jhep.2012.03.031. PubMed PMID: 22521342; PubMed Central PMCID: PMC3399030.

30. Koropatkin NM, Cameron EA, Martens EC. How glycan metabolism shapes the human gut microbiota. Nat Rev Micro. 2012;10(5):323-35. Epub 2012/04/12. doi: Doi 10.1038/Nrmicro2746. PubMed PMID: ISI:000302938700010; PubMed Central PMCID: PMC4005082.

31. Ilver D, Arnqvist A, Ogren J, Frick I-M, Kersulyte D, Incecik ET, et al. Helicobacter pylori Adhesin Binding Fucosylated Histo-Blood Group Antigens Revealed by Retagging. Science. 1998;279(5349):373-7. doi: 10.1126/science.279.5349.373.

32. Lindesmith L, Moe C, Marionneau S, Ruvoen N, Jiang X, Lindbland L, et al. Human susceptibility and resistance to Norwalk virus infection. Nature Medicine. 2003;9(5):548-53. Epub 2003/04/15. doi: 10.1038/nm860. PubMed PMID: WOS:000182610600035.

33. Ruiz-Palacios GM, Cervantes LE, Ramos P, Chavez-Munguia B, Newburg DS. Campylobacter jejuni Binds Intestinal H(O) Antigen (Fucα1, 2Galβ1, 4GlcNAc), and Fucosyloligosaccharides of Human Milk Inhibit Its Binding and Infection. Journal of Biological Chemistry. 2003;278(16):14112-20. Epub 2003/02/04. doi: 10.1074/jbc.M207744200. PubMed PMID: 12562767.

34. Hurd EA, Holmen JM, Hansson GC, Domino SE. Gastrointestinal mucins of *Fut2*-null mice lack terminal fucosylation without affecting colonization by Candida albicans. Glycobiology. 2005;15(10):1002-7. doi: 10.1093/glycob/cwi089. PubMed PMID: ISI:000232103500013.

35. Kindberg E, Hejdeman B, Bratt G, Wahren B, Lindblom B, Hinkula J, et al. A nonsense mutation (428G→A) in the fucosyltransferase *FUT2* gene affects the progression of HIV-1 infection. AIDS. 2006;20(5):685-9. PubMed PMID: WOS:000236421000007.

36. Hooper LV, Midtvedt T, Gordon JI. How host-microbial interactions shape the nutrient environment of the mammalian intestine. Annual Review of Nutrition. 2002;22:283-307. doi: 10.1146/annurev.nutr.22.011602.092259. PubMed PMID: WOS:000179917900014.

37. Hooper LV, Gordon JI. Glycans as legislators of host–microbial interactions: spanning the spectrum from symbiosis to pathogenicity. Glycobiology. 2001;11(2):1R-10R. doi: 10.1093/glycob/11.2.1R.

38. Nanthakumar NN, Dai D, Newburg DS, Walker WA. The role of indigenous microflora in the development of murine intestinal fucosyl- and sialyltransferases. The FASEB Journal. 2003;17(1):44-6. doi: 10.1096/fj.02-0031fje.

39. Meng D, Newburg DS, Young C, Baker A, Tonkonogy SL, Sartor RB, et al. Bacterial symbionts induce a *FUT2*-dependent fucosylated niche on colonic epithelium via ERK and JNK signaling. Am J Physiol Gastrointest Liver Physiol. 2007;293(4):G780-7. doi: 10.1152/ajpgi.00010.2007.

40. Liu Y, Koda Y, Soejima M, Pang H, Schlaphoff T, du Toit ED, et al. Extensive polymorphism of the *FUT2* gene in an African (Xhosa) population of South Africa. Human Genetics. 1998;103(2):204-10. doi: 10.1007/s004390050808.

41. Koda Y, Tachida H, Soejima M, Takenaka O, Kimura H. Ancient Origin of the Null Allele se428 of the Human ABO-Secretor Locus (*FUT2*). Journal of Molecular Evolution. 2000;50(3):243-8. doi: 10.1007/s002399910028.

42. Koda Y, Tachida H, Pang H, Liu Y, Soejima M, Ghaderi AA, et al. Contrasting Patterns of Polymorphisms at the ABO-Secretor Gene (*FUT2*) and Plasma alpha(1,3)Fucosyltransferase Gene (FUT6) in Human Populations. Genetics. 2001;158(2):747-56.

43. Pang H, Koda Y, Soejima M, Fujitani N, Ogaki T, Saito A, et al. Polymorphism of the human ABO-secretor locus (*FUT2*) in four populations in Asia: indication of distinct Asian subpopulations. Annals of Human Genetics. 2001;65:429-37. PubMed PMID: WOS:000173275600003.

44. Henry S, Mollicone R, Fernandez P, Samuelsson B, Oriol R, Larson G. Molecular basis for erythrocyte Le(a+b+) and salivary ABH partial-secretor phenotypes: expression of a *FUT2* secretor allele with an A→T mutation at nucleotide 385 correlates with reduced α(1,2) fucosyltransferase activity. Glycoconjugate Journal. 1996;13(6):985-93. doi: 10.1007/bf01053194.

45. Wacklin P, Tuimala J, Nikkilä J, Sebastian T, Mäkivuokko H, Alakulppi N, et al. Faecal Microbiota Composition in Adults Is Associated with the *FUT2* Gene Determining the Secretor Status. PLoS One. 2014;9(4):e94863. Epub 2014/04/16. doi:

10.1371/journal.pone.0094863. PubMed PMID: 24733310; PubMed Central PMCID: PMC3986271.

46. McGovern DPB, Jones MR, Taylor KD, Marciante K, Yan X, Dubinsky M, et al. Fucosyltransferase 2 (*FUT2*) non-secretor status is associated with Crohn's disease. Hum Mol Genet. 2010:1-9. Epub June 22, 2010. doi: 10.1093/hmg/ddq248.

47. Morrow AL, Ruiz-Palacios GM, Jiang X, Newburg DS. Human-Milk Glycans That Inhibit Pathogen Binding Protect Breast-feeding Infants against Infectious Diarrhea. J Nutr. 2005;135(5):1304-7.

48. Newburg DS, Pickering LK, McCluer RH, Cleary TG. Fucosylated Oligosaccharides of Human Milk Protect Suckling Mice from Heat-Stabile Enterotoxin of Escherichia coli. The Journal of Infectious Diseases. 1990;162(5):1075-80.

49. Turner J-E, Stockinger B, Helmby H. IL-22 Mediates Goblet Cell Hyperplasia and Worm Expulsion in Intestinal Helminth Infection. PLoS Pathog. 2013;9(10):e1003698. doi: 10.1371/journal.ppat.1003698.

50. Pham Tu Anh N, Clare S, Goulding D, Arasteh Julia M, Stares Mark D, Browne Hilary P, et al. Epithelial IL-22RA1-Mediated Fucosylation Promotes Intestinal Colonization Resistance to an Opportunistic Pathogen. Cell Host & Microbe. 2014;16(4):504-16. Epub 2014/09/30. doi: 10.1016/j.chom.2014.08.017. PubMed PMID: PMC4190086; PubMed Central PMCID: PMC4190086.

51. Pickard JM, Maurice CF, Kinnebrew MA, Abt MC, Schenten D, Golovkina TV, et al. Rapid fucosylation of intestinal epithelium sustains host-commensal symbiosis in sickness. Nature. 2014;advance online publication. doi: 10.1038/nature13823

http://www.nature.com/nature/journal/vaop/ncurrent/abs/nature13823.html#supplementary-information.

52. Goto Y, Obata T, Kunisawa J, Sato S, Ivanov II, Lamichhane A, et al. Innate lymphoid cells regulate intestinal epithelial cell glycosylation. Science. 2014;345(6202). doi: 10.1126/science.1254009.

53. Frank DN, Zhu W, Sartor RB, Li E. Investigating the biological and clinical significance of human dysbioses. Trends in Microbiology. 2011;19(9):427-34. doi: 10.1016/j.tim.2011.06.005.

54. Shanahan F. Crohn's disease. The Lancet. 2002;359(9300):62-9. Epub 2002/01/26. doi: Doi: 10.1016/s0140-6736(02)07284-7. PubMed PMID: 11809204.

55. Pacheco AR, Curtis MM, Ritchie JM, Munera D, Waldor MK, Moreira CG, et al. Fucose sensing regulates bacterial intestinal colonization. Nature. 2012;advance online publication. doi: http://www.nature.com/nature/journal/vaop/ncurrent/abs/nature11623.html#supplementary-information.

56. Chow WL, Lee YK. Free fucose is a danger signal to human intestinal epithelial cells. British Journal of Nutrition. 2008;99(3):449-54. doi: 10.1017/s0007114507812062. PubMed PMID: WOS:000253583000001.

57. Watanabe M, Kinoshita H, Nitta M, Yukishita R, Kawai Y, Kimura K, et al. Identification of a new adhesin-like protein from Lactobacillus mucosae ME-340 with specific affinity to the human blood group A and B antigens. Journal of Applied Microbiology. 2010;109(3):927-35. doi: 10.1111/j.1365-2672.2010.04719.x.

58. Saadi AT, Weir DM, Poxton IR, Stewart J, Essery SD, Caroline Blackwell C, et al. Isolation of an adhesin from Staphylococcus aureus that binds Lewis blood group antigen and its relevance to sudden infant death syndrome. FEMS Immunology & Medical Microbiology. 1994;8(4):315-20. doi: 10.1111/j.1574-695X.1994.tb00458.x.

59. Uchida H, Kawai Y, Kinoshita H, Kitazawa H, Miura K, Shiiba K, et al. Lactic Acid Bacteria (LAB) Bind to Human B- or H-Antigens Expressed on Intestinal Mucosa. Bioscience, Biotechnology, and Biochemistry. 2006;70(12):3073-6.

60. Kinoshita H, Wakahara N, Watanabe M, Kawasaki T, Matsuo H, Kawai Y, et al. Cell surface glyceraldehyde-3-phosphate dehydrogenase (GAPDH) of Lactobacillus plantarum LA 318 recognizes human A and B blood group antigens. Research in Microbiology. 2008;159(9-10):685-91. doi: DOI: 10.1016/j.resmic.2008.07.005.

61. Shannon CE. A mathematical theory of communication. Bell System Technical Journal. 1948;27. doi: citeulike-article-id:1584479.

62. Webb CO. Exploring the phylogenetic structure of ecological communities: An example for rain forest trees. Am Nat. 2000;156:145-55.

63. Ter Braak CJF. CANONICAL CORRESPONDENCE-ANALYSIS - A NEW EIGENVECTOR TECHNIQUE FOR MULTIVARIATE DIRECT GRADIENT ANALYSIS. Ecology. 1986;67(5):1167-79. PubMed PMID: WOS:A1986E197900005.

64. Legendre P, Anderson MJ. Distance-based redundancy analysis: Testing multispecies responses in multifactorial ecological experiments. Ecological Monographs. 1999;69(1):1-24. PubMed PMID: WOS:000078484900001.

65. Legendre P, Legendre L. Numerical ecology. Second English edition. Developments in Environmental Modelling. 1998;20:i-xv, 1-853. PubMed PMID: ZOOREC:ZOOR13500057538.

66. Ubeda C, Lipuma L, Gobourne A, Viale A, Leiner I, Equinda M, et al. Familial transmission rather than defective innate immunity shapes the distinct intestinal microbiota of TLR-deficient mice. The Journal of experimental medicine. 2012;209(8):1445-56. doi: 10.1084/jem.20120504.

67. Friedman J, Alm EJ. Inferring Correlation Networks from Genomic Survey Data. PLoS Comput Biol. 2012;8(9):e1002687. doi: 10.1371/journal.pcbi.1002687.

68. Brin S, Page L. The anatomy of a large-scale hypertextual Web search engine. Computer Networks and Isdn Systems. 1998;30(1-7):107-17. doi: 10.1016/s0169-7552(98)00110-x. PubMed PMID: WOS:000073360600013.

69. Freeman LC. CENTRALITY IN SOCIAL NETWORKS CONCEPTUAL CLARIFICATION. Social Networks. 1979;1(3):215-39. doi: 10.1016/0378-8733(78)90021-7. PubMed PMID: WOS:A1979GL19400002.

70. Newman MEJ, Girvan M. Finding and evaluating community structure in networks. Physical Review E. 2004;69(2). doi: 026113

10.1103/PhysRevE.69.026113. PubMed PMID: WOS:000220255500019.

71. Albert R, Jeong H, Barabasi A-L. Error and attack tolerance of complex networks. Nature. 2000;406(6794):378-82.

72. Ségurel L, Gao Z, Przeworski M. Ancestry runs deeper than blood: The evolutionary history of ABO points to cryptic variation of functional importance. Bioessays. 2013;35(10):862-7. doi: 10.1002/bies.201300030.

73. Johnsen JM, Teschke M, Pavlidis P, McGee BM, Tautz D, Ginsburg D, et al. Selection on cis-Regulatory Variation at B4galnt2 and Its Influence on von Willebrand Factor in House Mice. Molecular Biology and Evolution. 2009;26(3):567-78. Epub 2008/12/18. doi: 10.1093/molbev/msn284. PubMed PMID: ISI:000263420900009; PubMed Central PMCID: PMC2727395.

74. Linnenbrink M, Johnsen JM, Montero I, Brzezinski CR, Harr B, Baines JF. Long-term balancing selection at the blood group-related gene B4galnt2 in the genus Mus (Rodentia; Muridae). Molecular Biology and Evolution. 2011;28(11):2999-3003. Epub 2011/06/10. doi: 10.1093/molbev/msr150. PubMed PMID: 21652612.

75. Hooper LV, Gordon JI. Commensal Host-Bacterial Relationships in the Gut. Science. 2001;292(5519):1115-8. doi: 10.1126/science.1058709.

76. Hooper LV, Wong MH, Thelin A, Hansson L, Falk PG, Gordon JI. Molecular Analysis of Commensal Host-Microbial Relationships in the Intestine. Science. 2001;291(5505):881-4. doi: 10.1126/science.291.5505.881.

77. Freitas M, Axelsson LG, Cayuela C, Midtvedt T, Trugnan G. Microbial-host interactions specifically control the glycosylation pattern in intestinal mouse mucosa. Histochem Cell Biol. 2002;118(2):149-61. doi: 10.1007/s00418-002-0432-0. PubMed PMID: WOS:000177938800007.

78. Robbe C, Capon C, Maes E, Rousset M, Zweibaum A, Zanetta J-P, et al. Evidence of Regio-specific Glycosylation in Human Intestinal Mucins. Journal of Biological Chemistry. 2003;278(47):46337-48. doi: 10.1074/jbc.M302529200.

79. Johansson MEV, Larsson JMH, Hansson GC. The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host–microbial interactions. Proceedings of the National Academy of Sciences. 2011;108(Supplement 1):4659-65. doi: 10.1073/pnas.1006451107.

80. Ng KM, Ferreyra JA, Higginbottom SK, Lynch JB, Kashyap PC, Gopinath S, et al. Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. Nature. 2013;advance online publication(7469):96-9. Epub 2013/09/03. doi: 10.1038/nature12503

http://www.nature.com/nature/journal/vaop/ncurrent/abs/nature12503.html#supplementary-information. PubMed PMID: 23995682; PubMed Central PMCID: PMC3825626.

81. Zaneveld JR, Lozupone C, Gordon JI, Knight R. Ribosomal RNA diversity predicts genome diversity in gut bacteria and their relatives. Nucl Acids Res. 2010:gkq066. doi: 10.1093/nar/gkq066.

82. Violle C, Nemergut DR, Pu Z, Jiang L. Phylogenetic limiting similarity and competitive exclusion. Ecology Letters. 2011;14(8):782-7. doi: 10.1111/j.1461-0248.2011.01644.x. PubMed PMID: 21672121.

83. Valiente-Banuet A, Verdú M. Facilitation can increase the phylogenetic diversity of plant communities. Ecology Letters. 2007;10(11):1029-36. doi: 10.1111/j.1461-0248.2007.01100.x.

84. Jones EI, Nuismer SL, Gomulkiewicz R. Revisiting Darwin's conundrum reveals a twist on the relationship between phylogenetic distance and invasibility. Proceedings of the National Academy of Sciences. 2013;110(51):20627-32. doi: 10.1073/pnas.1310247110.

85. Tilman D. Biodiversity: Population Versus Ecosystem Stability. Ecology. 1996;77(2):350-63. doi: 10.2307/2265614.

86. Pfisterer AB, Schmid B. Diversity-dependent production can decrease the stability of ecosystem functioning. Nature. 2002;416(6876):84-6.

87. Isbell F, Calcagno V, Hector A, Connolly J, Harpole WS, Reich PB, et al. High plant diversity is needed to maintain ecosystem services. Nature. 2011;advance online publication. doi: http://www.nature.com/nature/journal/vaop/ncurrent/abs/nature10282.html#supplementary-information.

88. Doak DF, Bigger D, Harding EK, Marvier MA, O'Malley RE, Thomson D. The statistical inevitability of stability-diversity relationships in community ecology. American Naturalist. 1998;151(3):264-76. doi: 10.1086/286117. PubMed PMID: WOS:000072128600006.

89. Yachi S, Loreau M. Biodiversity and ecosystem productivity in a fluctuating environment: The insurance hypothesis. Proceedings of the National Academy of Sciences of the United States of America. 1999;96(4):1463-8. doi: 10.1073/pnas.96.4.1463. PubMed PMID: WOS:000078698400054.

90. Loreau M. Linking biodiversity and ecosystems: towards a unifying ecological theory2010 2010-01-12 00:00:00. 49-60 p.

91. Viverge D, Grimmonprez L, Cassanas G, Bardet L, Solere M. Discriminant Carbohydrate Components of Human Milk According to Donor Secretor Types. Journal of Pediatric Gastroenterology and Nutrition. 1990;11(3):365-70.

92. Rogier EW, Frantz AL, Bruno MEC, Wedlund L, Cohen DA, Stromberg AJ, et al. Secretory antibodies in breast milk promote long-term intestinal homeostasis by regulating the gut microbiota and host gene expression. Proceedings of the National Academy of Sciences. 2014. doi: 10.1073/pnas.1315792111.

93. Jiang X, Huang P, Zhong W, Tan M, Farkas T, Morrow Ardythe L, et al. Human Milk Contains Elements That Block Binding of Noroviruses to Human Histo–Blood Group Antigens in Saliva. The Journal of Infectious Diseases. 2004;190(10):1850-9. doi: doi:10.1086/425159. PubMed PMID: 15499543.

94. Weiss GA, Chassard C, Hennet T. Selective proliferation of intestinal Barnesiella under fucosyllactose supplementation in mice. Br J Nutr. 2014:1-9. Epub 2014/01/15. doi: S0007114513004200 [pii]

10.1017/S0007114513004200 [doi]. PubMed PMID: 24411010.

95. Domino SE, Hurd EA. LacZ expression in *Fut2*-LacZ reporter mice reveals estrogen-regulated endocervical glandular expression during estrous cycle, hormone replacement, and pregnancy. Glycobiology. 2004;14(2):169-75. doi: 10.1093/glycob/cwh019.

96. Hurd EA, Domino SE. Increased susceptibility of secretor factor gene *Fut2*-null mice to experimental vaginal candidiasis. Infection and Immunity. 2004;72(7):4279-81. doi: 10.1128/iai.72.7.4279-4281.2004. PubMed PMID: WOS:000222282800066.

97. Gajer P, Brotman RM, Bai G, Sakamoto J, Schütte UME, Zhong X, et al. Temporal Dynamics of the Human Vaginal Microbiota. Science Translational Medicine. 2012;4(132):132ra52. doi: 10.1126/scitranslmed.3003605.

98. Palmer C, Bik EM, DiGiulio DB, Relman DA, Brown PO. Development of the Human Infant Intestinal Microbiota. PLoS Biol. 2007;5(7):e177.

99. Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, Knight R, et al. Succession of microbial consortia in the developing infant gut microbiome. Proceedings of the National Academy of Sciences. 2011;108(Supplement 1):4578-85. doi: 10.1073/pnas.1000081107.

100. Pantoja-Feliciano IG, Clemente JC, Costello EK, Perez ME, Blaser MJ, Knight R, et al. Biphasic assembly of the murine intestinal microbiota during early development. ISME J. 2013;7(6):1112-5. doi: 10.1038/ismej.2013.15.

101. Sjogren YM, Tomicic S, Lundberg A, Bottcher MF, Bjorksten B, Sverremark-Ekstrom E, et al. Influence of early gut microbiota on the maturation of childhood mucosal and systemic immune responses. Clinical and Experimental Allergy. 2009;39(12):1842-51. doi: 10.1111/j.1365-2222.2009.03326.x. PubMed PMID: WOS:000271774300009.

102. Wesemann DR, Portuguese AJ, Meyers RM, Gallagher MP, Cluff-Jones K, Magee JM, et al. Microbial colonization influences early B-lineage development in the gut lamina propria. Nature. 2013;501(7465):112-5. doi: 10.1038/nature12496

http://www.nature.com/nature/journal/v501/n7465/abs/nature12496.html#supplementary-information.

103. Russell SL, Gold MJ, Hartmann M, Willing BP, Thorson L, Wlodarska M, et al. Early life antibiotic-driven changes in microbiota enhance susceptibility to allergic asthma. EMBO

Rep. 2012;advance online publication. doi: http://www.nature.com/embor/journal/vaop/ncurrent/suppinfo/embor201232a_S1.html.

104. Olszak T, An D, Zeissig S, Vera MP, Richter J, Franke A, et al. Microbial Exposure During Early Life Has Persistent Effects on Natural Killer T Cell Function. Science. 2012. doi: 10.1126/science.1219328.

105. Jeong H, Tombor B, Albert R, Oltvai ZN, Barabasi AL. The large-scale organization of metabolic networks. Nature. 2000;407(6804):651-4. doi: http://www.nature.com/nature/journal/v407/n6804/suppinfo/407651a0_S1.html.

106. Memmott J, Waser NM, Price MV. Tolerance of pollination networks to species extinctions. Proceedings of the Royal Society B-Biological Sciences. 2004;271(1557):2605-11. doi: 10.1098/rspb.2004.2909. PubMed PMID: WOS:000226419100011.

107. Iyer S, Killingback T, Sundaram B, Wang Z. Attack Robustness and Centrality of Complex Networks. PLoS One. 2013;8(4):e59613. doi: 10.1371/journal.pone.0059613.

108. Widder S, Besemer K, Singer GA, Ceola S, Bertuzzo E, Quince C, et al. Fluvial network organization imprints on microbial co-occurrence networks. Proceedings of the National Academy of Sciences. 2014;111(35):12799-804. doi: 10.1073/pnas.1411723111.

109. Tong M, Li X, Wegener Parfrey L, Roth B, Ippoliti A, Wei B, et al. A Modular Organization of the Human Intestinal Mucosal Microbiota and Its Association with Inflammatory Bowel Disease. PLoS One. 2013;8(11):e80702. doi: 10.1371/journal.pone.0080702.

110. Bastolla U, Fortuna MA, Pascual-Garcia A, Ferrera A, Luque B, Bascompte J. The architecture of mutualistic networks minimizes competition and increases biodiversity. Nature. 2009;458(7241):1018-20. doi: http://www.nature.com/nature/journal/v458/n7241/suppinfo/nature07950_S1.html.

111. Suweis S, Simini F, Banavar JR, Maritan A. Emergence of structural and dynamical properties of ecological mutualistic networks. Nature. 2013;500(7463):449-52. doi: 10.1038/nature12438

http://www.nature.com/nature/journal/v500/n7463/abs/nature12438.html#supplementary-information.

112. Coyne MJ, Chatzidaki-Livanis M, Paoletti LC, Comstock LE. Role of glycan synthesis in colonization of the mammalian gut by the bacterial symbiont Bacteroides fragilis.

Proceedings of the National Academy of Sciences. 2008;105(35):13099-104. doi: 10.1073/pnas.0804220105.

113. Petchey OL, Eklof A, Borrvall C, Ebenman B. Trophically unique species are vulnerable to cascading extinction. American Naturalist. 2008;171(5):568-79. doi: 10.1086/587068. PubMed PMID: WOS:000255212900004.

114. Berg S, Christianou M, Jonsson T, Ebenman B. Using sensitivity analysis to identify keystone species and keystone links in size-based food webs. Oikos. 2011;120(4):510-9. doi: 10.1111/j.1600-0706.2010.18864.x.

115. Curtsdotter A, Binzer A, Brose U, de Castro F, Ebenman B, Eklöf A, et al. Robustness to secondary extinctions: Comparing trait-based sequential deletions in static and dynamic food webs. Basic and Applied Ecology. 2011;12(7):571-80. doi: http://dx.doi.org/10.1016/j.baae.2011.09.008.

116. Iwamori M, Domino SE. Tissue-specific loss of fucosylated glycolipids in mice with targeted deletion of alpha(1,2)fucosyltransferase genes. Biochemical Journal. 2004;380:75-81. doi: 10.1042/bj20031668. PubMed PMID: WOS:000221735000008.

117. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: Open Source, Platform-independent, Community-supported Software for Describing and Comparing Microbial Communities. Appl Environ Microbiol. 2009;75(23):7537-41. Epub October 2, 2009. doi: 10.1128/aem.01541-09. PubMed PMID: 19801464; PubMed Central PMCID: PMC2786419.

118. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26(19):2460-1. Epub 2010/08/17. doi: 10.1093/bioinformatics/btq461. PubMed PMID: 20709691.

119. Cole JR, Chai B, Marsh TL, Farris RJ, Wang Q, Kulam SA, et al. The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. Nucl Acids Res. 2003;31(1):442-3. doi: 10.1093/nar/gkg039.

120. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. Appl Environ Microbiol. 2007;73(16):5261-7. Epub 2007/06/26. doi: 10.1128/aem.00062-07. PubMed PMID: 17586664; PubMed Central PMCID: PMC1950982.

121. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, et al. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucl Acids Res. 2007;35(21):7188-96. doi: 10.1093/nar/gkm864.

122. Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLoS One. 2010;5(3):e9490. Epub 2010/03/13. doi: 10.1371/journal.pone.0009490. PubMed PMID: 20224823; PubMed Central PMCID: PMC2835736.

123. Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, et al. Picante: R tools for integrating phylogenies and ecology. Bioinformatics. 2010;26(11):1463-4. Epub 2010/04/17. doi: 10.1093/bioinformatics/btq166. PubMed PMID: 20395285.

124. Oksanen J, Blanchet FG, Kindt R, Legendre P, O'Hara RB, Simpson GL, et al. vegan: Community Ecology Package. 1.17-6 ed: http://CRAN.R-project.org; 2011.

125. Gotelli NJ. NULL MODEL ANALYSIS OF SPECIES CO-OCCURRENCE PATTERNS. Ecology. 2000;81(9):2606-21. doi: 10.1890/0012-9658(2000)081[2606:nmaosc]2.0.co;2.

126. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. Applied and Environmental Microbiology. 2005;71(12):8228-35. doi: 10.1128/aem.71.12.8228-8235.2005. PubMed PMID: WOS:000234417600073.

127. McArdle BH, Anderson MJ. Fitting multivariate models to community data: A comment on distance-based redundancy analysis. Ecology. 2001;82(1):290-7. doi: Doi 10.1890/0012-9658(2001)082[0290:Fmmtcd]2.0.Co;2. PubMed PMID: ISI:000166488200024.

128. Bartoń K. MuMIn: multi-model inference, R package version 1.9.13. 2013.

129. Pinheiro J, Bates D, DebRoy S, Sarkar D, Team RDC. nlme: Linear and Nonlinear Mixed Effects Models. http://CRAN.R-project.org; 2011.

130. De Cáceres M, Legendre P, Moretti M. Improving indicator species analysis by combining groups of sites. Oikos. 2010;119(10):1674-84. doi: 10.1111/j.1600-0706.2010.18334.x.

131. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological). 1995;57(1):289-300. PubMed PMID: ISI:A1995QE45300017.

132. Csardi G, Nepusz T. The igraph Software Package for Complex Network Research. InterJournal. 2006;Complex Systems:1695. doi: citeulike-article-id:3443126.

133. Viger F, Latapy M. Efficient and Simple Generation of Random Simple Connected Graphs with Prescribed Degree Sequence. In: Wang L, editor. Computing and Combinatorics. Lecture Notes in Computer Science. 3595: Springer Berlin Heidelberg; 2005. p. 440-9.

134. Erdős P, Rényi A. On random graphs. Publicationes Mathematicae Debrecen. 1959;6:290-7.

135. Goh KI, Kahng B, Kim D. Universal behavior of load distribution in scale-free networks. Physical Review Letters. 2001;87(27). doi: 10.1103/PhysRevLett.87.278701. PubMed PMID: WOS:000173040800062.

Supplementary Material

Supplementary Figures:

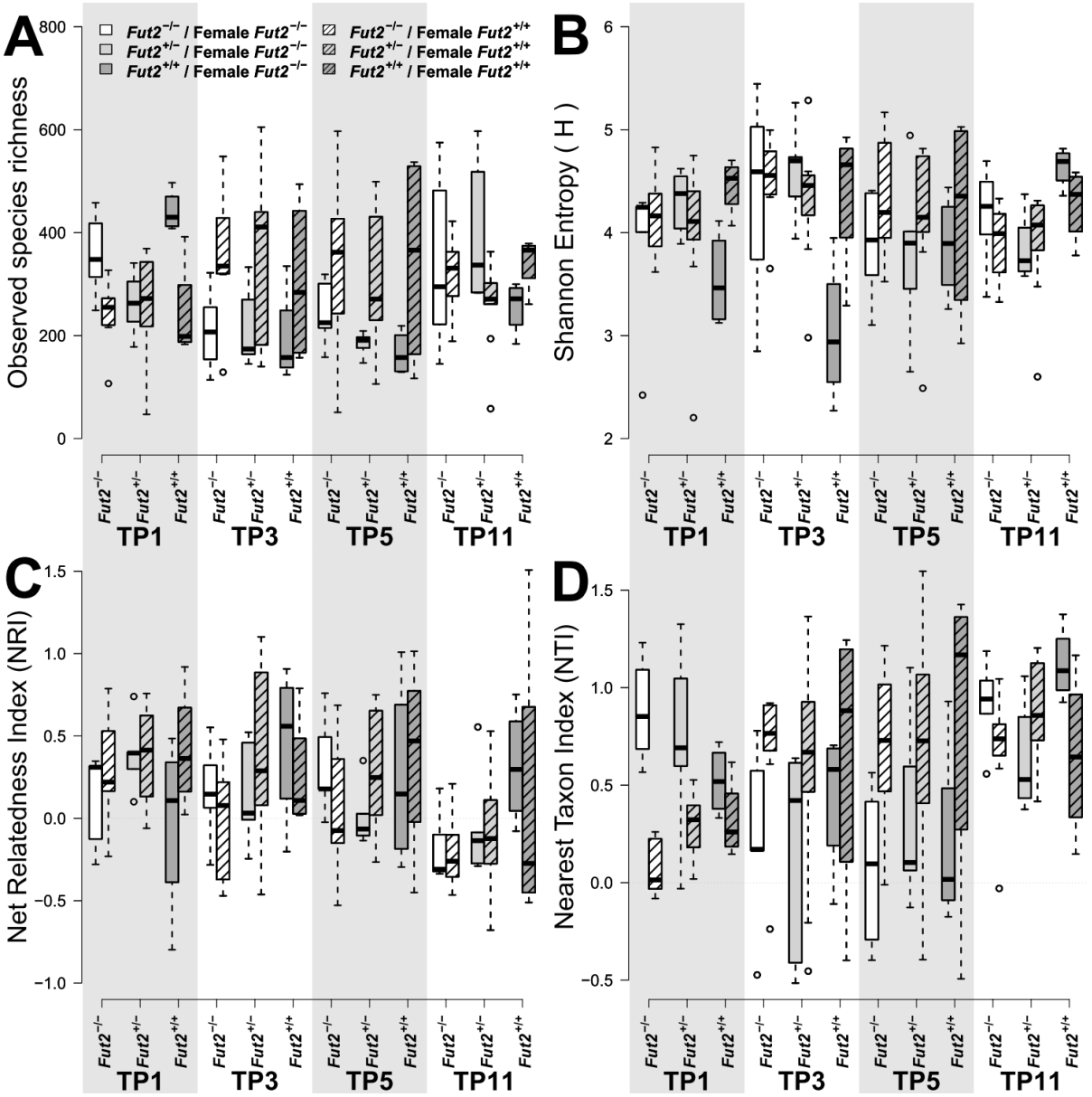


Fig.S1: Alpha diversity differences of fecal bacterial communities corresponding to the results in Table S6, focusing on the number of species (A), their distribution (B), and their phylogenetic relatedness (NRI/NTI; C, D).

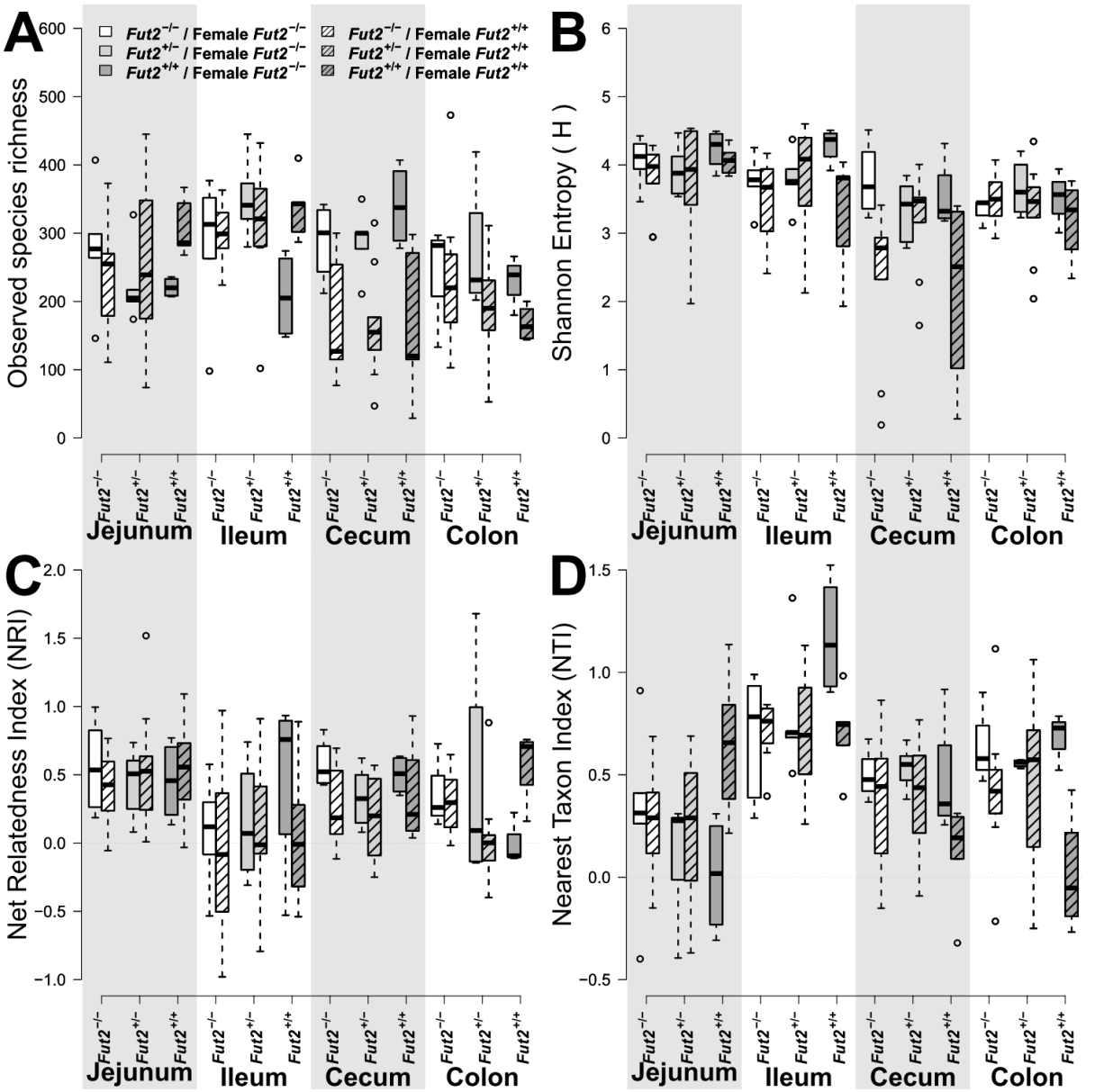


Fig.S2: Alpha diversity differences of mucosal associated bacterial communities corresponding to the results in Table S7 focusing on the number of species (A), their distribution (B), and their phylogenetic relatedness (NRI/NTI; C, D).

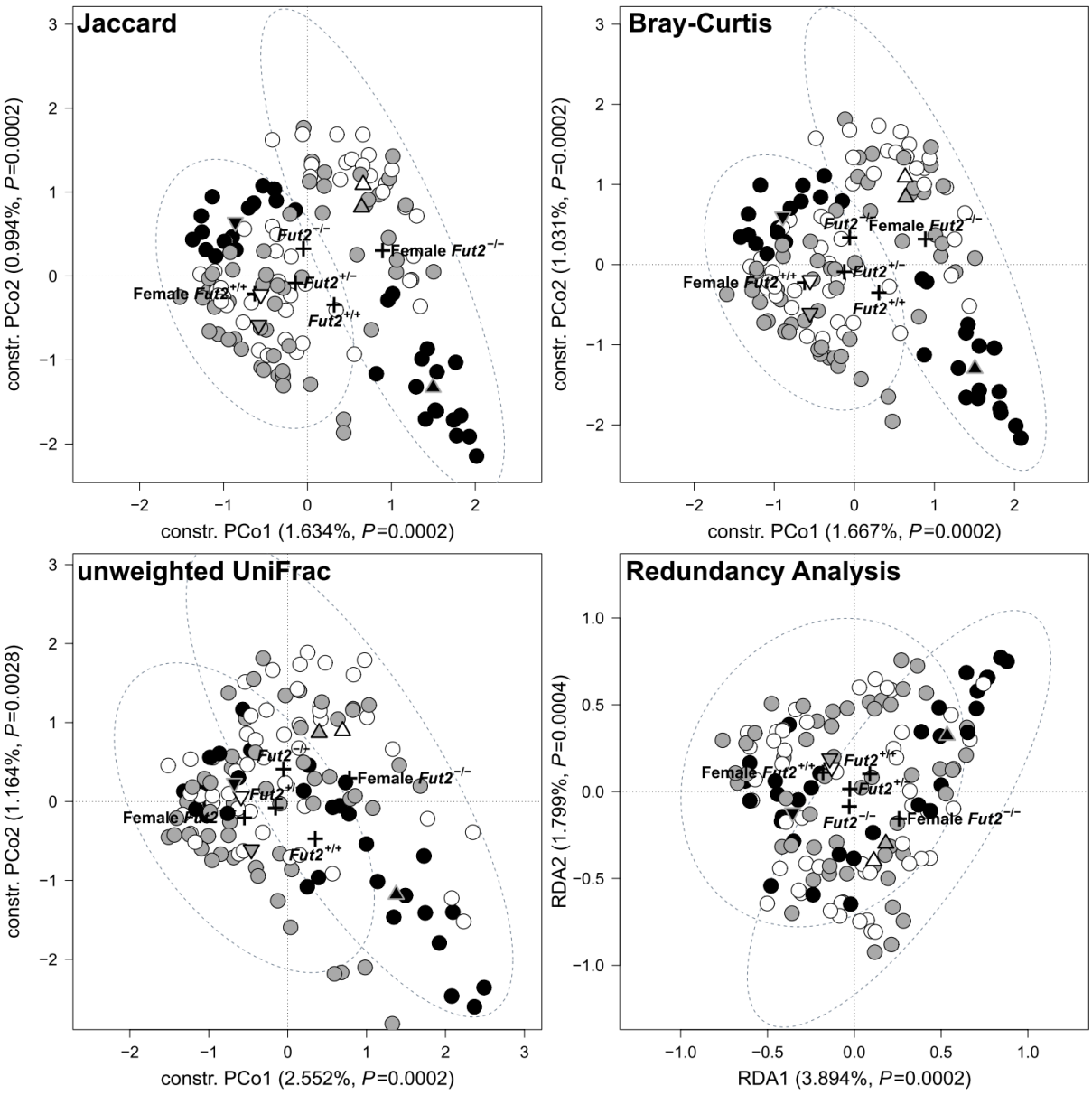


Fig.S3: Community clustering among *Fut2* genotypes and breeding direction over the whole time course regarding community composition (Jaccard), structure (Bray-Curtis, Redundancy Analysis/Euclidean) and phylogenetic composition (unweighted UniFrac) by constrained Principle Coordinate Analysis.

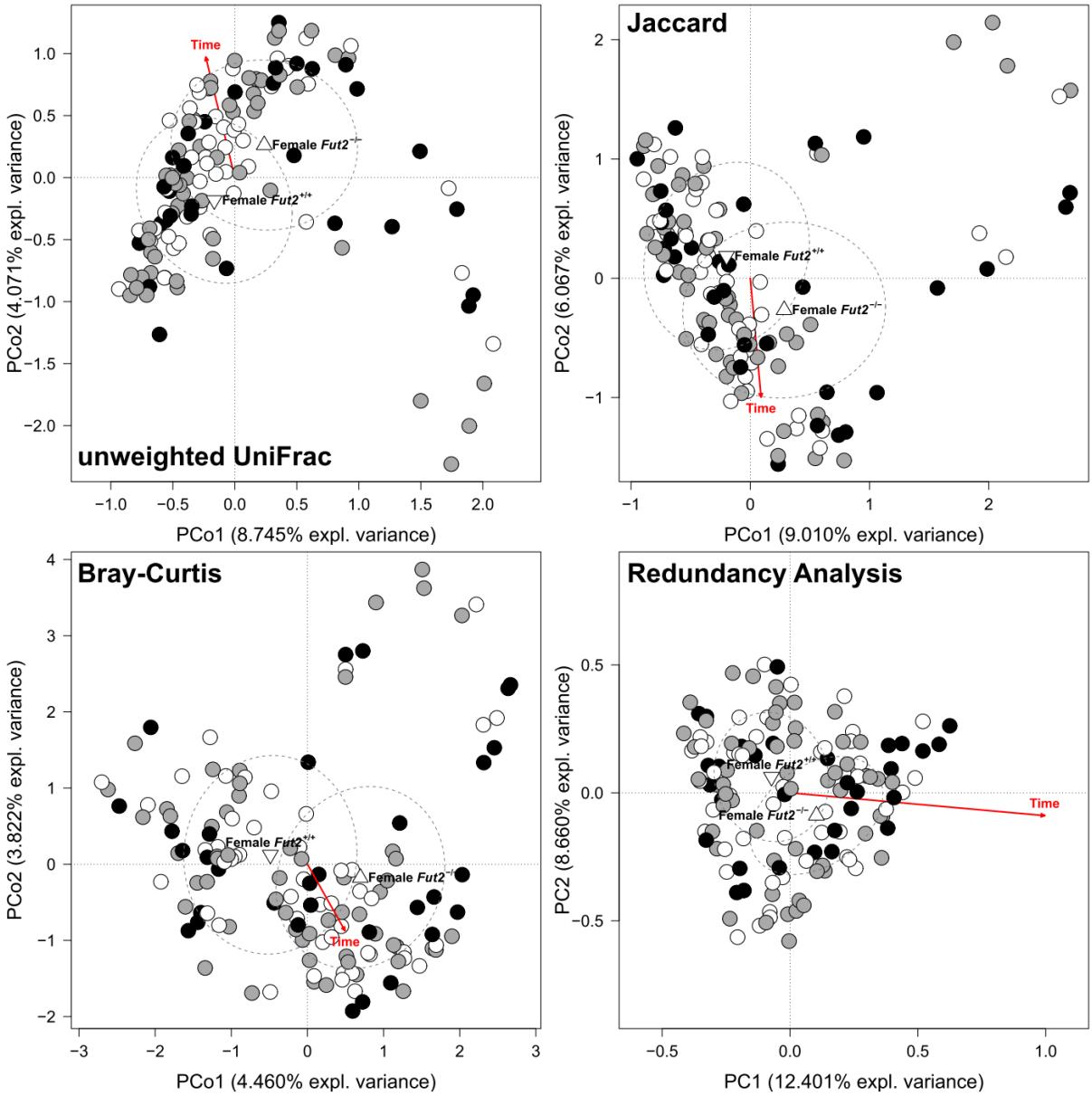


Fig.S4: Principle coordinate analysis based on phylogenetic composition (unweighted UniFrac: breeding direction- R^2=0.0878, P<0.0001; time points- R^2=0.0552, P=0.0243), community composition (Jaccard: breeding direction- R^2= 0.0843, P<0.0001; time points- R^2=0.2231, P<0.0001), and community structure (Bray-Curtis: breeding direction- R^2=0.1134, P<0.0001; time points- R^2=0.2602, P<0.0001; Redundancy Analysis/Euclidean: breeding direction- R^2=0.1013, P<0.0001; time points- R^2=0.2055, P<0.0001).

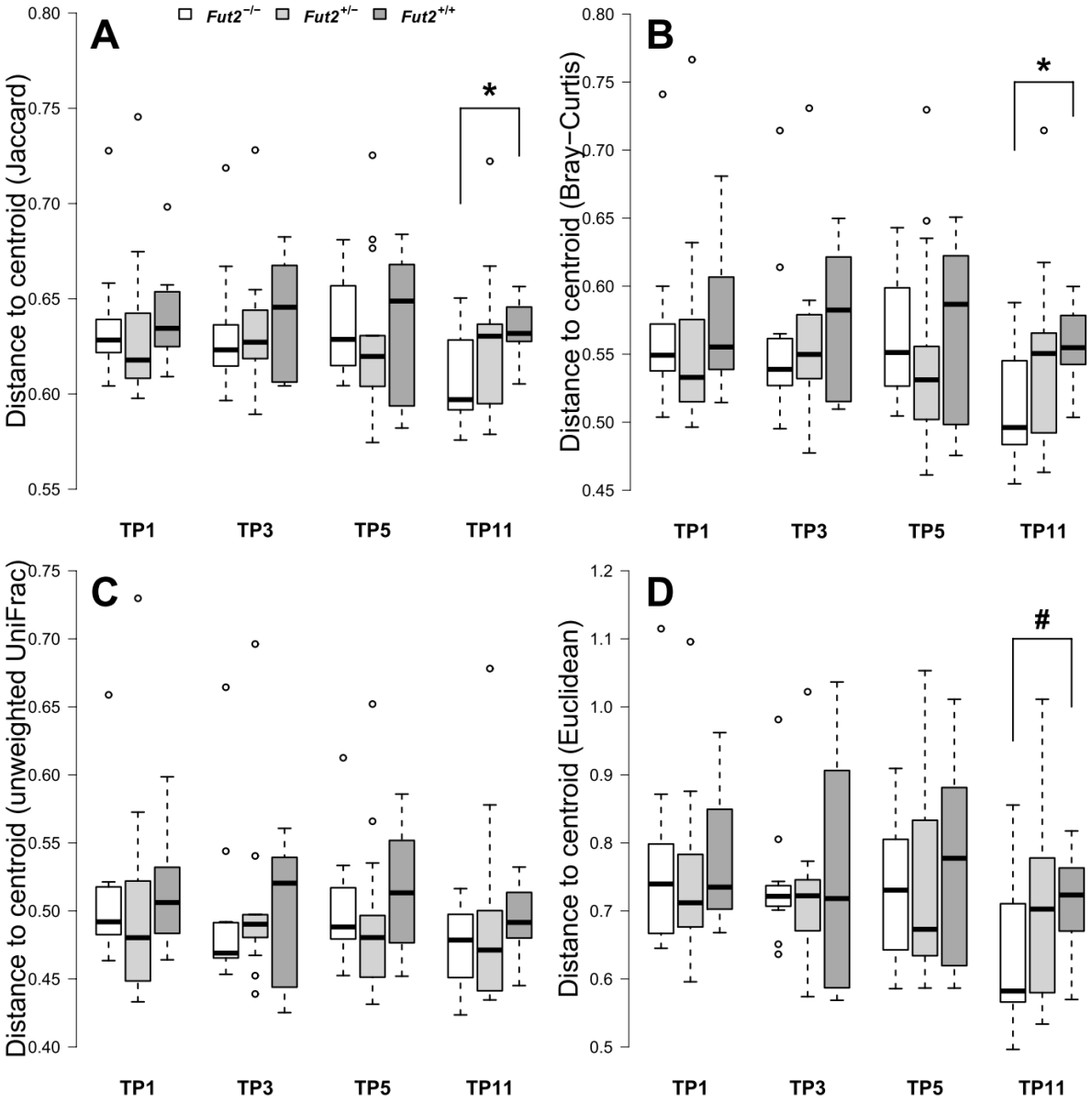


Fig.S5: Community variability among *Fut2* genotypes over time regarding community composition (A), structure (B, D) and phylogenetic composition (C).

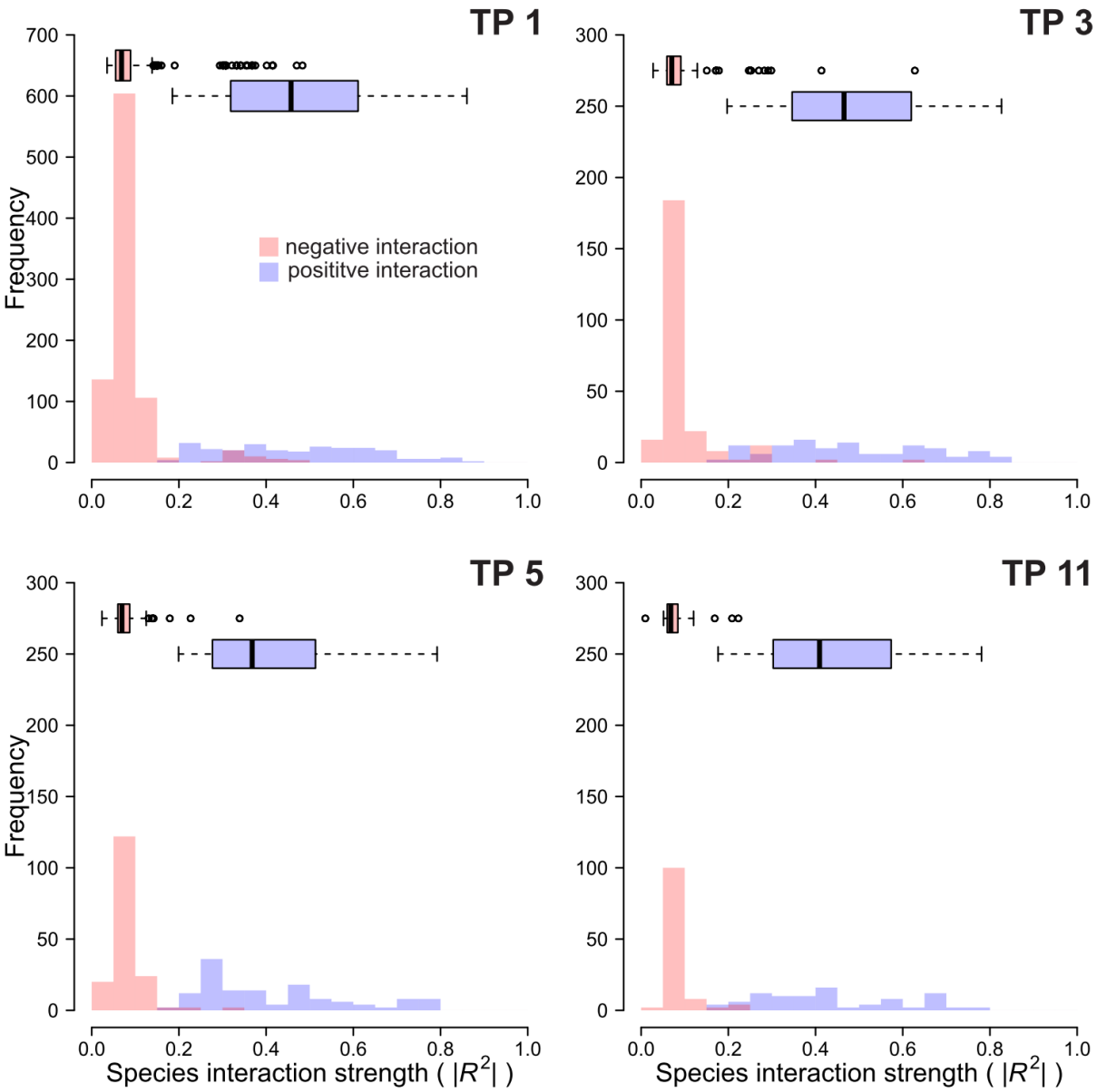


Fig.S6: Spectrum of genera interactions based on pairwise SparCC correlations with P ≤ 0.005.

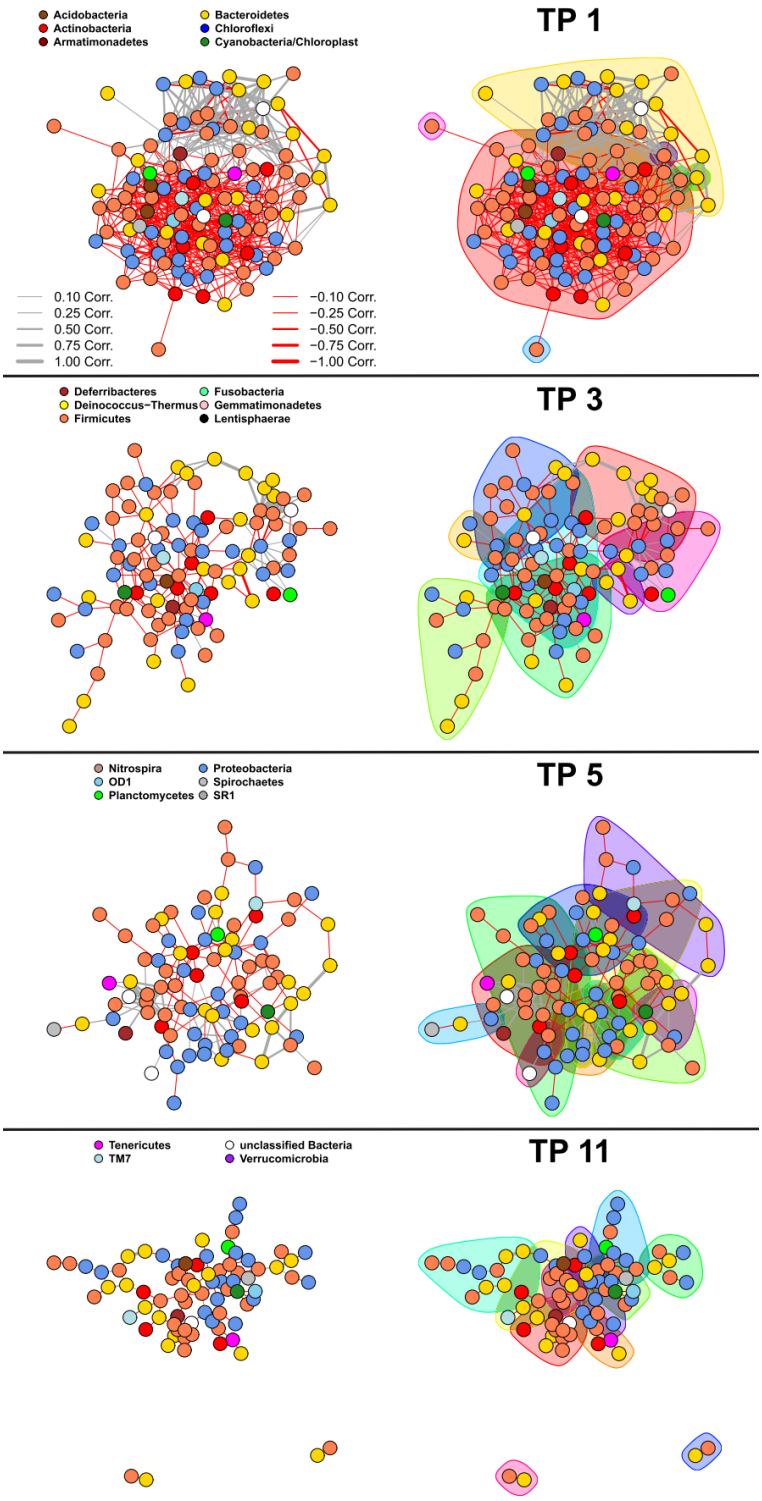


Fig.S7: Analysis of network modularity based on the importance of single interactions [1]. The network layout is based on the Kamada-Kawai-Algorithm.

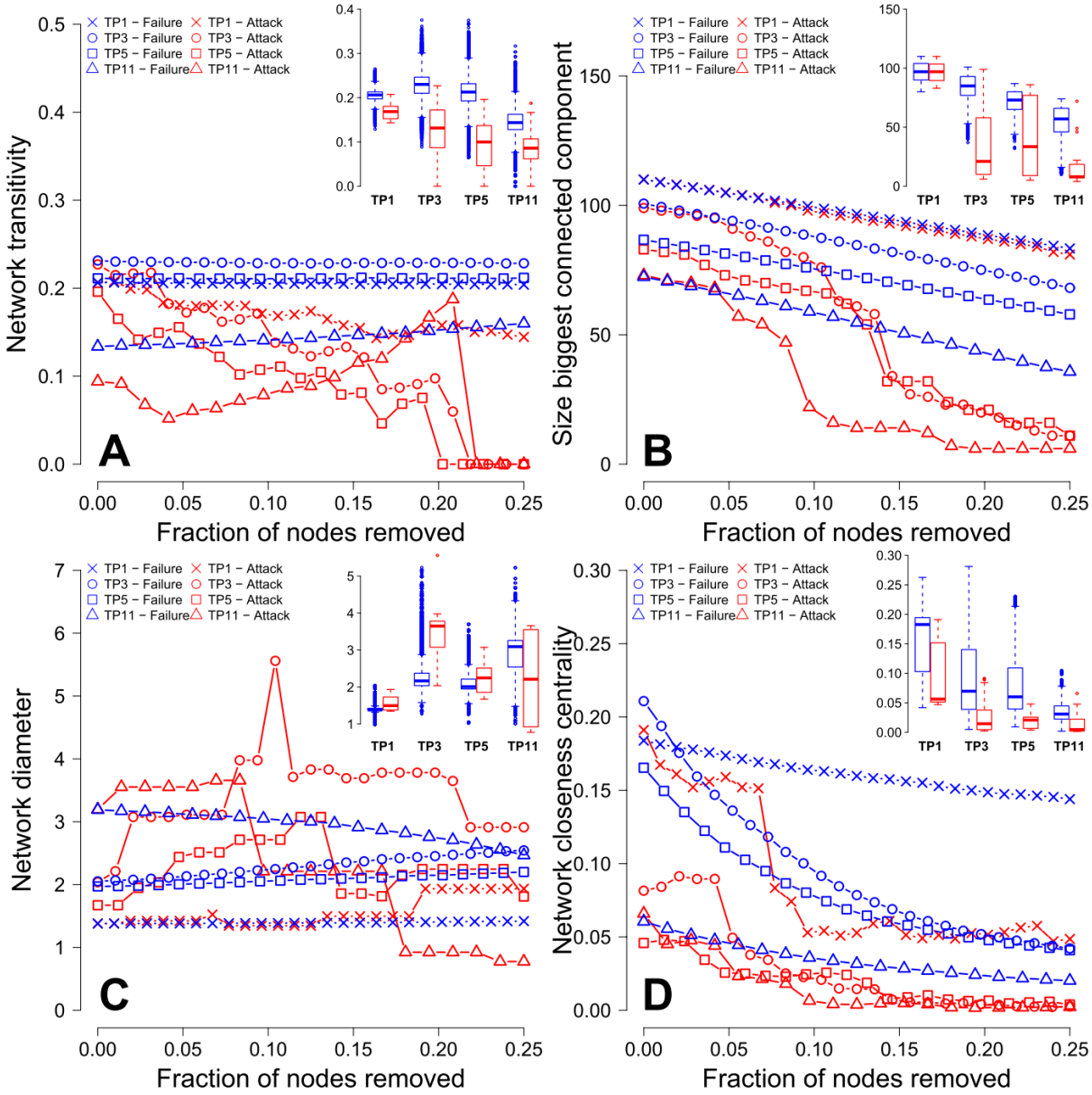


Fig.S8: Analysis of network robustness based on sequential random removal (network failure) and targeted attack on the most integrated genera (highest number of connections) measured as the decay of the networks based on the transitivity of the networks (A), average size of the biggest connected subnetwork (B), longest path within the networks (diameter, C), and average shortest paths within the networks (closeness centrality, D).

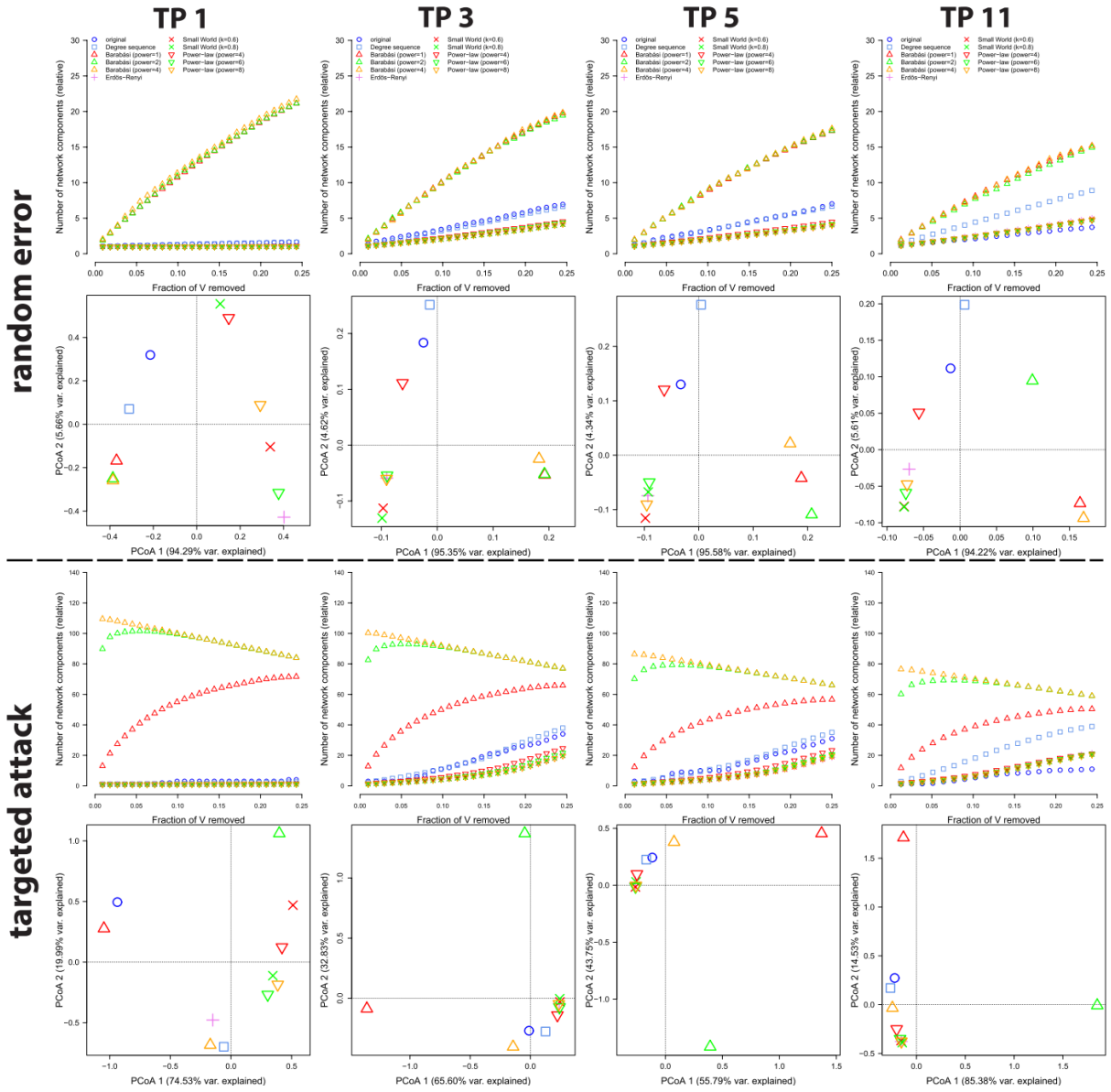


Fig.S9: Analysis of network robustness based on sequential random removal (network failure) and targeted attack on the most integrated genera (highest degree). Random networks of different characteristics, but similar size were constructed for each time point 100 times and sequential removal of up to 25% of vertices was performed by removal of the most connected node (targeted attack) or iteratively 1000 times for the measurement of network decay by random error (mean over 1000 iterations). PCoAs visualize the relative distance between decay profiles (1-|| Pearson *r* ||). Network decay is based on the relative average number of subnetworks (number of network components, see Table S11).

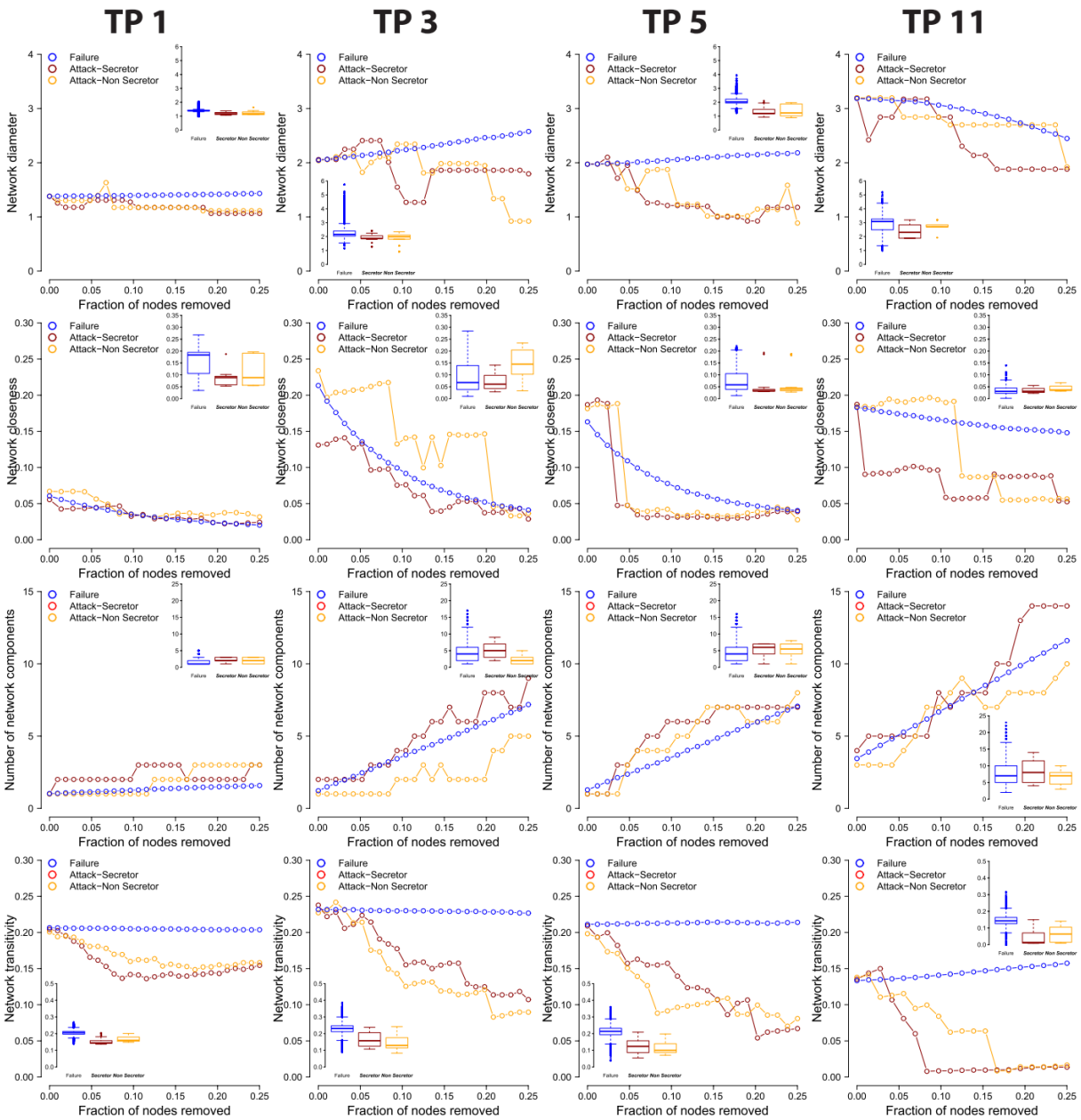


Fig.S10: Network characteristics (network diameter, closeness centrality, number of subnetworks, transitivity) under random node removal (blue, average of 1000 iterations) and targeted removal of secretor associated bacteria (red) and non secretor associated bacteria (orange), in order of their association strength.

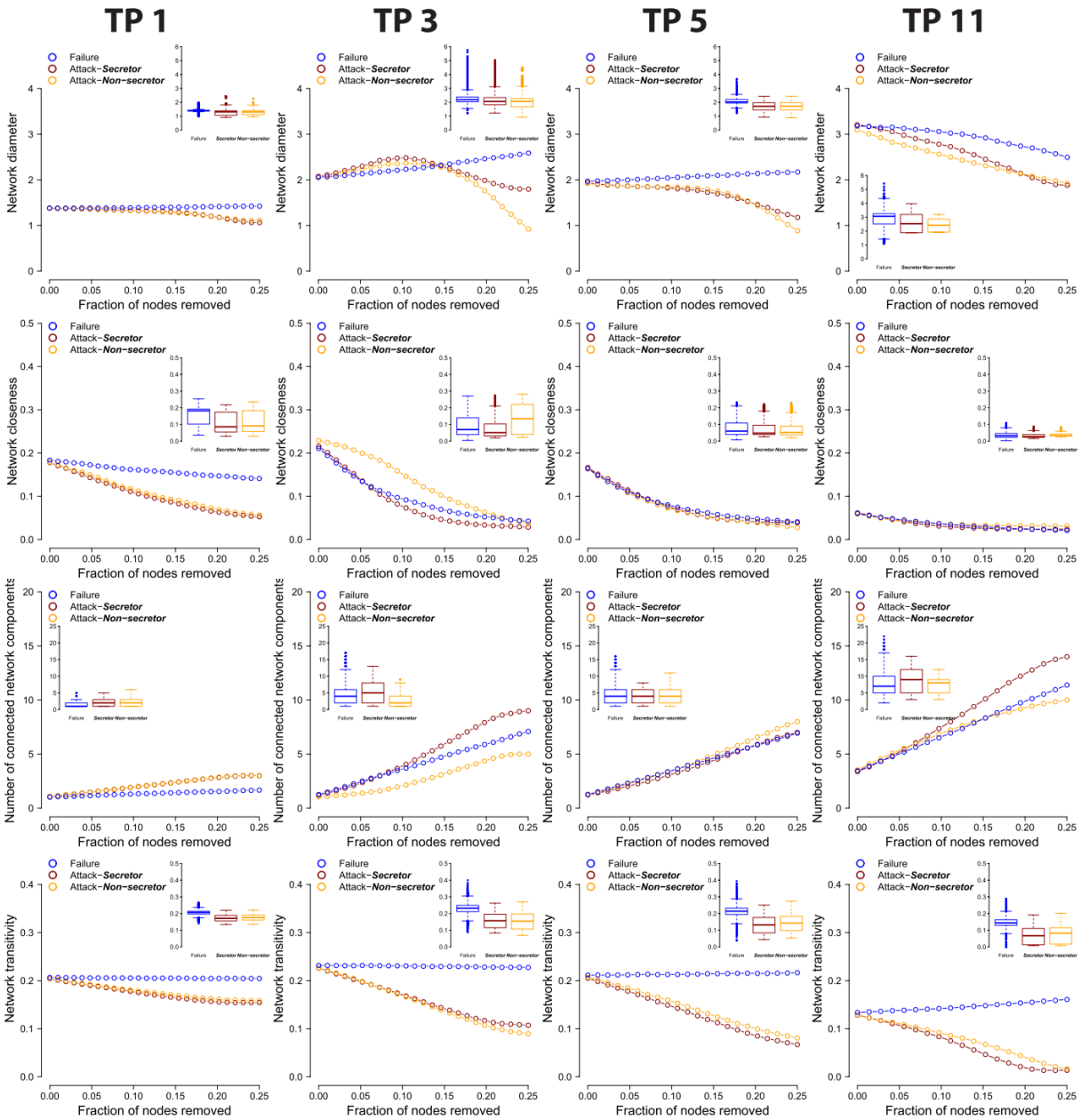


Fig.S11: Average network characteristics (network diameter, closeness centrality, number of subnetworks, transitivity) under random node removal (blue) and targeted removal the upper 25% of secretor associated bacteria (red) and non secretor associated bacteria (orange), based on the average of 1000 iterations.

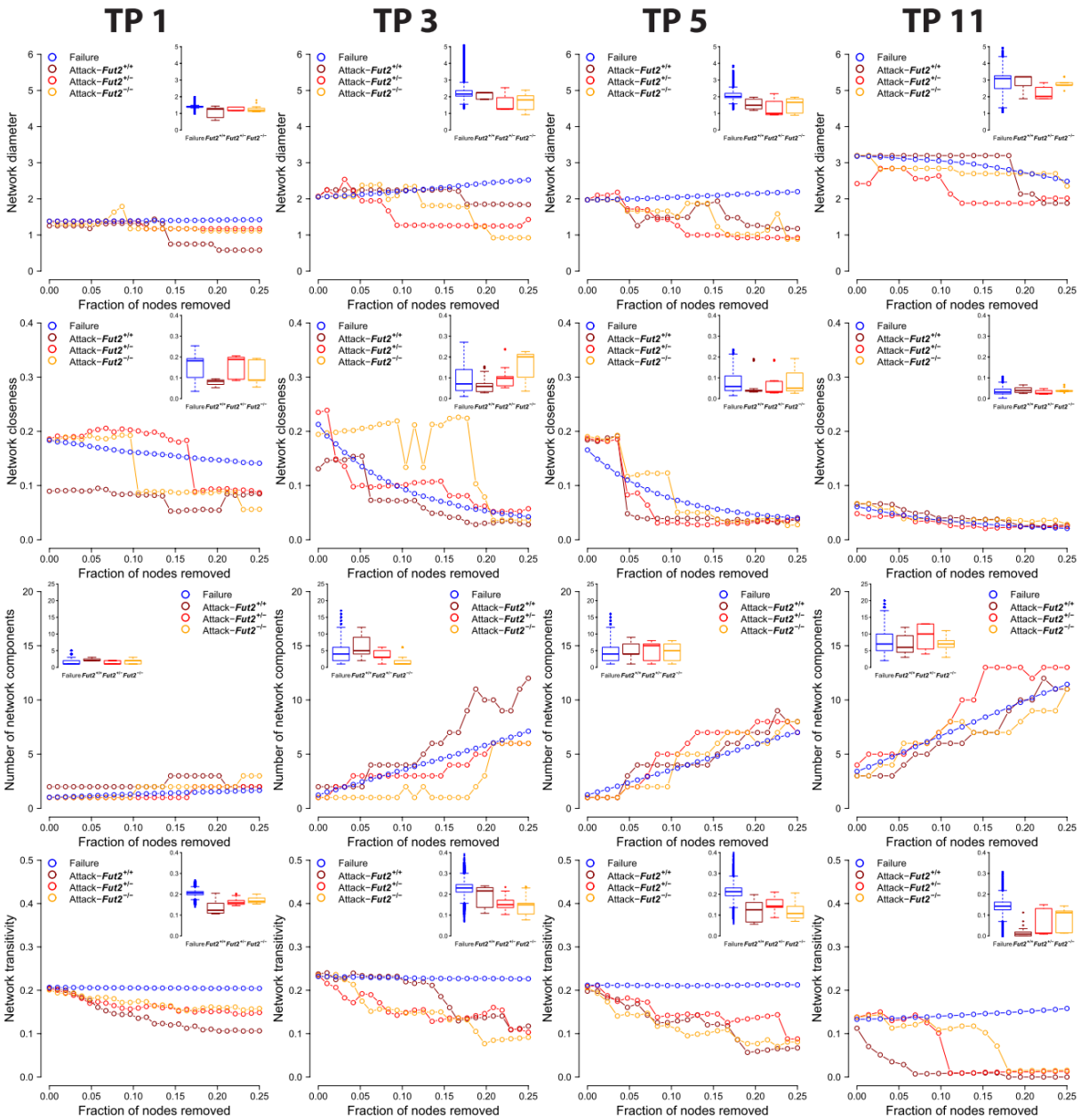


Fig.S12: Network characteristics (number of subnetworks, size of the biggest subnetwork, transitivity) under random node removal (blue) and targeted removal of bacteria associated to the *Fut2* $^{-/-}$ (darkred), *Fut2* $^{+/-}$ (red) or *Fut2* $^{-/-}$ associated bacteria (orange).

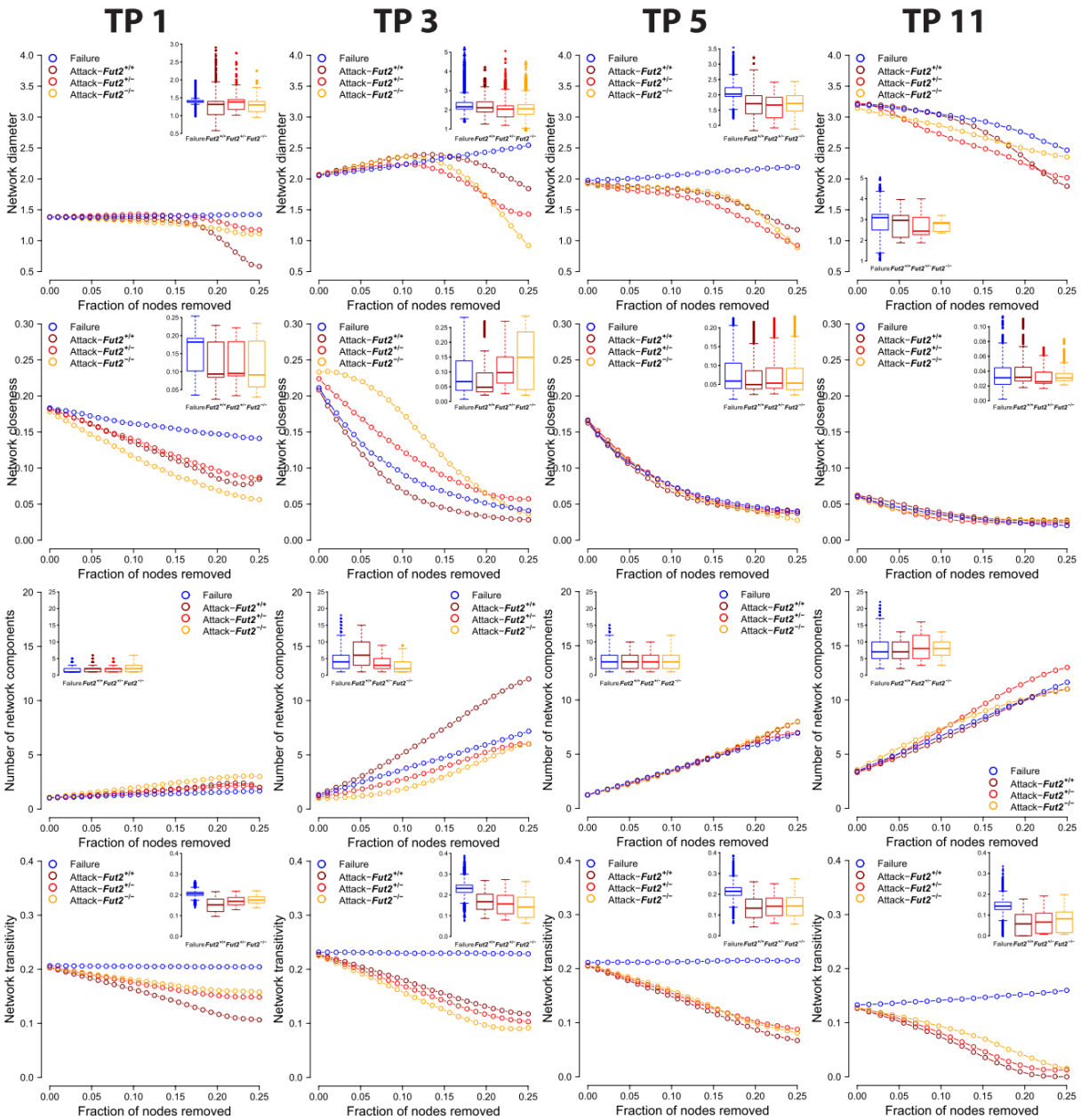


Fig.S13: Network characteristics (number of subnetworks, size of the biggest subnetwork, transitivity) under random node removal (blue) and targeted removal of the top 25% of bacteria associated to the *Fut2* $^{-/-}$ (darkred), *Fut2* $^{+/-}$ (red) or *Fut2* $^{-/-}$ associated bacteria (orange) based on 1000 iterations.

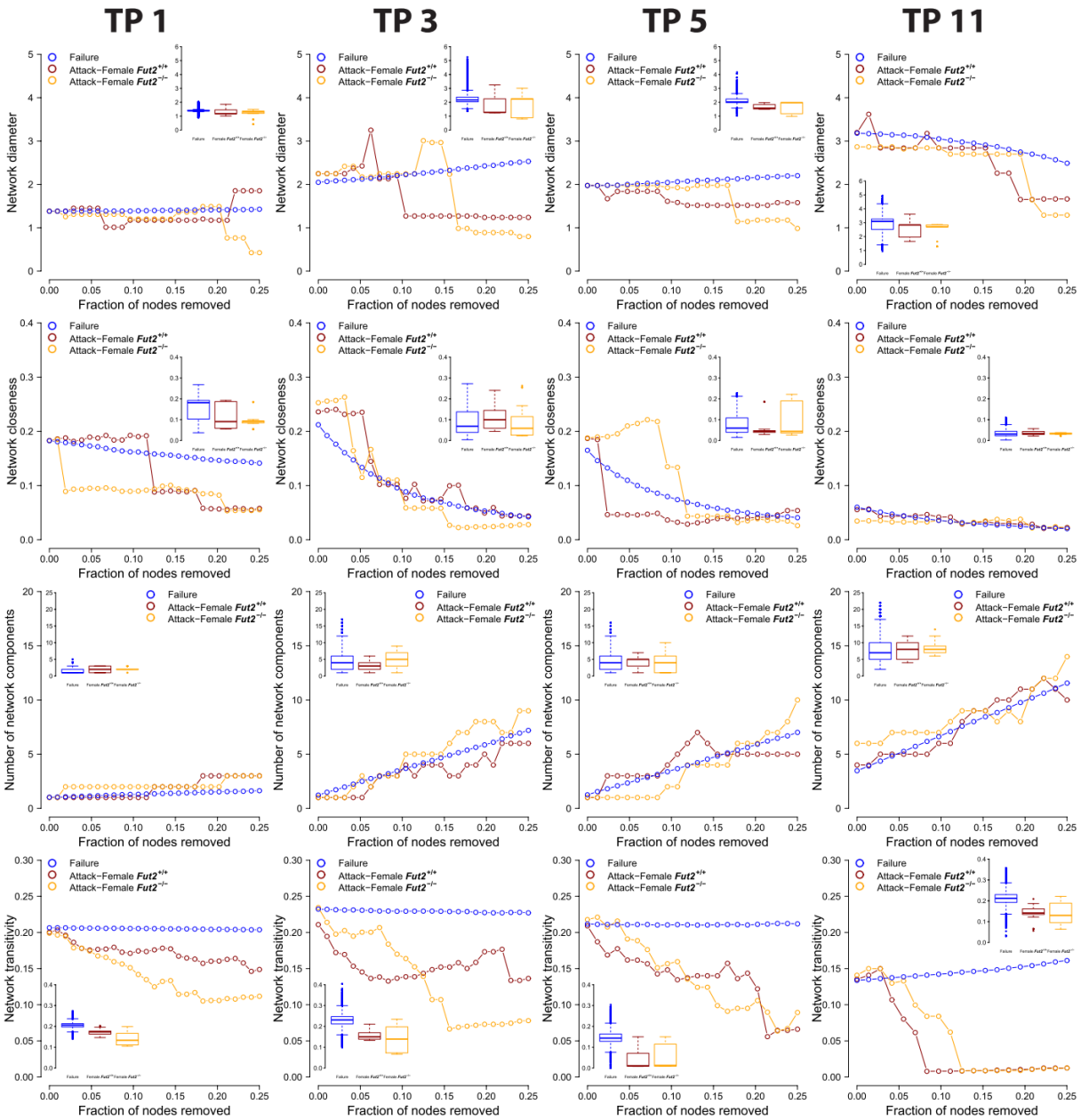


Fig.S14: Network characteristics (number of subnetworks, size of the biggest subnetwork, transitivity) under random node removal (blue) and targeted removal of bacteria associated to the *Fut2* $^{-/-}$ grand dam breeding line (red) or bacteria associated to *Fut2* $^{+/+}$ grand dam breeding line (orange).

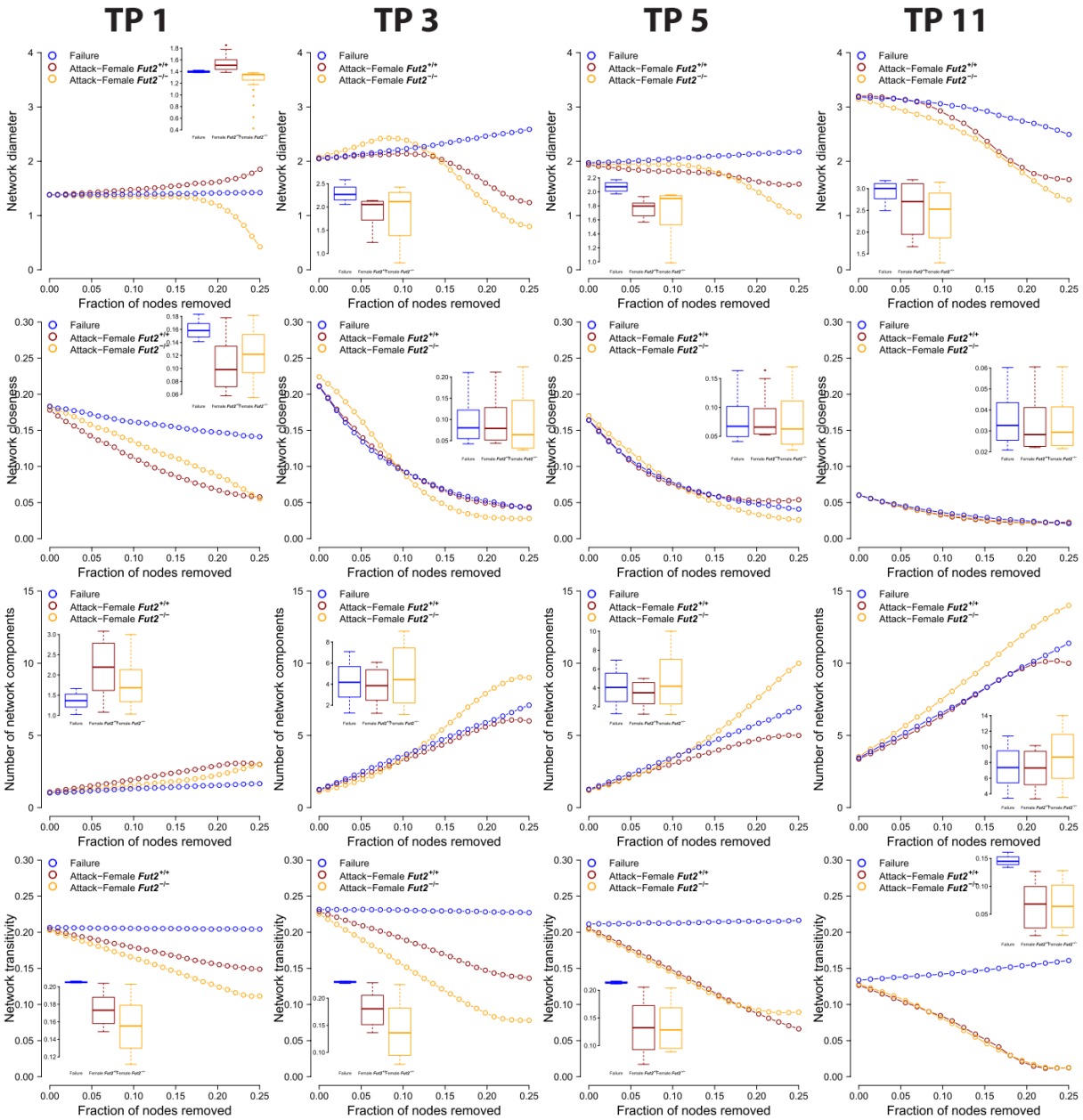


Fig.S15: Network characteristics (number of subnetworks, size of the biggest subnetwork, transitivity) under random node removal (blue) and targeted removal of the top 25% of bacteria associated to the *Fut2* $^{-/-}$ grand dam breeding line (red) or bacteria associated to *Fut2* $^{+/+}$ grand dam breeding line (orange) based on 1000 iterations.

Supplementary Tables

Table S1: Analysis of the major bacterial phyla among the different gastrointestinal locations incorporating the *Fut2* genotype/secretor status, or the mouse breeding lineage (founded by *Fut2* $^{-/-}$ or *Fut2* $^{+/+}$ grand dam).

Phylum	Tissue	Factor	*DF*	*F*-Value	P-Value
Firmicutes	Ileum#	*Intercept*	1,23	392.907	<0.0001
		Fut2	2,23	2.909	0.0747
	Jejununm‡	*Intercept*	1,21	90.712	<0.0001
	Cecum	*Intercept*	1,27	224.717	<0.0001
		Direction	1,9	2.998	0.1174
	Colon‡	*Intercept*	1,26	429.929	<0.0001
Bacteroidetes	Ileum*	*Intercept*	1,23	46.461	<0.0001
		Fut2	2,23	2.928	0.0736
	Jejununm***	*Intercept*	1,21	60.061	<0.0001
	Cecum	*Intercept*	1,27	794.952	<0.0001
		Direction	1,9	5.517	0.0434
	Colon	*Intercept*	1,26	274.419	<0.0001
Proteobacteria†	Ileum	*Intercept*	1,23	321.686	<0.0001
		Direction	1,9	0.611	0.4546
		Secretor	1,23	1.726	0.2019
		Direction:*Secretor*	1,23	7.236	0.0131
	Jejununm	*Intercept*	1,19	387.581	<0.0001
		Fut2	2,19	11.360	0.0006
	Cecum	*Intercept*	1,26	1400.570	<0.0001
		Secretor	1,26	6.613	0.0162
	Colon	*Intercept*	1,26	146.536	<0.0001

* X^2 transformed; # $X^{1/4}$ transformed; † log(X+1) transformed; ‡ $X^{1/2}$ transformed; *** X^3 transformed

Table S2: Indicator genera associated to the mucosal tissue (gray shading highlights taxa with multiple associations to the same factor among tissues).

Tissue	Genus	Association	*IndVal*	*P*-Value	P_{FDR}
Jejunum	*uncl. Clostridia*	*Fut2* $^{+/-}$	0.5578	0.0485	0.5057
	Acinetobacter	*Fut2* $^{+/+}$	0.7554	0.0013	0.0657
	Herbaspirillum	*Fut2* $^{+/+}$	0.7186	0.0219	0.3912
	Propionibacterium	*Fut2* $^{+/+}$	0.7556	0.0020	0.0657
	Staphylococcus	*Fut2* $^{+/+}$	0.6325	0.0268	0.3912
	Streptococcus	*Fut2* $^{+/+}$	0.7185	0.0027	0.0657
	Syntrophococcus	*Fut2* $^{+/+}$	0.7039	0.0456	0.5057

Ileum	*Bacteroides*	*Fut2* $^{+/-}$	0.7608	0.0363	0.9922
	uncl. Bacteroidetes	*Fut2* $^{+/-}$	0.6531	0.0493	0.9922
	uncl. Planococcaceae	*Fut2* $^{+/+}$	0.5411	0.0426	0.9922
Cecum	*Paraprevotella*	*Fut2* $^{-/-}$	0.5345	0.0405	0.5366
	Anaerotruncus	*Fut2* $^{+/+}$	0.6662	0.0140	0.3374
	Lactobacillus	*Fut2* $^{+/+}$	0.7294	0.0040	0.2120
	Odoribacter	*Fut2* $^{+/+}$	0.6765	0.0191	0.3374
Colon	*TM7 genus incertae sedis*	*Fut2* $^{-/-}$	0.5533	0.0369	0.7341
	Streptophyta	*Fut2* $^{+/+}$	0.5799	0.0189	0.7341
	Turicibacter	*Fut2* $^{+/+}$	0.5359	0.0419	0.7341
GIT	*Dorea*	*Fut2* $^{-/-}$	0.4598	0.0444	0.4140
	uncl. Erysipelotrichaceae	*Fut2* $^{+/-}$	0.3660	0.0398	0.4140
	Parabacteroides	*Fut2* $^{+/-}$	0.5981	0.0195	0.2691
	Parasutterella	*Fut2* $^{+/-}$	0.7275	0.0480	0.4140
	Acinetobacter	*Fut2* $^{+/+}$	0.5211	0.0121	0.2208
	Odoribacter	*Fut2* $^{+/+}$	0.6713	0.0001	0.0069
	Propionibacterium	*Fut2* $^{+/+}$	0.4360	0.0032	0.1104
	Staphylococcus	*Fut2* $^{+/+}$	0.4023	0.0128	0.2208
Jejunum	*Butyricicoccus*	Non secretor	0.7607	0.0430	0.7942
	uncl. Sphingobacteriales	Non secretor	0.6621	0.0322	0.7942
Ileum	*Acetanaerobacterium*	Non secretor	0.5805	0.0394	1.0000
Cecum	*Dorea*	Non secretor	0.6145	0.0070	0.2491
	Paraprevotella	Non secretor	0.5345	0.0136	0.2491
	Helicobacter	Secretor	0.8218	0.0141	0.2491
	Syntrophococcus	Secretor	0.7717	0.0384	0.5087
Colon	*Shewanella*	Non secretor	0.5534	0.0345	0.8777
	TM7 genus incertae sedis	Non secretor	0.5694	0.0183	0.8777
GIT	*Butyricicoccus*	Non secretor	0.6528	0.0304	0.4195
	Dorea	Non secretor	0.5323	0.0129	0.3519
	uncl. Sphingobacteriales	Non secretor	0.5496	0.0482	0.5542
	Odoribacter	Secretor	0.7081	0.0194	0.3519
	Parabacteroides	Secretor	0.6536	0.0204	0.3519
	Syntrophococcus	Secretor	0.7006	0.0044	0.3036
Jejunum	*uncl. Prevotellaceae*	*Fut2* $^{-/-}$ dam	0.8161	0.0074	0.1880
	Propionibacterium	*Fut2* $^{-/-}$ dam	0.6773	0.0103	0.1880
	uncl. Ruminococcaceae	*Fut2* $^{-/-}$ dam	0.8828	0.0078	0.1880
	Staphylococcus	*Fut2* $^{-/-}$ dam	0.7099	0.0052	0.1880
	Syntrophococcus	*Fut2* $^{-/-}$ dam	0.7929	0.0221	0.3226
	Clostridium Cluster XlVa	*Fut2* $^{+/+}$ dam	0.7633	0.0407	0.4951
Ileum	*Acetanaerobacterium*	*Fut2* $^{-/-}$ dam	0.7038	0.0009	0.0418
	Asaccharobacter	*Fut2* $^{-/-}$ dam	0.7401	0.0108	0.1435
	uncl. Bacteroidetes	*Fut2* $^{-/-}$ dam	0.7767	0.0356	0.2547
	uncl. Clostridiales	*Fut2* $^{-/-}$ dam	0.7958	0.0248	0.2113
	Odoribacter	*Fut2* $^{-/-}$ dam	0.7579	0.0099	0.1435
	Oscillibacter	*Fut2* $^{-/-}$ dam	0.8483	0.0028	0.0837
	uncl. Proteobacteria	*Fut2* $^{-/-}$ dam	0.6767	0.0300	0.2325
	uncl. Ruminococcaceae	*Fut2* $^{-/-}$ dam	0.8158	0.0246	0.2113
	Staphylococcus	*Fut2* $^{-/-}$ dam	0.7338	0.0003	0.0279
	TM7 genus incertae sedis	*Fut2* $^{-/-}$ dam	0.6785	0.0064	0.1190
	uncl. Bacteria	*Fut2* $^{+/+}$ dam	0.8352	0.0250	0.2113
	uncl. Firmicutes	*Fut2* $^{+/+}$ dam	0.9145	0.0036	0.0837
	Rheinheimera	*Fut2* $^{+/+}$ dam	0.6255	0.0218	0.2113
Cecum	*uncl. Bacteroidales*	*Fut2* $^{-/-}$ dam	0.7490	0.0385	0.5101
	uncl. Prevotellaceae	*Fut2* $^{-/-}$ dam	0.8120	0.0058	0.3074
	Clostridium Cluster IV	*Fut2* $^{+/+}$ dam	0.5774	0.0326	0.5101

	Mucispirillum	*Fut2* $^{+/+}$ dam	0.8225	0.0316	0.5101
Colon	*Anaerotruncus*	*Fut2* $^{-/-}$ dam	0.6989	0.0205	0.2508
	Flavonifractor	*Fut2* $^{-/-}$ dam	0.5609	0.0212	0.2508
	Lachnospiracea incertae sedis	*Fut2* $^{-/-}$ dam	0.7455	0.0169	0.2508
	Pseudoflavonifractor	*Fut2* $^{-/-}$ dam	0.7592	0.0472	0.3363
	uncl. Ruminococcaceae	*Fut2* $^{-/-}$ dam	0.7832	0.0222	0.2508
	TM7 genus incertae sedis	*Fut2* $^{-/-}$ dam	0.5694	0.0176	0.2508
	Clostridium Cluster IV	*Fut2* $^{+/+}$ dam	0.6279	0.0341	0.2776
	Dorea	*Fut2* $^{+/+}$ dam	0.5898	0.0264	0.2508
GIT	*Anaerotruncus*	*Fut2* $^{-/-}$ dam	0.5484	0.0113	0.0715
	uncl. Bacteroidetes	*Fut2* $^{-/-}$ dam	0.7511	0.0049	0.0483
	Odoribacter	*Fut2* $^{-/-}$ dam	0.7373	0.0114	0.0715
	Prevotella	*Fut2* $^{-/-}$ dam	0.6645	0.0018	0.0248
	uncl. Prevotellaceae	*Fut2* $^{-/-}$ dam	0.7560	0.0001	0.0034
	Propionibacterium	*Fut2* $^{-/-}$ dam	0.4332	0.0064	0.0491
	Pseudoflavonifractor	*Fut2* $^{-/-}$ dam	0.6850	0.0044	0.0483
	uncl. Ruminococcaceae	*Fut2* $^{-/-}$ dam	0.7853	0.0008	0.0138
	Staphylococcus	*Fut2* $^{-/-}$ dam	0.5581	0.0001	0.0034
	Syntrophococcus	*Fut2* $^{-/-}$ dam	0.6748	0.0401	0.1824
	TM7 genus incertae sedis	*Fut2* $^{-/-}$ dam	0.5239	0.0007	0.0138
	Butyrivibrio	*Fut2* $^{+/+}$ dam	0.4478	0.0349	0.1720
	Clostridium Cluster IV	*Fut2* $^{+/+}$ dam	0.4482	0.0152	0.0807
	Clostridium Cluster XIVa	*Fut2* $^{+/+}$ dam	0.6923	0.0473	0.1920
	Dorea	*Fut2* $^{+/+}$ dam	0.5104	0.0423	0.1824
	uncl. Firmicutes	*Fut2* $^{+/+}$ dam	0.8957	0.0064	0.0491
	Lachnobacterium	*Fut2* $^{+/+}$ dam	0.4249	0.0131	0.0753

Table S3: Indicator species level OTU association to secretor status (gray shading highlights taxa with multiple associations to the same factor among time points).

TP	OTU-ID	RDP9 Classification	Association	*IndVal*	*P*-Value	P_{FDR}
1	Otu00160	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Alistipes;*	Secretor	0.6953	0.0491	0.3971
	Otu00298	*Bacteroidetes;Bacteroidia;Bacteroidales;Prevotellaceae;Paraprevotella;*	Nonsecretor	0.6195	0.0385	0.3971
	Otu00417	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;*	Nonsecretor	0.6407	0.0228	0.3971
	Otu00537	*Bacteroidetes;Bacteroidia;Bacteroidales;Marinilabiaceae;Anaerophaga;*	Nonsecretor	0.5561	0.0311	0.3971
	Otu00622	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;*	Nonsecretor	0.6652	0.0190	0.3971
	Otu00737	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Barnesiella;*	Nonsecretor	0.5311	0.0436	0.3971
	Otu01015	*Bacteroidetes;Bacteroidia;Bacteroidales;Bacteroidaceae;Bacteroides;*	Nonsecretor	0.5370	0.0499	0.3971
	Otu01091	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;unclassified;*	Nonsecretor	0.5311	0.0439	0.3971
3	Otu00010	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Robinsoniella;*	Nonsecretor	0.6130	0.0469	0.3830
	Otu00077	*Firmicutes;Bacilli;Lactobacillales;Lactobacillaceae;Lactobacillus;*	Nonsecretor	0.6691	0.0350	0.3813
	Otu00193	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Roseburia;*	Nonsecretor	0.6455	0.0027	0.3813
	Otu00205	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Lachnospiracea incertae sedis;*	Nonsecretor	0.4890	0.0357	0.3813
	Otu00248	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	Secretor	0.6753	0.0252	0.3813
	Otu00305	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;*	Nonsecretor	0.7252	0.0054	0.3813
	Otu00320	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Alistipes;*	Nonsecretor	0.6722	0.0396	0.3813
	Otu00353	*Bacteroidetes;Bacteroidia;Bacteroidales;Marinilabiaceae;Anaerophaga;*	Nonsecretor	0.5938	0.0434	0.3813
	Otu00355	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;*	Nonsecretor	0.6459	0.0234	0.3813
	Otu00479	*Bacteroidetes;Bacteroidia;Bacteroidales;Marinilabiaceae;Anaerophaga;*	Nonsecretor	0.6055	0.0213	0.3813
	Otu00505	*Bacteroidetes;Bacteroidia;Bacteroidales;Marinilabiaceae;Anaerophaga;*	Nonsecretor	0.5452	0.0296	0.3813
	Otu00555	*Bacteroidetes;Bacteroidia;Bacteroidales;*	Nonsecretor	0.6157	0.0128	0.3813
	Otu00585	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	Nonsecretor	0.5774	0.0127	0.3813
	Otu00603	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Barnesiella;*	Nonsecretor	0.5416	0.0400	0.3813
	Otu00613	*Proteobacteria;Gammaproteobacteria;Thiotrichales;Piscirickettsiaceae;*	Nonsecretor	0.5416	0.0441	0.3813
	Otu00619	*Bacteroidetes;Bacteroidia;Bacteroidales;Marinilabiaceae;Anaerophaga;*	Nonsecretor	0.5938	0.0272	0.3813
	Otu00652	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	Nonsecretor	0.5416	0.0440	0.3813
	Otu00979	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	Nonsecretor	0.5528	0.0351	0.3813
5	Otu00056	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Howardella;*	Nonsecretor	0.7358	0.0399	0.3365
	Otu00117	*Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Hydrogenoanaerobacterium;*	Nonsecretor	0.4986	0.0414	0.3365
	Otu00131	*Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;*	Nonsecretor	0.6555	0.0084	0.3365

	Otu00132	*Bacteroidetes;Bacteroidia;Bacteroidales;Prevotellaceae;Prevotella;*	Nonsecretor	0.6618	0.0438	0.3365
	Otu00176	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	Nonsecretor	0.6842	0.0359	0.3365
	Otu00239	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Alistipes;*	Nonsecretor	0.6318	0.0231	0.3365
	Otu00313	*Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Ruminococcus;*	Nonsecretor	0.4880	0.0397	0.3365
	Otu00346	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Barnesiella;*	Nonsecretor	0.6129	0.0115	0.3365
	Otu00368	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;*	Nonsecretor	0.5647	0.0308	0.3365
	Otu00414	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	Secretor	0.6030	0.0489	0.3365
	Otu00415	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Lachnobacterium;*	Nonsecretor	0.5452	0.0347	0.3365
	Otu00467	*Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Hydrogenoanaerobacterium;*	Nonsecretor	0.5482	0.0353	0.3365
	Otu00656	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Lachnobacterium;*	Nonsecretor	0.5118	0.0457	0.3365
	Otu00815	*Bacteroidetes;Sphingobacteria;Sphingobacteriales;Flammeovirgaceae;Limibacter;*	Nonsecretor	0.5261	0.0471	0.3365
	Otu01066	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Alistipes;*	Nonsecretor	0.5848	0.0388	0.3365
	Otu01074	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Alistipes;*	Nonsecretor	0.5482	0.0335	0.3365
	Otu01158	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	Nonsecretor	0.5774	0.0085	0.3365
11	Otu00089	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Odoribacter;*	Secretor	0.6662	0.0436	0.3460
	Otu00127	*Bacteroidetes;Bacteroidia;Bacteroidales;Marinilabiaceae;Anaerophaga;*	Nonsecretor	0.6946	0.0185	0.2573
	Otu00215	*Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Sporobacter;*	Nonsecretor	0.5881	0.0475	0.3460
	Otu00224	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Lachnobacterium;*	Nonsecretor	0.6577	0.0146	0.2234
	Otu00241	*Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Pseudoflavonifractor;*	Nonsecretor	0.5848	0.0362	0.3460
	Otu00308	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	Nonsecretor	0.6407	0.0224	0.2856
	Otu00313	*Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Ruminococcus;*	Nonsecretor	0.5606	0.0278	0.3272
	Otu00323	*Bacteroidetes;Sphingobacteria;Sphingobacteriales;Flammeovirgaceae;Limibacter;*	Nonsecretor	0.5811	0.0300	0.3278
	Otu00324	*Bacteroidetes;Sphingobacteria;Sphingobacteriales;*	Nonsecretor	0.7298	0.0055	0.2234
	Otu00363	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Barnesiella;*	Nonsecretor	0.6096	0.0385	0.3460
	Otu00383	*Bacteroidetes;Bacteroidia;Bacteroidales;Prevotellaceae;Prevotella;*	Nonsecretor	0.6129	0.0116	0.2234
	Otu00394	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	Nonsecretor	0.6463	0.0473	0.3460
	Otu00431	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Tannerella;*	Nonsecretor	0.6180	0.0112	0.2234
	Otu00503	*Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Pseudoflavonifractor;*	Nonsecretor	0.5416	0.0431	0.3460
	Otu00571	*Bacteroidetes;Sphingobacteria;Sphingobacteriales;Cytophagaceae;Meniscus;*	Nonsecretor	0.5774	0.0096	0.2234
	Otu00614	*Bacteroidetes;Bacteroidia;Bacteroidales;Marinilabiaceae;Anaerophaga;*	Nonsecretor	0.5774	0.0095	0.2234
	Otu00615	*Bacteroidetes;Bacteroidia;Bacteroidales;Prevotellaceae;Xylanibacter;*	Nonsecretor	0.6129	0.0138	0.2234
	Otu00690	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	Nonsecretor	0.5848	0.0123	0.2234
	Otu00796	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	Nonsecretor	0.6455	0.0021	0.2234
	Otu01373	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	Nonsecretor	0.5774	0.0110	0.2234
	Otu01543	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Barnesiella;*	Nonsecretor	0.5482	0.0356	0.3460

1-11	Otu00205	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Lachnospiracea incertae sedis;*	Nonsecretor	0.4384	0.0148	0.1265
	Otu00313	*Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Ruminococcus;*	Nonsecretor	0.4460	0.0036	0.0504
	Otu00324	*Bacteroidetes;Sphingobacteria;Sphingobacteriales;*	Nonsecretor	0.5370	0.0022	0.0504
	Otu00363	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Barnesiella;*	Nonsecretor	0.4046	0.0363	0.1265
	Otu00414	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	Secretor	0.4001	0.0416	0.1265
	Otu00454	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	Secretor	0.4160	0.0488	0.1265
	Otu00532	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Alistipes;*	Nonsecretor	0.3699	0.0489	0.1265
	Otu00614	*Bacteroidetes;Bacteroidia;Bacteroidales;Marinilabiaceae;Anaerophaga;*	Nonsecretor	0.4114	0.0378	0.1265
	Otu00615	*Bacteroidetes;Bacteroidia;Bacteroidales;Prevotellaceae;Xylanibacter;*	Nonsecretor	0.3565	0.0412	0.1265
	Otu00781	*Bacteroidetes;Bacteroidia;Bacteroidales;Bacteroidaceae;Bacteroides;*	Nonsecretor	0.3838	0.0246	0.1265

Table S4: Indicator species level OTU association to *Fut2* genotype (gray shading highlights taxa with multiple associations to the same factor among time points).

TP	OTU-ID	RDP9 Classification	Association	*IndVal*	*P*-Value	P_{FDR}
1	Otu00159	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	*Fut2* $^{+/+}$	0.5777	0.0232	0.7332
	Otu00163	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{+/-}$/*Fut2* $^{+/+}$	0.7829	0.0208	0.7332
	Otu00218	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{+/+}$	0.5983	0.0069	0.7332
	Otu00248	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	*Fut2* $^{+/+}$	0.5557	0.0484	0.7332
	Otu00270	*Bacteroidetes;Sphingobacteria;Sphingobacteriales;Cytophagaceae;Meniscus;*	*Fut2* $^{+/+}$	0.5525	0.0323	0.7332
	Otu00464	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Alistipes;*	*Fut2* $^{+/-}$	0.5655	0.0385	0.7332
	Otu00919	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{+/+}$	0.5809	0.017	0.7332
	Otu00982	*Proteobacteria;Betaproteobacteria;Burkholderiales;Sutterellaceae;Parasutterella;*	*Fut2* $^{+/+}$	0.5669	0.0386	0.7332
	Otu01202	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;unclassified;*	*Fut2* $^{+/+}$	0.5612	0.0299	0.7332
	Otu01782	*Bacteroidetes;Bacteroidia;Bacteroidales;Bacteroidaceae;Bacteroides;*	*Fut2* $^{+/+}$	0.5809	0.0189	0.7332
3	Otu00045	*Firmicutes;Clostridia;Clostridiales;Clostridiaceae_1;Anaerobacter;*	*Fut2* $^{-/-}$/*Fut2* $^{+/-}$	0.7646	0.0451	0.6477
	Otu00163	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{-/-}$/*Fut2* $^{+/-}$	0.7689	0.0342	0.6477
	Otu00193	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Roseburia;*	*Fut2* $^{-/-}$	0.6455	0.0086	0.6477
	Otu00225	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Tannerella;*	*Fut2* $^{+/-}$	0.5583	0.0379	0.6477

	Otu00239	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Alistipes;*	*Fut2* $^{-/-}$/*Fut2* $^{+/-}$	0.787	0.0123	0.6477
	Otu00248	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	*Fut2* $^{+/-}$/*Fut2* $^{+/+}$	0.6896	0.021	0.6477
	Otu00255	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Butyrivibrio;*	*Fut2* $^{+/-}$	0.5655	0.0425	0.6477
	Otu00278	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Butyrivibrio;*	*Fut2* $^{+/+}$	0.5351	0.048	0.6477
	Otu00305	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;*	*Fut2* $^{-/-}$	0.7057	0.0136	0.6477
	Otu00449	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	*Fut2* $^{+/-}$	0.6124	0.0373	0.6477
	Otu00465	*Bacteroidetes;Bacteroidia;Bacteroidales;*	*Fut2* $^{+/+}$	0.5728	0.0262	0.6477
	Otu00555	*Bacteroidetes;Bacteroidia;Bacteroidales;*	*Fut2* $^{-/-}$	0.6004	0.0314	0.6477
	Otu00585	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{-/-}$	0.5774	0.0128	0.6477
	Otu00590	*Bacteroidetes;Bacteroidia;Bacteroidales;Bacteroidaceae;Bacteroides;*	*Fut2* $^{+/+}$	0.5669	0.0352	0.6477
	Otu00612	*Bacteroidetes;*	*Fut2* $^{+/+}$	0.6379	0.0074	0.6477
	Otu00613	*Proteobacteria;Gammaproteobacteria;Thiotrichales;Piscirickettsiaceae;*	*Fut2* $^{-/-}$	0.5239	0.0409	0.6477
	Otu00654	*Proteobacteria;Betaproteobacteria;Burkholderiales;Sutterellaceae;Parasutterella;*	*Fut2* $^{+/+}$	0.5728	0.0259	0.6477
	Otu00866	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{+/+}$	0.5525	0.0327	0.6477
	Otu00888	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Robinsoniella;*	*Fut2* $^{+/-}$	0.5976	0.0161	0.6477
	Otu01041	*Bacteroidetes;Bacteroidia;Bacteroidales;Marinilabiaceae;Anaerophaga;*	*Fut2* $^{+/+}$	0.6757	0.0024	0.6477
	Otu01403	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Alistipes;*	*Fut2* $^{+/-}$	0.5533	0.0467	0.6477
5	Otu00020	*Bacteroidetes;Bacteroidia;Bacteroidales;Bacteroidaceae;Bacteroides;*	*Fut2* $^{+/-}$/*Fut2* $^{+/+}$	0.9606	0.0067	0.6751
	Otu00056	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Howardella;*	*Fut2* $^{-/-}$/*Fut2* $^{+/-}$	0.7837	0.0264	0.6751
	Otu00131	*Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;unclassified;*	*Fut2* $^{-/-}$	0.6007	0.0427	0.7016
	Otu00143	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Barnesiella;*	*Fut2* $^{+/+}$	0.7395	0.002	0.5139
	Otu00163	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{-/-}$/*Fut2* $^{+/-}$	0.767	0.0236	0.6751
	Otu00346	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Barnesiella;*	*Fut2* $^{-/-}$	0.5964	0.0239	0.6751
	Otu00351	*Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Pseudoflavonifractor;*	*Fut2* $^{+/+}$	0.5557	0.0271	0.6751
	Otu00385	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{-/-}$/*Fut2* $^{+/+}$	0.611	0.0495	0.7016
	Otu00396	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{+/-}$	0.632	0.0169	0.6751
	Otu00414	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	*Fut2* $^{+/+}$	0.6521	0.0156	0.6751
	Otu00452	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Alistipes;*	*Fut2* $^{+/+}$	0.5669	0.0358	0.7016
	Otu00487	*Firmicutes;Erysipelotrichia;Erysipelotrichales;Erysipelotrichaceae;Erysipelotrichaceae incertae sedis;*	*Fut2* $^{+/-}$	0.5583	0.0393	0.7016
	Otu00505	*Bacteroidetes;Bacteroidia;Bacteroidales;Marinilabiaceae;Anaerophaga;*	*Fut2* $^{+/+}$	0.5401	0.0289	0.6751
	Otu00538	*Bacteroidetes;Bacteroidia;Bacteroidales;*	*Fut2* $^{+/+}$	0.5742	0.0481	0.7016
	Otu00668	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Alistipes;*	*Fut2* $^{+/-}$	0.5976	0.0173	0.6751
	Otu01158	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{-/-}$	0.5774	0.0136	0.6751
11	Otu00062	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Odoribacter;*	*Fut2* $^{+/+}$	0.7579	0.0052	0.4927

	Otu00070	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Tannerella;*	*Fut2* $^{-/-}$/*Fut2* $^{+/+}$	0.7998	0.0245	0.4927
	Otu00073	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Odoribacter;*	*Fut2* $^{-/-}$/*Fut2* $^{+/+}$	0.7518	0.0348	0.4993
	Otu00082	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Lachnospiracea incertae sedis;*	*Fut2* $^{+/+}$	0.7477	0.0031	0.4927
	Otu00089	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Odoribacter;*	*Fut2* $^{+/-}$/*Fut2* $^{+/+}$	0.6989	0.0417	0.5169
	Otu00100	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Robinsoniella;*	*Fut2* $^{+/+}$	0.6094	0.0095	0.4927
	Otu00163	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{-/-}$/*Fut2* $^{+/-}$	0.7071	0.041	0.5169
	Otu00169	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Robinsoniella;*	*Fut2* $^{+/-}$	0.6028	0.0461	0.5169
	Otu00218	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{-/-}$/*Fut2* $^{+/-}$	0.7338	0.0241	0.4927
	Otu00262	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Barnesiella;*	*Fut2* $^{-/-}$/*Fut2* $^{+/+}$	0.6325	0.018	0.4927
	Otu00273	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Robinsoniella;*	*Fut2* $^{+/+}$	0.5539	0.0474	0.5169
	Otu00324	*Bacteroidetes;Sphingobacteria;Sphingobacteriales;*	*Fut2* $^{-/-}$	0.6912	0.0156	0.4927
	Otu00383	*Bacteroidetes;Bacteroidia;Bacteroidales;Prevotellaceae;Prevotella;*	*Fut2* $^{-/-}$	0.5964	0.0239	0.4927
	Otu00394	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	*Fut2* $^{-/-}$/*Fut2* $^{+/-}$	0.7071	0.0352	0.4993
	Otu00431	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Tannerella;*	*Fut2* $^{-/-}$	0.5774	0.028	0.4927
	Otu00493	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Lachnobacterium;*	*Fut2* $^{+/+}$	0.6877	0.0035	0.4927
	Otu00540	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Syntrophococcus;*	*Fut2* $^{+/+}$	0.5669	0.0361	0.4993
	Otu00555	*Bacteroidetes;Bacteroidia;Bacteroidales;unclassified;unclassified;*	*Fut2* $^{+/+}$	0.6059	0.0235	0.4927
	Otu00566	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	*Fut2* $^{-/-}$/*Fut2* $^{+/+}$	0.6325	0.0241	0.4927
	Otu00571	*Bacteroidetes;Sphingobacteria;Sphingobacteriales;Cytophagaceae;Meniscus;*	*Fut2* $^{-/-}$	0.5774	0.0119	0.4927
	Otu00600	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Tannerella;*	*Fut2* $^{+/+}$	0.5482	0.0467	0.5169
	Otu00614	*Bacteroidetes;Bacteroidia;Bacteroidales;Marinilabiaceae;Anaerophaga;*	*Fut2* $^{-/-}$	0.5774	0.0111	0.4927
	Otu00651	*Bacteroidetes;Bacteroidia;Bacteroidales;unclassified;unclassified;*	*Fut2* $^{+/+}$	0.5441	0.0352	0.4993
	Otu00690	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{-/-}$	0.557	0.0346	0.4993
	Otu00739	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;*	*Fut2* $^{+/+}$	0.5752	0.0278	0.4927
	Otu00796	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	*Fut2* $^{-/-}$	0.6455	0.008	0.4927
	Otu00822	*Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Flavonifractor;*	*Fut2* $^{+/+}$	0.5351	0.0492	0.5169
	Otu00989	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Tannerella;*	*Fut2* $^{+/+}$	0.5401	0.0255	0.4927
	Otu01021	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;*	*Fut2* $^{+/+}$	0.5612	0.0282	0.4927
	Otu01373	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	*Fut2* $^{-/-}$	0.5774	0.0205	0.4927
1-11	Otu00078	*Firmicutes;Bacilli;Lactobacillales;Lactobacillaceae;Lactobacillus;*	*Fut2* $^{-/-}$/*Fut2* $^{+/-}$	0.5237	0.0204	0.2286
	Otu00079	*Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Pelagibaca;*	*Fut2* $^{-/-}$/*Fut2* $^{+/+}$	0.4916	0.0285	0.2286
	Otu00113	*Proteobacteria;Alphaproteobacteria;Rhizobiales;Methylocystaceae;Terasakiella;*	*Fut2* $^{+/+}$	0.4657	0.0047	0.1762
	Otu00313	*Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Ruminococcus;*	*Fut2* $^{-/-}$	0.4193	0.0282	0.2286
	Otu00324	*Bacteroidetes;Sphingobacteria;Sphingobacteriales;*	*Fut2* $^{-/-}$/*Fut2* $^{+/-}$	0.5111	0.0422	0.2286
	Otu00414	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	*Fut2* $^{+/-}$/*Fut2* $^{+/+}$	0.4255	0.0255	0.2286

	Otu00454	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{+/-}$/*Fut2* $^{+/+}$	0.4385	0.0432	0.2286
	Otu00592	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{+/+}$	0.448	0.0022	0.165
	Otu00619	*Bacteroidetes;Bacteroidia;Bacteroidales;Marinilabiaceae;Anaerophaga;*	*Fut2* $^{-/-}$/*Fut2* $^{+/-}$	0.435	0.0333	0.2286

Table S5: Indicator species level OTU association to breeding direction (gray shading highlights taxa with multiple associations to the same factor among time points).

TP	OTU-ID	RDP9 Classification	Association	*IndVal*	*P*-Value	P_{FDR}
1	Otu00002	*Bacteroidetes;Sphingobacteria;Sphingobacteriales;Cytophagaceae;Meniscus;*	*Fut2* $^{-/-}$ dam	0.9338	0.0002	0.0135
	Otu00003	*Bacteroidetes;Sphingobacteria;Sphingobacteriales;Cytophagaceae;Meniscus;*	*Fut2* $^{-/-}$ dam	0.8657	0.0012	0.0609
	Otu00050	*Proteobacteria;GammaproteoEnterobacteriales;Enterobacteriaceae;Escherichia/Shigella;*	*Fut2* $^{-/-}$ dam	0.7244	0.0143	0.1409
	Otu00079	*Proteobacteria;AlphaproteoRhodobacterales;Rhodobacteraceae;Pelagibaca;*	*Fut2* $^{-/-}$ dam	0.7206	0.0030	0.0670
	Otu00082	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Lachnospiracea incertae sedis;*	*Fut2* $^{-/-}$ dam	0.5310	0.0367	0.1827
	Otu00104	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{-/-}$ dam	0.6983	0.0027	0.0670
	Otu00110	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{-/-}$ dam	0.7559	0.0002	0.0135
	Otu00135	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{-/-}$ dam	0.6665	0.0147	0.1409
	Otu00143	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Barnesiella;*	*Fut2* $^{-/-}$ dam	0.5731	0.0267	0.1465
	Otu00148	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Odoribacter;*	*Fut2* $^{-/-}$ dam	0.5345	0.0209	0.1409
	Otu00206	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Clostridium_XlVa;*	*Fut2* $^{-/-}$ dam	0.5705	0.0364	0.1827
	Otu00218	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{-/-}$ dam	0.5345	0.0217	0.1409
	Otu00225	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Tannerella;*	*Fut2* $^{-/-}$ dam	0.7057	0.0031	0.0670
	Otu00243	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	*Fut2* $^{-/-}$ dam	0.6523	0.0087	0.1177
	Otu00260	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	*Fut2* $^{-/-}$ dam	0.6547	0.0024	0.0670
	Otu00268	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	*Fut2* $^{-/-}$ dam	0.5976	0.0062	0.1000
	Otu00270	*Bacteroidetes;Sphingobacteria;Sphingobacteriales;Cytophagaceae;Meniscus;*	*Fut2* $^{-/-}$ dam	0.5345	0.0217	0.1409
	Otu00280	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Dorea;*	*Fut2* $^{-/-}$ dam	0.5976	0.0068	0.1000
	Otu00344	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	*Fut2* $^{-/-}$ dam	0.5345	0.0191	0.1409
	Otu00539	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Tannerella;*	*Fut2* $^{-/-}$ dam	0.6547	0.0021	0.0670
	Otu00554	*Bacteroidetes;*	*Fut2* $^{-/-}$ dam	0.5345	0.0198	0.1409
	Otu00592	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{-/-}$ dam	0.6547	0.0033	0.0670

	Otu00630	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	*Fut2* $^{-/-}$ dam	0.5976	0.0069	0.1000
	Otu00683	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Barnesiella;*	*Fut2* $^{-/-}$ dam	0.5597	0.0306	0.1635
	Otu00781	*Bacteroidetes;Bacteroidia;Bacteroidales;Bacteroidaceae;Bacteroides;*	*Fut2* $^{-/-}$ dam	0.5345	0.0204	0.1409
	Otu00919	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{-/-}$ dam	0.5345	0.0237	0.1409
	Otu00930	*Bacteroidetes;Bacteroidia;Bacteroidales;Bacteroidaceae;Bacteroides;*	*Fut2* $^{-/-}$ dam	0.5756	0.0214	0.1409
	Otu00982	*Proteobacteria;BetaproteoBurkholderiales;Sutterellaceae;Parasutterella;*	*Fut2* $^{-/-}$ dam	0.5345	0.0237	0.1409
	Otu01782	*Bacteroidetes;Bacteroidia;Bacteroidales;Bacteroidaceae;Bacteroides;*	*Fut2* $^{-/-}$ dam	0.5345	0.0237	0.1409
	Otu00038	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Robinsoniella;*	*Fut2* $^{+/+}$ dam	0.6574	0.0475	0.2114
	Otu00051	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Alistipes;*	*Fut2* $^{+/+}$ dam	0.8076	0.0142	0.1409
	Otu00059	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	*Fut2* $^{+/+}$ dam	0.9159	0.0001	0.0135
	Otu00062	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Odoribacter;*	*Fut2* $^{+/+}$ dam	0.6530	0.0479	0.2114
	Otu00075	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Lachnospiracea incertae sedis;*	*Fut2* $^{+/+}$ dam	0.8341	0.0063	0.1000
	Otu00077	*Firmicutes;Bacilli;Lactobacillales;Lactobacillaceae;Lactobacillus;*	*Fut2* $^{+/+}$ dam	0.7535	0.0228	0.1409
	Otu00131	*Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;unclassified;*	*Fut2* $^{+/+}$ dam	0.5780	0.0396	0.1827
	Otu00137	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{+/+}$ dam	0.6164	0.0155	0.1409
	Otu00160	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Alistipes;*	*Fut2* $^{+/+}$ dam	0.7004	0.0373	0.1827
	Otu00170	*Bacteroidetes;Bacteroidia;Bacteroidales;Bacteroidaceae;Bacteroides;*	*Fut2* $^{+/+}$ dam	0.6486	0.0258	0.1455
	Otu00220	*Bacteroidetes;Bacteroidia;Bacteroidales;Marinilabiaceae;Anaerophaga;*	*Fut2* $^{+/+}$ dam	0.7379	0.0165	0.1409
	Otu00233	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Alistipes;*	*Fut2* $^{+/+}$ dam	0.6029	0.0390	0.1827
	Otu00251	*Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Pseudoflavonifractor;*	*Fut2* $^{+/+}$ dam	0.6325	0.0126	0.1409
	Otu00290	*Bacteroidetes;Bacteroidia;Bacteroidales;Bacteroidaceae;Bacteroides;*	*Fut2* $^{+/+}$ dam	0.6469	0.0243	0.1409
	Otu00365	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{+/+}$ dam	0.7422	0.0093	0.1180
	Otu00427	*Bacteroidetes;Bacteroidia;Bacteroidales;Marinilabiaceae;Anaerophaga;*	*Fut2* $^{+/+}$ dam	0.6669	0.0173	0.1409
	Otu00436	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;unclassified;*	*Fut2* $^{+/+}$ dam	0.6003	0.0380	0.1827
3	Otu00036	*Bacteroidetes;Bacteroidia;Bacteroidales;Prevotellaceae;Prevotella;*	*Fut2* $^{-/-}$ dam	0.8033	0.0061	0.0790
	Otu00042	*Proteobacteria;BetaproteoBurkholderiales;Sutterellaceae;Parasutterella;*	*Fut2* $^{-/-}$ dam	0.8718	0.0002	0.0144
	Otu00044	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Robinsoniella;*	*Fut2* $^{-/-}$ dam	0.7192	0.0076	0.0790
	Otu00071	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{-/-}$ dam	0.7709	0.0054	0.0790
	Otu00111	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Barnesiella;*	*Fut2* $^{-/-}$ dam	0.5821	0.0229	0.1422
	Otu00143	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Barnesiella;*	*Fut2* $^{-/-}$ dam	0.7559	0.0001	0.0144
	Otu00164	*Firmicutes;Erysipelotrichia;Erysipelotrichales;Erysipelotrichaceae;Erysipelotrichaceae incertae sedis;*	*Fut2* $^{-/-}$ dam	0.5634	0.0260	0.1486
	Otu00167	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	*Fut2* $^{-/-}$ dam	0.7509	0.0029	0.0696
	Otu00176	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	*Fut2* $^{-/-}$ dam	0.6592	0.0407	0.1968
	Otu00189	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Robinsoniella;*	*Fut2* $^{-/-}$ dam	0.5492	0.0241	0.1422

Otu00229	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	*Fut2* $^{-/-}$ dam	0.7239	0.0079	0.0790
Otu00243	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	*Fut2* $^{-/-}$ dam	0.5629	0.0488	0.2022
Otu00268	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	*Fut2* $^{-/-}$ dam	0.7071	0.0006	0.0240
Otu00271	*Bacteroidetes;*	*Fut2* $^{-/-}$ dam	0.6367	0.0446	0.2022
Otu00325	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Robinsoniella;*	*Fut2* $^{-/-}$ dam	0.5976	0.0088	0.0812
Otu00330	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{-/-}$ dam	0.5345	0.0200	0.1422
Otu00404	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	*Fut2* $^{-/-}$ dam	0.6506	0.0042	0.0790
Otu00431	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Tannerella;*	*Fut2* $^{-/-}$ dam	0.5655	0.0298	0.1625
Otu00498	*Bacteroidetes;Bacteroidia;Bacteroidales;Marinilabiaceae;Anaerophaga;*	*Fut2* $^{-/-}$ dam	0.6131	0.0200	0.1422
Otu00517	*Bacteroidetes;Bacteroidia;Bacteroidales;Marinilabiaceae;Anaerophaga;*	*Fut2* $^{-/-}$ dam	0.5976	0.0075	0.0790
Otu00592	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{-/-}$ dam	0.5345	0.0214	0.1422
Otu00631	*Bacteroidetes;Bacteroidia;Bacteroidales;Marinilabiaceae;Anaerophaga;*	*Fut2* $^{-/-}$ dam	0.6873	0.0027	0.0696
Otu00654	*Proteobacteria;BetaproteoBurkholderiales;Sutterellaceae;Parasutterella;*	*Fut2* $^{-/-}$ dam	0.5345	0.0243	0.1422
Otu00757	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Barnesiella;*	*Fut2* $^{-/-}$ dam	0.5058	0.0493	0.2022
Otu00813	*Bacteroidetes;Bacteroidia;Bacteroidales;Marinilabiaceae;Anaerophaga;*	*Fut2* $^{-/-}$ dam	0.5096	0.0354	0.1807
Otu00964	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	*Fut2* $^{-/-}$ dam	0.5345	0.0203	0.1422
Otu00982	*Proteobacteria;BetaproteoBurkholderiales;Sutterellaceae;Parasutterella;*	*Fut2* $^{-/-}$ dam	0.5976	0.0069	0.0790
Otu01027	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Robinsoniella;*	*Fut2* $^{-/-}$ dam	0.5345	0.0203	0.1422
Otu01041	*Bacteroidetes;Bacteroidia;Bacteroidales;Marinilabiaceae;Anaerophaga;*	*Fut2* $^{-/-}$ dam	0.5976	0.0063	0.0790
Otu01308	*Bacteroidetes;Bacteroidia;Bacteroidales;*	*Fut2* $^{-/-}$ dam	0.5976	0.0075	0.0790
Otu01368	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;*	*Fut2* $^{-/-}$ dam	0.5345	0.0223	0.1422
Otu01467	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	*Fut2* $^{-/-}$ dam	0.5345	0.0203	0.1422
Otu01591	*Bacteroidetes;Sphingobacteria;Sphingobacteriales;Cytophagaceae;Meniscus;*	*Fut2* $^{-/-}$ dam	0.5345	0.0225	0.1422
Otu02794	*Bacteroidetes;Bacteroidia;Bacteroidales;Marinilabiaceae;Anaerophaga;*	*Fut2* $^{-/-}$ dam	0.5345	0.0213	0.1422
Otu00005	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Robinsoniella;*	*Fut2* $^{+/+}$ dam	0.8435	0.0023	0.0690
Otu00008	*Proteobacteria;EpsilonproteoCampylobacterales;Helicobacteraceae;Helicobacter;*	*Fut2* $^{+/+}$ dam	0.7896	0.0072	0.0790
Otu00028	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Robinsoniella;*	*Fut2* $^{+/+}$ dam	0.6500	0.0327	0.1744
Otu00031	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Robinsoniella;*	*Fut2* $^{+/+}$ dam	0.7862	0.0032	0.0698
Otu00038	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Robinsoniella;*	*Fut2* $^{+/+}$ dam	0.8322	0.0003	0.0144
Otu00051	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Alistipes;*	*Fut2* $^{+/+}$ dam	0.8726	0.0003	0.0144
Otu00059	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	*Fut2* $^{+/+}$ dam	0.7325	0.0083	0.0797
Otu00065	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Lachnobacterium;*	*Fut2* $^{+/+}$ dam	0.5916	0.0243	0.1422
Otu00075	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Lachnospiracea incertae sedis;*	*Fut2* $^{+/+}$ dam	0.7948	0.0057	0.0790
Otu00101	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Robinsoniella;*	*Fut2* $^{+/+}$ dam	0.5477	0.0495	0.2022
Otu00170	*Bacteroidetes;Bacteroidia;Bacteroidales;Bacteroidaceae;Bacteroides;*	*Fut2* $^{+/+}$ dam	0.7231	0.0079	0.0790

	Otu00185	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Alistipes;*	*Fut2* $^{+/+}$ dam	0.6861	0.0410	0.1968
	Otu00192	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Proteiniphilum;*	*Fut2* $^{+/+}$ dam	0.5979	0.0366	0.1830
	Otu00204	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Alistipes;*	*Fut2* $^{+/+}$ dam	0.8367	0.0002	0.0144
	Otu00224	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Lachnobacterium;*	*Fut2* $^{+/+}$ dam	0.7416	0.0013	0.0446
	Otu00232	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Alistipes;*	*Fut2* $^{+/+}$ dam	0.6831	0.0341	0.1779
	Otu00263	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Robinsoniella;*	*Fut2* $^{+/+}$ dam	0.5916	0.0270	0.1507
	Otu00310	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Alistipes;*	*Fut2* $^{+/+}$ dam	0.6319	0.0483	0.2022
	Otu00448	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Tannerella;*	*Fut2* $^{+/+}$ dam	0.6410	0.0179	0.1422
	Otu00491	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Robinsoniella;*	*Fut2* $^{+/+}$ dam	0.6325	0.0113	0.1004
	Otu00515	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;unclassified;*	*Fut2* $^{+/+}$ dam	0.6543	0.0438	0.2022
	Otu00540	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Syntrophococcus;*	*Fut2* $^{+/+}$ dam	0.6708	0.0046	0.0790
	Otu01198	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Lachnobacterium;*	*Fut2* $^{+/+}$ dam	0.5477	0.0462	0.2022
5	Otu00029	*Bacteroidetes;Bacteroidia;Bacteroidales;Prevotellaceae;Prevotella;*	*Fut2* $^{-/-}$ dam	0.7996	0.0081	0.1463
	Otu00036	*Bacteroidetes;Bacteroidia;Bacteroidales;Prevotellaceae;Prevotella;*	*Fut2* $^{-/-}$ dam	0.8149	0.0039	0.1320
	Otu00044	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Robinsoniella;*	*Fut2* $^{-/-}$ dam	0.7962	0.0011	0.0726
	Otu00083	*Bacteroidetes;Bacteroidia;Bacteroidales;Prevotellaceae;Prevotella;*	*Fut2* $^{-/-}$ dam	0.5345	0.0210	0.1643
	Otu00130	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{-/-}$ dam	0.7470	0.0120	0.1463
	Otu00143	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Barnesiella;*	*Fut2* $^{-/-}$ dam	0.6561	0.0126	0.1463
	Otu00148	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Odoribacter;*	*Fut2* $^{-/-}$ dam	0.7559	0.0001	0.0198
	Otu00152	*Bacteroidetes;Bacteroidia;Bacteroidales;Prevotellaceae;Prevotella;*	*Fut2* $^{-/-}$ dam	0.7087	0.0167	0.1574
	Otu00264	*Actinobacteria;Actinobacteria;Actinomycetales;Nocardiaceae;Millisia;*	*Fut2* $^{-/-}$ dam	0.5976	0.0094	0.1463
	Otu00291	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{-/-}$ dam	0.6445	0.0068	0.1463
	Otu00294	*Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Clostridium IV;*	*Fut2* $^{-/-}$ dam	0.5976	0.0072	0.1463
	Otu00325	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Robinsoniella;*	*Fut2* $^{-/-}$ dam	0.5891	0.0133	0.1463
	Otu00374	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{-/-}$ dam	0.5655	0.0291	0.1746
	Otu00376	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Robinsoniella;*	*Fut2* $^{-/-}$ dam	0.5345	0.0234	0.1643
	Otu00484	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;*	*Fut2* $^{-/-}$ dam	0.5345	0.0204	0.1643
	Otu00485	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Robinsoniella;*	*Fut2* $^{-/-}$ dam	0.5809	0.0144	0.1500
	Otu00574	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{-/-}$ dam	0.5756	0.0237	0.1643
	Otu00584	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Alistipes;*	*Fut2* $^{-/-}$ dam	0.5655	0.0269	0.1664
	Otu00021	*Bacteroidetes;Bacteroidia;Bacteroidales;Bacteroidaceae;Bacteroides;*	*Fut2* $^{+/+}$ dam	0.9499	0.0011	0.0726
	Otu00038	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Robinsoniella;*	*Fut2* $^{+/+}$ dam	0.7745	0.0040	0.1320
	Otu00053	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Alistipes;*	*Fut2* $^{+/+}$ dam	0.7829	0.0189	0.1643
	Otu00059	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	*Fut2* $^{+/+}$ dam	0.7187	0.0113	0.1463
	Otu00065	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Lachnobacterium;*	*Fut2* $^{+/+}$ dam	0.6325	0.0166	0.1574

	Otu00101	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Robinsoniella;*	*Fut2* $^{+/+}$ dam	0.5916	0.0249	0.1643
	Otu00104	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{+/+}$ dam	0.6129	0.0319	0.1858
	Otu00160	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Alistipes;*	*Fut2* $^{+/+}$ dam	0.6831	0.0209	0.1643
	Otu00165	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;*	*Fut2* $^{+/+}$ dam	0.7227	0.0035	0.1320
	Otu00177	*Bacteroidetes;Bacteroidia;Bacteroidales;Bacteroidaceae;Bacteroides;*	*Fut2* $^{+/+}$ dam	0.5477	0.0476	0.2421
	Otu00188	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{+/+}$ dam	0.7135	0.0128	0.1463
	Otu00204	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Alistipes;*	*Fut2* $^{+/+}$ dam	0.6533	0.0368	0.2082
	Otu00224	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Lachnobacterium;*	*Fut2* $^{+/+}$ dam	0.6325	0.0124	0.1463
	Otu00228	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Marvinbryantia;*	*Fut2* $^{+/+}$ dam	0.6003	0.0384	0.2112
	Otu00232	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Alistipes;*	*Fut2* $^{+/+}$ dam	0.5916	0.0245	0.1643
	Otu00320	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Alistipes;*	*Fut2* $^{+/+}$ dam	0.6325	0.0097	0.1463
	Otu00350	*Bacteroidetes;Bacteroidia;Bacteroidales;Bacteroidaceae;Bacteroides;*	*Fut2* $^{+/+}$ dam	0.6325	0.0115	0.1463
	Otu00354	*Bacteroidetes;Bacteroidia;Bacteroidales;Bacteroidaceae;Bacteroides;*	*Fut2* $^{+/+}$ dam	0.5916	0.0268	0.1664
	Otu00449	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	*Fut2* $^{+/+}$ dam	0.5916	0.0249	0.1643
	Otu00498	*Bacteroidetes;Bacteroidia;Bacteroidales;Marinilabiaceae;Anaerophaga;*	*Fut2* $^{+/+}$ dam	0.5875	0.0473	0.2421
11	Otu00071	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{-/-}$ dam	0.8583	0.0003	0.0108
	Otu00082	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Lachnospiracea incertae sedis;*	*Fut2* $^{-/-}$ dam	0.5767	0.0478	0.2074
	Otu00083	*Bacteroidetes;Bacteroidia;Bacteroidales;Prevotellaceae;Prevotella;*	*Fut2* $^{-/-}$ dam	0.7559	0.0002	0.0108
	Otu00090	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Barnesiella;*	*Fut2* $^{-/-}$ dam	0.6863	0.0175	0.1191
	Otu00094	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Robinsoniella;*	*Fut2* $^{-/-}$ dam	0.5976	0.0072	0.0792
	Otu00100	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Robinsoniella;*	*Fut2* $^{-/-}$ dam	0.5345	0.0225	0.1191
	Otu00102	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	*Fut2* $^{-/-}$ dam	0.8406	0.0006	0.0163
	Otu00104	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{-/-}$ dam	0.8338	0.0002	0.0108
	Otu00110	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{-/-}$ dam	0.7496	0.0004	0.0124
	Otu00111	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Barnesiella;*	*Fut2* $^{-/-}$ dam	0.7958	0.0017	0.0362
	Otu00139	*Bacteroidetes;Bacteroidia;Bacteroidales;Marinilabiaceae;Anaerophaga;*	*Fut2* $^{-/-}$ dam	0.7895	0.0015	0.0362
	Otu00143	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Barnesiella;*	*Fut2* $^{-/-}$ dam	0.8018	0.0001	0.0108
	Otu00144	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{-/-}$ dam	0.8321	0.0019	0.0362
	Otu00152	*Bacteroidetes;Bacteroidia;Bacteroidales;Prevotellaceae;Prevotella;*	*Fut2* $^{-/-}$ dam	0.7574	0.0111	0.0926
	Otu00168	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Robinsoniella;*	*Fut2* $^{-/-}$ dam	0.5976	0.0072	0.0792
	Otu00189	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Robinsoniella;*	*Fut2* $^{-/-}$ dam	0.6547	0.0025	0.0417
	Otu00225	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Tannerella;*	*Fut2* $^{-/-}$ dam	0.7220	0.0286	0.1410
	Otu00246	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Alistipes;*	*Fut2* $^{-/-}$ dam	0.6313	0.0391	0.1844
	Otu00268	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	*Fut2* $^{-/-}$ dam	0.8018	0.0001	0.0108
	Otu00273	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Robinsoniella;*	*Fut2* $^{-/-}$ dam	0.5976	0.0085	0.0838

	Otu00294	*Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Clostridium IV;*	*Fut2* $^{-/-}$ dam	0.5345	0.0210	0.1191
	Otu00330	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{-/-}$ dam	0.5756	0.0157	0.1191
	Otu00343	*Bacteroidetes;Bacteroidia;Bacteroidales;Marinilabiaceae;Alkaliflexus;*	*Fut2* $^{-/-}$ dam	0.6276	0.0500	0.2127
	Otu00367	*Actinobacteria;Actinobacteria;Actinomycetales;Nocardiaceae;Millisia;*	*Fut2* $^{-/-}$ dam	0.5345	0.0232	0.1199
	Otu00371	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{-/-}$ dam	0.5345	0.0215	0.1191
	Otu00457	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Tannerella;*	*Fut2* $^{-/-}$ dam	0.6278	0.0102	0.0885
	Otu00539	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Tannerella;*	*Fut2* $^{-/-}$ dam	0.6547	0.0020	0.0362
	Otu00566	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	*Fut2* $^{-/-}$ dam	0.5976	0.0354	0.1707
	Otu00574	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{-/-}$ dam	0.5345	0.0207	0.1191
	Otu00592	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{-/-}$ dam	0.5345	0.0215	0.1191
	Otu00600	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Tannerella;*	*Fut2* $^{-/-}$ dam	0.5976	0.0067	0.0792
	Otu00651	*Bacteroidetes;Bacteroidia;Bacteroidales;unclassified;unclassified;*	*Fut2* $^{-/-}$ dam	0.5976	0.0090	0.0849
	Otu00866	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{-/-}$ dam	0.5345	0.0206	0.1191
	Otu00877	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Tannerella;*	*Fut2* $^{-/-}$ dam	0.6742	0.0034	0.0492
	Otu00989	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Tannerella;*	*Fut2* $^{-/-}$ dam	0.5976	0.0073	0.0792
	Otu01215	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	*Fut2* $^{-/-}$ dam	0.5345	0.0205	0.1191
	Otu01308	*Bacteroidetes;Bacteroidia;Bacteroidales;*	*Fut2* $^{-/-}$ dam	0.5345	0.0202	0.1191
	Otu01454	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{-/-}$ dam	0.5345	0.0202	0.1191
	Otu01729	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Lachnobacterium;*	*Fut2* $^{-/-}$ dam	0.5345	0.0208	0.1191
	Otu01886	*Proteobacteria;AlphaproteoRhizobiales;Beijerinckiaceae;Methylovirgula;*	*Fut2* $^{-/-}$ dam	0.5345	0.0225	0.1191
	Otu00059	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	*Fut2* $^{+/+}$ dam	0.8062	0.0003	0.0108
	Otu00088	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Robinsoniella;*	*Fut2* $^{+/+}$ dam	0.7071	0.0044	0.0597
	Otu00124	*Proteobacteria;AlphaproteoRhizobiales;Hyphomicrobiaceae;Zhangella;*	*Fut2* $^{+/+}$ dam	0.6708	0.0100	0.0885
	Otu00188	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{+/+}$ dam	0.6699	0.0432	0.1994
	Otu00192	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Proteiniphilum;*	*Fut2* $^{+/+}$ dam	0.7654	0.0030	0.0465
	Otu00287	*Firmicutes;Clostridia;Clostridiales;Heliobacteriaceae;Heliobacillus;*	*Fut2* $^{+/+}$ dam	0.5916	0.0248	0.1251
	Otu00296	*Bacteroidetes;*	*Fut2* $^{+/+}$ dam	0.6325	0.0083	0.0838
	Otu00320	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Alistipes;*	*Fut2* $^{+/+}$ dam	0.6431	0.0458	0.2028
	Otu00347	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Alistipes;*	*Fut2* $^{+/+}$ dam	0.5477	0.0450	0.2028
	Otu00394	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	*Fut2* $^{+/+}$ dam	0.6862	0.0190	0.1191
	Otu00423	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;*	*Fut2* $^{+/+}$ dam	0.6325	0.0122	0.0980
1-11	Otu00044	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Robinsoniella;*	*Fut2* $^{-/-}$ dam	0.7035	0.0001	0.0011
	Otu00047	*Bacteroidetes;Bacteroidia;Bacteroidales;Marinilabiaceae;Anaerophaga;*	*Fut2* $^{-/-}$ dam	0.4640	0.0266	0.0367
	Otu00050	*Proteobacteria;GammaproteoEnterobacteriales;Enterobacteriaceae;Escherichia/Shigella;*	*Fut2* $^{-/-}$ dam	0.4636	0.0029	0.0062
	Otu00071	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{-/-}$ dam	0.6843	0.0002	0.0014

Otu00079	*Proteobacteria;AlphaproteoRhodobacterales;Rhodobacteraceae;Pelagibaca;*	*Fut2* $^{-/-}$ dam	0.4775	0.0211	0.0313
Otu00082	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Lachnospiracea incertae sedis;*	*Fut2* $^{-/-}$ dam	0.5647	0.0013	0.0035
Otu00083	*Bacteroidetes;Bacteroidia;Bacteroidales;Prevotellaceae;Prevotella;*	*Fut2* $^{-/-}$ dam	0.5669	0.0001	0.0011
Otu00104	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{-/-}$ dam	0.5774	0.0028	0.0062
Otu00110	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{-/-}$ dam	0.5519	0.0009	0.0028
Otu00111	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Barnesiella;*	*Fut2* $^{-/-}$ dam	0.5787	0.0017	0.0043
Otu00143	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Barnesiella;*	*Fut2* $^{-/-}$ dam	0.6980	0.0001	0.0011
Otu00148	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Odoribacter;*	*Fut2* $^{-/-}$ dam	0.5000	0.0001	0.0011
Otu00151	*Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Oscillibacter;*	*Fut2* $^{-/-}$ dam	0.4444	0.0342	0.0447
Otu00152	*Bacteroidetes;Bacteroidia;Bacteroidales;Prevotellaceae;Prevotella;*	*Fut2* $^{-/-}$ dam	0.6259	0.0002	0.0014
Otu00189	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Robinsoniella;*	*Fut2* $^{-/-}$ dam	0.4867	0.0002	0.0014
Otu00206	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Clostridium XlVa;*	*Fut2* $^{-/-}$ dam	0.4874	0.0027	0.0062
Otu00225	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Tannerella;*	*Fut2* $^{-/-}$ dam	0.5878	0.0005	0.0021
Otu00243	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	*Fut2* $^{-/-}$ dam	0.6194	0.0002	0.0014
Otu00268	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	*Fut2* $^{-/-}$ dam	0.6376	0.0001	0.0011
Otu00291	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{-/-}$ dam	0.5282	0.0040	0.0080
Otu00325	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Robinsoniella;*	*Fut2* $^{-/-}$ dam	0.4665	0.0006	0.0023
Otu00330	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{-/-}$ dam	0.4145	0.0094	0.0171
Otu00383	*Bacteroidetes;Bacteroidia;Bacteroidales;Prevotellaceae;Prevotella;*	*Fut2* $^{-/-}$ dam	0.4298	0.0321	0.0431
Otu00486	*Bacteroidetes;Bacteroidia;Bacteroidales;*	*Fut2* $^{-/-}$ dam	0.4736	0.0033	0.0069
Otu00539	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Tannerella;*	*Fut2* $^{-/-}$ dam	0.4866	0.0004	0.0021
Otu00544	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{-/-}$ dam	0.4088	0.0358	0.0456
Otu00548	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{-/-}$ dam	0.3665	0.0493	0.0575
Otu00554	*Bacteroidetes;*	*Fut2* $^{-/-}$ dam	0.4147	0.0251	0.0356
Otu00565	*Bacteroidetes;Sphingobacteria;Sphingobacteriales;Cytophagaceae;Meniscus;*	*Fut2* $^{-/-}$ dam	0.3814	0.0471	0.0563
Otu00592	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{-/-}$ dam	0.4668	0.0006	0.0023
Otu00631	*Bacteroidetes;Bacteroidia;Bacteroidales;Marinilabiaceae;Anaerophaga;*	*Fut2* $^{-/-}$ dam	0.4381	0.0006	0.0023
Otu00652	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	*Fut2* $^{-/-}$ dam	0.3928	0.0456	0.0552
Otu00655	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Tannerella;*	*Fut2* $^{-/-}$ dam	0.3878	0.0088	0.0163
Otu00683	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Barnesiella;*	*Fut2* $^{-/-}$ dam	0.4011	0.0135	0.0220
Otu00038	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Robinsoniella;*	*Fut2* $^{+/+}$ dam	0.7370	0.0001	0.0011
Otu00059	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	*Fut2* $^{+/+}$ dam	0.7963	0.0001	0.0011
Otu00066	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Acetitomaculum;*	*Fut2* $^{+/+}$ dam	0.4928	0.0009	0.0028
Otu00081	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Robinsoniella;*	*Fut2* $^{+/+}$ dam	0.4455	0.0074	0.0145
Otu00101	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Robinsoniella;*	*Fut2* $^{+/+}$ dam	0.5189	0.0003	0.0017

Otu00124	*Proteobacteria;AlphaproteoRhizobiales;Hyphomicrobiaceae;Zhangella;*	*Fut2* $^{+/+}$ dam	0.5201	0.0002	0.0014
Otu00131	*Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;unclassified;*	*Fut2* $^{+/+}$ dam	0.5174	0.0009	0.0028
Otu00140	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Roseburia;*	*Fut2* $^{+/+}$ dam	0.4289	0.0197	0.0305
Otu00170	*Bacteroidetes;Bacteroidia;Bacteroidales;Bacteroidaceae;Bacteroides;*	*Fut2* $^{+/+}$ dam	0.5596	0.0029	0.0062
Otu00177	*Bacteroidetes;Bacteroidia;Bacteroidales;Bacteroidaceae;Bacteroides;*	*Fut2* $^{+/+}$ dam	0.5765	0.0005	0.0021
Otu00192	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Proteiniphilum;*	*Fut2* $^{+/+}$ dam	0.6256	0.0001	0.0011
Otu00201	*Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Oscillibacter;*	*Fut2* $^{+/+}$ dam	0.4330	0.0014	0.0037
Otu00204	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Alistipes;*	*Fut2* $^{+/+}$ dam	0.6221	0.0005	0.0021
Otu00214	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Lachnobacterium;*	*Fut2* $^{+/+}$ dam	0.4183	0.0016	0.0041
Otu00224	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Lachnobacterium;*	*Fut2* $^{+/+}$ dam	0.5832	0.0003	0.0017
Otu00226	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Acetitomaculum;*	*Fut2* $^{+/+}$ dam	0.4094	0.0199	0.0305
Otu00228	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Marvinbryantia;*	*Fut2* $^{+/+}$ dam	0.5478	0.0022	0.0053
Otu00232	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Alistipes;*	*Fut2* $^{+/+}$ dam	0.5936	0.0008	0.0027
Otu00252	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;*	*Fut2* $^{+/+}$ dam	0.5441	0.0013	0.0035
Otu00255	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Butyrivibrio;*	*Fut2* $^{+/+}$ dam	0.4472	0.0007	0.0025
Otu00263	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Robinsoniella;*	*Fut2* $^{+/+}$ dam	0.5493	0.0008	0.0027
Otu00277	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Marvinbryantia;*	*Fut2* $^{+/+}$ dam	0.4207	0.0116	0.0199
Otu00326	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Alistipes;*	*Fut2* $^{+/+}$ dam	0.4743	0.0003	0.0017
Otu00345	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Alistipes;*	*Fut2* $^{+/+}$ dam	0.5381	0.0026	0.0061
Otu00350	*Bacteroidetes;Bacteroidia;Bacteroidales;Bacteroidaceae;Bacteroides;*	*Fut2* $^{+/+}$ dam	0.5563	0.0011	0.0033
Otu00352	*Bacteroidetes;Bacteroidia;Bacteroidales;Bacteroidaceae;Bacteroides;*	*Fut2* $^{+/+}$ dam	0.4974	0.0038	0.0078
Otu00354	*Bacteroidetes;Bacteroidia;Bacteroidales;Bacteroidaceae;Bacteroides;*	*Fut2* $^{+/+}$ dam	0.4853	0.0130	0.0220
Otu00368	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;unclassified;*	*Fut2* $^{+/+}$ dam	0.4483	0.0210	0.0313
Otu00375	*Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Oscillibacter;*	*Fut2* $^{+/+}$ dam	0.4330	0.0005	0.0021
Otu00392	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Alistipes;*	*Fut2* $^{+/+}$ dam	0.4183	0.0021	0.0051
Otu00394	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenella;*	*Fut2* $^{+/+}$ dam	0.4316	0.0175	0.0281
Otu00406	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Robinsoniella;*	*Fut2* $^{+/+}$ dam	0.3916	0.0088	0.0163
Otu00412	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Roseburia;*	*Fut2* $^{+/+}$ dam	0.4330	0.0012	0.0035
Otu00468	*Bacteroidetes;Bacteroidia;Bacteroidales;Prevotellaceae;Paraprevotella;*	*Fut2* $^{+/+}$ dam	0.3955	0.0483	0.0570
Otu00491	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Robinsoniella;*	*Fut2* $^{+/+}$ dam	0.5000	0.0001	0.0011
Otu00515	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;unclassified;*	*Fut2* $^{+/+}$ dam	0.4478	0.0079	0.0152
Otu00531	*Bacteroidetes;Bacteroidia;Bacteroidales;Marinilabiaceae;Anaerophaga;*	*Fut2* $^{+/+}$ dam	0.3515	0.0454	0.0552
Otu00536	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;*	*Fut2* $^{+/+}$ dam	0.3888	0.0109	0.0194
Otu00538	*Bacteroidetes;Bacteroidia;Bacteroidales;*	*Fut2* $^{+/+}$ dam	0.4008	0.0111	0.0194
Otu00540	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Syntrophococcus;*	*Fut2* $^{+/+}$ dam	0.4729	0.0005	0.0021

	Otu00571	*Bacteroidetes;Sphingobacteria;Sphingobacteriales;Cytophagaceae;Meniscus;*	*Fut2* $^{+/+}$ dam	0.3715	0.0254	0.0356
	Otu00573	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Tannerella;*	*Fut2* $^{+/+}$ dam	0.3498	0.0405	0.0502
	Otu00576	*Bacteroidetes;Bacteroidia;Bacteroidales;Marinilabiaceae;Anaerophaga;*	*Fut2* $^{+/+}$ dam	0.4144	0.0135	0.0220
	Otu00582	*Bacteroidetes;*	*Fut2* $^{+/+}$ dam	0.4099	0.0216	0.0316
	Otu00586	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;unclassified;*	*Fut2* $^{+/+}$ dam	0.3779	0.0350	0.0451
	Otu00609	*Bacteroidetes;Bacteroidia;Bacteroidales;Marinilabiaceae;Anaerophaga;*	*Fut2* $^{+/+}$ dam	0.3911	0.0187	0.0296
	Otu00656	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Lachnobacterium;*	*Fut2* $^{+/+}$ dam	0.3985	0.0327	0.0433
	Otu00657	*Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;*	*Fut2* $^{+/+}$ dam	0.3985	0.0320	0.0431
	Otu00719	*Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Lachnobacterium;*	*Fut2* $^{+/+}$ dam	0.3758	0.0231	0.0333
	Otu00750	*Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Alistipes;*	*Fut2* $^{+/+}$ dam	0.3623	0.0380	0.0477

Table S6: Analysis of alpha diversity for each respective time point.

Alpha diversity	Time	Model Factor	*DF*	*F*-Value	*P*-Value
Species Richness observed	TP1	*Intercept*	1,2	204.763	<0.0001
		Fut2	2,2	2.472	0.1098
		Direction	1,8	4.258	**0.0730**
		Fut2:Direction	2,2	3.817	**0.0394**
	TP3	*Intercept*	1,24	79.444	<0.0001
		Direction	1,8	4.032	**0.0795**
	TP5*	*Intercept*	1,24	260.387	<0.0001
		Direction	1,8	3.119	0.1154
	TP11	Intercept	1,24	151.493	<0.0001
Shannon Entropy	TP1**	*Intercept*	1,24	258.911	<0.0001
	TP3	*Intercept*	1,22	1131.119	<0.0001
		Fut2	2,22	2.889	**0.0769**
	TP5	*Intercept*	1,24	847.781	<0.0001
	TP11	*Intercept*	1,22	1912.295	<0.0001
		Fut2	2,22	5.300	**0.0132**
Net Relatedness Index (NRI)	TP1	*Intercept*	1,24	24.167	**0.0001**
		Direction	1,8	2.656	0.1418
	TP3	*Intercept*	1,23	5.878	**0.0236**
		Secretor	1,23	3.323	**0.0813**
	TP5	*Intercept*	1,24	5.109	**0.0332**
	TP11	*Intercept*	1,24	0.321	0.5765
Nearest Taxon Index (NTI)	TP1	*Intercept*	1,22	52.758	**<0.0001**
		Secretor	1,22	0.029	0.8672
		Direction	1,8	16.269	**0.0038**
		Secretor:Direction	1,22	7.260	**0.0132**
	TP3	*Intercept*	1,24	9.463	**0.0052**
	TP5	*Intercept*	1,24	9.836	**0.0045**
		Direction	1,8	3.150	0.1138
	TP11	*Intercept*	1,24	146.348	**<0.0001**

* $X^{1/2}$ transformed; ** X^{3} transformed

Table S7: Analysis of alpha diversity for each respective location of the gastrointestinal tract.

Alpha diversity	Tissue	Model Factors	*DF*	*F*-Value	*P*-Value
Species Richness observed	Jejunum	*Intercept*	1,21	102.78	< 0.0001
	Ileum	*Intercept*	1,25	119.744	< 0.0001
		Direction	1,9	10.614	0.0099
	Cecum	*Intercept*	1,27	331.427	< 0.0001
	Colon	*Intercept*	1,26	347.451	< 0.0001
Shannon Entropy	Jejunum	*Intercept*	1,21	1149.763	< 0.0001
	Ileum*	*Intercept*	1,25	175.859	< 0.0001
		Direction	1,9	9.885	0.0119
	Cecum	*Intercept*	1,27	2044.081	< 0.0001
	Colon*	*Intercept*	1,26	224.544	< 0.0001
Net Relatedness Index	Jejunum	*Intercept*	1,21	10.116	0.0045

(NRI)	Ileum	*Intercept*	1,25	32.91	< 0.0001
		Direction	1,9	2.85	0.1257
	Cecum	*Intercept*	1,27	53.195	< 0.0001
	Colon	*Intercept*	1,26	0.571	0.4568
Nearest Taxon Index (NTI)	Jejunum	*Intercept*	1,21	47.35	< 0.0001
		Direction	1,8	3.712	0.0902
	Ileum	*Intercept*	1,25	72.999	< 0.0001
		Direction	1,9	3.55	0.0922
	Cecum	*Intercept*	1,27	14.791	0.0007
	Colon	*Intercept*	1,26	133.855	< 0.0001

* X^2 transformed

Table S8: Test for community difference between *Fut2* genotype and breeding direction among fecal time points, based on shared abundance (Bray-Curtis), shared presence (Jaccard), the phylogenetic relatedness (unweighted UniFrac) and distribution of species (Euclidean/Redundancy Analysis).

Distance	**Time point 1**	*F*-Value	*P*-Value	R^2	adj. R^2	**Time point 3**	*F*-Value	*P*-Value	R^2	adj. R^2
Bray-Curtis	*Fut2*	0.912	0.7404	0.178	0.031	*Fut2*	0.9847	0.4970	0.176	0.029
	Direction	1.879	**0.0024**			Direction	1.8125	**0.0006**		
	Fut2:Direction	1.178	**0.1000**			*Fut2*:Direction	1.0952	0.1684		
Jaccard	*Fut2*	0.956	0.7428	0.163	0.014	*Fut2*	0.9911	0.5272	0.163	0.013
	Direction	1.433	**0.0016**			Direction	1.3665	**0.0002**		
	Fut2:Direction	1.057	0.1696			*Fut2*:Direction	1.0415	0.1842		
UniFrac (unweighted)	*Fut2*	0.962	0.5582	0.178	0.031	*Fut2*	1.0245	0.3548	0.167	0.019
	Direction	1.677	**0.0096**			Direction	1.6726	**0.0044**		
	Fut2:Direction	1.230	**0.0584**			*Fut2*:Direction	0.9527	0.6246		
Redundancy Analysis	*Fut2*	0.878	0.6994	0.193	0.049	*Fut2*	1.0875	0.2662	0.225	0.087
	Direction	2.323	**0.0022**			Direction	3.0691	**0.0006**		
	Fut2:Direction	1.312	**0.0822**			*Fut2*:Direction	1.4402	**0.0362**		
Distance	**Time point 5**	*F*-Value	*P*-Value	R^2	adj. R^2	**Time point 11**	*F*-Value	*P*-Value	R^2	adj. R^2
Bray-Curtis	*Fut2*	0.851	0.9712	0.165	0.016	*Fut2*	1.0775	0.1604	0.171	0.024
	Direction	1.586	**0.0018**			Direction	1.7021	**0.0002**		
	Fut2:Direction	1.118	**0.1070**			*Fut2*:Direction	0.9679	0.6256		
Jaccard	*Fut2*	0.922	0.9712	0.159	0.009	*Fut2*	1.0459	0.1326	0.162	0.013
	Direction	1.337	**0.0006**			Direction	1.3721	**0.0002**		
	Fut2:Direction	1.058	0.1216			*Fut2*:Direction	0.9812	0.6610		
UniFrac (unweighted)	*Fut2*	0.895	0.8618	0.166	0.018	*Fut2*	1.1230	0.0882	0.169	0.021
	Direction	1.597	**0.0026**			Direction	1.5453	**0.0010**		
	Fut2:Direction	1.100	0.1588			*Fut2*:Direction	0.9570	0.6706		
Redundancy Analysis	*Fut2*	0.837	0.8408	0.184	0.038	*Fut2*	1.0446	0.3312	0.163	0.013
	Direction	1.828	**0.0086**			Direction	1.6299	**0.0178**		
	Fut2:Direction	1.406	**0.0260**			*Fut2*:Direction	0.8622	0.8200		

Table S9: Community differences of mucosal associated microbial communities between among genotypes and breeding directions, based on shared abundance (Bray-Curtis), shared presence (Jaccard), the phylogenetic relatedness (unweighted UniFrac) and distribution of species (Euclidean/Redundancy Analysis).

Distance	**Jejunum**	*F*-Values	*P*-Values	R^2	**Ileum**	*F*-Values	*P*-Values	R^2
Bray-Curtis	*Fut2*	1.015	0.3982	0.176	*Fut2*	1.015	0.3910	0.173
	Direction	1.424	**0.0600**		Direction	2.254	**0.0010**	
	Fut2:Direction	0.945	0.5918		*Fut2*:Direction	1.002	0.4292	
Jaccard	*Fut2*	1.009	0.3832	0.172	*Fut2*	0.973	0.5868	0.155
	Direction	1.278	**0.0434**		Direction	1.625	**0.0002**	
	Fut2:Direction	0.938	0.7184		*Fut2*:Direction	0.967	0.6312	
UniFrac (unweighted)	*Fut2*	0.939	0.6146	0.175	*Fut2*	0.963	0.5340	0.165
	Direction	1.441	**0.0450**		Direction	1.858	**0.0060**	
	Fut2:Direction	0.999	0.4406		*Fut2*:Direction	1.075	0.2770	
Redundancy Analysis	*Fut2*	0.974	0.4756	0.165	*Fut2*	1.497	0.1322	0.260
	Direction	1.360	0.1056		Direction	4.725	**0.0020**	
	Fut2:Direction	0.807	0.8440		*Fut2*:Direction	1.421	0.1552	
Distance	**Cecum**	*F*-Values	*P*-Values	R^2	**Colon**	*F*-Values	*P*-Values	R^2
Bray-Curtis	*Fut2*	0.994	0.5048	0.152	*Fut2*	0.914	0.8168	0.158
	Direction	1.792	**0.0002**		Direction	2.016	**0.0002**	
	Fut2:Direction	0.973	0.6368		*Fut2*:Direction	0.976	0.5754	
Jaccard	*Fut2*	0.984	0.6356	0.144	*Fut2*	0.955	0.7764	0.152
	Direction	1.433	**0.0002**		Direction	1.638	**0.0002**	
	Fut2:Direction	0.991	0.5610		*Fut2*:Direction	0.994	0.5178	
UniFrac (unweighted)	*Fut2*	0.927	0.8664	0.143	*Fut2*	0.943	0.7208	0.152
	Direction	1.551	**0.0002**		Direction	1.727	**0.0002**	
	Fut2:Direction	0.966	0.6682		*Fut2*:Direction	0.980	0.5620	
Redundancy Analysis	*Fut2*	1.221	0.0552	0.154	*Fut2*	0.882	0.6800	0.139
	Direction	1.588	**0.0072**		Direction	1.503	**0.0600**	
	Fut2:Direction	0.889	0.7982		*Fut2*:Direction	0.871	0.7076	

Table S10: Analysis of the single principle coordinates (all time points combined) according to the influence of *Fut2* genotype, secretor status, and breeding direction.

Metric	Dimension	Factors	*DF*	*F*-Value	*P*-Value
UniFrac	PCo1 (8.745%)	Direction	1,32	3.825	0.0593
	PCo2 (4.071%)	Direction	1,32	7.421	**0.0104**
		Time point* (poly)#	2,98	14.691	**< 0.0001**
		Direction:Time point (poly)	2,98	4.641	**0.0119**
Bray-Curtis	PCo1 (4.460%)	*Fut2*	2,28	0.524	0.5980
		Direction	1,28	17.659	**0.0002**
		Time span† (poly)	2,90	386.593	**< 0.0001**
		Fut2:Direction	2,28	2.701	0.0846
		Fut2:Time span (poly)	4,90	18.073	**< 0.0001**
		Direction:Time span (poly)	2,90	9.077	**0.0003**
		Fut2:Direction:Time span (poly)	4,90	6.071	**0.0002**
	PCo2 (3.821%)	*Secretor*	1,32	0.174	0.6793
		Time point	1,100	54.477	**< 0.0001**
		Secretor:Time point	1,100	5.029	**0.0271**
Jaccard	PCo1 (3.324%)	Direction	1,32	15.923	**0.0004**
		Time point (poly)	2,98	3.286	**0.0415**
		Direction:Time point (poly)	2,98	5.996	**0.0035**
	PCo2 (2.985%)	*Secretor*	1,32	0.110	0.7423
		Time point	1,100	45.830	**< 0.0001**
		Secretor:Time point	1,100	6.060	**0.0155**
Euclidean	PC1 (12.401%)	*Fut2*	2,28	1.546	0.2307
		Direction	1,28	14.928	**0.0006**
		Time point	1,98	90.423	**< 0.0001**
		Fut2:Direction	2,28	22.118	**< 0.0001**
		Direction:Time point	1,98	17.477	**0.0001**
		Fut2:Time point	2,98	7.167	**0.0012**
	PC2 (8.660%)	Direction	1,32	10.006	**0.0034**

* Time point- coded as order of observation (0-3); † Time span coded as time in weeks (0, 2, 4, 10); # fitted as second order polynomial

Table S11: Association of network indices to the association strength to *Fut2* genotype, secretor status, and breeding condition.

		Degree		
Time point	Association	ρ	P	P_{Hommel}
TP1	*Fut2* $^{-/-}$	-0.0701406	0.4948025	0.6681699
TP3		0.2953080	0.0208603	0.1251616
TP5		**0.3811911**	**0.0044570**	**0.0401132**
TP11		0.2841240	0.0433210	0.1993368
TP1	*Fut2* $^{+/-}$	-0.1428704	0.1627020	0.4881060
TP3		**0.3592057**	**0.0044691**	**0.0402220**
TP5		0.2792299	0.0408778	0.1993368
TP11		**0.3737919**	**0.0068936**	0.0620428
TP1	*Fut2* $^{+/+}$	-0.0440741	0.6681699	0.6681699
TP3		0.2261119	0.0797347	0.3189389
TP5		0.3138371	0.0208338	0.1250026
TP11		**0.4555520**	**0.0007818**	**0.0093822**
TP1	Non-secretor	-0.0747999	0.4665131	0.4665131
TP3		0.2962293	0.0204464	0.0613391
TP5		**0.3905822**	**0.0035006**	**0.0245039**
TP11		**0.3136227**	**0.0250221**	**0.0750663**
TP1	Secretor	-0.1258338	0.2193937	0.4387874
TP3		**0.3203014**	**0.0118505**	**0.0474021**
TP5		0.3222445	0.0174840	0.0524521
TP11		**0.4450805**	**0.0010657**	**0.0085255**
TP1	*Fut2* $^{-/-}$ dam	-0.0355422	0.7296274	0.7296274
TP3		**0.3481352**	**0.0059714**	**0.0358281**
TP5		0.3201075	0.0182887	0.0548661
TP11		**0.3493869**	**0.0119704**	**0.0478817**
TP1	*Fut2* $^{+/+}$ dam	-0.1677084	0.1005950	0.2011900
TP3		0.2830848	0.0270596	0.0811788
TP5		0.3309004	0.0145255	0.0541192
TP11		**0.4625994**	**0.0006312**	**0.0050496**

Table S12: Comparison of network disintegration (based on the number of subnetworks) between empirical and simulated random graphs via Kolmogorov-Smirnov test.

		Failure		Attack	
Random graph model	Time point	D	$P_{Bonferroni}$	D	$P_{Bonferroni}$
Degree Sequence (similar degree distribution)	1	0.07407	1.00000	0.62963	**0.00018**
	3	0.08000	1.00000	0.16000	1.00000
	5	0.04545	1.00000	0.13636	1.00000
	11	0.68421	**0.00062**	0.78947	**0.00006**
Barabási (power=1, same number of vertices)	1	1.00000	**1.50364×10^{-11}**	1.00000	**1.50364×10^{-11}**
	3	0.76000	**1.00566×10^{-6}**	0.84000	**1.74637×10^{-7}**
	5	0.72727	**0.00003**	0.81818	**3.21452×10^{-6}**
	11	0.89474	**1.59116×10^{-7}**	1.00000	**4.48224×10^{-8}**
Barabási (power=2, same number of vertices)	1	1.00000	**1.50364×10^{-11}**	1.00000	**1.50364×10^{-11}**
	3	0.76000	**1.00566×10^{-6}**	1.00000	**1.11104×10^{-10}**
	5	0.72727	**0.00003**	1.00000	**2.23157×10^{-9}**
	11	0.89474	**1.59116×10^{-7}**	1.00000	**4.48224×10^{-8}**
Barabási (power=4, same number of vertices)	1	1.00000	**1.50364×10^{-11}**	1.00000	**1.50364×10^{-11}**
	3	0.76000	**1.00566×10^{-6}**	1.00000	**1.11104×10^{-10}**
	5	0.72727	**0.00003**	1.00000	**2.23157×10^{-9}**
	11	0.89474	**1.59116×10^{-7}**	1.00000	**4.48224×10^{-8}**
Erdös-Renyi (similar number, of vertices and edges)	1	1.00000	**1.50364×10^{-11}**	0.70370	0.00001
	3	0.52000	**0.00768**	0.40000	0.14652
	5	0.50000	**0.02916**	0.45455	0.08492
	11	0.31579	1.00000	0.42105	0.27554
Small World (preferential reattachment k=0.6, similar number of vertices)	1	1.00000	**1.50364×10^{-11}**	0.70370	**0.00001**
	3	0.52000	**0.00768**	0.40000	0.14652
	5	0.50000	**0.02916**	0.45455	0.08492
	11	0.31579	1.00000	0.42105	0.27554
Small World (preferential reattachment k=0.8, similar number of vertices)	1	1.00000	**1.50364×10^{-11}**	0.70370	**0.00001**
	3	0.52000	**0.00768**	0.40000	0.14652
	5	0.50000	**0.02916**	0.40909	0.20141
	11	0.31579	1.00000	0.42105	0.27554
Power-law (degree distribution power=4, similar number of vertices)	1	0.96296	**1.07064×10^{-10}**	0.70370	**0.00001**
	3	0.44000	0.05935	0.28000	1.00000
	5	0.40909	0.19739	0.31818	0.86150
	11	0.26316	1.00000	0.47368	0.11262
Power-law (degree distribution power=6, similar number of vertices)	1	1.00000	**1.50364×10^{-11}**	0.70370	**0.00001**
	3	0.48000	**0.02246**	0.36000	0.31329
	5	0.45455	0.08019	0.45455	0.08492
	11	0.26316	1.00000	0.42105	0.27554
Power-law (degree distribution power=8, similar number of vertices)	1	1.00000	**1.50364×10^{-11}**	0.70370	**0.00001**
	3	0.52000	**0.00768**	0.40000	0.14652
	5	0.50000	**0.02916**	0.45455	0.08492
	11	0.26316	1.00000	0.42105	0.27554

Supplementary references:

1. Newman MEJ, Girvan M (2004) Finding and evaluating community structure in networks. Physical Review E 69.

Chapter IV: Expression of the blood-group-related gene *B4galnt2* alters susceptibility to *Salmonella* infection

Host pathogen co-evolution has shaped the expression and function of many genes, and not only limited to those with a direct link to the immune system. Blood group related antigens and their glycosyltransferases show strong signatures of balancing selection [265, 267, 268]. This conserved variation and the prominent role of glycans in cell-cell contact and cell physiology imply a role of these genes in host-pathogen recognition [404] and co-evolution [176, 177, 265]. A well-described example in humans is the polymorphism in the *FUT2* gene (*α-1,2-fucosyltransferase*), which is described elsewhere in this thesis. Overall, host glycans contribute strongly to the intestinal microenvironment for symbiotic microbes by providing carbohydrate sources or attachment sites [221, 395, 405], but in the same way also mediate pathogenic interactions [406, 407].

Mice show a naturally occurring polymorphism in a *cis*-regulatory region of the *B4galnt2* gene, altering its expression and repertoire of glycosylation targets. This variation has been maintained in the mouse lineage for several million years [267]. Intestinal expression of the blood group glycosyltransferase *β-1,4-N-acetylgalactosaminyltransferase 2* (*B4galnt2)* is conserved across vertebrate species [270] and directs the biosynthesis of a carbohydrate antigen similar to blood group A, termed Sd(a) / Lewis^X [269]. However, mice possess an allele that confers a tissue specific switch of *B4galnt2* expression from gut to blood vessels [271] that leads to aberrant glycosylation of the vascular coagulation factor *von Willebrand factor*. This allele is termed "*Modifier of von Willebrand Factor-1*" (*Mvwf1*) [273] and leads to an accelerated clearance of *von Willebrand factor* and prolonged bleeding. *Mvwf1* was first described in the RIIIS/J inbred mouse strain [273]. Subsequent studies revealed common RIIIS/J-like *B4galnt2* alleles that facilitate a tissue-specific expression switch from gut (epithelial) to blood vessel (endothelial) expression in wild mouse populations and species, maintained for millions of years [267]. This points towards a protective role in pathogen resistance that may compensate the effects of prolonged bleeding [268]. A role of *B4galnt2* modified glycans in host-microbe interactions is supported by observations of community alterations in the intestinal microbiota in *B4galnt2* deficient mice [274].

To investigate the role of *B4galnt2* expression in the context of intestinal infection and its interaction with the microbial community, we challenged mice expressing *B4galnt2* in various tissue-specific patterns with the intestinal pathogen *Salmonella enterica* serovar Typhimurium (*S.* Typhimurium). Examining histological and molecular markers of inflammation along with bacterial profiles before and during *S.* Typhimurium infection, we observed compositional changes of the intestinal microbiota according to the expression of

B4galnt2 glycans. Mice deficient in intestinal *B4galnt2* expression developed significantly less inflammation after *S.* Typhimurium infection, in concert with reduced induction of pro-inflammatory cytokine expression and infiltration of immune cells. *B4galnt2* associated intestinal microbial community profiles were further predictive of susceptibility to *S*. Typhimurium infection. Furthermore, we find that vascular *B4galnt2* expression leads to decreased *Salmonella* colonization and increased inflammatory cytokine expression. The effect of *B4galnt2* gut expression on the microbial community was further transferable to wild type gnotobiotic mice (*B4galnt2* $^{+/+}$), which reduced the inflammatory response after *S*. Typhimurium infection in mice which received a *B4galnt2* $^{-/-}$ fecal microbiome.

In conclusion, this study provides insight into the physiological role of *B4galnt2* expression and a potential trade off scenario occurring in the wild, centered on the relative fitness costs of extended bleeding after injuries and decreased susceptibility to intestinal infection.

Publications:

Rausch P, Steck N, Suwandi A, Seidel J, Künzel S, Bhullar K, Basic M, Bleich A, Johnsen J, Vallance B, Baines J, Grassl G. (2015). Expression of the blood-group-related gene *B4galnt2* alters susceptibility to *Salmonella* infection.
Accepted at Plos Pathogens

Expression of the blood-group-related gene *B4galnt2* alters susceptibility to Salmonella infection

Philipp Rausch[1,3,8], Natalie Steck[1,2,8], Abdulhadi Suwandi[1], Janice A. Seidel[1], Sven Künzel[3], Kirandeep Bhullar[4], Marijana Basic[5], Andre Bleich[5], Jill M. Johnsen[6,7], Bruce A. Vallance[4], John F. Baines[1,3,8]*, Guntram A. Grassl[1,2,8,9]*

[1]Institute for Experimental Medicine, Christian-Albrechts-University of Kiel, Kiel, Germany

[2]Models of Inflammation, Research Center Borstel, Borstel, Germany

[3]Max Planck Institute for Evolutionary Biology, Plön, Germany

[4]Department of Pediatrics, Division of Gastroenterology, Child and Family Research Institute, University of British Columbia, Vancouver, British Columbia, Canada

[5]Institute for Laboratory Animal Science, Hannover Medical School, Hannover, Germany

[6]Research Institute, Puget Sound Blood Center, Seattle, WA, USA

[7]Department of Medicine, University of Washington, Seattle, WA, USA

[8]These authors contributed equally to this work

[9]Present address: Institute of Medical Microbiology and Hospital Epidemiology, Hannover Medical School, Hannover, Germany and German Center for Infection Research (DZIF), Hannover, Germany

*Correspondence to co-senior authors: grassl.guntram@mh-hannover.de, phone +49 511 532 4540, Fax +49 511 532 4366; baines@evolbio.mpg.de, phone +49 4522 763 367, Fax +49 4537 763 281

Running title

B4galnt2 expression and *Salmonella* infection

Abstract

Glycans play important roles in host-microbe interactions. Tissue-specific expression patterns of the blood group glycosyltransferase *β-1,4-N-acetylgalactosaminyltransferase 2* (*B4galnt2*) are variable in wild mouse populations, and loss of *B4galnt2* expression is associated with altered intestinal microbiota. We hypothesized that variation in *B4galnt2* expression alters susceptibility to intestinal pathogens. To test this, we challenged mice genetically engineered to express different *B4galnt2* tissue-specific patterns with a *Salmonella* Typhimurium infection model. We found *B4galnt2* intestinal expression was strongly associated with bacterial community composition and increased *Salmonella* susceptibility as evidenced by increased intestinal inflammatory cytokines and infiltrating immune cells. Fecal transfer experiments demonstrated a crucial role of the *B4galnt2*-dependent microbiota in conferring susceptibility to intestinal inflammation, while epithelial *B4galnt2* expression facilitated epithelial invasion of *S.* Typhimurium. These data support a critical role for *B4galnt2* in gastrointestinal infections. We speculate that *B4galnt2*-specific differences in host susceptibility to intestinal pathogens underlie the strong signatures of balancing selection observed at the *B4galnt2* locus in wild mouse populations.

Author Summary

Human blood groups are among the oldest known genetic polymorphisms. It has been proposed that blood group variation is a byproduct of pathogen-driven selection, including in the gastrointestinal tract where blood-group-related genes are often variably expressed. The *B4galnt2* gene is responsible for the synthesis of the Sd(a)/Cad carbohydrate blood group antigen and displays variable tissue-specific expression patterns in wild mouse populations. Using an established model for *Salmonella* Typhimurium induced colitis, we found that loss of *B4galnt2* expression in the intestinal epithelium decreases susceptibility to infection. These effects were strongly associated with the influence of *B4galnt2* expression on the intestinal microbiota, whereby microbial diversity prior to infection was highly predictive of inflammation and resistance to *Salmonella* Typhimurium infection. Additionally, *B4galnt2* expression in blood vessels also distinctly influenced intestinal phenotypes and *Salmonella* susceptibility. These data lend new insights into bacterial community diversity as an “extended phenotype” that can be mediated by host genetic variation at blood-group-related genes. This work further provides strong experimental evidence in support of a scenario of complex selection on the *B4galnt2* tissue-specific expression variants via host-microbe relationships and susceptibility to infectious disease.

Introduction

The luminal surface of the intestinal mucosa is covered by distinct layers of highly glycosylated mucus that form a physical barrier between the intestinal microbial community and the host's tissues. In addition to their important roles in host metabolism and signaling, glycans are known to contribute to the composition and physiology of the intestinal microbiota, thereby playing an important role in regulating microbe-host interactions [1]. Host glycans can contribute to a beneficial microenvironment for symbiotic microbes by providing carbohydrate sources or by serving as attachment sites [1-3], but glycans can in the same way also mediate pathogenic interactions [4, 5]. The patterns of intestinal carbohydrate structures, which vary along sites of the gastrointestinal tract, are the product of a combination of host glycosyltransferase expression programs as well as microbial influences [6, 7].

The genes responsible for synthesizing carbohydrate blood group antigens frequently display signatures of balancing selection and are implicated in the co-evolution of hosts and their pathogens [8]. A well-described example is the *FUT2* gene, which encodes an *α-1,2-fucosyltransferase* that directs the expression of the H antigen in mucosal tissues and bodily secretions. Homozygosity for loss-of-function *FUT2* mutations leads to loss of expression of ABO and H blood group glycans in secretions and is known as the "nonsecretor" phenotype, which is common in human populations [9]. Nonsecretor status has been implicated as a detrimental genetic risk factor for inflammatory disorders such as Crohn's disease [10] and primary sclerosing cholangitis [11], while being positively associated with resistance to intestinal pathogens [12-14]. Glycosylation of the epithelium has recently been recognized as a direct immune cell mediated response to infection as a means to restore the protective functions of the microbial community and to ensure tissue homeostasis [15-17]. Glycans can also mediate host species preferences among pathogens, for example the presence of particular Helicobacter species in the canine gastric mucosa [18].

Gastrointestinal (GI) expression of the blood group glycosyltransferase *β-1,4-N-acetylgalactosaminyltransferase 2* (*B4galnt2*), which directs biosynthesis of a carbohydrate antigen similar to blood group A termed the Sd(a) [19] is conserved across vertebrates [20]. However, in mice there is a common allele which confers a tissue specific switch in *B4galnt2* expression from gut to blood vessels [21]. This allele is termed "*Modifier of von Willebrand Factor-1*" (*Mvwf1*) [22] because *B4galnt2* vascular expression leads to aberrant glycosylation of the vascular-derived blood coagulation factor von Willebrand factor (VWF), resulting in accelerated VWF clearance from circulation [23]. *Mvwf1* was first described in the RIIIS/J inbred mouse strain [22], and subsequent studies revealed RIIIS/J-like *B4galnt2* alleles, which confer the *B4galnt2* tissue-specific switch from gut (epithelial) to blood vessel

(endothelial) expression, to be common in wild mouse populations [24]. Further, this variation appears to have been maintained in the mouse lineage for several million years despite the presumed detrimental effect of prolonged bleeding time, possibly due to a protective role in host-pathogen interactions [25]. A role for *B4galnt2*-glycans in intestinal host-microbe interactions is supported by the observation of significant alterations in the intestinal microbiota in *B4galnt2*-deficient mice [26]. Taken together, the prevalence of alleles conferring the tissue-specific switch in *B4galnt2* expression in mice, the strong signatures of selection observed at the *B4galnt2* locus in wild mouse populations and the altered resident microbiota found in *B4galnt2*-deficient mice support the hypothesis that variant tissue-specific *B4galnt2* expression alters susceptibility to enteric infections in mice.

To investigate the role of variant host *B4galnt2* expression in the context of intestinal infection, we challenged mice engineered to express *B4galnt2* in various tissue-specific patterns with a mouse model of the intestinal pathogen *Salmonella enterica* serovar Typhimurium (*S.* Typhimurium). Prior to- and during the course of infection, we examined histological and molecular markers of inflammation along with bacterial community profiles. We found that the composition of the intestinal microbiota was consistently influenced by the expression of *B4galnt2*-glycans, and that *B4galnt2*-associated intestinal microbial community profiles were predictive of- and responsible for susceptibility to *S.* Typhimurium infection. We demonstrate that mice deficient in intestinal *B4galnt2* expression developed significantly less pathology after S. Typhimurium infection, in concert with attenuated induction of pro-inflammatory cytokines and infiltration of immune cells. Furthermore, we find that vascular *B4galnt2* expression leads to decreased *Salmonella* colonization and increased inflammatory cytokine expression. Overall, our study elucidates a new role for this key host carbohydrate blood group antigen in the interplay between the host, commensals, and susceptibility to pathogen infections.

Results

B4galnt2 expression influences susceptibility to *S.* Typhimurium-induced colitis

To test the hypothesis that expression of intestinal *B4galnt2* glycans influences host susceptibility to enteric pathogens, we used an established model for S. Typhimurium induced colitis [27]. Mice were bred to carry the desired combinations of alleles which express *B4galnt2* in the intestinal epithelium ("*B6*": referring to the endogenous C57BL6/J allele), vascular endothelium ("*RIII*": referring to the RIIIS/J-derived *Mvwf1* bacterial artificial chromosome transgene [21]), or lack a functional *B4galnt2* gene due to a targeted knock-out allele ("*B6* $^{-/-}$": referring to the *B4galnt2* knock-out [23]). Twenty-four hours after streptomycin

pre-treatment, mice were orally infected with *S.* Typhimurium SL1344 ("acute" infection, examined after 24 hours [28]) or the attenuated *ΔaroA* mutant ("chronic" infection, examined after 14 days [29]). None of the animals showed signs of inflammation or other pathology prior to infection. After infection in both the acute and chronic *Salmonella* models, mice expressing *B4galnt2* in the intestinal epithelium (*B6* $^{+/-}$ / *RIII* $^{-}$ and *B6* $^{+/-}$ / *RIII* $^{+}$) exhibited higher numbers of detached epithelial cells and neutrophils within the cecal lumen, increased inflammatory cell infiltration [29, 30] within the intestinal mucosa, and worsened submucosal edema in the ceca (Figure 1A). The dramatic reduction of cecum weight in infected *B6* $^{+/-}$ mice compared to *B6* $^{-/-}$ mice in acute *Salmonella* infection one day post infection (p.i.) indicated more severe disease [27] (Figure 1B). Accordingly, mice that did not express *B4galnt2* in the intestinal epithelium (*B6* $^{-/-}$) developed significantly less cecal inflammation in both the acute and chronic infection model (Figure 1C).

In order to evaluate *Salmonella* colonization, colony forming units (CFUs) were quantified from homogenized ceca. While *Salmonella* burdens were comparable between different *B4galnt2* intestinal epithelial-expressing genotypes (*B6*), *RIII* $^{+}$ (*B4galnt2*-endothelial expressing) animals exhibited lower *Salmonella* colonization in the acute *Salmonella* infections (Figure 1D). These results demonstrate a significant influence of intestinal epithelial *B4galnt2* expression on susceptibility to *Salmonella*-induced colitis, and an independent effect of vessel-specific *B4galnt2* expression on *Salmonella* burden. In contrast, infection of mice without prior streptomycin treatment resulted in equal bacterial organ colonization, organ weights, and elicited no intestinal inflammation regardless of the genotype of mice (Figure S1). Due to the marked differences between mouse *B4galnt2* genotypes in the acute infection model, we performed further studies only in this model.

B4galnt2-GalNAc residues have been shown to be detectable on the apical surface of intestinal epithelial cells [23, 26]. Immunohistochemical co-staining with Dolichos biflorus agglutinin (DBA) specifically detecting *B4galnt2*-derived β-1-4 linked GalNAc residues [21, 23] and MUCIN 2 (MUC2), the major secreted mucus protein in the large intestine, demonstrated a partial co-localization in goblet cells (Figure 2A, S2A). While MUC2 is considered to be glycosylated by B4GALNT2 [31], GalNAc residues were also detected in the intestinal mucosa of *Muc2*-deficient mice (Figure S2B), indicating the presence of other B4GALNT2-glycosylated substrates such as glycolipids [32, 33] and other glycoproteins [34-36]. To determine if *B4galnt2*-mediated glycosylation altered overall mucus thickness, which could make it easier for bacteria to cross the mucus layer and reach the epithelium, intestinal tissue of uninfected mice were fixed with Carnoy's fixative, stained with alcian blue and the thickness of the dense inner mucus layer was determined. Although mucus thickness was not significantly affected by the lack of intestinal *B4galnt2* expression (*B6* $^{-/-}$), it did show

slight differences between *RIII* $^{+}$ and *RIII* $^{-}$ (Figure 2B and 2C). Furthermore, less DBA lectin staining was observed in the cecal mucosa of *S.* Typhimurium infected mice on day one p.i. compared to uninfected mice (Figure 2D). In contrast to the DBA staining (GalNAc), the detection of N-Acetylglucosamine (GlcNAc) residues recognized by Wheat Germ Agglutinin (WGA) showed no clear difference after infection, suggesting the alteration of mucosal DBA lectin-reactive carbohydrate profiles that occur in response to *S.* Typhimurium infection did not affect substrates glycosylated by WGA-reactive GlcNAc (Figure 2D). *B4galnt2* gene expression was also down regulated upon infection (Figure 2E) which further corroborates the lectin staining results.

To test the direct effect of *B4galnt2* expression on *Salmonella's* interaction with the cecal epithelium, we performed both FISH staining of cecal sections 1 day p.i. as well as in vitro experiments with the intestinal epithelial Mode-K cell line and siRNA-mediated knockdown of *B4galnt2* expression. Bacteria were stained by FISH using the Gam42a probe, which stains γ-Proteobacteria. In our experience virtually all Gam42a positive bacteria reaching the tissue in the streptomycin model at day 1 p.i. are *Salmonella*. Bacteria were counted if they were adherent to epithelial cells or invaded into the tissue in ten high power fields per cecal section. While adherent *Salmonella* were not significantly different in *B6* $^{+/-}$ mice compared to *B6* $^{-/-}$ mice, significantly more *Salmonella* were found to have invaded into the tissue of *B6* $^{+/-}$ mice (Figure 3A). To further investigate whether *B4galnt2* expression influences the interaction of *Salmonella* with epithelial cells, we used the intestinal epithelial Mode-K cell line and siRNA-mediated knockdown of *B4galnt2* (knockdown efficiency: 96%; Figure 3B). Adhesion and invasion assays showed that knockdown of *B4galnt2* expression does not significantly influence adhesion of *Salmonella* to epithelial cells (Figure 3C). However, invasion of *S.* Typhimurium into *B4galnt2*-expressing cells is slightly, but significantly increased relative to *B4galnt2*-knockdown cells (Figure 3C). This data shows that epithelial expression of *B4galnt2*- both in vitro and in vivo- directly facilitates invasion by *Salmonella*.

Intestinal epithelial *B4galnt2* glycans are associated with elevated cytokine levels and higher numbers of inflammatory/immune cells after *S.* Typhimurium-induced colitis

We analyzed the transcript levels of pro-inflammatory cytokine genes in cecal tissues both prior to and after *S.* Typhimurium infection, focusing on those cytokines known to be induced early in *Salmonella*-triggered inflammation and associated with control of infection [37, 38]. The transcripts for the cytokines *Tumor necrosis factor-α* (*Tnf*-α), Interleukin-6 (*Il-6*), *Interferon-γ* (*Ifn-γ*) and Monocyte chemotactic protein-1 (Mcp-1) were elevated in all mice after infection, but to a significantly higher degree in *B6* $^{+/-}$ mice compared to *B6* $^{-/-}$ mice one day p.i. (Figure 4A-D; *Tnf-α*: *Z*=-2.123, *P*=0.0336; *Il-6*: *Z*=-2.458, *P*=0.0138; *Ifn-γ*: *Z*=-2.417,

P=0.0147; *Mcp-1*: *Z*=-2.219, *P*=0.0261; Wilcoxon test via Monte-Carlo resampling). Protein levels of Lipocalin-2 (LCN-2), a molecule implicated in antimicrobial defense and innate immunity [39, 40], were also increased in cecal tissue homogenates in *B6* $^{+/-}$ mice compared to *B6* $^{-/-}$ mice after infection (Figure 4E, Table S1). Furthermore, vascular endothelial *B4galnt2* expressing animals (*RIII* $^{+}$) exhibited increased *Il-6* expression (*Z*=-1.932, *P*=0.0528), but decreased LCN-2 production (Table S1), suggesting a role for vascular *B4galnt2* expression in the host immune response to intestinal infection (Figure 3).

We also analyzed cecal tissue sections for the presence of cells positive for CD68, which is strongly expressed by monocytes and macrophages, and CD3, which is expressed on mature T cells. Immunohistochemical staining and subsequent quantification of cell numbers revealed no difference in cell numbers according to endothelial (*RIII* $^{+}$) *B4galnt2* expression, but significantly fewer CD68 $^{+}$ and CD3 $^{+}$ cells were observed in the cecal tissues of *B6* $^{-/-}$ mice (Figure 5A, B, S3A, Table S1) after infection. The presence of neutrophils was further investigated by myeloperoxidase (MPO) staining. In line with our previous results, *B6* $^{-/-}$ had fewer MPO positive cells in the intestinal mucosa (lumen and edema) compared to *B6* $^{+/-}$ mice (Figure 4C, S3B) one day p.i., which was further quantified by the relative fluorescence signal intensity (*P*=0.0001; Figure 4D, Table S1). Overall, we detected differences in the abundance of CD68 $^{+}$ and CD3 $^{+}$ cells after infection with respect to the expression of *B4galnt2* in the intestinal epithelium, but almost no differences with respect to vascular endothelial expression.

Bacterial diversity within and between mice is influenced by intestinal epithelial expression of *B4galnt2*

To examine the effect of *B4galnt2* genotype on the intestinal microbiota in the context of infection, pyrosequencing of the 16S rRNA gene in fecal samples was performed for each individual before and after streptomycin treatment, and after S. Typhimurium infection. This resulted in a total of 122,818 sequences, with an average of 998.52 ± 13.49 SD reads per sample after normalization (Good's coverage of OTUs: 92.46 ± 9.05% SD).

To obtain a detailed picture of the interaction of microbial communities with host factors, we first assessed within-sample (alpha) diversity at multiple complementary levels including species richness (Chao1), distribution (Shannon H), and two phylogenetic measures including Nearest Taxon Distance (NTI) and the Net Relatedness Index (NRI) [41]. Species diversity within and between bacterial communities was strongly influenced by the administration of streptomycin and *S.* Typhimurium (Figure S4). Prior to streptomycin treatment and infection, the richness and evenness of operational taxonomic units (OTUs) show no significant differences according to *B4galnt2* genotype (Figure 6A, 6B Table 1) in

concordance with the results of Staubach *et al.* [26]. Phylogenetic clustering among close relatives (NTI) is significantly increased in animals with *B4galnt2* expression in the endothelium (*RIII* $^{+}$), while clustering of large phylogenetic groups (NRI) shows no discernable patterns (Figure S5, Table 1).

After S. Typhimurium infection, the number of species and the evenness of their distribution showed a clear decrease with inflammation (Figure 6C, S5C). Phylogenetic clustering of deep branches, on the other hand, is only weakly influenced by genotype and inflammation after *S.* Typhimurium infection (Figure S5E, Table 1), while terminal phylogenetic clustering (NTI) shows a strong negative correlation to inflammation (Figure S5, Table 1). In addition, the abundance of *S.* Typhimurium detected by 16S rRNA gene sequencing is influenced by *B6*- and *RIII* genotype, especially the low abundance observed in the *RIII* $^{+}$/*B6* $^{-/-}$ genotype (Figure S6), which is consistent with the observations based on colony forming units (Figure 1D; see above).

Next, we attempted to determine which aspects of microbial communities may be associated with infection susceptibility by correlating diversity measurements prior to antibiotic treatment to the outcome of infection (inflammation score, *S.* Typhimurium load). Species richness, distribution, and the amount of phylogenetic clustering displayed a significant relationship to the severity of infection outcome, whereby pathology is predicted with relatively high power (Table 2). Furthermore, epithelial *B4galnt2* expression (*i.e. B6*) significantly increases predictive power (Figure 6D, Figure S7) and may therefore be an important factor modifying the involvement of the microbiota during pathogenesis. Specifically, species loss (ΔChao1) caused by the streptomycin and S. Typhimurium infection, which is higher in phylogenetically clustered and species rich communities (ΔChao1~NTI before infection, ρ=-0.4216, P=0.006435, ΔChao1~Chao1 before infection, ρ=-0.9854, $P < 2.2 \times 10^{-16}$; Spearman rank correlation) may explain why high species diversity before treatment is correlated to a high inflammatory response (Table 2). Community resistance, measured here as the community turnover (Δ unweighted UniFrac) between the pre- and post-infection time points, is higher in *B6* $^{-/-}$ mice (*i.e.* lower Δ unweighted UniFrac; Figure 6E) and shows a strong positive correlation with inflammation and species diversity (Figure 6F, G, Figure S8, Table S2). Interestingly, the community turnover between the untreated and streptomycin treated communities (before infection) is not associated to the final *Salmonella* load or severity of inflammation. Thus, *B4galnt2* expression in the gut epithelium influences the diversity and resistance of bacterial communities, which in turn is associated with the outcome of infection. Furthermore, these results also underscore the metastable character of highly diverse communities, as was already implied by May in 1972 [42].

To infer whether differences between the bacterial communities of mice with different *B4galnt2* expression patterns may contribute to susceptibility, we performed beta diversity analyses. Accordingly, diversity between communities was measured based on different characteristics in untreated animals, including OTU- presence/absence (Jaccard/JA), -abundance (Bray-Curtis/BC) and -distribution (Redundancy Analysis/RDA), in addition to the presence/absence- (unweighted UniFrac/UW-UF)) and abundance of phylogenetic branches (weighted UniFrac/W-UF). This yielded similar community differences with respect to *B6* genotype in nearly all measures (Figure 7A, Figure S9, Table S3) and importantly, confirms the previous findings of Staubach *et al.* 2012 [26] with the current cohort of mice, which were re-derived and housed in a different animal facility. In addition, the bacterial communities among *B6* $^{+/-}$ animals displayed far less inter-individual variation in their community composition than *B6* $^{-/-}$ animals (Figure S9, S10).

Differences in community structure after S. Typhimurium infection were also evaluated and correlated with inflammation score as an additional variable. This showed that differences in communities with respect to *B4galnt2* genotype are also present after infection. Furthermore, the communities changed their species composition with increasing inflammation, which appeared to be most prominent in the microbiota of *B6* $^{+/-}$ animals (RDA: *B6*-$F_{1,38}$=3.4908, P=0.0022, inflammation- $F_{1,38}$=5.0547, P=0.0002, adjusted R^2=0.1406; Figure 7B, Figure S8, Table S3). Lastly, the inter-individual distance among *B6* $^{-/-}$ also remained higher after *S*. Typhimurium infection (Figure S9, S10).

Indicator species and genera characterize the bacterial communities according to intestinal epithelial expression of *B4galnt2*.

To investigate the drivers of community differentiation between *B4galnt2* genotypes, we employed indicator species analysis. Before treatment and subsequent infection, several genera and species were associated with *B4galnt2* expression (*B6* $^{+/-}$) in the gut, including members of the *Bacteroidales* (*Bacteroides*, *Prevotella*, *Prevotellaceae*) and *Parasutterella* (Proteobacteria), while *Turicibacter* (Firmicutes) and other members of the *Bacteroidales* (*Barnesiella*, *Porphyromonas*, *Porphyromonadaceae*) were indicative of mice lacking *B4galnt2* expression in the gut (*B6* $^{-/-}$; Figure 7C, D, Table S4, Table S5). In addition, *Turicibacter*, *Erysipelotrichaceae*, and *Marvinbryantia* (Firmicutes) are associated to endothelial expression of *B4galnt2* glycans (*RIII* $^{+}$; Figure 7C, D, Table S4). To further understand the nature of potential interactions among indicator taxa, we performed a targeted correlation network analysis using Spearman rank correlations of the indicator genera to the remaining community members. Interestingly, the genera displaying differential preferences with respect to *B4galnt2* genotype were also negatively correlated with one another, suggesting competitive exclusion mediated by the presence/absence of *B4galnt2*

glycans (*Turicibacter-Bacteroides*: ρ=-0.485, *P*=0.0013; *Turicibacter-uncl*. *Prevotellaceae*: ρ=-0.447, *P*=0.0034). Further, only *Turicibacter*, which is an indicator for the lack of *B4galnt2* expression in the gut, is directly correlated to the indicators of *B6* $^{+/-}$ genotype while *uncl*. *Porphyromonadaceae* (*B6* $^{-/-}$ indicator) are only associated to *Turicibacter* abundance (Figure 8A). Through this analysis we additionally found *Parabacteroides* as negatively associated to *Bacteroides* and *Prevotellaceae*, suggesting either competition for *B4galnt2* glycans or a secondary indicator for their absence (Figure 8A, Table S6). Furthermore, we detected associations of taxa post infection, such as an overabundance of *Salmonella* and *Cyanobacteria* in *B6* $^{+/-}$, and *uncl*. *Bacteroidales* and uncl. Firmicutes in *B6* $^{-/-}$ mice. Interestingly, we found taxa associated to *B4galnt2* expression in the gut overlapping with a previous study by Staubach *et al.* (2012), such as *Barnesiella* and *Porphyromonadaceae* (Table S4, S5) [26], which further strengthens the evidence for interactions with *B4galnt2* given the independence of these cohorts of mice (see above). Lastly, we explored the dataset for individual taxon associations with inflammation, revealing *Turicibacter* and *Salmonella* to be positively associated to inflammation, potentially benefiting from the inflammatory reactions at the epithelial barrier. Other indicators for the absence of *B4galnt2* glycans like *Parabacteroides* or *Prophyromonadaceae*, however, decline with increasing inflammation (Table S7). Only the *uncl*. *Erysipelotrichaceae*, which are secondary indicators for the absence of *B4galnt2* glycans in the epithelium (see Figure 8A, Table S6, S7), are potential probiotic bacteria whose abundance prior to treatment decreases with inflammation (ρ=-0.320, *P*=0.0417). The analysis of the complete co-occurrence network revealed strong dependencies among community members before treatment (Figure 8B). Specifically, we found a higher incidence of weak negative interactions (competition), and a low number of very strong positive interactions (Figure 8B). The co-occurrence network after *S.* Typhimurium infection shows a comparable distribution of positive and negative interactions, as observed before infection (Figure S11A). Further, it reveals the widespread impact of *Salmonella* (indicator of *B6* $^{+/-}$) on the microbial community, as its position is highly central and strongly influences several other highly integrated parts of the community (Figure S11B).

Increased susceptibility of *B6* $^{+/-}$ mice to *S.* Typhimurium triggered inflammation is dependent on microbiota composition.

In order to determine whether the microbiota composition contributes to the elevated susceptibility of *B6* $^{+/-}$ mice to inflammation, we transplanted feces from *B6* $^{+/-}$ and *B6* $^{-/-}$ donor mice into germfree C57BL/6J (*B6* $^{+/+}$) recipient mice. 21 days post fecal transplantation, mice were treated with streptomycin and 24 hours later infected with *S.* Typhimurium. Cecum weight and *S.* Typhimurium colonization (CFU count) do not differ significantly between the

fecal donor genotypes (Figure 9A, 9B, and 9C). However, the extent of tissue inflammation caused by *S.* Typhimurium infection was significantly lower in mice transplanted with microbiota from *B6* $^{-/-}$ mice due to decreased mucosal damage and decreased submucosal edema (Figure 9A and 9D). These results demonstrate that the differences in microbiota composition from *B6* $^{+/-}$ and *B6* $^{-/-}$ mice are responsible for the lower susceptibility of *B6* $^{-/-}$ mice to *Salmonella* induced inflammation.

Discussion

Infectious diseases are one of the strongest selective forces on many levels of biological complexity. Over time, a steady cycle of adaptation and counter-adaptation has left molecular traces in the genomes of many organisms including humans [43]. The most prominently affected members are genes associated with the immune system, *e.g.* MHC [44], however, others including blood-group-related genes display similar signatures of selection [3, 8, 45-47]. In this study, we investigated intestinal infection as a potential driver of selection at *B4galnt2* observed in the wild by studying the effect of variant tissue-specific expression of *B4galnt2* on host-microbiota interactions and susceptibility to intestinal infection with *Salmonella*. This revealed strong evidence for the influence of *B4galnt2*-specific host glycosylation on microbial community composition and a role in pathogen resistance.

Our experiments revealed less intestinal pathology, lower inflammatory responses, and changes in microbial community structure and composition in animals lacking *B4galnt2* expression in their intestinal epithelium. Host mucosal glycans can directly interact with the microbiota by serving as specific attachment sites or as nutrient sources for some microorganisms. Thus, host mucosal carbohydrates can influence, directly and indirectly, the establishment of overlapping competitive niches, which serve as a barrier against potential pathogens (*i.e.* “colonization resistance”) [48, 49]. We found *B4galnt2*-expression-dependent characteristics of the intestinal microbiota, such as species and phylogenetic diversity, which predict the colonization success of *S.* Typhimurium and the severity of the accompanying intestinal inflammation. In our experiment, species-rich and phylogenetically clustered microbial communities appear to be more vulnerable to *Salmonella* infection, and ultimately inflammation. Before the seminal works of May and others [42, 50, 51], high diversity habitats were synonymous with high stability and productivity [52, 53]. However, the diversity-stability debate remains unresolved [54-56]. High diversity only has a stabilizing effect if reactions of community members are asynchronous, which balances the reduction of one species by the complementary increase of other community members [57-59]. This “portfolio”- [60] or

"insurance" effect [61] dampens perturbations by a release of inter-species competition, or by differential susceptibility to the environmental stressors [54]. Diverse communities also exhibit an intrinsically higher tendency of community change, as a large number of species (especially rare species) are prone to becoming lost through environmental perturbation and stochastic events due to their limited relative abundance [62]. This is observed in *e.g.* grassland communities, where compositional instability increases with community diversity [63]. Thus, the comparably high number of strong positive interactions in the bacterial communities of this study (see Figure 8B) may therefore explain the tendency of exacerbated species loss and inflammation after disturbance, as the stabilizing effects of competitive release are lower [57, 60, 64, 65]. Furthermore, evolutionary relatedness among community members has a strong influence on community reactions and productivity. Closely related species (*e.g.* phylogenetically clustered) presumably overlap in their niches and functional capacities [66, 67] and react in similar ways to environmental stressors, which dampens the insurance effect (*i.e.* "negative insurance effect") as observed in the investigated microbial communities [68, 69].

Antibiotic treatment usually has long lasting effects, but previous studies show that a certain degree of resilience occurs through short-term repopulation of dormant bacteria [49, 70]. The disturbance in microbial communities appears to be buffered in mice not expressing *B4galnt2* glycans in the epithelium, possibly by conferring "colonization resistance" via a higher potential to compete with invading *Salmonella* and by dampening the effects of community disturbance [67, 71, 72]. Thus, in the context of a diminished and disturbed microbial community after streptomycin treatment [73], it is likely that the more resilient/resistant communities in *B4galnt2* $^{-/-}$ mice maintain a greater potential for rapid recovery [48, 70, 74]. We further postulate that *B4galnt2* genotype-dependent host-microbe interactions modulate the host's immune response, contributing to less severe pathology and increased pathogen clearance in mice lacking intestinal epithelial *B4galnt2* expression.

Commensal gut bacteria benefit from the intestinal mucus and its diverse glycan residues, as they offer a complex repertoire of binding sites and carbohydrate sources independent of the host diet [1, 3, 75, 76]. The indicator species identified for mucosal *B4galnt2* expression, *Prevotella* and *Bacteroides*, are known to digest and bind a large spectrum of glycans [77]. These bacteria of high metabolic potential show signs of niche competition with the genus *Turicibacter*, an indicator for *B4galnt2*-deficient mice. *Turicibacter*, *e.g. Turicibacter sanguinis*, is a known member of the human and murine gut microbiome, but can only utilize a narrow range of carbohydrates [78]. As suggested by Dimitriu *et al.* (2013) [79], the trade-off between low metabolic capacity and competitive abilities [78, 80] with the potential for fast colonization might explain the association of

Turicibacter with *B6* $^{-/-}$ mice and the co-increase with *S.* Typhimurium [81-83]. It was also suggested that *Turicibacter* possesses immune modulatory characteristics (increasing iNK T cell, and marginal zone B cell abundance [84]), and may thus help to lower the susceptibility to gut inflammation in *B6* $^{-/-}$ compared to *B6* $^{+/-}$ mice in the face of equivalent *Salmonella* burdens [79]. However, *Turicibacter* could also benefit from existing tissue inflammation, as several genomic features such as laminin, internalin, or a collagen binding pilus allow this genus to act as an opportunistic pathogen, and thus explain its association with tissue inflammation [78, 80]. Similarly, *Barnesiella* shows repeated association to the absence of *B4galnt2* glycans [26]. This genus has the potential to counteract inflammatory responses and thus appears to play a central role in the gut microbiome [85].

Co-staining of MUC2 and DBA lectin demonstrated a partial co-localization in goblet cells, suggesting that MUC2 is glycosylated by B4GALNT2 in agreement with previously published data [31]. However, *B4galnt2* glycans were also detectable in the cecal mucosa of *Muc2*-deficient mice (Figure 2A, Figure S2A), indicating additional intestinal targets of B4GALNT2 glycosylation. Other glycosylation targets for B4GALNT2 are Sd(a)/Cad antigens, which have been shown to be present in colonic mucins [34, 36], glycolipids and glycoproteins [32, 33, 35, 86]. Intestinal mucin glycans, including blood group α-1-2 fucosylated receptors, have been proposed as attachment sites for *Salmonella* [87, 88], but *Salmonella* does not appear to directly bind *B4galnt2*-GalNAc residues in vitro [4]. The glycan profile may also change in animals not expressing *B4galnt2* in addition to the lack of β1-4-GalNac residues/Sd(a), whereby the increase or decrease of other residues may offer new nutrient sources or attachment sites for bacteria or immune cells [35, 89]. Nevertheless, we found slightly increased invasion into epithelial cells in vivo and in vitro when *B4galnt2* is expressed. However, our fecal transfer experiments demonstrate that the altered bacterial community of *B6* $^{-/-}$ mice confers resistance towards *Salmonella* induced inflammation. Thus, it is likely that indirect mechanisms, such as the microbial community and its capability of glycan liberation, subsequent changes in nutrient or microbe abundances [90] and the type of interactions [72], are responsible for the higher susceptibility of mice expressing *B4galnt2* in the intestinal epithelium to *S.* Typhimurium infection.

Our study reveals an increased production of pro-inflammatory mediators, higher numbers of immune/inflammatory cells, and more severe colitis after *S.* Typhimurium infection in the ceca of mice expressing *B4galnt2* in the intestinal epithelium. Although endothelial *B4galnt2* expression did not impact the development of colitis as judged by histology, *RIII* $^{+}$ mice had lower pathogen burden in the cecum and lower levels of *Mcp-1* and LCN-2 compared to *RIII* $^{-}$ mice, supporting a role for vascular *B4galnt2* in host immune defense in the face of intestinal pathogens. Functionally, carbohydrate differentiation

antigens play an important role in the homing and differentiation of intraepithelial lymphocytes in the small intestine, indicating a plausible phenotype that may result from the expression of *B4galnt2* in endothelial cells [91-94]. The recruitment of neutrophils and CD3 $^{+}$ cells [35], as well as leukocyte infiltration, were reported to be influenced through the glycosylation of selectin receptors [95] and could be associated with the elimination of carbohydrate ligands for selectins. *B4galnt2* expression in gastrointestinal cancers has been shown to reduce metastatic dissemination, adding to the role of the Sd(a) antigen in cell motility [96, 97]. Further studies focusing on the role of endothelial *B4galnt2* expression are needed to understand the impact of *B4galnt2*-GalNAc residues in host immune responses and its potential role for homing of immune cells to the intestine.

In summary, we demonstrate that different patterns of tissue-specific *B4galnt2* expression not only influence intestinal microbial communities, but also change host susceptibility and immunological responses to *S.* Typhimurium infection [45, 98]. Thus, a complex scenario including *B4galnt2*-dependent changes in microbial communities, vascular immune phenotypes, bleeding tendencies and susceptibility to intestinal infections likely contributes to the maintenance of variation at *B4galnt2* in wild mouse populations.

Material & Methods

Animal models of variant *B4galnt2* tissue-specific expression: All genetically engineered mouse lines used in the study were backcrossed >20 generations to a C57BL/6J background prior to breeding of the experimental animals. C57BL/6J (*B6* $^{+/+}$) mice were purchased from The Jackson Laboratory. Mice heterozygous for the *B4galnt2* knock-out allele (*B6* $^{+/-}$) [23] and RIIIS/J-*B4galnt2* BAC transgenic (*RIII* $^{+}$) mice which exhibit the *Mvwf1* phenotype [21] were re-derived at the University Clinic Eppendorf, Hamburg, Germany. Intercross of *B6* $^{+/-}$ × *B6* $^{+/-}$*RIII* $^{+}$ generated heterozygous *B6* $^{+/-}$/*RIII* $^{+}$, *B6* $^{-/-}$/*RIII* $^{+}$, *B6* $^{+/-}$/*RIII* $^{-}$ and *B6* $^{-/-}$/*RIII* $^{-}$ offspring, which were raised and housed together as littermates under specific pathogen-free conditions in individually ventilated cages at the animal facility of the University of Kiel, Germany. Standard chow (ssniff, Soest, Germany) and water were provided *ad libitum*. Germ-free C57BL/6J mice were produced at the gnotobiotic facility of the Hannover Medical School. Experiments were conducted in the animal facility of the Leibniz Research Center Borstel, Germany and at the animal facility of University Hospital Schleswig-Holstein Kiel.

Ethics statement: All experiments were conducted consistent with the ethical requirements of the Animal Care Committee of the Ministry of Energy, Agriculture, the Environment and Rural Areas of Schleswig-Holstein, Germany and in direct accordance with the German Animal Protection Law. The protocols were approved by the Ministry of Energy, Agriculture,

the Environment and Rural Areas of Schleswig-Holstein, Germany (Protocol: V312-72241.123-3 and V312-7224.123-3).

***Salmonella* infection of mice:** Streptomycin (20mg per mouse) (Sigma-Aldrich, Hamburg, Germany) was given by oral gavage to mice aged 10-14 weeks. 24 hours after antibiotic administration, mice were infected with either *S.* Typhimurium SL1344 (acute infection; [28]) or the attenuated *S.* Typhimurium Δ*aroA* (chronic infection; [29]) at a dose of 3×10^6 bacteria in 100 µL HEPES buffer (100 mM, pH 8.0; PAA, Cölbe, Germany). Control mice (mock-infection) were given 100 µL HEPES buffer. Bacterial loads were determined by plating serial dilutions of homogenized organs on Luria Bertani agar (Roth, Karlsruhe, Germany) containing streptomycin (100 µg/mL).

siRNA knockdown and tissue culture infections: Mouse intestinal epithelial Mode-K cells were grown in DMEM supplemented with 5% fetal bovine serum (Biochrom, Berlin, Germany) and 1% HEPES (GE Healthcare, Frankfurt, Germany). For the siRNA knockdown of *B4galnt2* 1×10^5 cells per well were seeded in a 24 well plate containing 10nM siRNA and lipofectamine (Life Technologies, Darmstadt, Germany) according to manufacturer's instructions for reverse transfection. As a negative control cells were treated with scrambled siRNA. 24h post transfection cells were infected with an MOI 50 of wildtype *S.* Typhimurium grown to late-logarithmic phase. 30 min p.i., cells were washed and extracellular bacteria were killed by addition of medium containing gentamicin (100 µg/ml). Cells were lysed at various timepoints (30 min, 1 h and 4 h) and the number of adherent and invaded bacteria was determined by plating serial dilutions.

Fluorescence in situ Hybridization (FISH) Staining: Cecal tissues were fixed in Carnoy's fixative overnight, embedded in paraffin, and then cut in 5 µm sections on glass slides. Sections were deparaffinized and incubated with a Texas red-conjugated EUB338 general bacterial probe (GCTGCCTCCCGTAGGAGT) and an AlexaFluor 488 conjugated Gam42a probe (GCCTTCCCACATCGTTT) that recognizes bacteria that belong to the γ-Proteobacteria class (37°C, O/N, dark). Tissue samples were washed with hybridization buffer (0.9 M NaCl, 0.1 M Tris pH 7.2, 0.1% SDS). This step was repeated with FISH Washing Buffer (0.9 M NaCl, 0.1 M Tris pH 7.2) with gentle shaking for 15 minutes. Sections were washed with water and mounted using Prolong GOLD with DAPI (Molecular Probes) and imaged using an AxioImager microscope equipped with an AxioCam HRm camera operating through AxioVision software. High power field (HPF) (630X) was used for enumerating intracellular and extracellular *S.* Typhimurium.

Staining of acidic mucus and mucus thickness: Carnoy's-fixed paraffin-embedded tissues were sectioned (5 µm), deparaffinized, and stained with 1% Alcian Blue (Sigma-

Aldrich, Hamburg, Germany) solution (in 1% acetic acid) for 10 min, counterstained in nuclear fast red solution (1%), dehydrated, and mounted for examination. Photographs were taken at an original magnification of 100× and mucus thickness was measured at six random locations per section using NIS-Element Software (Nikon, Dusseldorf, Germany).

Fecal transplantation experiments: Fresh feces from *B6* $^{+/-}$ or *B6* $^{-/-}$ mice was sampled and immediately homogenized (1:10 w/v) in transfer buffer (sterile phosphate buffered saline containing 0.05% cysteine HCl (Sigma-Aldrich)). After centrifugation, the supernatant was collected and 200 µL were orally gavaged into germ-free adult C57BL/6J recipient mice. 21 days post transplantation mice were treated with streptomycin and 24 hours later infected with *S.* Typhimurium.

Histopathological analysis: Tissues were fixed in 10% neutral buffered formalin overnight and embedded in paraffin. 5 µm sections were deparaffinized and stained with haematoxylin and eosine (H&E). Histological scores in the ceca of infected mice were determined as previously described [30]. Briefly, pathological changes were assessed by evaluating various parameters such as presence of luminal cells, infiltrating immune cells, crypt abscesses and the formation of edema in the respective layer of the intestinal bowel wall including the surface epithelium, mucosa and submucosa. details

Immunohistochemistry: Formalin fixed tissue sections (5 µm) were deparaffinized and rehydrated. After antigen retrieval with 10 mM sodium citrate buffer (pH 6.0) and blocking with 2% normal goat serum, specimens were incubated with antibodies specific for *S.* Typhimurium (Clone B395M, Dunn Laboratories, Asbach, Germany), CD3 (Abcam, Cambridge, UK), CD68 (Abcam, Cambridge, UK), myeloperoxidase (MPO) (Thermo Fisher Scientific, Schwerte, Germany), and MUC2 (Santa Cruz, Dallas, TX, USA) followed by fluorescently labeled secondary antibodies (Molecular Probes, Invitrogen, Carlsbad, CA, USA) or with fluorescently labelled DBA (dolichus biflorus agglutinin) and WGA (wheat germ agglutinin) lectins (Vector laboratories, Burlingame, CA, USA). Counterstaining of nuclei was performed using 4,6-Diamidin-2-phenylindol (DAPI) (Invitrogen, Carlsbad, CA, USA). Images were obtained using a Leica SP5 confocal microscope (Leica, Wetzlar, Germany).

Lipocalin-2 enzyme-linked immunosorbent assay (ELISA): Lipocalin-2 concentrations in the supernatant of tissue homogenates were determined with a mouse specific ELISA Development Kit by R&D Systems (R&D Systems, Wiesbaden, Germany) according to the manufacturer's instructions.

Real-time quantitative polymerase chain reaction (RT-qPCR): RNA was extracted from cecal tips by using the High Pure RNA Tissue Kit (Roche Diagnostics, Mannheim, Germany) and reverse transcription was conducted with the Transcriptor High Fidelity cDNA Synthesis

Kit (Roche Diagnostics, Mannheim, Germany) according to the manufacturer's instructions. RT-qPCR was performed with Quantitect SYBR-Green Mastermix (QIAGEN, Hilden, Germany) for the following genes: *Ifn-γ*, fw TCAAGTGGCATAGATGTGGAAGAA, rev TGGCTCTGCAGGATTTTCATG; *Tnf-α*,fw CCACCACGCTCTTCTGTCTAC, rev AGGGTCTGGGCCATAGAACT; *Il-6*, fw GAGGATACCACTCCCAACAGACC, rev AAGTGCATCATCGTTGTTCATACA; *Mcp-1*, fw CCTGCTGTTCACAGTTGCC, rev ATTGGGATCATCTTGCTGGT; *B4galnt2*, fw TGGCAAGTCCTACCATGAGG, rev GTCTGCAGAAGTGGCTGGA; *Gapdh*, fw ATTGTCAGCAATGCATCCTG, rev ATGGACTGTGGTCATGAGCC; *Hprt*, fw AGTGTTGGATACAGGCCAGAC, rev CGTGATTCAAATCCCTGAAGT. Relative gene expression was calculated using geNORM and the 2-ΔΔCt method, with *Gapdh* and *Hprt* as housekeeping genes [99].

DNA extraction and 16S rRNA gene sequencing: DNA was extracted from fecal samples (stored at -80°C) using the PowerSoil® DNA Isolation Kit (MO Bio Laboratories, Carlsbad, CA) following the manufacturer's protocol. The 16S rRNA gene was amplified using barcoded primers flanking the V1 and V2 hypervariable regions (27F-338R) and were sequenced following the methods describe in Rausch *et al.* 2011 [100].

Sequence processing and quality control: Raw sequences were trimmed by mothur 1.31.2 requiring no ambiguous bases, a mean quality score within a window of 50 base pairs of ≥35 and a minimum length of 200 nucleotides for the coupled V1-V2 region [101]. Chimeric sequences were determined using USEARCH 4.25 (database informed UCHIME algorithm) [102]. Sequences were confirmed as bacterial using the RDP classifier with ≥ 60% bootstrap threshold [103]. For all downstream analyses of diversity and habitat association, we took a random subset of 1000 sequences per sample to normalize the read distribution (Good's Coverage; no treatment: 85.67 ± 6.61% SD; Streptomycin: 97.38 ± 3.13% SD; *S*. Typhimurium: 98.36 ± 1.74% SD). These sequences were aligned to the curated SILVA seed database using the NAST alignment procedure as implemented in mothur and subsequently OTU binning was carried out via average distance clustering [104]. Phylogenetic tree construction on representative OTU sequences (average distant sequence of the OTU) was done by FastTree 2.1 using the CAT substitution model with gamma correction [105]. Raw sequence data can be accessed online under the accession number PRJEB5269 at the European Nucleotide Archive.

Statistical analysis: Species diversity indices (Chao1 species richness, Shannon-Weaver index), as well as the phylogenetic distance at the tips of the phylogenetic tree (Nearest Taxon Index, NTI) and its deep branches (Net Relatedness Index ,NRI) were calculated in R [106-108]. The phylogenetic measures of beta diversity (unweighted- and weighted UniFrac) and metrics based on shared OTU presence (Jaccard) or abundance (Bray-Curtis) were

calculated in “vegan” [109-111]. Statistical analysis of community composition based on different beta diversity metrics was performed with Principal Coordinate Analysis (PCoA) and non-parametric multivariate analysis of variance and multivariate dispersion as implemented in the “vegan” package for R with 105 permutations. For constrained ordination (Redundancy Analysis) the OTU table was Hellinger-transformed and RDA was carried out following Legendre and Legendre [112]. Significance of factors and axes was ascertained using a permutative ANOVA approach (5000 permutations). Linear mixed models (LMM, cage as random factor) were applied to alpha diversity measures and optimized with model selection by AIC criterion, normality of model residuals and refitting of the final model under Restricted Maximum Likelihood (REML) [113]. The R^2 values of the final mixed model were calculated using the MuMIN package for R [114, 115]. Lipocalin-2 levels, fluorescence signals, inflammation scores, CFU counts, and cecum weights were analyzed in a Linear model framework with parameter selection to minimize the AIC value and no significant reduction of fit. For the comparison of expression values among genotypes we employed a Wilcoxon test with Monte-Carlo resampling [116]. *Salmonella* counts (Gam24a$^+$ cells) in Mode-K cell cultures were analyzed using an LMM with the independent rounds of experiments as random factor to incorporate experimental variation. Indicator species analysis was based on 10^5 permutations using the indicator value to assess the association for each taxon [117]. All *P*-values of the genera and OTU associations were adjusted by the Benjamini-Hochberg procedure. Taxon co-occurence networks were calculated by SPARCC based on 105 permutations and significant associations ($P < 0.05$) were included in the network construction [118].

Acknowledgements

We thank Yeojun Yun and E. Susanne Quabius for experimental support; Katja Cloppenborg-Schmidt and Janin Braun for excellent technical assistance.

References

1. Koropatkin NM, Cameron EA, Martens EC. How glycan metabolism shapes the human gut microbiota. Nat Rev Micro. 2012;10(5):323-35. Epub 2012/04/12. doi: Doi 10.1038/Nrmicro2746. PubMed PMID: ISI:000302938700010; PubMed Central PMCID: PMC4005082.

2. Sonnenburg JL, Xu J, Leip DD, Chen CH, Westover BP, Weatherford J, et al. Glycan foraging in vivo by an intestine-adapted bacterial symbiont. Science. 2005;307(5717):1955-9.

3. Varki A, Freeze HH, Gagneux P. Evolution of Glycan Diversity. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al., editors. Essentials of Glycobiology. 2010/03/20 ed. Cold Spring Harbor NY: The Consortium of Glycobiology Editors, La Jolla, California; 2009.

4. Giannasca KT, Giannasca PJ, Neutra MR. Adherence of *Salmonella* typhimurium to Caco-2 cells: identification of a glycoconjugate receptor. Infection and immunity. 1996;64(1):135-45. Epub 1996/01/01. PubMed PMID: 8557331; PubMed Central PMCID: PMC173738.

5. Kobayashi M, Lee H, Nakayama J, Fukuda M. Roles of gastric mucin-type O-glycans in the pathogenesis of Helicobacter pylori infection. Glycobiology. 2009;19(5):453-61. Epub 2009/01/20. doi: 10.1093/glycob/cwp004. PubMed PMID: WOS:000265096000001; PubMed Central PMCID: PMC2667159.

6. Henry SM. Molecular diversity in the biosynthesis of GI tract glycoconjugates. A blood-group-related chart of microorganism receptors. Transfusion clinique et biologique : journal de la Societe francaise de transfusion sanguine. 2001;8(3):226-30. Epub 2001/08/14. PubMed PMID: 11499965.

7. Moran AP, Gupta A, Joshi L. Sweet-talk: role of host glycosylation in bacterial pathogenesis of the gastrointestinal tract. Gut. 2011;60(10):1412-25. Epub 2011/01/14. doi: 10.1136/gut.2010.212704. PubMed PMID: 21228430.

8. Fumagalli M, Cagliani R, Pozzoli U, Riva S, Comi GP, Menozzi G, et al. Widespread balancing selection and pathogen-driven selection at blood group antigen genes. Genome Res. 2009;19(2):199-212. Epub 2008/11/11. doi: 10.1101/gr.082768.108. PubMed PMID: WOS:000263132600004; PubMed Central PMCID: PMC2652214.

9. Ferrer-Admetlla A, Sikora M, Laayouni H, Esteve A, Roubinet F, Blancher A, et al. A Natural History of FUT2 Polymorphism in Humans. Molecular Biology and Evolution. 2009;26(9):1993-2003. Epub 2009/06/03. doi: 10.1093/molbev/msp108. PubMed PMID: WOS:000269001500006.

10. McGovern DP, Jones MR, Taylor KD, Marciante K, Yan X, Dubinsky M, et al. Fucosyltransferase 2 (FUT2) non-secretor status is associated with Crohn's disease. Human molecular genetics. 2010;19(17):3468-76. Epub 2010/06/24. doi: 10.1093/hmg/ddq248. PubMed PMID: 20570966; PubMed Central PMCID: PMC2916706.

11. Folseraas T, Melum E, Rausch P, Juran BD, Ellinghaus E, Shiryaev A, et al. Extended analysis of a genome-wide association study in primary sclerosing cholangitis

detects multiple novel risk loci. Journal of hepatology. 2012;57(2):366-75. Epub 2012/04/24. doi: 10.1016/j.jhep.2012.03.031. PubMed PMID: 22521342; PubMed Central PMCID: PMC3399030.

12. Lindesmith L, Moe C, Marionneau S, Ruvoen N, Jiang X, Lindbland L, et al. Human susceptibility and resistance to Norwalk virus infection. Nature Medicine. 2003;9(5):548-53. Epub 2003/04/15. doi: 10.1038/nm860. PubMed PMID: WOS:000182610600035.

13. Ruiz-Palacios GM, Cervantes LE, Ramos P, Chavez-Munguia B, Newburg DS. Campylobacter jejuni Binds Intestinal H(O) Antigen (Fucα1, 2Galβ1, 4GlcNAc), and Fucosyloligosaccharides of Human Milk Inhibit Its Binding and Infection. Journal of Biological Chemistry. 2003;278(16):14112-20. Epub 2003/02/04. doi: 10.1074/jbc.M207744200. PubMed PMID: 12562767.

14. Magalhaes A, Gomes J, David L, Haas R, Boren T, Reis C. FUT2-Null Mice Show Impaired BabA-Mediated Adhesion of H. pylori to Gastric Mucosa. Helicobacter. 2009;14(4):371-. PubMed PMID: WOS:000268269300177.

15. Pham Tu Anh N, Clare S, Goulding D, Arasteh Julia M, Stares Mark D, Browne Hilary P, et al. Epithelial IL-22RA1-Mediated Fucosylation Promotes Intestinal Colonization Resistance to an Opportunistic Pathogen. Cell Host & Microbe. 2014;16(4):504-16. Epub 2014/09/30. doi: 10.1016/j.chom.2014.08.017. PubMed PMID: PMC4190086; PubMed Central PMCID: PMC4190086.

16. Goto Y, Obata T, Kunisawa J, Sato S, Ivanov II, Lamichhane A, et al. Innate lymphoid cells regulate intestinal epithelial cell glycosylation. Science. 2014;345(6202). doi: 10.1126/science.1254009.

17. Pickard JM, Maurice CF, Kinnebrew MA, Abt MC, Schenten D, Golovkina TV, et al. Rapid fucosylation of intestinal epithelium sustains host-commensal symbiosis in sickness. Nature. 2014;advance online publication. doi: 10.1038/nature13823

http://www.nature.com/nature/journal/vaop/ncurrent/abs/nature13823.html#supplementary-information.

18. Amorim I, Freitas DP, Magalhães A, Faria F, Lopes C, Faustino AM, et al. A comparison of Helicobacter pylori and non-Helicobacter pylori Helicobacter spp. Binding to Canine Gastric Mucosa with Defined Gastric Glycophenotype. Helicobacter. 2014;19(4):249-59. doi: 10.1111/hel.12125.

19. Lo Presti L, Cabuy E, Chiricolo M, Dall'Olio F. Molecular cloning of the human beta1,4 N-acetylgalactosaminyltransferase responsible for the biosynthesis of the Sd(a) histo-blood

group antigen: the sequence predicts a very long cytoplasmic domain. Journal of biochemistry. 2003;134(5):675-82. Epub 2003/12/23. PubMed PMID: 14688233.

20. Stuckenholz C, Lu L, Thakur P, Kaminski N, Bahary N. FACS-Assisted Microarray Profiling Implicates Novel Genes and Pathways in Zebrafish Gastrointestinal Tract Development. Gastroenterology. 2009;137(4):1321-32. Epub 2009/07/01. doi: http://dx.doi.org/10.1053/j.gastro.2009.06.050. PubMed PMID: 19563808; PubMed Central PMCID: PMC2785077.

21. Johnsen JM, Levy GG, Westrick RJ, Tucker PK, Ginsburg D. The endothelial-specific regulatory mutation, Mvwf1, is a common mouse founder allele. Mammalian Genome. 2008;19(1):32-40. doi: 10.1007/s00335-007-9079-4. PubMed PMID: WOS:000252483800005.

22. Mohlke KL, Nichols WC, Westrick RJ, Novak EK, Cooney KA, Swank RT, et al. A novel modifier gene for plasma von Willebrand factor level maps to distal mouse chromosome 11. Proceedings of the National Academy of Sciences. 1996;93(26):15352-7. Epub 1996/12/24. PubMed PMID: 8986815; PubMed Central PMCID: PMC26408.

23. Mohlke KL, Purkayastha AA, Westrick RJ, Smith PL, Petryniak B, Lowe JB, et al. Mvwf, a Dominant Modifier of Murine von Willebrand Factor, Results from Altered Lineage-Specific Expression of a Glycosyltransferase. Cell. 1999;96(1):111-20. Epub 1999/02/16. doi: http://dx.doi.org/10.1016/S0092-8674(00)80964-2. PubMed PMID: 9989502.

24. Johnsen JM, Teschke M, Pavlidis P, McGee BM, Tautz D, Ginsburg D, et al. Selection on cis-Regulatory Variation at *B4galnt2* and Its Influence on von Willebrand Factor in House Mice. Molecular Biology and Evolution. 2009;26(3):567-78. Epub 2008/12/18. doi: 10.1093/molbev/msn284. PubMed PMID: ISI:000263420900009; PubMed Central PMCID: PMC2727395.

25. Linnenbrink M, Johnsen JM, Montero I, Brzezinski CR, Harr B, Baines JF. Long-term balancing selection at the blood group-related gene *B4galnt2* in the genus Mus (Rodentia; Muridae). Molecular Biology and Evolution. 2011;28(11):2999-3003. Epub 2011/06/10. doi: 10.1093/molbev/msr150. PubMed PMID: 21652612.

26. Staubach F, Künzel S, Baines AC, Yee A, McGee BM, Bäckhed F, et al. Expression of the blood-group-related glycosyltransferase *B4galnt2* influences the intestinal microbiota in mice. ISME J. 2012;6(7):1345-55. Epub 2012/01/27. doi: http://www.nature.com/ismej/journal/vaop/ncurrent/suppinfo/ismej2011204s1.html. PubMed PMID: 22278669; PubMed Central PMCID: PMC3379640.

27. Barthel M, Hapfelmeier S, Quintanilla-Martínez L, Kremer M, Rohde M, Hogardt M, et al. Pretreatment of Mice with Streptomycin Provides a *Salmonella* enterica Serovar Typhimurium Colitis Model That Allows Analysis of Both Pathogen and Host. Infection and Immunity. 2003;71(5):2839-58. doi: 10.1128/iai.71.5.2839-2858.2003. PubMed PMID: ISI:000182501500061.

28. Hoiseth SK, Stocker BA. Genes aroA and serC of *Salmonella* typhimurium constitute an operon. Journal of Bacteriology. 1985;163(1):355-61. Epub 1985/07/01. PubMed PMID: 2989248; PubMed Central PMCID: PMC219121.

29. Grassl GA, Valdez Y, Bergstrom KSB, Vallance BA, Finlay BB. Chronic Enteric *Salmonella* Infection in Mice Leads to Severe and Persistent Intestinal Fibrosis. Gastroenterology. 2008;134(3):768-80.e2. Epub 2008/03/08. doi: http://dx.doi.org/10.1053/j.gastro.2007.12.043. PubMed PMID: 18325390.

30. Coburn B, Li Y, Owen D, Vallance BA, Finlay BB. *Salmonella* enterica serovar Typhimurium pathogenicity island 2 is necessary for complete virulence in a mouse model of infectious enterocolitis. Infection and immunity. 2005;73(6):3219-27. Epub 2005/05/24. doi: 10.1128/iai.73.6.3219-3227.2005. PubMed PMID: 15908346; PubMed Central PMCID: PMC1111876.

31. Wei X, Yang Z, Rey Federico E, Ridaura Vanessa K, Davidson Nicholas O, Gordon Jeffrey I, et al. Fatty Acid Synthase Modulates Intestinal Barrier Function through Palmitoylation of Mucin 2. Cell Host & Microbe. 2012;11(2):140-52. Epub 2012/02/22. doi: http://dx.doi.org/10.1016/j.chom.2011.12.006. PubMed PMID: 22341463; PubMed Central PMCID: PMC3285413.

32. Blanchard D, Piller F, Gillard B, Marcus D, Cartron JP. Identification of a novel ganglioside on erythrocytes with blood group Cad specificty. Journal of Biological Chemistry. 1985;260(13):7813-6.

33. Piller F, Blanchard D, Huet M, Cartron JP. Identification of a α-NeuAc-(2→3)-β-d-galactopyranosyl N-acetyl-β-d-galactosaminyltransferase in human kidney. Carbohydrate Research. 1986;149(1):171-84.

34. Dohi T, Yuyama Y, Natori Y, Smith PL, Lowe JB, Oshima M. Detection of N-acetylgalactosaminyltransferase mRNA which determines expression of Sda blood group carbohydrate structure in human gastrointestinal mucosa and cancer. International Journal of Cancer. 1996;67(5):626-31. doi: 10.1002/(sici)1097-0215(19960904)67:5<626::aid-ijc6>3.0.co;2-w.

35. Dall'Olio F, Malagolini N, Chiricolo M, Trinchera M, Harduin-Lepers A. The expanding roles of the Sda/Cad carbohydrate antigen and its cognate glycosyltransferase *B4GALNT2*. Biochimica et Biophysica Acta (BBA) - General Subjects. 2014;1840(1):443-53. Epub 2013/10/12. doi: http://dx.doi.org/10.1016/j.bbagen.2013.09.036. PubMed PMID: 24112972.

36. Capon C, Maes E, Michalski JC, Leffler H, Kim YS. Sd(a)-antigen-like structures carried on core 3 are prominent features of glycans from the mucin of normal human descending colon. Biochem J. 2001;358(3):657-64. Epub 2001/10/02. PubMed PMID: 11577689; PubMed Central PMCID: PMC1222115.

37. Godinez I, Haneda T, Raffatellu M, George MD, Paixao TA, Rolan HG, et al. T cells help to amplify inflammatory responses induced by *Salmonella* enterica serotype Typhimurium in the intestinal mucosa. Infect Immun. 2008;76(5):2008-17. Epub 2008/03/19. doi: IAI.01691-07 [pii]

10.1128/IAI.01691-07. PubMed PMID: 18347048; PubMed Central PMCID: PMC2346712.

38. de Jong HK, Parry CM, van der Poll T, Wiersinga WJ. Host-pathogen interaction in invasive Salmonellosis. PLoS Pathog. 2012;8(10):e1002933. Epub 2012/10/12. doi: 10.1371/journal.ppat.1002933

PPATHOGENS-D-12-00748 [pii]. PubMed PMID: 23055923; PubMed Central PMCID: PMC3464234.

39. Chassaing B, Srinivasan G, Delgado MA, Young AN, Gewirtz AT, Vijay-Kumar M. Fecal Lipocalin 2, a Sensitive and Broadly Dynamic Non-Invasive Biomarker for Intestinal Inflammation. PLoS One. 2012;7(9):e44328. Epub 2012/09/08. doi: 10.1371/journal.pone.0044328. PubMed PMID: 22957064; PubMed Central PMCID: PMC3434182.

40. Raffatellu M, George MD, Akiyama Y, Hornsby MJ, Nuccio SP, Paixao TA, et al. Lipocalin-2 resistance confers an advantage to *Salmonella* enterica serotype Typhimurium for growth and survival in the inflamed intestine. Cell Host Microbe. 2009;5(5):476-86. Epub 2009/05/21. doi: S1931-3128(09)00108-5 [pii]

10.1016/j.chom.2009.03.011. PubMed PMID: 19454351; PubMed Central PMCID: PMC2768556.

41. Webb CO, Ackerly DD, McPeek MA, Donoghue MJ. Phylogenies and community ecology. Annu Rev Ecol Syst. 2002;33(1):475-505. doi: doi:10.1146/annurev.ecolsys.33.010802.150448. PubMed PMID: ISI:000180007000018.

42. May RM. Will a Large Complex System be Stable? Nature. 1972;238(5364):413-4.

43. Fumagalli M, Sironi M, Pozzoli U, Ferrer-Admettla A, Pattini L, Nielsen R. Signatures of Environmental Genetic Adaptation Pinpoint Pathogens as the Main Selective Pressure through Human Evolution. PLoS Genet. 2011;7(11):e1002355. Epub 2011/11/11. doi: 10.1371/journal.pgen.1002355. PubMed PMID: 22072984; PubMed Central PMCID: PMC3207877.

44. Apanius V, Penn D, Slev PR, Ruff LR, Potts WK. The nature of selection on the major histocompatibility complex. Crit Rev Immunol. 1997;17(2):179-224. Epub 1997/01/01. PubMed PMID: ISI:A1997WR04600004.

45. Gagneux P, Varki A. Evolutionary considerations in relating oligosaccharide diversity to biological function. Glycobiology. 1999;9(8):747-55. Epub 1999/07/16. PubMed PMID: 10406840.

46. Varki A. Nothing in Glycobiology Makes Sense, except in the Light of Evolution. Cell. 2006;126(5):841-5. Epub 2006/09/09. doi: 10.1016/j.cell.2006.08.022. PubMed PMID: 16959563.

47. Andres AM, Hubisz MJ, Indap A, Torgerson DG, Degenhardt JD, Boyko AR, et al. Targets of balancing selection in the human genome. Mol Biol Evol. 2009;26(12):2755-64. Epub 2009/08/29. doi: 10.1093/molbev/msp190. PubMed PMID: MEDLINE:19713326; PubMed Central PMCID: PMC2782326.

48. Stecher B, Robbiani R, Walker AW, Westendorf AM, Barthel M, Kremer M, et al. *Salmonella* enterica Serovar Typhimurium Exploits Inflammation to Compete with the Intestinal Microbiota. PLoS Biol. 2007;5(10):e244. Epub 2007/09/01. doi: 10.1371/journal.pbio.0050244. PubMed PMID: 17760501; PubMed Central PMCID: PMC1951780.

49. Endt K, Stecher B, Chaffron S, Slack E, Tchitchek N, Benecke A, et al. The Microbiota Mediates Pathogen Clearance from the Gut Lumen after Non-Typhoidal *Salmonella* Diarrhea. PLoS Pathog. 2010;6(9):e1001097. doi: 10.1371/journal.ppat.1001097.

50. Levins R. COMPLEX SYSTEMS1970. 73-88 p.

51. Gardner MR, Ashby WR. CONNECTANCE OF LARGE DYNAMIC (CYBERNETIC) SYSTEMS - CRITICAL VALUES FOR STABILITY. Nature. 1970;228(5273):784-&. doi: 10.1038/228784a0. PubMed PMID: WOS:A1970H798400060.

52. Macarthur R. FLUCTUATIONS OF ANIMAL POPULATIONS, AND A MEASURE OF COMMUNITY STABILITY. Ecology. 1955;36(3):533-6. doi: 10.2307/1929601. PubMed PMID: WOS:A1955WR61800032.

53. Odum EP. Fundamentals of ecology1953. v+384p. Illus.-v+p. Illus. p.

54. Loreau M. Linking biodiversity and ecosystems: towards a unifying ecological theory2010 2010-01-12 00:00:00. 49-60 p.

55. Holling CS. Resilience and Stability of Ecological Systems. Annu Rev Ecol Syst. 1973;4(ArticleType: research-article / Full publication date: 1973 / Copyright © 1973 Annual Reviews):1-23. doi: 10.2307/2096802.

56. Ives AR, Carpenter SR. Stability and Diversity of Ecosystems. Science. 2007;317(5834):58-62. doi: 10.1126/science.1133258.

57. Tilman D. Biodiversity: Population Versus Ecosystem Stability. Ecology. 1996;77(2):350-63. doi: 10.2307/2265614.

58. Pfisterer AB, Schmid B. Diversity-dependent production can decrease the stability of ecosystem functioning. Nature. 2002;416(6876):84-6.

59. Isbell F, Calcagno V, Hector A, Connolly J, Harpole WS, Reich PB, et al. High plant diversity is needed to maintain ecosystem services. Nature. 2011;advance online publication. doi: http://www.nature.com/nature/journal/vaop/ncurrent/abs/nature10282.html#supplementary-information.

60. Doak DF, Bigger D, Harding EK, Marvier MA, O'Malley RE, Thomson D. The statistical inevitability of stability-diversity relationships in community ecology. American Naturalist. 1998;151(3):264-76. doi: 10.1086/286117. PubMed PMID: WOS:000072128600006.

61. Yachi S, Loreau M. Biodiversity and ecosystem productivity in a fluctuating environment: The insurance hypothesis. Proceedings of the National Academy of Sciences of the United States of America. 1999;96(4):1463-8. doi: 10.1073/pnas.96.4.1463. PubMed PMID: WOS:000078698400054.

62. Hubbell SP. The unified neutral theory of biodiversity and biogeography. Monographs in Population Biology. 2001;32:i-xiv, 1-375. PubMed PMID: ZOOREC:ZOOR13700052777.

63. Sankaran M, McNaughton SJ. Determinants of biodiversity regulate compositional stability of communities. Nature. 1999;401(6754):691-3.

64. McNaughton SJ. STABILITY AND DIVERSITY OF ECOLOGICAL COMMUNITIES. Nature. 1978;274(5668):251-3. doi: 10.1038/274251a0. PubMed PMID: WOS:A1978FG65000039.

65. McCann K, Hastings A, Huxel GR. Weak trophic interactions and the balance of nature. Nature. 1998;395(6704):794-8.

66. Cadotte MW, Cardinale BJ, Oakley TH. Evolutionary history and the effect of biodiversity on plant productivity. Proceedings of the National Academy of Sciences. 2008;105(44):17012-7. doi: 10.1073/pnas.0805962105.

67. Cadotte MW, Dinnage R, Tilman D. Phylogenetic diversity promotes ecosystem stability. Ecology. 2012;93(sp8):S223-S33. doi: 10.1890/11-0426.1. PubMed PMID: ISI:000307302400018.

68. Petchey OL, Casey T, Jiang L, McPhearson PT, Price J. Species richness, environmental fluctuations, and temporal change in total community biomass. Oikos. 2002;99(2):231-40. doi: 10.1034/j.1600-0706.2002.990203.x. PubMed PMID: WOS:000179715200003.

69. Zhang Q-G, Zhang D-Y. Species richness destabilizes ecosystem functioning in experimental aquatic microcosms. Oikos. 2006;112(1):218-26. doi: 10.1111/j.0030-1299.2006.14220.x.

70. Bohnhoff M, Drake BL, Miller CP. Effect of Streptomycin on Susceptibility of Intestinal Tract to Experimental *Salmonella* Infection. Experimental Biology and Medicine. 1954;86(1):132-7. Epub 1954/05/01. doi: 10.3181/00379727-86-21030. PubMed PMID: 13177610.

71. Srivastava DS, Cadotte MW, MacDonald AAM, Marushia RG, Mirotchnick N. Phylogenetic diversity and the functioning of ecosystems. Ecology Letters. 2012;15(7):637-48. Epub 2012/05/16. doi: 10.1111/j.1461-0248.2012.01795.x. PubMed PMID: 22583836.

72. Jones EI, Nuismer SL, Gomulkiewicz R. Revisiting Darwin's conundrum reveals a twist on the relationship between phylogenetic distance and invasibility. Proceedings of the National Academy of Sciences. 2013;110(51):20627-32. doi: 10.1073/pnas.1310247110.

73. Dethlefsen L, Huse S, Sogin ML, Relman DA. The Pervasive Effects of an Antibiotic on the Human Gut Microbiota, as Revealed by Deep 16S rRNA Sequencing. Plos Biology. 2008;6(11):e280. Epub 2008/11/21. doi: :10.1371/journal.pbio.0060280. PubMed PMID: BIOSIS:PREV200900045692; PubMed Central PMCID: PMC2586385.

74. Stecher B, Chaffron S, Käppeli R, Hapfelmeier S, Freedrich S, Weber TC, et al. Like Will to Like: Abundances of Closely Related Species Can Predict Susceptibility to Intestinal Colonization by Pathogenic and Commensal Bacteria. PLoS Pathog. 2010;6(1):e1000711. Epub 2010/01/12. doi: 10.1371/journal.ppat.1000711. PubMed PMID: 20062525; PubMed Central PMCID: PMC2796170.

75. Sonnenburg JL, Xu J, Leip DD, Chen C-H, Westover BP, Weatherford J, et al. Glycan Foraging in Vivo by an Intestine-Adapted Bacterial Symbiont. Science. 2005;307(5717):1955-9. Epub 2005/03/26. doi: 10.1126/science.1109051. PubMed PMID: 15790854.

76. Holmén Larsson JM, Karlsson H, Sjövall H, Hansson GC. A complex, but uniform O-glycosylation of the human MUC2 mucin from colonic biopsies analyzed by nanoLC/MSn. Glycobiology. 2009;19(7):756-66. doi: 10.1093/glycob/cwp048.

77. Kaoutari AE, Armougom F, Gordon JI, Raoult D, Henrissat B. The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. Nat Rev Micro. 2013;11(7):497-504. Epub 2013/06/12. doi: 10.1038/nrmicro3050

http://www.nature.com/nrmicro/journal/v11/n7/abs/nrmicro3050.html#supplementary-information. PubMed PMID: 23748339.

78. Bosshard PP, Zbinden R, Altwegg M. Turicibacter sanguinis gen. nov., sp. nov., a novel anaerobic, Gram-positive bacterium. International Journal of Systematic and Evolutionary Microbiology. 2002;52(4):1263-6. Epub 2002/08/01. doi: 10.1099/ijs.0.02056-0. PubMed PMID: 12148638.

79. Dimitriu PA, Boyce G, Samarakoon A, Hartmann M, Johnson P, Mohn WW. Temporal stability of the mouse gut microbiota in relation to innate and adaptive immunity. Environmental Microbiology Reports. 2013;5(2):200-10. Epub 2013/04/16. doi: 10.1111/j.1758-2229.2012.00393.x. PubMed PMID: 23584963.

80. Cuív PÓ, Klaassens ES, Durkin AS, Harkins DM, Foster L, McCorrison J, et al. Draft Genome Sequence of Turicibacter sanguinis PC909, Isolated from Human Feces. Journal of Bacteriology. 2011;193(5):1288-9. Epub 2010/12/25. doi: 10.1128/jb.01328-10. PubMed PMID: 21183674; PubMed Central PMCID: PMC3067595.

81. Wang GH. Plant traits and soil chemical variables during a secondary vegetation succession in abandoned fields on the Loess Plateau. Acta Botanica Sinica. 2002;44(8):990-8. PubMed PMID: WOS:000177681900019.

82. Suter M, Edwards PJ. Convergent succession of plant communities is linked to species' functional traits. Perspectives in Plant Ecology, Evolution and Systematics. 2013;15(4):217-25. doi: http://dx.doi.org/10.1016/j.ppees.2013.05.001.

83. Lohbeck M, Poorter L, Martínez-Ramos M, Rodriguez-Velázquez J, van Breugel M, Bongers F. Changing drivers of species dominance during tropical forest succession. Functional Ecology. 2014;28(4):n/a-n/a. doi: 10.1111/1365-2435.12240. PubMed PMID: ISI:000340673900028.

84. Presley LL, Wei B, Braun J, Borneman J. Bacteria Associated with Immunoregulatory Cells in Mice. Applied and Environmental Microbiology. 2010;76(3):936-41. Epub 2009/12/17. doi: 10.1128/aem.01561-09. PubMed PMID: 20008175; PubMed Central PMCID: PMC2813032.

85. Weiss GA, Chassard C, Hennet T. Selective proliferation of intestinal Barnesiella under fucosyllactose supplementation in mice. Br J Nutr. 2014:1-9. Epub 2014/01/15. doi: S0007114513004200 [pii]

10.1017/S0007114513004200 [doi]. PubMed PMID: 24411010.

86. Drouilhet L, Mansanet C, Sarry J, Tabet K, Bardou P, Woloszyn F, et al. The Highly Prolific Phenotype of Lacaune Sheep Is Associated with an Ectopic Expression of the *B4GALNT2* Gene within the Ovary. PLoS Genet. 2013;9(9):e1003809. Epub 2013/10/03. doi: 10.1371/journal.pgen.1003809. PubMed PMID: 24086150; PubMed Central PMCID: PMC3784507.

87. Vimal DB, Khullar M, Gupta S, Ganguly NK. Intestinal mucins: the binding sites for *Salmonella* typhimurium. Molecular and cellular biochemistry. 2000;204(1-2):107-17. Epub 2000/03/16. PubMed PMID: 10718631.

88. Chessa D, Winter MG, Jakomin M, Baumler AJ. *Salmonella* enterica serotype Typhimurium Std fimbriae bind terminal alpha(1,2)fucose residues in the cecal mucosa. Molecular microbiology. 2009;71(4):864-75. Epub 2009/02/03. doi: 10.1111/j.1365-2958.2008.06566.x. PubMed PMID: 19183274.

89. Groux-Degroote S, Wavelet C, Krzewinski-Recchi M-A, Portier L, Mortuaire M, Mihalache A, et al. *B4GALNT2* gene expression controls the biosynthesis of Sda and sialyl Lewis X antigens in healthy and cancer human gastrointestinal tract. The International Journal of Biochemistry & Cell Biology. 2014;53(0):442-9. doi: http://dx.doi.org/10.1016/j.biocel.2014.06.009.

90. Ng KM, Ferreyra JA, Higginbottom SK, Lynch JB, Kashyap PC, Gopinath S, et al. Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. Nature. 2013;advance online publication(7469):96-9. Epub 2013/09/03. doi: 10.1038/nature12503

http://www.nature.com/nature/journal/vaop/ncurrent/abs/nature12503.html#supplementary-information. PubMed PMID: 23995682; PubMed Central PMCID: PMC3825626.

91. Lefrancois L. Carbohydrate differentiation antigens of murine T cells: expression on intestinal lymphocytes and intestinal epithelium. Journal of immunology (Baltimore, Md : 1950). 1987;138(10):3375-84. Epub 1987/05/15. PubMed PMID: 2437191.

92. Zarbock A, Ley K, McEver RP, Hidalgo A. Leukocyte ligands for endothelial selectins: specialized glycoconjugates that mediate rolling and signaling under flow. Blood. 2011;118(26):6743-51. Epub 2011/10/25. doi: 10.1182/blood-2011-07-343566. PubMed PMID: 22021370; PubMed Central PMCID: PMC3245201.

93. Kobayashi M, Fukuda M, Nakayama J. Role of Sulfated O-Glycans Expressed by High Endothelial Venule-Like Vessels in Pathogenesis of Chronic Inflammatory Gastrointestinal Diseases. Biological & Pharmaceutical Bulletin. 2009;32(5):774-9. Epub 2009/05/08. PubMed PMID: WOS:000266047300003; PubMed Central PMCID: PMC2718737.

94. Gauguet J-M, Rosen SD, Marth JD, von Andrian UH. Core 2 branching β1,6-N-acetylglucosaminyltransferase and high endothelial cell N-acetylglucosamine-6-sulfotransferase exert differential control over B- and T-lymphocyte homing to peripheral lymph nodes. Blood. 2004;104(13):4104-12. Epub 2004/08/21. doi: 10.1182/blood-2004-05-1986. PubMed PMID: 15319280.

95. Lowe JB. Glycan-dependent leukocyte adhesion and recruitment in inflammation. Current opinion in cell biology. 2003;15(5):531-8. Epub 2003/10/02. PubMed PMID: 14519387.

96. Kawamura YI, Kawashima R, Fukunaga R, Hirai K, Toyama-Sorimachi N, Tokuhara M, et al. Introduction of Sd(a) carbohydrate antigen in gastrointestinal cancer cells eliminates selectin ligands and inhibits metastasis. Cancer research. 2005;65(14):6220-7. Epub 2005/07/19. doi: 10.1158/0008-5472.can-05-0639. PubMed PMID: 16024623.

97. Kawamura YI, Adachi Y, Curiel DT, Kawashima R, Kannagi R, Nishimoto N, et al. Therapeutic adenoviral gene transfer of a glycosyltransferase for prevention of peritoneal

dissemination and metastasis of gastric cancer. Cancer Gene Ther. 2014;21(10):427-33. Epub 2014/09/13. doi: 10.1038/cgt.2014.46. PubMed PMID: 25213663.

98. Littman Dan R, Pamer Eric G. Role of the Commensal Microbiota in Normal and Pathogenic Host Immune Responses. Cell Host & Microbe. 2011;10(4):311-23. Epub 2011/10/25. doi: 10.1016/j.chom.2011.10.004. PubMed PMID: 22018232; PubMed Central PMCID: PMC3202012.

99. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biology. 2002;3(7):research0034.1 - research.11. Epub 2002/08/20. PubMed PMID: doi:10.1186/gb-2002-3-7-research0034; PubMed Central PMCID: PMC126239.

100. Rausch P, Rehman A, Künzel S, Häsler R, Ott SJ, Schreiber S, et al. Colonic mucosa-associated microbiota is influenced by an interaction of Crohn disease and FUT2 (Secretor) genotype. Proceedings of the National Academy of Sciences. 2011;108(47):19030-5. Epub 2011/11/10. doi: 10.1073/pnas.1106408108. PubMed PMID: 22068912; PubMed Central PMCID: PMC3223430.

101. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: Open Source, Platform-independent, Community-supported Software for Describing and Comparing Microbial Communities. Appl Environ Microbiol. 2009;75(23):7537-41. Epub October 2, 2009. doi: 10.1128/aem.01541-09. PubMed PMID: 19801464; PubMed Central PMCID: PMC2786419.

102. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26(19):2460-1. Epub 2010/08/17. doi: 10.1093/bioinformatics/btq461. PubMed PMID: 20709691.

103. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. Appl Environ Microbiol. 2007;73(16):5261-7. Epub 2007/06/26. doi: 10.1128/aem.00062-07. PubMed PMID: 17586664; PubMed Central PMCID: PMC1950982.

104. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, et al. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucl Acids Res. 2007;35(21):7188-96. doi: 10.1093/nar/gkm864.

105. Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLoS One. 2010;5(3):e9490. Epub 2010/03/13. doi:

10.1371/journal.pone.0009490. PubMed PMID: 20224823; PubMed Central PMCID: PMC2835736.

106. Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, et al. Picante: R tools for integrating phylogenies and ecology. Bioinformatics. 2010;26(11):1463-4. Epub 2010/04/17. doi: 10.1093/bioinformatics/btq166. PubMed PMID: 20395285.

107. Oksanen J, Blanchet FG, Kindt R, Legendre P, O'Hara RB, Simpson GL, et al. vegan: Community Ecology Package. 1.17-6 ed: http://CRAN.R-project.org; 2011.

108. Team RDC. R: A language and environment for statistical computing. R Foundation for Statistical Computing. 2012.

109. McArdle BH, Anderson MJ. Fitting multivariate models to community data: A comment on distance-based redundancy analysis. Ecology. 2001;82(1):290-7. doi: Doi 10.1890/0012-9658(2001)082[0290:Fmmtcd]2.0.Co;2. PubMed PMID: ISI:000166488200024.

110. Anderson MJ. Distance-Based Tests for Homogeneity of Multivariate Dispersions. Biometrics. 2006;62(1):245-53. Epub 2006/03/18. doi: 10.1111/j.1541-0420.2005.00440.x. PubMed PMID: 16542252.

111. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. Applied and Environmental Microbiology. 2005;71(12):8228-35. doi: 10.1128/aem.71.12.8228-8235.2005. PubMed PMID: WOS:000234417600073.

112. Legendre P, Legendre L. Numerical ecology. Second English edition. Developments in Environmental Modelling. 1998;20:i-xv, 1-853. PubMed PMID: ZOOREC:ZOOR13500057538.

113. Pinheiro J, Bates D, DebRoy S, Sarkar D, Team RDC. nlme: Linear and Nonlinear Mixed Effects Models. http://CRAN.R-project.org; 2011.

114. Bartoń K. MuMIn: multi-model inference, R package version 1.9.13. 2013.

115. Magee L. R2 Measures Based on Wald and Likelihood Ratio Joint Significance Tests. The American Statistician. 1990;44(3):250-3. doi: 10.2307/2685352.

116. Hothorn T, Hornik K, Van de Wiel MA, Zeileis A. A Lego system for conditional inference. American Statistician. 2006;60(3):257-63. doi: 10.1198/000313006x118430. PubMed PMID: WOS:000239411800006.

117. De Cáceres M, Legendre P, Moretti M. Improving indicator species analysis by combining groups of sites. Oikos. 2010;119(10):1674-84. doi: 10.1111/j.1600-0706.2010.18334.x.

118. Friedman J, Alm EJ. Inferring Correlation Networks from Genomic Survey Data. PLoS Comput Biol. 2012;8(9):e1002687. doi: 10.1371/journal.pcbi.1002687.

Figures

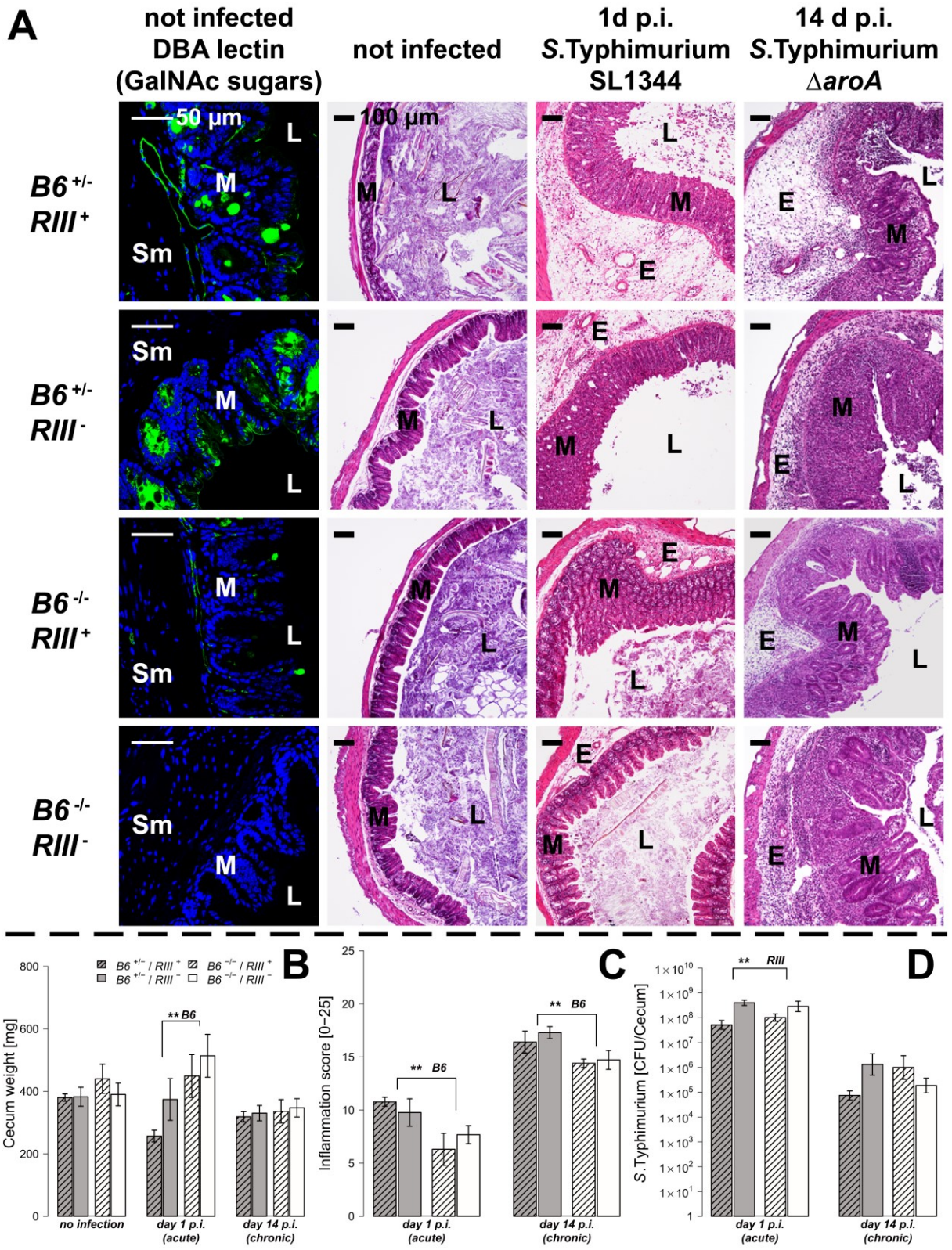


Figure 1: Tissue-specific expression of *B4galnt2* glycans influence susceptibility to *S.* Typhimurium-induced colitis. Mice were treated with streptomycin 24 h prior to infection

with *S.* Typhimurium strain SL1344 for 24 h (acute) or with the attenuated strain *S.* Typhimurium Δ*aroA* for 14 days (chronic). (A) *B4galnt2* expression phenotype is characterized by GalNAc residues, stained for by Dolichus biflorus agglutinin (DBA). H&E staining of cecal sections illustrated higher numbers of cells in the lumen (L), an increased influx of inflammatory cells to mucosa (M) and submucosa (Sm), epithelial cell desquamation and the formation of submucosal edema (E) upon infection with *S.* Typhimurium (bar=100 µm). (B) Cecal weight indicated a significant influence of intestinal *B4galnt2* glycans on *S.* Typhimurium induced colitis in the acute model (*B6*: $F_{1,49}$=8.709, *P*=0.0048; Linear model). (C) Histological scoring revealed higher inflammation in *B6* $^{+/-}$ compared to *B6* $^{-/-}$ mice (*B6*: $F_{1,49}$=13.242, *P*=0.0007; Linear model of X^4 transformed inflammation scores). (D) Intestinal *S*. Typhimurium colonization was determined in tissue homogenates (*RIII*: $F_{1,49}$=10.537, *P*=0.0021; Linear model of log(CFU)). Data are presented as mean ± SEM, N=9-19 per group in the acute model, N=5-7 in the chronic model (# *P*<0.100, * *P*<0.050, ** *P*<0.010, *** *P*<0.001).

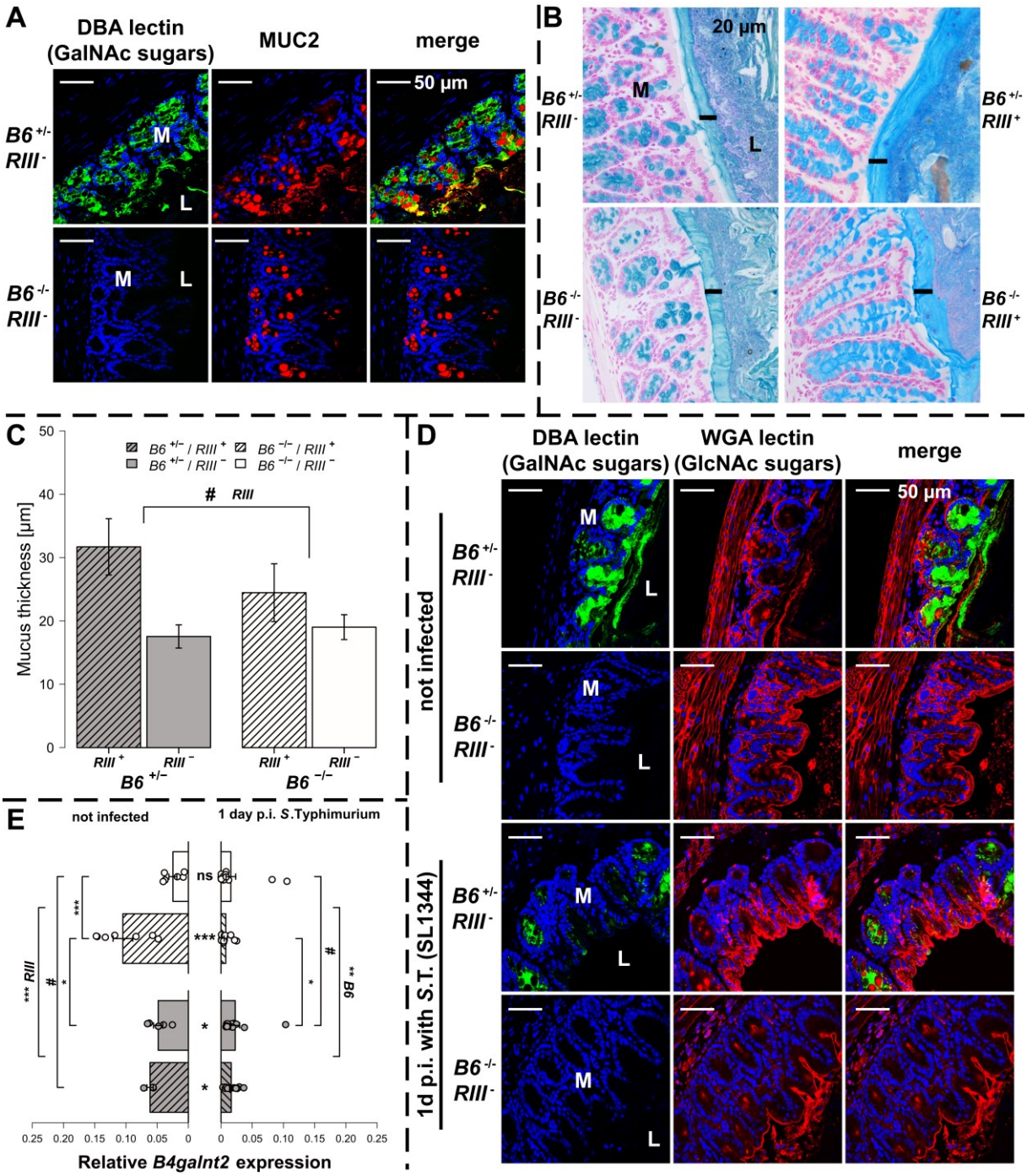


Figure 2: *B4galnt2* glycosylation in *S.* Typhimurium-induced colitis. (A) MUC2 (red) and DBA lectin (green) staining in formalin fixed cecal tissue sections (B) Acidic mucus was stained with alcian blue in Carnoy's-fixed tissue sections (bar=20 µm). (C) Mucus thickness was determined at five different regions within one animal from which mean values were analysed (N=3-5; *Z*=-1.807, *P*=0.0816; Wilcoxon test via Monte-Carlo resampling). (D) *B4galnt2* glycan residues (GalNAc) were stained with fluorescein labeled DBA (green) in formalin fixed cecal tissue sections before and 1 day p.i. with *S.* Typhimurium. GlcNAc residues were stained with Alexa633 labeled Wheat Germ Agglutinin (WGA) (red). (E) Relative expression of *B4galnt2* before and after infection with *S.* Typhimurium showing

significant differences between *B6* and *RIII* genotypes before infection (*B6*: $F_{1,17}$=0.216, *P*=0.64779; *RIII*: $F_{1,17}$=23.959, *P*=0.00014; *B6*/*RIII*: $F_{1,17}$=7.687, *P*=0.01304 [pairwise comparisons- $B6^{-/-}/RIII^{+}|B6^{-/-}/RIII^{-}$: *P*=0.00018, $B6^{+/-}/RIII^{+}|B6^{-/-}/RIII^{-}$: *P*=0.08626, $B6^{-/-}/RIII^{+}|B6^{+/-}/RIII^{-}$: *P*=0.02673]; Linear model and Tukey post-hoc test) and *B6* genotype differences after infection ($F_{1,53}$=11.787, *P*=0.001165). Infection has additional influence on *B4galnt2* expression (*Z*=5.268, *P*<0.00001, Wilcoxon test via Monte-Carlo resampling), which is also genotype specific ($B6^{+/-}/RIII^{+}$: *Z*=2.6458, $P_{Bonferroni}$=0.01192, $B6^{+/-}/RIII^{-}$: *Z*=2.6122, $P_{Bonferroni}$=0.02832; $B6^{-/-}/RIII^{+}$: *Z*=3.5496, $P_{Bonferroni}$=0.00016, $B6^{-/-}/RIII^{-}$: *Z*=2.0642, $P_{Bonferroni}$=0.16132; Wilcoxon test via Monte-Carlo resampling; # *P*<0.100, * *P*<0.050, ** *P*<0.010, *** *P*<0.001; error bars indicate SEM).

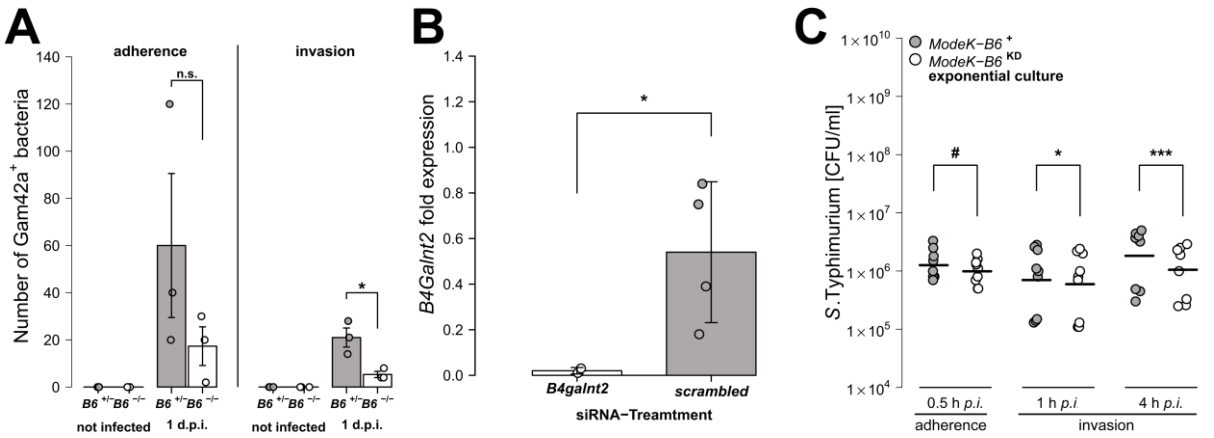


Figure 3: Epithelial *B4galnt2*-expression increases invasion by *S.* Typhimurium. (A) Carnoy's fixed cecal sections were stained by FISH (Gam42a probe) to visualize bacteria. Measurement of *Salmonella* adherence ($t_{2.286}$=-1.349, *P*=0.2954, unpaired t-test) and mucosa invasion ($t_{2.430}$=-3.681, *P*=0.0491; bacterial counts in 10 high power fields per individual, N=3; unpaired t-test). (B) *B4galnt2* expression in Mode-K cells after transfection with *B4galnt2* specific siRNA and scrambled siRNA relative to untreated cells ($t_{3.025}$=-3.3601, *P*=0.0432; unpaired t-test). (C) *Salmonella* invasion of Mode-K cell cultures transfected with *B4galnt2* specific and scrambled siRNA, infected with *S*. Typhimurium. There is no significant effect of *B4galnt2* expression for adhesion of the bacteria to Mode-K cells (0.5 h: $F_{1,14}$=3.133, *P*=0.0985; LMM with experiment as random factor; see "Statistical anaysis"), while invasion of *S.* Typhimurium into Mode-K cells expressing *B4galnt2* was slightly better than into cells with *B4galnt2* knockdown (1 h: $F_{1,14}$=7.644, *P*=0.0152, 4 h: $F_{1,14}$=26.336, *P*=0.0002; LMM with experiment as random factor; # *P*<0.100, * *P*<0.050, ** *P*<0.010, *** *P*<0.001; error bars indicate SEM).

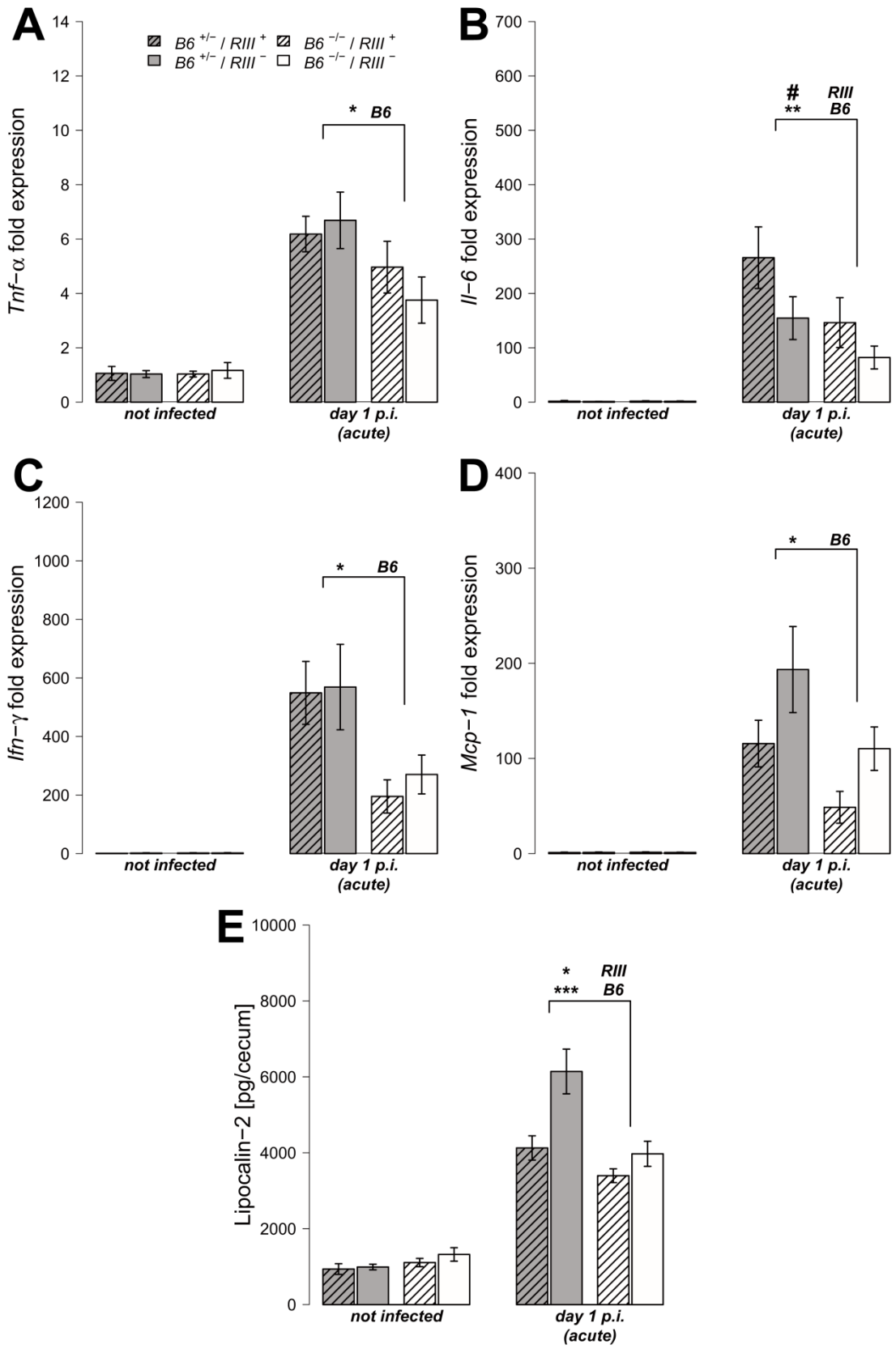


Figure 4: *B4galnt2*-dependent immune response after S. Typhimurium infection. (A-D) Relative gene expression of *Tnf-α*, *Il-6*, *Inf-γ* and *Mcp-1* was determined by RT-qPCR analysis. Values were normalized to *Gapdh* and *Hprt* and calculated as fold expression compared to the non-infected samples of each respective genotype. (E) Lipocalin-2 levels

were measured by ELISA in supernatants of cecal homogenates (N=3-11 per group) before- and one day p.i. with *S.* Typhimurium, showing a clear increase with infection (*Z*=-2.219, *P*=0.0261; Wilcoxon test via Monte-Carlo resampling) and differences between *B6* and *RIII* genotypes (Table S1, # *P*<0.100, * *P*<0.050, ** *P*<0.010, *** *P*<0.001; error bars indicate SEM).

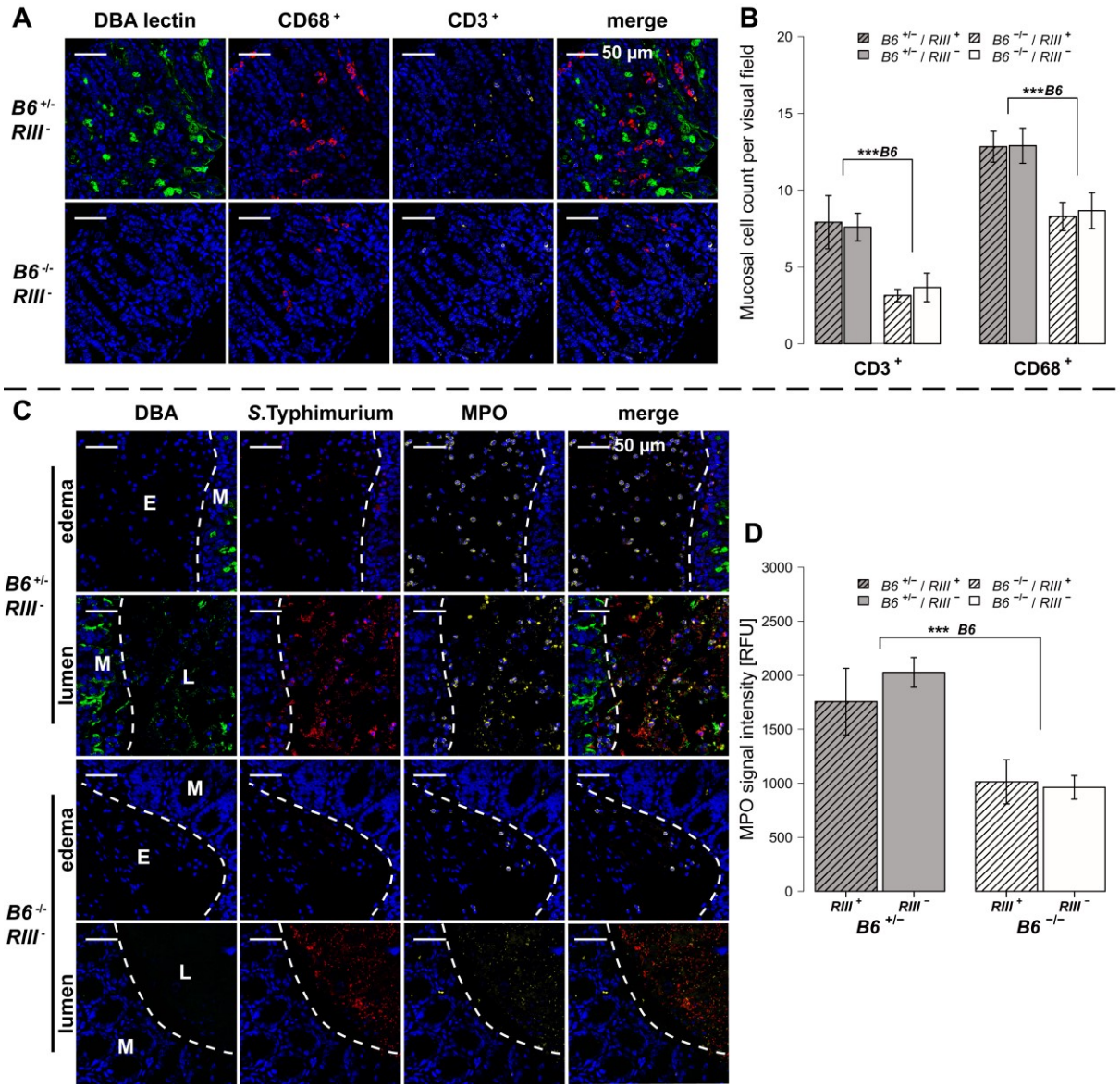


Figure 5: *B4galnt2*-dependent infiltration of immune cells after *S.* Typhimurium infection. (A and B) Immunofluorescence staining and enumeration of positive cells per vision field showed that *B6* $^{+/-}$ mice have higher numbers of CD68 (red) and CD3 (white) positive cells in the cecal mucosa 1 day p.i. (N=5-7). Nuclei were stained with DAPI (blue) and *B4galnt2* glycans by using fluorescein labeled DBA (green). (C) Myeloperoxidase (MPO)

positive cells (white) and *S.* Typhimurium (red) were determined by immunofluorescence staining. (D) MPO signal in lumen and edema was quantified and expressed as relative fluorescence units (RFU) (N=7; Linear model; # $P<0.100$, * $P<0.050$, ** $P<0.010$, *** $P<0.001$, error bars indicate SEM).

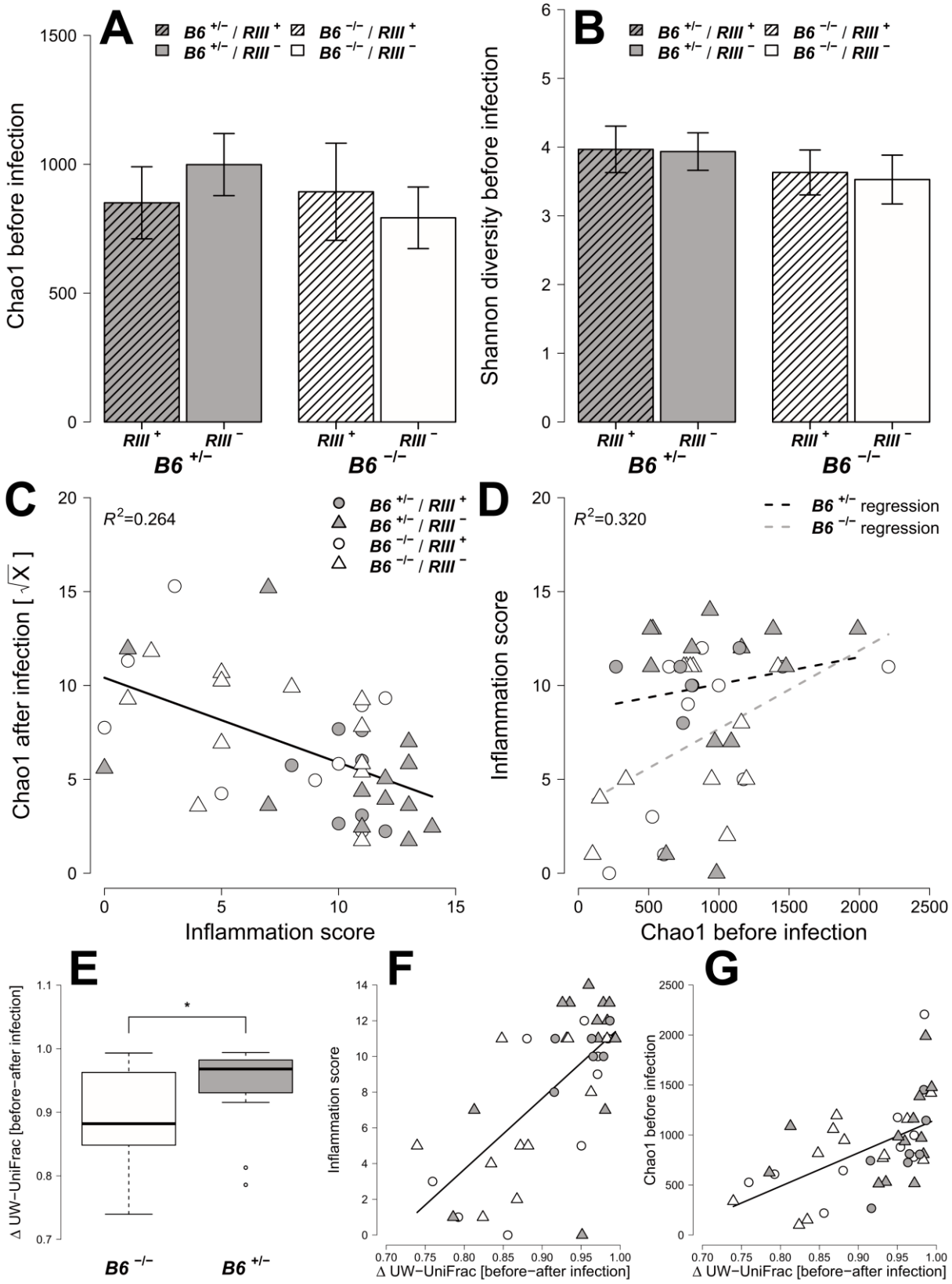


Figure 6: Analysis of microbial alpha diversity among genotypes and their association with intestinal inflammation. Microbial diversity was estimated from 97% species level OTUs and focused on the mean species richness (A; Chao1), and mean abundance based diversity (B; Shannon H), in the untreated animals. (C) The bacterial species richness is (i)

decreasing with increasing inflammation ($F_{1,22}$=14.2123, P=0.0011; LMM), (ii) but highly predictive of inflammation with differences among *B4galnt2* genotypes (D; Chao1 $F_{1,21}$=9.8274, P=0.005, *B6*: $F_{1,21}$=9.2976, P=0.0061, see also Table 2). The predictive power of alpha diversity for the outcome of infection is significantly improved by incorporating the *B6* genotype (Chao1: $R^2_{adjusted}$=0,320, ΔAIC=-5.936, LR=7.9360, $P_{LR\text{-}Test}$=0.0048; Shannon H: $R^2_{adjusted}$=0,271, ΔAIC=-6.1811, LR=8.1811, $P_{LR\text{-}Test}$=0.0042; NTI: $R^2_{adjusted}$=0.2625, ΔAIC=-8.8842, LR=10.8842, $P_{LR\text{-}Test}$=0.001). The turnover of bacterial communities (Δ unweighted UniFrac) over the course of the experiment is strongest in animals expressing *B4galnt2* in the epithelium (E; Z=-2.3213, P=0.01978; Wilcoxon test via Monte-Carlo resampling), and is highest in animals with strong inflammation (F; ρ=0.5894, P=0.00005; Spearman rank correlation). The community disturbance is also highest in animals with a high species richness before treatment (G; ρ=0.6040, P=0.000042; Spearman rank correlation; # P<0.100, * P<0.050, ** P<0.010, *** P<0.001, error bars indicate SEM; only results of best models are shown and pairwise tests are indicated).

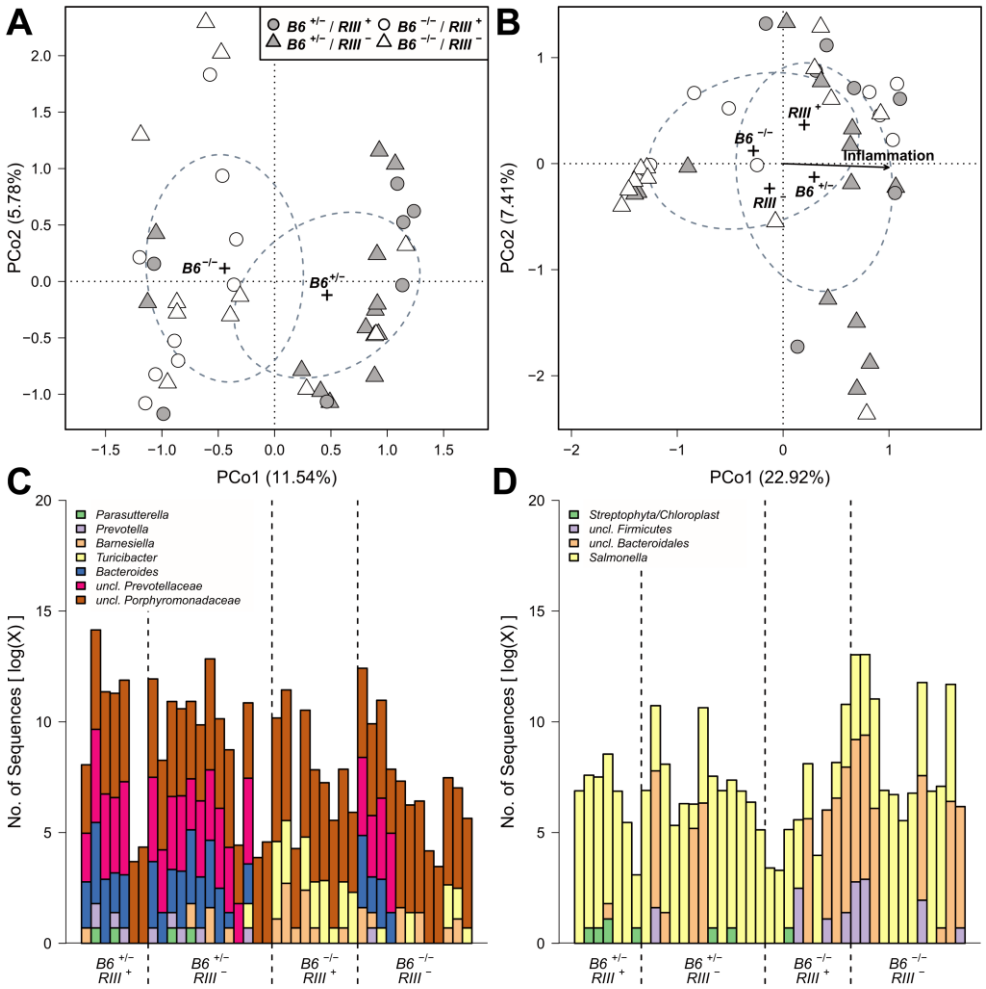


Figure 7: Treatment wise Principle Coordinate Analysis (unweighted UniFrac) of untreated- and *S.* Typhimurium inoculated mice and distribution of indicator bacteria among mice. The significant sample clusters and correlations are shown, displaying a strong influence of epithelial *B4galnt2* expression on the microbial community composition (no treatment (A): R^2=0.1480, P=0.0019; *Salmonella* treatment (B): *B6*: R^2=0.0607, P=0.08669, *RIII*: R^2=0.0781, P=0.040, inflammation: R^2=0.5531, P<0.0001). Abundance distribution of indicator genera before (C) and after *S*. Typhimurium infection (D) for *B4galnt2* gut expression among mice.

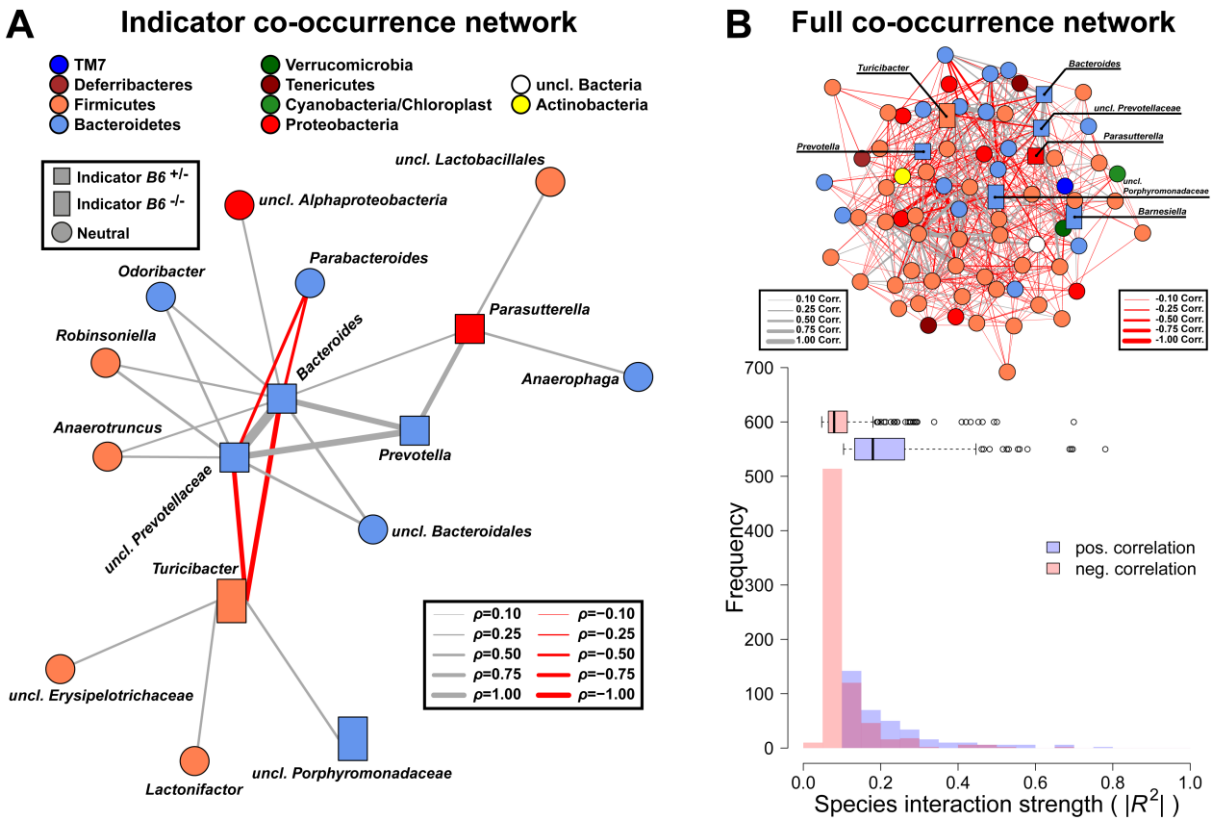


Figure 8: Targeted co-occurrence network analysis of indicator genera and overall network analysis. (A) Indicator genera for *B6* genotypes were correlated to abundances of the remaining community members to investigate proximate interactions among indicator genera and the surrounding community (interactions are Spearman correlations see Table S6; square - *B6* $^{+/-}$ indicator, rectangle - *B6* $^{-/-}$ indicator, circle - no indicator/neutral). (B) Microbial co-occurrence network based on genera abundances (only significant associations shown), with indicator species highlighted. Microbial communities show significant higher interaction strength among positive interactions (*i.e.* potential mutualistic; SPF: W=489396, $P < 2.20 \times 10^{-16}$; Wilcoxon test). However, the higher frequency of negative weak interactions overall has a stabilizing effect preventing the communities from collapsing (positive/negative interactions=0.482; # P<0.100, * P<0.050, ** P<0.010, *** P<0.001).

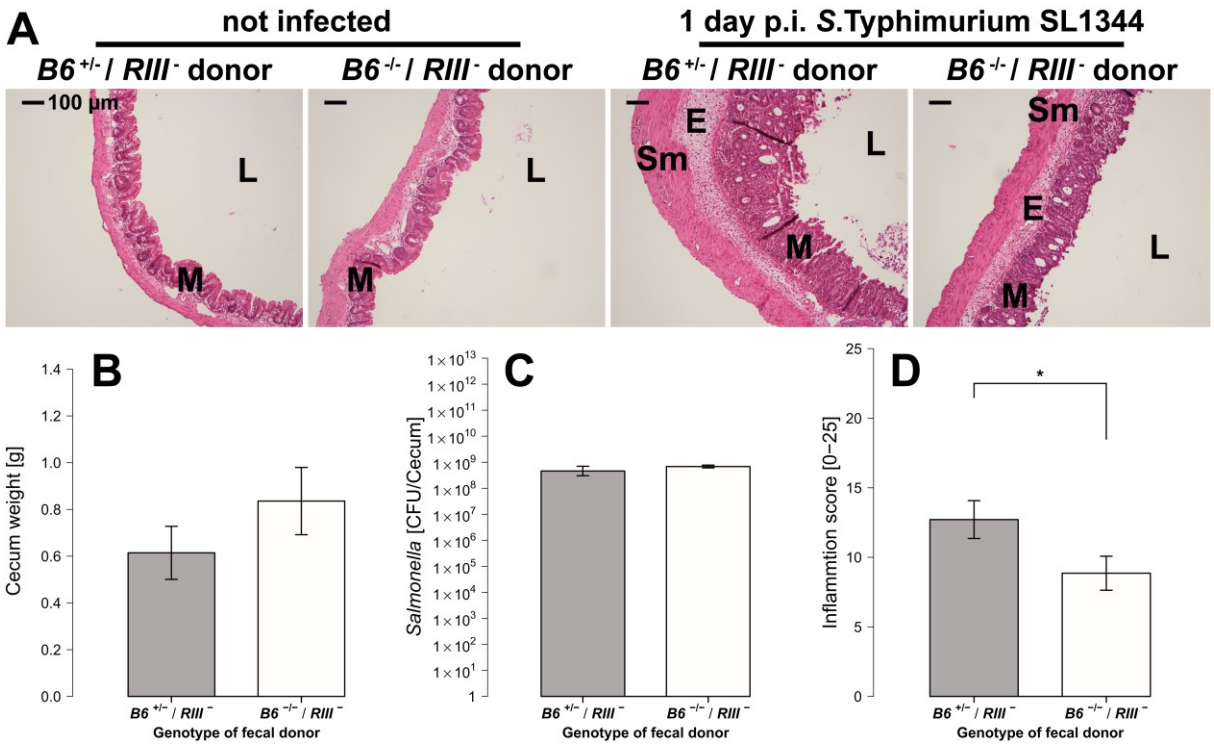


Figure 9: *B4galnt2*-dependent microbiota composition is responsible for enhanced susceptibility to inflammation. (A) Representative H&E staining of cecal sections with higher number of luminal cells (L), increased influx of inflammatory cell populations into the mucosa (M) and epithelial cell desquamation and submucosal edema (E) upon infection with *S.* Typhimurium (bar=100 µm). (B) Cecum weight (Z=1.087, P=0.3013, (C) and *Salmonella* abundance in the cecum (Z=0.447, P=0.7098) do not differ between donor genotypes (N=7 infected and N=3 uninfected controls per donor genotype). (D) Histological inflammation is significantly reduced in mice that received a *B6* $^{-/-}$ microbiome (Z=-2.074, P=0.0459; Wilcoxon test via Monte-Carlo resampling, # P<0.100, * P<0.050, ** P<0.010, *** P<0.001, error bars indicate SEM).

Tables

Table 1: Results of the alpha diversity analyses before and after infection with S. Typhimurium (best models after REML fitting).

Time point	Metric	Factor	DF	*F*-value	*P*-value	Adjusted R^2
before treatment	Shannon H (X^2)	Intercept	1,22	38.7456	<0.0001	0.0354
		RIII	1,22	1.6823	0.2081	
	Chao1	Intercept	1,23	95.9510	<0.0001	0.0737
		Gender	1,13	3.0484	0.1044	
	NRI	Intercept	1,23	50.9385	<0.0001	0.0239
		Gender	1,13	1.1028	0.3128	
	NTI	Intercept	1,22	365.5594	<0.0001	0.1234
		RIII	1,22	5.3731	**0.0301**	
1 d.p.i.	Shannon H	Intercept	1,22	126.3060	<0.0001	0.2538
		Inflammation	1,22	13.7716	**0.0012**	
	Chao1 ($X^{1/2}$)	Intercept	1,22	101.5123	<0.0001	0.2644
		Inflammation	1,22	14.2123	**0.0011**	
	NRI	Intercept	1,18	123.7569	<0.0001	0.2019
		RIII	1,18	1.9857	0.1758	
		poly(Inflammation)*	2,18	1.3985	0.2725	
		RIII:poly(Inflammation)	2,18	2.2966	0.1293	
	NTI	Intercept	1,22	100.8313	<0.0001	0.1184
		Inflammation	1,22	5.3981	**0.0298**	

* quadratic polynomial fit

Table 2: Prediction of inflammatory response by different aspects of alpha diversity (best models after REML fitting).

Factor	*DF*	*F*-Value	*P*-Value	adjusted R^2
Intercept	1,21	22.3707	0.0001	0.3200
Chao1	1,21	9.8274	0.0050	
B6	1,21	9.2976	0.0061	
Intercept	1,21	19.5089	0.0002	0.2707
Shannon H	1,21	4.4470	0.0471	
B6	1,21	10.5759	0.0038	
Intercept	1,21	27.2684	<0.0001	0.2625
NTI	1,21	3.4459	0.0775	
B6	1,21	12.1853	0.0022	
Intercept	1,21	27.5336	<0.0001	0.2212
NRI	1,21	1.2906	0.2687	
B6	1,21	10.1947	0.0044	
Intercept	1,21	21.8733	0.0001	0.3505
ΔChao1 [before-after *S*. T. infection]	1,21	13.6973	0.0013	
B6	1,21	8.8243	0.0073	

Supplementary Material

Supplementary Figures:

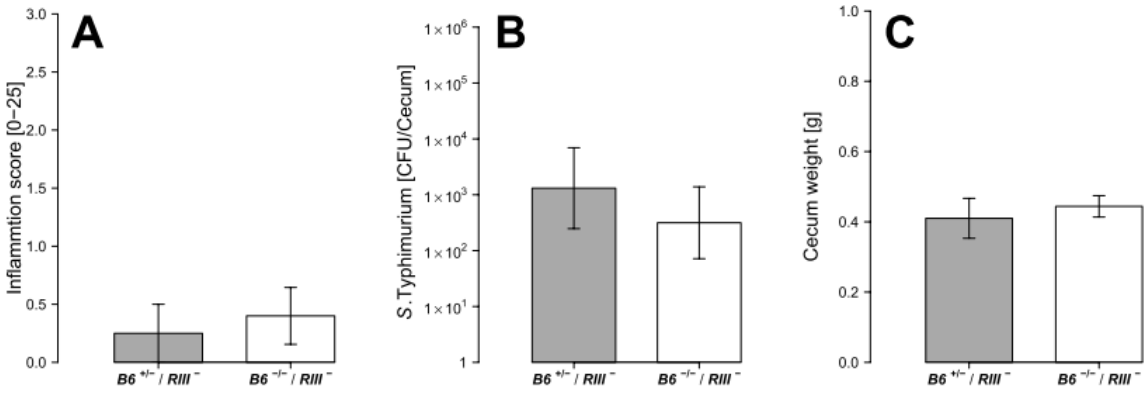


Figure S1: Inflammation after chronic infection with *S.* Typhimurium ∆*aroA*. (**A**) We find no difference between mice differing in *B4galnt2* expression in histological inflammation (Z=0.447, P=1.000), *Salmonella* load (**B**; Z=-0.747, P=0.5658) and (**C**) cecum weight (Z=0.490, P=0.7311) (Wilcoxon test via Monte-Carlo resampling ;# P<0.100, * P<0.050, ** P<0.010, *** P<0.001).

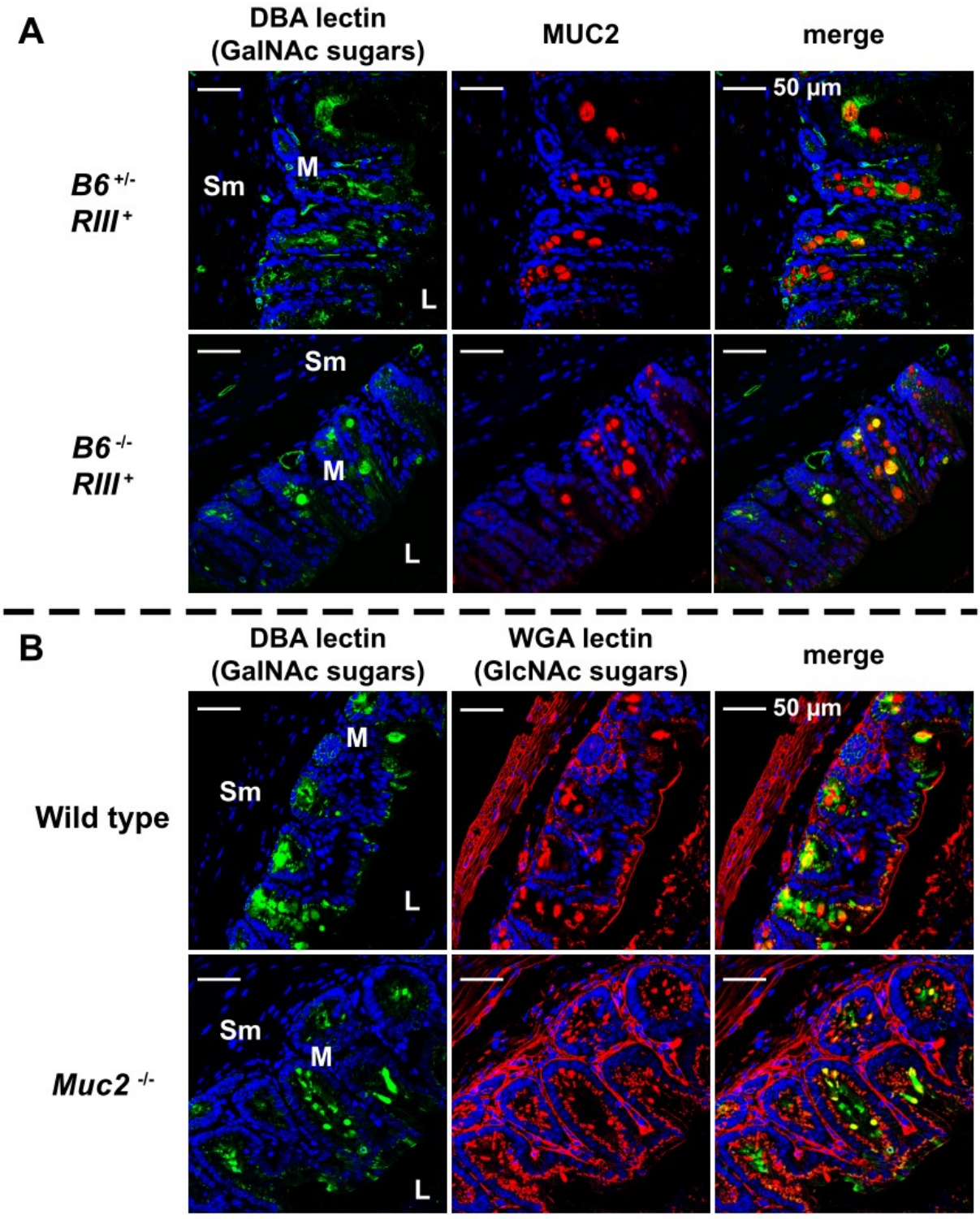


Figure S2: *B4galnt2* glycosylation dynamics in the intestinal mucosa (addition to Figure 2A and 2E). (**A**) Mucin-2 (MUC-2) and *B4galnt2* glycan residues (GalNAc) were stained with fluorescein labeled DBA in formalin fixed cecal tissue sections (Sm-submucosa , M-mucosa, L-lumen). (**B**) *B4galnt2* glycan residues (GalNAc) were stained with fluorescein labeled DBA in formalin fixed cecal tissue sections before and 1 days p.i. with *S.* Typhimurium. GlcNAc residues were stained with Alexa633 labeled Wheat Germ Agglutinin (WGA).

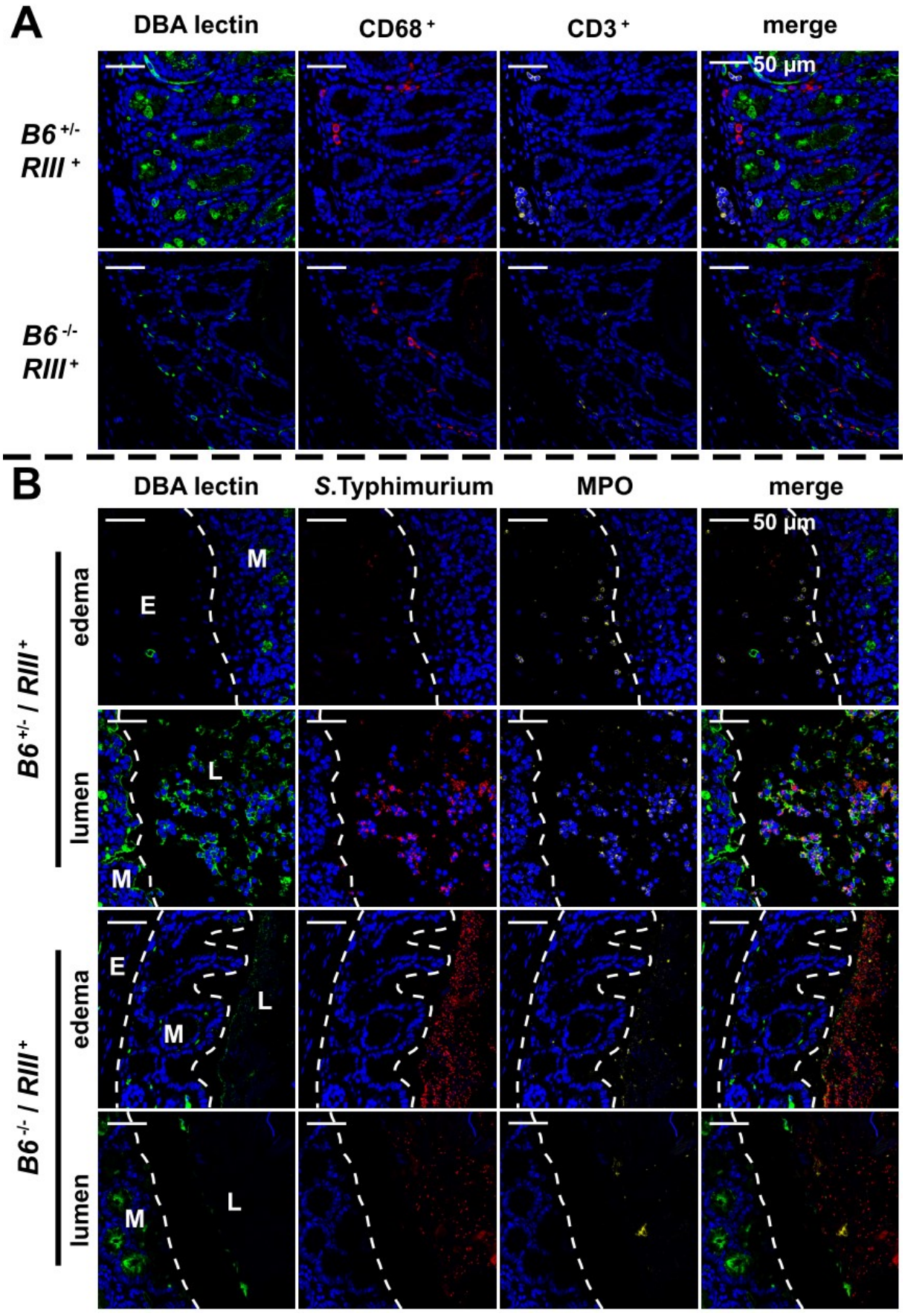


Figure S3: *B4galnt2*-dependent infiltration of immune cells after *S.* Typhimurium infection (addition to Figure 5A and 5C). (**A**) Immunofluorescence staining and

enumeration of positive cells per vision field showed that *B6* $^{+/-}$ mice have higher numbers of CD68 $^{+}$ and CD3 $^{+}$ cells in the cecal mucosa 1d p.i. (N=5-7; E-edema, M-mucosa, L-lumen). Nuclei were counterstained with DAPI and *B4galnt2* glycans by using fluorescein labeled DBA. (**B**) Myeloperoxidase (MPO) positive cells and *S.* Typhimurium were determined by immunofluorescence staining in formalin fixed cecal sections (5 µm).

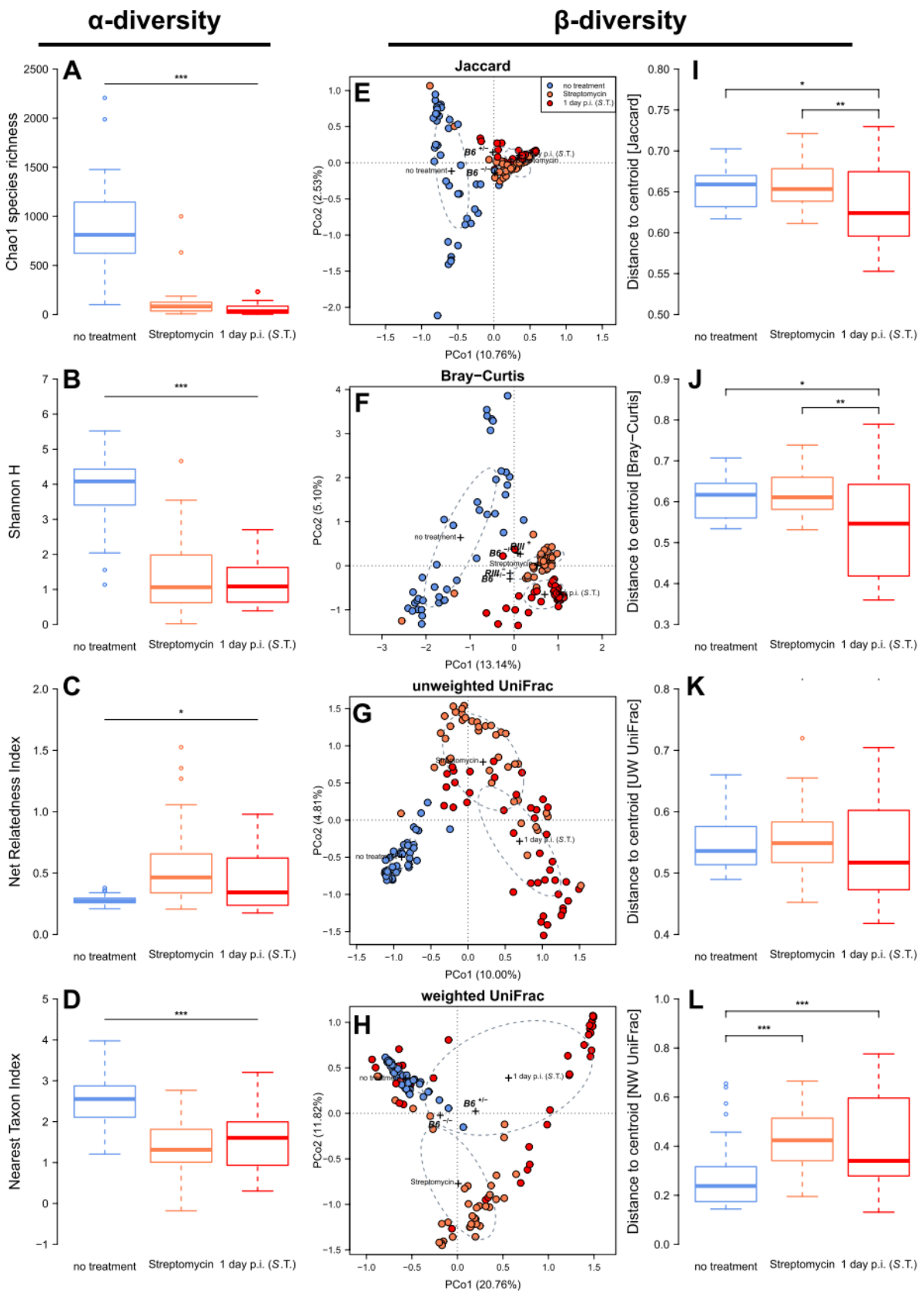


Figure S4: Analyses of microbial alpha diversity and beta diversity among treatments. Microbial diversity was estimated from 97% species level OTUs and focused on species richness (**A**; Chao1: χ^2=78.940, P<2.2 × 10^{-16}; Kruskal-Wallis test), species distribution (**B**;

Shannon H: χ^2=65.997, P=4.666 × 10^{-15}; Kruskal-Wallis test), and distant and close phylogenetic relatedness (**C**; NRI: χ^2=6.4166, P=0.04043; **D**; NTI: χ^2=50.4593, P=1.104 × 10^{-11}; Kruskal-Wallis test). Community changes among treatments were measured by the Jaccard distance (**E**; *adonis*: $F_{2,120}$=9.577, R^2=0.13765, P<0.0001), Bray-Curtis (**F**; *adonis*: $F_{2,120}$=12.055, R^2=0.1673, P<0.0001), UW-UF (**G**; *adonis*: $F_{2,120}$=13.932, R^2=0.18845, P<0.0001), and W-UF (**H**; *adonis*: $F_{2,120}$=20.615, R^2=0.25572, P<0.0001). Within treatment community variability (**I-L**) was also strongly influenced by the treatment regime (J-$F_{2,120}$=5.5668, P=0.0054; BC- $F_{2,120}$=9.1942, P=0.0004; W-UF: $F_{2,120}$=11.832, P<0.0001; UW-UF: $F_{2,120}$= 1.7496, P=0.1804).

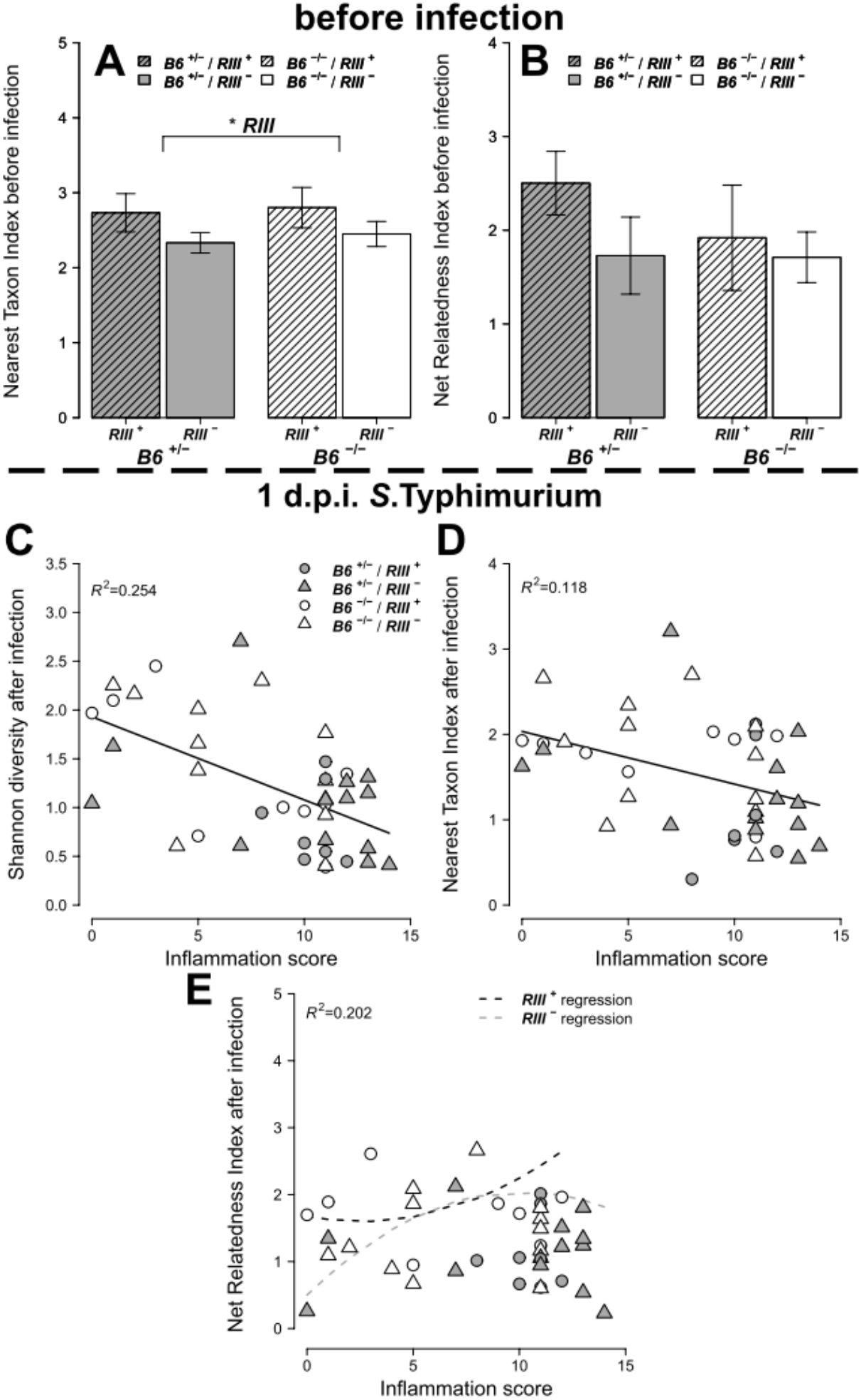

before infection
A
B
Nearest Taxon Index before infection
Net Relatedness Index before infection
* RIII
B6 +/- / RIII +
B6 +/- / RIII -
B6 -/- / RIII +
B6 -/- / RIII -
RIII +
RIII -
B6 +/-
B6 -/-
1 d.p.i. S.Typhimurium
C
D
E
Shannon diversity after infection
Nearest Taxon Index after infection
Net Relatedness Index after infection
Inflammation score
R2=0.254
R2=0.118
R2=0.202
RIII + regression
RIII - regression

Figure S5: Analysis of microbial alpha diversity among genotypes and their influence on intestinal inflammation. Microbial diversity was estimated from 97% species level OTUs and focused on species distribution (Shannon H: **C**), and close and distant phylogenetic relatedness (NTI: **A**, **D**; NRI: **B**, **E**), in the untreated state (**A**, **B**) and 1 day post infection with *S*. Typhimurium (**C-E**; Table 1 for the respective statistics).

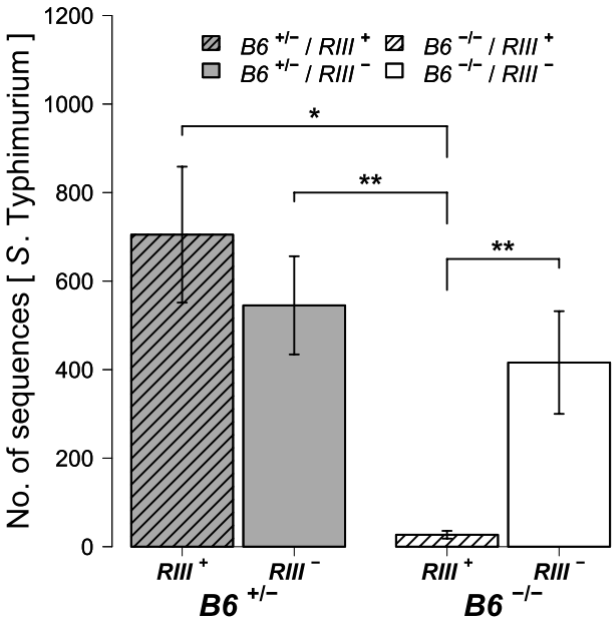


Figure S6: Salmonella abundance among *B4galnt2* genotypes based on sequence abundance. *Salmonella* abundance significantly differed between *B6* and *RIII* genotypes (*B6*: $F_{1,20}$=5.32081, P=0.0319; *RIII*: $F_{1,20}$=6.91949, P=0.0160; *B6/RIII*: $F_{1,20}$=2.74565, P=0.1131, adj.R^2=0.28114; LMM) with the lowest abundance in *RIII* $^{+}$/*B6* $^{-/-}$ animals (Tukey pairwise comparisons: *RIII* $^{+}$/*B6* $^{-/-}$-*RIII* $^{-}$/*B6* $^{-/-}$: Z=-3.102, P=0.00979; *RIII* $^{+}$/*B6* $^{-/-}$-*RIII* $^{-}$/*B6* $^{+/-}$: Z=-3.430, P=0.00341; *RIII* $^{+}$/*B6* $^{+/-}$-*RIII* $^{+}$/*B6* $^{-/-}$: Z=2.582, P=0.04698).

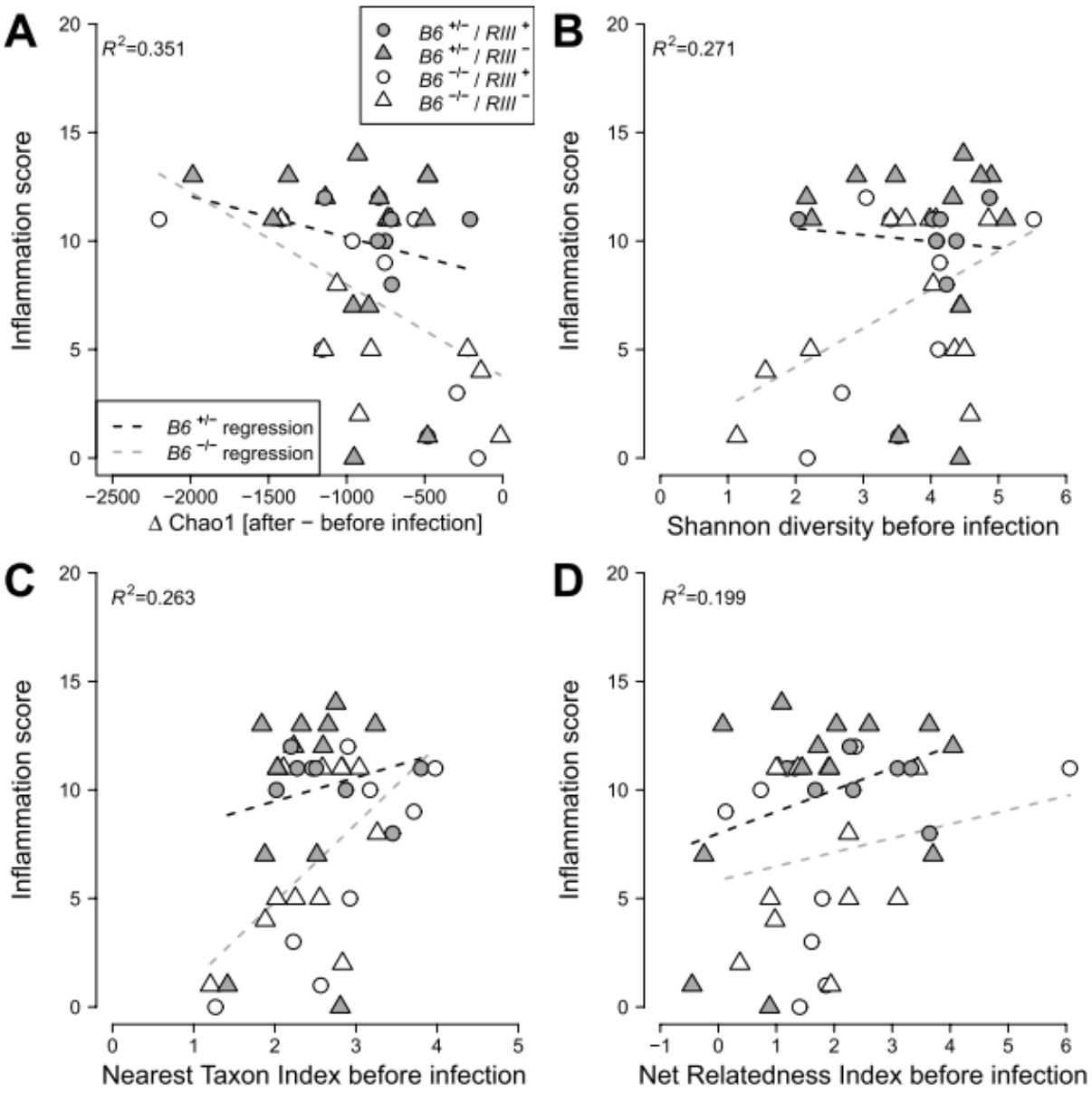


Figure S7: Prediction of infection outcome by alpha diversity. The severity of histological inflammation was significantly predictable by the change of species richness inflicted by *S*. Typhimurium infection and streptomycin treatment (**A**, ΔChao1), by the eveness of species distribution before infection (**B**, Shannon H), and clusteredness of closely related phylogenetic groups before infection (**C**, NTI). Phylogenetic clustering of distantly related species before infecation shows no significant association to the resulting inflammation (**D**, NRI, see Table 2).

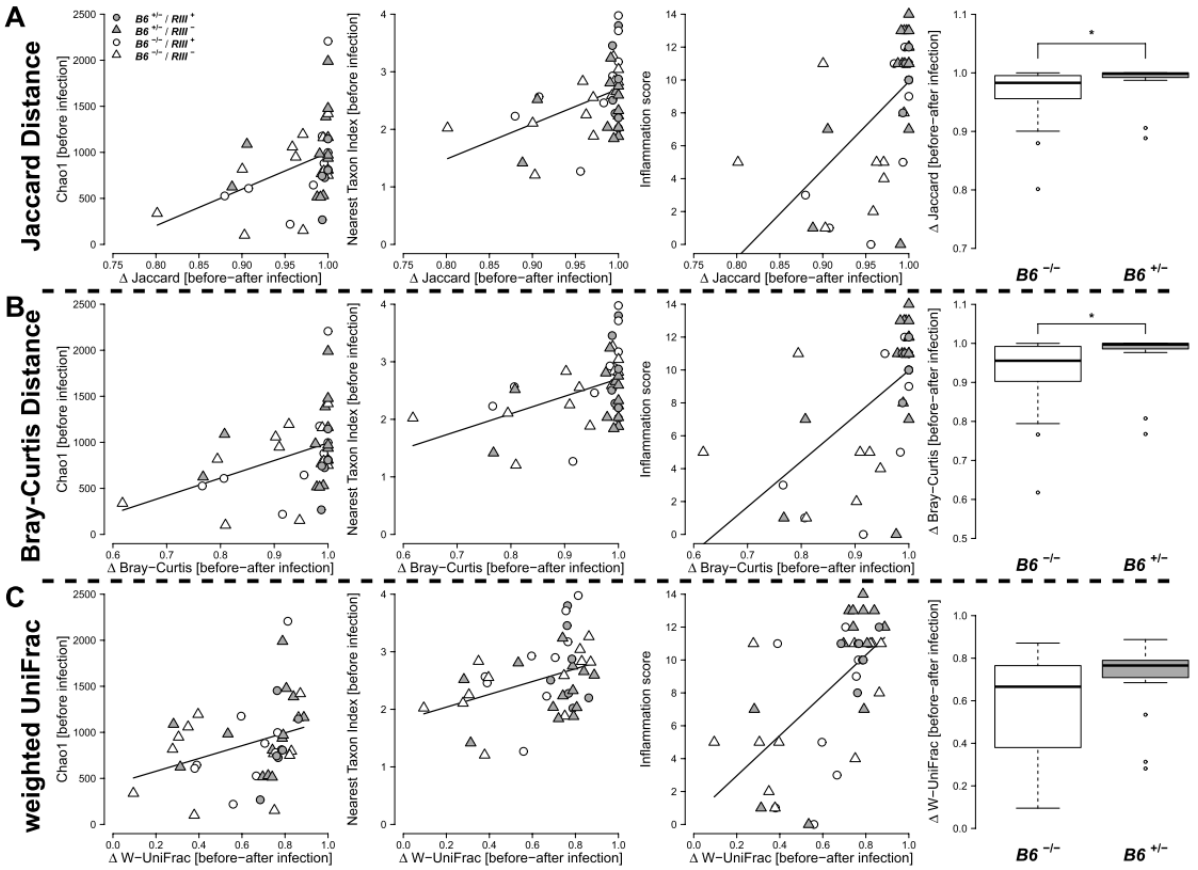


Figure S8: Analyses of community disturbance. The community distances between animals before and after treatment were used as a measure of community disturbance. This disturbance signifies an increased species turnover (higher distance) in animals with a diverse microbial community measured in different ways, considering species number, distribution and phylogenetic relatedness (see Table S2). Community turnover also correlates strongly with severity of inflammation, and increased *Salmonella* load (see Table S2). Furthermore animals lacking epithelial *B4galnt2* expression have on average less disturbance/higher resilience than mice with gut epithelial expression (Δ W-UniFrac: Z=-1.6171, P=0.1090; Δ Jaccard: Z=-2.2731, P=0.02311; Δ Bray-Curtis: Z=-2.2998, P=0.0205; Wilcoxon test via Monte-Carlo resampling; see also Figure 7).

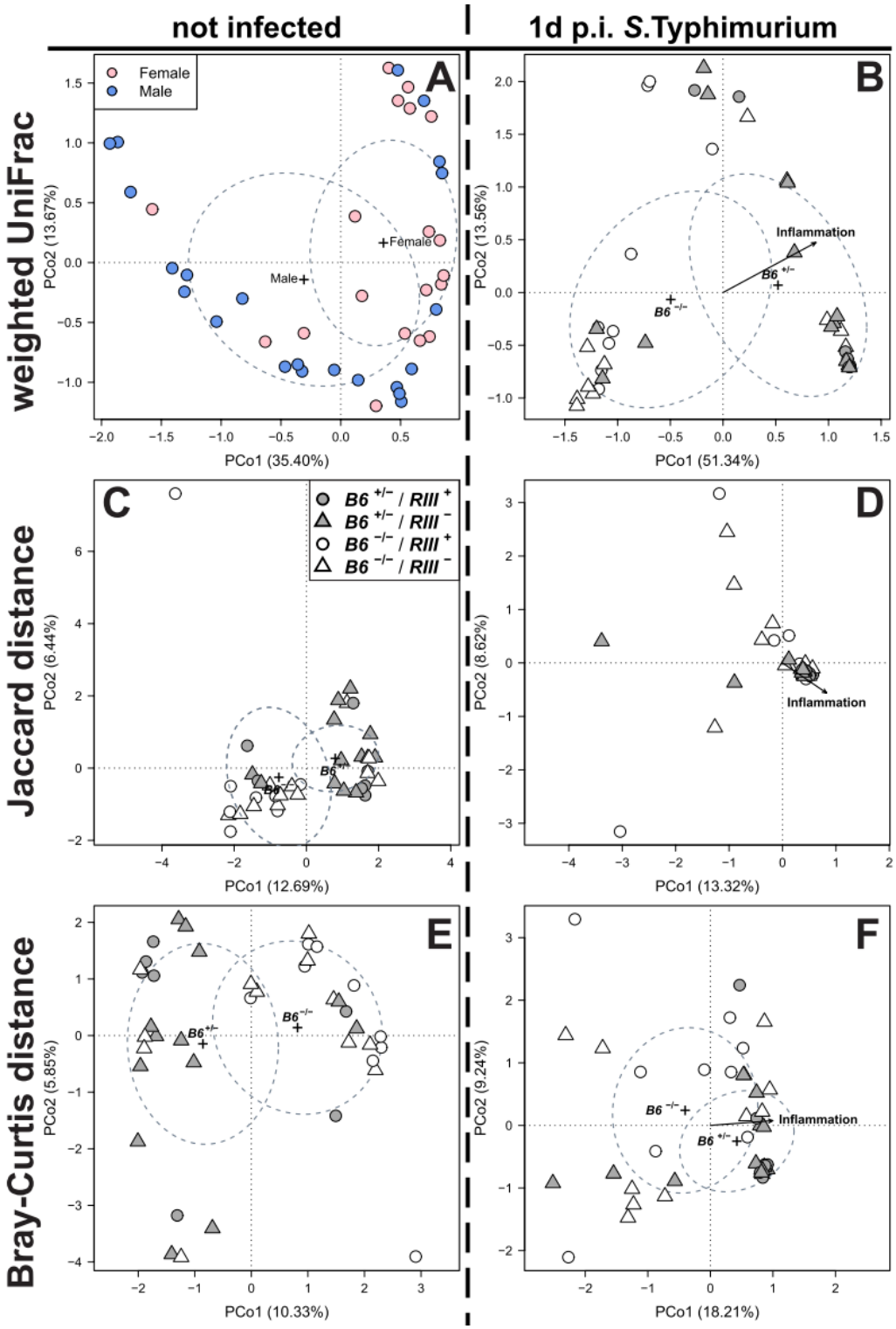


Figure S9: Principal coordinate Analyses of different beta diversity measures. PCoAs of phylogenetically informed (**A**, **B**) and species based (**C-F**) metrics of beta diversity, that show clustering of microbial communities by epithelial *B4galnt2* expression (**C**: R^2=0.1478, P=0.0011; **E**: R^2=0.1373, P=0.0020) and sex (**A**: R^2=0.0884, P=0.0260) before any treatment. After *S*.Typhimurium infection the community structures show strong and consistent correlation to histological inflammation (**B**: R^2=0.5054, P<0.0001; **D**: R^2=0.3167, P=0.0006; F: R^2=0.4935, P=0.0002) and significant discrimination among epithelial and

endothelial *B4galnt2* expression patterns (**B**: *B6*-R^2=0.1272, *P*=0.005199; **F**: *B6*-R^2=0.0951, *P*=0.01430).

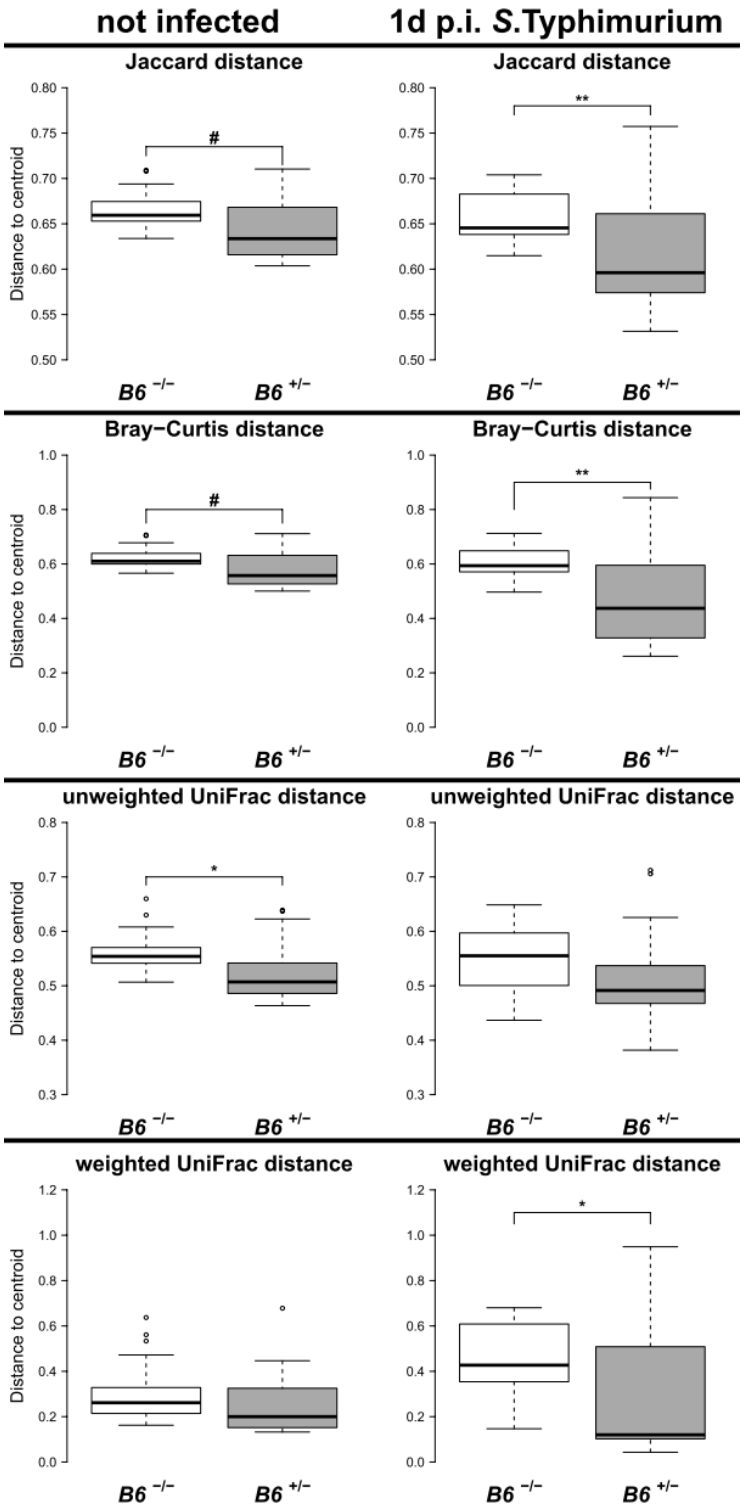


Figure S10: Community variability between genotypes. Comparison of bacterial community distances (beta diversity) between animals with and without epithelial *B4GalNt2*

expression, before and after *S.* Typhimurium infection (not infected- Jaccard: $F_{1,39}$=4.1584, P=0.04779; Bray-Curtis: $F_{1,39}$=3.961, P=0.05379, UW-UF: $F_{1,39}$=5.414, P=0.0246; W-UF: $F_{1,39}$=1.235, P=0.2732; 1d p.i. *S.* Typhimurium- Jaccard: $F_{1,39}$=7.614, P=0.006399; Bray-Curtis: $F_{1,39}$=9.1036, P=0.003399; UW-UF: $F_{1,39}$=2.3871, P=0.1334; W-UF: $F_{1,39}$=4.7569, P=0.03379). The beta diversity within genotypes was approximated by the distance of each sample to the centroid of its respective cluster (*B6* $^{+/-}$ or *B6* $^{-/-}$).

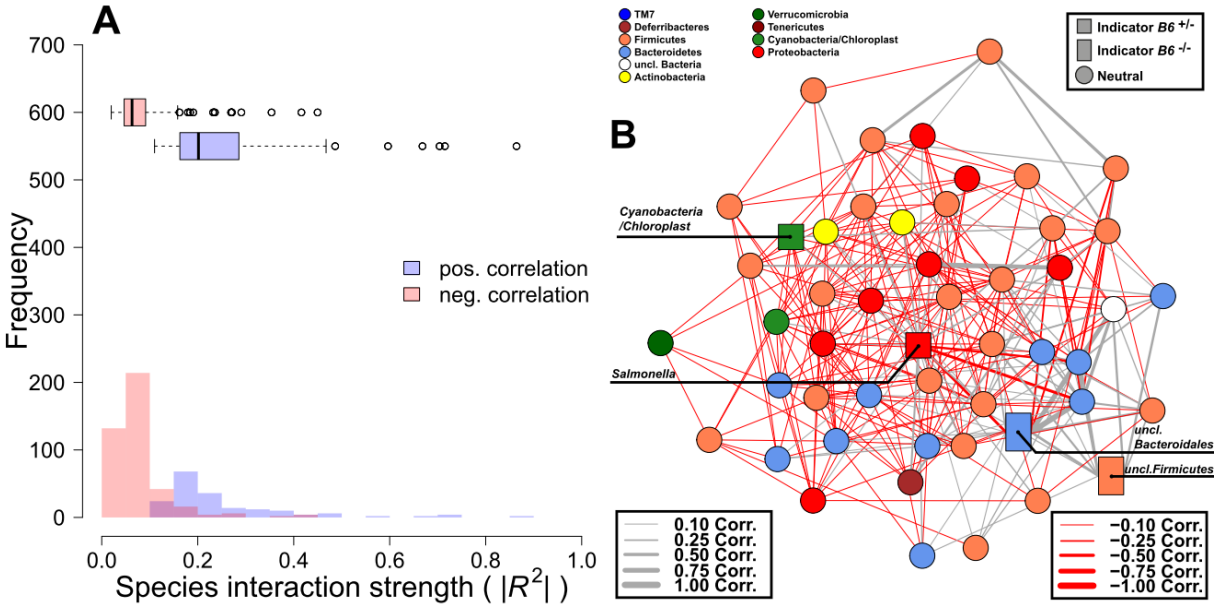


Figure S11: Co-occurrence network after streptomycin and *S.* Typhimurium infection. (**A**) Distribution of pairwise genera correlations after *Salmonella* infection, with a higher number of weak negative interactions, but higher positive interaction strength (positive/negative interactions=0.4381;W=74056, $P < 2.20 \times 10^{-16}$; Wilcoxon test). (**B**) Genera co-occrurence network with highlighted indicators for *B6* genotypes. The network also visualizes the central and strong influence of *Salmonella* on other community members (square - *B6* $^{+/-}$ indicator, rectangle - *B6* $^{-/-}$ indicator, circle - no indicator/neutral; see Table S6).

Supplementary tables:

Table S1: Statistical analyses of CFU counts, cecum weights, inflammation markers, and gene expression.

	not infected				1 day p.i.				14 days p.i.			
Measurement	Factors	*Df*	*F*-Value	*P*-Value	Factors	*Df*	*F*-Value	*P*-Value	Factors	*Df*	*F*-Value	*P*-Value
Lipocalin-2 (Lcn-2)**	*1*	16	NA	NA	*B6*	1,28	17.494	**0.0003**	-	-	-	-
					RIII	1,28	7.271	**0.0117**	-	-	-	-
Colony Forming Units*	-	-	-	-	*RIII*	1,49	10.537	**0.0021**	*B6*	1,23	0.0098	0.92207
	-	-	-	-					*RIII*	1,23	1.1159	0.30176
	-	-	-	-					*B6* : *RIII*	1,23	7.3680	**0.01237**
CD3 cells*	-	-	-	-	*B6*	1,22	20.170	**0.0002**	-	-	-	-
CD68 cells	-	-	-	-	*B6*	1,22	19.060	**0.0003**	-	-	-	-
MPO (RFU signal)	-	-	-	-	*B6*	1,26	20.300	**0.0001**	-	-	-	-

* log(X); ** $X^{1/4}$; *** X^2 data transformations; NA - no data available

Table S2: Analyses of community resistance/turnover as community distance between pre- and post-infection time points in SPF raised mice.

Distance	Factor	*ρ*	*P*-Value	Comparison	*Z*	*P*-Value
Bray-Curtis (no-1 d.p.i.)	Inflammation	0.6131	**0.000020**	*B6*	-2.2998	**0.0205**
	Salmonella	0.6393	**0.000007**	*RIII*	-1.0139	0.3204
	Chao1 (no treatment)	0.5062	**0.000733**			
	Shannon H (no treatment)	0.4332	**0.004670**			
	NRI (no treatment)	0.0922	0.566500			
	NTI (no treatment)	0.3631	**0.019610**			
	ΔChao1 (1 d.p.i.-no treat.)	-0.5911	**0.000047**			
Jaccard (no-1 d.p.i.)	Inflammation	0.6092	**0.000024**	*B6*	-2.2731	**0.0231**
	Salmonella	0.6236	**0.000013**	*RIII*	-1.0139	0.3227
	Chao1 (no treatment)	0.5266	**0.000405**			
	Shannon H (no treatment)	0.4484	**0.003283**			
	NRI (no treatment)	0.0768	0.633100			
	NTI (no treatment)	0.3785	**0.014680**			
	ΔChao1 (1 d.p.i.-no treat.)	-0.6098	**0.000023**			
unweighted UniFrac (no-1 d.p.i.)	Inflammation	0.5894	**0.000050**	*B6*	-2.3213	**0.0198**
	Salmonella	0.6024	**0.000031**	*RIII*	-0.3742	0.7216
	Chao1 (no treatment)	0.6040	**0.000042**			
	Shannon H (no treatment)	0.5057	**0.000869**			
	NRI (no treatment)	0.1030	0.520500			
	NTI (no treatment)	0.2920	0.064310			
	ΔChao1 (1 d.p.i.-no treat.)	-0.6897	**0.000001**			
weighted UniFrac (no-1 d.p.i.)	Inflammation	0.5429	**0.000245**	*B6*	-1.6171	0.1090
	Salmonella	0.7412	**0.000000**	*RIII*	-0.1871	0.8649
	Chao1 (no treatment)	0.4469	**0.003701**			
	Shannon H (no treatment)	0.4132	**0.007664**			
	NRI (no treatment)	0.2145	0.177600			
	NTI (no treatment)	0.3334	**0.033660**			
	ΔChao1 (1 d.p.i.-no treat.)	-0.5148	**0.000680**			

Table S3: Results of distance based redundancy analysis on different beta diversity metrics before and after *S.* Typhimurium infection.

Time point	Distance	Factor	*DF*	*F*-Value	*P*-Value	R^2	adj. R^2
before treatment	UW-UniFrac	*B6*	1,39	1.6603	**0.0008**	0.0408	0.0162
	NW-UniFrac	*B6*	1,39	1.3449	0.1336	0.0333	0.0085
	Jaccard	*B6*	1,39	1.6369	**0.0006**	0.0403	0.0157
	Bray-Curtis	*B6*	1,39	2.2698	**0.0010**	0.0550	0.0308
	RDA	*B6*	1,39	2.3381	**0.0028**	0.0566	0.0324
1 d.p.i.	UW-UniFrac	*B6*	1,39	1.4054	**0.0402**	0.0348	0.0100
	NW-UniFrac	*B6*	1,39	4.5133	**0.0028**	0.1078	0.0849
	Jaccard	*B6*	1,39	1.5109	**0.0114**	0.0373	0.0126
	Bray-Curtis	*B6*	1,39	2.7568	**0.0034**	0.0664	0.0424
	RDA	*B6*	1,39	3.1621	**0.0058**	0.0750	0.0513
	UW-UniFrac	*RIII*	1,37	1.2393	0.0942	0.1351	0.0650
		B6	1,37	1.5350	**0.0180**		
		Inflammation	1,37	3.0064	**0.0002**		
	NW-UniFrac	*RIII*	1,37	1.2393	0.0996	0.1351	0.0650
		B6	1,37	1.5350	**0.0186**		

	Inflammation	1,37	3.0064	**0.0002**		
Jaccard	*RIII*	1,37	1.0892	0.2424	0.1197	0.0483
	B6	1,37	1.6055	**0.0094**		
	Inflammation	1,37	2.3346	**0.0002**		
Bray-Curtis	*RIII*	1,37	1.2220	0.1866	0.1901	0.1244
	B6	1,37	2.9999	**0.0020**		
	Inflammation	1,37	4.4022	**0.0002**		
RDA	*RIII*	1,37	1.5673	0.1046	0.2140	0.1503
	B6	1,37	3.3826	**0.0050**		
	Inflammation	1,37	5.1250	**0.0002**		

Table S4: Indicator species analysis based on consensus genera for *B4Galnt2* expression patterns in SPF mice (*B6*, *RIII*), before and after *S. Typhimurium* treatment.

Time point	Classification (RDP 9, modified by P.Schloss)	Factor	*r.g.*	*P*-Value	*P*-Value (FDR)
before treatment	*Barnesiella*	$B6^{-/-}$	0.31314	0.02480	0.25624
	unclassified Porphyromonadaceae	$B6^{-/-}$	0.41552	0.00390	0.06044
	Turicibacter	$B6^{-/-}$	0.42849	0.00030	**0.01860**
	Bacteroides	$B6^{+/-}$	0.42503	0.00280	0.05786
	Parasutterella	$B6^{+/-}$	0.33333	0.04970	0.44016
	Prevotella	$B6^{+/-}$	0.41335	0.00880	0.10911
	unclassified Prevotellaceae	$B6^{+/-}$	0.49092	0.00160	**0.04960**
1 d.p.i.	*unclassified Bacteroidales*	$B6^{-/-}$	0.33913	0.03040	0.37631
	unclassified Firmicutes	$B6^{-/-}$	0.31522	0.03580	0.37631
	Salmonella	$B6^{+/-}$	0.42981	0.00580	0.22618
	Streptophyta	$B6^{+/-}$	0.33049	0.03860	0.37631
before treatment	*unclassified Erysipelotrichaceae*	$RIII^{+}$	0.43033	0.00640	0.19838
	Marvinbryantia	$RIII^{+}$	0.34085	0.02800	0.57861
	Turicibacter	$RIII^{+}$	0.40214	0.00110	0.06819
1 d.p.i.	*unclassified Lachnospiraceae*	$RIII^{+}$	0.19198	0.01900	0.66215
	Turicibacter	$RIII^{+}$	0.20854	0.04350	0.66215

Table S5: Indicator species analysis based on species level OTUs for *B4Galnt2* genotypes in SPF mice (*B6*, *RIII*), before and after *S. Typhimurium* treatment.

Time point	Classification (RDP 9, modified by P.Schloss)	Factor	*r.g.*	*P*-Value	*P*-Value (FDR)
before treatment	*0010-Firmicutes(100);Erysipelotrichia(100);Erysipelotrichales(100);Erysipelotrichaceae(100);Turicibacter(100);*	$B6^{-/-}$	0.4151	0.0005	0.2095
	0013-Bacteroidetes(100);Bacteroidia(94);Bacteroidales(94);Porphyromonadaceae(87);Barnesiella(66);	$B6^{-/-}$	0.4734	0.0005	0.2095
	0050-Firmicutes(100);Clostridia(100);Clostridiales(100);Lachnospiraceae(100);Robinsoniella(96);	$B6^{-/-}$	0.4043	0.0034	0.4383
	0053-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Porphyromonadaceae(100);Butyricimonas(66);	$B6^{-/-}$	0.4427	0.0013	0.2607
	0101-Bacteroidetes(100);Bacteroidia(97);Bacteroidales(97);Porphyromonadaceae(97);Paludibacter(96);	$B6^{-/-}$	0.3598	0.0072	0.5538
	0125-Firmicutes(100);Clostridia(100);Clostridiales(100);Lachnospiraceae(100);Robinsoniella(96);	$B6^{-/-}$	0.2386	0.0449	0.8453
	0157-Firmicutes(100);Clostridia(100);Clostridiales(100);Ruminococcaceae(98);Hydrogenoanaerobacterium(91);	$B6^{-/-}$	0.4036	0.0082	0.5608
	0189-Firmicutes(100);Clostridia(100);Clostridiales(100);Lachnospiraceae(100);Lachnobacterium(98);	$B6^{-/-}$	0.2080	0.0073	0.5538
	0195-Bacteroidetes(100);Bacteroidia(89);Bacteroidales(89);Porphyromonadaceae(89);Paludibacter(53);	$B6^{-/-}$	0.3612	0.0146	0.7134
	0196-Bacteroidetes(100);Flavobacteria(53);Flavobacteriales(53);Flavobacteriaceae(53);Pseudozobellia(53);	$B6^{-/-}$	0.3631	0.0061	0.5538
	0215-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Porphyromonadaceae(100);Tannerella(59);	$B6^{-/-}$	0.4028	0.0029	0.4050
	0216-Bacteroidetes(100);Bacteroidia(97);Bacteroidales(97);Marinilabiaceae(83);Anaerophaga(83);	$B6^{-/-}$	0.3159	0.0450	0.8453
	0252-Bacteroidetes(100);Bacteroidia(92);Bacteroidales(92);Marinilabiaceae(63);Anaerophaga(63);	$B6^{-/-}$	0.3349	0.0067	0.5538
	0276-Firmicutes(100);Clostridia(100);Clostridiales(100);Lachnospiraceae(100);Lachnospiracea incertae sedis(96);	$B6^{-/-}$	0.3711	0.0084	0.5608
	0286-Firmicutes(100);Clostridia(100);Clostridiales(100);Peptococcaceae 1(96);Peptococcus(96);	$B6^{-/-}$	0.2774	0.0464	0.8453
	0287-Bacteroidetes(100);Sphingobacteria(55);Sphingobacteriales(55);Flammeovirgaceae(55);Limibacter(55);	$B6^{-/-}$	0.3494	0.0076	0.5538
	0293-Bacteroidetes(100);Bacteroidia(61);Bacteroidales(61);	$B6^{-/-}$	0.3617	0.0204	0.8453
	0317-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Porphyromonadaceae(100);Barnesiella(71);	$B6^{-/-}$	0.4350	0.0020	0.3352
	0356-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Porphyromonadaceae(100);Paludibacter(86);	$B6^{-/-}$	0.2958	0.0487	0.8453
	0358-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Marinilabiaceae(93);Anaerophaga(93);	$B6^{-/-}$	0.3329	0.0460	0.8453
	0361-Firmicutes(100);Clostridia(100);Clostridiales(100);Lachnospiraceae(100);Acetitomaculum(100);	$B6^{-/-}$	0.3213	0.0488	0.8453
	0370-Firmicutes(100);Clostridia(100);Clostridiales(100);Lachnospiraceae(100);Robinsoniella(93);	$B6^{-/-}$	0.3392	0.0236	0.8453
	0391-Firmicutes(100);Clostridia(100);Clostridiales(100);Ruminococcaceae(100);Flavonifractor(100);	$B6^{-/-}$	0.3437	0.0324	0.8453
	0420-Firmicutes(100);Erysipelotrichia(100);Erysipelotrichales(100);Erysipelotrichaceae(100);Turicibacter(100);	$B6^{-/-}$	0.3859	0.0226	0.8453
	0424-Firmicutes(100);Clostridia(100);Clostridiales(100);Peptococcaceae 1(100);Peptococcus(100);	$B6^{-/-}$	0.3456	0.0175	0.7926
	0512-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Porphyromonadaceae(89);	$B6^{-/-}$	0.2872	0.0494	0.8453
	0525-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Porphyromonadaceae(78);Paludibacter(78);	$B6^{-/-}$	0.3361	0.0487	0.8453
	0678-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Porphyromonadaceae(100);Barnesiella(100);	$B6^{-/-}$	0.4082	0.0216	0.8453
	0815-Firmicutes(100);Clostridia(100);Clostridiales(100);Ruminococcaceae(100);Butyricicoccus(100);	$B6^{-/-}$	0.3676	0.0448	0.8453
	0019-Bacteroidetes(100);Bacteroidia(97);Bacteroidales(97);Porphyromonadaceae(96);Tannerella(95);	$B6^{+/-}$	0.3155	0.0393	0.8453
	0031-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Prevotellaceae(100);Prevotella(85);	$B6^{+/-}$	0.4956	0.0009	0.2514
	0034-Bacteroidetes(100);Sphingobacteria(93);Sphingobacteriales(93);Flammeovirgaceae(93);Limibacter(93);	$B6^{+/-}$	0.3795	0.0145	0.7134
	0074-Bacteroidetes(100);Bacteroidia(96);Bacteroidales(96);Porphyromonadaceae(96);Paludibacter(92);	$B6^{+/-}$	0.3838	0.0073	0.5538

0079-Bacteroidetes(100);Bacteroidia(70);Bacteroidales(70);Porphyromonadaceae(66);Paludibacter(66);	*B6* $^{+/-}$	0.3530	0.0247	0.8453
0082-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Bacteroidaceae(100);Bacteroides(100);	*B6* $^{+/-}$	0.3912	0.0076	0.5538
0090-Bacteroidetes(100);Bacteroidia(97);Bacteroidales(97);Porphyromonadaceae(92);Tannerella(92);	*B6* $^{+/-}$	0.3549	0.0149	0.7134
0092-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Rikenellaceae(100);Rikenella(100);	***B6*** $^{+/-}$	**0.4292**	**0.0001**	**0.0838**
0100-Firmicutes(100);Clostridia(100);Clostridiales(100);Lachnospiraceae(99);Robinsoniella(98);	*B6* $^{+/-}$	0.2492	0.0060	0.5538
0117-Bacteroidetes(100);Sphingobacteria(100);Sphingobacteriales(100);Flammeovirgaceae(100);Limibacter(100);	*B6* $^{+/-}$	0.3921	0.0069	0.5538
0120-Bacteroidetes(100);Bacteroidia(98);Bacteroidales(98);Porphyromonadaceae(98);Paludibacter(98);	*B6* $^{+/-}$	0.3845	0.0111	0.6001
0147-Bacteroidetes(100);Sphingobacteria(54);Sphingobacteriales(54);Flammeovirgaceae(54);Limibacter(54);	*B6* $^{+/-}$	0.3258	0.0229	0.8453
0149-Bacteroidetes(100);Bacteroidia(83);Bacteroidales(83);Porphyromonadaceae(83);Paludibacter(81);	*B6* $^{+/-}$	0.4141	0.0052	0.5538
0150-Bacteroidetes(100);Sphingobacteria(90);Sphingobacteriales(90);Flammeovirgaceae(90);Limibacter(90);	*B6* $^{+/-}$	0.3965	0.0014	0.2607
0153-Firmicutes(100);Clostridia(100);Clostridiales(100);Lachnospiraceae(100);Syntrophococcus(77);	*B6* $^{+/-}$	0.1881	0.0477	0.8453
0159-Bacteroidetes(100);Bacteroidia(80);Bacteroidales(80);Porphyromonadaceae(63);Paludibacter(56);	*B6* $^{+/-}$	0.3678	0.0183	0.8070
0161-Bacteroidetes(100);Bacteroidia(74);Bacteroidales(74);Marinilabiaceae(53);Anaerophaga(53);	*B6* $^{+/-}$	0.3256	0.0227	0.8453
0165-Bacteroidetes(98);Bacteroidia(52);Bacteroidales(52);	***B6*** $^{+/-}$	**0.5930**	**0.0001**	**0.0838**
0168-Bacteroidetes(100);Bacteroidia(66);Bacteroidales(66);Porphyromonadaceae(61);	*B6* $^{+/-}$	0.3764	0.0161	0.7495
0201-Bacteroidetes(100);Bacteroidia(76);Bacteroidales(76);Rikenellaceae(61);Rikenella(61);	*B6* $^{+/-}$	0.3816	0.0108	0.6001
0203-Bacteroidetes(100);Sphingobacteria(54);Sphingobacteriales(54);Flammeovirgaceae(54);Limibacter(54);	*B6* $^{+/-}$	0.4490	0.0026	0.3961
0212-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Porphyromonadaceae(100);Paludibacter(97);	*B6* $^{+/-}$	0.3074	0.0446	0.8453
0244-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Marinilabiaceae(76);Anaerophaga(76);	*B6* $^{+/-}$	0.3409	0.0087	0.5608
0316-Firmicutes(100);Clostridia(100);Clostridiales(100);Lachnospiraceae(100);Roseburia(77);	*B6* $^{+/-}$	0.2837	0.0095	0.5663
0325-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Prevotellaceae(100);Prevotella(71);	*B6* $^{+/-}$	0.3295	0.0327	0.8453
0384-Bacteroidetes(100);Bacteroidia(54);Bacteroidales(54);Marinilabiaceae(54);Anaerophaga(54);	*B6* $^{+/-}$	0.3145	0.0219	0.8453
0393-Bacteroidetes(100);Bacteroidia(54);Bacteroidales(54);	*B6* $^{+/-}$	0.3248	0.0462	0.8453
0394-Bacteroidetes(100);Bacteroidia(93);Bacteroidales(93);Rikenellaceae(93);Rikenella(93);	*B6* $^{+/-}$	0.4201	0.0092	0.5663
0433-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Marinilabiaceae(82);Anaerophaga(82);	*B6* $^{+/-}$	0.3466	0.0313	0.8453
0442-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Prevotellaceae(55);unclassified;	*B6* $^{+/-}$	0.3756	0.0147	0.7134
0443-Bacteroidetes(100);Bacteroidia(82);Bacteroidales(82);Porphyromonadaceae(82);Paludibacter(73);	*B6* $^{+/-}$	0.2247	0.0416	0.8453
0460-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Porphyromonadaceae(70);unclassified;	*B6* $^{+/-}$	0.2847	0.0471	0.8453
0463-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Bacteroidaceae(100);Bacteroides(100);	*B6* $^{+/-}$	0.3725	0.0198	0.8453
0496-Bacteroidetes(100);Bacteroidia(78);Bacteroidales(78);Rikenellaceae(78);Rikenella(78);	*B6* $^{+/-}$	0.4750	0.0007	0.2346
0502-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Porphyromonadaceae(100);Tannerella(100);	*B6* $^{+/-}$	0.3333	0.0496	0.8453
0528-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Porphyromonadaceae(88);Paludibacter(88);	*B6* $^{+/-}$	0.3922	0.0098	0.5663
0557-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Marinilabiaceae(88);Alkaliflexus(88);	*B6* $^{+/-}$	0.5000	0.0013	0.2607
0590-Bacteroidetes(100);Bacteroidia(86);Bacteroidales(86);Marinilabiaceae(58);Anaerophaga(58);	*B6* $^{+/-}$	0.3145	0.0466	0.8453
0630-Bacteroidetes(100);Sphingobacteria(100);Sphingobacteriales(100);Flammeovirgaceae(100);Limibacter(100);	*B6* $^{+/-}$	0.3540	0.0245	0.8453
0633-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Rikenellaceae(100);Alistipes(100);	*B6* $^{+/-}$	0.3333	0.0499	0.8453
0668-Bacteroidetes(100);Sphingobacteria(100);Sphingobacteriales(100);Flammeovirgaceae(100);Limibacter(100);	*B6* $^{+/-}$	0.3145	0.0466	0.8453
0696-Proteobacteria(100);BetaproteoBurkholderiales(100);Sutterellaceae(100);Parasutterella(100);	*B6* $^{+/-}$	0.3333	0.0453	0.8453
0769-Firmicutes(100);Clostridia(100);Clostridiales(100);Lachnospiraceae(100);Howardella(100);	*B6* $^{+/-}$	0.3131	0.0488	0.8453

	0813-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Porphyromonadaceae(100);Paludibacter(100);	*B6* $^{+/-}$	0.3131	0.0472	0.8453
	0856-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Rikenellaceae(61);Rikenella(61);	*B6* $^{+/-}$	0.3333	0.0499	0.8453
	0906-Bacteroidetes(100);	*B6* $^{+/-}$	0.3333	0.0466	0.8453
	0912-Firmicutes(100);Clostridia(100);Clostridiales(100);Ruminococcaceae(100);Oscillibacter(100);	*B6* $^{+/-}$	0.3333	0.0484	0.8453
	0951-Bacteroidetes(100);FlavoFlavobacteriales(100);Flavobacteriaceae(100);Flagellimonas(75);	*B6* $^{+/-}$	0.3333	0.0430	0.8453
	0992-Firmicutes(100);Clostridia(100);Clostridiales(100);Lachnospiraceae(100);Butyrivibrio(75);	*B6* $^{+/-}$	0.3333	0.0499	0.8453
	1002-Firmicutes(100);Clostridia(100);Clostridiales(100);Ruminococcaceae(100);Pseudoflavonifractor(100);	*B6* $^{+/-}$	0.3333	0.0492	0.8453
	1013-Bacteroidetes(100);Sphingobacteria(75);Sphingobacteriales(75);Flammeovirgaceae(75);Limibacter(75);	*B6* $^{+/-}$	0.3333	0.0498	0.8453
1 d.p.i.	*0002-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Rikenellaceae(95);Rikenella(95);*	*B6* $^{-/-}$	0.3264	0.0330	1.0000
	0001-Proteobacteria(100);GammaproteoEnterobacteriales(100);Enterobacteriaceae(100);Salmonella(100);	*B6* $^{+/-}$	0.4243	0.0065	1.0000
	0008-Proteobacteria(100);GammaproteoEnterobacteriales(100);Enterobacteriaceae(100);Salmonella(68);	*B6* $^{+/-}$	0.4241	0.0038	1.0000
	0072-Firmicutes(100);Clostridia(100);Clostridiales(100);Clostridiaceae 1(100);Clostridium sensu stricto(100);	*B6* $^{+/-}$	0.3383	0.0282	1.0000
before treatment	*0012-Bacteroidetes(100);Sphingobacteria(95);Sphingobacteriales(95);Cytophagaceae(95);Meniscus(95);*	*RIII* $^{-}$	0.3034	0.0256	1.0000
	0024-Bacteroidetes(100);Bacteroidia(96);Bacteroidales(96);Rikenellaceae(91);Rikenella(91);	*RIII* $^{-}$	0.3546	0.0240	1.0000
	0160-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Porphyromonadaceae(89);Tannerella(89);	*RIII* $^{-}$	0.3725	0.0199	1.0000
	0168-Bacteroidetes(100);Bacteroidia(66);Bacteroidales(66);Porphyromonadaceae(61);	*RIII* $^{-}$	0.3351	0.0378	1.0000
	0170-Bacteroidetes(100);Bacteroidia(95);Bacteroidales(95);Marinilabiaceae(88);Anaerophaga(88);	*RIII* $^{-}$	0.3274	0.0284	1.0000
	0324-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Prevotellaceae(100);Hallella(100);	*RIII* $^{-}$	0.3588	0.0361	1.0000
	0346-Bacteroidetes(100);Bacteroidia(87);Bacteroidales(87);Rikenellaceae(61);Rikenella(61);	*RIII* $^{-}$	0.3171	0.0279	1.0000
	0363-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Porphyromonadaceae(100);Paludibacter(93);	*RIII* $^{-}$	0.3052	0.0315	1.0000
	0010-Firmicutes(100);Erysipelotrichia(100);Erysipelotrichales(100);Erysipelotrichaceae(100);Turicibacter(100);	*RIII* $^{+}$	0.4010	0.0013	1.0000
	0025-Firmicutes(100);Bacilli(100);Lactobacillales(100);Lactobacillaceae(100);Lactobacillus(100);	*RIII* $^{+}$	0.4979	0.0004	0.6703
	0066-Firmicutes(100);Clostridia(100);Clostridiales(100);Lachnospiraceae(100);Robinsoniella(89);	*RIII* $^{+}$	0.3395	0.0406	1.0000
	0118-Firmicutes(100);Clostridia(100);Clostridiales(100);Ruminococcaceae(52);	*RIII* $^{+}$	0.3155	0.0437	1.0000
	0230-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Porphyromonadaceae(100);Paludibacter(86);	*RIII* $^{+}$	0.3399	0.0289	1.0000
	0337-Firmicutes(100);Clostridia(100);Clostridiales(100);Lachnospiraceae(100);Robinsoniella(100);	*RIII* $^{+}$	0.2768	0.0149	1.0000
	0378-Firmicutes(100);Erysipelotrichia(100);Erysipelotrichales(100);Erysipelotrichaceae(100);Allobaculum(70);	*RIII* $^{+}$	0.3780	0.0181	1.0000
	0390-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Porphyromonadaceae(100);Paludibacter(93);	*RIII* $^{+}$	0.3940	0.0177	1.0000
	0445-Firmicutes(100);Clostridia(100);Clostridiales(100);Lachnospiraceae(100);Robinsoniella(100);	*RIII* $^{+}$	0.3245	0.0175	1.0000
	0447-Firmicutes(100);Clostridia(100);Clostridiales(100);Lachnospiraceae(100);Roseburia(100);	*RIII* $^{+}$	0.4001	0.0054	1.0000
	0590-Bacteroidetes(100);Bacteroidia(86);Bacteroidales(86);Marinilabiaceae(58);Anaerophaga(58);	*RIII* $^{+}$	0.3560	0.0188	1.0000
	0668-Bacteroidetes(100);SphingoSphingobacteriales(100);Flammeovirgaceae(100);Limibacter(100);	*RIII* $^{+}$	0.3560	0.0188	1.0000
	1000-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Porphyromonadaceae(100);Barnesiella(75);	*RIII* $^{+}$	0.3216	0.0495	1.0000
1 d.p.i.	*0005-Firmicutes(100);Erysipelotrichia(100);Erysipelotrichales(100);Erysipelotrichaceae(100);Turicibacter(100);*	*RIII* $^{+}$	0.2099	0.0142	1.0000
	0136-Firmicutes(100);Clostridia(100);Clostridiales(100);Clostridiaceae 1(100);Clostridium sensu stricto(100);	*RIII* $^{+}$	0.2071	0.0067	1.0000

Table S6: Correlation of indicator genera to the rest of the pre-infection microbial community based on Spearman rank correlations (see Figure S10).

Indicator	Correlated genera	*ρ*	*P*-Value	*P*-Value (FDR)
Bacteroides	*Parabacteroides*	-0.5811	0.0001	0.0025
	Turicibacter	-0.4847	0.0013	0.0198
	Anaerotruncus	0.3992	0.0097	0.0728
	Parasutterella	0.4042	0.0088	0.0728
	Robinsoniella	0.4278	0.0053	0.0494
	uncl. Alphaproteobacteria	0.4282	0.0052	0.0494
	Odoribacter	0.4690	0.0020	0.0248
	Prevotella	0.5452	0.0002	0.0043
	uncl. Bacteroidales	0.5536	0.0002	0.0043
	uncl. Prevotellaceae	0.9070	<0.00001	<0.00001
Parasutterella	*Prevotella*	0.4491	0.0032	0.0807
	Anaerophaga	0.4809	0.0015	0.0547
	uncl. Lactobacillales	0.4809	0.0015	0.0547
Prevotella	*Parasutterella*	0.4491	0.0032	0.0807
	Bacteroides	0.5452	0.0002	0.0085
	uncl. Prevotellaceae	0.6216	<0.00001	0.0011
uncl. Prevotellaceae	*Parabacteroides*	-0.6046	<0.00001	0.0007
	Turicibacter	-0.4474	0.0034	0.0360
	Anaerotruncus	0.4338	0.0046	0.0431
	Odoribacter	0.4865	0.0013	0.0157
	Robinsoniella	0.5249	0.0004	0.0064
	uncl. Bacteroidales	0.5480	0.0002	0.0039
	Prevotella	0.6216	<0.00001	0.0005
	Bacteroides	0.9070	<0.00001	<0.00001
Turicibacter	*Bacteroides*	-0.4847	0.0013	0.0546
	uncl. Prevotellaceae	-0.4474	0.0034	0.0601
	uncl. Erysipelotrichaceae	0.4398	0.0040	0.0601
	Lactonifactor	0.4473	0.0034	0.0601
	uncl. Porphyromonadaceae	0.4810	0.0015	0.0546

Table S7: Correlation of consensus genera- and species level OTU abundance before and after *S.* Typhimurium infection to the final histological inflammation score.

Time point	RDP9 Classification (modified by P. Schloss)	Abundance	*ρ*	*P*-Value	*P* (FDR)
before treatment	*uncl. Erysipelotrichaceae*	5	-0.3196	0.04169	0.5995
1 d.p.i.	*Acetanaerobacterium*	9	-0.3241	0.03869	0.1439
	uncl. Bacteria	40	-0.5697	0.00010	0.0010
	uncl. Bacteroidales	5899	-0.5598	0.00014	0.0010
	uncl. Bacteroidetes	3032	-0.6429	0.00001	0.0002
	uncl. Clostridiales	92	-0.3932	0.01098	0.0535
	uncl. Firmicutes	59	-0.5952	0.00004	0.0008
	Parabacteroides	1349	-0.4823	0.00141	0.0078
	uncl. Porphyromonadaceae	691	-0.5572	0.00015	0.0010
	uncl. Ruminococcaceae	50	-0.3401	0.02960	0.1283
	Salmonella	17259	0.5685	0.00011	0.0010
	Turicibacter	1165	0.3212	0.04059	0.1439
before treatment	*0195-Bacteroidetes(100);Bacteroidia(89);Bacteroidales(89);Porphyromonadaceae(89);Paludibacter(53);*	*34*	*-0.5176*	*0.00053*	*0.8241*
	0318-Firmicutes(100);Bacilli(100);Lactobacillales(100);Lactobacillaceae(100);Lactobacillus(100);	*17*	*-0.4900*	*0.00115*	*0.8241*
	0620-Bacteroidetes(100);Bacteroidia(86);Bacteroidales(86);Rikenellaceae(86);Rikenella(86);	*7*	*-0.4539*	*0.00287*	*0.8241*
	0312-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Porphyromonadaceae(95);Paludibacter(95);	*17*	*-0.4499*	*0.00317*	*0.8241*
	0618-Bacteroidetes(100);Bacteroidia(86);Bacteroidales(86);	*7*	*-0.3644*	*0.01915*	*0.8241*
	0406-Bacteroidetes(100);Bacteroidia(92);Bacteroidales(92);Porphyromonadaceae(59);Paludibacter(59);	*12*	*-0.3640*	*0.01931*	*0.8241*
	0500-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Rikenellaceae(56);Rikenella(56);	*9*	*-0.3568*	*0.02204*	*0.8241*
	0796-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Porphyromonadaceae(80);Paludibacter(61);	*5*	*-0.3547*	*0.02287*	*0.8241*
	0273-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Marinilabiaceae(91);Anaerophaga(91);	*22*	*-0.3510*	*0.02442*	*0.8241*
	0732-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Marinilabiaceae(84);Anaerophaga(84);	*6*	*-0.3493*	*0.02517*	*0.8241*
	0692-Bacteroidetes(100);SphingoSphingobacteriales(100);Cytophagaceae(100);Meniscus(100);	*6*	*-0.3416*	*0.02883*	*0.8241*
	0494-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Rikenellaceae(78);Rikenella(78);	*9*	*-0.3386*	*0.03036*	*0.8241*
	0378-Firmicutes(100);Erysipelotrichia(100);Erysipelotrichales(100);Erysipelotrichaceae(100);Allobaculum(70);	*13*	*-0.3383*	*0.03050*	*0.8241*
	0320-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Porphyromonadaceae(77);	*17*	*-0.3257*	*0.03770*	*0.8241*
	0730-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Porphyromonadaceae(100);Paludibacter(100);	*6*	*-0.3233*	*0.03924*	*0.8241*
	0634-Firmicutes(100);Clostridia(100);Clostridiales(100);Lachnospiraceae(100);Robinsoniella(100);	*7*	*-0.3120*	*0.04706*	*0.8241*
	0217-Bacteroidetes(100);Bacteroidia(94);Bacteroidales(94);Porphyromonadaceae(73);Tannerella(73);	*29*	*-0.3103*	*0.04833*	*0.8241*
	0809-Firmicutes(100);Clostridia(100);Clostridiales(100);Lachnospiraceae(100);Robinsoniella(100);	*5*	*0.3089*	*0.04941*	*0.8241*
	0840-Firmicutes(100);Clostridia(100);Clostridiales(100);Ruminococcaceae(80);Butyricicoccus(80);	*5*	*0.3089*	*0.04941*	*0.8241*
	0099-Firmicutes(100);Clostridia(100);Clostridiales(100);Ruminococcaceae(100);Oscillibacter(99);	*94*	*0.3168*	*0.04360*	*0.8241*
	0586-Firmicutes(100);Clostridia(100);Clostridiales(100);Lachnospiraceae(100);Clostridium XlVa(86);	*7*	*0.3181*	*0.04271*	*0.8241*
	0272-Firmicutes(100);Clostridia(100);Clostridiales(100);Lachnospiraceae(100);Robinsoniella(96);	*22*	*0.3210*	*0.04070*	*0.8241*
	0451-Firmicutes(100);Clostridia(100);Clostridiales(100);Ruminococcaceae(100);Anaerotruncus(91);	*11*	*0.3213*	*0.04051*	*0.8241*
	0460-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Porphyromonadaceae(70);	*10*	*0.3253*	*0.03797*	*0.8241*
	0432-Firmicutes(100);Clostridia(100);Clostridiales(100);Lachnospiraceae(100);Robinsoniella(82);	*11*	*0.3267*	*0.03710*	*0.8241*

	0640-Firmicutes(100);Clostridia(100);Clostridiales(100);Lachnospiraceae(100);Lachnospiracea incertae sedis(86);	*7*	*0.3470*	*0.02625*	*0.8241*
	0693-Firmicutes(100);Clostridia(100);Clostridiales(100);Lachnospiraceae(100);Catonella(100);	*6*	*0.3483*	*0.02563*	*0.8241*
	0313-Firmicutes(100);Clostridia(100);Clostridiales(100);Lachnospiraceae(100);Anaerostipes(83);	*17*	*0.3543*	*0.02302*	*0.8241*
	0343-Firmicutes(100);Clostridia(100);Clostridiales(100);Lachnospiraceae(100);Robinsoniella(87);	*15*	*0.3555*	*0.02255*	*0.8241*
	0772-Firmicutes(100);Clostridia(100);Clostridiales(100);Ruminococcaceae(100);Pseudoflavonifractor(100);	*5*	*0.3634*	*0.01953*	*0.8241*
	0065-Firmicutes(100);Clostridia(100);Clostridiales(100);Lachnospiraceae(100);Robinsoniella(87);	*200*	*0.3744*	*0.01587*	*0.8241*
	0226-Firmicutes(100);Clostridia(100);Clostridiales(100);Lachnospiraceae(100);Blautia(97);	*27*	*0.3747*	*0.01580*	*0.8241*
	0042-Firmicutes(100);Clostridia(100);Clostridiales(100);Lachnospiraceae(100);Robinsoniella(98);	*398*	*0.4456*	*0.00351*	*0.8241*
1 d.p.i.	*0087-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Rikenellaceae(62);Rikenella(62);*	*127*	*-0.5791*	*0.00007*	*0.0113*
	0111-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Rikenellaceae(75);Rikenella(75);	*82*	*-0.5640*	*0.00012*	*0.0128*
	0002-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Rikenellaceae(95);Rikenella(95);	*13695*	*-0.5510*	*0.00019*	*0.0147*
	0122-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Porphyromonadaceae(100);Barnesiella(99);	*65*	*-0.4986*	*0.00091*	*0.0565*
	0142-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Rikenellaceae(83);Rikenella(83);	*51*	*-0.4550*	*0.00280*	*0.1245*
	0032-Firmicutes(100);Bacilli(98);Lactobacillales(98);Carnobacteriaceae(98);Isobaculum(86);	*604*	*-0.4462*	*0.00346*	*0.1245*
	0278-Bacteroidetes(100);Bacteroidia(72);Bacteroidales(72);Porphyromonadaceae(62);Paludibacter(58);	*21*	*-0.4458*	*0.00349*	*0.1245*
	0013-Bacteroidetes(100);Bacteroidia(94);Bacteroidales(94);Porphyromonadaceae(87);Barnesiella(66);	*2084*	*-0.4444*	*0.00360*	*0.1245*
	0068-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Porphyromonadaceae(100);Barnesiella(98);	*181*	*-0.4096*	*0.00782*	*0.2254*
	0176-Bacteroidetes(80);	*39*	*-0.4045*	*0.00872*	*0.2254*
	0077-Bacteroidetes(100);Bacteroidia(75);Bacteroidales(75);Porphyromonadaceae(51);	*163*	*-0.4013*	*0.00931*	*0.2254*
	0006-Bacteroidetes(100);Bacteroidia(95);Bacteroidales(95);Porphyromonadaceae(91);Paludibacter(87);	*4936*	*-0.4007*	*0.00942*	*0.2254*
	0015-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Porphyromonadaceae(98);Paludibacter(98);	*1762*	*-0.3683*	*0.01782*	*0.3198*
	0017-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Porphyromonadaceae(100);Parabacteroides(100);	*1590*	*-0.3675*	*0.01811*	*0.3198*
	0240-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Porphyromonadaceae(96);Parabacteroides(80);	*25*	*-0.3596*	*0.02093*	*0.3198*
	0083-Firmicutes(100);Bacilli(95);Lactobacillales(95);Carnobacteriaceae(95);Isobaculum(95);	*136*	*-0.3576*	*0.02173*	*0.3198*
	0132-Firmicutes(100);Bacilli(75);Lactobacillales(75);Carnobacteriaceae(75);Isobaculum(75);	*58*	*-0.3576*	*0.02173*	*0.3198*
	0354-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Porphyromonadaceae(100);Paludibacter(100);	*15*	*-0.3546*	*0.02292*	*0.3198*
	0108-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Porphyromonadaceae(93);Tannerella(75);	*88*	*-0.3545*	*0.02295*	*0.3198*
	0026-Firmicutes(100);Clostridia(100);Clostridiales(100);Clostridiaceae 4(61);Geosporobacter(61);	*724*	*-0.3528*	*0.02365*	*0.3198*
	0119-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Porphyromonadaceae(73);Paludibacter(73);	*68*	*-0.3482*	*0.02569*	*0.3328*
	0071-Firmicutes(100);Bacilli(99);Lactobacillales(99);Carnobacteriaceae(99);Pisciglobus(97);	*173*	*-0.3413*	*0.02897*	*0.3604*
	0430-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Porphyromonadaceae(82);Paludibacter(82);	*11*	*-0.3311*	*0.03446*	*0.3759*
	0274-Bacteroidetes(100);Bacteroidia(100);Bacteroidales(100);Porphyromonadaceae(100);Parabacteroides(77);	*21*	*-0.3305*	*0.03483*	*0.3759*
	0023-Bacteroidetes(100);Bacteroidia(99);Bacteroidales(99);Porphyromonadaceae(99);Paludibacter(96);	*839*	*-0.3304*	*0.03485*	*0.3759*
	0450-Bacteroidetes(100);Bacteroidia(55);Bacteroidales(55);Rikenellaceae(55);Rikenella(55);	*11*	*-0.3265*	*0.03724*	*0.3782*
	0322-Bacteroidetes(100);Bacteroidia(77);Bacteroidales(77);Rikenellaceae(59);Rikenella(59);	*17*	*-0.3107*	*0.04804*	*0.3782*
	0005-Firmicutes(100);Erysipelotrichia(100);Erysipelotrichales(100);Erysipelotrichaceae(100);Turicibacter(100);	*5193*	*0.3724*	*0.01649*	*0.3198*
	0001-ProteoGammaproteoEnterobacteriales(100);Enterobacteriaceae(100);Salmonella(100);	*15356*	*0.5894*	*0.00005*	*0.0113*

References

1. Margulies, M., et al., *Genome sequencing in microfabricated high-density picolitre reactors.* Nature, 2005. **437**(7057): p. 376-380.

2. Shendure, J., et al., *Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome.* Science, 2005. **309**(5741): p. 1728-1732.

3. Drmanac, R., et al., *Human Genome Sequencing Using Unchained Base Reads on Self-Assembling DNA Nanoarrays.* Science, 2010. **327**(5961): p. 78-81.

4. Grice, E.A. and J.A. Segre, *The Human Microbiome: Our Second Genome.* Annual Review of Genomics and Human Genetics, 2012. **13**(1): p. 151-170.

5. O'Hara, A.M. and F. Shanahan, *The gut flora as a forgotten organ.* Embo Reports, 2006. **7**(7): p. 688-693.

6. Olsen, G.J. and C.R. Woese, *RIBOSOMAL-RNA - A KEY TO PHYLOGENY.* FASEB Journal, 1993. **7**(1): p. 113-123.

7. Sagan, L., *On the origin of mitosing cells.* J Theor Biol, 1967. **14**(3): p. 255-74.

8. Spang, A., et al., *Complex archaea that bridge the gap between prokaryotes and eukaryotes.* Nature, 2015. **521**(7551): p. 173-179.

9. Canfield, D.E., K.S. Habicht, and B. Thamdrup, *The Archean Sulfur Cycle and the Early History of Atmospheric Oxygen.* Science, 2000. **288**(5466): p. 658-661.

10. Konhauser, K.O., et al., *Aerobic bacterial pyrite oxidation and acid rock drainage during the Great Oxidation Event.* Nature, 2011. **478**(7369): p. 369-373.

11. Crowe, S.A., et al., *Atmospheric oxygenation three billion years ago.* Nature, 2013. **501**(7468): p. 535-538.

12. Ley, R.E., et al., *Worlds within worlds: evolution of the vertebrate gut microbiota.* Nat Rev Micro, 2008. **6**(10): p. 776-788.

13. Denef, V.J., R.S. Mueller, and J.F. Banfield, *AMD biofilms: using model communities to study microbial evolution and ecological complexity in nature.* ISME J, 2010. **4**(5): p. 599-610.

14. Jørgensen, B.B., M.F. Isaksen, and H.W. Jannasch, *Bacterial Sulfate Reduction Above 100°C in Deep-Sea Hydrothermal Vent Sediments.* Science, 1992. **258**(5089): p. 1756-1757.

15. Glud, R.N., et al., *High rates of microbial carbon turnover in sediments in the deepest oceanic trench on Earth.* Nature Geosci, 2013. **6**(4): p. 284-288.

16. Lin, L.-H., et al., *Long-Term Sustainability of a High-Energy, Low-Diversity Crustal Biome.* Science, 2006. **314**(5798): p. 479-482.

17. Flores, G.E., et al., *Microbial Biogeography of Public Restroom Surfaces.* PLoS One, 2011. **6**(11): p. e28132.

18. Kembel, S.W., et al., *Architectural design influences the diversity and structure of the built environment microbiome.* ISME J, 2012. **6**(8): p. 1469-1479.

19. Martiny, A.C., et al., *Long-term succession of structure and diversity of a biofilm formed in a model drinking water distribution system.* Appl. Environ. Microbiol., 2003. **69**: p. 6899-6907.

20. Smillie, C.S., et al., *Ecology drives a global network of gene exchange connecting the human microbiome.* Nature, 2011. **advance online publication**.

21. Mendes, R., et al., *Deciphering the Rhizosphere Microbiome for Disease-Suppressive Bacteria.* Science, 2011. **332**(6033): p. 1097-1100.

22. Endt, K., et al., *The Microbiota Mediates Pathogen Clearance from the Gut Lumen after Non-Typhoidal Salmonella Diarrhea.* PLoS Pathog, 2010. **6**(9): p. e1001097.

23. Callaway, R.M., et al., *Soil biota and exotic plant invasion.* Nature, 2004. **427**(6976): p. 731-733.

24. Rosenberg, E., et al., *The role of microorganisms in coral health, disease and evolution.* Nat Rev Micro, 2007. **5**(5): p. 355-362.

25. Tyson, G., et al., *Community structure and metabolism through reconstruction of microbial genomes from the environment.* Nature, 2004. **428**: p. 37 - 43.

26. Schloss, P.D. and J. Handelsman, *Toward a Census of Bacteria in Soil.* PLoS Comput Biol, 2006. **2**(7): p. e92.

27. Ettema, C.H. and D.A. Wardle, *Spatial soil ecology.* Trends in Ecology & Evolution, 2002. **17**(4): p. 177-183.

28. Hartmann, M., et al., *Resistance and resilience of the forest soil microbiome to logging-associated compaction.* ISME J, 2014. **8**(1): p. 226-244.

29. Philippot, L., et al., *Loss in microbial diversity affects nitrogen cycling in soil.* ISME J, 2013. **7**(8): p. 1609-1619.

30. Torsvik, V. and L. Øvreås, *Microbial diversity and function in soil: from genes to ecosystems.* Current Opinion in Microbiology, 2002. **5**(3): p. 240-245.

31. Giovannoni, S.J., et al., *Proteorhodopsin in the ubiquitous marine bacterium SAR11.* Nature, 2005. **438**(7064): p. 82-85.

32. Chisholm, S.W., et al., *A novel free-living prochlorophyte abundant in the oceanic euphotic zone.* Nature, 1988. **334**(6180): p. 340-343.

33. Moore, L.R., G. Rocap, and S.W. Chisholm, *Physiology and molecular phylogeny of coexisting Prochlorococcus ecotypes.* Nature, 1998. **393**(6684): p. 464-467.

34. Whitman, W.B., D.C. Coleman, and W.J. Wiebe, *Prokaryotes: The unseen majority.* Proceedings of the National Academy of Sciences, 1998. **95**(12): p. 6578-6583.

35. Hanson, C.A., et al., *Beyond biogeographic patterns: processes shaping the microbial landscape.* Nat Rev Micro, 2012. **10**(7): p. 497-506.

36. Vellend, M., *CONCEPTUAL SYNTHESIS IN COMMUNITY ECOLOGY.* Quarterly Review of Biology, 2010. **85**(2): p. 183-206.

37. Baas-Becking, L.G.M., *Geobiologie of Inleiding Tot de Milieukunde*. 1934.

38. Silverman, M. and M. Simon, *Flagellar rotation and the mechanism of bacterial motility.* Nature, 1974. **249**(5452): p. 73-74.

39. Merz, A.J., M. So, and M.P. Sheetz, *Pilus retraction powers bacterial twitching motility.* Nature, 2000. **407**(6800): p. 98-102.

40. Mignot, T., *The elusive engine in Myxococcus xanthus gliding motility.* Cellular and Molecular Life Sciences, 2007. **64**(21): p. 2733-2745.

41. Be'er, A., et al., *Paenibacillus dendritiformis Bacterial Colony Growth Depends on Surfactant but Not on Bacterial Motion.* Journal of Bacteriology, 2009. **191**(18): p. 5758-5764.

42. Grossart, H.-P., et al., *Bacteria dispersal by hitchhiking on zooplankton.* Proceedings of the National Academy of Sciences, 2010. **107**(26): p. 11959-11964.

43. DeLeon-Rodriguez, N., et al., *Microbiome of the upper troposphere: Species composition and prevalence, effects of tropical storms, and atmospheric implications.* Proceedings of the National Academy of Sciences, 2013. **110**(7): p. 2575-2580.

44. Gibbons, S.M., et al., *Evidence for a persistent microbial seed bank throughout the global ocean.* Proceedings of the National Academy of Sciences, 2013.

45. Roberts, M.S. and F.M. Cohan, *Recombination and Migration Rates in Natural Populations of Bacillus subtilis and Bacillus mojavensis.* Evolution, 1995. **49**(6): p. 1081-1094.

46. Ramirez, K.S., et al., *Biogeographic patterns in below-ground diversity in New York City's Central Park are similar to those observed globally.* Proceedings of the Royal Society B: Biological Sciences, 2014. **281**(1795).

47. Fierer, N., et al., *Microbes do not follow the elevational diversity patterns of plants and animals.* Ecology, 2010. **92**(4): p. 797-804.

48. Bryant, J.A., et al., *Microbes on mountainsides: Contrasting elevational patterns of bacterial and plant diversity.* Proceedings of the National Academy of Sciences, 2008. **105**(Supplement 1): p. 11505-11511.

49. Nemergut, D.R., et al., *Microbial community succession in an unvegetated, recently deglaciated soil.* Microb Ecol, 2007. **53**(1): p. 110-22.

50. Sigler, W.V. and J. Zeyer, *Microbial Diversity and Activity along the Forefields of Two Receding Glaciers.* Microbial Ecology, 2002. **43**(4): p. 397-407.

51. Fierer, N., et al., *Changes through time: integrating microorganisms into the study of succession.* Research in Microbiology, 2010. **161**(8): p. 635-642.

52. Bell, T., *Experimental tests of the bacterial distance-decay relationship.* ISME J, 2010.

53. Redford, A.J., et al., *The ecology of the phyllosphere: geographic and phylogenetic variability in the distribution of bacteria on tree leaves.* Environmental Microbiology, 2010. **12**(11): p. 2885-2893.

54. King, A.J., et al., *Biogeography and habitat modelling of high-alpine bacteria.* Nat Commun, 2010. **1**(5): p. 53.

55. Fierer, N. and R.B. Jackson, *The diversity and biogeography of soil bacterial communities.* Proc Natl Acad Sci, 2006. **103**(3): p. 626-631.

56. Martiny, J.B.H., et al., *Microbial biogeography: putting microorganisms on the map.* Nat Rev Micro, 2006. **4**(2): p. 102-112.

57. Fierer, N., et al., *Reconstructing the Microbial Diversity and Function of Pre-Agricultural Tallgrass Prairie Soils in the United States.* Science, 2013. **342**(6158): p. 621-624.

58. Pommier, T., et al., *Global patterns of diversity and community structure in marine bacterioplankton.* Molecular Ecology, 2007. **16**(4): p. 867-880.

59. Cho, J.C. and J.M. Tiedje, *Biogeography and degree of endemicity of fluorescent Pseudomonas strains in soil.* Appl. Environ. Microbiol., 2000. **66**: p. 5448-5456.

60. Oakley, B.B., et al., *Evolutionary divergence and biogeography of sympatric niche-differentiated bacterial populations.* ISME J, 2010. **4**(4): p. 488-497.

61. Dini-Andreote, F., et al., *Dynamics of bacterial community succession in a salt marsh chronosequence: evidences for temporal niche partitioning.* ISME J, 2014. **8**(10): p. 1989-2001.

62. Gilbert, J.A., et al., *Defining seasonal marine microbial community dynamics.* ISME J, 2012. **6**(2): p. 298-308.

63. Zhang, N., et al., *Soil microbial community changes and their linkages with ecosystem carbon exchange under asymmetrically diurnal warming.* Soil Biology and Biochemistry, 2011. **43**(10): p. 2053-2059.

64. Lipson, D.A., *Relationships between temperature responses and bacterial community structure along seasonal and altitudinal gradients*. Vol. 59. 2007. 418-427.

65. Schadt, C.W., et al., *Seasonal Dynamics of Previously Unknown Fungal Lineages in Tundra Soils.* Science, 2003. **301**(5638): p. 1359-1361.

66. Faith, J.J., et al., *The Long-Term Stability of the Human Gut Microbiota.* Science, 2013. **341**(6141): p. 1237439.

67. Verleyen, E., et al., *The importance of dispersal related and local factors in shaping the taxonomic structure of diatom metacommunities.* Oikos, 2009. **118**(8): p. 1239-1249.

68. Kort, R., et al., *Shaping the oral microbiota through intimate kissing.* Microbiome, 2014. **2**(1): p. 41.

69. Seedorf, H., et al., *Bacteria from Diverse Habitats Colonize and Compete in the Mouse Gut.* Cell, 2014. **159**(2): p. 253-266.

70. David, L.A., et al., *Diet rapidly and reproducibly alters the human gut microbiome.* Nature, 2014. **505**(7484): p. 559-563.

71. Muegge, B.D., et al., *Diet Drives Convergence in Gut Microbiome Functions Across Mammalian Phylogeny and Within Humans.* Science, 2011. **332**(6032): p. 970-974.

72. Leamy, L., et al., *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, 2014. **15**(12): p. 552.

73. McKnite, A.M., et al., *Murine Gut Microbiota Is Defined by Host Genetics and Modulates Variation of Metabolic Traits.* PLoS One, 2012. **7**(6).

74. Benson, A.K., et al., *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, 2010. **107**(44): p. 18933-18938.

75. Lankau, E.W., P.-Y. Hong, and R.I. Mackie, *Ecological drift and local exposures drive enteric bacterial community differences within species of Galápagos iguanas.* Molecular Ecology, 2012. **21**(7): p. 1779-1788.

76. Peiffer, J.A., et al., *Diversity and heritability of the maize rhizosphere microbiome under field conditions.* Proceedings of the National Academy of Sciences, 2013. **110**(16): p. 6548-6553.

77. Horton, M.W., et al., *Genome-wide association study of Arabidopsis thaliana leaf microbial community.* Nat Commun, 2014. **5**: p. 5320.

78. Lundberg, D.S., et al., *Defining the core Arabidopsis thaliana root microbiome.* Nature, 2012. **488**(7409): p. 86-90.

79. Bulgarelli, D., et al., *Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota.* Nature, 2012. **488**(7409): p. 91-95.

80. Backhed, F., et al., *Host-bacterial mutualism in the human intestine.* Science, 2005. **307**(5717): p. 1915-1920.

81. Whitman, W.B., D.C. Coleman, and W.J. Wiebe, *Prokaryotes: The unseen majority.* Proceedings of the National Academy of Sciences of the United States of America, 1998. **95**(12): p. 6578-6583.

82. Dethlefsen, L., et al., *Assembly of the human intestinal microbiota.* Trends in Ecology & Evolution, 2006. **21**(9): p. 517-523.

83. Shade, A. and J. Handelsman, *Beyond the Venn diagram: the hunt for a core microbiome.* Environmental Microbiology, 2011: p. no-no.

84. Turnbaugh, P.J., et al., *A core gut microbiome in obese and lean twins.* Nature, 2009. **457**(7228): p. 480-U7.

85. Qin, J., et al., *A human gut microbial gene catalogue established by metagenomic sequencing.* Nature, 2010. **464**(7285): p. 59-65.

86. Consortium, T.H.M.P., *Structure, function and diversity of the healthy human microbiome.* Nature, 2012. **486**(7402): p. 207-214.

87. Li, J., et al., *An integrated catalog of reference genes in the human gut microbiome.* Nat Biotech, 2014. **32**(8): p. 834-841.

88. Flint, H.J., et al., *Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis.* Nat Rev Micro, 2008. **6**(2): p. 121-131.

89. Douglas, A.E., *Nutritional Interactions in Insect-Microbial Symbioses: Aphids and Their Symbiotic Bacteria Buchnera.* Annual Review of Entomology, 1998. **43**(1): p. 17-37.

90. Salem, H., et al., *Vitamin supplementation by gut symbionts ensures metabolic homeostasis in an insect host.* Proceedings of the Royal Society B: Biological Sciences, 2014. **281**(1796).

91. Chung, H., et al., *Gut Immune Maturation Depends on Colonization with a Host-Specific Microbiota.* Cell, 2012. **149**(7): p. 1578-1593.

92. Oliver, K.M., et al., *Facultative bacterial symbionts in aphids confer resistance to parasitic wasps.* Proceedings of the National Academy of Sciences, 2003. **100**(4): p. 1803-1807.

93. Koehler, S. and M. Kaltenpoth, *Maternal and Environmental Effects on Symbiont-Mediated Antimicrobial Defense.* Journal of Chemical Ecology, 2013. **39**(7): p. 978-988.

94. Bright, M. and S. Bulgheresi, *A complex journey: transmission of microbial symbionts.* Nat Rev Micro, 2010. **8**(3): p. 218-230.

95. Moran, N.A., J.P. McCutcheon, and A. Nakabachi, *Genomics and Evolution of Heritable Bacterial Symbionts*, in *Annual Review of Genetics*. 2008. p. 165-190.

96. Zilber-Rosenberg, I. and E. Rosenberg, *Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution.* Fems Microbiology Reviews, 2008. **32**(5): p. 723-735.

97. Kaltenpoth, M., et al., *Partner choice and fidelity stabilize coevolution in a Cretaceous-age defensive symbiosis.* Proceedings of the National Academy of Sciences, 2014. **111**(17): p. 6359-6364.

98. Linnenbrink, M., et al., *The role of biogeography in shaping diversity of the intestinal microbiota in house mice.* Molecular Ecology, 2013. **22**(7): p. 1904-16.

99. Ochman, H., et al., *Evolutionary Relationships of Wild Hominids Recapitulated by Gut Microbial Communities.* PLoS Biol, 2010. **8**(11): p. e1000546.

100. Ley, R.E., et al., *Evolution of mammals and their gut microbes.* Science, 2008. **320**(5883): p. 1647-1651.

101. Xu, J., et al., *Evolution of symbiotic bacteria in the distal human intestine.* Plos Biology, 2007. **5**(7): p. 1574-1586.

102. Hansen, E.E., et al., *Pan-genome of the dominant human gut-associated archaeon, Methanobrevibacter smithii, studied in twins.* Proceedings of the National Academy of Sciences, 2011.

103. Lozupone, C., et al., *Identifying genomic and metabolic features that can underlie early successional and opportunistic lifestyles of human gut symbionts.* Genome Research, 2012. **22**(10): p. 1974-1984.

104. Yatsunenko, T., et al., *Human gut microbiome viewed across age and geography.* Nature, 2012. **486**(7402): p. 222-7.

105. Brucker, R.M. and S.R. Bordenstein, *The Hologenomic Basis of Speciation: Gut Bacteria Cause Hybrid Lethality in the Genus Nasonia.* Science, 2013. **341**(6146): p. 667-9.

106. Miller, W.J., L. Ehrman, and D. Schneider, *Infectious Speciation Revisited: Impact of Symbiont-Depletion on Female Fitness and Mating Behavior of Drosophila paulistorum.* PLoS Pathog, 2010. **6**(12): p. e1001214.

107. Brucker, R.M. and S.R. Bordenstein, *Response to Comment on "The hologenomic basis of speciation: Gut bacteria cause hybrid lethality in the genus Nasonia".* Science, 2014. **345**(6200): p. 1011.

108. Chandler, J.A. and M. Turelli, *Comment on "The hologenomic basis of speciation: Gut bacteria cause hybrid lethality in the genus Nasonia".* Science, 2014. **345**(6200): p. 1011.

109. Fujishige, N.A., et al., *Rhizobium common nod genes are required for biofilm formation.* Molecular Microbiology, 2008. **67**(3): p. 504-515.

110. Nyholm, S.V. and M. McFall-Ngai, *The winnowing: establishing the squid-vibrio symbiosis.* Nat Rev Micro, 2004. **2**(8): p. 632-642.

111. Gage, D.J., *Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation of temperate legumes.* Microbiology and Molecular Biology Reviews, 2004. **68**(2): p. 280-+.

112. Alegado, R.A., et al., *A bacterial sulfonolipid triggers multicellular development in the closest living relatives of animals.* Elife, 2012. **1**.

113. Nichols, S.A., M.J. Dayel, and N. King, *Genomic, phylogenetic, and cell biological insights into metazoan origins*. Animal evolution: genomes, fossils, and trees., ed. M.J. Telford and D.T.J. Littlewood. 2009. 24-32.

114. Funkhouser, L.J. and S.R. Bordenstein, *Mom Knows Best: The Universality of Maternal Microbial Transmission.* PLoS Biol, 2013. **11**(8): p. e1001631.

115. Aagaard, K., et al., *The Placenta Harbors a Unique Microbiome.* Science Translational Medicine, 2014. **6**(237): p. 237ra65.

116. Franzenburg, S., et al., *Bacterial colonization of Hydra hatchlings follows a robust temporal pattern.* ISME J, 2013.

117. Koenig, J.E., et al., *Succession of microbial consortia in the developing infant gut microbiome.* Proceedings of the National Academy of Sciences, 2011. **108**(Supplement 1): p. 4578-4585.

118. Cox, Laura M., et al., *Altering the Intestinal Microbiota during a Critical Developmental Window Has Lasting Metabolic Consequences.* Cell, 2014. **158**(4): p. 705-721.

119. La Rosa, P.S., et al., *Patterned progression of bacterial populations in the premature infant gut.* Proceedings of the National Academy of Sciences, 2014. **111**(34): p. 12522-12527.

120. Adlerberth, I., *Factors Influencing the Establishment of the Intestinal Microbiota in Infancy*, in *Personalized Nutrition for the Diverse Needs of Infants and Children*, D.M. Bier, J.B. German, and B. Lonnerdal, Editors. 2008, Karger: Basel. p. 13-33.

121. Dominguez-Bello, M.G., et al., *Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns.* Proceedings of the National Academy of Sciences, 2010. **107**(26): p. 11971-11975.

122. Fallani, M., et al., *Intestinal Microbiota of 6-week-old Infants Across Europe: Geographic Influence Beyond Delivery Mode, Breast-feeding, and Antibiotics.* Journal of Pediatric Gastroenterology and Nutrition, 2010. **51**(1): p. 77-84 10.1097/MPG.0b013e3181d1b11e.

123. Cash, H.L., et al., *Symbiotic bacteria direct expression of an intestinal bactericidal lectin.* Science, 2006. **313**(5790): p. 1126-1130.

124. Goodrich, Julia K., et al., *Human Genetics Shape the Gut Microbiome.* Cell, 2014. **159**(4): p. 789-799.

125. Zivkovic, A.M., et al., *Human milk glycobiome and its impact on the infant gastrointestinal microbiota.* Proceedings of the National Academy of Sciences, 2011. **108**(Supplement 1): p. 4653-4658.

126. Rogier, E.W., et al., *Secretory antibodies in breast milk promote long-term intestinal homeostasis by regulating the gut microbiota and host gene expression.* Proceedings of the National Academy of Sciences, 2014.

127. Linden, S., et al., *Role of ABO secretor status in mucosal innate immunity and H. pylori infection.* PLoS Pathogens, 2008. **4**(1).

128. Sjogren, Y.M., et al., *Influence of early gut microbiota on the maturation of childhood mucosal and systemic immune responses.* Clinical and Experimental Allergy, 2009. **39**(12): p. 1842-1851.

129. Presley, L.L., et al., *Bacteria Associated with Immunoregulatory Cells in Mice.* Applied and Environmental Microbiology, 2010. **76**(3): p. 936-941.

130. Vijay-Kumar, M., et al., *Metabolic Syndrome and Altered Gut Microbiota in Mice Lacking Toll-Like Receptor 5.* Science, 2010. **328**(5975): p. 228-231.

131. Wu, H.-J., et al., *Gut-Residing Segmented Filamentous Bacteria Drive Autoimmune Arthritis via T Helper 17 Cells.* Immunity, 2010. **32**(6): p. 815-827.

132. Kanther, M., et al., *Microbial Colonization Induces Dynamic Temporal and Spatial Patterns of NF-κB Activation in the Zebrafish Digestive Tract.* Gastroenterology, 2011. **141**(1): p. 197-207.

133. Kofoed, E.M. and R.E. Vance, *Innate immune recognition of bacterial ligands by NAIPs determines inflammasome specificity.* Nature, 2011. **advance online publication**.

134. Lathrop, S.K., et al., *Peripheral education of the immune system by colonic commensal microbiota.* Nature, 2011. **478**(7368): p. 250-254.

135. Rehman, A., et al., *Nod2 is essential for temporal development of intestinal microbial communities.* Gut, 2011. **60**(10): p. 1354-62.

136. Round, J.L., et al., *The Toll-Like Receptor 2 Pathway Establishes Colonization by a Commensal of the Human Microbiota.* Science, 2011. **332**(6032): p. 974-977.

137. Fernandez, L., et al., *The human milk microbiota: Origin and potential roles in health and disease.* Pharmacological Research, 2013. **69**(1): p. 1-10.

138. Stearns, J.C., et al., *Bacterial biogeography of the human digestive tract.* Sci. Rep., 2011. **1**.

139. Grice, E.A., et al., *Topographical and Temporal Diversity of the Human Skin Microbiome.* Science, 2009. **324**(5931): p. 1190-1192.

140. Gajer, P., et al., *Temporal Dynamics of the Human Vaginal Microbiota.* Science Translational Medicine, 2012. **4**(132): p. 132ra52.

141. Costello, E.K., et al., *Bacterial Community Variation in Human Body Habitats Across Space and Time.* Science, 2009. **326**(5960): p. 1694-1697.

142. Dethlefsen, L., et al., *The Pervasive Effects of an Antibiotic on the Human Gut Microbiota, as Revealed by Deep 16S rRNA Sequencing.* Plos Biology, 2008. **6**(11): p. e280.

143. Fierer, N., et al., *The influence of sex, handedness, and washing on the diversity of hand surface bacteria.* Proceedings of the National Academy of Sciences, 2008. **105**(46): p. 17994-17999.

144. Lang, J.M., J.A. Eisen, and A.M. Zivkovic, *The microbes we eat: abundance and taxonomy of microbes consumed in a day's worth of meals for three diet types.* PeerJ, 2014. **2**: p. e659.

145. Venkataraman, A., et al., *Application of a Neutral Community Model To Assess Structuring of the Human Lung Microbiome.* mBio, 2015. **6**(1).

146. Hubbell, S.P., *The unified neutral theory of biodiversity and biogeography.* Monographs in Population Biology, 2001. **32**: p. i-xiv, 1-375.

147. Whiteson, K.L., et al., *The Upper Respiratory Tract as a Microbial Source for Pulmonary Infections in Cystic Fibrosis: Parallels from Island Biogeography.* American Journal of Respiratory and Critical Care Medicine, 2014. **189**(11): p. 1309-1315.

148. Bell, T., *Larger islands house more bacterial taxa.* Science, 2005. **308**: p. 1884.

149. Simberloff, D.S. and E.O. Wilson, *Experimental Zoogeography of Islands: The Colonization of Empty Islands.* Ecology, 1969. **50**(2): p. 278-296.

150. Bell, T., et al., *Larger Islands House More Bacterial Taxa.* Science, 2005. **308**(5730): p. 1884.

151. Kier, G., et al., *A global assessment of endemism and species richness across island and mainland regions.* Proceedings of the National Academy of Sciences, 2009. **106**(23): p. 9322-9327.

152. Walter, J. and R.E. Ley, *The Human Gut Microbiome: Ecology and Recent Evolutionary Changes.* Annual Review of Microbiology, 2011. **65**(1): p. null.

153. Mahler, D.L., et al., *Exceptional Convergence on the Macroevolutionary Landscape in Island Lizard Radiations.* Science, 2013. **341**(6143): p. 292-295.

154. Koren, O., et al., *A Guide to Enterotypes across the Human Body: Meta-Analysis of Microbial Community Structures in Human Microbiome Datasets.* PLoS Comput Biol, 2013. **9**(1): p. e1002863.

155. Wang, J., et al., *Dietary history contributes to enterotype-like clustering and functional metagenomic content in the intestinal microbiome of wild mice.* Proceedings of the National Academy of Sciences, 2014. **111**(26): p. E2703-E2710.

156. Arumugam, M., et al., *Enterotypes of the human gut microbiome.* Nature, 2011. **473**(7346): p. 174-80.

157. Moeller, A.H., et al., *Chimpanzees and humans harbour compositionally similar gut enterotypes.* Nat Commun, 2012. **3**: p. 1179.

158. Turnbaugh, P., et al., *An obesity-associated gut microbiome with increased capacity for energy harvest.* Nature, 2006. **444**(7122): p. 1027 - 1031.

159. Ding, T. and P.D. Schloss, *Dynamics and associations of microbial community types across the human body.* Nature, 2014. **509**(7500): p. 357-60.

160. Lahti, L., et al., *Tipping elements in the human intestinal ecosystem.* Nat Commun, 2014. **5**.

161. Drouilhet, L., et al., *The Highly Prolific Phenotype of Lacaune Sheep Is Associated with an Ectopic Expression of the B4GALNT2 Gene within the Ovary.* PLoS Genet, 2013. **9**(9): p. e1003809.

162. Dharmesh, S.M. and J.U. Baenziger, *Estrogen modulates expression of the glycosyltransferases that synthesize sulfated oligosaccharides on lutropin.* Proceedings of the National Academy of Sciences, 1993. **90**(23): p. 11127-11131.

163. Domino, S.E. and E.A. Hurd, *LacZ expression in Fut2-LacZ reporter mice reveals estrogen-regulated endocervical glandular expression during estrous cycle, hormone replacement, and pregnancy.* Glycobiology, 2004. **14**(2): p. 169-175.

164. Varki, A. and J.B. Lowe, *Biological Roles of Glycans.* 2009.

165. Bishop, J.R. and P. Gagneux, *Evolution of carbohydrate antigens—microbial forces shaping host glycomes?* Glycobiology, 2007. **17**(5): p. 23R-34R.

166. Freeze, H.H. and V. Westphal, *Balancing N-linked glycosylation to avoid disease.* Biochimie, 2001. **83**(8): p. 791-799.

167. Lewis, A.L., et al., *Innovations in host and microbial sialic acid biosynthesis revealed by phylogenomic prediction of nonulosonic acid structure.* Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(32): p. 13552-13557.

168. Breton, C., R. Oriol, and A. Imberty, *Conserved structural features in eukaryotic and prokaryotic fucosyltransferases.* Glycobiology, 1998. **8**(1): p. 87-94.

169. Javaud, C., et al., *The Fucosyltransferase Gene Family: An Amazing Summary of the Underlying Mechanisms of Gene Evolution.* Genetica, 2003. **118**(2-3): p. 157-170.

170. Oriol, R., et al., *Divergent evolution of fucosyltransferase genes from vertebrates, invertebrates, and bacteria.* Glycobiology, 1999. **9**(4): p. 323-334.

171. Cummings, R.D., *The repertoire of glycan determinants in the human glycome.* Mol Biosyst, 2009. **5**(10): p. 1087-104.

172. Dall'Olio, F., et al., *The expanding roles of the Sda/Cad carbohydrate antigen and its cognate glycosyltransferase B4GALNT2.* Biochimica et Biophysica Acta (BBA) - General Subjects, 2014. **1840**(1): p. 443-453.

173. Groux-Degroote, S., et al., *B4GALNT2 gene expression controls the biosynthesis of Sda and sialyl Lewis X antigens in healthy and cancer human gastrointestinal tract.* The International Journal of Biochemistry & Cell Biology, 2014. **53**(0): p. 442-449.

174. Watkins, W.M., et al., *Regulation of expression of carbohydrate blood group antigens.* Biochimie, 1988. **70**(11): p. 1597-1611.

175. Van Valen, L., *A new evolutionary law.* Evol Theor, 1973. **1**: p. 1 - 30.

176. Varki, A., *Nothing in Glycobiology Makes Sense, except in the Light of Evolution.* Cell, 2006. **126**(5): p. 841-845.

177. Varki, A., *Evolutionary forces shaping the Golgi glycosylation machinery: why cell surface glycans are universal to living cells.* Cold Spring Harb Perspect Biol, 2011. **3**(6).

178. Lau, K.S., et al., *Complex N-Glycan Number and Degree of Branching Cooperate to Regulate Cell Proliferation and Differentiation.* Cell, 2007. **129**(1): p. 123-134.

179. Sharon, N. and H. Lis, *History of lectins: from hemagglutinins to biological recognition molecules.* Glycobiology, 2004. **14**(11): p. 53R-62R.

180. Taylor, M.E. and K. Drickamer, *Structural insights into what glycan arrays tell us about how glycan-binding proteins interact with their ligands.* Glycobiology, 2009. **19**(11): p. 1155-1162.

181. Schwarzkopf, M., et al., *Sialylation is essential for early development in mice.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(8): p. 5267-5270.

182. Esko, J.D. and S.B. Selleck, *Order out of chaos: Assembly of ligand binding sites in heparan sulfate.* Annual Review of Biochemistry, 2002. **71**: p. 435-471.

183. Marek, K.W., I.K. Vijay, and J.D. Marth, *A recessive deletion in the GlcNAc-1-phosphotransferase gene results in peri-implantation embryonic lethality.* Glycobiology, 1999. **9**(11): p. 1263-1271.

184. Shi, S. and P. Stanley, *Protein O-fucosyltransferase 1 is an essential component of Notch signaling pathways.* Proceedings of the National Academy of Sciences, 2003. **100**(9): p. 5234-5239.

185. Martinko, J., et al., *Primate ABO glycosyltransferases: Evidence for trans-species evolution.* Immunogenetics, 1993. **37**(4): p. 274-278.

186. Ségurel, L., et al., *The ABO blood group is a trans-species polymorphism in primates.* Proceedings of the National Academy of Sciences, 2012. **109**(45): p. 18493-18498.

187. Calafell, F., et al., *Evolutionary dynamics of the human ABO gene.* Human Genetics, 2008. **124**(2): p. 123-135.

188. Yamamoto, F., et al., *An integrative evolution theory of histo-blood group ABO and related genes.* Scientific Reports, 2014. **4**.

189. Saitou, N. and F. Yamamoto, *Evolution of primate ABO blood group genes and their homologous genes.* Molecular Biology and Evolution, 1997. **14**(4): p. 399-411.

190. O'hUigin, C., A. Sato, and J. Klein, *Evidence for convergent evolution of A and B blood group antigens in primates.* Human Genetics, 1997. **101**(2): p. 141-148.

191. Diamond, D.C., et al., *Sequence comparison of baboon ABO histo-blood group alleles: Lesions found in O alleles differ between human and baboon.* Blood Cells Molecules and Diseases, 1997. **23**(13): p. 242-251.

192. Ségurel, L., Z. Gao, and M. Przeworski, *Ancestry runs deeper than blood: The evolutionary history of ABO points to cryptic variation of functional importance.* Bioessays, 2013. **35**(10): p. 862-867.

193. Landsteiner, K., *Über Agglutinationserscheinungen normalen menschlichen Blutes.* Wiener Klinische Wochenschrift, 1901. **14**: p. 1132- 1134.

194. Landsteiner, K., *Zur Kenntnis der antifermentativen, lytischen und agglutinierenden Wirkungen des Blutserums und der Lymphe.* Zbl Bakt, 1900. **27**(10): p. 357-362.

195. Leffler, E.M., et al., *Multiple Instances of Ancient Balancing Selection Shared Between Humans and Chimpanzees.* Science, 2013. **339**(6127): p. 1578-1582.

196. Kobata, A., E.F. Grollman, and V. Ginsburg, *An enzymic basis for blood type A in humans.* Archives of Biochemistry and Biophysics, 1968. **124**(1-3): p. 609-&.

197. Ginsburg, V., A. Kobata, and E.F. Grollman, *Studies on Enzymic Basis for Blood Types in Man.* Abstracts of Papers of the American Chemical Society, 1969(SEP): p. CA45-&.

198. Race, C., D. Ziderman, and W.M. Watkins, *An Alpha-D-Galactosyltransferase Assoicated with Blood-Group B Character.* Biochemical Journal, 1968. **107**(5): p. 733-&.

199. Aminoff, D. and W.T.J. Morgan, *Hexosamine components of the human blood group substances.* Nature, 1948. **162**(4119): p. 579-580.

200. Greenwell, P., *Blood group antigens: Molecules seeking a function?* Glycoconjugate Journal, 1997. **14**(2): p. 159-173.

201. Pendu, J.L., *A HYPOTHESIS ON THE DUAL SIGNIFICANCE OF ABH, LEWIS AND RELATED ANTIGENS.* International Journal of Immunogenetics, 1989. **16**(1): p. 53-61.

202. Seymour, R.M., et al., *Evolution of the human ABO polymorphism by two complementary selective pressures.* Proceedings of the Royal Society B-Biological Sciences, 2004. **271**(1543): p. 1065-1072.

203. Marionneau, S., et al., *ABH and Lewis histo-blood group antigens, a model for the meaning of oligosaccharide diversity in the face of a changing world.* Biochimie, 2001. **83**(7): p. 565-573.

204. Berger, S.A., N.A. Young, and S.C. Edberg, *RELATIONSHIP BETWEEN INFECTIOUS-DISEASES AND HUMAN-BLOOD TYPE.* European Journal of Clinical Microbiology & Infectious Diseases, 1989. **8**(8): p. 681-689.

205. Winstone, N.E., A.J. Henderson, and B.N. Brooke, *Blood-Groups and Secretor Status in Ulcerative Colitis.* Lancet, 1960. **2**(JUL9): p. 64-65.

206. Smith, R.S. and S.C. Truelove, *BLOOD GROUPS AND SECRETOR STATUS IN ULCERATIVE COLITIS.* British Medical Journal, 1961. **1**(522): p. 870-&.

207. Evans, D.A.P., et al., *Influence of ABO blood groups and secretor status on bleeding and on perforation of duodenal ulcer.* Gut, 1968. **9**(3): p. 319-322.

208. Morrow, A.L., et al., *Fucosyltransferase 2 Non-Secretor and Low Secretor Status Predicts Severe Outcomes in Premature Infants.* The Journal of Pediatrics, 2011. **158**(5): p. 745-751.

209. McGovern, D.P.B., et al., *Fucosyltransferase 2 (FUT2) non-secretor status is associated with Crohn's disease.* Hum. Mol. Genet., 2010: p. 1-9.

210. Miyoshi, J., et al., *Ectopic expression of blood type antigens in inflamed mucosa with higher incidence of FUT2 secretor status in colonic Crohn's disease.* Journal of Gastroenterology, 2011. **46**(9): p. 1056-1063.

211. Pavillard, E.R.J., H. Stegemann, and D. Rowley, *OPSONIC POWER OF HUMAN ANTI-B SERUM FOR BLOOD GROUP B-ACTIVE BACTERIA.* Aust J Exp Biol Med, 1964. **42**(1): p. 62-73.

212. Yang, N. and B. Boettcher, *DEVELOPMENT OF HUMAN ABO BLOOD GROUP-A ANTIGEN ON ESCHERICHIA-COLI Y1089 AND Y1090.* Immunology and Cell Biology, 1992. **70**: p. 411-416.

213. Springer, G.F. and R.E. Horton, *ERYTHROCYTE SENSITIZATION BY BLOOD GROUP-SPECIFIC BACTERIAL ANTIGENS.* Journal of General Physiology, 1964. **47**(6): p. 1229-&.

214. Andersson, M., et al., *STRUCTURAL STUDIES OF THE O-ANTIGENIC POLYSACCHARIDE OF ESCHERICHIA-COLI O86, WHICH POSSESSES BLOOD-GROUP-B ACTIVITY.* Carbohydrate Research, 1989. **185**(2): p. 211-223.

215. Springer, G.F., R.E. Horton, and M. Forbes, *ORIGIN OF ANTI-HUMAN BLOOD GROUP-B AGGLUTININS IN WHITE LEGHORN CHICKS.* Journal of Experimental Medicine, 1959. **110**(2): p. 221-244.

216. Springer, G.F. and R.E. Horton, *BLOOD GROUP ISOANTIBODY STIMULATION IN MAN BY FEEDING BLOOD GROUP-ACTIVE BACTERIA.* Journal of Clinical Investigation, 1969. **48**(7): p. 1280-&.

217. Galili, U., et al., *Interaction between human natural anti-alpha-galactosyl immunoglobulin G and bacteria of the human flora.* Infection and Immunity, 1988. **56**(7): p. 1730-1737.

218. Yilmaz, B., et al., *Gut Microbiota Elicits a Protective Immune Response against Malaria Transmission.* Cell, 2014. **159**(6): p. 1277-1289.

219. Ghaderi, D., et al., *Sexual selection by female immunity against paternal antigens can fix loss of function alleles.* Proceedings of the National Academy of Sciences of the United States of America, 2011. **108**(43): p. 17743-17748.

220. Liu, Y., et al., *Extensive polymorphism of the FUT2 gene in an African (Xhosa) population of South Africa.* Human Genetics, 1998. **103**(2): p. 204-210.

221. Varki, A., H.H. Freeze, and P. Gagneux, *Evolution of Glycan Diversity*, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2009, The Consortium of Glycobiology Editors, La Jolla, California: Cold Spring Harbor NY.

222. Varki, A., *Uniquely human evolution of sialic acid genetics and biology.* Proceedings of the National Academy of Sciences, 2010. **107**(Supplement 2): p. 8939-8946.

223. Wang, X., et al., *Specific inactivation of two immunomodulatory SIGLEC genes during human evolution.* Proceedings of the National Academy of Sciences, 2012. **109**(25): p. 9935-9940.

224. Mengerink, K.J. and V.D. Vacquier, *Glycobiology of sperm-egg interactions in deuterostomes.* Glycobiology, 2001. **11**(4): p. 37R-43R.

225. Zerfaoui, M., et al., *α(1,2)-Fucosylation prevents sialyl Lewis x expression and E-selectin-mediated adhesion of fucosyltransferase VII-transfected cells.* European Journal of Biochemistry, 2000. **267**(1): p. 53-61.

226. Laubli, H., et al., *L-selectin facilitation of metastasis involves temporal induction of Fut7-dependent ligands at sites of tumor cell arrest.* Cancer Res, 2006. **66**(3): p. 1536-42.

227. Aplin, J.D. and C.J.P. Jones, *Fucose, placental evolution and the glycocode.* Glycobiology, 2011.

228. Gloria-Bottini, F., et al., *ABH secretor genetic polymorphism: evidence of intrauterine selection.* European Journal of Obstetrics & Gynecology and Reproductive Biology, 2011. **154**(1): p. 20-23.

229. Li, P.-T., et al., *Localization of B4GALNT2 and its role in mouse embryo attachment.* Fertility and Sterility, 2012. **97**(5): p. 1206-1212.e3.

230. Ruiz-Palacios, G.M., et al., *Campylobacter jejuni Binds Intestinal H(O) Antigen (Fucα1, 2Galβ1, 4GlcNAc), and Fucosyloligosaccharides of Human Milk Inhibit Its Binding and Infection.* Journal of Biological Chemistry, 2003. **278**(16): p. 14112-14120.

231. Voynow, J.A., S.J. Gendler, and M.C. Rose, *Regulation of mucin genes in chronic inflammatory airway diseases.* American Journal of Respiratory Cell and Molecular Biology, 2006. **34**(6): p. 661-665.

232. Baumann, H. and J. Gauldie, *The acute phase response.* Immunology Today, 1994. **15**(2): p. 74-80.

233. Yamamoto, F.-i., et al., *Molecular genetic basis of the histo-blood group ABO system.* Nature, 1990. **345**(6272): p. 229-233.

234. Rossez, Y., et al., *Almost all human gastric mucin O-glycans harbor blood group A, B or H antigens and are potential binding sites for Helicobacter pylori.* Glycobiology, 2012. **22**(9): p. 1193-1206.

235. Robbe, C., et al., *Structural diversity and specific distribution of O-glycans in normal human mucins along the intestinal tract.* Biochem. J., 2004. **384**(2): p. 307-316.

236. Robbe, C., et al., *Evidence of Regio-specific Glycosylation in Human Intestinal Mucins.* Journal of Biological Chemistry, 2003. **278**(47): p. 46337-46348.

237. Newburg, D.S., et al., *Fucosylated Oligosaccharides of Human Milk Protect Suckling Mice from Heat-Stabile Enterotoxin of Escherichia coli.* The Journal of Infectious Diseases, 1990. **162**(5): p. 1075-1080.

238. Morrow, A.L., et al., *Human-Milk Glycans That Inhibit Pathogen Binding Protect Breast-feeding Infants against Infectious Diarrhea.* J. Nutr., 2005. **135**(5): p. 1304-1307.

239. Kelly, R.J., et al., *Sequence and Expression of a Candidate for the Human Secretor Blood Group α(1,2)Fucosyltransferase Gene (FUT2).* Journal of Biological Chemistry, 1995. **270**(9): p. 4640-4649.

240. Iwamori, M. and S.E. Domino, *Tissue-specific loss of fucosylated glycolipids in mice with targeted deletion of alpha(1,2)fucosyltransferase genes.* Biochemical Journal, 2004. **380**: p. 75-81.

241. Reguigne-Arnould, I., et al., *Relative positions of two clusters of human α-L-fucosyltransferases in 19q (FUT1–FUT2) and 19p (FUT6–FUT3–FUT5) within the microsatellite genetic map of chromosome 19.* Cytogenetic and Genome Research, 1995. **71**(2): p. 158-162.

242. Martinez-Duncker, I., et al., *A new superfamily of protein-O-fucosyltransferases, α2-fucosyltransferases, and α6-fucosyltransferases: phylogeny and identification of conserved peptide motifs.* Glycobiology, 2003. **13**(12): p. 1C-5C.

243. Mollicone, R., et al., *Activity, Splice Variants, Conserved Peptide Motifs, and Phylogeny of Two New α1,3-Fucosyltransferase Families (FUT10 and FUT11).* Journal of Biological Chemistry, 2009. **284**(7): p. 4723-4738.

244. Candelier, J.J., et al., *Expression of fucosyltransferases in skin, conjunctiva, and cornea during human development.* Histochemistry and Cell Biology, 2000. **114**(2): p. 113-124.

245. Kudo, T. and H. Narimatsu, *Fucosyltransferase 3. GDP-Fucose Lactosamine α1,3/4-Fucosyltransferase. Lea and Leb Histo-Blood Groups (FUT3, Lewis Enzyme)*, in *Handbook of Glycosyltransferases and Related Genes*, N. Taniguchi, et al., Editors. 2014, Springer Japan. p. 531-539.

246. Kannagi, R., *Fucosyltransferase 6. GDP-Fucose Lactosamine α3-Fucosyltransferase (FUT6)*, in *Handbook of Glycosyltransferases and Related Genes*, N. Taniguchi, et al., Editors. 2014, Springer Japan. p. 559-571.

247. Fernandez-Mateos, P., et al., *Point Mutations and Deletion Responsible for the Bombay H null and the Reunion H weak Blood Groups.* Vox Sanguinis, 1998. **75**(1): p. 37-46.

248. Ferrer-Admetlla, A., et al., *A Natural History of FUT2 Polymorphism in Humans.* Molecular Biology and Evolution, 2009. **26**(9): p. 1993-2003.

249. Mollicone, R., et al., *Molecular basis for Lewis alpha(1,3/1,4)-fucosyltransferase gene deficiency (FUT3) found in Lewis-negative Indonesian pedigrees.* Journal of Biological Chemistry, 1994. **269**(33): p. 20987-20994.

250. Elmgren, A., et al., *Significance of Individual Point Mutations, T202C and C314T, in the Human Lewis (FUT3) Gene for Expression of Lewis Antigens by the Human α(1,3/1,4)-Fucosyltransferase, Fuc-TIII.* Journal of Biological Chemistry, 1997. **272**(35): p. 21994-21998.

251. Li, F.R., et al., *[The genetic diversity in the full coding region of human FUT5 gene in a Chinese Han population].* Zhonghua Yi Xue Yi Chuan Xue Za Zhi, 2010. **27**(4): p. 473-6.

252. Mollicone, R., et al., *Molecular genetics of α-L-fucosyltransferase genes (H, Se, Le, FUT4, FUT5 and FUT6).* Transfusion Clinique Et Biologique, 1994. **1**(2): p. 91-97.

253. Mollicone, R., et al., *Molecular basis for plasma alpha(1,3)-fucosyltransferase gene deficiency (FUT6).* J Biol Chem, 1994. **269**(17): p. 12662-71.

254. Bengtson, P., et al., *Polymorphonuclear Leukocytes from Individuals Carrying the G329A Mutation in the α1,3-Fucosyltransferase VII Gene (FUT7) Roll on E- and P-Selectins.* The Journal of Immunology, 2002. **169**(7): p. 3940-3946.

255. Bengtson, P., et al., *Identification of a Missense Mutation (G329A; Arg110→ Gln) in the Human FUT7 Gene.* Journal of Biological Chemistry, 2001. **276**(34): p. 31575-31582.

256. Ihara, H., et al., *Fucosyltransferase 8. GDP-Fucose N-Glycan Core α6-Fucosyltransferase (FUT8)*, in *Handbook of Glycosyltransferases and Related Genes*, N. Taniguchi, et al., Editors. 2014, Springer Japan. p. 581-596.

257. Sikora, M., et al., *A variant in the gene FUT9 is associated with susceptibility to placental malaria infection.* Human Molecular Genetics, 2009. **18**(16): p. 3136-3144.

258. Li, M., et al., *Mutations in POFUT1, Encoding Protein O-fucosyltransferase 1, Cause Generalized Dowling-Degos Disease.* The American Journal of Human Genetics, 2013. **92**(6): p. 895-903.

259. Kakuda, S. and R. Haltiwanger, *Fucosyltransferases 12, 13: Protein O-Fucosyltransferases 1 and 2 (POFUT1, POFUT2)*, in *Handbook of Glycosyltransferases and Related Genes*, N. Taniguchi, et al., Editors. 2014, Springer Japan. p. 623-633.

260. Svensson, L., A. Petersson, and S.M. Henry, *Secretor genotyping for A385T, G428A, C571T, C628T, 685delTGG, G849A, and other mutations from a single PCR.* Transfusion, 2000. **40**(7): p. 856-860.

261. Pang, H., et al., *Polymorphism of the human ABO-secretor locus (FUT2) in four populations in Asia: indication of distinct Asian subpopulations.* Annals of Human Genetics, 2001. **65**: p. 429-437.

262. Koda, Y., et al., *Ancient Origin of the Null Allele se^{428} of the Human ABO-Secretor Locus (FUT2).* Journal of Molecular Evolution, 2000. **50**(3): p. 243-248.

263. Koda, Y., et al., *Contrasting Patterns of Polymorphisms at the ABO-Secretor Gene (FUT2) and Plasma alpha(1,3)Fucosyltransferase Gene (FUT6) in Human Populations.* Genetics, 2001. **158**(2): p. 747-756.

264. Viverge, D., et al., *Discriminant Carbohydrate Components of Human Milk According to Donor Secretor Types.* Journal of Pediatric Gastroenterology and Nutrition, 1990. **11**(3): p. 365-370.

265. Fumagalli, M., et al., *Widespread balancing selection and pathogen-driven selection at blood group antigen genes.* Genome Research, 2009. **19**(2): p. 199-212.

266. Guillon, P., et al., *Association between expression of the H histo-blood group antigen, α1,2fucosyltransferases polymorphism of wild rabbits, and sensitivity to rabbit hemorrhagic disease virus.* Glycobiology, 2009. **19**(1): p. 21-28.

267. Johnsen, J.M., et al., *Selection on cis-Regulatory Variation at B4galnt2 and Its Influence on von Willebrand Factor in House Mice.* Molecular Biology and Evolution, 2009. **26**(3): p. 567-578.

268. Linnenbrink, M., et al., *Long-term balancing selection at the blood group-related gene B4galnt2 in the genus Mus (Rodentia; Muridae).* Molecular Biology and Evolution, 2011. **28**(11): p. 2999-3003.

269. Lo Presti, L., et al., *Molecular cloning of the human beta1,4 N-acetylgalactosaminyltransferase responsible for the biosynthesis of the Sd(a) histo-blood group antigen: the sequence predicts a very long cytoplasmic domain.* J Biochem, 2003. **134**(5): p. 675-82.

270. Stuckenholz, C., et al., *FACS-Assisted Microarray Profiling Implicates Novel Genes and Pathways in Zebrafish Gastrointestinal Tract Development.* Gastroenterology, 2009. **137**(4): p. 1321-1332.

271. Johnsen, J.M., et al., *The endothelial-specific regulatory mutation, Mvwf1, is a common mouse founder allele.* Mammalian Genome, 2008. **19**(1): p. 32-40.

272. Mohlke, K.L., et al., *Mvwf, a Dominant Modifier of Murine von Willebrand Factor, Results from Altered Lineage-Specific Expression of a Glycosyltransferase.* Cell, 1999. **96**(1): p. 111-120.

273. Mohlke, K.L., et al., *A novel modifier gene for plasma von Willebrand factor level maps to distal mouse chromosome 11.* Proceedings of the National Academy of Sciences, 1996. **93**(26): p. 15352-15357.

274. Staubach, F., et al., *Expression of the blood-group-related glycosyltransferase B4galnt2 influences the intestinal microbiota in mice.* ISME J, 2012. **6**(7): p. 1345-55.

275. Ilver, D., et al., *Helicobacter pylori Adhesin Binding Fucosylated Histo-Blood Group Antigens Revealed by Retagging.* Science, 1998. **279**(5349): p. 373-377.

276. Lindesmith, L., et al., *Human susceptibility and resistance to Norwalk virus infection.* Nature Medicine, 2003. **9**(5): p. 548-553.

277. Hurd, E.A., et al., *Gastrointestinal mucins of Fut2-null mice lack terminal fucosylation without affecting colonization by Candida albicans.* Glycobiology, 2005. **15**(10): p. 1002-1007.

278. Kindberg, E., et al., *A nonsense mutation (428G→A) in the fucosyltransferase FUT2 gene affects the progression of HIV-1 infection.* AIDS, 2006. **20**(5): p. 685-689.

279. Marionneau, S., et al., *Influence of the combined ABO, FUT2, and FUT3 polymorphism on susceptibility to Norwalk virus attachment.* Journal of Infectious Diseases, 2005. **192**(6): p. 1071-1077.

280. Andres, A.M., et al., *Targets of balancing selection in the human genome.* Mol Biol Evol, 2009. **26**(12): p. 2755-64.

281. Johansson, M.E.V., J.M.H. Larsson, and G.C. Hansson, *The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host–microbial interactions.* Proceedings of the National Academy of Sciences, 2011. **108**(Supplement 1): p. 4659-4665.

282. Johansson, M.E.V., et al., *The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria.* Proceedings of the National Academy of Sciences, 2008. **105**(39): p. 15064-15069.

283. Schütte, A., et al., *Microbial-induced meprin β cleavage in MUC2 mucin and a functional CFTR channel are required to release anchored small intestinal mucus.* Proceedings of the National Academy of Sciences, 2014. **111**(34): p. 12396-12401.

284. Johansson, M.E.V., et al., *Bacteria penetrate the normally impenetrable inner colon mucus layer in both murine colitis models and patients with ulcerative colitis.* Gut, 2014. **63**(2): p. 281-291.

285. Cantarel, B.L., V. Lombard, and B. Henrissat, *Complex Carbohydrate Utilization by the Healthy Human Microbiome.* PLoS One, 2012. **7**(6): p. e28742.

286. Frese, S.A., et al., *The Evolution of Host Specialization in the Vertebrate Gut Symbiont Lactobacillus reuteri.* PLoS Genet, 2011. **7**(2): p. e1001314.

287. Johansson, M.E.V., et al., *Bacteria Penetrate the Inner Mucus Layer before Inflammation in the Dextran Sulfate Colitis Model.* PLoS One, 2010. **5**(8): p. e12238.

288. Hoskins, L.C. and E.T. Boulding, *DEGRADATION OF BLOOD-GROUP ANTIGENS IN HUMAN COLON ECOSYSTEMS .2. A GENE INTERACTION IN MAN THAT AFFECTS FECAL POPULATION-DENSITY OF CERTAIN ENTERIC BACTERIA.* Journal of Clinical Investigation, 1976. **57**(1): p. 74-82.

289. Hoskins, L.C., et al., *MUCIN DEGRADATION IN HUMAN-COLON ECOSYSTEMS - ISOLATION AND PROPERTIES OF FECAL STRAINS THAT DEGRADE ABH BLOOD-GROUP ANTIGENS AND OLIGOSACCHARIDES FROM MUCIN GLYCOPROTEINS.* Journal of Clinical Investigation, 1985. **75**(3): p. 944-953.

290. Connaris, S. and P. Greenwell, *Glycosidases in mucin-dwelling protozoans.* Glycoconjugate Journal, 1997. **14**(7): p. 879-882.

291. Irwin, J.A., et al., *Glycosidase activity in the excretory-secretory products of the liver fluke, Fasciola hepatica.* Parasitology, 2004. **129**(Pt 4): p. 465-72.

292. Brunham, R.C., F.A. Plummer, and R.S. Stephens, *BACTERIAL ANTIGENIC VARIATION, HOST IMMUNE-RESPONSE, AND PATHOGEN-HOST COEVOLUTION.* Infection and Immunity, 1993. **61**(6): p. 2273-2276.

293. Macfarlane, S., E.J. Woodmansey, and G.T. Macfarlane, *Colonization of Mucin by Human Intestinal Bacteria and Establishment of Biofilm Communities in a Two-Stage Continuous Culture System.* Applied and Environmental Microbiology, 2005. **71**(11): p. 7483-7492.

294. Rho, J.-h., et al., *A Novel Mechanism for Desulfation of Mucin: Identification and Cloning of a Mucin-Desulfating Glycosidase (Sulfoglycosidase) from Prevotella Strain RS2.* J. Bacteriol., 2005. **187**(5): p. 1543-1551.

295. Corfield, A.P., et al., *The roles of enteric bacterial sialidase, sialate O-acetyl esterase and glycosulfatase in the degradation of human colonic mucin.* Glycoconj J, 1993. **10**(1): p. 72-81.

296. Roberton, A.M., et al., *A Novel Bacterial Mucinase, Glycosulfatase, Is Associated with Bacterial Vaginosis.* Journal of Clinical Microbiology, 2005. **43**(11): p. 5504-5508.

297. Gut, H., S.J. King, and M.A. Walsh, *Structural and functional studies of Streptococcus pneumoniae neuraminidase B: An intramolecular trans-sialidase.* FEBS Letters, 2008. **582**(23–24): p. 3348-3352.

298. Soong, G., et al., *Bacterial neuraminidase facilitates mucosal infection by participating in biofilm production.* J Clin Invest, 2006. **116**(8): p. 2297-2305.

299. Lee, S.M., et al., *Bacterial colonization factors control specificity and stability of the gut microbiota.* Nature, 2013. **501**(7467): p. 426-429.

300. Ng, K.M., et al., *Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens.* Nature, 2013. **advance online publication**(7469): p. 96-9.

301. Larsbrink, J., et al., *A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes.* Nature, 2014. **advance online publication**.

302. Sonnenburg, J.L., et al., *Glycan Foraging in Vivo by an Intestine-Adapted Bacterial Symbiont.* Science, 2005. **307**(5717): p. 1955-1959.

303. Bry, L., et al., *A Model of Host-Microbial Interactions in an Open Mammalian Ecosystem.* Science, 1996. **273**(5280): p. 1380-1383.

304. Pacheco, A.R., et al., *Fucose sensing regulates bacterial intestinal colonization.* Nature, 2012. **advance online publication**.

305. Freitas, M., et al., *Microbial–host interactions specifically control the glycosylation pattern in intestinal mouse mucosa.* Histochemistry and Cell Biology, 2002. **118**(2): p. 149-161.

306. Freitas, M., et al., *Indigenous microbes and their soluble factors differentially modulate intestinal glycosylation steps in vivo.* Histochemistry and Cell Biology, 2005. **124**(5): p. 423-433.

307. Freitas, M. and C. Cayuela, *Microbial Modulation of Host Intestinal Glycosylation Patterns*. Vol. 12. 2011.

308. Meng, D., et al., *Bacterial symbionts induce a FUT2-dependent fucosylated niche on colonic epithelium via ERK and JNK signaling.* Am J Physiol Gastrointest Liver Physiol, 2007. **293**(4): p. G780-787.

309. Bergstrom, A., et al., *Nature of bacterial colonization influences transcription of mucin genes in mice during the first week of life.* BMC Research Notes, 2012. **5**(1): p. 402.

310. Benus, R.F.J., et al., *Association between Faecalibacterium prausnitzii and dietary fibre in colonic fermentation in healthy human subjects.* British Journal of Nutrition, 2010. **104**(05): p. 693-700.

311. Segain, J.-P., et al., *Butyrate inhibits inflammatory responses through NFκB inhibition: implications for Crohn's disease.* Gut, 2000. **47**(3): p. 397-403.

312. Chang, P.V., et al., *The microbial metabolite butyrate regulates intestinal macrophage function via histone deacetylase inhibition.* Proceedings of the National Academy of Sciences, 2014. **111**(6): p. 2247-2252.

313. Furusawa, Y., et al., *Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells.* Nature, 2013. **504**(7480): p. 446-450.

314. Hinnebusch, B.F., et al., *The Effects of Short-Chain Fatty Acids on Human Colon Cancer Cell Phenotype Are Associated with Histone Hyperacetylation.* The Journal of Nutrition, 2002. **132**(5): p. 1012-1017.

315. Rombeau, J.L. and S.A. Kripke, *Metabolic and Intestinal Effects of Short-Chain Fatty Acids.* Journal of Parenteral and Enteral Nutrition, 1990. **14**(5 suppl): p. 181S-185S.

316. Duncan, S.H., et al., *Acetate Utilization and Butyryl Coenzyme A (CoA):Acetate-CoA Transferase in Butyrate-Producing Bacteria from the Human Large Intestine.* Applied and Environmental Microbiology, 2002. **68**(10): p. 5186-5190.

317. Duncan, S.H., et al., *Contribution of acetate to butyrate formation by human faecal bacteria.* British Journal of Nutrition, 2004. **91**(6): p. 915-923.

318. Fukuda, S., et al., *Bifidobacteria can protect from enteropathogenic infection through production of acetate.* Nature, 2011. **469**(7331): p. 543-547.

319. Samuel, B.S., et al., *Genomic and metabolic adaptations of Methanobrevibacter smithii to the human gut.* Proceedings of the National Academy of Sciences, 2007. **104**(25): p. 10643-10648.

320. Weiss, G.A., C. Chassard, and T. Hennet, *Selective proliferation of intestinal Barnesiella under fucosyllactose supplementation in mice.* Br J Nutr, 2014: p. 1-9.

321. Jiang, X., et al., *Human Milk Contains Elements That Block Binding of Noroviruses to Human Histo–Blood Group Antigens in Saliva.* The Journal of Infectious Diseases, 2004. **190**(10): p. 1850-1859.

322. Pickard, J.M., et al., *Rapid fucosylation of intestinal epithelium sustains host-commensal symbiosis in sickness.* Nature, 2014. **advance online publication**.

323. Hooper, L.V., et al., *A molecular sensor that allows a gut commensal to control its nutrient foundation in a competitive ecosystem.* Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(17): p. 9833-9838.

324. Mahdavi, J., et al., *A novel O-linked glycan modulates Campylobacter jejuni major outer membrane protein-mediated adhesion to human histo-blood group antigens and chicken colonization.* Open Biology, 2014. **4**(1).

325. van Gestel, J., et al., *Density of founder cells affects spatial pattern formation and cooperation in Bacillus subtilis biofilms.* ISME J, 2014.

326. Severi, E., D. Hood, and G. Thomas, *Sialic acid utilization by bacterial pathogens.* Microbiology, 2007. **153**(Pt 9): p. 2817 - 2822.

327. Jones, M.K., et al., *Enteric bacteria promote human and mouse norovirus infection of B cells.* Science, 2014. **346**(6210): p. 755-759.

328. Hayes, K.S., et al., *Exploitation of the Intestinal Microflora by the Parasitic Nematode Trichuris muris.* Science, 2010. **328**(5984): p. 1391-1394.

329. Barr, J.J., et al., *Bacteriophage adhering to mucus provide a non–host-derived immunity.* Proceedings of the National Academy of Sciences, 2013.

330. Chessa, D., et al., *Salmonella enterica serotype Typhimurium Std fimbriae bind terminal α(1,2)fucose residues in the cecal mucosa.* Molecular Microbiology, 2009. **71**(4): p. 864-875.

331. McFall-Ngai, M., et al., *Mannose adhesin-glycan interactions in the Euprymna scolopes Vibrio fischeri symbiosis*. New Developments in Marine Biotechnology, ed. Y. LeGal and H.O. Halvorson. 1998. 273-276.

332. Celli, J.P., et al., *Helicobacter pylori moves through mucus by reducing mucin viscoelasticity.* Proceedings of the National Academy of Sciences, 2009. **106**(34): p. 14321-14326.

333. Mazmanian, S.K., J.L. Round, and D.L. Kasper, *A microbial symbiosis factor prevents intestinal inflammatory disease.* Nature, 2008. **453**(7195): p. 620-625.

334. Coyne, M.J., et al., *Role of glycan synthesis in colonization of the mammalian gut by the bacterial symbiont Bacteroides fragilis.* Proceedings of the National Academy of Sciences, 2008. **105**(35): p. 13099-13104.

335. Huang, J.Y., S.M. Lee, and S.K. Mazmanian, *The human commensal Bacteroides fragilis binds intestinal mucin.* Anaerobe, 2011. **17**(4): p. 137-141.

336. Chow, W.L. and Y.K. Lee, *Free fucose is a danger signal to human intestinal epithelial cells.* British Journal of Nutrition, 2008. **99**(3): p. 449-454.

337. Hooper, L.V. and J.I. Gordon, *Glycans as legislators of host–microbial interactions: spanning the spectrum from symbiosis to pathogenicity.* Glycobiology, 2001. **11**(2): p. 1R-10R.

338. Hooper, L.V., T. Midtvedt, and J.I. Gordon, *How host-microbial interactions shape the nutrient environment of the mammalian intestine.* Annual Review of Nutrition, 2002. **22**: p. 283-307.

339. Rodríguez-Díaz, J., A. Rubio-del-Campo, and M.J. Yebra, *Lactobacillus casei Ferments the N-Acetylglucosamine Moiety of Fucosyl-α-1,3-N-Acetylglucosamine and Excretes l-Fucose.* Applied and Environmental Microbiology, 2012. **78**(13): p. 4613-4619.

340. Olszak, T., et al., *Microbial Exposure During Early Life Has Persistent Effects on Natural Killer T Cell Function.* Science, 2012.

341. Gollwitzer, E.S., et al., *Lung microbiota promotes tolerance to allergens in neonates via PD-L1.* Nat Med, 2014. **20**(6): p. 642-7.

342. Turner, J.R., *Intestinal mucosal barrier function in health and disease.* Nat Rev Immunol, 2009. **9**(11): p. 799-809.

343. Sonnenburg, J.L., L.T. Angenent, and J.I. Gordon, *Getting a grip on things: how do communities of bacterial symbionts become established in our intestine?* Nat Immunol, 2004. **5**(6): p. 569-573.

344. Jostins, L., et al., *Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease.* Nature, 2012. **491**(7422): p. 119-124.

345. Ott, S.J., et al., *Reduction in diversity of the colonic mucosa associated bacterial microflora in patients with active inflammatory bowel disease.* Gut, 2004. **53**(5): p. 685-693.

346. Bach, J.-F., *The Effect of Infections on Susceptibility to Autoimmune and Allergic Diseases.* New England Journal of Medicine, 2002. **347**(12): p. 911-920.

347. Hampe, J., et al., *Association of inflammatory bowel disease with indicators for childhood antigen and infection exposure.* International Journal of Colorectal Disease, 2003. **18**(5): p. 413-417.

348. Gent, A.E., et al., *Inflammatory Bowel-Disease and Domestic Hygiene in Childhood.* Gastroenterology, 1994. **106**(4): p. A686-A686.

349. Hugot, J.P., et al., *Crohn's disease: the cold chain hypothesis.* Lancet, 2003. **362**(9400): p. 2012-2015.

350. Chassaing, B., et al., *Dietary emulsifiers impact the mouse gut microbiota promoting colitis and metabolic syndrome.* Nature, 2015. **519**(7541): p. 92-96.

351. Suez, J., et al., *Artificial sweeteners induce glucose intolerance by altering the gut microbiota.* Nature, 2014. **514**(7521): p. 181-186.

352. Thia, K.T., et al., *An Update on the Epidemiology of Inflammatory Bowel Disease in Asia.* Am J Gastroenterol, 2008. **103**(12): p. 3167-3182.

353. Burisch, J., et al., *East–West gradient in the incidence of inflammatory bowel disease in Europe: the ECCO-EpiCom inception cohort.* Gut, 2014. **63**(4): p. 588-597.

354. Murray, C.J.L., et al., *Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010.* The Lancet, 2012. **380**(9859): p. 2197-2223.

355. Lozano, R., et al., *Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010.* The Lancet, 2012. **380**(9859): p. 2095-2128.

356. Ogura, Y., et al., *A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease.* Nature, 2001. **411**(6837): p. 603-606.

357. Hampe, J., et al., *Evidence for a NOD2-independent susceptibility locus for inflammatory bowel disease on chromosome 16p.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(1): p. 321-326.

358. Maloy, K.J. and F. Powrie, *Intestinal homeostasis and its breakdown in inflammatory bowel disease.* Nature, 2011. **474**(7351): p. 298-306.

359. Khor, B., A. Gardet, and R.J. Xavier, *Genetics and pathogenesis of inflammatory bowel disease.* Nature, 2011. **474**(7351): p. 307-317.

360. Arnold, J.N., et al., *The Impact of Glycosylation on the Biological Function and Structure of Human Immunoglobulins.* Annual Review of Immunology, 2007. **25**(1): p. 21-50.

361. Smith, P.L., et al., *Conditional control of selectin ligand expression and global fucosylation events in mice with a targeted mutation at the FX locus.* J Cell Biol, 2002. **158**(4): p. 801-15.

362. Franke, A., et al., *Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci.* Nat Genet, 2010. **42**(12): p. 1118-1125.

363. McGovern, D.P.B., et al., *Genome-wide association identifies multiple ulcerative colitis susceptibility loci.* Nat Genet, 2010. **42**(4): p. 332-337.

364. Glynn, L.E. and E.J. Holborow, *RELATION BETWEEN BLOOD GROUPS, SECRETOR STATUS AND SUSCEPTIBILITY TO RHEUMATIC FEVER.* Arthritis and Rheumatism, 1961. **4**(2): p. 203-&.

365. Van der Sluis, M., et al., *Muc2-Deficient Mice Spontaneously Develop Colitis, Indicating That MUC2 Is Critical for Colonic Protection.* Gastroenterology, 2006. **131**(1): p. 117-129.

366. Fu, J., et al., *Loss of intestinal core 1–derived O-glycans causes spontaneous colitis in mice.* The Journal of Clinical Investigation, 2011. **121**(4): p. 1657-1666.

367. Miyahara, K., et al., *Serum Glycan Markers for Evaluation of Disease Activity and Prediction of Clinical Course in Patients with Ulcerative Colitis.* PLoS ONE, 2013. **8**(10): p. e74861.

368. McCarthy, C., et al., *The role and importance of glycosylation of acute phase proteins with focus on alpha-1 antitrypsin in acute and chronic inflammatory conditions.* J Proteome Res, 2014. **13**(7): p. 3131-43.

369. Chrostek, L., et al., *Sialic acid level reflects the disturbances of glycosylation and acute-phase reaction in rheumatic diseases.* Rheumatol Int, 2014. **34**(3): p. 393-9.

370. Larsson, J.M., et al., *Altered O-glycosylation profile of MUC2 mucin occurs in active ulcerative colitis and is associated with increased inflammation.* Inflamm Bowel Dis, 2011. **17**(11): p. 2299-307.

371. Hurd, E.A. and S.E. Domino, *Increased susceptibility of secretor factor gene Fut2-null mice to experimental vaginal candidiasis.* Infection and Immunity, 2004. **72**(7): p. 4279-4281.

372. Blackwell, C.C., et al., *Non-secretion of ABO antigens predisposing to infection by Neisseria meningitidis and Streptococcus pneumoniae.* Lancet, 1986. **2**(8501): p. 284-285.

373. Haverkorn, M.J. and W.R. Goslings, *Streptococci, ABO blood groups, and secretor status.* American Journal of Human Genetics, 1969. **21**(4): p. 360-375.

374. Blackwell, C.C., et al., *Non-secretion of ABO blood-group antigens predisposing to infection by Haemophilus influenzae.* Lancet, 1986. **2**(8508): p. 687-687.

375. Magalhaes, A., et al., *Fut2-null mice display an altered glycosylation profile and impaired BabA-mediated Helicobacter pylori adhesion to gastric mucosa.* Glycobiology, 2009. **19**(12): p. 1525-1536.

376. Rausch, P., et al., *Colonic mucosa-associated microbiota is influenced by an interaction of Crohn disease and FUT2 (Secretor) genotype.* Proceedings of the National Academy of Sciences, 2011. **108**(47): p. 19030-5.

377. Wacklin, P., et al., *Secretor Genotype FUT2 Is Strongly Associated with the Composition of Bifidobacteria in the Human Intestine.* PLoS One, 2011. **6**(5): p. e20113.

378. Wacklin, P., et al., *Faecal Microbiota Composition in Adults Is Associated with the FUT2 Gene Determining the Secretor Status.* PLoS One, 2014. **9**(4): p. e94863.

379. Tong, M., et al., *Reprograming of gut microbiome energy metabolism by the FUT2 Crohn's disease risk polymorphism.* ISME J, 2014. **8**(11): p. 2193-2206.

380. Kashyap, P.C., et al., *Genetically dictated change in host mucus carbohydrate landscape exerts a diet-dependent effect on the gut microbiota.* Proceedings of the National Academy of Sciences, 2013. **110**(42): p. 17059-64.

381. Li, M., et al., *Symbiotic gut microbes modulate human metabolic phenotypes.* Proceedings of the National Academy of Sciences, 2008. **105**(6): p. 2117-2122.

382. Ivanov, I.I., et al., *Induction of Intestinal Th17 Cells by Segmented Filamentous Bacteria.* Cell, 2009. **139**(3): p. 485-498.

383. Rivas, M.A., et al., *Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease.* Nat Genet, 2011. **43**(11): p. 1066-1073.

384. Loftus Jr, E.V., *Clinical epidemiology of inflammatory bowel disease: incidence, prevalence, and environmental influences.* Gastroenterology, 2004. **126**(6): p. 1504-1517.

385. Hashimoto, T., et al., *ACE2 links amino acid malnutrition to microbial ecology and intestinal inflammation.* Nature, 2012. **487**(7408): p. 477-481.

386. Gao, Z., et al., *Substantial Alterations of the Cutaneous Bacterial Biota in Psoriatic Lesions.* PLoS One, 2008. **3**(7): p. e2719.

387. Koren, O., et al., *Host Remodeling of the Gut Microbiome and Metabolic Changes during Pregnancy.* Cell, 2012. **150**(3): p. 470-480.

388. Yoshimoto, S., et al., *Obesity-induced gut microbial metabolite promotes liver cancer through senescence secretome.* Nature, 2013. **499**(7456): p. 97-101.

389. Couturier-Maillard, A., et al., *NOD2-mediated dysbiosis predisposes mice to transmissible colitis and colorectal cancer.* The Journal of Clinical Investigation, 2013. **123**(2): p. 700-711.

390. Tanaka, T., et al., *Genome-wide Association Study of Vitamin B6, Vitamin B12, Folate, and Homocysteine Blood Concentrations.* American Journal of Human Genetics, 2009. **84**(4): p. 477-482.

391. Karlsen, T.H. and K.M. Boberg, *Update on primary sclerosing cholangitis.* Journal of Hepatology, 2013. **59**(3): p. 571-582.

392. Franke, A., et al., *Sequence variants in IL10, ARPC2 and multiple other loci contribute to ulcerative colitis susceptibility.* Nat Genet, 2008. **40**(11): p. 1319-1323.

393. Devkota, S., et al., *Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in Il10-/- mice.* Nature, 2012. **487**(7405): p. 104-108.

394. Martens, E.C., H.C. Chiang, and J.I. Gordon, *Mucosal Glycan Foraging Enhances Fitness and Transmission of a Saccharolytic Human Gut Bacterial Symbiont.* Cell Host & Microbe, 2008. **4**(5): p. 447-457.

395. Koropatkin, N.M., E.A. Cameron, and E.C. Martens, *How glycan metabolism shapes the human gut microbiota.* Nat Rev Micro, 2012. **10**(5): p. 323-335.

396. Lozupone, C.A., et al., *Diversity, stability and resilience of the human gut microbiota.* Nature, 2012. **489**(7415): p. 220-230.

397. Folseraas, T., et al., *Extended analysis of a genome-wide association study in primary sclerosing cholangitis detects multiple novel risk loci.* Journal of Hepatology, 2012. **57**(0): p. 366-75.

398. Nanthakumar, N.N., et al., *The role of indigenous microflora in the development of murine intestinal fucosyl- and sialyltransferases.* The FASEB Journal, 2003. **17**(1): p. 44-46.

399. Knight, R., et al., *Unlocking the potential of metagenomics through replicated experimental design.* Nat Biotech, 2012. **30**(6): p. 513-520.

400. Bazzaz, F.A., *Physiological ecology of plant succession.* Annual Review of Ecology and Systematics, 1979. **10**: p. 351-371.

401. Spor, A., O. Koren, and R. Ley, *Unravelling the effects of the environment and host genotype on the gut microbiome.* Nat Rev Micro, 2011. **9**(4): p. 279-290.

402. Henry, S., et al., *Molecular basis for erythrocyte Le(a+b+) and salivary ABH partial-secretor phenotypes: expression of a FUT2 secretor allele with an A→T mutation at nucleotide 385 correlates with reduced α(1,2) fucosyltransferase activity.* Glycoconjugate Journal, 1996. **13**(6): p. 985-993.

403. Frank, D.N., et al., *Investigating the biological and clinical significance of human dysbioses.* Trends in Microbiology, 2011. **19**(9): p. 427-434.

404. Aspholm-Hurtig, M., et al., *Functional Adaptation of BabA, the H. pylori ABO Blood Group Antigen Binding Adhesin.* Science, 2004. **305**(5683): p. 519-522.

405. Sonnenburg, J.L., et al., *Glycan foraging in vivo by an intestine-adapted bacterial symbiont.* Science, 2005. **307**(5717): p. 1955-9.

406. Giannasca, K.T., P.J. Giannasca, and M.R. Neutra, *Adherence of Salmonella typhimurium to Caco-2 cells: identification of a glycoconjugate receptor.* Infect Immun, 1996. **64**(1): p. 135-45.

407. Kobayashi, M., et al., *Roles of gastric mucin-type O-glycans in the pathogenesis of Helicobacter pylori infection.* Glycobiology, 2009. **19**(5): p. 453-461.

Conclusions

In this thesis I investigated several important aspects of intestinal microbial communities in humans and mice. This included the influence of chronic- and acute diseases on the microbiome, but also genetic and environmental effects that influence the composition and dynamics of host associated microbial communities.

In the first chapter of this thesis, I explored population differences in the mucosa associated microbial community in patients with inflammatory bowel diseases and healthy controls. I was able to identify population-specific and common disease signatures, even though differences among populations were a dominating pattern. However, local and common disease signatures were more pronounced in the active microbial community, which we investigated by analyzing the 16S rRNA gene (DNA) and its transcripts (RNA) in parallel. These analyses reveal the impact of environmental factors on the human associated microbiome, especially in a disturbed state. Several comparisons of microbial communities among human populations have been performed so far, but these were mainly restricted to relatively defined geographic distribution within a population sample. Future studies should increase the sample size and include a broader sampling of healthy and diseased individuals across geographical gradients to validate those biogeographic patterns observed. By broadening the sampling effort and increasing the resolution of physical distance between sampled individuals, it should be possible to explore similarity distance-decay relationships in human microbial communities. Data about diet, age, diseases, and genetic information could help to disentangle neutral community dynamics, from host- and life-style dependent community adaptation. This would enable us to identify population specific parts of the microbial communities, which might influence the susceptibility to local environmental stressors, like infections.

Chapters two and three demonstrate how ABO blood group antigens influence the bacterial communities in health and disease. We were able to identify consistent *FUT2* associated patterns of microbial community characteristics (*e.g.* alpha diversity, beta diversity) in two different human populations (German, Norwegian) and tissues (colonic mucosa, bile). I also investigated the role of *Fut2* on the bacterial community development in a knockout mouse model over an 11 week time course. I established several breeding lines which were founded either by a *Fut2* $^{+/+}$ or a *Fut2* $^{-/-}$ dam. This allowed me to observe an increasing effect of *Fut2* genotype with time and differences in community stability between genotypes. Furthermore maternal/grand maternal genotype effects on the offspring's microbial community were very prominent. Multiple disturbances, such as different infections or community instabilities over a lifetime, differing immune responses, and a slightly more labile barrier are as single factors not of high importance. However, an accumulation of those

slight imbalances could be the driving force to change low baseline responses of the mucosa to its native microbial community into chronic and exacerbated inflammation. The higher risk of nonsecretors to develop Crohn Disease may in part be a side effect of ongoing host-pathogen coevolution between a wide spectrum of infectious agents, the ABO histo-blood-group genes and environmental changes associated with modern living.

In chapter four, I investigated multiple putative disturbances simultaneously, *i.e.* a host genotype-dependent (*B4galnt2*) change in the intestinal microbiota, antibiotic treatment and infectious colitis. This revealed a more resilient and less susceptible microbial community in mice lacking *B4galnt2* gut expression. This provides strong experimental evidence for a scenario of complex selection on the *B4galnt2* gene, which balances beneficial host-microbe relationships and lowered susceptibility to enteric pathogens, with the costs of compromised blood homeostasis. In conclusion, my thesis demonstrates the importance of blood-group-related antigens for the microbiome and susceptibility to chronic- and acute diseases.

Acknowledgements

First and most importantly, I am deeply grateful to John Baines, who took me in as a Master student and supported and taught me ever since. You showed me the ropes, not only giving advice on my research, but also for life and career. Thank you John!

Thanks to Guntram Grassl, for his insight, humor, and a bit of Bavarian attitude in northern Germany.

I thank Sven Künzel, for his pragmatism, his genuine spirit, and his amazing laboratory Kung Fu skills, which made so many projects possible. Many thanks to our incredible technicians, Katja Cloppenborg and Silke Carstensen, who kept me busy and the data flowing.

Thanks to my dear PhD colleagues and friends Jun Wang, Meriem Bellouhane, and Marie Vallier, who were always ready to discuss or laugh about the big and the small issues, and offered a helping hand, if girl/man power was needed. Thanks to Natalie Steck, for her great work, the many discussions, and for a sympathetic ear for a struggling PhD student.

Big THANKS to Abdulhadi Suwandi (*aka* Addi), Jan Schubert and our interns Theresa Arlt and Sarah Lemcke, who were all of great help.

I thank my thesis committee members, Philip Rosenstiel and Thomas Röder, who gave critical input on my work and guided me through the PhD.

Thanks to Tina Henrich, Joshka Kaufmann, Martin Kalbe, and Chaitanya Gokhale, for many cigarette breaks and the constant flow of coffee and ideas.

Thanks to Kerstin Mehnert, for running the IMPRS and for keeping us PhD students in check. Thanks to the whole staff of the MPI! You made it a fantastic experience to work here. Being surrounded by so many bright and capable people, who offer help and support no matter what, is a pleasure and inspiration.

I would also like to thank the “Excellence Cluster Inflammation at Interfaces” and the International Max-Planck-Research School at the MPI in Plön for support and funding.

Last but not least, I want to thank my family that gave me strength through their unlimited support and love!

Danke, dass Ihr an mich glaubt!

Curriculum Vitae

Name:	Philipp Rausch
Date of birth:	26.04.1986
Place of birth:	Greiz, Germany
Current residence:	Kiel, Germany
	Greiz, Germany
Education:	
08/1996-06/2004	Staatliches Gymnasium Greiz
10/2005-09/2008	Studies in general biology at the University of Bayreuth (Bachelor of Science) Thesis: “Der Einfluss des MHC auf die Gruppenzusammensetzung von Europäischen Wildkaninchen (Oryctolagus cuniculus L.)” (Prof. Dr. Dietrich v. Holst)
10/2008-09/2010	Master student in the “Evolution, Ecology, and Systematics”-Program of the Ludwig-Maximilians-University & Volkswagen Stiftung (Master of Science) Thesis: “The influence of blood group antigens on the human intestinal microbiota” (Prof. Dr. John F. Baines, Prof. Dr. John Parsch; Dr. Stephan Hutter)
11/2010-current	PhD student in the Excellence Cluster “Inflammation at Interfaces” Christian-Albrechts-University, Kiel; International Max-Planck Research School (IMPRS) at the Max-Planck-Institute for Evolutionary Biology, Plön
Awards and Grants:	
09/2008	poster price in the section evolutionary biology at 101st DZG Meeting Jena
02/2009	EES travel grant for an internship at the Department of Prof. Dr. J. Höglund (Uppsala)
04/2009	EES grant for the organization of a joint student conference EES-LMU and OEN-Potsdam
3/2010	EES thesis grant
4/2013	DFG Priority Program (SPP1656) with Prof. Dr. John F. Baines, Prof. Dr. Guntram Grassl, and Dr. Natalie Steck

Publications:

1. Oppelt C, Starkloff A, **Rausch P**, Von Holst D, Rödel HG. Major histocompatibility complex variation and age-specific endoparasite load in subadult European rabbits. Mol Ecol. 2010;19(19):4155-67. doi: 10.1111/j.1365-294X.2010.04766.x. PubMed PMID: 20723049.
2. **Rausch P**, Rehman A, Künzel S, Häsler R, Ott SJ, Schreiber S, et al. Colonic mucosa-associated microbiota is influenced by an interaction of Crohn disease and *FUT2* (*Secretor*) genotype. Proceedings of the National Academy of Sciences. 2011;108(47):19030-5. Epub 2011/11/10. doi: 10.1073/pnas.1106408108. PubMed PMID: 22068912; PubMed Central PMCID: PMC3223430.
3. Folseraas T, Melum E, **Rausch P**, Juran BD, Ellinghaus E, Shiryaev A, et al. Extended analysis of a genome-wide association study in primary sclerosing cholangitis detects multiple novel risk loci. Journal of Hepatology. 2012;57(0):366-75. Epub 2012/04/24. doi: 10.1016/j.jhep.2012.03.031. PubMed PMID: 22521342; PubMed Central PMCID: PMC3399030.
4. Rehman A, **Rausch P**, Wang J, Skieceviciene J, Kiudelis G, Bhagalia K, et al. Geographical patterns of the standing and active human gut microbiome in health and IBD. Gut. 2015. doi: 10.1136/gutjnl-2014-308341.
5. **Rausch P**, Steck N, Suwandi A, Seidel J, Künzel S, Bhullar K, Basic M, Bleich A, Johnsen J, Vallance B, Baines J, Grassl G.
   Expression of the blood-group-related gene *B4galnt2* alters susceptibility to *Salmonella* infection. Accepted at Plos Pathogens (2015)

Declaration

I hereby declare,

(i) that apart from my supervisor's guidance, the content and design of this thesis is completely my own work. Contributions of other authors are listed in the following section.

(ii) this thesis has not been submitted either partially or completely as part of a doctoral degree to another examining institution. No other materials are published or submitted for publication than indicated in this thesis.

(iii) this thesis was prepared in compliance with the "Rules of Good Scientific Practice" of the German Research Foundation (DFG).

Author Contributions:

Chapter I: "Geographical patterns of the standing and active human gut microbiome in health and IBD"

Philipp Rausch, Jun Wang, and John F. Baines analysed the data; Philipp Rausch, Philip Rosentiel, and John F. Baines wrote the paper; Ateequr Rehman, Stefan Schreiber, and Stephan Ott designed the research; Ateequr Rehman, Philipp Rausch, Jurgita Skieceviciene, Gediminas Kiudelis, Ketan Bhagalia, Deepak Amarapurkar, Limas Kupcinskas, Stephan Ott and Philip Rosenstiel performed the research.

Chapter II: "Colonic mucosa-associated microbiota is influenced by an interaction of Crohn disease and *FUT2* (*Secretor*) genotype"

Philipp Rausch, Ateequr Rehman, Stephan Ott, Philip Rosenstiel, and John F. Baines analyzed the data; Philipp Rausch, Philip Rosenstiel, Andre Franke, and John F. Baines wrote the paper; Ateequr Rehman, Stefan Schreiber, and John F. Baines designed the research; Philipp Rausch, Ateequr Rehman, Sven Künzel, Robert Häsler, and Andre Franke performed the research.

"Extended analysis of a genome-wide association study in primary sclerosing cholangitis detects multiple novel risk loci"

Trine Folseraas, Espen Melum, Philipp Rausch, Alexey Shiryaev, Kristian Holm, Jon K Laerdahl, Brian D Juran, Eva Ellinghaus, David Ellinghaus, John F Baines analyzed the data; Trine Folseraas, Espen Melum, John F. Baines, Andre Franke, and Tom H Karlsen coordinated the project and supervised data-analysis; Trine Folseraas, Espen Melum, Philipp Rausch, Alexey Shiryaev, Eva Ellinghaus, John F. Baines, Andre Franke, and Tom H Karlsen wrote the paper; remaining co-authors contributed to the patient ascertainment, sample collection and/or clinical data.

Chapter III: "Dependence of microbial community development on the host and maternal *α-1,2-fucosyltransferase* gene"

Philipp Rausch analyzed the data; Philipp Rausch and John F. Baines wrote the paper; Philipp Rausch, Philip Rosenstiel, and John F. Baines designed the research; Philipp Rausch and Sven Künzel performed the research.

Chapter IV: "Expression of the blood-group-related gene *B4galnt2* alters susceptibility to *Salmonella* infection"

Philipp Rausch, Natalie Steck analyzed the data; Philipp Rausch, Natalie Steck, John F Baines, and Guntram Grassl wrote the paper; Natalie Steck, John F Baines, Guntram Grassl, Abdulhadi Suwandi, Philipp Rausch , and Jill M Johnsen designed the research; Natalie Steck, Abdulhadi Suwandi, Guntram Grassl, Janice A Seidel, Sven Künzel, Kirandeep Bhullar, Bruce A Vallance, Marijana Basic, and Andre Bleich performed the research

…… ……

Philipp Rausch Prof. Dr. John F. Baines