MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Self-organized criticality in the gut microbiome"

verfasst von / submitted by

Franziska Bauchinger, BSc

angestrebter akademischer Grad / in partial fulfillment of the requirements for the degree of

Master of Science (MSc)

Wien, 2015 / Vienna 2015

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 299
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Interdisziplinäres Masterstudium Environmental Sciences
Betreut von / Supervisor:	Univ.-Prof. Mag. Dr. Thomas Rattei

Acknowledgment

Abstract

We introduce a theoretical model for time evolution of the gut microbiome. The model captures the temporal dynamics and simulates the composition within the microbial community.

For parametrization we fitted the model to experimental data by using generalized simulated annealing. The experimental data and data from simulations show typical features of self-organized criticality, a concept explaining how general characteristic behavior arises in systems of many interacting elements [1]. Specifically, species lifetime distributions follow a power-law and the mean power spectral densities of the time-series show pink noise in their distribution.

We computed power spectral densities for the individual microbial species and find that the microbial community stratifies into subgroups of species exhibiting different kinds of noise, namely white, pink and brown noise. The noise a subgroup exhibits correlates with species abundance. The subgroups also differ in how strong model parameters, namely immigration probability, extinction probability and microbe-microbe interaction strength, influence their time evolution.

Species reaching only low abundances typically exhibit white noise. Those species experience high turnover and are strongly influenced by stochastic external fluctuations, influences incorporated in immigration probability.

Species showing the highest abundances and high temporal persistence within the community exhibit brown noise. They are strongly affected by external deterministic drivers, which is mainly the host of the microbial community. This influence is comprised in extinction probability.

Species showing intermediate abundances exhibit pink noise. They are neither very strongly affected by stochastic external fluctuations nor by the influence of the host. The time evolution of species exhibiting pink noise is likely mostly shaped by internal structure, the interactions within the microbial community.

The gut microbiome as a system also exhibits pink noise. We therefore argue that it shows self-organized critical behavior and that the structures within the microbial community strongly shape the time development and dynamics of this system.

Our work provides insight into the systematic behavior of the gut microbiome and the influence of community structure. It can be used as a starting point for further research, e.g. to examine the influence of internal structure on various microbial communities and their dynamics or to investigate diseased states of the gut microbiome and potentially associated patterns.

Zusammenfassung

Wir stellen ein theoretisches Model der zeitlichen Entwicklung der Darmflora vor. Das Model erfasst die zeitliche Dynamik und simuliert die Zusammensetzung der mikrobiellen Gemeinschaft.

Zur Parametrisierung passten wir unser Model durch "generalized simulated annealing" an experimentelle Daten an. Sowohl experimentelle Daten als auch Daten aus der Simulation zeigen typische Merkmale selbstorganisierter Kritikalität, ein Konzept zur Erklärung von generellem charakteristischem Verhalten, dass in Systemen mit vielen interagierenden Elementen auftritt [1]. Diese Merkmale sind Verteilungen in den Lebenszeiten der Spezies die Potenzgesetzen folgen, sowie mittlere spektrale Leistungsdichten der Zeitreihen die rosa Rauschen in ihren Verteilungen zeigen.

Wir errechneten spektrale Leistungsdichten für die einzelnen mikrobiellen Spezies und finden eine Stratifizierung der mikrobiellen Gemeinschaft in Untergruppen, die unterschiedliche Typen von Rauschen aufweisen. Diese Untergruppen weisen entweder weißes, rosa oder braunes Rauschen auf. Das Rauschen, dass eine Untergruppe aufweist, korreliert mit der Abundanz der Spezies. Die Untergruppen unterscheiden sich auch darin, wie stark Model-Parameter ihre zeitliche Entwicklung beeinflussen. Diese Model-Parameter sind Immigrationswahrscheinlichkeit, Extinktionswahrscheinlichkeit und Stärke der Interaktion zwischen Mikroben.

Gering abundante Spezies zeigen typischerweise weißes Rauschen. Diese Spezies erfahren häufige Umwälzungen und sind stark beeinflusst von stochastischen externen Fluktuationen. Diese Einflüsse sind in der Immigrationswahrscheinlichkeit enthalten.

Spezies, die die höchsten Abundanzen und hohe zeitliche Persistenz aufweisen, zeigen braunes Rauschen. Sie sind stark beeinflusst von externen deterministischen Treibern, hauptsächlich dem Host der mikrobiellen Gemeinschaft. Diese Einflüsse sind in der Extinktionswahrscheinlichkeit enthalten.

Spezies mit mittleren Abundanzen zeigen rosa Rauschen. Sie sind weder durch stochastische externe Fluktuationen noch durch den Host besonders stark beeinflusst. Die zeitliche Entwicklung dieser Untergruppe von Spezies wird wahrscheinlich in erster Linie durch interne Strukturen geformt, die Interaktionen innerhalb der mikrobiellen Gemeinschaft.

Die Darmflora als System weist ebenfalls rosa Rauschen auf. Wir argumentieren daher, dass die Darmflora selborganisiert kritisches Verhalten zeigt und dass die Strukturen innerhalb der mikrobiellen Gemeinschaft die zeitliche Entwicklung und Dynamik dieses Systems stark formen.

Unsere Arbeit liefert Erkenntnisse über das systemische Verhalten der Darmflora und den Einfluss der mikrobiellen Gemeinschaftsstruktur. Sie kann als Ausgangspunkt für weiter Forschung dienen, z.B. um den Einfluss interner

Strukturen auf verschiedene mikrobielle Gemeinschaften und deren Dynamik zu ermitteln oder um Krankheitszustände der Darmflora und mögliche assoziierte Muster zu erforschen.

Contents

1 Introduction and Background

1.1 The gut microbiome

Recent research supports the idea that the human body functions as a habitat for myriads of microbes, consequentially making it an environment composed of by far more microbial than human genes. Efforts have been undertaken by projects as HMP (Human Microbiome Project) [2] or MetaHIT (Metagenomics of the Human Intestinal Tract) [3] to catalog large parts of these various human microbiomes, such as skin, mouth or gut. Of all these human body sites inhabited by microbes the intestinal tract is the most populated one. An estimated number of 10^{13} bacteria essentially makes it one of the most densely populated known ecosystems [4]. Several 100-1000 microbial species or OTUs (operational taxonomic units) make up this diverse community [5]. Interestingly, these species belong to only a few of all described bacterial phyla, predominantly *Bacteroidetes*, *Firmicutes* and, to a lesser extent, *Proteobacteria* and *Actinobacteria* [6]. Comparing the intestinal microbial community of different humans reveals great variability in species composition [7] making ones gut microbiome a unique characteristic very much like a fingerprint. This unique gut community is an invaluable symbiont for its host, performing many important functions including, but not limited to, metabolic functions, stimulation of the immune system and inhibition of pathogens through competitive exclusion [8].

The respective roles of genetics and environmental factors on the development and composition of the gut microbiome are yet unclear. For instance, it has been shown that the development of the gut microbiome in infants is highly dependent on the mode of delivery. The resemblance of the intestinal microbiota composition between infants and their mothers is significantly higher for vaginally delivered infants than for infants delivered by C-section. This difference decreases with the age of the infants and diet becomes a more prominent factor in the development of the gut microbiome [9]. In general, bacterial diversity increases with age and in older children the biological mothers fecal microbiota is no more similar to her children than their biological fathers. The intestinal microbial communities of monozygotic twins are also no more similar than those of dizygotic twins, whilst the fecal microbiota of co-habiting adults are significantly more similar than those of members from different families. Those findings of *Yatsunenko et al.* [10] suggest a rather low heritability of the microbiome and furthermore emphasize the importance of shared environmental influences. Overall it seems that predominantly geography and cultural traditions as well as age explain the variety in microbial composition of fecal samples.

There is also substantial variation in the species composition of one individual occurring over time, with especially the gut microbiome of infants

undergoing frequent transitions [8]. Although differences in the microbial composition of newborns might be pronounced, the functional maturation during the first three years shows common patterns and features, even in infants from geographically distinct regions [10]. The microbiota of adults is, particularly viewed over longer timescales, more stable and the fractional abundance of OTUs is relatively constant. Even though, relative and absolute species abundances are fluctuating on a short timescale and changes in diet, medication or varying environmental influences may result in composition shifts [11].

High variability, especially over short timescales, in the intestinal microbial community at species level poses a difficult challenge when trying to define the makeup of a healthy gut microbiome. *Arumugam et al.* [12] brought forward the idea that there are different enterotypes existing which together form something of a core gut microbiome. However, this thought has also been challenged [13], suggesting that microbial variation in the gut is continuous, not stratified. While defining a typical or common gut microbiome on the basis of species composition seems therefore infeasible, the functional gene profiles across individuals are quite similar, hinting towards a functional core microbiome [8]. Hence functional composition, though naturally going through transition during human development, might also be the more appropriate measure for defining a healthy gut microbiome [14].

A dysbiosis in the gut microbiota, i.e. disruption of the microbial community, has been associated with various diseases like allergies, inflammatory bowel diseases (IBD), obesity, type 2 diabetes, gastric cancer and even autism [15]. Even though, it is still unclear whether the dysbiosis is caused by the associated disease or contributed to its development. Moreover, treatment of a disease with e.g. antibiotics can lead to substantial shifts in the microbial community making the host more susceptible to pathogenic invasion [16]. Considering additionally the naturally high variability of the gut microbiota across individuals and time makes apparent the importance of distinguishing between naturally occurring fluctuations inherent to the system and changes that might lead to or stem from diseases. The question remains how resilient the gut microbial community is, what determines its resilience and of what nature and how strong perturbations have to be in order to alter the composition [8]. Broader understanding of the gut microbiome as an ecosystem and the dynamics within this system is needed to improve application of more specific therapies. Currently pro- and praebiotics, dietary measures and faecal microbiome transplantation are increasingly applied to directly or indirectly influence the community composition for medical purposes [14]. Yet our limited knowledge makes it difficult to design more specifically targeted therapies. In microbial ecology in principal and especially concerning the intestinal microbial community modeling approaches can provide valuable input. The incredible variability of this microbiome makes theoretical prediction and development of targeted thera-

2

pies difficult. Gut microbiome models can help to test assumptions, predict possible outcomes and infer connections between microbial composition and disease or health [17].

1.2 Self-organized criticality

The idea of self-organized criticality (SOC) was first introduced by Bak, Tang and Wiesenfeld in 1988 [18] as an explanation for emergent complex behavior in far-from-equilibrium systems. The word itself is comprised of two parts. "Self-organization" refers to the emergence of complex behavior or structure in a system of many interacting elements. "Criticality" refers to a system at a critical point, where little input of energy can have dispro-portionately large effects [19]. An intuitive metaphor for how SOC works and what characterizes systems showing SOC can be found in the book *"Self-Organized Criticality. Emergent Complex Behavior in Physical and Bi-ological Systems"* from 1998 by H.J.Jensen [1]: When attempting to push a piano across a floor and continuously applying the same amount of force, at first nothing is going to happen and the piano won't move. But once a crit-ical point is reached and the applied force overcomes the friction between piano and floor, the piano is going to jump forward. However, it is impossi-ble to predict the size of this jump forward. Even though the force applied is always the same, most of the time the piano will only move a little bit, but sometimes it will make a big jump forward. This is exactly the characteristic behavior of a self-organized critical system. The input to the system is small but continuous and once a critical point is reached, the system is going to react to the input. Due to the interactions between the individual elements of the system this reaction can sometimes propagate throughout the whole system, but most of the time the reaction is going to be small. Those event sizes, or "avalanches" as they are often referred to, show a self-similar distribution. This reflects the fact that most events are small and only a few events are big and affecting a large proportion of the system. This scale-freeness in lifetimes of system features, reaching from avalanches in sand piles to earthquakes or extinction events in biological communities, has been associated with a variety of complex systems [19] and is also one of the key characteristics of self-organized criticality [1]. Closely linked to this observation is also the most prominent property associated with SOC, a typical configuration of its spectral density termed pink noise or 1/f-noise. When performing a spectral density analysis of a fluctuating temporal sig-nal, the spectral density function of a process takes the form of $P(f) \propto f^\beta$ [20]. The power of the fluctuations in the signal $P(f)$ and their frequency f show a log-linear relationship where the value of β would be the observable slope in a log-log plot. As described in [21], β allows a qualitative charac-

terization of the underlying process and the special case of $\beta \sim -1$ is of particular interest in the context of SOC. An exponent $\beta \sim -1$ indicates that different event sizes and time-scales influence the system equally and the system exhibits fractal scaling. SOC is thought to be one possible mechanism underlying scale-freeness or fractal scaling in complex systems [1]. In biology especially 1/f-noise has attracted quite some interest, as it can be found on cellular and organic level in living organisms (e.g. self-discharge of neurons or heart rate and blood pressure) [22] as well as underlying various ecological fluctuations (e.g. population dynamics) [23]. The idea of various biological systems being self-organized critical seems accordingly reasonable. Applied to biological communities it would suggest that particularly interactions between individuals and species are of importance and shape the community structure. Noise dynamics are inherent to the system and noise is not solely a stochastic phenomenon, but to some extent an intrinsic feature. Indeed, indications for SOC based on noise spectra have been found in various systems and fields, reaching from the classical sand-pile model featured in the paper from 1988 [18] across earthquakes [24] and extinction events in fossil records [25] to human brain oscillations [26], stock markets [27] and social media [21]. However, the concept of SOC also poses difficulties. One of the main critical points is the lack of a formal and generally accepted definition. Consequentially, the characterization of SOC is qualitative rather than quantitative. Restricted parameter ranges that lead to SOC behavior additionally attract criticism [1]. Particularly in biology, where processes are strongly characterized by interactions of the protagonists, SOC still remains a promising field of research with the potential to shed some light on the emergence of complexity and the universality of fractals and scale-invariance. The gut microbial community is a biological system where interactions between the protagonists, microbes as well as the host, are very important for system functions. There is also a continuous flow-through of matter and energy, i.e. food intake and excretion, perturbing the system and introducing stochasticity. These characteristics make it a system where the concept of SOC could be particularly useful.

1.3 Modeling approaches to community structure

Modeling biological communities poses the problem of finding a balance between simplification and detail. When simplifying a model too much it won't capture the main dynamics of the modeled system. On the other hand, adding too much detail and too many variables to a model complicates deductive reasoning. With this in mind, a broad division in biological models can be made between models following the neutral theory and the ecological niche theory, respectively. Neutral theory, first introduced by S.P.Hubbell in [28] seeks to describe formation and maintenance of biodi-

4

versity patterns in ecological communities. It assumes that species-specific traits play a minor role in the development of biodiversity due to the strong influence of stochasticity. Application of the neutral theory to e.g. a model of the human microbiome evolution [29] emphasized differences in evolution due to the extent of parental contribution. Yet the assumptions taken are profound simplifications not applicable to the general case. According to Harris et al. [30], who applied neutral theory to 278 human gut microbiome data sets, neutral theory doesn't hold on the whole community level, suggesting that species-specific traits play an important role. Within certain taxonomic groups neutral theory and therefore stochasticity seem to have a bigger influence.

A different approach to formation and maintenance of microbial communities is ecological niche theory. As opposed to neutral theory, it assumes that species-specific traits are responsible for arising biodiversity and that different ecological niches in an ecosystem are occupied by species that differ in their characteristics [31]. Modeling of biological community dynamics based on ecological niche theory is, for example, done using generalized Lotka-Volterra equations [32]. Time-evolution of a number of N species is modeled in the form of:

$$\frac{dx_i(t)}{dt} = x_i(t)(b_i + \sum_{j=1}^{N} a_{ij}x_j(t))$$

where the abundance of species x_i at time t is calculated taking into account the specific growth rate b_i and the interactions between species i and species j a_{ij} that alter growth of x_i [33]. Generalized Lotka-Volterra equations have been used to model antibiotic administration to the gut microbiome and *C.difficile* infection, granting some insight into the underlying microbial network [34]. Another study by *Bucci et al.* [35], based on two coupled stochastic differential equations, also focused on a gut microbiome community disturbed by antibiotic treatment. It shows how antibiotic administration can lead to a hysteresis effect in the microbiota composition making recovery back to a healthy intestinal microbial community difficult. However, modeling not only a few taxa but a whole intestinal community complicates interpretation of modeling results.

Besides these population-level models, individual-based models are receiving more and more attention in the field of microbial ecology [36]. These models can be either well-mixed and not integrating space or they are spatial models that do take into account space. Individual-based models are usually computationally more demanding. They do have the benefit of accounting for interpopulation variability and they are able to link individual and population behavior allowing the emergence of population dynamics from modeling the individual level [37]. Individual-based modeling in microbial ecology has, among other areas, been successfully applied to biofilms

[38] and bioreactors [39].

The main concepts of self-organized criticality have also been incorporated into an individual-based model by Solé, Alonso and McKane in 2002 [40]. The authors describe the applicability of their model in various cases, for example as a predator-prey model or a neutral model. They also put forward a multispecies variety of the model with a focus on interactions, which shows typical features of SOC. This work showed that a model for SOC is able to capture system dynamics in macrobiological systems and is a useful tool for studying the community structure. Applying the concept to complex microbial communities could provide valuable information about their development and systematic behavior. We therefore chose this model by Solé, Alonso and McKane as a basis for our work on the gut microbial community.

1.4   Biological and microbial networks

When we aim to describe a biological or microbial system or community, a key instrument to represent and visualize these systems is the use of networks and, as Albert-László Barabási put it: "...networks are the prerequisite for describing any complex system..." [41]. The study of networks and their properties originated in mathematical graph theory, but proofed to be useful in various fields. Networks in principle are composed of nodes, which can be any entity or also trait and are connected with each other via directed or undirected links [42]. There are various network characteristics that can be analyzed and potentially linked to system features and dynamics, for example node degree, connectivity or betweenness centrality (for further details see 2.2). Of specific interest in biological systems are interaction networks between species. Those networks might for instance be foodwebs [43], pollination networks [44] or even social networks [45]. Causal networks derived from experimental data can reveal interaction patterns within a certain community. Experimental data composed of presence and absence data from species can however not be used to build causal networks. In this case, only data correlation is possible to build a correlational network. While it is possible and feasible in macrobiology to collect experimental evidence for causal interactions between different species or individuals, this poses a problem in microbiology. For microbial networks it is seldom possible to provide experimental evidence for causal interactions at the community level. Co-presence patterns are instead used to infer statistical correlations. They do not perfectly represent the actual biological interactions, but can be utilized to infer ecological trends and roles [46]. However, using presence and absence data can also lead to wrong conclusions. Common presence is not necessarily due to a common environmental driver. Double zeros in presence and absence data are even

more difficult to interpret. Various unrelated reasons can result in common absence, also pure stochasticity.

A model simulating a microbial community can close this gap to some extent as it allows computing of a causal interaction network. Our knowledge of causal interactions in the gut microbial community is limited. It is an area where modeling community composition and dynamics can provide insight. Inferring network structure from simulations can help understand the systematic behavior of the gut microbiome and its inherent dynamics. This interaction network from a modeled microbial community can reveal important interrelations between community structure and system dynamics.

1.5   Parametrization and model fitting

When biological parameters are not readily available, fitting a theoretical model to experimental data is one option. Moreover, this procedure allows for model validation. To assess the goodness-of-fit, the similarity of simulation and data needs to be queried. Various measures for calculating the similarity between time-series data are available: distance or divergence measures (e.g. Euclidean or Manhattan distance), correlation measures (e.g. Pearson's cross-correlation coefficient), transformation approaches based on principle component analysis (PCA) or Fourier transformation, or metric approaches (converting the time-series data into a number of parameters that sufficiently describe the data) [47]. Biological systems are mostly high dimensional with many interacting players generating multivariate time series. Due to their high dimensionality, a common approach is data compression, e.g. with PCA [48] or empirical mode decomposition (EMD) [49], to reduce the dimensions and then calculate similarity measures.

Time-series data from microbial gut communities, however, needs a somewhat different angle. Data fitting on the multivariate time series itself is in this case not feasible, because the variability in the gut composition and the fluctuation patterns between individuals are very high. A better suited approach is data fitting by means of measures that are able to capture the overall dynamics and structure of the data. For microbial communities, rank-abundance distributions showed to be a viable measure [50]. Those distributions display the species or OTUs ranked according to their relative abundance (see Figure 1).

After reducing the dimensions of the multivariate time-series one can apply a divergence measure to estimate the similarity between simulated and experimental data. Divergence measures are non-symmetric and used to compute the distance between two or more probability distributions (in this case rank-abundance distributions). For comparing two probability distributions a variety of divergence measures are available [51]. A divergence measure used in a various fields ranging from population biology [52] to

7

molecular biology [53] and also socioecology [54] is the Kullback-Leibler divergence. The general form of the Kullback-Leibler divergence is:

$$KL(P, Q) = \sum_i P(i) ln \frac{P(i)}{Q(i)}$$

KL(P,Q)...Kullback-Leibler divergence between P and Q
P,Q...probability distributions with i instances

[62]

where P is typically observed, experimental data and Q theoretical or modeled data. To validate the computed Kullback-Leibler divergence a goodness-of-fit measure suggested in [55] can be applied:

$$R^2 = 1 - \frac{KL(P,Q)}{KL(P,M)}$$

KL()...computed Kullback-Leibler divergence
P,Q...probability distributions
M...a vector with coordinates all equal to the mean of P

Finding model parameters that give a good approximation of the experimental data that we aim to describe is the next step after deciding on a divergence measure. Model parametrization is especially difficult when dealing with complex microbial communities. A vast parameter space is available and the objective is to find parameters which best describe the experimental data and therefore show the highest similarity in terms of the divergence measure. However, this is not a straightforward task. The parameter space is characterized as a rugged landscape (see figure 2 for an example) meaning that many combinations of parameters are, in terms of the landscape, locally optimal solutions, but globally there are better solutions. Ultimately, model parametrization aims for a global optimization. There are various techniques to find the global optimum of a complex optimization problem. One approach is called branch and bound, where upper and lower bounds are applied to a feasible region in the parameter space and then this feasible region is divided into several subregions. From the optimal solutions of all the subregions the best solution is selected. There are also clustering methods, which are multistart methods that cluster the starting points to avoid using redundant starting points that would end up in the same local optimum. Evolutionary algorithms are also used for global optimization. Those algorithms take their inspiration from natural selection. Different populations of starting points are formed and the solutions with the highest fitness in each population are then recombined. Solutions are also mutated by introducing small changes [56].

Another widely used approach to global optimization is simulated annealing, which is applied in fields as diverse as restoration ecology [57], protein structure prediction [58] or waste management [59]. Simulated annealing takes its inspiration from metallurgy, where the initial temperature of the metal and the cooling schedule are determining the energetic state of the cooled, solid metal. In simulated annealing, a parameter change is in principal accepted when it leads to a better solution. However, an artificial temperature term is introduced and the lower the temperature, the higher the probability that also a parameter change that results in a worse solution is accepted. Introducing this temperature term enables simulated annealing to navigate out of a local optimum and find the global optimum [60].
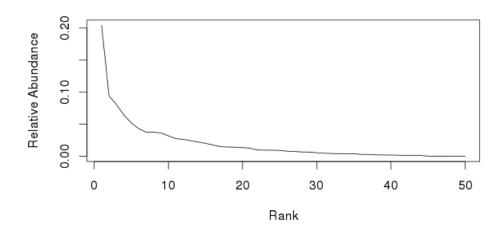
Figure 1: **Rank-Abundance Curve.** Species are ranked according to their relative abundance. The X-axis depicts the rank of the species, on the Y-axis the relative abundance is shown.
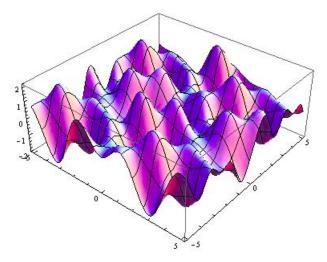

Figure 2: **Rugged landscape of the parameter space.** The goodness of a parameter combination is on the y-axis, while x- and z-axis represent adjustable parameters. There are locally and globally optimal solutions and also different parameter combinations that lead to comparably good solutions. Landscape computed in *Mathematica* [61]

1.6 Summary

With our work we aim to describe the temporal development and structural dynamics of the gut microbial community. The gut microbiome is a highly diverse and complex biological system. Temporal variability during the lifetime of a host, but also the high variability between hosts make it difficult to describe a typical intestinal microbial community. A healthy gut microbiome provides various important functions [8] for its host and dysbiosis has been linked to several diseases [15]. It is accordingly of crucial importance to understand the underlying structure and driving forces shaping this microbial community. Modeling the temporal dynamics and community structure is one possibly approach to aid this objective. Model approaches for biological communities reach from generalized Lotka-Volterra equations [34] to neutral models [29] and dynamic individual-based models [38].

We adapted a model for self-organized criticality (SOC) in macrobiological systems [40] to describe the gut microbial community. SOC is a concept that aims to describe arising complexity in far-from-equilibrium systems. Those systems are characterized by a constant flow-through of energy and the importance of the interaction network among the systems elements [1]. The gut microbiome is as well strongly influenced by the continuous flow-through of matter and energy and shaped by microbe-microbe and host-microbe interactions. Typical patterns found in systems exhibiting SOC are scale-free distributions of system feature lifetimes (e.g. earthquakes or extinction events in fossil records) and pink noise in the power spectral densities [18]. A challenge when modeling complex systems like the gut microbial community is model parametrization and verification. We chose to compute a Kullback-Leibler divergence [62] to compare multivariate time-series of microbial abundances from experimental data and data from simulations. For model parametrization we use generalized simulated annealing [60], a method for finding the global maximum of a multidimensional function.

2 Material and Methods

2.1 A model for self-organized criticality in the gut microbiome

The dynamical change and evolution of the gut microbiota is influenced and characterized by turn-over of the microbial mass, immigration and extinction of individual microbes, interactions between individuals that can either inhibit or facilitate growth, interactions between microbes and the host and characteristics of the host, e.g. diet patterns. A model aiming at simulating the time-evolution of this community should take into account the ability of microbial species to reproduce and colonize available space, which is in our model contained in a species-specific immigration probability. A species resilience and durability in the given environment as well as the stochastic loss of individual microbes due to the guts inherent continuous in- and outflow is combined in a species-specific extinction probability. Extinction probability also encompasses the hosts influence on microbial growth and resilience. Ecological interactions between microbes that shape the microbial community are accessible as species-specific interaction coefficients, reaching from mutualism and competition to predation.

We adapted a model for self-organized criticality in macrobiology introduced by Solé, Alonso and McKane in 2002: *Self-organized instability in complex ecosystems* [40]. We generalized it to include different types of symbiotic relationships between microbes, such as mutualism or competition. The model is individual-based, it uses discrete time-steps and deterministic mixing. During one given time step, an individual microbe can potentially interact with any other microbe and immigration as well as extinction solely depend on the immigration and extinction probabilities associated with the species and neither on biomass nor on spatial proximity. We further assume a finite number of N available sites for microbial colonization. Each site can be occupied by exactly one individual of one microbial species from the finite species pool S.

2.1.1 Simulation rules

Time evolution is simulated in a step-wise manner. During each time step $\triangle t$ every site N is visited and immigration and extinction rules are applied.

(I) **Immigration**
   occupation of an empty site by an individual of a randomly chosen species, i.e. A ϵ Σ(S), with probability μ_A

$$0 \xrightarrow{\mu_A} A$$

(I) **Extinction**
   a site occupied by an individual of species A can become empty with probability ε_A

$$A \xrightarrow{\varepsilon_A} 0$$

[40]

Every occupied site can now undergo interaction. The microbe on this site can interact with a randomly chosen second microbe belonging to a different species and occupying another site. Depending on the species-specific interaction coefficients of those two individuals, Ω_{AB} (the influence species B has on species A) and Ω_{BA} (the influence species A has on species B), there are different interaction possibilities:

(II) **Interaction**

(a) $\Omega_{AB}, \Omega_{BA} < 0$ and $\Omega_{AB} < \Omega_{BA}$

$A + B \xrightarrow{\Omega} 2B$
where $\Omega = \Omega_{BA} + abs(\Omega_{AB})$

(b) $\Omega_{AB} < 0, \Omega_{BA} = 0$

$A + B \xrightarrow{\Omega} 2B$
where $\Omega = \Omega_{BA} + abs(\Omega_{AB})$

(c) $\Omega_{AB} < 0, \Omega_{BA} > 0$

$A + B \xrightarrow{\Omega} 2B$
where $\Omega = \Omega_{BA} + abs(\Omega_{AB})$

(d) $\Omega_{AB} > 0$, $\Omega_{BA} = 0$
choose a third, random site

- if third site is empty

$$A + B + 0 \xrightarrow{\Omega} 2A + B$$
where $\Omega = \Omega_{BA} + abs(\Omega_{AB})$

- if third site is occupied by an individual of species C

$$A + B + C \xrightarrow{\Omega} 2A + B \qquad\qquad \text{if } \Omega_{CA} < \Omega$$
where $\Omega = \Omega_{BA} + abs(\Omega_{AB})$

(e) $\Omega_{AB}, \Omega_{BA} > 0$ and $\Omega_{AB} > \Omega_{BA}$
choose a third, random site

- if third site is empty

$$A + B + 0 \xrightarrow{\Omega} 2A + B$$
where $\Omega = \Omega_{BA} + abs(\Omega_{AB})$

- if third site is occupied by an individual of species C

$$A + B + C \xrightarrow{\Omega} 2A + B \qquad\qquad \text{if } \Omega_{CA} < \Omega$$
where $\Omega = \Omega_{BA} + abs(\Omega_{AB})$

For visualization of a model time-step see figure 3.

2.1.2 Post-processing of the simulated data

After a given amount of t time steps only a number of randomly filtered time steps are used for comparison to experimental data. We introduce this layer to account for stochasticity during experiments and collection of data, such as irregular sampling intervals, differing sequencing depth, various sequencing errors.
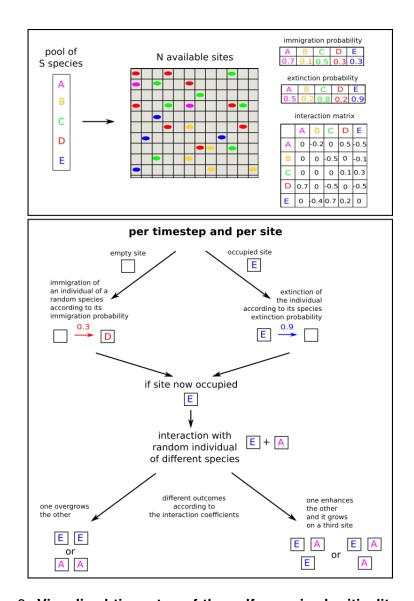
Figure 3: **Visualized time step of the self-organized criticality model.** (upper panel) Individuals from a vector of S species inhabit some of the N available sites. The properties of all S species are contained in the immigration probability vector μ, the extinction probability vector ϵ and the S*S interaction matrix Ω. (lower panel) Every time step, an empty site can become occupied by an individual of a randomly chosen species according to its immigration probability and an occupied site can become empty with the extinction probability of the inhabiting individual. Secondly, the individual on an occupied site can interact with another individual of a randomly chosen species according to their interaction coefficients

2.2 Community structure and network analysis

As it is yet unknown what topography underlies actual microbial networks, we chose two different, opposing approaches for structuring the interaction matrix Ω. This way, potentially occurring differences in system dynamics due to network structure should be easier to detect. On the one hand, Erdős–Rényi represents a random network with the degrees of the network nodes being normally distributed [63]. On the other hand, Klemm and Eguiluz introduced a network structure that does not only show a scale-free degree distribution, but also includes a high clustering coefficient and exhibits small-world properties [64].

The local community structure of interaction networks has a large impact on the emerging community properties, such as robustness and resilience [66]. While the interaction structure for simple microbial communities has been investigated for several case studies (e.g. [67], [68]), no reference system for intermediate to highly complex microbial communities such as the gut exists to our knowledge [2].

To account for this effect of interaction structure on the community dynamics we used the two different methods for structuring the community matrix Ω:

- **Erdős–Rényi**
  a random network, with the degrees (number of connections) of the network nodes (representing the species) being normally distributed [63], see figures 4 and 5

- **Klemm-Eguiluz**
  a small-world network, with a scale-free degree distribution and showing modularity [64], see figures 6 and 7
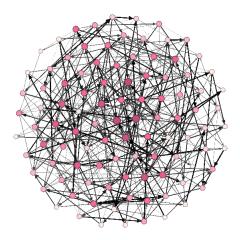
Figure 4: **Network from a community matrix with an underlying Erdős–Rényi structure.** Circles represent species, the size of the circle and the color intensity indicate the node degree; arrows represent a directed interaction between species, the thicker the arrow the more positive is the interaction coefficient. Network computed with *Gephi* [65]
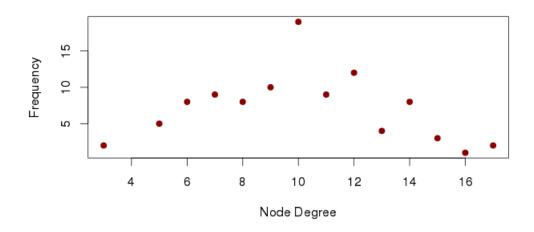

Figure 5: **Node degree distribution of the Erdős–Rényi network.** Node degrees are approximately normally distributed, the mean node degree is ~ 10
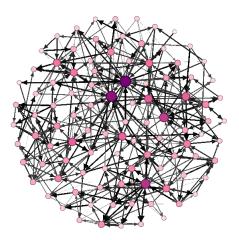
Figure 6: **Network from a community matrix with an underlying Klemm-Eguiluz structure.** Circles represent species, the size of the circle and the color intensity indicate the node degree; arrows represent a directed interaction between species, the thicker the arrow the more positive is the interaction coefficient. Network computed with *Gephi* [65]
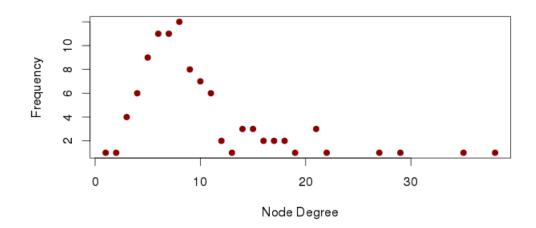

Figure 7: **Node degree distribution of the Klemm-Eguiluz network.** Node degrees show a skewed distribution, the mean node degree is also ~ 10

18

- **Connectivity**

  The connectivity of an interaction matrix is given by

  $$C = \frac{E}{N*(N-1)}$$

  E...number of possible edges (connections between nodes)
  N...number of nodes

  We set it to a value of 0.02, according to predicted networks from CoNet [69] (personal communication with Karoline Faust).

The following topological network measures were used to illustrate and compare the interaction structure of the community matrix Ω:

- **Node characteristics**

  – **Node Degree**

    The degree of a node k is the number of its connections (edges) to other nodes.

- **Local and wider neighborhood**

  – **Clustering Coefficient**

    The local clustering coefficient or transitivity gives the probability that the adjacent nodes of a node are connected to each other.

    $$T_n = \frac{\sum_{k_n}(e_{k_i k_j})}{k_n*(k_n-1)}$$

    n...node of a graph
    k_n...neighboring nodes of node n
    $e_{k_i k_j}$...edge from node k_j to node k_i

- **Navigability**

  – **Betweenness Centrality**
  Betweenness centrality BC is roughly defined by how many shortest paths lead through a given node.

  $$BC_n = \sum_{i \neq n \neq j} (\frac{\sigma_{inj}}{\sigma_{ij}})$$

  σ_{inj}...*number of shortest paths from node i to j passing through n*
  σ_{ij}...*number of shortest paths from node i to j*

  – **Closeness Centrality**
  Closeness centrality CC measures, how many steps it takes to get from one given node to every other node.

  $$CC_n = \frac{1}{\sum_{i \neq n} d(n,i)}$$

  d(n,i)...distance between node n and node i

[70]

We compared the computed networks with the *igraph* package in the R environment [70].

2.3   Divergence measure and model fitting

With our work we set out to capture the main driving forces and dynamics of the intestinal microbial community and establish a method to validate the model with respect to experimentally gained data. In order to quantify the overall similarity between simulated and experimental data, we use a Kullback-Leibler divergence as proposed in [62] and adapted it to:

$$KL(P,Q) = \sum_j \sum_i (P_{j,i} * log(\frac{P_{j,i}}{Q_{j,i}}) - P_{j,i} + Q_{j,i})$$

j...time-point
i...rank
P...mean ranks of experimental data
Q...mean ranks of simulated time-series

A Kullback-Leibler divergence gives an estimation on how much information is lost when using one probability distribution to approximate another probability distribution. We use rank-abundance distributions computed from the experimental as well as the simulated data as the probability distributions for the Kullback-Leibler divergence. Rank-abundance distributions are a commonly used way of characterizing microbial communities and representing their inherent dynamics [50]. This method ranks the species according to their relative abundance, resulting in a descending curve from the most abundant to the least abundant species (Figure 1).
We also applied a goodness of fit measure (R^2) to the computed divergence, as suggested in [55], in order to verify the calculated divergence:

$$R^2 = 1 - \frac{KL(P,Q)}{KL(P,M)}$$

KL()...computed Kullback-Leibler divergence
P...mean ranks of experimental data
Q...mean ranks of simulated time-series
M...a vector with coordinates all equal to the mean of P

For a parametrization of the SOC-model we applied model fitting. We used generalized simulated annealing, a heuristic optimization procedure, as introduced in [71] and implemented in the *GenSA* package [60] for locating a good approximation to the global optimum:

$$g_{q_v}(\triangle x(t)) \propto \frac{[T_{q_v}(t)]^{-\frac{D}{3-q_v}}}{[1+(q_v-1)\frac{(\triangle x(t))^2}{[T_{q_v}(t)]^{\frac{2}{3-q_v}}}]^{\frac{1}{q_v-1}+\frac{D-1}{2}}}$$

g_{q_v} is a distorted Cauchy-Lorentz visiting distribution, which is used to generate a trial jump distance $\triangle x(t)$ of the parameter $x(t)$ under the artificial temperature $T_{q_v}(t)$. This trial jump is accepted by the optimizer function if it improves the solution, i.e. if it decreases the Kullback-Leibler divergence measure. A trial jump that increases the divergence measure might be accepted according to an acceptance probability:

$$p_{q_v} = min\{1, [1-(1-q_a)\beta \triangle E]^{\frac{1}{1-q_v}}\}$$

Inspired by annealing in metallurgy, generalized simulated annealing introduces an artificial temperature $T_{q_v}(t)$, which influences the trial jump distance $\triangle x(t)$ and decreases with each artificial time-step t according to:

$$T_{q_v}(t) = T_{q_v}(1)\frac{2^{q_v-1}-1}{(1-t)^{q_v-1}-1} \text{ [60]}$$

Generalized simulated annealing enables the optimizer function to gradually minimize the Kullback-Leibler divergence by improving the parameter-set while still allowing backward steps (in terms of the divergence measure).

For fitting the parameters of the SOC-model to experimental data, 25 parallel simulations with the same parameter-set were computed. The relative abundances of this 25 computed time-series were then used to calculate the mean ranks of each time-step, which were needed for computing the Kullback-Leibler divergence. The heuristic optimization was stopped after an R^2 > 0.85 was reached. After this parametrization step, the best parameter-sets determined by the optimization function were used to simulate longer time-evolution, which was further analyzed and compared to the experimental data. For a detailed visualization of the parametrization procedure, see figures 8, 9 and 10.

All scripts for model implementation, parametrization and statistical analysis where developed in R [72].
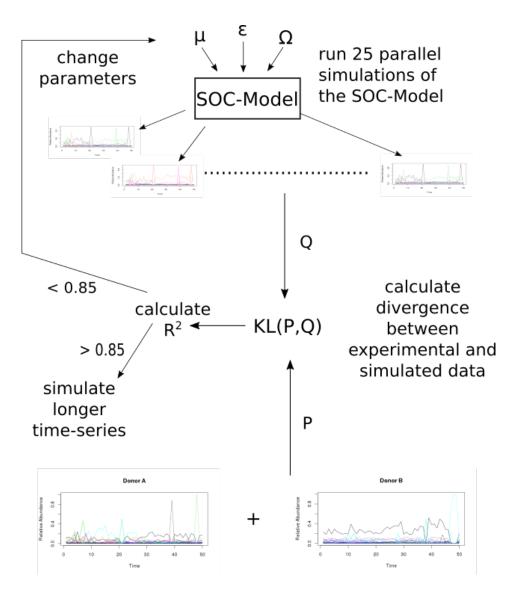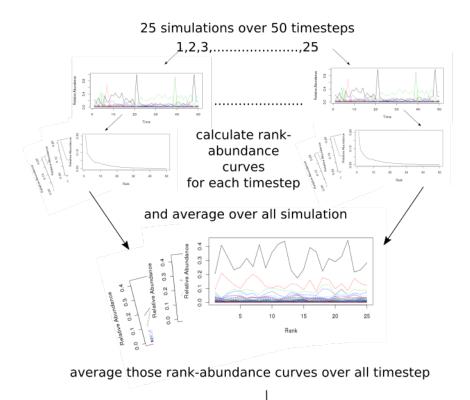
Figure 8: **Parametrization.** For parametrization averaged rank-abundances from 25 parallel model runs as well as averaged rank-abundances from the experimental data were computed and used as probability distributions for the Kullback-Leibler divergence measure. A goodness of fit measure, R^2, was calculated and used as a threshold for the optimization function
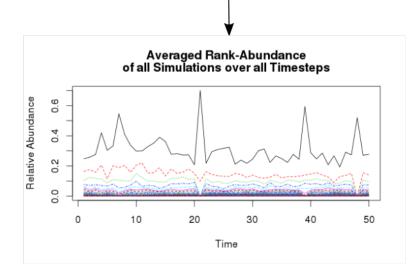
Figure 9: **Calculation of the averaged rank-abundance curve.** The relative abundances of 25 simulation runs were used to generate rank-abundance curves for each time-step of all 25 simulations. Those curves were then first averaged over all simulations and finally averaged over all time-steps resulting in the time evolution of the relative abundance of species ranks
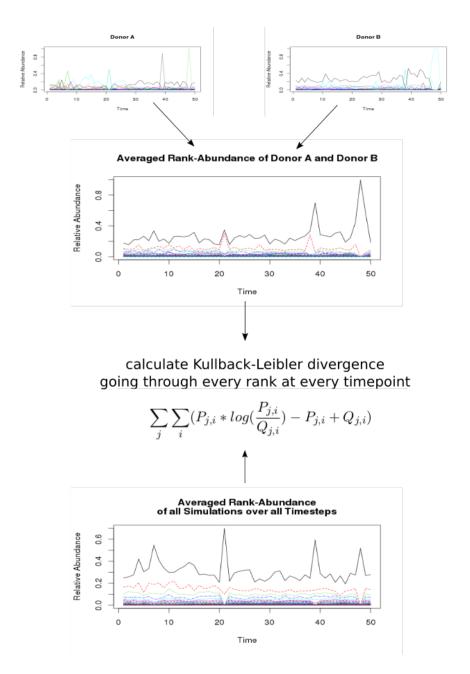
Figure 10: **Comparison of simulated and experimental data.** The averaged rank-abundances over time from simulated as well as experimental data, computed as shown in figure 9, are used to calculate the Kullback-Leibler divergence measure.

2.4 Experimental data

The experimental data used for parametrization and comparison to the simulated time evolution comprises longitudinal samples of 2 subjects over the course of one year [11]. In this study, high-throughput sequencing of amplified 16S rRNA was used to terize each sample. Reads were grouped into operational taxonomic units (OTUs) at 97% sequence similarity. After quality filtering, data from 4321 OTUs of 299 and 272 gut samples, respectively, were obtained from subject A and B.

For parametrization of the SOC-model we selected a subsample of the 150 most abundant OTUs, according to mean abundance over all samples. Those OTUs represent 80% of the sampled microbial population in terms of abundance. We performed an autocorrelation analysis to reveal possible recurring patterns in the longitudinal data and to determine the number of samples we then used for model fitting. Consistent with this analysis, we chose 50 time-steps with regular sampling intervals from both subjects for parametrization. In figures 11 and 12 relative abundances of the 150 most abundant OTUs over the chosen 50 time-steps are shown.
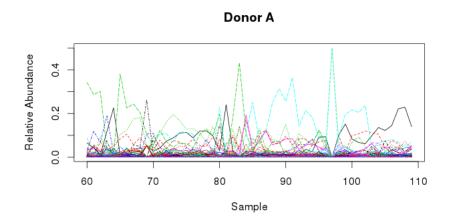
Figure 11: **Relative abundances of OTUs in samples from donor A.** Relative abundances of the 150 most abundant OTUs in samples from donor A over the 50 samples chosen for model fitting
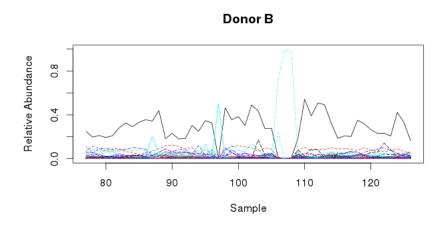

Figure 12: **Relative abundances of OTUs in samples from donor B.** Relative abundances of the 150 most abundant OTUs in samples from donor B over the 50 samples chosen for model fitting

27

2.5 Characterization of microbial time evolution

We examined the simulated and experimental time-series data for the following standard ecological properties:

- **Species Richness**

  Number of species present in a sample

- **Alpha-Diversity (Shannon Index)**

  Alpha diversity takes both species richness and evenness into account to compute a measure of diversity

  $$H_S = -\sum_{i}^{S} \frac{n_i}{N} * ln(\frac{n_i}{N})$$

  H_S...*alpha diversity for a community comprised of S species*
  N...total number of individuals
  n_i...*number of individuals of species i*

We also examined the data for properties that point towards self-organized criticality:

- **Scale-Invariance**

  We examined the lifetime distribution of OTUs in the experimental and simulated data as well as the lifetime distribution of individuals in the simulated data for scale-invariance. Scale invariant system features approximately follow a power-law distribution, therefore a linear model was fit to a log-log plot of the examined properties to evaluate whether they show scale-invariance [19].

- **Pink Noise**

  Pink noise, or 1/f noise, which is strongly associated with self-organized criticality [18], can be evaluated via a spectral density analysis that we performed with the function *spectrum* of the *stats-package* in the R environment [72]. The spectral density analysis relates the power spectra $P(f)$ of a signal to its frequency f as follows: $P(f) \propto f^{\beta}$ [20]. Pink noise is associated with an exponent β close to -1. By contrast an exponent close to 0 points towards a process showing white noise and an exponent close to -2 is associated with brown noise [21]. We analyzed the mean power spectra of all species combined as well as the power spectra of the individual species.

3   Results

3.1   Model parametrization for simulation of the gut microbiome

As biological parameters are not readily available, we fitted our theoretical model to experimental data. We computed a Kullback-Leibler divergence [62] between experimental and simulated data to estimate the similarity between them. For model parametrization we used generalized simulated annealing with the R package *GenSA* [60]. At first the optimizer function explores the available parameter space, as is shown in figure 13. During these fitting iterations, the computed Kullback-Leibler divergence varies strongly. After exploring the parameter space, the optimizer function starts with a more targeted parameter optimization, as is shown in figure 14. We used a goodness-of-fit measure (R^2) to validate the computed Kullback-Leibler divergence and set a threshold to R^2 = 0.85. If the goodness-of-fit measure was above this threshold, we used the fitted parameters to simulate time evolution of the microbial gut community. We compared simulations with fitted parameters from fitting runs that reached the targeted optimization regime and from fitting runs that reached R^2 > 0.85 during the parameter space evaluation. Simulations with fitted parameters from both optimization regimes resulted in very similar temporal dynamics. Due to this observation we chose to accept all parameter sets that reached a goodness-of-fit measure above the chosen threshold, even if the solution was found during parameter space evaluation.
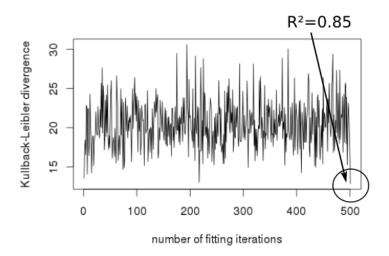
Figure 13: **Change of Kullback-Leibler divergence during one fitting run.** The optimizer function explores the available parameter space, an R^2 = 0.85 is reached after 501 iterations
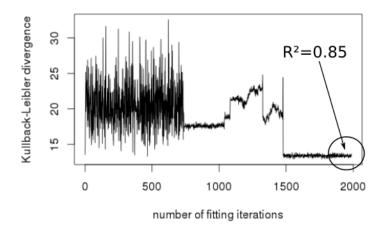

Figure 14: **Change of Kullback-Leibler divergence during one fitting run.** The optimizer function first explores the available parameter space and then starts targeted optimization, an R^2 = 0.85 is reached after 1987 iterations

3.2 Analysis of experimental and simulated data

3.2.1 Temporal dynamics of the microbial community

To gain intuition about the data, experimental as well as simulated, we compared the time evolution of relative species abundances in a qualitative manner.

Figure 15 and figure 16 show the time evolution of the 150 most abundant OTUs in experimental survey data [11]. All time-points represent samples taken from one of two different subjects (donor A and donor B). In the experimental time-course temporal fluctuation within single OTUs occur frequently and some species experience peaks of high relative abundance. The higher abundant species form a subgroup that is relatively stable in composition over time. The majority of species are however very low in relative abundance, with the overall mean being 0.007 for samples from donor A as well as donor B. Those species experience a high turnover and introduce a lot of fluctuation into the community.
Some time-points are effectively dominated by species showing extremely high relative abundances (between 0.8 and 1) in specific samples. To make the variability over time in the relative abundances more visible, those species were omitted in figures 15 and 16. The microbiome exhibits self-similar temporal dynamics as shown in figure 17. This means that on different time-scales and within different community sizes, time evolution shows similar patterns.

In simulated data, with random community structure (Erdős-Rényi topography underlying the interaction network) as well as clustered community structure (Klemm-Eguiluz topography underlying the interaction network), also high fluctuations on species level can be seen (figures 18 and 19). Again the majority of species is very low in abundance, with the overall means in relative abundance from both simulation approaches being exactly identical to the abundance means from experimental data, 0.007. Within both simulation approaches one species always developed very high relative abundances (omitted in figures 18 and 19). As simulation parameters associated with one species were fixed throughout the simulation, this pattern was consistent. This dominance structure in simulations lead to lower relative abundances as compared to the experimental data.

In both experimental and simulated data, individual species experience high fluctuations over time. Community composition within the higher abundant species, however, stays relatively constant. The majority of species shows very low relative abundances, with the mean in relative abundances being 0.007, identical for experimental as well as simulated data. This indi-

cates that the overall trend in time evolution observable in experimental data was well captured in simulated data. In experimental data single species show peaks of very high abundance in certain time-points. This pattern was not captured in simulations, but resulted in one or two species in each simulation dominating in terms of relative abundance. For computing the similarity between experimental and simulated data we used rank-abundances. Several time-points in the experimental data showed a dominating rank, which lead to the dominance structure we now observe in the simulated data.
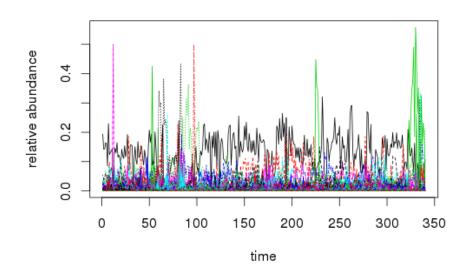

Figure 15: **Relative species abundances of 150 most abundant OTUs in all samples from donor A.** Individual species experience high fluctuation over time, with visible peaks of higher abundances in certain time-points. Mean relative abundance is 0.007. Species with peaks showing very high abundances were omitted to make dynamics more visible
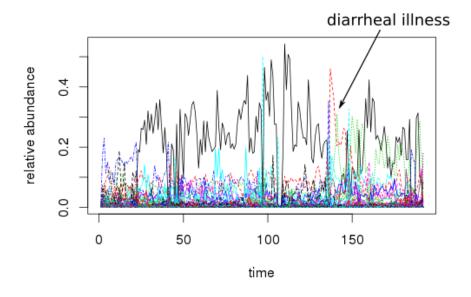
Figure 16: **Relative species abundances of 150 most abundant OTUs in all samples from donor B.** Individual species experience high fluctuation over time, with visible peaks of higher abundances in certain time-points. Diarrheal illness of the donor is only slightly visible in species composition. Mean relative abundance is 0.007. Species with peaks showing very high abundances were omitted to make dynamics more visible
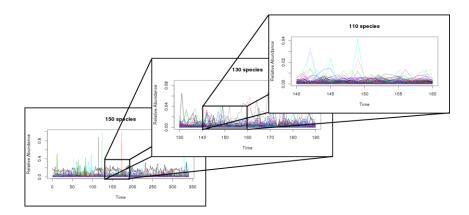

Figure 17: **Self-similar temporal dynamics in relative species abundances of donor A.** On different time-scales and within different community sizes, time evolution shows similar patterns
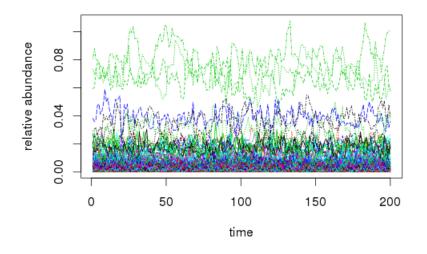
33

Figure 18: **Relative species abundances over time, simulation with random community structure.** Individual species experience high fluctuation over time. Mean relative abundance is 0.007. Species with very high abundance were left out to make dynamics more visible
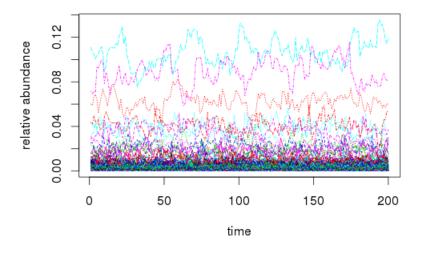

Figure 19: **Relative species abundances over time, simulation with clustered community structure.** Individual species experience high fluctuation over time . Mean relative abundance is 0.007. Species with very high abundance were left out to make dynamics more visible

3.2.2 Analyzing ecological properties of the gut microbiome

We examined typical ecological properties of experimental and simulated data. Those properties were species richness and alpha diversity, which can be used to determine how well the model captures overall features of the experimental data.

In experimental data, the microbial community is characterized by relatively constant species richness and alpha diversity in both donor A and donor B. At certain time-points, the community was shortly dominated by a single OTU in terms of relative abundance. These time-points correspondingly show a species richness of below 10 (figures 20 and 22).
A diarrheal illness of donor B caused a disruption in the community structure that resulted in a less diverse community. This is clearly visible around time-point 140 in figure 22. It is also reflected in a lower mean species richness of 108 in donor B's gut microbiome. Mean species richness in the gut microbial community of donor A is 131. The diarrheal illness is not as prevalent in the time evolution shown in figure 16, as mostly species with lower abundances were affected.
Alpha diversity, incorporating both species richness and evenness in the microbiome, shows a pattern similar to species richness in both donors (figures 21 and 23). Again, the less diverse community caused by diarrheal illness of donor B is observable. Mean alpha diversity is also slightly higher in samples from donor A (3.5) than in samples from donor B (3).
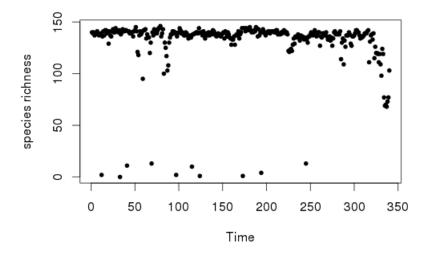
Figure 20: **Species richness over time in the gut microbial community of donor A.** Species richness is relatively constant over time, but the microbial community is dominated by single OTUs at certain time-points
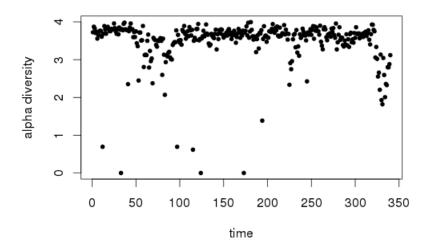

Figure 21: **Alpha diversity over time in the gut microbial community of donor A.** Alpha diversity follows a similar pattern as species richness and is relatively constant over time, but the microbial community is dominated by single OTUs at certain time-points
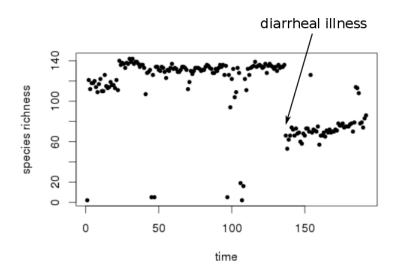
Figure 22: **Species richness over time in samples from donor B.** Species richness is relatively constant over time, but the microbial community is dominated by single OTUs at certain time-points. Clearly visible is a drop in community diversity due to a diarrheal illness of the donor around time-point 140
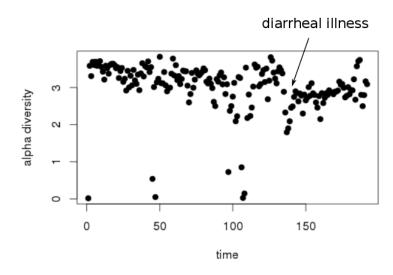

Figure 23: **Alpha diversity over time in samples from donor B.** Alpha diversity follows a similar pattern as species richness and is relatively constant over time, but the microbial community is dominated by single OTUs at certain time-points. Again observable is the drop in community diversity due to a diarrheal illness of the donor around time-point 140

37

In simulated data, species richness as well as alpha diversity are very constant within a single simulation run. There is, however, less variation between simulation runs in simulations with random community structure (figures 24 and 25) than in simulations with clustered community structure (figures 26 and 27). Mean species richness within one simulation varies between 121 and 129 in simulations with random community structure and between 87 and 130 in simulations with clustered community structure. Mean alpha diversity in simulations with random community structure varies between 2.2 and 2.8 and in simulations with clustered community structure between 0.9 and 3.

Due to one species in simulations always dominating the community, alpha diversity is generally lower than in experimental data. This is particularly pronounced in one simulation from a clustered community structure (black dots in figures 26 and 27). The dominating species in this simulation shows relative abundances around 0.85 and leads to lower species richness and alpha diversity.

We assume that those differences in variability between simulation runs are an effect of the different topographies that underlie the interaction matrices. The node degrees in interaction matrices of simulations with random community structure are normally distributed and lead to simulations showing very similar features and dynamics. The structure added to the node degree distribution in interaction matrices of simulations with clustered community structure (skewed distribution, small-world properties) allows for more variation between simulation runs and leads to differences in community dynamics.
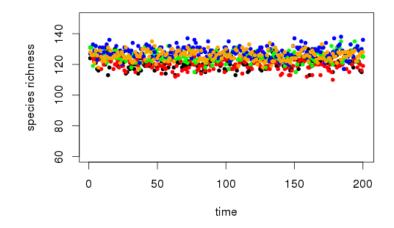
Figure 24: **Species richness over time of simulations with random community structure.** The different colors represent simulations from different fitting runs. Node degrees are normally distributed in interaction matrices of all simulations. This leads to very similar patterns in community diversity and is visible in the similarity of species richness.
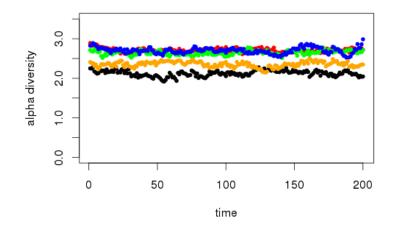

Figure 25: **Alpha diversity over time of simulations with random community structure.** The different colors represent simulations from different fitting runs. Node degrees are normally distributed in interaction matrices of all simulations. This leads to very similar patterns in community diversity and is also visible in the similarity of alpha diversity
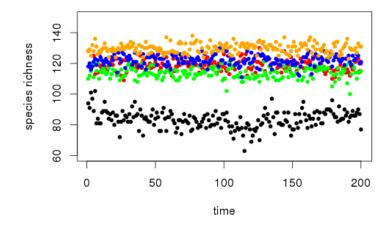
39

Figure 26: **Species richness over time of simulations with clustered community structure.** Different colors represent simulations from different fitting runs. A skewed distribution of node degrees and small-world properties in the interaction network lead to greater variability between individual fitting runs. This leads to the observable differences in species richness.
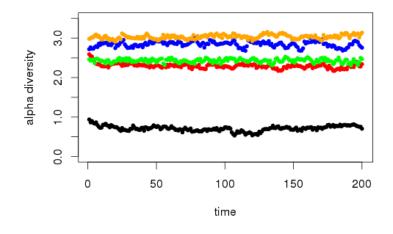

Figure 27: **Alpha diversity over time of simulations with clustered community structure.** Different colors represent simulations from different fitting runs. A skewed distribution of node degrees and small-world properties in the interaction network lead to greater variability between individual fitting runs. This also leads to the observable differences in alpha diversity.

In experimental and simulated data, the diversity of the microbial community is relatively stable over time. Mean species richness in simulated data overall ranges from 87 to 130, which is comparable to the mean species richness of 108 and 131 found in experimental data. Species richness and alpha diversity differ over time in the experimental data, but they also differ between donor A and donor B. This variability between two different microbial communities is better captured by simulations with clustered community structure. Those simulations show higher variability between individual fitting runs than simulations with random community structure.

To our knowledge, there is yet no reference system for the interaction structure of complex microbial communities. Our findings suggest that the high variability between the gut microbial communities of different hosts results from differently clustered community structures.

Differences between experimental data and simulations that are visible in the time evolution of the community (see section 3.2.1) are also reflected in species richness and alpha diversity. In terms of relative abundance, the microbial community in the experimental data is dominated by a single species in certain time-points. The microbial communities in simulated data are, however, dominated by one very highly abundant species over all time-points.

We chose rank-abundances for computing the similarity between experimental and simulated data. This measure does not capture this very pronounced short-term fluctuations within individual species present in the experimental data. Instead, parameters leading to very high abundances are fitted to the first rank, which leads to a different dominance structure in the data from simulations. This differing dominance structure also leads to a lower and more constant alpha diversity in all simulations as compared to the experimental data. The reduced diversity of the gut microbiome of donor B due to diarrheal illness was not captured in simulations, as the combined data from donor A and donor B was used for model parametrization.

3.2.3 Power-laws in lifetime distributions of intestinal microbial species

We investigated the distribution of species lifetimes over the time evolution in experimental and simulated data. The duration of a species lifetime can be seen as an event in the temporal development of a system. Plotting the lifetime of all microbiota on a logarithmic scale might reveal certain properties of the system. Specifically, a power-law distribution of those lifetimes is an important feature associated with self-organized criticality [1].

In experimental data, lifetime distributions of species in the microbial community from both donor A and donor B follow a power-law. The linear model fitted to the data gives a slope of -2.29, $p < 0.0001$ and R^2 = 0.83 (see figure 28, for statistics see table 1).

Species lifetime distributions from both simulation approaches follow a power-law as well. We also examined the influence of the community organization on the lifetime distribution of simulated gut microbiota. The linear model fitted to simulations with random community structure gives a slope of -1.31, $p < 0.0001$ and R^2 = 0.80 (see figure 29, for statistics see table 1). The linear model fitted to simulations with clustered community structure gives a slope of -1.53, $p < 0.0001$ and R^2 = 0.89 (see figure 30, for statistics see table 1).

Both in the experimental data as well as data from simulations species lifetimes show a power-law distribution. The slope of the linear model fit is however steeper in the experimental data. This suggests that in the experimental data species experience a higher number of extinction events than the species in the simulated data. This might be due to external influences affecting the microbial community, which were not explicitly implemented in the model. Species-specific model parameters were constant over time. They do not include possible changes in the environment, such as dietary changes or immune responses of the host, that might lead to certain species experiencing a drop in abundance.
The simulations with a clustered community structure showed a slope in the linear model fit closer to experimental data than the simulations with a random community structure. This indicates that the interaction network of the microbiota in the experimental data is most likely not random, but highly structured.

Power-laws in distributions of system features are associated with scale-freeness and self-organization [19]. The power-law distributions we found in lifetimes of the gut microbial community are a first indication that it is a self-organized critical system.
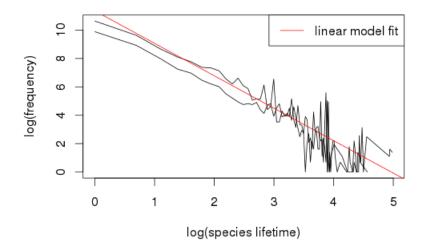
Figure 28: **Species lifetime distribution in experimental data.** Distribution of species lifetimes in a log-log plot follow a power-law with a slope of -2.29. The red line shows the linear model fit, for statistics see table 1
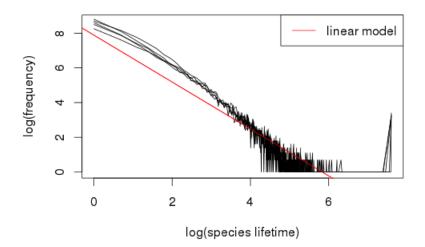

Figure 29: **Species lifetime distribution in simulations with random community structure.** Distribution of species lifetimes in a log-log plot follow a power-law. The slope of -1.31 is less steep than in experimental data. The red line shows the linear model fit, for statistics see table 1

Figure 30: **Species lifetime distribution in simulations with clustered community.** Distribution of species lifetimes in a log-log plot follow a power-law. The slope of -1.53 is also less steep than in experimental data. The red line shows the linear model, for statistics see table 1

Table 1: Lifetime distributions: linear model statistics

	slope	p	adjusted R^2
experimental data	-2.29	< 0.0001 ***	0.83
random community structure	-1.31	< 0.0001 ***	0.80
clustered community structure	-1.53	< 0.0001 ***	0.89

3.2.4 Pink noise in power spectral densities of the gut microbiome

On experimental data and data from simulations we performed spectral density analysis. Power spectral densities show the distribution of variability within a temporal signal. Examining these power spectral densities on a logarithmic scale can give information about system properties. If a linear model fitted to this distribution results in a slope $\beta \sim -1$, the system exhibits 1/f noise (= pink noise). Pink noise is the most important system feature associated with self-organized criticality [18] and characteristic for systems that are shaped mostly by internal structuring. A slope $\beta \sim 0$ suggests that a system exhibits white noise, which is an indication for a strong influence of stochasticity. A slope $\beta \sim -2$ suggests a system exhibiting brown noise, an indication for a strong deterministic driving force [21].

We investigated the distribution of power spectral densities in individual species of experimental data. This revealed that species differ in the type of noise they exhibit. Many species stratify along this measure into subgroups and exhibit either white, pink or brown noise. The largest of those subgroups of species exhibits white noise (N = 4163). Only a small number of species exhibit brown noise (N = 56). When closely examining the time evolution of this subgroup, they showed similar behavior to species exhibiting white noise. Due to many time-points where those species exhibiting brown noise where not found, the computation of power spectral densities might be difficult. We do believe that the classification of those species into the brown noise subgroup is an artifact. They do not represent the deterministic dynamics classically associated with brown noise. In the further analysis we therefore focus on the subgroups of species exhibiting either white or pink noise.

The noise type these subgroups exhibit also correlates with mean relative abundance of the species (see figure 31). The subgroup of species exhibiting white noise shows significantly lower mean species abundances than the subgroup of species exhibiting pink noise. White noise is associated with high stochastic influence and it seems intuitive that the low abundant species are those exhibiting white noise. They experience high turnover and strong temporal fluctuations and get frequently washed out of the system. This is observable in their shorter lifetimes (see figure 32). It can also be seen in their higher ratio between mean abundance and standard deviation of abundance (sd/mean, see figure 33). The higher this ratio, the higher is the fluctuation in abundance over time relative to the mean abundance. A higher sd/mean ratio is an indication for stronger stochastic influences.

The subgroup exhibiting pink noise consists of higher abundant species. Pink noise is associated with systems where the interaction patterns between elements are driving the temporal evolution and stochasticity plays a comparably smaller role. This is also in agreement with the patterns we

find in species exhibiting pink noise in our experimental data. They are more persistent and are less frequently washed out. This results in longer species lifetimes (see figure 32) and lower sd/mean ratios (see figure 33).

We also examined mean power spectral densities in experimental data. A linear model fitted to those mean power spectral densities gives an exponent β = -0.97 (see figure 34, for statistics see table 2). The majority of species in the gut microbiome might be strongly influenced by stochasticity and therefore exhibit white noise, but the microbial community as a whole exhibits pink noise. This suggests that the subgroup of species that exhibit pink noise influences overall system behavior the most. Those species, which are more persistent in the community, seem to strongly shape the gut microbiome.
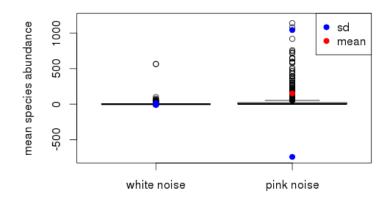

Figure 31: **Mean relative species abundances in experimental data.** Differences of mean species abundances between subgroups of species exhibiting white (N = 4163) and pink noise (619), respectively. Species exhibiting white noise show significantly lower mean relative abundances than species exhibiting pink noise; Wilcox-test: $p < 0.0001$ ***
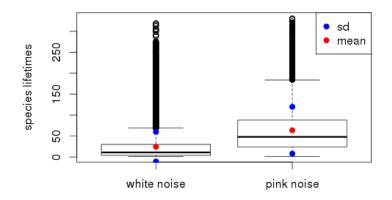
Figure 32: **Species lifetimes in experimental data.** Differences in lifetimes between subgroups of species exhibiting white noise (N = 4163) and pink noise (N = 619), respectively. Species exhibiting white noise show significantly shorter lifetimes than species exhibiting pink noise; Wilcox-test: p < 0.0001 ***
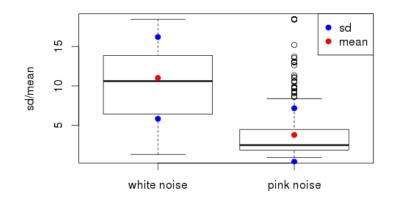

Figure 33: **Ratio between standard deviation and mean (sd/mean) of species abundances in experimental data.** Differences in sd/mean ratio between subgroups of species exhibiting white noise (N = 4163) and pink noise (N = 619), respectively. The ratio is significantly higher for species exhibiting white noise, meaning that fluctuations in abundance relative to the mean abundance are higher; Wilcox-test: p < 0.0001 ***
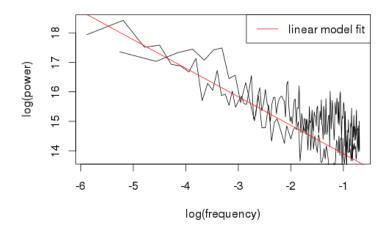
47

Figure 34: **Mean power spectra in experimental data.** Mean power spectral densities from experimental data in a log-log plot. The slope β = -0.97 indicates pink noise. The red line shows the linear model, for statistics see table 2

Table 2: Mean power spectral densities: linear model statistics

	β	p	adjusted R^2
experimental data	-0.97	< 0.0001 ***	0.74
random community structure	-1.19	< 0.0001 ***	0.97
clustered community structure	-1.17	< 0.0001 ***	0.97

In data from simulations, many species also stratified in terms of the noise they exhibit. We find subgroups of species exhibiting either white, pink or brown noise. In both simulation approaches these subgroups correlate with the mean species abundance (see figures 35 and 36). We see that low abundant species that frequently get washed out exhibit white noise. The high turnover within this subgroup of species is most likely due to the high flow-through in the gut microbiome. This leads to strong stochastic fluctuations in the abundances of those species. In the data from simulations this is also reflected in their short lifetimes (see figures 37 and 38) and

48

their high sd/mean ratios (see figures 39 and 40).

Species showing the highest abundances (see figures 35 and 36) fall into the subgroup of species exhibiting brown noise. Brown noise is thought to be exhibited by systems that are influenced by deterministic drivers. In the gut microbiome, the influence of the host presumably represents this deterministic drivers and seems to have the strongest effect on species showing brown noise. They are almost continuously present and get hardly washed out of the system. This is apparent in their long lifetimes (see figures 37 and 38) and low sd/mean ratio (see figures 39 and 40).

Species that have intermediate abundances (see figures 35 and 36) exhibit pink noise. Systems exhibiting pink noise are characterized mainly by the interaction structure between their elements. We assume that this also holds for the gut microbiome. Microbe-microbe interactions would therefore have the strongest effect on the temporal evolution of species exhibiting pink noise. Stochastic influences like flow-through and deterministic drivers like host-microbe interactions have less of an effect on this subgroup of species. When looking at their intermediate lifetimes (see figures 37 and 38) and sd/mean ratios (see figures 39 and 40), we can observe these reduced influences.

Those described patterns are very similar for simulations with random community structure and simulations with clustered community structure. An apparent difference in simulation approaches is the number of species within the subgroups exhibiting a certain noise type. These numbers are much smaller for simulations with random community structure. Simulations with clustered community structure have a higher diversity in interaction patterns between species. This added diversity seems to result in more species stratifying into subgroups exhibiting a certain noise type.

We could also find correlations between the noise type a subgroup exhibits and model parameters. These results will be discussed in section 3.2.5.
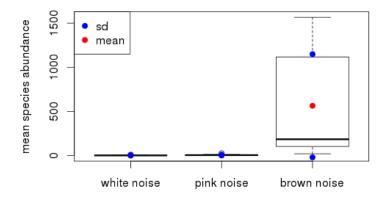
Figure 35: **Mean species abundances in simulations with random community structure.** Differences in mean abundances of species showing white noise (N = 49), pink noise (N = 63) and brown noise (N = 13), respectively. Species exhibiting different noise types show significant differences in their mean abundances; Kruskal-Wallis test: $p < 0.0001$ ***

Figure 36: **Mean species abundances in simulations with clustered community structure.** Differences in mean abundances of species showing white noise (N = 135), pink noise (N = 155) and brown noise (N = 52), respectively. Species exhibiting different noise types show significant differences in their mean abundances; Kruskal-Wallis test: $p < 0.0001$ ***
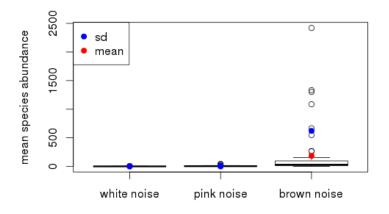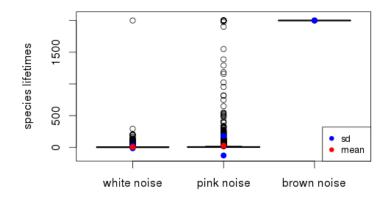
Figure 37: **Species lifetimes in simulations with random community structure.** Differences in lifetimes between subgroups of species exhibiting white noise (N = 49), pink noise (N = 63) and brown noise (N = 13), respectively. Species exhibiting different noise types vary significantly in their lifetimes; Kruskal-Wallis-test: $p < 0.0001$ ***
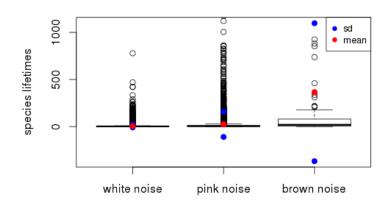

Figure 38: **Species lifetimes in simulations with clustered community structure.** Differences in lifetimes between subgroups of species exhibiting white noise (N = 135), pink noise (N = 155) and brown noise (N = 52), respectively. Species exhibiting different noise types vary significantly in their lifetimes; Kruskal-Wallis-test: $p < 0.0001$ ***
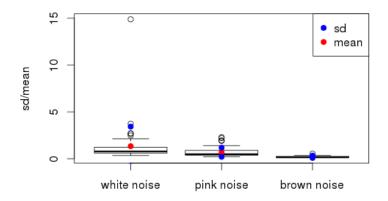
Figure 39: **Ratio between standard deviation and mean (sd/mean) of species abundances in simulations with random community structure.** Differences in sd/mean ratio of species showing white noise (N = 49), pink noise (N = 63) and brown noise (N = 13), respectively. Subgroups vary significantly in their sd/mean ratios; Kruskal-Wallis test: $p < 0.0001$ ***
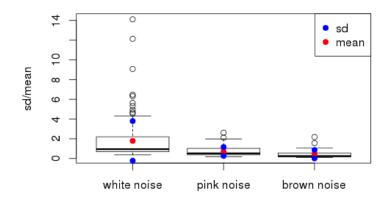

Figure 40: **Ratio between standard deviation and mean (sd/mean) of species abundances in simulations with clustered community structure.** Differences in sd/mean ratio of species showing white noise (N = 135), pink noise (N = 155) and brown noise (N = 52), respectively. Subgroups vary significantly in their sd/mean ratios; Kruskal-Wallis test: $p < 0.0001$ ***

In the data from simulation, we again examined mean power spectral densities. Linear models fitted to mean power spectral densities from both simulation approaches resulted in an exponent β close to -1. For simulations with random community structure β = -1.19 and for simulations with clustered community structure β = -1.17 (see figures 41 and 42, for statistics see table 2). Both exponents indicate that simulations overall exhibit pink noise and that species exhibiting pink noise are mainly shaping the temporal evolution of the gut microbial community. We argue that this means, microbe-microbe interactions have the strongest influence on community structure and behavior and make the gut microbiome robust against stochastic fluctuations and influences from deterministic drivers.
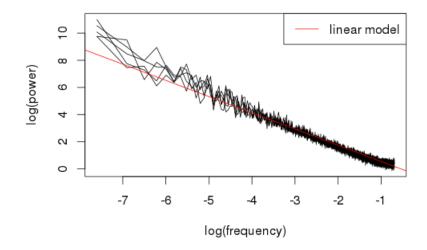

Figure 41: **Mean power spectra of simulations with random community structure.** Mean power spectral densities from different realizations of simulations with random community structure in a log-log plot. The slope β = -1.19 indicates pink noise. The red line shows the linear model, for statistics see table 2

Figure 42: **Mean power spectra of simulations with clustered community structure.** Mean power spectral densities from different realizations of simulations with clustered community structure in a log-log plot. The slope β = -1.17 indicates pink noise. The red line shows the linear model, for statistics see table 2
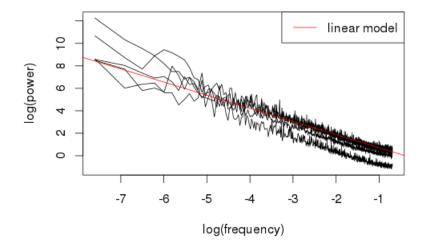
In both simulated and experimental data we observe that species stratify into subgroups exhibiting a certain noise type. The mean power spectral densities show that also both simulated data and experimental data overall exhibit pink noise. We do find species exhibiting brown noise in data from simulations. In the experimental data, however, we assume that species showing brown noise is only an artifact. In the model, we did not take into account frequently changing external influences, such as dietary changes or immune responses of the host. These additional sources of stochastic fluctuation influencing the whole microbial community might be the reason why species exhibiting real brown noise are not present in experimental data.

The results from examining power spectral densities suggest that neither stochastic fluctuations (flow-through) nor deterministic external drivers (host-microbe interactions) have the strongest effect on time evolution in the gut microbial community. The pink noise exhibited by the gut microbiome as a system indicates that system inherent structures, the microbe-microbe interactions, shape the system behavior.

54

3.2.5 How immigration and extinction shape noise and abundance in the simulated microbial community

We further examined how model parameters influence dynamical behavior of species in simulated data and to what extent those parameters are linked to the type of noise a species exhibits.

- Immigration probability

  Immigration probabilities were in both simulation approaches uniformly distributed among species. Immigration probability mainly represents a species ability to colonize available space and niches within the gut. This ability is highly influenced by the strength of stochastic influences, such as continuous flow-through. The better a species colonization ability, the less susceptible it is to stochastic fluctuations. Immigration probabilities are in neither of our two simulation approaches significantly different between subgroups of species exhibiting different types of noise. There are, however, trends visible within subgroups of species exhibiting white or pink noise. Within those subgroups, species with a better colonization ability, which are less influenced by stochasticity, reach higher abundance. We see these trends for both simulation approaches in the distribution of immigration probabilities within those subgroups (see figures 43 and 44). It is particularly visible in simulations with random community structure (figure 43) as there are less species that stratified into the subgroups. Temporal evolution in the subgroup of species exhibiting brown noise was in neither of the simulation approaches strongly shaped by their ability to colonize available space. Correspondingly we see no correlation between immigration probability and abundance in those species.

  These results suggest that a species ability to colonize space and niches within the gut is to some extent shaping the temporal dynamics of certain subgroups of species. Namely, we see a correlation between immigration probability and abundance in species exhibiting pink and white noise, respectively. On species showing the highest abundances within the gut microbiome, which exhibit brown noise, this property has no visible effect.

Figure 43: **Immigration probabilities in simulations with random community structure.** Mean species abundances and corresponding immigration probabilities of species showing either white or pink noise. Within both subgroups we see a trend that mean species abundance increases with increasing immigration probability
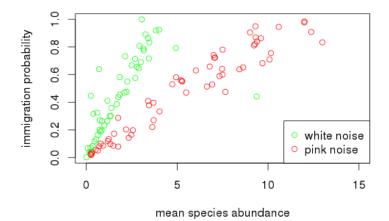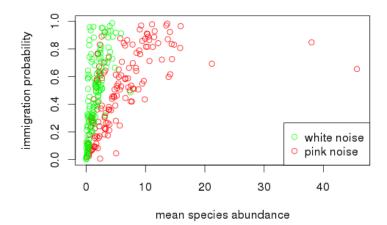

Figure 44: **Immigration probabilities in simulations with clustered community structure.** Mean species abundances and corresponding immigration probabilities of species showing either white or pink noise. Again we see a trend within both subgroups that mean species abundance increases with increasing immigration probability

56

- Extinction probability

  Extinction probabilities were in both simulation approaches uniformly distributed among species. Extinction probability reflects how well a species is adapted to its environment. This particularly encompasses how well it interacts with the host and how strong the hosts impact is on the species temporal dynamics. For both simulation approaches we find that strength of the hosts impact differs significantly between subgroups of species exhibiting different noise types. The external influence of the host is weakest in species showing white noise and strongest in species showing brown noise (see figures 45 and 46). We also find that the correlation between external influence and abundance is strongest in species exhibiting brown noise and weaker in the subgroups exhibiting pink or white noise. A linear model fitted to a log-log plot of mean species abundance and extinction probability visualizes that species exhibiting brown noise follow the trend line more than other species (see figures 47 and 48, for statistics of the linear model fits see table 3). Particularly the temporal evolution of low abundant species exhibiting white noise is strongly shaped by stochastic fluctuations. This could explain why the hosts external influence has only little effect on their behavior.
  The differences in external influence on species from different subgroups is more pronounced in simulations with random community structure. In those simulations less species stratified into noise subgroups. The model parameters of species within one subgroup are also very similar. We suggest that the added topography in the interaction network of simulations with clustered community structure also allows for a higher variability in model parameters that still lead to a species exhibiting a certain noise type.

  The influence of external drivers that shape the gut microbial community, particularly host-microbe interactions, varies greatly between different subgroups of species. Low abundant species exhibiting white noise are less influenced by the host than very high abundant species that exhibit brown noise. In the high abundant species that are most affected by the hosts influence we also find a clearer correlation between strength of external influence and species abundance.

Figure 45: **Extinction probabilities in simulations simulations with random community structure.** Differences in extinction probabilities of species showing white noise (N = 49), pink noise (N = 63) and brown noise (N = 13), respectively. The subgroups differ significantly in their extinction probabilities; Kruskal-Wallis test: p-value < 0.0001 ***
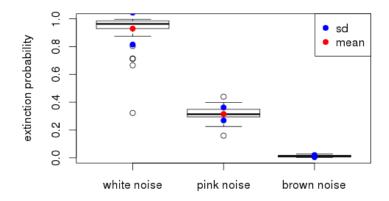

Figure 46: **Extinction probabilities in simulations with clustered community structure.** Differences in extinction probabilities of species showing white noise (N = 135), pink noise (N = 155) and brown noise (N = 52), respectively. The subgroups differ significantly in their extinction probabilities; Kruskal-Wallis test: p-value < 0.0001 ***
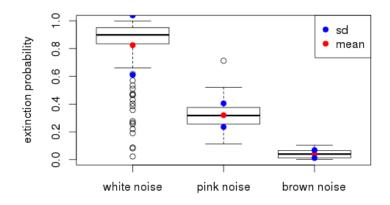
Figure 47: **Extinction probabilities in simulations with random community structure.** Mean species abundances and corresponding extinction probabilities of species showing either white, pink or brown noise on a logarithmic scale. The black line represents a linear model fit. Species exhibiting brown noise show a trend corresponding to the linear model
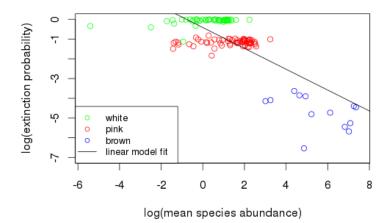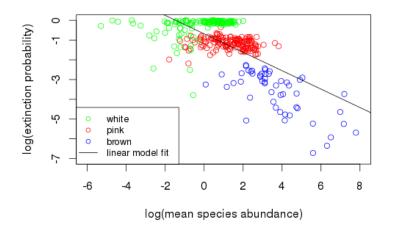

Figure 48: **Extinction probabilities in simulations with clustered community structure.** Mean species abundances and corresponding extinction probabilities of species showing either white, pink or brown noise on a logarithmic scale. The black line represents a linear model fit. Species exhibiting brown noise show a trend corresponding to the linear model

Table 3: Extinction probabilities: linear model statistics

	slope	p	adjusted R^2
random community structure	-0.53	< 0.0001 ***	0.58
clustered community structure	-0.47	< 0.0001 ***	0.46

3.2.6 How community structure shapes noise and abundance in the simulated gut microbiome

From experimental data of the gut microbial community it is hardly possible to infer causal interactions between microbes. Simulating the time evolution of the gut microbiome has the benefit that we can investigate the structure of the simulated community. This can give information about links between network properties and the respective community organization with the systems temporal dynamics.

The two different simulation approaches we used, differ in the topography that underlies the microbial interaction network. The Erdős-Rényi topography is characterized by a normal distribution in the node degrees of the network [63], the community structure is random. The Klemm-Eguiluz topography is characterized by a skewed distribution in the node degrees, which leads to a clustered community structure that shows small-world properties [64]. The difference in node degree distribution are shown in figure 49. These differences in simulation approaches are also visible in other topological measures of the interaction network that we examined. Namely, clustering coefficient and betweenness centrality show a greater variability in simulations with a clustered community structure (Klemm-Eguiluz, see figures 50 and 51). We also see higher variability in species richness and alpha diversity between individual simulations with clustered community structure than between individual simulations with random community structure (see 3.2.2). A clustered community structure leads to a broader range of possible interaction patterns which seem to allow for a higher variability in abundance distributions.

Figure 49: **Node degrees in simulations with different underlying network topography.** Differences in node degrees in simulations from networks with underlying Erdős-Rényi and Klemm-Eguiluz topography, respectively (each 5 simulations of 150 species resulting in measures from 750 species). The variance in node degrees is greater for Klemm-Eguiluz simulations due to the node degree distribution being skewed; Wilcoxon test: $p < 0.0001$ ***

Figure 50: **Clustering coefficients in simulations with different underlying network topography.** Differences in clustering coefficients in simulations from networks with underlying Erdős-Rényi and Klemm-Eguiluz topography, respectively (each 5 simulations of 150 species resulting in measures from 750 species). The variance in clustering coefficients is greater for Klemm-Eguiluz simulations; Wilcoxon test: $p < 0.0001$ ***
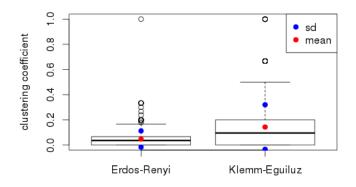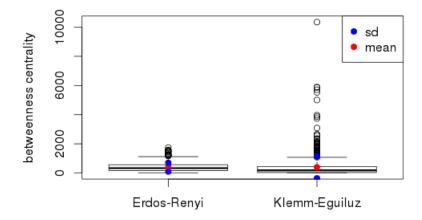
Figure 51: **Betweenness centrality in simulations with different underlying network topography.** Differences in betweenness centrality in simulations from networks with underlying Erdős-Rényi and Klemm-Eguiluz topography, respectively (each 5 simulations of 150 species resulting in measures from 750 species). The variance in betweenness centrality is greater for Klemm-Eguiluz simulations; Wilcoxon test: $p < 0.0001$ ***

We find power-laws in lifetime distributions and pink noise in mean power spectral densities in both experimental and simulated data. These are strong indications that the gut microbial community is a system exhibiting self-organized critical behavior. Self-organized criticality is thought to be found in systems strongly characterized by the interaction structure between system elements [1]. Therefore we assume that time evolution of the gut microbiome is also highly shaped by microbe-microbe interactions. Particularly, species showing pink noise in their individual power spectral densities should be affected by their interactions with other microbes. To visualize this influence of community structure on system behavior we examined topological measures of the interaction network. In simulations with clustered community structure, we find differences in betweenness centrality between subgroups of species exhibiting pink and brown noise, respectively (see figure 52). We also see an effect of community organization in simulations with random community structure. In these simulations, species exhibiting pink and brown noise, respectively, differ in their node degrees (see figure 53).

The topological measures we examined can be used to gain information about the navigability and the neighborhood within the network. However, they are static measures that have certain limitations. As both experimental and simulated data show clear indications for self-organized criticality, we assume that community structure has a stronger effect on system behavior than is visible in those topological measures. To draw systematic conclusions about how community structure influences time evolution in the gut microbiome and its connection to noise types different species exhibit, a more dynamic approach could be useful. Map equations could for example help to reveal patterns of information flow within the microbial network [73]. Another approach would be, to look at the community structure from a different angle and in more detail. We could investigate which microbes specifically interact with each other and what types of noise they exhibit. Patterns in this interaction structures within a subgroup of species could help unravel the role of community structure in time evolution of the gut microbiome.

Figure 52: **Betweenness centrality in simulations with clustered community structure.** Differences in betweenness centrality between species exhibiting pink and brown noise, respectively. Betweenness centrality is significantly higher in species exhibiting brown noise; Wilcoxon test: $p = 0.027$

Figure 53: **Node degrees in simulations with random community structure.** Differences in node degree between species exhibiting pink and brown noise, respectively. Betweenness centrality is significantly higher in species exhibiting brown noise; Wilcoxon test: $p = 0.0065$ *
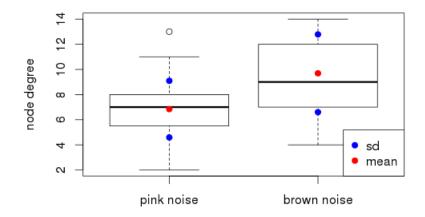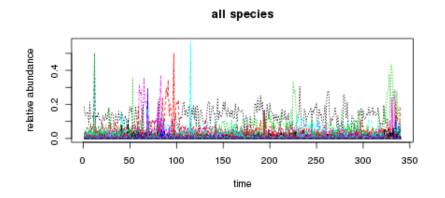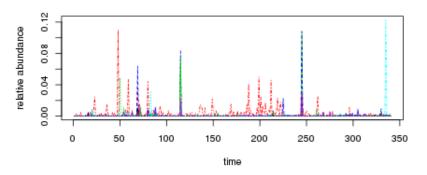
4 Discussion

In experimental and simulated data we performed spectral density analysis on the temporal abundance data from individual species. Many microbial species in experimental data stratified into two subgroups of species showing either pink or white noise in their power spectral densities. These two subgroups correlate with species abundance. As shown in figure 54 and figure 55 species exhibiting white noise have lower abundances than species exhibiting pink noise. The lifetimes of those subgroups of species also differ significantly (see section 3.2.4, figure 32) and higher abundant species exhibiting pink noise are generally more persistent in the microbial community.

The stratification of species into subgroups showing different noise types in their power spectral densities was also very strong in simulations with clustered community structure. Here we also find that species exhibiting white noise are very low in abundance, while species exhibiting pink noise show higher abundances (see figure 56). The subgroup of species exhibiting brown noise shows the highest abundances over time. Those subgroups again show different temporal persistence in the gut microbiome, which is reflected in their different lifetimes (see figure 38).

We argue that in both experimental and simulated data the subgroup of species showing higher lifetimes forms a stable part in the community. Composition of this stable part might vary and form enterotypes, as suggested in [12], or it is more of a functional core within the microbial community. Apart from this stable subgroup in the microbial community there is also a highly variable part that experiences frequent turnover. In the experimental and simulated data we analyzed, this part of the community shows significantly shorter lifetimes and exhibits white noise.

Figure 54: **Relative species abundances in all samples from donor A.** Species are grouped according to the noise type they exhibit: white noise (N=2420) and pink noise (N=177), respectively. Species with peaks showing very high abundances were left out to make dynamics more visible. Species exhibiting white noise are significantly less abundant than species exhibiting pink noise
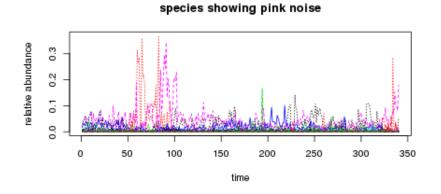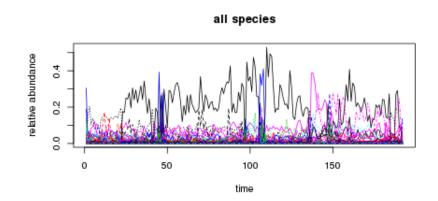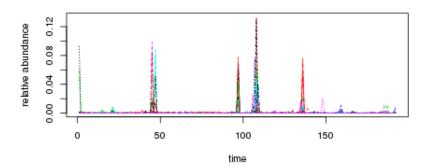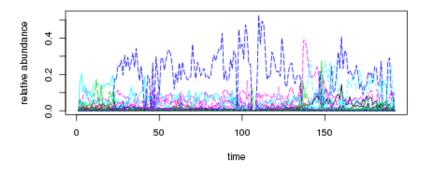
Figure 55: **Relative species abundances in all samples from
donor B.** Species are grouped according to the noise type they exhibit:
white noise (N=1743) and pink noise (N=442), respectively. Species with
peaks showing very high abundances were left out to make dynamics more
visible. Species exhibiting white noise are significantly less abundant than
species exhibiting pink noise

67

Figure 56: **Relative species abundances over time from a simulation with clustered community structure.** Species are grouped according to the noise type they exhibit: white noise (N=11), pink noise (N=34), brown noise (N=14), species with very high abundance were left out to make dynamics more visible. Very low abundant species exhibit white noise, species showing intermediate abundances exhibit pink noise and high abundant species exhibit brown noise

In the simulated data we also looked for trends within subgroups of species exhibiting a certain noise type. Specifically, we examined how external drivers influence those subgroups. In the theoretical model, external influences are captured in immigration and extinction probability. A species ability to colonize available space in the gut is incorporated in the immigration probability. It also includes how strong a species is influenced by the high flow-through in the gut. The influence the host has on the temporal evolution of a microbial species, e.g. through host-microbe interactions, mainly constitutes the extinction probability.
We find that external drivers seem to have strong effects on species exhibiting white and brown noise. Flow-through and constraints on colonization ability seem to have a large effect on species exhibiting white noise. Species exhibiting brown noise are apparently strongly affected by host influences. The subgroup of species exhibiting pink noise seems to be more robust against both those influences. We show this in figure 57 (upper plot) for simulations with clustered community structure. It is a log-log plot with immigration and extinction probability, respectively, on the x-axis and mean species abundance on the y-axis.

In terms of immigration probability, we see that the mean abundance of species exhibiting white noise is higher for higher immigration probabilities. The mean abundances of species exhibiting brown noise are higher for lower extinction probabilities. The subgroup of species exhibiting pink noise seems to be neither strongly affected by flow-through in the gut microbiome nor by host influences. This is also observable in figure 57 (upper plot). Species exhibiting pink noise show a slight trend in immigration probability, but no visible correlation of mean abundance with extinction probability.

We observe a similar pattern in simulations with random community structure (see figure 57, lower plot). Here, however, mean abundance of species exhibiting pink noise shows a stronger correlation with immigration probability.

Pink noise is strongly associated with self-organized criticality and suggests that a systems internal structure, i.e. interactions between the elements, is shaping its behavior [1]. Based on this concept and our observations we argue that time evolution of species exhibiting pink noise in the gut microbial community is most strongly influenced by their interactions with other microbial species. This internal structure makes them more robust against external drivers, be they high flow-through over time or the influences of the host. In simulations with random community structure, the variability in interaction patterns is limited. That in these simulations the influence of external drivers (immigration probability in figure 57, lower plot) is stronger than in simulations with clustered community structure emphasizes our argument.

Even in species that are most strongly influenced by the host we see some impact of community structure. In figure 58 we show that within species exhibiting brown noise in simulations with clustered community structure, there is a range of mean in-going interaction coefficients. The more other species positively influence the growth of a certain species, the more positive the measure is. It seems that abundance differences in those very high abundant species are to some extent due to microbe-microbe interactions and can not only be attributed to the influence of the host.

We also find many species exhibiting pink noise in the experimental data. The time evolution of those species is most likely strongly shaped by their interactions with other species as well. This is emphasized by the pink noise we find in mean power spectral densities of experimental data and also simulated data and the power-law distributions of the species lifetimes. We also do not find species that exhibit brown noise in experimental data, which is further evidence that the hosts influence on the gut microbial community is limited.

Based on our analysis we argue, that the gut microbial community is a system exhibiting self-organized critical behavior and that external drivers do

69

have limited influence on the gut microbiome. We suggest that the time evolution of this microbial community is mainly shaped by its internal structure.

Table 4: Extinction and immigration probabilities: linear model statistics

	slope	p	adjusted R^2
random community structure			
extinction probability	-1.15	< 0.0001 ***	0.62
immigration probability	1.2	< 0.0001 ***	0.42
clustered community structure			
extinction probability	-0.94	< 0.0001 ***	0.53
immigration probability	1.03	< 0.0001 ***	0.33

Figure 57: **Distribution of immigration and extinction probabilities on a logarithmic scale.** Immigration and extinction probabilities, respectively, and corresponding mean abundance of species showing different types of noise on a logarithmic scale.
(upper plot) simulations with clustered community structure
(lower plot) simulations with random community structure
Species exhibiting brown noise (blue circles) show increasing abundance with decreasing extinction probability (blue trend line); species exhibiting white noise (green triangles) show increasing abundance with increasing immigration probability (green trend line); species exhibiting pink noise (red circles and triangles) show a correlation with immigration probability but no visible correlation with extinction probability.
Linear model statistics in table 4

71

Figure 58: **Mean interaction coefficient from in-going interactions in simulations with clustered community structure.** The variability of mean in-going interaction coefficients in species exhibiting brown noise suggests that their differences in abundance are to some extent due to community structure

Our model simulates temporal dynamics within the intestinal microbial community and captures its systematic behavior. It provided insight into the development of community composition and temporal fluctuations. But we also faced limitations. The model does not take into account stochastic external perturbations or host-microbe interactions and species-specific parameters were not variable over time, but represented a stable environment. We simulated only a small part of OTUs in the experimental data resulting in information loss on the whole community. For more robust results, taking more than two experimental data sets for parametrization could also be useful. Analyses of the community structure also proofed to be very difficult. Being able to show the patterns and dynamics in an interacting community would, however, be very important.

Nevertheless, our theoretical model and the results from our analysis can be used for further research in various directions. In our simulations, model parameters associated with a particular species of the gut microbial community did not change over time. Strong perturbation of the system, for example due to antibiotic treatment or a permanent change in diet, can however lead to a shift in microbial community composition. After a per-

turbation species might be better adapted to the environment than before-hand while other species end up less well adapted. It might also lead to a shift in the noise regime of certain species. Species that exhibited white noise before a perturbation might now be better adapted to the new conditions and exhibit pink noise after the perturbation. This could further result in changed interaction structures, as evidenced by the noise pattern. As shown by Stein et al. in [34], a strong external perturbation can cause different species to rise in abundance. Due to community structure, their higher abundance also facilitates other species. This would mean they are less susceptible to stochastic fluctuations and might establish themselves in the intestinal microbial community. A strong external perturbation could result in a pathogenic microbial species establishing itself in the gut microbiome. Beneficial species on the other hand might be driven into a white noise regime due to competitive exclusion. External perturbation could be integrated into our SOC-model to investigate how it affects community dynamics. Perturbing species in a simulation and varying species parameters over time could also help understand how a dysbiosis of the microbial community develops.

We also think that simulating different microbial communities with this model can help understand patterns and trends in systematic behavior and detailed analysis could aid mechanistic understanding. The gut microbiome shows strong characteristics of self-organized criticality. In other microbial communities species exhibiting brown noise might be found in higher number, suggesting a stronger influence from an external driver, e.g. a host.

The gut microbiome could also be simulated in more detail by explicitly implementing host-microbe interactions to analyze their specific effects on microbial dynamics. We also suggest that our model could be used to model the reduced gut communities of e.g. mouse models and link back model parameters to experimental data. Vice versa, predictions and parametrizations could be tested in lab models such as mouse models.

Comparison of model fittings to experimental data from healthy and diseased subjects could provide understanding of systematic differences. As suggested in [74] systems shifting between two dynamical regimes show specific patterns that could be used as early-warning signals. This concept might be extendable to power spectral densities and self-organized critical systems. We argue that analyses of power spectral densities of healthy and diseased microbial gut communities could broaden our understanding of dysbiosis and its contribution to disease. It could also aid in identifying specific markers for diseases, be they single species, groups of species or patterns found in the temporal signal of the community. Finding patterns in system features could further provide a valuable tool for possible prediction of diseases, such as inflammatory bowel disease or gastric cancer.

5 Conclusion

We apply an individual-based model to the simulation of temporal dynamics in the gut microbiome. Our results show on a systematic level that the influence of external effects and internal structure on time evolution vary between species. We further find that the gut microbiome exhibits typical characteristics of self-organized criticality, particularly pink noise. The pink noise it exhibits indicates that the system behavior is strongly influenced by the internal structure, the interaction network of the microbial community. These findings can help to better understand structure and dynamics of the gut microbial community and aid in the development of general treatment strategies for diseased states of the microbiota. We make a first step of introducing noise in power spectral densities as a marker for microbial community organization. We argue that a systematic survey in a broad range of data sets will uncover the mechanistic basis and the impact of microbial community organization for emerging community properties, such as dysbiosis or persistent unfavorable microbiota composition responsible for human disease state.

References

[1] H.J.Jensen 1998 Self-Organized Criticality. Emergent Complex Behavior in Physical and Biological Systems *Cambridge University Press*

[2] The Human Microbiome Project Consortium 2012 Structure, function and diversity of the healthy human microbiome *Nature* 486, 207-214

[3] J.Qin et al. 2010 A human gut microbial gene catalogue established by metagenomic sequencing *Nature* 464, 59-65

[4] C.T.Peterson, V.Sharma, L.Elmén, S.N. Peterson 2014 Immune homeostasis, dysbiosis and therapeutic modulation of the gut microbiota *Clinical and Experimental Immunology* 179, 363-377

[5] Q.Aziz, J.Doré, A.Emmanuel, F.Guarner, E.M.M.Quigley 2013 Gut microbiota and gastrointestinal health: current concepts and future directions *Neurogastroenterology and Motility* 25, 4-15

[6] S.Schippa, M.P.Conte 2014 Dysbiotic Events in Gut Microbiota: Impact on Human Health *Nutrients* 6, 5786-5805

[7] A.K.Benson, S.A. Kelly, R. Legge et al. 2010 Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors *PNAS* 107, 18933-18938

[8] C.A.Lozupone, J.I.Stombaugh, J.I.Gordon, J.K.Jansson, R.Knight 2012 Diversity, stability and resilience of the human gut microbiota *Nature* 489, 220-230

[9] F.Bäckhed et al. 2015 Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life *Cell Host and Microbe* 17, 690-703

[10] T.Yatsunenko et al. 2012 Human gut microbiome viewed across age and geography *Nature* 486, 222-228

[11] L.A.David, A.C.Materna, J.Friedman, M.I.Campos-Baptista, M.C.Blackburn, A.Perrotta, S.E.Erdman, E.J.Alm 2014 Host lifestyle affects human microbiota on daily timescales *Genome Biology* 15:R89

[12] M.Arumugam et al. 2011 Enterotypes of the human gut microbiome *Nature* 473, 174-180

[13] D.Knights, T.L.Ward, C.E.McKinlay, H.Miller, A.Gonzalez, D.McDonald, R.Knight 2014 Rethinking "Enterotypes *Cell Host and Microbe* 16, 433-437

[14] F.Bäckhed, C.M.Fraser, Y.Ringel, M.E.Sanders, R.B.Sartor, P.M.Sherman, J.Versalovic, V.Young, B.B.Finlay 2012 Defining a Healthy Human Gut Microbiome: Current Concepts, Future Directions, and Clinical Applications *Cell Host and Microbe* 12, 611-622

[15] J.C.Clemente, L.K.Ursell, L.W.Parfrey, R.Knight 2012 The Impact of the Gut Microbiota on Human Health: An Integrative View *Cell* 148, 1258-1270

[16] C.G.Buffie, I.Jarchum, M.Equinda, L.Lipuma, A.Gobourne, A.Viale, C.Ubeda, J.Xavier, E.G.Pamer Profound Alterations of Intestinal Microbiota following a Single Dose of Clindamycin Results in Sustained Susceptibility to *Clostridium difficile*-Induced Colitis *Infection and Immunity* 80(1), 62-73

[17] D.Knights, L.W.Parfrey, J.Zaneveld, C.Lozupone, R.Knight 2011 Human-associated microbial signatures: examining their predictive value *Cell Host Microbe* 10(4), 292-296

[18] P.Bak, C.Tang, K.Wiesenfeld 1988 Self-organized criticality *Physical Review A* 38, 364-374

[19] D.Sornette 2006 Critical Phenomena in Natural Sciences *Springer Complexity*

[20] P.Bak 1996 How Nature Works *Springer Verlag*

[21] M.Aguilera, I.Morer, X.E.Barandiaran, M.G.Bedia 2013 Quantifying Political Self-Organization in Social Media. Fractal patterns in the Spanish 15M movement on Twitter *European Conference on Artificial Life - General Track* 395-402

[22] T.Musha 1997 1/f Fluctuations in Biological Systems *Proceedings of IEEE* 6, 2692-2697

[23] J.M.Halley, P.Inchausti 2004 The increasing importance of 1/f-noises as models of ecological variability *Fluctuation and Noise Letters* 4(2), R1-R26

[24] A.Sornette, D.Sornette 1989 Self-Organized Criticality and Earthquakes *Europhysics Letters* 9(3), 197-202

[25] R.V.Solé, J.Bascompte, S.C.Manrubia 1996 Extinction: Bad Genes or Weak Chaos? *Philosophical Transactions of the Royal Society* 263, 1407-1413

[26] K.Linkenkaer-Hansen, V.V.Nikouline, J.M.Palva, R.J.Ilmoniemi 2001 Long-Range Temporal Correlations and Scaling Behavior in Human Brain Oscillations *The Journal of Neuroscience* 21(4), 1370-1377

[27] M.Bartolozzi, D.B.Leinweber, A.W.Thomas 2004 Self-Organized Criticality and Stock Market Dynamics: An Empirical Study *Physica A* 350, 451-465

[28] S.P.Hubbell 2001 The Unified Neutral Theory of Biodiversity and Biogeography *Princeton University Press*

[29] Q.Zeng, J.Sukumaran, S.Wu, A.Rodrigo 2015 Neutral Models of Microbiome Evolution *PLoS Computational Biology* 11(7), e1004365

[30] K.Harris, T.L.Parsons, U.Z.Ijaz, L.Lahti, I.Holmes, C.Quince 2014 Linking statistical and ecological theory: Hubbell's unified neutral theory of biodiversity as a hierarchical Dirichlet process *Proceedings of the IEEE* arXiv:1410.4038 [q-bio.PE]

[31] J.B.Reece, S.A.Wasserman, L.A.Urry, P.V.Minorsky, M.L.Cain, R.B.Jackson 2014 Campbell Biology, Tenth Edition *Pearson Education*

[32] V.Volterra 1928 Variations and Fluctuations of the Number of Individuals in Animal Species living together *ICES Journal of Marine Science* 3(1), 3-51

[33] K.Faust, J.Raes 2012 Microbial Interactions: from networks to models *Nature* 10, 538-550

[34] R.R.Stein, V.Bucci, N.C.Toussaint, C.G.Buffie, G.Rätsch, E.G.Pamer, C.Sander, J.B.Xavier 2014 Ecological Modeling from Time-Series Inference: Insight into Dynamics and Stability of Intestinal Microbiota *PLoS Computational Biology* 9(12), e1003388

[35] V.Bucci, S.Bradde, G.Biroli, J.B.Xavier 2012 Social Interaction, Noise and Antibiotic-Mediated Switches in the Intestinal Microbiota *PLoS Computational Biology* 8(4), e1002497

[36] F.L.Hellweger, V.Bucci 2009 A bunch of tiny individuals - Individual-based modeling for microbes *Ecological Modelling* 220, 8-22

[37] V.Bucci, J.B.Xavier 2014 Towards Predicitve Models of the Human Gut Microbiome *Journal of Molecular Biology* 426, 3907-3916

[38] L.A.Lardon, B.V.Merkey, S.Martins, A.Dötsch, C.Picioreanu, J.U.Kreft, B.F.Smets iDynoMiCS: next-generation individual-based modelling of biofilms *Environmental Microbiology* 13(9), 2416-2434

[39] T.Skybová, M.Přibyl, J.Pocedič, P.Hasal 2012 Mathematical modeling of wastewater decolorization in a trickle-bed bioreactor *Journal of Biotechnology* 157, 512-523

77

[40] R.V.Solé, D.Alonso, A.McKane 2002 Self-organized instability in complex ecosystems *Philosophical Transactions of the Royal Society* 357, 667-681

[41] A.L.Barabási 2002 Linked: how everything is connected to everything else and what it means for business, science, and everyday life *Penguin Group*

[42] G.E.Mobus, M.C.Kalton 2015 Principles of Systems Science *Springer Complexity*

[43] J.A.Dunne, R.J.Williams, N.D.Martinez 2002 Network structure and biodiversity loss in food webs: robustness increases with connectance *Ecology Letters* 5(4), 558-567

[44] J.M.Olesen, J.Bascompte, Y.L.Dupont, P.Jordano 2007 The modularity of population networks *PNAS* 104(50), 19891–19896

[45] J.Krause, D.Lusseau, R.James 2009 Animal social networks: an introduction *Behavioral Ecology and Sociobiology* 63(7), 967-973

[46] D.Berry, S.Widder 2014 Deciphering microbial interactions and detecting keystone species with co-occurrence networks *Frontiers in Microbiology* 5, 219

[47] S.Lhermitte, J.Verbesselt, W.W.Verstraeten, P.Coppin 2011 A comparison of time series similarity measures for classification and change detection of ecosystem dynamics *Remote Sensing of Environment* 115, 3129-3152

[48] K.Yang, C.Shahabi 2004 A PCA-based similarity measure for multivariate time series *MMDB* 1, 65-74

[49] J.Wang, Y.Zhu, S.Li, D.Wan, P.Zhang 2014 Multivariate Time Series Similarity Searching *The Scientific World Journal* 1, 851017

[50] W.T.Sloan, S.Woodcock, M.Lunn, I.M.Head, T.P.Curtis Modeling Taxa-Abundance Distributions in Microbial Communities using Environmental Sequence Data *Microbial Ecology* 53, 443-455

[51] M.Basseville 2013 Divergence measures for statistical data processing *Signal Processing* 93, 621-633

[52] P.Damos 2015 Mixing times towards demographic equilibrium in insect populations with temperature variable age structures *Theoretical Population Biology* 103, 93-102

[53] M.Arenas, A.Sánchez-Cobos, U.Bastolla 2015 Maximum-Likelihood Phylogenetic Inference with Selection on Protein Folding Stability *Molecular Biology and Evolution* 32(8), 2195-2207

[54] D.Kasthurirathna, M.Piraveenan Emergence of scale-free characteristics in socio-ecological systems with bounded rationality *Scientific Reports* 11(5), 10448

[55] A.C.Cameron, F.A.G. Windmeijer 1997 An R-squared measure of goodness of fit for some common nonlinear regression models *Journal of Econometrics* 77, 329-342

[56] P.Gray, W.Hart, L.Painton, C.Phillips, M.Trahan, J.Wagner A Survey of Global Optimization Techniques *http://www.cs.sandia.gov/opt/survey/main.html*

[57] F.Molnár Jr., C.Caragine, T.Caraco, G.Korniss 2013 Restoration Ecology: Two-Sex Dynamics and Cost Minimization *PLoS One* 8(10), e77332

[58] M.C.R.Melo, R.C.Bernardi, T.V.A.Fernandes, P.G.Pascutti 2012 GSAFold: A new application of GSA to protein structure prediction *Proteins* 80(9), 2305-2310

[59] J.Song, J.He, M.Zhu, D.Tang, Y.Zhang, S.Ye, D.Shen, P.Zou 2014 Simulated Annealing Based Hybrid Forecast for Improving Daily Municipal Solid Waste Generation Prediction *The Scientific World Journal* 1, 834357

[60] Y.Xiang, S.Gubian, B.Suomela, J.Hoeng 2013 Generalized Simulated Annealing for Global Optimization: The GenSA Package *The R Journal* 5, 13-29, URL http://journal.r-project.org/archive/2013-1/xiang-gubian-suomela-etal.pdf

[61] Wolfram Research, Inc. 2012 Mathematica, Version 9.0 *Champaign, IL*

[62] S.Kullback, R.A.Leibler 1951 On Information and Sufficiency *The Annals of Mathematical Statistics* 22, 79-86

[63] P.Erdős, A.Rényi 1959 On random graphs I. *Publicationes Mathematicae* 6, 290–297

[64] K.Klemm, V.M.Eguíluz 2002 Growing scale-free networks with small-world behavior *Physical Review E* 56, 057102

[65] M.Bastian, S.Heymann, M.Jacomy 2009 Gephi: an open source software for exploring and manipulating networks *International AAAI Conference on Weblogs and Social Media*

[66] M.E.J.Newman 2003 The Structure and Function of Complex Networks *SIAM Review* 45(2), 167-256

[67] F.Ju, T.Zhang 2015 Bacterial assembly and temporal dynamics in activated sludge of a full-scale municipal wastewater treatment plant *The ISME Journal* 9, 683-695

[68] B.J.Baker, J.F.Banfield 2003 Microbial communities in acid mine drainage *FEMS Microbiology Ecology* 44, 139-152

[69] K.Faust, J.F.Sathirapongsasuti, J.Izard, N.Segata, D.Gevers, J.Raes, C.Huttenhower Microbial Co-Occurrence Relationships in the Human Microbiome *PLoS Computational Biology* 8(7), e1002606

[70] G.Csardi, T.Nepusz 2006 The igraph software package for complex network research *InterJournal* Complex Systems, 1695, URL http://igraph.org

[71] C.Tsallis, D.A.Stariolo 1996 Generalized Simulated Annealing *Physica A* 233, 395-406

[72] R Core Team 2012 R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0* URL http://www.R-project.org/

[73] M.Rosvall, C.T.Bergstrom 2008 Maps of random walks on complex networks reveal community structure *PNAS* 105(4)

[74] M.Scheffer, J.Bascompte, W.A.Brock, V.Brovkin, S.T.Carpenter, V.Dakos, H.Held, E.H.van Nes, M.Rietkerk, G.Sugihara Early-warning signals for critical transitions *Nature* 461, 53-59