bioRxiv preprint doi: https://doi.org/10.1101/367243; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The Neutral Metaorganism

Michael Sieber[a,*], Lucía Pita[b], Nancy Weiland-Bräuer[c], Philipp Dirksen[a,d], Jun Wang[e], Benedikt Mortzfeld[d], Sören Franzenburg[f], Ruth A. Schmitz[c], John F. Baines[a,g], Sebastian Fraune[d], Ute Hentschel[b,d], Hinrich Schulenburg[a,d], Thomas C. G. Bosch[d], and Arne Traulsen[a]

[a]*Max Planck Institute for Evolutionary Biology, Plön, Germany*
[b]*GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany*
[c]*Institute for General Microbiology, Christian-Albrechts-University Kiel, Kiel, Germany*
[d]*Zoological Institute, Christian-Albrechts-University Kiel, Kiel, Germany*
[e]*Institute of Microbiology, Chinese Academy of Science, Beijing, China*
[f]*Institute for Clinical Molecular Biology, Christian-Albrechts-University Kiel, Kiel, Germany*
[g]*Institute for Experimental Medicine, Christian-Albrechts-University Kiel, Kiel, Germany*

Almost all animals and plants are inhabited by diverse communities of microorganisms, the microbiota. The hosts often pose highly selective environments, which only a subset of the environmentally available microbes are able to colonize. From the host's perspective it seems useful to shape the community composition of these allowed microbes to promote a beneficial host-microbe symbiosis. In contrast to this, neutral models assume that the structure of the microbiota is entirely shaped by random population dynamics and dispersal. We here show that microbiota community structure from a wide range of host organisms, in particular including previously understudied invertebrates, is in almost all cases consistent with neutral expectations. Moreover, we demonstrate that apparent discrepancies with the neutral model can be due to transient community states rather than host selection. In conclusion, even though hosts are often assumed to control microbiota composition to ensure beneficial interactions, our broad-scale analysis highlights that following colonization, it could rather be neutral processes that determine microbial community structure.

The microbial communities living in and on animals can affect many important host functions, including metabolism [1, 2, 3], the immune system [4, 5, 6], and even behaviour [7, 8]. The extent and direction of this microbe-mediated influence is often linked to the presence or absence of species and their relative abundances. It is thus paramount to understand how host-associated microbial communities are assembled.

A classical explanation for the emergence of a particular ecological community structure posits that every species is defined by distinct traits and occupies a specific ecological niche. An implicit hypothesis underlying much of microbiome research is that hosts have the potential to actively shape their associated microbial communities by providing niches for useful microbes [9]. This implies that individual hosts could select for a potentially very specific community structure.

While the assumption that the metaorganism – the host together with its associated microbes [10, 11] – is an actively shaped symbiotic unit is appealing, it may bias interpretation of microbiota-host analyses. An example is the widely reported connection between the structure of the human gut microbiota and obesity, which turns out to be very hard to distinguish from random noise [12]. More recently it has also been demonstrated that the genetic background of the host does not significantly shape human microbiome composition [13]. Indeed, it has been argued that selective processes should not form the null hypothesis for explaining the allegedly cooperative host-microbe symbiosis [14].

Instead, neutral models have been proposed as a valuable tool for finding patterns in the tremendous complexity of ecological communities that may have a deeper mechanistic cause. Neutral models assume ecological equivalence between species. Thus, the community structure within a single host is the outcome of purely stochastic population dynamics, immigration and local extinctions [15]. The Unified Neutral Theory of Biodiversity [16, 17] extends the neutral framework from a single site to multiple sites, each

*e-mail: sieber@evolbio.mpg.de

bioRxiv preprint doi: https://doi.org/10.1101/367243; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

harbouring its own local community (Fig. 1). Diversity within the local communities is maintained by immigration from a common source community, so that the whole setup resembles a mainland-island structure [18]. Neutral theory has been applied to numerous ecological systems and sparked a lot of controversy along the way [19].

Despite these controversies and with the growing availability of microbial community data, the neutral theory has been adapted and applied to the microbial world [20, 21, 22, 23, 24]. In particular, the local community vs. metacommunity structure of neutral theory naturally extends to host-associated microbial communities. Here, the hosts are viewed as ecosystems and their microbiota are treated as local communities [25, 11]. This has led to several recent studies addressing the question whether the microbiota conform to the patterns predicted by the neutral theory. The rank-abundance patterns of the microbiota from three domesticated vertebrates for example were found to be largely consistent with the neutral expectation [26, 27], despite a significant non-neutral signal in the genomic data [26]. A good fit of the neutral model was also found for the microbiota of young individuals of the zebrafish *Danio rerio* [28]. Interestingly, in this study neutrality decreased as hosts aged, indicating an increasing influence of selective processes over developmental time, potentially linked to the activation of a fully functioning adaptive immune response. The human microbiota were found to be predominantly non-neutral across various body sites [29], and in this study the few neutral communities were mostly associated with the skin and the urological tract. Another study, on the other hand, found the composition of the skin microbiota of healthy human subjects in large Chinese cities to be better explained by non-neutral processes [30]. Contrasting results can also be obtained depending on the state of the host, with the healthy human lung microbiota being largely consistent with a neutral model, while microbes recovered from diseased lungs diverged from neutrality [31]. Neutral assembly processes in invertebrate hosts are understudied, but a recent study shows the microbial communities associated with the fruit fly *Drosophila melanogaster* to be consistent with the predictions of a neutral model [32].

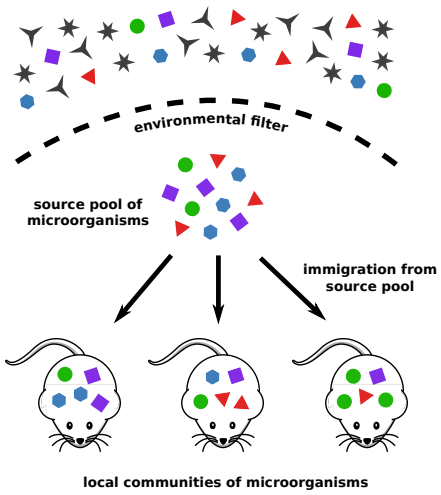

Figure 1 Sketch of the setup of the neutral model. Only a subset of the available microorganisms can pass through the environmental filter, forming a common source pool for all hosts. This selective step is not described by the neutral model. The microbiota of individual hosts form local communities and the microbial intra–host population dynamics are neutral. Local diversity is maintained by immigration from the source pool.

While these studies have drawn awareness to the potentially important contribution of neutral processes in shaping the microbiota, the use of several different neutral models and focus on one or are a few, mostly vertebrate host organisms, makes it hard to draw more general conclusions. Here, we aim to overcome these limitations by consistently applying a neutral model to a variety of distinct host systems, including, but not limited to, a wide range of invertebrate hosts. We included host species from a range of eukaryotic multicellular organisms with different life styles, ranging from early branching groups such as sponges, to the house mouse with its fully developed adaptive immune system. The use of a consistent modelling framework and the diversity of hosts in our study allows us to identify general characteristics of hosts that are more likely to look neutral. In particular, we hypothesize that for short–lived host species we may underestimate the level of neutrality, since microbiota analyses commonly take snapshots of community structure early in the neutralization process. Based on the neutral null hypothesis, we further identified members of the microbiota which consistently deviate from the neutral expectation, possibly indicating the action of selection on these particular microbes.

Results

We employ the neutral model presented by Sloan et al. [20], which is particularly suited for large-sized microbial populations. It describes the stochastic population dynamics within a local community, corresponding to the microbiota of a specific host. To maintain local diversity, which would otherwise reduce

bioRxiv preprint doi: https://doi.org/10.1101/367243; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

to a single species through ecological drift, immigration from a fixed source community occurs with rate m. This source community is not equivalent to the environmental pool of microorganisms, but rather it can be interpreted as the collection of all microbial species that can pass through the environmental filter of the host (Fig. 1). Thus, the neutral model does not make any assumptions about the selective processes that occur before and during colonization of the host and in particular, it is not concerned with any host traits that may restrict the range of colonizing microbes [33].

This model allows to derive a relatively simple expression for the expected long-term stationary community composition. Specifically, it predicts the relationship between the mean relative abundance of a taxon across all local communities, i.e. the metacommunity, and the probability of actually observing this taxon in any single community. Using well-established methods from population genetics, it has been shown that this relationship is determined by a beta distribution [20]. The only free parameter of this model is the immigration rate m, which can be calibrated by a nonlinear fit. The goodness of the fit then indicates how well the prediction of the neutral model compares to the empirical data. Concretely, we use the coefficient of determination R^2 as a quantitative measure of how consistent the data is with the neutral model. See the *Material and Methods* for details of the neutral model and the fitting procedure.

Neutrality of the microbiota

We fitted the theoretical neutral expectation to the published microbiota composition of eight different host species across four phyla and, for comparison, three environmental microbial communities (see Table S1 for an overview and references).

With the species *Sarcotragus fasciculatus*, *Ircinia oros* and *Carteriospongia foliascens*, our study includes examples from the oldest extant sister group to all other animals, the sponges (Porifera) [34]. Despite their filter-feeding life-style, sponges harbor a very diverse and highly specific microbiota [35], which mediates the functional role of the sponges in the ecosystem [36]. This is complemented by the jellyfish *Aurelia aurita*, the starlet sea anemone *Nematostella vectensis*, and the fresh-water polyp *Hydra vulgaris*. These hosts are examples from another early branching phylum, the Cnidaria, which comprise a basic innate immune system thought to play an important role in controlling one of the oldest known symbiotic host-microbe relationships [37]. We also included the nematode *Caenorhabditis elegans* as one of the best–studied multicellular organisms, whose microbiota has only more recently come into the focus [38, 39, 40, 41]. Finally, the house mouse *Mus musculus* with its potent adaptive immune system [42] offers a well–studied intestinal microbiota [43, 44].

In all cases, community composition was determined by the relative abundances of operational taxonomic units (OTUs), obtained by standard high-throughput 16S rRNA sequencing techniques [45]. To ensure equal sample sizes, all OTU abundance tables were rarified to the same read depth (1000 reads per sample). We analyzed the effect of the rarefaction on the neutral fit for datasets where more reads were available and found it to be of little relevance for most communities (Fig. S5). The estimated dispersal

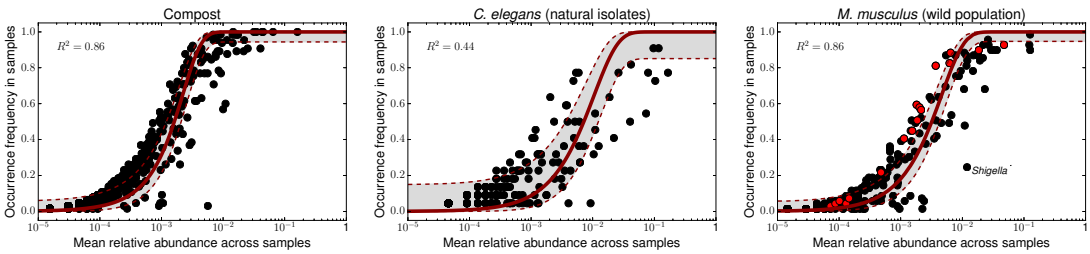

Figure 2 The relationship between mean relative abundance of taxa across all samples and the frequency with which they are detected in individual samples for three representative datasets. Each dot represents a taxon and the solid line is the best-fitting neutral community expectation. The dashed lines and grey area depict the 95% confidence bands. The left panel shows a microbial community sampled from a compost as an example of an abiotic environment. The centre panel shows data for the natural microbiota of *C. elegans* nematode populations sampled from the same compost. The right panel shows the microbiota of a wild *M. musculus* mouse population. Here, the red dots indicate members of the microbial family Ruminococcaceae, which are predominantly overrepresented compared to the neutral expectation. A genus that is found in much fewer mice than expected is *Shigella*, being present in less than a quarter of the hosts despite a neutral expectation of nearly 100% prevalence.

bioRxiv preprint doi: https://doi.org/10.1101/367243; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

parameter m from the neutral best-fits for all host populations showed considerable variation, but there is no correlation between m and the level of neutrality (Fig. S4A and Tab. S1). Additionally, we found no correlation between sample size and neutrality, or between microbiota diversity and neutrality (Fig. S4).

A comparison of the theoretical and observed relationship between mean relative abundances and occurrence frequencies of OTUs for three illustrative examples is shown in Fig. 2 (see Fig. S1 for an overview of all datasets). Taxa that lie above the neutral expectation are found more often than would be expected from their relative abundance in the microbiota metacommunity, implying a possible role of positive selection. In contrast, taxa below the neutral prediction are found in fewer hosts than expected by their relative abundance across all hosts, which may be due to negative host-mediated selection.

The composition of microbial communities derived from compost and seawater are very consistent with the neutral expectation (Fig. 2A and S1), which aligns with previous studies showing that random population dynamics and immigration play a major role in shaping microbial communities in abiotic environments [21, 23, 24]. Interestingly, the sediment community showed a relatively low consistency with the neutral model. This dataset is also the only one that showed a substantial effect of rarefaction, with neutrality increasing to the same level as the seawater community, when more reads are included (Fig. S5). This may indicate a greater influence of rare taxa, potentially leading us to underestimate the level of neutrality in the rarefied sediment dataset.

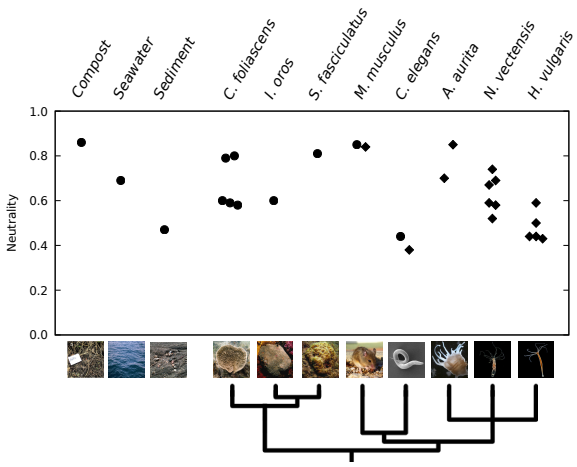

Figure 3 Overview of how consistent a range of host-associated and environmental communities are with the neutral model. Circles denote natural populations and diamonds denote laboratory populations. The data for *C. foliascens* is from several different natural populations, while the data for *H. vulgaris* and *N. vectensis* is from different time points. Spread of points along the x-axis is added to increase visibility. Phylogeny generated with *phyloT* based on NCBI taxonomy. Sponge photographs courtesy of Susanna López-Legentil (UNC Wilmington).

Biological hosts with their potential for active selection of specific taxa, could on the other hand be expected to have their resident microbial communities under tighter control, reducing the relative importance of stochastic processes after colonization, which would be reflected in a worse fit of the neutral expectation. The nematode *C. elegans* for example naturally consumes bacteria as its food source and possesses an innate immune system, potentially to defend against the threat of ingested pathogenic microbes [46]. This suggests that these worms have some control over their microbiota and indeed, for worms that were sampled at the same locations where the compost samples were taken, the best fit of the neutral expectation is considerably worse (Fig. 2B). We found a similar level of neutrality for the microbiota of laboratory worm populations grown exclusively on *E. coli* (Fig. 3), indicating a similar influence of neutral processes under the more controlled laboratory environments.

Vertebrate hosts with their more sophisticated and complex immune system are expected to diverge even further from the neutral expectation. However, intriguingly, the neutral model showed a very good fit to the microbiota of a wild *M. musculus* mouse population (Fig. 2C). In fact, the best fit was just as good as for the abiotic compost environment. A notable non-neutral microbe is *Shigella*, which is found in significantly fewer mice than expected from the neutral model (Fig. 2C). Invasive *Shigella* causes intestinal inflammations in humans, but mice mount an effective defense against it, thereby preventing acute infections [47]. This illustrates how potentially pathogenic microbes, which are actively selected against by the host, are typically found below the neutral expectation. The high level of neutrality in the natural mouse population is corroborated by a laboratory population of *M. musculus* (Fig. 3).

We found similar high levels of neutrality for sponges which as aquatic filter feeders are exposed to the full range of marine microbes. Like nematodes, they comprise a basic immune system [48], and especially for *S. fasciculatus* and some populations of *C. foliascens*, the data is in very good agreement with the neutral expectation (Fig. S1). In contrast, the results for the three aquatic polyps are less consistent. The two laboratory populations of *A. aurita* were in very good agreement with the neutral prediction, while the

bioRxiv preprint doi: https://doi.org/10.1101/367243; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

neutrality of samples taken at different developmental stages of *H. vulgaris* and *N. vectensis* varied from poor to good (Fig. 3).

Comparing the neutral expectation with data from a wide range of host species showed that a good fit of the neutral model is in fact the norm rather than the exception (Fig. 3). In particular, there was no substantial difference in neutrality between microbial communities living in abiotic environments, such as compost and seawater, and the microbiota from a diverse array of host species. We thus conclude that neutral processes may play a substantial role in the assembly of microbial communities associated with hosts as different as mice and marine sponges.

Why then, do in particular the microbiota of *C. elegans* diverge from the neutral expectation, suggesting a greater influence of selective processes? The most commonly assumed cause of such deviations is selection by the host, for example through its immune system. However, it seems intriguing that nematodes with a comparatively simple immune system would be much better at selecting microbes than mice with their sophisticated immune systems. Therefore, it is important to assess other possible violations of the underlying model assumptions.

Change of neutrality over time

A crucial aspect of the neutral prediction is that it corresponds to the expected *long-term* stationary distribution of the microbial community. If the community is in a transient state, however, it can appear non-neutral even if it is the product of a purely neutral process. This effect is illustrated in Figure 4, showing a simulation of the population dynamics of 50 neutral communities (*Material and Methods*). For every community, only a few random colonisers were picked from the source community, which yields random, but non-neutral, initial communities. We then compared the neutral expectation to the in-silico communities many times during the simulation, showing that they become "neutralized" through random population dynamics and immigration until the community composition reaches its long-term equilibrium, or the host dies.

Thus, in principle, a bad fit of the neutral expectation may be due to a transient non-equilibrium state of the community, rather than actual non-neutral processes. This effect would be especially pronounced for short-lived host species such as *C. elegans*, where initial colonization can lead to communities that appear non-neutral even for ecologically equivalent (i.e. neutral) colonizers [49]. The microbiota of such hosts may never effectively converge to the neutral long-term expectation, even if they are dominated by stochastic dynamics and immigration. For longer–lived species with a substantial amount of vertical transmission of the microbiota, such as sponges and mice, we expect these transient effects to play a smaller role.

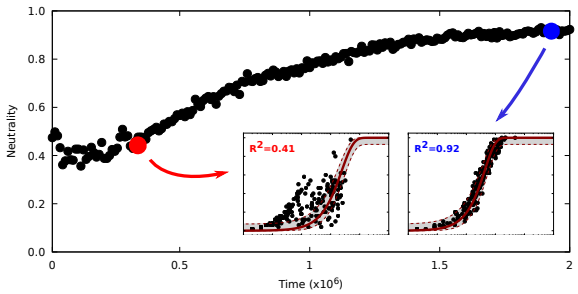

Figure 4 Goodness of fit of the neutral long-term expectation to neutrally simulated communities over time, where each dot represents the best fit at that timepoint. The two insets show the best fits of the neutral model to snapshots of the communities at an early timepoint (indicated by the red dot) and a late timepoint (indicated by the blue dot). Scale and meaning of the axes for the insets are the same as in Fig. 2. Each community was initialised with a few random colonizers, which gives a non-neutral initial community ($R^2 \approx 0.4$).

If community assembly is governed by purely neutral processes, we would expect the fit of the neutral model to become better over time. Decreasing neutrality, which has been reported for the zebrafish *Danio rerio* [28], would on the other hand be a good indicator of selective processes. A similar increase in the relative importance of selective processes over stochastic effects has been described for the ecological succession of microbial communities in salt marshes [50]. However, while the microbiota from the *Hydra* and *Nematostella* populations show clear developmental patterns [51], we found consistency with the neutral model to remain relatively constant over time (Fig. S3).

Identifying non-neutral taxa

To identify those members from the bulk of the microbiota that diverge from the neutral expectation is of great practical importance, as this may reveal potential selective processes. In the following, we classified

bioRxiv preprint doi: https://doi.org/10.1101/367243; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

all taxa outside of the 95% confidence bands around the neutral prediction as not consistent with the neutral expectation.

We found that the distribution of non-neutral taxa around the neutral prediction is roughly symmetrical in all cases, and no microbial order deviates systematically above or below the neutral expectation (Fig. S2). In particular, within microbial orders that contain a substantial amount of non-neutral taxa, such as the Clostridiales and Bacteroidales in the mouse microbiota, genera were roughly evenly over- and under-represented compared to the neutral null hypothesis. This is consistent with the hypothesis that the observed deviations from neutrality are at least in part due to transient neutral processes rather than selection.

To further distinguish random deviations from neutrality and actual non-neutral candidates, we also applied a more conservative definition of non-neutrality. For this, a particular taxon was classified as non-neutral only when it consistently diverged from the neutral expectation in the same direction in independent host populations. Applying this definition to the natural and laboratory populations of *C. elegans* and *M. musculus* yields a reduced subset of microbial taxa, which lie above or below the neutral prediction in both populations.

This reveals that of the *C. elegans* microbiota, only the genus *Ochrobactrum* is consistently under-represented (Tab. S2). It is found in only ca. 40% of the natural isolates and 70% of the laboratory worms, despite its relatively high mean abundance across all worms and a neutral expectation of 100% prevalence. This underrepresentation of *Ochrobactrum* is intriguing, as it had previously been identified as enriched in worms and able to persist in the nematode's gut even under starvation conditions [39]. However, this observation is consistent with a transient community state, where *Ochrobactrum* has not yet colonized all worms, despite being very successful once it has entered the worm.

In the *M. musculus* microbiota of natural and laboratory populations, over a third of the over-represented taxa were from the bacterial family *Ruminococcaceae*. Members of this family were never consistently found below the neutral expectation (Fig. 2C and Tab. S3). This family includes several physiologically relevant genera involved in the utilization of plant polysaccharides such as starch and cellulose [52], and high-fat diets low in plant-derived materials have been found to decrease the proportion of the *Ruminococcaceae* in mouse guts [53]. That we were able to consistently identify a bacterial family that had previously been linked to important metabolic functions as over-represented in the overall neutral mice microbiota, shows the utility of the neutral model as a null hypothesis and a tool of finding meaningful patterns in the vast complexity of the microbiota. An objective definition of a core microbiome is not straightforward. Taxa such as the *Ruminococcaceae* that are present significantly more often than predicted by chance are good candidates fulfilling clear objective criteria to be defined as a core taxon.

Comparison with a random community assembly model

The observed high levels of neutrality do not necessarily imply that the microbiota are simply a random collection of microbes. A good fit of the expected distribution indicates that the simple neutral model is sufficient to describe the relationship between abundance of a species and its occurrence frequency. But this can not rule out other processes, niche-related or stochastic, that lead to the same community structure [54].

A different model is obtained by assembling a local community by simply drawing randomly from the fixed source community. In this case, the probability of observing a specific species in a local community is determined by a binomial distribution. Comparing the best fits of the neutral expectation and of a binomial distribution using the Aikake information criterion (AIC) indicates that the neutral model fits the data better in all cases (Tab S1). This indicates that the microbiota are not merely random samples of the microbes present in the metacommunity, and that stochastic population dynamics and dispersal play important roles in shaping communities.

Discussion

We set out to quantify how consistent the microbiota of a range of different host organisms are with the expectation from a neutral model. We found that the structures of the host-associated microbial communities are often in surprisingly good agreement with the neutral expectation. This leads us to conclude that upon colonization, neutral processes could play a significant role in the assembly of the microbiota. A comparison with environmental microbial communities suggests that biological hosts are not better at structuring the microbial community than abiotic environments. Taking into account a more complete simulation analysis of the neutral model further indicates that the divergence from the neutral expectation observed for

bioRxiv preprint doi: https://doi.org/10.1101/367243; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

some host species may ultimately be due to transient stages of the communities, rather than being the result of actual selective processes.

This finding has major implications for practical problems, such as our ability to manipulate the microbiota to restore or improve a community structure that is associated with a healthy host. A dominance of deterministic, niche-related factors would suggest that undesirable community structures may be amenable to intervention by manipulating certain host traits or the community itself. But if microbiota dynamics are mostly neutral, intervention strategies to modify microbiota structure might be in vain as there are no distinguishable states of the microbiota.

However, even a high consistency with the neutral expectation does not rule out the presence of selective abiotic or biotic processes, as non-neutral models have been shown to yield predictions consistent with neutral theory under certain conditions [54]. In particular, biological hosts, like any other abiotic environment, will select for a subset of the environmentally "available" microbes, and yet the resulting microbiota may still look mostly neutral. In fact, comparing the microbes present in seawater to the species present in individual sponges reveals that there is almost no overlap (Fig. S6A). This indicates that sponges are a highly selective environment, yet their microbiota are very consistent with the neutral expectation. A similar selective filter is found for *C. elegans*, where only a fraction of the microbes present in the environment successfully colonize the worms (Fig. S6B). This suggests that there are host traits that act as environmental filters [33] and they often will do so in a predictable, i.e. deterministic way. In this sense the microbiological tenet that "*everything is everywhere*, but *the environment selects*" also applies when the environment happens to be another biological organism. However, our results highlight that the processes that are at play *after* the environmental filter has been passed can still be highly consistent with the neutral model. It has also been suggested that this interplay between selective environments and neutral processes acting on the allowed species may have contributed to some of the contrasting results found in previous studies [55].

Another factor to consider is that the data analyzed in this study is based on 16S rRNA gene based taxonomic profiling. Given the known examples of functional redundancy present e.g. in the human gut microbiota [56], viewing communities from the perspective of functional (e.g. metabolic) categories may provide an alternative picture. This ties in with recent studies demonstrating that taxonomic and functional microbial community composition are shaped by largely independent processes [57, 58]. Thus, the dominance of neutral factors which we observe on the taxonomic level may in principle be complemented by non-neutral selection of functional groups mediated by the host.

Conclusion

The neutral assumption that species do not interact is an anathema to many ecologists, and for the microbiota in particular it is assumed that the functioning of the community is enhanced by cooperative interactions. Recent results however suggest that interactions within the mouse gut microbiota are indeed predominantly competitive and very weak [59]. Together with our finding of the microbiota being highly consistent with neutral expectations for several different host species, this lends support to taking the neutral null hypothesis as a key component of our understanding of host-microbe interactions. Awareness of the potentially significant role of neutrality in shaping the microbiota is crucial to avoid being lead astray by randomness in view of the incredible complexity of host-microbiota symbioses and to be able to uncover general principles of microbiome assembly.

bioRxiv preprint doi: https://doi.org/10.1101/367243; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Methods

The Neutral Model

Here, we summarize the model presented in [20], considering only the purely neutral special case with equivalent growth rates for all species. A local community contains a fixed number of N individuals and changes to the relative abundance of a particular species requires individuals to die. The death rate δ is the same for all species and a dead individual is replaced by reproduction of a local individual (with probability $1-m$), or by immigration from a fixed source community (with probability m). The relative abundance of a species in this source community is denoted by p_i. With this the probabilities that the number of individuals of species i increases by one, decreases by one or stays the same are given by

$$\begin{aligned} P(N_i+1) &= \frac{N-N_i}{N}\left(m\,p_i+(1-m)\frac{N_i}{N-1}\right), \\ P(N_i-1) &= \frac{N_i}{N}\left(m\,(1-p_i)+(1-m)\frac{N-N_i}{N-1}\right), \\ P(N_i) &= \frac{N_i}{N}\left(m\,p_i+(1-m)\frac{N_i-1}{N-1}\right) \\ &\quad+\frac{N-N_i}{N}\left(m\,(1-p_i)+(1-m)\frac{N-N_i-1}{N-1}\right). \end{aligned} \tag{1}$$

These transition probabilities correspond to Hubbell's original neutral model, but the following continuous approximation derived by Sloan et al. [20] can be efficiently applied to very large population sizes and allows for a relatively simple analytical solution. In particular, this allows for a calibration with the high-throughput 16S rRNA sequencing data obtained from microbial communities.

If N is large enough, such as in microbial communities, we can assume the relative abundance $x_i = N_i/N$ of species i to be continuous. Applying methods from population genetics [60] this leads to a Fokker-Planck equation for the probability density function $\phi_i(x_i,t)$ of x_i,

$$\frac{\partial\phi_i}{\partial t} = -\frac{\partial}{\partial x_i}\left(M_{\delta x}\,\phi_i\right)+\frac{1}{2}\frac{\partial^2}{\partial x_i^2}\left(V_{\delta x}\,\phi_i\right), \tag{2}$$

where $M_{\delta x}$ and $V_{\delta x}$ are the expected rates of change in frequency and variability, respectively. Assuming the time intervals between individual death-birth events are short, they can be approximated by

$$M_{\delta x} = \frac{m(p_i-x_i)}{N} \tag{3}$$

and

$$V_{\delta x} = \frac{2(1-x_i)x_i+m(p_i-x_i)(1-2x_i)}{N^2}. \tag{4}$$

Equation (2) together with (3) and (4) describes the neutral population dynamics of a local microbial community. This is not in general amenable to analytical treatment, but one can approximate the long-term equilibrium solution of this equation. Namely, the potential solution to $\partial\phi_i/\partial t = 0$ leads to $M_{\delta x}\,\phi_i - \frac{1}{2}\frac{\partial}{\partial x_i}\left(V_{\delta x}\,\phi_i\right) = \text{const.}$ with the solution approximately given by the beta distribution

$$\phi_i(x_i;N,p_i,m) = c\,(1-x_i)^{Nm(1-p_i)-1}\,x_i^{Nmp_i-1}. \tag{5}$$

This can be connected to empirical observations by realizing that, for a given detection threshold d, the probability of actually observing a species in a local community is given by the truncated cumulative probability density function

$$P(\text{species } i \text{ observed}) = \int_d^1 \phi_i(x_i;N,p_i,m)\,dx. \tag{6}$$

Here, N is given by the number of reads per sample and the relative abundance p_i of the focal species in the source community can be approximated by the mean relative abundance of the species across all samples. The probability of immigration m is thus the only free parameter and used to fit the predicted long-term distribution to the observed occurence frequencies of the microbiota.

bioRxiv preprint doi: https://doi.org/10.1101/367243; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Fitting procedure

The fitting process applied to all datasets works as follows. First, the OTU abundance table was rarefied to the same read depth (1000 reads per sample), which was determined by the lowest available read depth from all datasets. Rarefaction was performed using the *PySurvey* package (pypi.org/project/PySurvey). The expected observation probability (6) and a binomial model were then fitted to the observed mean relative OTU abundances $\overline{p_i}$ and occurence frequencies f_i obtained from this rarefied table using non-linear least squares minimization with the *lmfit* package (pypi.org/project/lmfit) [61]. As a measure of the goodness of fit we then calculated the standard coefficient of determination using the ratio of the sum of squared residuals and the total sum of squares:

$$R^2 = 1 - \frac{\sum_i (f_i - \Phi_i)^2}{\sum_i (f_i - \overline{f_i})^2}, \tag{7}$$

where Φ_i is the expected ocurrence frequency obtained from the best-fit neutral prediction. Rarefaction and fitting was repeated 1000 times and the R^2 reported throughout is the mean across individual best-fits for each dataset. Additionally, for comparison of the neutral and binomial model, the mean Aikake information criterion (AIC) was calculated from all rarefactions.

Neutral model simulations

We simulated the neutral process for 50 local communities, each containing $N = 10000$ individuals, with a source community containing 200 species and an immigration probability of $m = 0.05$.

Code availibility

The Python code for fitting and simulating the neutral model will be made available on github and in the meantime is available from the authors on request.

Data availibility

All data has been published previously, see Table S1 for the corresponding references.

Acknowledgments

This work was supported by the Collaborative Research Centre 1182 “Origin and Function of Metaorganisms” granted by the Deutsche Forschungsgemeinschaft DFG. TCGB gratefully appreciates support from the Canadian Institute for Advanced Research (CIFAR).

bioRxiv preprint doi: https://doi.org/10.1101/367243; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

References

[1] Fredrik Bäckhed, Hao Ding, Ting Wang, Lora V Hooper, Gou Young Koh, Andras Nagy, Clay F Semenkovich, and Jeffrey I Gordon. The gut microbiota as an environmental factor that regulates fat storage. *Proceedings of the National Academy of Sciences*, 101(44):15718–15723, 2004.

[2] Thomas Greiner and Fredrik Bäckhed. Effects of the gut microbiota on obesity and glucose homeostasis. *Trends in Endocrinology & Metabolism*, 22(4):117–123, 2011.

[3] Vanessa K Ridaura, Jeremiah J Faith, Federico E Rey, Jiye Cheng, Alexis E Duncan, Andrew L Kau, Nicholas W Griffin, Vincent Lombard, Bernard Henrissat, James R Bain, Muehlbauer M J, Ilkayeva O, Semenkovich C F, Funai K, Hayashi D K, Lyle B J, Martini MC, Ursell L K, Clemente J C, Van Treuren W, Walters W A, Knight R, Newgard C B, Heath A C, and Gordon J I. Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science*, 341(6150):1241214, 2013.

[4] Yasmine Belkaid and Timothy W Hand. Role of the microbiota in immunity and inflammation. *Cell*, 157(1):121–141, 2014.

[5] Thomas CG Bosch. Rethinking the role of immunity: lessons from *Hydra*. *Trends in Immunology*, 35(10):495–502, 2014.

[6] Christoph A Thaiss, Niv Zmora, Maayan Levy, and Eran Elinav. The microbiome and innate immunity. *Nature*, 535(7610):65–74, 2016.

[7] Helen E Vuong, Jessica M Yano, Thomas C Fung, and Elaine Y Hsiao. The microbiome and host behavior. *Annual Review of Neuroscience*, 40:21–49, 2017.

[8] Andrea P Murillo-Rincon, Alexander Klimovich, Eileen Pemöller, Jan Taubenheim, Benedikt Mortzfeld, René Augustin, and Thomas CG Bosch. Spontaneous body contractions are modulated by the microbiome of *Hydra*. *Scientific Reports*, 7(1):15937, 2017.

[9] Michael Shapira. Gut microbiotas and host evolution: scaling up symbiosis. *Trends in Ecology & Evolution*, 31(7):539–549, 2016.

[10] Thomas CG Bosch and Margaret J McFall-Ngai. Metaorganisms as the new frontier. *Zoology*, 114(4):185–190, 2011.

[11] Margaret McFall-Ngai, Michael G Hadfield, Thomas C G Bosch, Hannah V Carey, Tomislav Domazet-Lošo, Angela E Douglas, Nicole Dubilier, Gerard Eberl, Tadashi Fukami, Scott F Gilbert, U Hentschel, N King, S Kjelleberg, A H Knoll, N Kremer, S K Mazmanian, J L Metcalf, K Nealson, N E Pierce, J F Rawls, A Reid, E G Ruby, M Rumpho, J G Sanders, D Tautz, and J J Wernegreen. Animals in a bacterial world, a new imperative for the life sciences. *Proceedings of the National Academy of Sciences*, 110(9):3229–3236, 2013.

[12] Marc A Sze and Patrick D Schloss. Looking for a signal in the noise: revisiting obesity and the microbiome. *mBio*, 7(4):e01018–16, 2016.

[13] Daphna Rothschild, Omer Weissbrod, Elad Barkan, Alexander Kurilshikov, Tal Korem, David Zeevi, I. Costea, Paul, Anastasia Godneva, N. Kalka, Iris, Noam Bar, Smadar Shilo, Dar Lador, Vich Vila, Arnau, Niv Zmora, Meirav Pevsner-Fischer, David Israeli, Noa Kosower, Gal Malka, Chen Wolf, Bat, Tali Avnit-Sagi, Maya Lotan-Pompan, Adina Weinberger, Zamir Halpern, Shai Carmi, Jingyuan Fu, Cisca Wijmenga, Alexandra Zhernakova, Eran Elinav, and Eran Segal. Environment dominates over host genetics in shaping human gut microbiota. *Nature*, 555:210–215, 2018.

[14] Nancy A Moran and Daniel B Sloan. The hologenome concept: helpful or hollow? *PLoS Biology*, 13(12):e1002311, 2015.

[15] H Caswell. Community structure: A neutral model analysis. *Ecological Monographs*, 46(3):327–354, 1976.

[16] Stephen P Hubbell. *The unified neutral theory of biodiversity and biogeography*. Monographs in Population Biology. Princeton University Press, Princeton, NJ, USA, 2001.

bioRxiv preprint doi: https://doi.org/10.1101/367243; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[17] Igor Volkov, Jayanth R Banavar, Stephen P Hubbell, and Amos Maritan. Neutral theory and relative species abundance in ecology. *Nature*, 424(6952):1035–1037, 2003.

[18] R. Macarthur and E. O. Wilson. *The theory of island biogeography*. Princeton University Press, Princeton, NJ, 1967.

[19] James Rosindell, Stephen P Hubbell, Fangliang He, Luke J Harmon, and Rampal S Etienne. The case for ecological neutral theory. *Trends in Ecology & Evolution*, 27(4):203–208, 2012.

[20] William T Sloan, Mary Lunn, Stephen Woodcock, Ian M Head, Sean Nee, and Thomas P Curtis. Quantifying the roles of immigration and chance in shaping prokaryote community structure. *Environmental Microbiology*, 8(4):732–740, 2006.

[21] Stephen Woodcock, Christopher J Van Der Gast, Thomas Bell, Mary Lunn, Thomas P Curtis, Ian M Head, and William T Sloan. Neutral assembly of bacterial communities. *FEMS Microbiology Ecology*, 62(2):171–180, 2007.

[22] Alex J Dumbrell, Michaela Nelson, Thorunn Helgason, Calvin Dytham, and Alastair H Fitter. Relative roles of niche and neutral processes in structuring a soil microbial community. *The ISME Journal*, 4(3):337–345, 2010.

[23] Irina Dana Ofiţeru, Mary Lunn, Thomas P Curtis, George F Wells, Craig S Criddle, Christopher A Francis, and William T Sloan. Combined niche and neutral effects in a microbial wastewater treatment community. *Proceedings of the National Academy of Sciences*, 107(35):15345–15350, 2010.

[24] Jack E Lee, Hannah L Buckley, Rampal S Etienne, and Gavin Lear. Both species sorting and neutral processes drive assembly of bacterial communities in aquatic microcosms. *FEMS Microbiology Ecology*, 86(2):288–302, 2013.

[25] E K Costello, K Stagaman, L Dethlefsen, B J Bohannan, and D A Relman. The application of ecological theory toward an understanding of the human microbiome. *Science*, 336(6086):1255–1262, 2012.

[26] Patricio Jeraldo, Maksim Sipos, Nicholas Chia, Jennifer M Brulc, A Singh Dhillon, Michael E Konkel, Charles L Larson, Karen E Nelson, Ani Qu, Lawrence B Schook, F Yang, B A White, and N Goldenfeld. Quantification of the relative roles of niche and neutral processes in structuring gastrointestinal microbiomes. *Proceedings of the National Academy of Sciences*, 109(25):9692–9698, 2012.

[27] Claudia Sala, Silvia Vitali, Enrico Giampieri, Ìtalo Faria do Valle, Daniel Remondini, Paolo Garagnani, Matteo Bersanelli, Ettore Mosca, Luciano Milanesi, and Gastone Castellani. Stochastic neutral modelling of the gut microbiota's relative species abundance from next generation sequencing data. *BMC Bioinformatics*, 17(Suppl 2):S16, 2016.

[28] A R Burns, W Z Stephens, K Stagaman, S Wong, J F Rawls, K Guillemin, and B J Bohannan. Contribution of neutral processes to the assembly of gut microbial communities in the zebrafish over host development. *The ISME Journal*, 10(3):655–664, 2016.

[29] Lianwei Li and Zhanshan Sam Ma. Testing the neutral theory of biodiversity with human microbiome datasets. *Scientific Reports*, 6:31448, 2016.

[30] Hye-Jin Kim, Hanbyul Kim, Jin Ju Kim, Nu Ri Myeong, Taeyune Kim, Taehun Park, Eunjoo Kim, Ji-yei Choi, Johnhwan Lee, Susun An, et al. Fragile skin microbiomes in megacities are assembled by a predominantly niche-based process. *Science Advances*, 4(3):e1701581, 2018.

[31] Arvind Venkataraman, Christine M Bassis, James M Beck, Vincent B Young, Jeffrey L Curtis, Gary B Huffnagle, and Thomas M Schmidt. Application of a neutral community model to assess structuring of the human lung microbiome. *mBio*, 6(1):e02284–14, 2015.

[32] K L Adair, M Wilson, A Bost, and A E Douglas. Microbial community assembly in wild populations of the fruit fly *Drosophila melanogaster*. *The ISME Journal*, 2018.

[33] Karen L Adair and Angela E Douglas. Making a microbiome: the many determinants of host-associated microbial community composition. *Current Opinion in Microbiology*, 35:23–29, 2017.

bioRxiv preprint doi: https://doi.org/10.1101/367243; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[34] R Feuda, M Dohrmann, W Pett, H Philippe, O Rota-Stabelli, N Lartillot, G Wörheide, and D Pisani. Improved modeling of compositional heterogeneity supports sponges as sister to all other animals. *Current Biology*, 27(24):3864–3870, 2017.

[35] Torsten Thomas, Lucas Moitinho-Silva, Miguel Lurgi, Johannes R Björk, Cole Easson, Carmen Astudillo-García, Julie B Olson, Patrick M Erwin, Susanna López-Legentil, Heidi Luter, A Chaves-Fonnegra, R Costa, P J Schupp, L Steindler, D Erpenbeck, J Gilbert, R Knight, G Ackermann, J Victor Lopez, M W Taylor, R W Thacker, J M Montoya, U Hentschel, , and N S Webster. Diversity, structure and convergent evolution of the global sponge microbiome. *Nature Communications*, 7:11870, 2016.

[36] Lucia Pita, Laura Rix, Beate M Slaby, Andrea Franke, and Ute Hentschel. The sponge holobiont in a changing ocean: from microbes to ecosystems. *Microbiome*, 6(1):46, 2018.

[37] Bosch, Thomas CG. Cnidarian-microbe interactions and the origin of innate immunity in metazoans. *Annual Review of Microbiology*, 67:499–518, 2013.

[38] Buck S Samuel, Holli Rowedder, Christian Braendle, Marie-Anne Félix, and Gary Ruvkun. *Caenorhabditis elegans* responses to bacteria from its natural habitats. *Proceedings of the National Academy of Sciences*, 113(27):E3941–E3949, 2016.

[39] Philipp Dirksen, Sarah Arnaud Marsh, Ines Braker, Nele Heitland, Sophia Wagner, Rania Nakad, Sebastian Mader, Carola Petersen, Vienna Kowallik, Philip Rosenstiel, M A Félix, and H Schulenburg. The native microbiome of the nematode *Caenorhabditis elegans*: gateway to a new host-microbiome model. *BMC Biology*, 14(1):38, 2016.

[40] Maureen Berg, Ben Stenuit, Joshua Ho, Andrew Wang, Caitlin Parke, Matthew Knight, Lisa Alvarez-Cohen, and Michael Shapira. Assembly of the *Caenorhabditis elegans* gut microbiota from diverse soil microbial environments. *The ISME Journal*, 10(8):1998–2009, 2016.

[41] Fan Zhang, Maureen Berg, Katja Dierking, Marie-Anne Félix, Michael Shapira, Buck S Samuel, and Hinrich Schulenburg. *Caenorhabditis elegans* as a model for microbiome research. *Frontiers in Microbiology*, 8:485, 2017.

[42] Stephen Abolins, Elizabeth C King, Luke Lazarou, Laura Weldon, Louise Hughes, Paul Drescher, John G Raynes, Julius CR Hafalla, Mark E Viney, and Eleanor M Riley. The comparative immunology of wild and laboratory mice, *Mus musculus domesticus*. *Nature Communications*, 8:14811, 2017.

[43] Miriam Linnenbrink, Jun Wang, Emilie A Hardouin, Sven Künzel, Dirk Metzler, and John F Baines. The role of biogeography in shaping diversity of the intestinal microbiota in house mice. *Molecular Ecology*, 22(7):1904–1916, 2013.

[44] Jun Wang, Shirin Kalyan, Natalie Steck, Leslie M Turner, Bettina Harr, Sven Künzel, Marie Vallier, Robert Häsler, Andre Franke, Hans-Heinrich Oberg, et al. Analysis of intestinal microbiota in hybrid house mice reveals evolutionary divergence in a vertebrate hologenome. *Nature Communications*, 6:6440, 2015.

[45] J Gregory Caporaso, Christian L Lauber, William A Walters, Donna Berg-Lyons, James Huntley, Noah Fierer, Sarah M Owens, Jason Betley, Louise Fraser, Markus Bauer, N. Gormley, J. A. Gilbert, G. Smith, and Knight R. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME Journal*, 6(8):1621–1624, 2012.

[46] Javier E Irazoqui, Jonathan M Urbach, and Frederick M Ausubel. Evolution of host innate defence: insights from *Caenorhabditis elegans* and primitive invertebrates. *Nature Reviews Immunology*, 10(1):47–58, 2010.

[47] Maria-Isabel Fernandez, Béatrice Regnault, Céline Mulet, Myriam Tanguy, Philippe Jay, Philippe J Sansonetti, and Thierry Pédron. Maturation of paneth cells induces the refractory state of newborn mice to *Shigella* infection. *The Journal of Immunology*, 180(7):4924–4930, 2008.

[48] Werner EG Müller and Isabel M Müller. Origin of the metazoan immune system: Identification of the molecules and their functions in sponges. *Integrative and Comparative Biology*, 43(2):281–292, 2003.

bioRxiv preprint doi: https://doi.org/10.1101/367243; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[49] Nicole M Vega and Jeff Gore. Stochastic assembly produces heterogeneous communities in the *Caenorhabditis elegans* intestine. *PLoS Biology*, 15(3):e2000633, 2017.

[50] Francisco Dini-Andreote, James C Stegen, Jan Dirk van Elsas, and Joana Falcão Salles. Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession. *Proceedings of the National Academy of Sciences*, 112(11):E1326–E1332, 2015.

[51] Benedikt M Mortzfeld, Szymon Urbanski, Adam M Reitzel, Sven Künzel, Ulrich Technau, and Sebastian Fraune. Response of bacterial colonization in *Nematostella vectensis* to development, environment and biogeography. *Environmental Microbiology*, 18(6):1764–1781, 2016.

[52] Harry J Flint, Edward A Bayer, Marco T Rincon, Raphael Lamed, and Bryan A White. Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. *Nature Reviews Microbiology*, 6(2):121–131, 2008.

[53] Hannelore Daniel, Amin Moghaddas Gholami, David Berry, Charles Desmarchelier, Hannes Hahne, Gunnar Loh, Stanislas Mondot, Patricia Lepage, Michael Rothballer, Alesia Walker, C Böhm, M Wenning, M Wagner, Blaut, P Schmitt-Kopplin, B Kuster, D Haller, and T Clavel. High-fat diet alters gut microbiota physiology in mice. *The ISME Journal*, 8(2):295–308, 2014.

[54] Steven C Walker. When and why do non-neutral metacommunities appear neutral? *Theoretical Population Biology*, 71(3):318–331, 2007.

[55] Ekaterina Avershina, Ola Storrø, Torbjørn Øien, Roar Johnsen, Phil Pope, and Knut Rudi. Major faecal microbiota shifts in composition and diversity with age in a geographically restricted cohort of mothers and their children. *FEMS Microbiology Ecology*, 87(1):280–290, 2014.

[56] Catherine A Lozupone, Jesse I Stombaugh, Jeffrey I Gordon, Janet K Jansson, and Rob Knight. Diversity, stability and resilience of the human gut microbiota. *Nature*, 489(7415):220–230, 2012.

[57] Stilianos Louca, Laura Wegener Parfrey, and Michael Doebeli. Decoupling function and taxonomy in the global ocean microbiome. *Science*, 353(6305):1272–1277, 2016.

[58] Stilianos Louca, Saulo MS Jacques, Aliny PF Pires, Juliana S Leal, Diane S Srivastava, Laura Wegener Parfrey, Vinicius F Farjalla, and Michael Doebeli. High taxonomic variability despite stable functional structure across microbial communities. *Nature Ecology & Evolution*, 1(1):0015, 2016.

[59] K Z Coyte, J Schluter, and K R Foster. The ecology of the microbiome: Networks, competition, and stability. *Science*, 350(6261):663–666, 2015.

[60] J. F. Crow and M. Kimura. *An Introduction to Population Genetics Theory*. Harper and Row, New York, 1970.

[61] M Newville, T Stensitzki, D Allen, and A Ingargiola. Lmfit: Non-linear least-squares minimization and curve-fitting for python. Zenodo, 2014.

[62] S. Franzenburg, S. Fraune, P.M. Altrock, S. Kuenzel, J.F. Baines, A. Traulsen, and T. C. G. Bosch. Bacterial colonization of hydra hatchlings follows a robust temporal pattern. *ISME Journal*, 7:781–790, 2013.

bioRxiv preprint doi: https://doi.org/10.1101/367243; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Supporting Information

bioRxiv preprint doi: https://doi.org/10.1101/367243; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table S1 The datasets included in this study and the results of the model fits. Data is for rarefaction level of 1000 reads per sample.

Host species	#samples	#OTUs	best-fit m	R^2 neutral model	AIC neutral model	AIC binomial model	Reference
Caenorhabditis elegans							[39]
natural samples	22	193	0.03 ± 0.004	0.44	-673.2	-452.5	
lab samples	34	106	0.007 ± 0.001	0.38	-330.0	-208.2	
Ircinia oros	11	1121	0.41 ± 0.03	0.6	-5391.47	-4966.26	[35]
Sarcotragus fasciculatus	12	735	0.83 ± 0.08	0.81	-3552.82	-3165.09	[35]
Carteriospongia foliascens							[35]
Australia, Davies Reef	15	939	0.36 ± 0.03	0.6	-5435.40	-5007.39	
Australia, Fantome Island	14	728	0.81 ± 0.1	0.79	-3740.45	-3336.08	
Australia, Orpheus Island	15	750	0.86 ± 0.1	0.8	-3987.01	-3539.84	
Australia, Green Island	13	774	0.47 ± 0.05	0.59	-3846.0	-3542.79	
Australia, Torres Strait	7	336	0.78 ± 0.15	0.58	-1410.75	-1244.0	
Mus musculus							[44]
natural samples	69	281	0.11 ± 0.01	0.85	-1217.97	-998.15	
lab samples	54	136	0.18 ± 0.02	0.84	-536.30	-452.48	
Nematostella vectensis							[51]
1 dpf	6	226	0.23 ± 0.02	0.52	-738.68	-655.06	
4 dpf	15	149	0.33 ± 0.06	0.67	-521.62	-479.29	
40 dpf	12	155	0.09 ± 0.01	0.59	-502.39	-413.96	
123 dpf	12	195	0.39 ± 0.06	0.69	-678.72	-623.49	
385 dpf	20	225	0.17 ± 0.02	0.58	-766.93	-679.74	
401 dpf	8	120	0.7 ± 0.15	0.74	-446.34	-402.59	
Hydra vulgaris							[62]
0.5 wah	8	699	0.6 ± 0.07	0.5	-2527.91	-2289.35	
2.5 wah	8	248	0.28 ± 0.05	0.44	-820.92	-756.19	
5 wah	8	242	0.53 ± 0.1	0.43	-1129.85	-1041.28	
9 wah	8	257	0.5 ± 0.1	0.44	-833.46	-765.88	
15 wah	8	140	0.45 ± 0.12	0.59	-466.20	-432.55	
Aurelia aurita							submitted
control	5	163	0.91 ± 0.07	0.69	-624.96	-538.45	
quorum quenching (QQ)	18	391	0.62 ± 0.06	0.85	-1716.09	-1532.6	
Environment							
Compost	65	587	0.48 ± 0.03	0.86	-2690.26	-2491.77	[39]
Seawater	16	2518	0.62 ± 0.03	0.7	-12345.33	-11113.17	[35]
Sediment	12	3796	0.77 ± 0.05	0.47	-16658.27	-14665.48	[35]

bioRxiv preprint doi: https://doi.org/10.1101/367243; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

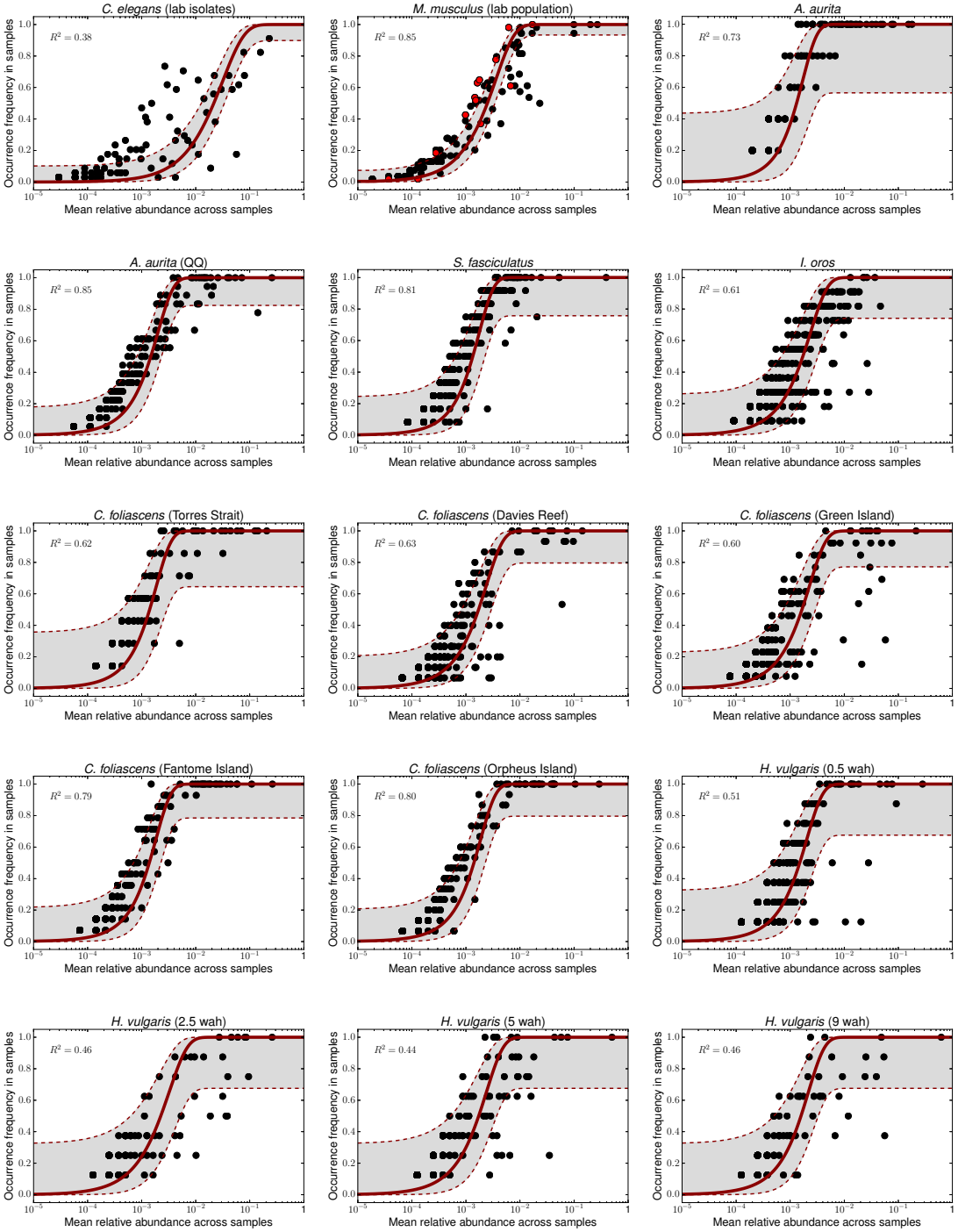

Figure S1 The relationship between mean relative abundance of taxa across all samples and the frequency with which they are detected in individual samples. Each dot represents a taxon and the solid line is the best-fitting neutral community expectation. The dashed lines and grey area depict the 95% confidence bands.

bioRxiv preprint doi: https://doi.org/10.1101/367243; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

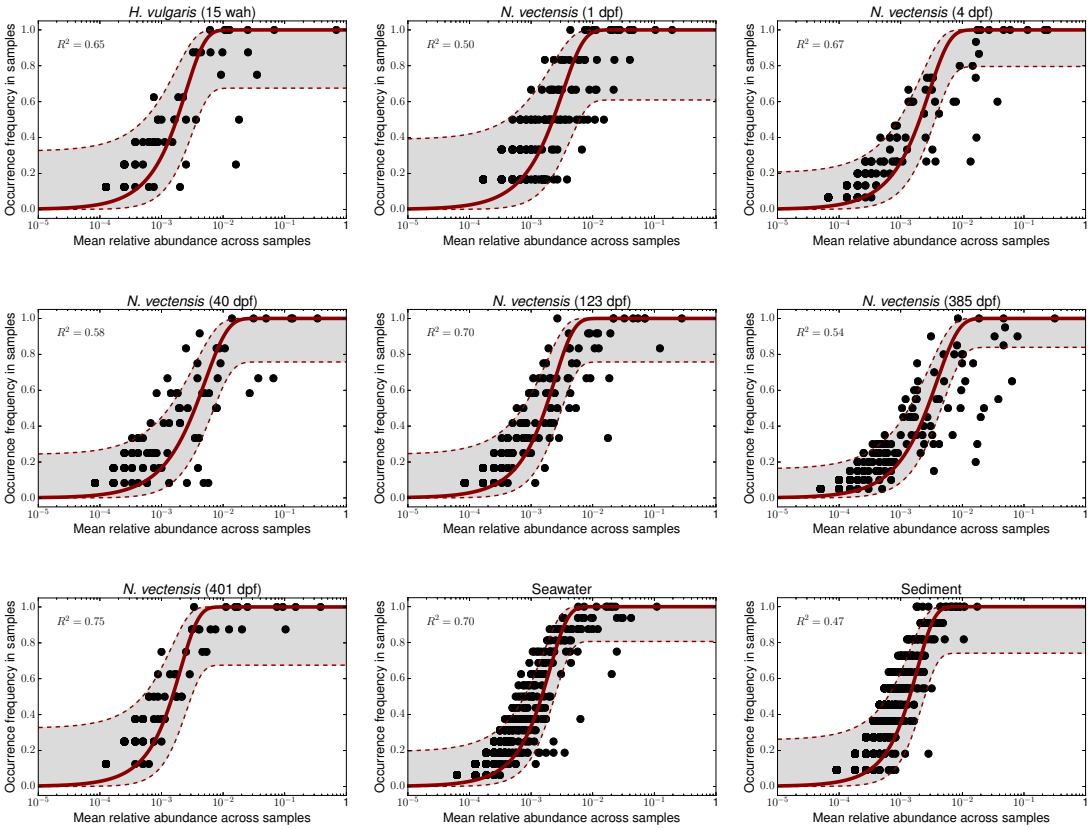

Figure S1 he relationship between mean relative abundance of taxa across all samples and the frequency with which they are detected in individual samples. Each dot represents a taxon and the solid line is the best-fitting neutral community expectation. The dashed lines and grey area depict the 95% confidence bands.

bioRxiv preprint doi: https://doi.org/10.1101/367243; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

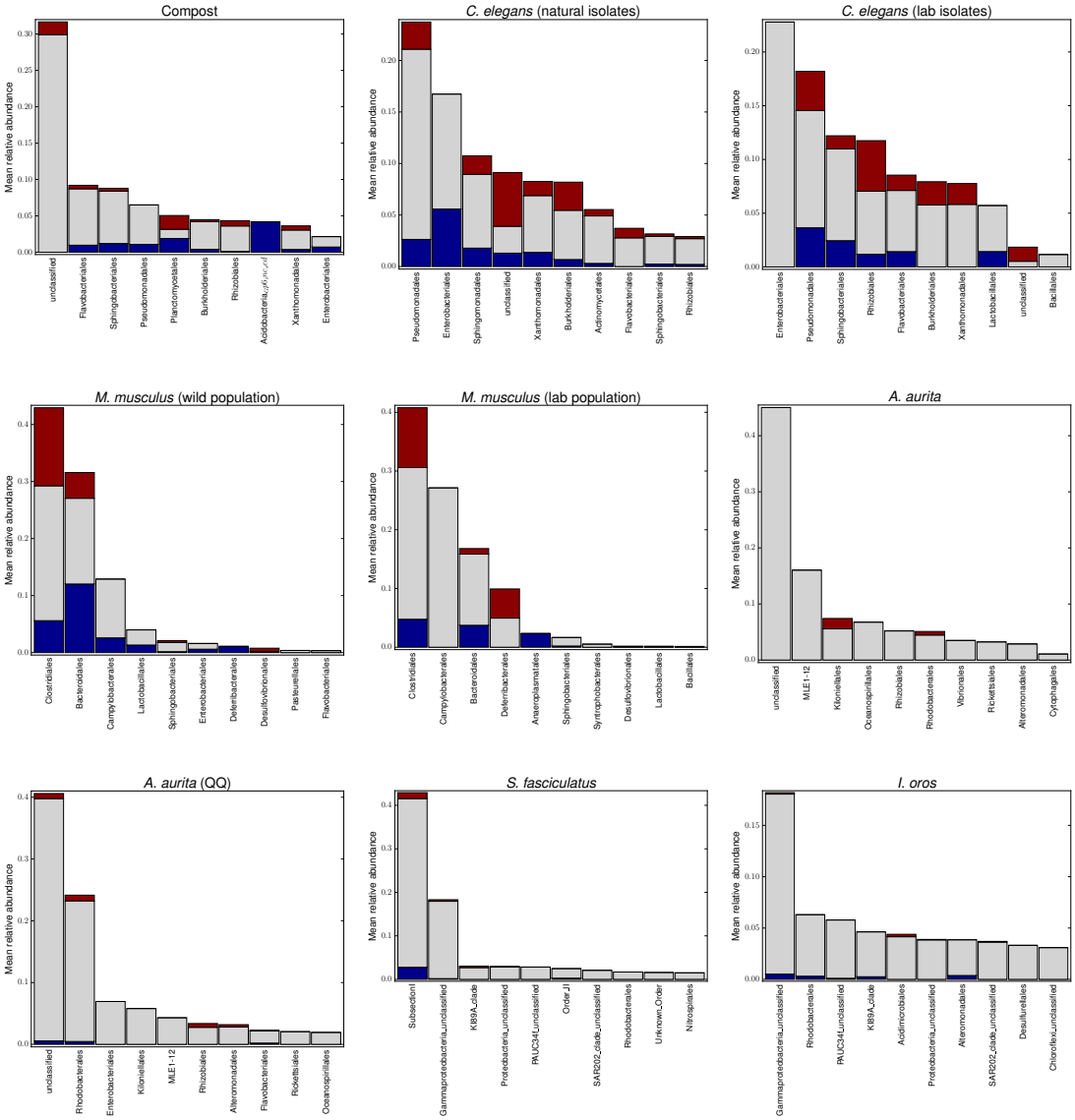

Figure S2 Bars show the mean relative abundance of the ten most abundant microbial orders in a specific host population. The colored sections indicate the fractions of OTUs within that order which were found above the neutral expectation (red), within the neutral expectation (grey) and below the neutral expectiation (blue).

bioRxiv preprint doi: https://doi.org/10.1101/367243; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

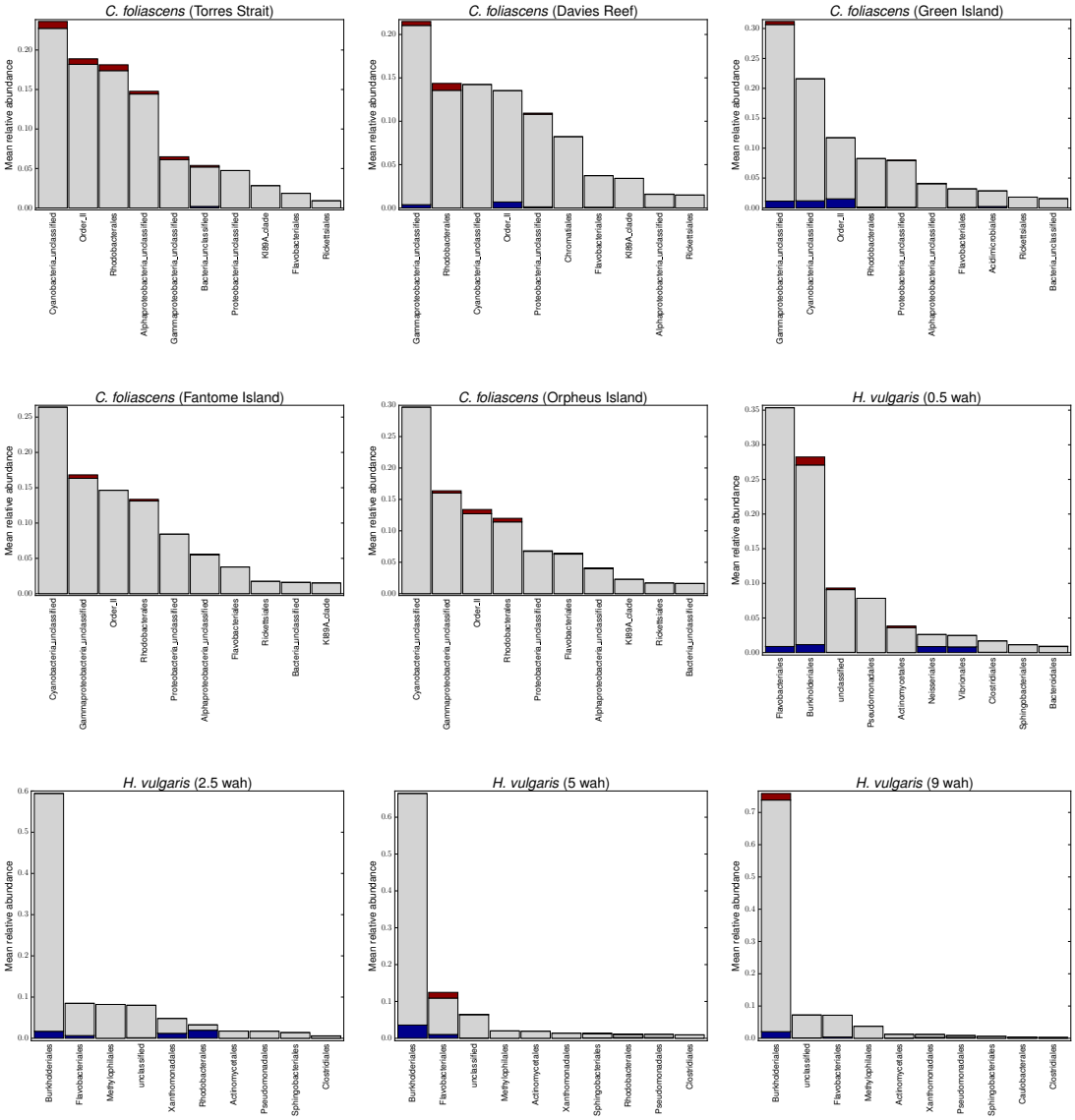

Figure S2 Bars show the mean relative abundance of the ten most abundant microbial orders in a specific host population. The colored sections indicate the fractions of OTUs within that order which were found above the neutral expectation (red), within the neutral expectation (grey) and below the neutral expectiation (blue).

bioRxiv preprint doi: https://doi.org/10.1101/367243; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

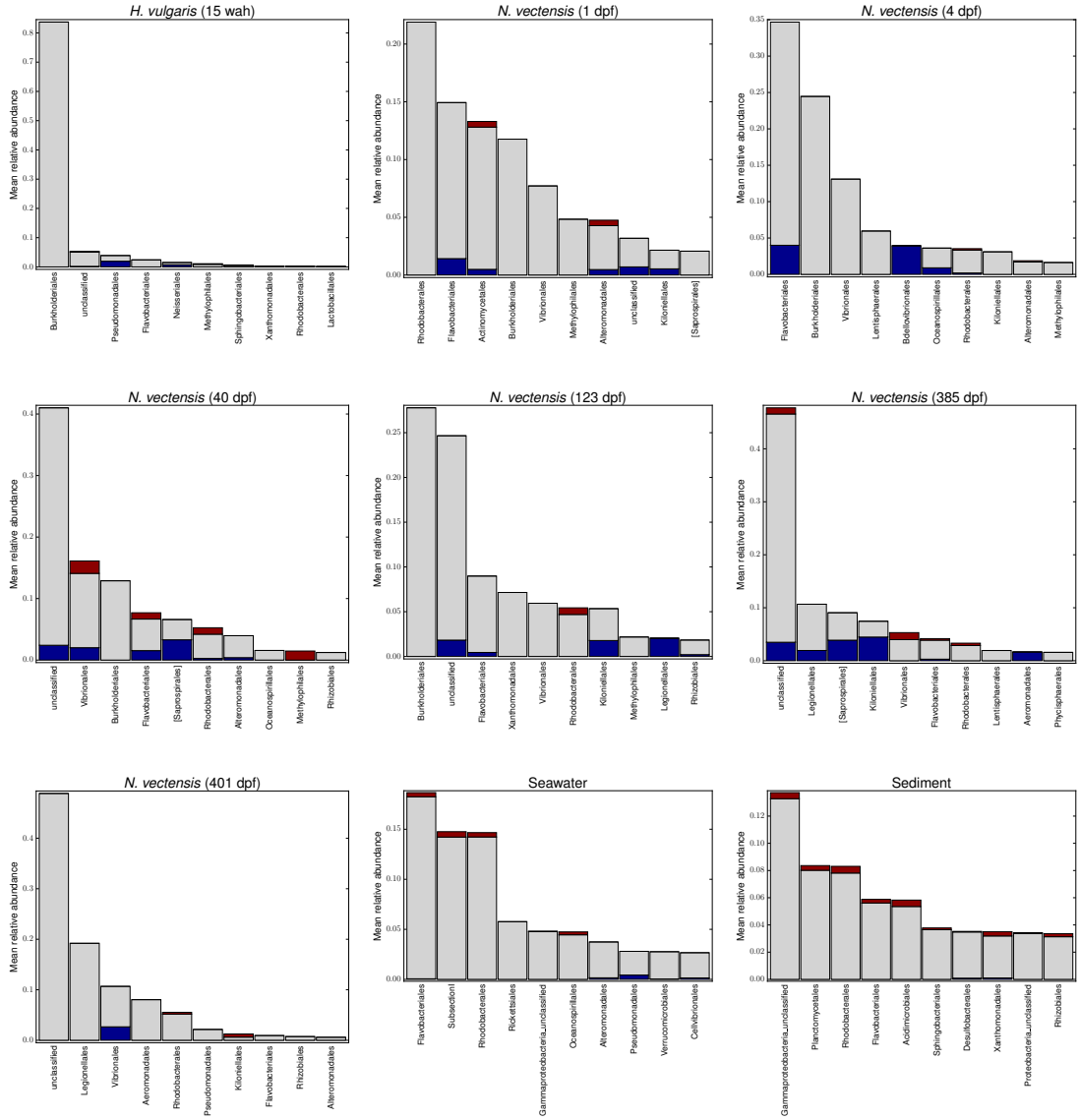

Figure S2 Bars show the mean relative abundance of the ten most abundant microbial orders in a specific host population. The colored sections indicate the fractions of OTUs within that order which were found above the neutral expectation (red), within the neutral expectation (grey) and below the neutral expectiation (blue).

bioRxiv preprint doi: https://doi.org/10.1101/367243; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table S2 Non-neutral genera found in the *C. elegans* microbiota

Domain	Phylum	Class	Order	Family	Genus	
Bacteria	Bacteroidetes	Flavobacteria	Flavobacteriales	Flavobacteriaceae	unclassified	over-represented
Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Hyphomicrobiaceae	Devosia	
Bacteria	Proteobacteria	Alphaproteobacteria	unclassified	unclassified	unclassified	
Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Alcaligenaceae	unclassified	
Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	unclassified	
Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	unclassified	unclassified	
Bacteria	Proteobacteria	Betaproteobacteria	unclassified	unclassified	unclassified	
Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	unclassified	
Bacteria	Proteobacteria	Gammaproteobacteria	unclassified	unclassified	unclassified	
Bacteria	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	unclassified	
Bacteria	unclassified	unclassified	unclassified	unclassified	unclassified	
Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Brucellaceae	Ochrobactrum	under-represented

Table S3 Non-neutral genera found in the *M. musculus* microbiota

Domain	Phylum	Class	Order	Family	Genus	
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Acetomicrobium	over-represented
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	Proteiniphilum	
Bacteria	Bacteroidetes	Flavobacteria	Flavobacteriales	Cryomorphaceae	Fluviicola	
Bacteria	Firmicutes	Clostridia	Clostridiales	Clostridiaceae 4	Caminicella	
Bacteria	Firmicutes	Clostridia	Clostridiales	Clostridiales_Incertae Sedis XIII	Anaerovorax	
Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Clostridium XIVb	
Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Coprococcus	
Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Dorea	
Bacteria	Firmicutes	Clostridia	Clostridiales	Peptococcaceae 1	Peptococcus	
Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Acetanaerobacterium	
Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Anaerotruncus	
Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Ethanoligenens	
Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Flavonifractor	
Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Hydrogenoanaerobacterium	
Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Sporobacter	
Bacteria	Firmicutes	Negativicutes	Selenomonadales	Veillonellaceae	Anaerospora	
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Marinilabiaceae	Anaerophaga	under-represented
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	Butyricimonas	
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	Paludibacter	
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	Tannerella	
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Rikenellaceae	Rikenella	
Bacteria	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Flammeovirgaceae	Sediminitomix	
Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Blautia	
Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Butyrivibrio	
Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Catonella	
Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Marvinbryantia	

bioRxiv preprint doi: https://doi.org/10.1101/367243; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

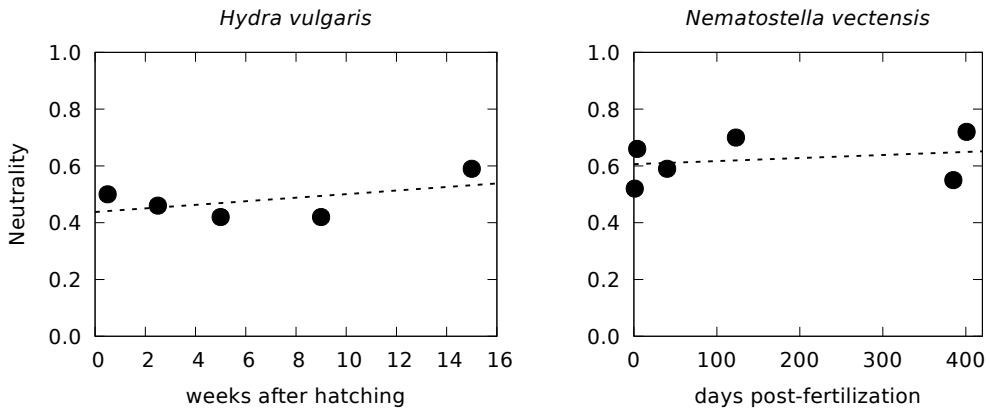

Figure S3 Neutrality of the microbiota of *Hydra vulgaris* and *Nematostella vectensis* at different time points during development and adult life stages. The dashed lines are the best linear fits, slopes are not significantly different from zero.

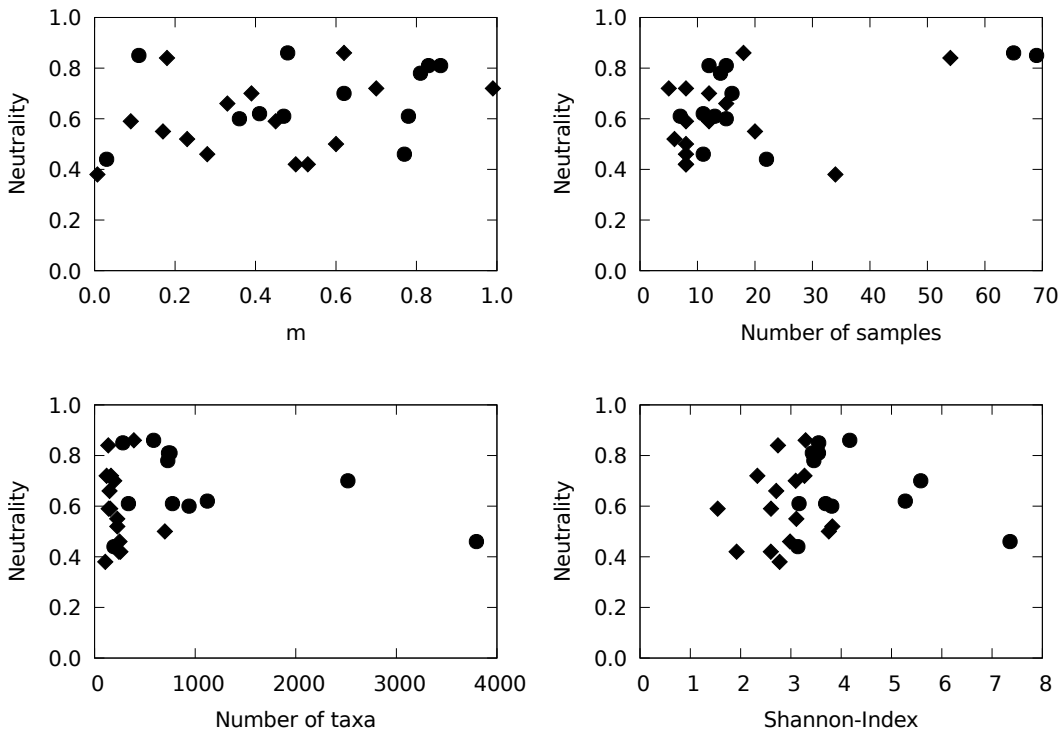

Figure S4 Consistency with the neutral model vs. the estimated dispersal parameter *m* (top left), the number of samples (top right), number of identified taxa (bottom left) and the Shannon-Index of diversity (bottom right). Circles denote natural populations and diamonds laboratory populations.

bioRxiv preprint doi: https://doi.org/10.1101/367243; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

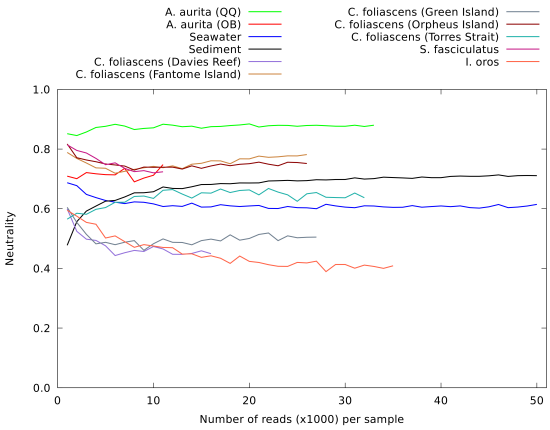

Figure S5 Effect of rarefaction on neutrality. Generally, consistency with the neutral model initially decreased with increasing read depth, until it leveled off as read depth increased further. For communities that showed a high consistency with the neutral model, varying read depths did not affect the results much, ranging from almost no effect at all for *A. aurita* to a slight drop in neutrality for the seawater samples from $R^2 = 0.7$ at 1000 reads/sample to $R^2 \approx 0.6$ at 50000 reads/sample. Only for the communities associated with the sponge *I. oros* and two of the *C. foliascens* populations did read depth show a more pronounced effect, where in both cases neutrality dropped from $R^2 \approx 0.6$ at 1000 reads/sample to $R^2 \approx 0.45$ when read depth was exceeding 10000 reads/sample. Interestingly, an opposite trend was observed for the sediment samples, where neutrality increased from $R^2 \approx 0.5$ at 1000 reads/sample to $R^2 \approx 0.7$ at 10000 reads/sample. The minimal to moderate changes in neutrality with increasing read depth for some datasets potentially reflects the influence of rare, non-neutral taxa, which are only detected with higher read depths.

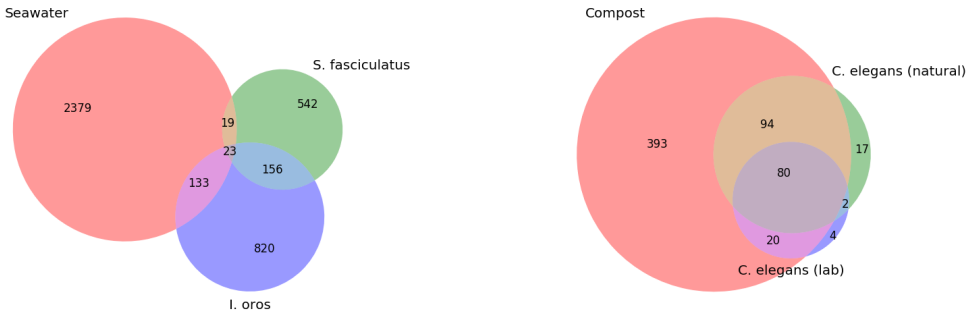

Figure S6 Overlap of taxa found in hosts and the environment. Left: For two sponge species, there is only a very small overlap between the sponge microbiota and the taxa found in seawater. Right: For *C. elegans*, only a subset of the environmentally available microbes is found in the worms.