ORIGINAL ARTICLE

Contribution of neutral processes to the assembly of gut microbial communities in the zebrafish over host development

Adam R Burns[1], W Zac Stephens[2], Keaton Stagaman[1], Sandi Wong[3,4], John F Rawls[3], Karen Guillemin[5] and Brendan JM Bohannan[1]

[1]Institute of Ecology and Evolution, University of Oregon, Eugene, OR, USA; [2]Department of Pathology, University of Utah, Salt Lake City, UT, USA; [3]Department of Molecular Genetics and Microbiology, Duke University, Durham, NC, USA; [4]Department of Microbiology and Immunology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA and [5]Institute of Molecular Biology, University of Oregon, Eugene, OR, USA

Despite their importance to host health and development, the communities of microorganisms associated with humans and other animals are characterized by a large degree of unexplained variation across individual hosts. The processes that drive such inter-individual variation are not well understood. To address this, we surveyed the microbial communities associated with the intestine of the zebrafish, *Danio rerio*, over developmental time. We compared our observations of community composition and distribution across hosts with that predicted by a neutral assembly model, which assumes that community assembly is driven solely by chance and dispersal. We found that as hosts develop from larvae to adults, the fit of the model to observed microbial distributions decreases, suggesting that the relative importance of non-neutral processes, such as microbe-microbe interactions, active dispersal, or selection by the host, increases as hosts mature. We also observed that taxa which depart in their distributions from the neutral prediction form ecologically distinct sub-groups, which are phylogenetically clustered with respect to the full metacommunity. These results demonstrate that neutral processes are sufficient to generate substantial variation in microbiota composition across individual hosts, and suggest that potentially unique or important taxa may be identified by their divergence from neutral distributions.

The ISME Journal (2016) **10,** 655–664; doi:10.1038/ismej.2015.142; published online 21 August 2015

Introduction

The microorganisms that reside on and inside humans and other vertebrate animals are remarkable, not only because of their importance to host health and development (Bates *et al.,* 2006; Fraune and Bosch, 2010; Sommer and Bäckhed, 2013), but also because they assemble into complex communities *de novo* in every new hatchling or infant host. The processes responsible for structuring these complex systems, often referred to as a host's *microbiota*, are not well understood despite a strong interest in manipulating them to improve human health. Recent advances in sequencing allow us to observe and describe microbial communities with unprecedented depth and accuracy, but using these data to make inferences about how they assemble remains challenging. One approach to addressing this issue is

to adopt a conceptual framework where animal hosts are viewed as ecosystems and their associated microbiota are treated as ecological communities (Dethlefsen *et al.,* 2006; Robinson *et al.,* 2010; Costello *et al.,* 2012). This approach is attractive because it allows researchers to borrow concepts and tools developed over decades of research in ecology.

In host-associated systems, there are a large number of specific factors that may contribute to community assembly. Many host-specific factors have been studied, including host species, genotype, diet and health (Rawls *et al.,* 2006; Turnbaugh *et al.,* 2006; Benson *et al.,* 2010; Goodrich *et al.,* 2014), as well as microbe-specific factors, including mutualistic and competitive interactions (de Muinck *et al.,* 2013; Levy and Borenstein, 2013). While the list of potential factors is long, they can be divided into two major categories: selective processes, in which microbes establish and thrive in an environment (in this case the host itself) due to differences in their relative ecological fitness; and neutral processes, which include the dynamics of passive dispersal (for example, sampling individuals from a source pool of available colonists) and the effects of

Correspondence: AR Burns, Institute of Ecology and Evolution, University of Oregon, 335 Pacific Hall, 5289, Eugene, OR 97403-5289, USA. Tel: +1 541 514 1539; Fax: +1 541 346 2364;
E-mail: arayburns@gmail.com

656

ecological drift (the stochastic loss and replacement of individuals; Chase and Myers, 2011). While considerable progress has been made to investigate the roles of specific interactions between microbes and their hosts, the relative roles of dispersal and ecological drift in shaping host-associated microbial communities have largely been ignored (but see Jeraldo *et al.,* 2012; Lankau *et al.,* 2012 and Venktaraman *et al.,* 2015). In contrast, these processes have been studied in the general field of ecology for decades, with a renewed surge of interest in recent years (Caswell, 1976; Hubbell, 2001; Rosindell *et al.,* 2011).

Neutral and other sampling-based theories provide an ideal starting place for investigating assembly patterns because of their relative simplicity. Neutral theory derives its name from its defining assumption of equivalent per-capita growth, death and dispersal rates of species, thus assuming species are 'neutral' in their ecological fitness. In the absence of such differences, community assembly is the result of the stochastic processes of dispersal and drift; organisms in the community are randomly lost, and are replaced at random by individuals from within the community or by dispersal of individuals from outside the community. While these assumptions of ecological equivalence may seem over-simplified, neutral models have successfully predicted the structures of many communities, including microbial communities (Woodcock *et al.,* 2007; Östman *et al.,* 2010; Ofiteru *et al.,* 2010; Venkataraman *et al.,* 2015). Such models are particularly useful in modeling microbial systems, where the immense diversity of communities makes characterizing the specific ecological traits of each individual taxon difficult. They also allow researchers to quantify the importance of processes which are difficult to observe directly, such as dispersal, but can nevertheless have large impacts on microbial communities (Kerr *et al.,* 2002; Lindström and Östman, 2011).

Given the variable nature of host-associated microbial communities, a comprehensive investigation of the role of neutral processes in structuring these communities requires a high degree of replication and control. In this regard, the intestinal microbiota of the zebrafish (*Danio rerio*) is an ideal experimental system. Zebrafish have historically been used to study vertebrate development, but have also recently emerged as an ideal model for studying interactions between vertebrate hosts and their associated microbial communities (Rawls *et al.,* 2006; Roeselers *et al.,* 2011; Yan *et al.,* 2012; Stephens *et al.,* 2015). This is in large part due to the feasibility of raising a large number of individuals from a single crossing and co-housing them throughout their lifespan in highly controlled environments, thereby minimizing the effects of inter-host variation and ensuring that all individuals are exposed to a shared source pool of microorganisms.

In the present study, we assess the ability of neutral models to explain the distribution of microorganisms among a population of zebrafish, and then determine the conditions leading to departures from neutral behavior. In doing so, we adopt a conceptual framework in which we consider the microorganisms associated with individual zebrafish hosts to be *local communities* that are a part of a broader *metacommunity* consisting of the microorganisms associated with all of the hosts in the population (Leibold *et al.,* 2004; Costello *et al.,* 2012). We hypothesized that the ability of hosts to differentially select their microbial inhabitants increases with developmental age, thereby decreasing the relative importance of neutral processes. Assuming that decreases in the fit of the neutral model are indicative of increased selection pressures, we expected that deviations from the neutral prediction should be compositionally and phylogenetically distinct, to the extent that ecological traits are phylogenetically conserved. In addressing these hypotheses, we also provide a framework for identifying communities and taxa of potential interest based on the degree to which they diverge from the predictions of neutral theory.

Materials and methods

Zebrafish microbiota longitudinal study
For the present study, we used a 16S rRNA gene sequence data set from a previously reported longitudinal study of the developing zebrafish intestinal microbiome (Stephens *et al.,* 2015). A brief description of the study design and sample collection follows, and readers are referred to Stephens *et al.* (2015) for additional details. A population of zebrafish resulting solely from a single mating pair was raised under identical conditions, to minimize both genetic and environmental heterogeneity, and sampled at multiple ages, conventionally measured by days post fertilization (dpf). Zebrafish embryos develop within sterile chorions and are not exposed to microorganisms in their environment until they hatch (between 2 and 3 dpf). This population was divided evenly among four replicate tanks (resulting in 70 fish per tank) before hatching to ensure a shared initial exposure and account for potential tank effects. These fish were then raised under standard laboratory rearing conditions. The intestines of individual fish from this population were aseptically removed (as per Milligan-Myhre *et al.,* 2011), and the associated microbial communities were characterized by 16S rRNA gene amplicon sequencing at important developmental milestones: 4 dpf (complete opening of the digestive tract), 10 dpf (after feeding began), 21, 28 and 35 dpf (activation of the adaptive immune system, Lam *et al.,* 2004), 75 dpf (sexual maturity), and finally 380 dpf (onset of senescence). At each time point, 20 fish (five from each tank) were

randomly selected for sampling, with the exception of the 75 dpf time point at which time 24 fish (three male and three female from each tank; before 75 dpf the sex of fish could not be confidently determined), and the 380 dpf time point, at which time 6 fish from each of three remaining tanks were sampled. In addition, the microbial communities of the surrounding water, tank surfaces and food for 4, 10 and 75 dpf time points were also sampled and characterized.

The zebrafish used in this study were raised under conventional laboratory conditions. This involved a number of husbandry changes. Before 21 dpf, the zebrafish were raised in a nursery tank with uncirculated water that was exchanged manually on a daily basis. Just before sampling at 21 dpf, the fish were transferred to a main facility system where water was continuously recirculated at a fixed rate through a sand and UV filter. The diet also changed over the course of the study: before 6 dpf fish were not fed and subsisted off their yolks alone, after which time fish were feed live *Paramecia* from 6 dpf to 10 dpf, live *Artemia* (brine shrimp) just after fish were sampled at 10–21 dpf, and a standard dry fish food mixture from 21 dpf onward. Between 75 and 380 dpf, the manufacturer of this standard diet changed, but the feeding schedule remained the same. All zebrafish experiments were conducted in conformity with the Public Health Service Policy on Humane Care and Use of Laboratory Animals using standard protocols approved by the Institutional Animal Care and Use Committees of the University of Oregon and the University of North Carolina at Chapel Hill.

The microbial communities sampled in this study were characterized by Illumina sequencing of the 16S rRNA gene. 16S rDNA sequences from the V4 region of the 16S rRNA gene, subsequent 97% similarity operation taxonomic unit (OTU) tables rarefied to 4250 sequences per sample, and taxonomic classifications were taken directly from Stephens *et al.* (2015).

Sloan neutral community model for prokaryotes

To determine the potential importance of neutral processes to community assembly, we assessed the fit of the Sloan Neutral Community Model for Prokaryotes to the distributions of microbial taxa in our data (Sloan *et al.,* 2006). This neutral model predicts the relationship between the frequency with which taxa occur in a set of local communities (in this case individual zebrafish intestinal communities) and their abundance across the wider metacommunity (the intestinal communities of all zebrafish sampled at a given time point). In general, the model predicts that taxa that are abundant in the metacommunity will also be widespread, since they are more likely to disperse by chance and be randomly sampled by an individual host, while rare taxa are more likely to be lost from individual hosts

due to ecological drift. In contrast to many other contemporary neutral models, namely the unified neutral theory of biodiversity (Hubbell, 2001), the neutral model used here does not incorporate the process of speciation. However, while microbial speciation and diversification are no doubt important in generating the diversity of microorganisms in this system at the broad, regional level, our explicit focus is on the assembly of host-associated communities over the course of host development. As such, it is highly unlikely that microbial diversification will occur over that time span to an extent that it would impact diversity among communities at the resolution we observe them (that is, 97% similarity in 16S gene sequences).

The Sloan neutral model is fit to the observed frequency of occurrence of OTUs (that is, the proportion of local communities in which each OTU is detected) and their abundance in the metacommunity (estimated in this case by the mean relative abundance across all local communities) by a single free parameter describing the migration rate, m. This estimated migration rate is the probability that a random loss of an individual in a local community will be replaced by dispersal from the metacommunity, as opposed to reproduction within the local community, and can thus be interpreted as a measure of dispersal limitation. The fitting of this parameter was performed in R using non-linear least-squares fitting and the minpack.lm package (Elzhov *et al.,* 2013; R Core Team, 2015). Binomial proportion 95% confidence intervals around the model predictions were calculated using the Wilson score interval in the HMisc package in R (Brown *et al.,* 2001; Harrell, 2014). We assessed the overall fit of the model to observed data by comparing the sum of squares of residuals, SS_{err}, with the total sum of squares, SS_{total}: model fit $= 1 - SS_{err}/SS_{total}$ (generalized R-squared; Östman *et al.,* 2010). The fit of the neutral model was also compared with the fit of a binomial distribution model to determine whether incorporating drift and dispersal limitations improve the fit of a model beyond just random sampling of the source metacommunity (Sloan *et al.,* 2007). Sampling from a binomial distribution represents the case where local communities are random subsets of the metacommunity in the absence of processes of drift and dispersal limitations. While generalized r-squared is a useful measure for comparing the fit of multiple data sets with a single model, it is a poor choice for comparing the fit of multiple models with a single data set (Spiess and Neumeyer, 2010). Therefore, to compare the fit of the neutral and binomial model, we compared the Akaike information criterion of each model. Computation of the Akaike information criterion was done in R. Calculation of 95% confidence intervals around all fitting statistics was done by bootstrapping with 1000 bootstrap replicates. The mean relative abundance and observed and predicted occurrence frequency of each OTU at each time

point can be found in Supplementary Data 1. Example R code used to fit the model and calculate goodness-of-fit statistics is included as a supplement (Supplementary Code 1).

To analyze deviations from the neutral model predictions, we compared the composition and diversity of neutrally and non-neutrally distributed OTUs. To accomplish this, samples belonging to the same age group were first pooled, and OTUs from this pool were subsequently sorted into three partitions depending on whether they occurred more frequently than ('above' partition), less frequently than ('below' partition) or within ('neutral' partition) the 95% confidence interval of the neutral model predictions. Each partition was then treated as a distinct community sample for further analysis, resulting in 21 total partitions (3 per each of the 7 age groups). To facilitate comparisons among partitions, each partition was rarefied to an equal number of OTUs corresponding to the number of OTUs in the smallest partition, unless otherwise noted.

Diversity and taxonomic analysis
To quantify the variation in phylogenetic composition, we calculated pairwise unweighted UniFrac distances among neutral and non-neutral metacommunity partitions (Lozupone and Knight, 2005). Differences in this distance among groups were assessed by permutational multivariate analysis of variance (MANOVA) using 1000 random permutations, while differences in the degree of variation within groups were assessed by an analysis of variance (ANOVA) of average distance to centroid within groups (multivariate homogeneity of groups dispersions test; Anderson, 2006). Non-metric multidimensional scaling of UniFrac distances was performed to visualize difference among neutral and non-neutral partitions. Calculation of the UniFrac distances was performed in R using the GUniFrac package (Chen, 2012), while permutational MANOVA, multivariate homogeneity of groups dispersion test, and non-metric multidimensional scaling were performed in R using the vegan package (Oksanen *et al.*, 2013).

To identify microbial taxonomic groups that distinguish neutral from non-neutral partitions of the metacommunity, we performed logistic regression with partition type (above, below, or within the neutral prediction) as a predictor and the presence or absence of each taxon as a binary response variable. To determine the significance of this relationship, we compared the deviation of the fitted regression model with that of an empty null model (chi-square test).

An indicator taxa analysis was performed to identify OTUs associated with either fish or tank environment (that is, water, surface and food) samples. Each OTU was assigned an indicator value based on their abundance and occurrence frequency in either intestinal or environmental samples; OTUs found frequently in high abundance in one sample type but not in the other would have a high indicator value for that sample type (Dufrêne and Legendre, 1997). Significance of this association with sample type was determined by comparing the observed value with the values from 1000 random permutations. Calculation of the indicator values and probability for OTUs was performed in R using the labdsv package (Roberts, 2013). Since a full set of environmental data was only available for the 4, 10 and 75 dpf age groups, this analysis was performed only for those time points. The initial results of this analysis can be found in Supplementary Data 1.

Phylogenetic sampling theory
To further examine and compare the phylogenetic structure of neutral and non-neutral partitions of the observed communities, we employed a phylogenetic sampling theory that analytically predicts the phylogenetic diversity in a local community assuming random sampling from the phylogenetic tree of the metacommunity (O'Dwyer *et al.*, 2012). Observed measures of phylogenetic diversity for individual samples can be compared with these predictions to determine the degree to which communities appear random with respect to phylogeny as opposed to over-dispersed or clustered. If the observed phylogenetic diversity is greater than the expected diversity, then we consider the community to be phylogenetically over dispersed, meaning that distantly related taxa were more likely to be sampled than closely related taxa. When the observed phylogenetic diversity is less than the expected diversity, then we consider the community to be phylogenetically clustered, meaning that closely related taxa were more likely to be sampled (Horner-Devine and Bohannan, 2006).

Implementation of the phylogenetic sampling theory was performed in R using methods described in O'Dwyer *et al.* (2012) and the picante package (Kembel *et al.*, 2010). Phylogenetic diversity was defined as the sum of the total phylogenetic branch length for a sample (Faith, 1992). Random sampling of the regional phylogenetic tree was modeled by binomial sampling. A strength of this approach is that it can be used to compare samples of unequal sizes. As such, this analysis was applied to un-rarefied data, and differences between observed and expected phylogenetic diversity were compared by calculating and comparing standardized deviations, or z-scores, for each partition.

Results

Relative importance of neutral processes decreases over host development
Overall, the frequency with which microbial taxa occurred in individual communities was well described by the neutral model (Figure 1;
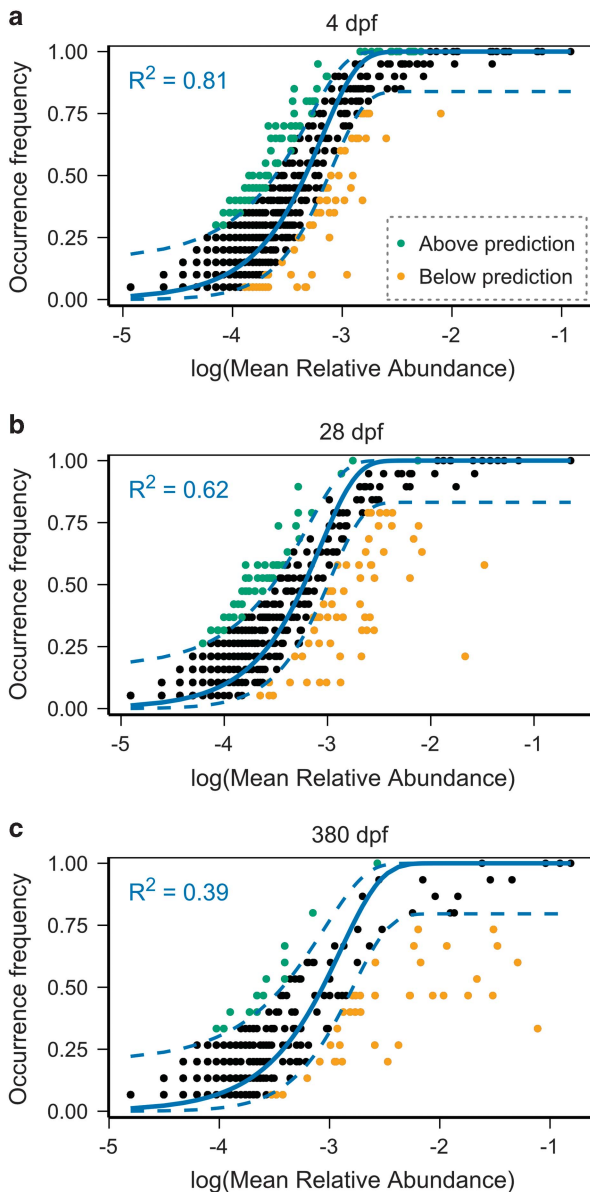
Figure 1 Fit of the neutral model. The predicted occurrence frequencies for 4 (**a**), 28 (**b**) and 380 (**c**) dpf zebrafish communities representing larval, juvenile, and adult developmental stages, respectively. OTUs that occur more frequently than predicted by the model are shown in green while those that occur less frequently than predicted are shown in orange. Dashed lines represent 95% confidence intervals around the model prediction (blue line).

Supplementary Figure 1). However, the fit of the model varied over host development, and was negatively correlated with host age (Spearman's $rho = -0.93$, $P = 0.007$; Figure 2a). In all cases, the neutral model outperformed a binomial distribution model, suggesting that the processes of passive dispersal and ecological drift have an impact above and beyond just random sampling of the source community (Figure 2b). Overall, estimated migration rates tended to be higher in younger than older fish, suggesting that communities become increasingly dispersal limited with age (Spearman's $rho = -0.86$, $P = 0.02$; Figure 2c).

Deviations from neutral predictions are ecologically distinct

For any age group of fish, there were a number of microbial taxa that occurred more or less frequently than predicted by the model given their overall abundance in the metacommunity (points above, in green, or below, in orange, the line in Figure 1). We would expect points that differ significantly from the neutral prediction to be indicative of taxa that are actively being selected for or against by the host. Specifically, points above the prediction represent taxa that are found more frequently than expected, suggesting that they are actively being maintained and selected for by the host, while points found below the prediction represent taxa found less frequently than expected, suggesting that they are either selected against by the host or are especially dispersal limited. We expect that these selective processes should be reflected in the taxonomic and phylogenetic composition of taxa that deviate from the neutral prediction. We tested this hypothesis explicitly and examined how these differences may be informative of the overall ecology of the intestinal community.

Taxa found above, below or within the prediction of the neutral model formed phylogenetically distinct partitions of the total metacommunity. For each age group, we separated the metacommunity into three partitions comprises those OTUs found above, below or not significantly different from the neutral prediction and calculated the phylogenetic dissimilarity among partitions (Supplementary Figure 2). We found that partitions clustered strongly based on whether and how they deviated from the neutral prediction (that is, above, below or within the neutral prediction) across host age (permutational MANOVA $r^2 = 0.19$, $P < 0.001$; Figure 3a). Thus, the phylogenetic composition of the sub-groups that diverge from neutral patterns remains relatively similar across host development, despite the composition of communities as a whole changing (Stephens *et al.*, 2015). Across age groups, non-neutral partitions of the metacommunity were also much more homogeneous than the neutral partitions, (ANOVA, $P < 0.01$; illustrated by the spread of points in Figure 3a). A possible consequence of the heterogeneity is that we identified very few taxonomic groups that strongly distinguished neutral partitions. Partitions above the neutral prediction were most strongly distinguished by the presence of Fusobacteria ($P < 0.001$) and γ-Proteobacteria ($P = 0.022$), in particular the families *Enterobacteriaceae* ($P = 0.003$) and *Aeromonadaceae* ($P < 0.001$), while partitions below the neutral prediction were distinguished by the presence of Actinobacteria ($P = 0.004$), Bacilli ($P = 0.004$) and Clostridia ($P = 0.031$) and the genera *Lactobacillus* ($P = 0.004$), *Staphylococcus* ($P = 0.037$) and *Stenotrophomonas* ($P = 0.012$).

The taxa comprising neutral and non-neutral partitions of the metacommunity also tended to be
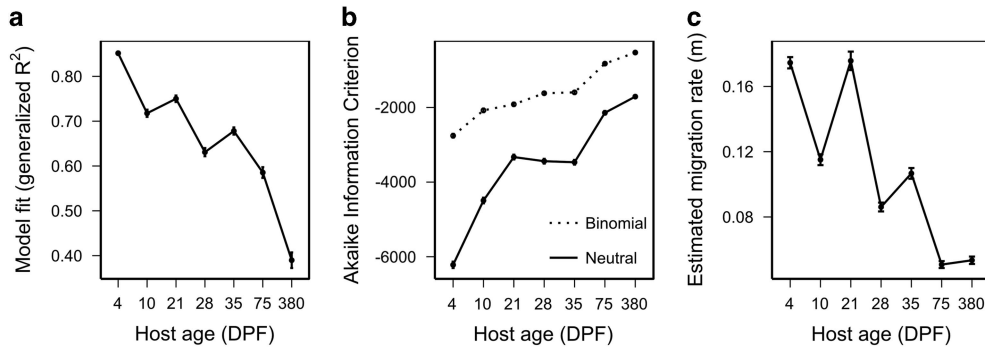
Figure 2 Neutral model fit decreases over host development. The goodness-of-fit of the Sloan neutral (**a**), comparison of the maximum likelihood fit of the neutral and binomial models (**b**), and the estimated migration rate (**c**) for zebrafish-associated communities.
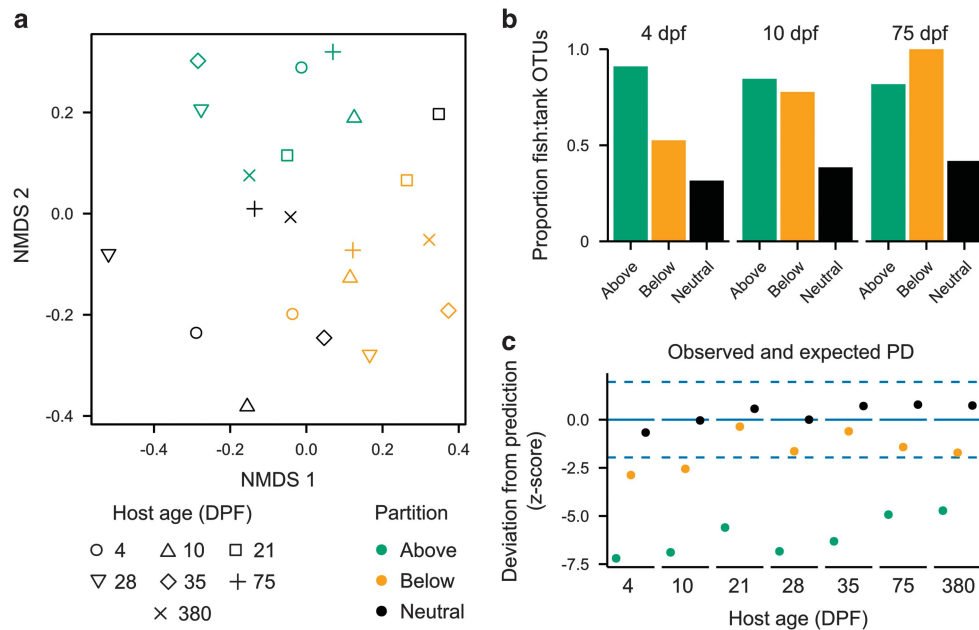

Figure 3 Neutral and non-neutral partitions of the metacommunity are compositionally and phylogenetically distinct. For each age group, communities were pooled and OTUs were then divided into separate partitions based on whether they were consistent with (in black) or deviated above (in green) or below (in orange) the neutral prediction (color coding is consistent for all panels). (**a**) Non-metric multidimensional scaling ordination based on UniFrac distances. (**b**) The proportion of fish associated with tank-associated OTUs in each partition following an indicator taxa analysis. Results are shown for 4, 10 and 75 dpf fish only as environmental samples were not available for the other time points. (**c**) The standardized difference, in units of standard deviations (z-score), between observed and expected phylogenetic diversity assuming random sampling for each partition. Solid blue lines represent the expected phylogenetic diversity for each age group while dashed lines represent 95% confidence intervals. The distance points either above or below the line represent the degree to which those partitions are phylogenetically over-dispersed or clustered, respectively.

associated with different environments. We first performed an indicator taxa analysis to identify OTUs that were significantly associated with either intestinal or environmental (tank water, surfaces or food) samples in our full data set (Supplementary Data 1). We then compared the proportion of OTUs significantly associated with fish to those significantly associated with the tank environment in each partition and found that this proportion was much higher above and below than it was within the model's prediction (Figure 3b). This pattern was consistent across host development. In other words, non-neutral partitions of the metacommunities were more likely to be comprises microbial taxa that were associated with zebrafish, while taxa largely

associated with the tank environment were more likely to be neutrally distributed across fish intestinal communities.

Finally, we found that non-neutral partitions were phylogenetically clustered with respect to the metacommunity as a whole. For this, we calculated the phylogenetic diversity of each partition and compared these observed values with the phylogenetic diversity expected if taxa were sampled randomly with respect to phylogeny (O'Dwyer *et al.*, 2012). As expected, the observed phylogenetic diversity for neutrally distributed groups was consistent with the random prediction across host development. In contrast, the phylogenetic diversity of the partitions that deviated above the neutral

prediction was consistently less than expected, indicative of phylogenetic clustering, while those partitions below the neutral prediction were phylogenetically clustered in larval fish but became less so as the hosts aged (Figure 3c). Assuming that more closely related microorganisms are on average more ecologically or functionally similar than more distantly related ones (Burns and Strauss, 2011), this result reinforces the conclusion that taxa which deviate from the neutral prediction, particularly those more widespread than expected, are portions of the microbiota that are more likely to be actively selected (for or against) by the host.

Discussion

The neutral model used in this study was able to predict the microbial distributions across communities by incorporating only the effects of random dispersal and demographic processes. Even in adult zebrafish, where the fit of the model was relatively poor compared with the younger fish, the distribution of OTUs in the metacommunity still followed the same basic trend of abundant taxa being widespread, consistent with neutral theory (Supplementary Figure 1). These findings illustrate an important point which is often ignored: not all of the variation among host-associated microbial communities need to be the result of differences among hosts or associated microorganisms. On the contrary, neutral processes of drift and dispersal are powerful enough on their own to generate a large amount of diversity both within and among hosts, and these processes can explain a significant portion of the structure of communities observed in this study. This is not to say that neutral processes are the only important factors, but they can act alongside and may even swamp the effects of non-neutral forces. These results also indicate that in addition to local ecological factors (for example, the environment of a zebrafish intestine and differential competitive fitness among microorganisms), host-associated microbial communities are heavily influenced by ecological dynamics occurring outside an individual host at a broader scale.

While the model's general success highlights the potential importance of neutral processes, it is also useful as a null model to identify the conditions under which the model's predictions fail, which can lead to a better understanding of specific additional factors structuring these communities. Within each age group there were a number of microbial taxa whose distributions deviated from neutral predictions. These taxa were not randomly distributed throughout the total metacommunity, implying that they are distinct in ways that are ecologically informative. The taxa whose deviations from the neutral pattern led them to be more widespread than expected are likely taxa that are specifically adapted to, and selected by, the host environment. This is

supported by the dominance of intestinal associated OTUs within non-neutral partitions and is consistent with these partitions being phylogenetically clustered, suggesting the host habitat selects microbial taxa based on a specific set of phylogenetically conserved traits (Figures 3b and c). Likewise, abundant taxa that occurred less frequently than expected may be characteristic of 'invasive' microorganisms and potential pathogens that are selected against by the zebrafish hosts overall, but are nevertheless able to proliferate in a few susceptible individuals. If true, this would explain why these taxa were more likely to be significantly associated with fish despite having distributions suggesting that they are being selected against. This is in contrast to neutrally distributed taxa, which were more likely to be associated with exogenous environmental tank samples as well as exhibiting greater phylogenetic variation and diversity across age groups. Such patterns suggest that these neutrally distributed taxa are less likely to be specifically adapted to the host and their presence in any given community is largely the result of their abundance in the surrounding metacommunity and source pool. It is worth emphasizing, however, that this does not mean that these taxa are functionally unimportant or even that they are not interacting intimately with their hosts. Rather the host environment is not differentially selecting them, and consequently their distributions are the result of neutral dispersal and drift.

As hosts aged the ability of the neutral model to predict the distribution of associated microbial taxa decreased, indicating that neutral processes become relatively less important as the host ages (Figure 2a). We suspect this pattern is largely the result of the development of the host. The 4-dpf time point, for example, occurs shortly after the intestinal tract of the zebrafish is fully opened and colonized by bacteria but before the fish develops an active adaptive immune response (between 21 and 35 dpf) and reaches sexual maturity (between 35 and 75 dpf). It is also probable that husbandry changes over the course of the experiment had an impact on this pattern. The strongest evidence of neutral dynamics occurred before the fish began eating (4 dpf) and while the fish were housed in nursery tanks unconnected to the main facility water system (from 4 to 21 dpf). At 21 dpf fish were not only moved from nursery tanks to the main facility system tanks, but also had their diet significantly changed (see Materials and methods). These changes to the host's physiology and environment gradually accumulate over development, and likely differentiate the ability of microbial taxa to establish and thrive within them. While it is difficult to disentangle whether the observed patterns are driven mostly by developmental or husbandry changes, we note that the decrease in the fit of the model continues between 28 and 380 dpf, during which time the zebrafish continue to develop (see above), but their housing

conditions remain unchanged. The decrease in the fit of the model was accompanied by a decrease in the estimated migration rate, which suggests that these changes in the hosts may also decrease the ability of microorganisms to disperse into and among hosts. This is further supported by our previous observation that communities associated with 4 dpf and 10 dpf fish were more similar to environmental communities than those associated with the older 75 dpf fish (Stephens *et al.*, 2015), as well as the observation that within-host diversity decreased over the same time span, which is a predicted consequence of decreased dispersal rates (Cadotte, 2006).

The patterns of neutral assembly in the zebrafish intestinal microbiota described here are consistent with and provide possible explanations for observed patterns in human-associated microbial communities. In general, large-scale studies of human intestinal microbiota have revealed high variation in community composition, both across individuals and within individuals over time (Costello *et al.*, 2009; The Human Microbiome Project Consortium, 2012; Yatsunenko *et al.*, 2012). Often, this variation is not easily explained by measured host factors, suggesting that much of it might be explained by neutral assembly processes. Our observation that communities associated with young fish were the most neutrally assembled could also explain observations that variation is greater among communities associated with young humans (Kurokawa *et al.*, 2007; Palmer *et al.*, 2007; Yatsuenko *et al.*, 2012), the general variable nature of infant microbiota over time (Koenig *et al.*, 2011), and observations that the infant microbiota is heavily influenced by exogenous microbial communities, specifically those of the mother (Dominguez-Bello *et al.*, 2010; Funkhouser and Bordenstein, 2013).

These results may also be extended more broadly beyond animal associated communities. Using a similar conceptual framework to the one used here (multiple local communities sampling from a broader metacommunity), Jabot *et al.* (2008) found that the distribution of young saplings in a tropical forest was better fit by a neutral model than that of older trees. Likewise, Dini-Andreote *et al.* (2015) found that the relative importance of stochastic processes decreased over the succession of microbial salt marsh communities. The consistency of this pattern across these different systems may be indicative of more fundamental ecological processes. Even if the structure of communities is ultimately determined by differential selection, many communities in nature may exist in transitory, non-equilibrium states such that these selective processes do not have the opportunity to fully play out and manifest their effects (Manceau *et al.*, 2015).

Despite extensive research on microbial communities associated with animal hosts, it has remained difficult to explain the high levels of variation in these systems. We addressed this question by adopting a framework that recognizes that these questions are ecological in nature and can be addressed through the use of established ecological theory. Because this framework is grounded in ecological theory, it provides hypotheses that can be tested in an explicit manner. For example, it would be interesting to see whether the neutral and non-neutral partitions of the metacommunity are physically delineated in the intestine, wherein we might expect neutral taxa to be found in the lumen while deviations from neutral patterns might be more intimately attached to the epithelial layer where interactions between host and bacterial cells may be more likely to occur. Additionally, it is possible that non-neutral behavior in these communities is driven by differences among taxa in dispersal rates, in which case partitioning the communities on the basis of differences in immigration rates would likely improve neutral predictions (Janzen *et al.*, 2015). It might also be fruitful to compare the neutral patterns seen in healthy hosts with those seen in diseased, infected or diet-altered individuals, which we predict will be characterized by deviations from neutral predictions. Similarly, we predict that infectious or pathogenic microorganisms could be identified by their deviations from neutral predictions, occurring much less frequently than expected given their relative abundance in a metacommunity.

Ultimately, one of the goals of studying host-associated communities is to better understand how they might be altered or manipulated to improve health and prevent disease. At their core, our results demonstrate a relationship between the abundance of a microorganism and how widespread that microorganism is in a population of hosts. In other words, the distribution of microorganisms in these systems is the result of both local factors specific to individual hosts and those processes occurring at a broader metacommunity scale linking multiple hosts. Attempts to manipulate a host's microbiota must therefore focus on understanding not only the communities within an individual host, but also the communities of microorganisms present around them.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

models used in this study. We also thank Rose Sockol (University of Oregon) for help in generating and maintaining zebrafish populations.

References

Anderson MJ. (2006). Distance-based tests for homogeneity of multivariate dispersions. *Biometrics* **62**: 245–253.

Bates JM, Mittge E, Kuhlman J, Baden KN, Cheesman SE, Guillemin K. (2006). Distinct signals from the microbiota promote different aspects of zebrafish gut differentiation. *Dev Biol* **297**: 374–386.

Benson AK, Kelly SA, Legge R, Ma F, Low SJ, Kim J *et al.* (2010). Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proc Natl Acad Sci USA* **107**: 18933–18938.

Brown LD, Cai TT, DasGupta A. (2001). Interval estimation for a binomial proportion. *Stat Sci* **16**: 101–133.

Burns JH, Strauss SY. (2011). More closely related species are more ecologically similar in an experimental test. *Proc Natl Acad Sci USA* **108**: 5302–5307.

Cadotte MW. (2006). Dispersal and species diversity: a meta-analysis. *Am Nat* **167**: 913–924.

Caswell H. (1976). Community structure: a neutral model analysis. *Ecol Monogr* **46**: 327–354.

Chase JM, Myers JA. (2011). Disentangling the importance of ecological niches from stochastic processes across scales. *Philos Trans R Soc Lond B Biol Sci* **366**: 2351–2363.

Chen J. (2012). GUniFrac: Generalized UniFrac distances. R package version 1.0 http://CRAN.R-project.org/package=GUniFrac.

Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R. (2009). Bacterial community variation in human body habitats across space and time. *Science* **326**: 1694–1697.

Costello EK, Stagaman K, Dethlefsen L, Bohannan BJM, Relman DA. (2012). The application of ecological theory toward an understanding of the human microbiome. *Science* **336**: 1255–1262.

de Muinck EJ, Stenseth NC, Sachse D, vander Roost J, Rønningen KS, Rudi K *et al.* (2013). Context-dependent competition in a model gut bacterial community. *PLoS One* **8**: 1–16.

Dethlefsen L, Eckburg P, Bik E, Relman D. (2006). Assembly of the human intestinal microbiota. *Trends Ecol Evol* **21**: 517–523.

Dini-Andreote F, Stegen JC, van Elsas JD, Salles JF. (2015). Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession. *Proc Natl Acad Sci USA* **112**: e1326–e1332.

Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, Fierer N *et al.* (2010). Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc Natl Acad Sci USA* **107**: 11971–11975.

Dufrêne M, Legendre P. (1997). Species assemblages and indicator species: The need for a flexible asymmetrical approach. *Ecol Monogr* **67**: 345–366.

Elzhov TV, Mullen KM, Spiess AN, Bolker B. (2013). minpack.lm: R interface to the Levenberg-Marquardt nonlinear least-squares algorithm found in MINPACK, plus support for bounds. *R package version 1.1-8.* http://CRAN.R-project.org/package=minpack.lm.

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

Fraune S, Bosch TCG. (2010). Why bacteria matter in animal development and evolution. *BioEssays* **32**: 571–580.

Funkhouser LJ, Bordenstein SR. (2013). Mom knows best: the universality of maternal microbial transmission. *PLoS Biol* **11**: e1001631.

Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R *et al.* (2014). Human genetics shape the gut microbiome. *Cell* **159**: 789–799.

Harrell FE. (2014). Hmisc: Harrell Miscellaneous. R package version 3.14-5 http://CRAN.R-project.org/package=Hmisc.

Horner-Devine MC, Bohannan BJM. (2006). Phylogenetic clustering and overdispersion in bacterial communities. *Ecology* **87**: S100–S108.

Hubbell SP. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University PressPrinceton, New Jersey.

Jabot F, Etienne RS, Chave J. (2008). Reconciling neutral community models and environmental filtering: theory and an empirical test. *Oikos* **117**: 1308–1320.

Janzen T, Haegeman B, Etienne RS. (2015). A sampling formula for ecological communities with multiple dispersal syndromes. *J Theor Biol* **374**: 94–106.

Jeraldo P, Sipos M, Chia N, Brulc JM, Dhillon AS, Konkel ME *et al.* (2012). Quantification of the relative roles of niche and neutral processes in structuring gastrointestinal microbiomes. *Proc Natl Acad Sci USA* **109**: 9692–9698.

Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD *et al.* (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* **26**: 1463–1464.

Kerr B, Riley MA, Feldman MW, Bohannan BJM. (2002). Local dispersal promotes biodiversity in a real-life game of rock-paper-scissors. *Nature* **418**: 171–174.

Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, Knight R *et al.* (2011). Succession of microbial consortia in the developing infant gut microbiome. *Proc Natl Acad Sci USA* **108**: 4578–4585.

Kurokawa K, Itoh T, Kuwahara T, Oshima K, Toh H, Toyoda A *et al.* (2007). Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res* **14**: 169–181.

Lam S, Chua H, Gong Z, Lam T, Sin Y. (2004). Development and maturation of the immune system in zebrafish, Danio rerio: a gene expression profiling, in situ hybridization and immunological study. *Dev Comp Immunol* **28**: 9–28.

Lankau EW, Hong PY, MacKie RI. (2012). Ecological drift and local exposures drive enteric bacterial community differences within species of Galápagos iguanas. *Mol Ecol* **21**: 1779–1788.

Leibold MA, Holyoak M, Mouquet N, Amarasekare P, Chase JM, Hoopes MF *et al.* (2004). The metacommunity concept: a framework for multi-scale community ecology. *Ecol Lett* **7**: 601–613.

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

Lindström ES, Östman Ö. (2011). The importance of dispersal for bacterial community composition and functioning. *PLoS One* **6**: e0025883.

Lozupone C, Knight R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* **71**: 8228–8235.

Manceau M, Lambert A, Morlon H. (2015). Phylogenies support out-of-equilibrium models of biodiversity. *Ecol Lett* **18**: 347–356.

Milligan-Myhre K, Charette JR, Phennicie RT, Stephens WZ, Rawls JF, Guillemin K *et al. et al.* (2011). *Study of Host-Microbe Interactions in Zebrafish*. 3rd edn Elsevier Inc.

O'Dwyer JP, Kembel SW, Green JL. (2012). Phylogenetic diversity theory sheds light on the structure of microbial communities. *PLoS Comput Biol* **8**: e1002832.

Ofiteru ID, Lunn M, Curtis TP, Wells GF, Criddle CS, Francis CA *et al.* (2010). Combined niche and neutral effects in a microbial wastewater treatment community. *Proc Natl Acad Sci USA* **107**: 15345–15350.

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB *et al.* (2013). vegan: Community Ecology Package. R package version 2.0-10 http://CRAN.R-project.org/package=vegan.

Östman Ö, Drakare S, Kritzberg ES, Langenheder S, Logue JB, Lindström ES. (2010). Regional invariance among microbial communities. *Ecol Lett* **13**: 118–127.

Palmer C, Bik EM, DiGiulio DB, Relman DA, Brown PO. (2007). Development of the human infant intestinal microbiota. *PLoS Biol* **5**: 1556–1573.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria. Available from http://www.R-project.org/.

Rawls JF, Mahowald MA, Ley RE, Gordon JI. (2006). Reciprocal Gut Microbiota Transplants from Zebrafish and Mice to Germ-free Recipients Reveal Host Habitat Selection. *Cell* **127**: 423–433.

Roberts DW. (2013). labdsv: Ordination and Multivariate Analysis for Ecology. R package version 1.6-1 http://CRAN.R-project.org/package=labdsv.

Robinson CJ, Bohannan BJM, Young VB. (2010). From structure to function: the ecology of host-associated microbial communities. *Microbiol Mol Biol Rev* **74**: 453–476.

Roeselers G, Mittge EK, Stephens WZ, Parichy DM, Cavanaugh CM, Guillemin K *et al.* (2011). Evidence for a core gut microbiota in the zebrafish. *ISME J* **5**: 1595–1608.

Rosindell J, Hubbell SP, Etienne RS. (2011). The unified neutral theory of biodiversity and biogeography at age ten. *Trends Ecol Evol* **26**: 340–348.

Sloan WT, Lunn M, Woodcock S, Head IM, Nee S, Curtis TP. (2006). Quantifying the roles of immigration and chance in shaping prokaryote community structure. *Environ Microbiol* **8**: 732–740.

Sloan WT, Woodcock S, Lunn M, Head IM, Curtis TP. (2007). Modeling taxa-abundance distributions in microbial communities using environmental sequence data. *Microb Ecol* **53**: 443–455.

Sommer F, Bäckhed F. (2013). The gut microbiota–masters of host development and physiology. *Nat Rev Microbiol* **11**: 227–238.

Spiess AN, Neumeyer N. (2010). An evaluation of R2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: a Monte Carlo approach. *BMC Pharmacol* **10**: 6.

Stephens WZ, Burns AR, Stagaman K, Wong S, Rawls JF, Guillemin K *et al.* (2015). The composition of the zebrafish intestinal microbial community varies across development. *ISME J* (in press).

The Human Microbiome Project Consortium (2012). Structure, function and diversity of the healthy human microbiome. *Nature* **486**: 207–214.

Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**: 1027–1031.

Venkataraman A, Bassis CM, Beck JM, Young VB, Curtis JL, Huffnagle GB, Schmidt TM. (2015). Application of a neutral community model to assess structuring of the human lung microbiome. *mBio* **6**: e02284–14.

Woodcock S, Van Der Gast CJ, Bell T, Lunn M, Curtis TP, Head IM *et al.* (2007). Neutral assembly of bacterial communities. *FEMS Microbiol Ecol* **62**: 171–180.

Yan Q, van der Gast CJ, Yu Y. (2012). Bacterial community assembly and turnover within the intestines of developing zebrafish. *PLoS One* **7**: e30603.

Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M *et al.* (2012). Human gut microbiome viewed across age and geography. *Nature* **486**: 222–227.

Supplementary Information accompanies this paper on The ISME Journal website (http://www.nature.com/ismej)