



OPEN

Robust bacterial co-occurrence community structures are independent of *r*- and *K*-selection history

Jakob Peder Pettersen¹, Madeleine S. Gundersen¹ & Eivind Almaas^{1,2}✉

Selection for bacteria which are *K*-strategists instead of *r*-strategists has been shown to improve fish health and survival in aquaculture. We considered an experiment where microcosms were inoculated with natural seawater and the selection regime was switched from *K*-selection (by continuous feeding) to *r*-selection (by pulse feeding) and vice versa. We found the networks of significant co-occurrences to contain clusters of taxonomically related bacteria having positive associations. Comparing this with the time dynamics, we found that the clusters most likely were results of similar niche preferences of the involved bacteria. In particular, the distinction between *r*- or *K*-strategists was evident. Each selection regime seemed to give rise to a specific pattern, to which the community converges regardless of its prehistory. Furthermore, the results proved robust to parameter choices in the analysis, such as the filtering threshold, level of random noise, replacing absolute abundances with relative abundances, and the choice of similarity measure. Even though our data and approaches cannot directly predict ecological interactions, our approach provides insights on how the selection regime affects the composition of the microbial community, providing a basis for aquaculture experiments targeted at eliminating opportunistic fish pathogens.

In aquaculture, the fish is in close contact with its environmental microbiome¹. Fish larvae are at an especially vulnerable life stage, with high death rates causing economic problems in the aquaculture industry^{2,3}. Research during the last decades has uncovered that the bacterial composition of the larval environment affects their survival, and both detrimental and favourable host-microbe interactions have been identified. This interplay strongly suggests that one may manipulate the larval environmental microbiome to improve their health and survival^{4,5}.

A viable approach for microbiome control is to select against opportunistic pathogens and select for favourable bacteria. This approach is based on the concept of *r*- and *K*-strategists, introduced in microbial ecology by Andrews and Harris⁶. Most opportunistic bacteria are *r*-strategists, meaning that they grow rapidly when resources are in surplus. If the opportunistic strain is pathogenic, such environments facilitate its proliferation and may subsequently lead to fish disease. However, *r*-strategists compete poorly when the environment is resource-limited. In such a competitive environment, the slow-growing *K*-strategists will quickly dominate due to their high resource-acquiring affinities and high yields^{1,7}. Thus, resource availability is a crucial variable to manage in aquaculture. Stable resource availability promotes *K*-selection, whereas fluctuating availability promotes *r*-selection⁷.

Since it is unclear, due to lack of experimental evidence, whether selecting for *K*-strategists will make a recurring set of bacteria co-occur or whether competition results in co-exclusion among the *K*-strategists, we wanted to investigate this problem using co-occurrence network analysis. Furthermore, we also wanted to study the extent to which *r*-strategists co-occur with *K*-strategists. Earlier studies have suggested that using similarity measures for network inference could determine bacterial niches, describe a microbial community's response to environmental disturbances, predict ecological keystone organisms, and explain changes in a microbial community over time^{8–12}. We hypothesized that such a tool could partition bacteria based on their growth preferences and be useful to characterise and identify which bacteria are *r*-strategists and which are *K*-strategists. When a microbial community is subjected to external disturbances, it may change composition permanently, it may be

¹Department of Biotechnology and Food Science, NTNU - Norwegian University of Science and Technology, Trondheim, Norway. ²Department of Public Health and General Practice, K.G. Jebsen Center for Genetic Epidemiology, NTNU - Norwegian University of Science and Technology, Trondheim, Norway. ✉email: eivind.almaas@ntnu.no

resistant (insensitive to the disturbance), or it may “remember” its original state and be resilient (return to its original state after initially changing)¹³. We were interested in investigating whether *r*- and *K*-selection will give the microbial communities memory, and whether the selection regime would provide resistance or resilience against changing external factors.

Given our research questions, we found the dataset from Gundersen *et al.*¹⁴ to be particularly useful. This is a 2×2 factorial crossover microcosm experiment that tested varying feeding regime and resource availability (high/low). Briefly, half of the microcosms were pulse-fed resources which promotes the growth of *r*-strategists, whereas the other half received a steady, continuous supply of nutrients promoting *K*-strategists. We hereafter refer to these selection regimes as *r*- and *K*-selection. The bacterial communities were sampled and characterised through 16S ribosomal RNA gene sequencing (16S-RNA)^{9,15,16} at 18 time-points over a 50 day period.

What made the Gundersen dataset¹⁴ particularly suited for our analysis, was that approximately halfway into the experiment (i.e. between day 28 and 29), the *r*- and *K*-selection regimes were switched such that each microcosm was subjected to both selection regimes. While the original paper on the dataset focused on studying the effect of the selection regimes and resource availability on ecological community assembly, our work uses network analysis to gain a more detailed understanding of the community dynamics than solely comparing samples can provide.

Results

To investigate the bacterial community structure and dynamics in *r*- and *K*-selected communities, we assessed the co-occurrence patterns of 1,537 operational taxonomic units¹⁷ (OTUs) observed in the microcosm experiment. This 16S-rRNA gene dataset consisted of 202 samples from 12 microcosms cultivated over 50 days. Note that, 6 of the microcosm were *r*-selected, and 6 were *K*-selected. Between day 28 and 29, the *r*/*K*-selection regime was switched such that *r*-selected communities now were *K*-selected (the RK-group) and vice versa (the KR group). Furthermore, the microcosms varied by the amount of resources supplied, high (H) and low (L). However, exploratory analysis of the dataset did not indicate any relevant effect of the resource supply, and hereafter we will only focus on the *r*- and *K*-selection regimes.

Similarity measurements and network inference. We assessed the co-occurrence patterns between the OTUs using two similarity measurements and varying levels of random noise, OTU filtering, and type of abundance (relative/absolute). In contrast to many other 16S-rRNA microbiome datasets, we estimated the bacterial community's absolute abundances using flow cytometry.

Here, we present the results for the Spearman correlation measure with a low level of random noise, low OTU filtering threshold and absolute abundances (see the Methods' section for more details). We decided to focus on the rank-based Spearman correlation because it is widely applied for detecting associations^{16,18,19}. We will later discuss the robustness of these results by contrasting and comparing with other similarity measures and parameter choices.

From an ecological perspective, an interaction between two microbes is an effect which one microorganism has on another. This includes cross-feeding, biofilm formation, and parasitism^{9,20,21}. However, in further discussion, unless stated otherwise, we will use the term interaction in a network-theoretic perspective, where we apply a “guilt by association” heuristic. This means that, we define two OTUs to have a positive interaction if they co-occur in the same samples to a larger degree than expected by random chance. Conversely, we define two OTUs to have a negative interaction if they co-occur more rarely than expected by random chance^{16,22,23}. Even if there cannot be any direct ecological interactions between the bacteria in different microcosms, the network concept of interactions still enables us to infer associations across samples collected from different microcosms.

We wanted to create a network of the pairwise associations between the OTUs and thus had to determine which edges to include. Selecting a hard threshold for the *q*-value for an interaction to be statistically significant (for instance at $q \leq 0.05$), is not an easy choice²⁴. We, therefore, illustrate the number of significant interactions over a range of threshold *q*-values up to 0.05 (Fig. 1). From this figure, we see that there is no obvious cutoff.

There were in total 3250 interactions having $q \leq 0.05$, of which 1679 were $0.05 \geq q \geq 10^{-4}$ and 639 with $q < 10^{-10}$. Therefore, we determined 500 edges to be a reasonable balance between selecting high-significance edges and a network with lower average node connections. With this setting, the effective *q*-value threshold became $5.0 \cdot 10^{-13}$. The resulting network modules were labelled using the walktrap algorithm with 20 steps²⁵ (Fig. 2).

Phylogenetic clustering within the modules. The co-occurrence network analysis clustered the OTUs into four distinct modules (Fig. 2). Also, two OTUs were not assigned any module (shown as colourless nodes inside module 3) as they were connected to the rest of the network with negative interactions only. Module 3 and 4 stood out as the most interesting modules for two reasons. First, they had the largest number of nodes. Second, there were negative links between the modules, suggesting mutual exclusion between modules. For these two main modules we observed a large number of positive links within the modules, but negative edges between them. Therefore we wondered whether the modules were phylogenetically clustered. Indeed, we observed a clear pattern between OTU module membership and phylogenetic classification (Fig. 3 and Supplementary Table S1). Module 3 consisted primarily of *Alphaproteobacteria* (including *Rhodobacteraceae*) and *Flavobacteria*, whereas most OTUs in module 4 were *Gammaproteobacteria* (including *Colwellia* and *Vibrio*). Hence, we see that within each module, the OTUs were most often phylogenetically related.

Temporal trajectories of the microbial communities. After having observed the network modules, we were interested in understanding the co-occurrence structures and how it influenced community dynamics.

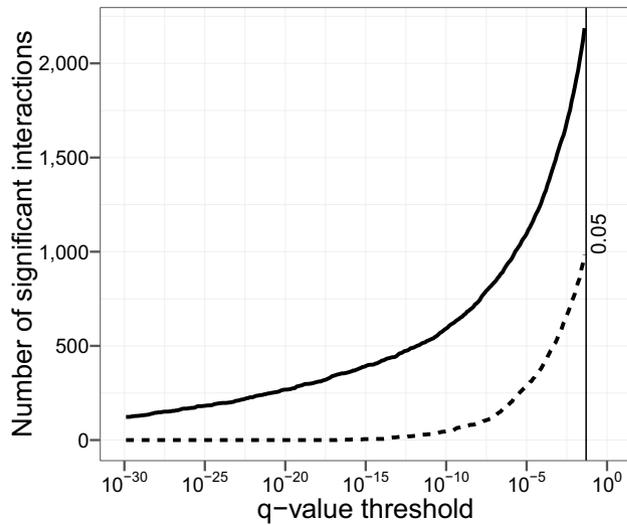


Figure 1. The cumulative number of significant interactions as a function of the critical q -value threshold considered. The solid line signifies positive interactions detected, while the dashed line represents the number of negative interactions.

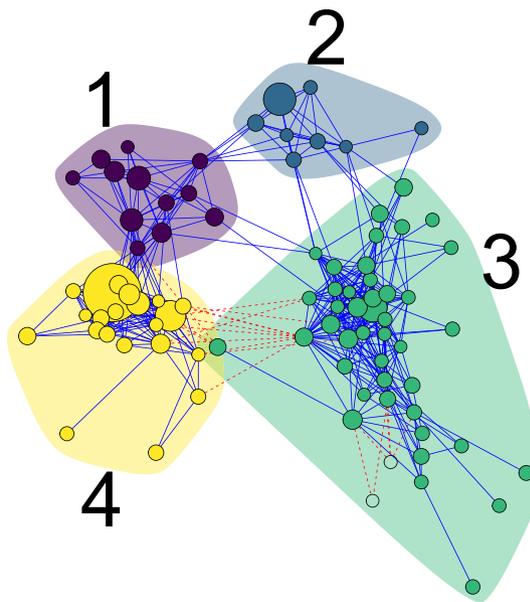


Figure 2. Module-labelled network of the 500 most significant interactions in the r/K -selection-switch dataset. Each of the 86 nodes is an OTU, while each edge corresponds to a statistically significant association between the OTUs. Blue solid edges indicate positive interactions, whereas red dashed edges indicate negative interactions. The node-sizes are scaled logarithmically according to overall mean abundance.

To investigate the community dynamics within the microbial communities, we plotted the Bray-Curtis PCoA-ordinations of the samples and observed the successional trajectories of each microcosm (Fig. 4).

From this time trajectory plot, we compared the panels diagonally and observed that microcosms undergoing K -selection converged towards the upper-left area in the plot, whereas microcosms under r -selection converged the middle-right area. This effect seemed independent of the experimental period and of pre-existing experimental conditions. A PERMANOVA analysis showed that the current selection regime was most important for the community composition ($R^2 = 0.344$ and $p < 10^{-6}$) compared to the minor effect of the overall selection group ($R^2 = 0.084$ and $p = 0.017$), see Methods for details. Consequently, in this respect the communities did not seem to have any memory-effect that gave rise to resistance against changes in composition. As the r/K -selection regimes resulted in clustered communities, we aimed at investigating how the network arose from these dynamics.

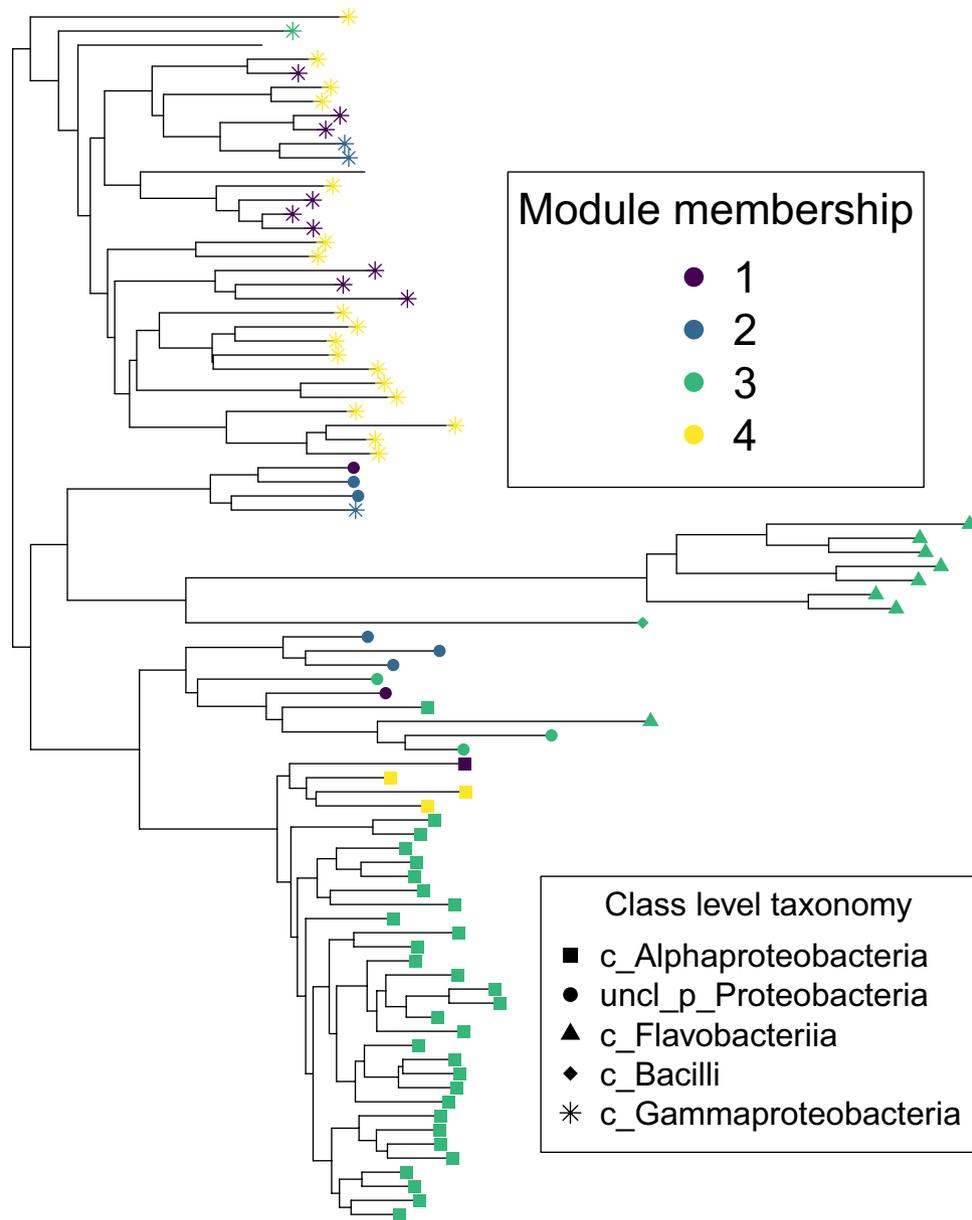


Figure 3. The phylogenetic tree of the 86 OTUs from Fig. 2. together with the class level taxonomical assignment. Point colour indicates module membership, whereas the shape indicates class level taxonomical assignment. Notice that there are some inconsistencies between the phylogenetic tree and the assigned taxonomy.

r/K-strategist network patterns. We further investigated what influenced the dynamics of the community, and conversely the dynamics' contributions to the overall network network in Fig. 2. For this, we visualised the rank-based z-scores of the OTU abundances (see Methods for details) for selected days during the experiment for high resource supply (Fig. 5). For low resource supply, the results were very similar and is thus not discussed any further (see Supplementary Fig. S1 for further details).

There were some obvious patterns that were apparent when investigating the temporal networks, especially with regards to module 3 and 4. The abundance of the OTUs in module 3 increased during *K*-selection, while the ones in module 4 had the opposite trend and had high abundances during *r*-selection. Hence, within each module, the OTUs had coordinated abundance patterns leading to positive inferred interactions. On the other hand, between module 3 and 4, the abundance patterns were anti-coordinated such that we obtained negative interactions.

We expected the dataset to display two time periods of instability: The first at the start of the experiment when the microbial community would adapt to lab-culture conditions, and the second disturbance instability after switching the selection regime, after day 28. During these unsteady periods, we expected more instability and less coordination between OTUs belonging to the same module. This in turn, would contribute to negative

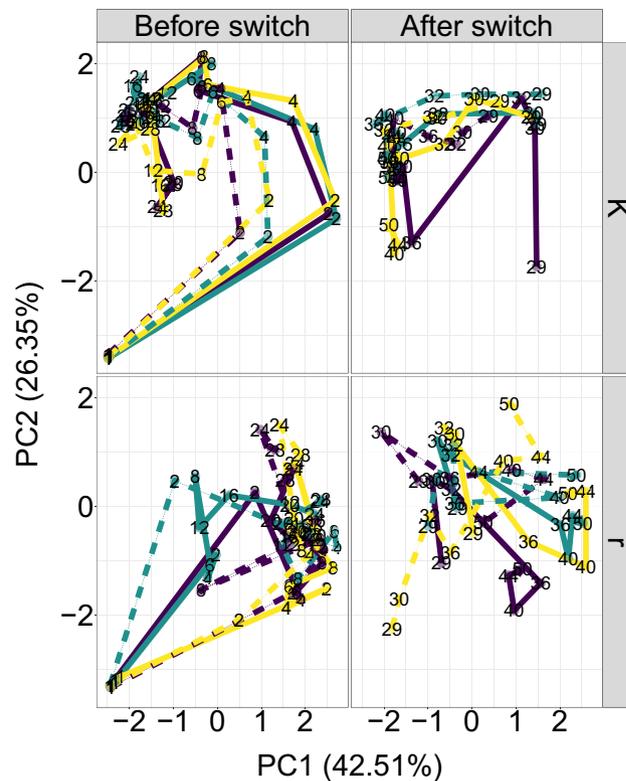


Figure 4. PCoA ordination of Bray-Curtis distances between samples showing the time trajectories for each microcosm. The single ordination was faceted vertically based on the state of selection regime at the time of sampling (being *r* or *K*-selection), and horizontally to highlight temporal trends. Solid and dotted lines indicate high (H) and low (L) resource supply, respectively. The labels indicate the day of sampling, whereas the line colours are purely to visually distinguish the replicate time series.

interactions or weaken the positive ones. However, this expectation was only partially fulfilled because we observed negative edges within modules in Fig. 5 also outside the two predicted periods of instability. One potential factor contributing to instability at the beginning of the experiment, was the fact that the oligotrophic seawater was introduced to high amount of nutrients, favouring *r*-strategists to proliferate even if the feeding was continuous.

Network robustness. If our results were different when changing parameters, the conclusions would be less likely to give us any real insight into how the communities actually behave. Therefore, we checked the robustness of the chosen parameters, changing one at a time while keeping all other parameters constant. Increasing the levels of random noise from low to medium (see Methods section) did not give any substantial difference in terms of significant interactions. Some cosmetic changes were visible due to different color labeling of communities and orientations of plots (details in Supplementary Section S2). Exchanging estimated absolute abundances with relative ones gave higher proportion of negative interactions and different assignments of OTUs into modules (see Supplementary Section S3). However, the greater trends in the results stay the same, such as clustering based on phylogeny and the considerable change in the community behaviour after switching selection regime.

Selecting a more stringent OTU filtering cutoff only has a minor consequence on the results, at the level of cosmetic changes in the plots (see Supplementary Section S4 for details). On the other hand, we notice a more pronounced effect when replacing the Spearman correlation by Pearson correlation. This is not surprising, since Spearman is non-parametric and Pearson measures degree of linear co-occurrence. In this case (Pearson), we got far fewer negative significant interactions for the same *q*-value, none of which are among the 500 most significant ones. Still, modules of phylogenetically related OTUs are present, and the selection regime still seems to explain the modules (Supplementary Section S5).

Discussion

In literature, challenges of microbial datasets such as sparsity, compositionality and habitat filtering have been addressed and solutions proposed for finding ecological interactions^{22,26–29}. Despite the fact that predictions from ecological interaction-inference tools have been successfully validated in some cases^{30,31}, any universally accepted gold standard of finding ecological microbial interactions is not yet agreed upon. Furthermore, some reviews assessing existing methods for inferring ecological interactions have demonstrated that current methods have far too low predictive power, and more refined approaches specifically designed to cope with difficulties in microbial datasets have failed to perform better than the basic ones^{19,32}. Hence, we believe that our choice

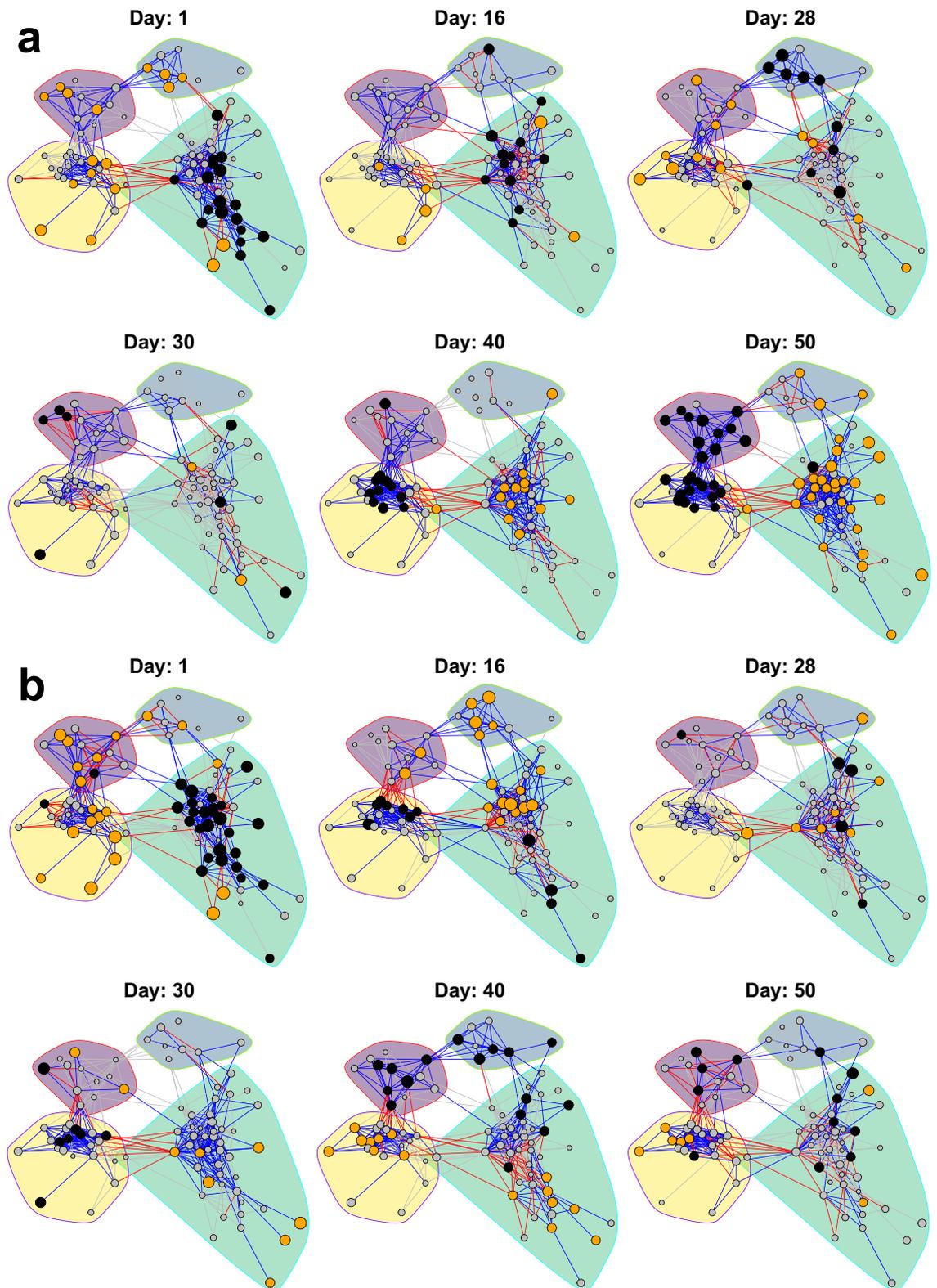


Figure 5. Dynamic visualisation of the network in Fig. 2, for (a) the RK selection group and (b) the KR selection group for high (H) resource supply. Nodes are coloured according to the corresponding OTUs' abundance compared to its overall mean for all sampling days, represented by its z -score. Orange, grey and black nodes mean higher, about the same or lower abundance than its mean, respectively. The edges are coloured by the product of the nodes' z -scores. This means that blue and red edges contribute to positive and negative association across the time series, respectively. The grey edges indicate that no major contribution to neither positive nor negative association is made. As we want to emphasize the orange and black nodes, the nodes with higher absolute z -scores are larger.

of using the relatively simple ReBoot²² procedure is reasonable, even though the approach in and of itself is somewhat coarse-grained.

We observed that our correlation networks clustered the OTUs according to taxonomy and niche preferences as a result of selection for *K*- and *r*-strategists. The finding that taxonomy and niche preferences dominate co-occurrence patterns is in line with work by Chaffron *et al.*³³ who produced similar results from samples stored in a ribosomal RNA database. Along the same line, Bock *et al.*³⁴ also noted that many of the interactions in a correlation network occur between closely related species when studying bacterial and protist communities in European lakes. Bacteria with similar niches are expected to be competitors and, hence, have negative interactions with each other. However, the effect of habitat filtering will create positive correlations between species with similar niches that are often stronger than those arising from ecological competition^{10,35,36}. The same reasoning goes for taxonomical relatedness, as closely related organisms often belong to the same niche and have similar functions. This favours positive interactions within modules, whereas we, to a lesser extent got negative interactions between modules where the growth requirements are different.

Moreover, we have not undertaken any attempt to deal with indirect interactions. This means that two OTUs can appear with a strong (correlation) link even though they have no direct effect on each other, but instead interact with a third OTU. Consequently, it is challenging to determine causality when working with inferred interactions. Also, such indirect effects can be caused by environmental variables and biological entities not taken into account, such as protists and bacteriophages.

For reasons mentioned above, our results are not meant to directly represent real ecological interactions. Nevertheless, our results are interesting from a fish-health perspective, as they show that selection regime can control community composition. In terms of *r*- and *K*-selection, literature consider the orders *Alteromonadales* and *Vibrionales* represented in module 4 as *r*-strategists, whereas the *Rhodobacteraceae* in module 3 are considered *K*-strategists³⁷. Additionally, *Vibrio* strains are known to cause disease in fish, whereas *Rhodobacteraceae* bacteria have been shown to protect against *Vibrio* infections through competition^{38–41}. This agrees with and extends prior knowledge that *K*-selection is a potent tool for improving fish health and survival^{1,7}.

The long-term behaviour of the community did not appear to depend on its prehistory. Potentially, this means that changing the microbiota from a detrimental to a healthy state in a running aquaculture facility requires the same measures as ensuring a healthy microbiota for a new facility. Furthermore, the trustworthiness of the results is strengthened by their robustness to changes in parameter settings, such as filtering cut-off, amount of random noise, type of abundance, and similarity measure.

This experiment was conducted in an artificial setting without any fish of which the health could be tracked. Furthermore, we do not know whether up-scaling and broader exposure could change the workings of the microbial community. Therefore, follow-up studies could be implemented in realistic aquaculture settings, perhaps such as a RAS facility, to investigate whether switching between *K*-selection and *r*-selection will yield the same community dynamics as described in this paper. Additionally, such an experiment would provide opportunity to investigate possible connections between the state of the microbial community and the health of the fish.

We acknowledge that there exist alternative approaches one could follow. For instance, treating the OTUs as discrete units is a bit misleading. As the results show, closely related OTUs often occur together, so it could make more sense to treat the bacteria as a taxonomical continuum. A novel approach based on amplicon sequence variants (ASVs) avoids the clustering of OTUs altogether by considering each individual unique read as an own entity⁴². The phylogenetic relatedness between ASVs could then be used as a constraint for finding co-occurrence patterns. In addition, incorporating environmental information, such as organic nutrient load, salinity, and temperature, would be useful because this allows us to better predict how the desired *K*-selection should be obtained. Joint Species Distribution Models (JSDMs)^{43,44} might have this useful potential to account for both species interaction, environmental factors, and taxonomical relatedness. However, its use in microbial ecology is still in its early stages and time dynamics are not yet embedded into the framework^{45,46}.

Methods

Selection-switch experiment. The dataset used for this article is previously published¹⁴, but we include a brief summary for completeness: Natural seawater was collected and used to inoculate microcosms in a 2 × 2 factorial crossover design with 3 replicates conducted for 50 days, which were sampled 18 times during the experiment. Half of the microcosms were given high (H) resource supply, whereas the other half were given low (L) resource supply. The factor of resource supply level was constant throughout the experiment. The other factor was the selection regime, which meant that the microcosms were either given continuous supply of nutrients (favouring *K*-selection, and hence the designation K) or being pulse-fed with nutrients after diluting the contents of the microcosms with growth medium (favouring *r*-selection, designated R). The active selection regime was switched at the experimental halfway point (between days 28 and 29), yielding two selection groups designated as RK and KR.

DNA was extracted from the collected samples, and the V3-V4 region of the bacterial 16S-rRNA gene was amplified with PCR using broad-coverage primers and the index sequences were ligated. The amplicon library was pooled and sequenced with two runs on an Illumina MiSeq machine. The reads are available at the European Nucleotide Archive with accession number ERS7182426-ERS7182513.

The USEARCH pipeline⁴⁷ (v11) was used to remove low-quality reads and cluster the reads into OTUs at 97% similarity level. Finally, the taxonomy of the OTUs was determined by the SINTAX classifier using data from the RPD training set (v.16) where the confidence threshold was set to 80%.

Quantification of bacterial density. For each sample, the bacterial density was quantified using flow cytometry (BC Accuri C6)¹⁴. In brief, the bacterial communities were diluted in 0.1x TE buffer, mixed with 2x

SYBR Green II RNA gel stain (ThermoFisher Scientific) and incubated in the dark at room temperature for 15 minutes. Then, each sample was measured for 2.5 minutes at $35 \mu\text{L min}^{-1}$ with an FL1-H (533/30 nm) threshold of 3000. We gated the bacterial population as those events with an FL1-A $> 10^4$ and FSC-A $< 10^5$. The raw flow cytometry data files are available at <https://doi.org/10.6084/m9.figshare.15104409>.

Alignment and phylogenetic tree. The selection-switch dataset was acquired directly from the authors¹⁴. This dataset consists of a total of 206 samples. Two of these samples were taken from the communities from which the reactors were inoculated, whereas the other samples were taken from the microcosms with 17 time points \times 4 regimes \times 3 replicates. We discarded the inoculum samples for further analysis. The OTU reference sequences were aligned with SINA version 1.6.1⁴⁸ using the SILVA Release 138 NR 99 SSU dataset⁴⁹. Using this alignment, the phylogenetic tree was constructed by neighbour-joining using MEGA X⁵⁰ with default parameters.

Filtering and preprocessing. The mean number of reads per sample was 63,460 with standard deviation 31,411. For our analysis, we wanted to estimate the abundance of each OTU as accurately as possible and therefore skipped any correction for unequal sequencing depth. Read counts for each OTU in each sample were divided by the total number of reads for the sample, generating relative abundances. Thereafter, all OTUs having a maximum abundance (across all samples) below a certain threshold, were removed. Three levels of filtering thresholds (as count proportions) were applied: High level at $5 \cdot 10^{-3}$, medium level at $1 \cdot 10^{-3}$ and low level at $5 \cdot 10^{-4}$. The purpose of the filtering was to remove rare OTUs in order to avoid noise and spurious correlations¹¹. For obtaining estimates of absolute abundances, the relative abundances were scaled by the estimate of total bacterial cell density for each sample. The *phyloseq* package (version 1.36.0)⁵¹ and the R programming language (version 4.1.1)⁵² facilitated this procedure. In addition, we wrote an R-package named *micInt* (version 0.18.0, available at <https://github.com/AlmaasLab/micInt>) to facilitate and provide a pipeline for the analysis.

Similarity measures and addition of noise. For this study, we used two similarity measures, the Pearson correlation and the Spearman correlation. A similarity measure, as referred to in this article, can be thought of as a function $f: \mathbb{R}^n \times \mathbb{R}^n \rightarrow D$ where $D = [-1, 1]$. In this regard, $f(\mathbf{x}, \mathbf{y})$ is the similarity of two abundance vectors \mathbf{x} and \mathbf{y} belonging to different OTUs, where $f(\mathbf{x}, \mathbf{y}) = 1$ indicates perfect correlation, $f(\mathbf{x}, \mathbf{y}) = 0$ indicates no correlation and $f(\mathbf{x}, \mathbf{y}) = -1$ indicates perfect negative correlation. Noise was added to distort patterns of double zeros, which otherwise could result in spurious correlations. Given two vectors \mathbf{x} and \mathbf{y} of abundances, normally distributed noise was added to each of the abundance vectors, and the similarity measure has invoked thereafter: Given a similarity measure f , the similarity between the abundance vectors after adding noise is given by:

$$f^*(\mathbf{x}, \mathbf{y}) = f(\mathbf{x} + \boldsymbol{\varepsilon}_x, \mathbf{y} + \boldsymbol{\varepsilon}_y), \quad (1)$$

where $\boldsymbol{\varepsilon}_x$ and $\boldsymbol{\varepsilon}_y$ are random vector where all components are independent and normally distributed with mean zero and variance γ^2 . The level of noise γ was determined by the smallest non-zero relative abundance x_{\min} in the dataset and a fixed constant s called the *magnitude factor*, such that $\gamma = s \cdot x_{\min}$. For no noise, $s = 0$, for low noise $s = 1$, for middle noise $s = 10$ and for high noise $s = 100$.

Network creation. Significance of the pairwise OTU associations were determined by the ReBoot procedure introduced by Faust *et al.*²² and shares the underlying algorithm used in the CoNet Cytoscape package⁵³. This approach accepts a dataset of microbial abundances and a similarity measure, and evaluates for each pair of OTUs in the dataset the null hypothesis H_0 : “The association between the OTUs is caused by chance”. By bootstrapping over the samples, the similarity score of each pair of OTUs is estimated, forming a bootstrap distribution. By randomly permuting the pairwise abundances of OTUs and finding the pairwise similarity scores, a bootstrap distribution is formed. The bootstrap and permutation distribution are then compared with a two-sided Z-test (based on the normal distribution) to evaluate whether the difference is statistically significant. For this, the z -value, p -value and q -value (calculated by the Benjamini-Hochberg-Yekutieli procedure⁵⁴) are provided for each pair of OTUs in the dataset. Our ReBoot approach is based on the R-package *cCrepe* (version 1.28.0)⁵⁵, but is integrated into the *micInt* package with the following major changes:

- The original ReBoot uses renormalization of the permuted abundances to keep the sum-to-constant constraint. Whereas this is reasonable to do with relative abundances, our modified version enables turning this feature off when we analyse data with absolute abundances.
- Optimizations have been made to memory use and CPU consumption to enable analyses of large datasets.
- In contrast to the usual ReBoot procedure, networks generated by the different similarity measures are not merged by p -value, but kept as they are.

For our analysis the number of bootstrap and permutation iterations was set to 1000. All OTUs being absent in more than $n \cdot 10^{-4}$ samples, where n is the total number of samples, were excluded through the *errthresh* argument but still kept for renormalization (if turned on). The associations were made across all samples, even the ones belonging to a different selection group or resource supply.

Dynamic PCoA visualization. All samples in the dataset were used for PCoA ordination, where the Bray-Curtis distance metric between the samples was applied before creating the decomposition. After the ordination was computed, the samples were divided into four facets based on their combination of current selection regime and resource supply. Finally, all samples belonging to the same microcosm were connected by a line in chronological order and the line was given a separate style based on the resource supply and coloured to visually distinguish it from the two other replicate microcosm within the same facet.

Permutational multivariate analysis of variance. Sequential PERmutational Multivariate Analysis of VAriance (PERMANOVA) of the samples was conducted on the absolute abundances, where only the samples from day 28 and 50 were included. These sample points correspond to time just before the experimental selection-regime crossover and a point at the end of the experiment. These days were selected because they were the most likely to capture the composition of stable communities in contrast to transient ones. The procedure was carried out by the function `adonis` from the R package `vegan` (version 2.5-7) with 10^6 permutations. The dependent data given to the function was the matrix of one minus the Spearman correlation of the samples (in order to resample dissimilarity), while the independent variables were the selection group (first variable) and the current selection regime (second variable).

Network visualization. The networks were plotted by the R package `igraph` (version 1.2.6)⁵⁶. Network modules were found by the `walktrap`²⁵ algorithm implemented in `igraph` with the setting `steps=20`, including the positive edges only. Later, the negative edges were added and the networks plotted with the community labelling.

The time dynamics of the networks were visualised by taking the former network and adjusting the node colour and size, as well as the edge colour. For this, a certain combination of selection group (i.e RK) and resource supply (i.e H) was chosen. Further, let $x_{i,j,k}$ be the abundance of OTU k at sampling day i in microcosm j . As there are three replicates, we have that $j = 1, 2, 3$. If the underlying network was created by Pearson correlation, we denote the day mean $x_{i,..k}$ as the average over the replicates, this is:

$$x_{i,..k} = \frac{x_{i,1,k} + x_{i,2,k} + x_{i,3,k}}{3}. \quad (2)$$

The time series mean of OTU k , $x_{,..k}$ is the mean of these daily means over all sampling days,

$$x_{,..k} = \frac{\sum_{i=1}^N x_{i,..k}}{N}, \quad (3)$$

where N denotes the number of sampling days. Furthermore, we have the associated standard deviation σ_k as given by:

$$\sigma_k = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_{i,..k} - x_{,..k})^2}. \quad (4)$$

The z -value of the abundance of OTU k at day i is then:

$$z_{i,k} = \frac{x_{i,..k} - x_{,..k}}{\sigma_k}. \quad (5)$$

This value is used in the mapping of the node sizes and colours. The node for OTU k at sampling day i has the size $a + b \cdot |z_{i,k}|$, where a and b are constants. Furthermore, the same node is coloured:

- Black if $z_{i,k} < -1$. This indicates that the OTU that day had a lower abundance than the average.
- Grey if $-1 \leq z_{i,k} \leq 1$. This indicates that the OTU that day had about the same abundance as the average.
- Orange if $z_{i,k} > 1$. This indicates that the OTU that day had a higher abundance than the average.

Furthermore, the edge colour are dependent on the product of the two participating nodes. Hence, the edge between OTU k and OTU l at day i will have the colour:

- Red if $z_{i,k} \cdot z_{i,l} < -0.3$. This shows a contribution to a negative interaction.
- Gray if $-0.3 \leq z_{i,k} \cdot z_{i,l} \leq 0.3$. This shows no major contribution of neither a positive nor negative interaction.
- Blue if $z_{i,k} \cdot z_{i,l} > 0.3$. This shows a contribution to a positive interaction.

Our approach is motivated by the fact that the Pearson correlation $\rho_{k,l}$ of the day means of OTU k and OTU l is given by:

$$\rho_{k,l} = \frac{1}{N} \sum_{i=1}^N z_{i,k} \cdot z_{i,l}. \quad (6)$$

For the Spearman correlation, the visualization is based on the rank of each of the OTU abundance values in a sample. Hence, instead of using the raw abundances $x_{i,j,k}$ in the calculation of the day mean, the ranks $r_{i,j,k}$

are used instead, and all subsequent calculations and mappings are the same. In a scenario when there is only one replicate, the quantity $\rho_{k,l}$ would then be the Spearman correlation of the abundances of OTU k and OTU l .

Data availability

The micInt package used for the analysis is available at GitHub <https://github.com/AlmaasLab/micInt>. The raw flow cytometry data files are available at <https://doi.org/10.6084/m9.figshare.15104409>. The scripts used to analyse the data and create the figures are available on GitHub at <https://github.com/yaccos/Microbial-co-occurrence>.

Received: 4 August 2021; Accepted: 19 November 2021

Published online: 06 December 2021

References

- De Schryver, P. & Vadstein, O. Ecological theory as a foundation to control pathogenic invasion in aquaculture. *ISME J.* **8**, 2360–2368. <https://doi.org/10.1038/ismej.2014.84> (2014).
- Hjeltnes, B., Bang-Jensen, B., Bornø, G., Haukaas, A. & Walde, C. S. The health situation in norwegian aquaculture 2018 (2019).
- Macpherson, H. L., Bergh, O. & Birkbeck, T. H. An aerolysin-like enterotoxin from vibrio splendidus may be involved in intestinal tract damage and mortalities in turbot, *Scophthalmus maximus* (L.), and cod, *Gadus morhua* L., larvae. *J. Fish Dis.* **35**, 153–167 (2012).
- Derome, N., Gauthier, J., Boutin, S. & Llewellyn, M. *Bacterial Opportunistic Pathogens of Fish* 81–108 (Springer, Cham, 2016).
- May, T. *et al.* Reducing mortality associated with opportunistic infections in atlantic salmon salmo salar fry using hydrogen peroxide and peracetic acid. *Aquacul. Res.* (2021).
- Andrews, J. H. & Harris, R. F. r-selection and k-selection and microbial ecology. *Adv. Microbial. Ecol.* **9**, 99–147 (1986).
- Vadstein, O., Attramadal, K. J. K., Bakke, I. & Olsen, Y. K-selection as microbial community management strategy: A method for improved viability of larvae in aquaculture. *Front. Microbiol.* <https://doi.org/10.3389/fmicb.2018.02730> (2018).
- Barberán, A., Bates, S. T., Casamayor, E. O. & Fierer, N. Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J.* **6**, 343–351. <https://doi.org/10.1038/ismej.2011.119> (2012).
- Faust, K. & Raes, J. Microbial interactions: From networks to models. *Nat. Rev. Microbiol.* **10**, 538–550. <https://doi.org/10.1038/nrmicro2832> (2012).
- Röttgers, L. & Faust, K. From hairballs to hypotheses—biological insights from microbial networks. *FEMS Microbiol. Rev.* **42**, 761–780. <https://doi.org/10.1093/femsre/fuy030> (2018).
- Faust, K., Lahti, L., Gonze, D., de Vos, W. M. & Raes, J. Metagenomics meets time series analysis: Unraveling microbial community dynamics. *Curr. Opin. Microbiol.* **25**, 56–66. <https://doi.org/10.1016/j.mib.2015.04.004> (2015).
- David, L. A. *et al.* Host lifestyle affects human microbiota on daily timescales. *Genome Biol.* **15**, R89. <https://doi.org/10.1186/gb-2014-15-7-r89> (2014).
- Shade, A. *et al.* Fundamentals of microbial community resistance and resilience. *Front. Microbiol.* **3**, 417. <https://doi.org/10.3389/fmicb.2012.00417> (2012).
- Gundersen, M. S., Morelan, I. A., Andersen, T., Bakke, I. & Vadstein, O. The effect of periodic disturbances and carrying capacity on the significance of selection and drift in complex bacterial communities. *ISME Commun.* **1**, 53. <https://doi.org/10.1038/s43705-021-00058-4> (2021).
- Tringe, S. G. *et al.* Comparative metagenomics of microbial communities. *Science* **308**, 554–557. <https://doi.org/10.1126/science.1107851> (2005).
- Layeghifard, M., Hwang, D. M. & Guttman, D. S. Disentangling interactions in the microbiome: A network perspective. *Trends Microbiol.* **25**, 217–228. <https://doi.org/10.1016/j.tim.2016.11.008> (2017).
- Nguyen, N.-P., Warnow, T., Pop, M. & White, B. A perspective on 16s rna operational taxonomic unit clustering using sequence similarity. *npj Biofilms Microbiomes* <https://doi.org/10.1038/npjbiofilms.2016.4> (2016).
- Springer, *Spearman Rank Correlation Coefficient*, 502–505 (Springer, 2008).
- Weiss, S. *et al.* Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. *ISME J.* **10**, 1669–1681. <https://doi.org/10.1038/ismej.2015.235> (2016).
- Braga, R. M., Dourado, M. N. & Araújo, W. L. Microbial interactions: Ecology in a molecular perspective. *Braz. J. Microbiol.* **47**, 86–98. <https://doi.org/10.1016/j.bjm.2016.10.005> (2016).
- Tshikantwa, T. S., Ullah, M. W., He, F. & Yang, G. Current trends and potential applications of microbial interactions for human welfare. *Front. Microbiol.* **9**, 1156. <https://doi.org/10.3389/fmicb.2018.01156> (2018).
- Faust, K. *et al.* Microbial co-occurrence relationships in the human microbiome. *PLoS Comput. Biol.* <https://doi.org/10.1371/journal.pcbi.1002606> (2012).
- Berry, D. & Widder, S. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Front. Microbiol.* <https://doi.org/10.3389/fmicb.2014.00219> (2014).
- Betensky, R. A. The p-value requires context, not a threshold. *Am. Stat.* **73**, 115–117. <https://doi.org/10.1080/00031305.2018.1529624> (2019).
- Pons, P. *et al.* Computing communities in large networks using random walks. *Lect. Notes Comput. Sci.* **3733**, 284–293 (2005).
- Friedman, J. & Alm, E. J. Inferring correlation networks from genomic survey data. *PLoS Comput. Biol.* **8**, 1–11. <https://doi.org/10.1371/journal.pcbi.1002687> (2012).
- Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V. & Egozcue, J. J. Microbiome datasets are compositional: And this is not optional. *Front. Microbiol.* <https://doi.org/10.3389/fmicb.2017.02224> (2017).
- Kurtz, Z. D. *et al.* Sparse and compositionally robust inference of microbial ecological networks. *PLoS Comput. Biol.* **11**, e1004226. <https://doi.org/10.1371/journal.pcbi.1004226> (2015).
- Brisson, V., Schmidt, J., Northen, T. R., Vogel, J. P. & Gaudin, A. A new method to correct for habitat filtering in microbial correlation networks. *Front. Microbiol.* **10**, 585. <https://doi.org/10.3389/fmicb.2019.00585> (2019).
- Stein, R. R. *et al.* Ecological modeling from time-series inference: Insight into dynamics and stability of intestinal microbiota. *PLoS Comput. Biol.* <https://doi.org/10.1371/journal.pcbi.1003388> (2013).
- Lima-Mendez, G. *et al.* Ocean plankton determinants of community structure in the global plankton interactome. *Science (New York, N.Y.)* **348**, 1262073. <https://doi.org/10.1126/science.1262073> (2015).
- Hirano, H. & Takemoto, K. Difficulty in inferring microbial community structure based on co-occurrence network approaches. *BMC Bioinform.* <https://doi.org/10.1186/s12859-019-2915-1> (2019).
- Chaffron, S., Rehrauer, H., Pernthaler, J. & von Mering, C. A global network of coexisting microbes from environmental and whole-genome sequence data. *Genome Res.* **20**, 947–959. <https://doi.org/10.1101/gr.104521.109> (2010).
- Bock, C. *et al.* Factors shaping community patterns of protists and bacteria on a European scale. *Environ. Microbiol.* <https://doi.org/10.1111/1462-2920.14992> (2020).

35. Fisher, C. K. & Mehta, P. Identifying keystone species in the human gut microbiome from metagenomic timeseries using sparse linear regression. *PLoS One* <https://doi.org/10.1371/journal.pone.0102451> (2014).
36. Cazelles, K., Araujo, M. B., Mouquet, N. & Gravel, D. A theory for species co-occurrence in interaction networks. *Theor. Ecol.* **9**, 39–48. <https://doi.org/10.1007/s12080-015-0281-9> (2016).
37. Lauro, F. M. *et al.* The genomic basis of trophic strategy in marine bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 15527–15533. <https://doi.org/10.1073/pnas.0903507106> (2009).
38. D'Alvise, P. W., Melchiorson, J., Porsby, C. H., Nielsen, K. F. & Gram, L. Inactivation of vibrio anguillarum by attached and planktonic roseobacter cells. *Appl. Environ. Microbiol.* **76**, 2366–2370. <https://doi.org/10.1128/AEM.02717-09> (2010).
39. Planas, M. *et al.* Probiotic effect in vivo of roseobacter strain 27–4 against vibrio (listonella) anguillarum infections in turbot (scophthalmus maximus l.) larvae. *Aquaculture* **255**, 323–333. <https://doi.org/10.1016/j.aquaculture.2005.11.039> (2006).
40. D'Alvise, P. W., Lillebø, S., Wergeland, H. I., Gram, L. & Bergh, Ø. Protection of cod larvae from vibriosis by phaeobacter spp.: A comparison of strains and introduction times. *Aquaculture* **384–387**, 82–86. <https://doi.org/10.1016/j.aquaculture.2012.12.013> (2013).
41. Porsby, C., Nielsen, K. & Gram, L. Phaeobacter and ruegeria species of the roseobacter clade colonize separate niches in a danish turbot (scophthalmus maximus)-rearing farm and antagonize vibrio anguillarum under different growth conditions. *Appl. Environ. Microbiol.* **74**, 7356–7364. <https://doi.org/10.1128/AEM.01738-08> (2008) (Cited By 112).
42. Callahan, B. J., McMurdie, P. J. & Holmes, S. P. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J.* **11**, 2639–2643. <https://doi.org/10.1038/ismej.2017.119> (2017).
43. Warton, D. I. *et al.* So many variables: Joint modeling in community ecology. *Trends Ecol. Evol.* **30**, 766–779. <https://doi.org/10.1016/j.tree.2015.09.007> (2015).
44. Ovaskainen, O. *et al.* How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecol. Lett.* **20**, 561–576. <https://doi.org/10.1111/ele.12757> (2017).
45. Björk, J. R., Hui, F. K. C., O'Hara, R. B. & Montoya, J. M. Uncovering the drivers of host-associated microbiota with joint species distribution modelling. *Mol. Ecol.* **27**, 2714–2724. <https://doi.org/10.1111/mec.14718> (2018).
46. Leite, M. F. & Kuramae, E. E. You must choose, but choose wisely: Model-based approaches for microbial community analysis. *Soil Biol. Biochem.* **151**, 108042. <https://doi.org/10.1016/j.soilbio.2020.108042> (2020).
47. Edgar, R. C. Search and clustering orders of magnitude faster than blast. *Bioinformatics* **26**, 2460–2461. <https://doi.org/10.1093/bioinformatics/btq461> (2010).
48. Pruesse, E., Peplies, J. & Glöckner, F. O. Sina: Accurate high-throughput multiple sequence alignment of ribosomal rna genes. *Bioinformatics (Oxford, England)* **28**, 1823–1829. <https://doi.org/10.1093/bioinformatics/bts252> (2012).
49. Quast, C. *et al.* The silva ribosomal rna gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596. <https://doi.org/10.1093/nar/gks1219> (2013).
50. Kumar, S., Stecher, G., Li, M., Nnyaz, C. & Tamura, K. Mega x: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549. <https://doi.org/10.1093/molbev/msy096> (2018).
51. McMurdie, P. J. & Holmes, S. Phyloseq: An r package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **8**, e61217 (2013).
52. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2017).
53. Faust, K. & Raes, J. Conet app: Inference of biological association networks using cytoscape. *F1000Res* **5**, 1519. <https://doi.org/10.12688/f1000research.9050.2> (2016).
54. Benjamini, Y. & Yekutieli, D. The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* **29**, 1165–1188 (2001).
55. Schwager, E., Bielski, C. & Weingart, G. *ccrease: ccrease_and_nc.score*, r package version 1.14.0 edn. (2014).
56. Csardi, G. & Nepusz, T. The igraph software package for complex network research. *Int. J. Complex Syst.* 1695 (2006).

Acknowledgements

Thanks to Olav Vadstein for being helpful with explaining the experimental procedures and supply input on how the data could be interpreted and processed. J.P.P. and E.A. are grateful to ERA CoBioTech project CoolWine and Norwegian Research Council grant 283862 for funding.

Author contributions

E.A. and J.P.P. conceived the project. M.S.G provided the data and comments to how to interpret it. J.P.P wrote the software, did the analysis, created visualisations and wrote the first draft of the paper. All authors contributed to and accepted the final version of the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-03018-z>.

Correspondence and requests for materials should be addressed to E.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021