

# Scale-dependent signatures of microbial co-occurrence revealed via multilayer network analysis

Geut Galai <sup>1,®</sup>, Dafna Arbel<sup>1,®</sup>, Keren Klass<sup>2</sup>, Ido Grinshpan <sup>1,3</sup>, Itzhak Mizrahi <sup>1,3</sup>,  
and Shai Pilosof <sup>1,4,\*</sup>

<sup>1</sup>*Department of Life Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel*

<sup>2</sup>*The Alexander Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel*

<sup>3</sup>*National Institute of Biotechnology in the Negev, Ben-Gurion University of the Negev, Be'er Sheva, Israel*

<sup>4</sup>*The Goldman Sonnenfeldt School of Sustainability and Climate Change, Ben-Gurion University of the Negev, Be'er Sheva, Israel*

\* *Corresponding author: pilos@bgu.ac.il*

<sup>®</sup> *Equal contribution.*

## Abstract

The composition of microbial communities underlies many ecosystem functions. Co-occurrence networks offer insights into the complexity of microbial interactions, particularly in highly diverse environments where direct observation is challenging. However, identifying the scale at which local and non-local processes structure co-occurrence networks remains challenging because it requires simultaneously analyzing network structure within and between local networks. Using the rumen microbiome of 1,012 cows in seven European farms, we built a multilayer network in which microbes (nodes) are linked within and between farms (layers), effectively creating a meta-co-occurrence network. At the local scale, we tested for a non-random transitive signature (the tendency of a microbe to close triangles) because transitivity indicates environmental similarities or interaction dependence. At the non-local scale, we tested for partner fidelity because if microbes tend to conserve their interaction partners more than expected by chance, co-occurrence is not determined locally. We further tested at which scale (farm vs. regional) the partition of microbes to modules of densely co-occurring microbes was prominent. Via comparison to shuffled networks, we discovered that, although core microbes appeared across the entire system, there were non-random signatures of local and non-local co-occurrence patterns. Microbes had transitive interactions and tended to conserved partners across farms. Microbes in four farms were clustered separately (each farm in a module), providing a signature for a strong but not complete influence of local processes. The genetic structure of cow populations partly explains modularity, indicating that environmental filtering is an important force shaping co-occurrence. Our results provide evidence for the scale at which interconnected co-occurrence networks are structured using a clear set of hypotheses. Our developed approach can also be used to explore local and non-local signatures in species interaction metanetworks or metacommunities.

## 1 Introduction

The composition and interaction structure (who interacts with whom) of microbial communities determine their stability and function [1,2]. For example, different gut microbiome compositions are associated with animals' digestion [3,4]. Therefore, a central goal of microbial ecology is to understand the mechanisms underlying microbiome composition and structure. However, it is highly challenging in microbial systems to observe interactions (e.g., competition, cross-feeding) when diversity is high under natural conditions. Hence, researchers use co-occurrence networks to describe the structure of microbial communities. While co-occurrence is not direct evidence for interactions, it is a prerequisite [5–8]. In a co-occurrence network, a link represents a statistically significant association across space (or time). Such associations emanate from three primary processes: dispersal limitation, environmental filtering, and biotic interactions [9–11]. Therefore, the structure of microbial co-occurrence networks is governed by a mixture of processes operating at different scales, as is also true for other network types such as food webs [12]. Identifying the signatures of local vs. non-local processes that determine the structure of co-occurrence networks remains a challenge and is the primary goal of this work.

At local scales, co-occurrence depends on direct and indirect interactions with other microbes. For instance, competition is a selective force that occurs at a local scale because it typically requires physical proximity [13]. Cross-feeding networks in which microbes depend on each other directly or indirectly via the production and consumption of metabolites is another key example [14]. At the non-local scale, processes such as environmental filtering affect local microbe composition [15,16]. For instance, physical-chemical conditions [17,18] and temperature [16]. Therefore, co-occurrence patterns can vary in space, for example, due to variability in environmental conditions or host characteristics [17,18]. Moreover, local communities exchange species with a 'regional' species pool and are interconnected via colonization-extinction dynamics due to microbe dispersal, generating meta-communities [19,20].

Ecological networks, including co-occurrence networks, contain signatures—that is, observable patterns indicative of the processes and factors that generated the network [1,21]. For example, in a modular structure, modules represent clusters of microbes that co-occur strongly with each other but weakly with microbes from other modules. A modular structure may be formed because microbes from different modules occupy different ecological niches [22,23], or because they strongly interact.

One approach to detect scale signatures in structure is to analyze local co-occurrence networks separately and compare them [18,24]. However, this approach effectively treats these networks as separate, disconnected communities. Alternatively, one can generate a single network by pooling all local occurrence information [17]. Nevertheless, pooling microbe occurrences across zones ignores local information. Because microbe co-occurrence is governed by a mixture of processes operating at different scales, it is necessary to analyze network structure within and between local networks simultaneously. That is, we need a mathematical object that separates local networks on the one hand, and connects them on the other. Multilayer networks [25] are a very powerful tool that can be applied to represent multiple local co-occurrence networks that are interlinked, effectively creating a meta-co-occurrence network. This framework allowed us to investigate, for the first time, local and non-local network signatures of co-occurrence simultaneously.

We tested our questions on the ecologically relevant and highly complex microbial ecosystem of the rumen microbiome [26]. The rumen ecosystem is a confined microbial environment in terms of abiotic conditions. This is particularly advantageous because it allows for focused examination of interactions

within this ecosystem, reducing biases that result from a mismatch between spatial scales of analysis and interactions [8]. Moreover, the rumen ecosystem’s considerable importance to food security drives a comprehensive monitoring of on-farm parameters, including diet composition and occasionally even host genetics which enables us to establish connections between ecological patterns and their potential determinants. We chose a rumen microbiome data set of 1,012 cows from seven European farms of different localities that were monitored for diet and host genetics [27]. We focused our analysis on the ‘common core’ microbiome—a subset of microbes that occur in a minimum proportion of hosts [28]. Using core microbes allowed us to focus on microbes that are more likely to influence other microbes and drive co-occurrence patterns [29]. Core microbes were also associated with cow phenotypes (e.g., methane emissions and dairy production) [27].

At the local scale, we hypothesized that there would be repeatable signatures of co-occurrence across farms because the function of the rumen microbiome was consistent regardless of location. Specifically, we tested for a non-random transitive signature such that if microbe A co-occurred with B, and B with C, then microbe A would also co-occur with C. Transitivity is a local measure of network clustering. Non-random transitivity would indicate that local co-occurrence structure was driven by either similar environmental conditions (A, B, and C occupy the same niche) or that there was a signature of interaction dependence [13]. For example, rock-paper-scissors competition dynamics can lead to the coexistence of three species [30].

At the non-local scale, we performed two analyses. At the microbe (node) level, we hypothesized that the extent to which a focal microbe will interact with the same partners across farms would be determined by a mix of non-local and local factors. We tested this hypothesis using partner fidelity [31]. If microbes tend to conserve their interaction partners more than expected by chance, then co-occurrence is not a local feature but rather repeatable at higher scales. At the mesoscale network level, we tested at which scale—farm, country, or regional (across countries)—the partition of microbes to modules was prominent. We hypothesized that the overlap between modules and farms would contain signatures of scale and the genetic background of the cow hosts.

Via comparison to shuffled networks, we discovered that, although core microbes appeared across the entire system, there were non-random signatures of local and non-local co-occurrence patterns. Microbes had transitive interactions, and they tended to conserved partners across farms. Notably, the modular structure of the meta-co-occurrence network could be partly explained by the genetic structure of cow populations.

## 2 Methods

### 2.1 Data

We used data from Wallace et al. [27]. Rumen microbiome samples were collected from 1012 cows. The cows were divided into two cow breeds located in four different countries in Europe: 816 Holstein dairy cows in the United Kingdom (farms UK1 and UK2) and Italy (farms IT1, IT2, and IT3), and 196 Nordic Red dairy cows in Finland (farm FI1) and Sweden (farm SE1). Cows were fed highly controlled diets. The Holsteins received a maize silage-based diet, while the Nordic Reds received a nutritionally equivalent diet based on grass silage as forage (Table S1). In the original study [27], the farmers did not reveal the farm identity or location. Hence, the regions the farms are located in are available, but not their exact coordinates. This does not affect our study because the signatures

we obtained are not related to distance. Unlike wildlife, cows and the microbes within them do not disperse freely across space.

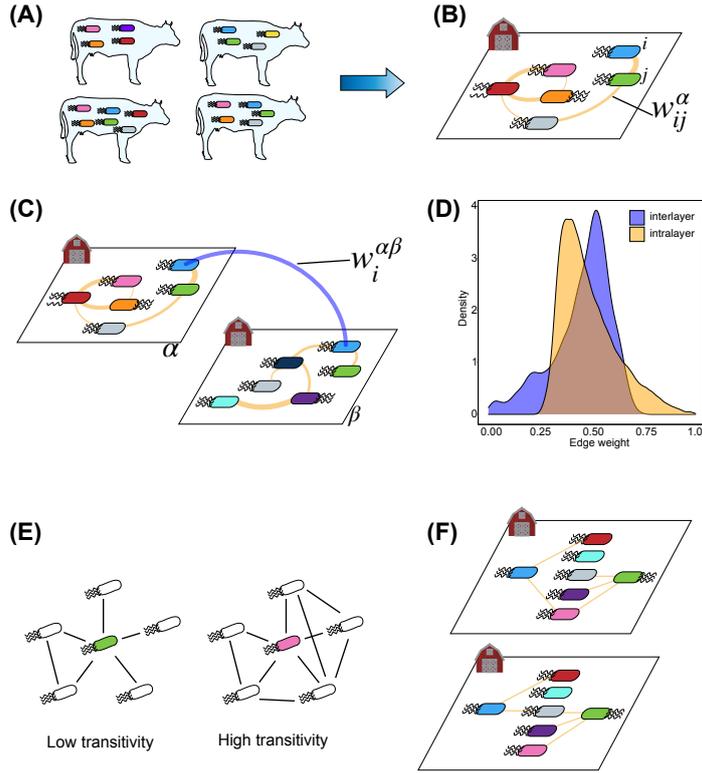
Each rumen sample was analyzed for paired amplicon reads [27]. The paired reads were then run through DADA2 using default parameters as pre-merged reads. The taxonomic assignments for the reads were performed using the SILVA database (silva\_nr\_v138). Because the data was for sequences of the V4 region, only Prokaryotes and Archaea are included. The complete data set included 12,109 microbe ASVs, clustered at 100% similarity. The microbiome samples were composed primarily of bacteria (99%), and the rest were archaea, or unknown.

We used the ‘common core’ microbes in each farm [27,28] to focus on the signal produced by the key microbes. Wallace et al. [27] defined the common core microbes as those occurring at  $\geq 50\%$  of cows within a farm, remaining with 454 microbes. Even at this strict threshold, the core microbes strongly predicted the cow phenotype (e.g., methane emission, lactation). Therefore, this can be considered as the ‘functional core’ [28]. Nevertheless, a 50% threshold is stringent as it includes only 3.7% of ASVs, which omits variation in occurrence that can influence signatures in the meta-co-occurrence network. We opted not to deviate drastically from the functional core because it is biologically meaningful. Hence, we focused our analysis on microbes occurring in at least 30% of the cows. We also analyzed other thresholds: 5%, 10%, and 20%. These results, sensitivity analysis, and further discussion on the core microbiome selection are in the Supplementary Information.

## 2.2 Construction of farm-level co-occurrence networks

Within each farm, we constructed a co-occurrence network using a combinatorial model that provides the probability that two ASVs significantly positively or negatively co-occur [32] (implemented with R package `co-occur` v1.3 [33]) (Fig. 1A-B). Statistical significance of co-occurrence is determined per ASV-pair by calculating the combinatorial probability that the two ASVs will randomly co-occur with a probability equal to, or more than, observed in the data. The calculation considers their prevalence in cows and the number of cows. If this probability is smaller than 5%, we considered an intralayer co-occurrence link (Fig. 1B). We note that there are several methods to detect co-occurrence in microbial communities, each with its advantages and disadvantages [6]. In reality, no method will detect the ‘true’ co-occurrence. We used Veech’s method [32] for three reasons. First, it has a ‘built-in’ null hypothesis whereby p values are calculated using a distribution-free probabilistic model [32]. It, therefore, avoids potential biases due to randomization-based null models [34]. Second, our data were binary (presence/absence of a microbe in cows). Third, it has been widely used and found to be a robust method, including in studies of microbial communities [35,36].

We drew a link for microbe pairs with statistically significant positive co-occurrences. Calculation of co-occurrence uses the cows, so these are not included in the resulting network. Hence, in the farm-level co-occurrence networks (layers in the multilayer network), nodes were ASVs (Fig. 1B). The combinatorial model was solely used to determine the existence of co-occurrence links ( $w_{ij}^\alpha \in \{0, 1\}$ ). We used these binary networks to calculate transitivity and partner fidelity, indices that use binary networks (see below).



**Fig. 1: Network construction.** (A) Occurrence data consisted of microbes in cows. The four cows in this toy example belong to the same farm. (B) For each farm, we calculated the probability that two microbes will co-occur across the cows. Co-occurrence calculation creates a microbial co-occurrence network without cows. Microbes are connected whenever the co-occurrence is statistically significant. (C) We connected the farm co-occurrence networks to form a *multilayer co-occurrence network*. In this network, layers (depicted as polygons) are farm-level networks (panel B). Intralayer link weight,  $w_{ij}^{\alpha}$  (depicted as link width), is the Jaccard similarity in the cow identities where the microbe pair occurs (Methods). Interlayer links ( $w_i^{\alpha\beta}$ ; purple) connect a microbe to itself between farms. The weight of interlayer links is the UniFrac similarity between the co-occurring partners of each focal microbe. (D) Density plots of intralayer (orange) and interlayer (purple) link weights. (E) Example for transitivity (see Methods for details on calculations). (F) Example for partner fidelity. The blue bacteria has low fidelity as it changes partners across farms, while the green bacteria has high fidelity as it maintains partners (see Methods for details on calculations).

### 2.3 Clustering coefficient

We tested transitivity (Fig. 1E) in co-occurrence by calculating the clustering coefficient. The clustering coefficient of a microbe  $i$ ,  $CC_i$ , is defined as the number of closed triangles (three connected nodes, including  $i$ ) out of all possible triangles:  $CC_i = 2N_i / (k_i(k_i - 1))$ , where  $N_i$  is the total number of links between  $i$ 's neighbors and  $k_i$  is the degree (number of links) of  $i$ . If none of the co-occurring partners of a focal microbe co-occur with one another,  $CC_i$  will be 0. In contrast, if all the neighbors of a microbe co-occur among themselves,  $CC_i$  will be 1. Because  $CC_i$  can be biased when the degree is low (it is easier to close triangles), we only calculated CC for ASVs with ten or more partners [37].

### 2.4 Partner fidelity

Partner fidelity captures the tendency of an ASV to co-occur with the same ASVs in different layers (farms) (Fig. 1F). Mathematically, this translates to how similar the partners of a focal ASV are across layers [31]. For each focal ASV  $i$ , we created a table in which columns are the farms in which

the ASV occurs, and rows are the focal ASV’s co-occurring partners within each farm (see diagram in Fig. S7). Table entries were 1 if a partner occurred in a farm or 0 otherwise. We used the table to calculate the Jaccard similarity in partners between farms (performed with R package `vegan` v2.5-7 [38]). We used the mean of the resulting similarity matrix as a measure of partner fidelity that we termed  $P_i^J$ . This index ranges from 0 to 1 when the identities of co-occurring partners of a microbe  $i$  are completely discordant or exactly the same, respectively.

However, some microbes are highly similar in their sequence and can effectively be phylogenetically close strains. Therefore, microbe identity (the ASV label in the data set) may not reveal the complete picture. To account for phylogeny, we calculated partner fidelity as we did for  $P_j^U$ , using UniFrac [39] instead of Jaccard. We used the R package `GUniFrac` version 1.4 [40]. We termed this index  $P_i^U$ . UniFrac is a similarity index that considers the degree of divergence between ASV sequences instead of ASV identity. If the lineages of a microbe’s partners in different farms are distinct (no branches shared),  $P_i^U$  approximates 0; in contrast, if they share the maximum possible branch length,  $P_i^U$  is 1. For UniFrac, we first constructed a phylogenetic tree. We used the genetic sequences to reconstruct an initial phylogenetic tree using a neighbor-joining method (using the R package `ape` v5.6-2 [41]). Specifically, the conversion from sequence to distance is calculated according to the Tamura and Nei model [42]. The likelihood of this initial tree was computed given the original data, and its parameters were optimized (using the R package `phangorn` v2.8-1 [43]). The significance of the fitting was tested using ANOVA and AIC (R package `stats` v4.1.2 [44]).

## 2.5 Network shuffling for statistical tests

We tested for non-random co-occurrence patterns by creating 500 shuffled networks for each farm. We randomly distributed the microbes between cows within each farm, conserving the number of microbes a cow hosts. This is a conservative algorithm because allowing the number of microbes per cow to vary would make it easier to detect statistically significant results (increasing false positive rates) [45]. This strengthens our results, ensuring that the non-random patterns revealed are statistically strong and biologically meaningful. After redistributing microbes, we recalculated the farm’s co-occurrence network. This process was repeated 500 times per farm, resulting in 3,500 shuffled networks.

## 2.6 Calculating significance using z-scores

We tested for non-random patterns in transitivity by comparing  $CC_i$  of each ASV  $i$  to that obtained for the same ASV in the layer across 500 shuffled networks using a z-score, following [31]. Z-scores are calculated as

$$z_i = \frac{CC_i^{empirical} - \overline{CC_i^{shuffled}}}{\sigma^{shuffled}}, \quad (1)$$

where  $\overline{CC_i^{shuffled}}$  and  $\sigma^{shuffled}$  are the mean and standard deviation of the transitivity obtained from the 500 shuffled networks. Hence, a positive (negative) z-score suggests that transitivity is higher (lower) than expected if microbes were randomly distributed across cows. The significance of each empirical value was determined at the 0.05 level: a z-score  $> 1.96$  or  $< -1.96$  indicates that  $CC_i$  is greater or lower than the random expectation, respectively.

Network shuffling also alters the partners of each microbe in a farm and, therefore, serves as an adequate null model for testing partner fidelity. The calculation of z-scores and significance for  $P_i^J$

and  $P_i^U$  was done exactly the same as for  $CC_i$ .

## 2.7 Construction of a meta-co-occurrence multilayer network

For the modularity analysis, we connected the farm-level co-occurrence networks to form a multilayer network (Fig. 1C). Intralayer link weights ( $w_{ij}^\alpha$ ) encode the similarity in the cow identities of each pair of microbes  $i$  and  $j$  within layer  $\alpha$ . The weight of any intralayer link  $w_{ij}^\alpha$  connecting an ASV  $i$  to  $j$  within a farm  $\alpha$  was calculated as

$$w_{ij}^\alpha = \frac{n_{ij}}{n_i + n_j - n_{ij}}, \quad (2)$$

where  $n_i$  and  $n_j$  are the number of cows in which microbes  $i$  and  $j$  occur, respectively, and  $n_{ij}$  are the number of cows in which both occur. Therefore, when the two microbes co-occur in the same individual cows  $w_{ij}^\alpha = 1$ . Intralayer links were undirected such that  $w_{ij}^\alpha = w_{ji}^\alpha$  (Fig. 1B,C).

To create a multilayer co-occurrence network, we connected between instances of the same ASV across layers (farms) using interlayer links (Fig. 1C). Interlayer links have been applied multiple times in ecological multilayer networks and are the major advantage of using multilayer networks [25,46–49]. We calculated the value of an interlayer link  $w_i^{\alpha\beta}$  connecting an ASV  $i$  to itself between layers  $\alpha$  and  $\beta$  as the UniFrac similarity between the partners of  $i$ . We preferred UniFrac to Jaccard because it better represents microbe biology. Note that the value of an interlayer link is not the value of  $P_i^U$  because  $P_i^U$  is averaged across all the layers in which  $i$  occurs, whereas the interlayer link is calculated per two layers. Interlayer links were undirected ( $w_i^{\alpha\beta} = w_i^{\beta\alpha}$ )

The interlayer link definition encodes the extent to which local vs global processes affect local co-occurrence and maps local co-occurrence patterns to the meta-community level. If the node has identical co-occurrence partners, the interlayer link is 1, giving high weight to global processes. Alternatively, if a microbe tends not to conserve its co-occurring partners, the interlayer link value approximates 0, reflecting that mainly local processes determine co-occurrence (if all interlayer links are 0, then farms are completely disconnected). Using a similarity measure for both the intra- and interlayer links places them on the same scale (bounded 0 to 1; Fig. 1D), ensuring that the modularity algorithm is not a-priori biased towards intralayer or interlayer processes [46,50].

## 2.8 Modularity analysis

We used Infomap via the R package ‘infomapecology’ [50] to capture the modular structure of the meta-co-occurrence multilayer network. Infomap is a flow-based algorithm designed explicitly for multilayer networks. The advantage of using an algorithm dedicated to multilayer networks is that it allows detecting modules that cross layers [25,50] and, therefore, contain microbes from different farms. Hence, a module is a group of ASVs that co-occur more with each other than with other ASVs within and across farms.

Briefly, Infomap detects the optimal network partition based on the movement of a random walker on the network (see [50–52] for details). For any given network partition, the random walker moves across nodes in proportion to the weight of the edges. The amount of information it costs to describe the walk is quantified using the objective function  $L$  called the map equation. The optimal network partition is the one that minimizes  $L$  [51]. In multilayer networks, nodes representing observable entities such as ASVs are called *physical nodes*, and nodes describing the occurrence of ASVs in the layers are called *state nodes*. The random walker moves from state node to state node within and across the layers on

intra- and inter-links, respectively. Therefore, the multilayer network representation is not merely an extended network with unique nodes in all layers, and a module can encompass multiple farms. In addition, a physical node (ASV) can be assigned to different modules in different layers.

We tested for statistical significance in modularity by comparing the value of the observed network to those obtained from 500 shuffled multilayer networks. We recalculated the interlayer links for each set of seven network layers from the 500 shuffled iterations to create a shuffled multilayer co-occurrence network.

## 2.9 Analysis of cow population genetics

We analyzed population genetic structure for cows for which we had genetic and microbiome data (935 out of 1012 cows: 741 Holsteins and 194 Nordic Reds). The data set had 67785 shared SNPs. First, we computed a genetic similarity matrix. We used a frequency-weighted allele-sharing similarity (FWASS) approach, in which a greater weight is assigned to shared rare alleles than to shared common alleles when computing genetic similarity between a given pair of individuals [53,54]. We then used the genetic similarity matrix to compute hierarchical population genetic structure by computing hierarchical clustering using the Ward method.

## 2.10 Code and data

All data and code will be available upon acceptance.

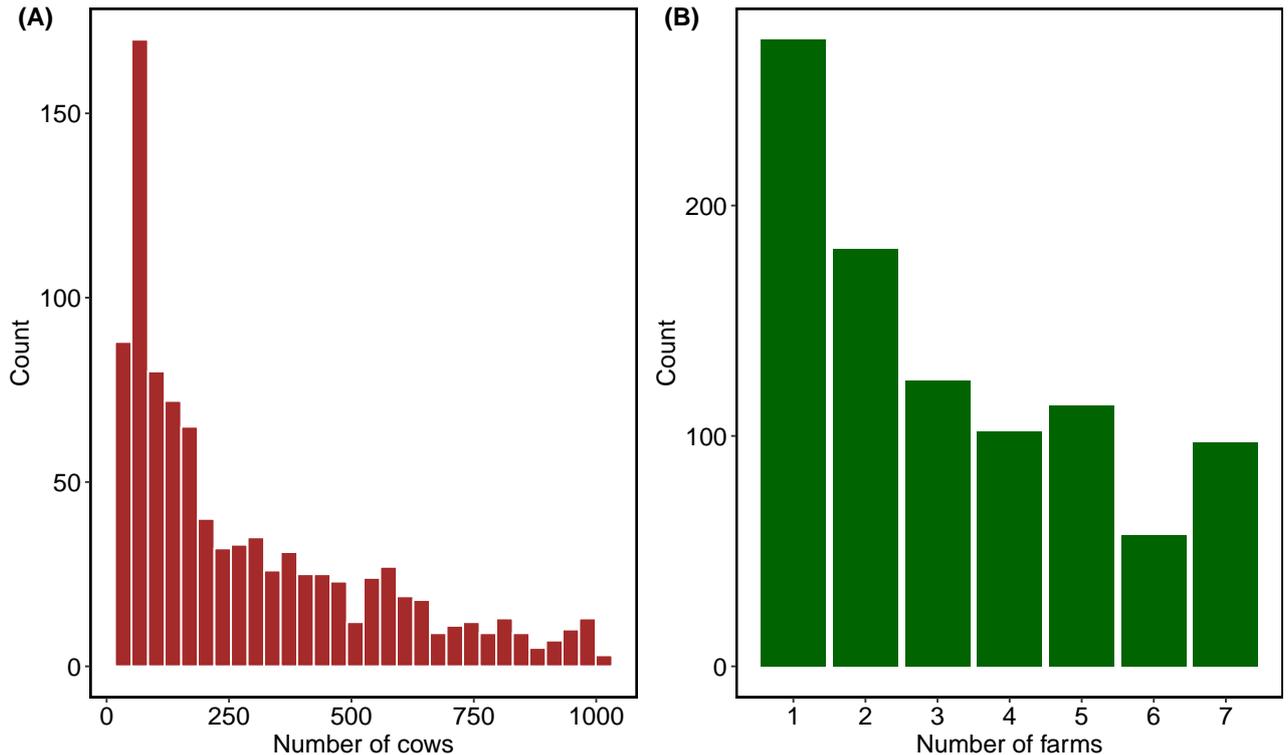
# 3 Results

## 3.1 Microbe diversity

Our threshold criteria resulted in 946 microbes (931 bacteria and 15 archaea) across the 1012 cows. The vast majority belonged to the Bacteroidetes (52.7%) and Firmicutes (27.8%) phyla. At the family level, microbes belonged mostly to Prevotellaceae (37.6%), followed by Lachnospiraceae (12.9%) (Fig. S1). The number of microbes ranged from 42 to 505 per cow across farms (mean±sd: 269.7±104; Figs. S2,S3). Farms varied in the number of microbes and cows they contained (Table 1). In addition, the occurrence of microbes in farms and cows was skewed with most microbes occurring in few farms and cows and a tail of microbes that occur across all the meta-community (Fig 2).

**Table 1: Summary statistics of farms.** a: number of cows; b: number of microbial ASVs (richness); c: mean number [range] of ASVs per cow; d: number of co-occurrence links. Density is calculated as the number of links divided by the number of possible links:  $2L/(S(S-1))$ .

Farm	Cows <sup>a</sup>	ASVs( <i>S</i> ) <sup>b</sup>	ASVs per cow <sup>c</sup>	Links <sup>d</sup> ( <i>L</i> )	Density
FI1	100	329	185.8 [88-300]	23298	0.49
SE1	96	461	259.4 [42-389]	36783	0.35
IT1	185	307	179.8 [89-292]	20465	0.45
IT2	176	566	349.6 [129-505]	63310	0.42
IT3	48	217	125.8 [73-188]	2776	0.13
UK1	243	490	280.3 [100-442]	71101	0.61
UK2	164	630	369.1 [145-498]	51551	0.27



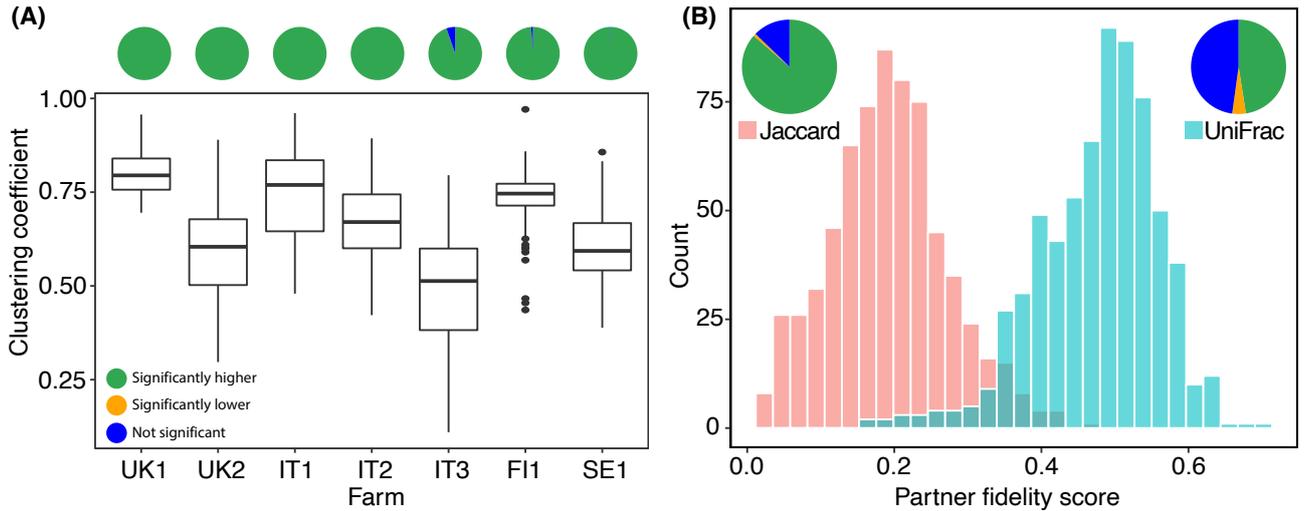
**Fig. 2: Microbe distribution across cows and farms is skewed.** The histograms depict the number of cows (A) and farms (B) in which microbes occur. Most microbes occur in a few cows while a few (the tail of the distribution in (A)) occur in almost all cows. Over 70% occur in at least two farms.

### 3.2 Microbe co-occurrence is transitive and repeatable across farms

We first investigated co-occurrence at a local scale (within farms). The mean  $CC_i$  ranged from 0.497-0.8 in the seven farms (Fig. 3A). Transitivity was significantly larger than random for the vast majority of microbes in all farms (Fig. 3A). Transitivity is a local signature of co-occurrence but 674 out of 946 microbes ( $\approx 70\%$ ) occur in more than one farm (Fig 2B). We therefore asked if the occurrence of microbes in different farms leads to repeatable patterns of co-occurrence, which we quantified using partner fidelity. We first used Jaccard similarity as a partner fidelity score ( $P_i^J$ ). The mean  $P_i^J$  across the microbes that occurred in two or more farms was  $0.19 \pm 0.08$  (mean $\pm$ sd) with a maximum of  $\approx 0.4$  (Fig. 3B). In 86.4% of microbes  $P_i^J$  was higher than expected from a random distribution of microbes across cows. Considering phylogeny (using  $P_i^U$ ) increased partner fidelity, as expected. The mean  $P_i^U$  score was  $0.477 \pm 0.08$ . Comparison to shuffled networks showed that almost 50% of microbes (284 out of 649) tended to have smaller genetic distances between their co-occurring partners across farms than expected at random (Fig. 3B). The higher values and lower proportion of significant partner fidelity in  $P_i^U$  are expected because the probability of conserving the same ASV labels (used in  $P_i^J$ ) is low due to the high number of ASVs.

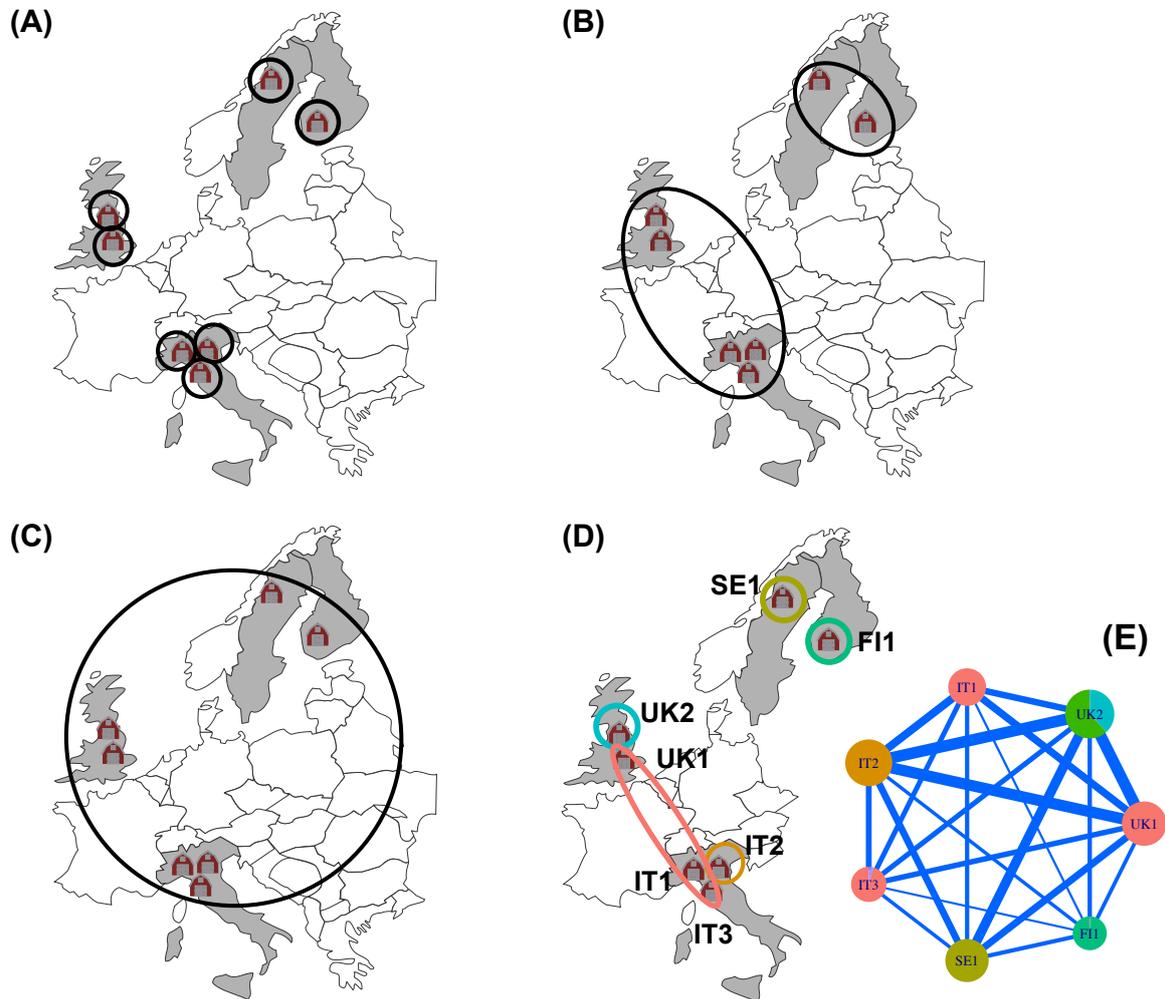
### 3.3 The meta-co-occurrence network contains a mix of local and non-local signatures

Partner fidelity indicates that patterns of co-occurrence can be determined and detected at a non-local scale. To determine the scale at which signatures of co-occurrence are most prominent, we used modularity analysis to detect clusters (modules) of highly connected microbes. Our alternative hypotheses are as follows. If local processes dominate, then modules will correspond to local farms



**Fig. 3: Microbe transitivity and partner fidelity are higher than expected by random. (A)** Box plots depict the values of the clustering coefficient  $CC_i$  across all microbes within a farm. **(B)** Distributions of 649 microbes' partner fidelity calculated with Jaccard ( $P_i^J$ ) and UniFrac ( $P_i^U$ ). The pie charts in both panels depict the proportion of microbes whose  $CC_i$ ,  $P_i^J$  and  $P_i^U$  is significantly greater (green), lower (orange) or non-significant (blue). Significance was calculated with a z-score test comparing the observed index of each microbe to that obtained from 500 shuffled networks of each farm (see Methods for the description of the null model). Microbes are predominantly involved in transitive co-occurrence within each farm (A) and maintain their co-occurrence partners between farms (B) more than expected at random.

(Fig. 4A). At the other extreme, if global factors dominate, then the network will have a single module that encompasses all microbes in all farms (Fig. 4C). This scenario is likely because some microbes occur in all farms, linking them (Fig. 2B). Between these two extremes, other partitions are possible. Specifically, if host genetic background is a strong determinant of co-occurrence then we should expect two modules (Fig. 4B), corresponding to the northern (SE, FI) and southern (UK, IT) farms because they differ in cow breed (Table S1). Testing these alternative hypotheses requires a mathematical object that separates network farms on the one hand, and connects them on the other. We solved this problem using a multilayer network, which represents a meta-co-occurrence [25,46] (Fig. 1C).

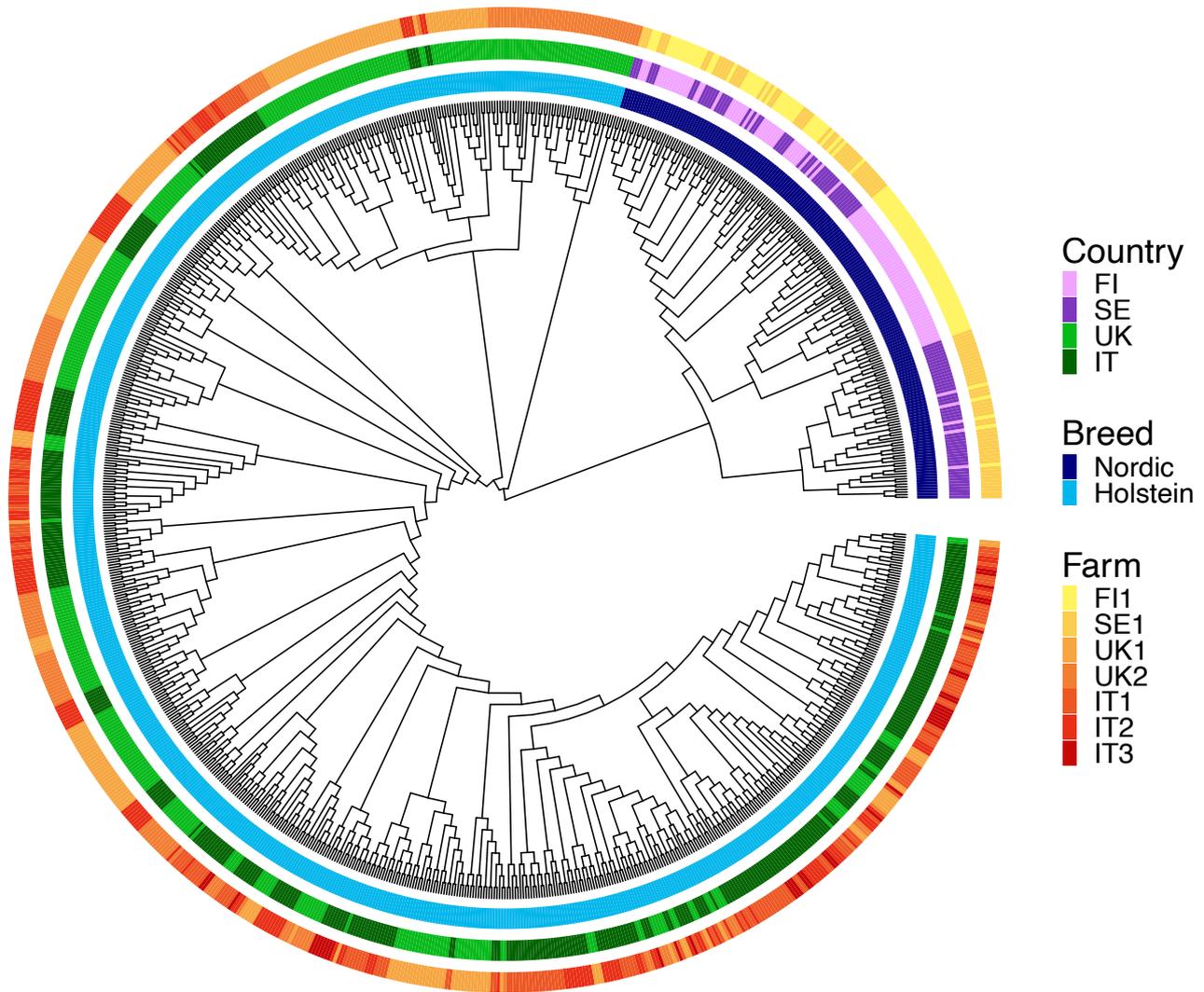


**Fig. 4: Hypotheses for modularity analysis.** Farms are depicted with icons and modules with colored circles. **(A)** Local processes dominate. In this hypothesis, each farm (layer in the network) will be assigned to a module of its own. **(B)** Host genetic background determines modularity. The two modules correspond to two cow breeds (Methods, Table S1). **(C)** Global processes determine co-occurrence. In this scenario, all the farms are assigned to a single module that encompasses all the network. **(D)** The results of the analysis show a mix of these hypotheses. Colored circles are modules. While the Swedish and Finnish farms are separated from the Italian and Uk farms, they are also each in a module of its own. Moreover, UK2 is in its own module. However, two Italian farms and one UK farm are connected. **(E)** The layer-level network. Each node is a farm and its size is proportional to the number of microbes occurring in the farm (Table 1). Links indicate that at least one microbe connects the two farms via an interlayer link. Link width is proportional to the number of interlayer links in the multilayer co-occurrence network. The nodes are depicted as pie charts that detail the percentage of microbes in that farm in each module, while every color is a different module. Module colors in E correspond to those in D.

The meta-co-occurrence network contained 10 modules. Four of these modules contained only 0.48%-2.86% of nodes within a farm and therefore presented an extremely weak signal. To focus on the strongest signal, we removed modules that contained less than 3% of the microbes within a farm (Fig. S4). With the exception of UK2, virtually all the microbes in a given farm belonged to the same module (Fig. S4). Hence, in UK2 strong local processes may generate structure within the farm. Comparison to shuffled networks showed that the modular structure was significantly non-random. More than 99% of the shuffled networks were consistently partitioned into a single large module that contains all farms. Furthermore, partitioning of the empirical network provided a more optimal solution than shuffled networks ( $P < 0.002$ , Fig. S6).

The partitioning of the observed network into modules represented a mix of our alternative hypotheses. Farms SE1, FI1, IT2, and UK2 did not share modules with other farms, suggesting that local processes dominate co-occurrence structure. However, IT1, IT3 and UK1 were clustered in a single module. The assignment of two farms to the same module is likely when they share microbes. A follow-up beta-diversity analysis indicated that this is not the case (Fig. S5). For example, although UK1 and UK2 are in separate modules, they share many microbes (UniFrac similarity 0.73; see link weights in Fig. 4E). Therefore, although farms do share many microbes, they still belong to different modules, indicating that local processes generate structure.

To test the hypothesis that the modular structure is determined by the breed of the cow (Fig. 4B) we analyzed the population genetic structure of the cows. The hierarchical population genetic structure corresponded well, though not perfectly, with our hypothesis (Fig. 5). *At level 1*, the two breeds were clearly sorted into two clusters, with all the Nordic Reds from Finland and Sweden forming one cluster, and all Holsteins from both the UK and Italy forming a second cluster. This level corresponds to the module partitioning between the northern (SE, FI) and southern (UK, IT) farms. *At level 2*, we found an additional structure by country of origin within the Nordic Reds: 95/100 of the individuals from Finland and 46/94 individuals from Sweden formed one cluster, and 48/94 Swedish cows together with the remaining five Finnish cows formed a second cluster. However, the structure at this second level was less prominent and can only partially explain the separate modules of SE1 and FI1. In addition, there was no clear signature for separation between UK and IT at this level. *The lower levels of the hierarchy (levels 3-11)* show increasingly fine-scale population genetic structure within each breed, but without correspondence between the genetic structure and farm. Therefore, cow genetic population structure does not explain why microbes from UK2 and IT2 belong to a single module and why microbes in IT2 and UK2 did not share modules with the other farms in UK and IT.



**Fig. 5: Cow genetic population structure.** The genetic similarity was calculated using SNPs data (see Methods), followed by clustering using Ward hierarchical clustering according to the genetic distance. Clustering was performed on 935 cows for which both SNP and microbiome data was available. Each leaf in the tree is a cow and colors indicate the cows' breed, country and farm.

#### 4 Discussion

Highly diverse microbial communities are often studied using co-occurrence networks because direct physical interactions are impossible to observe. Nevertheless, the relative contributions of local and non-local processes to the structure of co-occurrence networks still need to be studied. We addressed this challenge using a meta-cooccurrence network analyzed as a multilayer network. We showed that co-occurrence patterns were repeatable across farms, with a strong phylogenetic component. We further found a non-random modular signature determined by a mix of local and non-local processes. At a local scale, co-occurrence was nonrandomly transitive in virtually all microbes regardless of the farm. Although several studies have quantified transitivity, its interpretation has been vague [6,16,55]. From a statistical perspective, a transitivity value is only informative with proper comparisons with a null model. We solved the latter issue by comparing the transitivity of each microbe to random expectations. The high and non-random transitivity we found results partly from a high core microbe threshold because a high threshold includes few microbes that occur in many cows (Fig. S9). However,

this is not the only explanation. From a biological perspective, transitivity can result from multiple mechanisms that are difficult to disentangle but could serve as a hypothesis for further studies. The coexistence of three microbes (transitivity is triangle closure) can result from shared habitat preferences (environmental filtering). This is likely because, within a farm, cow hosts were genetically similar (same breed) and fed the same diet. However, theory predicts that when two microbes depend on metabolites produced by a third one, they will compete with each other [6,56]. Therefore, faster-growing bacteria should prevail without stabilizing mechanisms that maintain coexistence. Our results suggest that stabilizing mechanisms, such as fluctuating resources, niche separation, and host-pathogen dynamics, allow microbe coexistence. Identifying these mechanisms and their interplay with environmental filtering is currently an open question.

At the non-local scale, a focal microbe tends to maintain phylogenetically similar partners across farms. About half of the microbes conserved phylogenetically similar partners, more than expected from a random occurrence of microbes in cows. There are two non-mutually exclusive explanations for this pattern. First, if biotic interactions determine co-occurrence, a phylogenetic signal would appear among a microbe’s partners across farms because it maintains partners that fulfill similar functions (assuming that phylogeny is a proxy for function). Second, although the farms are located in different countries, there are similar environmental conditions across all farms that filter similar microbial functions that are cohesive with phylogenies and, therefore, enforce the occurrence of phylogenetically similar microbes.

The structure of co-occurrence networks was shown to depend on environmental conditions [57–60]. However, previous studies compared disconnected networks assembled from samples originating in different environments. In contrast, we leveraged a multilayer approach to test for multiple alternative hypotheses. Unraveling signatures of scale-dependent processes in a meta-co-occurrence network can only be done using a multilayer network with interlayer edges because unless nodes are linked between layers, any modularity algorithm will analyze each layer independently, not being able to detect modules that span layers [50,52]. Although farms were similar in their microbial composition, they were still partitioned into different modules. Therefore, local processes have a substantial effect on modularity. This result is in line with a previous study that showed mixed effects of climate and local interactions on microbial communities of the pitcher plant *Sarracenia purpurea* [60]. Here, host genetics can serve as an environmental filter for microbes, determining occurrence [19]. We showed that the modular structure of the meta-co-occurrence network can be partially explained by cow population genetic structure. Specifically, module separation between the northern (SE1 and FI1) and the southern farms and between the northern farms corresponded to cow population genetic structure. One confounding factor is that the farms in the UK and Italy received maize silage, while the farms in Sweden and Finland received grass silage. Nevertheless, while diet type could affect microbiome composition, this is highly unlikely because the diet was nutritionally equivalent in all farms. Finally, the module separation between FI1 and SE1 could be explained by different cow housing conditions in these two farms. Other factors, such as climate, can also influence microbial composition via effects on the host environment. For instance, in a controlled experiment, the structure of a microbial network was affected by environmental temperature [16].

Our study has several limitations. First, we used presence-absence data. We opted not to use reads as abundance estimates due to known limitations [61]. Nevertheless, microbes with low prevalence but

high occurrence could influence co-occurrence structure. Second, we used a threshold to define the ‘common core’ [28]. While such a threshold is necessary and common practice in co-occurrence studies [28], it may influence the observed patterns. We performed a sensitivity analysis and detected a quantitative effect of threshold choice. However, the effect of the threshold was qualitatively predictable (see discussion in SI). It is also worth noting that this study was conducted on endo-symbiotic microbes by using data sampled from a mammalian gut. Studying samples of free-living microbes could reveal a different structure and is a relevant direction for further research.

Despite these limitations, our study significantly contributes to community and microbial ecology. First, we demonstrate an approach to creating and analyzing the structure of a meta-co-occurrence network with an informative set of alternative hypotheses. This approach is not limited to co-occurrence networks. Future studies can use it to explore local and non-local signatures in species interaction (meta)networks or metacommunities. Identifying such signatures remains little studied [12]. Second, understanding the mechanisms underlying microbe distribution and interactions is an open question in microbial ecology. We discovered local and non-local signatures in co-occurrence structure, which were linked to the population genetic structure of the animal host. Finally, the methodology and results presented here are relevant to community ecology, where community assembly theory is mainly developed for macroscopic organisms.

## **Acknowledgments**

We thank Prof. Itamar Giladi and Rohit Sahasrabudde for valuable comments.

## **Conflict of interest**

The authors declare no conflict of interest

## **Ethics statement**

This article does not present research with ethical considerations.

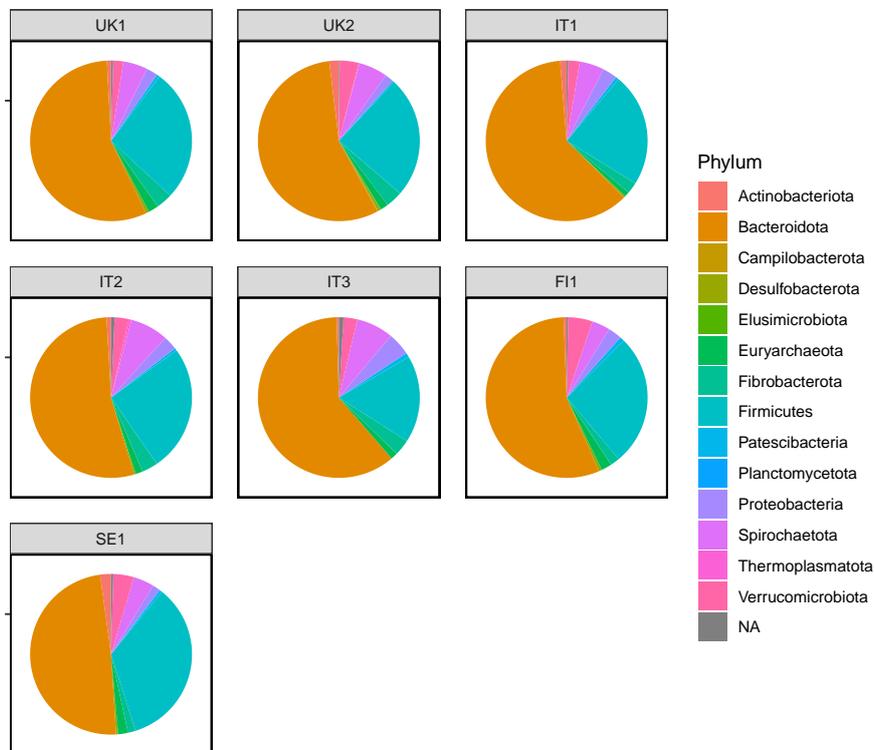
## **Funding statement**

SP was supported by the Israel Science Foundation (ISF grant 1281/20) and Human Frontiers Science Program (HFSP award number RGY0064/2022). IM was supported by the Deutsche Forschungsgemeinschaft (German-Israeli Project Cooperation, DIP, 2476/2 -1), ERC (866530) and ISF (1947/19).

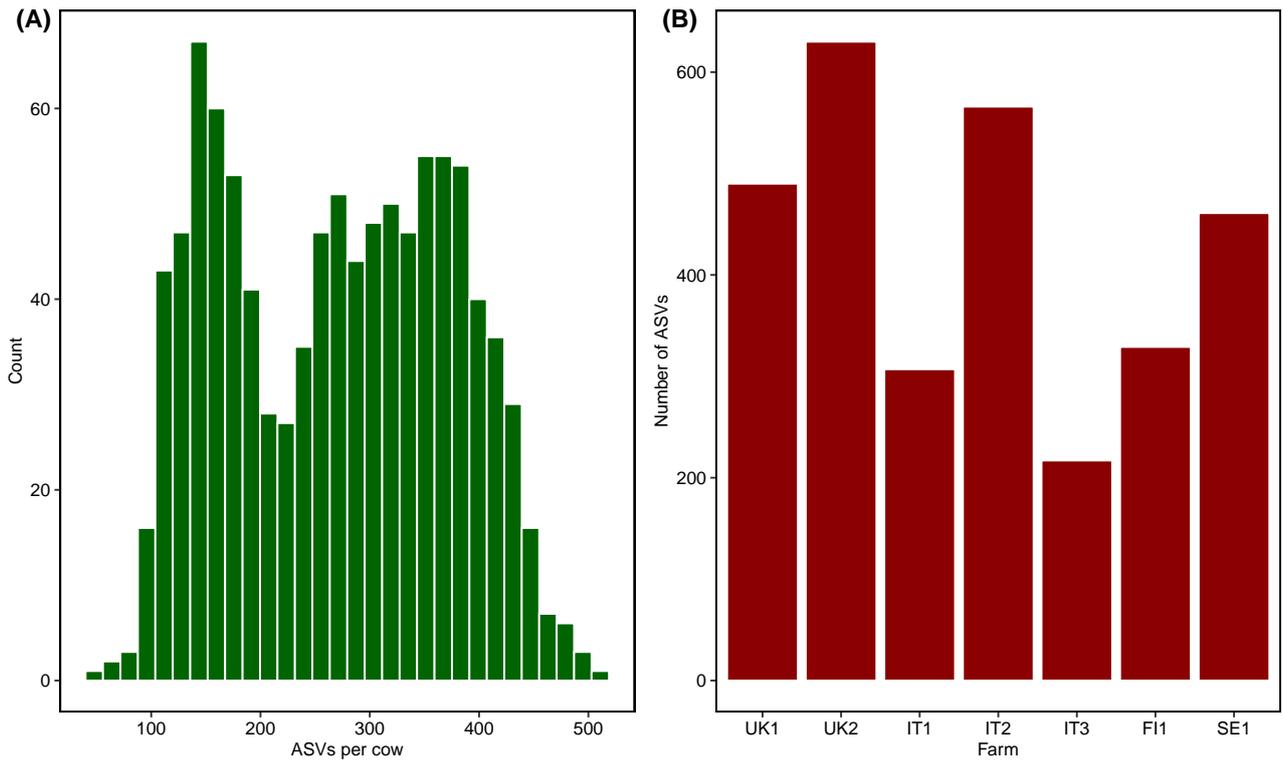
## Supplementary Information

**Table S1: Details of Wallace et al.'s [27] study design.**

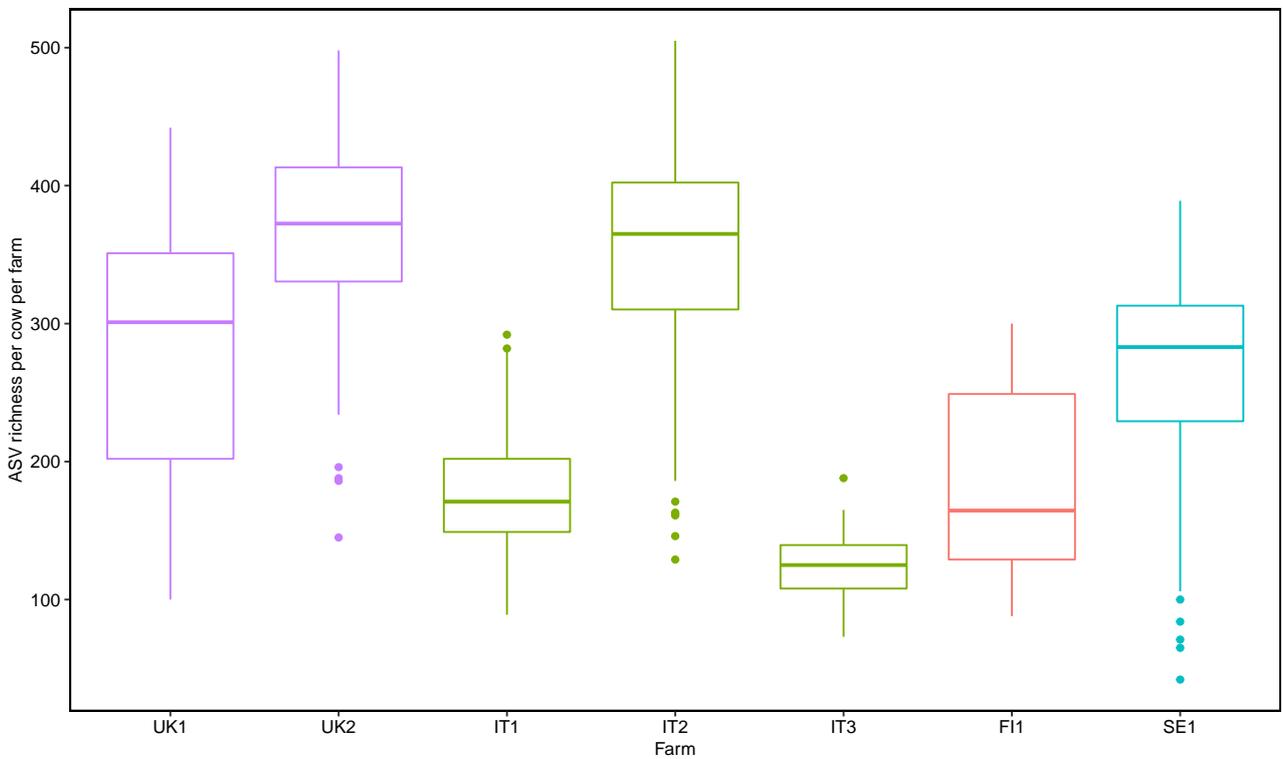
Farm	Breed	Diet	Housing
UK1 UK2 IT1 IT2 IT3	Holstein dairy cows	Maize silage, grass silage or grass hay, and concentrates	Group-housed
SE1 FI1	Nordic Red dairy cows	Grass silage and concentrates	Group-housed Individual standings



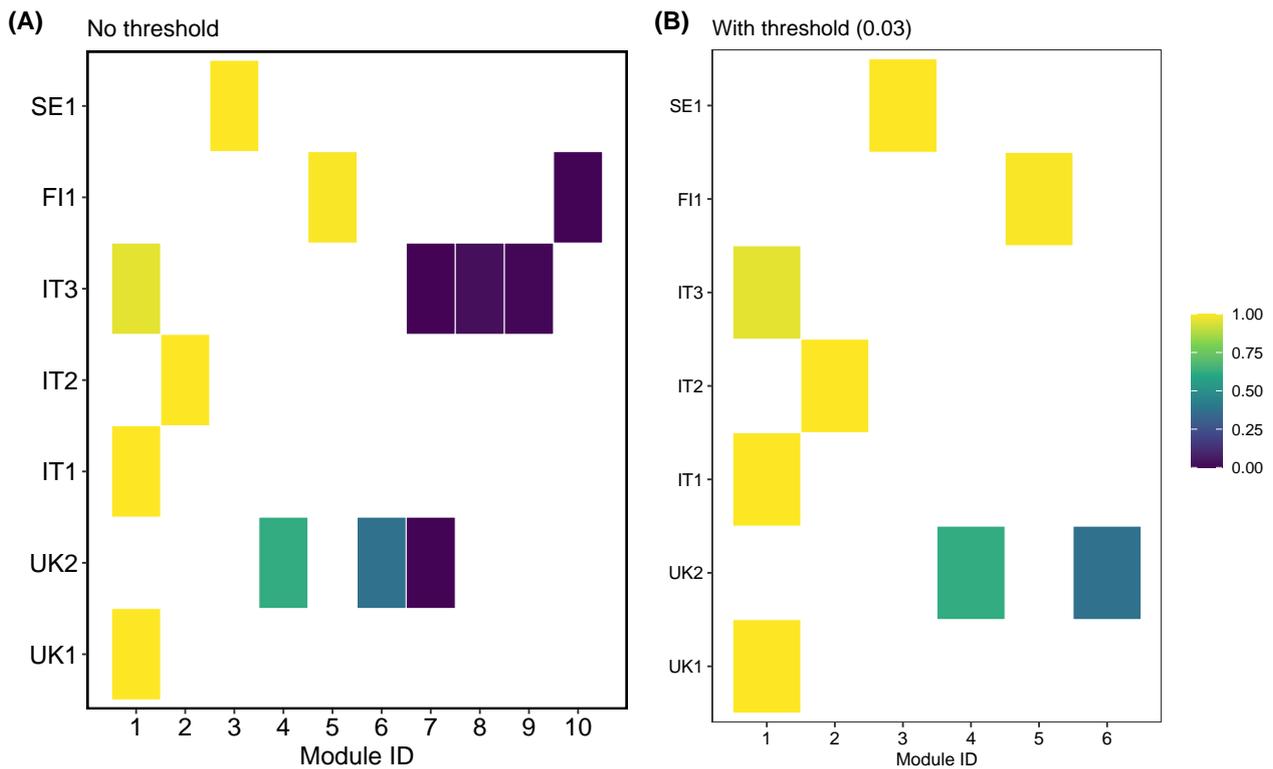
**Fig. S1: ASV taxonomy dominated by Bacteroidetes and Firmicutes.** The relative richness—number of ASVs that belong to a certain Phylum out of all ASVs—in each farm. The two largest phyla are Bacteroidetes and Firmicutes.



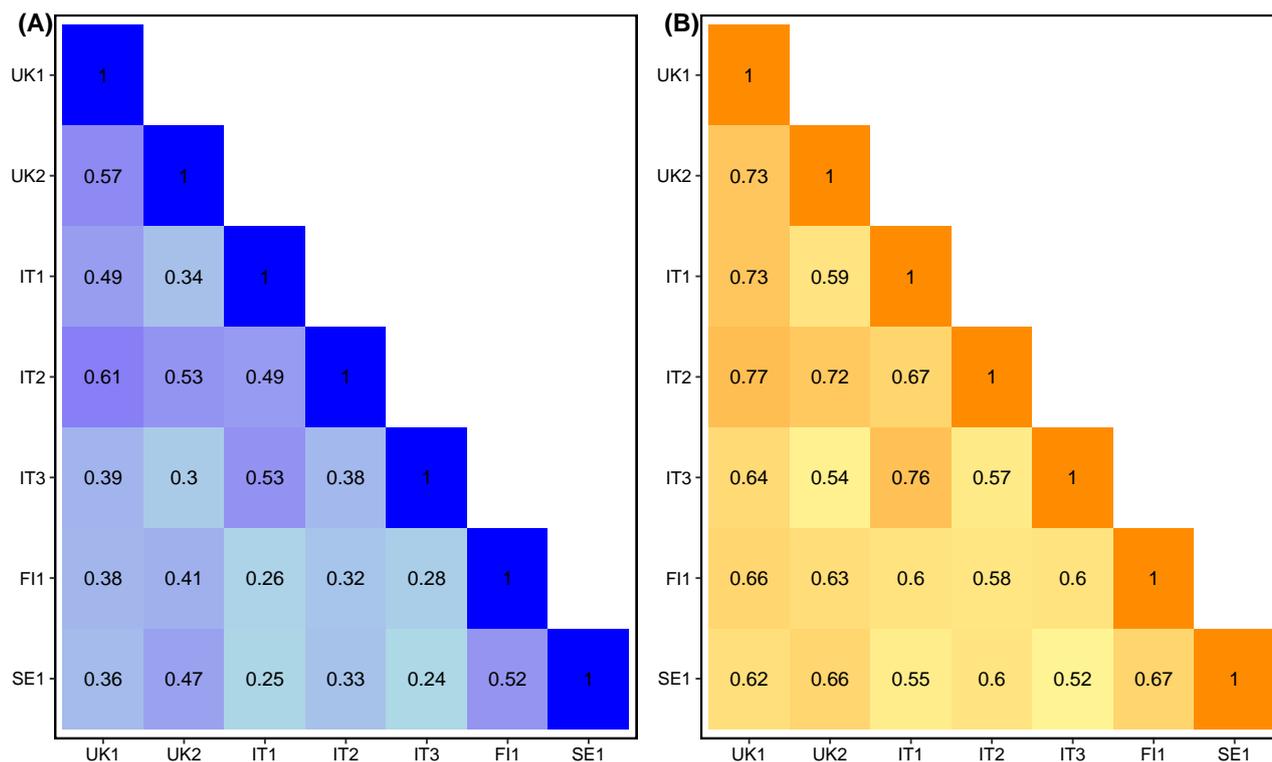
**Fig. S2: ASV distribution.** (A) A histogram showing the number of ASVs per cow (across farms). (B) Bar plots of the number of ASVs per farm (across cows).



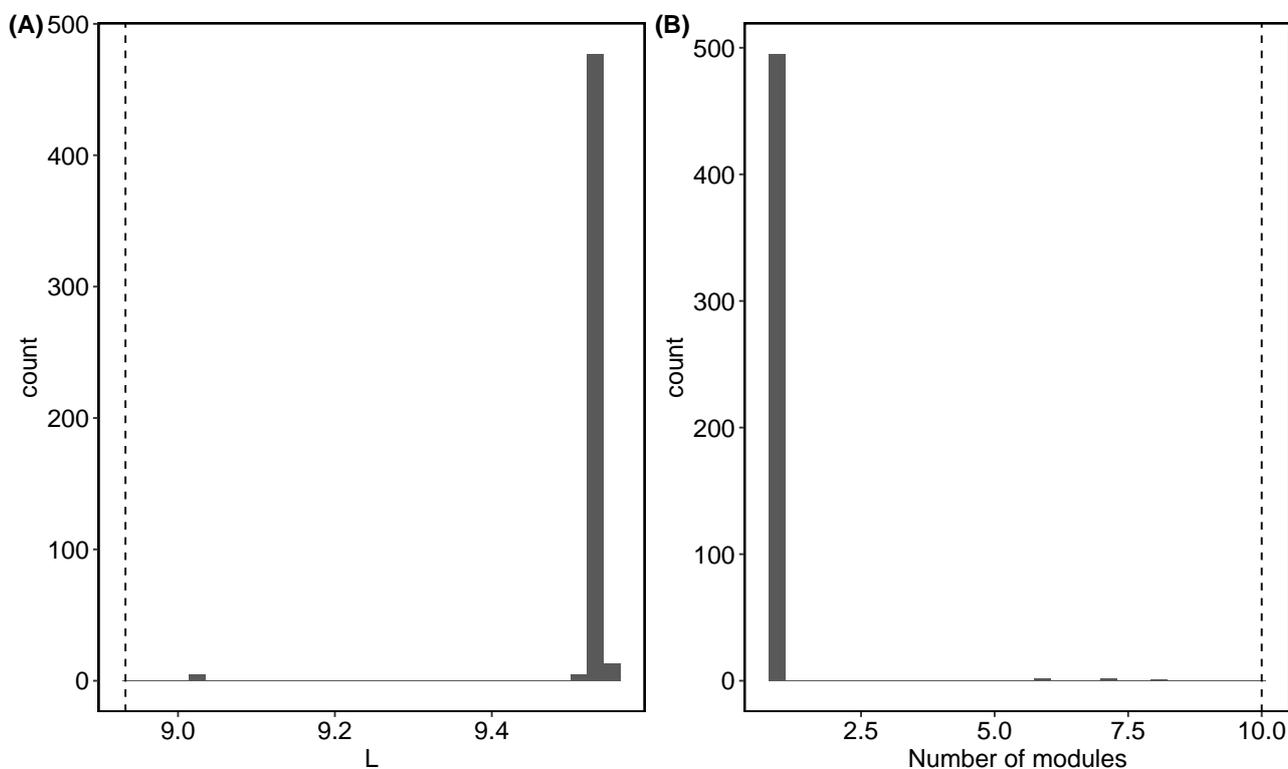
**Fig. S3: ASV richness.** Box plots of the distribution of number of ASVs in cows, within each farm. Colors depict countries.



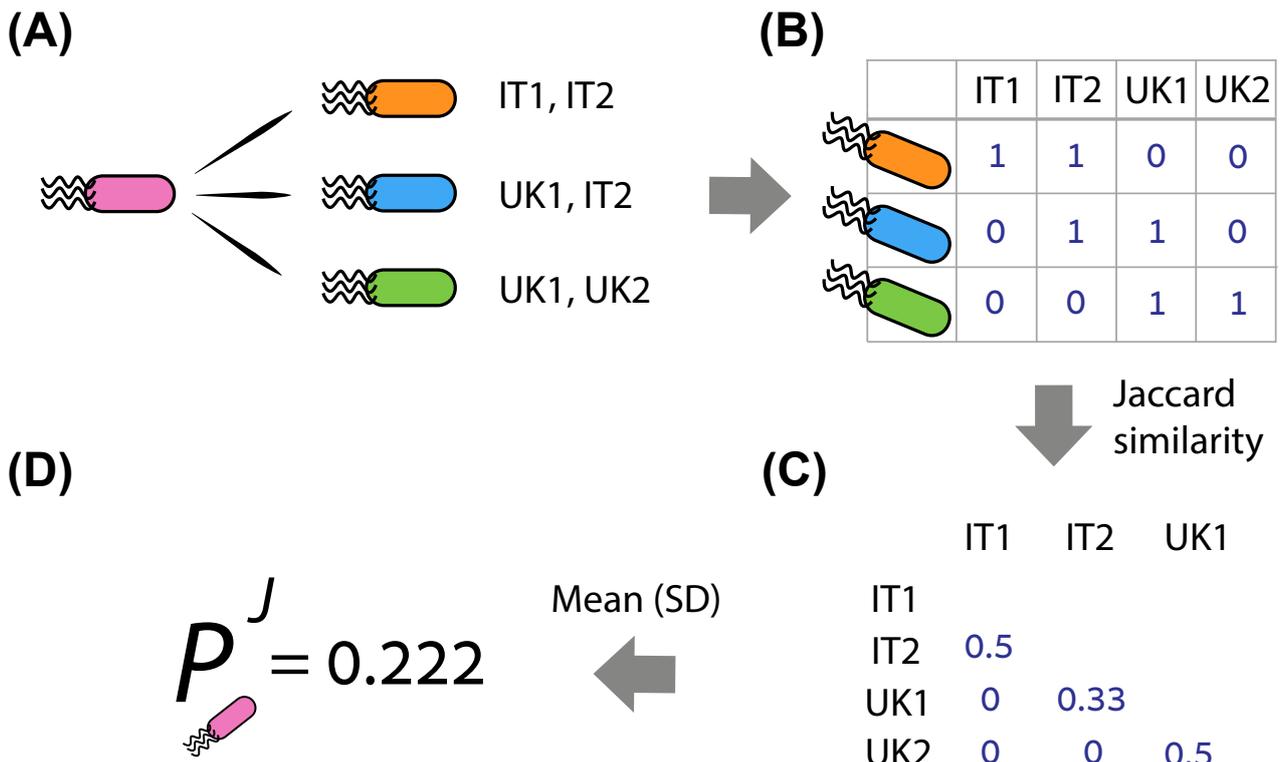
**Fig. S4: Partition of the multilayer network to modules.** Each matrix shows the module IDs (x-axis) to which nodes from a given farm (y-axis) belong. Color scale corresponds to the proportion of ASVs from each farm that were assigned to a module (rows sums to 1). **(A)** All the modules. For example, IT3 was partitioned to five modules (1,2,7,8,9), 3 of which contain very few nodes (the purple ones). **(B)** The same modules, after removing modules that contained less than 3% of the microbes within the farm. Removing this noise clarifies that almost all the nodes in SE1 and in FI1 were included in separate modules (one per farm).



**Fig. S5: ASV similarity across farms.** The similarity between farms in ASV composition, calculated with Jaccard **(A)** and Unifrac **(B)**.



**Fig. S6: Modularity results are significantly different than random.** **(A)** The value of the objective function  $L$  was consistently lower in the observed network compared to 500 shuffled ones, indicating that it was more modular than expected from a random distribution of microbes in cows. **(B)** The shuffled networks produced predominantly a single module. Dashed line indicates the values in the observed network.



**Fig. S7: Partner fidelity calculation process.** Partner fidelity calculation is done for each microbe separately by comparing the set of co-occurring partners it has in one layer (farm) to the set of partners it has in another layer using a beta-diversity index. (A) In this diagram partner fidelity ( $P_i^J$ ) was calculated for the pink microbe by first (B) identifying the partners it has in each layer, then (C) calculating for each pair of layers how similar are the partners in one layer compared to the other, using Jaccard and UniFrac. These UniFrac Values were used as interlayer edges to connect (in this case) the pink microbe in one layer to itself in the other layer for each pair of layers (farms). (D) Lastly, the mean value of all pairwise similarities across layers was calculated and defined as the partner fidelity of the pink microbe.

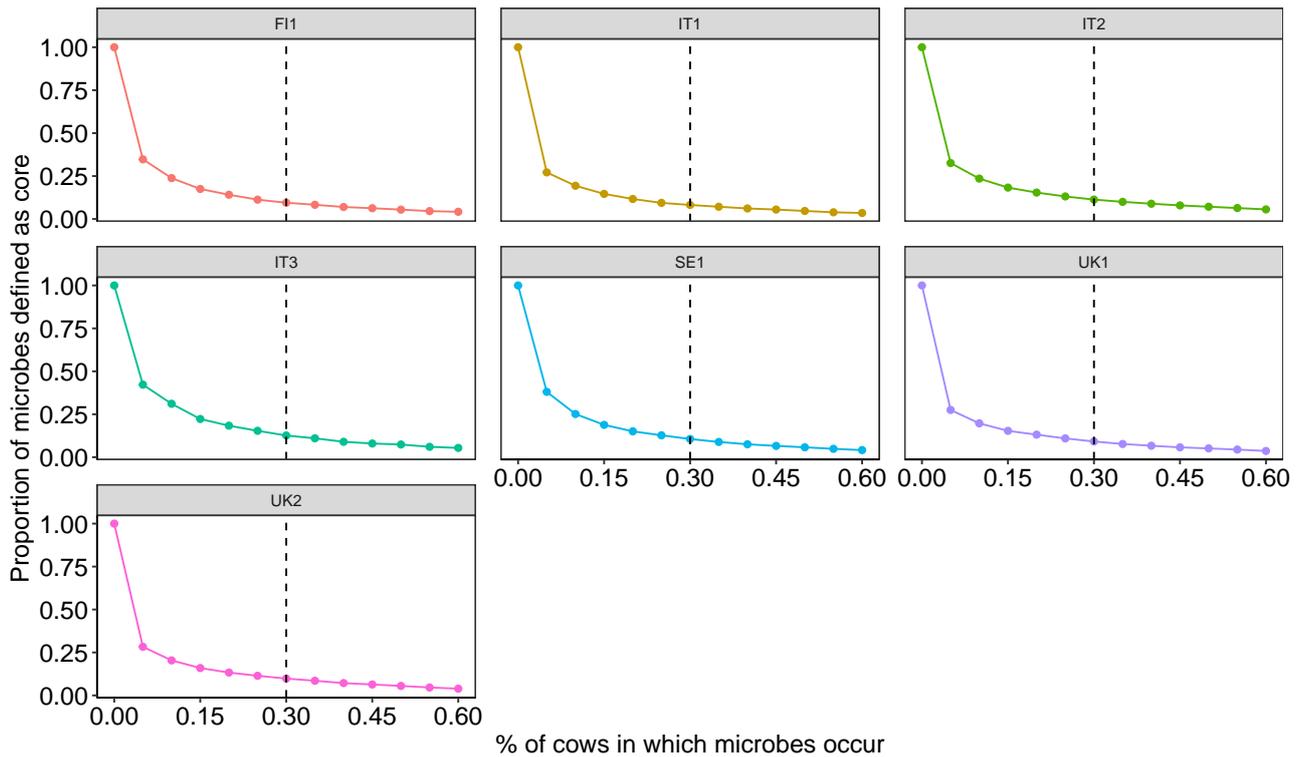
## S1 Sensitivity to thresholds of core microbes

To focus on the signal produced by the key microbes in the community, we used the common core microbes in each farm [27,28]. We did not work with abundance or relative abundance due to known issues with using reads as abundance estimates [61]. Therefore, core microbes were defined as ASVs occurring in a certain proportion of cows within each farm (the ‘common core’ definition of [28]). A general problem in deciding on a relevant threshold for setting core microbes is the trade off between environmental adaptability to biological variation [28]. On the one hand, including too many ASVs can skew the signal because typically most microbes are rare or occur in too few samples (here, cows). In that case, network structure may not be representative of the relevant gut biota because low-prevalence microbes may have little effect on community function. Moreover, computing co-occurrence networks is extremely extensive, even when run on High Performance Computing systems. Including too many ASVs may restrict the amount of analyses we can practically perform. On the other hand, including too few ASVs will cause loss of ecologically meaningful variation. There is obviously no one true answer to this problem. Our approach was therefore to start with prior knowledge of this data set. In the original paper, Wallace et al. [27] defined core microbes as those occurring at  $\geq 50\%$  of cows within a farm, remaining with 454 ASVs out of  $\approx 12,000$  (3.7%). They showed that even at this strict threshold, the remaining microbes can strongly predict the cow phenotype (e.g., methane emission, lactation). Therefore, this can be considered as the ‘functional core’ [28].

High-occupancy typically suggests that microbes are well adapted to their environment and carry functions that lead to high prevalence [28]. We therefore opted not to deviate drastically from the functional core because it is biologically meaningful. Nevertheless, a 50% threshold is extremely strict as it includes only 3.7% of ASVs. Additionally, the nature of the question we asked requires more variation because the functional core, being so restricted, will be highly repetitive across cows and farms. Therefore, we conducted a sensitivity analysis measuring the percentage of ASVs present in cows in each farm as a function of different threshold levels (Fig. S8). While the sharpest decline was at 5%, it was a large deviation from the functional core. In addition, this threshold had too many bacteria for practical analysis that included shuffled networks. Given that after this drop, the curve did not change drastically, we opted to work with a threshold of 30%, remaining with 946 ASVs. This allowed us to relax Wallace et al.’s [27] definition while doubling the number of microbes in the analysis. In addition, we repeated the analyses (without the shuffled networks) for the 5%, 10% and 20% threshold levels as described hereafter.

Sensitivity analysis showed that, qualitatively, the general pattern of  $CC_i$  was maintained across farms. Lower threshold include more microbes, lowering the probability that microbes will close triangles. Therefore, within each farm,  $CC_i$  increased with threshold (Fig. S9).

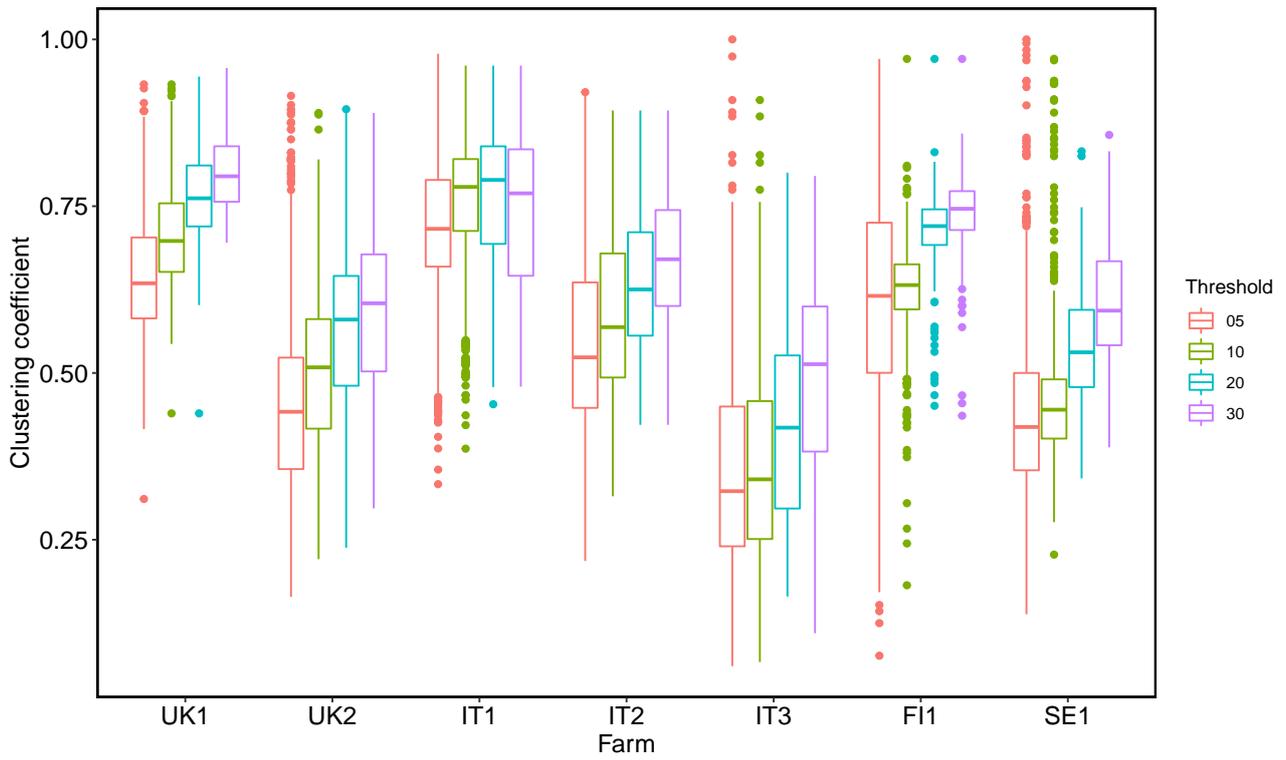
The qualitative differences between  $P_i^J$  and  $P_i^U$  was maintained as with the 30% threshold. That is, UniFrac had generally higher similarity than Jaccard. (Fig. S10). Including more microbes (lower thresholds) increases the number of intralayer edges compared to interlayer edges (Fig. S11A-C). This is because decreasing the threshold increases the number of microbes that occur in a single or few farms. These microbes have intralayer edges but not interlayer edges (which connect microbes to themselves). Quantitatively, these intralayer edges become weak compared to the interlayer edges because by definition these microbes occur in a few cows (that is the meaning of the threshold). The combination of many weak intralayer edges with few stronger interlayer edges results in many small



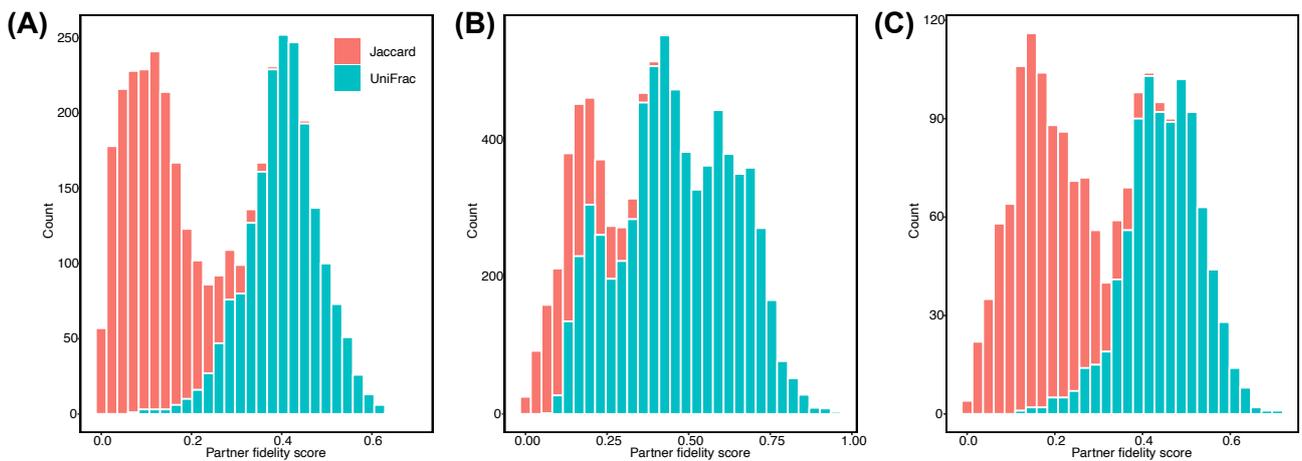
**Fig. S8: Sensitivity analysis for core microbes.** Within each farm we included ASVs that occurred in a given proportion of the cows (x-axis), and quantified the proportion of ASVs that were defined as core microbes. The vertical dashed line marks the threshold with which we work in the main text.

modules within each farm and a single large module between farms (Fig. S11D-F). This module by definition contains those microbes that occur in all farms, which are likely to be the more prevalent microbes (those included in the analysis in the main text). This single module indicates global processes. These are likely the processes that effectively determine the ‘functional core’. Note that in the 20% threshold, the farms are start to show tendency to be separated from the rest, resembling the results in the main text.

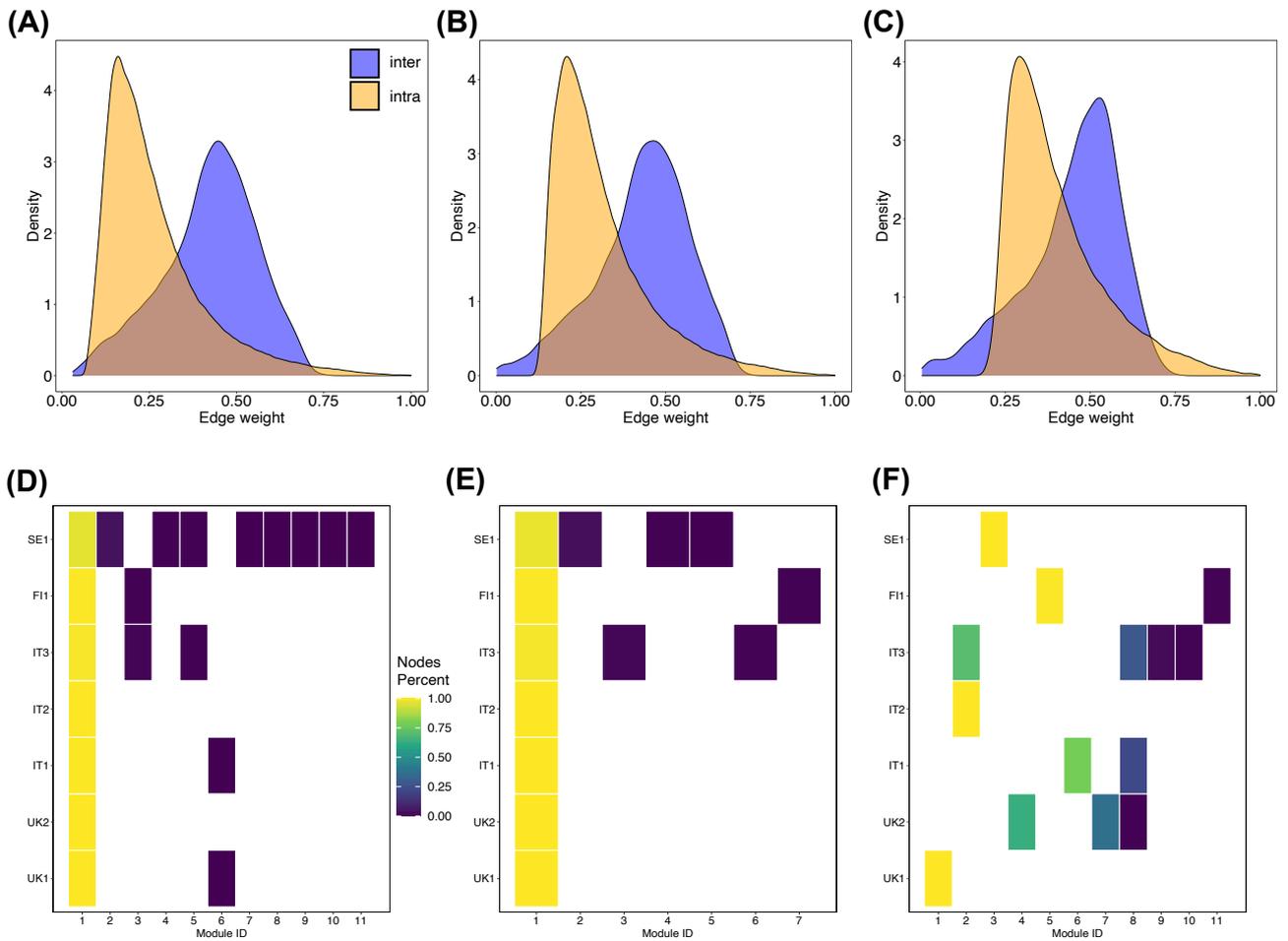
Overall, the sensitivity analysis shows that the patterns we obtained in the 30% threshold, which is close to the ‘functional core’ found in Wallace et al. [27] provides a good balance between a biologically meaningful set of microbes and ability to discern the scale of the processes that define it.



**Fig. S9: Sensitivity analysis of transitivity.** Box plots depict the values of the clustering coefficient  $CC_i$  across all microbes within a farm for different thresholds of the common core (depicted in colors). As expected, the lower the threshold, the lower the  $CC_i$  because lower thresholds include more microbes.



**Fig. S10: Sensitivity analysis of partner fidelity.** Distributions of  $P_i^J$  (red) and  $P_i^U$  (blue) for the 5% (A), 10% (B) and 20% (C) thresholds.



**Fig. S11: Sensitivity analysis of modularity.** (A)-(C) Distributions of the intralayer and interlayer edges for the 5%, 10% and 20% thresholds, respectively. With higher thresholds, the number of intralayer edges decreases (y-axis) and they become stronger (x-axis). Conversely, the number of interlayer edges increases and there is a better balance between the two edge types. (D)-(F) The modules for the same order of thresholds. There is a single large module encompassing all farms. Within each farm, there are many small modules, resulting from an excessive number of nodes. This trend decreases with increasing thresholds.

## References

1. Pascual, M. & Dunne, J. A. *Ecological Networks: Linking Structure to Dynamics in Food Webs* (Oxford University Press, 2006).
2. Kéfi, S. *et al.* Advancing our understanding of ecological stability. *Ecol. Lett.* **22**, 1349–1356. doi:[10.1111/ele.13340](https://doi.org/10.1111/ele.13340) (2019).
3. Shabat, S. K. B. *et al.* Specific microbiome-dependent mechanisms underlie the energy harvest efficiency of ruminants. *ISME J.* **10**, 2958–2972. doi:[10.1038/ismej.2016.62](https://doi.org/10.1038/ismej.2016.62) (2016).
4. Bergamaschi, M. *et al.* Gut microbiome composition differences among breeds impact feed efficiency in swine. *Microbiome* **8**, 110. doi:[10.1186/s40168-020-00888-9](https://doi.org/10.1186/s40168-020-00888-9) (2020).
5. Berry, D. & Widder, S. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Front. Microbiol.* **5**, 219. doi:[10.3389/fmicb.2014.00219](https://doi.org/10.3389/fmicb.2014.00219) (2014).
6. Röttjers, L. & Faust, K. From hairballs to hypotheses—biological insights from microbial networks. *FEMS Microbiol. Rev.* **42**, 761–780. doi:[10.1093/femsre/fuy030](https://doi.org/10.1093/femsre/fuy030) (2018).
7. Blanchet, F. G., Cazelles, K. & Gravel, D. Co-occurrence is not evidence of ecological interactions. *Ecol. Lett.* **23**, 1050–1063. doi:[10.1111/ele.13525](https://doi.org/10.1111/ele.13525) (2020).
8. Goberna, M. & Verdú, M. Cautionary notes on the use of co-occurrence networks in soil ecology. *Soil Biol. Biochem.* **166**, 108534. doi:[10.1016/j.soilbio.2021.108534](https://doi.org/10.1016/j.soilbio.2021.108534) (2022).
9. Román, S., Ortiz-Álvarez, R., Romano, C., Casamayor, E. O. & Martín, D. Microbial community structure and functionality in the deep sea floor: Evaluating the causes of spatial heterogeneity in a submarine canyon system (NW Mediterranean, Spain). *Front. Mar. Sci.* **6**. doi:[10.3389/fmars.2019.00108](https://doi.org/10.3389/fmars.2019.00108) (2019).
10. Ortiz-Álvarez, R., Cáliz, J., Camarero, L. & Casamayor, E. O. Regional community assembly drivers and microbial environmental sources shaping bacterioplankton in an alpine lacustrine district (Pyrenees, Spain). *Environ. Microbiol.* **22**, 297–309. doi:[10.1111/1462-2920.14848](https://doi.org/10.1111/1462-2920.14848) (2020).
11. Ontiveros Vicente J. *et al.* Biological Microbial Interactions from Cooccurrence Networks in a High Mountain Lacustrine District. *mSphere* **0**, e00918–21. doi:[10.1128/msphere.00918-21](https://doi.org/10.1128/msphere.00918-21) (2022).
12. Saravia, L. A., Marina, T. I., Kristensen, N. P., De Troch, M. & Momo, F. R. Ecological network assembly: How the regional metaweb influences local food webs. *J. Anim. Ecol.* **91**, 630–642. doi:[10.1111/1365-2656.13652](https://doi.org/10.1111/1365-2656.13652) (2022).
13. Araújo, M. B. & Rozenfeld, A. The geographic scaling of biotic interactions. *Ecography* **37**, 406–415. doi:[10.1111/j.1600-0587.2013.00643.x](https://doi.org/10.1111/j.1600-0587.2013.00643.x) (2014).
14. Goldford, J. E. *et al.* Emergent simplicity in microbial community assembly. *Science* **361**, 469–474. doi:[10.1126/science.aat1168](https://doi.org/10.1126/science.aat1168) (2018).
15. Ladau, J. & Elie-Fadrosh, E. A. Spatial, Temporal, and Phylogenetic Scales of Microbial Ecology. *Trends Microbiol.* **27**, 662–669. doi:[10.1016/j.tim.2019.03.003](https://doi.org/10.1016/j.tim.2019.03.003) (2019).
16. Yuan, M. M. *et al.* Climate warming enhances microbial network complexity and stability. *Nat. Clim. Chang.* **11**, 343–348. doi:[10.1038/s41558-021-00989-9](https://doi.org/10.1038/s41558-021-00989-9) (2021).

17. Ma, B. *et al.* Geographic patterns of co-occurrence network topological features for soil microbiota at continental scale in eastern China. *ISME J.* **10**, 1891–1901. doi:[10.1038/ismej.2015.261](https://doi.org/10.1038/ismej.2015.261) (2016).
18. Mandakovic, D. *et al.* Structure and co-occurrence patterns in microbial communities under acute environmental stress reveal ecological factors fostering resilience. *Sci. Rep.* **8**, 5875. doi:[10.1038/s41598-018-23931-0](https://doi.org/10.1038/s41598-018-23931-0) (2018).
19. Mittelbach, G. G. & Schemske, D. W. Ecological and evolutionary perspectives on community assembly. *Trends Ecol. Evol.* **30**, 241–247. doi:[10.1016/j.tree.2015.02.008](https://doi.org/10.1016/j.tree.2015.02.008) (2015).
20. Miller, E. T., Svanbäck, R. & Bohannan, B. J. M. Microbiomes as Metacommunities: Understanding Host-Associated Microbes through Metacommunity Ecology. *Trends Ecol. Evol.* **33**, 926–935. doi:[10.1016/j.tree.2018.09.002](https://doi.org/10.1016/j.tree.2018.09.002) (2018).
21. Bascompte, J. Networks in ecology. *Basic Appl. Ecol.* **8**, 485–490. doi:[10.1016/j.baae.2007.06.003](https://doi.org/10.1016/j.baae.2007.06.003) (2007).
22. Jiao, S. *et al.* Bacterial communities in oil contaminated soils: Biogeography and co-occurrence patterns. *Soil Biol. Biochem.* **98**, 64–73. doi:[10.1016/j.soilbio.2016.04.005](https://doi.org/10.1016/j.soilbio.2016.04.005) (2016).
23. Jones, C. M. & Hallin, S. Geospatial variation in co-occurrence networks of nitrifying microbial guilds. *Mol. Ecol.* **28**, 293–306. doi:[10.1111/mec.14893](https://doi.org/10.1111/mec.14893) (2019).
24. Li, J. *et al.* Distinct mechanisms shape soil bacterial and fungal co-occurrence networks in a mountain ecosystem. *FEMS Microbiol. Ecol.* **96**. doi:[10.1093/femsec/fiaa030](https://doi.org/10.1093/femsec/fiaa030) (2020).
25. Pilosof, S., Porter, M. A., Pascual, M. & Kéfi, S. The multilayer nature of ecological networks. *Nat Ecol Evol* **1**, 0101. doi:[10.1038/s41559-017-0101](https://doi.org/10.1038/s41559-017-0101) (2017).
26. Mizrahi, I., Wallace, R. J. & Morais, S. The rumen microbiome: balancing food security and environmental impacts. *Nat. Rev. Microbiol.* **19**, 553–566. doi:[10.1038/s41579-021-00543-6](https://doi.org/10.1038/s41579-021-00543-6) (2021).
27. Wallace, R. J. *et al.* A heritable subset of the core rumen microbiome dictates dairy cow productivity and emissions. *Sci Adv* **5**, eaav8391. doi:[10.1126/sciadv.aav8391](https://doi.org/10.1126/sciadv.aav8391) (2019).
28. Perlman, D. *et al.* Concepts and Consequences of a Core Gut Microbiota for Animal Growth and Development. *Annu Rev Anim Biosci* **10**, 177–201. doi:[10.1146/annurev-animal-013020-020412](https://doi.org/10.1146/annurev-animal-013020-020412) (2022).
29. Kokou, F. *et al.* Core gut microbial communities are maintained by beneficial interactions and strain variability in fish. *Nat Microbiol* **4**, 2456–2465. doi:[10.1038/s41564-019-0560-0](https://doi.org/10.1038/s41564-019-0560-0) (2019).
30. Hibbing, M. E., Fuqua, C., Parsek, M. R. & Peterson, S. B. Bacterial competition: surviving and thriving in the microbial jungle. *Nat. Rev. Microbiol.* **8**, 15–25. doi:[10.1038/nrmicro2259](https://doi.org/10.1038/nrmicro2259) (2010).
31. Trøjelsgaard, K., Jordano, P., Carstensen, D. W. & Olesen, J. M. Geographical variation in mutualistic networks: similarity, turnover and partner fidelity. *Proc. Biol. Sci.* **282**, 20142925. doi:[10.1098/rspb.2014.2925](https://doi.org/10.1098/rspb.2014.2925) (2015).
32. Veech, J. A. A probabilistic model for analysing species co-occurrence. *Glob. Ecol. Biogeogr.* **22**, 252–260. doi:[10.1111/j.1466-8238.2012.00789.x](https://doi.org/10.1111/j.1466-8238.2012.00789.x) (2013).

33. Griffith, D. M., Veech, J. A. & Marsh, C. J. cooccur: Probabilistic Species Co-Occurrence Analysis in R. *J. Stat. Softw.* **69**, 1–17. doi:[10.18637/jss.v069.c02](https://doi.org/10.18637/jss.v069.c02) (2016).
34. Mainali, K. P., Slud, E., Singer, M. C. & Fagan, W. F. A better index for analysis of co-occurrence and similarity. *Sci Adv* **8**, eabj9204. doi:[10.1126/sciadv.abj9204](https://doi.org/10.1126/sciadv.abj9204) (2022).
35. Mainali, K. P. *et al.* Statistical analysis of co-occurrence patterns in microbial presence-absence datasets. *PLoS One* **12**, e0187132. doi:[10.1371/journal.pone.0187132](https://doi.org/10.1371/journal.pone.0187132) (2017).
36. Adair, K. L., Wilson, M., Bost, A. & Douglas, A. E. Microbial community assembly in wild populations of the fruit fly *Drosophila melanogaster*. *ISME J.* **12**, 959–972. doi:[10.1038/s41396-017-0020-x](https://doi.org/10.1038/s41396-017-0020-x) (2018).
37. Delmas, E. *et al.* Analysing ecological networks of species interactions: Analyzing ecological networks. *Biol Rev* **94**, 16–36. doi:[10.1111/brv.12433](https://doi.org/10.1111/brv.12433) (2019).
38. Oksanen, J. *et al.* vegan: Community Ecology Package (version 2.5-7). [https://CRAN.R-project.org/package=](https://CRAN.R-project.org/package=vegan) (2020).
39. Lozupone, C. & Knight, R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* **71**, 8228–8235. doi:[10.1128/AEM.71.12.8228-8235.2005](https://doi.org/10.1128/AEM.71.12.8228-8235.2005) (2005).
40. Chen, J. *et al.* Associating microbiome composition with environmental covariates using generalized UniFrac distances. *Bioinformatics* **28**, 2106–2113. doi:[10.1093/bioinformatics/bts342](https://doi.org/10.1093/bioinformatics/bts342) (2012).
41. Paradis & Schliep. *ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R*
42. Tamura, K & Nei, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* **10**, 512–526. doi:[10.1093/oxfordjournals.molbev.a040023](https://doi.org/10.1093/oxfordjournals.molbev.a040023) (1993).
43. Schliep, K. P. *phangorn: phylogenetic analysis in R* 2011. doi:[10.1093/bioinformatics/btq706](https://doi.org/10.1093/bioinformatics/btq706).
44. R Core Team. *R: A Language and Environment for Statistical Computing* Vienna, Austria, 2021.
45. Gotelli, N. J. Null Model Analysis of Species Co-Occurrence Patterns. *Ecology* **81**, 2606–2621. doi:[10.2307/177478](https://doi.org/10.2307/177478) (2000).
46. Hutchinson, M. C. *et al.* Seeing the forest for the trees: Putting multilayer networks to work for community ecology. *Funct. Ecol.* **33** (ed Godoy, O.) 206–217. doi:[10.1111/1365-2435.13237](https://doi.org/10.1111/1365-2435.13237) (2019).
47. Hervías-Parejo, S *et al.* Species functional traits and abundance as drivers of multiplex ecological networks: first empirical quantification of inter-layer edge weights. *Proc. Biol. Sci.* **287**, 20202127. doi:[10.1098/rspb.2020.2127](https://doi.org/10.1098/rspb.2020.2127) (2020).
48. Vitali, A. *et al.* Invasive species modulate the structure and stability of a multilayer mutualistic network. *Proc. Biol. Sci.* **290**, 20230132. doi:[10.1098/rspb.2023.0132](https://doi.org/10.1098/rspb.2023.0132) (2023).
49. Shapiro, J. T. *et al.* Multilayer networks of plasmid genetic similarity reveal potential pathways of gene transmission. *ISME J.*, 1–11. doi:[10.1038/s41396-023-01373-5](https://doi.org/10.1038/s41396-023-01373-5) (2023).

50. Farage, C., Edler, D., Eklöf, A., Rosvall, M. & Pilosof, S. Identifying flow modules in ecological networks using Infomap. *Methods Ecol. Evol.* **12**, 778–786. doi:[10.1111/2041-210x.13569](https://doi.org/10.1111/2041-210x.13569) (2021).
51. Rosvall, M. & Bergstrom, C. T. Maps of random walks on complex networks reveal community structure. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 1118–1123. doi:[10.1073/pnas.0706851105](https://doi.org/10.1073/pnas.0706851105) (2008).
52. De Domenico, M., Lancichinetti, A., Arenas, A. & Rosvall, M. Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems. *Phys. Rev. X* **5**, 011027. doi:[10.1103/PhysRevX.5.011027](https://doi.org/10.1103/PhysRevX.5.011027) (2015).
53. Greenbaum, G., Templeton, A. R. & Bar-David, S. Inference and Analysis of Population Structure Using Genetic Data and Network Theory. *Genetics* **202**, 1299–1312. doi:[10.1534/genetics.115.182626](https://doi.org/10.1534/genetics.115.182626) (2016).
54. Greenbaum, G., Rubin, A., Templeton, A. R. & Rosenberg, N. A. Network-based hierarchical population structure analysis for large genomic data sets. *Genome Res.* **29**, 2020–2033. doi:[10.1101/gr.250092.119](https://doi.org/10.1101/gr.250092.119) (2019).
55. Faust, K. *et al.* Cross-biome comparison of microbial association networks. *Front. Microbiol.* **6**, 1200. doi:[10.3389/fmicb.2015.01200](https://doi.org/10.3389/fmicb.2015.01200) (2015).
56. Chesson, P. Mechanisms of maintenance of species diversity. *Annu. Rev. Ecol. Syst.* **31**, 343–366 (2000).
57. Araújo, M. B., Rozenfeld, A., Rahbek, C. & Marquet, P. A. Using species co-occurrence networks to assess the impacts of climate change. *Ecography* **34**, 897–908. doi:[10.1111/j.1600-0587.2011.06919.x](https://doi.org/10.1111/j.1600-0587.2011.06919.x) (2011).
58. Tu, Q. *et al.* Biogeographic patterns of microbial co-occurrence ecological networks in six American forests. *Soil Biol. Biochem.* **148**, 107897. doi:[10.1016/j.soilbio.2020.107897](https://doi.org/10.1016/j.soilbio.2020.107897) (2020).
59. Baldassano, S. N. & Bassett, D. S. Topological distortion and reorganized modular structure of gut microbial co-occurrence networks in inflammatory bowel disease. *Sci. Rep.* **6**, 26087. doi:[10.1038/srep26087](https://doi.org/10.1038/srep26087) (2016).
60. Freedman, Z. B. *et al.* Environment-host-microbial interactions shape the *Sarracenia purpurea* microbiome at the continental scale. *Ecology* **102**, e03308. doi:[10.1002/ecy.3308](https://doi.org/10.1002/ecy.3308) (2021).
61. Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V. & Egozcue, J. J. Microbiome Datasets Are Compositional: And This Is Not Optional. *Front. Microbiol.* **8**, 2224. doi:[10.3389/fmicb.2017.02224](https://doi.org/10.3389/fmicb.2017.02224) (2017).