Differential assembly of core and non-core host-microbe network structures along a land-use change gradient

Matan Markfeld [1], Georgia Titcomb [2], Toky Maheriniaina Randriamoria[3], Garima Sehgal[4], Nuzha Baksh[4], Ashley Kerrigan[4], Voahangy Soarimalala [3], Charles L. Nunn [5,6], and Shai Pilosof [1,7]

[1]Department of Life Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel
[2]Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, USA
[3]Association Vahatra, Antananarivo, Madagascar
[4]Department of Ecology, Evolution, and Marine Biology, University of California Santa Barbara, Santa Barbara, USA
[5]Department of Evolutionary Anthropology, Duke University, Durham, NC, USA
[6]Duke Global Health Institute, Durham, NC, USA
[7]The Goldman Sonnenfeldt School of Sustainability and Climate Change, Ben-Gurion University of the Negev, Beer-Sheva, Israel

Corresponding authors: *markfmat@post.bgu.ac.il*; *pilos@bgu.ac.il*

Keywords

Microbiome, community assembly, microbial ecology, land-use change, rat, host-microbe network, modularity, Madagascar

Abstract

Microbial communities are fundamental to host health, yet their assembly dynamics under environmental change remain poorly understood. We analyzed individual-level host-microbe networks in the non-native wild black rats (*Rattus rattus*) across a land-use gradient in Madagascar. By applying a moving prevalence threshold, we distinguished between core and non-core microbes and compared the assembly drivers shaping their network structures. Non-core microbes formed fragmented, modular networks shaped mainly by heterogeneous selection, reflecting environmental filtering. In contrast, core microbes exhibited stable, less modular networks driven primarily by stochastic ecological drift. These distinct assembly processes persisted across thresholds, highlighting fundamental differences in microbial structuring. Land-use change significantly influenced the modular structure of non-core microbes but had minimal effects on core microbes, demonstrating the differential sensitivity of microbial groups to environmental variation. This study advances our understanding of host-microbe interactions and provides a framework for assessing microbiome assembly under anthropogenic change.

Introduction

The structural organization of species interactions plays a pivotal role in shaping biodiversity patterns, maintaining community stability, and supporting ecosystem functions [1–3]. As a result, ecological research has long focused on unraveling these interaction patterns and the mechanisms underlying their emergence. However, widespread transformation and degradation of natural ecosystems in recent decades have disrupted ecological communities and their intricate interactions [4,5]. An example is the relationship between animal hosts and their associated microbial communities [6,7]. Microbial communities are essential for host fitness and functioning, influencing critical processes such as metabolism, immune responses, and disease resistance [8–10]. Despite the recognized importance of microbiomes, the mechanisms driving their assembly and structuring remain poorly understood, particularly in wild populations under environmental change.

Anthropogenic land-use change, including the conversion of natural areas into agricultural or urban ecosystems, alters the environment and the conditions shaping the host microbiome [6,11]. Different land-use types create distinct environmental conditions, such as changes in host community composition and density, variations in vegetation cover and composition, and shifts in human activity, including agricultural practices. These variations can impact the host microbiome through both deterministic (e.g., environmental filtering) and stochastic (e.g., dispersal and ecological drift) processes [12–14].

The processes that shape microbiome assembly play a critical role in structuring the host-microbiome interaction network, producing patterns that reflect the interplay between environmental conditions and microbial dynamics [15]. For instance, microbes that optimize digestion or nutrient absorption under specific environmental conditions set by host diet may be deterministically selected [16]. Over time, such selection pressure could lead to more associations with particular microbial taxa than with others in the community [17]. Structural patterns are therefore expected to align closely with environmental conditions, reflecting the ecological context in which they arise. Conversely, stochastic processes can dominate when selective pressures are weaker [18]. In such scenarios, network structure can be shaped by random demographic variations, such as fluctuations in microbial population sizes or the accidental extinction of certain taxa [18]. Additionally, stochastic dispersal events during the early stages of microbiome assembly can have lasting effects on community composition and structure through priority effects, where the order and timing of microbial colonization influence which taxa establish successfully [19,20]. These random processes may result in substantial variability in microbial community structure, even under similar environmental conditions.

Host-microbe interactions are also affected by microbe function and prevalence. The microbiome comprises diverse groups of microbes that can be categorized into different functional or ecological groups. One key distinction is between core and non-core microbes [21–23]. The most widely used definition is the 'common core' microbiome, which refers to microbial taxa that are consistently found across a host population above a defined prevalence threshold [24]. While high-prevalence taxa may indicate stable and essential functional relevance to the host (such as predicting key host phenotypes like methane emission and lactation in cows [25]), their prevalence alone does not imply they are necessary for host function [22]. Transient and variable non-core microbes

may provide flexible responses to changing conditions or introduce novel capabilities [21,26]. Examining how the structure of host-microbe interactions varies across different thresholds of core and non-core microbes can enhance our ability to disentangle the ecological mechanisms shaping those interactions.

Network analysis offers a powerful approach to studying the associations between hosts and their partners (e.g., mammals and parasites or bacteria and phages) by mapping their interactions and revealing structural relationships [7,27]. However, there are only a few studies that have explored these structural patterns in host-microbiome bipartite networks (e.g., [15,28]). A key structural property of such interaction networks is modularity, in which the network is partitioned into groups (modules) composed of species that interact strongly within a module and weakly outside their module [29]. In host-microbe networks, modules could represent functional units of microbial communities that perform specific metabolic processes, enhancing system stability and bolstering network resilience against perturbations [15]. Furthermore, modular structures can be signatures of underlying processes—such as coevolution, competition, or shared niches [15,30,31]. However, the processes driving modularity remain under-explored [30,32]. Identifying these processes in host-microbe networks, for both core and non-core microbes, will provide valuable insights into microbiome assembly dynamics and function. Doing so under environmental gradients will further help elucidate how microbiomes respond to anthropogenic changes.

In this study, we investigate the structure of the host-microbe network of wild black rat (*Rattus rattus*) populations across a land-use change gradient in rural northern Madagascar. *R. rattus* is a non-native, generalist omnivore that is the most abundant species in the region. It poses a threat to endemic species and is considered a nuisance to humans [33]. The region consists of a mosaic of human villages, including small agricultural communities that rely on farming rice, vanilla, and other crops. These land uses transition into more forested areas and a national park. Such land-use gradients, where human activity is at the interface with natural ecosystems, are common in rural low-income regions and provide an important context for studying wildlife-associated microbial communities.

We address two fundamental questions: (1) How do deterministic and stochastic processes shape host-microbe community structures along a gradient of land-use change? (2) How do these processes differ between core and non-core microbial groups? Differences in the distribution of core and non-core microbes suggest that these groups possess distinct traits, leading us to hypothesize that their community assembly is driven by different processes. Specifically, core microbes, which are consistently present across hosts, are expected to be structured by homogeneous selection, where hosts favor specific vital taxa, resulting in similar microbial communities across individuals, or alternatively, by more stochastic processes. In contrast, given that non-core microbes are more variable and sensitive to environmental changes, we propose that their occurrence in hosts is primarily shaped by deterministic processes. Particularly, we predicted that heterogeneous selection, where environmental differences filter microbial taxa differently, would lead to shifts in their modular structure along the land-use gradient (**Figure 1**). However, defining core versus non-core microbes is challenging, as studies typically rely on arbitrary thresholds that lack clear biological justification [22]. To address this issue, we apply a

moving threshold approach, examining microbial prevalence along a continuous gradient rather than using a strict cutoff. This allowed us to assess how assembly processes varied continuously with microbial prevalence in shaping host-microbe community structures.

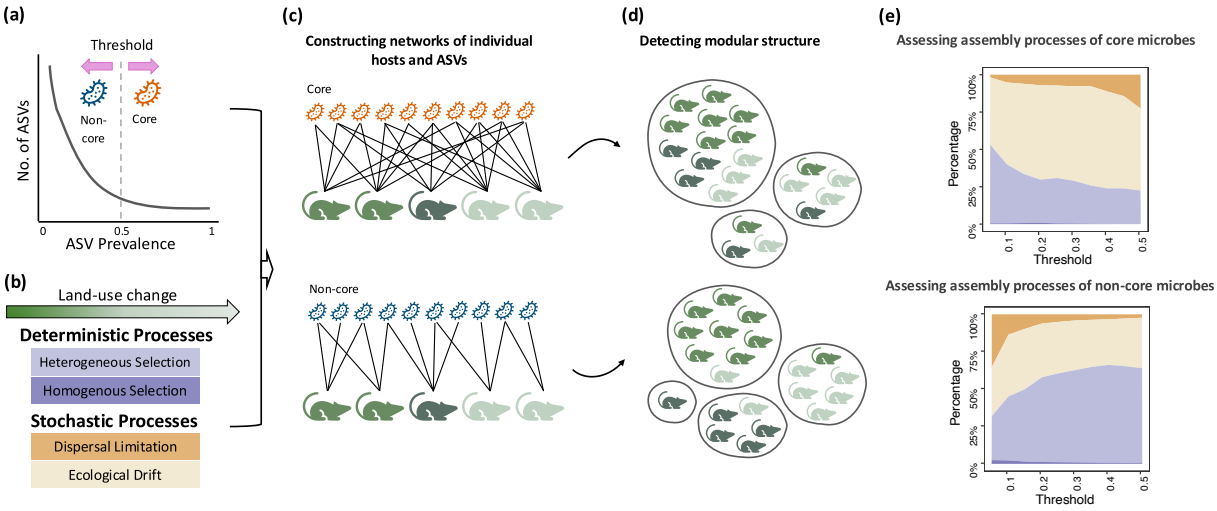

Figure 1: Conceptual scheme of methods and hypotheses. **(a)** ASVs are classified into core and non-core groups. The core threshold is defined within a prevalence range of 0.05–0.5. **(b)** Land-use change influences host-microbe assembly through deterministic processes (heterogeneous and homogeneous selection) and stochastic processes (dispersal limitation and ecological drift). **(c)** A bipartite network is constructed to link individual hosts with core and non-core ASVs. Host colors represent different land-use types. **(d)** Modules of hosts are identified based on their strong associations with similar ASVs. **(e)** Quantification of the relative importance of stochastic and deterministic processes for comparisons between hosts clustered within the same module and those in different modules (see *Materials and methods* and **Figure S1** for details). Colors correspond to the processes in panel (b).

Materials and methods

Study site and rat sampling

Small mammals were collected in three villages in the SAVA region of northeast Madagascar, in the area surrounding Marojejy National Park (**SI note 1**). The park encompasses natural moist evergreen forests that range from low elevation to high mountain peaks over 2000 m. In each village, seven sites were sampled along a degradation gradient: (1) semi-intact forest inside the national park, (2) secondary forest, (3) brushy regrowth (savoka), (4) agroforest (vanilla plantation), (5) mixed agriculture (sugarcane/coffee plantation), (6) flooded rice, and (7) village (**Table S1**). Sites in each village setting were located ~500 m apart. In each site a grid of 121 live traps (11x11, with 10 m distance between traps) was established, in addition to two lines of 11 pitfall traps. Each site was sampled for six consecutive nights three times during the sampling period. The sampling was carried out between October 2019 and September 2020 in the village M; between November 2020 and September 2021 in the village S; and between October 2021 and August 2022 in the village A. Fecal pellets were collected from each live-trapped animal and immediately stored in Zymo DNA/RNA Shield.

We focused our host-microbe network on the black rat (*Rattus rattus*), a non-native species

introduced to Madagascar probably a few thousand years ago, and the most abundant species in the region [33]. It is a large generalist omnivore with a broad and flexible diet [34]. All these characteristics make it one of the main agricultural pests, a major threat to native ecological communities in the local area, and potentially an important factor in the transmission of zoonotic diseases [33,35].

DNA extraction and 16S rRNA gene amplicon sequencing

DNA was extracted from ∼1g feces collected from trapped small mammals using Zymo Quick-DNA Fecal/Soil Microbe Miniprep kits (cat #D6010) using manufacturer protocols. 16S metabarcoding was conducted using 515F–806R primers to target the V4 region of the 16S SSU rRNA [36]. Each primer included an Illumina adapter, barcode, primer pad, and linker. Reactions were carried out in 25uL volumes consisting of 10uL of 1.25uM forward and reverse primer, 2uL of DNA, and 13uL of Platinum Hot Start PCR mastermix (ThermoFisher Scientific, cat #13000014). Reaction conditions were as follows: 95°C for 3min, 35x 98°C for 30secs, 58°C for 30secs, and 72°C for 30secs, followed by a final extension at 72°C for 5min. Concentrations were measured using Promega One Quantifluor kits on a Teccan platereader. Samples were then normalized to 7ng/uL prior to pooling. The product was cleaned using magnetic beads (bead:DNA ratio was 0.8:1) and sequenced at UC Santa Barbara Biological Nanostructures Laboratory on an Illumina MiSeq (v3 chemistry, 2x300bp, 24M reads).

Sequences were demultiplexed using cutadapt (v.3.4) with zero error tolerance [37]. We then performed quality filtering steps using the *dada2* package in R [38]. Specifically, we filtered and trimmed amplicons (minimum length = 100, 15% PhiX removed), inferred and removed errors, dereplicated sequences, inferred amplicon sequencing variants (ASVs) using the pseudo-pooling method, merged pairs, and removed chimeras. We assigned taxonomic identifications to ASVs using the *assignTaxonomy* function in *dada2*, using the SILVA nr99 SSU reference database (v.138.1).

Sequence data processing

We filtered out very rare ASVs with a relative abundance lower than 0.1% in a sample or those that occur in less than 1% of all individuals. Additionally, we removed all ASVs that were identified as 'Chloroplast' or 'Mitochondria'. Finally, we excluded 21 out of 876 samples with fewer than 5000 total reads from our analysis. Filtering procedures resulted in 855 individual hosts and 1,951 ASVs from an original total of 10,358.

Core classification and network construction

We categorized bacterial ASVs into core and non-core groups based on their prevalence (**Figure S2**). Typically, this distinction is made before analysis by applying a predefined prevalence threshold [22,39]. However, no objective boundary separates core and non-core microbes. To account for this, we implemented a moving threshold of increasing prevalence ranging from 0.05 to 0.5, in increments of 0.05. ASVs with prevalence values above or below the threshold were classified as core and non-core, respectively. (**Figure 1a**). For each threshold value we constructed a bipartite network in which links represented the occurrence of a bacterial ASV in a rat individual, with link weights as the ASV's relative read abundance. While microbial prevalence does not always directly reflect function, core and non-core groups exhibited

substantial taxonomic differences at the family level (**SI note 2**). This suggests that their distinction is based not only on prevalence but also on potential functional differences, under the assumption that functional traits are phylogenetically conserved [40]. For convenience, we report some of the results using the 20% threshold (based on the ASV prevalence distribution), but our results hold across thresholds.

Modularity analysis

We assessed modularity in the host-microbe network using Infomap. Infomap detects the optimal network partition based on the movement of a random walker on the network [41,42]. For any given network partition, the random walker moves across nodes. 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, which minimizes L [41], clusters hosts and microbes into modules that are more densely connected. The modules can contain hosts from a single or multiple sites, and each site may consist of a single or multiple different modules (**Figure 1d**). In this way, modules capture variation in host microbiome within and across land uses.

Community assembly processes within and between modules

To quantify the relative contributions of deterministic and stochastic processes to microbial community assembly, we applied the framework proposed by [43,44]. This method is widely used in the study of microbial assembly [18,45–47]. Specifically, we inferred assembly processes across all pairwise comparisons of individual hosts using a combination of the β-Nearest Taxon Index (βNTI) and the Raup-Crick metric (RC) (A visual guide is presented in **Figure S1**).

βNTI quantifies the phylogenetic turnover between pairs of communities, assuming that different environmental conditions lead to the selection of phylogenetically distinct communities. Therefore, it provides insights into the influence of selection on community assembly [18,43]. To calculate βNTI, we constructed a phylogenetic tree based on the 16S sequence data using the Maximum Likelihood method. We then calculated the β-Mean-Nearest-Taxon-Distance (βMNTD) for each pair of microbe communities (i.e., individual hosts). βMNTD is the average phylogenetic distance between each ASV in one individual host and its closest relative ASV in another individual host. High βMNTD values mean that individual hosts have very different microbial communities from one another. We compared the observed βMNTD to a null distribution of βMNTD values generated by shuffling taxa labels across the tips of the phylogenetic tree to randomize phylogenetic relationships among species. We then calculated βNTI as the number of standard deviations the observed βMNTD deviated from the mean of the null distribution (as in z-scores):

$$\beta\text{NTI} = \frac{\beta\text{MNTD}_{\text{obs}} - \overline{\beta\text{MNTD}_{\text{null}}}}{\text{sd}(\beta\text{MNTD}_{\text{null}})} \tag{1}$$

The Raup-Crick metric (RC) quantifies the probability that two microbial communities differ in species composition while accounting for differences in richness [48]. It compares the observed number of shared ASVs (without considering ASV relatedness) between two hosts to a null distribution, which is generated by randomly sampling from the species pool, with ASV selection probabilities proportional to their frequencies. The metric is calculated as the count of observed

values that are more than or equal to the null similarity, scaled between -1 and +1:

$$RC = 2\left(\frac{\#(\text{observed shared ASVs} \leq \text{null shared ASVs})}{\#\text{total null iterations}} - 0.5\right) \quad (2)$$

where values approaching 1 indicate that two communities are less similar than expected by chance, values near 0 indicate similarity consistent with random assembly, and values approaching -1 suggest greater similarity than expected.

For each of the core and non-core groups, we inferred community assembly processes as follows [43,44]. $|\beta\text{NTI}| > 2$ indicates dominance by deterministic processes such as environmental filtering, as the observed phylogenetic turnover deviates significantly from that expected under random turnover. Positive values ($\beta\text{NTI} > 2$) suggest *heterogeneous selection*, whereby environmental differences among hosts filter microbial taxa differently, leading to higher-than-random phylogenetic turnover. In contrast, negative values ($\beta\text{NTI} < 2$) indicate *homogeneous selection*, where similar conditions across hosts favor the same microbial taxa, resulting in lower-than-random phylogenetic turnover.

A $|\beta\text{NTI}| < 2$, indicates weak selection, whereby stochastic processes dominate. Within this range, if $RC > +0.95$, dispersal limitation is inferred, meaning that while communities remain phylogenetically similar, higher-than-random taxonomic turnover suggests that microbial dispersal is restricted, preventing taxa from colonizing suitable hosts. Conversely, if $RC < -0.95$, we infer homogenizing dispersal, where frequent microbial movement between hosts reduces taxonomic differences, leading to more uniform community compositions. When $|\beta\text{NTI}| < 2$ and $|RC| < 0.95$, the process is classified as ecological drift—a stochastic process where random fluctuations, rather than selection or dispersal, drive community composition (**Figure S1**).

We measured the relative contribution of each process to community assembly by calculating the proportion of all rat individuals' pairwise comparisons where the process was inferred. To evaluate whether specific assembly processes generate the modular structure, we determined the relative importance of each process separately for comparisons between hosts clustered within the same module and those in different modules (**Figure 1d-e**).

Quantifying the effect of land-use on host-microbe network structure

To explore how land-use change impacts the network's modular structure, we used transformation-based redundancy analysis (tb-RDA), a constrained ordination technique that examines variation in response variables explained by environmental predictors [49]. A Hellinger-transformed matrix of the number of hosts in each module at each site served as the response variable. The Hellinger transformation converts raw count data into the square root of relative abundances, thus preserving ecological distances while reducing the disproportionate influence of rare instances. To represent land-use change, we included three environmental gradients—vegetation, elevation, and distance to the village center—as explanatory variables (see **SI note 3** for details). Vegetation was characterized using the first two principal components (PCs) from a PCA of eight vegetation attributes. PC1 was primarily associated with high herbaceous cover and low tree cover, while PC2 distinguished between different tree types. Distance to the village center served as a proxy for human disturbance. The variables were standardized to a zero mean and unit variance. The number of hosts at each site was included as a conditioning term to account for the influence

of host abundance before analyzing the effects of the main predictors. We excluded trapping sites located inside villages from the analysis due to missing vegetation data. Additionally, we removed the largest module in each network, as it spanned all sites and homogenized the data. The significance of the overall tb-RDA model was tested using a Monte Carlo permutation test with 999 runs. The significance of each explanatory variable was evaluated with a Monte Carlo procedure (using the anova.cca function with term margins), applying Holm's correction to adjust p-values for multiple testing.

Software packages

All analyses were conducted in R (version 4.2.1) [50]. We used the R package *infomapecology* [42] (version 2.7.1) to calculate modularity. We calculated the βNTI and *RC* indices and null models using the R packages *iCAMP* [51] and *Vegan* [52], respectively, *RC*, with 500 iterations used for generating the null distributions. We used the *Vegan* [52] package for the tb-RDA analysis.

Results

Core and non-core microbial networks are modular

As is common in microbial communities, the community was dominated by non-core microbes, and even at the lowest threshold, non-core ASVs accounted for >70% of all ASVs (**Figure 2a(i)**). The connectance (i.e., the proportion of realized links out of all possible links) of the core microbial network was consistently higher than that of the non-core network across all core thresholds, ranging from 0.13 to 0.65 for the core group compared to 0.02 to 0.05 for the non-core group (**Figure 2a(ii)**). This reflects the fact that core microbes occur in many more hosts than non-core microbes.

Both core and non-core host-microbe networks exhibited modular structures across all core thresholds. However, the non-core network was more fragmented, with a greater number of smaller modules (**Figure 2a(iii-iv)**). For instance, at a core threshold of 0.2, the core group network consisted of 11 modules with an average module size of 77 hosts, whereas the non-core group network comprised 89 modules with an average module size of 9 hosts (**Figure 2b**). In all core networks there was a single large module that spanned all sites and included 35-50% of hosts (**Figure 2a(v)**). For the non-core group, this large module emerges and rapidly expands (31-63% of hosts) as the core threshold increases beyond 0.15. At a core threshold of 0.2 the largest module covers 48% of hosts in the core group and 45% of hosts in the non-core group (**Figure 2b**).

The taxonomic composition varied significantly between the core and non-core groups across all core thresholds due to differences in the relative abundance of microbial families (**SI note 2**). While the most abundant core and non-core microbial families remained consistent across all core thresholds, the non-core group exhibited greater taxonomic variation between modules (**Figure 3**). Specifically, the dominant families varied between modules in the non-core group, whereas in the core group, taxonomic composition remained more stable.

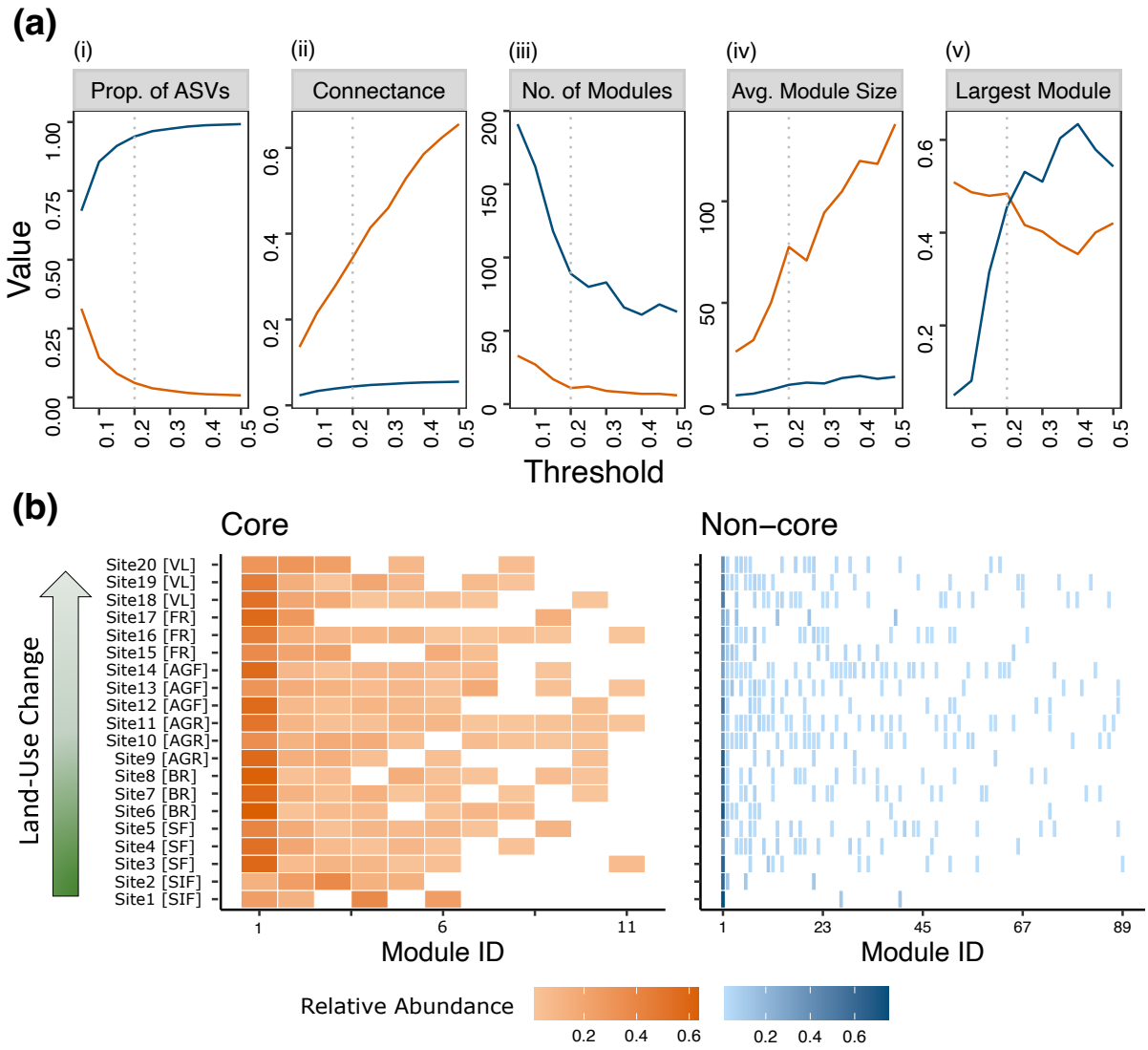

Figure 2: Modular structure of host-microbe networks. **(a)** Network properties of the core (orange) and non-core (blue) microbial groups across a range of core thresholds. (i) The proportion of ASVs classified as core and non-core at each core threshold, out of 1951 ASVs. (ii) Network connectance (proportion of realized host-ASV links out of all possible ones). (iii) The number of modules identified in the network. (iv) The average module size (i.e., the average number of hosts per module). (v) The proportional size of the largest module in the network. The gray dashed vertical line indicates a core threshold of 0.2. **(b)** Modular structure of the networks at a core threshold of 0.2. Each column represents a module, and each row corresponds to a site. The cells indicate the presence of hosts from a site within a module, with the color intensity reflecting the relative abundance of hosts, out of all the hosts in the site (each row sums to 1). Sites are arranged along a land-use gradient, from semi-intact forest [SIF], secondary forest [SF], brushy regrowth [BR], agriculture [AGR], agroforest [AGF], flooded rice [FR], to the village [VL] (**Table S1**). The non-core microbial network is fragmented into more and smaller modules, compared to the core microbial network.

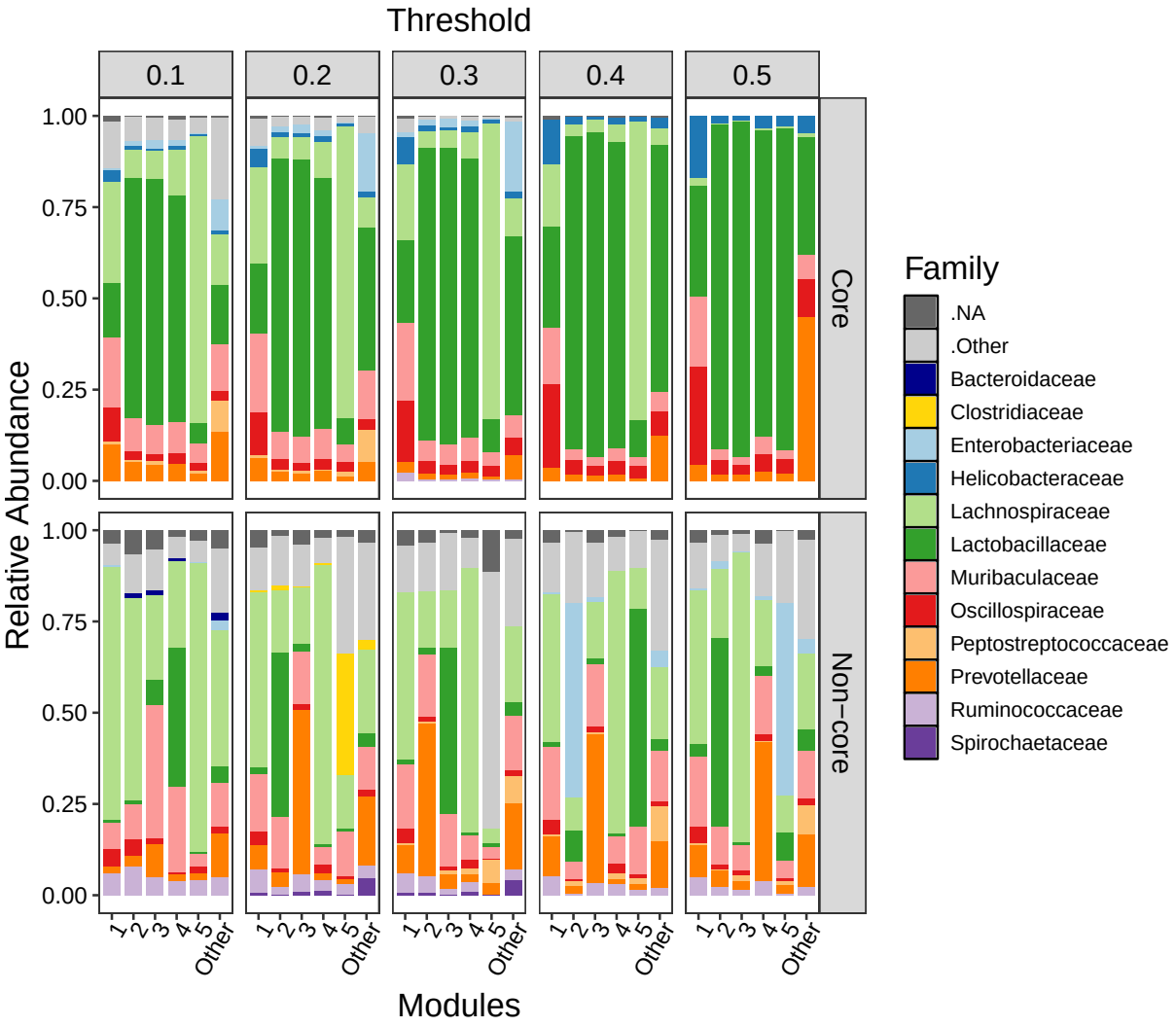

Figure 3: Microbial taxa variation across modules and thresholds. Rows represent core and non-core groups across different core thresholds, as shown in the columns. Bars indicate the relative abundance of microbial families. Reads from all hosts within each module were aggregated, and data are presented for the five largest modules (ordered 1–5 by size). All remaining modules are grouped under "Other." Each color represents a bacterial family, with only the eight most abundant families from both core and non-core groups displayed. Less abundant families are combined and shown in gray, while ASVs with no identified family are represented in black. The core and non-core groups differed significantly in taxonomic composition across all thresholds, with non-core groups showing greater variation across modules.

Distinct processes drive the modular structure of core and non-core microbial groups

The observed differences in modular structure between microbial groups suggest that they are shaped by distinct ecological processes. To explore this further, we analyzed the assembly processes driving variation within and between modules in each group. Our findings reveal that distinct ecological mechanisms govern the modular organization of core and non-core microbial groups (**Figure 4**). In core groups, microbiome variation between hosts within the same module and across different modules is predominantly driven by ecological drift, a pattern consistent across all core thresholds. For instance, at a core threshold of 0.2, drift accounted for 63.2% of pairwise comparisons between hosts in different modules and 73.5% of comparisons within modules. This pattern highlights the stable and ubiquitous nature of core microbes across environmental conditions. The reduced influence of selection further amplifies the role of stochastic processes, such as ecological drift, in driving microbiome variation and shaping modular organization.

In contrast, selection primarily drives the assembly of non-core microbes. For instance, at a core threshold of 0.2, variation between modules was explained predominantly by heterogeneous selection (56.9%), followed by drift (36%), dispersal limitation (6.2%), and homogenous selection (0.9%). In comparison, variation within modules was largely attributed to drift, ranging from 49.9% to 67%. At a core threshold of 0.2, within-module variation was driven by 31.3% heterogeneous selection, 59.2% drift, 5.8% dispersal limitation, and 3.7% homogeneous selection. This pattern suggests that local environmental conditions and host-specific factors select (i.e., heterogeneous selection) for phylogenetically distinct microbial communities, which are then clustered into separate modules.

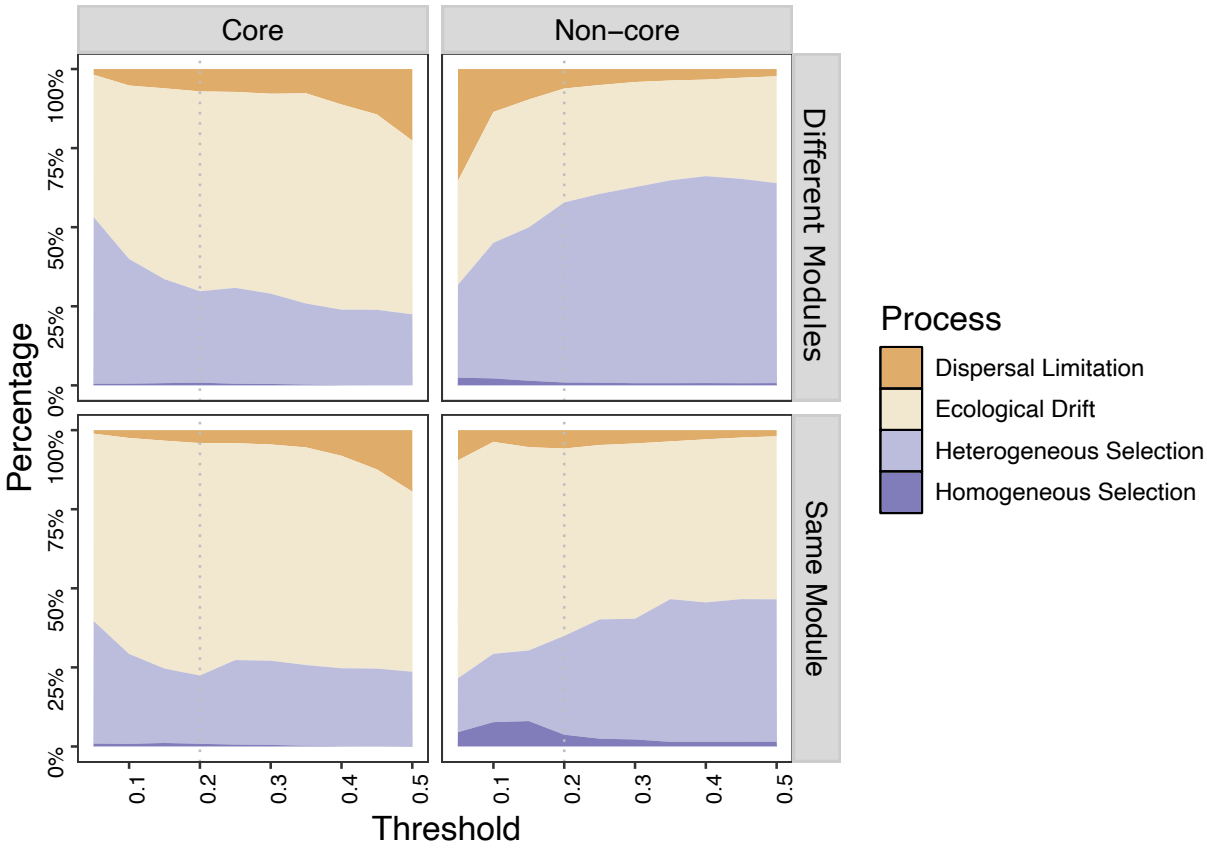

Figure 4: Assembly processes of core and non-core microbial groups within and across modules. Proportion of pairwise comparisons (y axis) between hosts clustered within the same module or in different modules across a range of core thresholds (x axis). Each color represents a specific assembly process. We inferred community assembly processes across all pairwise host comparisons using the β-Nearest Taxon Index (βNTI) and the Raup-Crick metric (RC) (see Materials and Methods). The relative contribution of each process was measured by the proportion of pairwise comparisons where it was inferred. The vertical gray dashed line indicates a core threshold of 0.2.

Land-use change drives the modular structure of the non-core microbial network

To evaluate the impact of land-use change on the network's modular structure, we examined how the similarity of modules across sites aligned with the land-use gradient. In line with our hypotheses, we found that environmental characteristics associated with different land-use types consistently influenced module similarity across sites for the non-core microbial group. This was shown by a significant tb-RDA across nearly all thresholds (0.1–0.5) (**Figure 5a**). These findings also correspond to our finding that heterogeneous selection drives the non-core modular structure of microbial communities. The variation explained by the model (adjusted R^2) ranged from 6.1% to 11.7%. For most thresholds, at least one explanatory variable was significant in the model, with site vegetation and elevation being the most frequently significant variables (**Figure 5b**). For instance, at a core threshold of 0.2, the model ($F = 1.41$, $p < 0.01$) explained 10% of the variation in module composition, with vegetation and elevation being the dominant factors (**Figure 5c**).

A different pattern emerged for the core microbial group, as the tb-RDA was significant just at the core thresholds of 0.05 and 0.5, explaining 9.3% and 18.3% of the variation, respectively

(**Figure 5a**). Distance from the village center was the only marginally significant variable at the core threshold of 0.5. At a core threshold of 0.2, the model was not significant ($F = 1.21$, $p = 0.218$), explaining just 4.9% of the variation. This weak effect of land-use change aligns with the earlier finding that deterministic processes play a relatively minor role in the core group assembly.

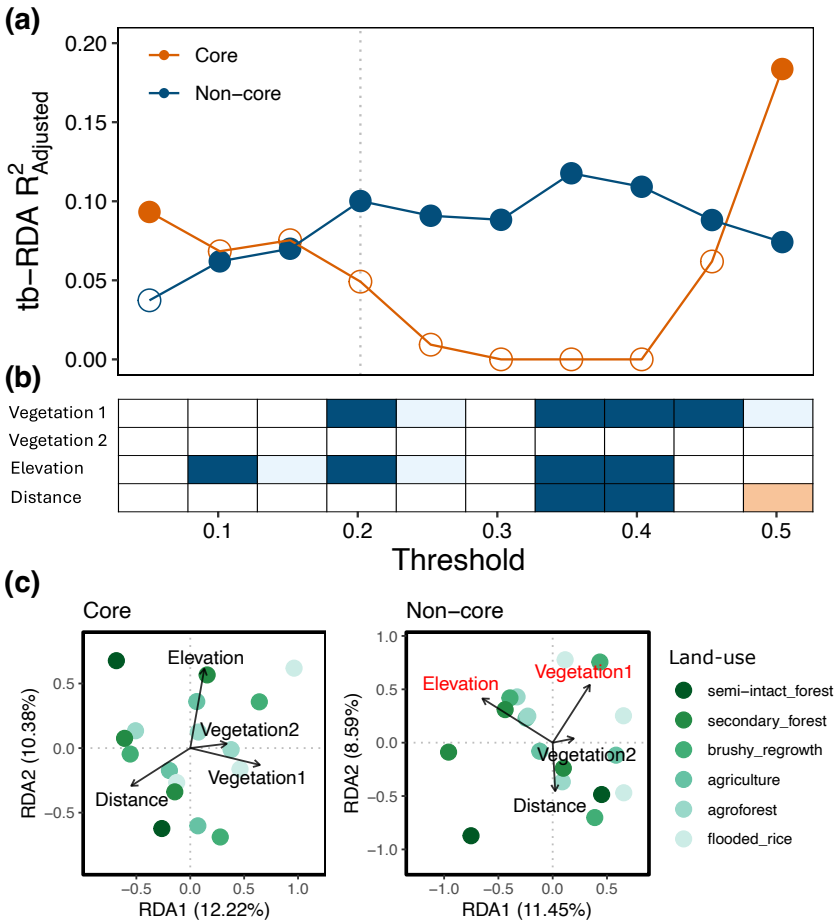

Figure 5: Similarity in composition of modules across environmental gradients. (a) Variation in module composition explained (adjusted R^2) by the transformation-based redundancy analysis (tb-RDA) across a range of core thresholds for core (orange) and non-core (blue) microbial groups. Filled and non-filled circles represent significant ($p < 0.05$) and non-significant models, respectively. **(b)** The significance of explanatory variables. Filled cells indicate significant variables at specific core thresholds, with colors corresponding to the microbial group, and their shades indicating significance (dark, $p < 0.05$) or marginal significance (light, $p < 0.1$). The gray dashed line indicates a core threshold of 0.2. **(c)** An example for the tb-RDA ordination of network modules at a core threshold of 0.2. Each point represents a site, with colors indicating land-use type. Significant variables are highlighted in red. Vegetation1 is positively associated with herbaceous cover and height and negatively associated with tree cover.

Discussion

Understanding how microbial communities assemble and respond to environmental change is a fundamental question in ecology, with implications for host health and ecosystem function [53,54]. In this study, we used a network-based approach to investigate the assembly processes

shaping host-microbe interactions in non-native black rats that are replacing endemic species along a land-use gradient in Madagascar. By applying a moving threshold to classify core and non-core microbes, we captured nuances in microbiome structure and its drivers, while also demonstrating the robustness of our main findings across arbitrary analytical thresholds. Our results suggest that core and non-core microbial communities are governed by distinct processes: non-core microbes were shaped primarily by deterministic forces, particularly heterogeneous selection driven by environmental variation along the land-use gradient, whereas core microbes were structured mainly by stochastic processes, with ecological drift playing a dominant role.

Communities of rat hosts and non-core microbes were highly modular, with a greater number of smaller modules compared to core microbes. This pattern aligns with the expectation that non-core microbes, which are more transient, respond strongly to selective pressures that, in turn, structure the host-microbe network. The microbiomes of rats in different modules varied primarily due to heterogeneous selection, whereas those within the same module were phylogenetically similar and shaped mainly by ecological drift. The strong influence of heterogeneous selection between modules suggests that environmental differences across land-use types act as filters, structuring non-core microbial communities by selecting for taxa best suited to specific conditions and potentially providing hosts with beneficial functions. These taxa are then clustered into modules reflecting the host-microbe interactions under distinct conditions. This is further supported by the statistically significant effect of land-use—particularly vegetation and elevation—on module composition along the gradient: sites with more similar environmental conditions tend to have more similar module composition, meaning that hosts in these areas share more similar microbes. The adaptable nature of the non-core microbiome may facilitate the success of rats as effective invaders across diverse land-use types [55].

Microbial communities comprise diverse microbes with varying attributes, such as differences in prevalence, suggesting that multiple ecological processes shape their assembly [13,56]. However, classifying microbes as core or non-core is often done arbitrarily, relying on threshold definitions [22]. A flexible threshold approach, as applied in our study, offers a more comprehensive understanding of microbial assembly and helps clarify the ecological forces shaping host–microbe networks. Our results indicate that while patterns shifted gradually along the core threshold gradient, they remained qualitatively consistent across different thresholds. This suggests that, at least in our system, selecting a core threshold above a relatively low prevalence ($\sim$0.2%) does not fundamentally alter the underlying assembly processes of microbial communities. In other words, microbes with a prevalence below 0.2% appear to be the most sensitive to selective pressures imposed by the environment.

These patterns can be further understood by examining specific microbial families. The most abundant families, such as *Lachnospiraceae*, *Muribaculaceae*, *Prevotellaceae*, and *Ruminococcaceae*, remain relatively consistent across all thresholds, supporting the host by aiding digestion and producing essential metabolites [57–59]. However, their distribution is not uniform, with certain families disproportionately abundant in specific modules. For example, *Prevotellaceae* plays a key role in fermentative metabolism, breaking down carbohydrates and proteins [60]. Specifically, the genus *Prevotella* helps digest plant-based polysaccharides and is associated with high-fiber diets [61]. This family is most abundant in a module that primarily consists of hosts from

herbaceous-dominated sites, suggesting a link between these microbes and the host diet. Additionally, many rarer microbial families are restricted to specific modules and sites (e.g., *Clostridiaceae* and *Spirochaetaceae*). These microbes may either provide hosts with specialized functions adapted to particular environmental conditions or simply reflect distinct environmental microbial communities [12,62].

Unlike non-core microbes, core microbial communities exhibited a less fragmented network structure, characterized by fewer but larger modules that persisted across the land-use gradient. At all core thresholds, ecological drift was the primary driver of host-microbe network structure, influencing microbiome composition both within and between modules. Species turnover is common in microbial communities, as random fluctuations in community composition occur even in the absence of deterministic forces [18]. Consequently, hosts with relatively higher taxonomic but not phylogenetic turnover were grouped into different modules. This suggests that despite compositional shifts in taxa (ASVs), the microbiome maintained its phylogenetic structure and potentially its functional integrity both within and across modules. This pattern supports the idea that core microbes play essential roles in host metabolism and overall function [23,25].

Notably, the composition of modules remained largely unchanged despite land-use variation, indicating that the core microbiome remained relatively stable across different environmental conditions. For example, the most abundant families—*Lactobacillaceae*, *Muribaculaceae*, and *Oscillospiraceae*—known for their essential roles in metabolism and fiber digestion [63,64], were consistently present across all modules and land-use types.

However, the impact of land-use on structuring the host-microbe network remains inconclusive for both core and non-core microbes, a pattern demonstrated by two observations. First, a single large module encompassed about half of the rat individuals and spanned all sites across most thresholds, indicating that many hosts clustered together despite varying environmental conditions. Second, the ordination test explained only $\sim$10% of the variation in module composition among non-core microbes across sites and was not significant for core microbes. These findings suggest that factors not included in our models, such as direct and indirect interactions with pathogens, may exert a stronger influence on selection pressures than land-use change [6,65].

One potential factor reducing the signal in the patterns we found is the fact that rats are mobile and can move within and between sites [34], potentially obscuring direct site-specific effects. The spatial mismatch between site-level environmental measurements and the finer scale at which microbes respond to their surroundings may cause environmental variables to inadequately capture the conditions experienced by microbial communities [66], ultimately contributing to the low variation explained in our RDA model. Directly measuring the host's diet could help address this issue by providing more precise and relevant information about the gut environment, which directly influences microbial community composition [16,67]. In addition, we lack direct and detailed information on microbial function. Future studies would benefit from precise functional measurements, obtained via whole-genome or shotgun sequencing, which provide a more direct and comprehensive understanding of microbiome function [68,69], to investigate whether these functional characteristics vary with land-use.

In conclusion, our findings indicate that core and non-core microbes are shaped by distinct

ecological processes, with non-core microbes responding more strongly to environmental selection, while core microbes remain relatively stable and are primarily influenced by stochastic forces. These insights contribute to a deeper understanding of microbiome ecology in wild animal populations, where environmental variability plays a key role in shaping microbial communities. By integrating network analysis with a flexible moving threshold to define core and non-core microbes, we provide new insights on microbiome assembly. This approach offers a powerful framework for disentangling the complex interactions between hosts, microbes, and their environment, with broad applicability across diverse systems and contexts, including land-use change, climate change, and biological invasions.

Acknowledgments

This work was supported by the Israel Science Foundation (grant 1281/20 to SP), the U.S. National Science Foundation (grant 2308460 to GT and CLN), the U.S.-Israel Binational Science Foundation (grant 2022721 to SP), the NIH-NSF-NIFA Ecology and Evolution of Infectious Diseases program (award R01-TW011493 to CLN), a Duke University Provost's Collaboratory Award to CLN, and the Human Frontier Science Program (grant RGY0064/2022 to SP). We are grateful to all members of our Madagascar Health team for their support and constructive feedback. We warmly thank the Ministère de l'Environnement et du Développement Durable, which has issued the research permits to enable carrying out the field work. We thank the Duke Lemur Center-SAVA Conservation for providing logistical assistance. We are also deeply appreciative of the communities and local authorities that hosted us and the many field assistants from the villages where the research was conducted.

Author contributions

Conceptualization: MM, SP; Data sampling: VS, TMR; Sample creation, preparation, and lab analysis: GT, GS, NB, AK; Molecular and bioinformatic analysis: GT; Formal analysis: MM; Funding acquisition: CLN, SP, GT; Writing – original draft: MM, SP; Writing – review and editing: MM, SP, GT, CLN, VS, TMR; Supervision: SP.

Data and code availability statement

Data and code are available on the GitHub repository https://github.com/Ecological-Complexity-Lab/Microbiome_Structure_Madagascar

References

1. Kéfi, S. *et al.* Advancing our understanding of ecological stability. *Ecol. Lett.* **22,** 1349–1356 (2019).
2. Bascompte, J. Disentangling the web of life. *Science* **325,** 416–419 (2009).
3. Thébault, E. & Fontaine, C. Stability of ecological communities and the architecture of mutualistic and trophic networks. *Science* **329,** 853–856 (2010).
4. Newbold, T. *et al.* Global effects of land use on local terrestrial biodiversity. *Nature* **520,** 45–50 (2015).
5. Tylianakis, J. M. & Morris, R. J. Ecological Networks Across Environmental Gradients. *Annu. Rev. Ecol. Evol. Syst.* **48,** 25–48 (2017).
6. Bernardo-Cravo, A. P., Schmeller, D. S., Chatzinotas, A., Vredenburg, V. T. & Loyau, A. Environmental Factors and Host Microbiomes Shape Host-Pathogen Dynamics. *Trends Parasitol.* **36,** 616–633 (2020).
7. Runghen, R., Poulin, R., Monlleó-Borrull, C. & Llopis-Belenguer, C. Network Analysis: Ten Years Shining Light on Host–Parasite Interactions. *Trends Parasitol.* **37,** 445–455 (2021).
8. McLaren, M. R. & Callahan, B. J. Pathogen resistance may be the principal evolutionary advantage provided by the microbiome. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **375,** 20190592 (2020).
9. Rosenberg, E. & Zilber-Rosenberg, I. The hologenome concept of evolution after 10 years. *Microbiome* **6,** 78 (2018).
10. Rosshart, S. P. *et al.* Wild mouse gut Microbiota promotes host fitness and improves disease resistance. *Cell* **171,** 1015–1028.e13 (2017).
11. Fackelmann, G. *et al.* Human encroachment into wildlife gut microbiomes. *Commun Biol* **4,** 800 (2021).
12. 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 (2018).
13. Karkman, A., Lehtimäki, J. & Ruokolainen, L. The ecology of human microbiota: dynamics and diversity in health and disease. *Ann. N. Y. Acad. Sci.* **1399,** 78–92 (2017).
14. Costello, E. K., Stagaman, K., Dethlefsen, L., Bohannan, B. J. M. & Relman, D. A. The application of ecological theory toward an understanding of the human microbiome. *Science* **336,** 1255–1262 (2012).
15. Lurgi, M., Thomas, T., Wemheuer, B., Webster, N. S. & Montoya, J. M. Modularity and predicted functions of the global sponge-microbiome network. *Nat. Commun.* **10,** 992 (2019).
16. Stapleton, T. E., Lindsey, L. M., Sundar, H. & Dearing, M. D. Rodents consuming the same toxic diet harbor a unique functional core microbiome. *Anim. Microbiome* **6,** 43 (2024).
17. Brice, K. L. *et al.* The Koala (Phascolarctos cinereus) faecal microbiome differs with diet in a wild population. *PeerJ* **7** (2019).

18. Zhou Jizhong & Ning Daliang. Stochastic Community Assembly: Does It Matter in Microbial Ecology? *Microbiol. Mol. Biol. Rev.* **81,** 10.1128/mmbr.00002–17 (2017).

19. Debray, R. *et al.* Priority effects in microbiome assembly. *Nat. Rev. Microbiol.* **20,** 109–121 (2022).

20. Toju, H., Vannette, R. L., Gauthier, M.-P. L., Dhami, M. K. & Fukami, T. Priority effects can persist across floral generations in nectar microbial metacommunities. *Oikos* **127,** 345–352 (2018).

21. Shade, A. & Handelsman, J. Beyond the Venn diagram: the hunt for a core microbiome: The hunt for a core microbiome. *Environ. Microbiol.* **14,** 4–12 (2012).

22. Risely, A. Applying the core microbiome to understand host-microbe systems. *J. Anim. Ecol.* **89,** 1549–1558 (2020).

23. Perlman, D. *et al.* Concepts and Consequences of a Core Gut Microbiota for Animal Growth and Development. *Annu Rev Anim Biosci* **10,** 177–201 (2022).

24. Neu, A. T., Allen, E. E. & Roy, K. Defining and quantifying the core microbiome: Challenges and prospects. *Proc. Natl. Acad. Sci. U. S. A.* **118** (2021).

25. Wallace, R. J. *et al.* A heritable subset of the core rumen microbiome dictates dairy cow productivity and emissions. *Sci. Adv.* **5,** eaav8391 (2019).

26. Jousset, A. *et al.* Where less may be more: how the rare biosphere pulls ecosystems strings. *ISME J.* **11,** 853–862 (2017).

27. Weitz, J. S. *et al.* Phage-bacteria infection networks. *Trends Microbiol.* **21,** 82–91 (2013).

28. Massol, F., Macke, E., Callens, M. & Decaestecker, E. A methodological framework to analyse determinants of host-microbiota networks, with an application to the relationships between Daphnia magna's gut microbiota and bacterioplankton. *J. Anim. Ecol.* **90,** 102–119 (2021).

29. Olesen, J. M., Bascompte, J., Dupont, Y. L. & Jordano, P. The modularity of pollination networks. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 19891–19896 (2007).

30. Dormann, C. F., Fründ, J. & Schaefer, H. M. Identifying causes of patterns in ecological networks: Opportunities and limitations. *Annu. Rev. Ecol. Evol. Syst.* **48,** 559–584 (2017).

31. Segar, S. T. *et al.* The Role of Evolution in Shaping Ecological Networks. *Trends Ecol. Evol.* **35,** 454–466 (2020).

32. Röttjers, L. & Faust, K. From hairballs to hypotheses–biological insights from microbial networks. *FEMS Microbiol. Rev.* **42,** 761–780 (2018).

33. Goodman, S. M. Rattus on Madagascar and the Dilemma of Protecting the Endemic Rodent Fauna. *Conserv. Biol.* **9,** 450–453 (1995).

34. Dammhahn, M., Randriamoria, T. M. & Goodman, S. M. Broad and flexible stable isotope niches in invasive non-native Rattus spp. in anthropogenic and natural habitats of central eastern Madagascar. *BMC Ecol.* **17,** 16 (2017).

35. Rahelinirina, S. *et al.* First isolation and direct evidence for the existence of large small-mammal reservoirs of Leptospira sp. in Madagascar. *PLoS One* **5,** e14111 (2010).

36. Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* **6,** 1621–1624 (2012).

37. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17,** 10–12 (2011).

38. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13,** 581–583 (2016).

39. Galai, G. *et al.* Regional processes shape the structure of rumen microbial co-occurrence networks. *Ecography (Cop.)* (2024).

40. Martiny, J. B. H., Jones, S. E., Lennon, J. T. & Martiny, A. C. Microbiomes in light of traits: A phylogenetic perspective. *Science* **350,** aac9323 (2015).

41. 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 (2008).

42. Farage, C., Edler, D., Eklöf, A., Rosvall, M. & Pilosof, S. Identifying flow modules in ecological networks using Infomap. *Methods Ecol. Evol.* (2021).

43. Stegen, J. C., Lin, X., Konopka, A. E. & Fredrickson, J. K. Stochastic and deterministic assembly processes in subsurface microbial communities. *ISME J.* **6,** 1653–1664 (2012).

44. Stegen, J. C. *et al.* Quantifying community assembly processes and identifying features that impose them. *ISME J.* **7,** 2069–2079 (2013).

45. Arnault, G., Mony, C. & Vandenkoornhuyse, P. Plant microbiota dysbiosis and the Anna Karenina Principle. *Trends Plant Sci.* **28,** 18–30 (2023).

46. Jiao, S., Yang, Y., Xu, Y., Zhang, J. & Lu, Y. Balance between community assembly processes mediates species coexistence in agricultural soil microbiomes across eastern China. *ISME J.* **14,** 202–216 (2020).

47. Peng, Z. *et al.* Land conversion to agriculture induces taxonomic homogenization of soil microbial communities globally. *Nat. Commun.* **15,** 3624 (2024).

48. Stegen, J. C., Lin, X., Fredrickson, J. K. & Konopka, A. E. Estimating and mapping ecological processes influencing microbial community assembly. *Front. Microbiol.* **6,** 370 (2015).

49. Legendre, P., Legendre, L. & Legendre, L. F. *Numerical Ecology* 3rd ed. (Elsevier, 2012).

50. R Core Team. *R: A Language and Environment for Statistical Computing* Vienna, Austria, 2024.

51. Ning, D. *et al.* A quantitative framework reveals ecological drivers of grassland microbial community assembly in response to warming. *Nat. Commun.* **11,** 4717 (2020).

52. Oksanen, J. *et al. vegan: Community Ecology Package* 2024.

53. Nemergut, D. R. *et al.* Patterns and processes of microbial community assembly. *Microbiol. Mol. Biol. Rev.* **77,** 342–356 (2013).

54. Berg, G. *et al.* Microbiome definition re-visited: old concepts and new challenges. *Microbiome* **8** (2020).

55. Wagener, C., Mohanty, N. P. & Measey, J. *The gut microbiome facilitates ecological adaptation in an invasive vertebrate* 2020.

56. Araujo, G., Montoya, J. M., Thomas, T., Webster, N. S. & Lurgi, M. A mechanistic framework for complex microbe-host symbioses. *Trends Microbiol.* **33,** 96–111 (2024).

57. Chung, Y. W., Gwak, H.-J., Moon, S., Rho, M. & Ryu, J.-H. Functional dynamics of bacterial species in the mouse gut microbiome revealed by metagenomic and metatranscriptomic analyses. *PLoS One* **15,** e0227886 (2020).

58. Meehan, C. J. & Beiko, R. G. A phylogenomic view of ecological specialization in the Lachnospiraceae, a family of digestive tract-associated bacteria. *Genome Biol. Evol.* **6,** 703–713 (2014).

59. Biddle, A., Stewart, L., Blanchard, J. & Leschine, S. Untangling the genetic basis of fibrolytic specialization by Lachnospiraceae and Ruminococcaceae in diverse gut communities. *Diversity (Basel)* **5,** 627–640 (2013).

60. Betancur-Murillo, C. L., Aguilar-Marín, S. B. & Jovel, J. Prevotella: A key player in ruminal metabolism. *Microorganisms* **11,** 1 (2022).

61. Gellman, R. H. *et al.* Hadza Prevotella require diet-derived microbiota-accessible carbohydrates to persist in mice. *Cell Rep.* **42,** 113233 (2023).

62. Raulo, A. *et al.* Social and environmental transmission spread different sets of gut microbes in wild mice. *Nat Ecol Evol* (2024).

63. Ardalani, O, Phaneuf, P. V., Mohite, O. S., Nielsen, L. K. & Palsson, B. O. Pangenome reconstruction of Lactobacillaceae metabolism predicts species-specific metabolic traits. *mSystems* **9,** e0015624 (2024).

64. Zhu, Y. *et al.* Exploration of the Muribaculaceae family in the gut Microbiota: Diversity, metabolism, and function. *Nutrients* **16,** 2660 (2024).

65. Schmid, D. W. *et al.* A framework for testing the impact of co-infections on host gut microbiomes. *Anim Microbiome* **4,** 48 (2022).

66. Ladau, J. & Eloe-Fadrosh, E. A. Spatial, Temporal, and Phylogenetic Scales of Microbial Ecology. *Trends Microbiol.* **27,** 662–669 (2019).

67. Carmody, R. N. *et al.* Diet dominates host genotype in shaping the murine gut microbiota. *Cell Host Microbe* **17,** 72–84 (2015).

68. Lemanceau, P., Blouin, M., Muller, D. & Moënne-Loccoz, Y. Let the Core Microbiota Be Functional. *Trends Plant Sci.* **22,** 583–595 (2017).

69. Shafquat, A., Joice, R., Simmons, S. L. & Huttenhower, C. Functional and phylogenetic assembly of microbial communities in the human microbiome. *Trends Microbiol.* **22,** 261–266 (2014).

Supplementary information

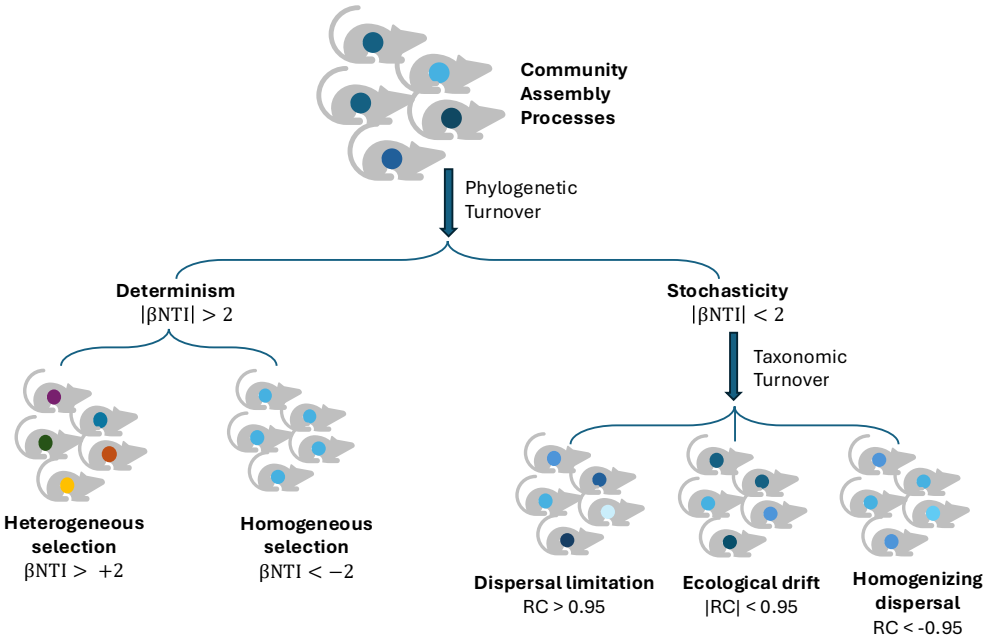

Figure S1: Inference of Community Assembly Processes. For core and non-core groups, we inferred community assembly processes using βNTI and RC following [43,44]. Values of $|\beta\text{NTI}| > 2$ indicate deterministic processes: positive values suggest heterogeneous selection, while negative values indicate homogeneous selection. When $|\beta\text{NTI}| < 2$, stochastic processes dominate. Within this range, dispersal limitation is inferred if $RC > +0.95$, while homogenizing dispersal is inferred if $RC < -0.95$. When both metrics fall within neutral ranges, ecological drift governs community composition through random fluctuations. Colors represent distinct microbial community compositions, with similar shades indicating phylogenetically related communities.

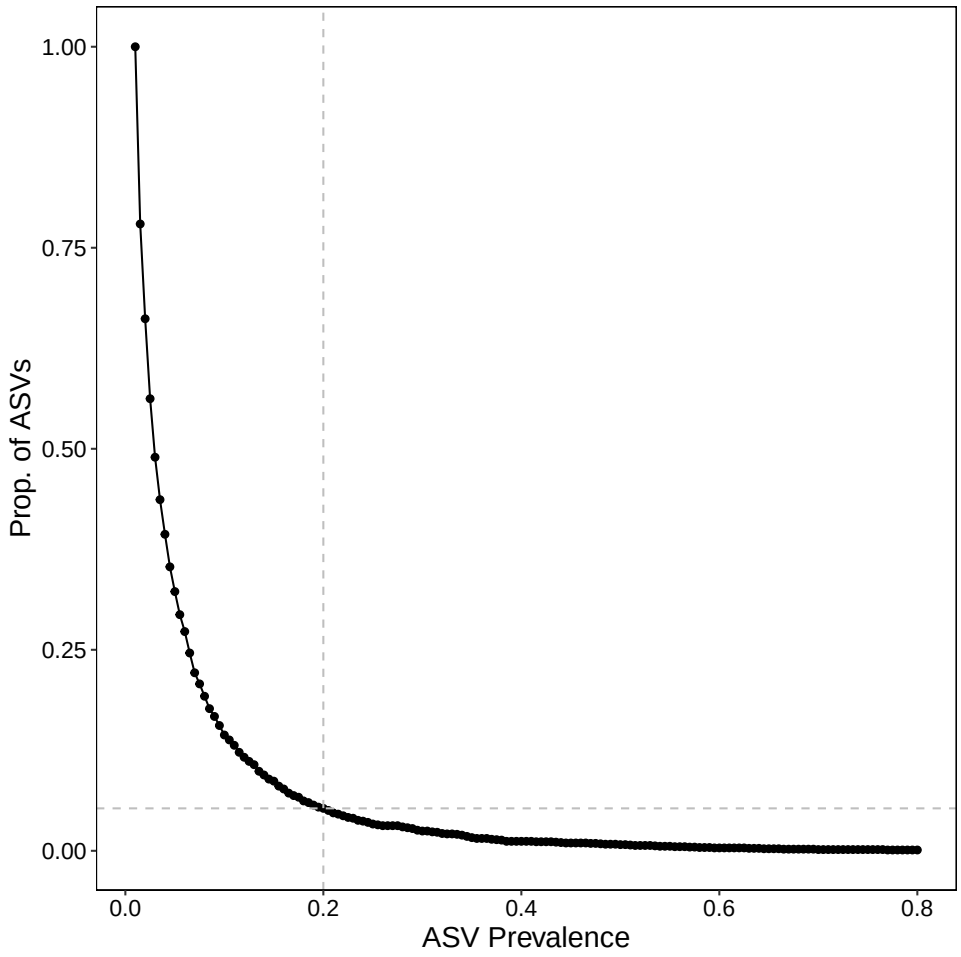

Figure S2: Proportion of ASVs in every prevalence threshold. The proportion of ASVs with prevalence exceeding thresholds ranging from 0.01 to 0.8, in increments of 0.01. For instance, when the threshold is 0, all the ASVs are included, but only 5.3% of the ASVs occur in at least 20% of the hosts (threshold of 0.2). The gray dashed lines indicate the proportion of ASVs with a prevalence greater than 0.2.

Table S1: Summary of site characteristics.

Site	Village	Land-use Type	No. of ASVs	No. of Rats	No. of Other SM	Vegetation PC1	Vegetation PC2	Distance to Village Center (m)	Elevation
1	A	semi-intact forest	611	8	24	-4.665	1.544	2162.7	476.2
2	M	semi-intact forest	521	8	37	-3.915	1.440	3650.7	275.6
3	A	secondary forest	1180	43	39	-1.669	-1.266	1614.1	534.2
4	M	secondary forest	1140	30	46	-1.979	-0.005	2887.1	202.0
5	S	secondary forest	1341	54	46	-0.919	-2.090	2215.4	783.8
6	A	brushy regrowth	1067	30	55	2.766	-0.063	1249.9	388.3
7	M	brushy regrowth	1462	60	65	-0.165	-2.082	2419.4	148.4
8	S	brushy regrowth	1398	53	99	2.115	-0.442	734.1	593.3
9	A	agriculture	932	21	59	-0.230	-0.282	1007.3	372.2
10	M	agriculture	1382	63	93	1.672	-0.308	1800.9	131.0
11	S	agriculture	1475	73	69	1.905	0.624	1014.5	601.0
12	A	agroforest	1209	37	69	1.401	-0.227	617.5	265.3
13	M	agroforest	1298	60	75	1.625	0.371	600.6	95.0
14	S	agroforest	1429	69	38	0.896	0.031	2016.8	737.8
15	A	flooded rice	860	19	44	2.042	1.170	513.1	105.5
16	M	flooded rice	1468	69	82	1.758	0.367	1352.6	118.1
17	S	flooded rice	524	7	75	1.110	2.066	496.0	481.4
18	A	village	1359	72	14	-	-	0.0	135.8
19	M	village	1315	46	14	-	-	0.0	93.7
20	S	village	1191	33	22	-	-	0.0	551.3

SI note 1: Study site and small mammals sampling

Small mammals were collected in the vicinity of three villages in the SAVA region of northeast Madagascar, in the surroundings of Marojejy National Park. The village M (14.477049° S, 49.8147° E) was sampled between October 2019 and September 2020. A second village, S (14.607567° S, 49.647759° E), was sampled between November 2020 and September 2021, while a third village, A (14.397276° S, 49.8820° E), was sampled between October 2021 and August 2022. In each village, seven sites were sampled along a degradation gradient: (1) semi-intact forest inside the national park, (2) secondary forest, (3) brushy regrowth (savoka), (4) agroforest (vanilla plantation), (5) mixed agriculture (sugarcane/coffee plantation), (6) flooded rice, and (7) the village (**Table S1**). At the semi-intact forest site in the village S, no rats were captured, so the site was excluded from the analysis, resulting in a total of 20 sites. Sites in each village setting were located ∼500 m apart.

For sampling small mammals a 110 m X 110 m grid of 121 live traps (11x11) was established, including 97 Sherman (H. B. Sherman Traps, Inc., Tallahassee, Florida, model LFA and XLK), and 24 Tomahawk (Tomahawk Live Trap, Hazelhurst, Wisconsin, model 201), placed 10 m apart and baited with peanut butter. Additionally, two pitfall lines were installed between 20-50 m outside of the grid, running in parallel to the grid edge. Each pitfall line was 100 m in length, with 11 buckets dug into the ground and placed every 10 m, and an 80 cm high vertical plastic fencing oriented and stapled to vertical stakes and bisecting each bucket. Each plot was sampled for six consecutive nights three times during the sampling period.

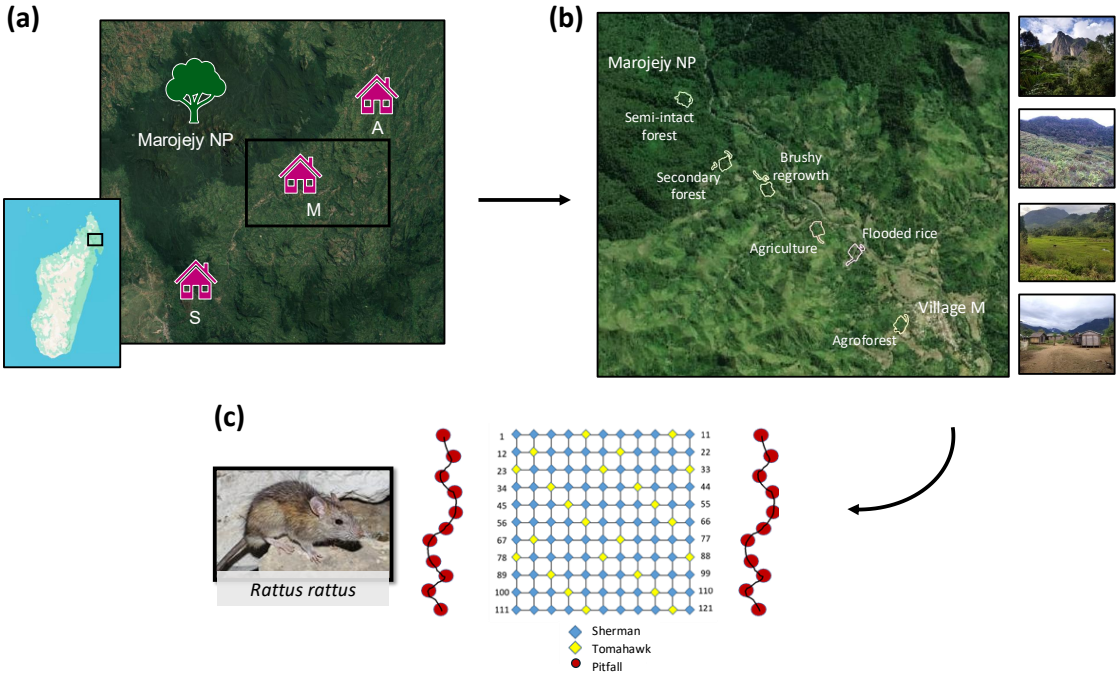

Figure S3: Study site and sampling scheme. **(a)** Sampling was conducted in northeastern Madagascar, across three villages near Marojejy National Park. **(b)** In each village, seven distinct land-use types were sampled. The map illustrates the village M as an example. The images depict typical landscapes from top to bottom: semi-intact forest, brushy regrowth, flooded rice fields, and village plots. **(c)** In each plot, an 11×11 trapping grid consisting of a mix of Sherman and Tomahawk traps was installed, along with two pitfall lines. A total of 855 *Rattus rattus* individuals were captured.

SI note 2: Comparing microbial taxonomy of core and non-core groups

To examine taxonomic variation between core and non-core groups, we aggregated ASVs at the family level for each core threshold. ASVs with unidentified families were excluded, resulting in the analysis of 1,770 ASVs (90.72% of all ASVs) classified into 55 families. For each microbial group at each core threshold, we summed the read counts for each family across all hosts. Relative read abundance was then calculated by dividing by the total number of reads for each microbial group and threshold combination. To assess differences in taxonomy, we calculated the Bray-Curtis Dissimilarity Index between core and non-core groups at each threshold. The groups differed significantly, with the Bray-Curtis index increasing from 0.35 at a threshold of 0.1 to 0.84 at a threshold of 0.45, indicating that core microbes belong to different families than non-core microbes (**Figure S4**). A similar pattern was found for the relative abundance of microbial families (**Figure S5**).

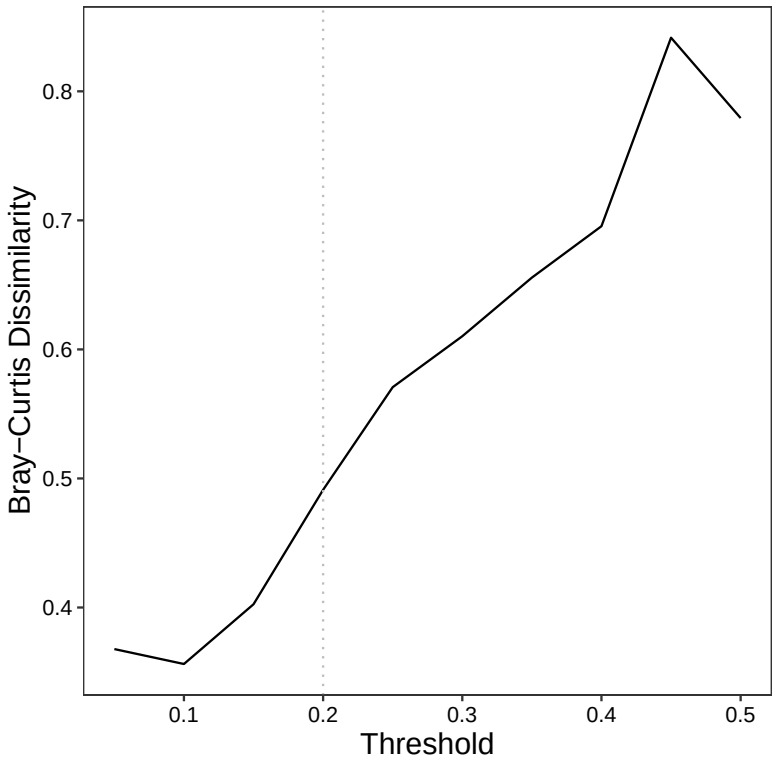

Figure S4: Bray-Curtis Dissimilarity Index between core and non-core microbial groups. Bray-Curtis was calculated at the aggregated family level across all core thresholds.

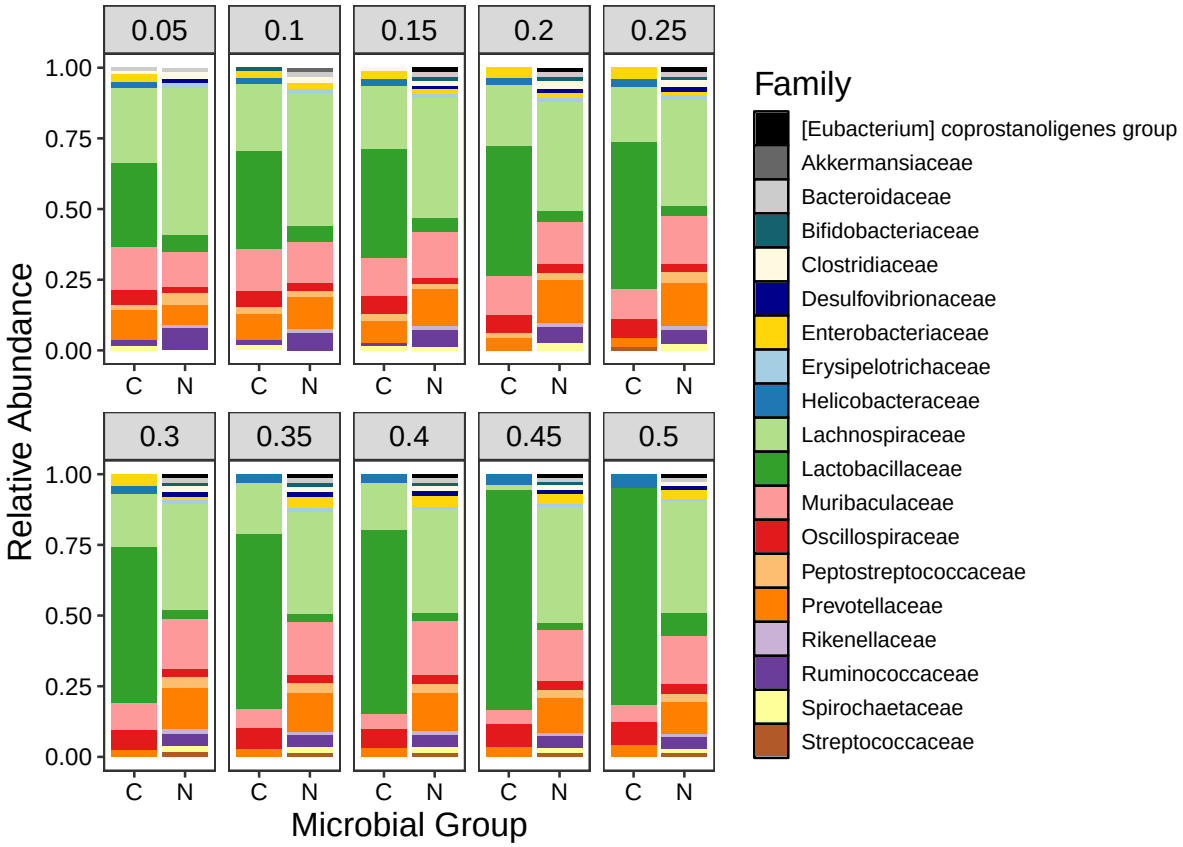

Figure S5: Taxonomical classification of core and non-core microbial groups at the family level. For better visualization, only families with a relative abundance greater than 1% are included. [C]=Core; [N]=Non-core.

SI note 3: Measuring environmental gradients across sites

To explore land-use change we measured three environmental gradients across sites: vegetation, elevation, and the distance from the village center (**Table S1**). These variables collectively capture much of the natural and anthropogenic variation along the land-use gradient. Elevation was recorded at the center of each site. The distance to the nearest village was calculated as the shortest distance from the village center to the site center, using the distHaversine function from the *geosphere* R package.

In addition, we measured vegetation attributes in each site using 16 plots (5m $\times$ 5m) within the sampling grid, conducting measurements three times during the sampling period. At each plot, we assessed eight vegetation attributes: (1) Number of trees [n_trees] (2) Number of dead logs [n_logs] (3) Tree diameter at breast height [tree_dbh] (4) Tree height (5) Percent canopy cover [m_canopy_cv] (6) Number of Liana sp. [n_liana] (7) Herbaceous height [m_herb_ht] (8) Percent herbaceous cover [m_herb_cv]. We averaged the measurements across all plots to calculate mean values for each site. To explore vegetation variation between sites, we conducted a principal component analysis (PCA). Prior to analysis, all variables were centered at 0 and rescaled to have unit variance. The first two principal components explained 83.39% of the variation across sites (PC1: 67.06%, PC2: 16.33%). Vegetation PC1 divides the more natural sites (Semi-intact forest and Secondary forest) from the more disturbed sites. Accordingly, PC1 is positively correlated with herbaceous cover and height and negatively correlated with tree-related variables.

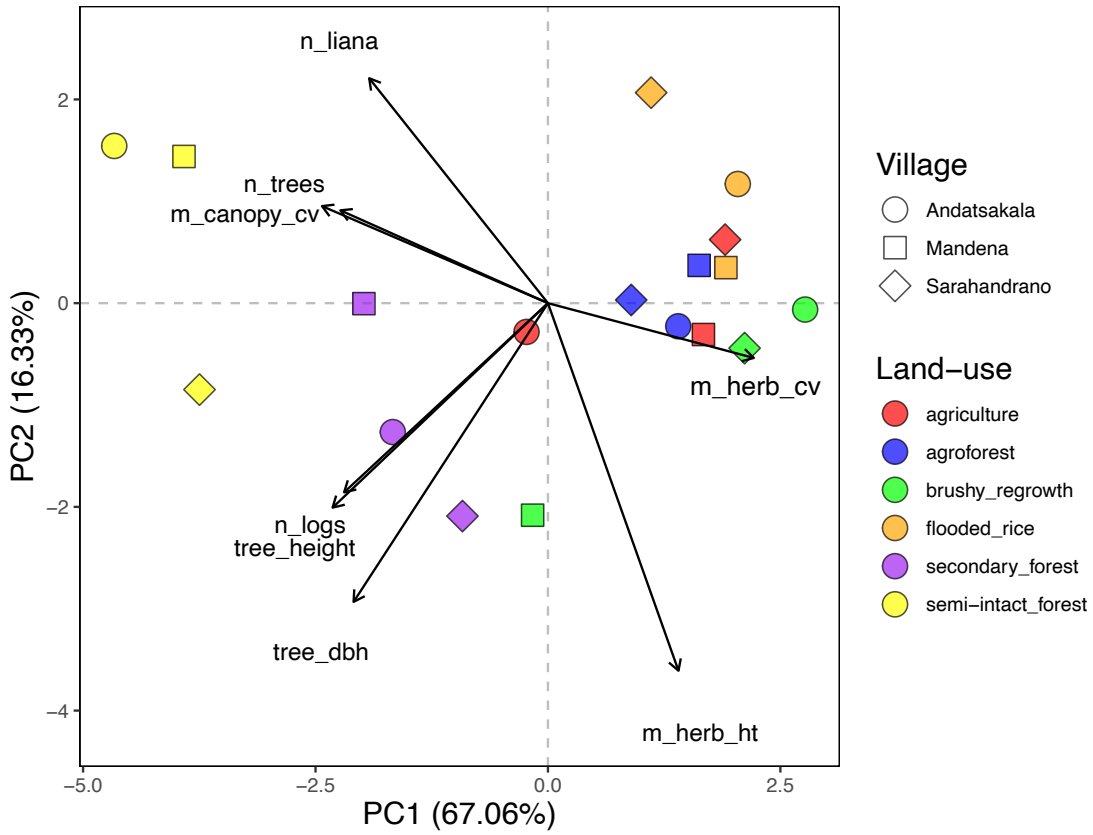

Figure S6: Vegetation PCA across land-use types. Each point in the ordination plot represents a site, with the shape indicating the village and the color indicating the land-use type. The length and direction of the arrows indicate the contribution of each vegetation variable to the first two PCA components (PC1 and PC2).