nature ecology & evolution

Article

https://doi.org/10.1038/s41559-023-02149-y

Multidimensional specialization and generalization are pervasive in soil prokaryotes

Received: 8 November 2022

Accepted: 4 July 2023

Damian J. Hernandez [1,2] ✉, Kasey N. Kiesewetter [1,2], Brianna K. Almeida [1], Daniel Revillini [1] & Michelle E. Afkhami [1]

Habitat specialization underpins biological processes from species distributions to speciation. However, organisms are often described as specialists or generalists based on a single niche axis, despite facing complex, multidimensional environments. Here, we analysed 236 environmental soil microbiomes across the United States and demonstrate that 90% of >1,200 prokaryotes followed one of two trajectories: specialization on all niche axes (multidimensional specialization) or generalization on all axes (multidimensional generalization). We then documented that this pervasive multidimensional specialization/generalization had many ecological and evolutionary consequences. First, multidimensional specialization and generalization are highly conserved with very few transitions between these two trajectories. Second, multidimensional generalists dominated communities because they were 73 times more abundant than specialists. Lastly, multidimensional specialists played important roles in community structure with ~220% more connections in microbiome networks. These results indicate that multidimensional generalization and specialization are evolutionarily stable with multidimensional generalists supporting larger populations and multidimensional specialists playing important roles within communities, probably stemming from their overrepresentation among pollutant detoxifiers and nutrient cyclers. Taken together, we demonstrate that the vast majority of soil prokaryotes are restricted to one of two multidimensional niche trajectories, multidimensional specialization or multidimensional generalization, which then has far-reaching consequences for evolutionary transitions, microbial dominance and community roles.

In nature, organisms navigate complex environments by embracing diverse conditions (generalists) or utilizing a smaller portion of available resources and habitats (specialists). The extent to which organisms specialize or generalize is central to many ecological and evolutionary processes such as species distributions, rates of speciation and resilience to disturbances[1,2]. However, studies of specialization and generalization have historically focused on only a single environmental axis which overlooks the reality that organisms experience complex, heterogeneous environments that change on many axes through space and time. Omitting environmental complexity disregards relationships across niche axes and the impact they have on organisms with important implications for their ecology, evolution and conservation.

---

[1] Department of Biology, University of Miami, Coral Gables, FL, USA. [2] These authors contributed equally: Damian J. Hernandez, Kasey N. Kiesewetter. ✉e-mail: d.hernandez18@umiami.edu

For example, the more restrictive habitat requirements of multidimensional specialists (organisms that specialize across many niche dimensions) compared with single-axis specialists could make multidimensional specialists especially susceptible to the disturbances and intensifying stress of the Anthropocene. Despite growing interest in multidimensional specialization and generalization[2–7], there are few, if any, empirical tests of the prevalence of multidimensional specialization and generalization and their consequent effects on evolutionary trajectories, species dominance and ecological communities.

Although plant and animal studies have been important to our foundational knowledge of specialization and generalization[1,8,9], investigating multidimensional niche processes in macro-organisms can be prohibitively labour intensive for even just a few species or axes[4,5,10]. By contrast, relatively recent advances in next-generation sequencing now allow surveys of entire, natural microbial communities (hundreds to thousands of species) across multiple environmental axes, locations and scales with relatively low time and resource costs. Therefore, the recent rise of microbiome studies provides a promising new avenue for efficiently investigating multidimensional specialization and multidimensional generalization and their ecological and evolutionary consequences for thousands of taxa.

Here, we quantify multidimensional specialization and generalization in thousands of co-occurring microbes at local and continental scales. To do this, we analysed the niche breadths of >1,200 prokaryotes from soil microbiomes across the continental United States along environmental axes that include some of the most important abiotic factors known to shape prokaryotic soil communities (soil pH, litter depth, soil moisture and soil temperature as well as per cent soil nitrogen, per cent soil carbon and carbon/nitrogen ratio for a subset of sites)[11]. We confirm that these soil parameters are meaningful niche axes for prokaryotes in our analyses by demonstrating that they explain approximately 67% of community variation (Methods). This large-scale investigation was made possible by the launch (January 2021) of the National Ecological Observatory Network (NEON), which is the National Science Foundation's flagship ecological repository of biological, climatic and environmental information across the continental United States and is already among the world's largest repositories of soil microbiome data. In this study, we determined (1) the frequencies of multidimensional specialization (specialization across all characterized niche dimensions) and multidimensional generalization (generalization across all characterized niche dimensions) within microbial communities. We then explored the ecology and evolution of multidimensional specialists and generalists, asking (2) whether evolutionary transitions from multidimensional specialist to multidimensional generalist or vice versa are more common, (3) which group is dominant within microbial communities and (4) which group plays more central roles in their communities. Our study reveals a new ecological principle of prokaryotic niches in which nearly all soil taxa follow only one of two opposing trajectories—multidimensional specialization or multidimensional generalization—and highlights how constraining taxa to these two trajectories has meaningful consequences for microbial ecology and evolution.

Ubiquity of multidimensional specialization and generalization

Our evaluation of 236 microbial communities from 30 sites across the United States (Extended Data Fig. 1) demonstrated that multidimensional shaping of ecological niches of microbes is ubiquitous, with multidimensional generalization occurring more commonly than multidimensional specialization. We calculated niche breadths across all axes using the standard metric of 'proportional similarity', which accounts for species resource use and how common those resources are in the environment[12,13]. Microbial taxa ('species' as identified by mapping to the GreenGenes database) displayed a bimodal distribution in niche breadth that we then categorize into 'specialist' and 'generalist' categories based on the local minima between the two peaks in niche breadth (a heuristic delineation between low and high niche breadth; Methods and Fig. 1e,f). Not only did ~90% of prokaryotes (1,090 of 1,230 taxa) in the 236 communities show consistent degrees of specialization or generalization across all the axes investigated, but these relationships were stronger than environmental correlations among axes and robust to different analysis decisions. Specifically, we found that ~57% of prokaryotes (697 of 1,230 taxa) were multidimensional generalists and ~32% were multidimensional specialists across the four main environmental axes (soil pH, moisture, temperature and litter depth). The bimodal distribution with ~90% of taxa being multidimensional specialists or generalists was robust to other filtering cut-offs (for example, including taxa that occur in only 1% or 5% of samples; Supplementary Data 1), and the lower number of multidimensional specialists is unlikely to result from insufficient sequencing depth because rarefaction curves in our analyses consistently plateaued (Extended Data Fig. 2a). Also, when using the more lenient filtering criteria of 1% and 5% occupancy, the bimodal distribution of multidimensional generalists and multidimensional specialists still holds with 48.1% and 52.4% of taxa identified as multidimensional generalists and 42.3% and 36.9% of taxa identified as multidimensional specialists for the 1% and 5% occupancy cut-offs, respectively. In all cases, mixed specialization/generalization was unusual with only ~10% of taxa showing a mixture of generalization and specialization across dimensions regardless of filtering cut-off. Our results were also robust to using an alternative niche breadth metric that accounts for the similarities among habitats taxa occupy and the range of environmental conditions (niche range and Levin's niche breadth; Methods and Supplementary Data 1). These results highlight that multidimensional generalization and multidimensional specialization are opposing niche trajectories for soil prokaryotes, supporting that when taxa generalize or specialize across one niche dimension, generalization/specialization impacts the other niche axes.

In addition, microbial niche breadth on one environmental axis explained ~80% of the variation in all other niche breadths (Spearman's $\rho = 0.94 \pm 0.005$, mean ± s.e.m.; Fig. 1), further indicating that niche specialization and generalization occur together along multiple environmental axes. This conclusion was supported by several additional lines of evidence. First, the relationships among niche breadths on different environmental axes when calculated at the continental scale were more than ten times stronger on average (range: 2–54) than the correlations among environmental axes (niche breadth Spearman's $\rho = 0.94 \pm 0.005$ versus environmental $|\text{Spearman's } \rho| = 0.37 \pm 0.078$, mean ± s.e.m.; $W = 36$, $P = 0.0022$; Fig. 1b–d). Second, multidimensional specialization and generalization were still ubiquitous when analyses were done at a higher taxonomic resolution with >14,000 exact sequence variant (ESV) taxa (Spearman's $\rho = 0.82 \pm 0.010$; Extended Data Fig. 3a–d). Third, multidimensional specialization and generalization remained pervasive when per cent carbon, per cent nitrogen and carbon to nitrogen ratios (additional major determinants of soil prokaryotic community composition available for a subset of sites) were included in our analyses (Spearman's ρ for analysis with seven niche axes = 0.96 ± 0.001; Extended Data Fig. 3e–h). These results show that multidimensional generalization and specialization are biologically important rather than simple byproducts of correlations between environmental axes, and are consistent across different taxonomic scales and different types of niche axes.

In addition to their ubiquity at the continental scale, multidimensional specialization and multidimensional generalization were also prevalent in soils at local geographic scales. When we determined the niche breadth of taxa using the 21 individual sites with sufficient within-site replication, 75.2% ± 1.8% of taxa were specialized or generalized across all environmental axes with multidimensional generalists approximately three times more common than multidimensional specialists, again indicating the importance of multidimensional shaping

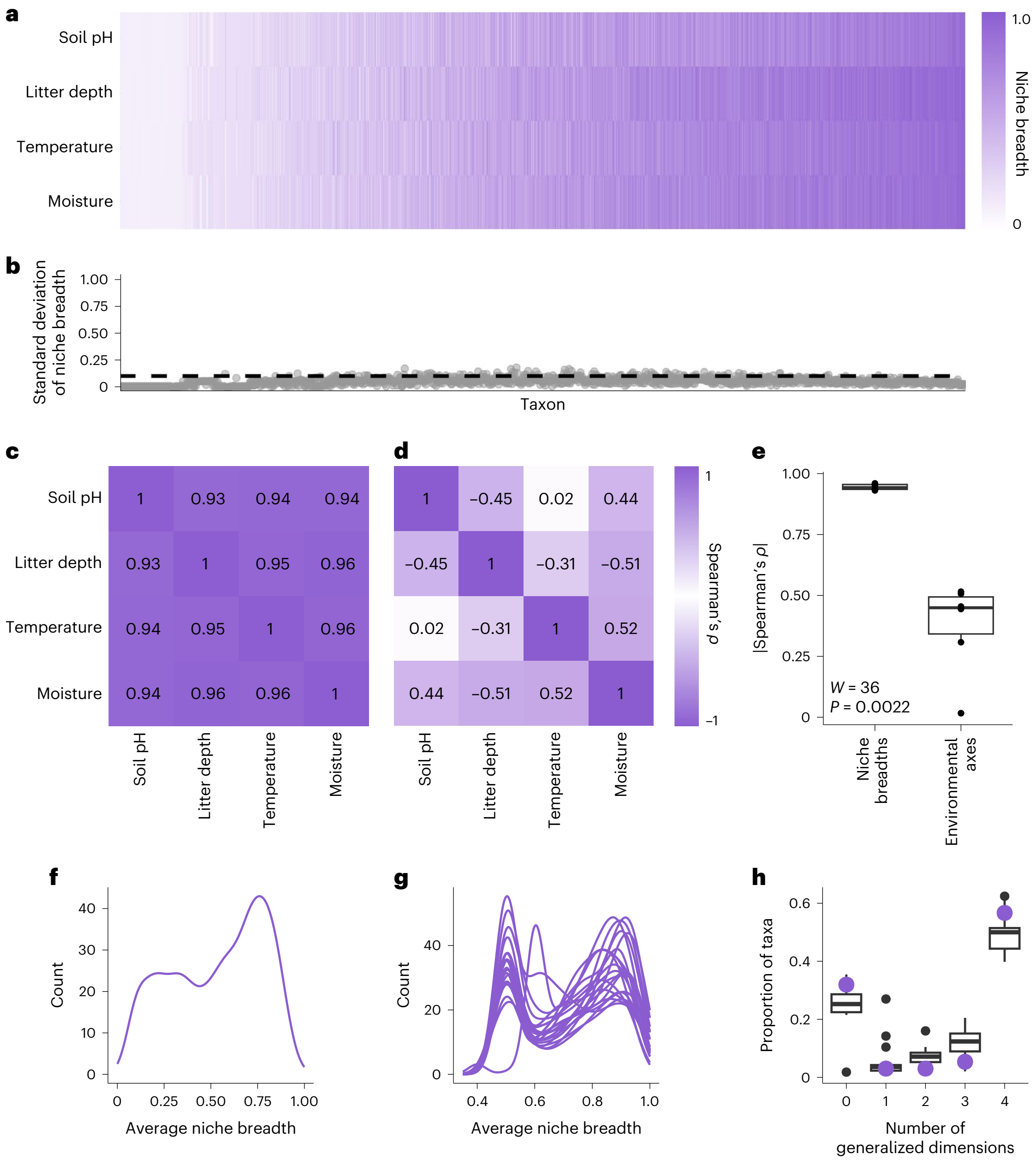


Fig. 1 | Multidimensional generalization and specialization are widespread in prokaryotes and exceed what can be explained by environmental correlations. a, Heatmap of niche breadths highly correlated for 1,230 prokaryotic taxa (*x* axis) along four environmental axes (also true with seven axes; Extended Data Fig. 3), indicating that most taxa (57%) are generalists across all axes and most of the remaining taxa (32%) are specialists across all axes. Taxa are ordered by average niche breadth increasing from left to right for visualization. **b**, Standard deviation of niche breadth across all four axes demonstrates how consistent niche breadth is across niche dimensions. The dashed line represents a standard deviation of 0.1 in proportional similarity. Taxa are in the same order as in the heatmap. **c**, Heatmap of Spearman's *ρ* from correlations between niche breadths of 1,230 microbial taxa along different axes. **d**, Heatmap of Spearman's *ρ* from correlations between environmental axes across 236 individual plots. **e**, Comparison of the absolute values of Spearman's *ρ* from correlations between niche breadths and correlations between environmental axes, demonstrating that niche breadth correlations are significantly stronger than correlations in environmental variation among axes. Significance determined by two-tailed Mann–Whitney *U*-test. Box plots show the median (middle line) and interquartile range (box). **f**, Average niche breadths of taxa (mean breadth calculated across all niche dimensions) showed a bimodal distribution between specialist and generalist at a continental scale. **g**, Similarly, the average niche breadth of taxa calculated within each individual site also demonstrated a consistently bimodal distribution between specialist and generalist across sites. **h**, Distribution of multidimensional specialists/generalists is consistent across scales. The box plot shows the distribution of 1,230 taxa that generalize on *n* dimensions at the local scale (across 21 independent sites) with the purple points displaying the proportion of 1,230 taxa that generalize on *n* dimensions at the continental scale. Box plots show the median (middle line) and interquartile range (box).

of ecological niches and that multidimensional generalization is more common than multidimensional specialization. Further, >80% of sites (17 of 21, *t*-tests corrected for multiple comparisons) showed significantly stronger niche breadth relationships (5.83 ± 1.68 times greater across all sites) than environmental correlations. In short, a substantial part of the multidimensional specialization and generalization is occurring independently from relationships between environmental axes (Extended Data Fig. 4). Our continental- and local-scale analyses together emphasize that multidimensional specialization and generalization is a widespread, scale-independent ecological phenomenon in prokaryotic soil communities.

Evolution of multidimensional specialization and generalization

Importantly, we found that this pervasive multidimensional specialization and generalization also have diverse implications for prokaryotic ecology and evolution, including for evolutionary transitions between generalist and specialist states, species dominance and organisms' roles within microbial networks.

First, the bifurcating trajectories of multidimensional specialists and generalists were further supported by low transition rates between specialist and generalist states, with specialist-to-generalist and generalist-to-specialist transitions occurring ~90% less often than expected by chance. Specifically, permutational tests comparing observed transition rates with null transition distributions demonstrated that both specialist-to-generalist and generalist-to-specialist transitions occurred ~90% less often than expected by chance across all 100 phylogenetic trees tested (false discovery rate (FDR) < 0.05 (specialist-to-generalist transitions), Methods; FDR < 0.05 (generalist-to-specialist transitions), Fig. 2a and Extended Data Fig. 5). Limited transitions between niche trajectories could be due to generalization and specialization requiring specific, opposing heritable ecological strategies and adaptations. In other words, the relatively few transitions between niche trajectories may reflect a fitness landscape in which most prokaryotes occupy one of two types of peaks (a specialist or generalist peak) with few intermediate alternatives. Mapping other dichotomous ecological frameworks (for example, copiotrophy versus oligotrophy, *r*/*K* selection theory) to these two alternative strategies would be important for building generalizable frameworks of prokaryotic ecology by improving the predictive ability of how prokaryotes navigate their environments (for example, a taxon identified as generalist would be predicted to invest more in growth rates and replication over resilience). We find signatures of this dichotomy in the phylogenetic conservation of niche breadth with extreme specialists/generalists far more likely to have close relatives that are also as specialized/generalized (Fig. 2b and Methods), which is what would be expected if following specialist/generalist strategies makes it difficult to switch trajectories. In fact, the relationship between conservation of niche breadth and the actual value of niche breadth follows a more quadratic (37.4% ± 0.02% adjusted R^2, mean ± s.e.m.) relationship rather than a linear relationship (25.4% ± 0.01% adjusted R^2, mean ± s.e.m.) in all 100 phylogenetic trees tested (Fig. 2c and Extended Data Fig. 6), indicating that the more specialized taxa are, the less likely they will evolve into generalists and vice versa.

These results suggest that multidimensional specialization and generalization may require incompatible strategies and adaptations so that becoming more specialized makes it more difficult to generalize (and vice versa). For example, shifts to a multidimensional specialist lifestyle may be difficult for generalists because specialization can require changes in multiple genes across the organism's genome and complex genetic regulations susceptible to mismatches in gene–gene interactions. These challenges could cause failure to specialize or to survive after specialization. Conversely, specialization may require adaptive strategies to persist that are incompatible with a generalized niche. For example, multidimensional specialists could experience a trade-off in which, instead of improving fitness by having large populations—explored below—they improve fitness by increasing their endurance through investing in, for example, more durable spores or other stress-tolerant traits[1,14] such as the endospore-forming bacteria order Clostridiales[15] that is enriched for multidimensional specialists. Other obstacles could also exist that limit specialist-to-generalist transitions such as Dollo parsimony because regaining lost functions can be difficult[16–18] and multidimensional specialization could result in a loss of genes/traits essential to persisting in multiple environments, thereby hindering transitions to generalist identities. This possibility is reinforced by previous work demonstrating that specialist microbes tend to have smaller genomes than generalists[19,20].

Overall, our findings highlight multidimensional generalization and specialization as an intrinsic feature of soil prokaryotes that probably represent two opposing evolutionary trajectories. Because

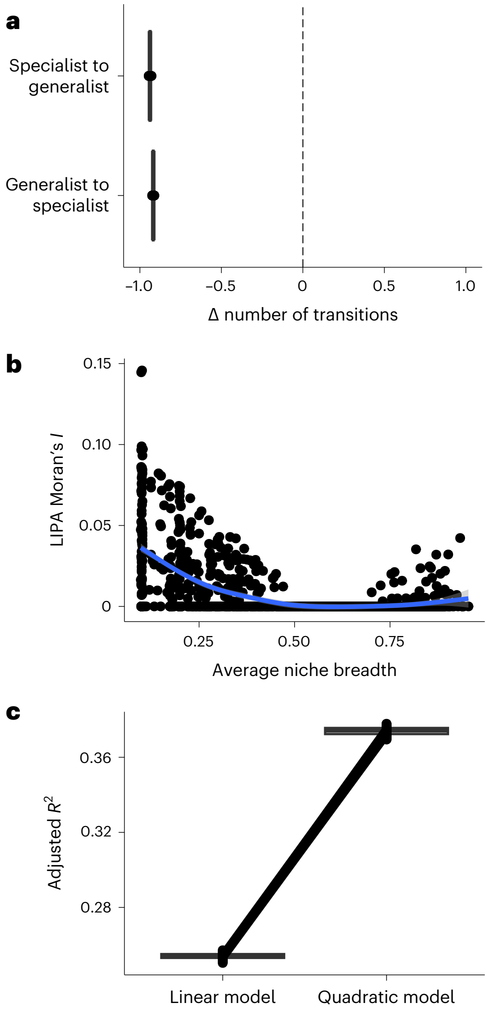


Fig. 2 | Multidimensional niche trajectories are phylogenetically conserved. **a**, Per cent difference (Δ) between number of transitions in 100 observed trees versus the null expectations (average number of transitions in the corresponding 1,000 randomized trees). The dashed line indicates 0 transitions from the null expectation. Distributions of transitions in randomized trees were consistent across all 100 trees (Kolmogorov–Smirnov test) and are displayed in Extended Data Fig. 5. Each point is overlayed on a box plot showing the median (middle line) and interquartile range (box). **b**, LIPA Moran's *I* (local measure of phylogenetic conservation of average niche breadth) in 1 of the 100 observed trees for all 1,230 taxa (see Extended Data Fig. 6 for the other 99 relationships which are all qualitatively and statistically the same as this example). The more positive the LIPA Moran's *I* value the more similar niche breadth is between closely related taxa. Each point represents one taxon. The blue line is a LOESS fit. Taxa with non-significant LIPA Moran's *I* have an *I* of 0. **c**, Box plot comparing the adjusted R^2 values of linear and quadratic fits between average niche breadth and LIPA Moran's *I*. A stronger quadratic fit indicates stronger phylogenetic conservation of niche breadth at the extremes (the more generalized/specialized a taxon is, the greater the likelihood close relatives will be just as generalized/specialized). Each point represents a tree with each model type (linear or quadratic) consisting of 100 trees (200 trees in total). Each line connects linear and quadratic models of the same tree. Two-tailed Wilcoxon signed-rank test, $V = 0$, $P < 2.20 \times 10^{-16}$.

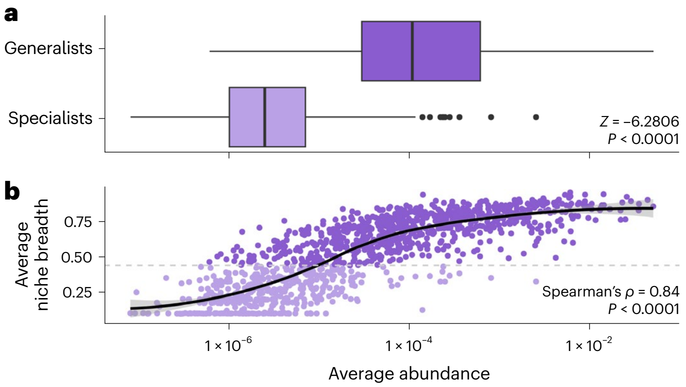


Fig. 3 | Multidimensional generalists are more dominant within communities. a, Mean abundances of generalist (dark purple) and specialist (light purple) taxa. Significance calculated with a two-sided permutational test ($Z = -6.2806$, $P = 3.37 \times 10^{-10}$). Box plots show the median (middle line) and interquartile range (box). **b**, Average relative abundances of 1,230 taxa regressed against average niche breadth. The line is fitted with LOESS smoothing, and the shaded region around the line is the 95% confidence interval. The dashed horizontal line indicates the local minima in the bimodal distribution of average niche breadth used to indicate specialists (light purple) and generalists (dark purple) in **b**. The direction of the relationship is determined using a Spearman's correlation test and significance is calculated using a two-tailed permutational test in which abundances are randomized 10,000 times ($P < 2.20 \times 10^{-16}$).

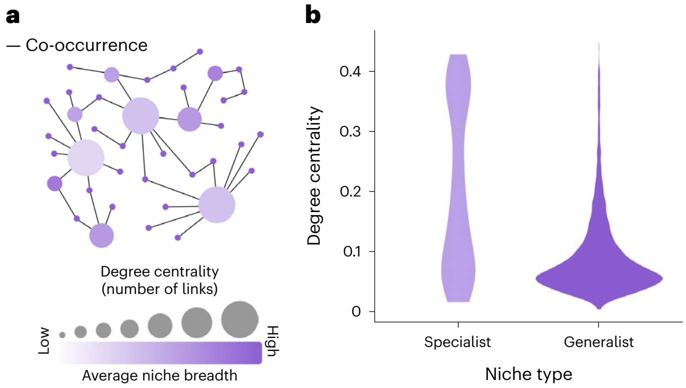


Fig. 4 | Multidimensional specialists are more central to microbiome networks. a, Schematic of analysis performed. We calculated the normalized degree centrality (number of links out of all possible links within a microbiome network) and regressed centrality against the average niche breadth of each taxon. **b**, Violin plot of normalized degree centrality of multidimensional specialists and multidimensional generalists. Taxa with smaller niche breadths (more specialized) are disproportionately more central than taxa with wider niche breadths (more generalized) even after accounting for site identity ($P < 2.20 \times 10^{-16}$, permutational linear model).

multidimensional niche breadth relationships shape evolution, we asked whether multidimensional niche breadth impacts prokaryotic ecology through community dominance and structure.

Multidimensional generalists dominate microbial communities

Second, when investigating ecological consequences of multidimensional generalization/specialization, we found that multidimensional generalists were 73 times more dominant on average than multidimensional specialists ($Z = -6.28$, $P < 0.0001$; Fig. 3). Specifically, we compared the mean and maximum relative abundances (indicators of dominance/performance within communities) to determine whether the size of a taxon's niche breadth explains how dominant that taxon is relative to other taxa. Prokaryotes with wider niche breadths were more abundant than those with narrower breadths for both mean and maximum relative abundance (Spearman's $\rho = 0.84$ and 0.66 ($P < 0.0001$ and $P < 0.0001$) for mean and maximum relative abundance, respectively; Fig. 3). The dominance of generalists highlights how flexible the vast majority of the soil prokaryotic community should be to changes in environmental conditions because they can persist across a wide range of environmental conditions in many different niche dimensions.

The dominance of multidimensional generalists was also highly robust to a wide range of biological factors, analysis decisions and spatial scales (Extended Data Figs. 7 and 8). For instance, this relationship was maintained when analysing communities at higher taxonomic resolutions (>14,000 ESV; Spearman's $\rho = 0.62$ and 0.33 ($P < 0.0001$ and $P < 0.0001$) for mean and maximum relative abundance, respectively; Extended Data Fig. 7a,b) and when including the three additional resource axes: per cent carbon, per cent nitrogen and carbon to nitrogen ratios (Spearman's $\rho = 0.85$ and 0.70 ($P < 0.0001$ and $P < 0.0001$) for mean and maximum relative abundance, respectively; Extended Data Fig. 7c,d). Because the dominance of multidimensional generalists could be overestimated if abundances of specialists are downweighted by absences outside their range, we also conducted analyses accounting for the size of each taxon's niche breadth and the number of habitats in which a taxon is present. The goal of these two analyses was to demonstrate that our conclusions were robust when accounting for the potential of an abundance–occupancy bias. Multidimensional generalists were still dominant even after accounting for a taxon's niche breadth (mean relative abundance/mean niche breadth) with multidimensional generalists 27 times more abundant than specialists on average ($Z = 6.3857$, $P < 0.0001$). We also found that when evaluating abundances only where taxa occur, multidimensional generalist taxa were still four times more dominant on average than specialist taxa ($Z = -6.1038$, $P < 0.0001$). Finally, the higher abundances of multidimensional generalists compared with specialists occurred not only at the continental scale, but also locally (permutational analysis of variance (ANOVA); $P < 0.0001$ and $P < 0.0001$ for mean and maximum relative abundance; Extended Data Fig. 8). In fact, niche breadth was a more important predictor of a taxon's relative abundance than sampling site, with niche breadth explaining >15 times more variation in both relative abundance metrics ($\omega^2_{\text{Niche Breadth}} = 0.52$, $\omega^2_{\text{Site}} = 0.03$). Overall, our analyses indicate that multidimensional generalists are more dominant than multidimensional specialists regardless of spatial scales and taxonomic resolution, emphasizing that many soil prokaryotes are likely to be resilient to environmental changes because multidimensional generalists persist across a wide range of conditions in many different niche dimensions, and that larger population sizes may be an important part of the multidimensional generalists' ecological strategy.

Multidimensional specialists and community function

Finally, multidimensional specialists are central to community networks and functions. Microbiome network analysis revealed that multidimensional specialists are more highly connected across the broader microbiome community than multidimensional generalists, indicating that multidimensional specialists are often 'hub taxa' and may have keystone roles within their communities. Specifically, when we constructed microbiome co-occurrence networks for each site and assessed each taxon's ('species') connectedness (degree centrality), specialized taxa had a ~220% greater degree centrality than generalist taxa ($P < 0.0001$, permutational ANOVA; Fig. 4). Interestingly, specialist taxa also appear to shape the overall structure of their communities because microbiomes with greater frequencies of multidimensional specialists have significantly higher network clustering (average clustering coefficients; $F_{1,19} = 13.19$, $P = 0.0018$), making them highly connected, tightly knit communities. 'Hub taxa' in microbiomes are often considered keystone

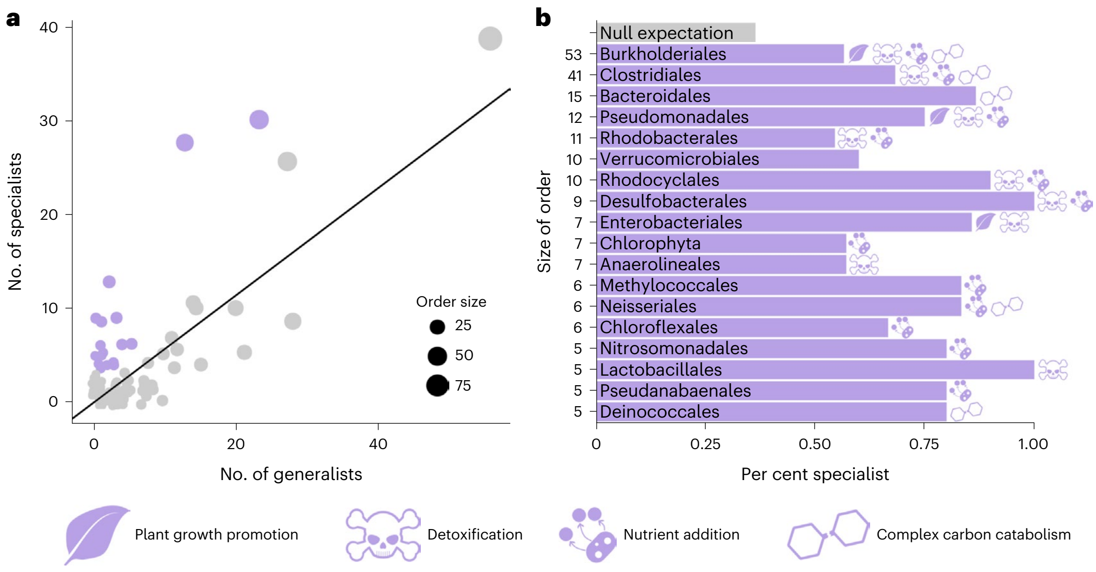


Fig. 5 | Multidimensional specialists are overrepresented in many nutrient-cycling and detoxifying orders of soil prokaryotes. a, Each dot represents an order and their sizes represent the number of taxa identified in our dataset. Purple dots are highlighted to identify orders of interest for further multidimensional specialist research (that is, the order contained at least five taxa and twofold higher ratios of multidimensional specialists to multidimensional generalists). The solid line represents the observed ratio of multidimensional specialists to multidimensional generalists across the entire dataset of prokaryotic communities, 0.571. A jitter is applied to points to better highlight the number of orders in our dataset. **b**, Bar chart of the purple highlighted orders in **a**. The numbers on the *y* axis are the sizes of the orders. The *x* axis is the proportion of the orders identified as multidimensional specialists. The grey 'null expectation' is the percentage of specialists in the whole dataset (36.34%). Icons to the right of columns represent functions of specialist taxa within these orders that could be verified by a literature search (literature search results are provided in Supplementary Data 1). 'Plant growth promotion' includes bacteria that increase plant growth or improve plant defence. 'Detoxification' includes heavy metal immobilization, xenobiotic degradation and hydrocarbon degradation. 'Nutrient addition' includes important biogeochemical cycling processes such as carbon fixation, denitrification, sulfate reduction and nitrogen fixation. 'Complex carbon catabolism' includes breakdown of complex, difficult to metabolize carbon sources such as lignin, chitin and cellulose. The numbers of multidimensional specialists and generalists for all orders are provided in Supplementary Data 1.

species[21,22] (organisms that play a disproportionate role in structuring communities[22–24]), which we have recently empirically demonstrated in the field[25]. For example, the removal of a 'hub' microbe in leaf endophyte and epiphyte communities destabilized communities with greater variability in community composition when the hub is absent than when it is present[21], and highly connected microbes (taxa with a high degree of centrality) shaped soil microbiome assembly in nature, repeatedly increasing biodiversity and deterministically structuring community composition during succession[25]. Therefore, although multidimensional specialists are often fewer and less abundant in communities than multidimensional generalists, their central placement within microbiome networks highlights that they probably play critical roles in these communities. Because central placement within microbiome networks and rare microbes have been documented to be important for determining microbial community variation at global scales[26] and for supplying unique, but critical, services within communities[24,27], the fact that multidimensional specialists have both these characteristics and were associated with changes in microbiome-wide network properties implicates them as structurally important taxa within communities.

Multidimensional specialists do indeed appear to play central functional roles within their communities. We found that many prokaryotic orders involved in nutrient-cycling and/or detoxification had unexpectedly high proportions of multidimensional specialists (Fig. 5 and Supplementary Data 1) further highlighting multidimensional specialists as structurally important to their community networks. For example, 90% of Rhodocyclales taxa identified in our dataset are multidimensional specialists and, alongside other orders with overrepresentation of multidimensional specialists (for example, Desulfobacterales and Burkholderiales), have been implicated in nitrogen-cycling[28], sulfur-cycling[29] and detoxification[29,30]. Clostridiales, another order with overrepresentation of multidimensional specialists, is associated with nutrient-cycling[31] and plant symbiosis[32]. Also, many small orders of nutrient cyclers contained high numbers of multidimensional specialists; for instance, all three taxa of the green sulfur bacteria Chlorobiales[33] identified in our dataset were multidimensional specialists. In essence, although multidimensional specialists may be rare, their persistence in communities and their central roles in microbial networks probably reflect their important functions in microbiomes, including crucial roles in providing nutrients and detoxifying environments.

Conclusions

In conclusion, our study highlights multidimensional generalization and specialization as an important ecological principle in prokaryotic communities by demonstrating that virtually all soil prokaryotes follow two opposing trajectories in multidimensional niche space with cascading consequences for evolutionary transitions, taxon dominance and microbial roles within communities. Given that these microbes undergird many ecosystem functions and services (for example, nutrient-cycling, carbon sequestration, supporting primary producers), widespread multidimensional specialization and generalization influences natural processes by shaping prokaryotic ecology and evolution. For example, ecosystem services reliant on multidimensional generalists may be more robust to environmental instability[3,34–36] compared with those reliant on multidimensional specialists because multidimensional specialists may be more dependent on maintaining a complex set of environmental conditions. The discovery of multidimensional generalization and specialization as a ubiquitous feature of soil prokaryote communities has also sparked many questions for future work. First, investigation into the underlying mechanisms and processes that constrain prokaryotes to those two trajectories would be profitable. In particular, we advocate for studies asking: 'Do multidimensional generalists have more genetic diversity in their populations allowing them to occupy greater multidimensional niche space?', 'Can specialization on one axis restrict the environmental conditions a taxon is exposed to thereby leading to specialization across additional axes through adaptation to those restricted conditions?' and/or 'To what extent do pleiotropic and epistatic interactions among genes underpin multidimensional changes in the niche?'. Although agglomerative strategies such as bulk physiological measurements (for example, carbon flux in soil cores) or meta-genomes/meta-transcriptomes of whole soils do not allow fine-enough resolution to address these types of questions, the advent of new microfluidic sequencing and culturing technologies[37–39] could provide the fine-tuned resolution necessary to study the genetics and physiology of not just populations, but the actual individual microbes that make up those populations. In addition, some abiotic dimensions, such as oxygen availability, can vary at microhabitat levels and change dramatically along a single granule of soil[40,41]. As a result, future work analysing these fine-scale niche dimensions would be interesting to determine whether the multidimensional relationships we found for 'macrohabitat' dimensions are also important at the microhabitat scale. Microhabitat studies could provide valuable insight into whether multidimensional specialists utilize microhabitats to avoid competitive exclusion by the community-dominating multidimensional generalists. Second, although multidimensional generalists were more common and more dominant in our study, multidimensional specialists did make up a meaningful part of the overall communities (32%) and were more often hub taxa, making their ecology especially interesting for future investigation. To understand how these multidimensional specialists may be structuring their communities, we propose that single-cell sequencing approaches could be used to profile functional expression of these taxa in situ and microfluidic culturing approaches can be used to isolate these putative keystones for phenotyping and experimental manipulations. Third, we have analysed multidimensional niche breadth relationships along abiotic dimensions (for example, pH, temperature), but, moving forward, it would be valuable to also analyse biotic dimensions (for example, host plant breadth) to determine whether multidimensional specialization and generalization is also common for the biotic niche of prokaryotes. Unlike abiotic niche dimensions, biotic niches can be actively shaped by adaptation in the partner organisms, which may require different ecological strategies outside multidimensional specialization and generalization. Recent work in fungi and oomycetes suggests that specialization and generalization may not be strongly correlated between abiotic and biotic niche dimensions[42–44], but this has not been tested in soil prokaryotes. Because multidimensional specialization creates more constraints on where organisms can persist, multidimensional specialists may be at greater risk from accelerating habitat loss and environmental change in the Anthropocene[10,45], highlighting how ecosystem management that accounts for multidimensional environments could benefit microbial specialists and their communities. Further, because of multidimensional specialists' central role in their communities, the loss or decline of these taxa could perturb the entire prokaryotic community, especially in cases in which they provide unique, but critical, functions. Thus, future studies testing predictions of multidimensional specialist/generalist resilience and consequent effects on ecosystem function and stability could be especially important for understanding microbial roles in ecosystem responses to global change.

Methods

All analyses and data preparation were conducted in R (v.4.0.2) using the packages described in Supplementary Data 1 unless indicated otherwise.

Microbial sequence and environmental data collection

For this study, we analysed prokaryotic soil communities and environmental/biogeochemical data from 236 plots in NEON, which is the National Science Foundation's flagship ecological repository of biological, climatic and environmental information across the continental United States. NEON provides long-term, standardized data needed to understand ecological principles of the natural world[46]. NEON's study sites are split into three hierarchical groupings (Extended Data Fig. 1): site (broadest), plot and subplot (narrowest). NEON is already among the world's largest repositories of soil microbiome data, collecting prokaryotic community and biogeochemical data from soil cores at each subplot which are further subdivided into 'organic' and 'mineral' layers (if present) and analysed separately. Sample collection and raw data processing are described in the 'NEON User Guide to Microbe Marker Gene Sequences' (DP1.10108.001; DP1.20280.001; DP1.20282.001)[47].

To obtain data on the prokaryotic community, we downloaded raw, demultiplexed prokaryotic amplicon sequencing data from the NEON database[48] using scripts in Supplementary Data 2. NEON samples were collected from field sites at peak greenness/productivity to standardize across habitats. Microbial genomes were extracted by NEON using homogenization and lysis bead beating, and DNA was extracted using the DNEasy PowerSoil kit following the standard operating procedures described[49,50]. To survey prokaryotic communities, the hypervariable V4 region of 16S ribosomal RNA from extracted microbial genomes was amplified using standard Earth Microbiome Project primers, 515F (GTGYCAGCMGCCGCGGTAA) and 806R (GTGYCAGCMGCCGCGG-TAA)[47,51,52]. Amplicons were sequenced on the Illumina MiSeq platform as described in the Argonne National Laboratory (2015) and Battelle Ecology (2018) standard operating procedures[49,50].

To compare prokaryotic communities and abundances in different environments, we also obtained environmental data known to greatly shape microbial communities: soil pH, soil temperature, litter depth, soil moisture, per cent soil nitrogen, per cent soil carbon and carbon to nitrogen ratio (Supplementary Data 1)[11,53]. All 236 plots (30 sites) had data on soil pH, soil temperature, litter depth and soil moisture; however, a subset of the data (84 plots, 10 sites) had additional soil chemical characteristics (per cent soil carbon, per cent soil nitrogen and carbon to nitrogen ratio; Extended Data Fig. 1). As a result, we analysed niche breadth twice: once across the 236-plot dataset and once across the subsetted 84-plot dataset that had additional soil chemical information. Analysing niche breadth with both the full 236-plot dataset and the subsetted 84-plot dataset allowed us to analyse first more prokaryotic communities (across fewer environmental axes) and second more environmental axes (with fewer prokaryotic communities). In the first dataset, our environmental axes explain 34% and 64% of community variation without and with spatial structure (including the site ID of the plot as a factor), respectively (distance-based redundancy analysis). In the second dataset, 52% and 67% of community variation is explained by the environmental axes without and with spatial structure, respectively (distance-based redundancy analysis). Taken together, these analyses, alongside the literature[11,54], demonstrate that the environmental axes we selected are important components of prokaryotic niches.

Microbial sequence processing

To convert raw prokaryotic sequencing data to relative abundances of prokaryotes (Supplementary Data 2), we processed microbial sequencing data through QIIME2 (v.2019.1) to remove sequencing adapters and chimeras, denoise single-end reads and classify operational taxonomic units (OTUs)[55]. In short, we denoised microbial sequencing data using Dada2, which categorized reads into ESVs[56]. We normalized the abundances of ESVs by dividing the observed number of denoised reads for a variant by the total number of denoised reads in a sample. We further grouped ESVs into 'species' using a naive Bayes classifier against the 97% taxonomy reference sequence database from GreenGenes (v.13.5)[57]. We constructed the classifier by using the fit-classifier-naive-bayes function[58] within QIIME2's feature-classifier plug-in on the 97% taxonomy reference database. We then used the above classifier on the ESVs by using the classify-sklearn function (feature-classifier plug-in)[55]. This allowed grouping ESVs into taxa from 'kingdom' to 'species' levels. We refer to 'species' ('level 7' in QIIME2's terminology) as OTUs from here onward. Sample rarefaction curves plateaued indicating that further sequencing would be unlikely to identify additional taxa (Extended Data Fig. 2). To compare communities across samples, we averaged the reads for each OTU across subplots within a plot and repeated this for any environmental data that was also collected at the subplot level to avoid overrepresentation of plots that were sampled more often at the subplot level.

Calculating niche breadth

We first filtered the datasets, removing OTUs that were not present in at least 10% of all plots (in ≤23 of 236 plots for the larger site dataset or in ≤8 of 84 plots for the smaller site dataset). This 10% cut-off filter was used to avoid spurious niche breadths resulting from poorly represented taxa with a lack of data that prevents accurate estimation of niche breadth. Likewise, we also calculated niche breadth for taxa within a site to account for the potential that differences in niche breadth seen at the continental scale are a result of large dispersal limits or other local geographic characteristics as opposed to environmental parameters. We calculated OTU niche breadth (NB_i) using proportional similarity[12]:

$$\mathrm{NB}_i = 1 - 0.5 \sum_j |p_{ij} - q_j|$$

Proportional similarity constrains niche breadth between values of the smallest q_j and 1 with higher values indicating wider niche breadths. Proportional similarity quantifies the habitat preference of taxa by determining whether a taxon uses habitats in proportion to their availability (generalization) or occurs disproportionately within one or a few habitats (specialization). To do this, proportional similarity compares the proportion of a taxon's population in each habitat with how common that habitat is. For instance, if the proportions of a taxon's population mirrors how common habitats are, then that taxon has no habitat preference and is generalized. Proportional similarity takes into account the proportion of taxon i (p_i) present in habitat j and the proportion of all habitats that are habitat j (q_j). Proportional similarity offers two advantages: (1) because NB_i can only reach a maximum of 1, it is possible to compare niche breadths across multiple environments (this is particularly important in our analyses when niche breadth is calculated at each site) because different environments will be on the same scale; and (2) proportional similarity accounts for how rare/common a habitat is in the community's environment so that, if the proportion of a taxon's population in a given habitat is almost equal to how common that habitat is (there is no preference), the difference between p_{ij} and q_j approaches 0 for every habitat (j) thus resulting in niche breadths closer to 1 (ref. 12). To identify 'habitats' along our environmental axes, we binned each axis into ten bins using the functions cut and cut2 in the base and Hmisc R packages[59,60] (our results were consistent up to our highest tested number of bins, 30; Extended Data Fig. 2b–d). We also repeated our analyses using two other common niche breadth metrics—niche range and Levin's niche breadth—finding consistent results (results and statistics are provided in Supplementary Data 1). We also confirm that multidimensional niche breadths are robust across different time points with (1) the same multidimensional niche breadth relationships present in the two most sampled time points in the NEON dataset (July and October 2017) and (2) the niche breadth of taxa consistent between these two time points (Spearman's $\rho = 0.76$).

Relationships between niche breadth across axes

To determine whether niche breadth is related across multiple axes (that is, whether multidimensional niche specialization and/or generalization are common or if niche breadths have no relationship among niche axes), we correlated niche breadths of each taxon on one environmental axis (for example, soil pH) with niche breadths on all other environmental axes using Spearman correlations. Importantly, we also examined whether multidimensional specialization and generalization were due to correlation in the environmental axes themselves (for example, if organisms' multidimensional generalization on soil temperature and soil moisture niche axes is because soil temperature and moisture are strongly correlated). To do this, we determined the relationship between environmental axes using Spearman correlations and compared the absolute values of those Spearman's coefficients with the Spearman's coefficients for the niche breadth relationships using a Mann–Whitney *U*-test. To test whether effects were the same at local scales, we repeated these analyses at each site using niche breadths calculated at the site level with the aovp function (lmPerm package). aovp uses permutation tests to calculate significance in an ANOVA model (without requiring a normal distribution). This test allowed us to determine whether relationships between niche breadths were stronger or weaker than relationships between environmental axes after accounting for variation associated with a sample's originating site. To evaluate the robustness of our results to taxonomic resolution, we repeated these analyses at the higher taxonomic resolution of ESVs.

Comparing taxon dominance of multidimensional specialists and generalists

To assess whether multidimensional specialists or generalists are more dominant in communities, we determined the average and maximum relative abundances of all taxa across the 236 communities and compared differences between multidimensional specialists (all niche breadths below the local minimum in average niche breadth distributions) and multidimensional generalists (all niche breadths above the local minimum) for each abundance metric with permutational tests (independence_test function, coin package). After the permutational test determined whether these specialists and generalists had different abundances, we ran Spearman's correlations tests of the two abundance metrics against niche breadth to determine: (1) whether relative abundance increases or decreases with niche breadth and (2) how much of the variation in abundance differences could be attributed to niche breadth. We also repeated this analysis at the higher taxonomic resolution of ESVs to ensure results were robust. We conducted several other analyses to further check that the results were robust. For instance, because effects on abundance may result from dispersal limits at a continental scale, we ran a similar test at the local 'site' level. Specifically, we tested whether niche breadth (calculated for each taxon at each site) explained the two abundance metrics after accounting for variation associated with which site a sample originates by using the aovp function (lmPerm package). In addition, because dominance of generalists could be overestimated if abundances of specialists are downweighted by absences outside their range, we also conducted two analyses accounting for the potential of an abundance–occupancy bias by accounting for sizes of taxons' niche breadths and the number of habitats in which a taxon is present.

Calculating transition rates between multidimensional generalist and specialist taxa

To determine whether multidimensional generalists and specialists are more or less likely to transition to the opposite state (for example, generalist transitions to specialist) than expected by chance, we constructed phylogenetic trees using the microbial sequences from the NEON sequencing data and calculated transitions with stochastic character mapping. Because OTUs consist of multiple ESVs and observations may change based on which variant is used to represent an OTU, we constructed 100 phylogenetic trees in which a randomly chosen variant within an OTU is used to represent that OTU. We performed multiple sequence alignments using the CLUSTALΩ[61] algorithm with default parameters. We then converted the alignments to distance matrices with the dist.ml function (phangorn package) and used the Jukes and Cantor 1969 (JC69) substitution model which assumes equal frequencies of nucleotides and equal mutation rates between nucleotides[62]. We built trees using the neighbour-joining method (NJ function in phangorn package) and rooted these trees at the midpoint (midpoint function in phangorn package)[63–65]. For each of the 100 constructed trees, we compared the observed transition rates from specialist to generalist (and vice versa) under an 'equal rates' model using the make.simmap function (phytools package) against the transition rates of trees in which 'specialist' and 'generalist' status were randomized 1,000 times without replacement. This allowed us to determine: (1) whether the observed transition rates in the 100 observed trees were different from expected by random chance and (2) how much the observed transition rates changed compared with random expectations. We used an 'equal rates' model, which assumes that transitions between specialist and generalist status (and vice versa) occur at equal rates[66]. We calculated one-tailed *P* values for the observed number of transitions from generalist to specialist status and vice versa for each of the 100 observed trees by calculating the number of transitions in 1,000 randomized trees that were less than the number of transitions in the observed tree and then dividing that sum by 1,000 (the number of permutations). We then corrected for multiple comparisons by calculating the FDR (Benjamini–Hochberg correction).

To measure whether more specialized/generalized taxa have more specialized/generalized relatives than intermediately specialized/generalized taxa, we calculated Local Indicators of Phylogenetic Association (LIPA) Moran's *I* of average niche breadth across all 100 previously constructed trees. LIPA Moran's *I* is the same formula for Local Indicators of Spatial Association, but, instead of being applied on spatial distances, it is applied to phylogenetic distances[67]. For each taxon, LIPA Moran's *I* can be calculated for its average niche breadth to quantify whether niche breadth in that area of the tree is a hotspot of phylogenetic clustering in niche breadth by applying a weighting constant of d_{ij}^{-1} where d_{ij} is the phylogenetic distance between focal taxon (i) and all other taxa in the phylogeny (j). We then determined whether Local Moran's *I* (niche breadth conservation) is constant throughout the entire phylogeny or highest at the two niche breadth extremes by regressing Local Moran's *I* against average niche breadth. We then compared, for each observed tree, the adjusted *R*-squared of two generalized linear models: (1) one in which Local Moran's *I* was the response variable and average niche breadth was the predictor variable, and (2) a quadratic model with the same variables and an additional predictor variable of average niche breadth squared.

Assessing relationships between niche breadth and network centrality

To determine whether multidimensional specialists or generalists have central roles in their ecological communities, we used network theory to assess if specialist or generalist taxa are more highly connected to other taxa within their microbiome community network and thus more likely to be hub taxa within microbiomes. We constructed co-occurrence networks using the FastSpar package[68] for each site with at least ten plots (21 NEON sites spread across the United States). Networks had 283 ± 13.7 species (mean ± s.e.m.). FastSpar is an optimized reimplementation of the SparCC algorithm which infers correlations between taxa while limiting the occurrence of spurious correlations inherent in analyses of compositional datasets (for example, relative abundances of taxa in communities)[69]. Links within co-occurrence networks are significant correlations between taxa abundances (as identified by FastSpar), which can represent interactions between connected taxa and/or shared habitat preferences[70–73]. As a result, analysing the number of links a taxon has with others in a network (degree centrality) has been used in the literature to provide information about whether a taxon is a keystone species and/or a habitat indicator species[21,30,74–76]. We calculated the degree centrality of each taxon in each network using the degree_centrality function in the networkx package (Python). The degree_centrality function normalizes the number of links a taxon has by the number of other taxa in a network $\frac{\text{Number of a taxon's links}}{\text{Number of nodes in a network}-1}$ which allows for comparing degree centralities of taxa across networks of varying sizes. To identify relationships between specialization and degree centrality, we regressed the degree centrality of taxa against their average niche breadth at that site and blocked by the site from which degree centrality and average niche breadth are collected. We perform this regression using the permutational strategy in the aovp function (lmPerm package). We also analysed global network structure of the networks and regressed average clustering coefficient (how tightly knit is the network) with the proportion of the network that are specialists using a generalized linear model.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All raw sequencing and environmental data are publicly available through the NEON database (DP1.10081.001). Scripts to download data from NEON and process sequencing data into ESVs and OTUs are available in Supplementary Data 2. OTU abundances from 'kingdom' to 'species' levels are available in Supplementary Data 1. We used the GreenGenes database (v.13.5) for taxonomic assignments.

Code availability

Code to replicate our analyses and a 'project' folder containing all the intermediate files and statistical summaries from RMarkdown scripts are available at Zenodo (https://doi.org/10.5281/zenodo.7747186).

References

1. Sexton, J. P., Montiel, J., Shay, J. E., Stephens, M. R. & Slatyer, R. A. Evolution of ecological niche breadth. *Annu. Rev. Ecol. Evol. Syst.* **48**, 183–206 (2017).
2. Carscadden, K. A. et al. Niche breadth: causes and consequences for ecology, evolution, and conservation. *Q. Rev. Biol.* **95**, 179–214 (2020).
3. Muller, E. E. L. Determining microbial niche breadth in the environment for better ecosystem fate predictions. *mSystems* **4**, e00080-19 (2019).
4. Bonetti, M. F. & Wiens, J. J. Evolution of climatic niche specialization: a phylogenetic analysis in amphibians. *Proc. Biol. Sci.* **281**, 20133229 (2014).
5. Julliard, R., Clavel, J., Devictor, V., Jiguet, F. & Couvet, D. Spatial segregation of specialists and generalists in bird communities. *Ecol. Lett.* **9**, 1237–1244 (2006).
6. Devictor, V. et al. Defining and measuring ecological specialization. *J. Appl. Ecol.* **47**, 15–25 (2010).
7. Dehling, D. M., Jordano, P., Schaefer, H. M., Böhning-Gaese, K. & Schleuning, M. Morphology predicts species' functional roles and their degree of specialization in plant–frugivore interactions. *Proc. Biol. Sci.* **283**, 20152444 (2016).
8. Hardy, N. B. & Otto, S. P. Specialization and generalization in the diversification of phytophagous insects: tests of the musical chairs and oscillation hypotheses. *Proc. Biol. Sci.* **281**, 20132960 (2014).
9. Grime, J. P. Competitive exclusion in herbaceous vegetation. *Nature* **242**, 344–347 (1973).
10. Warren, M. S. et al. Rapid responses of British butterflies to opposing forces of climate and habitat change. *Nature* **414**, 65–69 (2001).
11. Fierer, N. Embracing the unknown: disentangling the complexities of the soil microbiome. *Nat. Rev. Microbiol.* **15**, 579–590 (2017).

12. Feinsinger, P., Spears, E. E. & Poole, R. W. A simple measure of niche breadth. *Ecology* **62**, 27–32 (1981).
13. Kim, S. L., Tinker, M. T., Estes, J. A. & Koch, P. L. Ontogenetic and among-individual variation in foraging strategies of northeast Pacific white sharks based on stable isotope analysis. *PLoS ONE* **7**, e45068 (2012).
14. Fierer, N., Bradford, M. A. & Jackson, R. B. Toward an ecological classification of soil bacteria. *Ecology* **88**, 1354–1364 (2007).
15. Setlow, P., Wang, S. & Li, Y.-Q. Germination of spores of the orders Bacillales and Clostridiales. *Annu. Rev. Microbiol.* **71**, 459–477 (2017).
16. Felsenstein, J. Parsimony in systematics: biological and statistical issues. *Annu. Rev. Ecol. Syst.* **14**, 313–333 (1983).
17. Ochman, H. & Moran, N. A. Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. *Science* **292**, 1096–1099 (2001).
18. Deeds, E. J., Hennessey, H. & Shakhnovich, E. I. Prokaryotic phylogenies inferred from protein structural domains. *Genome Res.* **15**, 393–402 (2005).
19. Sriswasdi, S., Yang, C. & Iwasaki, W. Generalist species drive microbial dispersion and evolution. *Nat. Commun.* **8**, 1162 (2017).
20. Barberán, A. et al. Why are some microbes more ubiquitous than others? Predicting the habitat breadth of soil bacteria. *Ecol. Lett.* **17**, 794–802 (2014).
21. Agler, M. T. et al. Microbial hub taxa link host and abiotic factors to plant microbiome variation. *PLoS Biol.* **14**, e1002352 (2016).
22. Banerjee, S., Schlaeppi, K. & van der Heijden, M. G. A. Keystone taxa as drivers of microbiome structure and functioning. *Nat. Rev. Microbiol.* **16**, 567–576 (2018).
23. Paine, R. T. Food web complexity and species diversity. *Am. Nat.* **100**, 65–75 (1966).
24. Xun, W. et al. Specialized metabolic functions of keystone taxa sustain soil microbiome stability. *Microbiome* **9**, 35 (2021).
25. Rawstern, A. H., Hernandez, D. J. & Afkhami, M. E. Hub taxa are keystone microbes during early succession. Preprint at *bioRxiv* https://doi.org/10.1101/2023.03.02.530218 (2023).
26. Ramirez, K. S. et al. Detecting macroecological patterns in bacterial communities across independent studies of global soils. *Nat. Microbiol.* **3**, 189–196 (2018).
27. Bittleston, L. S., Gralka, M., Leventhal, G. E., Mizrahi, I. & Cordero, O. X. Context-dependent dynamics lead to the assembly of functionally distinct microbial communities. *Nat. Commun.* **11**, 1440 (2020).
28. Tang, S. et al. Microbial coupling mechanisms of nitrogen removal in constructed wetlands: a review. *Bioresour. Technol.* **314**, 123759 (2020).
29. Deng, J., Xiao, T., Fan, W., Ning, Z. & Xiao, E. Relevance of the microbial community to Sb and As biogeochemical cycling in natural wetlands. *Sci. Total Environ.* **818**, 151826 (2022).
30. Banerjee, S. et al. Poor nutrient availability in opencast coalmine influences microbial community composition and diversity in exposed and underground soil profiles. *Appl. Soil Ecol.* **152**, 103544 (2020).
31. Jia, R. et al. Abundance and community succession of nitrogen-fixing bacteria in ferrihydrite enriched cultures of paddy soils is closely related to Fe(III)-reduction. *Sci. Total Environ.* **720**, 137633 (2020).
32. Xiao, X. et al. Two cultivated legume plants reveal the enrichment process of the microbiome in the rhizocompartments. *Mol. Ecol.* **26**, 1641–1651 (2017).
33. Ozaki, K., Thompson, K. J., Simister, R. L., Crowe, S. A. & Reinhard, C. T. Anoxygenic photosynthesis and the delayed oxygenation of Earth's atmosphere. *Nat. Commun.* **10**, 3026 (2019).
34. Norden, N., Chazdon, R. L., Chao, A., Jiang, Y.-H. & Vílchez-Alvarado, B. Resilience of tropical rain forests: tree community reassembly in secondary forests. *Ecol. Lett.* **12**, 385–394 (2009).
35. García, Y., Clara Castellanos, M. & Pausas, J. G. Differential pollinator response underlies plant reproductive resilience after fires. *Ann. Bot.* **122**, 961–971 (2018).
36. Memmott, J., Waser, N. M. & Price, M. V. Tolerance of pollination networks to species extinctions. *Proc. Biol. Sci.* **271**, 2605–2611 (2004).
37. Imdahl, F., Vafadarnejad, E., Homberger, C., Saliba, A.-E. & Vogel, J. Single-cell RNA-sequencing reports growth-condition-specific global transcriptomes of individual bacteria. *Nat. Microbiol.* **5**, 1202–1206 (2020).
38. Ma, P. et al. Bacterial droplet-based single-cell RNA-seq reveals antibiotic-associated heterogeneous cellular states. *Cell* **186**, 877–891.e14 (2023).
39. Yin, J. et al. A droplet-based microfluidic approach to isolating functional bacteria from gut microbiota. *Front. Cell. Infect. Microbiol.* **12**, 920986 (2022).
40. Young, I. M. & Crawford, J. W. Interactions and self-organization in the soil–microbe complex. *Science* **304**, 1634–1637 (2004).
41. Schlüter, S., Sammartino, S. & Koestel, J. Exploring the relationship between soil structure and soil functions via pore-scale imaging. *Geoderma* **370**, 114370 (2020).
42. Bebber, D. P. & Chaloner, T. M. Specialists, generalists and the shape of the ecological niche in fungi. *New Phytol.* **234**, 345–349 (2022).
43. Chaloner, T. M., Gurr, S. J. & Bebber, D. P. Geometry and evolution of the ecological niche in plant-associated microbes. *Nat. Commun.* **11**, 2955 (2020).
44. Davison, J. et al. Temperature and pH define the realised niche space of arbuscular mycorrhizal fungi. *New Phytol.* **231**, 763–776 (2021).
45. Slatyer, R. A., Hirst, M. & Sexton, J. P. Niche breadth predicts geographical range size: a general ecological pattern. *Ecol. Lett.* **16**, 1104–1114 (2013).
46. Keller, M., Schimel, D. S., Hargrove, W. W. & Hoffman, F. M. A continental strategy for the National Ecological Observatory Network. *Front. Ecol. Environ.* **6**, 282–284 (2008).
47. Stanish, L. & Parker, S. *NEON User Guide to Microbe Marker Gene Sequences (DP1.10108.001; DP1.20280.001; DP1.20282.001)* (2020); https://data.neonscience.org/documents/10179/2237401/NEON_markerGenes_userGuide_vD/7be3774b-b924-9640-4d89-b154d31993df?version=1.0&previewFileIndex=
48. National Ecological Observatory Network (NEON). Soil microbe marker gene sequences (DP1.10108.001) (2022); https://data.neonscience.org/data-products/DP1.10108.001
49. *MoBio PowerSoil-htp 96-Well Manual Extraction Method Using a Swing Bucket Centrifuge* (Argonne National Laboratory, 2015); https://data.neonscience.org/documents/10179/2655517/ANL_soilDnaExtractionSOP_2015/ce0f07df-ca4d-428e-96f8-8c5a7413cd17
50. *NEON DNA Extraction Standard Operating Procedure v.1* (Battelle Ecology, Inc., 2018); https://data.neonscience.org/documents/10179/2655517/BMI_dnaExtractionSOP_v1/34aa8f9c-819c-48ed-821a-a82168a9dd20
51. Parada, A. E., Needham, D. M. & Fuhrman, J. A. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ. Microbiol.* **18**, 1403–1414 (2016).
52. Apprill, A., McNally, S., Parsons, R. & Weber, L. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat. Microb. Ecol.* **75**, 129–137 (2015).

53. National Ecological Observatory Network (NEON). Soil physical and chemical properties, periodic (DP1.10086.001) (2022); https://data.neonscience.org/data-products/DP1.10086.001
54. Dastogeer, K. M. G., Tumpa, F. H., Sultana, A., Akter, M. A. & Chakraborty, A. Plant microbiome–an account of the factors that shape community composition and diversity. *Curr. Plant Biol.* **23**, 100161 (2020).
55. Bolyen, E. et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 852–857 (2019).
56. Callahan, B. J. et al. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
57. McDonald, D. et al. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* **6**, 610–618 (2012).
58. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
59. Labrousse, S. et al. Under the sea ice: exploring the relationship between sea ice and the foraging behaviour of southern elephant seals in East Antarctica. *Prog. Oceanogr.* **156**, 17–40 (2017).
60. Afkhami, M. E., McIntyre, P. J. & Strauss, S. Y. Mutualist-mediated effects on species' range limits across large geographic scales. *Ecol. Lett.* **17**, 1265–1273 (2014).
61. Sievers, F. & Higgins, D. G. Clustal omega. *Curr. Protoc. Bioinformatics* **48**, 3–13 (2014).
62. Jukes, T. H. & Cantor, C. R. in *Mammalian Protein Metabolism* (ed. Munro, H. N.) 21–132 (Academic Press, 1969).
63. Goffredi, S. K. et al. Genomic versatility and functional variation between two dominant heterotrophic symbionts of deep-sea *Osedax* worms. *ISME J.* **8**, 908–924 (2014).
64. Ogata, H. et al. Genome sequence of *Rickettsia bellii* illuminates the role of amoebae in gene exchanges between intracellular pathogens. *PLoS Genet.* **2**, e76 (2006).
65. Argimón, S. et al. A global resource for genomic predictions of antimicrobial resistance and surveillance of *Salmonella* Typhi at pathogenwatch. *Nat. Commun.* **12**, 2879 (2021).
66. Blum, M. G. B. & François, O. Which random processes describe the tree of life? A large-scale study of phylogenetic tree imbalance. *Syst. Biol.* **55**, 685–691 (2006).
67. Keck, F., Rimet, F., Bouchez, A. & Franc, A. phylosignal: an R package to measure, test, and explore the phylogenetic signal. *Ecol. Evol.* **6**, 2774–2780 (2016).
68. Watts, S. C., Ritchie, S. C., Inouye, M. & Holt, K. E. FastSpar: rapid and scalable correlation estimation for compositional data. *Bioinformatics* **35**, 1064–1066 (2019).
69. Friedman, J. & Alm, E. J. Inferring correlation networks from genomic survey data. *PLoS Comput. Biol.* **8**, e1002687 (2012).
70. Layeghifard, M., Hwang, D. M. & Guttman, D. S. Disentangling interactions in the microbiome: a network perspective. *Trends Microbiol.* **25**, 217–228 (2017).
71. Barberán, A., Bates, S. T., Casamayor, E. O. & Fierer, N. Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J.* **6**, 343–351 (2012).
72. Hernandez, D. J., David, A. S., Menges, E. S., Searcy, C. A. & Afkhami, M. E. Environmental stress destabilizes microbial networks. *ISME J.* **15**, 1722–1734 (2021).
73. Röttjers, L. & Faust, K. From hairballs to hypotheses–biological insights from microbial networks. *FEMS Microbiol. Rev.* **42**, 761–780 (2018).
74. van der Heijden, M. G. A. & Hartmann, M. Networking in the plant microbiome. *PLoS Biol.* **14**, e1002378 (2016).
75. Gough, E. K. et al. Linear growth faltering in infants is associated with *Acidaminococcus* sp. and community-level changes in the gut microbiota. *Microbiome* **3**, 24 (2015).
76. Jordán, F. Keystone species and food webs. *Philos. Trans. R. Soc. B* **364**, 1733–1741 (2009).

Acknowledgements

We thank the National Ecological Observatory Network for making their data publicly available. We also thank A. Zanne (University of Miami), S. Strauss (University of California, Davis), K. Crawford (University of Houston), M. Jayachandran (Florida International University) and A. Rawstern, A. Igwe and G. Ortiz of the Afkhami lab (University of Miami) as well as the labs of A. Wilson, C. Silveira and L. Müller (University of Miami) for their feedback on this manuscript. We also thank the editors A. McKay and L. Grinham for their feedback on this manuscript. We acknowledge funding support from the University of Miami to D.J.H. (Maytag Fellowship, Dean's Summer Research Fellowship, Dean's Dissertation Fellowship) and K.N.K. (Lisa D. Anness Fellowship) as well as funding from the United States Department of Agriculture to D.J.H. (National Institute of Food and Agriculture Predoctoral Fellowship 2022-67011-36456) and the National Science Foundation (NSF) to K.N.K. (Graduate Research Fellowship), B.K.A. (Graduate Research Fellowship) and M.E.A. (DEB-1922521 and NSF DEB-2030060).

Author contributions

D.J.H., K.N.K. and M.E.A. conceptualized the study. D.J.H. and K.N.K. analysed the data. D.J.H., K.N.K., B.K.A. and D.R. collected the data. D.J.H., K.N.K. and M.E.A. wrote the manuscript. D.J.H., K.N.K., B.K.A., D.R. and M.E.A. reviewed and edited the manuscript. D.J.H. and M.E.A. revised the manuscript. M.E.A. supervised the study.

Competing interests

The authors declare no competing interests.

Additional information

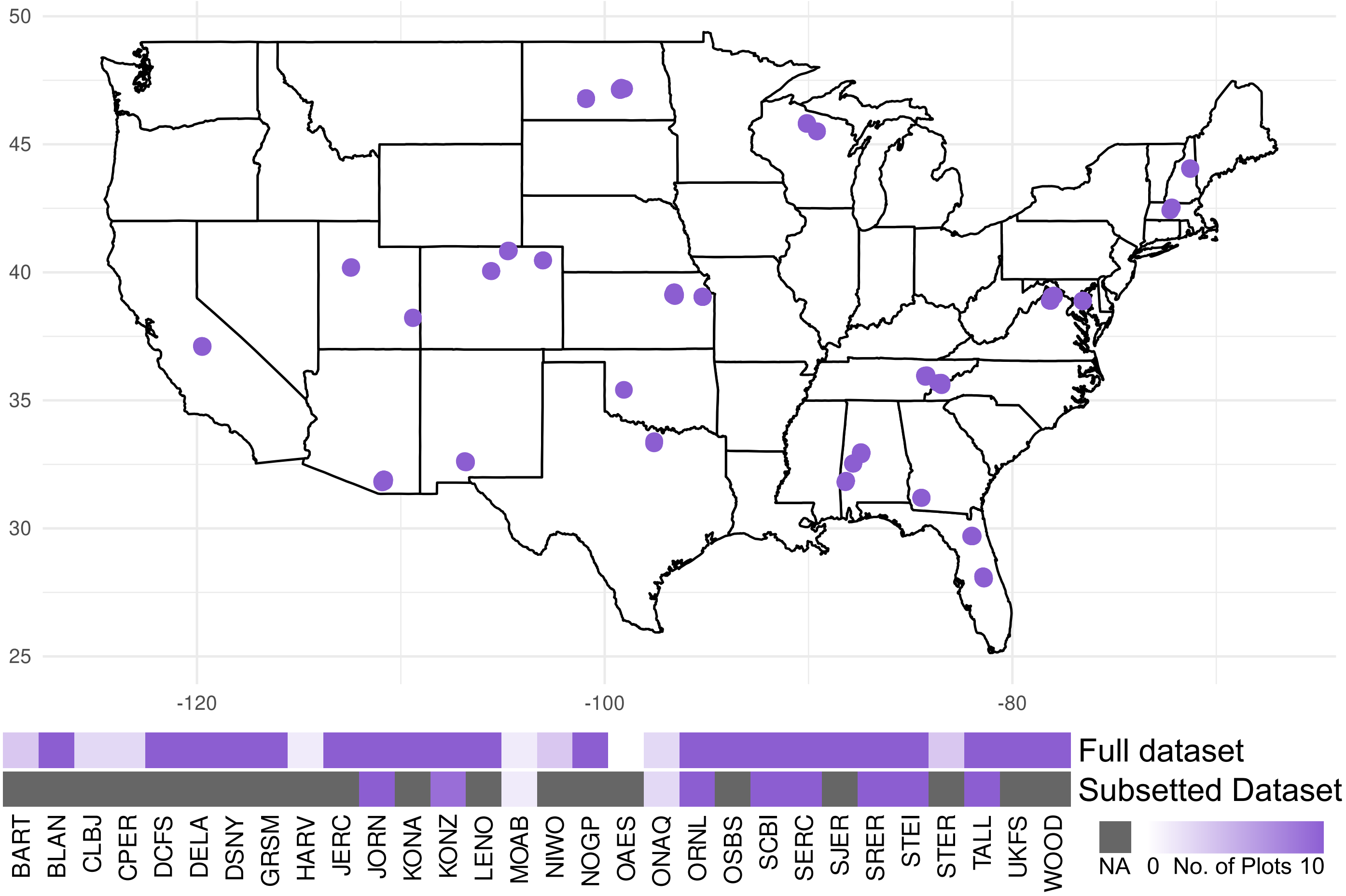
Extended data is available for this paper at https://doi.org/10.1038/s41559-023-02149-y.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41559-023-02149-y.

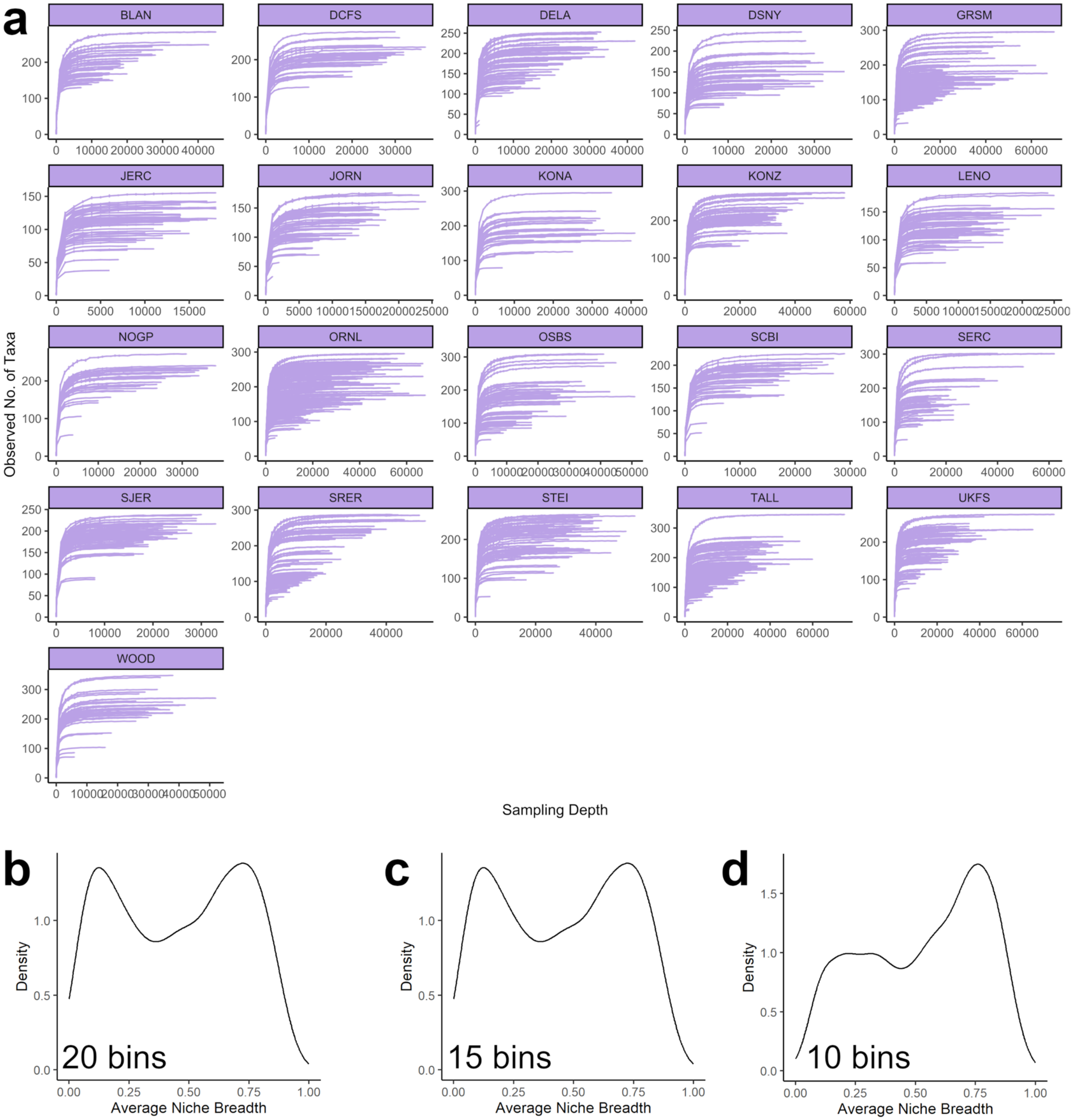
Correspondence and requests for materials should be addressed to Damian J. Hernandez.

Peer review information *Nature Ecology & Evolution* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

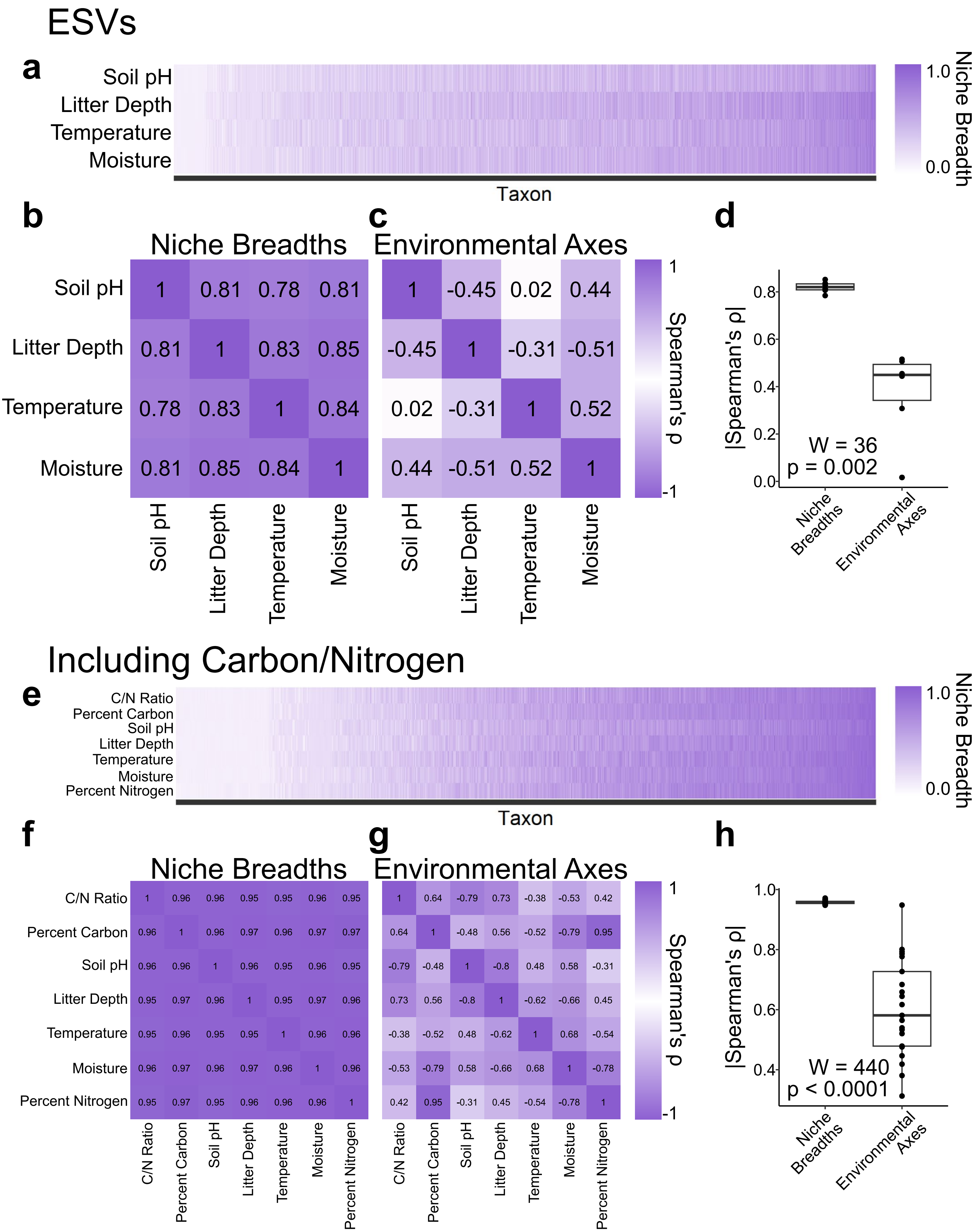
Reprints and permissions information is available at www.nature.com/reprints.



Extended Data Fig. 1 | Sampling design of NEON soil collections. Map of 30 NEON collection sites across the continental United States. There are 236 plots (up to 10 plots per site) in which complete data on soil pH, soil temperature, litter depth, and soil moisture are collected ('full dataset'). A subset of the 236 plots (84 plots across 10 sites) had additional biogeochemical data on per cent carbon, per cent nitrogen, and carbon/nitrogen ratios ('subsetted dataset'). Distribution of the number of plots at each site is displayed in the heatmap for both the full and subsetted datasets. Dark grey squares indicates data was not available (NA).

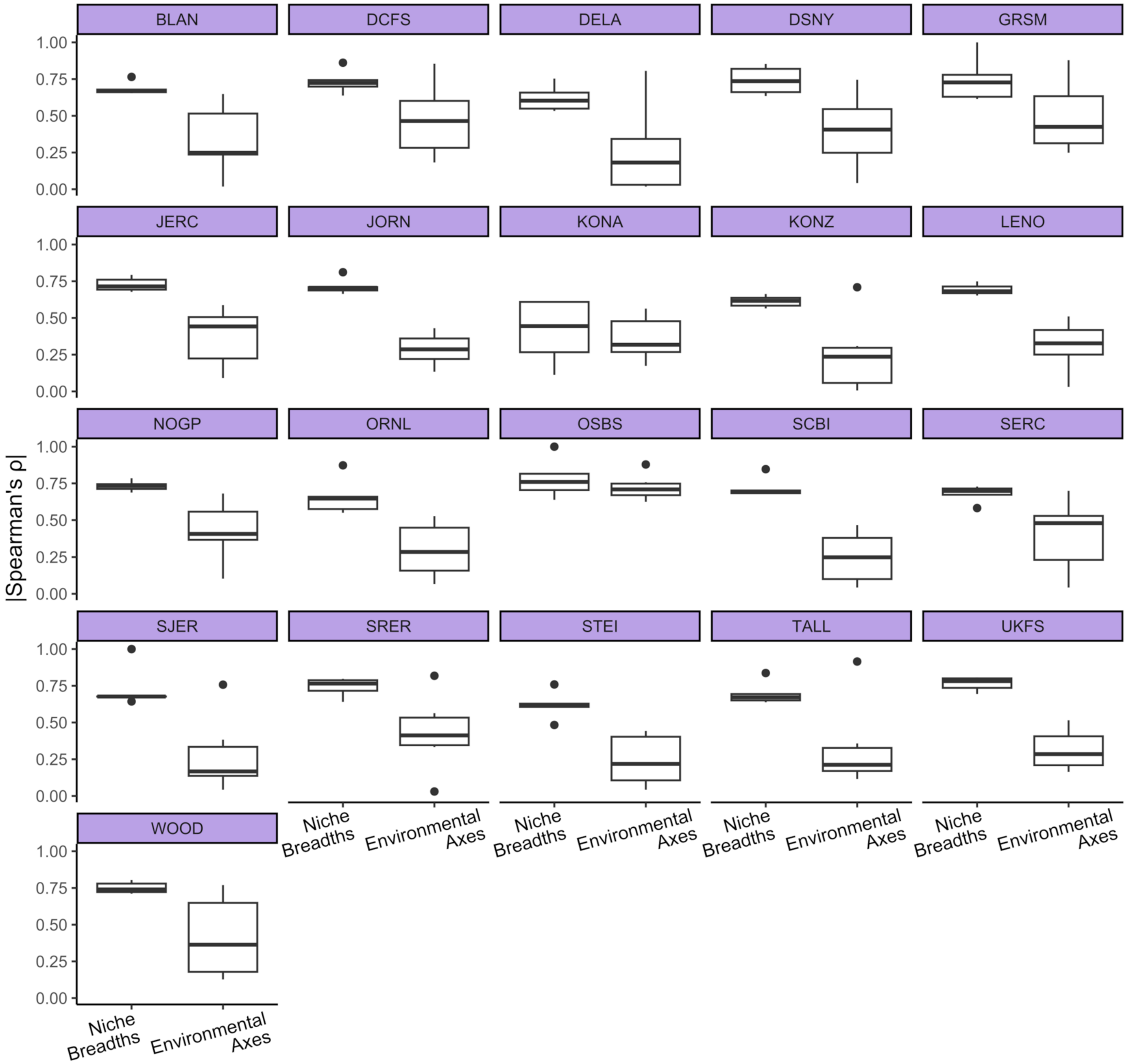


Extended Data Fig. 2 | Rarefaction curves of each sample split by site and robustness of bimodal niche breadth distributions to analysis decisions. a) All samples reached plateaus in their rarefaction curves indicating that we had enough sequencing depth to fully characterize communities. Each line represents the rarefaction curve of a sample. **b-d)** To calculate niche breadth along continuous axes, these axes must be broken into bins. Here, we show that bimodal distributions of niche breadth are robust to the important analysis decision of defining 'habitats' (that is, bins). In this manuscript, we present the results from the most conservative binning of 10 bins.

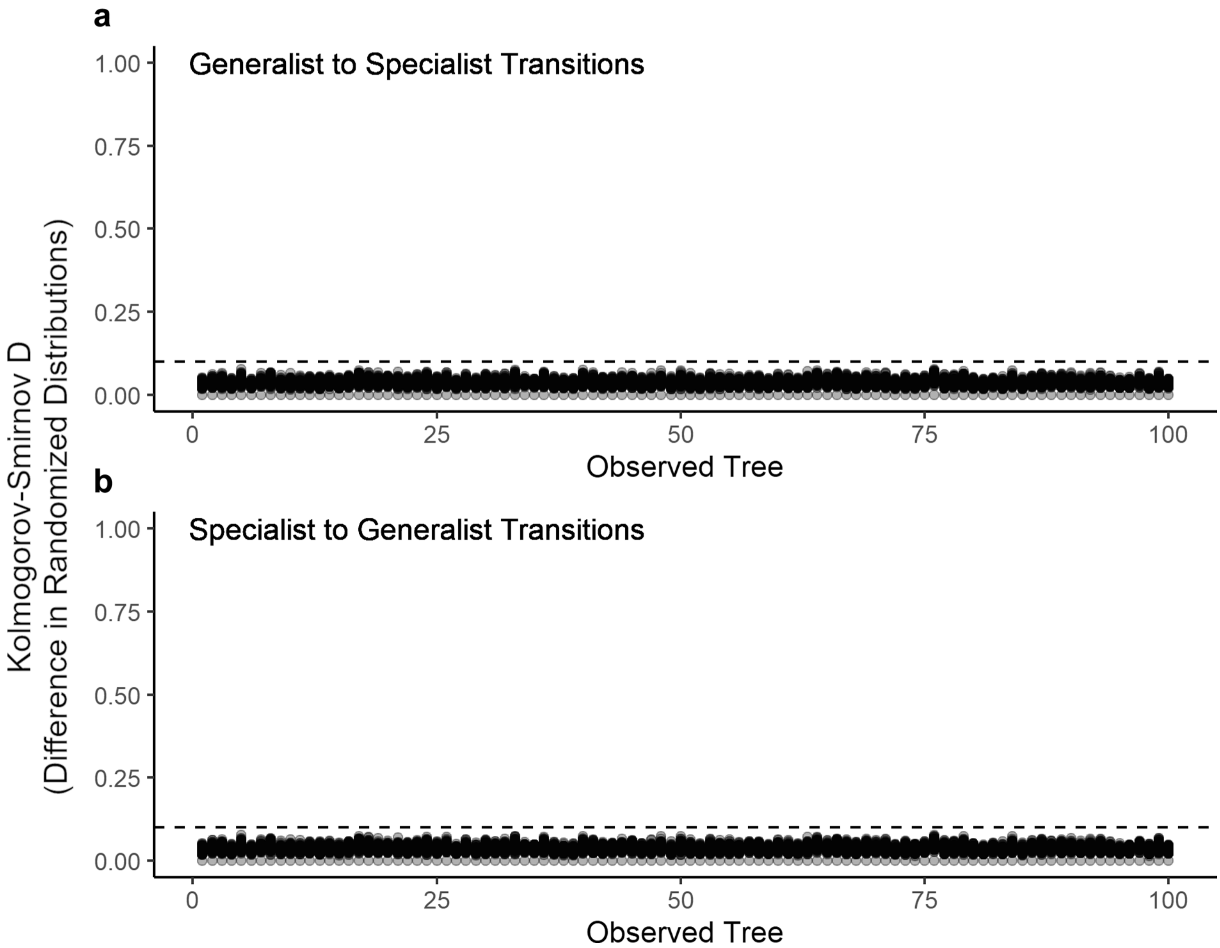


Extended Data Fig. 3 | See next page for caption.

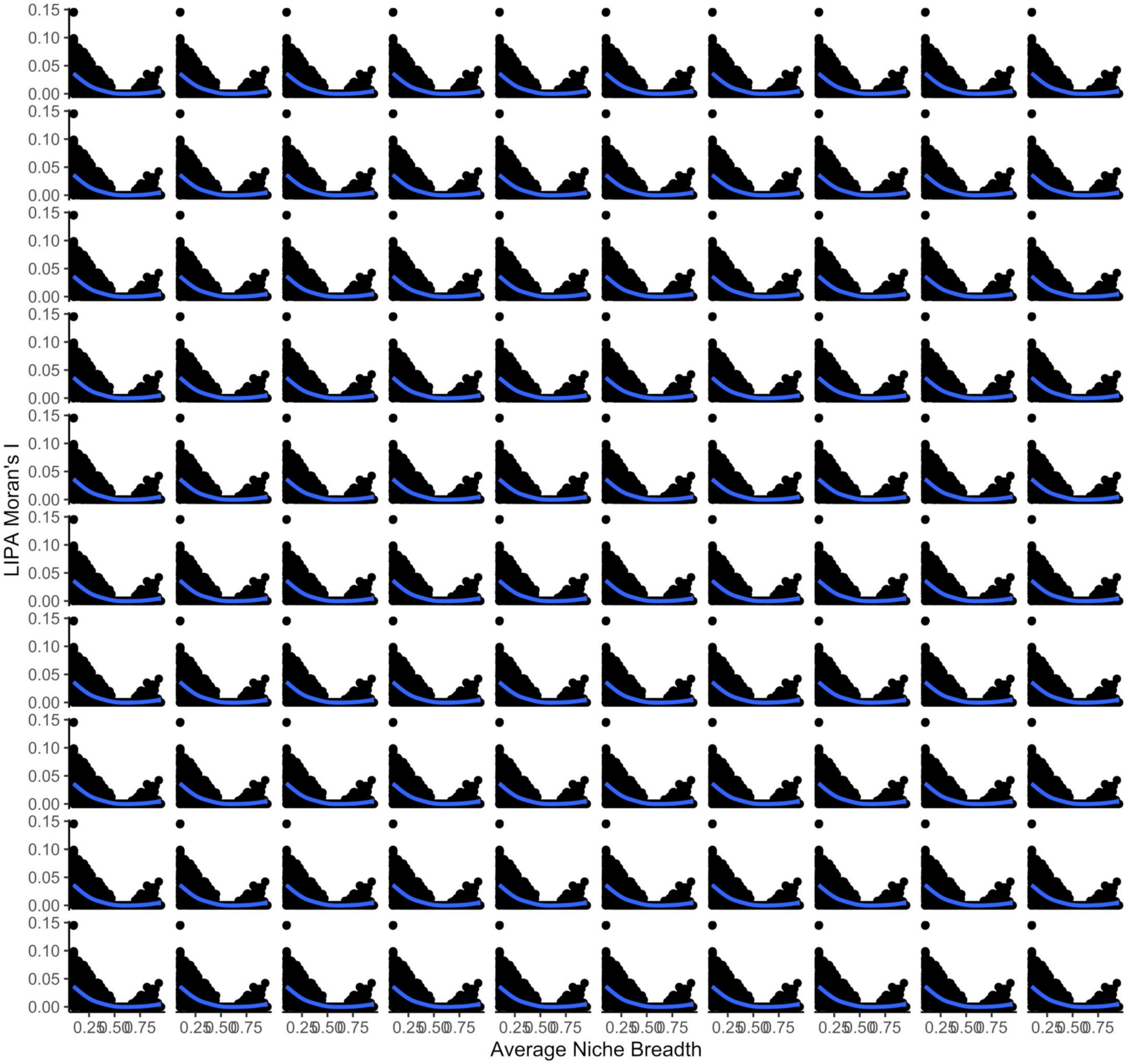
Extended Data Fig. 3 | Niche breadth and environmental correlations across niche dimensions for Exact Sequence Variants (A-D) and including carbon/nitrogen niche axes (E-H). ESV analyses are represented by subfigures A-D and taxa level analyses including carbon/nitrogen data are represented by subfigures E-H. **a)** Heatmap of 14,015 prokaryotic ESV taxa (x-axis) along environmental axes. ESVs are sorted from lowest to highest average niche breadth for visualization. **b)** Heatmap of Spearman's ρ from correlations between niche breadths of 14,015 ESVs along different axes. **c)** Heatmap of Spearman's ρ from correlations between environmental axes calculated across 236 individual plots. **d)** Comparison of the absolute values of Spearman's ρ from correlations between niche breadths and correlations between environmental axes, demonstrating that niche breadth correlations are significantly stronger than correlations in environmental variation among axes. Significance determined by two-tailed Mann-Whitney U test (W = 36, p = 0.002). Box plots show the median (middle line) and interquartile range (the box). **e)** Heatmap of 1085 prokaryotic taxa (x-axis) along seven environmental axes that include measures of carbon and nitrogen. Taxa are sorted from lowest to highest average niche breadth for visualization. **f)** Heatmap of Spearman's ρ from correlations between niche breadths of the 1085 taxa along the seven different axes. **g)** Heatmap of Spearman's ρ from correlations between the seven environmental axes across 84 individual plots. **h)** Comparison of the absolute values of Spearman's ρ from correlations between niche breadths and correlations between environmental axes, again demonstrating that niche breadth correlations are significantly stronger than correlations in environmental variation among axes. Significance determined by two-tailed Mann-Whitney U test (W = 440, $p = 7.43 \times 10^{-12}$). Box plots show the median (middle line) and interquartile range (the box).



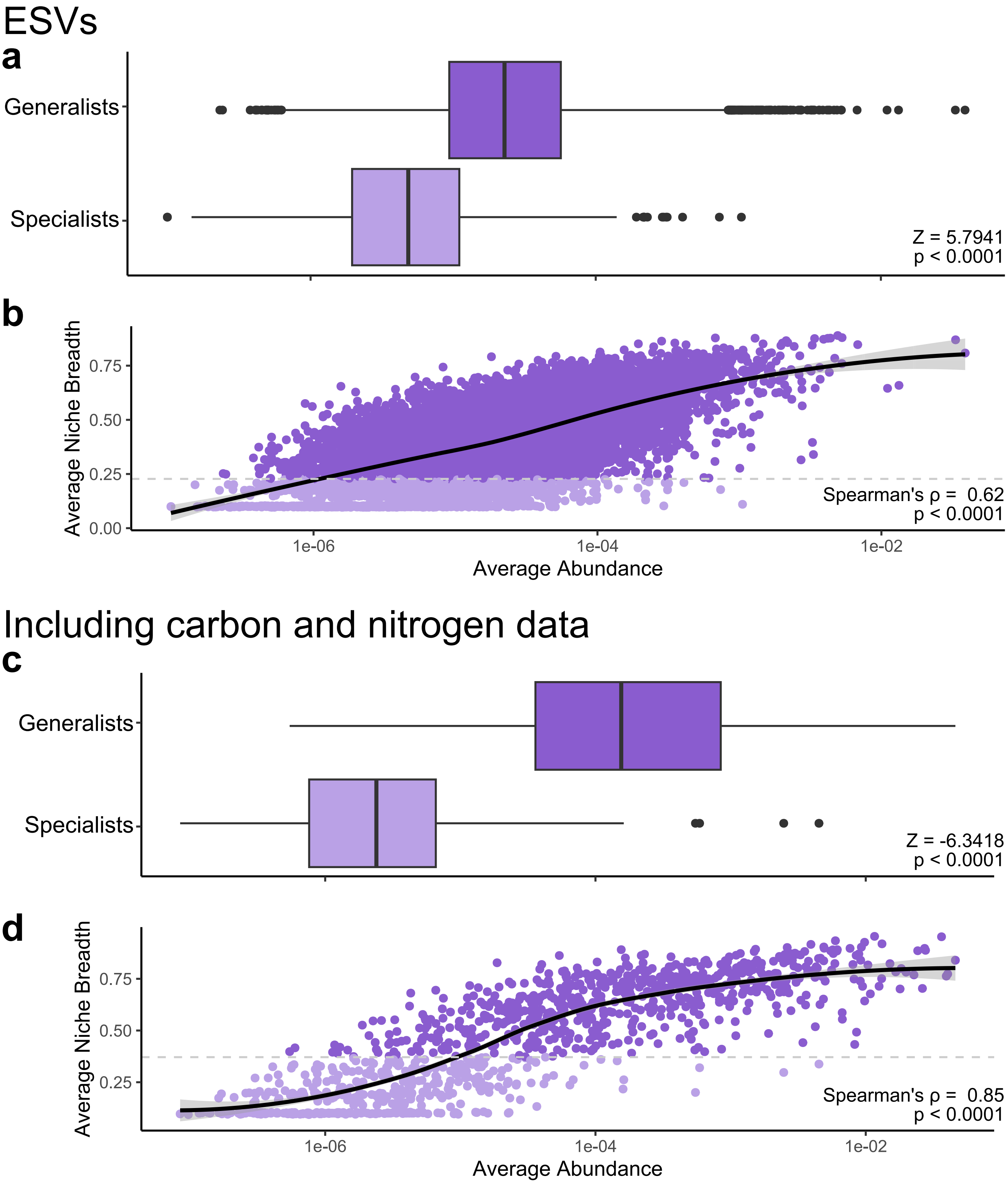
Extended Data Fig. 4 | Differences in magnitude of correlations between niche breadths and environmental axes. Boxplots of the magnitude of Spearman's coefficients between niche breadths and environmental axes at each of the 21 sites. When we account for the site from which data was collected, niche breadth relationships are still substantially stronger than environmental correlations ($p < 2.20 \times 10^{-16}$, permutational ANOVA accounting for origin site) with the type of the relationship (that is, relationship between niche breadths versus relationship between environmental axes) having an effect size >4 times stronger than site identity ($\omega_{\text{correlation type}}/\omega_{\text{site}} = 4.17$). Boxplots show the median (middle line) and interquartile range (the box).



Extended Data Fig. 5 | Randomized state transitions are consistent across all 100 observed representative trees. Kolomogorov-Smirnov statistics (a measure of how different the shape of two distributions are) of generalist-to-specialist (a) and specialist-to-generalist (b) transitions in all 100 observed representative trees (x-axis). Each point is the comparison of the randomized distribution of the focal representative tree (value of x-axis) against each other representative tree. All values are below a D of 0.1 (dashed horizontal line) indicating that our analyses are robust to changes in which ESV represents each OTU. The higher the D, the more different the distributions are from each other. The lower the D, the more similar the distributions are.

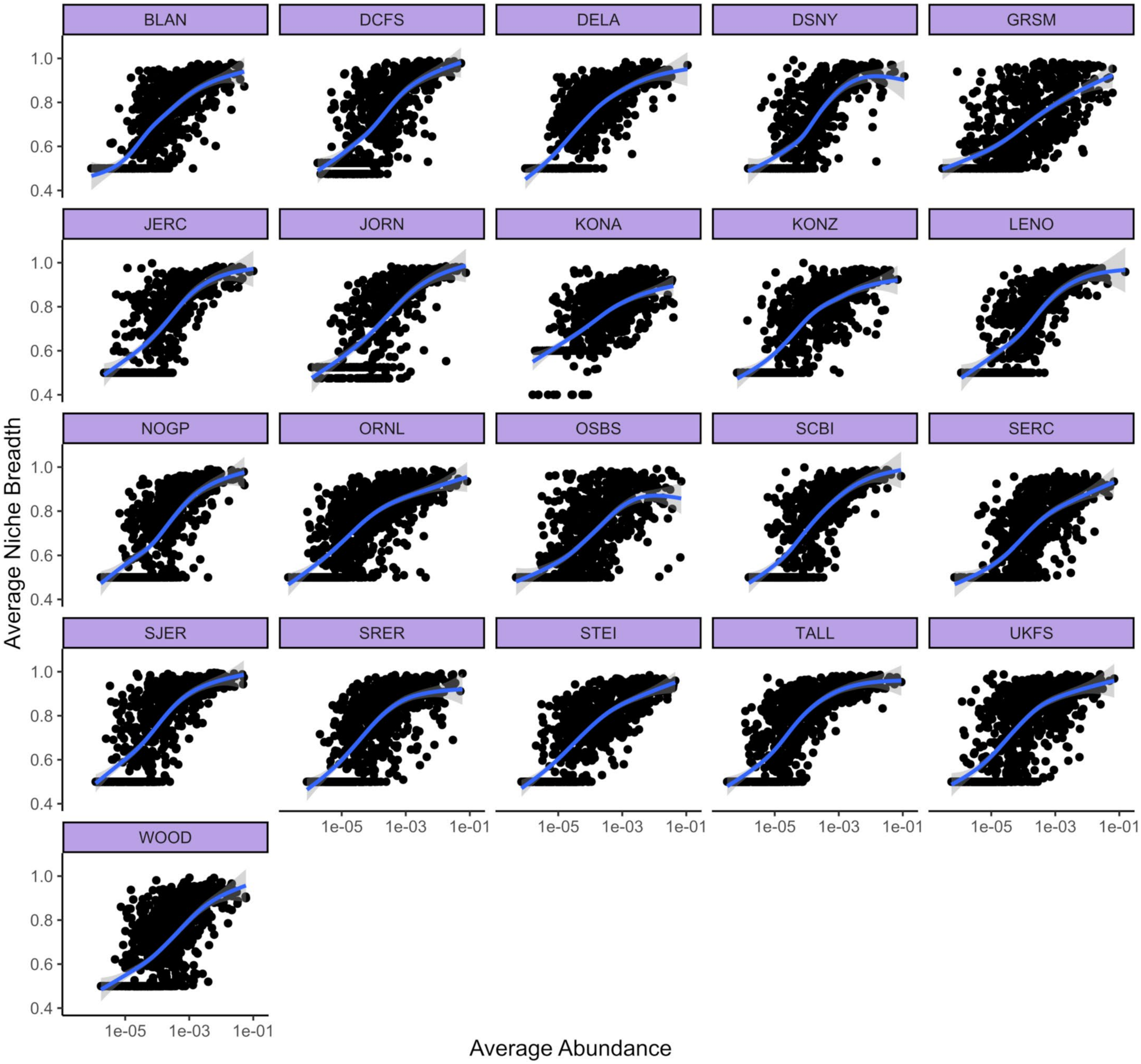


Extended Data Fig. 6 | Phylogenetic relationships of niche breadth in closely-related taxa. Scatter plots of LIPA Moran's I of average niche breadth for all 1230 taxa (points) in all 100 observed trees (each graph). A LOESS fit (blue line) is plotted to visualize if pattern follows linear or quadratic relationships (compared in Fig. 2). A higher LIPA Moran's I indicates more phylogenetic conservation of average niche breadth among closely related taxa. A quadratic relationship (a better fit than a linear model in all trees; Fig. 2) indicates that phylogenetic conservation of average niche breadth is strongest when taxa are highly specialized or highly generalized. The quadratic relationship further supports multidimensional specialization and generalization as opposing niche trajectories.



Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Relationship between average relative abundance and niche breadth of Exact Sequence Variants (A-B) and including carbon/nitrogen niche axes data (C-D). ESV analyses are represented by subfigures A-B and taxa level analyses including carbon/nitrogen data are represented by subfigures C-D. **a)** Average abundances of generalist (dark purple) and specialist (light purple) taxa (14,015 ESV taxa total). Significance calculated with a two-tailed permutational test ($Z = 5.79$, $p = 6.87 \times 10^{-9}$). Boxplots show the median (middle line) and interquartile range (the box). **b)** Average relative abundances of 14015 prokaryotic ESV taxa regressed against average niche breadth. Direction of the relationship is determined using a Spearman's correlation test and significance is calculated using a two-tailed permutational test ($p < 2.20 \times 10^{-16}$). **c)** Average abundances of generalist (dark purple) and specialist (light purple taxa (1085 taxa total). Significance calculated with a two-tailed permutational test ($Z = -6.34$, $p = 2.27 \times 10^{-10}$). Boxplots show the median (middle line) and interquartile range (the box). **d)** Average relative abundances of 1085 taxa regressed against average niche breadth. In B and D, Lines are fitted with LOESS smoothing, shaded regions around the lines are the 95% confidence intervals, and the x-axes are on a $\log_{10}$ scale. Dashed horizontal line indicates the local minima in the bimodal distribution of average niche breadth used to indicate specialists (light purple) and generalists (dark purple). Direction of relationships were determined using a Spearman's correlation test and significance was calculated using two-tailed permutational tests in which abundances were randomized 10,000 times ($p < 2.20 \times 10^{-16}$).



Extended Data Fig. 8 | Average relative abundance within a site is explained by a taxon's average niche breadth at that site. Average relative abundances of taxa at each of 21 sites regressed against average niche breadth in the corresponding site. Lines are fitted with LOESS smoothing, and shaded regions around the lines are the 95% confidence interval. The x-axes are displayed on a $\log_{10}$ scale.

nature portfolio

Corresponding author(s): Damian Hernandez

Last updated by author(s): 30 Jun 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	*Provide a description of all commercial, open source and custom code used to collect the data in this study, specifying the version used OR state that no software was used.*
Data analysis	All computational scripts and data is either available through NEON or at Zenodo (10.5281/zenodo.7747186). The following R (4.0.2) packages were used for data analyses: adephylo (1.1-13), ape (5.7), car (3.012), coin (1.42), dplyr (1.0.7), ggridges (0.5.4), ggplot2 (3.3.5), ggpubr (0.4.0), gplots (3.1.1), gridExtra (2.3), hilbertSimilarity (0.4.3), Hmisc (4.6), lmPerm (2.1.0), matrixStats (0.61.0) msa (1.26.0), patchwork (1.1.1), phangorn (2.8.1), phylobase (0.8.10), phyloseq (1.42.0), phylosignal (1.3), phytools (1.01), plyr (1.8.8), scales (1.1.1), seqinr (4.2-23), stringr (1.4.0), tidyr (1.1.4), vegan (2.57). The following Python (3.10.2) packages were used for data analyses: matplotlib (3.5.1), networkx (2.6.2), pandas (1.3.5), scipy (1.7.3), seaborn (0.11.2), statsmodels (0.13.2). All packages are listed in Supplementary File 1 alongside their purposes in the manuscript.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All raw sequencing and environmental data are publicly available through the NEON database (DP1.10081.001). Scripts to download data from NEON and process sequencing data into ESVs and OTUs are available in Supplementary File 2. OTU abundances from "kingdom" to "species" levels are available in Supplementary File 1. Code to replicate our analyses and a "project" folder containing all the intermediate files and statistical summaries from Rmarkdown scripts are available at Zenodo (10.5281/zenodo.7747186). We used the publicly available GreenGenes database (version 13.5) for taxonomic assignments.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	*Describe how sample size was determined, detailing any statistical methods used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.*
Data exclusions	*Describe any data exclusions. If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Replication	*Describe the measures taken to verify the reproducibility of the experimental findings. If all attempts at replication were successful, confirm this OR if there are any findings that were not replicated or cannot be reproduced, note this and describe why.*
Randomization	*Describe how samples/organisms/participants were allocated into experimental groups. If allocation was not random, describe how covariates were controlled OR if this is not relevant to your study, explain why.*
Blinding	*Describe whether the investigators were blinded to group allocation during data collection and/or analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	*Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).*
Research sample	*State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic*

Research sample	*information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.*
Sampling strategy	*Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.*
Data collection	*Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.*
Timing	*Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Non-participation	*State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.*
Randomization	*If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.*

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	An analysis of prokaryotic niche breadth across multiple niche dimensions using the publicly available NEON database.
Research sample	Environmental data such as soil pH, soil moisture, and nutrient content as well as microbiome sequencing data of the 16S hypervariable 4 region.
Sampling strategy	Sampling of multiple subplots within a plot within 30 sites across the continental United States. For our purposes, we average abundances across subplots per plot to avoid pseudoreplication from plots that are sampled more frequently by the National Ecological Observatory Network.
Data collection	All files are download using the scripts in Supplementary File 2. Data collection is described using the NEON data products as cited in the Manuscript text and in Supplementary File 1.
Timing and spatial scale	Continental scale.
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Reproducibility	All raw data is publicly accessible through NEON and our analyses are provided through Zenodo as a project folder (see data availability statement).
Randomization	*Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.*
Blinding	*Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Did the study involve field work? ☐ Yes ☒ No

Field work, collection and transport

Field conditions	*Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).*
Location	*State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).*
Access & import/export	*Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).*
Disturbance	*Describe any disturbance caused by the study and how it was minimized.*

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging