nature ecology & evolution

Article

https://doi.org/10.1038/s41559-023-02250-2

Identifying keystone species in microbial communities using deep learning

Received: 29 March 2023

Accepted: 16 October 2023

Published online: 16 November 2023

Check for updates

Xu-Wen Wang[1], Zheng Sun[1], Huijue Jia[2,3], Sebastian Michel-Mata[4], Marco Tulio Angulo[5], Lei Dai[6,7], Xuesong He[8,9], Scott T. Weiss[1] & Yang-Yu Liu[1,10]

Previous studies suggested that microbial communities can harbour keystone species whose removal can cause a dramatic shift in microbiome structure and functioning. Yet, an efficient method to systematically identify keystone species in microbial communities is still lacking. Here we propose a data-driven keystone species identification (DKI) framework based on deep learning to resolve this challenge. Our key idea is to implicitly learn the assembly rules of microbial communities from a particular habitat by training a deep-learning model using microbiome samples collected from this habitat. The well-trained deep-learning model enables us to quantify the community-specific keystoneness of each species in any microbiome sample from this habitat by conducting a thought experiment on species removal. We systematically validated this DKI framework using synthetic data and applied DKI to analyse real data. We found that those taxa with high median keystoneness across different communities display strong community specificity. The presented DKI framework demonstrates the power of machine learning in tackling a fundamental problem in community ecology, paving the way for the data-driven management of complex microbial communities.

The notion of keystone species has its roots in food web ecology[1,2]. Since Paine coined it in describing results from his pioneering field experiments in 1969, the notion of keystone species has been widely applied in the ecological literature. Such a broad application (and often abuse) has generated considerable confusion about what precisely a keystone species is[3]. Here, we adopt the original definition by Paine, that is, a keystone species is a species that has a disproportionately large effect on the stability of the community relative to its abundance[1,2]. Existing methods to identify keystone species for macro ecosystems can be classified into two approaches: experimental manipulations and statistical comparisons[4].

Previous studies also suggest that microbial communities harbour keystone species[5–9]. Yet, the keystone species identification approaches developed for macro ecosystems are challenging to apply to large, complex microbial communities[5–9]. For experimental manipulations, targeted removal of each species in a complex community is impossible with current antimicrobial techniques, not to mention the corresponding ethical concerns for host-associated microbial communities

[1]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. [2]School of Life Sciences, Fudan University, Shanghai, China. [3]Institute of Precision Medicine–Greater Bay Area (Guangzhou), Fudan University, Guangzhou, China. [4]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA. [5]Institute of Mathematics, Universidad Nacional Autónoma de México, Juriquilla, Mexico. [6]CAS Key Laboratory of Quantitative Engineering Biology, Shenzhen Institute of Synthetic Biology, Shenzhen Institute of Advanced Technology, Shenzhen, China. [7]University of Chinese Academy of Sciences, Beijing, China. [8]Department of Microbiology, The Forsyth Institute, Cambridge, MA, USA. [9]Department of Oral Medicine, Infection and Immunity, Harvard School of Dental Medicine, Boston, MA, USA. [10]Center for Artificial Intelligence and Modeling, The Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Champaign, IL, USA. e-mail: yyl@channing.harvard.edu

https://doi.org/10.1038/s41559-023-02250-2

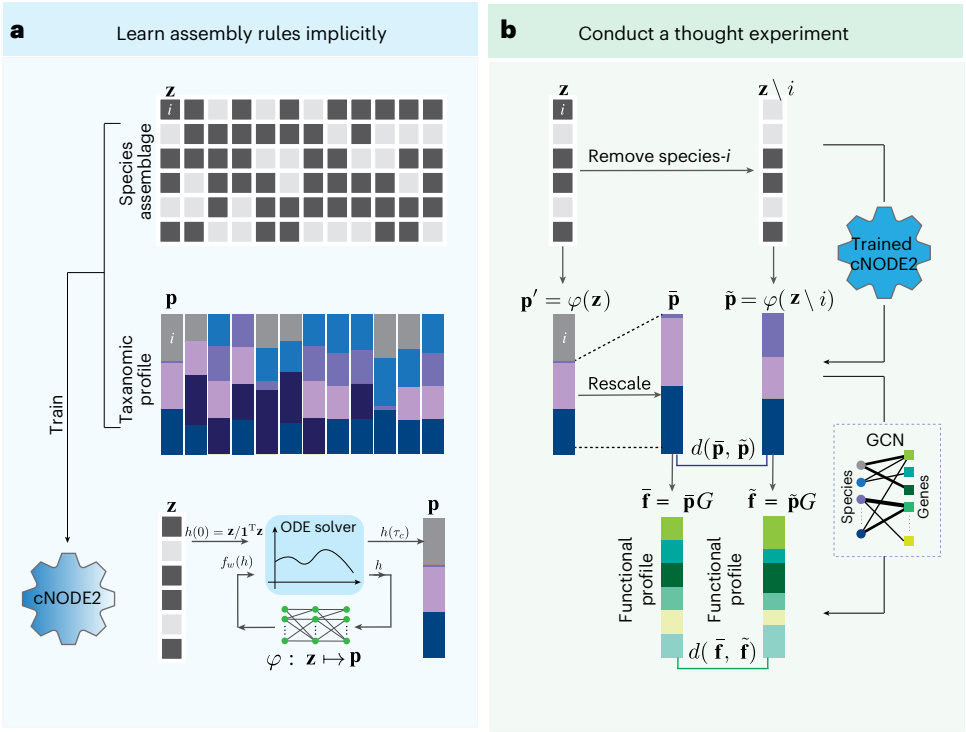

Fig. 1 | Workflow of the DKI framework. a, The species assemblage of a microbiome sample s is represented by a binary vector $\mathbf{z} \in \{0,1\}^N$ whose ith entry z_i is 1 (or 0) if species i is present in (or absent from) this sample. The microbial composition of this sample is characterized by a vector $\mathbf{p} \in \Delta^N$ whose ith entry p_i is the relative abundance of species i in this sample (each colour represents a species) and Δ^N is the probability simplex. A deep-learning model (cNODE2) is trained to learn the map: $\mathbf{z} \in \{0,1\}^N \mapsto \mathbf{p} \in \Delta^N$. ODE, ordinary differential equation. $\mathbf{1} = (1, \cdots, 1)^{\mathrm{T}}$. τ_c is virtual time, h is the variable and f_w is a linear function. **b**, We conduct a thought experiment on the removal of species i. In particular, for the community $s = (\mathbf{z}, \mathbf{p})$ with species collection $\mathbf{z}$ and microbial composition $\mathbf{p}$, we remove species i from $\mathbf{z}$ to form a new species collection $\tilde{\mathbf{z}} = \mathbf{z} \backslash i$. Then, for the new species collection $\tilde{\mathbf{z}}$, we use cNODE2 to predict its new composition $\tilde{\mathbf{p}} = \varphi(\tilde{\mathbf{z}})$. To quantify the impact of the removal of species i, we need to compare the new composition $\tilde{\mathbf{p}}$ with a null composition $\bar{\mathbf{p}}$ in the absence of species i (obtained by assuming that the removal of species i will not affect other species' abundances at all). The structural impact of the removal of species i on the community $s = (\mathbf{z}, \mathbf{p})$ can be defined as the distance or dissimilarity between $\bar{\mathbf{p}}$ and $\tilde{\mathbf{p}}$, that is, $d(\bar{\mathbf{p}}, \tilde{\mathbf{p}})$. Similarly, the functional impact of the removal of species i on the community $s = (\mathbf{z}, \mathbf{p})$ can be defined as the distance or dissimilarity between $\bar{\mathbf{f}}$ and $\tilde{\mathbf{f}}$ (each colour represents a gene), that is, $d(\bar{\mathbf{f}}, \tilde{\mathbf{f}})$. Here, the functional profile $\bar{\mathbf{f}}$ (or $\tilde{\mathbf{f}}$) can be computed by multiplying $\bar{\mathbf{p}}$ (or $\tilde{\mathbf{p}}$) with the incidence matrix of the GCN, respectively.

such as the human gut microbiome. As for statistical comparisons, finding two communities that differ by just one species is challenging, especially for complex host-associated microbial communities (for example, the human gut microbiome) with very personalized compositions[10,11]. Moreover, statistical comparisons can suffer from numerous confounding factors[12]. To resolve the above limitations, one may consider directly inferring a population dynamics model to predict the temporal behaviour of microbial communities and then identify keystone species through numerical simulations of targeted species removal. Yet, model misspecification and the requirement for high-quality absolute abundance data[13–15] for those dynamics inference methods limit their application for identifying keystone species in large, complex microbial communities.

A recent numerical study[16] claimed that those highly connected (that is, 'hubs') and high-betweenness-centrality species in the microbial correlation network are keystone species of microbial communities[6,17]. Despite the popularity and interpretability of the correlation network approach, we think this claim is problematic for at least two reasons. First, edges in microbial correlation networks do not represent direct ecological interactions but just statistically significant co-occurrences or mutual exclusions of species. Second, the impact of a species' removal naturally depends on the resident community. This underscores a fundamental challenge in keystone species identification—the community specificity, that is, a species may be a keystone in one community but not necessarily a keystone in another community, which is completely ignored based on degree, betweenness or any topological indices in the correlation (or ecological) network.

So far, very few microbial species have been experimentally confirmed as keystones[5,18–20]. An efficient method to systematically identify community-specific keystone species in complex microbial communities is still lacking[21–23]. In fact, we even lack a widely accepted operational definition of keystoneness—an index to quantify the role of a species to be a keystone. In this Article, we first proposed an operational definition of keystoneness for microbial species on the basis of commonly available relative abundance data. Then, we proposed a data-driven keystone species identification (DKI) framework to compute the keystoneness. The DKI framework does not assume any particular ecological model, naturally avoiding the model misspecification issue. Moreover, the DKI framework quantifies the keystoneness of each species for each community (sample). Hence, it naturally considers the community specificity of keystoneness.

Results

An operational definition of keystoneness for microbial communities

Consider a microbiome sample or microbial community $s = (\mathbf{z}, \mathbf{p})$, where the species assemblage of the community s is represented by a binary vector $\mathbf{z} \in \{0,1\}^N$ whose ith entry z_i is 1 (or 0) if species i is present in (or absent from) s. The microbial composition or taxonomic profile of this community is characterized by a compositional vector $\mathbf{p} \in \Delta^N$ whose ith entry p_i represents the relative abundance of species i in s

https://doi.org/10.1038/s41559-023-02250-2

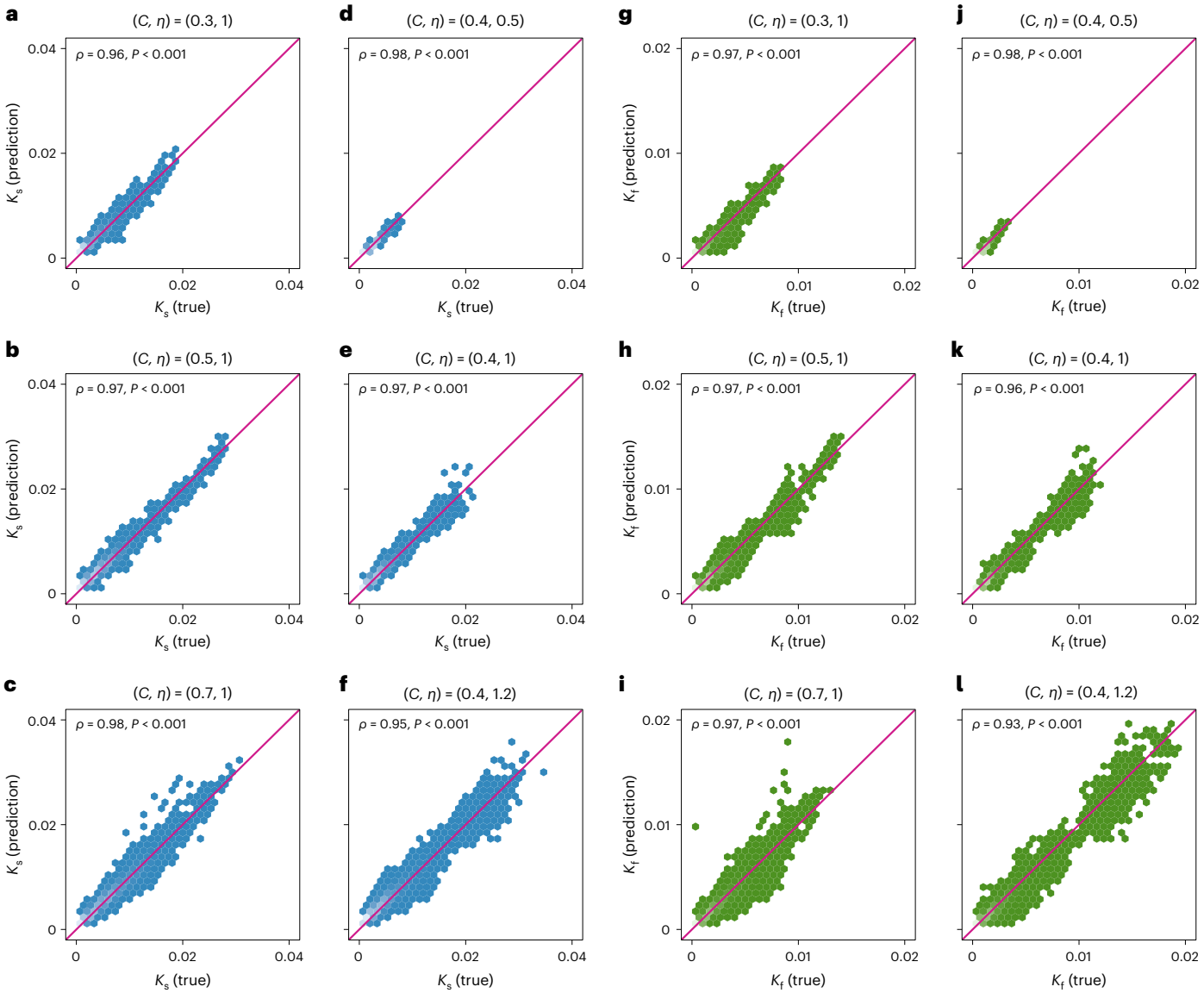

Fig. 2 | In silico validation of the DKI framework. Results obtained for pools of $N = 100$ species with GLV population dynamics. We generate 500 samples to validate the performance of DKI. The population dynamics is characterized by two parameters: the network connectivity $C > 0$ and the boosting strength $\eta > 0$. **a**–**f**, Hexbin plot of the predicted structural keystoneness and true structural keystoneness with network connectivity $C = 0.3$ (**a**), $C = 0.5$ (**b**) and $C = 0.7$ (**c**) or boosting strength $\eta = 0.5$ (**d**), $\eta = 1$ (**e**) and $\eta = 1.2$ (**f**). **g**–**l**, Hexbin plot of the predicted functional keystoneness and true functional keystoneness with network connectivity $C = 0.3$ (**g**), $C = 0.5$ (**h**) and $C = 0.7$ (**i**) or boosting strength $\eta = 0.5$ (**j**), $\eta = 1$ (**k**) and $\eta = 1.2$ (**l**). For different network connectivities, the characteristic interaction strength is $\sigma = 0.01$ and the boosting strength is $\eta = 1$. For different boosting strengths, the characteristic interaction strength is $\sigma = 0.01$ and the network connectivity is $C = 0.4$. Each panel shows the Spearman correlation (ρ) between the predicted and true keystoneness values, and the P value obtained with a two-sided t test. The red line in each panel represents the case of perfect regression, where the predicted keystoneness equals the true keystoneness.

and Δ^N is the probability simplex. Inspired by the keystoneness definition in macroecology[24], we define the keystoneness of species in microbial communities as the product of two components: the impact component and the biomass component.

More specifically, we define the structural keystoneness of species i in a community $s = (\mathbf{z}, \mathbf{p})$ as

$$K_s(i,s) \equiv d\left(\tilde{\mathbf{p}}, \bar{\mathbf{p}}\right)(1 - p_i), \quad (1)$$

where the impact component $d(\tilde{\mathbf{p}}, \bar{\mathbf{p}})$ quantifies the structural impact of the removal of species i on community s, while the biomass component $(1 - p_i)$ captures how disproportionate this impact is.

For the impact component, we quantify the impact of the removal of species i on the structure of community s as the dissimilarity between the taxonomic profiles $\tilde{\mathbf{p}}$ and $\bar{\mathbf{p}}$, that is, $d(\tilde{\mathbf{p}}, \bar{\mathbf{p}})$, where $\tilde{\mathbf{p}}$ is the new community composition after the removal of species i and $\bar{\mathbf{p}}$ is the null composition obtained by assuming that the removal of species i will not affect any other species (Methods).

Similarly, we define the functional keystoneness of species i in a community $s = (\mathbf{z}, \mathbf{p})$ as

$$K_f(i,s) \equiv d\left(\tilde{\mathbf{f}}, \bar{\mathbf{f}}\right)(1 - p_i). \quad (2)$$

Here, the dissimilarity between the new functional profile $\tilde{\mathbf{f}}$ and the null functional profile $\bar{\mathbf{f}}$, that is, $d(\tilde{\mathbf{f}}, \bar{\mathbf{f}})$, captures the impact of the removal of species i on the function of community s. $\tilde{\mathbf{f}}$ (or $\bar{\mathbf{f}}$) can be computed from $\tilde{\mathbf{p}}$ (or $\bar{\mathbf{p}}$) and the genomic content network (GCN)[25], respectively (Methods).

https://doi.org/10.1038/s41559-023-02250-2

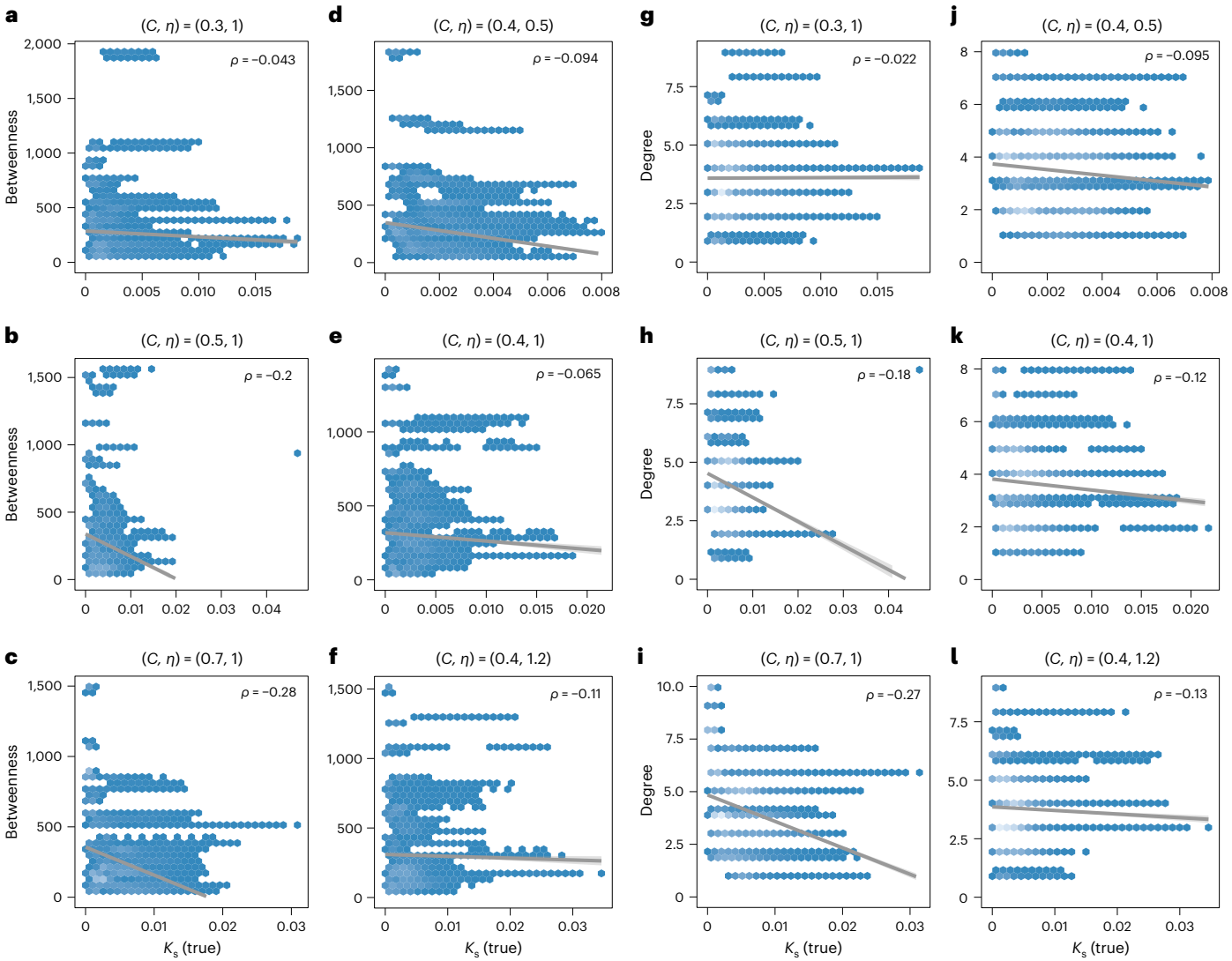

Fig. 3 | Traditional topological indices calculated from the undirected correlation network do not correlate with structural keystoneness. Synthetic samples (taxonomic profiles) were generated from the GLV model with $N = 100$ species in the species pool. The initial species collection of each sample (local community) consists of 50 species drawn randomly from the species pool (Supplementary Discussion 2). The structural keystoneness K_s of each species in each sample was calculated from the simulated species removal process in the GLV model. Two traditional topological indices (betweenness and degree) of each species were calculated from the correlation network of species abundances constructed using sparCC[41] with a threshold of 0.1. **a**–**c**, Hexbin plot of structural keystoneness versus betweenness for ecological network connectivity of $C = 0.3$ (**a**), $C = 0.5$ (**b**) and $C = 0.7$ (**c**) with characteristic interaction strength of $\sigma = 0.01$ and boosting strength of $\eta = 1$. **d**–**f**, Hexbin plot of structural keystoneness versus betweenness with $\eta = 0.5$ (**d**), $\eta = 1$ (**e**) and $\eta = 1.2$ (**f**) with $C = 0.4$ and $\sigma = 0.01$. **g**–**i**, Structural keystoneness versus degree with $C = 0.3$ (**g**), $C = 0.5$ (**h**) and $C = 0.7$ (**i**) with $\sigma = 0.01$ and $\eta = 1$. **j**–**l**, Structural keystoneness versus degree with $\eta = 0.5$ (**j**), $\eta = 1$ (**k**) and $\eta = 1.2$ (**l**) with $C = 0.4$ and $\sigma = 0.01$. The grey line in each panel is the best fit in linear regression.

We emphasize that the structural (or functional) keystoneness defined here is community-specific, which is fundamentally different from those topological indices used in the food web and other ecological systems[26].

The DKI framework

Consider a particular habitat (or meta-community) that harbours a pool of N different microbial species, denoted as $\Omega = \{1, \cdots, N\}$. Suppose we have a large set of microbiome samples $\mathcal{S} = \{1, \ldots, M\}$ collected from this habitat. A microbiome sample $s \in \mathcal{S}$ can be viewed as a local community of the habitat. We assume that the collected samples roughly represent the steady states of the local communities so that they can be used to learn the assembly rules of those communities.

The DKI framework consists of two phases. In the first phase (Fig. 1a), we implicitly learn the assembly rules of microbial communities in this habitat using a deep-learning method with $\mathcal{S}$ as the training data. This is achieved by learning a map from the species assemblage $\mathbf{z}$ of a sample $s = (\mathbf{z}, \mathbf{p})$ to its taxonomic profile $\mathbf{p}$, that is, $\varphi : \mathbf{z} \mapsto \mathbf{p}$. Various deep-learning methods, including multi-layer perceptron[27] or ResNet[28], can be used to learn such a map without using any population dynamic model, but with a few reasonable assumptions (for example, the universality of microbial dynamics, steady-state samples, no true multi-stability and enough training samples; see Supplementary Discussion 1 for details) to ensure the problem is mathematically well defined. Here, based on our previous work[29], we developed composition Neural Ordinary Differential Equation version 2.0 (cNODE2) to learn the map φ (for details see Supplementary Discussion 1). Learning this map φ will enable us to predict what will happen to the taxonomic profile of a local community upon removal of any species.

https://doi.org/10.1038/s41559-023-02250-2

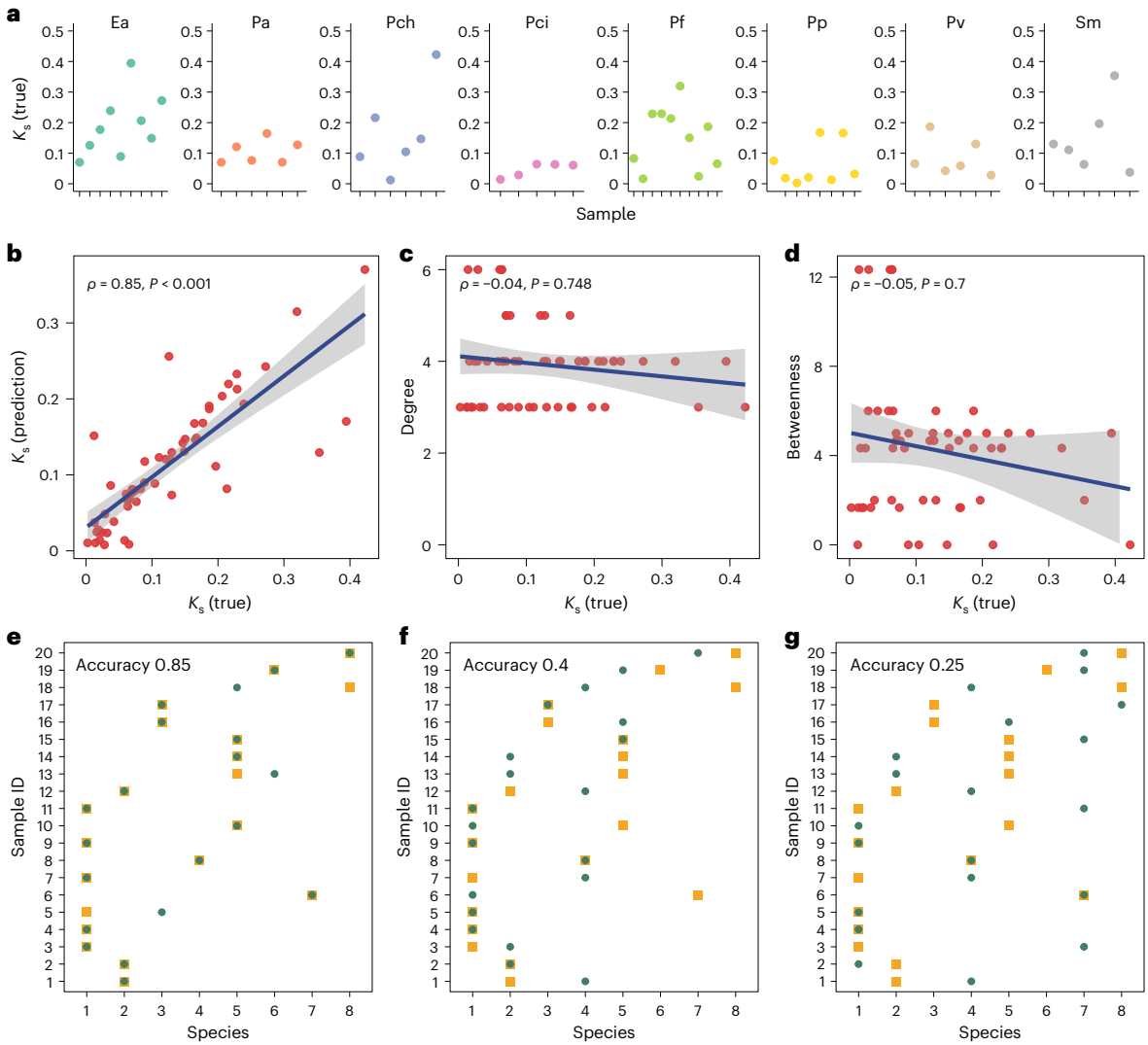

Fig. 4 | In vitro validation of cNODE in keystoneness prediction. We analysed data from a synthetic consortium of eight soil species: Ea, Sm, Pci, Pp, Pa, Pch, Pf and Pv to validate DKI. **a**, The true keystoneness of each species in different communities. **b**–**d**, The calculated Spearman correlations between the true keystoneness and the DKI-predicted keystoneness (**b**), degree (**c**) and betweenness (**d**). Shaded areas indicate 95% confidence intervals of the regression (blue line). **e**–**g**, The calculated true keystone species in each sample (yellow squares) and the DKI-predicted keystone (green circles) (**e**), degree-predicted (**f**) and betweenness-predicted keystone species (**g**). The true (predicted) keystone species is considered to be the species with the highest true (predicted) keystoneness. Two traditional topological indices (betweenness and degree) of each species were calculated from the correlation network of species abundances constructed using sparCC[41] with a threshold of 0.05. The *P* values for the Spearman correlation coefficients (ρ) were obtained with a two-sided *t* test.

In the second phase (Fig. 1b), to quantify the community-specific keystoneness of species i in a local community or microbiome sample s, we conduct a thought experiment of removing species i from s and use cNODE2 to compute the impact of the removal of species i on s. In particular, for $s = (\mathbf{z}, \mathbf{p})$ with species collection $\mathbf{z}$ and microbial composition $\mathbf{p}$, we remove species i from $\mathbf{z}$ to form a new species collection $\tilde{\mathbf{z}} = \mathbf{z}\backslash i$. Then, for the new species collection $\tilde{\mathbf{z}}$, we use cNODE2 to predict its composition $\tilde{\mathbf{p}} = \varphi(\tilde{\mathbf{z}})$. To quantify the impact of the removal of species i, we need to compare the new composition $\tilde{\mathbf{p}}$ with the null composition $\bar{\mathbf{p}}$ (with $\bar{p}_i = 0$ and $\bar{p}_j = p_j/\sum_{k\neq i} p_k$ for $j \neq i$). In reality, the map φ cannot be learned perfectly, and the composition prediction always contains some error. To take this into account, we can compute the null composition $\bar{\mathbf{p}}$ by renormalizing the relative abundances of the remaining species from the predicted composition of the original community, that is, $\bar{p}_j = p'_j/\sum_{k\neq i} p'_k$, where $\mathbf{p}' = (p'_k) = \varphi(\mathbf{z})$. This way, the prediction errors in $\tilde{\mathbf{p}}$ and $\bar{\mathbf{p}}$ will be cancelled to some extent, and hence the predicted keystoneness will be more accurate (Supplementary Fig. 1). From the predicted $\tilde{\mathbf{p}}$ and $\bar{\mathbf{p}}$, we can compute the functional profiles $\tilde{\mathbf{f}}$ and $\bar{\mathbf{f}}$, and then compute both the structural keystoneness and the functional keystoneness of species i in the community $s = (\mathbf{z}, \mathbf{p})$.

Validation of DKI framework using synthetic dataset

To demonstrate DKI's performance in keystoneness prediction, we generated synthetic data using the generalized Lotka–Volterra (GLV) model with $N = 100$ species in the species pool (meta-community). The initial species collection of each sample (local community) consists of 50 species randomly drawn from the species pool (Supplementary Discussion 2). We characterized the population dynamics of the meta-community using two parameters: (1) the connectivity C of the underlying ecological network (which encodes all the pairwise inter-species interactions), representing the probability that two species interact directly, and (2) the characteristic interaction strength σ representing the typical impact of one species over the

https://doi.org/10.1038/s41559-023-02250-2

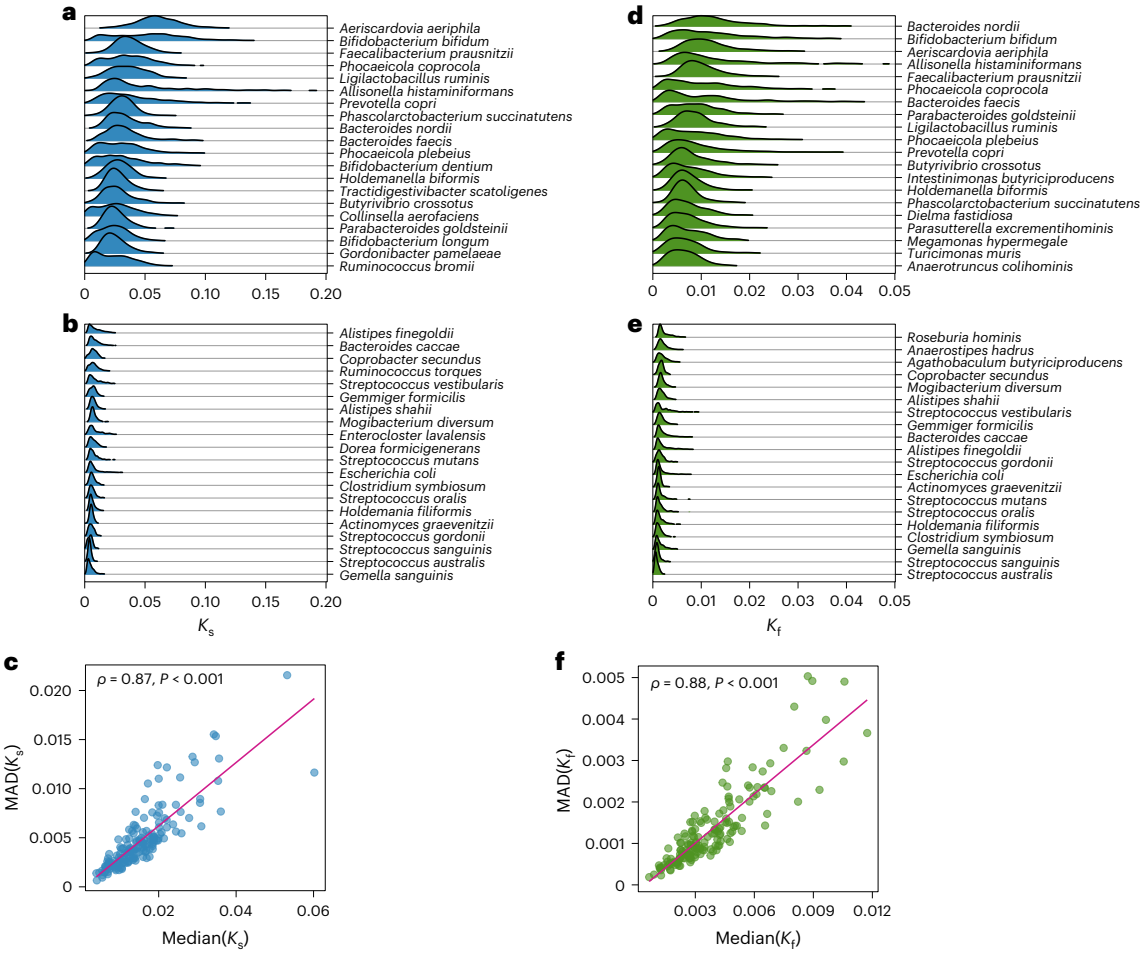

Fig. 5 | Keystone species in the human gut microbiome. We applied DKI to a large-scale human gut microbiome dataset collected in curatedMetagenomicData[32]. This dataset includes 2,815 faecal samples from healthy adults without antibiotics usage. In total, we have 1,103 species. **a**,**b**, The structural keystoneness distribution of the top 20 (**a**) and bottom 20 species (**b**) ranked by their median structural keystoneness. **c**, The Spearman correlation between the median structural keystoneness and median deviation of the keystoneness of each species. **d**,**e**, The functional keystoneness distribution of the top 20 (**d**) and bottom 20 species (**e**) ranked by their median functional keystoneness. **f**, The Spearman correlation between the median functional keystoneness and the median deviation of keystoneness of each species. In panels **a**, **b**, **d** and **e**, the top/bottom 20 species were selected among species present in at least 10% of total samples. The *P* values for the Spearman correlation coefficients (ρ) were obtained with a two-sided *t* test. The red line in **c** and **f** represents the best fit of linear regression.

per-capita growth rate of another species if they interact. To introduce keystone species into the local communities, inspired by a previous study[23], we amplified all the link weights (inter-species interactions) to $\widetilde{a}_{ij} = \theta_{ij} a_{ij}$, where θ_{ij} is drawn randomly from a log-normal distribution with mean 0 and standard deviation η. This will generate a few strong interactions, presumably leading to a few species with high keystoneness. Hereafter, we refer η as the characteristic amplification coefficient.

We first trained cNODE2 to minimize the loss function defined as the mean Bray–Curtis dissimilarity between the true and predicted compositions for all samples (Supplementary Discussion 1). Then, we evaluated DKI using all possible new species collections $\tilde{\mathbf{z}}$ obtained by removing each species present in each sample. We systematically evaluated the performance of DKI in predicting the structural keystoneness K_s using simulated data generated from the GLV model with different values for the parameter pair (C, η). We found that DKI can accurately predict the K_s of each species over different samples for a wide range of C or η values (Fig. 2a–f). The Spearman correlation ρ between the true K_s (calculated from the simulated species removal process in the GLV model) and the predicted K_s is around 0.97 with $P < 0.001$ (two-sided t test). See Supplementary Figs. 2 and 3 and Supplementary Discussion 1.3 for a discussion of the performance of DKI when varying other factors.

To calculate the functional keystoneness using the synthetic data, we generated a random GCN displaying nested structure (with a nestedness metric based on overlap and decreasing fill[30] (NODF) of 0.31) for 100 species and 500 genes. We found that DKI can also accurately predict the K_f of each species over different samples for a wide range of C or η values (Fig. 2g–l). The Spearman correlation ρ between the true K_f (calculated from the simulated species removal process in the GLV model and the randomly generated GCN) and the predicted K_f is around 0.96 with $P < 0.001$.

We emphasize that each species' structural (or functional) keystoneness is context dependent or community specific. Yet, existing methods, especially those based on topological indices of correlation (or ecological) networks constructed (or inferred) from a collection of samples, cannot offer community-specific keystoneness. Moreover, those topological measures do not correlate with each species' structural (or functional) keystoneness. To demonstrate this point, we generated synthetic data (Supplementary Discussion 2) and compared the structural keystoneness K_s (calculated from the simulated species removal process in the GLV model) with two classical topological

indices, that is, degree (the number of species connected with the species under consideration) and betweenness (the frequency of the species under consideration on the shortest paths connecting all pairs of other species) in the directed ecological network or the undirected correlation network. As shown in Fig. 3 and Supplementary Fig. 4, the two topological indices do not correlate with K_s at all, regardless of whether we use the correlation network (Fig. 3) or the ecological network (for details see Supplementary Fig. 4 and Supplementary Discussion 3).

Validation of DKI using in vitro synthetic microbial communities

We then validated our DKI framework using data from an in vitro study of synthetic microbial communities comprising up to eight soil bacterial species: *Enterobacter aerogenes* (Ea), *Serratia marcescens* (Sm), *Pseudomonas citronellolis* (Pci), *Pseudomonas putida* (Pp), *Pseudomonas aurantiaca* (Pa), *Pseudomonas chlororaphis* (Pch), *Pseudomonas fluorescens* (Pf) and *Pseudomonas veronii* (Pv). Those communities involved 101 different species combinations: all 8 solos, 28 duos, 56 trios, all 8 septets and 1 octet[31]. To validate the DKI, we used duos and trios (in total, 42 species combinations). For each sample with species collection $\mathbf{z}$, we examined whether there is a corresponding sample with species collection $\tilde{\mathbf{z}} = \mathbf{z}\backslash i$. In total, 56 sample pairs were identified to test the prediction of DKI. For each sample pair ($\mathbf{z}$, $\tilde{\mathbf{z}}$), we used the ground-truth sample $\tilde{\mathbf{z}}$ as the test set and the remaining samples to train the DKI. We found that the keystoneness of each species displays strong community specificity for most of the species (Fig. 4a). Then, we compared the predicted keystoneness of each species with its true keystoneness, finding that the keystoneness predicted by DKI is consistent with the true keystoneness (Spearman correlation of $\rho = 0.85$, $P < 0.001$; Fig. 4b). Importantly, the two topological indices (that is, degree and betweenness) do not correlate with K_s at all (Fig. 4c,d). To examine the sensitivity of each method in keystone identification, we considered the species with the highest true keystoneness as keystone species in each community, and the predicted keystone as the species with the highest predicted keystoneness or topological indices. We found that DKI yields the highest accuracy 0.85 (Fig. 4e) compared with degree (accuracy 0.4; Fig. 4f) and betweenness (accuracy 0.25; Fig. 4g).

Keystone species in the human microbiome

We applied the DKI framework to the human gut microbiome data collected in the curatedMetagenomicDatabase[32]. We focused on the metagenomic data of stool samples of healthy adults aged 18–65 years and without antibiotics usage. In total, we have 2,815 samples involving 1,103 species. In the National Center for Biotechnology Information Reference Sequence Database (RefSeq), species are defined on the basis of comprehensive, integrated, non-redundant and well-annotated reference sequences, including genomic, transcript and protein data.

We first trained cNODE2 by using all the 2,815 samples. Then, for each of the 2,815 samples, we computed the structural keystoneness K_s for each species present in the sample. For species present in at least 10% of the samples, we ranked them based on their median structural keystoneness: median(K_s). Figure 5a,b shows the top 20 and bottom 20 species, respectively. We found that those species with higher median(K_s), for example, *Prevotella copri*, tend to have a larger variation of their K_s across different samples, suggesting stronger community specificity (Fig. 5a), while those species with lower median(K_s), for example, *Alistipes finegoldii*, tend to have a smaller K_s variation, suggesting weaker community specificity (Fig. 5b). In addition, those species with the highest keystoneness tend to have a much larger biomass component than the impact component (Supplementary Fig. 5).

To systematically explore the community specificity of those species' structural keystoneness, we plotted their median keystoneness median(K_s) versus their median absolute deviation of structural keystoneness MAD(K_s) over all samples. We found that MAD(K_s) is highly correlated with median(K_s) with Spearman correlation of $\rho = 0.87$ ($P < 0.001$; Fig. 5c). This result indicates that taxa with low median structural keystoneness are unlikely to be keystone taxa in any community. By contrast, taxa with high median keystoneness have high keystoneness (and hence are probably keystone taxa) in some communities, but they can also have small keystoneness in other communities. Similar results were also observed from human oral microbiome and environmental microbiomes (for details see Supplementary Discussions 4–7 and Supplementary Figs. 6–8).

We noticed that the largest (structural) keystoneness value is still smaller than 0.2 (Fig. 5a), which is much lower than the upper bound of our keystoneness metric, that is, 1. This can be explained by the fact that many complex microbial communities (including the human microbiome) typically have high functional redundancy[25], meaning that many phylogenetically unrelated species carry similar genes and perform similar functions. A high level of functional redundancy can be related to the reliability with which an ecosystem will continue to deliver services under moderate species loss. Such functional redundancy has been considered to underlie the stability and resilience of microbial communities. Therefore, a complex microbial community with high functional redundancy will not be so fragile that removing one species will cause a collapse in the services it provides (Supplementary Fig. 9).

To compute the functional keystoneness, we constructed a reference GCN (for details see Supplementary Discussion 4). We found that, in general, a species' functional keystoneness is smaller than its structural keystoneness (Supplementary Fig. 10). This is also closely related to the concept of functional redundancy[25]. Similar to our results on structural keystoneness, we found that those species with higher median(K_f), for example, *Bifidobacterium bifidum*, tend to have a larger variation of their K_f across different samples, suggesting stronger community specificity (Fig. 5d), while those species with lower median(K_f), for example, *Roseburia hominis*, tend to have a smaller K_f variation, suggesting weaker community specificity (Fig. 5e). We found that median(K_f) and MAD (K_f) are strongly correlated with a Spearman correlation of $\rho = 0.88$ ($P < 0.001$; Fig. 5f).

On the basis of the ranking of the median structural keystoneness, we found that, among those top-ranking species, many have been identified as keystone species that carry unique functions and are essential for maintaining host–microbe haemostasis[33]. For example, Bifidobacteria are keystone microorganisms in gut microbiota associated with early life[34]; *Prevotella copri* is a keystone species of a healthy human intestinal mucosa[35]; *Faecalibacterium prausnitzii* is a keystone species that produces butyrate and whose reduced abundance has been associated with Crohn's disease[36]; *Bifidobacterium longum* is a minority species, but influences gut microbiota formation by breaking down complex carbohydrates and providing degradants for other bacterial groups to use[37]; *Ruminococcus bromii* plays key roles in promoting the synergistic utilization of resistant starch by initiating degradation of insoluble resistant starch particles[5,33] (Fig. 5a).

Interestingly, some of the species with high median structural keystoneness also have high median functional keystoneness, for example, *Faecalibacterium prausnitzii* and *Prevotella copri* (Fig. 5d). Based on the high median functional keystoneness, we also identified some potential keystone species that have been reported to perform important functions. For example, *Intestinimonas*-like bacteria are important butyrate producers that utilize *N*-ε-fructosyllysine and lysine in formula-fed infants and adults[38].

Discussion

The concept of keystone species has been extensively investigated in ecology. Despite the considerable confusion and difference[3], the operational definitions agree that a keystone species disproportionately affects its natural environment relative to its abundance. Systematically identifying keystone species in complex microbial communities

is very challenging owing to our limited knowledge of the population dynamics of those communities, as well as many logistical and ethical concerns regarding the manipulation of those communities. In this work, we propose a data-driven framework to systematically identify keystone species in complex microbial communities. This framework enables us to compute the structural and functional keystoneness of each species in a community for the first time. Our framework can be used to facilitate data-driven management of complex microbial communities.

We emphasize that the proposed framework is general enough and can be modified in many different ways. For example, instead of using cNODE2, one can use other deep-learning models (for example, multi-layer perceptron) to learn the map $\varphi : \boldsymbol{z} \mapsto \boldsymbol{p}$ (Supplementary Fig. 11). Moreover, instead of using Bray–Curtis dissimilarity, one can use other dissimilarity or distance measures (for example, the weighted UniFrac distance) to quantify the structural or functional impact of species' removal (Supplementary Fig. 12). One can also use other formulas that combine the impact component and the biomass component to quantify the keystoneness[24]. In addition, beyond studying the impact of a species' removal on the community-level functional profile, we can also focus on its impact on any specific microbial function and hence quantify its sensitivity. For instance, one can calculate the relative abundances of a function before and after a species' removal, respectively (Supplementary Fig. 13 and Methods).

Among all the five microbiome datasets analysed in this work, we found that the keystoneness value of in vitro synthetic microbial communities can be higher than 0.4. However, the keystoneness of the other four large real microbial communities is much lower than 1. This result could be interpreted as saying that there are hardly any keystone species for the studied microbiome samples in those four datasets. Importantly, this conclusion does not imply that there are hardly any keystone species for any microbial community, because of the strong community specificity of keystoneness. Indeed, one may design a synthetic community with a few keystone species present. However, for naturally observed microbial communities (be they host-associated or host-free), we believe that the chance of finding keystone species is quite low. In addition, the relative ranking of keystoneness might be more important, instead of the absolute value, in quantitatively identifying keystone species. For example, the species with highest median keystoneness across communities are more likely to be keystone species.

We admit that there are some caveats to our calculation/interpretation of functional keystoneness (Supplementary Discussion 8). All the functions are potential functions encoded in the microbial genomes. They do not have to be active. To study the true functions, it is necessary to leverage metaproteomics data. It is known that the (protein-level) functional profiles, or the selective expression of proteins, might depend on the community itself. That is, some species might express certain proteins to consume certain resources to avoid niche overlap with other species in the community. In other words, the protein-content network is community dependent, while the GCN is independent of the community[39].

Methods

Keystoneness calculation

Inspired by the keystoneness definition in macroecology[24], we defined the structural keystoneness of species in microbial communities as the product of two components: the impact component and the biomass component. The impact component is computed as the dissimilarity between the taxonomic profiles $\tilde{\mathbf{p}}$ and $\bar{\mathbf{p}}$, that is, $d(\tilde{\mathbf{p}}, \bar{\mathbf{p}})$. Here, $\tilde{\mathbf{p}}$ is the new community composition after the removal of species i, while $\bar{\mathbf{p}}$ is the null composition computed by simply setting $\bar{p}_i = 0$ and renormalizing the relative abundances of the remaining species, that is, $\bar{p}_j = p_j / \sum_{k \neq i} p_k$ for $j \neq i$. Note that $d(\tilde{\mathbf{p}}, \bar{\mathbf{p}})$ can be any distance or dissimilarity measure, for example, the Bray–Curtis dissimilarity. The biomass component is simply computed as $(1 - p_i)$.

Note that our structural keystoneness $K_s(i,s) = d(\tilde{\mathbf{p}}, \bar{\mathbf{p}})(1 - p_i)$ has clear lower and upper bounds. The lower bound of $K_s(i,s)$ is 0, representing the case that species i is not interacting with any other species in the community, hence its removal will have zero impact on the community. This is, of course, an extreme case. The upper bound of $K_s(i,s)$ is 1. This corresponds to another extreme case that the relative abundance p_i of species i is close to 0, and its removal will cause many other species to die out and only a few species survive, and the sum of those survived species' initial relative abundances is also close to 0. Hence, mathematically, our quantitative metric describes the qualitative keystoneness (in terms of species extinction or system collapse). In fact, our keystoneness definition is more generic in the sense that, as long as species i has a very low relative abundance and its removal will have a very large impact on the community composition (not necessarily leading to many species extinctions), then its keystoneness will be very high.

To compute the functional keystoneness $K_f(i,s) = d(\tilde{\mathbf{f}}, \bar{\mathbf{f}})(1 - p_i)$, we need to compute the functional profiles $\tilde{\mathbf{f}}$ and $\bar{\mathbf{f}}$ from their taxonomic profiles $\tilde{\mathbf{p}}$ and $\bar{\mathbf{p}}$, and the GCN[25]. Here, the GCN is a weighted bipartite graph connecting the N species to their genes (Supplementary Discussions 4 and 5). Suppose that there are, in total, M genes in the metagenome of the N species. The GCN can then be represented by an incidence matrix $G = (G_{ia}) \in \mathbb{R}^{N \times M}$, where a non-negative integer G_{ia} indicates the copy number of gene a in the genome of species i. The gene composition or functional profile $\mathbf{f} \in \Delta^M$ of a microbiome sample with taxonomic profile $\mathbf{p}$ can then be calculated as $\mathbf{f} = c\boldsymbol{p}G$, where $c = \left[\sum_{a=1}^{M} \sum_{i=1}^{N} p_i G_{ia}\right]^{-1}$ is a normalization constant.

Function sensitivity

Instead of studying the impact of a species' removal on the community-level functional profile, we can also focus on the impact on any specific function and hence quantify the sensitivity of the function. This study provides insight into microbial metabolites and associated functional pathways in controlling host physiology[33]. Here, we quantified the sensitivity of each specific function (that is, metabolic pathway) in the human gut microbiome by calculating its abundance change, that is, the abundance change in pathway a caused by the removal of species i from community s, given by $\delta_a^{(s,i)} = |f_a^{(s)} - f_a^{(s)}(\mathbf{z}\backslash i)| / (f_a^{(s)} + \epsilon)$, where $f_a^{(s)}$ and $f_a^{(s)}(\mathbf{z}\backslash i)$ are the relative abundances of pathway a before and after the removal of species i, respectively, and $\epsilon = 10^{-10}$. Then, we ranked the pathways based on their mean change over different species and samples/communities, that is, $\bar{\delta_a} = \sum_{i=1}^{N} (\sum_{j=1}^{M_i} \delta_a^{(s,i)} / M_i) / N$, where M_i is the prevalence of species i.

Reporting summary

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

Data availability

Gut microbiome data were collected from the curatedMetagenomicData[32] database. Oral microbiome data are available at the CNGB Sequence Archive (CNSA) of the China National GeneBank DataBase (CNGBdb) (CNSA CNP0000687 for the 4D-SZ cohort and CNP0001221 for the Yunnan cohort). Coral and soil microbiome data were collected from Qiita[40] (IDs 10895 and 2104). Data supporting our findings are provided at https://github.com/spxuw/DKI.

Code availability

The code used in this work is available at https://github.com/spxuw/DKI.

References

1. Paine, R. T. A note on trophic complexity and community stability. *Am. Nat.* **103**, 91–93 (1969).
2. Paine, R. T. Food web complexity and species diversity. *Am. Nat.* **100**, 65–75 (1966).

3. Cottee-Jones, H. E. W. & Whittaker, R. J. The keystone species concept: a critical appraisal. *Front Biogeogr.* **4**, 117–127 (2012).
4. Power, M. E. et al. Challenges in the quest for keystones: identifying keystone species is difficult—but essential to understanding how loss of species will affect ecosystems. *BioScience* **46**, 609–620 (1996).
5. Ze, X., Duncan, S. H., Louis, P. & Flint, H. J. *Ruminococcus bromii* is a keystone species for the degradation of resistant starch in the human colon. *ISME J.* **6**, 1535–1543 (2012).
6. Banerjee, S. et al. Network analysis reveals functional redundancy and keystone taxa amongst bacterial and fungal communities during organic matter decomposition in an arable soil. *Soil Biol. Biochem.* **97**, 188–198 (2016).
7. Trosvik, P. & de Muinck, E. J. Ecology of bacteria in the human gastrointestinal tract—identification of keystone and foundation taxa. *Microbiome* **3**, 44 (2015).
8. Xun, W. et al. Specialized metabolic functions of keystone taxa sustain soil microbiome stability. *Microbiome* **9**, 35 (2021).
9. LeBlanc, N. & Crouch, J. A. Prokaryotic taxa play keystone roles in the soil microbiome associated with woody perennial plants in the genus *Buxus*. *Ecol. Evol.* **9**, 11102–11111 (2019).
10. The Human Microbiome Project Consortium Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
11. Franzosa, E. A. et al. Identifying personal microbiomes using metagenomic codes. *Proc. Natl Acad. Sci. USA* **112**, E2930–E2938 (2015).
12. Vujkovic-Cvijin, I. et al. Host variables confound gut microbiota studies of human disease. *Nature* **587**, 448–454 (2020).
13. Stein, R. R. et al. Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota. *PLoS Comput. Biol.* **9**, e1003388 (2013).
14. Fisher, C. K. & Mehta, P. Identifying keystone species in the human gut microbiome from metagenomic timeseries using sparse linear regression. *PLoS ONE* **9**, e102451 (2014).
15. Bucci, V. et al. MDSINE: Microbial Dynamical Systems INference Engine for microbiome time-series analyses. *Genome Biol.* **17**, 121 (2016).
16. Berry, D. & Widder, S. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Front. Microbiol.* **5**, 219 (2014).
17. 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).
18. Garrett, W. S. et al. Enterobacteriaceae act in concert with the gut microbiota to induce spontaneous and maternally transmitted colitis. *Cell Host Microbe* **8**, 292–300 (2010).
19. Hajishengallis, G. et al. Low-abundance biofilm species orchestrates inflammatory periodontal disease through the commensal microbiota and complement. *Cell Host Microbe* **10**, 497–506 (2011).
20. Agler, M. T. et al. Microbial hub taxa link host and abiotic factors to plant microbiome variation. *PLoS Biol.* **14**, e1002352 (2016).
21. Banerjee, S., Schlaeppi, K. & van der Heijden, M. G. Reply to 'Can we predict microbial keystones?'. *Nat. Rev. Microbiol.* **17**, 194–194 (2019).
22. Röttjers, L. & Faust, K. Can we predict keystones? *Nat. Rev. Microbiol.* **17**, 193 (2019).
23. Amit, G. & Bashan, A. Top-down identification of keystone taxa in the microbiome. *Nat. Commun.* **14**, 3951 (2023).
24. Valls, A., Coll, M. & Christensen, V. Keystone species: toward an operational concept for marine biodiversity conservation. *Ecol. Monogr.* **85**, 29–47 (2015).
25. Tian, L. et al. Deciphering functional redundancy in the human microbiome. *Nat. Commun.* **11**, 6217 (2020).
26. Gouveia, C., Móréh, Á. & Jordán, F. Combining centrality indices: maximizing the predictability of keystone species in food webs. *Ecol. Indic.* **126**, 107617 (2021).
27. Kruse, R., Borgelt, C., Braune, C., Mostaghim, S. & Steinbrecher, M. *Computational Intelligence: A Methodological Introduction* (Springer, 2022).
28. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proc. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 770–778 (IEEE, 2016).
29. Michel-Mata, S., Wang, X.-W., Liu, Y.-Y. & Angulo, M. T. Predicting microbiome compositions from species assemblages through deep learning. *Imeta* **1**, e3 (2022).
30. Almeida-Neto, M., Guimaraes, P., Guimaraes, P. R. Jr, Loyola, R. D. & Ulrich, W. A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos* **117**, 1227–1239 (2008).
31. Friedman, J., Higgins, L. M. & Gore, J. Community structure follows simple assembly rules in microbial microcosms. *Nat. Ecol. Evol.* **1**, 0109 (2017).
32. Pasolli, E. et al. Accessible, curated metagenomic data through ExperimentHub. *Nat. Methods* **14**, 1023–1024 (2017).
33. Tudela, H., Claus, S. P. & Saleh, M. Next generation microbiome research: identification of keystone species in the metabolic regulation of host–gut microbiota interplay. *Front. Cell Dev. Biol.* **9**, 719072 (2021).
34. Alessandri, G., van Sinderen, D. & Ventura, M. The genus *Bifidobacterium*: from genomics to functionality of an important component of the mammalian gut microbiota. *Comput. Struct. Biotechnol. J.* **19**, 1472–1487 (2021).
35. Zhang, Z. et al. Spatial heterogeneity and co-occurrence patterns of human mucosal-associated intestinal microbiota. *ISME J.* **8**, 881–893 (2014).
36. Leylabadlo, H. E. et al. The critical role of *Faecalibacterium prausnitzii* in human health: an overview. *Microb. Pathog.* **149**, 104344 (2020).
37. Gotoh, A., Ojima, M. N. & Katayama, T. Minority species influences microbiota formation: the role of *Bifidobacterium* with extracellular glycosidases in bifidus flora formation in breastfed infant guts. *Microb. Biotechnol.* **12**, 259–264 (2019).
38. Bui, T. P. N. et al. *Intestinimonas*-like bacteria are important butyrate producers that utilize Nε-fructosyllysine and lysine in formula-fed infants and adults. *J. Funct. Foods* **70**, 103974 (2020).
39. Li, L. et al. Revealing proteome-level functional redundancy in the human gut microbiome using ultra-deep metaproteomics. *Nat. Commun.* **14**, 3428 (2023).
40. Gonzalez, A. et al. Qiita: rapid, web-enabled microbiome meta-analysis. *Nat. Methods* **15**, 796–798 (2018).
41. Friedman, J. & Alm, E. J. Inferring correlation networks from genomic survey data. *PLoS Comput. Biol.* **8**, e1002687 (2012).

Acknowledgements

Y.-Y.L. acknowledges funding support from the National Institutes of Health (R01AI141529, R01HD093761, RF1AG067744, UH3OD023268, U19AI095219 and U01HL089856) as well as the Office of the Assistant Secretary of Defense for Health Affairs, through the Traumatic Brain Injury and Psychological Health Research Program (Focused Program Award) under award no. W81XWH-22-S-TBIPH2, endorsed by the Department of Defense. X.-W.W. acknowledges funding support from the National Institutes of Health (K25HL166208). Z.S. acknowledges funding support from the National Institutes of Health (K99HL163519). M.T.A. acknowledges

the financial support provided by CONACyT grant No. A1-S-13909 and PAPIIT 104915.

Author contributions

Y.-Y.L. conceived and designed the project. X.-W.W. performed all the numerical calculations. X.-W.W. and Z.S. analysed real data. X.-W.W. and Y.-Y.L. wrote the manuscript. H.J., S.M.-M., M.T.A., L.D., X.H. and S.T.W. edited the manuscript. All authors approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41559-023-02250-2.

Correspondence and requests for materials should be addressed to Yang-Yu Liu.

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.

nature portfolio

Corresponding author(s): Yang-Yu Liu

Last updated by author(s): Sep 29, 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	R (version 4.1.2) package "curatedMetagenomicData" (version 3.2.3) was used to collected the gut microbiome dataset. Qiita was used to collected coral and soil microbiome. R pacakge "seqtime" (version 0.1.1) was used to generate simualted dataset.
Data analysis	Python (3.8.13) and Pytorch (version 1.7.1) were used to train cNODE2. R (version 4.1.2) and ggplot2 were used to visualize the results. R package "igraph" (1.4.2) was used to compute the betweenness. Correlation networks were constructed using "SparCC" implemented in CGLasso (https://github.com/huayingfang/CCLasso). Random phylogenetic tree was generated using R package "ape" (version 5.7-1).

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

The data and code used in this work are available at https://github.com/spxuw/DKI.

Human research participants

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

Reporting on sex and gender	Sex and gender analysis were not included in this study, as we are aiming to develop a deep learning framework to identify keystone species in microbiome, which is not depending on sex and gender.
Population characteristics	Environmental microbiome samples do not have characteristics. Gut microbiome samples were collected from healthy adults aged between 18 to 65. Oral microbiome were collected from adults with an average age of 30.27 (± 5.58) years old, including 1051 males and 1111 females.
Recruitment	Oral samples were self-collected by the volunteers during a physical examination. Other samples were collected from publicly available datasets.
Ethics oversight	Oral study was approved by the Institutional Review Board (IRB) of BGI-Shenzhen (Nos. BGI-IRB19121 and BGI- IRB17162) and the ethics committee of No.1 Affiliated People's Hospital of Kunming Medical University [(2017) Ethics review L No.14], China. Informed consent was obtained from each participant.

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

Ecological, evolutionary & environmental sciences study design

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

Study description	We proposed a Data-driven Keystone species Identification (DKI) framework based on deep learning .
Research sample	Human gut samples were collected from curatedMetagenomicData database. Human oral samples were collected from study: "Over 50,000 Metagenomically Assembled Draft Genomes for the Human Oral Microbiome Reveal New Taxa". Coral and soil samples were collected from Qiita (IDs: 10895 and 2104 In vitro soil data was provided by the original author.
Sampling strategy	For human gut and oral microbiome, we used the healthy samples.
Data collection	All the samples were from publicly available datasets.
Timing and spatial scale	All the samples were from publicly available datasets.
Data exclusions	We excluded the disease samples.
Reproducibility	Most of findings are reproducible across different datasets.
Randomization	We are aiming to develop a deep learning framework to identify keystone species in microbiome, so randomization is not relevant.
Blinding	The same as Randomlization, blinding is also not relevant.

Did the study involve field work? ☐ Yes ☒ No

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