Predicting microbiome compositions through deep learning

Sebastian Michel-Mata[1,2], Xu-Wen Wang[3], Yang-Yu Liu[3*], and Marco Tulio Angulo[4,**]

[1]Center for Applied Physics and Advanced Technology, Universidad Nacional Autónoma de México, Juriquilla 76230, México.

[2]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA

[3]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA.

[4]CONACyT - Institute of Mathematics, Universidad Nacional Autónoma de México, Juriquilla 76230, México.

Friday 18th June, 2021

Microbes can form complex communities that perform critical functions in maintaining the integrity of their environment[1,2] or the well-being of their hosts[3–6]. Successfully managing these microbial communities requires the ability to predict the community composition based on the species assemblage[7]. However, making such a prediction remains challenging because of our limited knowledge of the diverse physical[8], biochemical[9], and ecological[10,11] processes governing the microbial dynamics. To overcome this challenge, here we present a deep learning framework that automatically learns the map between species assemblages and community compositions from training data. First, we systematically validate our framework using synthetic data generated by classical population dynamics models. Then, we apply it to experimental data of both *in vitro* and *in vivo* communities, including ocean and soil microbial communities[12,13], *Drosophila melanogaster* gut microbiota[14], and human gut and oral microbiota[15]. Our results demonstrate how deep learning can enable us to understand better and potentially manage complex microbial communities.

Consider the pool $\Omega = \{1, \cdots, N\}$ of all microbial species that can inhabit an ecological habitat of interest, such as the human gut. A microbiome sample obtained from this habitat can be considered as a local community assembled from Ω with a particular *species assemblage*. The species assemblage of a sample is characterized by a binary vector $z \in \{0, 1\}^N$, where its i-th entry z_i satisfies $z_i = 1$ (or $z_i = 0$) if the i-th species is present (or absent) in this sample. Each sample is also associated with a *composition* vector $p \in \Delta^N$, where p_i is the relative abundance of the i-th species, and $\Delta^N = \left\{p \in \mathbb{R}^N_{\geq 0} | \sum_i p_i = 1\right\}$ is the probability simplex. Mathematically, our problem is to learn the map

$$\varphi : z \in \{0, 1\}^N \longmapsto p \in \Delta^N, \tag{1}$$

which assigns the composition vector $p = \varphi(z)$ based on the species assemblage z.

Knowing the above map would be instrumental in understanding the assembly rules of microbial communities[16]. However, learning this map is a fundamental challenge because the map depends on many physical, biochemical, and ecological processes influencing the dynamics of microbial communities. These processes include the spatial structure of the ecological habitat[8], the chemical gradients of available resources[9], and inter/intra-species interactions[11], to name a few. For large microbial communities like the human gut microbiota, our knowledge of all these processes is still rudimentary, hindering our ability to predict microbial compositions from species assemblages.

Methods

Here we show it is possible to predict the microbial composition from species assemblage without knowing the mechanistic details of the above processes. Our solution is a deep learning framework that learns the map φ directly from a dataset $\mathfrak{D}$ of S samples, each of which is associated with a pair (z, p), see Fig. 1.

Conditions for predicting compositions from species assemblages.

To ensure that the problem of learning φ from $\mathfrak{D}$ is mathematically well-posed, we make the following assumptions. *First*, we assume that the species pool in the habitat has universal dynamics[17], i.e., different local communities of this habitat can be described by the same population dynamics model with the same parameters. This assumption is necessary because, otherwise, the map φ simply does not exist, implying that predicting community compositions from species assemblages has to be done in a sample-specific manner, which is a daunting task. For *in vitro* communities, this assumption is satisfied if samples were collected from the same experiment or from multiple experiments but with very similar environmental conditions. For *in vivo* communities, empirical evidence indicates that the human gut and oral microbiota of healthy adults display very strong universal dynamics[17]. *Second*, we assume that the compositions of those collected samples represent steady states. This assumption is natural because for highly fluctuating microbial compositions the map φ is simply not well defined. We note that observational studies of host-associated microbial communities such as the human gut microbiota indicate that they remain close to stable steady states in the absence of drastic dietary change or antibiotic administrations[15,18,19]. *Finally*, we assume that for each species assemblage $z \in \{0, 1\}^N$ there is a unique steady-state composition $p \in \Delta^N$. In particular, this assumption requires that the true multi-stability does not exist for the species pool (or any subset of it) in this habitat. This assumption is required because, otherwise, the map φ is not injective and the prediction of community compositions becomes mathematically ill-defined. In practice, we expect that the above three assumptions can not be strictly satisfied. This means that any algorithm that predicts microbial compositions from species assemblages needs to be systematically tested to ensure its robustness against errors due to the violation of such approximations.

Limitations of traditional deep learning frameworks.

Under the above assumptions, a straightforward approach to learning the map φ from $\mathfrak{D}$ would be using deep neural networks[20,21] such as a feedforward Residual Network[22] (ResNet). As a very popular tool in image processing, ResNet is a cascade of $L \geq 1$ hidden layers where the state $h_\ell \in \mathbb{R}^N$ of the ℓ-th hidden layer satisfies $h_\ell = h_{\ell-1} + f_\theta(h_{\ell-1})$, $\ell = 1, \cdots, L$, for some parametrized function f_θ with parameters θ. These hidden layers are plugged to the input $h_0 = g_{\text{in}}(z)$ and the output $\hat{p} = g_{\text{out}}(h_L)$ layers, where g_{in} and g_{out} are some functions. Crucially, for our problem, any architecture must satisfy two restrictions: (1) vector $\hat{p}$ must be compositional (i.e., $\hat{p} \in \Delta^N$); and (2) the predicted relative abundance of any absent species must be identically zero (i.e., $z_i = 0$ should imply that $\hat{p}_i = 0$). Simultaneously satisfying both restrictions requires that the output layer is a normalization of the form $\hat{p}_i = z_i h_{L,i} / \sum_j z_j h_{L,j}$, and that f_θ is a non-negative function (because $h_L \geq 0$ is required to ensure the normalization is correct). We found that it is possible to train such a ResNet for predicting compositions in simple cases like small *in vitro* communities (Supplementary Note S2.1). But for large *in vivo* communities like the human gut microbiota, ResNet does not perform very well (Supplementary Fig. S1). This is likely due to the normalization of the output layer, which challenges the training of neural networks because of vanishing gradients[21]. The vanishing gradient problem is often solved by using other normalization layers such as the `softmax` or `sparsemax` layers[23]. However, we cannot use these layers in our problem because they do not satisfy the second restriction. We also note that ResNet becomes a universal approximation only in the limit $L \to \infty$, which again complicates the training[24].

A new deep learning framework.

To overcome the limitations of traditional deep learning frameworks based on neural networks (such as ResNet) in predicting microbial compositions from species assemblages, we developed cNODE (compositional Neural Ordinary Differential Equation), see Fig. 1b. The cNODE framework is based on the notion of Neural Ordinary Differential Equations, which can be interpreted as a continuous limit of ResNet where the hidden layers h's are replaced by an ordinary differential equation (ODE)[25]. In cNODE, an input species assemblage $z \in \{0, 1\}^N$ is first transformed into the initial condition $h(0) = z/\mathbb{1}^\intercal z \in \Delta^N$, where $\mathbb{1} = (1, \cdots, 1)^\intercal \in \mathbb{R}^N$ (left in Fig. 1b). This initial condition is used to solve the set of nonlinear ODEs

$$\frac{\mathrm{d}h(\tau)}{\mathrm{d}\tau} = h(\tau) \odot \big[f_\theta\big(h(\tau)\big) - \mathbb{1}\, h(\tau)^\intercal f_\theta\big(h(\tau)\big) \big]. \tag{2}$$

Here, the independent variable $\tau \geq 0$ represents a virtual "time". The expression $h \odot v$ is the entry-wise multiplication of the vectors $h, v \in \mathbb{R}^N$. The function $f_\theta : \Delta^N \to \mathbb{R}^N$ can be any continuous function parametrized by θ. For example, it can be the linear function $f_\theta(h) = \Theta h$ with parameter matrix $\Theta \in \mathbb{R}^{N \times N}$ (bottom in Fig. 1b), or a more complicated function represented by a feedforward deep neural network. Note that Eq. (2) can be considered as a very general form of the replicator equation —a canonical model in evolutionary game theory[26]— with f_θ representing the fitness function. By choosing a final integration "time" $\tau_c > 0$, Eq. (2) is numerically integrated to obtain the prediction $\hat{p} = h(\tau_c)$ that is the output of cNODE (right in Fig. 1b). We choose $\tau_c = 1$ without loss of generality, as τ in Eq. (2) can be rescaled by multiplying f_θ by a constant. The cNODE thus implements the map

$$\hat{\varphi}_\theta : z \in \{0, 1\}^N \longmapsto \hat{p} \in \Delta^N, \tag{3}$$

taking an input species assemblage z to the predicted composition $\hat{p}$ (see Supplementary Note S1 for implementation details). Note that Eq. (2) is key to cNODE because its architecture guarantees that the two restrictions imposed before are naturally satisfied. Namely, $\hat{p} \in \Delta^N$ because the conditions $h(0) \in \Delta^N$ and $\mathbb{1}^\intercal \mathrm{d}h/\mathrm{d}\tau = 0$ imply that $h(\tau) \in \Delta^N$ for all $\tau \geq 0$. Additionally, $z_i = 0$ implies $\hat{p}_i = 0$ because $h(0)$ and z have the same zero pattern, and the right-hand side of Eq. (2) is entry-wise multiplied by h.

We train cNODE by adjusting the parameters θ to approximate φ with $\hat{\varphi}_\theta$. To do this, we first choose a distance or dissimilarity function $d(p, q)$ to quantify how dissimilar are two compositions $p, q \in \Delta^N$. One can use any Minkowski distance or dissimilarity function. In the rest of this paper, we choose the Bray-Curtis[27] dissimilarity to present our results. Specifically, for a dataset $\mathfrak{D}_i \subseteq \mathfrak{D}$, we use the loss function

$$E(\mathfrak{D}_i) = \frac{1}{|\mathfrak{D}_i|} \sum_{(z,p) \in \mathfrak{D}_i} d(p, \hat{\varphi}_\theta(z)). \tag{4}$$

Second, we randomly split the dataset $\mathfrak{D}$ into training $\mathfrak{D}_1$ and test $\mathfrak{D}_2$ datasets. Next, we choose an adequate functional form for f_θ. In our experiments, we found that the linear function $f_\theta(h) = \Theta h$, $\Theta \in \mathbb{R}^{N \times N}$, provides accurate predictions for the composition of *in silico*, *in vitro*, and *in vivo* communities. Note that, despite f_θ is linear, the map $\hat{\varphi}_\theta$ is nonlinear because it is the solution of the nonlinear ODE of Eq. (2). Finally, we adjust the parameters θ by minimizing Eq. (4) on $\mathfrak{D}_1$ using a gradient-based meta-learning algorithm[28]. This learning algorithm enhances the generalizability of cNODE (Supplementary Note S1.2 and Supplementary Fig. S1). Once trained, we calculate

cNODE's test prediction error $E(\mathfrak{D}_2)$ that quantifies cNODE's performance in predicting the compositions of never-seen-before species assemblages. Test prediction errors could be due to a poor adjustment of the parameters (i.e., inaccurate prediction of the training set), low ability to generalize (i.e., inaccurate predictions of the test dataset), or violations of our three assumptions (universal dynamics, steady-state samples, no true multi-stability).

Fig. 1 demonstrates the application of cNODE to the fly gut microbiome samples collected in an experimental study[14]. In this study, germ-free flies (*Drosophila melanogaster*) were colonized with all possible combinations of five core species of fly gut bacteria, i.e., *Lactobacillus plantarum* (species-1), *Lactobacillus brevis* (species-2), *Acetobacter pasteurianus* (species-3), *Acetobacter tropicalis* (species-4), and *Acetobacter orientalis* (species-5). The dataset contains 41 replicates for the composition of each of the $2^N - 1 = 31$ local communities with different species assamblages. To apply cNODE, we aggregated all replicates and calculated their average composition, resulting in one "representative" sample per species assamblage (Supplementary Note S4). We also excluded the trivial samples with a single species, resulting in $S = 26$ samples. We trained cNODE by randomly choosing 21 of those samples (80%) as the training dataset (Fig. 1a). Once trained, cNODE accurately predicts microbial compositions in the test dataset of 5 species assemblages (Fig. 1c). For example, cNODE predicts that in the community with only species 3 and 4 present, species 3 will become nearly extinct, which agrees well with the experimental result (sample 26 in Fig. 1c).

Results

In silico validation of cNODE.

To systematically evaluate the performance cNODE, we generated *in silico* data for pools of $N = 100$ species with population dynamics given by the classical Generalized Lotka-Volterra (GLV) model[29]

$$\frac{\mathrm{d}x_i(t)}{\mathrm{d}t} = x_i(t)\left[r_i + \sum_{j=1}^{N} a_{ij}\, x_j(t)\right], \quad i = 1, \cdots, N. \tag{5}$$

Above, $x_i(t)$ denotes the abundance of the i-th species at time $t \geq 0$. The GLV model has as parameters the interaction matrix $A = (a_{ij}) \in \mathbb{R}^{N\times N}$, and the intrinsic growth-rate vector $r = (r_i) \in \mathbb{R}^N$. Here, a_{ij} denotes the inter- (if $j \neq i$) or intra- (if $j = i$) species *interaction strength* of species j to the per-capita growth rate of species i. The parameter r_i is the intrinsic growth rate of species i. Recall that the interaction matrix A determines the ecological network

$\mathcal{G}(A)$ underlying the species pool. Namely, this network has one node per species and edges $(j \to i) \in \mathcal{G}(A)$ if $a_{ij} \neq 0$. The *connectivity* $C \in [0, 1]$ of this network is the proportion of edges it has compared to the N^2 edges in a complete network. Despite its simplicity, the GLV model has been successfully applied to describe the population dynamics of microbial communities in diverse environments, from the soil[30] and lakes[31] to the human gut[32,33]. To validate cNODE, we generated synthetic microbiome samples as steady-state compositions of GLV models with random parameters by choosing $a_{ij} \sim \text{Bernoulli}(C)\text{Normal}(0, \sigma)$ if $i \neq j$, $a_{ii} = -1$, and $r_i \sim \text{Uniform}[0, 1]$, for different values of connectivity C and characteristic inter-species interaction strength $\sigma > 0$ (Supplementary Note S3).

Figure 2a shows the prediction error in synthetic training and test datasets, each of which has N samples generated by the GLV model of N species, with $\sigma = 0.5$ and different values of C. The prediction error in the training set, $E(\mathfrak{D}_1)$, keeps decreasing with the increasing number of training epochs, especially for high C values (as shown in dashed and dotted cyan lines in Fig. 2a). Interestingly, the prediction error in the test dataset, $E(\mathfrak{D}_2)$, reaches a plateau after enough number of training epochs regardless of the C values (see solid, dashed and dotted yellows lines in Fig. 2a), which is a clear evidence of an adequate training of cNODE with low overfitting. Note that the plateau of $E(\mathcal{D})$ increases with C. We confirm this result in datasets with different sizes of the training dataset (Fig. 2b). Moreover, we found that the plateau increases with increasing characteristic interaction strength σ (Fig. 2c). Fortunately, the increase of $E(\mathfrak{D}_2)$ (due to increasing C or σ) can be compensated by increasing the sample size of the training set $\mathfrak{D}_1$. Indeed, as shown in Fig. 2b,c, $E(\mathfrak{D}_2)$ decreases with increasing $|\mathfrak{D}_1|/N$.

To systematically evaluate the robustness of cNODE against violation of its three key assumptions, we performed three types of validations. In the first validation, we generated datasets that violate the assumption of universal dynamics. For this, given a "base" GLV model with parameters (A, r), we consider two forms of universality loss (Supplementary Note S3). *First*, samples are generated using a GLV with the same ecological network but with those non-zero interaction strengths a_{ij} replaced by $a_{ij} + \text{Normal}(0, \eta)$, where $\eta > 0$ characterizes the changes in the typical interaction strength. *Second*, samples are generated using a GLV with slightly different ecological networks obtained by randomly rewiring a proportion $\rho \in [0, 1]$ of their edges. We find that cNODE is robust to both forms of universality loss as its asymptotic prediction error changes continuously, maintaining a reasonably low prediction error up to $\eta = 0.4$ and $\rho = 0.1$ (Fig. 2d and Supplementary Fig. S2).

In the second validation, we evaluated the robustness of cNODE against measurement noises in

the relative abundance of species. For this, for each sample, we first change the relative abundance of the i-th species from p_i to $\max\{0, p_i + \text{Normal}(0, \varepsilon)\}$, where $\varepsilon \geq 0$ characterizes the measurement noise intensity. Then, we normalize the vector p to ensure it is still compositional, i.e., $p \in \Delta^N$. Due to the measurement noise, some species that were absent may be measured as present, and vice-versa. In this case, we find that cNODE performs adequately up to $\varepsilon = 0.025$ (Fig. 2f)

In the third validation, we generated datasets with true multi-stability by simulating a population dynamics model with nonlinear functional responses (Supplementary Notes S3). For each species collection, these functional responses generate two interior equilibria in different "regimes": one regime with low biomass, and the other regime with high biomass. We then train cNODE with datasets obtained by choosing a fraction $(1 - \mu)$ of samples from the first regime, and the rest from the second regime. We find that cNODE is robust enough to provide reasonable predictions up to $\mu = 0.2$ (Fig. 2d).

Evaluation of cNODE using real data.

We evaluated cNODE using six microbiome datasets of different habitats (Supplementary Note S4). The first dataset consists of $S = 275$ samples[34] of the ocean microbiome at phylum taxonomic level, resulting in $N = 73$ different taxa. The second dataset consists of $S = 26$ *in vivo* samples of *Drosophila melanogaster* gut microbiota of $N = 5$ species[14], as described in Fig. 1. The third dataset has $S = 93$ samples of *in vitro* communities of $N = 8$ soil bacterial species[12]. The fourth dataset contains $S = 113$ samples of the Central Park soil microbiome[13] at the phylum level ($N = 36$ phyla). The fifth dataset contains $S = 150$ samples of the human oral microbiome from the Human Microbiome Project[15] (HMP) at the genus level ($N = 73$ genera). The final dataset has $S = 106$ samples of the human gut microbiome from HMP at the genus level ($N = 58$ genera).

To evaluate cNODE, we performed the leave-one-out cross-validation on each dataset. The median test prediction errors were 0.06, 0.066, 0.079, 0.107, 0.211 and 0.242 for the six datasets, respectively (Fig. 3a). To understand the meaning of these errors, for each dataset we inspected five pairs $(p, \hat{p})$ corresponding to the observed and the predicted compositions of five samples. We chose the five samples based on their test prediction error. Specifically, we selected those samples with the minimal error, close to the first quartile, closer to the median, closer to the third quartile, and with the maximal error (columns in Fig. 3b-g, from left to right). We found that samples with errors below the third quartile provide acceptable predictions (left three columns in Fig. 3b-g), while samples with errors close to the third quartile or with the maximal error do demonstrate salient

differences between the observed and predicted compositions (right two columns in Fig. 3b-g). Note that in the sample with largest error of the human gut dataset (Fig. 3g, rightmost column), the observed composition is dominated by *Provotella* (pink) while the predicted sample is dominated by *Bacteroides* (blue). This drastic difference is likely due to different dietary patterns[35].

Discussion

cNODE is a deep learning framework to predict microbial compositions from species assemblages. We validated its performance using *in silico*, *in vitro*, and *in vivo* microbial communities. Several methods have been developed for predicting species abundances in microbial communities by modeling their population dynamics[12,32,36,37], but these methods typically require high-quality time-series data of absolute abundances that are difficult to obtain for large *in vivo* microbial communities. cNODE circumvents the need of absolute abundances or time-series data. The price to pay is that the trained function f_θ cannot be directly interpreted because the lack of identifiability inherent to compositional data[38,39]. We also note a recent statistical method to predict coexistence of ecological communities[40], but this method also requires absolute abundance measurements. cNODE can outperform this statistical method despite using only relative abundances (Supplementary Note S6). See also Supplementary Note S5 for a discussion of how our framework compares to other related works.

Deep learning techniques are actively applied to microbiome research[41–49] such as for classifying samples that shifted to a diseased state[50], predicting infection complications in immunocompromised patients[51], or predicting the temporal or spatial evolution of certain species collection[52,53]. However, to the best of our knowledge, the potential of deep learning for predicting the effect of changing species collection was not explored nor validated before. Our proposed framework based the notion of neural ODE is a baseline which could be improved by incorporating additional information. For example, incorporating available environmental information such as pH, temperature, age, BMI and diet of the host, could enhance the prediction accuracy. This would help to predict the species present in different environments. Adding "hidden variables" such as the unmeasured total biomass or unmeasured resources to our ODE will enhance the expressivity of the cNODE[54,55], but this may result in a more challenging training. Finally, if available, knowledge of the genetic similarity between species can be leveraged into the loss function by using the phylogenetic Wasserstein distance[56] that provides a well-defined gradientt[57].

Our framework does have limitations. For example, it cannot accurately predict the abundance

of taxa that have never been observed in the training dataset. Also, a limitation of our current architecture is that it assumes that true multistability does not exist —namely, a community with a given species assemblage permits only one stable steady-state, where each species in the collection has a positive abundance. For complex microbial communities such as the human gut microbiota, the highly personalized species collections makes it very difficult to decide if true multistability exists or not. Our framework could be extended to handle multi-stability by predicting a probability density function for the abundance of each species. In such a case, true multistability would correspond to predicting a multimodal density function.

We anticipate that a useful application of our framework is to predict if by adding some species collection to a local community we can bring the abundance of target species below the practical extinction threshold. Thus, given a local community containing the target (and potentially pathogenic) species, we could use a greedy optimization algorithm to identify a minimal collection of species to add such that our architecture predicts that they will decolonize the target species.

References

1. East, R. Soil Science comes to life. *Nature* **501,** S18 (2013).
2. Busby, P. E., Soman, C., Wagner, M. R., Friesen, M. L., Kremer, J., Bennett, A., Morsy, M., Eisen, J. A., Leach, J. E. & Dangl, J. L. Research priorities for harnessing plant microbiomes in sustainable agriculture. *PLoS Biology* **15,** e2001793 (2017).
3. Thursby, E. & Juge, N. Introduction to the human gut microbiota. *Biochemical Journal* **474,** 1823–1836 (2017).
4. Chittim, C. L., Irwin, S. M. & Balskus, E. P. Deciphering Human Gut Microbiota–Nutrient Interactions: A Role for Biochemistry. *Biochemistry* **57,** 2567–2577 (2018).
5. Cryan, J. F. & Dinan, T. G. Mind-altering microorganisms: the impact of the gut microbiota on brain and behaviour. *Nature Reviews NeuroScience* **13,** 701 (2012).
6. Arrieta, M.-C., Stiemsma, L. T., Amenyogbe, N., Brown, E. M. & Finlay, B. The intestinal microbiome in early life: health and disease. *Frontiers in Immunology* **5,** 427 (2014).
7. Widder, S., Allen, R. J., Pfeiffer, T., Curtis, T. P., Wiuf, C., Sloan, W. T., Cordero, O. X., Brown, S. P., Momeni, B., Shou, W., *et al.* Challenges in microbial ecology: building predictive understanding of community function and dynamics. *The ISME journal* **10,** 2557 (2016).
8. Tropini, C., Earle, K. A., Huang, K. C. & Sonnenburg, J. L. The gut microbiome: connecting spatial organization to function. *Cell host & microbe* **21,** 433–442 (2017).
9. Quinn, R. A., Comstock, W., Zhang, T., Morton, J. T., da Silva, R., Tran, A., Aksenov, A., Nothias, L.-F., Wangpraseurt, D., Melnik, A. V., *et al.* Niche partitioning of a pathogenic microbiome driven by chemical gradients. *Science advances* **4,** eaau1908 (2018).
10. Kuramitsu, H. K., He, X., Lux, R., Anderson, M. H. & Shi, W. Interspecies interactions within oral microbial communities. *Microbiol. Mol. Biol. Rev.* **71,** 653–670 (2007).
11. Coyte, K. Z., Schluter, J. & Foster, K. R. The ecology of the microbiome: networks, competition, and stability. *Science* **350,** 663–666 (2015).
12. Friedman, J., Higgins, L. M. & Gore, J. Community structure follows simple assembly rules in microbial microcosms. *Nature Ecology & Evolution* **1,** 0109 (2017).
13. Ramirez, K. S., Leff, J. W., Barberán, A., Bates, S. T., Betley, J., Crowther, T. W., Kelly, E. F., Oldfield, E. E., Shaw, E. A., Steenbock, C., *et al.* Biogeographic patterns in below-ground diversity in New York City's Central Park are similar to those observed globally. *Proceedings of the royal society B: biological Sciences* **281,** 20141988 (2014).
14. Gould, A. L., Zhang, V., Lamberti, L., Jones, E. W., Obadia, B., Korasidis, N., Gavryushkin, A., Carlson, J. M., Beerenwinkel, N. & Ludington, W. B. Microbiome interactions shape host fitness. *Proceedings of the National Academy of Sciences* **115,** E11951–E11960 (2018).
15. Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J. H., Chinwalla, A. T., Creasy, H. H., Earl, A. M., FitzGerald, M. G., Fulton, R. S., *et al.* Structure, function and diversity of the healthy human microbiome. *Nature* **486,** 207 (2012).
16. Suez, J., Zmora, N., Segal, E. & Elinav, E. The pros, cons, and many unknowns of probiotics. *Nature medicine* **25,** 716–729 (2019).
17. Bashan, A., Gibson, T. E., Friedman, J., Carey, V. J., Weiss, S. T., Hohmann, E. L. & Liu, Y.-Y. Universality of human microbial dynamics. *Nature* **534,** 259 (2016).

18. Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K. & Knight, R. Diversity, stability and resilience of the human gut microbiota. *Nature* **489,** 220 (2012).

19. Gibbons, S. M., Kearney, S. M., Smillie, C. S. & Alm, E. J. Two dynamic regimes in the human gut microbiome. *PLoS Computational Biology* **13,** 1–20 (Feb. 2017).

20. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521,** 436 (2015).

21. Goodfellow, I., Bengio, Y. & Courville, A. *Deep learning* (MIT press, 2016).

22. He, K., Zhang, X., Ren, S. & Sun, J. *Deep residual learning for image recognition* in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), 770–778.

23. Martins, A. & Astudillo, R. *From softmax to sparsemax: A sparse model of attention and multi-label classification* in *International Conference on Machine Learning* (2016), 1614–1623.

24. Lin, H. & Jegelka, S. *Resnet with one-neuron hidden layers is a universal approximator* in *Advances in Neural Information Processing Systems* (2018), 6169–6178.

25. Chen, T. Q., Rubanova, Y., Bettencourt, J. & Duvenaud, D. K. *Neural ordinary differential equations* in *Advances in neural information processing systems* (2018), 6571–6583.

26. Josef Hofbauer and, K. S. *The Theory of Evolution and Dynamical Systems: Mathematical Aspects of Selection* (London Mathematical Society Students Text, 1988).

27. Legendre, P. & Legendre, L. Numerical ecology. 3rd English ed. *Developments in environmental modelling* **24** (2012).

28. Nichol, A., Achiam, J. & Schulman, J. On first-order meta-learning algorithms. *arXiv preprint arXiv:1803.02999* (2018).

29. Case, T. J. *An Illustrated Guide to Theoretical Ecology* (Oxford Univ. Press, Oxford, 2000).

30. Moore, J. C., de Ruiter, P. C., Hunt, H. W., Coleman, D. C. & Freckman, D. W. Microcosms and soil ecology: critical linkages between fields studies and modelling food webs. *Ecology* **77,** 694–705 (1996).

31. Dam, P., Fonseca, L. L., Konstantinidis, K. T. & Voit, E. O. Dynamic models of the complex microbial metapopulation of lake mendota. *NPJ systems biology and applications* **2,** 1–7 (2016).

32. Stein, R. R., Bucci, V., Toussaint, N. C., Buffie, C. G., Rätsch, G., Pamer, E. G., Sander, C. & Xavier, J. B. Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota. *PLoS Computational Biology* **9** (2013).

33. Buffie, C. G., Bucci, V., Stein, R. R., McKenney, P. T., Ling, L., Gobourne, A., No, D., Liu, H., Kinnebrew, M., Viale, A., *et al.* Precision microbiome reconstitution restores bile acid mediated resistance to Clostridium difficile. *Nature* **517,** 205–208 (2015).

34. Laverock, B., Smith, C. J., Tait, K., Osborn, A. M., Widdicombe, S. & Gilbert, J. A. Bioturbating shrimp alter the structure and diversity of bacterial communities in coastal marine sediments. *The ISME journal* **4,** 1531–1544 (2010).

35. Wu, G. D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.-Y., Keilbaugh, S. A., Bewtra, M., Knights, D., Walters, W. A., Knight, R., *et al.* Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334,** 105–108 (2011).

36. Mounier, J., Monnet, C., Vallaeys, T., Arditi, R., Sarthou, A.-S., Hélias, A. & Irlinger, F. Microbial interactions within a cheese microbial community. *Applied and Environmental Microbiology* **74,** 172–181 (2008).

37. Venturelli, O. S., Carr, A. V., Fisher, G., Hsu, R. H., Lau, R., Bowen, B. P., Hromada, S., Northen, T. & Arkin, A. P. Deciphering microbial interactions in synthetic human gut microbiome communities. *Molecular Systems Biology* **14,** e8157 (2018).

38. Aitchison, J. Principles of compositional data analysis. *Lecture Notes-Monograph Series,* 73–81 (1994).

39. Cao, H.-T., Gibson, T. E., Bashan, A. & Liu, Y.-Y. Inferring human microbial dynamics from temporal metagenomics data: Pitfalls and lessons. *BioEssays* **39,** 1600188 (2017).

40. Maynard, D. S., Miller, Z. R. & Allesina, S. Predicting coexistence in experimental ecological communities. *bioRxiv,* 598326 (2019).

41. Reiman, D., Metwally, A. & Dai, Y. *Using convolutional neural networks to explore the microbiome* in *Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE* (2017), 4269–4272.

42. García-Jiménez, B., Muñoz, J., Cabello, S., Medina, J. & Wilkinson, M. D. Predicting microbiomes through a deep latent space. *Bioinformatics* (Nov. 2020).

43. Metwally, A. A., Yu, P. S., Reiman, D., Dai, Y., Finn, P. W. & Perkins, D. L. Utilizing longitudinal microbiome taxonomic profiles to predict food allergy via Long Short-Term Memory networks. *PLoS Computational Biology* **15,** e1006693 (2019).

44. Sharma, D., Paterson, A. D. & Xu, W. TaxoNN: ensemble of neural networks on stratified microbiome data for disease prediction. *Bioinformatics* **36,** 4544–4550 (2020).

45. Galkin, F., Mamoshina, P., Aliper, A., Putin, E., Moskalev, V., Gladyshev, V. N. & Zhavoronkov, A. Human gut microbiome aging clock based on taxonomic profiling and deep learning. *iScience* **23,** 101199 (2020).

46. Asgari, E., Garakani, K., McHardy, A. C. & Mofrad, M. R. MicroPheno: predicting environments and host phenotypes from 16S rRNA gene sequencing using a k-mer based representation of shallow sub-samples. *Bioinformatics* **34,** i32–i42 (2018).

47. Zhu, Q., Jiang, X., Zhu, Q., Pan, M. & He, T. Graph Embedding Deep Learning Guides Microbial Biomarkers' Identification. *Frontiers in genetics* **10,** 1182 (2019).

48. Le, V., Quinn, T. P., Tran, T. & Venkatesh, S. Deep in the Bowel: highly interpretable neural encoder-decoder networks predict gut metabolites from gut microbiome. *BMC genomics* **21,** 1–15 (2020).

49. LaPierre, N., Ju, C. J.-T., Zhou, G. & Wang, W. MetaPheno: A critical evaluation of deep learning and machine learning in metagenome-based disease prediction. *Methods* **166,** 74–82 (2019).

50. Yazdani, M., Taylor, B. C., Debelius, J. W., Li, W., Knight, R. & Smarr, L. *Using machine learning to identify major shifts in human gut microbiome protein family abundance in disease.* in *BigData* (2016), 1272–1280.

51. Espinoza, J. L. Machine learning for tackling microbiota data and infection complications in immunocompromised patients with cancer. *Journal of Internal Medicine* (2018).

52. Larsen, P. E., Field, D. & Gilbert, J. A. Predicting bacterial community assemblages using an artificial neural network approach. *Nature Methods* **9,** 621 (2012).

53. Zhou, G., Jiang, J.-Y., Ju, C. J.-T. & Wang, W. Prediction of microbial communities for urban metagenomics using neural network approach. *Human Genomics* **13,** 47 (2019).

54. Zhang, H., Gao, X., Unterman, J. & Arodz, T. Approximation capabilities of neural ordinary differential equations. *arXiv preprint arXiv:1907.12998* (2019).

55. Dupont, E., Doucet, A. & Teh, Y. W. Augmented neural odes. *arXiv preprint arXiv:1904.01681* (2019).

56. Evans, S. N. & Matsen, F. A. The phylogenetic Kantorovich–Rubinstein metric for environmental sequence samples. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **74,** 569–592 (2012).

57. Peyré, G., Cuturi, M., *et al.* Computational optimal transport. *Foundations and Trends® in Machine Learning* **11,** 355–607 (2019).

58. Innes, M. Flux: Elegant machine learning with julia. *Journal of Open Source Software* **3,** 602 (2018).

59. Glorot, X. & Bengio, Y. *Understanding the difficulty of training deep feedforward neural networks* in *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (2010), 249–256.

60. Rackauckas, C. & Nie, Q. Differentialequations.jl – a performant and feature-rich ecosystem for solving differential equations in julia. *Journal of Open Research Software* **5** (2017).

61. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

62. Ruder, S. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747* (2016).

63. FluxML. *MetaLearning.jl* <https://github.com/FluxML/model-zoo/blob/master/contrib/meta-learning/MetaLearning.jl> (2019).

64. Jost, C. & Ellner, S. P. Testing for predator dependence in predator-prey dynamics: a non-parametric approach. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **267,** 1611–1620 (2000).

65. Xiao, Y., Angulo, M. T., Friedman, J., Waldor, M. K., Weiss, S. T. & Liu, Y.-Y. Mapping the ecological networks of microbial communities. *Nature Communications* **8,** 2042 (2017).

66. Tung, J., Barreiro, L. B., Burns, M. B., Grenier, J.-C., Lynch, J., Grieneisen, L. E., Altmann, J., Alberts, S. C., Blekhman, R. & Archie, E. A. Social networks predict gut microbiome composition in wild baboons. *eLife* **4,** e05224 (2015).

Acknowledgments M.T.A. gratefully acknowledges the financial support from CONACyT project A1-S-13909, México. Y.-Y.L. acknowledges the funding support from National Institutes of Health (R01AI141529, R01HD093761, RF1AG067744 , UH3OD023268, U19AI095219, and U01HL089856).

Competing financial interests The authors declare no competing financial interests.

Author contributions M.T.A. and Y.-Y. L. conceived and designed the project. S.M.M. did the numerical analysis. S.M.M. and X.-W.W. performed the real data analysis. All authors analyzed the results. M.T.A. and Y.-Y.L. wrote the manuscript. S.M.M. and X.-W.W. edited the manuscript.

Data accessibility The data and code used in this work are available from the corresponding authors upon reasonable request.

Correspondence Correspondence and requests for materials should be addressed to M.T.A. (email: mangulo@im.unam.mx) or Y.-Y. L. (email: yyl@channing.harvard.edu).

Figures

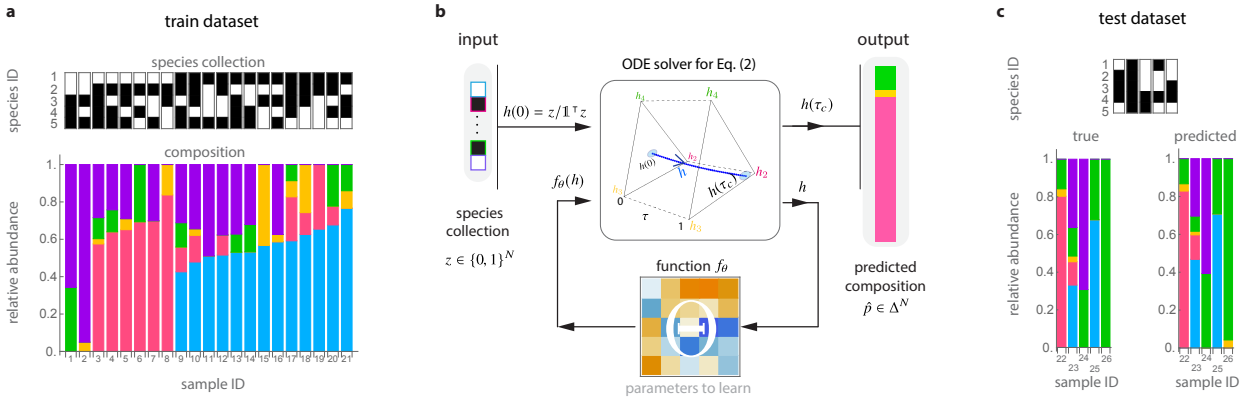


Figure 1: **A deep learning framework to predict microbiome compositions from species assemblages.** We illustrate this framework using experimental data from a pool of $N = 5$ bacterial species in *Drosophila melangaster* gut microbiota[14]: *Lactobacillus plantarum* (blue), *Lactobacillus brevis* (pink), *Acetobacter pasteurianus* (yellow), *Acetobacter tropicalis* (green), and *Acetobacter orientalis* (purple). **a.** We randomly split this dataset into training and test datasets: $\mathfrak{D}_1$ and $\mathfrak{D}_2$, which contain 80% and 20% of the samples, respectively. Each dataset contains pairs (z, p) of the species assamblage $z \in \{0,1\}^N$ (top) and composition $p \in \Delta^N$ (bottom) from each sample. **b.** To predict compositions from species assamblages, our cNODE framework consists of a solver for the ODE shown in Eq. (2), together with a chosen parametrized function f_θ. During training, the parameters θ are adjusted to learn to predict the composition $\hat{p} \in \Delta^N$ of the species assamblage $z \in \{0,1\}^N$ in $\mathfrak{D}_1$. **c.** After training, performance is evaluated by predicting the composition of never-seen-before species assamblages in the test dataset $\mathfrak{D}_2$.

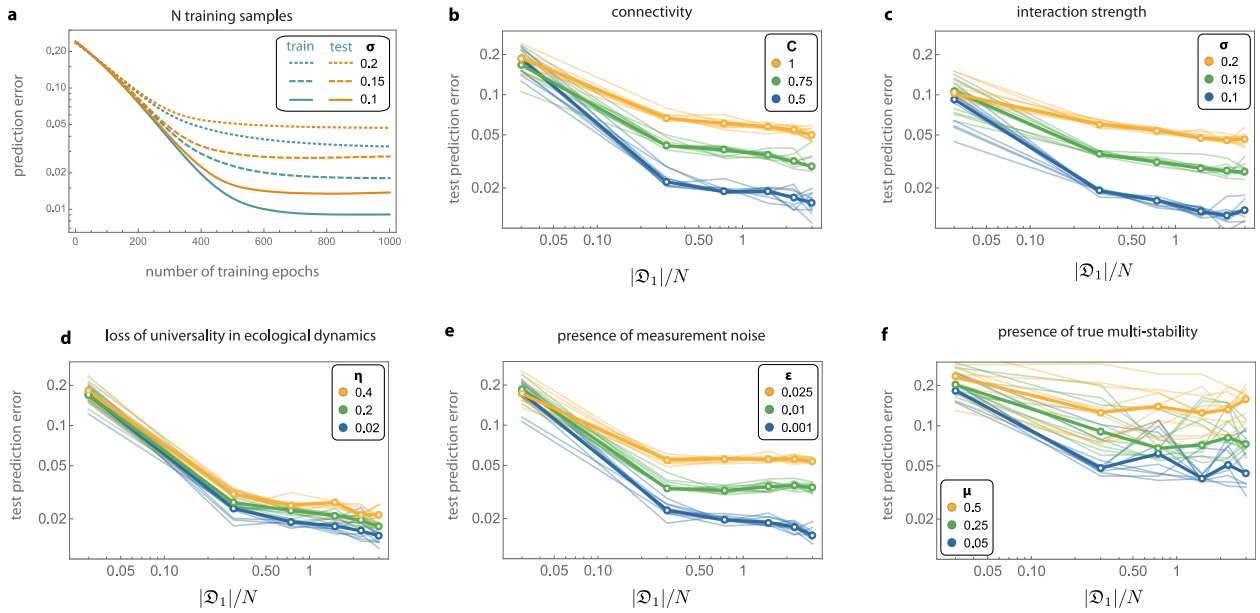


Figure 2: ***In silico* validation of cNODE using synthetic datasets.** Results are for synthetic communities of $N = 100$ species generated by the with Generalized Lotka-Volterra model (panels **a**-**e**) or a population dynamics model with nonlinear functional responses (panel **f**). **a.** Training cNODE with N samples obtained from GLV models with connectivity $C = 0.1$ (solid), $C = 0.15$ (dashed), $C = 0.2$ (dotted). **b.** Performance of cNODE for GLV datasets with $C = 0.5$ and different interaction strengths σ. **c.** Performance of cNODE for GLV datasets with $\sigma = 0.5$ and different connectivity C. **d.** Performance of cNODE for GLV datasets with non-universal dynamics, quantified by the value of η. For all datasets, $\sigma = 0.1$ and $C = 0.5$. **e.** Performance of cNODE for GLV datasets with measurement errors quantified by ε. For all datasets, $\sigma = 0.1$ and $C = 0.5$. **f.** Performance of cNODE for synthetic datasets with multiple interior equilibria, quantified by the probability $\mu \in [0, 1]$ of finding multiple equilibria. For all datasets, $C = 0.5, \sigma = 0.1$. In panels **b**-**f**, thin lines represent the prediction errors for ten validations of training cNODE with a different dataset. Mean errors are shown in thick lines.

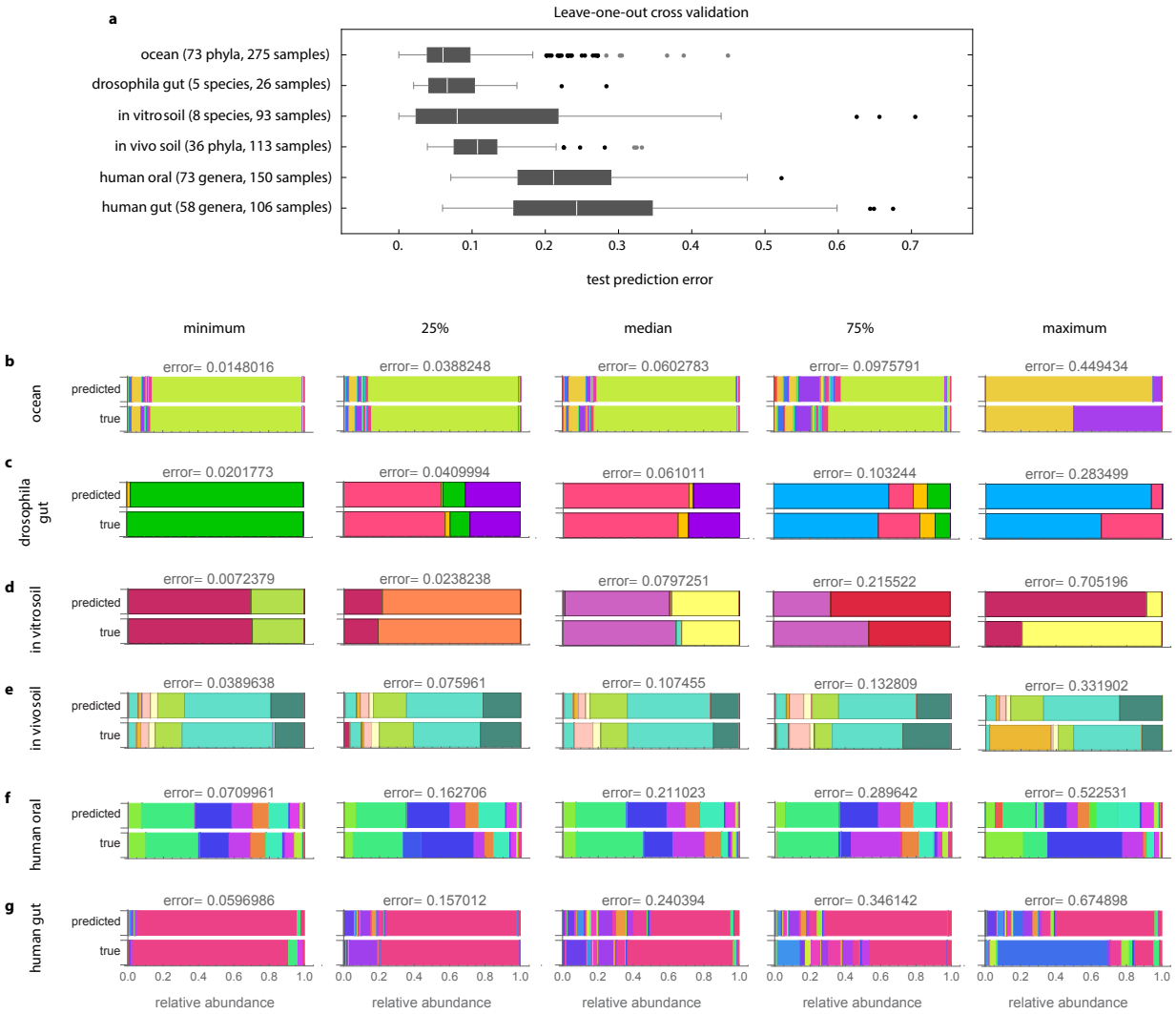


Figure 3: **Predicting the composition of real microbiomes. a.** Boxplots with the prediction error obtained from a leave-one-out crossvalidation of each dataset. **b-g**: For each dataset, we show true and predicted samples corresponding to the minimal prediction error, closer to the first quartile, median, closer to the third quartile, maximum prediction error (including outliers). Note all shown in panels **b-g** predictions are out-of-sample predictions.

Supplementary Notes

Contents

S1 Implementation of the compositional Neural Ordinary Differential Equation **1**
S1.1 Flux implementation of cNODE. 1
S1.2 Training cNODE. 1

S2 Comparison of two deep learning frameworks in predicting microbiome compositions **2**
S2.1 A ResNet architecture for predicting microbiome compositions. 2
S2.2 Performance comparison. 2

S3 Population dynamics for the *in silico* validations. **4**
S3.1 Generating datasets with universal dynamics. 4
S3.2 Generating datasets with non-universal dynamics. 4
S3.3 Adding measurement noise to a dataset. 5
S3.4 Generating datasets with multi-stability. 5

S4 Description of the experimental datasets. **6**
S4.1 In-vivo drosophila core gut microbiome. 6
S4.2 In-vitro soil community. 6
S4.3 -vivo soil microbiome. 6
S4.4 Human gut microbiome. 6

S5 Related work. **10**

S6 Comparing cNODE with Maynard et al.'s method. **12**

S1. Implementation of the compositional Neural Ordinary Differential Equation

S1.1 Flux implementation of cNODE.

We implemented cNODE using `Flux`[58], a library for machine learning in the Julia programming language with support for Neural Ordinary Differential Equations[25]. A complete implementation of cNODE is given in the file `cNODE.jl`.

Our implementation is based on a structure called `FitnessLayer` that contains cNODE's parameter $\Theta \in \mathbb{R}^{N \times N}$. This parameter is initialized using the Xavier's method[59]. When evaluated in a composition p, the `FitnessLayer` computes first $f_\theta(p) = \Theta p$. More complex functions f_θ can be easily incorporated in the code. Finally, the structure uses f_θ to calculate the right-hand side of the ODE in Eq. (2).

To predict the composition $\hat{p} \in \Delta^N$ associated with a species collection $z \in \{0, 1\}^N$, cNODE numerically solves such an the ODE in Eq.(2). In `Flux`, this is automated by the function `neural_ode`, which constructs the cNODE by building a NODE with the dynamics specified by the `FitnessLayer`.

The dynamics is numerically integrated over the interval $\tau \in [0, \tau_c]$ using the `Tsit5` method[60], which is the default integration method for nonstiff ODEs in Julia. We choose $\tau_c = 1$ without loss of generality, as τ in Eq. (2) can be rescaled by multiplying f_θ by a constant. After integration, the final value at time $\tau = 1.0$ is returned as the prediction of cNODE.

The loss function calculates the average Bray-Curtis dissimilarity between there true composition $p \in \Delta^N$ and the prediction $\hat{p} \in \Delta^N$ generated by cNODE.

S1.2 Training cNODE.

To train cNODE, we adjust the parameters θ to minimize the loss over the training set. We experimented with two training algorithms:

1. **The ADAM algorithm**[61]. ADAM is a widely used gradient-based stochastic optimization algorithm that often compares favorably to other gradient optimization algorithms[62]. We refer to [61, 62] for additional details.

2 **ADAM plus a first-order gradient-based meta-learning algorithm based on Reptile**[28]. In the meta-learning framework, lets consider a set $\mathcal{T}$ of different *tasks* that the network needs to perform, such as learning to predict in different datasets. Reptile works by sampling some task $\tau \in \mathcal{T}$, training on it to obtain the weights $\tilde{\theta}$, and then updating the initial weights θ towards $\tilde{\theta}$.

   To train cNODE, we used ADAM plus Reptile training[28]. More precisely, we defined each task as a random partition of the training set in mini-batches. In this way, training enhances cNODE's generalizing ability from the predictions regardless of any specific partition or any

sequence order of mini-batches. The algorithm we used is described below:

Algorithm 1: ADAM + Reptile for cNODE

Initialize θ, the vector of initial parameters
for *epoch = 1,2...* **do**
  Random partition of training set ρ
  Compute $\tilde{\theta} = \text{ADAM}_{\rho}^{k}(\theta)$, denoting k steps
  Update $\theta \leftarrow \theta + \text{ADAM}(\tilde{\theta} - \theta)$
end

Our implementation in the Julia language of the above algorithm is based on the model zoo of the library Flux[63].

S2. Comparison of two deep learning frameworks in predicting microbiome compositions

S2.1 A ResNet architecture for predicting microbiome compositions.

We tested the performance the classical ResNet architecture[25] for predicting microbiome compositions. More precisely, we used the input layer

$$h_0 = W_0 z + b_0,$$

where $W_0 \in \mathbb{R}^{N\times N}$ and $b_0 \in \mathbb{R}^N$ are parameters to adjust. We used $L = 3$ hidden layers of the form:

$$h_\ell = h_{\ell-1} + \text{ReLU}(W_{\ell-1}h_{\ell-1} + b_{\ell-1}), \quad \ell = 1, \cdots, L,$$

where $W_\ell \in \mathbb{R}^{N\times N}$ and $b_\ell \in \mathbb{R}^N$ are parameters to adjust. Finally, the output layer takes the form

$$\hat{p} = \left(\frac{h_{L,1}z_1}{\sum_j h_{L,j}z_j}, \cdots, \frac{h_{L,N}z_N}{\sum_j h_{L,N}z_N} \right).$$

A `Flux` implementation of this ResNet can be found in the file `ResNet.jl`.
We trained this ResNet architecture using the two methods described in S1.2.

S2.2 Performance comparison.

We compared the performance of cNODE and ResNet architecture for predicting the composition of four out of six real microbiome datasets described in Supplementary Note S4. These datasets contain a small set of species (between 5 and 58) and samples (between 26 and 113), allowing us to make computationally expensive leave-one-out cross validation analysis. Here, we also compared the effect of training both architectures with ADAM, and with ADAM plus the Reptile like metalearning algorithms.

To perform the comparison we used a leave-one-out cross validation on each of the four datasets (Supplementary Fig. S1). From these results, some remarks are in order:

1. For both *in-vivo* datasets, cNODE trained with Reptile outperforms all other algorithms.

2. The ResNet architecture trained using only ADAM can provide reasonably accurate prediction for simple small species collections (i.e., for the $N = 5$ species in Drosophila gut and the $N = 8$ species in the soil *in-vitro* community).

3. For cNODE, training using the Reptile metalearning algorithm decreases the prediction error in the test dataset. Interestingly, using Reptile does not alway decreases the prediction error in the train dataset. Therefore, the Reptile metalearning algorithm is performing as desired to enhance the generalizing ability of cNODE.

4. For ResNet, training with the Reptile metalearning algorithm can increase the prediction errors in both the training and test datasets when compared to training only with ADAM.

5. The ResNet architecture exhibits a higher variability in the training set when compared to cNODE. This suggests that the performance of ResNet is significantly influenced by the initialization parameters. In particular, training a ResNet with Reptile can significantly increase the variability of prediction errors (see, e.g., the soil *in vitro* dataset).

Overall, the above remarks indicate that cNODE training with Reptile outperforms the other architectures when predicting complex microbial communities like human gut microbiota or *in vivo* soil communities.

S3. Population dynamics for the *in silico* validations.

We generated *in-silico* species pools using the Generalized Lotka-Volterra (GLV) equations, a classical population dynamics model successfully applied to diverse microbial communities, from soils[30] and cheese[36], to the human body[32,33]. The GLV model takes the form

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = x(t) \odot [Ax(t) + r], \quad x(0) = x_0, \tag{S1}$$

where $x(t) = (x_1(t), \cdots, x_N(t))^\top \in \mathbb{R}_{\geq 0}$ and $x_i(t)$ denotes the absolute abundance of species i at time t. Above, $x \odot v$ is the entry-wise multiplication of vectors $x, v \in \mathbb{R}^N$. The GLV model has two parameters: the interaction matrix of the species pool $A = (a_{ij}) \in \mathbb{R}^{N\times N}$, and the intrinsic growth-rate of the species $r = (r_i) \in \mathbb{R}^N$. In particular, the j-th species has a positive impact on the i-th species if $a_{ij} > 0$, a negative impact if $a_{ij} < 0$, and no impact if $a_{ij} = 0$. Recall also that the interaction matrix determines the underlying ecological network $\mathcal{G}(A)$ of the species pool. This network has one vertex associated to each species and an edge $(j \to i) \in \mathcal{G}(A)$ if $a_{ij} \neq 0$.

To generate the relative abundance vector $p \in \Delta^N$ corresponding to a local community with species collection $z \in \{0, 1\}^N$, we follow four steps:

1. Set the parameters (A, r).
2. Set the initial abundance of species $x_0 \in \mathbb{R}^N_{\geq 0}$ as

$$x_{0,\,i} = \begin{cases} 0 & \text{if } z_i = 0, \\ \text{Uniform}[0, 1] & \text{otherwise,} \end{cases}$$

   for $i = 1, \cdots, N$.
3. Numerically integrate Eq. (S1) with initial condition x_0 until the system reaches a steady-state abundance $x^* = (x_1^*, \cdots, x_N^*) \in \mathbb{R}^N$. For the results presented in our paper, we choose a final integration time $t_f = 1000$.
4. Compute the relative abundance vector $p = (p_1, \cdots, p_N)^T \in \Delta^N$ as $p_i = x_i^* / \sum_j x_j^*$.

Using the above procedure, we generated a dataset $\mathfrak{D}$ by randomly sampling species collections $z \in \{0, 1\}^N$ and calculating the corresponding $\hat{p} \in \Delta^N$. Below we detail the construction of the three types of datasets used in the Main Text.

S3.1 Generating datasets with universal dynamics.

To generate a dataset with universal dynamics, we considered that all species collections have the same parameters (A, r). These parameters were generated as follows. The interaction matrix A of the community is obtained as the adjacency matrix of a directed weighted Erdös-Rényi random network with connectivity $C \in [0, 1]$. The edge-weights were chosen from a Normal distribution with zero mean and variance σ^2, where $\sigma > 0$ represents the "characteristic" inter-species interaction strength. The intrinsic growth r_i is chosen uniformly at random from the interval $[0, 1]$.

S3.2 Generating datasets with non-universal dynamics.

To generate a dataset with non-universal dynamics, we considered two possible sources for non-universality. First, the mechanisms of interaction between species may differ across local communities. In Eq. (S1), this translates as using different parameters a_{ij} for each non-zero interaction in

different local communities. Thus, in this case we replaced each $a_{ij} \neq 0$ by $a_{ij} + \eta \text{Normal}(0, 1)$, where $\eta > 0$ quantifies the *changes in the typical interaction strength*, and hence the "loss" of universality in this case.

Second, we considered that each local community may have a different ecological network. To model this case, we considered that the ecological network of each local community is obtained by *randomly rewiring* a proportion $\rho \in [0, 1]$ of the edges of a baseline ecological network $\mathcal{G}(A)$, thus shuffling a proportion of entries of the associated A matrix. Since $\rho = 0$ corresponds to universal dynamics, the magnitude of ρ quantifies the "loss" of universality in this case.

S3.3 Adding measurement noise to a dataset.

For a pair (z, p) in a dataset $\mathfrak{D}$, we added noise by replacing p_i by first adding a small noise $w_i = \max\{0, p_i + \varepsilon \text{Normal}(0, 1)\}$, and then normalizing to obtain the noisy measurement $p_i \leftarrow w_i / \sum_j w_j$. Here, the parameter $\varepsilon > 0$ controls the *measurement noise intensity*.

S3.4 Generating datasets with multi-stability.

To generate a dataset with true multi-stability, we calculated the steady-states from a population dynamics model with the following non-linear functional response:

$$\frac{\mathrm{d}x_i(t)}{\mathrm{d}t} = x_i(t) \left[r_i + \sum_{j=1}^{N} a_{ij} \frac{x_j(t)}{1 + h\, x_j(t)^2} \right], \quad i = 1, \cdots, N, \tag{S2}$$

where h denotes the handling time.

To generate steady-states with multi-stability, we first select a GLV model with a linear functional response (Eq. S1) and universal dynamics (Supplementary Notes S3), and compute the steady-state ξ^*. Note that the steady-state abundances satisfies the equation

$$r_i = -\sum_{j=1}^{N} a_{ij}\, \xi_j^*. \tag{S3}$$

The steady-states of Eq.(S2) satisfies the equation

$$\sum_{j=1}^{N} a_{ij} \frac{x_j}{1 + h\, x_j^2} + r_i = 0, \tag{S4}$$

so that we substitute Eq.S3 in Eq.S4 and solve for x_j the following quadratic equation:

$$h \xi_j^* x_j^2 - x_j + \xi_j^* = 0, \tag{S5}$$

for all j, and then we compute the relative abundance vector $p = (p_1, \cdots, p_N)^T \in \Delta^N$ as $p_i = x_i^* / \sum_j x_j^*$. To ensure that there are two real solutions for x_j, we chose $h = \frac{1}{4\xi_k^{*2}} > 0$ for some k.

The two steady-state abundances corresponds to a high and low total biomass regimes, respectively. To build the datasets, we chose a fraction (1-μ) from the first regime, and the rest from the second.

S4. Description of the experimental datasets.

S4.1 In-vivo drosophila core gut microbiome.

The drosophila dataset[14] contains the absolute abundance of the five species in each possible local community with different species collection. See Supplementary Table S1 for species IDs. There is five replicates for each of those species collections. We averaged those five replicates, discarded samples with a single species, and obtained the relative abundance of each of the remaining samples. This yielded 31 samples with different species collection.

ID	Genus	Species
1	Lactobacillus	plantarum
2	Lactobacillus	brevis
3	Acetobacter	pasteurianus
4	Acetobacter	tropicalis
5	Acetobacter	orientalis

Supplementary Table S1: Species IDs Drosophila gut microbiota.

S4.2 In-vitro soil community.

This laboratory community of $N = 8$ heterotrophic soil-dwelling bacterial species[12] described in Table S2. The available dataset contains 98 samples, including all solos, all duos, some trios, one septet and one octet.

ID	Genus	Species
1	Enterobacter	aerogenes
2	Pseudomonas	aurantiaca
3	Pseudomonas	chlororaphis
4	Pseudomonas	citronellolis
5	Pseudomonas	fluorescens
6	Pseudomonas	putida
7	Pseudomonas	veronii
8	Serratia	marcescens

Supplementary Table S2: Species IDs the in-vitro soil community.

S4.3 -vivo soil microbiome.

The soil dataset consists of soil microbiome across Central Park in New York City consist of 1160 samples. This data set is 16S rRNA gene-based with variable region V4. The data is available at https://qiita.ucsd.edu/ under study ID 2140 and the detailed description of this data set can be found in Ref. [13]. We used the function `summarize_taxa.py` QIIME 1 to summarize taxa to different taxonomic levels with defaults options. Supplementary Table S3 provides the IDs associated to each phylumm.

S4.4 Human gut microbiome.

A 16S rRNA gene-based data set from variable regions V3 to V5. The data are available at http://www.hmpdacc.org/HMQCP/. We selected the samples from the stool body site. For multiple samples from the same subject, we only keep one single sample of that subject. To guarantee the

ID	Kindgom	Phylum
1	Archaea	Crenarchaeota
2	Archaea	Euryarchaeota
3	Archaea	Parvarchaeota
4	Bacteria	others
5	Bacteria	Acidobacteria
6	Bacteria	Actinobacteria
7	Bacteria	Aquificae
8	Bacteria	Armatimonadetes
9	Bacteria	BHI80-139
10	Bacteria	Bacteroidetes
11	Bacteria	Chlorobi
12	Bacteria	Chloroflexi
13	Bacteria	Cyanobacteria
14	Bacteria	Elusimicrobia
15	Bacteria	FBP
16	Bacteria	Firmicutes
17	Bacteria	GN02
18	Bacteria	Gemmatimonadetes
19	Bacteria	Lentisphaerae
20	Bacteria	NC10
21	Bacteria	Nitrospirae
22	Bacteria	OD1
23	Bacteria	OP3
24	Bacteria	OP9
25	Bacteria	Planctomycetes
26	Bacteria	Proteobacteria
27	Bacteria	SBR1093
28	Bacteria	Spirochaetes
29	Bacteria	TM6
30	Bacteria	TM7
31	Bacteria	TPD-58
32	Bacteria	Tenericutes
33	Bacteria	Verrucomicrobia
34	Bacteria	WPS-2
35	Bacteria	WS1
36	Bacteria	WS6

Supplementary Table S3: Phylum IDs for soil dataset.

model can be trained sufficiently, we summarized the taxa into the genus level and removed the genus with fewer than 50 reads. See also Supplementary Table S4 for genus ID.

ID	Phylum	Class	Order	Family	Genus
1	Firmicutes	Clostridia	Clostridiales	Veillonellaceae	*Veillonella*
2	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	*Clostridium*
3	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	*Bacteroides*
4	Tenericutes	Erysipelotrichi	Erysipelotrichales	Erysipelotrichaceae	*Coprobacillus*
5	Firmicutes	Clostridia	Clostridiales	ClostridialesFamilyXIII.IncertaeSedis	
6	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	*Odoribacter*
7	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	*Lachnobacterium*
8	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Verrucomicrobiaceae	*Akkermansia*
9	Bacteroidetes				
10	Proteobacteria	Gammaproteobacteria	Pasteurellales	Pasteurellaceae	*Haemophilus*
11	Firmicutes	Clostridia	Clostridiales	Veillonellaceae	*Megasphaera*
12	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	*Streptococcus*
13	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	*Anaerotruncus*
14	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	*Parabacteroides*
15	Firmicutes	Clostridia	Clostridiales	Dehalobacteriaceae	*Dehalobacterium*
16	Tenericutes				
17	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	*Ruminococcus*
18	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	*Dorea*
19	Tenericutes	Erysipelotrichi	Erysipelotrichales	Erysipelotrichaceae	*Catenibacterium*
20	Proteobacteria	Deltaproteobacteria	Desulfovibrionales	Desulfovibrionaceae	*Desulfovibrio*
21	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	*Oscillospira*
22	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	*Clostridium*
23	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	*Escherichia*
24	Firmicutes	Bacilli	Turicibacterales	Turicibacteraceae	
25	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	
26	Actinobacteria	Actinobacteria	Coriobacteriales	Coriobacteriaceae	*Collinsella*
27	Firmicutes	Clostridia	Clostridiales	Veillonellaceae	*Acidaminococcus*
28	Firmicutes	Clostridia	Clostridiales	ClostridialesFamilyXIII.IncertaeSedis	*Eubacterium*
29	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	*Lactobacillus*
30	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	*Roseburia*
31	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	*Eubacterium*
32	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	*Lachnospira*
33	Bacteroidetes	Bacteroidia	Bacteroidales		
34	Proteobacteria	Deltaproteobacteria	Desulfovibrionales	Desulfovibrionaceae	*Bilophila*
35	Bacteroidetes	Bacteroidia	Bacteroidales	Prevotellaceae	*Prevotella*
36	Bacteroidetes	Bacteroidia	Bacteroidales	Rikenellaceae	*Alistipes*
37	Firmicutes	Clostridia	Clostridiales	Veillonellaceae	*Dialister*
38	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	*Ruminococcus*
39	Proteobacteria	Betaproteobacteria	Burkholderiales		
40	Tenericutes	Erysipelotrichi	Erysipelotrichales	Erysipelotrichaceae	*Holdemania*
41	Fusobacteria	Fusobacteria	Fusobacteriales	Fusobacteriaceae	*Fusobacterium*
42	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	*Eubacterium*
43	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	*Subdoligranulum*
44	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	*Faecalibacterium*
45	Tenericutes	Erysipelotrichi	Erysipelotrichales	Erysipelotrichaceae	*Clostridium*
46	Proteobacteria	Betaproteobacteria	Burkholderiales	Alcaligenaceae	*Sutterella*
47	Others				
48	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	*Clostridium*
49	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	*Coprococcus*
50	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	*Blautia*
51	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	*Bacteroides*
52	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	
53	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	*Bacteroides*
54	Firmicutes	Clostridia	Clostridiales	Veillonellaceae	*Phascolarctobacterium*
55	Firmicutes	Clostridia	Clostridiales		
56	Actinobacteria	Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	*Bifidobacterium*
57	Firmicutes	Clostridia	Clostridiales	Veillonellaceae	*Megamonas*
58	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	

Supplementary Table S4: Genus IDs for the human gut microbiota dataset.

S5. Related work.

Here we describe related methods, emphasizing some key differences with respect to our framework.

1. **Abundance prediction based on inferred population dynamics.** A classical method to predict species abundance in microbial ecosystems is modeling their population dynamics[12,32,33,36,37]. Typically, the model is a set of parametrized ODEs —such as the Generalized Lotka-Volterra equations— describing the changes over time in the *absolute* abundances of a set of species. The model is fitted to available temporal data of absolute abundance to infer parameters, such as intrinsic growth rates, inter-species interaction strengths, etc. Then, to predict the abundance of certain species collection, the fitted ODE model is solved starting from suitable initial condition. However, applying this method for large microbial communities like the human gut is challenging if not impossible because: 1) the absence of high-quality temporal data; and 2) typically only relative abundance of species are measured. Furthermore, because the very broad population dynamics that ecosystems display even at the scale of two species[64], it is always very challenging to choose an adequate parametrized ODE model for the population dynamics of the community.

2. **Predictions based on neural networks methods.** Larsen et al.[52] employed an artificial neural network to predict the *temporal* evolution of the composition of bacterial communities with a constant species collection. More precisely, they developed a bioclimatic model of relative microbial abundance that specifically incorporates interactions between biological units. They modeled the complex interactions between microbial taxa and their environment as an artificial neural network (ANN). This method is based on two key assumptions: (1) community patterns share mathematically describable relationships with environmental conditions; and (2) the ecosystem maintains a persistent microbial community. Note that the second assumption implies that this method can not be used to predict the impact of changing the species collections. Compared to method based on inferring population dynamics, this method has the advantage of not requiring to specify any model for the community dynamics. However, in contrast to our framework, this method cannot predict the effect of changing the species collection.

   Similarly, the recent work of Zhou et al.[53] uses a neural network to predict the temporal and *spatial* evolution of the composition of microbial communities with a constant species collection. More precisely, the authors formulated the prediction of microbial communities at unsampled locations as a multi-label classification task, where each location is considered as an instance and each label represents a microbe species. Based on a set of heterogeneous features extracted from the urban environment, they aimed to predict the presence or absence of a list of microbes species at a nearby location. Note, however, that this method cannot be immediately used to predict the effect of changing the species collection.

3. **Predictions based on statistical methods.** Recently, Maynard et al.[40] proposed a statistical method to predict species abundances from species collections. More precisely, based on measuring the absolute abundance of species at some steady-states of the ecosystem, this method assumes a linear model to predict all other steady-states. For the method to be applicable, it requires that the following assumptions are satisfied: (1) each species must be present in at least n distinct endpoints, not counting replicates; (2) each species must co-occur with each other species in at least one endpoint (that is, for every pair of species i and j, there must be some endpoint where i and j co-occur, possibly along with other species); and (3) for each i there must exist a perfect matching between the n species and

the endpoints in which they co-occur with i. As explained in the original manuscript[40], conditions (1) and (2) above requires that "coexistence among species must be reasonably widespread for [these] conditions to hold." This method may be challenging to apply for microbial communities because it requires measuring absolute abundances. Furthermore, because microbial communities tend to have nonlinear behaviors even at the scale of two species[64,65], the implicit assumption of linearity may fail to be satisfied. Finally, we note that cNODE does not require any of the above three assumptions to be applicable, although its prediction accuracy may be influenced by them.

Similarly, Tung et al.[66] use a linear regression method to predict species compositions from information of social networks of individuals. More precisely, the authors fit a classical linear mixed model to predict the relative abundance of species in a sample based on the following predictors: social group membership, age, sex, and read depth. We note here that the predictors in this approach are completely different than the predictors used in cNODE (i.e., it only uses species collections), and thus are not comparable.

S6. Comparing cNODE with Maynard et al.'s method.

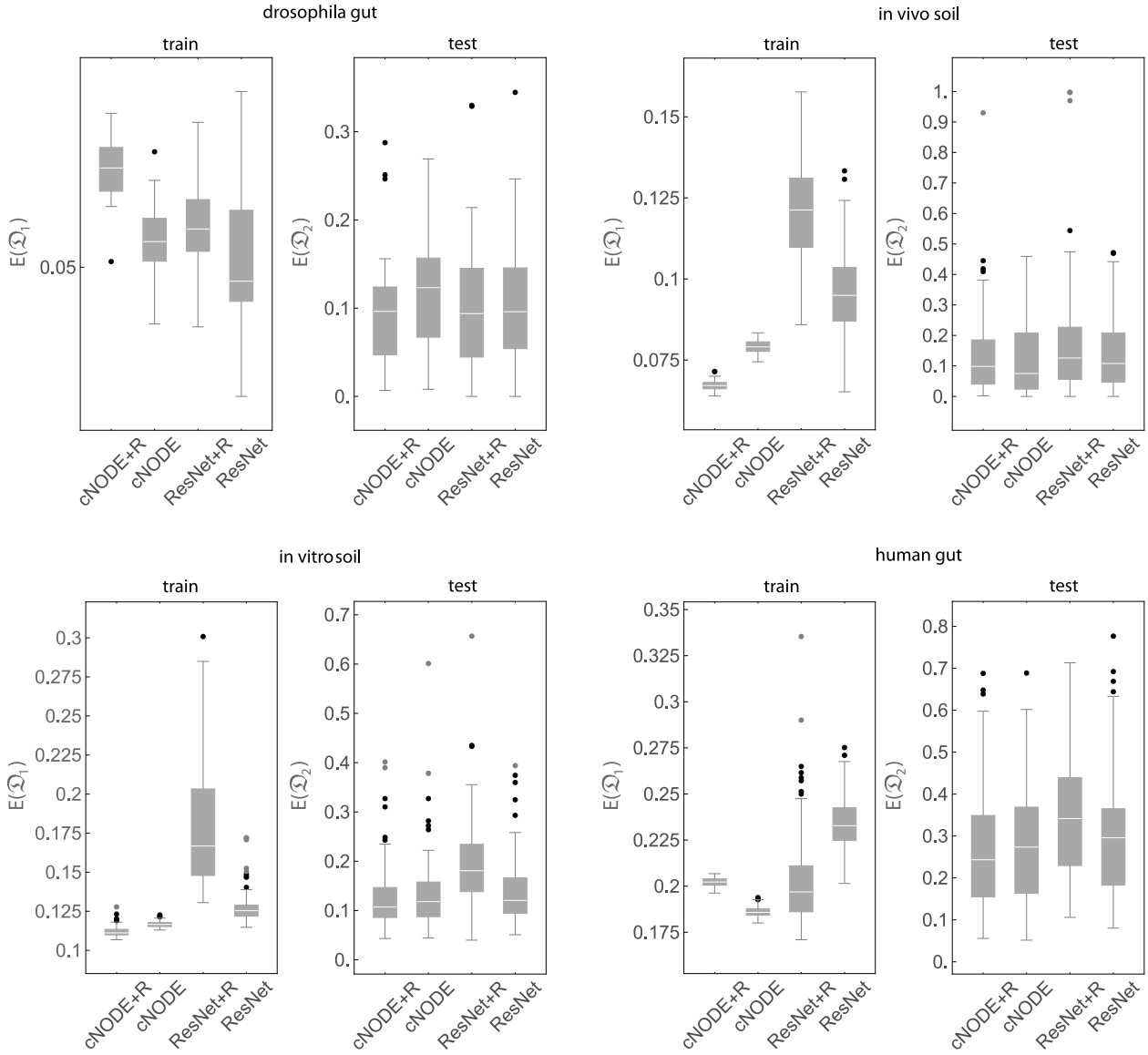
Here we compare the performance of cNODE with the method of Maynard et al. for predicting endpoints[40]. This last method is described with details in item 3 of Supplementary Note S5.

To perform the comparison, we generated *in silico* datasets of $N = 5$ species with Generalized Lotka-Volterra dynamics. More precisely, we generated datasets with universal dynamics, maintaining the connectivity $C = 0.5$ constant and changing the typical interaction strength as $\sigma \in \{0.1, 0.2, 0.3, 0.4\}$. For each value of σ, we generated 10 datasets containing all $S = 2^5 - 1$ samples with different species collections following the simulation method described in Supplementary Note S3. We repeated this simulation method three times, obtaining three repetitions for the abundance of each species collections that can be used in Maynard's method. Additionally, because Maynard's method requires absolute abundances, we kept in the datasets both the absolute abundance and relative abundance of each steady-state that is reached. Using these datasets, we constructed training datasets by randomly choosing 70% of the samples, and the rest of the samples as test datasets.

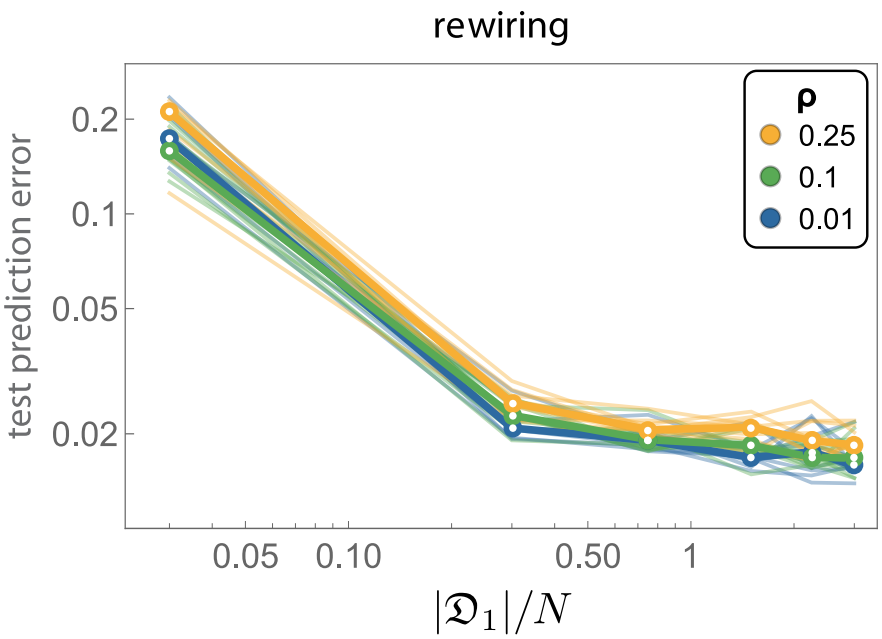
To adjust Maynard's method we used the default parameters that were selected for $N = 4$ species. After this, we obtained the corresponding absolute abundance predictions of the test dataset. Finally, to allow a comparison with cNODE that predicts relative abundances, we transformed each predicted absolute abundance into a predicted relative abundance, and then calculated the prediction error using the Bray-Curtis dissimilarity.

For cNODE, we used the exact same training dataset as for Maynard's method, the only difference being that we trained cNODE with relative abundances. We choose a inner learning rate of 0.001, an outer learning rate of 0.005, and trained cNODE for 500 epochs using mini batches of 5 samples. We then calculated the prediction error in the test dataset using the Bray-Curtis dissimilarity. We emphasize that cNODE uses less information that Maynard's method, in the sense that the total biomass of each sample is unknown in this case.

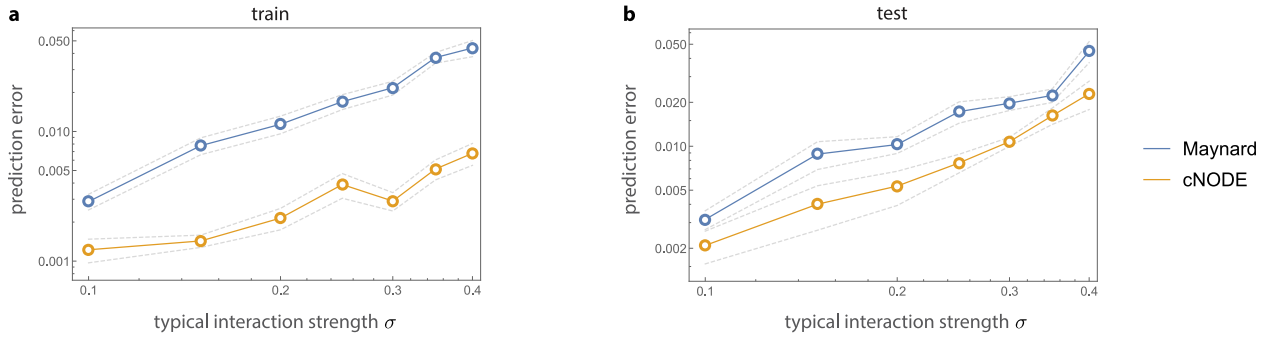
The results of the comparison of Maynard's method and cNODE are shown in Fig. S3. We find that, for the above conditions used to construct the datasets, cNODE outperforms Maynard's method in both the training and test datasets. Crucially, note that cNODE was trained using only relative abundance measurements. We do not claim this results holds for all datasets, as there might be cases where the assumptions required by Maynard's method are exactly satisfied but none of the assumptions of cNODE are satisfied.



Supplementary Figure S1: **Performance of the ResNet and cNODE architectures for predicting compositions in experimental microbiomes.** Vertical axis denotes prediction error.



Supplementary Figure S2: **Performance of the ResNet and cNODE architectures for predicting compositions in experimental microbiomes.** Vertical axis denotes prediction error.



Supplementary Figure S3: **Prediction errors of Maynard et al.'s method[40] and cNODE.** For an *in silico* dataset of $N = 5$ species with universal dynamics and different typical interaction strength. Circles denote mean error for 10 repetitions, and gray shadows indicate standard deviation of the mean. **a.** Train dataset. **b.** Test dataset.