Predicting microbiome compositions from species assemblages 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 25th June, 2021

Abstract

Microbes can form complex communities that perform critical functions in maintaining the integrity of their environment or their hosts' well-being. Rationally managing these microbial communities requires improving our ability to predict how different species assemblages affect the final species composition of the community. However, making such a prediction remains challenging because of our limited knowledge of the diverse physical, biochemical, and ecological processes governing 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 only, without knowledge of any of the above processes. 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, *Drosophila melanogaster* gut microbiota, and human gut and oral microbiota. In particular, we show how our framework learns to perform accurate out-of-sample predictions of complex community compositions from a small number of training samples. Our results demonstrate how deep learning can enable us to understand better and potentially manage complex microbial communities.

Introduction

Microbes can form complex multispecies communities that perform critical functions in maintaining the integrity of their environment[1,2] or the well-being of their hosts[3–6]. For example, microbial communities play key roles in nutrient cycling in soils[7] and crop growth[8]. In humans, the gut microbiota plays important roles in our nutrition[9], immune system response[10], pathogen resistance[11], and even our nervous central system response[5]. Still, species invasions (e.g., pathogens) and extinctions (e.g., due to antibiotic administration) produce changes in the species assemblages that may shift these communities to undesired compositions[12]. For instance, antibiotic administrations can shift the human gut microbiota to compositions making the host more susceptible to recurrent infections by pathogens[13]. Similarly, intentional changes in the species assemblages, such as by using fecal microbiota transplantations, can shift back these communities to desired "healthier" compositions[14,15]. Therefore, improving our ability to rationally manage these microbial communities requires that we can predict changes in the community composition based on changes in species assemblages[16]. Building these predictions would also reduce managing costs, helping us to predict which changes in the species' assemblages are more likely to yield a desired community composition. Unfortunately, making such a prediction remains challenging because of our limited knowledge of the diverse physical[17], biochemical[18], and ecological[19,20] processes governing the microbial dynamics.

To overcome the above challenge, here we present a deep learning framework that automatically learns the map between species assemblages and community compositions from training data only, without knowledge of the underlying microbial dynamics. We systematically validated our framework using synthetic data generated by classical ecological dynamics models, demonstrating its robustness to changes in the system dynamics and to measurement errors. Then, we applied our framework to experimental data of both *in vitro* and *in vivo* communities, including ocean and soil microbial communities[21,22], *Drosophila melanogaster* gut microbiota[23], and human gut and oral microbiota[24]. Across these diverse microbial communities, we show how our framework learns to predict accurate out-of-sample compositions given a few training samples. Our results show how deep learning can be an enabling ingredient for understanding and managing complex microbial communities.

Methods

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[25]. 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[17], the chemical gradients of available resources[18], and inter/intra-species interactions[20], 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.

Next, 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[26] (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 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 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, as well as certain environment microbiota, display strong universal dynamics[26]. *Second*, we assume that the compositions of those collected samples represent steady states of the microbial communities. This assumption is natural because the map φ is not well defined for highly fluctuating microbial compositions. 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[24,27,28]. *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 cannot be strictly satisfied. Therefore, 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[29,30] such as a feedforward Residual Network[31] (ResNet). As a top-rated 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 result is likely due to the normalization of the output layer, which challenges the training of neural networks because of vanishing gradients[30].

The vanishing gradient problem is often solved by using other normalization layers such as the `softmax` or `sparsemax` layers[32]. However, we cannot use these layers 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[33].

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)[34]. 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

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

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) is a general form of the replicator equation —a canonical model in evolutionary game theory[35]— 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[36] 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. 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[37]. 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).

Figure 1 shows the result of applying cNODE to fly gut microbiome samples collected in an experimental study[23]. In this study, germ-free flies (*Drosophila melanogaster*) were colonized with all possible combinations of $N = 5$ 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 assemblage of species 3 with species 4, 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 classic Generalized Lotka-Volterra (GLV) model[38]

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

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 $r = (r_i) \in \mathbb{R}^N$. The parameter 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. 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 successfully describes the population dynamics of microbial communities in diverse environments, from the soil[39] and lakes[40] to the human gut[11,41,42]. 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[43] 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[23], as described in Fig. 1. The third dataset has $S = 93$ samples of *in vitro* communities of $N = 8$ soil bacterial species[21]. The fourth dataset contains $S = 113$ samples of the Central Park soil microbiome[22] at the phylum level ($N = 36$ phyla). The fifth dataset contains $S = 150$ samples of the human oral microbiome from the Human Microbiome Project[24] (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 out-of-sample predicted composition 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[44].

Discussion

cNODE is a deep learning framework to predict microbial compositions from species assemblages only. We validated its performance using *in silico*, *in vitro*, and *in vivo* microbial communities, finding that cNODE learns to perform accurate out-of-sample predictions using a few training samples. Classic methods for predicting species abundances in microbial communities require inference based on population dynamics models[21,41,45,46]. However, these methods typically require high-quality time-series data of species absolute abundances, which can be difficult and expensive to obtain for *in vivo* microbial communities. cNODE circumvents needing absolute abundances or time-series data. However, compared to the classic methods, the cost to pay is that the trained function f_θ cannot be mechanistically interpreted because of the lack of identifiability inherent to

compositional data[47,48]. We also note a recent statistical method to predict the steady-state abundance in ecological communities[49]. 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 in microbiome research[50–58], such as for classifying samples that shifted to a diseased state[59], predicting infection complications in immunocompromised patients[60], or predicting the temporal or spatial evolution of certain species collection[61,62]. However, to the best of our knowledge, the potential of deep learning for predicting the effect of changing species assemblages was not explored nor validated before. Our proposed framework, based on the notion of neural ODE[34], is a baseline that could be improved by incorporating additional information. For example, incorporating available environmental information such as pH, temperature, age, BMI, and host's diet could enhance the prediction accuracy. This additional information 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[63,64], but this may result in 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[65] that provides a well-defined gradient[66].

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.

Our framework does have limitations. For example, cNODE 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 make it very difficult to decide if true multistability exists or not. Our framework could be extended to handle multistability 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.

In conclusion, the many species and the complex, uncertain dynamics that microbial communi-

ties exhibit have been fundamental obstacles in our ability to learn how they respond to alterations, such as removing or adding species. Moving this field forward may require losing some ability to interpret the mechanism behind their response. In this sense, deep learning methods could enable us to rationally manage and predict complex microbial communities' dynamics.

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. Fierer, N. Embracing the unknown: disentangling the complexities of the soil microbiome. *Nature Reviews Microbiology* **15,** 579–590 (2017).
8. Mueller, U. G. & Sachs, J. L. Engineering microbiomes to improve plant and animal health. *Trends in microbiology* **23,** 606–617 (2015).
9. Gill, S. R., Pop, M., DeBoy, R. T., Eckburg, P. B., Turnbaugh, P. J., Samuel, B. S., Gordon, J. I., Relman, D. A., Fraser-Liggett, C. M. & Nelson, K. E. Metagenomic analysis of the human distal gut microbiome. *science* **312,** 1355–1359 (2006).
10. Hooper, L. V., Littman, D. R. & Macpherson, A. J. Interactions between the microbiota and the immune system. *science* **336,** 1268–1273 (2012).
11. 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).
12. Costello, E. K., Stagaman, K., Dethlefsen, L., Bohannan, B. J. & Relman, D. A. The application of ecological theory toward an understanding of the human microbiome. *Science* **336,** 1255–1262 (2012).
13. Lemon, K. P., Armitage, G. C., Relman, D. A. & Fischbach, M. A. Microbiota-targeted therapies: an ecological perspective. *Science translational medicine* **4,** 137rv5–137rv5 (2012).
14. Borody, T. J., Paramsothy, S. & Agrawal, G. Fecal microbiota transplantation: indications, methods, evidence, and future directions. *Current gastroenterology reports* **15,** 1–7 (2013).
15. Xiao, Y., Angulo, M. T., Lao, S., Weiss, S. T. & Liu, Y.-Y. An ecological framework to understand the efficacy of fecal microbiota transplantation. *Nature communications* **11,** 1–17 (2020).
16. 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).
17. 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).

18. 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).

19. 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).

20. Coyte, K. Z., Schluter, J. & Foster, K. R. The ecology of the microbiome: networks, competition, and stability. *Science* **350,** 663–666 (2015).

21. Friedman, J., Higgins, L. M. & Gore, J. Community structure follows simple assembly rules in microbial microcosms. *Nature Ecology & Evolution* **1,** 0109 (2017).

22. 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).

23. 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).

24. 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).

25. Suez, J., Zmora, N., Segal, E. & Elinav, E. The pros, cons, and many unknowns of probiotics. *Nature medicine* **25,** 716–729 (2019).

26. 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).

27. 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).

28. 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).

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

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

31. 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.

32. 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.

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

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

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

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

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

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

39. 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).

40. 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).

41. 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).

42. Rao, C., Coyte, K. Z., Bainter, W., Geha, R. S., Martin, C. R. & Rakoff-Nahoum, S. Multi-kingdom ecological drivers of microbiota assembly in preterm infants. *Nature* **591,** 633–638 (2021).

43. 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).

44. 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).

45. 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).

46. 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).

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

48. 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).

49. Maynard, D. S., Miller, Z. R. & Allesina, S. Predicting coexistence in experimental ecological communities. *Nature ecology & evolution* **4,** 91–100 (2020).

50. 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.

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

52. 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).

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

54. 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).

55. 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).

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

57. 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).

58. 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).

59. 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.

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

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

62. 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).

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

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

65. 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).

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

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

68. 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.

69. 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).

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

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

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

73. 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).

74. 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).

75. 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

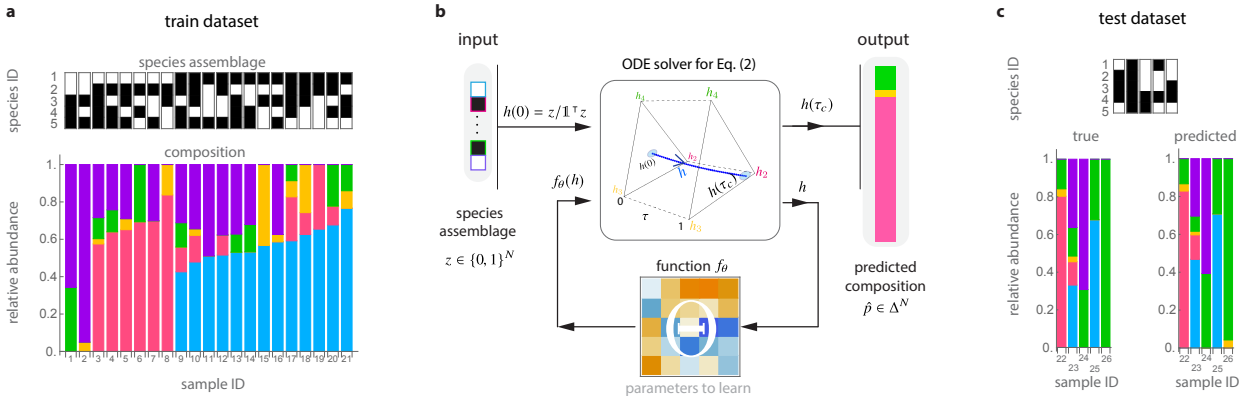

Figure 1: **A deep learning framework to predict microbiome compositions from species assemblages.** We illustrate this framework using experimental data from a pool of $N = 5$ bacterial species in *Drosophila melangaster* gut microbiota[23]: *Lactobacillus plantarum* (blue), *Lactobacillus brevis* (pink), *Acetobacter pasteurianus* (yellow), *Acetobacter tropicalis* (green), and *Acetobacter orientalis* (purple). **a.** We randomly split this dataset into training ($\mathfrak{D}_1$) and test ($\mathfrak{D}_2$) datasets, which contain 80% and 20% of the samples, respectively. Each dataset contains pairs (z, p) with the species assamblage $z \in \{0,1\}^N$ (top) and its corresponding 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, the performance is evaluated by predicting the composition of never-seen-before species assamblages in the test dataset $\mathfrak{D}_2$. In this experimental microbiota, cNODE learned to perform accurate predictions of the composition in the test dataset. For example, in the assemblage of species 3 and 4 (sample 26), cNODE correctly predicts that the composition is strongly dominated by a single species.

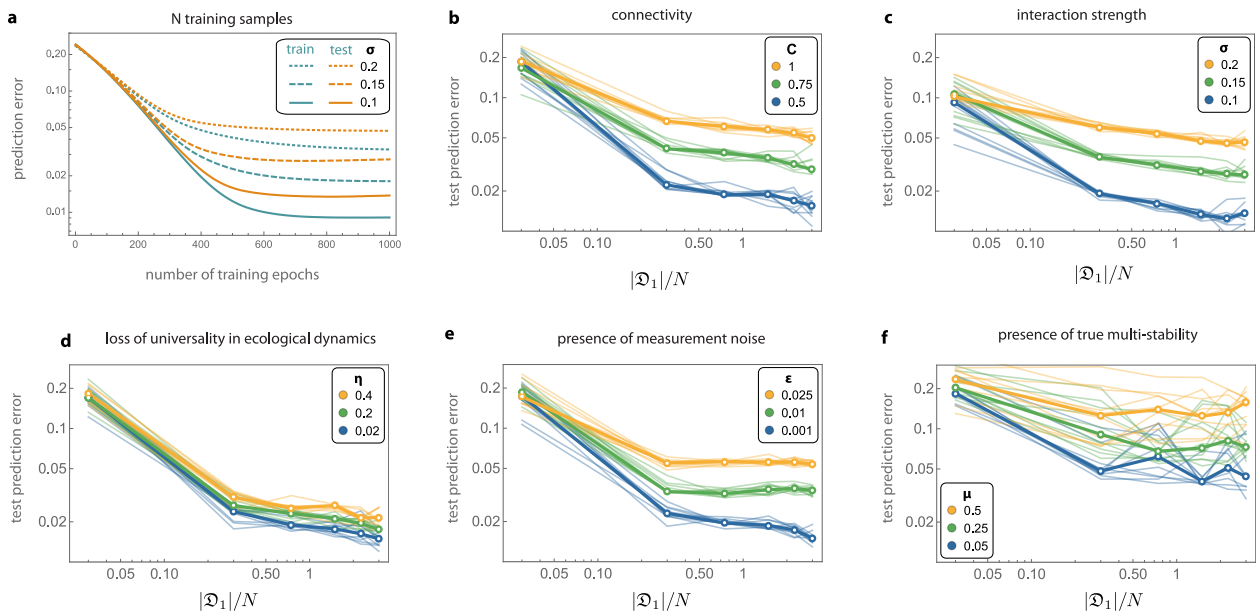

Figure 2: ***In silico* validation of cNODE using synthetic datasets.** Results are for synthetic communities of $N = 100$ species generated by the with Generalized Lotka-Volterra model (panels **a**-**e**) or a population dynamics model with nonlinear functional responses (panel **f**). **a.** Training 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.

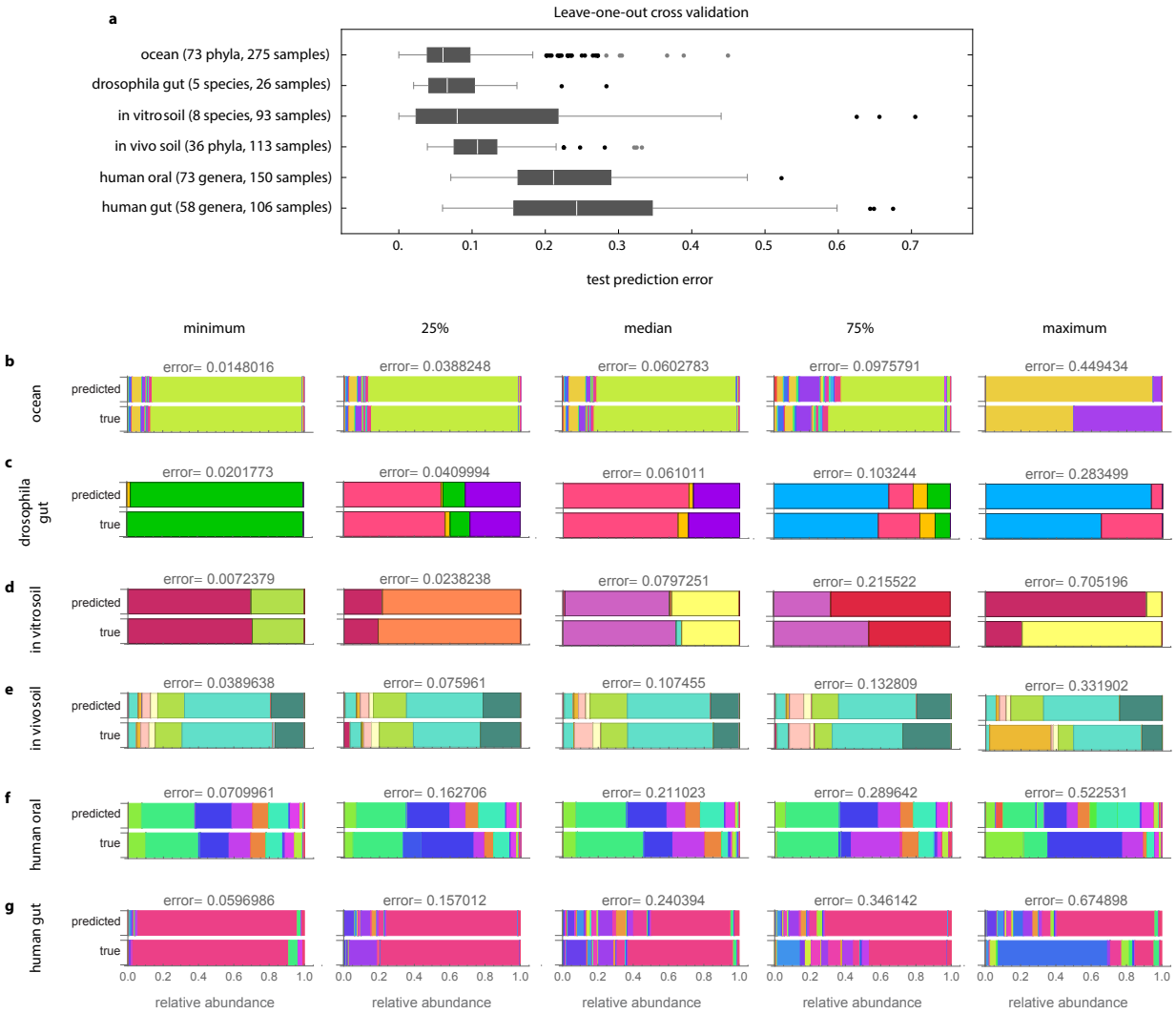

Figure 3: **Predicting the composition of real microbiomes. a.** Boxplots with the prediction error obtained from a leave-one-out crossvalidation of each dataset. **b-g**: For each dataset, we show true and predicted samples corresponding to the minimal prediction error, closer to the first quartile,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