UNIVERSITY OF CALIFORNIA SAN DIEGO

Predicting the Complexity and Progression of the Gut Microbiome Using Temporal Data and Deep Learning

A thesis submitted in partial satisfaction of the requirements
for the degree Masters of Science

in

Bioengineering

by

Michael Wiest

Committee in Charge:
Professor Karsten Zengler, Chair
Professor Gert Cauwenberghs, Co-chair
Professor Sheng Zhong

The Thesis of Michael Wiest is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

Co-Chair

_____

Chair

University of California San Diego

2019

Table of Contents

List of Figures

List of Tables

Acknowledgements

This research would not have been possible without Karsten. He generously afforded me the freedom to pursue something that nobody else in the lab was researching. I would also like to thank Cristal Zuniga and Manish Kumar for working tirelessly with me on this research project. I could not have asked for better collaborators.

Also I would like to thank Gert Cauwenberghs for the use of his lab's GPU servers to train my models.

ABSTRACT OF THE THESIS

Predicting the Complexity and Progression of the Gut Microbiome Using Temporal Data and Deep Learning

By

Michael Wiest

Master of Science in Bioengineering

University of California San Diego, 2019

Professor Karsten Zengler, Chair

The human microbiota exhibit a highly dynamic composition over the course of life and changes in the human gut microbiota have been associated with human health or disease. Reprogramming of the gut microbiota by interventions that counter these changes and promote long-lasting health has been an emerging topic in microbiome research. Predicting changes in the gut microbiome is therefore crucial for the nature and design of these interventions. Here, we report on a new method based on deep learning to forecast changes in the microbiome. We processed and analyzed nine time-course datasets of the human gut microbiome, identifying the main microorganisms present in these microbial communities at any given time. We then used an encoder-decoder neural network to train a model that successfully predicts the progression of the microbiome composition over time given only five time points of context data. Our results demonstrate the ability to predict the fate of the human gut microbiome into the future, providing the foundation for rational intervention design.

Introduction

As humans, from the moment we are born we are exposed to and become hosts of a rich and prolific consortium of bacteria, viruses, and fungi. Our microbial stowaways are so impactful on our health that they have been implicated in everything from metabolism to immune response.[1] This diverse set of microbes changes with us as we age,[2] become ill,[3] or while dieting[4] or travelling.[5] The human microbiome is ubiquitous and very diverse, not only from person-to-person, but also between different sites on the human body.[6] For example, within a given host, at the same sampling site fewer than 50% of genes are shared by greater than 50% of a host's microbes.[1] Current research suggests that the human microbiome is in general stable, consisting of a core group of microbes with individual capabilities.[7]

The unit of measurement when analyzing the microbiome has traditionally been the operational taxonomic unit (OTU), which is typically presented as the relative abundance of a particular fragment of 16S rRNA.[8] Numerous studies have demonstrated static snapshots of the human microbiome composition in terms of OTU abundance during healthy and diseased states.[6] Currently, there has been a growing interest in assessing the dynamics of the microbiota and its impact on human health and disease.[9] Gut microbiota have been implicated in contributing to many diseases, such as Irritable Bowel Disease (IBD),[10] type-2 diabetes,[11] and autism.[12] Moreover, therapies that restore, alter, or restart the gut microbiota have shown promising results in combating microbiome-related diseases.[13] A particular therapy, microbial transplants, have shown success across a variety of diseased states,[14] and have demonstrated improved outcomes after transplantation.[15] States of dysbiosis in a microbial system are not only clinically problematic, but also financially burdensome; *Clostridium difficile* infections alone are estimated to cost the U.S. Healthcare system 4.8 billion dollars annually.[16] Maintaining healthy gut microbiota and thus preventing dysbiosis in the first place would be highly desirable over drastic interventions, such as fecal microbial transplantation. However, gut microbiota homeostasis and dysbiosis prevention require the ability to predict changes in the microbiota before they take place.

Given the large number of potential interactions between microbes in a system, deconvoluting the exact role of a microorganism in a sample is currently a challenging task. Thankfully due to advancements in machine learning and artificial intelligence, computational tools are now available that master the task of predictions from large-scale data, in particular time series data.[17] Neural networks are machine learning tools that have been deployed successfully for self-driving cars, natural language processing, and many other tasks detecting patterns in high-dimensional data. Put simply, neural networks "learn" a mapping from a set of inputs ($\overline{X}$) to a set of output values (y). When training a neural network a "loss function" is used to evaluate the model's prediction accuracy. A loss function is a mathematical representation between the similarity of a guess and the ground-truth. Different loss functions are appropriate for different prediction tasks; cross-entropy loss (multi-class classification) and root mean squared error (continuously valued data) are just two examples.[18] For networks, where there is a known

output value, y, to predict (such as OTU abundance) the learning process proceeds thusly: (1) The model is given a set of inputs ($\overline{X}$) and makes an estimate of the corresponding output values ($\hat{y}$). (2) The model's estimated output ($\hat{y}$) is compared to the ground-truth value (y) and the accuracy is evaluated given a loss function. (3) Given the determined accuracy, the model reweights its internal representation of the data through a process called backpropagation.[19] (4) Steps 1-3 are repeated until the model has converged or the accuracy is satisfactory.

Neural networks have an advantage over traditional machine learning techniques such as support vector machines or linear regressions in that they can learn non-linear relationships between inputs and outputs and are able to handle long range temporal interactions.[20] In simpler neural networks, examples of training data are considered to be independent of one another. These networks are known as feed-forward networks (FFN). On the other hand, recurrent neural networks (RNN) are given training data that are explicitly ordered in time.[20] More formally, at a given timepoint, t, the model is supplied an input tensor $\overline{X^t}$ that corresponds to a target value of y^t. This input tensor can be one-dimensional (e.g. containing expression data over time for a single gene), three-dimensional (e.g. for still frames of a video over time), or 36-dimensional as used in this study. Additionally, the sequences supplied to an RNN have a defined order but no explicit notion of the size of the temporal distance between data points; an example of this is a string of nucleotides.[20]

Collection of time series data enables unraveling of microbial dynamics over time. In one yearlong study of the gut microbiome of two subjects, the authors found that the hosts' experiences (e.g. traveling, illness) had a significant impact on the composition of the host's microbiome.[5] Traveling abroad resulted in a reconfiguration of the gut microbiome; however, upon returning to the subject's home, the microbiome returned to its previous equilibrium. On the other hand, the host's microbiota was permanently altered when the subject was infected with *Salmonella*, without recovering the equilibrium state observed prior to infection.[5]

Here, we used time series data (OTU abundance) from the gut microbiome of nine human donors. These data were used to train and evaluate a single neural network (Fig 1) for predicting the dynamic profiles of nine different gut microbiomes. We also performed microbial association analysis[21] to estimate the associated partners in gut microbial communities of all nine donors. An initial modeling framework to predict the future composition of the human gut microbiome based on past microbiome data of a subject is presented.

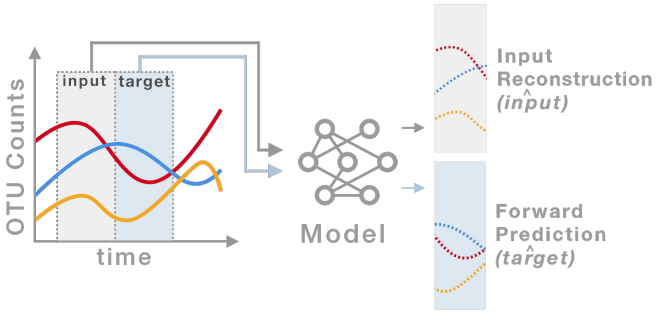

Figure 1 Example schematic of the neural network used to learn the dynamics of a microbial community. Data are sampled in time and supplied to the neural network. Over the course of training, the model learns to both predict forward in time and to reconstruct the supplied input sequence.

Results

Dataset Description and Preprocessing

Stool samples from nine individuals were collected over time periods ranging from months to years and downloaded from the microbiome repository QIITA.[22] Characteristics of the datasets for each donor used in this study are shown in Table 1. Datasets were suitable for modeling after normalization and completion (Materials and Methods).[23] Utilizing matrix completion circumvented problems of zero-counts traditionally encountered when computing OTU tables.[23] This is a necessary step when analyzing microbiome data and especially while training a neural network because erroneous zero-counts act adversarially against true zero-counts in the input data.

Metagenomic amplicon 16S sequence libraries were generated for all samples identically, resulting in metagenomes for 4,113 time points, consisting of between 3,000 and 177,000 OTUs in total. Greater than 90% of sample collection for donors occurred within zero or one days of the previous sampling date (Fig. S1). Fig. S2 highlights the properties of the datasets before and after normalization and completion. We observed that the microbiome composition was consistent across all datasets, containing up to 75% similar OTUs. OTU variance before and after normalization was estimated, finding a Pearson correlation above 0.83, p-value$= 4 \times 10^{-5}$. We found that the completion processes enriched OTUs such as *Stenotrophomonas*, *Holdemania*, and *Lactobacillus* (Fig. S2). Evaluation of typical regression models (linear, quadratic, cubic, and exponential) for all OTUs in each donors showed poor regression coefficients (R^2) below 0.15 (Fig. S1), demonstrating that microbial dynamics could not be predicted using simple regression models.

To reduce the dimensionality of the OTUs in the input data, OTUs were collapsed to the genus level. Collapsed OTUs present in each dataset were sorted by abundance and the top OTUs present across all

donors datasets were selected, resulting in 36 collapsed OTUs in total. This table of collapsed, completed, and normalized OTUs are what will be referred to as OTUs throughout the rest of the manuscript and are used for time series neural networks training and prediction. Only OTUs present in all donors were used because the neural network needs input data of the same shape from each donor (same number of input channels).

Table 1 Metadata about each of the donors included in this study.

Name	Age (years)	Sex	Number of Time Points	Training Data	Validation Data	Test Data	QIITA[22] Study ID
Donor$_0$	36	F	68	✓			1015†
Donor$_1$	33	M	274			✓	1015†
Donor$_2$	26	M	48	✓			1015†
Donor$_3$	63-68	M	145	✓	✓		10283[3]
Donor$_4$	33, 37	F	248	✓	✓		11052†[24]
Donor$_5$	0.3-3	F	680	✓	✓		11052†[24]
Donor$_6$	31-33, 35-38	M	2124	✓	✓		11052†[24]
Donor$_7$	26	M	335	✓	✓		2202[5]
Donor$_8$	36	M	191	✓	✓		2202[5]

*Donor$_4$ is the mother of Donor$_5$.
† Unpublished data.

Input Data Quality Control

The taxonomic assignment for all OTUs is available in Fig. S3. Throughout the remainder of the manuscript, OTUs will simply be referred to by their assigned number instead of the full taxonomic name as seen in Fig. S3. Given the amount of data-preprocessing necessary to make the time series data into the form being supplied to the neural network, we first performed quality control to ensure confidence in the data pipeline. We explored the characteristics of each donor by averaging their OTUs across the time axis yielding a static representation of the microbiota (Fig. 2A).

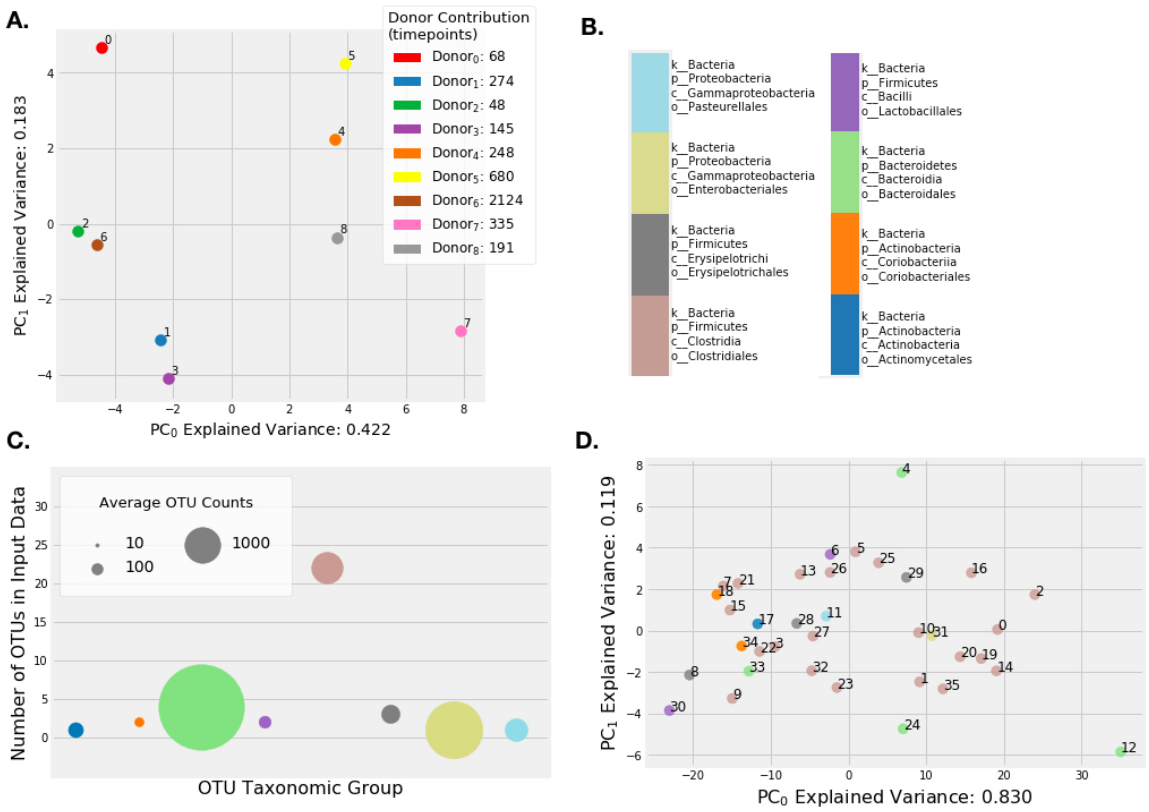

Figure 2 **Static analysis of donor and OTU properites.** All explained variances are listed as fractions, not percentages. **A. PCA of donor's OTUs averaged across time:** Each donor's OTUs plotted in principal coordinates. This gives a measure of similarity of each donor to the others with closer points being more similar. **B. Color-key following taxonomic grouping** This same coloration is used in Figure 3. **C. Bubble plot of OTU group prevalence:** After grouping at the first four taxonomic levels, Clostridia in *Firmicutes* have the most OTUs in the input data, which was also noted during microbial associations analysis in Fig. S4. However, *Bacteriodetes* have the highest average OTU count. **D. PCA performed on OTUs averaged across donors:** These results are the average of Fig. S5 taken across all donors. On average it is clear that certain OTUs, such as numbers 4, 12, and 30 are outliers from the other OTUs (given their position in the principal coordinates).

The results of Fig. 2 serve as a quality control of our data to check for errors in the data-preprocessing pipeline. Donor_0, Donor_4, and Donor_5 were all female and are separated from all the male donors by a clear line in the principal coordinates. Furthermore, Donor_4 and Donor_5 were very similar (Donor_5 is the daughter of Donor_4) (Fig. 2A). Additionally, proximity of Donor_2 and Donor_6 could be explained by the fact that they all traveled abroad together (unpublished data from QIITA).[22]

In Fig. 2B and C the OTUs are grouped at the first four taxonomic levels: kingdom, phylum, class, and order (Fig. 2). Averaged across all donors, many of the OTUs exhibit similar counts (are close to one another in principal coordinates). On the other hand, OTUs 4, 12, and 30 were consistent outliers given that their dynamics are dissimilar from those of other OTUs. Figs. S5 and S6 show how Donor_6 is the primary contributor to OTU 4's placement as an outlier. Of particular note is the fact that the PCA performed here accounts for greater than 96% of all the variance (Fig. 2C). The most prominent

phyla found in these donors are in general agreement with previously reported microbiome compositions (Fig. 2B).[25] Again, these results serve to validate the data processing pipeline implemented here and provide confidence in the data being supplied to the neural network.

Loss Over Training

After validating the OTU time series datasets, these data were used as inputs and targets to train the neural network. A train-validation split of 80:20 was used, splitting along the time axis for each donor. In other words, model was shown the first 80% of the data while learning and the last 20% of the data were held out (used to evaluate the model). For example, Donor_4 has 248 time points in total, so 198 of those were used for training and 50 for validation.

Because the model eventually learned to predict sequences of 30 time points, in order to be used for validation data 20% of a donor's time points needed to be of length 30 or more (the minimum number of total time points for a donor is therefore 150). Four donors' data did not contain enough time points for them to be used as validation data so those donors were only used to train the model (Table 1).

When training a neural network it is common to use a test dataset that the model has never seen to assess the model's generalizability to new data. Given the similarity between Donor_1 and Donor_6, Donor_1 was held-out as a test dataset while all other donors were used to train the model (using the aforementioned training-validation split) (Fig. 2). This decision was made because Donor_6 has the largest number of time points of data and therefore the model will likely be biased toward those data. Donor_1 exhibits less variance than Donor_2 and was therefore deemed an easier test dataset for the model (Fig. S6). Thus, eight of the nine datasets were used to train the neural network.

Here we implement a root mean squared error (RMSE) loss function because it is used for measuring the similarity of a continuous valued OTU estimate to that of a continuous valued OTU target. High values from the loss function imply that the model is less accurate in its predictions. Over the course of training the loss should decrease as the model learns a more appropriate mapping from its inputs to outputs.

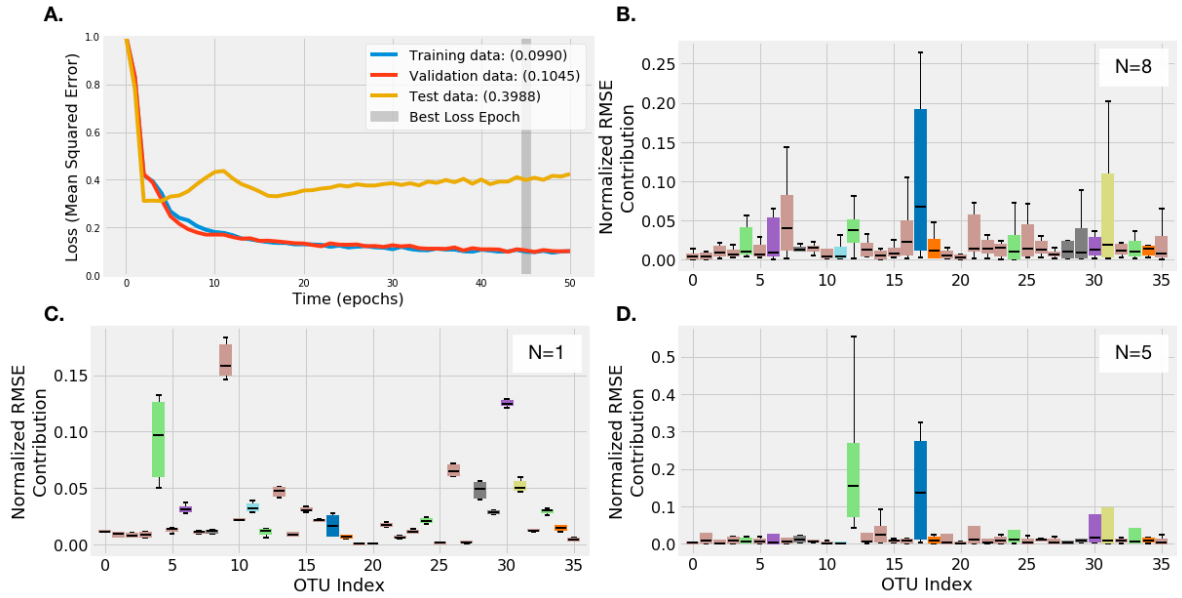

Figure 3 **A. Root Mean Squared Error (RMSE) over training:** As the loss (RMSE) decreases, the model's accuracy of prediction is by definition increasing. The highlighted portion has the lowest validation loss and is therefore considered to be the best version of the model that is used for evaluation. **B, C, D. Individual OTU contribution to loss:** Each bar represents the normalized contribution of each OTU to the overall loss. Higher values indicate less accurate predictions. Coloration is as used in Figure 2B. **B** corresponds to training data, **C** to testing data, and **D** to validation data.

The loss observed between the training and validation data are similar given that they are trained on the same data just at different time points (Fig. 3A). Interestingly, all datasets exhibit high loss on those OTUs that correspond to outliers as seen in Fig. 2A. These OTUs are dissimilar from other OTUs and are therefore harder to predict (see high loss associated with OTUs 4, 9, 12, and 30: Fig. 3, Fig. S7). The training data exhibits the highest variability of prediction accuracy (see large whiskers in Fig. 3B). This can be explained by the fact that the training data has the largest number of donors (eight), which leads to the larger variability (Fig. S7).

Neural Network Predictions

After the model was trained, it was used to generate predictions in time about the abundance and dynamics of a given set of OTUs. As during training, the model was shown (or "primed" with) a sequence of five time points that contain all 36 OTUs. The model then predicts into the future given the context established by the priming sequence.

Figure 4 All predictions here are performed over the validation period of data. Because each donor has a different number of time points of data, the validation data begin at different points in time for each donor. For clarity, only three of the possible 36 OTUs are plotted. **A-F. Model's predictions after five time points of context on validation data:** The model accurately predicts a large variety of different behaviors. **G. Model's predictions after five time points of context on testing data:** Donor_1 is the test set and qualitatively exhibits worse predictions. **H. Effect of priming sequence length on prediction accuracy:** On average the model exhibits similar ability to predict with greater than five time points of primer, but accuracy suffers below five time points (lower y-values are higher accuracy).

The neural network has developed an internal representation of the data allowing it to exhibit both periodic behavior (Fig. 4F) and also accurately predict when certain OTUs will decrease or increase in relative abundance (Fig. 4A). While only three OTUs are plotted in Fig. 4, the model predicted the changes for all 36 OTUs. Large fluctuations in microbe abundance were challenging for the model to

accurately predict (Fig. 4C). The challenge here is that the same OTUs that exhibit slightly oscillatory behavior in one donor can also oscillate wildly in other donors (Fig. 4E, F).

We further tested what is the minimal number of time points of context that the model needs to see in order to make accurate predictions. Figure 4H highlights the relationship between the length of the context sequence being shown to the model and its accuracy of prediction. The model's ability to predict suffers when the priming sequences are shorter than five time points, and appears independent of priming length when shown sequences of length five or above (excluding Donor_3). Throughout the course of training, the model was given five time points of priming data and trained to predict increasingly far into the future. Thus the fact that the model's predictions are less accurate when shown fewer than five time points of priming data is logical. Notably, the model's accuracy improves above five time points of primer when predicting on the test dataset (Fig. 4H blue line). The additional context likely assists in the model's predictive power for data in which it has lower confidence. For Donor_3 the model's accuracy suffers as the length of the context sequence increases beyond five.

Discussion

Here we present a pipeline by which human gut microbiome time series data can be used to predict the composition of the gut microbiome into the future. This same approach can be used with other sampling sites such as mouth, skin, etc. Despite the large number of sequencing runs needed to assemble these datasets, the data processing pipeline appears strong as many of the results are highly logical such as the fact that Donor_5's most similar neighbor is Donor_4 (Figure 2A) and that Donor_5 is Donor_4's daughter.

The model was not trained knowing that each donors' data were distinct from one another, or provided with certain metadata about sex, age, etc.. However, the model still managed to learn an accurate internal representation of the microbial community. The fact that a single model had the flexibility to predict changes across nine different donors supports its utilization as a general model for the gut microbiome. The success of the model lends credence to the fact that there are underlying trends in the data irrespective of donor. Moreover, the model was able to predict changes in microbiome composition through times of sickness and health of the donors without being told explicitly of the host's health state.

Despite the different microbial association networks present in each of the donors (Fig. S4), the model still learns a relationship between the input OTUs that is generalizable across donors. It is of particular note that the model is able to generate accurate predictions for Donor_5 (infant) despite the fact that there is only one example of a child's microbiome in the training set. The success of the model in translating its understanding of adult microbiomes to that of a child lends credence to the fact that this framework is highly scalable to a diverse range of microbial landscapes. The logical next step for

this research is to combine the results of the association graphs with the neural network. Providing the neural network with this type of structured input will allow for implementing a graph convolutional network. This type of network has been very successfully implemented on such structured data.[26]

Significantly, given the ability of this model to forecast changes in the microbiome, it could be used in concert with other machine learning models to predict whether or not an individual will trend toward a diseased state. This is clinically significant as it could allow for preventative measures for dysbiosis related diseases prior to the patient exhibiting symptoms. Furthermore, given this model's ability to generate accurate predictions of microbial consortia, it can be used to supplement other existing microbiome datasets by generating new data. This last point is impactful as it will allow for more robust modeling of disease states with few data samples.

A shortcoming of the model is in its bias toward Donor_6 (Figure 4D), compared to other donors. This is particularly evident when the data are highly oscillatory. This is due to the fact that the network was shown many more examples from Donor_6 given that those data account for over half of all training data shown to the neural network.

Among the OTUs that the model struggles to accurately predict, OTU 17 corresponds to a *Corynebacterium* which has been implicated as a major driver of dysbiosis in mucosal microbial systems.[27] OTU 12 (*Bacteroides*) has been shown to be a dominant organism in the adult human intestinal tract with known roles in digestion.[28] Given the role these microbes play in human microbial systems, further refining the models predictions of them is an important next step. A major drawback in the training of this model is in the volume input data. Typically when training neural networks millions of examples are utilized over the course of training, so the 4,113 time points utilized here are diminutive in comparison. Furthermore, these data are inherently biased as they are either taken from members of the scientific community or those adjacent to it.[22] Thus continuing to collect time series microbiome data from a large variety of individuals in varying states of health is paramount in refining the predictive capabilities of neural network-based models.

The results of using a neural network are highly encouraging because unlike other computational tools, this model does not fit any preexisting ecological models to the data.[9,29] Instead the neural network was allowed to learn for itself from the community composition how the OTUs interact with one another. As more and more time series microbial data become available, the accuracy of these models will likely increase. Unlike traditional predator-prey or differential equation based models, this network requires only observing system dynamics to make predictions. Thus, this same framework is scalable to include new data types in the future.

Materials and Methods

All code for data preprocessing, model generation, and model evaluation can be found in this Github repository:

https://github.com/michaelwiest/microbiome-rnn

Data Acquisition

Datasets for training and evaluating the neural networks were downloaded from QIITA with accession IDs: 550, 1015, 11052, 2202, and 10283.[22] Specifically, deblurred samples with a read length of 100 were selected.

Each of these datasets result in a matrix of shape: [Number of OTUs x Number of timepoints]

Data Preprocessing

Given that the samples taken from QIITA had no taxonomy data associated with them, the QIIME processing tool `assign_taxonomy` was used to populate taxonomy information. All default parameters were used when calling this function.[30,31]

Following taxonomy assignment, each dataset from QIITA was broken out into samples based on the subject and the area sampled (ie, `subject_a_stool`, `subject_b_skin`, etc.). Next, each sample was sorted by the field `collection_timestamp` present in the metadata. These steps can be found in the directory `data_preprocessing/`.

Prior to matrix completion, each OTU was grouped by summing the values according to the assigned taxonomy from the above steps. Specifically the OTUs were grouped up to the level of genus (ignoring species). This helps to reduce the number of input dimension to the model from on the order of 1000s to 100s.

All OTUs (row of data) with fewer than ten non-zero entries across all time points were removed. Next, each OTU was normalized to account for sequencing bias:

$$new\ val = \frac{old\ val \cdot median(column\ sums)}{column\ sum} \tag{1}$$

Where *column sum* is the sum of all of the filtered OTUs at each time point. Following normalization, matrix completion was implemented by DEICODE with parameters: $minval = 0.1$ and $iterations = 100$.[23] Code can be found in `data_preprocessing/filtering_normalization_completion.py`.

In order to reduce the very large number of genera present (order of 100s) only the top OTUs that are present in all donors are kept. This results in 36 OTUs being kept from each donor.

Regression analysis

Regression analysis was performed as attempt to benchmark and categorize the microbes behaviour over the course of time. We normalized the time series abundance data using z-scores. Subsequently, normalized trends were iteratively fitted to linear, quadratic, cubic, exponential, and two-term exponential regression models. We used the coefficient of determination, R^2, as a measure of the quality of the fit. The R^2 was determined by least-minimum squares fitting using MATLAB (The MathWorks Inc.).

Neural Network Training Procedure

All training of the neural networks was done using the Pytorch library for Python.[32]

In all networks Root Mean Squared Error (RMSE) was used as the loss function and the Adam Optimizer was used for model optimization.

Prior to data being fed through the model they are normalized by taking the centered-log-ratio (CLR) and z-scoring. For a given array $\overrightarrow{x}$, a z-scored value is given by:

$$x_i' = \frac{x_i \cdot \overline{x}}{\sigma_x} \tag{2}$$

Where $\overline{x}$ and σ_x are $\overrightarrow{x}$'s mean and standard deviation taken on a per-time point basis, respectively.

Lastly, a train-validation split of 80:20 was used. These data were split along the axis of time. That is the model was shown the first 80% of the data while learning and the last 20% of the data was held out. $Donor_1$ was used as a test dataset while all other donors were used to train the model (using the aforementioned training/validation split).

In other neural network architectures, such as a convolutional neural network, the data are structured in how they are passed to the network. For example, when doing image classification there is an explicit relationship between pixels given that they are spatially near one another. Potential methods for structuring the inputs are graph convolution or structuring based on taxonomy, but neither of those were used in this study.[26] As such, the 36 input channels (one for each input OTU) are not structured.

Encoder-decoder Network

The network that achieved the best results in an attention based encoder-decoder network.[33] This network is based off the results of Srivasta et al. where the model must learn to predict into the future and also recapitulate its input data.[17] In this implementation, the encoder and decoder modules are all long short term memory (LSTM) units.

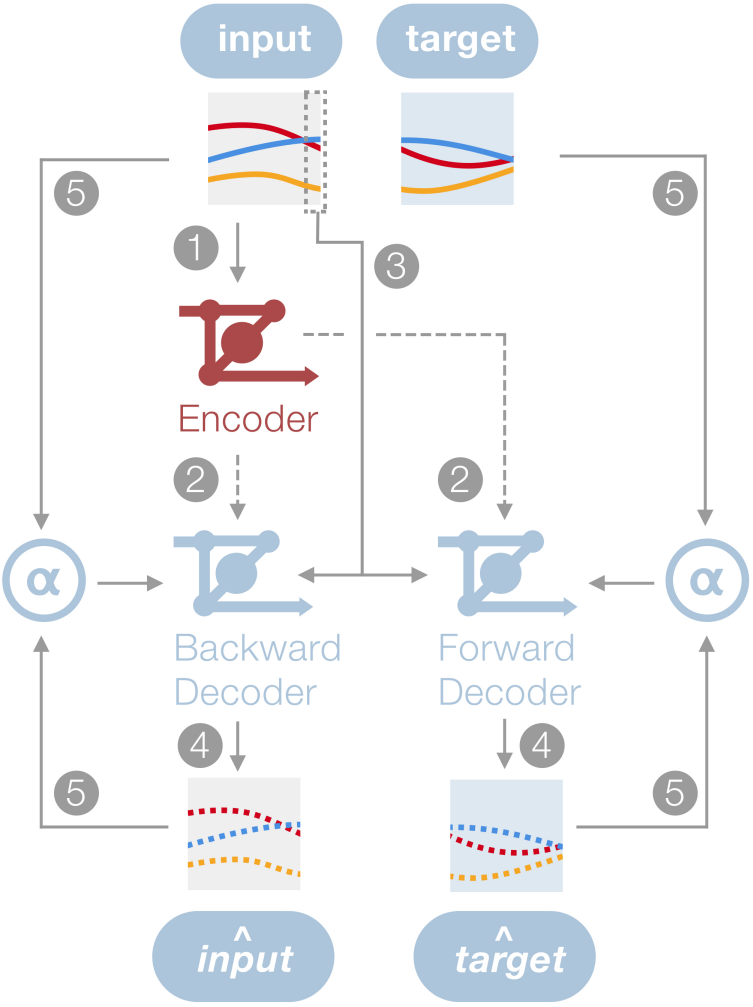

Figure 5 Architecture and training schematic of the encoder decoder. **1)** The input array is passed to the encoder LSTM. **2)** The hidden state of the encoder is copied to the forward and backward decoders. **3)** The last entry of the input sequence is passed to both decoder LSTMs. **4** Each of the decoders generates a new output estimate. **5)** Depending on the teacher forcing fraction (α), either the output of the decoder (from 4) or the actual value is passed to the decoder to generate the next prediction. This process of passing the predicted or actual values to the decoders proceeds until the output sequence is of the desired length. Once the $\hat{input}$ and $\hat{target}$ have been produced then the loss is generated and the error backpropogated through the network.

The input to the neural network are tensors of shape [batch size x 36 x sequence length], where the value 36 is the number of OTUs in the input data.

The training began by passing sequences of length 5 to the network and attempting to predict the following sequence of the same length. After every two epochs the length of the sequence being predicted was increased by one. This is in an effort to train the network to learn gradually longer range interactions among the input strains.

Generally it is believed that teacher forcing, the process of using the desired output to produce the decoders next guess leads to faster convergence than just purely using the output of the decoder;

furthermore some claim that teacher forcing is necessary for the network learning oscillatory behavior (which these data are as seen in Fig. 4).[34]

Evaluation of microbial associations

OTU counts data of gut microbiome of all nine donors have been deployed to compute the microbial association networks. A previously published statistical method, namely SPIEC-EASI[21](Sparse Inverse Covariance Estimation for Ecological Association Inference) has been employed for inference of microbial associations network. This method relies on the transformation of OTU counts data using central log-ratio (clr) that can be defined as the log-ratio of OTU counts and a geometric mean of all OTU counts (1981Aitchison):

$$clr(x) = [log(\frac{x_1}{g(x)}), ..., log(\frac{x_p}{g(x)})]$$

Where $g(x) = [\prod_{i=1}^{p} x_i]^{\frac{1}{p}}$ is the geometric mean a composition vector. x represents a composition vector of an abundance matrix. p denotes the number of OTUs in abundance matrix.Further, this data was used to generate the inverse covariance matrix, which is, in turn, employed to compute the taxon-taxon associations.[21] All microbial associations related calculations have been done using SpiecEasi package[21] on R platform (R version 3.4.4) (https://www.R-project.org/) in Rstudio environment (Version 1.1.453) (http://www.rstudio.com/). SpiecEasi relies on additional R packages like huge[35] and MASS.[36] Cytoscape (version 3.7.1)[37] was used to generate the networks using microbial associations data from SpiecEasi. All above tools have been installed and deployed on Linux platform (Ubuntu 18.04.1 LTS).

Robustness of microbial association networks was estimated as natural connectivity that has been demonstrated previously as a measure of robustness of complex network[38] and stability of microbial communities.[39] It is estimated based on remaining connectivity in the association network after removing each node from the network. Robustness or natural connectivity ($\overline{\lambda}$) can be defined as:

$$\overline{\lambda} = log(\frac{1}{N}\sum_{i=1}^{N} \exp \lambda_i)$$

λ represent the eigenvalue for each node, which can be defined as remaining connectivity in the network after removing each node. N represents a total number of nodes in the network, where each node represents an individual OTU.

References

[1] Andrew B Shreiner, John Y Kao, and Vincent B Young. The gut microbiome in health and in disease. *Current opinion in gastroenterology*, 31(1):69–75, 01 2015.

[2] Les Dethlefsen, Sue Huse, Mitchell L Sogin, and David A Relman. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16s rrna sequencing. *PLOS Biology*, 6(11):1–18, 11 2008.

[3] Xin Fang, Jonathan M. Monk, Sergey Nurk, Margarita Akseshina, Qiyun Zhu, Christopher Gemmell, Connor Gianetto-Hill, Nelly Leung, Richard Szubin, Jon Sanders, Paul L. Beck, Weizhong Li, William J. Sandborn, Scott D. Gray-Owen, Rob Knight, Emma Allen-Vercoe, Bernhard O. Palsson, and Larry Smarr. Metagenomics-based, strain-level analysis of escherichia coli from a time-series of microbiome samples from a crohn's disease patient. *Frontiers in Microbiology*, 9:2559, 2018.

[4] Leonardo Mancabelli, Christian Milani, Gabriele Andrea Lugli, Francesca Turroni, Chiara Ferrario, Douwe van Sinderen, and Marco Ventura. Meta-analysis of the human gut microbiome from urbanized and pre-agricultural populations. *Environmental Microbiology*, 19(4):1379–1390, 2019/03/08 2017.

[5] Lawrence A. David, Arne C. Materna, Jonathan Friedman, Maria I. Campos-Baptista, Matthew C. Blackburn, Allison Perrotta, Susan E. Erdman, and Eric J. Alm. Host lifestyle affects human microbiota on daily timescales. *Genome Biology*, 15(7):R89, Jul 2014.

[6] Aymé Spor, Omry Koren, and Ruth Ley. Unravelling the effects of the environment and host genotype on the gut microbiome. *Nature Reviews Microbiology*, 9:279 EP –, 03 2011.

[7] Amir Bashan, Travis E Gibson, Jonathan Friedman, Vincent J Carey, Scott T Weiss, Elizabeth L Hohmann, and Yang-Yu Liu. Universality of human microbial dynamics. *Nature*, 534(7606):259–262, 06 2016.

[8] Thomas Clavel, Ilias Lagkouvardos, and Andreas Hiergeist. Microbiome sequencing: challenges and opportunities for molecular medicine. *Expert Review of Molecular Diagnostics*, 16(7):795–805, 2016. PMID: 27125906.

[9] Travis E. Gibson and Georg K. Gerber. Robust and scalable models of microbiome dynamics. 2018.

[10] Daniel N. Frank, Allison L. St. Amand, Robert A. Feldman, Edgar C. Boedeker, Noam Harpaz, and Norman R. Pace. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proceedings of the National Academy of Sciences*, 104(34):13780–13785, 2007.

[11] Kristoffer Forslund, Falk Hildebrand, Trine Nielsen, Gwen Falony, Emmanuelle Le Chatelier, Shinichi Sunagawa, Edi Prifti, Sara Vieira-Silva, Valborg Gudmundsdottir, Helle Krogh Pedersen, Manimozhiyan Arumugam, Karsten Kristiansen, Anita Yvonne Voigt, Henrik Vestergaard, Rajna Hercog, Paul Igor Costea, Jens Roat Kultima, Junhua Li, Torben Jørgensen, Florence Levenez, Joël Dore, MetaHIT consortium, H. Bjørn Nielsen, Søren Brunak, Jeroen Raes, Torben Hansen, Jun Wang, S. Dusko Ehrlich, Peer Bork, and Oluf Pedersen. Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. *Nature*, 528:262 EP –, 12 2015.

[12] Qinrui Li, Ying Han, Angel Belle C. Dy, and Randi J. Hagerman. The gut microbiota and autism spectrum disorders. *Frontiers in Cellular Neuroscience*, 11:120, 2017.

[13] Magnus Simrén, Giovanni Barbara, Harry J Flint, Brennan M R Spiegel, Robin C Spiller, Stephen Vanner, Elena F Verdu, Peter J Whorwell, and Erwin G Zoetendal. Intestinal microbiota in functional bowel disorders: a rome foundation report. *Gut*, 62(1):159–176, 01 2013.

[14] Teruaki Nakatsuji, Tiffany H. Chen, Saisindhu Narala, Kimberly A. Chun, Aimee M. Two, Tong Yun, Faiza Shafiq, Paul F. Kotol, Amina Bouslimani, Alexey V. Melnik, Haythem Latif, Ji-Nu Kim, Alexandre Lockhart, Keli Artis, Gloria David, Patricia Taylor, Joanne Streib, Pieter C. Dorrestein, Alex Grier, Steven R. Gill, Karsten Zengler, Tissa R. Hata, Donald Y. M. Leung, and Richard L. Gallo. Antimicrobials from human skin commensal bacteria protect against staphylococcus aureus and are deficient in atopic dermatitis. *Science Translational Medicine*, 9(378), 2017.

[15] Wendy S. Garrett, Graham M. Lord, Shivesh Punit, Geanncarlo Lugo-Villarino, SarkisK. Mazmanian, Susumu Ito, Jonathan N. Glickman, and Laurie H. Glimcher. Communicable ulcerative colitis induced by t-bet deficiency in the innate immune system. *Cell*, 131(1):33 – 45, 2007.

[16] Erik R Dubberke and Margaret A Olsen. Burden of clostridium difficile on the healthcare system. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 55 Suppl 2(Suppl 2):S88–S92, 08 2012.

[17] Nitish Srivastava, Elman Mansimov, and Ruslan Salakhutdinov. Unsupervised learning of video representations using lstms. *CoRR*, abs/1502.04681, 2015.

[18] Katarzyna Janocha and Wojciech Marian Czarnecki. On loss functions for deep neural networks in classification. *CoRR*, abs/1702.05659, 2017.

[19] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4):541–551, Dec 1989.

[20] Zachary Chase Lipton. A critical review of recurrent neural networks for sequence learning. *CoRR*, abs/1506.00019, 2015.

[21] Zachary D. Kurtz, Christian L. Müller, Emily R. Miraldi, Dan R. Littman, Martin J. Blaser, and Richard A. Bonneau. Sparse and Compositionally Robust Inference of Microbial Ecological Networks. *PLoS Computational Biology*, 11(5):1–25, 2015.

[22] Qiita. `http://https://qiita.ucsd.edu`. Accessed: 2018-05-01.

[23] Cameron Martino, James T. Morton, Clarisse A. Marotz, Luke R. Thompson, Anupriya Tripathi, Rob Knight, and Karsten Zengler. A novel sparse compositional technique reveals microbial perturbations. *mSystems*, 4(1), 2019.

[24] J. Gregory Caporaso, Christian L. Lauber, Elizabeth K. Costello, Donna Berg-Lyons, Antonio Gonzalez, Jesse Stombaugh, Dan Knights, Pawel Gajer, Jacques Ravel, Noah Fierer, Jeffrey I. Gordon, and Rob Knight. Moving pictures of the human microbiome. *Genome Biology*, 12(5):R50, May 2011.

[25] The Human Microbiome Project Consortium, Curtis Huttenhower, Dirk Gevers, Rob Knight, Sahar Abubucker, Jonathan H. Badger, Asif T. Chinwalla, Heather H. Creasy, Ashlee M. Earl, Michael G. FitzGerald, Robert S. Fulton, Michelle G. Giglio, Kymberlie Hallsworth-Pepin, Elizabeth A. Lobos, Ramana Madupu, Vincent Magrini, John C. Martin, Makedonka Mitreva, Donna M. Muzny, Erica J. Sodergren, James Versalovic, Aye M. Wollam, Kim C. Worley, Jennifer R. Wortman, Sarah K. Young, Qiandong Zeng, Kjersti M. Aagaard, Olukemi O. Abolude, Emma Allen-Vercoe, Eric J. Alm, Lucia Alvarado, Gary L. Andersen, Scott Anderson, Elizabeth Appelbaum, Harindra M. Arachchi, Gary Armitage, Cesar A. Arze, Tulin Ayvaz, Carl C. Baker, Lisa Begg, Tsegahiwot Belachew, Veena Bhonagiri, Monika Bihan, Martin J. Blaser, Toby Bloom, Vivien Bonazzi, J. Paul Brooks, Gregory A. Buck, Christian J. Buhay, Dana A. Busam, Joseph L. Campbell, Shane R. Canon, Brandi L. Cantarel, Patrick S. G. Chain, I-Min A. Chen, Lei Chen, Shaila Chhibba, Ken Chu, Dawn M. Ciulla, Jose C. Clemente, Sandra W. Clifton, Sean Conlan, Jonathan Crabtree, Mary A. Cutting, Noam J. Davidovics, Catherine C. Davis, Todd Z. DeSantis, Carolyn Deal, Kimberley D. Delehaunty, Floyd E. Dewhirst, Elena Deych, Yan Ding, David J. Dooling, Shannon P. Dugan, Wm Michael Dunne, A. Scott Durkin, Robert C. Edgar, Rachel L. Erlich, Candace N. Farmer, Ruth M. Farrell, Karoline Faust, Michael Feldgarden, Victor M. Felix, Sheila Fisher, Anthony A. Fodor, Larry J. Forney, Leslie Foster, Valentina Di Francesco, Jonathan Friedman, Dennis C. Friedrich, Catrina C. Fronick, Lucinda L. Fulton, Hongyu Gao, Nathalia Garcia, Georgia Giannoukos, Christina Giblin, Maria Y. Giovanni, Jonathan M. Goldberg, Johannes Goll, Antonio Gonzalez, Allison Griggs, Sharvari Gujja, Susan Kinder Haake, Brian J. Haas, Holli A. Hamilton, Emily L. Harris, Theresa A. Hepburn, Brandi Herter, Diane E. Hoffmann, Michael E. Holder, Clinton Howarth, Katherine H. Huang, Susan M. Huse, Jacques Izard, Janet K. Jansson, Huaiyang Jiang, Catherine Jordan, Vandita Joshi, James A. Katancik, Wendy A. Keitel, Scott T. Kelley, Cristyn Kells, Nicholas B. King, Dan Knights, Heidi H. Kong, Omry Koren, Sergey Koren, Karthik C. Kota, Christie L. Kovar, Nikos C. Kyrpides, Patricio S. La Rosa, Sandra L. Lee, Katherine P. Lemon, Niall Lennon, Cecil M. Lewis, Lora Lewis, Ruth E. Ley, Kelvin Li, Konstantinos Liolios, Bo Liu, Yue Liu, Chien-Chi Lo, Catherine A. Lozupone, R. Dwayne Lunsford, Tessa Madden, Anup A. Mahurkar, Peter J. Mannon, Elaine R. Mardis, Victor M. Markowitz, Konstantinos Mavromatis, Jamison M. McCorrison, Daniel McDonald, Jean McEwen, Amy L. McGuire, Pamela McInnes, Teena Mehta, Kathie A. Mihindukulasuriya, Jason R. Miller, Patrick J. Minx, Irene Newsham, Chad Nusbaum, Michelle O'Laughlin, Joshua Orvis, Ioanna Pagani, Krishna Palaniappan, Shital M. Patel,

Matthew Pearson, Jane Peterson, Mircea Podar, Craig Pohl, Katherine S. Pollard, Mihai Pop, Margaret E. Priest, Lita M. Proctor, Xiang Qin, Jeroen Raes, Jacques Ravel, Jeffrey G. Reid, Mina Rho, Rosamond Rhodes, Kevin P. Riehle, Maria C. Rivera, Beltran Rodriguez-Mueller, Yu-Hui Rogers, Matthew C. Ross, Carsten Russ, Ravi K. Sanka, Pamela Sankar, J. Fah Sathirapongsasuti, Jeffery A. Schloss, Patrick D. Schloss, Thomas M. Schmidt, Matthew Scholz, Lynn Schriml, Alyxandria M. Schubert, Nicola Segata, Julia A. Segre, William D. Shannon, Richard R. Sharp, Thomas J. Sharpton, Narmada Shenoy, Nihar U. Sheth, Gina A. Simone, Indresh Singh, Christopher S. Smillie, Jack D. Sobel, Daniel D. Sommer, Paul Spicer, Granger G. Sutton, Sean M. Sykes, Diana G. Tabbaa, Mathangi Thiagarajan, Chad M. Tomlinson, Manolito Torralba, Todd J. Treangen, Rebecca M. Truty, Tatiana A. Vishnivetskaya, Jason Walker, Lu Wang, Zhengyuan Wang, Doyle V. Ward, Wesley Warren, Mark A. Watson, Christopher Wellington, Kris A. Wetterstrand, James R. White, Katarzyna Wilczek-Boney, YuanQing Wu, Kristine M. Wylie, Todd Wylie, Chandri Yandava, Liang Ye, Yuzhen Ye, Shibu Yooseph, Bonnie P. Youmans, Lan Zhang, Yanjiao Zhou, Yiming Zhu, Laurie Zoloth, Jeremy D. Zucker, Bruce W. Birren, Richard A. Gibbs, Sarah K. Highlander, Barbara A. Methé, Karen E. Nelson, Joseph F. Petrosino, George M. Weinstock, Richard K. Wilson, and Owen White. Structure, function and diversity of the healthy human microbiome. *Nature*, 486:207 EP –, 06 2012.

[26] Mikael Henaff, Joan Bruna, and Yann LeCun. Deep convolutional networks on graph-structured data. *CoRR*, abs/1506.05163, 2015.

[27] Nicole A. Abreu, Nabeetha A. Nagalingam, Yuanlin Song, Frederick C. Roediger, Steven D. Pletcher, Andrew N. Goldberg, and Susan V. Lynch. Sinus microbiome diversity depletion and corynebacterium tuberculostearicum enrichment mediates rhinosinusitis. *Science Translational Medicine*, 4(151):151ra124–151ra124, 2012.

[28] Jian Xu, Magnus K. Bjursell, Jason Himrod, Su Deng, Lynn K. Carmichael, Herbert C. Chiang, Lora V. Hooper, and Jeffrey I. Gordon. A genomic view of the human-bacteroides thetaiotaomicron symbiosis. *Science*, 299(5615):2074–2076, 2003.

[29] Feilun Wu, Allison J. Lopatkin, Daniel A. Needs, Charlotte T. Lee, Sayan Mukherjee, and Lingchong You. A unifying framework for interpreting and predicting mutualistic systems. *Nature Communications*, 10(1):242, 2019.

[30] Qiong Wang, George M Garrity, James M Tiedje, and James R Cole. Naïve bayesian classifier for rapid assignment of rrna sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, 73(16):5261–5267, 08 2007.

[31] Daniel McDonald, Morgan N Price, Julia Goodrich, Eric P Nawrocki, Todd Z DeSantis, Alexander Probst, Gary L Andersen, Rob Knight, and Philip Hugenholtz. An improved greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *The ISME Journal*, 6(3):610–618, 03 2012.

[32] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. In *NIPS-W*, 2017.

[33] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *CoRR*, abs/1409.0473, 2014.

[34] K. Doya. Bifurcations in the learning of recurrent neural networks. In *[Proceedings] 1992 IEEE International Symposium on Circuits and Systems*, volume 6, pages 2777–2780 vol.6, May 1992.

[35] Tuo Zhao, Han Liu, Kathryn Roeder, John Lafferty, and Larry Wasserman. The huge Package for High-dimensional Undirected Graph Estimation in {R}. *The Journal of Machine Learning Research*, 13:1059–1062, 2012.

[36] William N. Venables and Brian D. Ripley. *Modern Applied Statistics with S-Plus*. Springer, New York, 4th edition, 2002.

[37] Paul Shannon, Andrew Markiel, Owen Ozier, Nitin S Baliga, Jonathan T Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. Cytoscape : A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, 13:2498–2504, 2003.

[38] WU Jun, Mauricio Barahona, TAN Yue-Jin, and DENG Hong-Zhong. Natural Connectivity of Complex Networks. *Chinease Physics Letters*, 27(7):1–4, 2010.

[39] Laura Tipton, Christian L Müller, Zachary D Kurtz, Laurence Huang, Eric Kleerup, Alison Morris, Richard Bonneau, and Elodie Ghedin. Fungi stabilize connectivity in the lung and skin microbial ecosystems. *Microbiome*, 6(12):1–14, 2018.

[40] Manish Kumar, Parizad Babaei, Boyang Ji, and Jens Nielsen. Human gut microbiota and healthy aging: Recent developments and future prospective. *Nutrition and Healthy Aging*, Preprint(Preprint):1–14, 2016.

Author contributions statement

Author contributions: M.W., C.Z., and K.Z. designed research; M.W., B.A., C.A., and D.A. developed and trained the model; C.Z. and M.K. performed the statistical analysis, and M.W. analyzed data, and M.W. wrote the paper with input of all co-authors.

Additional information

To include, in this order: **Accession codes** (where applicable); **Competing financial interests** (mandatory statement). Karsten Zengler is responsible for submitting a competing financial interests statement on behalf of all authors of the paper.

Supplemental information

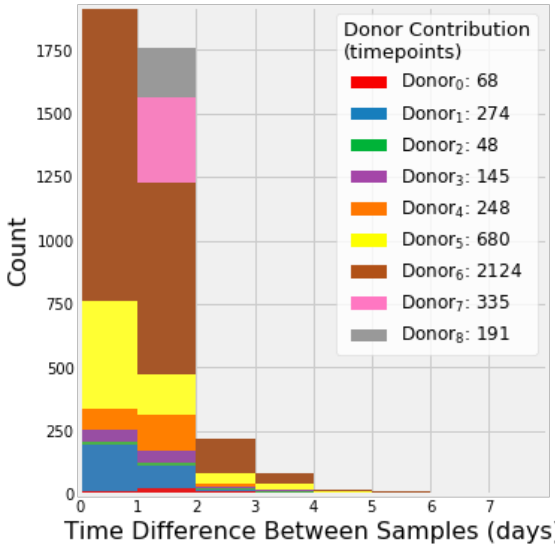

Figure S1 Histogram of the difference in time points between samples. The counts at zero are because the donor sampled him or herself in the same day. Greater than 90% of all sampling was done within one or zero days of the previous sampling day.

1.png 1.png

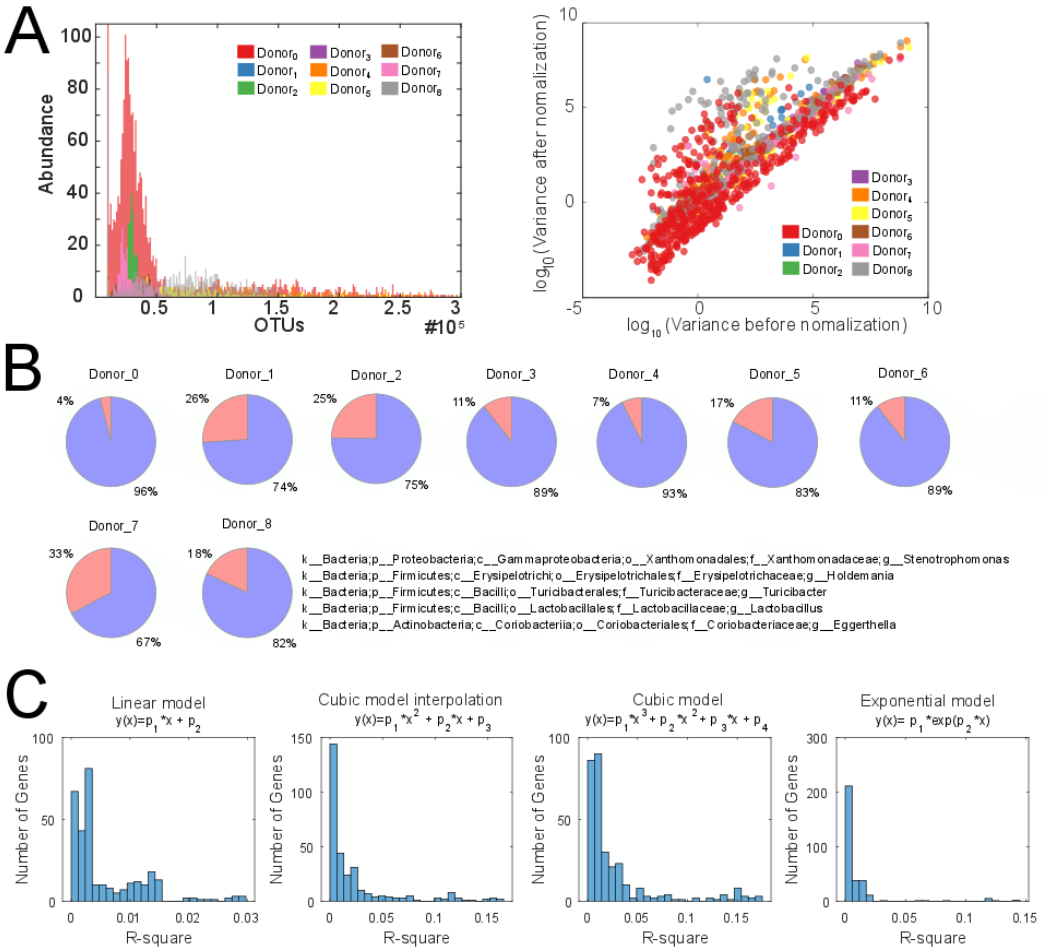

Figure S2 Data preprocessing. (A) Distribution of normalized counts across OTUs. (B) Matrix completion benchmarking and enriched OTUs. Completed and uncompleted datasets were sorted and a sub-sample of 150 OTUs was obtained. The OTUs were compared before and after completion, observing enrichment of 4% up to 33% depending on the donor. (C) Statistical regression analysis. Histograms show all regression coefficients R^2 calculated for the top 62 strains present in all the donors.

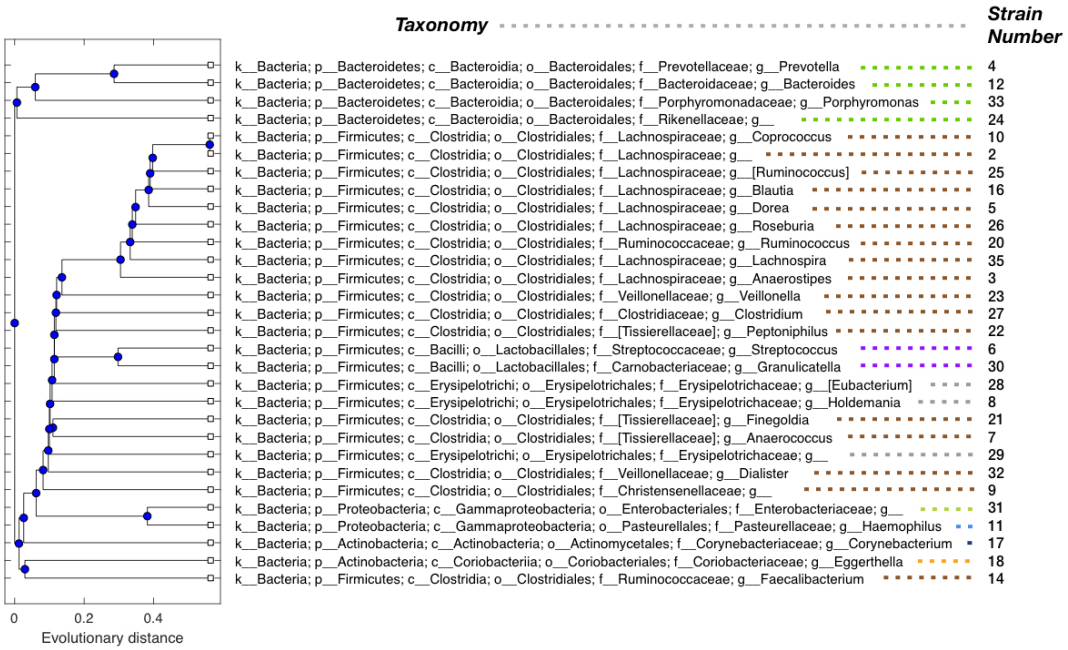

Figure S3 Phylogenetic relationship between input OTUs with available data. The strain number listed on the right is used throughout the manuscript. Six OTUs were unable to be phylogenetically aligned in the tree on the left because of no available data.

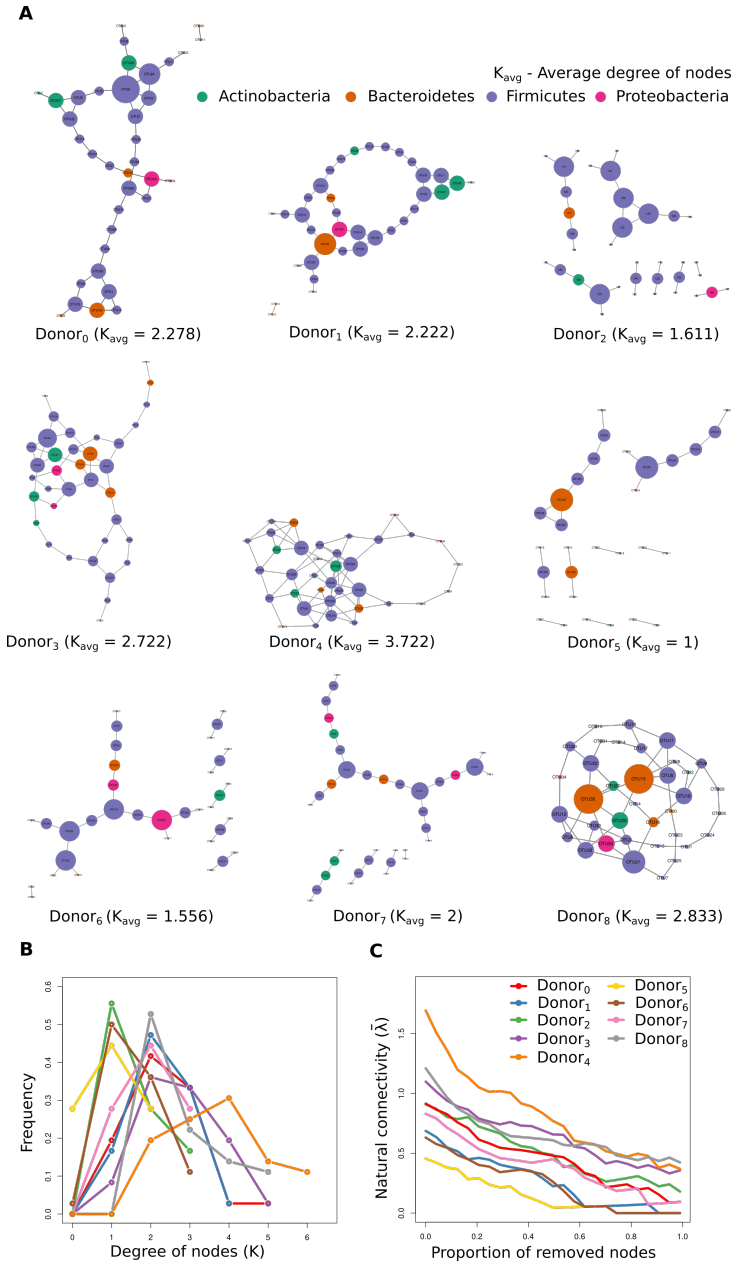

Figure S4 Evaluation of microbial associations **A.** Microbial association networks of the human gut microbiome. OTU count matrices were used to generate microbial association networks for all nine donors. Nodes and edges in the association network represent the individual OTUs (Figure S3) and predicted interactions (Materials and Methods), respectively. The size of each node represents the degree of that node, which is defined as number of edges connected with said node. Nodes have been colored by phyla in the network graphs. Nodes that are spatially close to each other demonstrate stronger associations than those that are further away. **B.** Distribution of degree of nodes (K). **C.** Robustness plots of all nine network graphs. Natural connectivity, or robustness, is considered as a measure of stability of microbial communities.[39] X-axis denotes percentage of removed nodes. Y-axis denotes natural connectivity of remaining network after sequential removal of nodes. Curves are color-coded by different donors as in Panel B. Natural connectivity determined by the remaining connectivity in the network after sequentially removing nodes (Materials and methods). A larger area under the curve represents a more robust network (C).

In addition to predicting the dynamics of individual OTUs over the course of time, we also estimated the associations between different OTUs. Microbial associations were estimated by inferring the association

networks using OTU count matrices (Materials and Methods; Figure S4). Figure S4 shows microbial association networks for all nine donors. In general, this analysis demonstrated that among all donors, *Firmicutes* were the most connected microbes to both themselves and to other phyla. Further, based on average number of association partners (average degree of nodes; K_{avg}) for each OTU, OTUs from $Donor_5$ were the least connected ($K_{avg} = 1$) compared to other donors. This can be explained by the fact that $Donor_5$ is an infant, whereas the others are adults (Table 1). It has been shown by previous studies that a child's gut microbiome is less diverse and less stable than those of adults.[40] Comparing the model's predictions and connectivity of OTUs in the association networks, it was found that the model was able to predict the abundance profiles despite the different OTU connectivity of the donors.

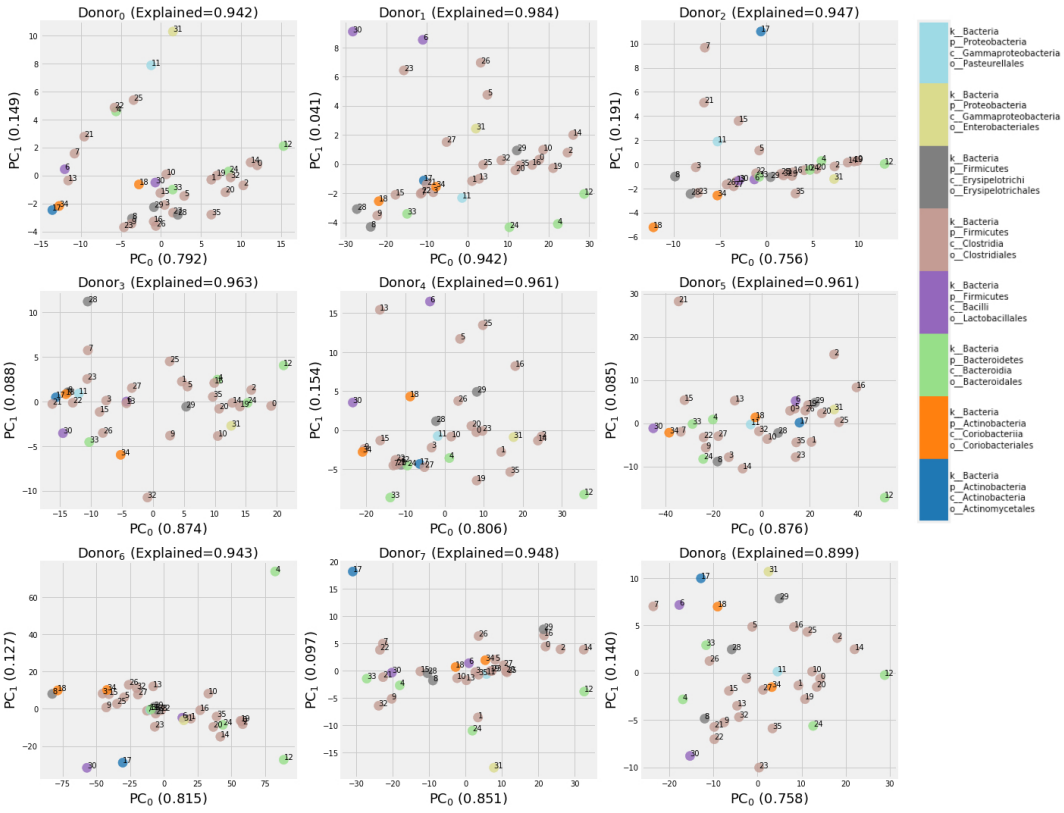

Figure S5 PCA performed across the time-axis for each donors OTUs. The results of Fig. 2C are this figure averaged across donors.

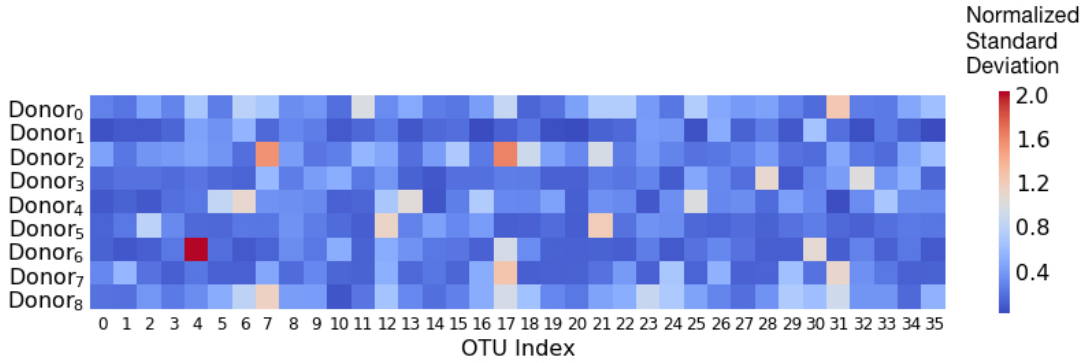

Figure S6 Standard deviation taken across time for each OTU and donor.

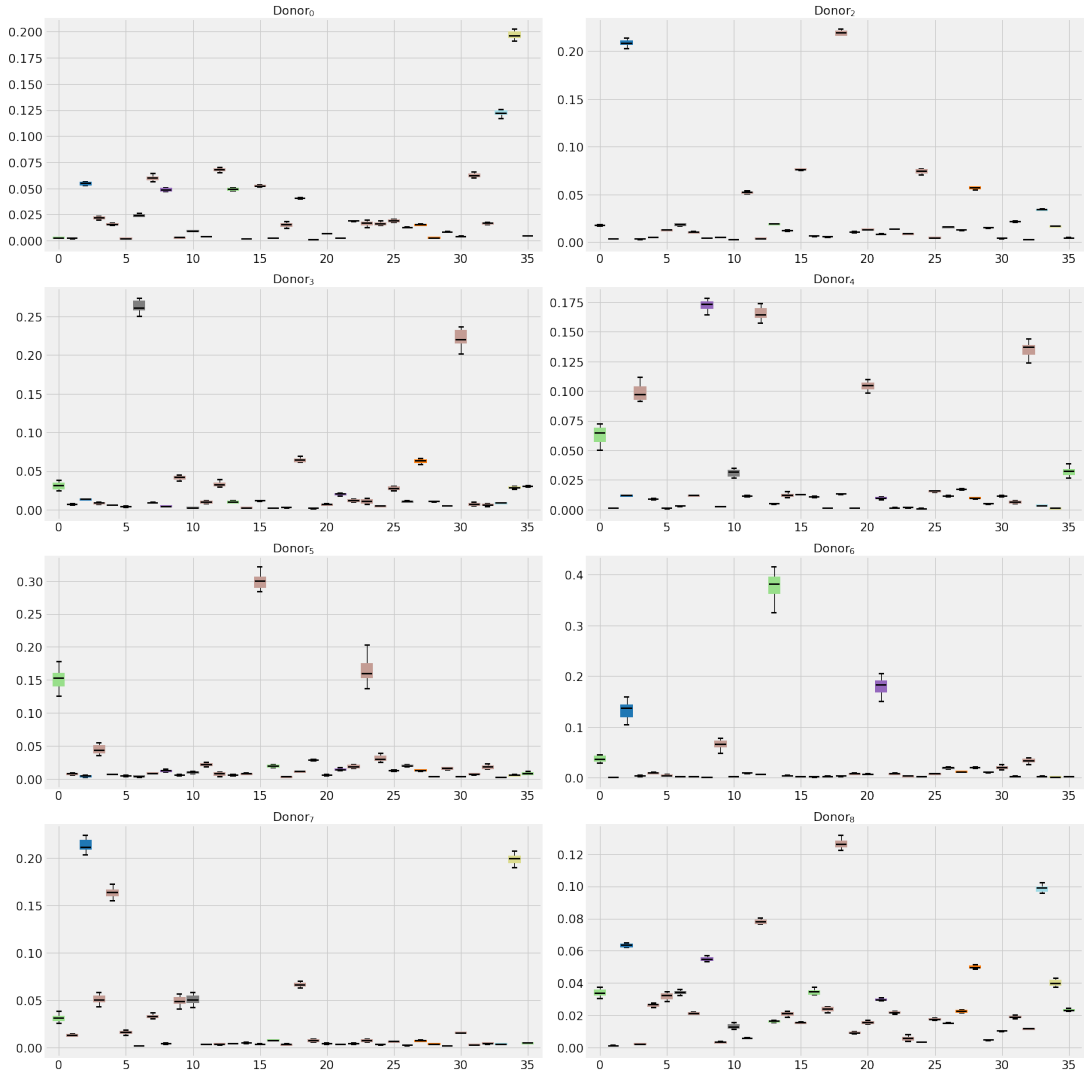

Figure S7 Prediction accuracy by OTU broken out by donor. These results correspond to Figure 3B.