Modeling the temporal dynamics of the gut microbial community in adults and infants

Liat Shenhav[1,5], Ori Furman[2,5], Itzhak Mizrahi[2], and Eran Halperin[1,3,4]

[1]Department of Computer Science, University of California Los Angeles, Los Angeles, CA, USA

[2]Life Sciences, Ben Gurion University

[3]Department of Human Genetics, University of California Los Angeles, Los Angeles, CA, USA

[4]Department of Anesthesiology and Perioperative Medicine, University of California Los Angeles, Los Angeles, CA, USA

[5]Co-first author

Abstract

The increasing appreciation for the human gut microbiome potential roles in diagnosis, treatment, and ultimately prevention of human disease requires an understanding of the temporal variability of the microbial community over the lifespan of an individual. Given the highly dynamic and complex nature of the human gut microbial community, the ability to differentiate between the OTUs behavior over time, and specifically, identify and predict the time-dependent OTUs, is crucial. However, the field has not yet settled upon the degree to which the microbial composition of the gut at a given time affects the microbial composition at a later time. Particularly, it has been recently suggested that only a minority of the OTUs depend on the microbial composition in earlier times. Here, we demonstrate that these earlier findings are severely biased by the sub-optimality of the statistical approaches used to model the microbial composition dependency in time. To that end, we introduce *MTV-LMM* , a linear mixed model method for the prediction of the microbial community temporal dynamics based on the community composition at previous time stamps. *MTV-LMM* thus allows identifying auto-regressive OTUs in time series microbiome datasets, which can be taken to further analysis of the microbiome trajectory over time. We evaluated the performance of *MTV-LMM* on three microbiome time series datasets, and we found that *MTV-LMM* significantly outperforms the existing methods for microbiome time-series modeling. Furthermore, we quantify the time-dependency of the microbiome in infants and adults and we demonstrate, for the first time, that a considerable proportion of OTUs has a significant auto-regressive component, in both infants and adults, and that time-dependency is a considerably more abundant phenomenon compared to previous report.

Introduction

There is an increasing recognition that the human gut microbiome is a contributor to many aspects of human physiology and health including obesity, non-alcoholic fatty liver disease, inflammatory diseases, cancer, metabolic diseases, aging, and neurodegenerative disorders [1–14]. This suggests that the human gut microbiome will play important roles in the diagnosis, treatment, and ultimately prevention of human disease. These applications require an understanding of the temporal variability of the microbiota over the lifespan of an individual, particularly since we now recognize that our microbiota is highly dynamic and that these dynamics are linked to ecological resilience and host health [15–17].

Due to the lack of data and insufficient methodology, we currently have major gaps in our understanding of fundamental issues related to the temporal behavior of the microbiome. Critically, we currently do not have a clear characterization of how and why our gut microbiome varies in time, and whether these dynamics are consistent across humans. It is also unclear whether we can define 'stable' or 'healthy' dynamics, or rather an unhealthy or abnormal time dynamics behavior, which could potentially reflect a health condition or an environmental factor affecting the individual, such as antibiotics exposure or diet. Moreover, the field has not yet settled upon whether gut microbial community structure varies continuously or whether it jumps between discrete community states, and whether these states are shared across individuals [18, 19]. Notably, recent work [20] suggests that human gut microbiome composition is dominated by environmental factors rather than by host genetics, emphasizing the dynamic nature of this ecosystem.

The need for understanding microbiome temporal dynamics and its interaction with host attributes have led to a rise in longitudinal studies that record the temporal variation of microbial communities in a wide range of environments including the human gut microbiome. These time series studies are enabling increasingly comprehensive analyses of the microbiome over time, which are beginning to provide insights into fundamental questions about microbiome dynamics [16, 17, 21].

One of the most fundamental questions that still remains unanswered, is to what degree the

bacterial community in the gut is deterministically determined based on its initial composition. More generally, it is unknown to what degree the microbial composition of the gut at a given time determines the microbial composition at a later time. Additionally, there is only preliminary evidence of long-term effects of early life events on the gut microbial community composition, and it is yet unclear whether these long-term effects traverse through a predefined set of potential trajectories [21, 22].

In order to answer these questions, it is important to quantify the dependency of the microbiome community at a given time based on past information. [23,24] . This task has been previously studied in theoretical settings. Specifically, the Lotka-Volterra models predict changes in community composition through defined species-species or species-resource interaction terms, and are popular for describing internal ecological dynamics such as external factors (e.g., diet and host behavior [17]), species-species interactions (e.g., cross-feeding or successional turnover), and host-species interactions (e.g., immune system regulation or host physiology) [25–27]. Lotka-Volterra models are deterministic and fairly straight-forward to interpret, but little is known about the relative importance of these purely autoregressive factors (a stochastic process in which future values are a weighted sum of past values) in driving gut microbial dynamics.

It has recently been suggested by Gibbons et al. [24] that the human gut microbial community has two dynamic regimes: auto-regressive and non-autoregressive. The auto-regressive regime includes operational taxonomic units (OTUs) that are affected by the community composition in previous time points, while the non auto-regressive regime includes OTUs that are completely stochastic, i.e, their appearance in a specific time does not depend on the past. Gibbons et al. use sparse vector autoregressive models (sVAR) [24] in order to cluster the OTUs into these two regimes. Surprisingly, according to [24], the auto-regressive components of the variance accounted for a minority of the total community variance. Additionally, Ridenhour et al. [28] suggest modeling the OTU read counts along time using Poisson-regression in order to infer the temporal interactions within the microbial community.

In this paper, we show that previous studies substantially underestimate the auto-regressive com-

ponent of the gut microbiome. In order to quantify the dependency of an OTU on past occurrences, we developed a method, Microbial community Temporal Variability Linear Mixed Model (*MTV-LMM*). Linear mixed models have been heavily used in statistical genetics and in genomics, in order to estimate the heritability of a trait [29, 30]. The heritability of a trait can be defined as the amount of variance of the phenotype explained by the genetic information. Analogously and similar to [20], we define the term time-explainability, which quantifies the overall association between the microbiome relative abundance in present time and microbiome composition in the past, after accounting for the association of the individual. Put differently, for each OTU, we quantify the temporal effect - the effect of the community relative abundance in the previous time points on the relative abundance in the current time point.

Our approach has a few notable advantages. First, unlike the sVAR model used in Gibbons et al., [24], our approach models all the individual hosts simultaneously, thus leveraging the information across an entire population, while adjusting for the individual's effect. This provides *MTV-LMM* an improved power to detect temporal dependencies, as well as the ability to quantify the dynamics' consistency across individuals. The Poisson-regression method suggested by Ridenhour et al. [28] is also utilizing the information from all individuals, but without accounting for the individual's effect which can result in an inflated auto-regressive component. Second, *MTV-LMM* can serve both as a feature selection method (selecting only the OTUs affected by time) and as a prediction model. The ability to differentiate between the OTU's behavior over time, and specifically, choosing the time-dependent OTUs, is crucial when fitting a time series model in order to study the microbial community trajectories and specifically, in finding keystone taxa that may be responsible for the temporal changes. Finally, we demonstrate that *MTV-LMM* can serve as a standalone prediction model that outperforms existing prediction methods by an order of magnitude.

We applied *MTV-LMM* to three datasets of temporal gut microbiome (David et al. [17], Caporaso et al. [16], and DABIMMUNE [21]). These datasets contain longitudinal relative abundance OTU data (using 16S rRNA gene sequencing). Our results show that compared to previous approaches, *MTV-LMM* has a substantially improved prediction accuracy of the OTU relative abundance levels

at the next timestamp, indicating that the underlying model used by *MTV-LMM* better matches the underlying biology of the gut microbiome. This can provide insights for future extensions of temporal ecologic models such as the Lotka-Volterra models [31].

Finally, using *MTV-LMM* we demonstrate that unlike previously thought, a considerable proportion of OTUs have non-negligible auto-regressive components, and that auto-regressiveness is a spectrum, that is, there are OTUs that are almost deterministically determined by the previous time points, others that are completely independent of previous time points, and others that have some, but not full dependency on the previous time points.

Results

A brief description of *MTV-LMM* We first informally describe the main idea and utility of *MTV-LMM* . A more comprehensive description can be found in the Methods. *MTV-LMM* is motivated by our assumption that the temporal changes in relative abundance levels, of a specific OTU j, are a time-homogeneous high-order Markov process. *MTV-LMM* model the transitions of this Markov process by fitting a sequential linear mixed model(LMM), in order to predict an OTU relative abundance level at a given time point, given the previous time points. Intuitively, the linear mixed model correlates the similarity between the entire microbial community, across different timestamps, with the similarity of the OTU abundance levels of the next time points. The input to *MTV-LMM* is the microbial community relative abundance at time points $\{1, \ldots, t-1\}$ and its output is the prediction of the relative abundance, for each OTU, at time t. In order to apply linear mixed models, *MTV-LMM* generates a *temporal kinship matrix*, which represents the similarity between every pair of samples across time, where a sample is represented by the vector of the relative abundances of the OTUs at a given time point for a given individual (see Methods). When predicting the relative abundance level of a specific OTU j at time t, the model uses both the global state of the entire microbial community in the last q time points, as well as the relative abundance levels of OTU j in the previous p time points. p and q are parameters determined by

the user, or they can be determined using a cross-validation approach; a more formal description of their role is provided in the Methods.

Model Evaluation. We evaluated the ability of *MTV-LMM* to capture the auto-regressive component of the microbial community using real time series data from three different data sets, each composed of longitudinal relative abundance OTU data. David et al. (2 adult donors - DA, DB - average 250-time points per individual), Caporaso et al. (2 adult donors - M3, F4 - average 231-time points per individual) and DIABIMMUNE data set (39 infant donors, average 28-time points per individual). In these data sets *MTV-LMM* p and q were estimated, using a training set, and are ranging between $0-3$ per OTU (see Methods).

We compared the results of *MTV-LMM* to the approaches that have been used in the past for temporal microbiome modeling, namely the AR(1) model, the sparse vector auto-regression model sVAR [24] model, and the ARIMA $(1,0,0)$ Poisson-regression [28]. Overall, *MTV-LMM* 's prediction accuracy is higher than AR's (Supplementary Table. 1) and significantly outperforms both the sVAR method and the Poisson-regression method across all data sets (Fig. 1). Therefore, we conclude that *MTV-LMM* has a better fit to the data, and thus *MTV-LMM* is more suitable to estimate the auto-regressive component in longitudinal OTU data.

Our estimated fraction of time-explained variance for over 96% of the OTUs are substantially greater than those reported previously using sVAR [24] and ARIMA Poisson regression [28]. The *MTV-LMM* explained, on average, 12.8% of the variance in DIABIMMUNE data set, 12.3% of the variance in Caporaso et al. (2011) and 10.4% David el al. In contrast, sVAR and ARIMA Poisson regression explained on average only 1.783% and 1.95% of the variance in DIABIMMUNE data set respectively, 0.58% and 1.5% of the variance in Caporaso et al. (2011), and 0.28% and 0.4% of the variance in David el al.

Time-explainability as a measure of the auto-regressive component in the microbial community. In order to address the fundamental questions regarding the gut microbiota tem-

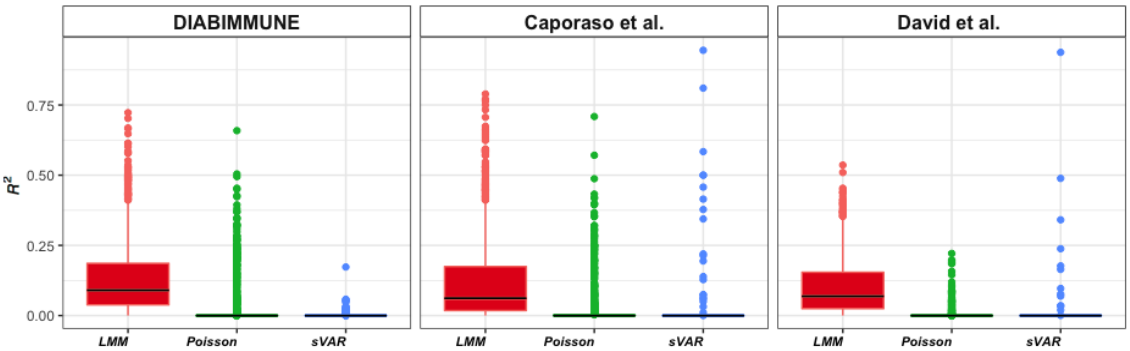

Figure 1. Superior prediction accuracy (R^2) and auto-regressive detection in *MTV-LMM* . The *MTV-LMM* predictions are in red, the predictions of the 'sVAR' and 'ARIMA Poisson-regression' methods are in green and blue respectively.

poral variation, we quantify its auto-regressive component. Namely, we quantify to what degree the relative abundance of different OTUs can be inferred based on the microbiome at previous time points. In statistical genetics, the fraction of phenotypic variance explained by genetic factors is called heritability and is typically evaluated under an LMM framework [29]. Intuitively, linear mixed models estimate heritability by measuring the correlation between the genetic similarity and the phenotypic similarity of pairs of individuals. *MTV-LMM* can be used to define an analogous concept that we term time-explainability, which corresponds to the fraction of OTU variance explained by the microbiome in previous time points. Similarly to the concept of heritability, in time-explainability, the linear mixed models measure the correlation between the similarity of the entire microbial community across different time points and different individuals, with the abundance of the OTU levels at the next time points (see Methods for details).

We next estimated the time-explainability of the OTUs in each data set, using the parameters $q = 1, p = 0$. The resulting model, including only the effect of the microbial community, corresponds to the formula: OTU_t = microbiome community$_{(t-1)}$ + individual effect$_{(t-1)}$ + unknown effects. Notably, time-explainability can be estimated fairly accurately, with an average 95% exact CI width of 23.7%, across all different time-explainability levels and across all data sets (Methods).

Our analysis identified a large portion of the OTUs examined with a statistically significant time-

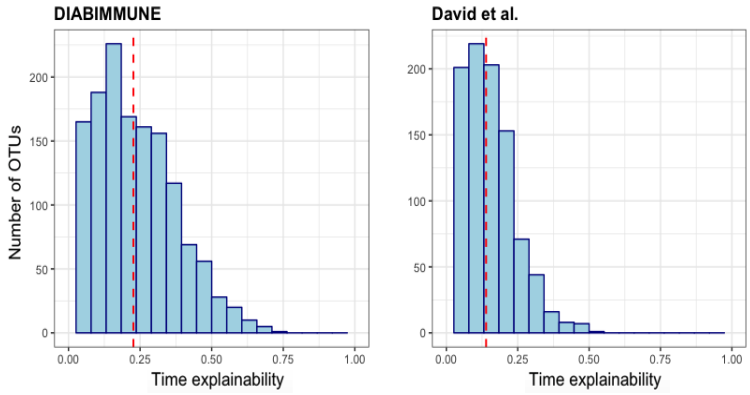

Figure 2. Time-explinability distribution in the DIABIMMUNE infant data set (left) and David et al. adult data set (right). The average time-explainability (denoted by a dashed line) in DIABIMMUNE cohort is 0.23 and in David et al. 0.14.

explainability component across data sets. Specifically, we found that over 85% of the OTUs included in the temporal kinship matrix are significantly explained by the time-explainability component, with estimated time-explainability average levels of 23% in the DIABIMMUNE infant data set (sd = 15%), 21% in Caporaso et al. (2011) data set (sd = 15%) and 14% in David el al. (sd = 10%) (Fig. 2, Supplementary Fig. 7). Notably, the time-explainability estimates, across data sets, are highly correlated to the fraction of variance explained by the predictions R^2 (Supplementary Fig. 8).

Non-autoregressive dynamics contain taxonomic structure. We aggregated the *MTV-LMM* time-explainability by the taxonomic order level, and found that in some orders (*non-autoregressive orders*) all OTUs are non auto-regressive, while in other orders *mixed orders* we observe the presence of both auto-regressive and non auto-regressive OTUs (Fig. 3), where an auto-regressive OTU is defined as one having a statistically significant positive time-explainability component.

Particularly, in the DIABIMMUNE infant data set, there are $7,244$ OTUs which are divided into 55 different orders. However, the OTUs recognized by *MTV-LMM* as auto-regressive (1387 out of 7244) are represented in only 19 orders out of the 55. The remaining 36 orders do not include any

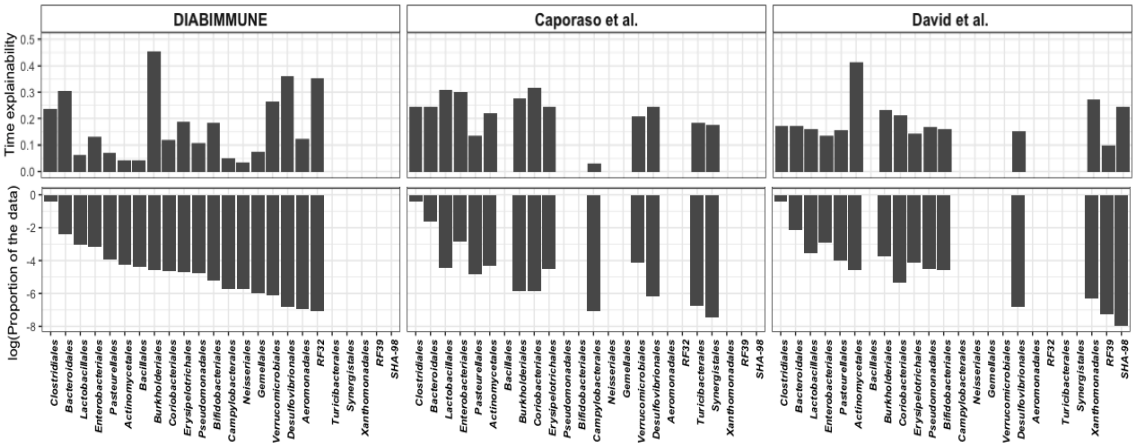

Figure 3. Differential time-explainability aggregated by the taxonomic level 'order' across all data sets. The x-axis is the order among the auto-regressive OTUs across all data sets. In the tow row, the y-axis is the average time-explainability and in the bottom row, the y-axis is the order proportion in log scale.

auto-regressive OTUs. Unlike the auto-regressive dynamics, these non-autoregressive dynamics carry a strong taxonomic signal (P-value = 0), that may indicate a niche/habitat filtering. This observation is consistent with the findings of Gibbons et al. [24], who found a strong phylogenetic structure in the non-autoregressive dynamics in the adult microbiome. We validated this finding using the adult microbiome data sets (Supplementary).

Notably, across all data sets, there is no significant correlation between the order dominance (number of OTUs in the order) and the magnitude of its time-explainability component (median $r = 0.12$). For example, in the DIABIMMUNE data set, the proportion of auto-regressive OTUs within the 19 mixed orders varies between 2% and 75%, where the average is approximately 20%. In the most dominant order (representing 68% of the OTUs), 'Clostridiales', approximately 20% of the OTUs are auto-regressive and the average time-explainability is 23%. In the second most dominant order, 'Bacteroidales', approximately 35% of the OTUs are auto-regressive and the average time-explainability is 31%. Surprisingly, in the rare order 'Bifidobacteriales', approximately 75% of the OTUs are auto-regressive, and the average time-explainability is 19% (Fig. 3).

As an example of the *MTV-LMM* ability to differentiate auto-regressive from non-auto-regressive

OTUs within the same order, we examined the 'Burkholderiales', a relatively rare order (less then 2% of the OTUs in the data), with 76 OTUs overall, where only 19 of which were recognized as auto-regressive by *MTV-LMM* . Indeed, by examining the temporal behavior of each non-auto regressive OTU in this order, we witnessed abrupt changes in the OTU abundance over time, while the maximal number of consecutive time points, in which the relative abundance was > 0 is very small. On the other hand, in the auto-regressive OTUs, we witnessed a consistent behavior over time, while the maximal number of consecutive time points, in which the relative abundance was positive is well over 10 (Supplementary Fig. 9).

Next, we analyzed the adult microbiome (Caporaso et al. and David et al.) and observed that the temporal presence-absence patterns of the OTUs, when divided into families based on the family taxonomic level, are highly associated with the auto-regressive component. Specifically, we found a significant linear correlation between the time-explainability and the family time-prevalence, defined as the median number of time points in which the family OTUs are present ($r = 0.74, 0.41$ respectively). This correlation was not as strong in the DIABIMMUNE infant data set ($r = 0.19$), which exhibit a higher variance in time-prevalence within families, emphasizing the differences between the relatively stable adult microbiome versus the developing microbiome of an infant. Notably, the auto-regressive OTUs, across data sets, are significantly more prevalent over time, in comparison to the non auto-regressive OTUs (p-value $< 10^{-16}$).

Communities with high temporal changes in alpha-diversity are prone to a higher auto-regressive component In our previous analysis, we found that the fraction of auto-regressive OTUs varies across different data sets. In order to shed light on the underlying factors that may lead to temporal dependencies, we compared the auto-regressive components across the different data sets by examining the number of auto-regressive OTUs, the time-explainability of each OTU and the correlation with the measured relative abundance. We find a positive correlation between the time-explainability and the first derivative of the alpha-diversity over time. Specifically, communities with high temporal changes in alpha-diversity are prone to a higher auto-regressive component.

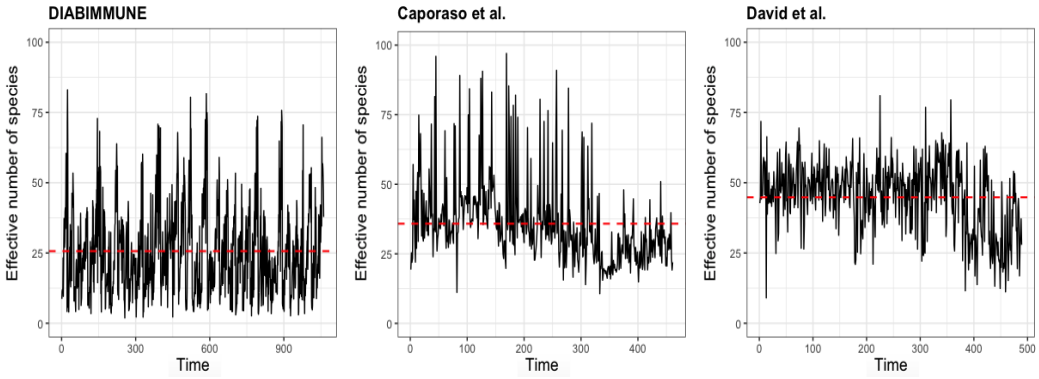

Figure 4. Black lines show the Shannon 'effective number of species' N_{eff} (a measure of alpha diversity) for each data set. The red dashed lines show the average N_{eff} for each data set. The average N_{eff} is between $26 - 45$ across the time series, which is high enough that compositional effects are expected to be negligible [32].

The DIABIMMUNE data set, which is fundamentally different from the other data sets, since it measures the development of the gut microbial community in infants, is characterized by significant temporal changes in alpha-diversity and its estimated time-explainability average levels are 23%. On the other hand, the David et al. data presented with a relatively stable alpha-diversity along time, and with estimated time-explainability average levels of 14%. These findings corroborate and further emphasize the strength of the *MTV-LMM* method in detecting the auto-regressive OTUs.

The auto-regressive component of an adult versus an infant microbiome . The colonization of the human gut begins at birth and is characterized by a succession of microbial consortia [33–36], where the diversity and richness of the microbiota reach adult levels in early childhood. A longitudinal observational study has recently been used to show that infant gut microbiota begins transitioning towards adult communities after weaning [37], implying that one of the most dominant factors associated with temporal variation in an infant microbiome is time. This observation is validated using our infant longitudinal data set (DIABIMMUNE) by applying PCA on the temporal kinship matrix (Fig. 5). Our analysis reveals that the first principal component (accounting for 26% of the overall variability) is associated with time. Specifically, there is a clear clustering of the time samples from the first six months of an infant's life and the

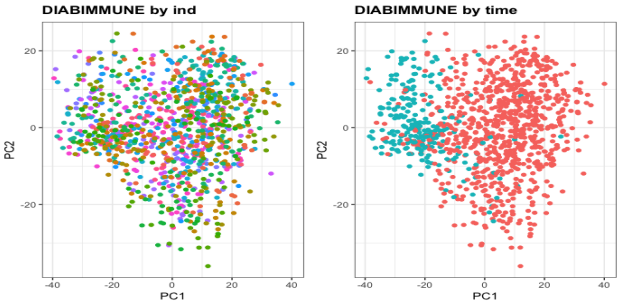

Figure 5. The first two PCs of the temporal kinship matrix color coded by individual (left) and by time - before and after six months (right) in the DIABIMMUNE data (39 infant donors).

rest of the time samples (months $7-36$) which can be correlated to weaning. As expected, we find a strong auto-regressive component in an infant microbiome, which is highly associated with temporal variation across individuals. Using *MTV-LMM* we utilized the high similarity between infants over time, and simultaneously quantify the auto-regressive component in their microbiome.

In contrast to the infant microbiome, the adult microbiome is considered relatively stable [16, 38], with considerable variation in the constituents of the microbial community between individuals. Specifically, it was previously suggested that each individual adult has a unique gut microbial signature [39–41], which is affected, among others factors, by environmental factors [20] and host lifestyle (i.e., antibiotics consumption, high-fat diets [17] etc.). In addition, [17] showed that over the course of one year, differences between individuals were much larger than variation within individuals. This observation was validated in our adult data sets (David et al. and Caporaso et al.) using the same PCA of the temporal kinship matrix which reveals that the first principal component, that accounts for $61,43\%$ of the overall variability respectively, is associated with the individual's identity (Fig. 6, Supplementary Fig. 5).

Using *MTV-LMM* we observe that despite the large similarity along time within adult individuals, there is also a non-negligible auto-regressive component in the adult microbiome, where its mechanism is potentially cognate between individuals. The fraction of variance explained by time across individuals can range from 6% up to 79%. These results shed more light on the temporal behavior of OTUs in adults since as opposed to the temporal behavior in infants that was known

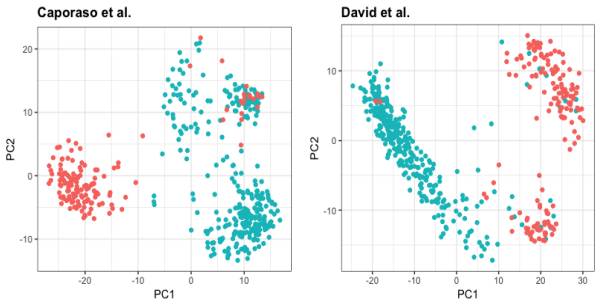

Figure 6. The first two PCs of the temporal kinship matrix color coded by individual in the adults data sets (a) Caporaso et al. (2 adult donors) (b) David et al. (2 adult donors).

to be highly affected by time, to the best of our knowledge, prior to this work in adults there was no evidence for a significant auto-regressive component.

Methods

The *MTV-LMM* Algorithm. *MTV-LMM* uses a linear mixed model (see [42] for a detailed review), a natural extension of standard linear regression, for the prediction of time series data. We describe the technical details of the linear mixed models below.

We assume that the relative abundance levels of a specific OTU j at time point t depend on a linear combination of the relative abundance levels of the OTUs in the microbial community at previous time points. We further assume that the temporal changes in relative abundance levels, of a specific OTU j, are a time-homogeneous high-order Markov process. We model the transitions of this Markov process using a linear mixed model, where we fit the p previous time points of OTU j as fixed effects and the q previous time points of all other OTUs as random effects. p and q are the temporal parameters of the model.

For simplicity of exposition, we present the generative linear mixed model that motivates the approach taken in *MTV-LMM* in two steps. In the first step we model the time series for each OTU in one individual host. In the second step we extend our model to N individuals, while

accounting for the hosts' effect.

We first describe the model assuming there is only one individual. We assume that the relative-abundance levels of m OTUs, denoted as the microbial community, have been measured at T time points. We get as input an $m \times T$ matrix M, where M_{jt} represents the relative-abundance levels of OTU j at time point t. Let $y^j = (M_{j,p+1}, \ldots, M_{jT})^t$ be a $(T-p) \times 1$ vector of OTU j relative abundance, across $T-p$ time points starting at time point $p+1$ and ending at time point T. Let X^j be a $(T-p) \times (p+1)$ matrix of p covariates, comprised of an intercept vector as well as the first p time lags of OTU j (i.e., the relative abundance of OTU j in the p time stamps prior to the one predicted). Formally, for $k = 1$ we have $X^j_{tk} = 1$, and for $p+1 \geq k > 1$ we have $X^j_{tk} = M_{j,t-k+1}$. For simplicity of exposition and to minimize the notation complexity, we assume for now that $p = 1$. Let W be an $(T-q) \times q \cdot m$ standardized relative abundance matrix, representing the first q time lags of the microbial community. For simplicity of exposition we describe the model in the case $q = 1$, and then $W_{tj} = M_{jt}$ (in the more general case, we have $W_{tj} = M_{\lceil j/q \rceil, t-(j \ mod \ q)}$).

With these notations, we assume the following linear model:

$$y^j = X^j \beta^j + W u^j + \epsilon^j, \tag{1}$$

where u^j and ϵ^j are independent random variables distributed as $u \sim N(0_m, \sigma^2_{u^j} I_m)$ and $\epsilon^j \sim N(0_{T-1}, \sigma^2_{\epsilon^j} I_{T-1})$. The parameters of the model are β^j (fixed effects), $\sigma^2_{u^j}$, and $\sigma^2_{\epsilon^j}$.

We note that environmental factors known to be correlated with OTU relative abundance levels (i.e., diet, antibiotic usage [17,20]) can be added to the model as fixed linear effects (i.e., added to the matrix X^j).

Since the relative abundance levels are highly variable across different OTUs, this model is not going to perform well in practice. Intuitively, we would like capture the information as to whether an OTU is present or absent, or potentially introduce a few levels (i.e., high abundance, medium, and low). We use the quantiles of each OTU to transform the matrix M into a matrix $\tilde{M}$, where

$\tilde{M}_{jt} \in \{0, 1, 2\}$ depending whether the abundance level is low (below 25% quantile), medium, or high (above 75% quantile). We subsequently replace the matrix W by a matrix $\tilde{W}$, which is constructed analogously to W, but using $\tilde{M}$ instead of M. Thus, our model can now be described as

$$y^j = X^j \beta^j + \tilde{W} u^j + \epsilon^j \tag{2}$$

So far, we described the model assuming we have time series data from one individual. We next extend the model to the case where time series data is available from multiple individuals. In this case, we assume that the relative-abundance levels of m OTUs, denoted as the microbial community, have been measured at T time points across N individuals. We assume the input consists of N matrices, $M^1, \ldots, M^N$, where matrix M_i corresponds to individual i, and it is of size $m \times T$. Therefore, the outcome vector y^j is now is now an $n \times 1$ vector. For example, when I describe the dimensions of ϵ_j, composed of N blocks, where $n = (T-1)N$, where block i corresponds to the time points of individual i. Formally, $y^j_k = M^{\lceil k/(T-1) \rceil}_{j,(k \ mod \ (T-1))}$. Similarly, we define X^j and $\tilde{W}$ as block matrices, with N different blocks, where each block correspond to individual i.

When applied to multiple individuals, Model (2) may overfit to the individual effects, e.g., due to the host genetics. In other words, since our goal is to model the changes in time, we need to condition these changes in time on the individual effects, that are unwanted confounders for our purposes. We therefore construct a matrix H by randomly permuting the rows of each block matrix i in $\tilde{W}$, where the permutation is conducted only within the same individual. Formally, we apply permutation $\pi_i \in S_{T-1}$ on the rows of each block matrix i, M^i, corresponding to individual i, where S_{T-1} is the set of all permutations of $(T-1)$ elements. In each π_i, we are simultaneously permuting the entire microbial community. Hence, matrix H corresponds to the data of each one of the individuals, but with no information about the time (since the data was shuffled across the different time stamps). With this addition, our final model is given by

$$y^j = X^j \beta^j + \tilde{W} u^j + Hr + \epsilon^j, \tag{3}$$

where $u^j \sim N(0_m, \sigma^2_{u^j} I_m)$ and $\epsilon^j \sim N(0_n, \sigma^2_{\epsilon^j} I_n)$, and $r \sim N(0_m, \sigma^2_r I_m)$. It is easy to verify that an equivalent mathematical representation of model 3 can be given by

$$y^j \sim N(X^j \beta^j, \sigma^2_{AR^j} K_1 + \sigma^2_{ind} K_2 + \sigma^2_{\epsilon^j} I), \tag{4}$$

where $\sigma^2_{AR^j} = m\sigma^2_{u^j}$, $K_1 = \frac{1}{m}\tilde{W}\tilde{W}^T$, $\sigma^2_{ind} = m\sigma^2_r$, $K_2 = \frac{1}{m}HH^T$. We will refer to K_1 as the *temporal kinship matrix*, or the temporal relationship matrix.

We note that for the simplicity of exposition we assumed so far that each sample has the same number of time points T, however in practice the number of samples may vary between the different individuals. It is easy to extend the above model to the case where individual i has T_i time points, however the notations become cumbersome; the implementation of *MTV-LMM* , however takes into account a variable number of time points across the different individuals.

Once the distribution of y^j is specified, one can proceed to estimate the fixed effects β^j and the variance of the random effects using maximum likelihood approaches. Specifically, the common approach for estimating variance components is known as restricted maximum likelihood (REML). We followed the procedure described in the GCTA software package [43], originally developed for genotype data, and adjusted it for the OTU data. GCTA implements the REML method via the average information (AI) algorithm.

Time-explainability We define the term 'time-explainability', denoted as χ, to be the variance of a specific OTU relative abundance explained by the microbial community in the previous time points. Formally, for OTU j we define

$$\chi^j = \frac{\sigma^2_{AR^j}}{\sigma^2_{AR^j} + \sigma^2_{ind} + \sigma^2_{\epsilon^j}}$$

The time-explainability was estimated using GCTA, using the temporal kinship matrix. Confidence intervals were computed using FIESTA [44]. In order to measure the accuracy of time-explainability estimation, the average confidence interval width was estimated by computing the confidence

interval widths for all auto-regressive OTUs and averaging the results. Additionally, we adjust the time-explainability P-values for multiple comparisons using the 'Benjamini-Hochberg' method [45].

Best linear Unbiased Predictor We now turn to the task of predicting y_t^j using the OTU relative abundance in time $t-1$ (or more generally in the last few time points). Using our model notation, we are given x^j and $\tilde{w}$, the covariates associated with a newly observed time point t in OTU j, we would like to predict y_t^j with the greatest possible accuracy. For a simple linear regression model, the answer is simply taking the covariate vector x and multiplying it by the estimated coefficients $\hat{\beta}$: $\hat{y}_t^j = x^T\hat{\beta}$. This practice yields unbiased estimates. However, when attempting prediction in the linear mixed model case, things are not so simple. One could adopt the same approach, but since the effects of the random components are not directly estimated, the vector of covariates $\tilde{w}$ will not contribute directly to the predicted value of y_t^j, and will only affect the variance of the prediction, resulting in an unbiased but inefficient estimate. Instead, one can use the correlation between the realized values of $\tilde{W}u$, to attempt a better guess at the realization of $\tilde{w}u$ for the new sample. This is achieved by computing the distribution of the outcome of the new sample conditional on the full dataset, by using the following property of the multivariate normal distribution. Assume we sampled $t-1$ time points from OTU j, but the relative abundance level for next time point t, y_t^j, is unknown. The conditional distribution of y_t^j given the relative abundance levels at all previous time points, y^j, is given by:

$$y_t^j|y^j \sim N(x^T\beta^j + \Sigma_{t,-t}\Sigma_{-t,-t}^{-1}(y^j - X^j\beta^j), \Sigma_{t,-t}\Sigma_{-t,-t}^{-1}\Sigma_{-t,t}), \tag{5}$$

where $\Sigma = \tilde{W}\tilde{W}^T\sigma_{u^j}^2 + HH^T\sigma_r^2 + I\sigma_{\epsilon^j}^2$ and positive/negative indices indicate the extraction/removal of rows or columns, respectively. Intuitively, we use information from the previous time points that have a high correlation with the new time point, to improve its prediction accuracy. The practice of using the conditional distribution is known as BLUP (Best linear Unbiased Predictor). Therefore, *MTV-LMM* could be used to learn OTU effects in a discovery set (OTU table at time points $1, \ldots, \tilde{t}$), and subsequently use these learned OTU effects to predict the temporal-community contribution in the next time point (OTU j at $\tilde{t}+1$). In our experiments, the initial $\tilde{t}$ was set to

be approximately 1/3 of the time series length in each data set.

Prediction accuracy The predictive ability of a model is commonly assessed using the prediction error variance, $PEV = Var(y^j - \hat{y}^j)$. The proportional reduction in relative abundance variance accounted for by the predictions (referred to as R^2 in this paper) can be quantified using

$$R^2 = \frac{Var(y^j) - Var(\hat{y}^j)}{Var(y^j)} = \frac{Cov(y^j, \hat{y}^j)^2}{Var(y^j)Var(\hat{y}^j)}$$

Model selection Given that the model presented in equation (3) can be extended to any arbitrary p and q, we tested four different variations of this model: 1. $p = 0$ and $q = 1$ (no fixed effect, one random effect based on 1-time lag), 2. $p = 1$ and $q = 1$ (one fixed effect based on 1-time lag, one random effect based on 1-time lag), 3. $p = 0$ and $q = 3$ (no fixed effect, one random effect based on 3-time lags) and 4. $p = 1$ and $q = 3$ (one fixed effect based on 1-time lag, one random effect based on 3-time lags).

We divide each data set to three parts - train, validation, and test, where each part is approximately 1/3 of the time series (sequentially). We train all four models presented above and use the validation set to select a model for each OTU j based on the highest correlation with the real relative abundance. We then compute sequential out-of-sample predictions on the test set with the selected model.

Phylogenetic analysis We performed the following phylogenetic analysis. First, in order to test the hypothesis that both auto-regressive and non auto-regressive dynamics carry a taxonomic signal, we fitted a linear mixed model, where the kinship matrix is the phylogenetic distance between pairs of OTUs and the outcomes are the time-explainability measurements for each OTU. Second, in order to test the hypothesis that only non auto-regressive dynamics carry a taxonomic signal, we conducted a permutation test by shuffling the order 'labels' of all the OTUs, over $100,000$ iterations, where in each iteration we counted the number of non auto-regressive orders.

We calculated an exact P-value for each data set. Lastly, we defined 'time-prevalence' as the median number of time points in which the OTU is present and calculate the linear correlation between the time-explainability and the time-prevalence within each taxonomic rank 'family'.

Alpha diversity measures To measure the alpha diversity, we used Shannon-Wiener index, which is defined as $H = -\sum p_j ln(p_j)$, where p_j is the proportional abundance of species j. Shannon-Wiener index accounts for both abundance and evenness of the species present. Additionally, we computed the 'effective number of species' (also known as true diversity), the number of equally-common species required to give a particular value of an index. The 'effective number of species' associated with a specific Shannon-Wiener index $'a'$ is equal to $exp(a)$.

Preliminary OTU screening according to temporal presence-absence patterns. For the calculation of the temporal kinship matrix we included OTUs using the following inclusion criteria. OTU was included if it was present in at least 0.05% of the time points (removes dominant zero abundance OTUs). In David et al. data we included 1051 (out of 2804), in Caporaso et al. data we 922 (out of 3436) and in the DIABIMMUNE data we included 1440 (out of 7244) OTUs.

Methods comparison We compared *MTV-LMM* to two existing methods: sVAR suggested by [24] and Poisson regression suggested by [28]. In the sVAR method, we followed the procedure described in [24], while running the model and computing the prediction for each individual separately. We then computed an aggregated prediction accuracy score for each OTU, by averaging the prediction accuracy of each individual, using the OTUs the 'survived' the preliminary screening as described above. In the Poisson regression method, we followed the procedure described in [28], while running the model for all the individuals simultaneously and calculating prediction accuracy for each OTU. We used the OTUs that 'survived' the screening suggested in [28] (eliminating any OTUs in the data for which there were a small number (< 6) of average reads per sample). In both models, the train set was 0.67 of the data and the test set was the remaining 0.33 of the data. In both cases we used the code supplied by the authors.

Data sets. We evaluated the performance of *MTV-LMM* using three data sets collected using 16S rRNA gene sequencing. All data sets are publicly available. The first data set was collected and studied by David et al. (2014) [17] (2 adult donors). The next data set was collected and studied by Caporaso et al. (2011) [16] (2 adult donors). The third data set was collected by the 'DIABIMMUNE' project and studied by Yassour et al. (2016 [21] (39 infant donors). In order to compare across studies and reduce technical variance between studies, closed reference Operational Taxonomic Units (OTUs) were clustered at 99% identity against the Greengenes database $v.13_5/8$ [46]. Open reference OTU picking was also run [47], in order to look for non-database OTUs that might contribute substantially to community dynamics. Time series OTU tables were normalized by random sub-sampling to contain $10,000$ reads per sample.

David et al. (2014) data set [17]. Stool samples from 2 healthy American adults were collected (donor A = DA and donor B = DB). DA collected gut microbiota samples between days 0 and 364 of the study (total 311 samples). DB primarily collected gut microbiota samples between study days 0 and 252 (total 180 samples). The V4 region of the 16S ribosomal RNA gene subunit was used to identify bacteria in a culture-independent manner. DNA was amplified using custom barcoded primers and sequenced with paired-end 100 bp reads on an Illumina GAIIx according to a previously published protocol [48]. 'OTU picking' and 'quality control' were performed essentially as described [17]. In this work, we used the OTUs shared across donors ($2,804$ OTUs).

Caporaso et al. (2011) data set [16]. Two healthy American adults, one male (M3) and one female (F4), were sampled daily at three body sites (gut (feces), mouth, and skin (left and right palms)). M3 was sampled for 15 months (total 332 samples) and F4 for 6 months (total 131 samples). Variable region 4 (V4) of 16S rRNA genes present in each community sample were amplified by PCR and subjected to multiplex sequencing on an Illumina Genome Analyzer IIx according to a previously published protocol [48]. 'OTU picking' and 'quality control' were performed essentially as described [16]. In this work, we used the OTUs shared across donors ($3,436$ OTUs).

DIABIMMUNE data set [21]. Monthly stool samples from 39 Finnish children aged 2 to 36 months (total of 1101 samples). To analyze the composition of the microbial communities in this cohort,

DNA from stool samples was isolated and amplified and sequenced the V4 region of the 16S rRNA gene. Sequences were sorted into operational taxonomic units (OTUs). 16S rRNA gene sequencing was performed essentially as previously described in [21]. In this work, we used all the OTUs in the sample ($7,244$ OTUs).

Discussion

We have presented *MTV-LMM* , a method for the analysis and modeling of time series microbial community data. *MTV-LMM* can be used to predict the relative abundance levels of OTUs in the future, and it can be used to assess the time-explainability of each of the OTUs, i.e., how much of the variability of the OTU relative abundance levels can be explained based on the past microbiome community structure. The time-explainability can be critical for selecting auto-regressive OTUs that are essential to our understanding of the microbiome behavior in longitudinal studies. Particularly, such OTUs could be used to characterize temporal trajectories of the microbiome community. This provides a stepping stone towards better understanding of basic ecology microbial community temporal dynamics as well as the use of microbiome data as a detection and diagnostic tool in various fields.

We have demonstrated that *MTV-LMM* significantly outperforms existing approaches for temporal modeling of the microbiome, both in terms of its prediction accuracy, as well as in its ability to identify time-dependent OTUs. We hypothesize that this improved performance stems from the fact that unlike previous approaches, *MTV-LMM* leverages the information across an entire population, while adjusting for the individual's effect. In addition, *MTV-LMM* uses linear mixed models which have been shown to be useful in other genomic contexts (particularly in genetics).

MTV-LMM is a flexible and computationally efficient tool, which can be easily adapted by researchers in order to select the core time-dependent OTUs, quantify their temporal effect given the microbial community and predict their future relative abundance. *MTV-LMM* can also be used to

estimate the strength of ecological interactions within the microbial community.

Using *MTV-LMM* we have demonstrated that auto-regressiveness is a spectrum, where some OTUs are almost deterministically determined by the community previous time points, others that have some, but not full dependency on the previous time points and lastly, some OTUs that are completely independent of previous time points. We further show that these auto-regressive characterization is highly related to alpha diversity and to the phylogenetic structure of the OTUs. Specifically, the phylogenetic orders harboring the auto-regressive OTUs are of mixed nature, containing both auto-regressive and non-autoregressive OTUs in differing proportions (Fig. 3). The ability of *MTV-LMM* to differentiate between OTUs within the same order is crucial in finding keystone taxa that may be responsible for the temporal changes observed in the different data sets.

By applying *MTV-LMM* to three datasets of temporal gut microbiome [16, 17, 21], we found that, unlike previously thought, a considerable proportion of OTUs have a non-negligible auto-regressive component, in both infants and adults. For the adult microbiome our results provide such evidence for the first time, and as we show previous approaches miss this observation as they underestimate the time-explainability of the OTUs. This highlights the sensitivity of our method and its ability to cope with both developing and stable datasets, as well as opens the possibility for microbiome prediction and potential manipulation in adult individuals.

As shown in the results section (Fig. 5), when applying PCA on the temporal kinship matrix in the infants' dataset, our analysis revealed a clear clustering separating the time samples from the first six months of an infant's life from the rest of the time samples (7-36 months). We infer that a major dietary change was the cause for this temporal clustering, most likely weaning. Indeed, our results correspond to a recent longitudinal observational study showing that infant gut microbiota begins transitioning towards adult communities after weaning, roughly at the age of six months [37]. Furthermore, our results are in agreement with the ones presented by [22] in relation to dietary effects, which validates the ability of *MTV-LMM* to detect the relevant OTUs affected by dietary changes in the developing microbiome.

It is important to realize that the time-explainability of a given OTU may strongly depend on the density of the sampling. For example, if the microbiome is sampled every two months as opposed to every month, the time-explainability may be reduced since more noise is added over time. The instrumental novelty of our method is the statistical power gained from the overall community behavior, as well as all the individuals in the sample, for the purpose of predicting a specific OTU temporal behavior . This suggests that internal interactions within the microbiome are of major importance in modulating individual microbes behaviour over time.

References

[1] Peter J Turnbaugh, Micah Hamady, Tanya Yatsunenko, Brandi L Cantarel, Alexis Duncan, Ruth E Ley, Mitchell L Sogin, William J Jones, Bruce A Roe, Jason P Affourtit, et al. A core gut microbiome in obese and lean twins. *nature*, 457(7228):480, 2009.

[2] Peter J Turnbaugh, Ruth E Ley, Michael A Mahowald, Vincent Magrini, Elaine R Mardis, and Jeffrey I Gordon. An obesity-associated gut microbiome with increased capacity for energy harvest. *nature*, 444(7122):1027–131, 2006.

[3] Ruth E Ley. Obesity and the human microbiome. *Current opinion in gastroenterology*, 26(1):5–11, 2010.

[4] Peter J Turnbaugh, Fredrik Bäckhed, Lucinda Fulton, and Jeffrey I Gordon. Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. *Cell host & microbe*, 3(4):213–223, 2008.

[5] Ruth E Ley, Fredrik Bäckhed, Peter Turnbaugh, Catherine A Lozupone, Robin D Knight, and Jeffrey I Gordon. Obesity alters gut microbial ecology. *Proceedings of the National Academy of Sciences of the United States of America*, 102(31):11070–11075, 2005.

[6] Omry Koren, Aymé Spor, Jenny Felin, Frida Fåk, Jesse Stombaugh, Valentina Tremaroli, Carl Johan Behre, Rob Knight, Björn Fagerberg, Ruth E Ley, et al. Human oral, gut, and plaque microbiota in patients with atherosclerosis. *Proceedings of the National Academy of Sciences*, 108(Supplement 1):4592–4598, 2011.

[7] Jose C Clemente, Luke K Ursell, Laura Wegener Parfrey, and Rob Knight. The impact of the gut microbiota on human health: an integrative view. *Cell*, 148(6):1258–1270, 2012.

[8] Vanessa K Ridaura, Jeremiah J Faith, Federico E Rey, Jiye Cheng, Alexis E Duncan, Andrew L Kau, Nicholas W Griffin, Vincent Lombard, Bernard Henrissat, James R Bain, et al.

Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science*, 341(6150):1241214, 2013.

[9] Emmanuelle Le Chatelier, Trine Nielsen, Junjie Qin, Edi Prifti, Falk Hildebrand, Gwen Falony, Mathieu Almeida, Manimozhiyan Arumugam, Jean-Michel Batto, Sean Kennedy, et al. Richness of human gut microbiome correlates with metabolic markers. *Nature*, 500(7464):541–546, 2013.

[10] Siobhan F Clarke, Eileen F Murphy, Kanishka Nilaweera, Paul R Ross, Fergus Shanahan, Paul W O'Toole, and Paul D Cotter. The gut microbiota and its relationship to diet and obesity: new insights. *Gut microbes*, 3(3):186–202, 2012.

[11] Ian B Jeffery, Eamonn MM Quigley, Lena Öhman, Magnus Simrén, and Paul W O'toole. The microbiota link to irritable bowel syndrome: an emerging story. *Gut Microbes*, 3(6):572–576, 2012.

[12] Julian R Marchesi, Bas E Dutilh, Neil Hall, Wilbert HM Peters, Rian Roelofs, Annemarie Boleij, and Harold Tjalsma. Towards the human colorectal cancer microbiome. *PloS one*, 6(5):e20447, 2011.

[13] Andrew L Kau, Philip P Ahern, Nicholas W Griffin, Andrew L Goodman, and Jeffrey I Gordon. Human nutrition, the gut microbiome, and immune system: envisioning the future. *Nature*, 474(7351):327, 2011.

[14] Junjie Qin, Yingrui Li, Zhiming Cai, Shenghui Li, Jianfeng Zhu, Fan Zhang, Suisha Liang, Wenwei Zhang, Yuanlin Guan, Dongqian Shen, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature*, 490(7418):55–60, 2012.

[15] Georg K Gerber. The dynamic microbiome. *FEBS letters*, 588(22):4131–4139, 2014.

[16] J Gregory Caporaso, Christian L Lauber, Elizabeth K Costello, Donna Berg-Lyons, Antonio Gonzalez, Jesse Stombaugh, Dan Knights, Pawel Gajer, Jacques Ravel, Noah Fierer, et al.

Moving pictures of the human microbiome. *Genome biology*, 12(5):R50, 2011.

[17] 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, 2014.

[18] Manimozhiyan Arumugam, Jeroen Raes, Eric Pelletier, Denis Le Paslier, Takuji Yamada, Daniel R Mende, Gabriel R Fernandes, Julien Tap, Thomas Bruls, Jean-Michel Batto, et al. Enterotypes of the human gut microbiome. *nature*, 473(7346):174, 2011.

[19] Dan Knights, Tonya L Ward, Christopher E McKinlay, Hannah Miller, Antonio Gonzalez, Daniel McDonald, and Rob Knight. Rethinking "enterotypes". *Cell host & microbe*, 16(4):433–437, 2014.

[20] Daphna Rothschild, Omer Weissbrod, Elad Barkan, Tal Korem, David Zeevi, Paul Igor Costea, Anastasia Godneva, Iris Nati Kalka, Noam Bar, Niv Zmora, et al. Environmental factors dominate over host genetics in shaping human gut microbiota composition. *bioRxiv*, page 150540, 2017.

[21] Moran Yassour, Tommi Vatanen, Heli Siljander, Anu-Maaria Hämäläinen, Taina Härkönen, Samppa J Ryhänen, Eric A Franzosa, Hera Vlamakis, Curtis Huttenhower, Dirk Gevers, et al. Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability. *Science translational medicine*, 8(343):343ra81–343ra81, 2016.

[22] Fredrik Bäckhed, Josefine Roswall, Yangqing Peng, Qiang Feng, Huijue Jia, Petia Kovatcheva-Datchary, Yin Li, Yan Xia, Hailiang Xie, Huanzi Zhong, et al. Dynamics and stabilization of the human gut microbiome during the first year of life. *Cell host & microbe*, 17(5):690–703, 2015.

[23] Tadashi Fukami. Historical contingency in community assembly: integrating niches, species pools, and priority effects. *Annual Review of Ecology, Evolution, and Systematics*, 46, 2015.

[24] Sean M Gibbons, Sean M Kearney, Chris S Smillie, and Eric J Alm. Two dynamic regimes in the human gut microbiome. *PLoS computational biology*, 13(2):e1005364, 2017.

[25] Shiri Freilich, Raphy Zarecki, Omer Eilam, Ella Shtifman Segal, Christopher S Henry, Martin Kupiec, Uri Gophna, Roded Sharan, and Eytan Ruppin. Competitive and cooperative metabolic interactions in bacterial communities. *Nature communications*, 2:589, 2011.

[26] Daniel A Peterson, Nathan P McNulty, Janaki L Guruge, and Jeffrey I Gordon. Iga response to symbiotic bacteria as a mediator of gut homeostasis. *Cell host & microbe*, 2(5):328–339, 2007.

[27] Harry J Flint, Sylvia H Duncan, Karen P Scott, and Petra Louis. Interactions and competition within the microbial community of the human colon: links between diet and health. *Environmental microbiology*, 9(5):1101–1111, 2007.

[28] Benjamin J Ridenhour, Sarah L Brooker, Janet E Williams, James T Van Leuven, Aaron W Miller, M Denise Dearing, and Christopher H Remien. Modeling time-series data from microbial communities. *bioRxiv*, page 071449, 2016.

[29] Jian Yang, Beben Benyamin, Brian P McEvoy, Scott Gordon, Anjali K Henders, Dale R Nyholt, Pamela A Madden, Andrew C Heath, Nicholas G Martin, Grant W Montgomery, et al. Common snps explain a large proportion of the heritability for human height. *Nature genetics*, 42(7):565–569, 2010.

[30] Peter M Visscher, Gibran Hemani, Anna AE Vinkhuyzen, Guo-Bo Chen, Sang Hong Lee, Naomi R Wray, Michael E Goddard, and Jian Yang. Statistical power to detect genetic (co) variance of complex traits using snp data in unrelated samples. *PLoS genetics*, 10(4):e1004269, 2014.

[31] Alan A Berryman. The orgins and evolution of predator-prey theory. *Ecology*, 73(5):1530–1535, 1992.

[32] Jonathan Friedman and Eric J Alm. Inferring correlation networks from genomic survey data. *PLoS computational biology*, 8(9):e1002687, 2012.

[33] Carlotta De Filippo, Duccio Cavalieri, Monica Di Paola, Matteo Ramazzotti, Jean Baptiste Poullet, Sebastien Massart, Silvia Collini, Giuseppe Pieraccini, and Paolo Lionetti. Impact of diet in shaping gut microbiota revealed by a comparative study in children from europe and rural africa. *Proceedings of the National Academy of Sciences*, 107(33):14691–14696, 2010.

[34] Meghan B Azad, Theodore Konya, Heather Maughan, David S Guttman, Catherine J Field, Radha S Chari, Malcolm R Sears, Allan B Becker, James A Scott, and Anita L Kozyrskyj. Gut microbiota of healthy canadian infants: profiles by mode of delivery and infant diet at 4 months. *Canadian Medical Association Journal*, 185(5):385–394, 2013.

[35] Ilseung Cho and Martin J Blaser. The human microbiome: at the interface of health and disease. *Nature reviews. Genetics*, 13(4):260, 2012.

[36] Noel T Mueller, Elizabeth Bakacs, Joan Combellick, Zoya Grigoryan, and Maria G Dominguez-Bello. The infant microbiome development: mom matters. *Trends in molecular medicine*, 21(2):109–117, 2015.

[37] Jeremy E Koenig, Aymé Spor, Nicholas Scalfone, Ashwana D Fricker, Jesse Stombaugh, Rob Knight, Largus T Angenent, and Ruth E Ley. Succession of microbial consortia in the developing infant gut microbiome. *Proceedings of the National Academy of Sciences*, 108(Supplement 1):4578–4585, 2011.

[38] Dionysios A Antonopoulos, Susan M Huse, Hilary G Morrison, Thomas M Schmidt, Mitchell L Sogin, and Vincent B Young. Reproducible community dynamics of the gastrointestinal microbiota following antibiotic perturbation. *Infection and immunity*, 77(6):2367–2375, 2009.

[39] Simon Lax, Daniel P Smith, Jarrad Hampton-Marcell, Sarah M Owens, Kim M Handley, Nicole M Scott, Sean M Gibbons, Peter Larsen, Benjamin D Shogan, Sophie Weiss, et al.

Longitudinal analysis of microbial interaction between humans and the indoor environment. *Science*, 345(6200):1048–1052, 2014.

[40] Noah Fierer, Christian L Lauber, Nick Zhou, Daniel McDonald, Elizabeth K Costello, and Rob Knight. Forensic identification using skin bacterial communities. *Proceedings of the National Academy of Sciences*, 107(14):6477–6481, 2010.

[41] Eric A Franzosa, Katherine Huang, James F Meadow, Dirk Gevers, Katherine P Lemon, Brendan JM Bohannan, and Curtis Huttenhower. Identifying personal microbiomes using metagenomic codes. *Proceedings of the National Academy of Sciences*, 112(22):E2930–E2938, 2015.

[42] Shayle R Searle, George Casella, and Charles E McCulloch. *Variance components*, volume 391. John Wiley & Sons, 2009.

[43] Jian Yang, S Hong Lee, Michael E Goddard, and Peter M Visscher. Gcta: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics*, 88(1):76–82, 2011.

[44] Regev Schweiger, Eyal Fisher, Elior Rahmani, Liat Shenhav, Saharon Rosset, and Eran Halperin. Using stochastic approximation techniques to efficiently construct confidence intervals for heritability. In *RECOMB*, pages 241–256, 2017.

[45] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, pages 289–300, 1995.

[46] 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, 2012.

[47] Jai Ram Rideout, Yan He, Jose A Navas-Molina, William A Walters, Luke K Ursell, Sean M

Gibbons, John Chase, Daniel McDonald, Antonio Gonzalez, Adam Robbins-Pianka, et al. Subsampled open-reference clustering creates consistent, comprehensive otu definitions and scales to billions of sequences. *PeerJ*, 2:e545, 2014.

[48] J Gregory Caporaso, Justin Kuczynski, Jesse Stombaugh, Kyle Bittinger, Frederic D Bushman, Elizabeth K Costello, Noah Fierer, Antonio Gonzalez Peña, Julia K Goodrich, Jeffrey I Gordon, et al. Qiime allows analysis of high-throughput community sequencing data. *Nature methods*, 7(5):335–336, 2010.