



Inference of Significant Microbial Interactions From Longitudinal Metagenomics Data

Xuefeng Gao, Bich-Tram Huynh, Didier Guillemot, Philippe Glaser, Lulla Opatowski

► To cite this version:

Xuefeng Gao, Bich-Tram Huynh, Didier Guillemot, Philippe Glaser, Lulla Opatowski. Inference of Significant Microbial Interactions From Longitudinal Metagenomics Data. *Frontiers in Microbiology*, 2018, 9, pp.2319. 10.3389/fmicb.2018.02319 . pasteur-02565633

HAL Id: pasteur-02565633

<https://pasteur.hal.science/pasteur-02565633>

Submitted on 6 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



Inference of Significant Microbial Interactions From Longitudinal Metagenomics Data

Xuefeng Gao^{1,2,3,4,5}, Bich-Tram Huynh^{1,2,3}, Didier Guillemot^{1,2,3}, Philippe Glaser^{6,7} and Lulla Opatowski^{1,2,3*}

¹ Inserm UMR 1181, Biostatistics, Biomathematics, Pharmacoeconomics and Infectious Diseases (B2PHI), Paris, France,

² Institut Pasteur, UMR 1181, B2PHI, Paris, France, ³ Université de Versailles St Quentin, UMR 1181, B2PHI, Paris, France,

⁴ Department of Gastroenterology and Hepatology, Shenzhen University General Hospital, Shenzhen, China, ⁵ Shenzhen University Clinical Medical Academy, Shenzhen, China, ⁶ Institut Pasteur, Unité Ecologie et Evolution de la Résistance aux Antibiotiques, Paris, France, ⁷ CNRS UMR3525, Paris, France

OPEN ACCESS

Edited by:

George Tsiamis,
University of Patras, Greece

Reviewed by:

Boyang Ji,
Chalmers University of Technology,
Sweden

Frank O'Neill Aylward,
Virginia Tech, United States

*Correspondence:

Lulla Opatowski
lulla.opatowski@pasteur.fr

Specialty section:

This article was submitted to
Systems Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 10 July 2018

Accepted: 11 September 2018

Published: 16 October 2018

Citation:

Gao X, Huynh B-T, Guillemot D,
Glaser P and Opatowski L (2018)
Inference of Significant Microbial
Interactions From Longitudinal
Metagenomics Data.
Front. Microbiol. 9:2319.
doi: 10.3389/fmicb.2018.02319

Data of next-generation sequencing (NGS) and their analysis have been facilitating advances in our understanding of microbial ecosystems such as human gut microbiota. However, inference of microbial interactions occurring within an ecosystem is still a challenge mainly due to sequencing data (e.g., 16S rDNA sequences) providing relative abundance of microbes instead of absolute cell count. In order to describe growth dynamics of microbial communities and estimate the involved microbial interactions, we introduce a procedure by integrating generalized Lotka-Volterra equations, forward stepwise regression and bootstrap aggregation. First, we successfully identify experimentally confirmed microbial interactions based on relative abundance data of a cheese microbial community. Then, we apply the procedure to time-series of 16S rDNA sequences of gut microbiomes of children who were progressing to Type 1 diabetes (T1D progressors), and compare their gut microbial interactions to a healthy control group. Our results suggest that the number of inferred microbial interactions increased over time during the first 3 years of life. More microbial interactions are found in the gut flora of healthy children than that of T1D progressors. The inhibitory effects from *Actinobacteria* and *Bacilli* to *Bacteroidia*, from *Bacteroidia* to *Clostridia*, and the beneficial effect from *Clostridia* to *Bacteroidia* are shared between healthy children and T1D progressors. An inhibition of *Clostridia* by *Gammaproteobacteria* is found in healthy children that maintains through their first 3 years of life. This suppression appears in T1D progressors during the first year of life, which transforms to a commensalism relationship at the age of 3 years old. *Gammaproteobacteria* is found exerting an inhibition on *Bacteroidia* in the T1D progressors, which is not identified in the healthy controls.

Keywords: microbial interactions, longitudinal taxonomic data, Lotka-Volterra equations, forward stepwise regression, bootstrap aggregation

INTRODUCTION

The human gut microbiota, an ecosystem comprising a diverse collection of interactive microbial species, plays a key role in nutrition, metabolism, physiology, and immune function in humans (Sekirot et al., 2010). An interaction between two microbes can either be neutral (0), positive (+), or negative (−) to the fitness of the interacting microbes. Based on the overall effects, an interaction

can be classified into commensalism (0/+), amensalism (0/−), mutualism (+/+), parasitism (+/−), competition (−/−), or no interactions (0/0). Interactions among microbial species characterize the composition and function of gut microbiota, thereby influencing the host's health (Rios-Covian et al., 2016). The microbial interactions in the gastrointestinal tract are complex, flexible, and capable of adapting to physiological perturbations (Sun and Chang, 2014). The change of one species may shift the relative abundances of other members in the community and affect the community's functional capacity. A better understanding of ecological dynamics and microbial interactions is essential to investigate the consequences of taxonomic perturbations.

By using NGS technologies, such as 16S rRNA gene profiling, it is now possible to follow the time-evolution of a microbial population by measuring abundance of bacterial species in a microbial community. However, analyzing the microbial community dynamics from genomic survey data is not straightforward, because sequencing data provide relative abundances of microbial species based on a fixed total number of sequences rather than absolute cell count. The high density of genomic survey data and its compositional nature are the most common challenges involved in inferring microbial interaction networks. Hence, the power of computational approaches allows to address this challenge. Mathematical and computational approaches make it possible to analyse highly complex microbial communities.

Correlation networks have proved useful for detecting biological interactions. There are many different techniques for computing correlation networks, for example Pearson correlation coefficient (Pearson, 1909), Spearman correlation coefficient (Spearman, 2010), Bray–Curtis distance (Roger Bray and Curtis, 1957), Local Similarity Analysis (Ruan et al., 2006; Beman et al., 2011; Steele et al., 2011; Xia et al., 2013), Maximal Information Coefficient (Reshef et al., 2011), SparCC (Friedman and Alm, 2012) based on Aitchison's log-ratio analysis (Anders and Huber, 2010) and CoNet which combines information from several different standard comparison metrics (Faust et al., 2012). Weiss et al. (2016) benchmarked the performance of these techniques in dealing with the relative abundance in simulated data specific to microbiome studies. Despite the correlation networks are useful in studying some overall biological relationships, they have significant limits in deciphering non-linear interactions such as between three or more species, and are unable to detect relationships like amensalism and partial-obligate-syntrophy (Morris et al., 2013; Weiss et al., 2016). Indeed, two species may be correlated even if they do not directly interact with each other.

The use of non-linear differential equations like Lotka-Volterra model is an alternative approach to study microbial interactions. The generalized Lotka-Volterra (gLV) equations are able to describe the time-dependent population dynamics and predict ecological relationships (i.e., mutualism, commensalism, parasitism, and competition) between members of different biological species. Several studies have attempted to use gLV equations studying microbial communities that consist of

multiple bacterial species (Stein et al., 2013; Fisher and Mehta, 2014; Marino et al., 2014; Bucci et al., 2016; Shaw et al., 2016). In particular, Mounier et al. used gLV equations describing the population dynamics of a cheese microbial community consisting of five microbial groups with experimental data of the absolute cell numbers (Mounier et al., 2008). This model predicted some microbial interactions that were confirmed by co-culture experiments afterwards.

Fisher and Mehta (2014) have proposed an approach named Learning Interactions from Microbial Time Series (LIMITS), which implements forward stepwise regression with median bootstrap aggregation for gLV equations inferring the species interactions in the human gut microbiomes. By combining forward stepwise regression with bootstrap aggregation, LIMITS was able to overcome a statistical issue termed “errors-in-variables” (accounting for measurement errors in the independent variables) and infer species interactions from time series relative abundances of species (Fuller, 1980). However, these inferred interactions were not experimentally confirmed.

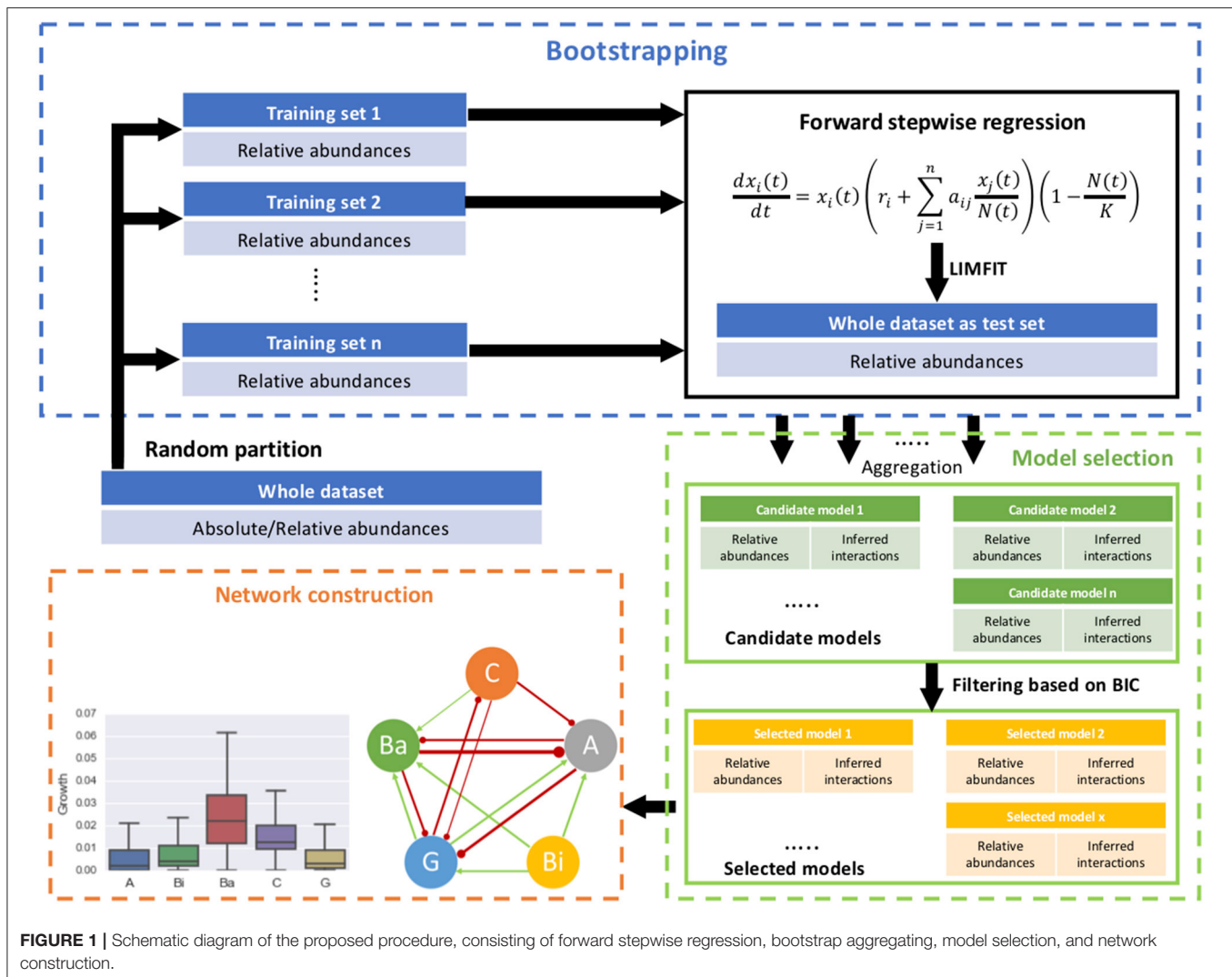
Although these tools brought some insights to infer microbial interactions, they have no proven ability to infer biologically confirmed microbial interactions from relative abundance data. Here, we propose a procedure by integrating gLV equations, forward stepwise regression, bootstrap aggregation and one-sample *t*-test. First, we test and validate this procedure by correctly identifying microbial interactions from relative abundances of a cheese microbial community (Mounier et al., 2008). For application, we then apply our procedure to infer microbial interactions within gut microbiota of healthy and type 1 diabetes (T1D) progressing infants (seroconverters for diabetes autoimmune) from a set of reported data (Kostic et al., 2015).

RESULTS

Procedure Validation on Experimentally Confirmed Data: Identification of Significant Microbial Interactions Within a Cheese Microbial Community

The procedure was tested through describing the dynamics of relative abundances of five microbial groups and estimating their interactions within a cheese microbial community. In the original study, the microbial population growth was measured in cell count, which was converted into relative abundances as input of our procedure. For initial conditions, the intrinsic growth rates and inter-species interaction coefficients were set at 1 generation/day and 0, respectively. The carrying capacity is 2×10^{10} CFU/g according to coculture studies (Mounier et al., 2008). A diagram of implementation of the procedure is given in **Figure 1**.

The gLV modeling succeeded in describing the changes of relative abundance of the microbial populations, and roughly predicted the growth trends of each member of this community (**Figure 2**). The inferred intrinsic growth rates of the microbes



and their significant interactions were shown in **Figure 3**. In particular, some of the inferred interactions (**Figure 3B**) were previously confirmed in the co-culture experiments, including the promotion effects from *G. candidum* to the bacterial group and *Leucobacter* sp., and the inhibitory effects of *G. candidum* on *D. hansenii* (Mounier et al., 2008). The estimated intrinsic growth rate of the bacterial group was low (**Figure 3A**), indicating that their growth was highly influenced by other members in the community. A benefit effect of *G. candidum* on the bacterial group was inferred, suggesting that the growth of the bacterial group partially relies on *G. candidum*. According to the co-culture study (Mounier et al., 2008), the growth of some bacteria, such as *Brevibacterium aurantiacum*, was actually relying on *G. candidum* was identified.

The sensitivity of our procedure was assessed by varying the number of bootstrapping samples and the size of training data (ranging from 100 to 1,000 bootstraps). The performance of this procedure was influenced by the number of bootstrap samples. We observed that some significant interactions were detected using a relatively low number of bootstrap samples. For instance,

the suppression of bacterial group by *D. hansenii* was identified as low as 100 bootstraps while it took 700 bootstraps to identify bacteria inhibiting *G. candidum* (**Supplementary Table S1**). The frequency of these interactions may be an indicator of their weight in shaping the community dynamics.

In summary, our procedure was able to identify ground-truth microbe-microbe interactions within a microbial community through relative abundance data.

Application on Time-Series 16S rRNA Gene Sequence Data: Gut Microbial Interaction Networks of Healthy Children vs. T1D Progressors

Next, we used this approach to study the gut microbial dynamics in healthy children and those who were progressing to T1D during their first 3 years of life. Kostic et al. (2015) examined the composition dynamics of gut microbiomes in 33 children genetically predisposed to T1D from birth until 3 years of age with monthly sampling. The authors observed a relative

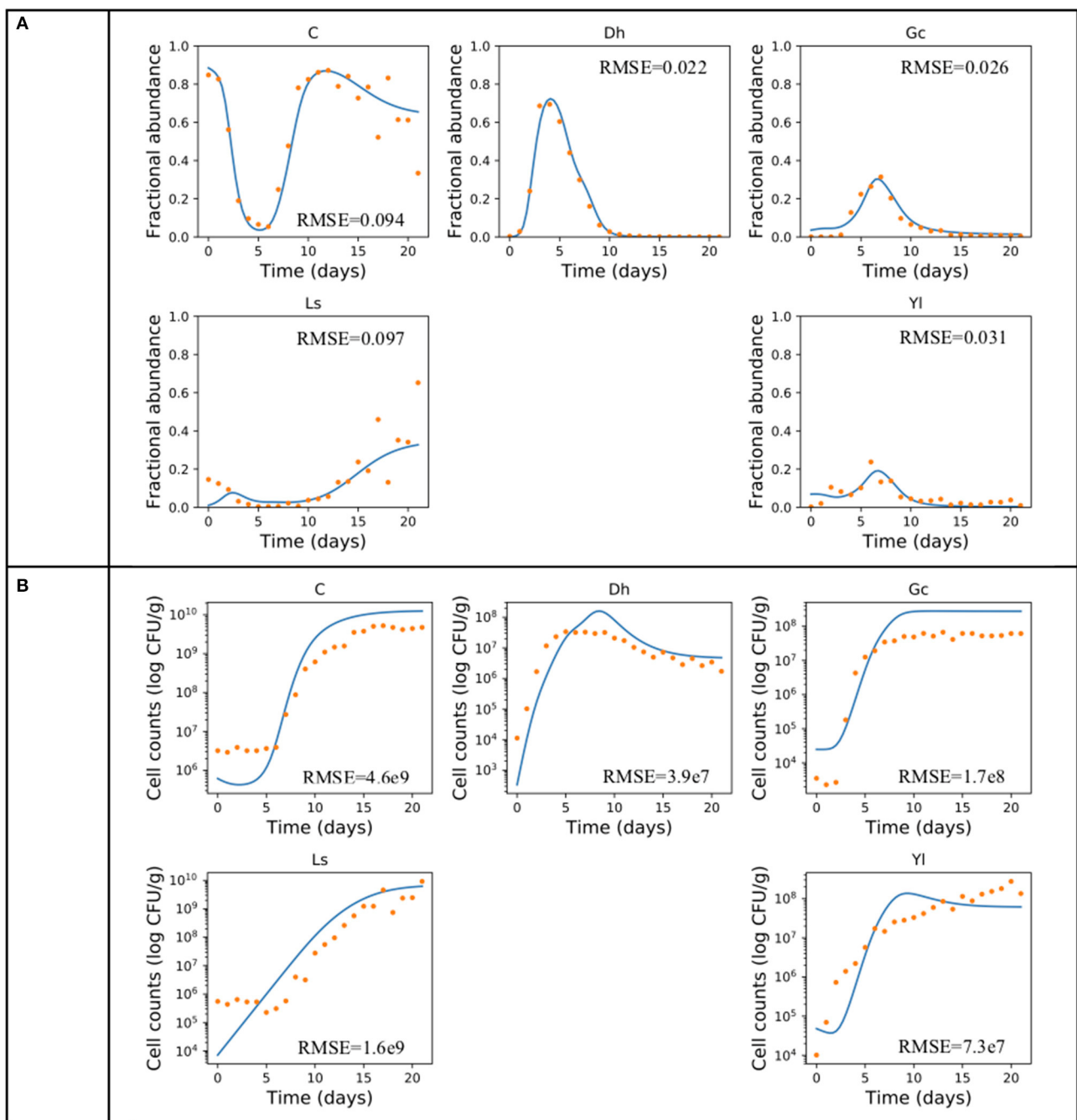


FIGURE 2 | Producing the growth dynamics of the cheese microbial community. An example of using gLV equations and forward stepwise regression describing the growth dynamics of the cheese microbial community in **(A)** relative abundance and **(B)** cell count. Experimental data (Mounier et al., 2008) is indicated by dots and the predicted population dynamics are indicated by curves. The fitting performance is measured in root-mean-square error (RMSE). A, *Actinobacteria*; Bi, *Bacilli*; Ba, *Bacteroidia*; C, *Clostridia*; G, *Gammaproteobacteria*.

reduction in alpha-diversity in the gut of children who progress to T1D compared to the seroconverters defined as positive for at least two autoantibodies (no T1D developers occurred during the follow-up) and in non-converters' gut. We applied the proposed procedure on the 16S rRNA sequencing data reported in (Kostic et al., 2015). We investigated the class

level gut bacterial interactions in T1D-associated and healthy infants at different age stages. The carrying capacity of intestinal microbial community was assumed to be $1e^{11}$ cells/g of feces [given the average bacteria number $3.4e^{10} \pm 3e^{10}$ cells/g feces (Savino et al., 2017)]. The resulting intrinsic growth rates and interaction coefficients of the five bacterial classes were given

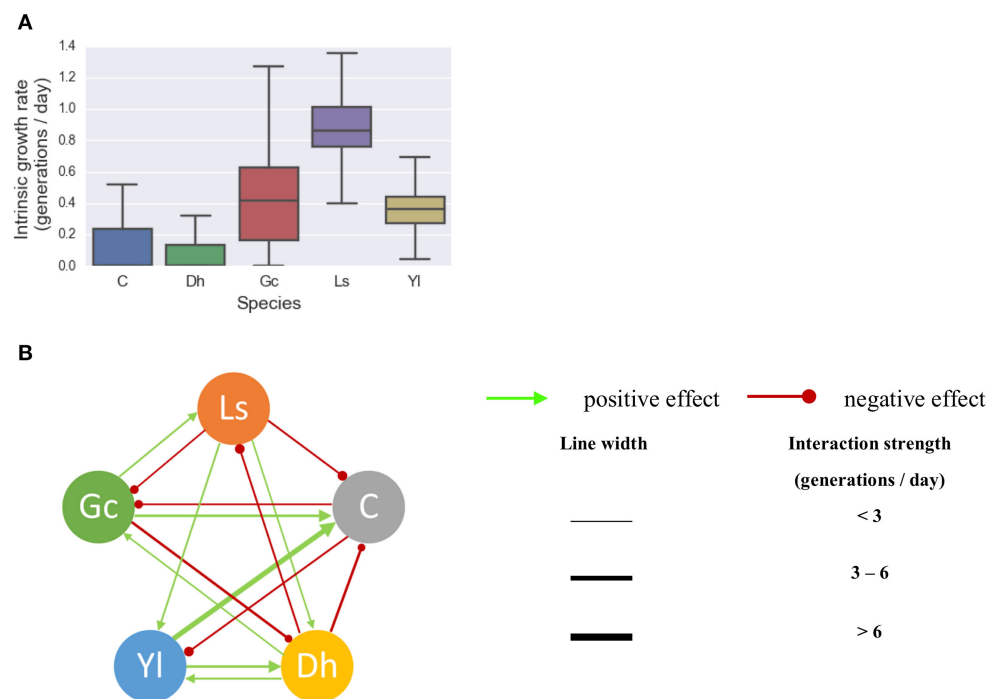


FIGURE 3 | Inferring significant microbial interactions within a cheese microbial community. **(A)** The intrinsic growth rates of the microbes and **(B)** their significant interactions were inferred from the longitudinal data of five microbes. 80% of the whole data set was randomly selected for training the gLV model, with 1,000 bootstrap samples. Resulted interactions were selected with one-sample t -test $P(a_{ij} \neq 0) > 95\%$. The line thickness is proportional to the strength of the interaction. Dh, *D. Hansenii*; Yl, *Y. Lipolytica*; Gc, *G. Candidum*; Ls, *Leucobacter sp.*; C, a bacterial group includes *Arthrobacter arilaitensis*, *Hafnia alvei*, *Corynebacterium casei*, *Brevibacterium aurantiacum*, and *Staphylococcus xylosus*.

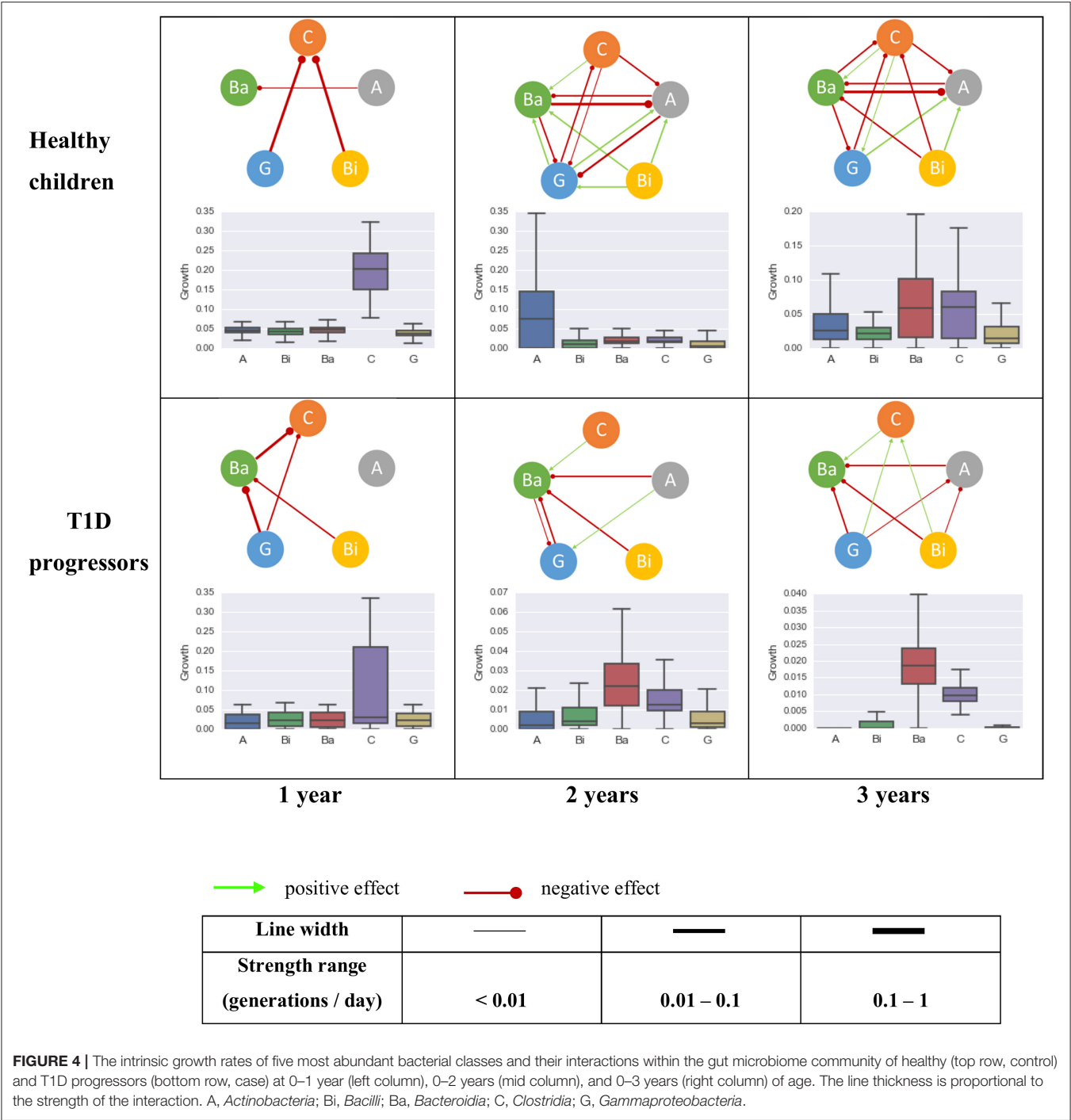
in **Supplementary Table S3** for the two groups at different age stages.

The gut microbial interaction network consisted of just a few amensalism interactions for the first year of life, but became more complex over time (**Figure 4**). There were more bacterial interactions identified in healthy children than T1D progressors after the age of 1 year old. Some interactions were shared between healthy controls and T1D progressors, including the inhibitory effects *Gammaproteobacteria* to *Clostridia* for 1 year of age, the inhibitory effect from *Actinobacteria* to *Bacteroidia* and the promotion effect from *Clostridia* to *Bacteroidia*.

Three interactions were identified uniquely in the healthy controls: inhibitions of *Actinobacteria* by *Bacteroidia* and *Clostridia*, and a promotion from *Gammaproteobacteria* to *Actinobacteria*. An inhibitory effect from *Gammaproteobacteria* to *Clostridia* was established within the first year of life which maintained through the following 3 years (top row of **Figure 4**). Intriguingly, the effect from *Gammaproteobacteria* to *Clostridia* was identified as amensalism during the first year of life for T1D progressors, which transformed to a benefit effect at their age of three. An inhibitory effect from *Gammaproteobacteria* to *Bacteroidia* was only observed in the T1D progression group (bottom row of **Figure 4**). The relationship between *Bacilli* and *Clostridia* was amensalism for healthy children, but commensalism for T1D progression children at age of three.

DISCUSSIONS

Investigating the gut microbial community composition and their interaction patterns are crucial to understand the role of gut microbiota in maintaining human health and causing diseases. However, a systematic approach for accessing microbial interactions has been lacking. Some methods have been attempted to infer microbial interaction networks, such as correlation networks and mathematical modeling. Although these attempts have contributed in creating interaction networks, none of their resulting microbial interactions have been experimentally confirmed. In the present study, we showed that a combination of gLV equations, forward stepwise regression, and bootstrap aggregation was able to infer microbial interactions from longitudinal relative abundance data. In order to test the performance this procedure, we used the dataset of a cheese microbial community with known microbial interactions (Mounier et al., 2008). Our proposed procedure was able to correctly infer some experimentally confirmed microbial interactions, such as the promotions of *G. candidum* on *Leucobacter sp.* and on a group of bacteria, and a suppression of *D. hansenii* by *G. candidum*. In addition, the estimated changes in cell counts by our procedure roughly recapitulated the growth trends of each microbial group in the community. To access the performance of this procedure, we tested the cheese dataset on MetaMIS software (Shaw et al., 2016),



and obtained comparable results. Our procedure identified 15 significant interactions among the microbe groups, and 14 of them share the same directions with the inferred interactions by MetaMIS (**Supplementary Table S2**). However, our proposed procedure also provided bacterial growth rates as they are important information to decipher microbial population dynamics. Despite the potential of our proposed method in dealing with compositional data, it could not explain all the

observations from the co-culture experiments. For example, our procedure predicted an inhibitory effect of *D. hansenii* on the bacterial group, which was inconsistent with the experimental observation (Mounier et al., 2008). In addition, the interaction coefficient may be not proportional to the microbial abundances (e.g., variable *a* in Equation 1 does not only depend on the abundance of the microbes). This level of complexity was not considered in our current procedure, but

it is an important mechanism needed to be addressed in the future.

One goal of comparative metagenomics is to identify meaningful changes in the microbiome's taxonomic and functional composition that are associated with health and disease. Disruptions of the process of gut microbial colonization during childhood have been shown to be associated with an increased risk of adiposity and pathogenesis of autoimmune disorders such as T1D (Huh et al., 2012; Blustein et al., 2013; Murri et al., 2013; Kostic et al., 2015). To address how alterations of the gut microbial community composition may contribute to childhood disease, we must first investigate the normal dynamics of the community in the growing infant. By comparing the gut taxonomic trajectories of children who progress to T1D compared to healthy controls, Kostic et al. (2015) have found general changes in abundance of the gut microbial community and the timing of these changes. They identified some microbes strongly correlated which was consistent across most healthy subjects (Kostic et al., 2015), such as a positive correlation between *Gammaproteobacteria* and *Actinobacteria* or a negative correlation between *Gammaproteobacteria* and *Clostridia* (Kostic et al., 2015). These correlations can be explained by the inferred interactions from our procedure, e.g., promotion effect from *Gammaproteobacteria* to *Actinobacteria* and inhibition effect from *Gammaproteobacteria* to *Clostridia*.

By applying our procedure to a longitudinal time-series study of gut microbiota in children (Kostic et al., 2015), we demonstrated that the gut bacterial interaction network (at class level) was getting more complex along with age, with more interactions inferred for the healthy children than the T1D progressors. Some microbial interactions were shared between healthy children and the T1D progressors, such as inhibitory effects from *Actinobacteria* and *Bacilli* to *Bacteroidia*. Although the composition of the microbiome community was changing along with child development, some of the bacterial interactions, in particular at higher phylogenetic levels such as class, remained stable after establishment. For example, the inhibitory effect from *Gammaproteobacteria* to *Clostridia* was established for healthy children cohort in the first year of life and maintained through the age of three. In T1D progressors, whereas the amensalism relationship between *Gammaproteobacteria* and *Clostridia* appeared during their first year of life, which became commensalism (i.e., *Gammaproteobacteria* enhancing the growth of *Clostridia*) at the age of three. The inhibition from *Gammaproteobacteria* to *Bacteroidia* maintained during the first 3 years of life for T1D progressors, which was not detected in the healthy controls.

It has been observed that the quantity of *Clostridium* (genus level, belongs to the class *Clostridia*) correlated positively and significantly with the plasma glucose level in diabetic children. In newborn infants, *Gammaproteobacteria* have been shown to cause a healthy level of inflammation in their intestines, protecting them from excessive inflammatory, and autoimmune disorders later in life (Mirpuri et al., 2014). From the data of (Kostic et al., 2015), the proportion of *Gammaproteobacteria* was indeed lower in the T1D progression children group than healthy controls during the first months of life (e.g., about

11% vs. 16% for the two groups within 180 days after birth). In addition, *Bacilli* was also found to promote the growth of *Clostridia* in the gut community of T1D progression children, whereas it contributed to inhibition of *Clostridia* for the healthy controls. Thus, our results suggest that the inhibitory effects of *Gammaproteobacteria* and *Bacilli* on *Clostridia* might play a role in regulating plasma glucose in early live and protect children from developing to T1D. Further studies are necessary to test this hypothesis.

CONCLUSIONS

Here we propose a procedure which was capable of inferring experimentally confirmed microbial interactions out of compositional data. By using our procedure, we identified some similarities as well as differences in the gut bacterial interaction networks between children toward T1D progression and healthy controls in their first 3 years of life. The interaction networks were getting more complex in the gut microbiome of children after their first year of life. The number of interactions was higher in healthy children than in T1D progressors. Some bacterial interactions were exclusively found in each group, which may be able to predict the T1D progression state. Our results provide potential new insights into the relationship between gut microbial interactions and infants' T1D progression. In the future, by incorporating microbiome's functional shift (Manor and Borenstein, 2017), the procedure presented above might help to interpret disease-specific changes of microbial interactions and accordingly help to predict disease progression.

METHODS

Generalized Lotka-Volterra Equation

We used gLV equations to describe an ecological community consisting of n microbiome taxa. For taxon i , the population dynamics is described as

$$\frac{dx_i(t)}{dt} = x_i(t) \left(r_i + \sum_{j=1}^n a_{ij} \frac{x_j(t)}{N(t)} \right) \left(1 - \frac{N(t)}{K} \right) \quad (1)$$

where $x_i(t)$ is the absolute abundance of taxon i at time t , $N(t)$ the community size at time t , n the number of species in the ecosystem, and K the carrying capacity of the community (a logistic growth component). r_i represents the intrinsic growth rate of taxon i , and a_{ij} the effect that taxon j has upon taxon i , which is proportional to the relative abundance of species j (i.e., $\frac{x_j(t)}{N(t)}$). We assumed the interactions between species i and j is not necessarily bilateral, which can take one of six possible forms based on the signs of a_{ij}/a_{ji} ($i \neq j$): $+/-$ (parasitism), $-/-$ (competition), $+/+$ (mutualism), $+/0$ (commensalism), $-/0$ (amensalism), and $0/0$ (neutral). For simplicity, we did not consider intra-species interactions, such as the competition for resources between the members of a same species. However, a carrying capacity of the environment was introduced. We fitted relative abundance data to the gLV equations using the Levenberg-Marquardt algorithm for non-linear least squares

minimization, implemented by LMFIT (Newville et al., 2016) which is a high-level interface to non-linear optimization and curve fitting tool for Python.

Forward Stepwise Regression

We used a forward stepwise regression method for the gLV model selecting interactions that explain most of the population dynamics of microbial species. This method started with no interaction coefficients in the model (i.e., only with growth terms), testing the addition of each interaction coefficient using a comparison criterion, adding the interaction if it improves the model the most, and repeating this process until no improvement of the model fitting. Stepwise regression is prone to overfitting the data, which arises from searching a large space of possible models. We use Bayesian information criterion (BIC) (Kass and Wasserman, 1995) as a selection criterion in order to solve the overfitting problem by introducing a penalty term for the number of parameters in the model. BIC is calculated by

$$BIC = U \cdot \ln \left(\frac{RSS}{U} \right) + \ln(U) \cdot V \quad (2)$$

where U is the number of data points, and V the number of variable parameters in the gLV model. Residual sum of squares (RSS) over the whole data set is given by

$$RSS = \sum_{t=1}^T \sum_{i=1}^n \left(\frac{x_i^t}{N^t} - \frac{\hat{x}_i^t}{N^t} \right)^2 \quad (3)$$

where $\frac{x_i}{N}$ and $\frac{\hat{x}_i}{N}$ are the actual observed and predicted relative abundance, respectively. T is the total number of time points, and n is the number of species in the ecosystem.

Therefore, a lower BIC implies either a better fit, fewer explanatory variables, or both.

Bootstrap Aggregating (Bagging)

In order to attenuate the instability problem of forward stepwise regression, we introduced bootstrap aggregating strategy for noise-reduction (Breiman, 1996) by randomly partitioning the data into two sets: training set and test set. A typical way to test for accuracy in models created by stepwise regression is to evaluate the models against a test set that was not used for model creation. Accuracy is then calculated as errors between the values predicted by the model and data in the test set. In order to assess both the descriptive and predictive power, we fitted the models to the training set but evaluate the models to the whole data set. The procedure is performed in the following steps:

1. At each iteration, partitioning the data into a training set and a test set.
2. The gLV model fitting on the training set and BIC calculated for the best parameters starting with only growth terms, and all interaction coefficients are assigned to a parameter set TEST $\{a_{ji}\}$.
3. Each single and pairwise interaction coefficient in the TEST $\{a_{ji}\}$ is separately added to current model, then a BIC is calculated over the whole data set. After scanning over all coefficients in the TEST $\{a_{ji}\}$, the one (or pairwise) that

generated the lowest BIC is transferred to a parameter set SELECT. If multiple models have similar BIC values that are close to the lowest BIC ($\Delta BIC < 2$), one of them is randomly selected. The new model then consists of all coefficients in SELECT $\{a_{ji}\}$.

4. Repeating step 3 for the new model to evaluate the model fitting after addition of (remaining) coefficients in the TEST $\{a_{ji}\}$ and update the lowest BIC accordingly.
5. Iteration stops when additional coefficients no longer reduce the lowest BIC. The resulted coefficients in the SELECT $\{a_{ji}\}$ comprise an estimate of the interaction network.

This forward stepwise regression process was repeated many times (up to 1,000 bootstraps in this study), each resulting in a set of interaction coefficients that compose a specific gLV model. The python code we used for generating candidate models is provided in **Supplementary Code S1**. These models were then aggregated. Next, we removed models that violated biological constraints (e.g., based on a minimum population size and/or a minimum number for each species in the community). After filtering, we selected models with minimum BIC values (the minimum BIC + 10% allowance), and their parameters (growth rates and interaction coefficients) were compiled. Finally, the significant interactions were determined through one-sample t -test with the value of $P(a_{ij} \neq 0) > 95\%$.

Data

Mounier et al. (2008) investigated interactions between yeast and bacteria within a cheese microbial community composed of three yeasts (*Debaryomyces hansenii* 1L25, *Geotrichum candidum* 3E17, and *Yarrowia lipolytica* 1E07) and six bacteria (*Arthrobacter arilaitensis* 3M03, *Brevibacterium aurantiacum* 2M23, *Corynebacterium casei* 2M01, *Hafnia alvei* 2E12, *Leucobacter* sp. strain 1L36, and *Staphylococcus xylosus* 1L18). Two experiments were conducted to measure the microbial dynamics during the development of the ecosystem. In the first one, the dynamic of the full ecosystem was studied. Cheeses were sampled in duplicate every day for 21 days for microbial enumeration together with measurements of lactose, lactate content, and pH. The second experiment aimed to investigate the effect of the absence of one, two or three yeasts (all combination were tested) in the ecosystem development. To test our procedure, we studied the development dynamics of this cheese microbiome community [data were obtained from (Mounier et al., 2008)].

Kostic et al. studied the impact of the gut microbiome dynamic on type 1 diabetes (T1D). They recruited between September 2008 and August 2010 in Finland and Estonia 33 newborns with positive cord blood testing for HLA DR-DQ, which are alleles conferring risk of T1D. The infants were followed-up until the age of 3 years with monthly stool samples. Data regarding infections, drugs consumption (in particular antibiotics), diet were collected. Serum samples were collected at 0 (cord blood), 3, 6, 12, 18, 24, and 36 months to test for 4 diabetes-associated autoantibodies. 16S rRNA sequencing were performed for accessing the fecal microbiota composition. Eleven children had at least two positive autoantibodies seroconversion during their follow-up but did not develop T1D (i.e., seroconverters).

Four children progressed to T1D by the end of the study. The 11 seroconverters were matched with the 22 healthy controls for gender, HLA genotype and country. We applied our procedure on a set of 16S rDNA sequences (in OTUs) of the gut microbiota of 11 T1D progression children (seven seroconverters and four who progressed to T1D) and 22 healthy controls between the age of 0–3 years old [data were obtained from **Supplementary Table S2** published in Kostic et al. (2015)]. In specific, we used the relative abundance of five bacterial classes: *Actinobacteria*, *Bacteroidia*, *Bacilli*, *Clostridia*, and *Gammaproteobacteria*.

AUTHOR CONTRIBUTIONS

XG and LO conceived the method and analyzed the data. XG and LO wrote the manuscript with help from B-TH, DG, and PG. LO and PG were responsible for management of the project.

FUNDING

This work was supported by the program Projets Transversaux de Recherche PTR 2015 (grant no PTR558). This work was also supported directly by internal resources of the French National Institute for Health and Medical Research (Inserm), the Institut Pasteur, the University of Versailles–Saint-Quentin-en-Yvelines (UVSQ) and the French Government's Investissement d'Avenir program, Laboratoire d'Excellence Integrative Biology of Emerging Infectious Diseases (grant no ANR-10-LABX-62-IBED).

REFERENCES

- Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11:R106. doi: 10.1186/gb-2010-11-10-r106
- Beman, J. M., Steele, J. A., and Fuhrman, J. A. (2011). Co-occurrence patterns for abundant marine archaeal and bacterial lineages in the deep chlorophyll maximum of coastal California. *ISME J.* 5, 1077–1085. doi: 10.1038/ismej.2010.204
- Blustein, J., Attina, T., Liu, M., Ryan, A. M., Cox, L. M., Blaser, M. J., et al. (2013). Association of caesarean delivery with child adiposity from age 6 weeks to 15 years. *Int. J. Obes.* 37, 900–906. doi: 10.1038/ijo.2013.49
- Breiman, L. (1996). Bagging predictors. *Mach. Learn.* 24, 123–140. doi: 10.1007/BF00058655
- Bucci, V., Tzen, B., Li, N., Simmons, M., Tanoue, T., Bogart, E., et al. (2016). MDSINE: microbial dynamical systems inference engine for microbiome time-series analyses. *Genome Biol.* 17:121. doi: 10.1186/s13059-016-0980-6
- Faust, K., Sathirapongsasuti, J. F., Izard, J., Segata, N., Gevers, D., Raes, J., et al. (2012). Microbial co-occurrence relationships in the human microbiome. *PLoS Comput. Biol.* 8:e1002606. doi: 10.1371/journal.pcbi.1002606
- Fisher, C. K., and Mehta, P. (2014). Identifying keystone species in the human gut microbiome from metagenomic timeseries using sparse linear regression. *PLoS ONE* 9:e102451. doi: 10.1371/journal.pone.0102451
- Friedman, J., and Alm, E. J. (2012). Inferring correlation networks from genomic survey data. *PLoS Comput. Biol.* 8:e1002687. doi: 10.1371/journal.pcbi.1002687
- Fuller, W. A. (1980). Properties of some estimators for the errors-in-variables model. *Ann. Stat.* 8, 407–422. doi: 10.1214/aos/1176344961
- Gao, X., Huynh, B.-T., Guillemot, D., Glaser, P., and Opatowski, L. (2018). Inference of significant microbial interactions from longitudinal metagenomics sequencing data. *bioRxiv [Preprint]*. doi: 10.1101/305326

ACKNOWLEDGMENTS

A previous version of this article was published as preprint on BioRxiv (Gao et al., 2018).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2018.02319/full#supplementary-material>

Supplementary Table S1 | Sensitivity analysis of the number of bootstrap samples and the size of training dataset. The strengths of significant interactions within a cheese microbial community (Mounier et al., 2008), resulted from 100 to 1,000 bootstraps with one-sample *t*-test $P(a_{ij} \neq 0) > 95\%$. Fifty to ninety percent of the whole data set was randomly selected for training the gLV model.

Supplementary Table S2 | Comparisons of microbial interactions inferred from our proposed procedure and those resulted from MetaMIS.

Supplementary Table S3 | The strengths of significant gut microbial interactions (class level) for the health children and the T1D progressors (Kostic et al., 2015), resulted from 1,000 bootstraps with one-sample *t*-test $P(a_{ij} \neq 0) > 95\%$. Eighty percent of the whole data set was randomly selected for training the gLV model.

Supplementary Code S1 | Python code for the procedure generating candidate models and an example of implementation with the cheese dataset (**Supplementary Data 1**). https://bitbucket.org/RyanGao/supplementary_code_1/get/Oa0cea8846c0.zip

Supplementary Data 1 | Cell counts of the cheese microbial community (Mounier et al., 2008). Abbreviations in the table: D, *D. Hansenii*; Y, *Y. Lipolytica*; G, *G. Candidum*; L, *Leucobacter sp.*; C, sum of *Arthrobacter arilaitensis*, *Hafnia alvei*, *Corynebacterium casei*, *Brevibacterium aurantiacum*, and *Staphylococcus xylosum*.

- Huh, S. Y., Rifas-Shiman, S. L., Zera, C. A., Edwards, J. W., Oken, E., Weiss, S. T., et al. (2012). Delivery by caesarean section and risk of obesity in preschool age children: a prospective cohort study. *Arch. Dis. Child.* 97, 610–616. doi: 10.1136/archdischild-2011-301141
- Kass, R. E., and Wasserman, L. (1995). A reference bayesian test for nested hypotheses and its relationship to the schwarz criterion. *J. Am. Stat. Assoc.* 90, 928–934.
- Kostic, A. D., Gevers, D., Siljander, H., Vatanen, T., Hyötyläinen, T., Hamalainen, A. M., et al. (2015). The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. *Cell Host Microbe* 17, 260–273. doi: 10.1016/j.chom.2015.01.001
- Manor, O., and Borenstein, E. (2017). Systematic characterization and analysis of the taxonomic drivers of functional shifts in the human microbiome. *Cell Host Microbe* 21, 254–267. doi: 10.1016/j.chom.2016.12.014
- Marino, S., Baxter, N. T., Huffnagle, G. B., Petrosino, J. F., and Schloss, P. D. (2014). Mathematical modeling of primary succession of murine intestinal microbiota. *Proc. Natl. Acad. Sci. U. S. A.* 111, 439–444. doi: 10.1073/pnas.1311322111
- Mirpuri, J., Raetz, M., Sturge, C. R., Wilhelm, C. L., Benson, A., Savani, R. C., et al. (2014). Proteobacteria-specific IgA regulates maturation of the intestinal microbiota. *Gut Microbes* 5, 28–39. doi: 10.4161/gmic.26489
- Morris, B. E., Henneberger, R., Huber, H., and Moissl-Eichinger, C. (2013). Microbial syntrophy: interaction for the common good. *FEMS Microbiol. Rev.* 37, 384–406. doi: 10.1111/1574-6976.12019
- Mounier, J., Monnet, C., Vallaëys, T., Arditi, R., Sarthou, A. S., Helias, A., et al. (2008). Microbial interactions within a cheese microbial community. *Appl. Environ. Microbiol.* 74, 172–181. doi: 10.1128/AEM.01338-07
- Murri, M., Leiva, I., Gomez-Zumaquero, J. M., Tinahones, F. J., Cardona, F., Soriguer, F., et al. (2013). Gut microbiota in children with type 1 diabetes

- differs from that in healthy children: a case-control study. *BMC Med.* 11:46. doi: 10.1186/1741-7015-11-46
- Newville, M., Stensitzki, T., Allen, D. B., Rawlik, M., Ingargiola, A., and Nelson, A., (2016). Lmfit: non-linear least-square minimization and curve-fitting for python. *Code Source Lib.* doi: 10.5281/zenodo.11813
- Pearson, K. (1909). Determination of the coefficient of correlation. *Science* 30, 23–25. doi: 10.1126/science.30.757.23
- Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., et al. (2011). Detecting novel associations in large data sets. *Science* 334, 1518–1524. doi: 10.1126/science.1205438
- Rios-Covian, D., Ruas-Madiedo, P., Margolles, A., Gueimonde, M., de Los Reyes-Gavilan, C. G., and Salazar, N. (2016). Intestinal short chain fatty acids and their link with diet and human health. *Front. Microbiol.* 7:185. doi: 10.3389/fmicb.2016.00185
- Roger Bray, J., and Curtis, J. T. (1957). An ordination of the upland forest communities of southern wisconsin. *Ecol. Monogr.* 27, 325–349. doi: 10.2307/1942268
- Ruan, Q., Dutta, D., Schwalbach, M. S., Steele, J. A., Fuhrman, J. A., and Sun, F. (2006). Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors. *Bioinformatics* 22, 2532–2538. doi: 10.1093/bioinformatics/btl417
- Savino, F., Quartieri, A., De Marco, A., Garro, M., Amaretti, A., Raimondi, S., et al. (2017). Comparison of formula-fed infants with and without colic revealed significant differences in total bacteria, *Enterobacteriaceae* and faecal ammonia. *Acta Paediatr.* 106, 573–578. doi: 10.1111/apa.13642
- Sekirov, I., Russell, S. L., Antunes, L. C., and Finlay, B. B. (2010). Gut microbiota in health and disease. *Physiol. Rev.* 90, 859–904. doi: 10.1152/physrev.00045.2009
- Shaw, G. T., Pao, Y. Y., and Wang, D. (2016). MetaMIS: a metagenomic microbial interaction simulator based on microbial community profiles. *BMC Bioinformatics* 17:488. doi: 10.1186/s12859-016-1359-0
- Spearman, C. (2010). The proof and measurement of association between two things. *Int. J. Epidemiol.* 39, 1137–1150. doi: 10.1093/ije/dyq191
- Steele, J. A., Countway, P. D., Xia, L., Vigil, P. D., Beman, J. M., Kim, D. Y., et al. (2011). Marine bacterial, archaeal and protistan association networks reveal ecological linkages. *ISME J.* 5, 1414–1425. doi: 10.1038/ismej.2011.24
- Stein, R. R., Bucci, V., Toussaint, N. C., Buffie, C. G., Ratsch, G., Pamer, E. G., et al. (2013). Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota. *PLoS Comput. Biol.* 9:e1003388. doi: 10.1371/journal.pcbi.1003388
- Sun, J., and Chang, E. B. (2014). Exploring gut microbes in human health and disease: pushing the envelope. *Genes Dis.* 1, 132–139. doi: 10.1016/j.gendis.2014.08.001
- Weiss, S., Van Treuren, W., Lozupone, C., Faust, K., Friedman, J., Deng, Y., et al. (2016). Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. *ISME J.* 10, 1669–1681. doi: 10.1038/ismej.2015.235
- Xia, L. C., Ai, D., Cram, J., Fuhrman, J. A., and Sun, F. (2013). Efficient statistical significance approximation for local similarity analysis of high-throughput time series data. *Bioinformatics* 29, 230–237. doi: 10.1093/bioinformatics/bts668

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Gao, Huynh, Guillemot, Glaser and Opatowski. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.