

METHODOLOGY

Open Access



Microbial network inference for longitudinal microbiome studies with LUPINE

Saritha Kodikara¹ and Kim-Anh Lê Cao^{1*}

Abstract

Background The microbiome is a complex ecosystem of interdependent taxa that has traditionally been studied through cross-sectional studies. However, longitudinal microbiome studies are becoming increasingly popular. These studies enable researchers to infer taxa associations towards the understanding of coexistence, competition, and collaboration between microbes across time. Traditional metrics for association analysis, such as correlation, are limited due to the data characteristics of microbiome data (sparse, compositional, multivariate). Several network inference methods have been proposed, but have been largely unexplored in a longitudinal setting.

Results We introduce LUPINE (LongitUdinal modelling with Partial least squares regression for NEtwork inference), a novel approach that leverages on conditional independence and low-dimensional data representation. This method is specifically designed to handle scenarios with small sample sizes and small number of time points. LUPINE is the first method of its kind to infer microbial networks across time, while considering information from all past time points and is thus able to capture dynamic microbial interactions that evolve over time. We validate LUPINE and its variant, LUPINE_single (for single time point analysis) in simulated data and four case studies, where we highlight LUPINE's ability to identify relevant taxa in each study context, across different experimental designs (mouse and human studies, with or without interventions, and short or long time courses). To detect changes in the networks across time and groups or in response to external disturbances, we used different metrics to compare the inferred networks.

Conclusions LUPINE is a simple yet innovative network inference methodology that is suitable for, but not limited to, analysing longitudinal microbiome data. The R code and data are publicly available for readers interested in applying these new methods to their studies.

Keywords Longitudinal, Network, 16S, Partial correlation

Introduction

The microbiome refers to a distinct microbial community that resides within a well-defined habitat, characterised by specific physio-chemical properties [1]. This ecosystem is composed of bacteria from hundreds to thousands

of different taxa [2], forming complex community structures that are challenging to decipher. Given the inherent dynamics of the microbiome, longitudinal studies are essential to gain insights into the impact, variations over time, and responses to external perturbations such as dietary changes or medications on the microbiome. The temporal dimension, while adding complexity, provides richer insights into the dynamic processes, patterns, and interactions within microbial communities, surpassing what cross-sectional studies can offer [3]. The advancements in sequencing technologies, coupled with reduced

*Correspondence:

Kim-Anh Lê Cao
kimanh.lecao@unimelb.edu.au

¹ Melbourne Integrative Genomics, School of Mathematics and Statistics,
The University of Melbourne, Royal Parade, 3052 Parkville, Victoria,
Australia



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

sequencing costs, have enabled researchers to conduct an increasing number of longitudinal microbiome studies.

One objective of microbiome studies is to accurately infer microbial networks from the datasets generated by sequencing technologies. These networks can be viewed as temporal or spatial snapshots of ecosystems [4]. In these networks, interactions signify significant associations between taxa, which are typically non-directional in nature [5]. The simplest way to detect associations is via correlation analysis, such as Pearson or Spearman correlation. However, these correlation methods are sub-optimal for microbiome data as they ignore the compositional structure of the data and lead to spurious results [6]. A more valid approach is to use the concept of partial correlation, which effectively measure pairwise associations between variables subject to a constant-sum constraint (i.e. compositional data, [7]). Existing compositional-aware microbiome network methods have currently focused on single time point studies, rather than longitudinal. Widely used single time point methods are either based on correlation approaches, such as SparCC [8] or precision based approaches, such as SpiecEasi [5]. In contrast to correlation-based methods, precision or partial correlation-based methods concentrate on direct associations by removing indirect associations.

Longitudinal microbiome studies enables researchers to create temporal snapshots of microbial networks for a comprehensive view of the microbiome. But network methods for such type of studies are still in their infancy and exhibit diverse analytical objectives [9]. Some methods aim to model relationships between taxa for each individual subject. Other methods aim to model collective microbial interactions across all time points. However, the assumption that microbial interaction stays constant through time, whether on an individual basis or across all subjects, proves limiting. This limitation is particularly evident when there are external interventions, such as alterations in diet or antibiotic usage, which disrupt the stability of these relationships. We have addressed this limitation with a sequential approach, that, to the best of our knowledge, is the first of its kind for longitudinal microbiome studies. We acknowledge the dynamic nature of microbial interactions, especially in the face of interventions, and introduce a flexible framework for the temporal evolution of these interactions.

LUPINE (LongitUdinal modelling with Partial least squares regression for NEtwork inference) combines one-dimensional approximation and partial correlation to measure the linear association between a pair of taxa, accounting for the effects of the other taxa. We developed two variants, one for single time point network inference, and another for longitudinal sequential network inference. The former focuses on inferring the network

at a specific time point, in the vein of SpiecEasi [5] and SparCC [8]. The latter, however, incorporates information from all previous time points. The inferred network from our method is in the format of a binary graph. In this network, the vertices represent different taxa, and the edges represent significant associations between them. An edge between two taxa indicates that there is a significant connection between them. This connection is visualised as a line between the two nodes in the graph.

In the ‘Materials and methods’ section, we explain the rationale of our approaches along with downstream analyses that can be conducted to compare inferred networks. In the ‘Results’ section, we first benchmark LUPINE and LUPINE_single with two state-of-the-art network inference methods (solely designed for single time point microbiome data) on simulated data. Next, through a case study, we demonstrate the robustness of LUPINE compared to LUPINE_single. We then showcase the applicability of LUPINE in real data with different experimental designs in four case studies. These experiments varied from human studies to mouse studies, from short to longer time courses, and ranged from cross-sectional to intervention studies. Finally, in the ‘Discussion’ section, we delve into the implications of our findings, special aspects we had to consider when developing LUPINE, and suggest avenues for future research.

Materials and methods

Method

Our network inference methodology is tailored for the analysis of longitudinal microbiome data, where we assume that the dynamics of taxa interactions are temporal. In order to uncover these intricate relationships, we leverage both past and current information from the data in a sequential manner. Including past information in the model enables insights into inherent variations in microbial interactions across diverse temporal scenarios, including interventions. Our methodology includes three variants that are described in the following sections:

- Single time point modelling: given a pair of taxa, we estimate pairwise partial correlations while accounting for the influence of the other taxa using one-dimensional approximation with principal component analysis (PCA).
- Two time point modelling: given a pair of taxa, we estimate pairwise partial correlations while accounting for the influence of the other taxa using projection to latent structures (PLS) regression [10]. PLS regression performs dimension reduction such that the covariance between the current and preceding time point data sets is maximised.

- Several time point modelling: given a pair of taxa, we estimate pairwise partial correlations while accounting for the influence of the other taxa using generalised PLS for multiple blocks of data ('blockPLS') regression [11, 12]. We use blockPLS to maximise the covariance between the current and any past time point datasets.

We assume that individuals within a specific group (e.g. control group) have a common network structure at a particular time point. Therefore, we consider each group separately for studies with multiple sample groups (as illustrated in case studies 1, 3 and 4) to focus on each of their unique microbial characteristics.

Notations. In the following, we denote X an $(n \times p)$ data matrix (where n is the number of samples and p is the number of taxa); X^i and X^j are columns in X for taxa i and j ; $X^{-(i,j)}$ is an $(n \times p - 2)$ data matrix excluding taxa i and j columns in X ; \mathbf{l} is a vector of library sizes for n samples. In situations where time is of interest (i.e. in longitudinal modelling), all of the above matrices and vectors include a subscript indicating the time point. For example, X_t represents an $(n \times p)$ data matrix at time point t . For simplicity, we will omit the hat notation (i.e. $\hat{\rho}$) of our estimates. However, note that all estimates approximate some population parameter. Additionally, $\|\cdot\|$ denotes the Euclidean norm.

Single time point modelling with PCA

We first introduce the single time point approach which we will generalise for multiple time points in the next sections. This type of analysis gives insights of microbial associations at a single time point. It can also be used when there is a large time gap between time points but only one particular time point is of interest.

To assess the strength of the association between two taxa, we use partial correlation—a statistical technique that measures association while controlling for other variables. For example, we estimate the partial correlation between a pair of taxa, denoted by i and j in Fig. 1, to assess the strength of their association while controlling for other taxa. Controlling for these taxa requires regressing taxa i and j against $p - 2$ taxa, which becomes unsolvable when the number of taxa p is larger than the number of samples n (e.g. in typical microbiome studies $p = 150 - 500$ and $n < 50$). Hence, the first step in our approach is to calculate a one-dimensional approximation of the control variables (e.g. all taxa except the pair i and j), through the first principal

component. Using simulations (see the 'Effect of using different number of components in the deflation' section in Appendix D), we demonstrate that using a single component produces more accurate network inference than using two or three components to approximate the control variables specifically when sample sizes are small. However, this may not be true for all datasets. Therefore, in our methods (i.e. LUPINE_single and LUPINE), we encourage users to explore different numbers of components when approximating the control variables in the model.

We first start with a generic formulation of PCA [13]. The objective function to obtain the first principal component is

$$\operatorname{argmax}_{\|\mathbf{w}\|=1} \operatorname{var}(\mathbf{w}X) \quad (1)$$

where \mathbf{w} is the p dimensional loading vector associated to the first principal component \mathbf{u} , with $\mathbf{u} = \mathbf{w}X$ under the constraint that \mathbf{w} is of unit (norm) 1.

First, we modify Eq. (1) to find the first principal component of the control taxa that excludes taxa i and j as

$$\operatorname{argmax}_{\|\mathbf{w}^{-(i,j)}\|=1} \operatorname{var}(\mathbf{w}^{-(i,j)}X^{-(i,j)}) \quad (2)$$

where $\mathbf{w}^{-(i,j)}$ is the $(p - 2)$ dimensional loading vector associated to the first principal component $\mathbf{u}^{-(i,j)}$ with $\mathbf{u}^{-(i,j)} = \mathbf{w}^{-(i,j)}X^{-(i,j)}$.

Next, we fit two log linear regressions on each of the taxa i and j . They are regressed against the first principal component, $\mathbf{u}^{-(i,j)}$, similar to principal component regression (PCR, [14]). PCR combines PCA and multiple linear regression, where the principal components of the explanatory variables serve as predictors rather than using the original variables directly in the regression. Here, we fit two log linear regressions to extract their residuals denoted by \mathbf{e}_i and \mathbf{e}_j :

$$\mathbf{e}_i = \log(X^i) - E(\log(X^i)|\mathbf{u}^{-(i,j)}); \quad \mathbf{e}_j = \log(X^j) - E(\log(X^j)|\mathbf{u}^{-(i,j)}) \quad (3)$$

where $E(\cdot|\mathbf{u}^{-(i,j)})$ is the conditional expectation with respect to the first principal component $\mathbf{u}^{-(i,j)}$. We use a log linear regression as the response variables are counts. This is similar to the bias-corrected version of ANCOM [15] that uses a linear regression framework based on log-transformed taxa counts. We add an offset of 1 to zero counts to ensure that the log transformation is valid. Additionally, in the 'Effect of normalisation' section in Appendix D, we assess the impact of using log-transformed counts in Eq. 2.

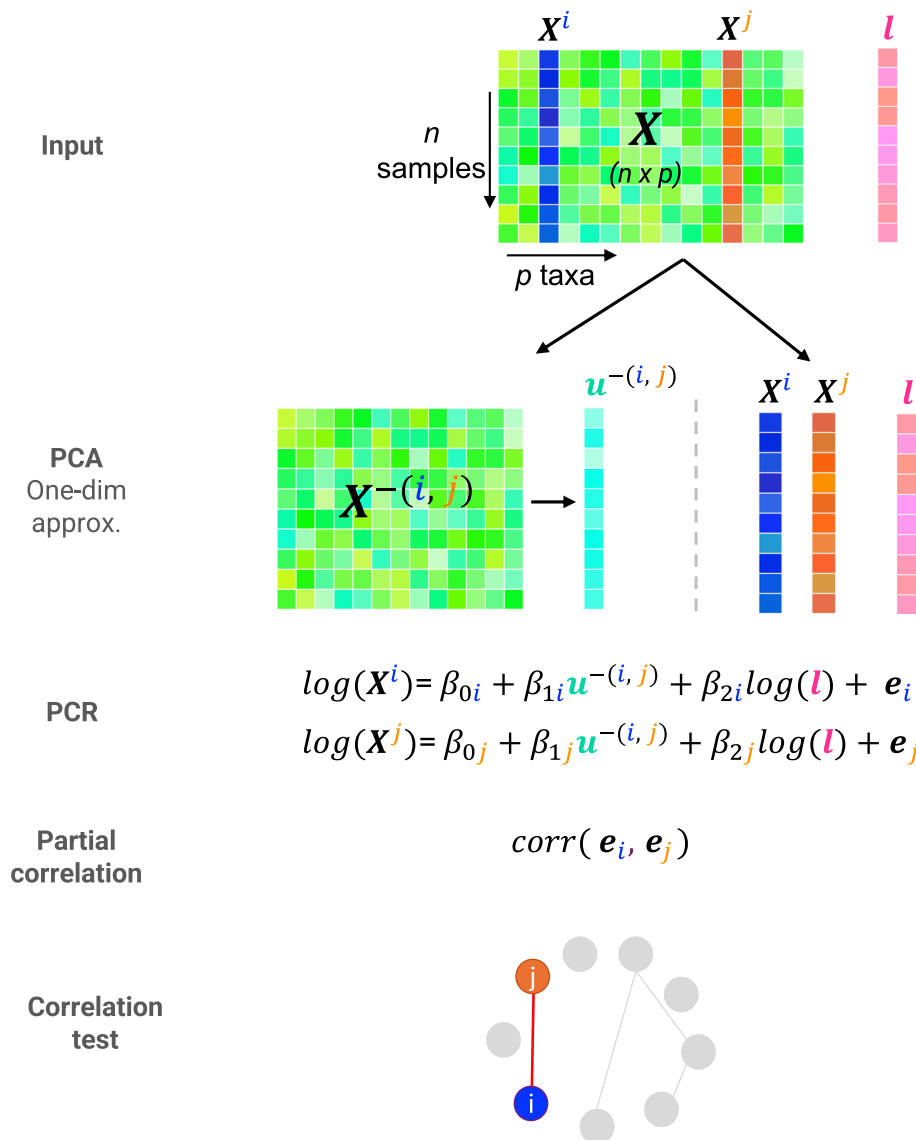


Fig. 1 Single time point modelling overview. Consider the $(n \times p)$ input data matrix X and l the library size of each sample. First, we use PCA to derive a one-dimensional approximation, $u^{-(i,j)}$ to control for taxa other than i and j from the X data matrix (i.e. $X^{-(i,j)}$). We then apply two independent log linear regression models (a.k.a. principal component regression) for X^i and X^j , regressed on component $u^{-(i,j)}$ and $\log(l)$ to obtain the residuals e_i and e_j . We estimate the partial correlation estimate for taxa i and j as the correlation between the two residuals. Finally, we test for the statistical significance of the true correlation between e_i and e_j to infer whether taxa i and j are connected in the network. We iterate this process for all pairs of taxa (i, j) to obtain a full binary network. The ‘Improvement in computational time’ section explains how to avoid repeated PCA approximations to improve computational time

In addition, to correct for library size (i.e. the variation in the total sequence reads), we modify Eq. 3 and include an offset, similar to the approach of [16, 17]:

We then quantify the strength of association between taxa i and j based on partial correlation using the residuals e_i and e_j . Partial correlations reveal direct (i.e. condi-

$$e_i = \log(X^i) - E(\log(X^i) | u^{-(i,j)}, \log(l)); \quad e_j = \log(X^j) - E(\log(X^j) | u^{-(i,j)}, \log(l)) \quad (4)$$

where $E(\cdot | u^{-(i,j)}, \log(l))$ is the conditional expectation with respect to the first principal component $u^{-(i,j)}$ and the log library size $\log(l)$.

tionally independent) associations between two variables while controlling for the other variables. We calculate the partial correlation between the pair of taxa i and j as

$$\pi_{e_i, e_j} = \text{corr}(\mathbf{e}_i, \mathbf{e}_j). \quad (5)$$

Finally, we test whether the true correlation ρ_{e_i, e_j} between \mathbf{e}_i and \mathbf{e}_j (estimated with π_{e_i, e_j}) is significantly different from zero. The correlation test assumes that when the correlation is zero, the test statistic $t = \pi_{e_i, e_j} \sqrt{\frac{n-2}{1-\pi_{e_i, e_j}^2}}$, follows a normal distribution. Thus, we test the hypothesis $H_0 : \rho_{e_i, e_j} = 0$ versus $H_1 : \rho_{e_i, e_j} \neq 0$ using a t-distribution with $n - 2$ degrees of freedom. We then consider that a pair of taxa is connected in a network if the associated p -value from the correlation test is less than a chosen significance level (e.g. 0.05). We apply this approach to all possible pairs of taxa to obtain a binary network summarising the pairwise connections between all taxa. In the ‘[Permutation test vs. Correlation test](#)’ section in Appendix D, we compare this analytical approach with a permutation-based p -value calculation.

To quantify the uncertainty of the correlation test p -value, we use bootstrapping, which helps determine the significance of a correlation. We resample the data set 1000 times and calculate the p -values from correlation tests for each iteration. This process generates a confidence interval for each p -value, allowing us to better assess the significance of the correlation. We demonstrate this approach in case study 4.

In the pseudo-algorithm 1 and Fig. 1, we outline the different steps in LUPINE_single. In the [Improvement in computational time](#) section, we explain how we improve computational time by avoiding repeated PCA approximations.

$$\mathbf{e}_{it} = \log(X_t^i) - E(\log(X_t^i) | \mathbf{u}_t^{-(i,j)}, \log(\mathbf{l}_t)); \quad \mathbf{e}_{jt} = \log(X_t^j) - E(\log(X_t^j) | \mathbf{u}_t^{-(i,j)}, \log(\mathbf{l}_t)) \quad (8)$$

Algorithm 1 Single time point scheme

Input: \mathbf{X} ($n \times p$): count matrix of n samples and p taxa
 \mathbf{l} : vector of n library sizes

Output: π_{e_i, e_j} (partial correlation) and p_{e_i, e_j} (p -value) for all pairs of taxa (i, j)

- 1: **for** each pair of taxa (i, j) **do**
- 2: Extract first principal component $\mathbf{u}^{-(i,j)}$ from $\mathbf{X}^{-(i,j)}$ using PCA
- 3: Extract residuals, \mathbf{e}_i and \mathbf{e}_j by fitting log linear regressions f_i and f_j as:
- 4: $f_i \rightarrow \log(X^i) \sim \mathbf{u}^{-(i,j)} + \log(\mathbf{l})$
- 5: $f_j \rightarrow \log(X^j) \sim \mathbf{u}^{-(i,j)} + \log(\mathbf{l})$
- 6: Compute partial correlation $\pi_{e_i, e_j} = \text{corr}(\mathbf{e}_i, \mathbf{e}_j)$
- 7: Test significance of ρ_{e_i, e_j} and extract p -value p_{e_i, e_j}
- 8: **end for**

Two time point modelling with PLS regression

Here, we use PLS to obtain a one-dimensional approximation between two datasets from two time points (Fig. 2). PLS is a projection-based method that maximises the covariance between the latent components associated to two data sets while managing correlated information

[10]. The PLS latent variables (or components), are linear combinations of variables from each data set. These latent variables aim to uncover subtle biological effects not evident otherwise if each data set is considered independently [12].

The objective function for the first dimension of PLS is

$$\underset{\|\mathbf{w}_{t-1}\|=1, \|\mathbf{w}_t\|=1}{\text{argmax}} \quad \text{Cov}(\mathbf{w}_{t-1} \mathbf{X}_{t-1}, \mathbf{w}_t \mathbf{X}_t) \quad (6)$$

where \mathbf{w}_{t-1} and \mathbf{w}_t are the two p dimensional loading vectors associated to the first latent components $\mathbf{u}_{t-1} = \mathbf{w}_{t-1} \mathbf{X}_{t-1}$ and $\mathbf{u}_t = \mathbf{w}_t \mathbf{X}_t$.

We modify Eq. 6 to find the first latent components of the controlling taxa at time point t that has a maximum covariance with the preceding time point latent component as

$$\underset{\|\mathbf{w}_{t-1}^{-(i,j)}\|=1, \|\mathbf{w}_t^{-(i,j)}\|=1}{\text{argmax}} \quad \text{Cov}(\mathbf{w}_{t-1}^{-(i,j)} \mathbf{X}_{t-1}^{-(i,j)}, \mathbf{w}_t^{-(i,j)} \mathbf{X}_t^{-(i,j)}) \quad (7)$$

where $\mathbf{w}_{t-1}^{-(i,j)}$ and $\mathbf{w}_t^{-(i,j)}$ are the two $(p-2)$ loading vectors associated to the first latent components $\mathbf{u}_{t-1}^{-(i,j)} = \mathbf{w}_{t-1}^{-(i,j)} \mathbf{X}_{t-1}^{-(i,j)}$ and $\mathbf{u}_t^{-(i,j)} = \mathbf{w}_t^{-(i,j)} \mathbf{X}_t^{-(i,j)}$.

To infer the network at time point t using the information from the preceding time point $(t-1)$, we fit two log linear regressions. These regressions are fitted on X_t^i and X_t^j that are regressed against the first latent component $\mathbf{u}_t^{-(i,j)}$ and the log library size $\log(\mathbf{l}_t)$. Note that $\mathbf{u}_t^{-(i,j)}$ differs from the single time point approach $\mathbf{u}^{-(i,j)}$ as we maximise the covariance with the preceding time point latent component $\mathbf{u}_{t-1}^{-(i,j)}$. The updated Eqs. 4 and 5 become

$$\pi_{e_{it}, e_{jt}} = \text{corr}(\mathbf{e}_{it}, \mathbf{e}_{jt}) \quad (9)$$

where $E(\cdot | \mathbf{u}_t^{-(i,j)}, \log(\mathbf{l}_t))$ is the conditional expectation with respect to the first latent component $\mathbf{u}_t^{-(i,j)}$ and the log library size, $\log(\mathbf{l}_t)$ at time point t ; and \mathbf{e}_{it} and \mathbf{e}_{jt} are the residuals from the two log linear models.

Similar to the single time point scheme described above, we test for the strength of association between taxa i and j with a correlation test and obtain a binary network summarising the pairwise connections between all taxa after controlling for all other taxa from current time point and previous time point.

Multiple time point modelling with blockPLS regression

We use blockPLS, an extension of PLS, to calculate a one-dimensional approximation that maximises the

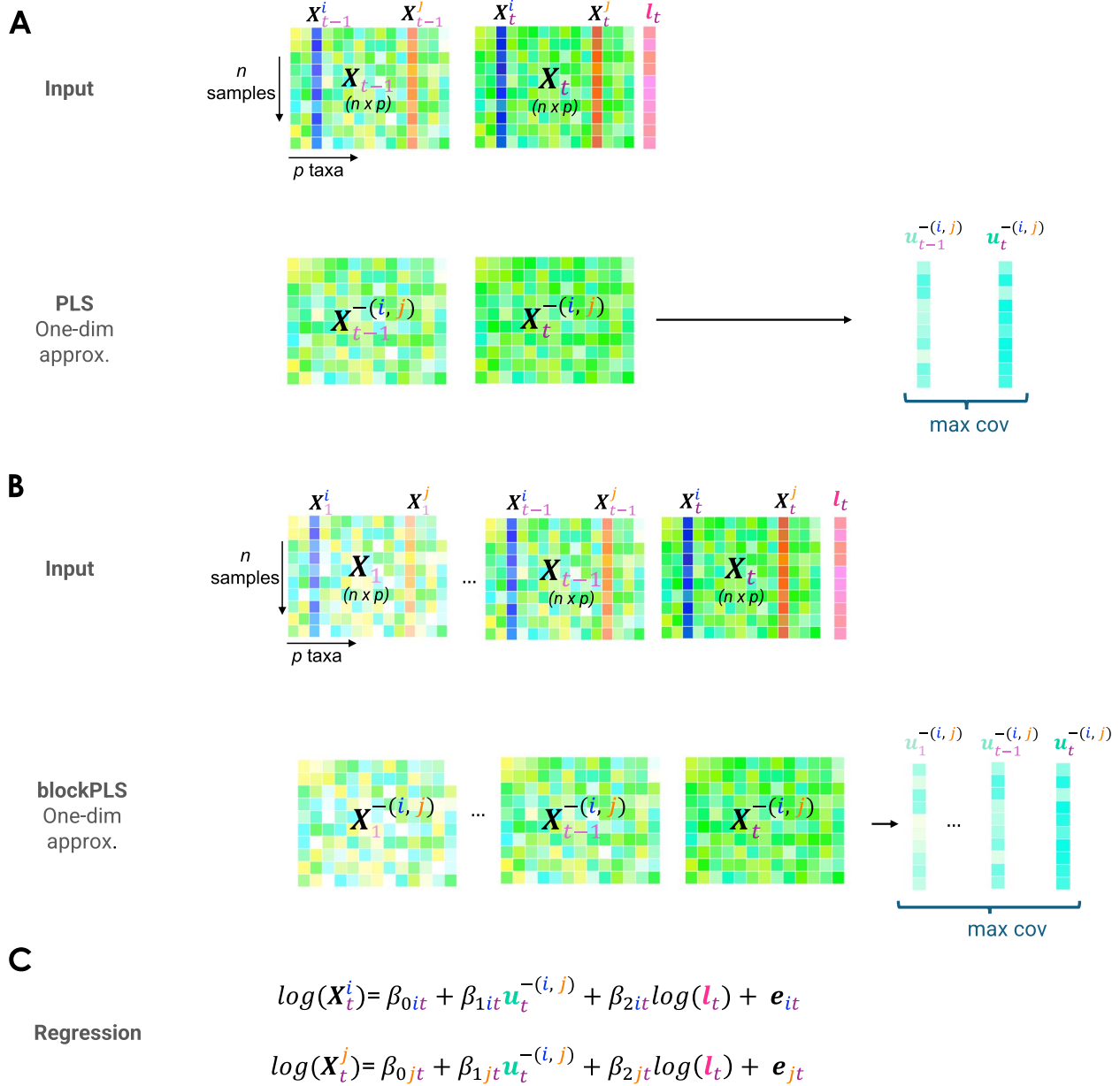


Fig. 2 Two or more time points modelling overview. Consider as input the $(n \times p)$ data matrices for **A** two time points or **B** more than two time points (denoted $X_1 \dots, X_t$) and their respective library size l_t at time t . Similar to single time point modelling presented in Fig. 1, first, we derive a one-dimensional approximation, $u_t^{-(i,j)}$ to control for taxa other than i and j . Instead of using PCA as previously for a single time point, we use either **A** PLS maximising the covariance with the control taxa at the preceding time $t - 1$, or **B** blockPLS maximising the covariance with the control taxa at all previous times $1, \dots, t - 1$. **C** We then fit two independent log linear models for X_t^i and X_t^j , regressed on $u_t^{-(i,j)}$ and $\log(l_t)$ to compute the residuals e_{it} and e_{jt} following either **A** or **B**. The remainder of the modelling process, including the partial correlation calculation and the correlation test, is identical to that in Fig. 1

covariance between latent components associated to the current time point and all past time points. BlockPLS is based on generalised PLS [11, 12] that involves regressing the response matrix (e.g. X_t) on multiple datasets (X_1, X_2, \dots, X_{t-1}).

The objective function for the first dimension of block-PLS is

$$\underset{\|w_1\|=1, \dots, \|w_t\|=1}{\operatorname{argmax}} \sum_{q,k=1, q \neq k}^t \operatorname{Cov}(w_q X_q, w_k X_k) \quad (10)$$

where \mathbf{w}_q is a p dimensional loading vector associated to the first latent component \mathbf{u}_q , where $\mathbf{u}_q = \mathbf{w}_q \mathbf{X}_q$.

We modify Eq. 10 to find the first latent components of all controlling taxa at time point t that has a maximum covariance with previous time points latent components as

$$\underset{\|\mathbf{w}_1^{-(i,j)}\|=1, \dots, \|\mathbf{w}_t^{-(i,j)}\|=1, q,k=1, q \neq k}{\operatorname{argmax}} \sum_{t=1}^t \operatorname{Cov}(\mathbf{w}_q^{-(i,j)} \mathbf{X}_q^{-(i,j)}, \mathbf{w}_k^{-(i,j)} \mathbf{X}_k^{-(i,j)}) \quad (11)$$

where $\mathbf{w}_q^{-(i,j)}$ is a $p-2$ dimensional loading vector associated to the first latent component $\mathbf{u}_q^{-(i,j)}$ at time point q , with $\mathbf{u}_q^{-(i,j)} = \mathbf{w}_q^{-(i,j)} \mathbf{X}_q^{-(i,j)}$.

To infer a network at any time point t using all time points until time point t , we fit two log linear regressions on the counts of \mathbf{X}_t^i and \mathbf{X}_t^j against the first latent compo-

$$\begin{bmatrix} \mathbf{u}_1^{-(i,j)} \\ \mathbf{u}_2^{-(i,j)} \\ \vdots \\ \mathbf{u}_n^{-(i,j)} \end{bmatrix}_{n \times 1} \approx \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}_{n \times p} \times p \begin{bmatrix} w_1 & w_2 & \cdots & w_i = 0 & \cdots & w_j = 0 & \cdots & w_p \end{bmatrix}_{1 \times p}^T \quad (13)$$

nent at t , $\mathbf{u}_t^{-(i,j)}$, and log library size at t , $\log(\mathbf{l}_t)$. However, we need to keep in mind that the latent component $\mathbf{u}_t^{-(i,j)}$ maximises the covariance with all previous time points latent variables, $[\mathbf{u}_1^{-(i,j)}, \dots, \mathbf{u}_{t-1}^{-(i,j)}]$, thus effectively taking into account all past time points. We then use Eqs. 8 and 9 to quantify and test the statistical significance of the partial correlation between taxa i and j .

Pseudo-algorithm 2 outlines the different steps in either two time points and multiple time points modelling.

Algorithm 2 Multi time point scheme

Input: $\mathbf{X}_t (n \times p)$: count matrix of n samples and p taxa at time t
 \mathbf{l}_t : vector of n library sizes at time t

Output: $\pi_{e_{it}, e_{jt}}$ (partial correlations) and $p_{e_{it}, e_{jt}}$ (p-value) for all pairs of taxa (i, j) at time t

```

1: for  $t = 2, \dots, t$  do
2:   for each pair of taxa  $(i, j)$  do
3:     if two time modelling then
4:       Extract first latent component  $\mathbf{u}_t^{-(i,j)}$  using PLS regression
         between  $\mathbf{X}_t^{-(i,j)}$  and  $\mathbf{X}_{t-1}^{-(i,j)}$ 
5:     else
6:       Extract first latent component  $\mathbf{u}_t^{-(i,j)}$  using blockPLS regression
         between  $\mathbf{X}_t^{-(i,j)}$  and  $\mathbf{X}_{t-1}^{-(i,j)}, \dots, \mathbf{X}_1^{-(i,j)}$ 
7:     end if
8:     Extract residuals,  $e_{it}$  and  $e_{jt}$  by fitting log linear regressions  $f_{it}$  and  $f_{jt}$  as:
9:        $f_{it} \rightarrow \log(\mathbf{X}_t^i) \sim \mathbf{u}_t^{-(i,j)} + \log(\mathbf{l}_t)$ 
10:       $f_{jt} \rightarrow \log(\mathbf{X}_t^j) \sim \mathbf{u}_t^{-(i,j)} + \log(\mathbf{l}_t)$ 
11:     Compute partial correlation  $\pi_{e_{it}, e_{jt}} = \operatorname{corr}(e_{it}, e_{jt})$ 
12:     Test significance of  $\pi_{e_{it}, e_{jt}}$  and extract p-value  $p_{e_{it}, e_{jt}}$ 
13:   end for
14: end for

```

Note that in both PCA and PLS versions we centre and scale each taxon to avoid scaling issues.

Improvement in computational time

Our approach requires to compute one-dimensional approximation for each pair of taxa (i, j) , i.e. $p * (p-1)/2$ times. Using matrix decomposition principles, we propose instead to approximate the full data matrix \mathbf{X} with the first one-dimension component \mathbf{u} that summarises most of the variation in \mathbf{X} , and its associated loading vector that are obtained from PCA, PLS, or blockPLS:

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}_{n \times p} \approx \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}_{n \times 1} \times \begin{bmatrix} w_1 & w_2 & \cdots & w_p \end{bmatrix}_{1 \times p} \quad (12)$$

Using the approximation from Eq. 12, we then can calculate the $\mathbf{u}^{-(i,j)}$ vector by setting the loading coefficients w_i and w_j of the pair of taxa (i, j) to zero:

This allows us to obtain the vectors \mathbf{u} and \mathbf{w} only once. We show the efficacy of this approximation in Appendix A.

Comparisons to other network inference methods

Existing longitudinal network methods either infer networks for a single individual or derive a single network from all data [9], which can be problematic when we expect changes in taxa associations over time. LUPINE infers networks sequentially to include the information learned across time. We compared our results with two state-of-the-art single time point network inference methods, SpiecEasi [18] and SparCC [18], using simulations where the true network is known. SpiecEasi and SparCC represent two distinct branches of network analysis. Specifically, SparCC employs a correlation-based approach, while SpiecEasi uses a conditional dependence approach. The SpiecEasi method further includes two methods based on different graphical model estimators, representing different types of network analysis: (i) neighbourhood selection and (ii) sparse inverse covariance selection, which we will refer to as ‘SpiecEasi_mb’ and ‘SpiecEasi_glasso’, respectively.

Measures for network comparisons

We used two distinct measures to identify similarities and dissimilarities among the inferred networks: a network distance measure for quantitatively evaluating the network topology (see Fig. 3B1–B2) and a node-wise measure for assessing the influential nodes in the network (see

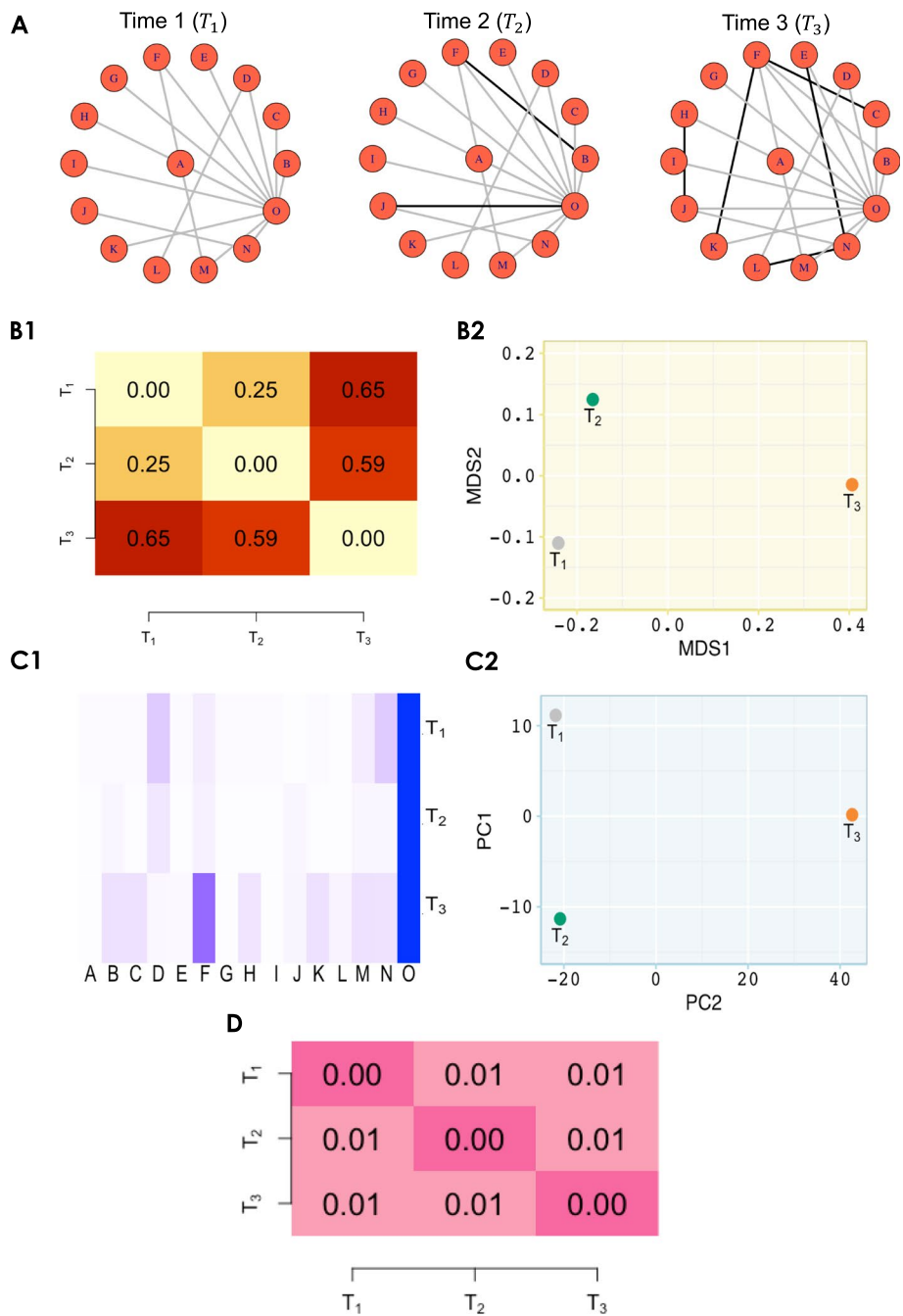


Fig. 3 Toy example for network comparison. **A** Consider three networks across three time points, where nodes represent taxa and edges represent significant association between taxa. Black edges denote new connections compared to the preceding time point. **B** Network dissimilarity through graph diffusion distance (GDD). **B1** Pairwise GDD distances reveal that the network at T_3 is dissimilar to the networks at T_1 and T_2 . **B2** Transformation of pairwise GDD into a lower-dimensional space using classical multidimensional scaling (MDS) represents the separation of time 3 network from the other two networks. **C** Network discrimination through influential nodes (i.e. taxa). Node influence is computed by integrating local, semi-local, and global measures. **C1** Heatmap showing the node influence (more influential = darker blue shades). For example, node 'O' is influential in all time points, but node 'F' is more influential at T_3 rather than T_1 and T_2 . **C2** Principal component analysis (PCA) plot to represent the networks based on their respective node influence. Similar to what we observed with the GDD distance, the largest source of variation is explained between the network at T_3 and the networks at earlier time points. **D** A heatmap of p -values from the correlation tests between each pair of networks indicates that all three networks are correlated

Fig. 3C1–C2). Additionally, we measured the differences between network pairs using hypothesis testing (Fig. 3D).

Pairwise distance between networks

We use graph diffusion distance (GDD) [19] to measure pairwise differences in network typologies. GDD assesses the average similarity between two networks by exploring the information flow and connectivity within their structures, based on the heat diffusion process on graphs. The GDD score is calculated by searching for a diffusion time that maximises the Frobenius norm of the difference between the diffusion kernels. By calculating the GDD for all pairs, we obtain a pairwise distance matrix representing the network distances between pairs of graphs. We then transform these pairwise distances into a lower-dimensional space using classical multidimensional scaling (MDS) [20]. MDS maintains pairwise dissimilarities between objects. As a result, in the MDS plot, points that represent each network are deemed similar if they are positioned closely and dissimilar if they are spaced further apart. This visual technique allows us to compare multiple networks at once (Fig. 3B2).

Node influence

To assess influential nodes in the network, we use the integrated value of influence (IVI) algorithm that integrates local (degree centrality and ClusterRank), semi-local (neighbourhood connectivity and local H-index), and global (betweenness centrality and collective influence) measures [21]. IVI combines these measures to capture both local prominence as well as the broader impact of the nodes. After calculating the IVI values for all nodes, we perform a PCA on the resulting matrix, where the rows represent different networks, and the columns represent different nodes (here taxa). While MDS aims to preserve pairwise distances or dissimilarities as much as possible, PCA prioritise explaining the maximum variation in the data and highlighting strong patterns (Fig. 3C2).

Correlation test between two networks

To evaluate correlations between two network adjacency matrices, we first calculate pairwise Hamming distances for each matrix [22]. This step allows us to compare differences in connections between nodes. Next, we apply the Mantel test [23], a permutation-based method, to assess whether the correlation between these distance matrices is statistically significant. By comparing the observed correlation to a distribution generated from random permutations, the Mantel test evaluates its significance, providing a robust measure of the relationship between the two matrices.

Simulation and case studies

In this section, we outline our simulation strategy and detail the four case studies analysed with LUPINE.

Simulation study

In our simulation, we used two realistic networks based on our high-fat high-sugar (HFHS) case study described in the [Case studies](#) section. The two networks were inferred using the SpiecEasi method [5] where concatenated the data across days for each diet. Due to computational constraints, we restricted our simulation to the sub-network of the *Bacteroidales* order, resulting in 54 nodes. The networks used across the two time periods are represented in Fig. 4. We simulated data using the network in Fig. 4A until day 5, then transitioned to the network in Fig. 4B from day 6 to day 10 to reflect a network change. We used a multivariate Poisson distribution to simulate the count data. For each simulation, a new correlation matrix was generated from the two networks. To explore the impact of sample size, we varied the sample size from 23 to 50, and then to 120. For each sample size, 50 data sets were simulated, each with 54 taxa and 10 time points. More details on the simulation strategy are provided in the [Appendix B](#).

Using simulated data, we also conducted a sensitivity analysis in [Appendix D](#) to evaluate the robustness of LUPINE and LUPINE_single under various conditions. The analysis was designed to determine whether the model's performance remains consistent when different parameters or assumptions are altered.

Case studies

We analysed three 16S rRNA amplicon datasets and one metagenomic dataset, the experimental designs of which are outlined in [Table 1](#). For the 16S datasets which were available in raw counts, we grouped data by 'group' and 'time point' and calculated the relative abundance of each taxa within these specific groups and time points. Any taxa with a relative abundance less than 0.1% across all groups and all time points were excluded. This filtering process, while significantly reducing the number of taxa, aligns with the goal of conducting a more focused analysis [24]. However, we only calculated the partial correlations for taxa that had a mean relative abundance exceeding 0.1% at a given time point and group. This approach helps us avoid identifying connections between taxa with low abundance at certain time points in specific groups. As a result, there were no edges in the network from the taxa with low abundance at that time point and group. When we fit a longitudinal model on a given time point, we only consider the taxa filtered on that specific time point and group, and across the past time points.

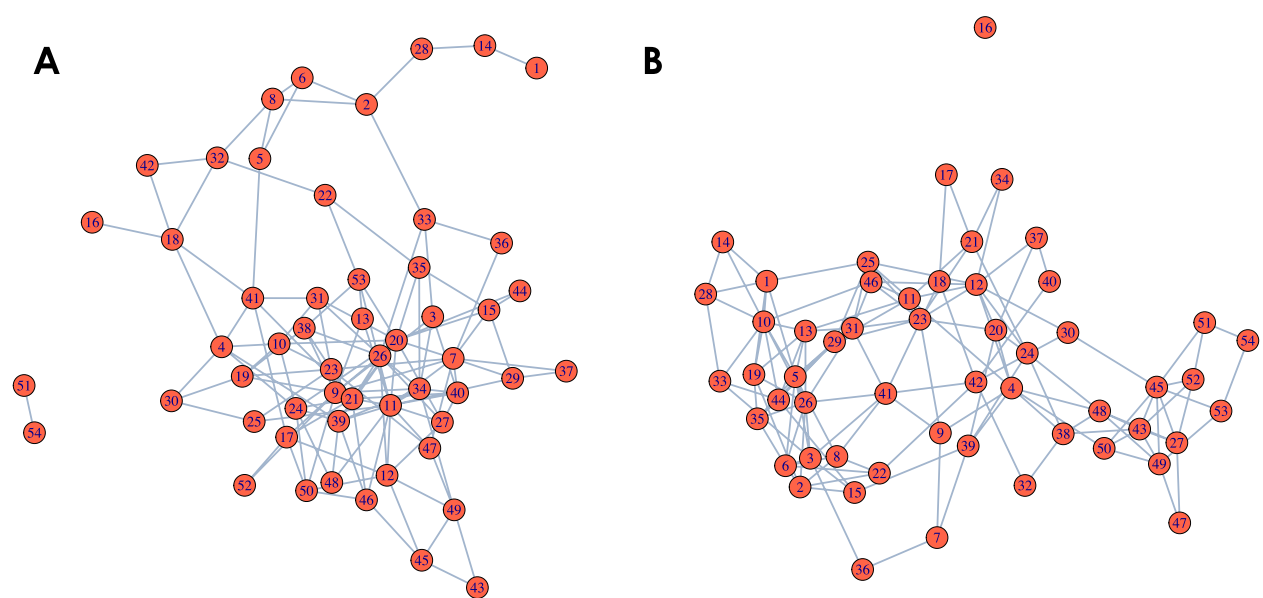


Fig. 4 Networks used in simulation. Inferred network from SpiecEasi for **A** normal diet and **B** HFHS diet among *Bacteroidales*. These networks form the basis of our simulation for the two time periods (i.e. from day 1 to 5 and from day 6 to 10)

Case study 1: HFHS (a four time point case-control mouse study)

We investigated the effect of a high-fat high-sugar (HFHS) diet on the mouse microbiome, as outlined in [25]. To assess the diet effect on the gut microbiome, 47 C57/B6 female black mice were fed with either an HFHS or a normal diet. Fecal samples were collected at days 0, 1, 4, and 7.

The raw data contained 1172 taxa across all four time points. We removed one outlier mouse from the normal diet group with a very large library size. After filtering, we modelled 102,107,105, and 91 taxa respectively for the normal diet for each day and 99, 147, 85, and 92 taxa for the HFHS diet.

Case study 2: VREfm (an eleven time point mouse study with two interventions)

Mu et al. [26] investigated the functional roles of the gut microbiome during vancomycin-resistant *Enterococcus faecium* (VREfm) colonisation. To assess the bacterial community composition in the murine gut, 9 C57BL/6 mice (co-housed wild-type males) were monitored. Fecal samples were collected over 14 days from the same three mice. Over the course of 14 days, the mice underwent two interventions: (1) administration of ceftriaxone treatment at a concentration of 0.5 g/liter in the drinking water over 2 days on days 6 and 7 and (2) colonisation with a dosage of (1×10^6) VREfm ST796 on day 9. The days before the antibiotic administration were considered as the naive phase of the experiment. The 2 days when

Table 1 Summary of experimental frameworks: key parameters for each case study

Name	Sample size	Number of taxa	Time points	Collection	Interventions
High-fat high-sugar (HFHS)	HFHS = 23 Normal = 24	212	4	Days (0, 1, 4, 7)	NA
Vancomycin-resistant <i>Enterococcus faecium</i> (VREfm) colonisation	9	126	11	Days (0–2, 5–9, 12–14)	Ceftriaxone antibiotic treatment (days 6–7) VREfm colonisation (day 9)
Diet study	Plant diet = 10 Animal diet = 10	311	15	Days (–4, –3, –2, –1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10)	Diet intervention (days 0–4)
Pre-diabetic study	MDE diet ^a = 96 PPT diet ^b = 93	91	2	Months (0,6)	Diet intervention (months 0–6)

Number of taxa are indicated after the filtering step described in the ‘Case studies’ section
^aMediterranean diet, ^bPersonalized postprandial glucose-targeting diet

the antibiotic was given were considered the antibiotic phase, and all other days after that were considered the VRE phase.

The data contained 3574 taxa across eleven time points. Day 8 was excluded from the analysis due to insufficient taxonomic counts for two mice. After filtering, we considered the following number of taxa in each phase: naive (53, 63, 58, 72), antibiotic (34, 22) and VRE (10, 15, 15, 15).

Case study 3: Diet (a fifteen time point case-control human study with one intervention)

David et al. [27] investigated the human gut microbiome responses to a short-term diet intervention. To evaluate the dynamics of the diet intervention, they assigned 10 participants to each diet intervention, namely the ‘plant-based’ and ‘animal-based’ diets for five consecutive days. The study participants were observed for 4 days before the diet intervention, a period referred to as the ‘baseline period’ to study their regular eating habits. After intervention, a 6-day ‘washout period’ was also observed to assess microbial recovery. Data were obtained from the MITRE repository [28]. The data contained 17,310 taxa across all time points. As taxonomic details were lacking, we used the sequence information along with dada2 reference database ‘silva_nr99_v138.1_wSpecies_train_set.fa.gz’ for taxonomic classification [29].

After filtering, we retained the following number of taxa in the plant-based diet group for each period: baseline (105, 100, 111, 103), intervention (118, 116, 116, 114, 114), and washout (121, 120, 90, 96, 115, 109); and in the animal-based diet group: baseline (120, 109, 128, 115), intervention (113, 113, 108, 105, 110), and washout (112, 104, 106, 105, 107, 123). Across the duration of the experiment, only 8–15 samples were obtained from each participant because of unsuccessful sampling at certain time points. Thus, we imputed the missing values using a cubic splines on the centred log ratio (clr) transformed data. This was motivated by the linear interpolation and cubic splines interpolation adopted in [24, 30]. Since, clr transformation accounts for library size differences, we used Eq. 3 instead of Eq. 4 in our modelling.

Case study 4: Pre-diabetic (large human case control study)

Shoer et al. [31] investigated the impact of a personalized postprandial glucose-targeting diet (PPT) and Mediterranean diet (MED) on gut microbiome on 200 pre-diabetic individuals. MED is the standard of care for pre-diabetes. Participants provided fecal samples before and after a 6-month diet intervention.

The data contained 378 fecal gut samples with 605 species-level genome bins (SGBs) given as unique relative abundance (URA, [32]). After removing *Archaea* and

unknown kingdoms, we were left with 505 features, but with many missing values. We only kept features with less than 25% missing values across all samples, resulting in 91 SGBs.

Results

We first evaluated our two approaches against two widely recognised network inference methods (described in the ‘[Comparisons to other network inference methods](#)’ section) for cross-sectional microbiome data. Using simulated data, our objective was to demonstrate the effectiveness and computational efficiency of our proposed approaches compared to commonly used single time point methods. We then highlighted the advantages of LUPINE over its single time point counterpart LUPINE_single in the HFHS study.

After this benchmark step, we applied LUPINE in four case studies. In the longitudinal case-control mouse study, we evaluated whether the difference in diets could be detected in our network models. In the mouse study with two interventions, we evaluated the effectiveness of LUPINE in detecting abrupt changes. In the human case-control study, we assessed the ability of LUPINE to handle inherent high variability typical of human studies across fifteen time points. Finally, in the metagenomic human case-control study, we assessed LUPINE’s capacity to generate biological inferences in a large human study with two time points.

Benchmarking analysis

Simulation study: LUPINE methods outperformed existing single time point methods

We first compared our proposed longitudinal method (LUPINE) and its single time point version (LUPINE_single), with two SpiecEasi models and SparCC. Note that all methods except LUPINE were designed for single time point modelling.

We assessed model performance across 50 simulated datasets, focusing on scenarios where the number of samples was smaller than the number of variables. We used two metrics: area under the receiver operating characteristic curve (AUC-ROC) and area under the precision-recall characteristic curve (AUC-PRC). True positives were defined as the edges present in the inferred network that also existed in the simulated (true) network at each time point. For LUPINE_single, LUPINE, and SparCC, we used p -values from the correlation test to infer edges between pairs of nodes. A higher value for $1 - p$ -value indicated a greater likelihood of an inferred edge, with the value reflecting the strength of the connection between a pair of taxa. For the two SpiecEasi methods, stability was used to measure the strength of the inferred connection.

The AUC-ROC captures the trade-off between sensitivity (true positive rate) and specificity (false positive rate) across different thresholds ($1 - p$ -value or stability), providing a broad overview of model performance. In contrast, the AUC-PRC focuses on precision (positive predictive value) versus recall (true positive rate), which is particularly informative for scenarios where the number of true edges is much smaller than the number of non-edges. By analysing these two metrics, we gained insights into the model's ability to distinguish true edges from spurious edges depending on network sparsity. Higher values in both metrics signified better model performance, with AUC-ROC evaluating overall discriminatory power and AUC-PRC evaluating precision and recall characteristics.

Our proposed methods and SparCC outperformed SpiecEasi with higher AUC-ROC values (Fig. 5A). However, all methods had similar AUC-PRC values (Fig. 5B). Additionally, both our approaches were computationally more efficient than SparCC (Fig. 5C). When the sample size increased, LUPINE and LUPINE_single led to superior model performance based on AUC-ROC and AUC-PRC values (Figs. 16 and 17 in Appendix C).

Next, we compared the inferred networks across time points, as explained in the 'Measures for network comparisons' section. For SparCC and our methods, taxa connections were determined via a p -value cut off (this was not required for SpiecEasi methods, as they inherently output a binary network).

To compare networks based on structure, we computed the pairwise GDDs for all inferred networks across time points, then averaged these distances for each time point across all simulations. In our simulation (described in the 'Simulation study' section), we generated two stable networks that were distinct, across two time periods (days 2–5 and 6–10). Thus, we expect the inferred networks to be similar within each period and dissimilar between the two periods. However, SpiecEasi displayed the largest variation between the inferred and the actual networks in the MDS plot of the pairwise GDDs (Fig. 6A2–A3). In contrast, our two proposed methods (Fig. 6A4–A5) and SparCC (Fig. 6A1) showed the largest between-period variation, regardless of whether the networks were true or inferred.

We applied the same procedure used for GDDs to conduct node influence based comparisons based on IVI values. Similar to the MDS plots, SparCC (Fig. 6B1) and LUPINE methods (Fig. 6B4–B5) explained the greatest between-period variation in the PCA plots. While all methods identified the most influential nodes (Fig. 6C1–C2), the two SpiecEasi methods showed less pronounced influential nodes due to their sparse network inference, compared to the other methods. Both LUPINE methods and SparCC successfully distinguished between the networks from days 2 to 5 and days 6 to 10 (Fig. 6D1–D5).

As expected, the correlations were significant within the day 2 to 5 and day 6 to 10 intervals, but not between days within each interval.

HFHS study: LUPINE highlighted more robust longitudinal network patterns than LUPINE_single

We compared LUPINE and LUPINE_single on the HFHS study. This study includes a small number of time points (Fig. 8A), where mice were subjected to either a HFHS or a normal diet.

Both LUPINE_single and LUPINE differentiated the diet groups in the first axis of the MDS and PCA plots (Fig. 7). For LUPINE_single, microbial networks from the normal diet group and the HFHS diet at day 0 were clustered, indicating a similar network structure at the onset of the two different diets (Fig. 7A2). However, we did not observe a close cluster of microbial networks within the normal diet networks in the PCA plot of the IVI scores (Fig. 7A2), suggesting that the node influence within the normal diet group is changing over time.

In contrast, for LUPINE, the groupings in the MDS plot (Fig. 7B1) closely matched to those observed in the PCA plot (Fig. 7B2), indicating consistent network results based on both GDD and IVI scores. Thus, these results suggest that LUPINE is able to model robust longitudinal patterns. Note that the initial day 0 microbial network is absent from the longitudinal scheme as we require at least one prior time point for inference. However, we found that both approaches successfully identified significant correlations within each diet group across days, while correlations between diet groups across days were insignificant, particularly later in the diet (Fig. 7A3–B3).

In this benchmarking section, we first demonstrated the superior model performance of both LUPINE and LUPINE_single using AUC-ROC and AUC-PRC values. We then highlighted the advantages of LUPINE in comparison to our own LUPINE_single approach. We showed that LUPINE led to robust longitudinal network patterns. The following sections focus on the biological interpretation of the microbial communities identified by LUPINE in four case studies.

HFHS study: LUPINE identified taxonomic orders that differentiate microbial networks between different diet groups in mice across four time points

We identified patterns in the microbial network plots that distinguished the two diet groups (Fig. 8B). In the normal diet group, we observed denser connections among nodes within the *Bacteroidales* order, particularly on days 4 and 7, compared to the HFHS diet group. Additionally, nodes within the *Lactobacillales* order exhibited a higher number of connections in the HFHS diet group, specifically with nodes within the *Erysipelotrichales* and *Clostridiales*.

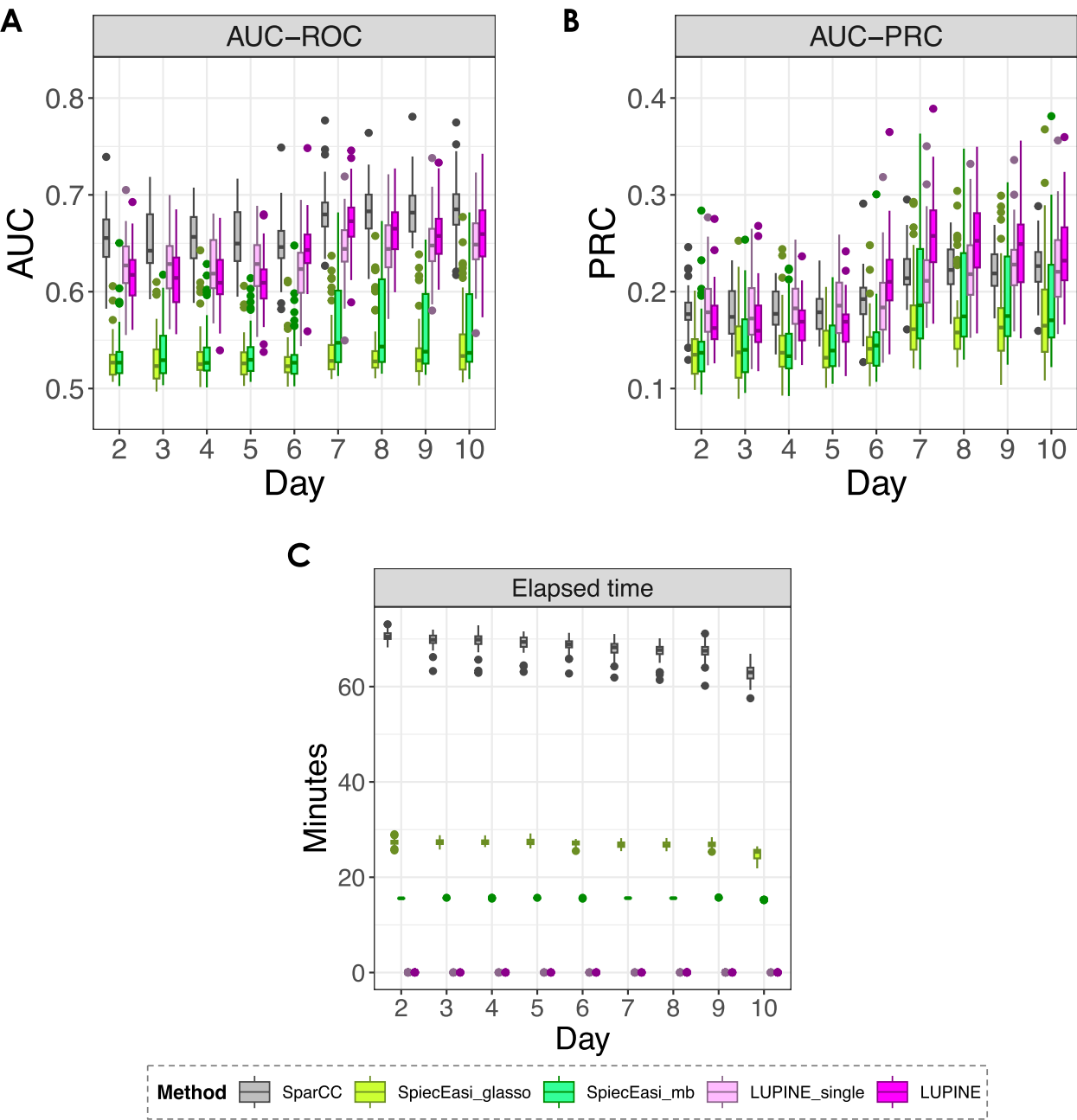


Fig. 5 Simulation results ($n = 23$). Box plots of the **A** area under the receiver operating characteristic (AUC-ROC), **B** precision-recall curve (AUC-PRC) over time for different methods. Although all methods performed similarly in terms of AUC-PRC values, both LUPINE models and SparCC outperformed the two SpiecEasi methods based on AUC-ROC values. **C** Elapsed time for each method, showing the superior computational performance of the two LUPINE methods

Similar inferences can also be made from the IVI scores in Fig. 8C1–C2: in the normal diet group, we observed decreased IVI scores across days for nodes within the *Clostridiales* order, with the majority having a zero IVI value. Compared to the normal diet group, the HFHS diet group networks had a higher node influence in *Lactobacillales* order (Fig. 8C2). While these findings do not imply

causation, they are consistent with previous studies that reported decreased relative abundances of *Bacteroidales* and enrichment of *Lactobacillales* in mice that had been fed with a high-fat diet [33, 34], suggesting a potential influence of diet on these taxonomic orders. Daniel et al. [35] also found that the proportions of *Lactobacillales* and *Erysipelotrichales* were higher in mice fed with high-fat diet.

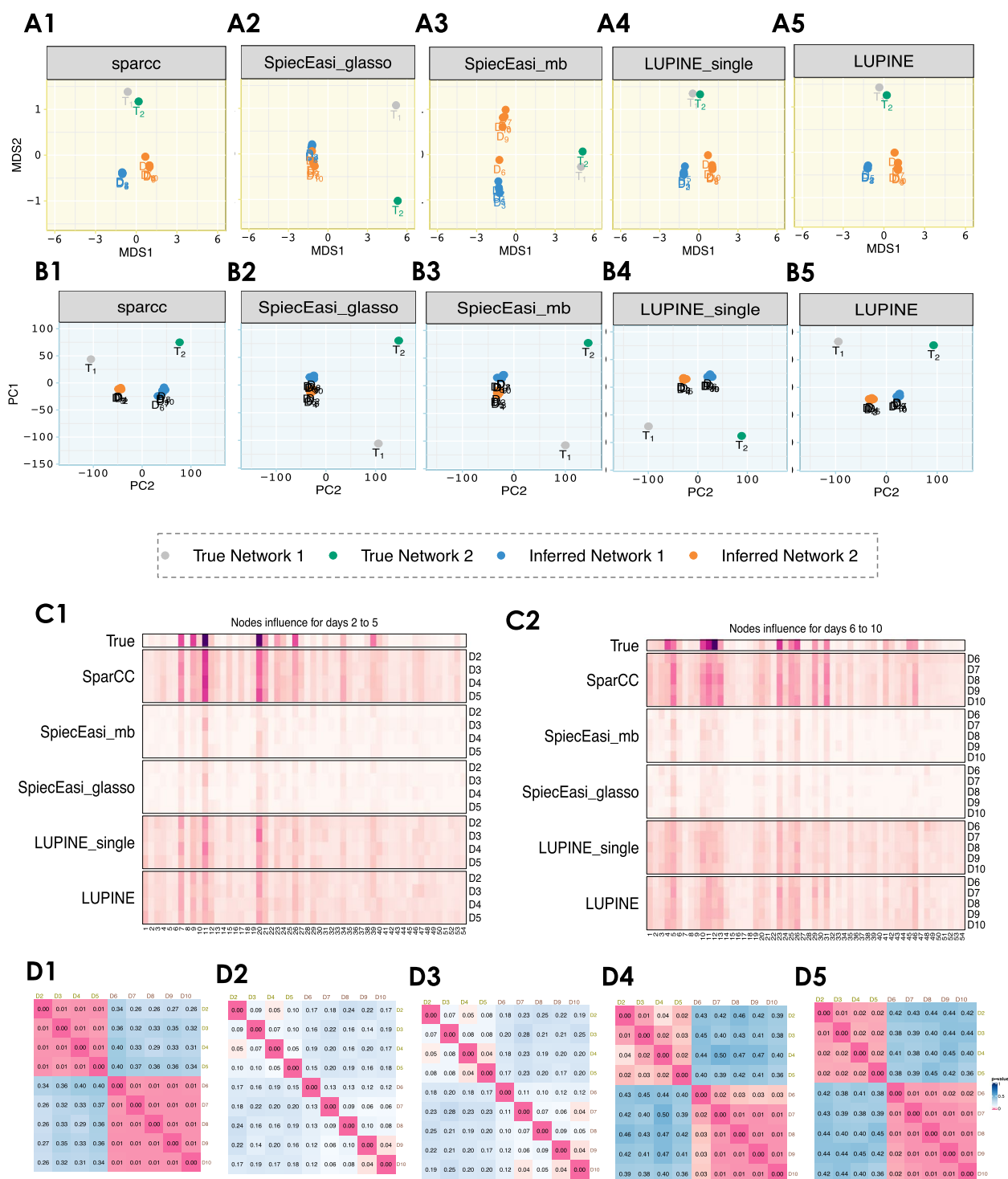


Fig. 6 Network comparisons for simulation results $n = 23$. **A1–A5** MDS plots of the pairwise GDD for the inferred networks from the different methods. All but SpiecEasi methods distinguish networks based on the two periods (days 2–5 and 6–10) in the first coordinate, regardless of the network being true or inferred. **B1–B5** PCA plots of the IVI scores for the inferred networks for each method. The largest node influence variation occurs across days, with a clear contrast between the two periods (days 2–5 and 6–10) for both LUPINE models and SparCC. Heatmaps of mean IVI scores for each node **C1** from day 2 to 5 and **C2** from day 6 to 10. The first row displays IVI scores for true networks from Fig. 4. All methods identified the most influential nodes. **D1–D5** Heatmaps of network correlation p -values indicate large differences between networks from days 2 to 5 from days 6 to 10 in both LUPINE models and SparCC

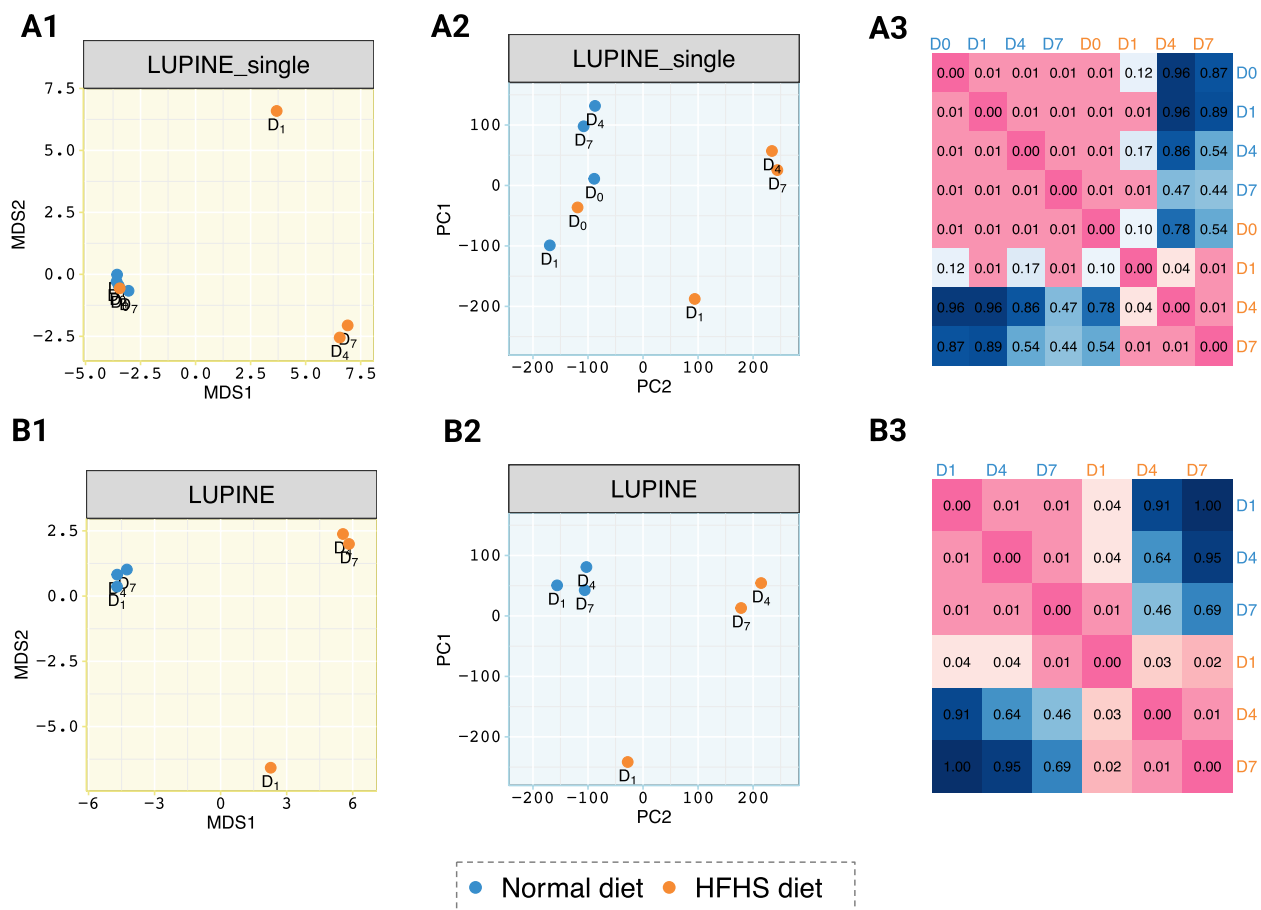


Fig. 7 Network comparison results for LUPINE_single and LUPINE. **A1–B1** MDS plot of pairwise GDD and **A2–B2** PCA plot of the IVI scores illustrating network similarities. **A** from LUPINE_single scheme and **B** from LUPINE longitudinal scheme. Compared to LUPINE_single, LUPINE exhibits consistent grouping patterns for both types of measures. **A3–B3** Heatmaps of network correlation *p*-values show that both approaches effectively identify significant differences between networks in the normal diet group and the later-stage networks in the HFHS diet group

Lactobacillus has also been extensively studied to prevent or treat type 2 diabetes mellitus [36], indicating its potential effect on high sugar diet.

To summarise, LUPINE analysis highlighted two different microbial community networks between normal and HFHS diets, with an increase in connections in the taxa nodes belonging to *Lactobacillales* in mice that were fed with a HFHS diet.

VREfm study: LUPINE highlighted changes in the network structure across two abrupt interventions in a mouse study with 11 time points

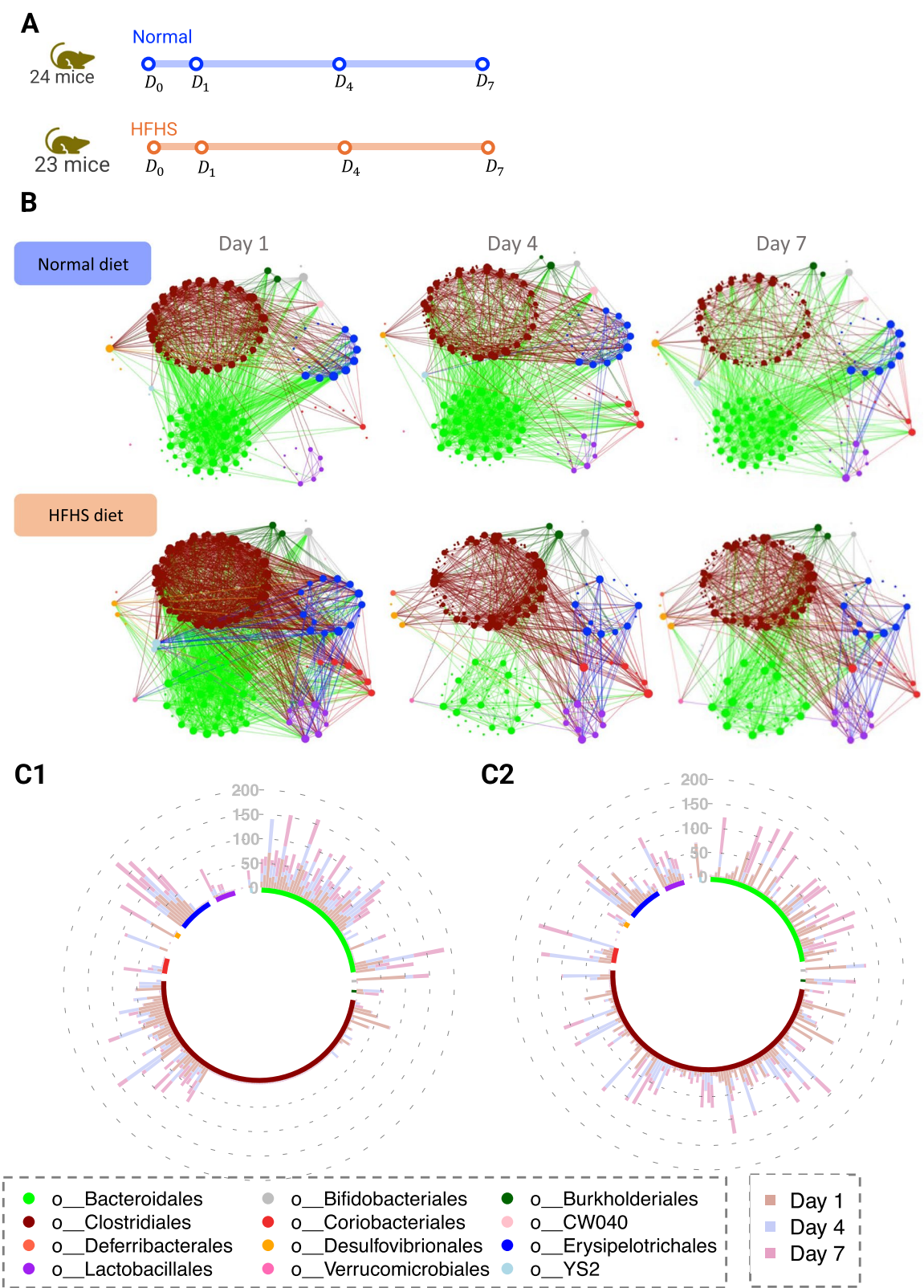
In the VREfm (Fig. 9A), mice underwent two interventions: antibiotic administration and VRE colonisation.

Similar to the first case study, this is a more controlled experiment compared to human studies. However, in this study, the time periods are divided into three phases: naive, antibiotic, and VRE, reflecting different stages of the experiment. In contrast to the first case study, where the objective was to identify the group differences the objective here is to explore network differences between the phases.

During the antibiotic phase, the number of edges in the inferred networks decreased in *Bacteroidales* and *Clostridiales* (Fig. 9B). *Bacteroidales* are an important microbe for short-chain fatty acids. Miao et al. [37] found that *Bacteroidales* almost disappear from mouse

(See figure on next page.)

Fig. 8 HFHS case study. **A** Graphical representation of the study timeline. **B** Inferred microbial networks over time for both normal group and HFHS diet groups. In the HFHS diet group, we observe a decrease in connections among nodes belonging to the *Bacteroidales* order over time. In the normal diet group, nodes belonging to the *Erysipelotrichales* order are more connected to those from the *Bacteroidales* order, but in the HFHS diet group, we observe that nodes belonging to the *Erysipelotrichales* order are more connected to those from the *Lactobacillales* order. **C** Circular stacked bar plot of the IVI scores for **C1** normal diet and **C2** HFHS diet groups show a reduction in IVI scores for nodes associated with the *Clostridiales* order in the normal diet group compared to the HFHS diet group



feces during ceftriaxone treatment. During the VRE phase, the number of edges increased, but this increase was not observed across all taxonomic orders. Specifically, we observed an increased number of edges in the nodes belonging to *Bacteroidales*. In contrast, and even at the end of the experiment, the nodes belonging to *Clostridiales* had very few connections compared those detecting in the naive phase. From the antibiotic phase to day 9, two nodes belonging to taxonomic order *Streptophyta* and *Rickettsiales* exhibited higher connections. This finding is in agreement with [38], who studied the disruption of the gut microbiota after the administration of broad-spectrum antibiotics and found a strong relative increase in bacteria of the genus *Streptococcus* across all subjects. Therefore, the increased connections of *Streptophyta* and *Rickettsiales* could potentially be attributed to the impact of antibiotics on the gut microbiota. From day 12, we also observed an increased number of connections in *Verrucomicrobiales*. The genome of *Akkermansia muciniphila*, a species within this taxonomic order, has been examined for its potential antibiotic resistance-associated genes [39].

The MDS plot of the pairwise GDD revealed a clear separation between the naive phase, antibiotic phase, and VRE phase networks (Fig. 9C). Within the VRE phase, we observed a temporal progression, indicating a unique network structure on day 9, which marked the initiation of VREfm colonisation. The PCA plot of the IVI scores (Fig. 9D) showed a similar grouping structure to the MDS plot. The similarity between the two plots indicated a consensus in the network grouping, regardless of the two metrics employed (as we showed previously in the HFHS study). The naive phase networks IVI scores were well separated from the other phases, accounting for 35% of total variation of the scores. After the naive phase, we observed a decrease in the influential nodes for the majority of taxa. However, taxa belonging to the *Enterobacterales* and *Lactobacillales* orders became more influential following the antibiotic phase (Fig. 9E). In [9], we also found that these taxa belonging to families *Enterobacteriaceae* and *Enterococcaceae* showed a group difference between the naive and the VRE phases. We also noted a decrease in the IVI scores for *Clostridiales* nodes

following the naive phase, a decrease that persisted until the end of the experiment. This taxonomic order was a key taxa group that has been studied for its ability to restrict gut colonisation by *Enterobacteriaceae* pathogens in mice, as shown by [40–42]. When comparing networks, we observed strong correlations within the naive phase, between the antibiotic phase and day 9, and during the later days of the VRE phase (Fig. 9F). Additionally, there was a weak correlation between the networks across phases.

Due to the weak correlation between the VRE phase and the other two phases, we evaluated how the amount of past information can affect the day 14 network structure (using data from day 9 onward vs. data from all past time points). The network based solely from the VRE phase data had only two fewer edges (Fig. 10B) than the network built based on all time points (Fig. 10A). The two networks differed only by two edges: one between *Bacteroidales* species and another between *Bacteroidales* and *Clostridiales* (Fig. 10C). This finding suggests that using all prior data or only VRE phase data has minimal impact on the network structure. However, users can adjust how many past time points they may wish to include in the model.

To summarise, LUPINE highlighted a reduction in the number of edges after antibiotic treatment, affecting nodes belonging to the *Clostridiales* order. Interestingly, the reduction in *Clostridiales* node connections persisted even after the antibiotic phase, suggesting that the VRE colonisation had an impact on the recovery of these connections. However, further studies are needed to assess whether these connections would be recovered without VRE intervention.

Diet study: LUPINE detected diet-specific and diet-stable taxonomic groups in a case-control human study spanning across 15 time points

The diet case-control study differs from the previous ones as it involves humans and numerous time points (Fig. 11A) and a small number of participants. This results in expected higher variability than mouse studies.

The plant-based diet group differed from the animal-based diet group, with the former showing a tighter

(See figure on next page.)

Fig. 9 VREfm case study. **A** Graphical representation of the study timeline. **B** Inferred networks across time. After antibiotic treatment on day 6, we observe a reduction in the associations in the nodes belonging to the *Clostridiales* order, which are not recovered until day 14. However, the connections among nodes of the *Bacteroidales* order appear to recover from day 12. **C** MDS plot of the pairwise GDD shows that the network structures are significantly different for each phase of the experiment. **D** PCA plot of the IVI scores exhibits a grouping pattern consistent with the MDS plot. **E** Line plots of the IVI scores for each taxa across time, grouped by their taxonomic order show a reduction of influential nodes after antibiotics treatment, indicating a less dense network structure. **F** The heatmap of Mantel *p*-values comparing network correlations indicates significant correlations within the naive phase, between the antibiotic phase and day 9, and across the last three days of the VRE phase

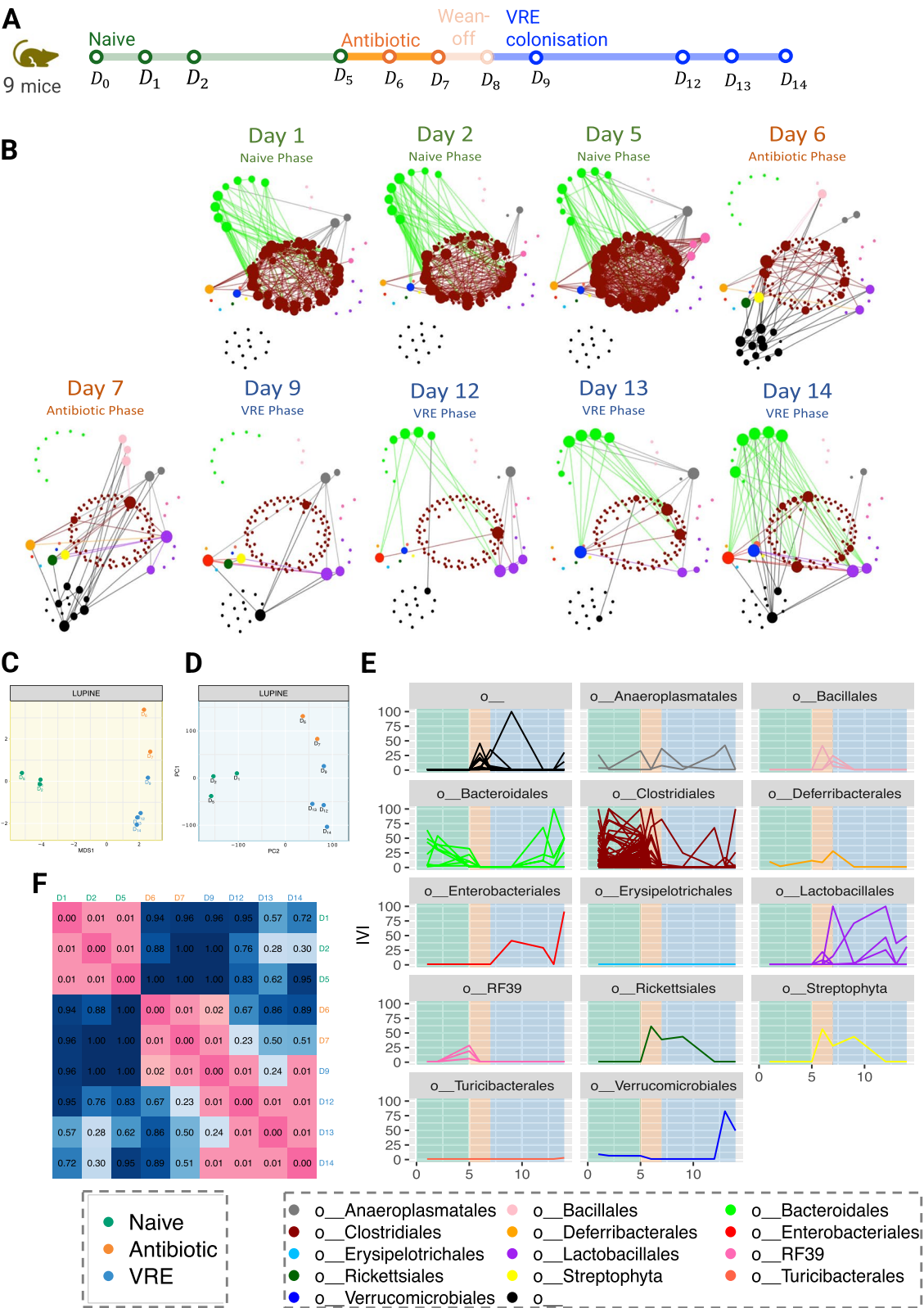


Fig. 9 (See legend on previous page.)

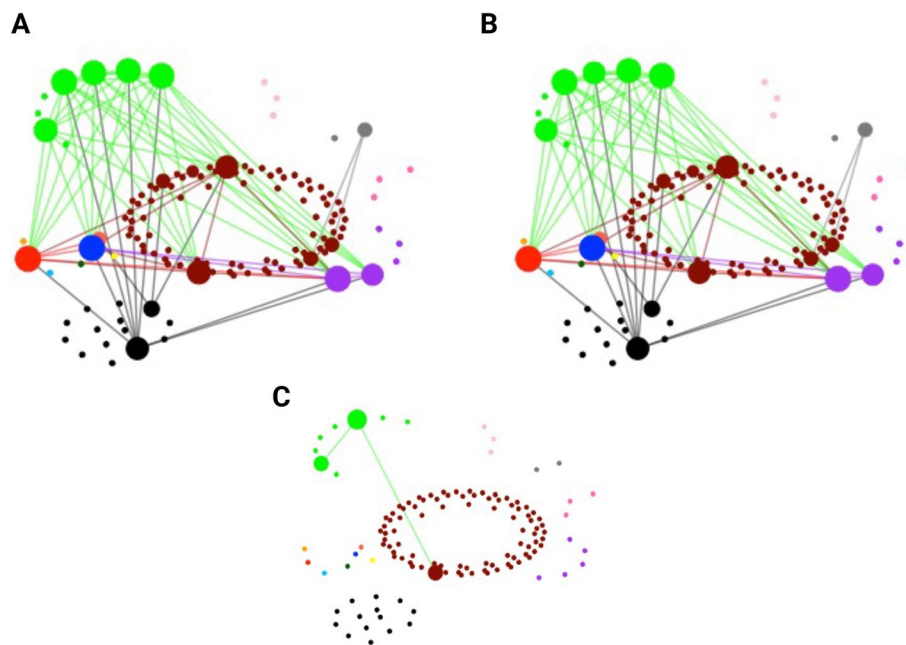


Fig. 10 Impact of past data on the day 14 microbial network in the VREfm case study. **A** Inferred microbial network at day 14 using all past time points. **B** Inferred microbial network at day 14 using only the VRE phase time points. **C** Difference between the two networks, highlighting only two edges present in A but absent in B. The two networks are largely similar

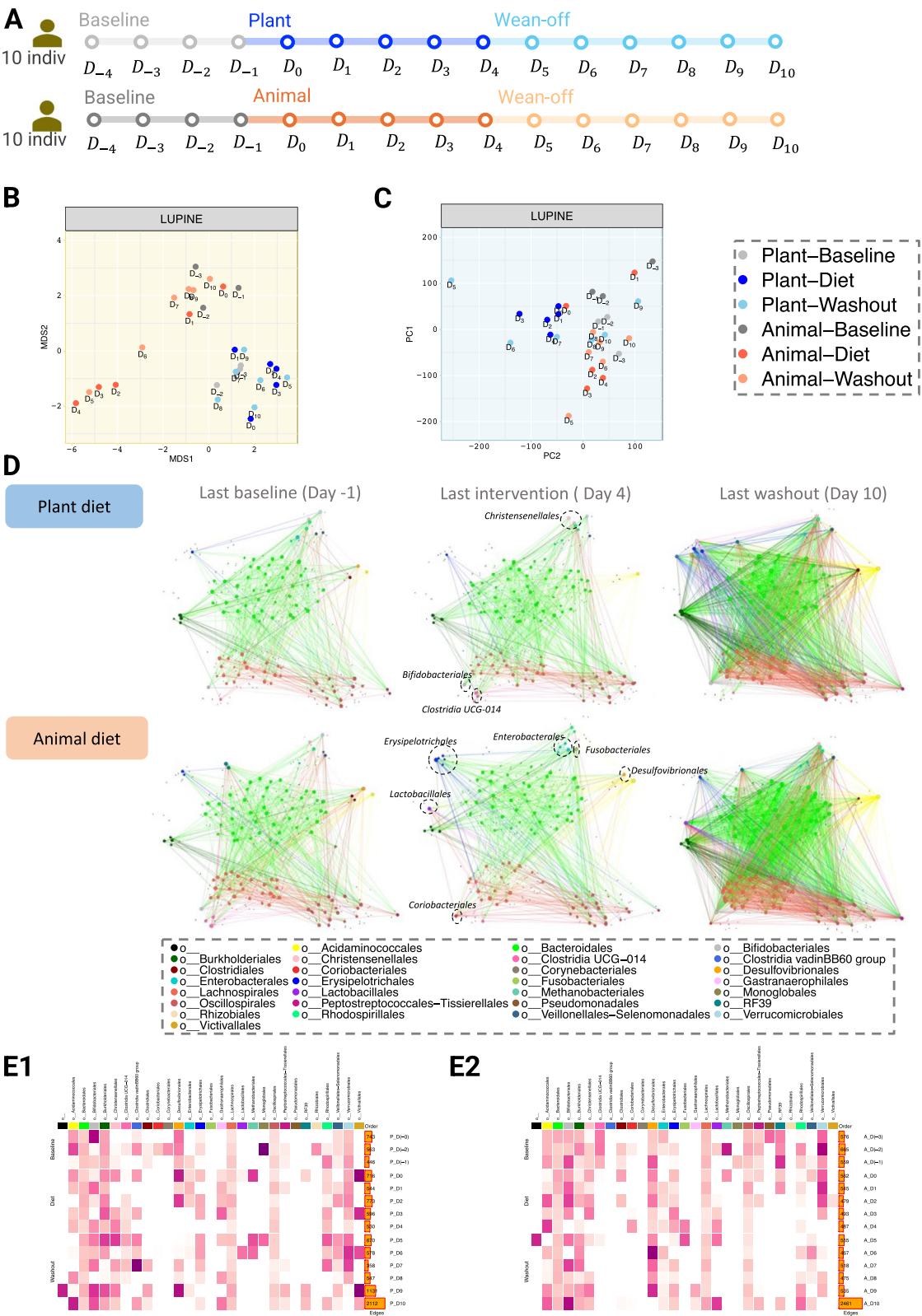
network grouping, indicating more significant changes in the animal diet intervention (Fig. 11B). In the animal diet group, two clusters were observed. The first included networks from days 2 to 4, and the second from days – 3 to 1 and 7 to 10, with the network on day 6 distinct from both clusters. This suggests that the network structure returned to its pre-intervention form two days into the washout period. A high variability was also noted across the animal diet group networks in the PCA plot (Fig. 11C). However, the PCA plot revealed a less distinct separation between the two network groupings within the animal diet group compared to the MDS plot. This indicates a larger difference in the inferred networks, due to information flow or structure, compared to the IVI score variation. Furthermore, we noticed a close alignment in the baseline inferred networks across both

groups. This implies a similar degree of node influence during the baseline phase for both groups.

We examined in more detail the six inferred networks from the two diet groups at three different days, each reflecting the final day of each phase: baseline, intervention, and washout (Fig. 11D). We highlighted few taxa nodes in the networks at day 4 that differed between the two diet groups. On day 4, we observed the nodes belonging to *Clostridia UCG 014* had a comparatively higher degree in the plant based diet group than in the animal based diet group. *Clostridia UCG 014* is a bacterial taxon belonging to the class *Clostridia* in the phylum *Firmicutes*. While [43] found that *Clostridia UCG 014* was reduced in the high-fat diet, [44] associated the decrease in *Clostridia UCG 014* with obesity. Additionally, [45] found a positive correlation between *Clostridia UCG 014* and blood glucose levels, and [46] inferred *Clostridia UCG 014* as one of the

(See figure on next page.)

Fig. 11 Diet case study. **A** Graphical representation of the study timeline. **B** MDS plot of pairwise GDD shows that distinct patterns emerge based on diet groups (tighter cluster for plant-based networks compared to animal-based networks). **C** PCA plot of the IVI scores show that inferred networks at baseline are similar regardless of diet, indicating similar node wise importance. **D** Inferred networks across two diet groups and three time points (last day of baseline, intervention and washout phase). On day 4, the plant-based network shows increased connections in nodes associated with *Christensenellales*, *Christensenellales*, and *Clostridia UCG 014* compared to the animal-based network, which exhibits increased connections in nodes related to *Erysipelotrichales*, *Lactobacillales*, *Coriobacteriales*, *Enterobacteriales*, *Fusobacteriales*, and *Desulfovibrionales*. **E** Heatmap of the average IVI score for each taxonomic order for **E1** plant based and **E2** animal based diet groups, with darker pink representing high average IVI value. Nodes from *Bacteroidetes*, *Lachnospirales*, and *Oscilospirales* consistently exhibit a non-zero IVI score, indicating their stable influence, unaffected by diet or daily variations



taxa potentially linked to cholesterol regulation. [47] found *Clostridia UCG 014* to be positively correlated with propionate in mice fecal samples. Propionate is a short-chain fatty acid that can be produced by gut bacteria through fermentation of dietary fiber. Additionally, *Clostridia UCG 014* was also found to co-exist with other fiber responsive bacteria such as *Ruminococcus* [48]. Nodes belonging to *Lactobacillales* had comparatively higher degree in the animal based diet group compared to the plant based diet group. The original study also found that the abundances of *Lactococcus lactis* and *Pediococcus acidilactici*, which belong to the taxonomy order *Lactobacillales*, significantly increased in the animal-based diet. This is in agreement with the review from [49] reporting an increased level of *Lactobacillus*, a member of the *Lactobacillales* order, resulting from the consumption of white meat protein (such as chicken and fish) and dairy products (like milk, cheese, yogurt, and kefir). By comparing these taxa with the average IVI scores across the intervention period, we observed some agreement between a high average IVI score (Fig. 11E1–E2) and a high number of connections in the networks in Fig. 11D. For instance, *Clostridia UCG 014* had a non zero IVI average in the plant based diet group from day 2, whereas the IVI score in the animal based diet group was zero specifically during the intervention period (day 0 to day 4). We also observed that the nodes belonging to taxa order *Lactobacillales* had a non zero IVI average in the animal based diet group from day 2, whereas the IVI score in the plant based diet group was zero specifically during days 1 to 5. In the heatmaps, the bar plots revealed a high number of edges on the last day (i.e. day 10) compared to all other days in both groups. No clear reasoning can be given for this observation as there was no change at this particular time point. However, this network did not stand out in our network comparison, indicating that even though the number of connections increased, the network structure and the influential nodes on day 10 were similar to those on day 9. A densely connected network at the last time point was also visible in the network plots in Fig. 11D. However, unlike the previous two case studies, we did not find diet specific correlations occurring in this particular case study (Fig. 18 in Appendix C).

To summarise, LUPINE applied to a case-control human study with intervention identified several taxonomic associations that were specific to plant based diet group or animal based diet group. We also identified few taxonomic orders such as *Bacteroidetes*, *Lachnospirales*, and *Oscillospirales* that were influential regardless of time or diet intervention. This suggests that these taxa may exhibit stability in the face of dietary changes. In fact, *Oscillospira* is currently being explored as a next-generation probiotic due to its potential health benefits [50].

Pre-diabetic study: LUPINE identified key taxonomic groups studied as potential therapeutic targets for diabetic patients in a large human metagenomic study conducted at two time points.

The pre-diabetic case study includes metagenomics data from a larger number of individuals in each group, across two time points (Fig. 12A), with pre-diabetic individuals following either a MED diet or a PPT diet.

Compared to other case studies, the networks inferred through LUPINE displayed a similar network structure in the pre-diabetic case study (Fig. 12B). This was further confirmed by a low *p*-value in the Mantel test, indicating a significant correlation between the network adjacency matrices. However, we identified differences in the two networks based on IVI scores (Fig. 12C). Specifically, several genera varied between the two diet groups. For example, nodes belonging to the genus *Faecalibacterium* had higher IVI scores in the PPT group compared to the MED group. Previous research by [51] showed that butyrate-producing bacteria, such as *Faecalibacterium prausnitzii*, were reduced in individuals with pre-diabetes compared to those with normal glucose tolerance. Several other studies have also reported a decline in butyrate-producing bacteria in pre-diabetic or diabetic groups [52, 53]. Butyrate has been shown to downregulate inflammation and increases mucosal barrier integrity [54]. Palacios et al. [55] suggested that increased butyrate levels could enhance glucose management, while [56] demonstrated that *Faecalibacterium prausnitzii* lowered fasting blood glucose, improved glucose tolerance, and reduced HbA_{1c} levels in pre-diabetic and diabetic mice. Thus, a higher number of connections for nodes belonging to *Faecalibacterium* may indicate positive effects of the PPT diet on pre-diabetic patients.

Similarly, nodes belonging to *Bifidobacterium* also had higher IVI scores in the PPT group than in the MED group. Chang et al. [57] found lower levels of *Bifidobacterium* in pre-diabetic patients compared to healthy individuals. *Bifidobacterium* was observed as a bacterium that is indirectly capable of promoting GLP-1 (glucagon-like peptide-1) production [58]. GLP-1 receptor agonists are used as a class of medication to treat type 2 diabetes [59]. Therefore, higher IVI scores for *Bifidobacterium* nodes may further suggest beneficial effects of the PPT diet on pre-diabetic patients. In contrast, nodes from *Dorea* had lower IVI scores in the PPT group than in the MED group. Several studies have reported higher abundance of *Dorea* in pre-diabetic individuals [60–62]. Additionally, [62] identified *Dorea* as one of the top 10 bacteria differentiating healthy individuals from diabetic patients (type 2 or pre-diabetic). Positive correlations were also found between *Dorea* abundance and fasting plasma glucose, C-peptide, BMI, and waist circumference

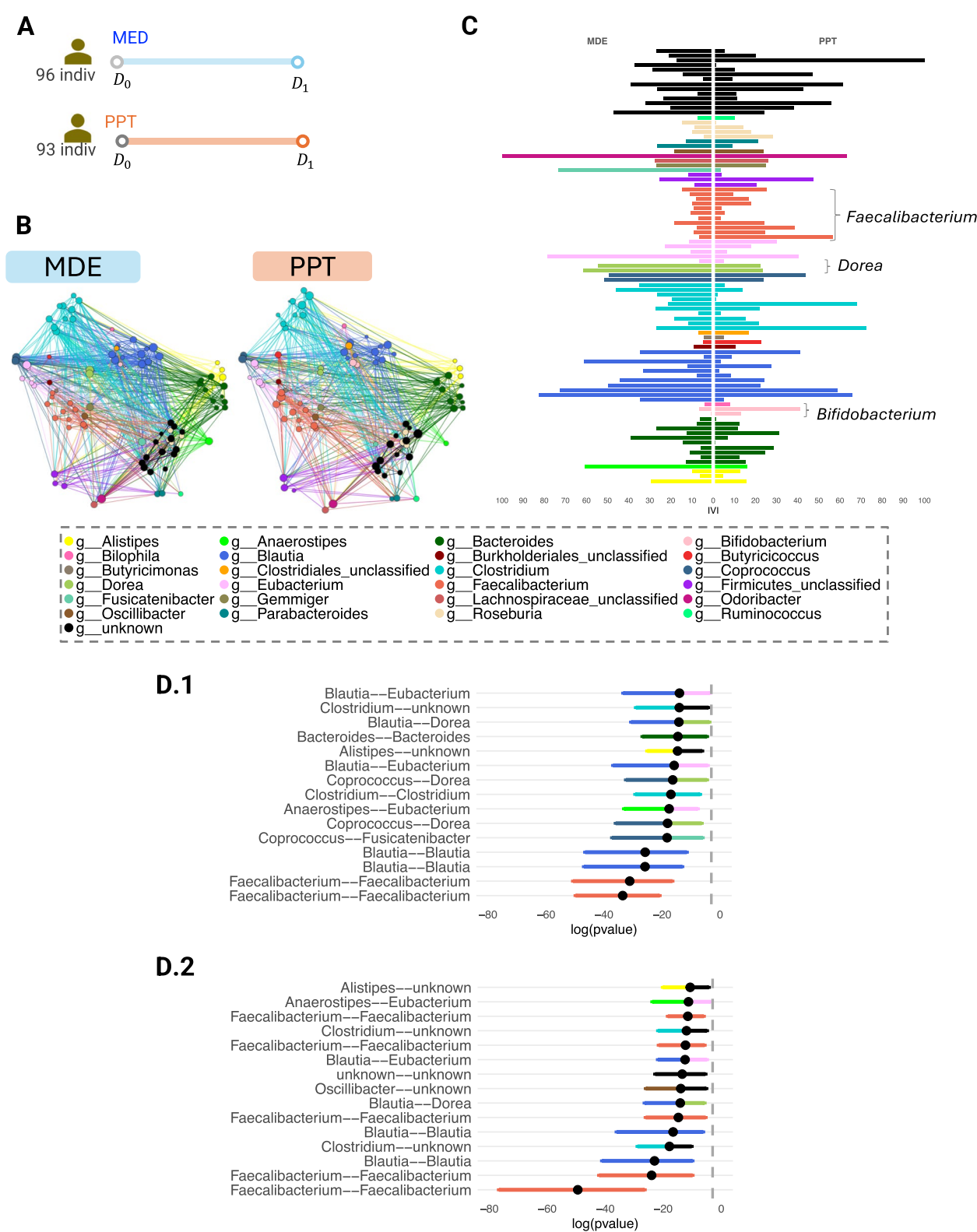


Fig. 12 Pre-diabetic case study. **A** Graphical representation of the study timeline. **B** Inferred networks post-intervention for MDE and PPT diets show similar network structures across both diets. **C** Bar plot of IVI scores for each SGB highlights subtle differences in some genus groups. Notably, nodes belonging to *Bifidobacterium* and *Faecalibacterium* exhibit higher IVI scores in the PPT diet compared to the MDE diet. Confidence interval plots for log transformed *p*-values from correlation tests: **D1** MDE diet **D2** PPT diet, show a higher number of connections between *Faecalibacterium* pairs in the PPT diet compared to the MDE diet. The dashed grey vertical line represents a *p*-value of 0.05

[60]. Hence, lower IVI scores for *Dorea* nodes may indicate a healthier profile.

We also observed differences between the two diet groups based on these bootstrap confidence intervals for the p -values from correlation tests between pairs of taxa (Fig. 12D1-D2). In the PPT diet group, most of the 15 lowest p -values indicating significant associations were between nodes belonging to *Faecalibacterium*. *Faecalibacterium prausnitzii*, a species from this genus, has been identified as a potential target for diabetes prevention or treatment strategies, such as prebiotics or probiotics, due to its role in producing butyrate, which enhances insulin homeostasis [63]. Therefore, increased connections involving *Faecalibacterium* could reflect the positive effects of the PPT diet. In the MDE diet group, nodes belonging to *Dorea* were associated with *Coprococcus* nodes twice and with a *Blautia* node once. In the PPT diet group, *Dorea* was only associated with a *Blautia* node in the top 15 associations. Takeuchi et al. [64] showed that organisms from the genera *Blautia*, *Dorea*, and *Coprococcus*, as listed in the Kyoto Encyclopedia of Genes and Genomes, exhibited the three highest positive correlations with fecal carbohydrates.

To summarise, LUPINE applied to a large case-control metagenomic human study indicated that both MDE and PPT resulted in similar networks. However, downstream analysis highlighted several taxonomic genera that were more prominent in the PPT diet, including *Faecalibacterium*. *Faecalibacterium prausnitzii*, a species within this genus, has been suggested as a therapeutic option for type 2 diabetes due to its potential to enhance insulin sensitivity, improve lipid metabolism, and reduce inflammation [65].

Discussion

We developed LUPINE to detect the stability of taxa associations within a microbial community over time and the responses of these microbial communities to external disturbances such as dietary changes and medication. To the best of our knowledge, LUPINE is the first sequential microbial network inference approach for a longitudinal setting.

In LUPINE, we combined the concepts of low-dimensional approximation with partial correlation to infer networks across time points. We then used GDD and IVI metrics to identify any abrupt network changes across time, groups, and key taxa nodes in each of the networks. Additionally, we tested the significant correlations between network adjacency matrices using the Mantel test. Note that a requirement for LUPINE is that samples match across time points, as LUPINE assumes correlation structure across time and individuals. In the case where different individuals are sampled across time, a better approach for

network inference is our variant LUPINE_single applied to each time point.

In our simulation study, we demonstrated that LUPINE and LUPINE_single improved model performance and were computationally more efficient compared to existing network modelling approaches SparCC and SpiecEasi that were designed for single time point analyses. We also demonstrated the applicability of LUPINE in case studies with different study designs. The two controlled mouse studies were either case-control or intervention studies, while the third and fourth studies were more complex human study incorporating elements of both a case-control and an intervention design in either long or short time courses. In all case studies, our LUPINE analyses extracted meaningful biological insights, including clustering patterns in the inferred networks that were coherent within each study design.

While developing LUPINE, we had to consider several aspects related to the characteristics of microbiome data. First, taxa filtering is an important data processing step in this kind of analysis. We only retained taxa with a mean relative abundance exceeding 0.1% for any group at any time point to obtain a consistent set of taxa being examined across various time points and groups. Further, to avoid identifying connections between taxa with low abundance at certain time points in specific groups, we then only examined taxa with a mean relative abundance greater than 0.1% within each group and time point in the inferred networks. Second, most experimental microbiome studies are typically low sample size, which are likely to result in false positive associations. We used a correlation test that is appropriate for small sample size when calculating test statistics. We chose to identify significant associations based on p -values with an arbitrary cutoff of 0.05, rather than a resampling based model selection approach (used in SpiecEasi, [5]) or a bootstrap-based p -value calculation approach (used in SparCC, [8]) that are computationally expensive. Adjustments for false discovery on p -values could also be considered to obtain sparser networks. However, our experience has shown that these adjustments may lead to empty networks in some situations.

We have identified several potential extensions of LUPINE that could enrich longitudinal analysis of microbiome data. Firstly, when considering studies with several groups (e.g. HFHS and Diet studies), LUPINE must be applied to each group separately if we assume that the true networks within these groups differ from one another. A potential extension of LUPINE could include all groups together to infer a common network across all groups. Secondly, LUPINE main focus is to identify associations between taxa at specific time points, while taking into account information from previous time points. Therefore, we do not model time gaps

between points. Future developments could include approaches that weight closer time points more heavily than distant ones. While we have only analysed one metagenomic study, which focused on taxonomy profiles based on species abundance, a future direction could be to extend this approach to include functional profiles derived from gene abundance information. An interesting extension of LUPINE to fully harness these complex data would be to establish connections between these two layers, akin to multilayer networks at each time point. This advancement could significantly broaden our understanding of microbiome dynamics, opening up new possibilities for research and discovery.

Appendix A: Assessing the efficacy of the approximation used to improve the computation time

In this section, we evaluate the adequacy of the approximation used to improve computational time using the normal diet data on day 0 from the HFHS case study (X_0). As a reminder, our method involves the computation of one-dimensional approximations for $p - 2$ taxa. In the simplest case of a single time point, we achieved this by systematically excluding pairs of taxa from the initial dataset X_0 , performing a principal component analysis (PCA) and extracting the first principal component. Given that there are $p \times (p - 1)/2$ possible ways of selecting the excluding taxa pair, we performed PCA $p \times (p - 1)/2$ on $X_0^{-(i,j)}$. $X_0^{-(i,j)}$ denote the data set X_0 after excluding the i th and j th taxa.

To improve computational efficiency and avoid repetitive PCA computation, we employed an approximation strategy for calculating the first principal component of $X_0^{-(i,j)}$. We first compute the loading vector for the entire dataset, X_0 , then, approximate the first principal components for $p - 2$ taxa by selectively nullifying the loading weights corresponding to the excluded taxa pair, then multiplying the resulting loading vector by X_0 .

To compare the original principal component with its approximation we used the concordance correlation coefficient. The concordance correlation coefficient evaluates the similarity between paired data by quantifying the deviation from the concordance line (45° line through the origin) [66]. A concordance correlation value close to 1 signifies that when plotting the original principal component against its approximated counterpart, the values closely adhere to the concordance line, thus indicating a high level of agreement between the original and the approximated principal component. In Fig. 13, we observe that all the concordance correlation values are above 0.99 between the approximated and the original principal component indicating a

high level of agreement between the two. In the remainder of the main article, we therefore chose to use approximated principal component to calculate the partial correlations.

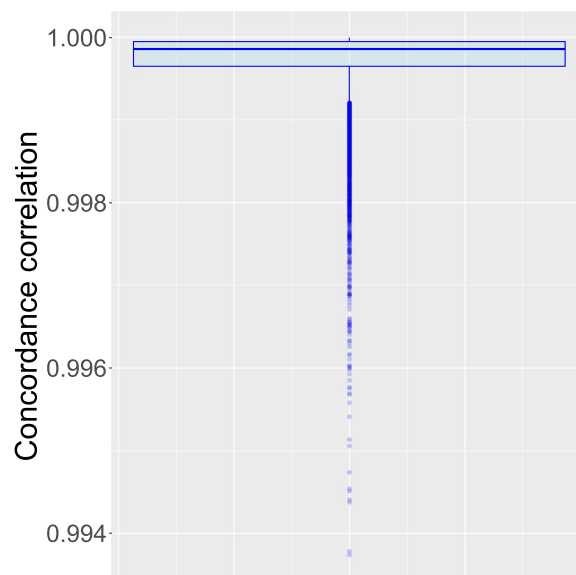


Fig. 13 Concordance correlation values for the normal diet data on day 0 from the HFHS case study. The concordance correlation value is used to assess the agreement between the original and the approximated principal component for all excluded taxa pairs. All the correlation values are above 0.99 indicating a high level of agreement between the two approaches

Appendix B: Simulation design

In this section, we describe the details of the simulation study for comparing network inference methods.

We generate data using a multivariate Poisson distribution and incorporate concepts inspired by birth and death processes [67]. At the initial time point, we use a copula-based multivariate Poisson distribution to generate 54 taxa counts for each individual. For each subsequent time point, we simulate the occurrence of new births using a multivariate Poisson process. We then update the taxa counts by adding these new births to the previous counts, which are reduced to simulate the concept of 'death'. To simulate 'death', we incorporate the First-order Integer-valued Autoregressive (INAR(1)) mechanism. This mechanism employs a binomial thinning operator to model non-negative integer-valued time series [68]. This operator represents the decrease in taxa population between two time points. To understand the idea of a binomial thinning operator, consider a population of size Y at a specific time t . If we observe the same population at time $t + 1$, the population may decrease due to the death of some entities between times t and $t + 1$. Assuming these deaths are independent events and that the probability of dying between t and $t + 1$ is uniform, denoted by

$1 - \alpha$, then the resulting count of survivors at $t + 1$ can be expressed as $\alpha \circ Y$ [69]. Thus, at a given time point t , taxa count matrix, C_t , is modelled as

$$C_t = \alpha \circ C_{t-1} + N_t \quad (14)$$

where C_{t-1} represents the taxa count matrix at time $t - 1$; α is the survival rate; and N_t is the new birth matrix at time t . The matrices C_t , C_{t-1} , and N_t correspond to n individuals and 54 taxa.

To preserve the correlation structure, we use a correlated binomial distribution for thinning, with $\alpha = 0.5$ until time point 6 and $\alpha = 0.1$ thereafter. We assume the expected value of each taxon i for individual j (λ_{ij}) in the multivariate Poisson distribution remains constant over the two time periods: days 1–5 and days 6–10. We generate these λ_{ij} values using two different Gamma distributions. For the first period (days 1–5), we generate λ_{ij} using a Gamma distribution with its parameters estimated from the mean taxon values in the normal diet group. For the second period (days 6–10), we generate another λ_{ij} using a different Gamma distribution with its parameters estimated from the mean taxon values in the HFHS diet group. To assess the goodness of fit between the observed mean taxon values and the fitted Gamma distribution, we generate quantile-quantile (Q-Q) plots (Figs. 14 and 15). These Q-Q plots confirm the goodness of fit for the majority of taxa, as evidenced by the close alignment between the points and the theoretically expected values of the distribution.

Finally, to mimic the library size effect in our simulation, we rarefy the samples to be between 5000 and 8000.



Fig. 14 Goodness fit of the mean taxon levels estimated from gamma distributions in the normal diet group. For each taxon, qq plots are generated from theoretical quantiles and sample quantiles. Overall the fit of gamma distribution is satisfactory except for a few that deviate more from the theoretical quantiles (i.e. *Taxa*₅)



Fig. 15 Goodness fit of the mean taxon levels estimated from gamma distributions in the HFHS diet group. For each taxon, qq plots are generated from theoretical quantiles and sample quantiles. Overall the fit of gamma distribution is satisfactory except for a few that deviate more from the theoretical quantiles (i.e. *Taxa*₁₁₇)

Appendix C: Supplementary figures

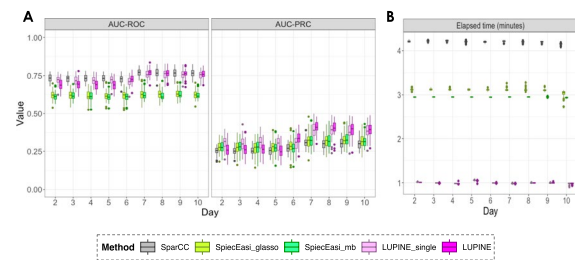


Fig. 16 **A** Box plots of the area under the receiver operating characteristic (AUC-ROC), precision-recall curve (AUC-PRC) and **B** elapsed time values across different network inference methods for sample size 50. Each boxplot represents a distinct method, distinguished by different colours. Similar to sample size 23, two LUPINE methods and SparCC outperformed the two SpiecEasi methods based on AUC-ROC values. However, two LUPINE methods outperformed SparCC, specifically in later days. Similar to sample size 23, elapsed time was superior in LUPINE methods

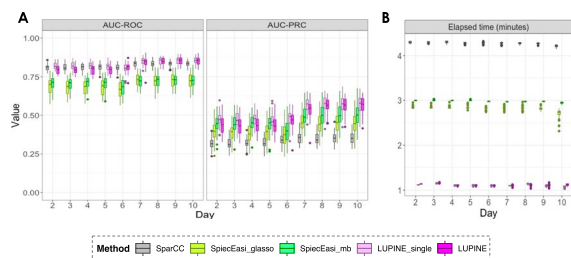


Fig. 17 **A** Box plots of the area under the receiver operating characteristic (AUC-ROC), precision-recall curve (AUC-PRC) and **B** elapsed time values across different network inference methods for sample size 120. Each boxplot represents a distinct method, distinguished by different colours. Similar to sample size 23 and 50, two LUPINE methods and SparCC outperformed the two SpiecEasi methods based on AUC-ROC values. However, two LUPINE methods outperformed all other methods based on AUC-PRC values. Elapsed time was again superior in LUPINE methods

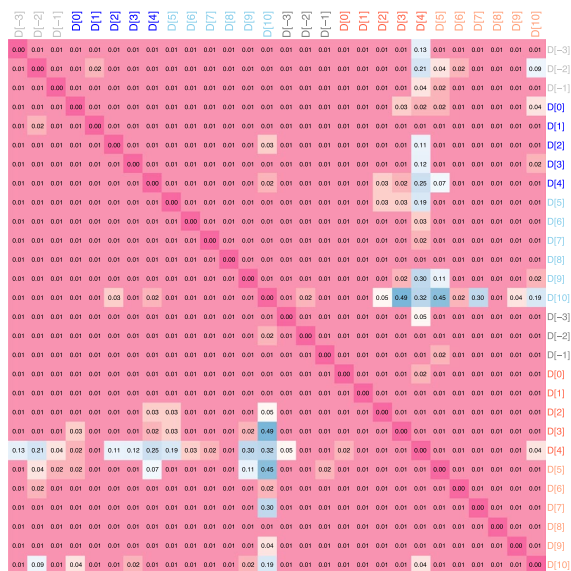


Fig. 18 Heatmap of Mantel p -values for diet case study. In both animal and plant diets, there is a significant correlation across networks, indicating the presence of a core network shared between the two diets

Appendix D: Sensitivity analysis

In this section, we performed a sensitivity analysis to assess the robustness of LUPINE and LUPINE_single across various conditions using simulated data. The key factors evaluated include the choice of statistical

test (correlation test vs. permutation-based test), normalisation techniques, number of components used in deflation, dimensionality reduction methods, edge proportions, and sample size. Our results showed that:

- Correlation-based inference were more efficient than permutation-based methods, specifically for large sample size.
- Normalisation methods showed no significant impact on model performance.
- Using one component for deflation yielded comparable or superior performance to using multiple components, especially for smaller sample sizes.
- Alternative dimension reduction methods to PCA, such as robust PCA (RPCA) and independent component analysis (ICA) produced similar outcomes, with PCA showing slight advantages for larger sample sizes.
- Model performance decreased as edge proportions increased, though AUC-PRC results suggested that significant edges still captured true correlations.
- As expected, larger sample sizes enhanced model performance due to improved signal detection.

These findings reinforce the stability of LUPINE and LUPINE_single under various configurations.

Permutation test vs. correlation test

In LUPINE, we proposed using a correlation test to infer associations between taxa. On simulated data, we compared the association inferences from a permutation-based test as follows. We first calculating the partial correlation between taxa i and j , as shown in Eq. 5 (referred to as the original correlation). Next, we permuted the residuals from taxon j (e_j) to eliminate any association between i and j and then calculated the correlation between e_i and the permuted e_j . This procedure was repeated 100 times, after which we calculated the proportion of instances where the absolute value of the original correlation was less than the absolute value of the permuted correlation. A lower proportion indicated a significant association. We compared this permutation method with our correlation test results on simulated data with $n = 23$ and $n = 120$ in the ‘Simulation study’ section. As seen in Fig. 19, correlation-based inference identified the true correlations more accurately than the permutation-based inference approach, particularly when considering the PRC values. The computational time for the permutation-based method increased exponentially with the number of nodes (i.e. taxa) as we need to perform permutations for each pair of taxa.

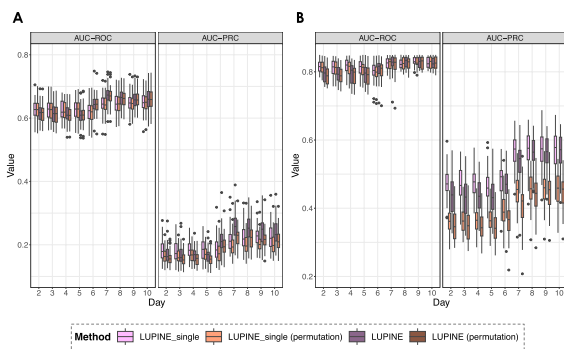


Fig. 19 Simulated data. Box plots of the area under the receiver operating characteristic (AUC-ROC), precision-recall curve (AUC-PRC) for **A** sample size = 23 **B** sample size = 120. Each boxplot represents LUPINE_single and LUPINE evaluated with a correlation test or a permutation test, distinguished by different colours. When sample size is 23, both evaluation methods perform similarly. However, correlation based association inference outperform the permutation based inference when sample size is large

Effect of normalisation

In our original LUPINE approach, we applied PCA to the count matrix, with taxa counts centred and scaled. For the linear regressions involving taxa i and j , we used the PCs as explanatory variables, and log-transformed the taxa counts from i and j to convert them to a continuous scale before using them as a response variable.

In this section, using simulated data, we evaluated the sensitivity of the results when PCA was performed on the log or clr transformed counts. Thus, in this subsection, we compared the sensitivity of LUPINE_single and LUPINE on simulated data with $n = 23$ and $n = 120$ (as described in the ‘Simulation study’ section) using the following three approaches:

1. (Default) In the linear regressions involving taxa i and j , the explanatory variables are the log library size and the first principal component from the centred and scaled counts (excluding taxa i and j), while the response variables are the log-transformed counts of taxa i and j .
2. As in 1, but count data are log-transformed in PCA.
3. Explanatory variable is the first principal component from the centred and scaled clr transformed counts (excluding taxa i and j) and the response variables are the clr transformed counts of taxa i and j .

In the simulation, we did not observe significant differences in the three approaches considered (Fig. 20).

Thus, in situations where count data are available, we recommend using our default version. However, users of LUPINE can choose between count or log/clr transformed data.

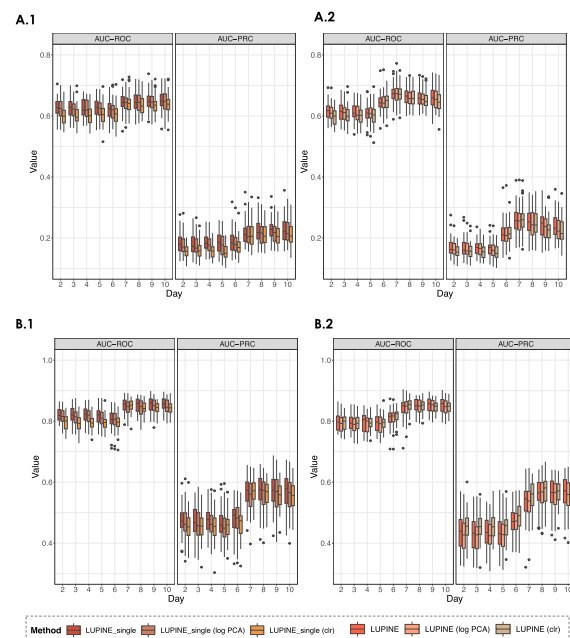


Fig. 20 Simulated data. Box plots of the area under the receiver operating characteristic (AUC-ROC), precision-recall curve (AUC-PRC) for **A.1** sample size = 23 for different versions of LUPINE_single; **A.2** sample size = 23 for different versions of LUPINE; **B.1** sample size = 120 for different versions of LUPINE_single; **B.2** sample size = 120 for different versions of LUPINE. Each boxplot represents LUPINE_single and LUPINE evaluated with different inputs (e.g. clr transformed vs raw counts), distinguished by different colours. For both sample sizes, all three approaches produce similar results

Effect of using different number of components in the deflation

In our original approach, we used one component in the deflation to account for the largest variation in the control taxa. Including more components, however, could potentially improve model performance. To assess this, we evaluated model performance using one to three components. Our simulation results showed that using just one component produced similar or better outcomes compared to using two or three components, across two sample sizes (Fig. 21). In particular, the one-component approach performed better when the sample size was small.

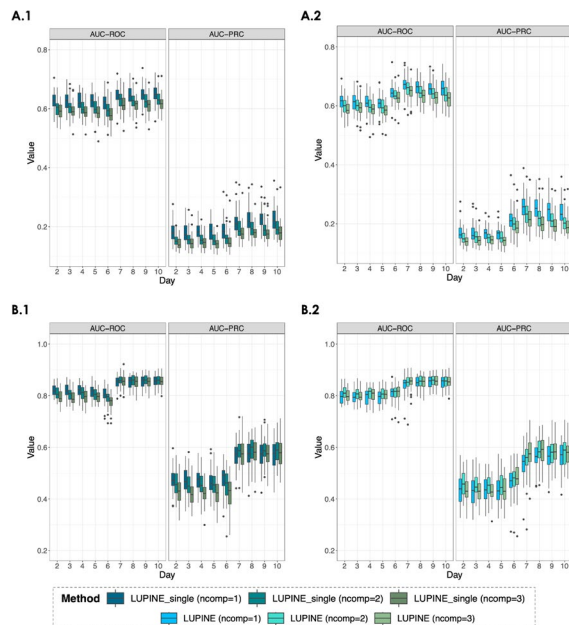


Fig. 21 Simulated data. Box plots of the area under the receiver operating characteristic (AUC-ROC), precision-recall curve (AUC-PRC) for **A.1** sample size = 23 for LUPINE_single; **A.2** sample size = 23 for LUPINE; **B.1** sample size = 120 for LUPINE_single; **B.2** sample size = 120 for LUPINE with one to three components used in deflation. Each box plot compares the performance of LUPINE_single and LUPINE with different numbers of components used in deflation. Colours differentiate between the number of components. Results are similar for both sample sizes when using one, two, or three components

Effect of using different dimension reduction methods in LUPINE_single

In LUPINE_single, we used PCA as the dimensionality reduction method. Armstrong et al. [70] suggested other dimension reduction methods that might be suitable for microbiome data. We focused on RPCA, as it was available in R and could be easily integrated into our package. We also considered ICA. We obtained similar performance for PCA, ICA, and RPCA on our simulated data. However, PCA showed slightly better performance with a larger sample size ($n = 120$) (Fig. 22). Thus, PCA is kept as the default option in the package, with alternatives for RPCA and ICA.

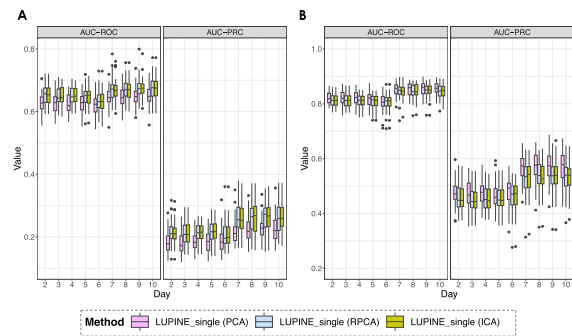


Fig. 22 Simulated data. Box plots of the area under the receiver operating characteristic (AUC-ROC), precision-recall curve (AUC-PRC) for **A** sample size = 23 **B** sample size = 120. Each boxplot represents LUPINE_single evaluated with a PCA, RPCA, and ICA, each distinguished by different colours. All three dimension reduction approaches produce similar results

Effect of number of edges

To evaluate the performance of the LUPINE_single and LUPINE models, we tested them using different edge proportions with a sample size of 23. Edge proportion was defined as the ratio of observed edges to the total number of possible edges. Model performance, measured by AUC-ROC, decreased as the edge proportion increased from 0.005 to 0.8 (Fig. 23A). This decline occurred because increasing the number of edges reduced the true partial correlations. Maintaining the positive definiteness of the partial correlation matrix required this adjustment (Fig. 23B). As partial correlations decreased, detecting significant correlations that differed from zero became more difficult, resulting in lower performance. However, performance based on AUC-PRC initially decreased and then improved, suggesting that significant edges were more likely to represent true correlations.

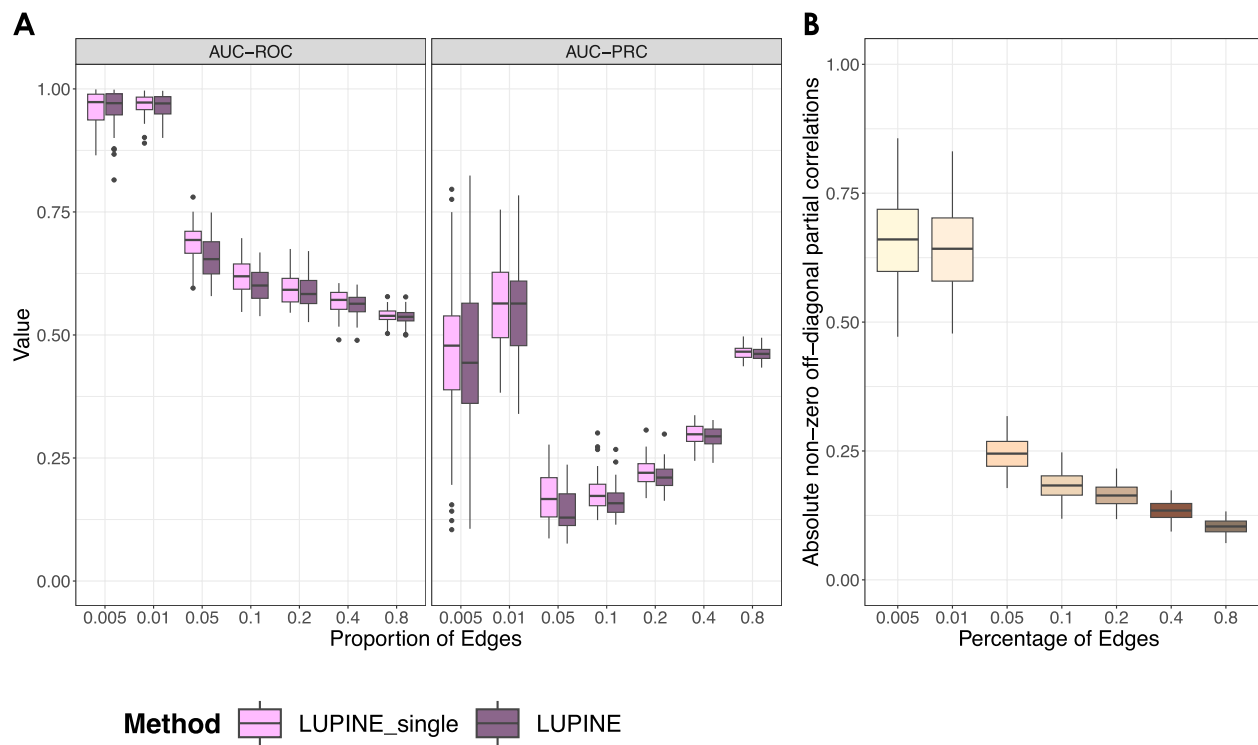


Fig. 23 Simulated data. Box plots of the **A** area under the receiver operating characteristic (AUC-ROC), precision-recall curve (AUC-PRC) for different percentages of edges with a sample size = 23 **B** true absolute partial off-diagonal correlations. As the number of edges increases, the true correlation values decrease to maintain the positive definite property of the partial correlation matrix. Consequently, as the number of edges increases, AUC-ROC performance declines

Effect of sample size

We tested model performance across different sample sizes using an edge proportion of 0.1. The choice of a 0.1 edge proportion was inspired by the edge proportions in the simulated network shown in Fig. 4A, which was inferred using SpiecEasi. Additionally, performance was generally worse at this edge proportion, as seen in the AUC-ROC and AUC-PRC values in Fig. 23.

LUPINE_single and LUPINE use a one-dimensional approximation of $p - 2$ taxa when calculating partial correlations. However, when $p < n$, it is also possible to estimate partial correlations without this approximation: we use all $p - 2$ controlled taxa in the regressions in Eq. 3.

As n approaches infinity, using all $p - 2$ taxa should converge to the actual partial correlation. Therefore, in this section, we evaluated LUPINE_single and LUPINE alongside a new approach that calculates partial correlations using all $p - 2$ taxa when $p < n$.

Increasing the sample size improved the signal-to-noise ratio, enhancing the performance of all methods (Fig. 24). For a sample size of 100, nearly double the number of features, the LUPINE approaches outperformed the method using $p - 2$ taxa in both AUC-ROC and AUC-PRC measures. However, as the sample size increased, the performance difference between the methods decreased, specifically for AUC-ROC values.

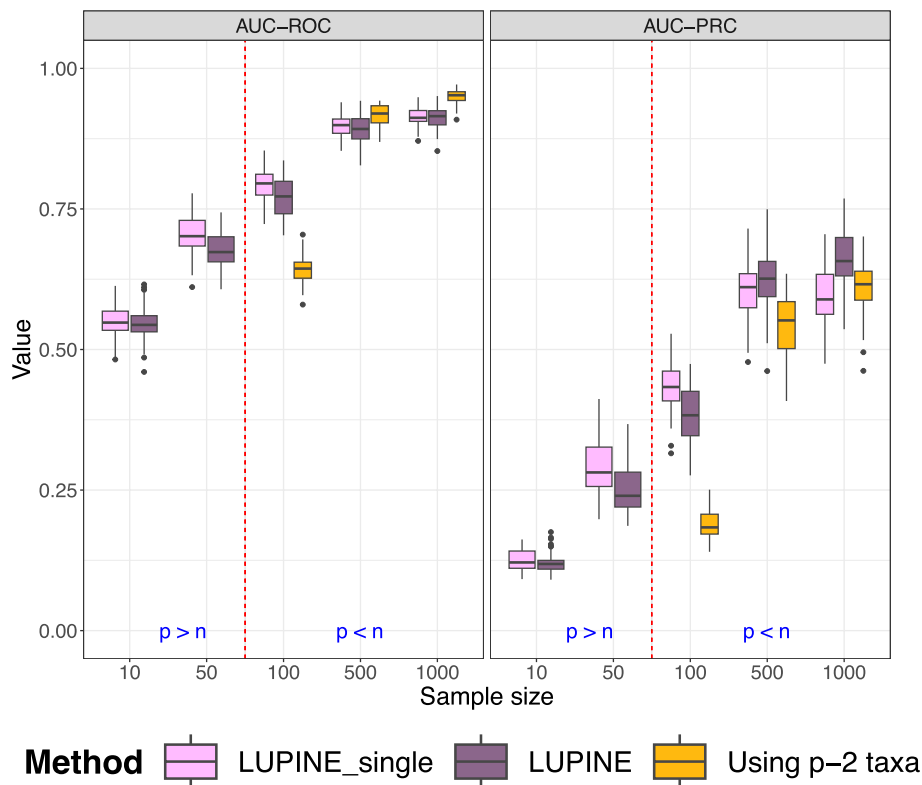


Fig. 24 Simulated data. Box plots illustrating the Area Under the Receiver Operating Characteristic (AUC-ROC) and Precision-Recall Curve (AUC-PRC) for LUPINE_single, LUPINE, and a comparative method using all $p - 2$ taxa for partial correlation estimation when $p < n$. The edge proportion is set at 0.1. Performance improves for all methods as sample size increases. At sample size of 100, LUPINE methods outperform the comparative approach in both AUC-ROC and AUC-PRC measures. This advantage decreases with larger sample sizes, especially for AUC-ROC

Acknowledgements

This work was supported by the National Health and Medical Research Council (NHMRC) Investigator Grant (GNT2025648).

Authors' contributions

Conceived and designed the study: SK KALC. Performed the analysis: SK. Wrote the paper: SK KALC.

Funding

National Health and Medical Research Council (NHMRC) investigator grant (GNT2025648)

Data availability

No datasets were generated. All the data analysed in the manuscript are publicly available on Github https://github.com/SarithaKodikara/LUPINE_manuscript.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 9 May 2024 Accepted: 17 January 2025

Published online: 03 March 2025

References

- Berg G, Rybakova D, Fischer D, Cernava T, Vergès MCC, Charles T, et al. Microbiome definition re-visited: old concepts and new challenges. *Microbiome*. 2020;8:1–22.
- Dicks LM, Geldenhuys J, Mikkelsen LS, Brandsborg E, Marcotte H. Our gut microbiota: a long walk to homeostasis. *Benefic Microbes*. 2018;9(1):3–20.
- Lyu R, Qu Y, Divaris K, Wu D. Methodological considerations in longitudinal analyses of microbiome data: a comprehensive review. *Genes*. 2023;15(1):51.
- Röttgers L, Faust K. From hairballs to hypotheses-biological insights from microbial networks. *FEMS Microbiol Rev*. 2018;42(6):761–80.
- Kurtz ZD, Müller CL, Miraldi ER, Littman DR, Blaser MJ, Bonneau RA. Sparse and compositionally robust inference of microbial ecological networks. *PLoS Comput Biol*. 2015;11(5):e1004226.
- Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome datasets are compositional: and this is not optional. *Front Microbiol*. 2017;8:2224.
- Erb I. Partial correlations in compositional data analysis. *Appl Comput Geosci*. 2020;6:100026.
- Friedman J, Alm EJ. Inferring correlation networks from genomic survey data. *PLoS Comput Biol*. 2012;8(9):e1002687.
- Kodikara S, Ellul S, Lê Cao KA. Statistical challenges in longitudinal microbiome data analysis. *Brief Bioinforma*. 2022;23(4):bbac273.
- Wold S, Sjöström M, Eriksson L. PLS-regression: a basic tool of chemometrics. *Chemometr Intell Lab Syst*. 2001;58(2):109–30.
- Tenenhaus A, Tenenhaus M. Regularized generalized canonical correlation analysis. *Psychometrika*. 2011;76:257–84.
- Lê Cao KA, Welham ZM. Multivariate data integration using R: methods and applications with the mixOmics package. New York: Chapman and Hall/CRC; 2021.
- Jolliffe IT. Principal component analysis for special types of data. New York: Springer; 2002.
- Jolliffe IT. A note on the use of principal components in regression. *J R Stat Soc Ser C Appl Stat*. 1982;31(3):300–3.
- Lin H, Peddada SD. Analysis of compositions of microbiomes with bias correction. *Nat Commun*. 2020;11(1):3514.
- Zhang X, Pei YF, Zhang L, Guo B, Pendegraft AH, Zhuang W, et al. Negative binomial mixed models for analyzing longitudinal microbiome data. *Front Microbiol*. 2018;9:1683.
- Zhang X, Guo B, Yi N. Zero-inflated Gaussian mixed models for analyzing longitudinal microbiome data. *PLoS ONE*. 2020;15(11):e0242073.
- Matchado MS, Lauber M, Reitmeier S, Kacprowski T, Baumbach J, Haller D, et al. Network analysis methods for studying microbial communities: a mini review. *Comput Struct Biotechnol J*. 2021;19:2687–98.
- Hammond DK, Gur Y, Johnson CR. Graph diffusion distance: a difference measure for weighted graphs based on the graph Laplacian exponential kernel. In: 2013 IEEE Global Conference on Signal and Information Processing. 2013. pp. 419–22. <https://doi.org/10.1109/GlobalSIP.2013.6736904>.
- Torgerson WS. Theory and methods of scaling. New York: Wiley; 1958.
- Salavaty A, Ramialison M, Currie PD. Integrated value of influence: an integrative method for the identification of the most influential nodes within networks. *Patterns*. 2020;1(5):100052.
- Waggener B, Waggener WN. Pulse code modulation techniques. Berlin: Springer Science & Business Media; 1995.
- Mantel N. The detection of disease clustering and a generalized regression approach. *Cancer Res*. 1967;27(2_Part_1):209–20.
- Sankaran K, Jeganathan P. mbtransfer: Microbiome intervention analysis with transfer functions and mirror statistics. *PLoS Comput Biol*. 2024;20(6):e1012196.
- Susin A, Wang Y, Lê Cao KA, Calle ML. Variable selection in microbiome compositional data analysis. *NAR Genomics Bioinforma*. 2020;2(2):lqaa029.
- Mu A, Carter GP, Li L, Isles NS, Vrbanc AF, Morton JT, et al. Microbe-metabolite associations linked to the rebounding murine gut microbiome postcolonization with vancomycin-resistant *Enterococcus faecium*. *Msystems*. 2020;5(4):10–1128.
- David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, et al. Diet rapidly and reproducibly alters the human gut microbiome. *Nature*. 2014;505(7484):559–63.
- Bogart E, Creswell R, Gerber GK. MITRE: inferring features from microbiota time-series data linked to host status. *Genome Biol*. 2019;20:1–15.
- Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods*. 2016;13(7):581–3.
- Ruiz-Perez D, Lugo-Martinez J, Bourguignon N, Mathee K, Lerner B, Bar-Joseph Z, et al. Dynamic bayesian networks for integrating multi-omics time series microbiome data. *Msystems*. 2021;6(2):10–1128.
- Shoer S, Shilo S, Godneva A, Ben-Yacov O, Rein M, Wolf BC, et al. Impact of dietary interventions on pre-diabetic oral and gut microbiome, metabolites and cytokines. *Nat Commun*. 2023;14(1):5384.
- Rothschild D, Leviatan S, Hanemann A, Cohen Y, Weissbrod O, Segal E. An atlas of robust microbiome associations with phenotypic traits based on large-scale cohorts from two continents. *PLoS ONE*. 2022;17(3):e0265756.
- Guo X, Li J, Tang R, Zhang G, Zeng H, Wood RJ, et al. High fat diet alters gut microbiota and the expression of paneth cell-antimicrobial peptides preceding changes of circulating inflammatory cytokines. *Mediat Inflamm*. 2017;2017(1):9474896.
- Yin J, Li Y, Han H, Chen S, Gao J, Liu G, et al. Melatonin reprogramming of gut microbiota improves lipid dysmetabolism in high-fat diet-fed mice. *J Pineal Res*. 2018;65(4):e12524.
- Daniel H, Gholami AM, Berry D, Desmarchelier C, Hahne H, Loh G, et al. High-fat diet alters gut microbiota physiology in mice. *ISME J*. 2014;8(2):295–308.
- Lee YS, Lee D, Park GS, Ko SH, Park J, Lee YK, et al. Lactobacillus plantarum HAC01 ameliorates type 2 diabetes in high-fat diet and streptozotocin-induced diabetic mice in association with modulating the gut microbiota. *Food Function*. 2021;12(14):6363–73.
- Miao Z, Cheng R, Zhang Y, Liang H, Jiang F, Shen X, et al. Antibiotics can cause weight loss by impairing gut microbiota in mice and the potent benefits of lactobacilli. *Biosci Biotechnol Biochem*. 2020;84(2):411–20.
- Lankelma JM, Belzer C, Hoogendijk AJ, De Vos AF, De Vos WM, Van Der Poll T, et al. Antibiotic-induced gut microbiota disruption decreases TNF- α release by mononuclear cells in healthy adults. *Clin Transl Gastroenterol*. 2016;7(8):e186.
- Van Passel MW, Kant R, Zoetendal EG, Plugge CM, Derrien M, Malfatti SA, et al. The genome of Akkermansia muciniphila, a dedicated intestinal

- mucin degrader, and its use in exploring intestinal metagenomes. *PLoS ONE*. 2011;6(3):e16876.
40. Byndloss MX, Olsan EE, Rivera-Chávez F, Tiffany CR, Cevallos SA, Lokken KL, et al. Microbiota-activated PPAR- γ signaling inhibits dysbiotic Enterobacteriaceae expansion. *Science*. 2017;357(6351):570–5.
 41. Kim YG, Sakamoto K, Seo SU, Pickard JM, Gilliland MG III, Pudlo NA, et al. Neonatal acquisition of Clostridia species protects against colonization by bacterial pathogens. *Science*. 2017;356(6335):315–9.
 42. Djukovic A, Garzón MJ, Canlet C, Cabral V, Lalaoui R, García-Garcera M, et al. Lactobacillus supports Clostridiales to restrict gut colonization by multidrug-resistant Enterobacteriaceae. *Nat Commun*. 2022;13(1):5617.
 43. Pessoa J, Belew GD, Barroso C, Egas C, Jones JG. The gut microbiome responds progressively to fat and/or sugar-rich diets and is differentially modified by dietary fat and sugar. *Nutrients*. 2023;15(9):2097.
 44. Li K, Epperly MW, Barreto GA, Greenberger JS, Methé BA. Longitudinal fecal microbiome study of total body irradiated mice treated with radiation mitigators identifies bacterial associations with survival. *Front Cell Infect Microbiol*. 2021;11:715396.
 45. He G, Chen T, Huang L. Tremella fuciformis polysaccharide reduces obesity in high-fat diet-fed mice by modulation of gut microbiota. *Front Microbiol*. 2022;13:1073350.
 46. Jiang X, Zhang B, Lan F, Zhong C, Jin J, Li X, et al. Host genetics and gut microbiota jointly regulate blood biochemical indicators in chickens. *Appl Microbiol Biotechnol*. 2023;107(24):7601–20.
 47. He Z, Ma Y, Chen X, Liu S, Xiao J, Wang Y, et al. Protective effects of intestinal gallic acid in neonatal dairy calves against extended-spectrum β -lactamase producing enteroaggregative Escherichia coli infection: modulating intestinal homeostasis and colitis. *Front Nutr*. 2022;9:864080.
 48. Ravelo AD, Arce-Cordero JA, Lobo RR, Liu T, Jeong KC, Faciola A. Effects of partially replacing dietary corn with sugars in a dual-flow continuous culture system on the ruminal microbiome. *Transl Anim Sci*. 2023;7(1):txad011.
 49. Lee C, Lee J, Eor JY, Kwak MJ, Huh CS, Kim Y. Effect of consumption of animal products on the gut microbiome composition and gut health. *Food Sci Anim Resour*. 2023;43(5):723.
 50. Yang J, Li Y, Wen Z, Liu W, Meng L, Huang H. Oscillospira—a candidate for the next-generation probiotics. *Gut Microbes*. 2021;13(1):1987783.
 51. Zhang X, Shen D, Fang Z, Jie Z, Qiu X, Zhang C, et al. Human gut microbiota changes reveal the progression of glucose intolerance. *PLoS ONE*. 2013;8(8):e71108.
 52. Bhute SS, Suryavanshi MV, Joshi SM, Yajnik CS, Shouche YS, Ghaskadbi SS. Gut microbial diversity assessment of Indian type-2-diabetics reveals alterations in eubacteria, archaea, and eukaryotes. *Front Microbiol*. 2017;8:214.
 53. Wu H, Tremaroli V, Schmidt C, Lundqvist A, Olsson LM, Krämer M, et al. The gut microbiota in prediabetes and diabetes: a population-based cross-sectional study. *Cell Metab*. 2020;32(3):379–90.
 54. Chen J, Vitetta L. The role of butyrate in attenuating pathobiont-induced hyperinflammation. *Immune Netw*. 2020;20(2):e15.
 55. Palacios T, Vitetta L, Coulson S, Madigan CD, Lam YY, Manuel R, et al. Targeting the intestinal microbiota to prevent type 2 diabetes and enhance the effect of metformin on glycaemia: a randomised controlled pilot study. *Nutrients*. 2020;12(7):2041.
 56. Kallassy J, Gagnon E, Rosenberg D, Silbart LK, McManus SA. Strains of Faecalibacterium prausnitzii and its extracts reduce blood glucose levels, percent HbA1c, and improve glucose tolerance without causing hypoglycemic side effects in diabetic and prediabetic mice. *BMJ Open Diabetes Res Care*. 2023;11(3):e003101.
 57. Chang WL, Chen YE, Tseng HT, Cheng CF, Wu JH, Hou YC. Gut microbiota in patients with prediabetes. *Nutrients*. 2024;16(8):1105.
 58. Li Y, Wu Y, Wu L, Qin L, Liu T. The effects of probiotic administration on patients with prediabetes: a meta-analysis and systematic review. *J Transl Med*. 2022;20(1):498.
 59. Hinnen D. Glucagon-like peptide 1 receptor agonists for type 2 diabetes. *Diabetes Spectr*. 2017;30(3):202–10.
 60. Allin KH, Tremaroli V, Caesar R, Jensen BA, Damgaard MT, Bahl MI, et al. Aberrant intestinal microbiota in individuals with prediabetes. *Diabetologia*. 2018;61:810–20.
 61. Pinna NK, Anjana RM, Saxena S, Dutta A, Gnanaprakash V, Rameshkumar G, et al. Trans-ethnic gut microbial signatures of prediabetic subjects from India and Denmark. *Genome Med*. 2021;13:1–20.
 62. Gravdal K, Kirste KH, Grzelak K, Kirubakaran GT, Leissner P, Saliou A, et al. Exploring the gut microbiota in patients with pre-diabetes and treatment naïve diabetes type 2—a pilot study. *BMC Endocr Disord*. 2023;23(1):179.
 63. Cui J, Ramesh G, Wu M, Jensen ET, Crago O, Bertoni AG, et al. Butyrate-producing bacteria and insulin homeostasis: the microbiome and insulin longitudinal evaluation study (MILES). *Diabetes*. 2022;71(11):2438–46.
 64. Takeuchi T, Kubota T, Nakanishi Y, Tsugawa H, Suda W, Kwon ATJ, et al. Gut microbial carbohydrate metabolism contributes to insulin resistance. *Nature*. 2023;621(7978):389–95.
 65. Xuan W, Ou Y, Chen W, Huang L, Wen C, Huang G, et al. Faecalibacterium prausnitzii improves lipid metabolism disorder and insulin resistance in type 2 diabetic mice. *Br J Biomed Sci*. 2023;80:10794.
 66. Lawrence I, Lin K. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*. 1989;45(1):255–68.
 67. Feller W. Die grundlagen der volterraschen theorie des kampfes ums dasein in wahrscheinlichkeitstheoretischer behandlung. *Acta Biotheor*. 1939;5(1):11–40.
 68. Simarmata D, Novkaniza F, Widyaningsih Y. A time series model: First-order integer-valued autoregressive (INAR (1)). In: AIP Conference Proceedings. vol. 1862. New York: AIP Publishing; 2017.
 69. Jazi MA, Alamatsaz M. Two new thinning operators and their applications. *Glob J Pure Appl Math*. 2012;8(1):13–28.
 70. Armstrong G, Rahman G, Martino C, McDonald D, Gonzalez A, Mishne G, et al. Applications and comparison of dimensionality reduction methods for microbiome data. *Front Bioinforma*. 2022;2:821861.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.