parafac4microbiome: Exploratory analysis of longitudinal microbiome data using Parallel Factor Analysis

G.R. van der Ploeg[1*], J.A. Westerhuis[1], A. Heintz-Buschart[1], A.K. Smilde[1]

[1] *Biosystems Data Analysis, Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, The Netherlands*

*Author for correspondence: Geert Roelof van der Ploeg g.r.ploeg@uva.nl

Geert Roelof van der Ploeg: https://orcid.org/0009-0007-5204-3386

Johan Westerhuis: https://orcid.org/0000-0002-6747-9779

Anna Heintz-Buschart: https://orcid.org/0000-0002-9780-1933

Age Smilde: https://orcid.org/0000-0002-3052-4644

Abstract

Studies investigating microbial temporal dynamics are increasingly common, leveraging longitudinal designs that collect microbial abundance data across multiple time points from the same subjects. Traditional exploratory approaches like Principal Component Analysis (PCA) fail to fully utilize this structure. By organizing data as a three-way array—subjects as rows, microbial abundances as columns, and time points as the third dimension—multi-way methods such as Parallel Factor Analysis (PARAFAC) can better capture temporal and structural patterns. This study demonstrates Parallel Factor Analysis (PARAFAC) as a method to explore longitudinal microbiome data using three exemplary studies.

In the first example, a long time series of in vitro microbiomes, PARAFAC identifies primary time-resolved variations. The second example, a longitudinal infant gut microbiome study, shows that PARAFAC can distinguish subject groups and enhance comparative analysis, even with moderate missing data. In the third example, a gingivitis intervention study of the oral microbiome, PARAFAC enables the identification of microbial subcommunities of interest through post-hoc clustering.

These examples highlight PARAFAC's broad applicability for analysing longitudinal microbiome data across diverse environments. The approach is implemented in the R package *parafac4microbiome*, available on CRAN, providing researchers with accessible tools for similar analyses.

Importance

Understanding how microbiomes change over time can give us valuable insights into their role in health and disease. Many traditional methods like Principal Component Analysis (PCA) miss important patterns in data collected over time, but Parallel Factor Analysis (PARAFAC) helps uncover these trends in a much clearer way. Using this approach, we were able to identify key changes in microbiomes across different settings, like lab experiments, the infant gut, and the mouth. PARAFAC also works well even when some data is missing, which is a common issue. To make this tool accessible, we have included it in a user-friendly R package, enabling other researchers to analyse microbiome dynamics in their own studies and explore how these changes might influence health and treatments.

Key Words

Multi-way, PARAFAC, Tensor Factorization, Decomposition, Dynamics

Background

The role of the microbiome in biological systems and in disease has been thoroughly explored [1–3]. However, most studies are cross-sectional while microbial temporal dynamics are required to understand host-microbiome and microbe-microbe interactions [4, 5]. Recently, studies that investigate these dynamics have become more frequent [6–8].

In a longitudinal microbiome study design, microbial abundance data are collected across multiple time points from the same subjects [9–11]. These data are often stored in a two-way data table, where the rows constitute the measurements (samples from multiple individuals over time), and the columns contain the microbial abundances (**Figure 1A**). Two-way ordination methods like Principal Component Analysis (PCA) [12, 13] or Multidimensional Scaling (MDS) [14] are then employed to explore the data (**Figure 1B**). However, this often results in hard-to-interpret visualizations because these methods treat every measurement as a separate object to be ordinated (**Figure 1D, E**). As such, there is a need for exploratory methods that take the crossed longitudinal study design into account.

The longitudinal study design can be used to create a three-way data array where the rows constitute the subjects, the columns contain the microbial abundances, and the third dimension contains the time points (**Figure 1A**). Three-way ordination methods like Parallel Factor Analysis (PARAFAC;

[15, 16]) can then be used to explore the data (**Figure 1C**). By making full use of the study design, the model can identify subject groups as well as their time profiles (**Figure 1D, F**).

In this work we show how Parallel Factor Analysis can be used to explore longitudinal microbiome data. We first give a mathematical description of the PARAFAC method and its main requirement, followed by a brief review of longitudinal microbiome data processing needed to meet this requirement and the principles of PARAFAC model selection. We then highlight the benefits of making a PARAFAC model using three exemplary datasets. In the first example study [6] we highlight the time-resolved variation that PARAFAC models. In the second study [7], we highlight how PARAFAC can find differences between subject groups and enhance comparative analysis despite a moderate amount of missing data. In the third study [8] we highlight a post-hoc clustering approach to identify microbes of interest with PARAFAC. The analyses and the example datasets with the resulting figures are implemented in the R package `parafac4microbiome`, which is available on CRAN at https://cran.rstudio.com/web/packages/parafac4microbiome/.

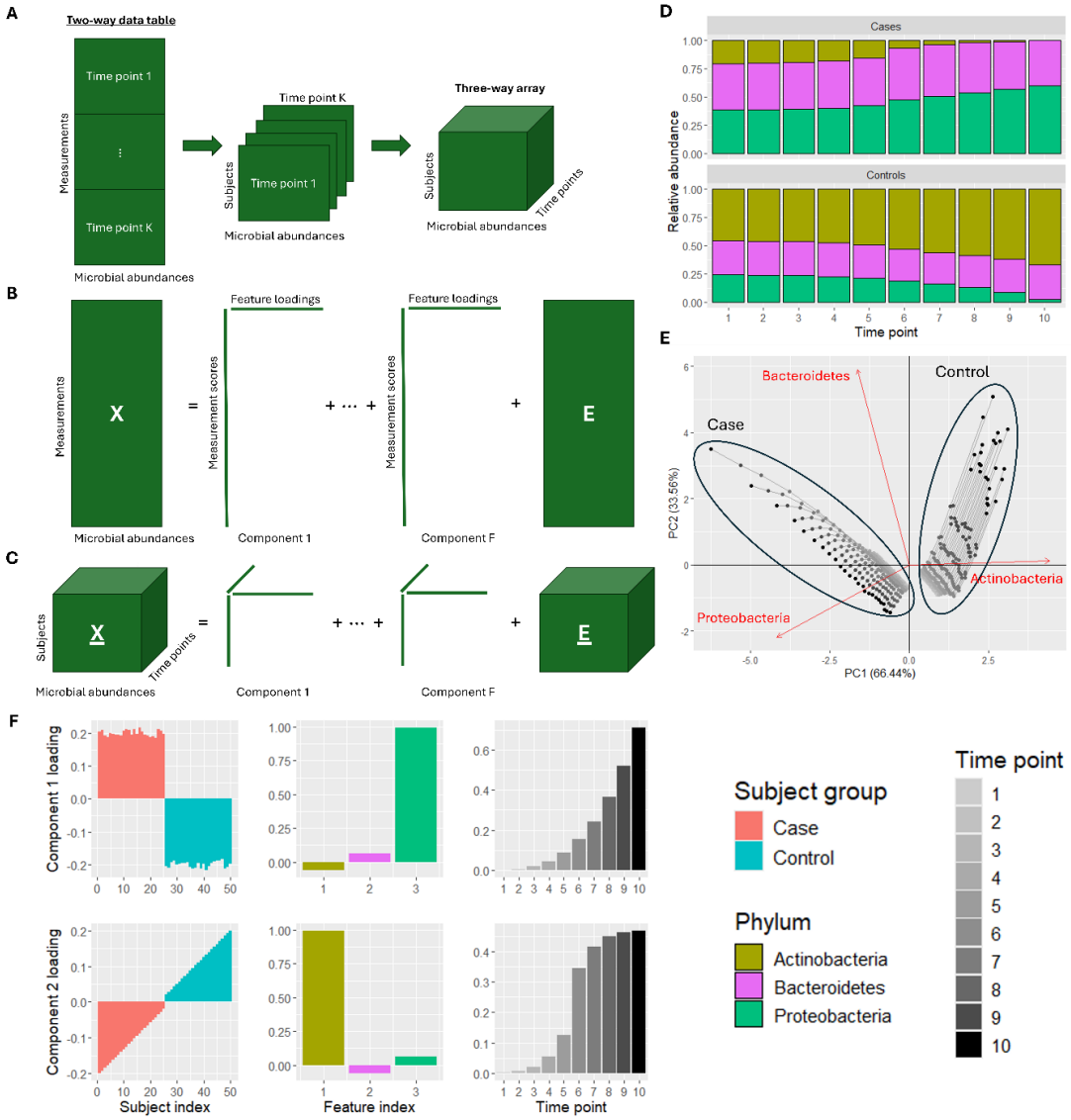


Figure 1: (A) Longitudinal microbiome data is often presented as a two-way data table with measurements (i.e. one sample for every subject-time point combination) in the rows and microbial abundances in the columns. This kind of data can be converted to a three-way array by putting all measurements for each time point in the third dimension. (B) Principal Component Analysis decomposes a two-way data table into components with scores for the measurements and loadings for the microbial relative abundances. This does not take the longitudinal study design into account. (C) Parallel factor analysis decomposes a three-way data array into components with three loading vectors each: one for the subjects, one for the microbial relative abundances, and one for time. (D) In a mock intervention dataset, we have obtained count data of three phyla (Actinobacteria, Bacteroidetes and Proteobacteria) corresponding to 25 case and 25 control subjects across 10 time points. At time point 1, all subjects look very similar to each other. At time point 10 the case subjects are depleted of Actinobacteria and enriched in Proteobacteria, while for the control subjects the opposite is true. The magnitude of the depletion and enrichment of Actinobacteria at time point 10 varies substantially within the case and control groups, respectively. The simulated counts corresponding to Bacteroidetes have no time-resolved subject group differences. (E) Biplot of a Principal Component Analysis of the mock dataset after centred log-ratio transformation, centring and scaling. Samples taken from the same subject are connected with a line. The first principal component, explaining 66.44% of the variation, separates the case and control samples based on the abundance of Actinobacteria. The second principal component, explaining 33.56% of the variation, separates the time points using the Proteobacteria and Bacteroidetes abundances. However, the time profiles per subject group are unclear. (F) Plot of the Parallel Factor Analysis model of the data, describing subject group differences and their time profiles. In component 1 the exponential growth of Proteobacteria in the case subjects is identified. In component 2 the sigmoid shaped growth curve of Actinobacteria in the control subjects is identified. The magnitude of the depletion and enrichment of Actinobateria within the case and control groups is visible as a gradient of loading values. Bacteroidetes has a loading close to zero in both components, correctly identifying it as having no discriminating time-resolved variation. This illustrates that PARAFAC can capture various time-resolved differences in microbial abundances between the subject groups in a more interpretable way than PCA.

Methods

Notation and definitions

We briefly define the mathematical notation that will be used throughout this paper. Scalars are indicated by lower-case italics such as a. Vectors are indicated by bold lower-case characters such as $\boldsymbol{b}$. Two-way matrices are indicated with bold capitalized characters such as $\boldsymbol{X}$. Underlined bold capitalized characters are used for three-way arrays such as $\underline{\boldsymbol{X}}$. The letters I, J, and K are reserved to indicate the dimension of the subject, microbial abundance, and time mode, respectively. Hence the element for subject $i \in (1 \dots I)$, microbial abundance $j \in (1 \dots J)$ at timepoint $k \in (1 \dots K)$ of a three-way array $\underline{\boldsymbol{X}}$ is called x_{ijk}. We create the data cubes of microbial relative abundances such that the subjects are in the first mode, the microbial taxa are in the second mode, and time points are in the third mode. We will not distinguish between the terms 'factor' and 'component', nor between the terms 'way' and 'mode'. Generally, we will use the words 'component' and 'mode' throughout this paper.

Mathematical definition of Parallel Factor Analysis

Parallel Factor Analysis (PARAFAC) [15, 16] is a factor decomposition method that generalizes Principal Component Analysis (PCA) to multi-way data. First, we give a brief mathematical summary of PCA to make the generalizing principle of PARAFAC clearer. We then describe the algorithmic implementation of this method and its properties.

A longitudinal microbiome dataset is often represented as a two-way matrix $\boldsymbol{X}$ $(IK \times J)$ with all time points for all subjects as rows and the microbial abundances as columns (**Figure 1B**). For convenience, we will refer to a subject-timepoint combination as a measurement $m \in (1 \dots M)$ where $M = IK$. A PCA model of this two-way dataset $\boldsymbol{X}$ $(M \times J)$ is then given by the two matrices $\boldsymbol{T}$ $(M \times F)$ and $\boldsymbol{P}$ $(J \times F)$ containing the scores and loadings of all F components respectively. The scores describe the differences between the measurements, and the loadings describe the microbial abundances that are different. A PCA model of such a two-way matrix can be formulated elementwise as

$$x_{mj} = \sum_{f=1}^{F} t_{mf} p_{jf} + e_{mj} \tag{1}$$

with x_{mj} the matrix element corresponding to measurement m and feature j, t_{mf} the score of measurement m in component f, p_{jf} the loading of feature j in component f and e_{mj} the residual of the matrix element. PCA is a bilinear model as we are using F sets of two vectors to approximate the dataset $\boldsymbol{X}$. This also reveals the main issue at hand: the measurements M are ordinated without taking the connection to the subjects I and time points K into account.

A three-way data array $\underline{\boldsymbol{X}}$ is defined as having I subjects in the rows, J microbial abundances in the columns and K time points in the third dimension (**Figure 1C**). A PARAFAC model of a three-way data array $\underline{\boldsymbol{X}}$ $(I \times J \times K)$ is given by three matrices $\boldsymbol{A}$ $(I \times F)$, $\boldsymbol{B}$ $(J \times F)$ and $\boldsymbol{C}$ $(K \times F)$ containing the loadings of all F components corresponding to the subject, feature, and time mode, respectively. The element-wise definition of the model is formulated as

$$x_{ijk} = \sum_{f=1}^{F} a_{if} b_{jf} c_{kf} + e_{ijk} \tag{2}$$

with x_{ijk} the matrix element corresponding to subject i, feature j at time point k, a_{if} the loading of subject i of component f, b_{jf} the loading of feature j of component f and c_{kf} the loading of time point k of component f and e_{ijk} the residual of the matrix element. In the elementwise definition, the generalization of PCA is readily visible: the loading for the time mode c_{kf} has been added to the calculation of element x_{ijk} to represent the time factor. PARAFAC is a trilinear model as we are using F sets of three vectors to approximate the data array $\boldsymbol{X}$.

Algorithmic implementation of Parallel Factor Analysis

The objective of the algorithmic implementation of PARAFAC is to minimise the sum-of-squares of the residuals through the loss function

$$L(a_{11}, a_{12}, \dots, c_{KF}) = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \left(x_{ijk} - \sum_{f=1}^{F} a_{if} b_{jf} c_{kf} \right)^2 \tag{3}$$

Various implementations of estimating the PARAFAC loading matrices exist [17]. In this paper we estimate the PARAFAC loading matrices through an iterative alternating least-squares algorithm (ALS; [16, 18]), as this implementation is well-documented and robust towards missing values or time points (**Table 1**).

Table 1: Algorithmic implementation of the alternating least squares based PARAFAC model estimation.

Parallel Factor Analysis (PARAFAC) ALS algorithm for a given three-way array $\underline{\boldsymbol{X}}$
(0) Decide on the number of components F
(1) Initialise $\boldsymbol{B}$ and $\boldsymbol{C}$
(2) Estimate $\boldsymbol{A}$ from $\underline{\boldsymbol{X}}, \boldsymbol{B}$ and $\boldsymbol{C}$ such that $L(a_{11}, a_{12}, \dots, c_{KF})$ is minimised
(3) Estimate $\boldsymbol{B}$ from $\underline{\boldsymbol{X}}, \boldsymbol{A}$ and $\boldsymbol{C}$ such that $L(a_{11}, a_{12}, \dots, c_{KF})$ is minimised
(4) Estimate $\boldsymbol{C}$ from $\underline{\boldsymbol{X}}, \boldsymbol{A}$ and $\boldsymbol{B}$ such that $L(a_{11}, a_{12}, \dots, c_{KF})$ is minimised
(5) Continue to step (2) until convergence.

Various approaches for initialisation of $\boldsymbol{B}$ and $\boldsymbol{C}$ have been proposed and are documented elsewhere [19]. In this paper, we used randomly initialised vectors to determine the appropriate number of components and singular value decomposition-based initialised vectors to investigate the stability of the loadings.

We use the PARAFAC algorithm supplied in the `parafac4microbiome` R package (v1.1.2, [20]). With our default settings, the model converges in step (5) when the change in sum-of-squares is lower than 1e-4, or terminates after 500 iterations of steps 2-4. In this paper, the convergence threshold is set lower to ensure that PARAFAC models corresponding to a global minimum are reported. Fitting a model for any of the processed datasets in this paper takes less than a minute of computational time on an average computer (Microsoft Windows 11 Home v10.0.22631 with an 12th Gen Intel® Core™ i5-12400F running 2500 MHz with 6 cores or 12 logical processors and 16 Gb RAM).

PARAFAC model constraints and properties

A PARAFAC decomposition is unique, apart from trivial rescaling and permutations of the factors. This means that estimated component matrices cannot be rotated without a loss of fit. This is a distinct advantage over Principal Component Analysis [16, 21–24]. Additionally, PARAFAC can capture any type of temporal dynamics – linear, exponential or otherwise - because it models time categorically. Thus, the time points are not connected like in a dynamic model. By treating time as a categorical variable, it is assumed that sampling time points are appropriately spaced, and each sampling time point receives a fair contribution to the model. A limitation of PARAFAC is that all subjects must be measured on the same time points. While the method cannot interpolate or extrapolate to time points that were not originally measured, it can interpolate data that is missing

174 in some subjects. Furthermore, the components are orthogonal in PCA while in PARAFAC the
175 components are non-orthogonal by default. The orthogonality constraint in PCA results in models
176 that force interpretation in terms of orthogonal factors which do not always align with the
177 underlying biology, whereas the absence of such a constraint for PARAFAC results in models that
178 can describe processes that are not necessarily independent. Non-orthogonality is an advantage
179 because it is more likely a realistic representation of the biological system of interest. The downside
180 of non-orthogonality is that components can be very similar to each other. As such, interpretation
181 should be done with care: examining one component at a time can lead to wrong conclusions and
182 plotting the loadings of two components against each other requires a transformation
183 (**Supplementary Methods**). Possible additional constraints such as orthogonality and non-negativity
184 of the loadings are described elsewhere [19, 20]. In this paper we will not constrain the PARAFAC
185 models.

186 *Processing of longitudinal microbiome data*

187 The main requirement of PARAFAC is that the residuals are distributed symmetrically around zero
188 [15, 16, 19]. In this paper we perform the following four processing steps for each dataset to meet
189 this requirement: centred log-ratio transformation, feature filtering, centring and scaling. In this
190 section we outline these operations for multi-way data arrays (**Figure 2**). In the supplied R package,
191 the function processDataCube() performs each of these steps according to user specifications.

192 Centred log-ratio transformation (CLR) is a popular method to create a version of the data that can
193 be ordinated and interpreted through Euclidean distances [25]. This transformation approach deals
194 with the compositionality inherent to microbiome data and makes the data continuous [26].
195 However, it is hampered by the zero-inflation commonly observed in microbiome data because
196 zeroes are not defined in log space [27]. Here we choose to CLR transform the datasets after adding
197 a pseudo-count of 1 to all counts (hereafter $\underline{\boldsymbol{X}^{clr}}$), which appears to yield similar models as other
198 pseudo-count strategies based on a brief sensitivity analysis (**Supplementary Figure 18**).

199 Due to the high dimensionality of the microbiome data, filtering of the features is a requirement to
200 model the data appropriately and quickly [28, 29]. However, in longitudinal microbiome data the
201 complete time trajectory of a feature should be considered prior to filtering. As such, we define the
202 sparsity of a feature as the fraction of measurements where the feature has zero counts. A sparsity
203 threshold can then be established to select features that are common, thus controlling the number

of zeroes in the data. Sparsity-based feature filtering is based on the count data and is applied after the centred log-ratio transformation step. Since sparsity filtering is heavily dependent on the sequencing depth and microbiome complexity, no overall sparsity threshold can be suggested here. Therefore, a manual inspection of the sparsity per feature histogram is required to find a good balance between removing too many features and keeping features with enough variation of interest. The calculateSparsity() function in the supplied R package can be used for this purpose.

After CLR transformation and feature filtering, we used centring to remove offsets in the data [30, 31]. This can be done in three different directions for three-way arrays: across the subject mode, across the feature mode and across the time mode. Centring across the subject mode removes the average subject time profile from every combination of microbial abundance and time point, i.e.

$$x_{ijk}^{cnt} = x_{ijk} - \frac{\sum_{i=1}^{I} x_{ijk}^{clr}}{I}. \tag{4}$$

Centring across the feature mode removes the average feature time profile from every combination of subject and time point, i.e.

$$x_{ijk}^{cnt} = x_{ijk} - \frac{\sum_{j=1}^{J} x_{ijk}^{clr}}{J}. \tag{5}$$

Centring across the time mode removes the average overall time profile from every combination of subject and microbial abundance, i.e.

$$x_{ijk}^{cnt} = x_{ijk} - \frac{\sum_{k=1}^{K} x_{ijk}^{clr}}{K}. \tag{6}$$

Here x_{ijk}^{cnt} refers to the centred value of the ijk-th element in the data cube across the various directions specified above, and x_{ijk}^{clr} refers to the CLR transformed value of the ijk-th element. Both centring across the subject mode and across the feature mode affects the interpretation of the time mode as deviations from the average profile will be modelled. Since we expect individual offsets to be present in the exemplary studies and wish to focus on differences between the individuals, we centre across the subject mode. As a result, the PARAFAC models will describe deviations from the average time profile.

After centring, we scaled the data to change the importance attached to different parts [30, 31]. For three-way arrays, this involves scaling a two-way matrix to standard deviation one. As such, there are two possibilities for three-way arrays: scaling within the subject mode and scaling within the

feature mode. Scaling within the subject mode makes the standard deviation for all counts across all subjects and microbial abundances per time point equal to one, i.e.

$$x_{ijk}^{scl} = \frac{x_{ijk}^{cnt}}{\sqrt{\sum_{j=1}^{J}\sum_{k=1}^{K}\frac{\left(x_{ijk}^{cnt}\right)^2}{JK}}} \,. \tag{7}$$

This makes all time points equally important for the modelling procedure. Scaling within the feature mode makes the standard deviation across all subjects and time points per microbial abundance equal to one, i.e.

$$x_{ijk}^{scl} = \frac{x_{ijk}^{cnt}}{\sqrt{\sum_{i=1}^{I}\sum_{k=1}^{K}\frac{\left(x_{ijk}^{cnt}\right)^2}{IK}}} \,. \tag{8}$$

Here x_{ijk}^{scl} refers to the scaled value of the ijk-th element in the data cube within the various directions specified above and x_{ijk}^{cnt} refers to the CLR transformed and centred value of the ijk-th element in the data cube. This makes all microbial abundances equally important for the modelling procedure. Since we are interested in groups of microbiota with different time profiles, we scale within the feature mode. The processing steps used for all datasets in this paper are outlined in **Figure 2**.

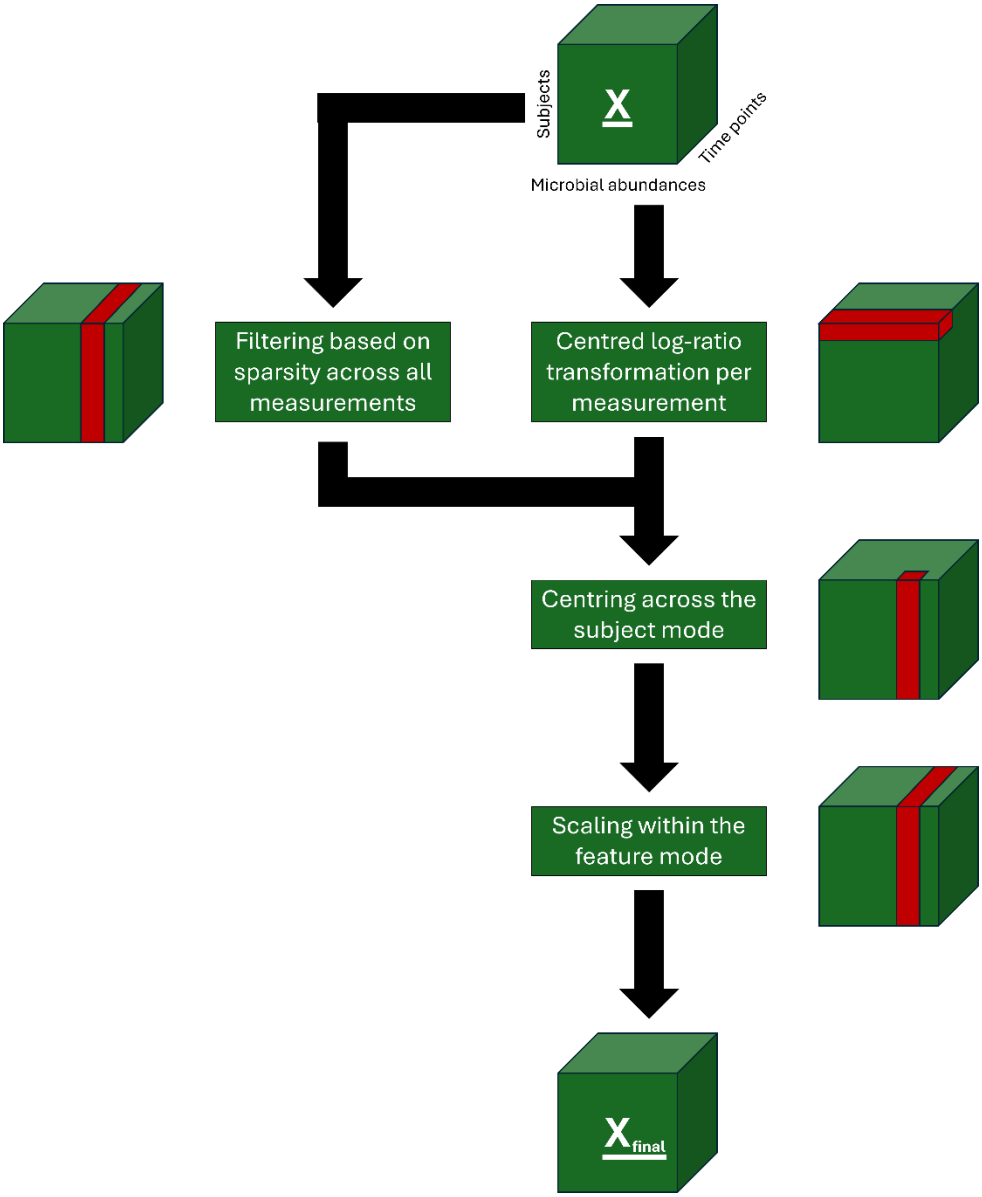


Figure 2: Workflow diagram of steps taken to process the microbiome count data for each dataset presented in this paper. The microbiome count data is indicated as $\underline{X}$ at the top of the figure. Per step, a small figure of the data cube is shown with the part of the cube that is used for the operation indicated in red. Sparsity-based feature filtering is based on the count data and is applied after the centred log-ratio transformation step. The final version of the data that is used as input for the PARAFAC modelling procedure is indicated as $\underline{X}_{final}$ at the bottom of the figure.

Example dataset: Fujita et al., 2023

The first example is a long time series of experimental in vitro microbiomes [6]. Culture conditions, microbiome sampling, 16S rRNA gene amplicon sequencing and data pre-processing are described in the original publication. We analyse only one experimental set-up, the aquatic microbiome culture in a peptone medium, which was performed in eight replicates and measured daily for 110 days.

256 The amplicon sequence variant (ASV) data and taxonomic information were processed using the
257 `parafac4microbiome` package (v1.0.3) in R (v4.4.1; [32]). The supplied dataset is a three-way array
258 of counts of size 8 replicates x 28 microbial (ASV) abundances x 110 time points. Features were kept
259 if they had ≤99% sparsity across the dataset (**Supplementary Figure 1**). This step resulted in 23
260 features being selected. We performed a centred log-ratio transformation with a pseudo-count of 1
261 to correct for compositionality [25, 26]. The data was then centred across the subject (here: replicate)
262 mode and scaled within the feature mode to focus the analysis on describing differences from the
263 average time trajectory. PARAFAC models of the processed data were created with parameters set
264 to ctol=1e-6 and maxit=500. All models converged successfully by reaching the convergence
265 tolerance threshold.

266 *Example dataset: Shao et al., 2019*

267 The second example is a longitudinal faecal microbiome study with samples collected from 314
268 vaginally born and 282 caesarean-section born infants [7]. Sampling was performed at four time
269 points post-partum: day 4, 7, 21, and a follow-up sample was taken between 4 to 12 months (referred
270 to as the infancy period). Recruitment, microbiome sampling and 16S rRNA gene amplicon
271 sequencing and pre-processing are described in the original publication.

272 Pre-processed sequencing data were obtained from ENA (accession numbers ERP115334 and
273 ERP024601). Taxonomic classification was performed using mOTUs (v2.5.1; [33]). The marker-
274 gene based operational taxonomic unit (mOTU) data and taxonomic information were processed
275 using the `parafac4microbiome` package (v1.0.3) in R (v4.4.1; [32]). The supplied dataset is a three-
276 way array of counts of size 395 subjects x 959 microbial (mOTU) abundances x 4 time points, keeping
277 missing measurements as a row of NAs. This allows PARAFAC to impute the missing values. In
278 total, 62.5% of the data was missing (**Supplementary Tables 1-2**). mOTUs were kept if the sparsity
279 in either birth mode group was ≤90% (**Supplementary Figure 5**). This conserves biologically relevant
280 mOTUs that only appear in one group and resulted in 91 of the 959 mOTUs being selected. The data
281 was then centred log-ratio transformed using a pseudo-count of 1 to correct for compositionality
282 [25, 26]. Subsequently, the data was centred across the subject mode and scaled within the feature
283 mode to focus the analysis on describing differences from the average time trajectory. PARAFAC
284 models of the processed data were created with parameters set to ctol=1e-8 and maxit=500 to ensure
285 proper exploration of the loss space given the large amount of missing data. All models converged
286 successfully by reaching the convergence tolerance threshold.

287 *Example dataset: Van der Ploeg et al., 2024*

288 The third example is a longitudinal oral microbiome gingivitis intervention study [8]. Full details
289 on the microbiome sampling and pre-processing are described in the original publication. We
290 analyse only one oral microbiome niche, the supragingival plaque at the upper jaw lingual surfaces,
291 which was measured in 41 subjects at 7 time points: day -14 (baseline), 0 (baseline, start of
292 intervention), 2, 5, 9, 14 (end of intervention) and 21 (resolution). Using quantitative light-induced
293 fluorescence [34–36], the subjects were stratified into low, medium and high responders based on
294 the amount of red fluorescent plaque on day 14.

295 The ASV data and taxonomic information were processed using the `parafac4microbiome` package
296 (v1.0.3) in R (v4.4.1; [32]). The supplied dataset is a three-way array of counts of size 41 subjects x
297 2253 microbial (ASV) abundances x 7 time points. In accordance with the original analysis, ASVs
298 were kept if the sparsity was ≤50% in any response group. This step resulted in 65 of the 2253 ASVs
299 being selected. Subsequently, a centred-log ratio transformation using a pseudo-count of 1 was
300 performed to correct for compositionality [25, 26]. The data was then centred across the subject
301 mode and scaled within the feature mode to focus the analysis on describing differences from the
302 average time trajectory. PARAFAC models of the processed data were created with parameters set
303 to ctol=1e-6 and maxit=500. All models converged successfully by reaching the convergence
304 tolerance threshold.

305 *Selecting the appropriate number of PARAFAC components*

306 Like in Principal Component Analysis, the number of biologically meaningful components that can
307 be extracted from longitudinal microbiome data is limited. After a given number of components,
308 PARAFAC will start to model noise. As such, one challenge of PARAFAC modelling for longitudinal
309 microbiome data is selecting the appropriate number of components. A model with the correct
310 number of components will (1) converge quickly, (2) not have highly similar loading vectors in any
311 mode, (3) describe meaningful biological phenomena and (4) describe these phenomena even when
312 one or more samples are removed from the data.

313 Various diagnostics and methods have been proposed and need to be manually examined to
314 determine the appropriate number of components for a PARAFAC model [37]. We propose a two-
315 step approach for selecting the correct number of components. First, we investigate quick
316 convergence and similarity of the loading vectors by randomly initializing a number of PARAFAC

317 models for 1-5 components. Next, we investigate the stability of the model using a jack-knife
318 procedure while keeping the initialized vectors the same using a singular value decomposition. In
319 the supplied R package, this procedure is automatically performed and visualized for inspection by
320 the user with the functions assessModelQuality() and assessModelStability(), respectively. All
321 output plots examined to select the number of components for each dataset are shown in
322 **Supplementary Figures**.

323 *Metrics for assessing the appropriate number of components*

324 For each dataset, we have investigated PARAFAC models with 1-5 components and a number of
325 random initializations, depending on the size of the data and how quickly the models converge. In
326 the supplied R package, this procedure is automatically performed and visualized for inspection by
327 the user with the function assessModelQuality(). We examine the following metrics of interest.

328 In PCA, the residual sum-of-squares are plotted against the number of components in a scree-plot
329 [38, 39]. The appropriate number of components is then the area of the plot where the residual sum-
330 of-squares flattens out. For PARAFAC this principle holds as well [19].

331 The alternating least-squares implementation of PARAFAC that is used throughout this paper
332 iteratively updates the loading vectors until the convergence threshold is reached. Therefore, the
333 number of iterations needed to converge is a useful metric to inspect for the correct number of
334 components. Randomly initialized PARAFAC models with the correct number of components
335 should converge quickly compared to models with an extra component modelling noise [37, 40].

336 The core consistency diagnostic (CORCONDIA) indicates how well the assumed tri-linearity of the
337 PARAFAC model holds [41]. The range of CORCONDIA is typically 0-100 but can be negative. The
338 number of components in PARAFAC models with a CORCONDIA of 50-100 can be considered
339 appropriate [19, 41].

340 The components of a PARAFAC model are non-orthogonal by default, which can sometimes lead
341 to very similar loading vectors if the model cannot be estimated properly. The Tucker congruence
342 coefficient identifies this similarity between pairs of loading vectors of the same mode [42, 43]. A
343 value in the range of 0.85-0.94 suggests that the loading vectors are fairly similar, whereas a value
344 of 0.95-1 means that the loading vectors are essentially equal [42]. The Tucker congruence
345 coefficient in the `parafac4microbiome` package uses the implementation from the multiway R
346 package (version 1.0-6; [20]).

347 *Assessing model stability through the jack-knifing of samples*

348 We expect a PARAFAC model with the correct number of components to describe meaningful
349 biological phenomena, even when one or several samples are removed from the data. We test this
350 by performing iterative jack-knifing of samples [44]. In the supplied R package, this procedure is
351 automatically performed and visualized for inspection by the user with the function
352 assessModelStability(). With this jack-knifing approach, one random sample is removed in every
353 iteration, after which a PARAFAC model is fitted to the data. We perform balanced jack-knifing,
354 removing one sample from every subject group in every iteration, if such subject metadata is
355 available. If this is impossible due to unequally sized groups, samples are randomly chosen. We
356 choose to initialise the modelling algorithm using a singular value decomposition for every iteration
357 to ensure that instability can only be caused by the removed sample(s). A PARAFAC model with
358 the correct number of components will be stable (i.e. have a small standard deviation for every
359 loading) across all jack-knifing iterations.

360 *Post-hoc clustering of PARAFAC loadings*

361 Clustering of the feature loadings can help identify groups of microbiota that are connected to a
362 particular time profile in the PARAFAC model. We performed the clustering procedure of the
363 feature loadings as follows. In accordance with the original analysis in van der Ploeg et al. [8],
364 features were not considered if the variation explained by the model for their entire time trajectory
365 was lower than the average or the congruence loading of the feature was lower than 0.4 [45]. The
366 remaining features were then clustered based on their modelled time trajectories using the K-
367 medoids algorithm from the cluster R package (v2.1.4; [46]) with 50 random starts to be robust
368 against outliers. The number of clusters was determined using the within-cluster sum of squares,
369 silhouette width and gap statistic metrics as reported by the factoextra R package (**Supplementary**
370 **Figure 17**; v1.0.7; [47]).

371 **Results**

372 *PARAFAC can identify the main time-resolved variation in longitudinal microbiome data*

373 The first example is a long time series of experimental in vitro microbiomes [6]. We analyse only
374 one experimental set-up, the aquatic microbiome culture in a peptone medium, which was
375 performed in eight replicates and measured daily for 110 days. With this example, we aim to
376 visualize the main time-resolved variation that can be modelled through PARAFAC.

377 We inspected the unprocessed count data by computing the relative abundances for every replicate
378 culture at every time point (**Figure 3A**). In most samples, 2 amplicon sequence variants (ASVs)
379 belonging to *Aeromonas* spp. comprised most of the counts. The exceptions were: the emergence of
380 *Cedecea*, *Burkholderia-Caballeronia-Paraburkholderia*, and *Pandoraea* species between days 50-110
381 in replicate culture 4, the emergence of *Clostridioides mangenotii* and a *Lachnoclostridium* species
382 between days 80-100 in replicate cultures 4 and 8, the disappearance of a *Salmonella*, *Propionivibrio*,
383 *Microbacter* and *Herbaspirillum* species between days 0-60 in replicate culture 3, the transient
384 appearance of a *Microbacter* species in replicate cultures 2 and 8, and a (small) increase in a
385 *Salmonella* species between days 100-110 in most replicate cultures. There was no clear time profile
386 visible that all replicates had in common.

387 We inspected PARAFAC modelling performance and stability for the processed count data using 1-
388 5 components. The core consistency diagnostic (CORCONDIA) scores dropped below 80 for 5
389 components. The three and four-component models explained 38.78% and 45.08% on average,
390 respectively (**Supplementary Figure 2**). Performing a jack-knifing procedure showed that the three-
391 component PARAFAC models were more stable than the four-component models (**Supplementary**
392 **Figures 3-4**). Hence, we selected the model explaining the most variation from all randomly
393 initialized models of three components. The selected model explained 39.0% of the variation (**Figure**
394 **3B**).

395 As expected, the selected PARAFAC model did not describe *Aeromonas* ASVs, because these do not
396 vary strongly over time. The first component described the emergence of a *Cedecea*, *Burkholderia-*
397 *Caballeronia-Paraburkholderia* and *Pandoraea* species between days 50-110, which is unique to
398 replicate 4. The second component described the disappearance of a *Salmonella*, *Propionivibrio*,
399 *Microbacter* and *Herbaspirillum* species between days 0-60, which is unique to replicate 3. The third
400 component described the emergence of *Clostridioides mangenotii* and a *Lachnoclostridium* species
401 between days 80-100 in replicates 4 and 8. Some of these microbiota may have low relative
402 abundances but can still describe meaningful differences between the replicates.

403 While this example highlights the strength of PARAFAC: identifying and modelling the largest
404 sources of variation in a longitudinal microbiome dataset, the PARAFAC model only described the
405 variation partially (39%). As a result, the transient appearance of a *Herbaspirillum* species in
406 replicate cultures 2, 3 and 8, nor the (small) increase in a *Salmonella* species between days 100-110
407 in most replicate cultures, were not modelled.

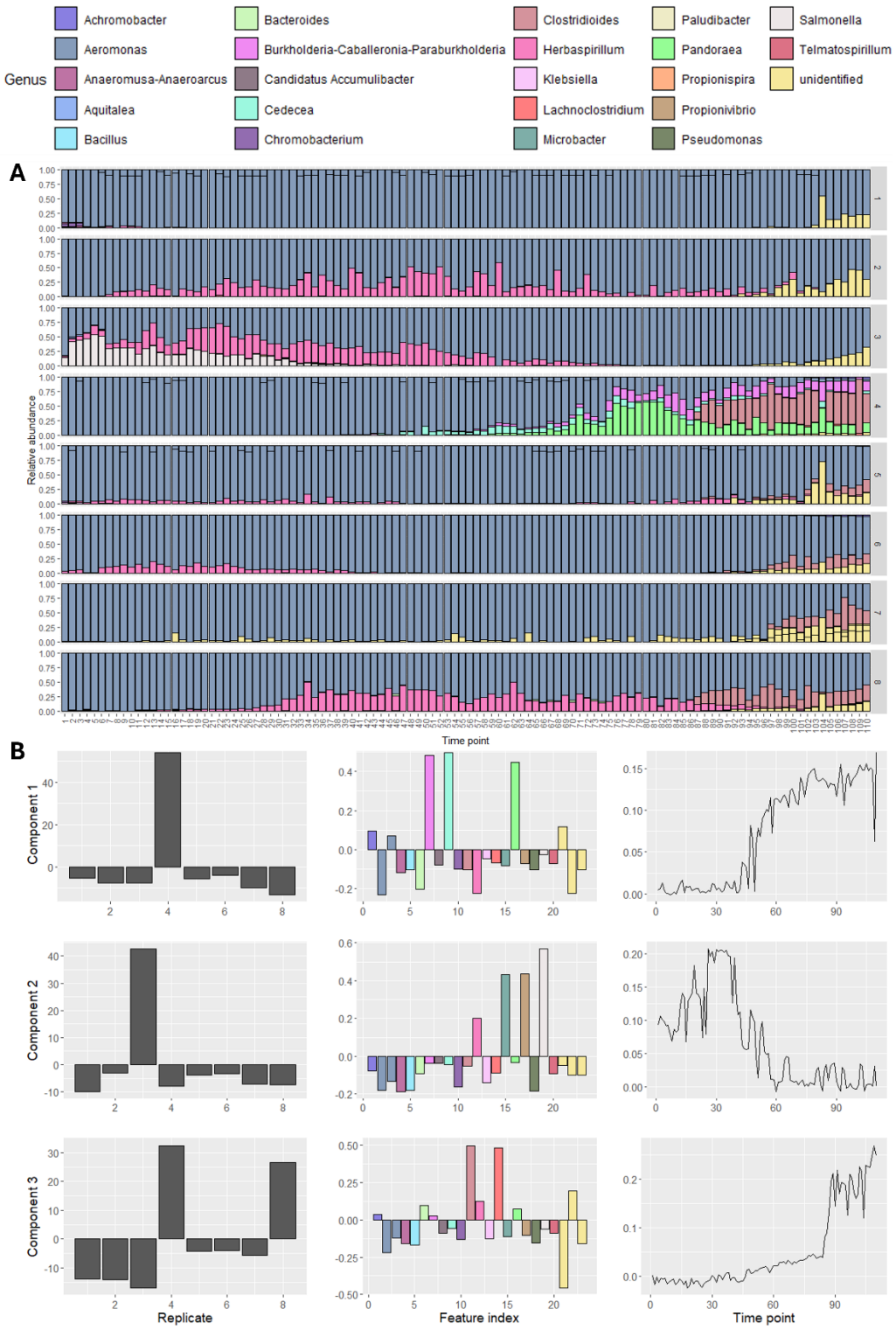


Figure 3: Overview of the Fujita et al., 2023 dataset. (A) Relative abundance plot of the dataset, split per replicate. In most samples, Aeromonas species comprise the vast majority of the counts. In the latter half of the samples in replicate 4 and the first half of the samples in replicate 3 other species are present that comprise the majority. There is no clear time profile visible that all replicates have in common. (B) The best three-component PARAFAC model of the processed data describing 39.0% of the variation. The components are shown in the rows, while the replicate, microbial abundance and time modes are shown in the columns. The first component described the emergence of a Cedecea, Burkholderia-Caballeronia-Paraburkholderia and Pandoraea species between days 50-110, which is unique to replicate 4. The second component described the disappearance of a Salmonella, Propionivibrio, Microbacter and Herbaspirillum species between days 0-60, which is unique to replicate 3. The third component described the emergence of Clostridioides mangenotii and a Lachnoclostridium species between days 80-100 in replicates 4 and 8.

419 *PARAFAC finds differences between subject groups and enhances comparative analyses*

420 The second example is a longitudinal infant faecal microbiome study [7]. Samples were collected
421 from 314 vaginally born and 282 caesarean-section born infants at four time points: 4, 7 and 21 days
422 after birth and a follow-up sample during infancy (between 4-12 months after birth). One of the
423 four measurements is missing for most infants in this cohort, and we keep the missing data in the
424 data cube for the PARAFAC algorithm to impute (**Supplementary Table 1-2**). With this example,
425 we show how PARAFAC can identify the differences between the subject groups and enhance the
426 comparative analysis done by the original authors despite a moderate amount of missing data.

427 The authors stated that there was a very clear distinction in the microbiomes of vaginally and
428 caesarean-section born children, also over time. To analyse this distinction, we inspected the
429 unprocessed count data by computing the average relative abundance for the vaginal and caesarean
430 groups at every time point (**Figure 4A**). This confirmed a clear difference at the phylum level
431 between the vaginal and caesarean groups at day 4, 7 and 21 postpartum, which disappeared in
432 infancy.

433 We inspected PARAFAC modelling performance and stability for the processed count data using 1-
434 5 components. The CORCONDIA scores dropped below 80 at 3-5 components (**Supplementary**
435 **Figure 6**). Performing a jack-knifing procedure showed that the two-component model was as stable
436 as the three-component model (**Supplementary figures 7-8**). Hence, we selected the model
437 explaining the most variation from all randomly initialized models of three components. The
438 selected model explained 9.40% of the variation (**Figure 4B**).

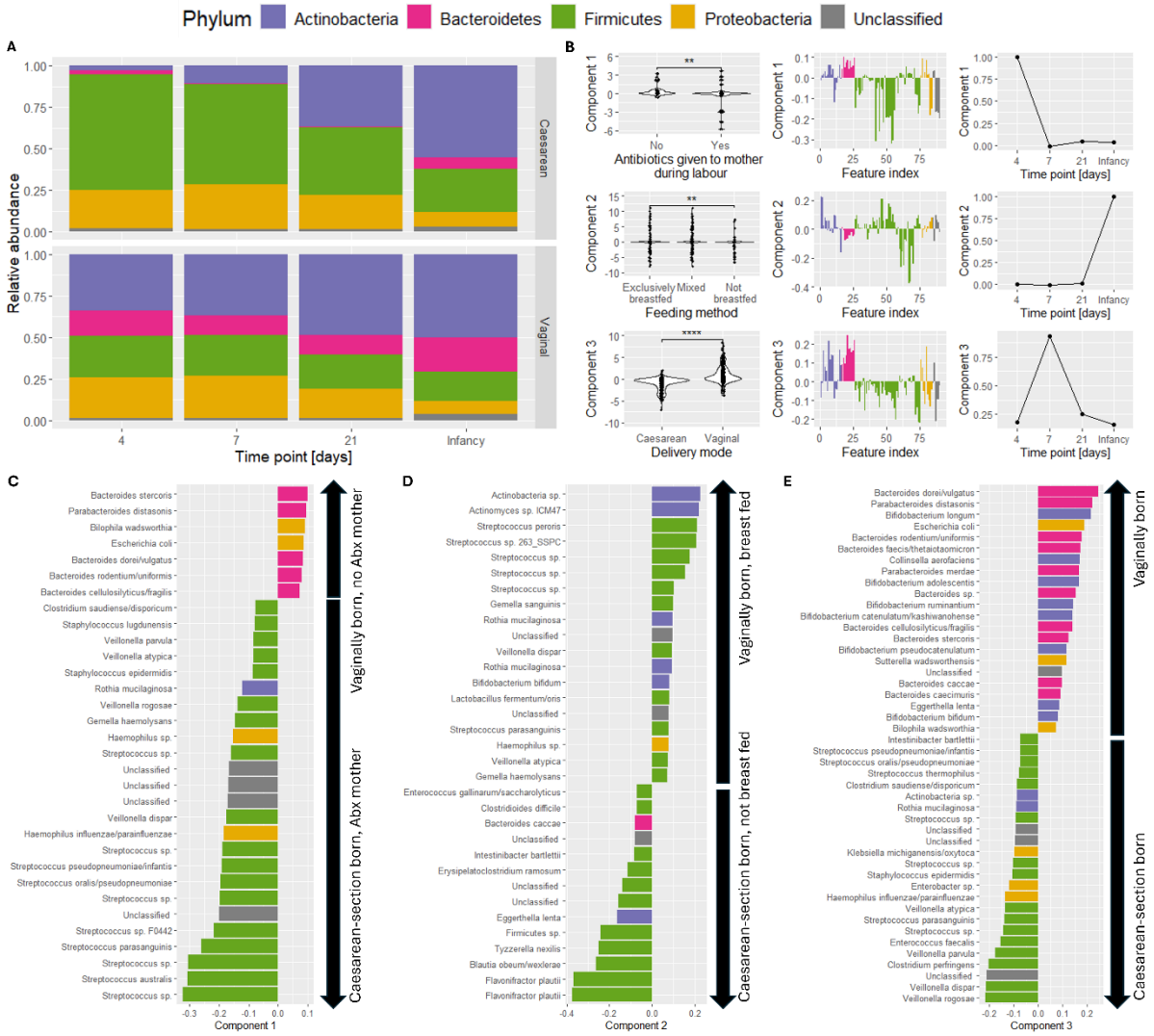


Figure 4: Overview of the Shao et al., 2019 dataset. (A) Overview of the average relative abundance per time point in vaginal and caesarean-section born infants. The birth mode groups are distinguishable between time points 4 and 21 postpartum and become similar at infancy. (B) The best three-component PARAFAC model of the processed data, explaining 9.40% of the variation. Components 1-3 all describe differences between the birth mode groups. Component 1, mainly present on day 4, additionally describes differences based on antibiotics being given to the mother during labour. Component 2, mainly present during infancy, additionally describes feeding mode. Component 3 describes only describes differences between the birth mode groups, which are present throughout the study but are largest on day 7. Wilcoxon rank-sum tests were used to compare the subject loadings against subject metadata. Full results of the tests are supplied in Supplementary Table 3. (C) Loadings of the microbiota in component 1. Positively loaded microbiota are associated with vaginally born infants and/or mothers having received antibiotics during labour. Negatively loaded microbiota are associated with caesarean-section born infants and/or mothers not having received antibiotics during labour. (D) Loadings of the microbiota in component 2. Positively loaded microbiota are associated with vaginally born and/or breast-fed infants. Negatively loaded microbiota are associated with caesarean-section born and/or not breast-fed infants. (E) Loadings of the microbiota in component 3. Positively loaded microbiota are associated with vaginally born infants. Negatively loaded microbiota are associated with caesarean-section born infants. Feature loadings in (C-E) are filtered to only show loadings with an absolute value higher than 0.075. For the unfiltered version of these plots, see Supplementary Figures 9-11.

All components of the selected model described differences in birth mode (p=2.4e-4, p=7.3e-9 and p=1.2e-46, respectively; Benjamini-Hochberg corrected Wilcoxon rank-sum test; **Supplementary Table 3**) and differences based on the *Bacteroides* profile as described by the original authors based on the median relative abundance of the genus (p=1.5e-4, p=7.3e-9, p=9.2e-59, respectively; [7]). In

460 addition, the first component described differences related to whether antibiotics were given to the
461 mother during labour (p=0.0059) and the second component described differences based on feeding
462 method (p=0.024).

463 Inspection of the positive loadings in the first component revealed that PARAFAC described the
464 genera that were enriched in vaginally born infants: *Bacteroides stercoris, Parabacteroides*
465 *distasonis, Bilophilia wadsworthia, Escherichia coli, Bacteroides dorei/vulgatus, Bacteroides*
466 *rodentium/uniformis and Bacteroides cellulosilyticus/fragilis* (**Figure 4C**). These genera were
467 depleted in caesarean-section born infants and infants whose mothers received antibiotics during
468 labour. The first component also described genera that were enriched in caesarean-section born
469 infants or infants whose mothers received antibiotics during labour: *Streptococcus australis,*
470 *Streptococcus parasanguinis, Streptococcus oralis,* amongst others. These genera were depleted in
471 vaginally born infants.

472 Inspection of the positive loadings in the second component revealed that PARAFAC described the
473 genera that were enriched in vaginally born or breast-fed infants: *Actinobacteria* sp., *Actinomyces*
474 sp. ICM47, and *Streptococcus peroris,* amongst others (**Figure 4D**). These genera were depleted in
475 caesarean-section born and formula-fed infants. It also described the genera that were enriched in
476 caesarean-section born or formula-fed infants: *Flavonifractor plautii, Blautia obeum/wexlerae* and
477 *Tyzzerella nexilis,* amongst others. These genera were depleted in vaginally born and breast-fed
478 infants. While not described in the original analysis, these results partially agree with other
479 literature [46–48].

480 Inspection of the positive loadings in the third component revealed that PARAFAC described the
481 genera that were enriched in vaginally born infants: *Bacteroides dorei/vulgatus, Parabacteroides*
482 *distasonis* and *Bifidobacterium longum,* amongst others (**Figure 4E**). These genera were depleted in
483 caesarean-section born infants. It also described the genera that were enriched in caesarean-section
484 born infants: *Veillonella rogosae, Veillonella dispar,* and *Clostridium perfringens,* amongst others.
485 These genera were depleted in vaginally born infants. This result is consistent with the original
486 analysis and other studies investigating the development of the infant gut microbiome in hospital
487 environments [48, 49].

488 Taken together, the PARAFAC model of this example dataset confirmed previously found
489 differences between the caesarean-section and vaginally born infants. It also found novel genera

490 that describe the differences between the birth mode groups and feeding mode groups. Additionally,
491 the moderate (62.5%) amount of missing data did not hamper the modelling procedure. In summary,
492 the PARAFAC model of this dataset confirms and enhances the results from the comparative
493 analysis done by the original authors.

494 *Post-hoc clustering of PARAFAC results can help identify microbial groups of interest*

495 The third example is a longitudinal oral microbiome gingivitis intervention study [8]. Here, we
496 analyse one oral microbiome niche, the supragingival plaque at the upper jaw lingual surfaces,
497 which was measured in 41 subjects at 7 time points: day -14 (baseline), 0 (baseline, start of
498 intervention), 2, 5, 9, 14 (end of intervention) and 21 (resolution). Using quantitative light-induced
499 fluorescence [34–36], the subjects were stratified into low, medium and high responders based on
500 the amount of red fluorescent dental plaque on day 14. With this example, we show how a post-hoc
501 clustering approach of a PARAFAC model can help identify groups of microbiota related to the
502 intervention.

503 We inspected PARAFAC modelling performance and stability for the processed count data using 1-
504 5 components. The CORCONDIA scores dropped below 80 for 5 components and the sum of squared
505 errors began to flatten between 2-3 components (**Supplementary Figure 13**). The three-component
506 models had very similar feature and time mode loadings between several components, making them
507 not suitable for describing the data (**Supplementary Figure 14**). This was confirmed by the jack-
508 knifing procedure (**Supplementary Figures 15-16**). Hence, we selected the model explaining the
509 most variation from all randomly initialized models of two components. The selected model
510 explained 19.21% of the variation (**Figure 5A**).

511

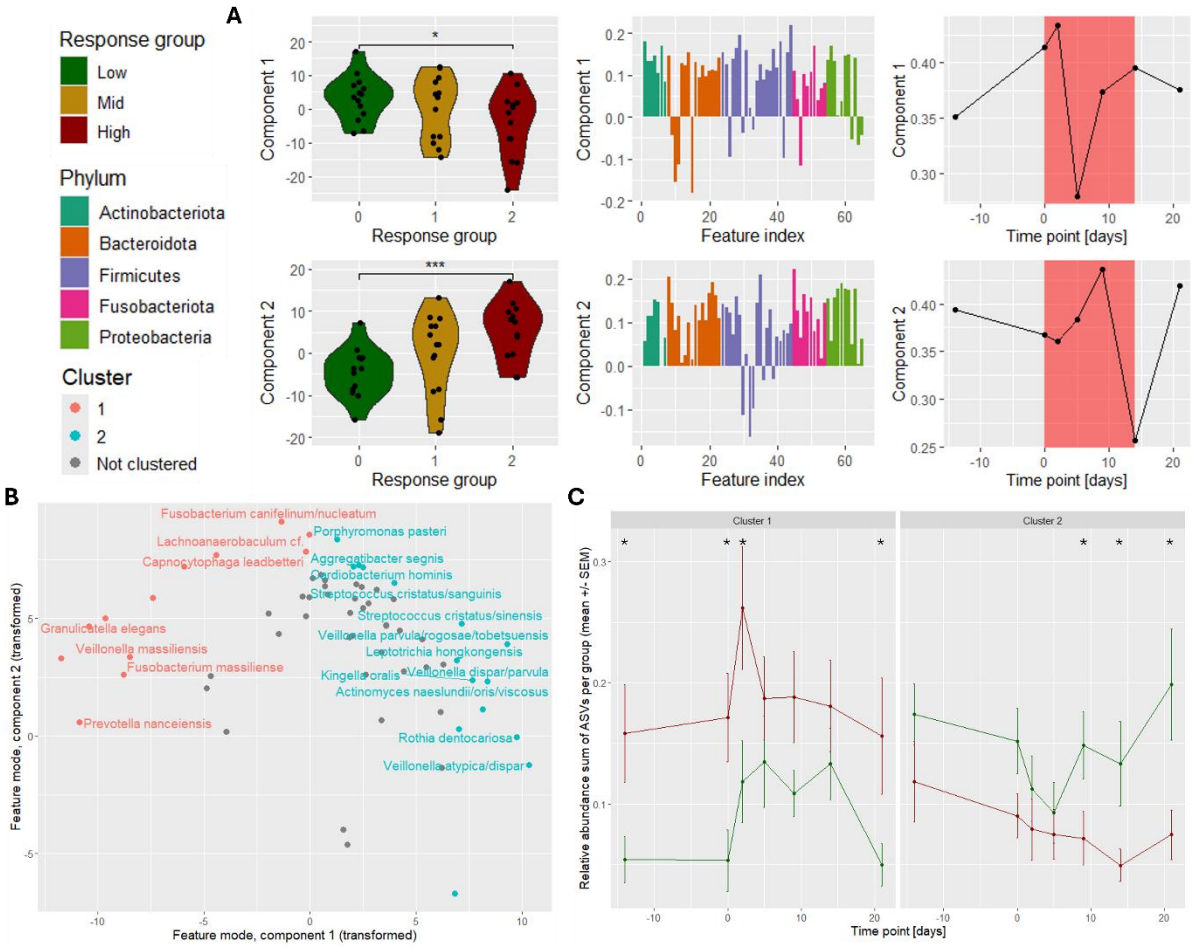


Figure 5: Overview of the van der Ploeg et al., 2024 dataset. (A) The best two-component PARAFAC model of the processed data, explaining 19.21% of the variation. Both components describe differences between the response groups that are relatively constant throughout the study. The time profile in component 1 describes that the differences between the response groups are larger on day 2 and much smaller on day 5. The time profile in component 2 describes that the differences between the response groups are larger on day 9 and much smaller on day 14. Wilcoxon rank-sum tests were used to compare the subject loadings against subject metadata. Full results of the tests are supplied in Supplementary Table 4. (B) Scatter plot showing the transformed feature loadings (Supplementary Methods) of PARAFAC component 1 versus component 2. Features were clustered into two groups after filtering. Only amplicon sequence variants (ASVs) with taxonomic annotations resolved to species level are labelled. (C) Overview of the relative abundance sum per ASV cluster in the low and high response groups. Error bars correspond to the standard error of the mean (SEM) across all samples per response group. The mean difference between the high and low response groups per time point was tested using a Benjamini-Hochberg corrected permutation test of 999 iterations (: p≤0.05; **: p≤0.01; ***: p≤0.001). Mid responders are not shown for visual clarity.*

To interpret the models, we tested the correlation of the subject loadings with the measured parameters of gingivitis. This approach revealed that both components of the PARAFAC model significantly described the amount of red fluorescent plaque over time (p=1.8e-5 and p=0.029, respectively; Benjamini-Hochberg corrected Pearson correlation test; **Supplementary Table 4**) and the percentage of sites that were covered by plaque (p=0.029 and p=0.029, respectively).

Following the original analysis [8], we identified microbial subcommunities with common responses by selecting and then clustering only well-modelled amplicon sequence variants (ASVs) based on their fitted response using a K-medoids algorithm (**Figure 5B**). The sum of the relative

533 abundances of the ASVs in each cluster was then determined to test the difference between the
534 microbiomes of individuals in the low and high response groups at every time point. This approach
535 revealed that the relative abundance sum of ASV cluster 1 was significantly higher in the high
536 responder individuals compared to the low responder individuals on days -14, 0, 2, and 21 (**Figure**
537 **5C, Supplementary Table 5**; $p \leq 0.05$; Benjamini-Hochberg corrected permutation test of mean
538 difference). Similarly, the relative abundance sum of ASV cluster 2 was found to be significantly
539 higher in low responder individuals compared to high responder individuals on days 9, 14 and 21.
540 Taken together, the PARAFAC models of plaque microbiomes described microbial subcommunities
541 with common dynamics that connect to the magnitude of the gingivitis response.

542 Discussion

543 Using three example studies [6–8], we show that the unsupervised modelling approach of Parallel
544 Factor Analysis (PARAFAC) can (1) identify the main time-resolved variation in longitudinal
545 microbiome data, (2) find differences between subject groups and enhance comparative analyses
546 and (3) helps identify feature groups of interest using a post-hoc clustering approach. In addition,
547 we show that PARAFAC is robust to missing measurements in the Shao et al., 2019 [7] example.
548 Robustness of the PARAFAC algorithm in the context of missing data was an expected result and is
549 well-documented [50, 51]. We also show in the Shao et al., 2019 [7] and van der Ploeg et al., 2024
550 [8] examples that the modelled variation is related to clinically relevant parameters. However, many
551 other multi-way approaches exist that are more flexible in their modelling procedure [31]. For
552 example, Tucker3 allows for interactions between the components [52]. PARAFAC2 allows for time
553 shifts that are unique per subject [19, 53]. Multi-way regression, in which the PARAFAC
554 components are supervised using a response variable, is also possible with N-way partial least squares
555 [54]. As highlighted in the Fujita et al. example study [6], these multi-way dimension reduction
556 methods require commonality between the subjects to be useful. While PARAFAC's treatment of
557 time as a categorical variable is useful to describe time profiles in small-scale microbiome studies,
558 large-scale microbiome studies should be analysed with models that use quantitative time factors,
559 such as General Additive Models [55] or (spline) Linear Mixed Effects Models [56, 57]. Further
560 research is needed to investigate the relevance of these methods in the context of longitudinal
561 microbiome data analysis.

562 In PARAFAC, the nature of the third mode in the multi-way array is not limited to time. For
563 example, a recent study has shown that soil treatment can be used as the third mode to identify

relationships between the abundances of mycorrhizal fungi and plant diversity and history [58]. Additionally, the principles shown in this work are generalizable to four or more modes if that better represents the data that was generated by the study design.

In this work we have not elaborated on the various options that exist for processing microbial abundance data. The sparsity filtering approach presented in this work does not consider microbiota disappearing as time progresses. The addition of a pseudo-count prior to a centred log-ratio (CLR) transformation is a complex and unresolved matter [59, 60]. It has been noted that this approach introduces a dependency with library size, depending on the number of zeroes in the data [59]. Other work [60] suggests that a sample-specific pseudo-count may resolve this issue, but further research is needed to examine this approach in the context of longitudinal microbiome data. Other options include setting a pseudo-count value of 0.5, 0.1 or replacing the zeroes with a random value between 0 and 1 drawn from a uniform distribution. We conducted a brief sensitivity analysis for these options (**Supplementary Figure S18**) and decided to use a pseudo-count of 1 as this treats all zeroes as relevant values while yielding similar models. In a related paper on tensor factorization of longitudinal microbiome data [61] the robust-centred log-ratio was suggested as an approach to transform the data and to correct for compositionality. This approach instead calculates the geometric mean by ignoring the zeroes and setting them to missing values after the transformation. This makes all zeroes available for imputation by the PARAFAC model, but ignores the different nature that zeroes can potentially have [27]. Further research is needed to investigate the appropriateness of the various processing methods for (longitudinal) microbiome count data in describing biological phenomena of interest.

Conclusions

We have shown that Parallel Factor Analysis is an applicable method for longitudinal microbiome data analysis across a wide range of microbial environments. Our results demonstrate that PARAFAC is able to (1) model the main time-resolved variation in longitudinal microbiome data, (2) find differences between subject groups and enhance comparative analyses despite a moderate amount of missing data, and (3) help identify feature groups of interest. The analyses and the example datasets with the resulting figures are implemented in the R package `parafac4microbiome`, which is available on CRAN at https://cran.rstudio.com/web/packages/parafac4microbiome/.

List of abbreviations

- ALS: Alternating Least-Squares
- ASV: Amplicon Sequence Variant
- CLR: Centred Log-Ratio
- CORCONDIA: Core Consistency Diagnostic
- mOTU: Marker-gene based Operational Taxonomic Unit
- PCA: Principal Component Analysis
- PARAFAC: Parallel Factor Analysis

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of the data and materials

The package used during the current study is available in the GitHub repository, https://doi.org/10.5281/zenodo.13862690. The datasets generated and/or analysed during the current study are available in the GitHub repository, https://doi.org/10.5281/zenodo.11072651. A STORMS (Strengthening The Organizing and Reporting of Microbiome Studies)[62] checklist is available at https://doi.org/10.5281/zenodo.11072651.

Competing interests

The authors declare that they have no competing interests.

Funding

GRvdP was funded by a grant from the University of Amsterdam, Research Priority Area on Personal Microbiome Health.

Authors' contributions

G.R. van der Ploeg: conceptualization, visualization, formal analysis, methodology, programming, writing – original draft; J.A. Westerhuis: conceptualization, methodology, supervision, formal analysis, writing – original draft; A. Heintz-Buschart: conceptualization, formal analysis,

supervision, project administration, writing – original draft; A.K. Smilde: conceptualization, methodology, supervision, funding acquisition, writing – review & editing.

Acknowledgements

We want to thank Jesse Alderliesten (UU/UvA) for their useful discussions and suggestions.

References

1. Lloyd-Price J, Arze C, Ananthakrishnan AN, Schirmer M, Avila-Pacheco J, Poon TW, et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature. 2019;569:655–62.

2. Pascal M, Perez-Gordo M, Caballero T, Escribese MM, Lopez Longo MN, Luengo O, et al. Microbiome and Allergic Diseases. Front Immunol. 2018;9.

3. Saus E, Iraola-Guzmán S, Willis JR, Brunet-Vega A, Gabaldón T. Microbiome and colorectal cancer: Roles in carcinogenesis and clinical potential. Mol Aspects Med. 2019;69:93–106.

4. Gerber GK. Longitudinal Microbiome Data Analysis. In: Metagenomics for Microbiology. Elsevier; 2015. p. 97–111.

5. Schmidt TSB, Raes J, Bork P. The Human Gut Microbiome: From Association to Modulation. Cell. 2018;172:1198–215.

6. Fujita H, Ushio M, Suzuki K, Abe MS, Yamamichi M, Iwayama K, et al. Alternative stable states, nonlinear behavior, and predictability of microbiome dynamics. Microbiome. 2023;11:63.

7. Shao Y, Forster SC, Tsaliki E, Vervier K, Strang A, Simpson N, et al. Stunted microbiota and opportunistic pathogen colonization in caesarean-section birth. Nature. 2019;574:117–21.

8. van der Ploeg GR, Brandt BW, Keijser BJF, van der Veen MH, Volgenant CMC, Zaura E, et al. Multi-way modelling of oral microbial dynamics and host-microbiome interactions during induced gingivitis. Npj Biofilms Microbiomes. 2024;10:1–12.

9. Diggle P. Analysis of longitudinal data. Oxford university press; 2002.

10. Fitzmaurice GM, Laird NM, Ware JH. Applied longitudinal analysis. John Wiley & Sons; 2012.

11. Kodikara S, Ellul S, Lê Cao K-A. Statistical challenges in longitudinal microbiome data analysis. Brief Bioinform. 2022;23:bbac273.

12. Bro R, Smilde AK. Principal component analysis. Anal Methods. 2014;6:2812–31.

13. Jolliffe I. Principal Component Analysis. In: Wiley StatsRef: Statistics Reference Online. John Wiley & Sons, Ltd; 2014.

14. Borg I, Groenen PJF. Modern Multidimensional Scaling: Theory and Applications. Springer Science & Business Media; 2007.

15. Carroll JD, Chang J-J. Analysis of individual differences in multidimensional scaling via an N-way generalization of “Eckart-Young” decomposition. Psychometrika. 1970;35:283–319.

654 16. Harshman RA. Foundations of the PARAFAC procedure: Models and conditions for an" explanatory"
655 multimodal factor analysis. 1970.

656 17. Tomasi G, Bro R. A comparison of algorithms for fitting the PARAFAC model. Comput Stat Data
657 Anal. 2006;50:1700–34.

658 18. Yates F. The analysis of replicated experiments when the field results are incomplete. Emp J Exp
659 Agric. 1933;1:129–42.

660 19. Bro R. Multi-way analysis in the food industry. Models Algorithms Appl Acad Proefschr Dinam.
661 1998.

662 20. Helwig NE, Helwig MNE. Package 'multiway.' 2019.

663 21. Kruskal JB. More factors than subjects, tests and treatments: An indeterminacy theorem for
664 canonical decomposition and individual differences scaling. Psychometrika. 1976;41:281–93.

665 22. Kruskal JB. Three-way arrays: rank and uniqueness of trilinear decompositions, with application to
666 arithmetic complexity and statistics. Linear Algebra Its Appl. 1977;18:95–138.

667 23. ten Berge JMF, Sidiropoulos ND. On uniqueness in candecomp/parafac. Psychometrika.
668 2002;67:399–409.

669 24. Kruskal JB. Rank, decomposition, and uniqueness for 3-way and N-way arrays. In: Multiway data
670 analysis. 1989. p. 7–18.

671 25. Aitchison J. The Statistical Analysis of Compositional Data. J R Stat Soc Ser B Methodol.
672 1982;44:139–60.

673 26. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome Datasets Are Compositional:
674 And This Is Not Optional. Front Microbiol. 2017;8:2224.

675 27. Silverman JD, Roche K, Mukherjee S, David LA. Naught all zeros in sequence count data are the
676 same. Comput Struct Biotechnol J. 2020;18:2789–98.

677 28. Paulson JN, Stine OC, Bravo HC, Pop M. Differential abundance analysis for microbial marker-gene
678 surveys. Nat Methods. 2013;10:1200–2.

679 29. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic biomarker
680 discovery and explanation. Genome Biol. 2011;12:R60.

681 30. Bro R, Smilde AK. Centering and scaling in component analysis. J Chemom. 2003;17:16–33.

682 31. Smilde AK, Bro R, Geladi P. Multi-way analysis: applications in the chemical sciences. John Wiley &
683 Sons; 2005.

684 32. R Core Team R. R: A language and environment for statistical computing. 2013.

685 33. Milanese A, Mende DR, Paoli L, Salazar G, Ruscheweyh H-J, Cuenca M, et al. Microbial abundance,
686 activity and population genomic profiling with mOTUs2. Nat Commun. 2019;10:1014.

34. Heinrich-Weltzien R, Kühnisch J, Van Der Veen M, De Josselin De Jong E, Stößer L. Quantitative light-induced fluorescence (QLF) - A potential method for the dental practitioner. Quintessence Int. 2003;34:181–8.

35. van der Veen MH, Volgenant CMC, Keijser B, ten Cate J (Bob) M, Crielaard W. Dynamics of red fluorescent dental plaque during experimental gingivitis—A cohort study. J Dent. 2016;48:71–6.

36. Volgenant CMC, Zaura E, Brandt BW, Buijs MJ, Tellez M, Malik G, et al. Red fluorescence of dental plaque in children —A cross-sectional study. J Dent. 2017;58:40–7.

37. Bro R. PARAFAC. Tutorial and applications. Chemom Intell Lab Syst. 1997;:23.

38. Cattell RB. The Scree Test For The Number Of Factors. Multivar Behav Res. 1966;1:245–76.

39. Horn JL, Engstrom R. Cattell's Scree Test In Relation To Bartlett's Chi-Square Test And Other Observations On The Number Of Factors Problem. Multivar Behav Res. 1979;14:283–300.

40. Wang K, Yu J, Bi Y, Li Y, Huang P, Hou D, et al. A novel bound constrained optimization method for three-way chemical data analysis with application to fluorescence excitation-emission matrix spectroscopy. Chemom Intell Lab Syst. 2020;203:104036.

41. Bro R, Kiers H. A new Efficient Method for Determining the Number of Components in PARAFAC Models. J Chemom. 2003;17:274–86.

42. Lorenzo-Seva U, ten Berge JMF. Tucker's congruence coefficient as a meaningful index of factor similarity. Methodol Eur J Res Methods Behav Soc Sci. 2006;2:57–64.

43. Tucker LR. A method for synthesis of factor analysis studies. Educational Testing Service Princeton, NJ; 1951.

44. Riu J, Bro R. Jack-knife technique for outlier detection and estimation of standard errors in PARAFAC models. Chemom Intell Lab Syst. 2003;65:35–49.

45. Lorho G, Westad F, Bro R. Generalized correlation loadings: Extending correlation loadings to congruence and to multi-way models. Chemom Intell Lab Syst. 2006;84:119–25.

46. Maechler M. Finding groups in data: Cluster analysis extended Rousseeuw et al. R Package Version. 2019;2:242–8.

47. Kassambara A, Mundt F. Package 'factoextra.' Extr Vis Results Multivar Data Anal. 2017;76.

48. Raveh-Sadka T, Firek B, Sharon I, Baker R, Brown CT, Thomas BC, et al. Evidence for persistent and shared bacterial strains against a background of largely unique gut colonization in hospitalized premature infants. ISME J. 2016;10:2817–30.

49. Stewart CJ, Skeath T, Nelson A, Fernstad SJ, Marrs EC, Perry JD, et al. Preterm gut microbiota and metabolome following discharge from intensive care. Sci Rep. 2015;5:17141.

50. Tomasi G, Bro R. PARAFAC and missing values. Chemom Intell Lab Syst. 2005;75:163–80.

51. Acar E, Dunlavy DM, Kolda TG, Mørup M. Scalable tensor factorizations for incomplete data. Chemom Intell Lab Syst. 2011;106:41–56.

722 52. Kroonenberg PM, De Leeuw J. Principal component analysis of three-mode data by means of
723 alternating least squares algorithms. Psychometrika. 1980;45:69–97.

724 53. Kiers HAL, Ten Berge JMF, Bro R. PARAFAC2—Part I. A direct fitting algorithm for the PARAFAC2
725 model. J Chemom. 1999;13:275–94.

726 54. Bro R. Multiway calibration. Multilinear PLS. J Chemom. 1996;10:47–61.

727 55. Wood SN. Generalized Additive Models: An Introduction with R, Second Edition. 2nd edition. New
728 York: Chapman and Hall/CRC; 2017.

729 56. Pinheiro J, Bates D. Mixed-effects models in S and S-PLUS. Springer science & business media; 2000.

730 57. Molenberghs G. Models for Discrete Longitudinal Data. 1st ed. New York, NY: Springer New York;
731 2006.

732 58. Albracht C, Solbach MD, Hennecke J, Bassi L, Van Der Ploeg GR, Eisenhauer N, et al. Common soil
733 history is more important than plant history for arbuscular mycorrhizal community assembly in an
734 experimental grassland diversity gradient. preprint. Microbiology; 2024.

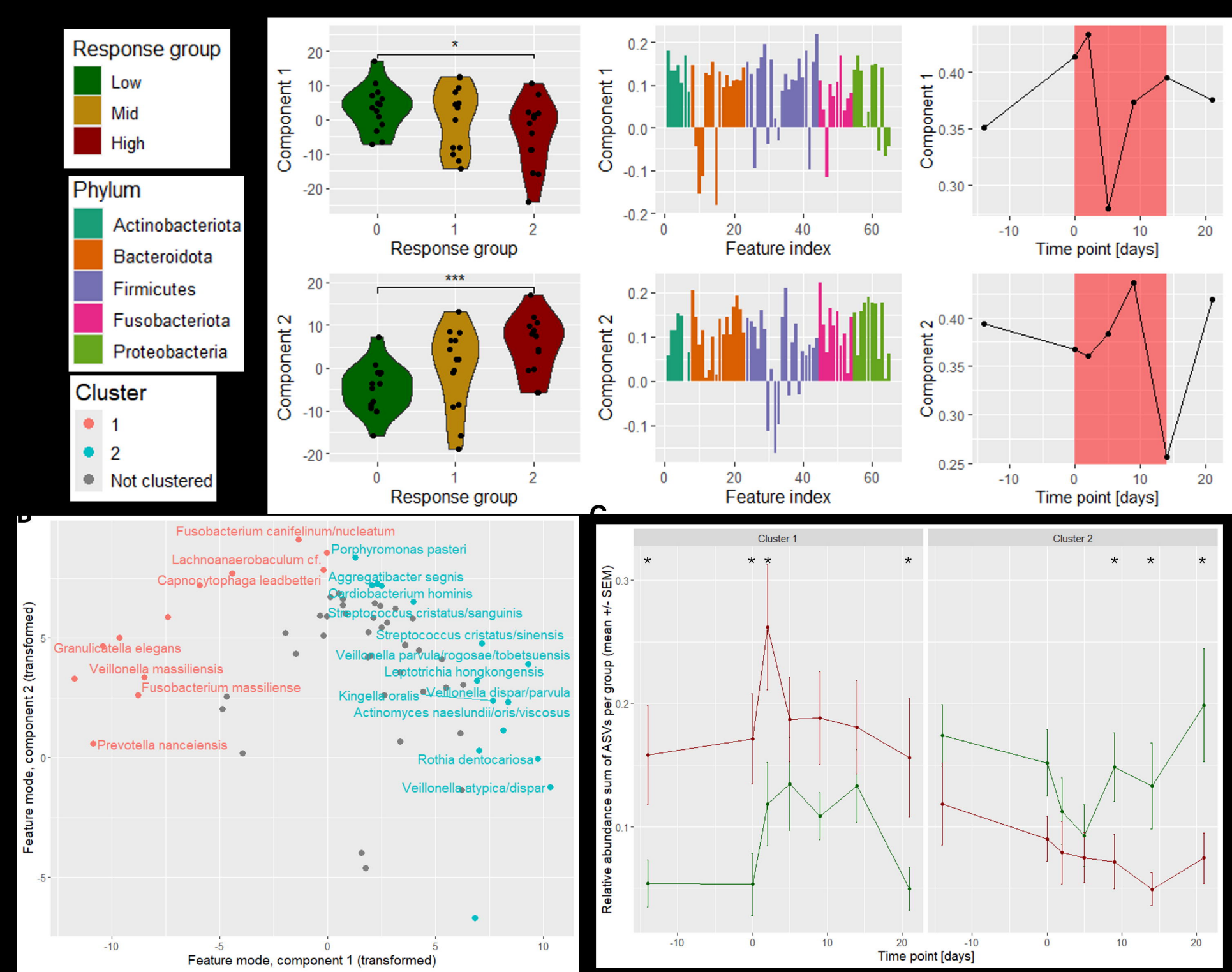
735 59. Te Beest DE, Nijhuis EH, Möhlmann TWR, Ter Braak CJF. Log-ratio analysis of microbiome data with
736 many zeroes is library size dependent. Mol Ecol Resour. 2021;21:1866–74.

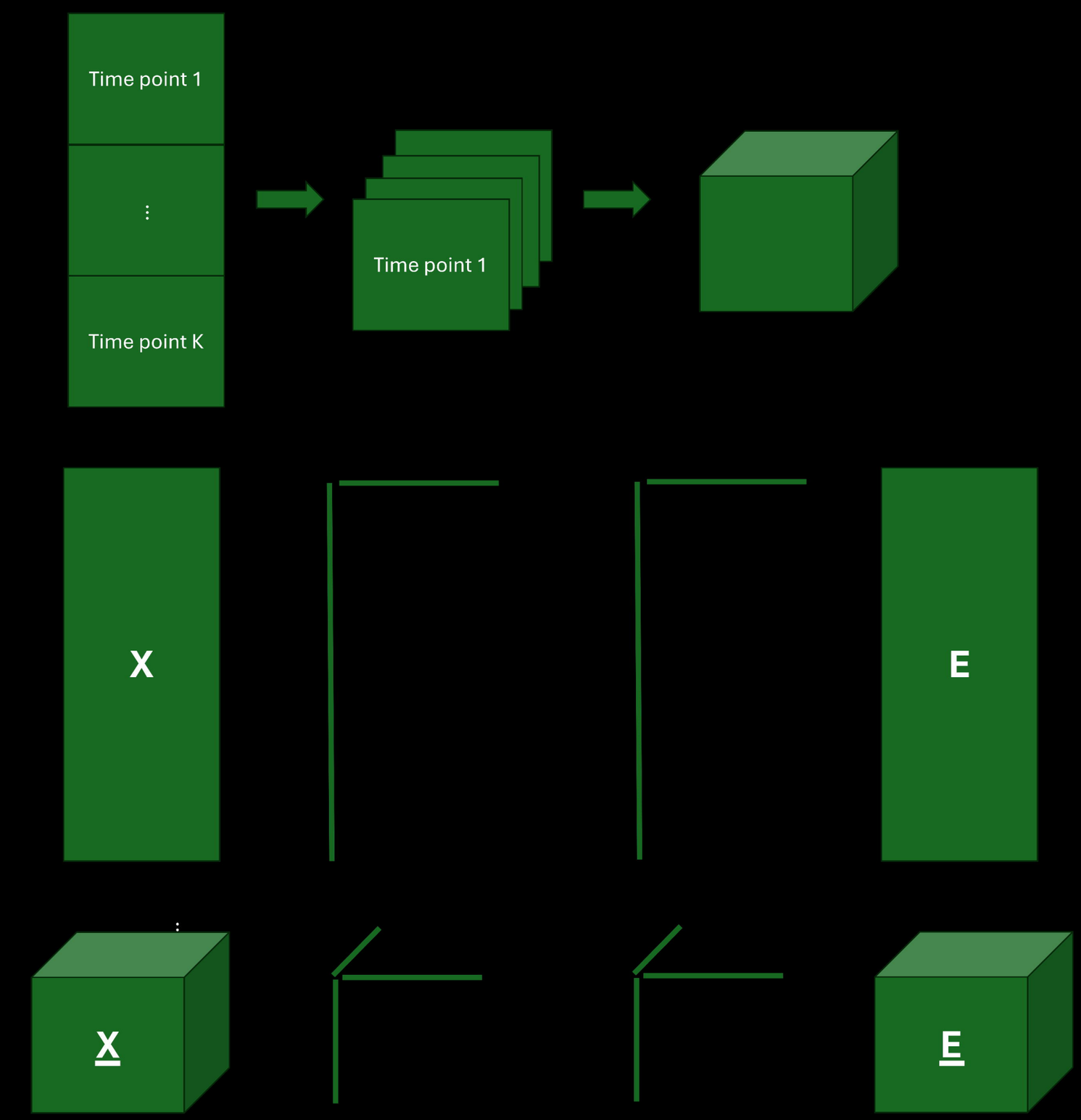
737 60. Erhard F. Estimating pseudocounts and fold changes for digital expression measurements.
738 Bioinformatics. 2018;34:4054–63.

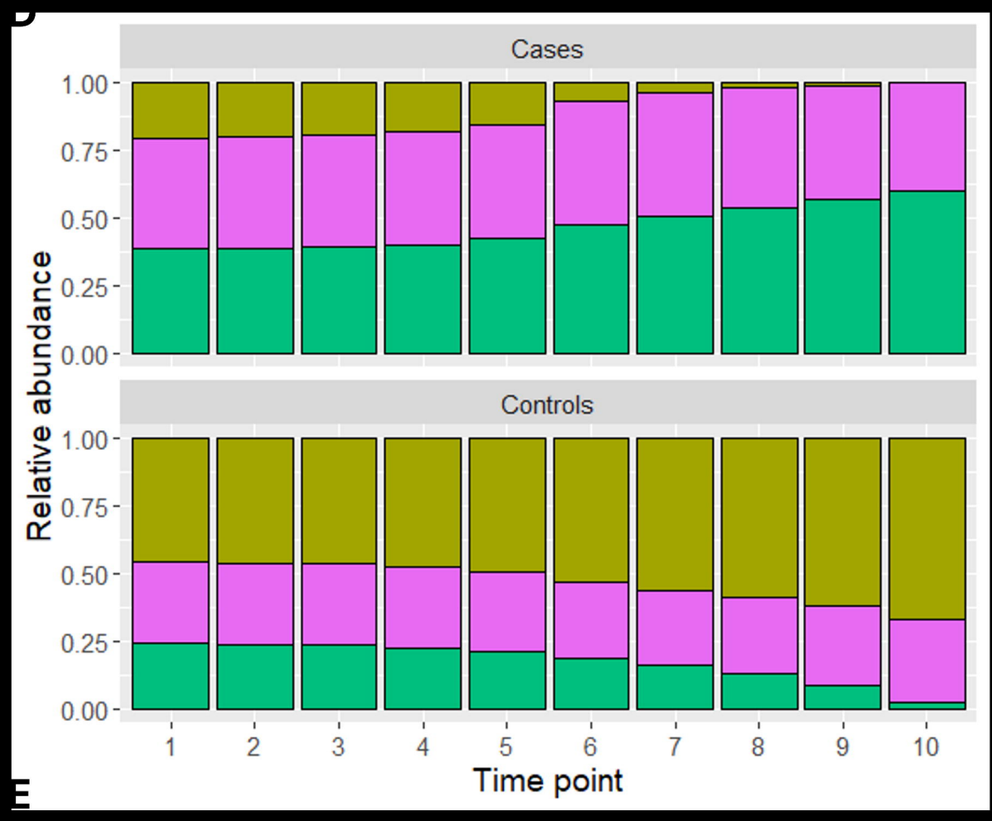
739 61. Martino C, Morton JT, Marotz CA, Thompson LR, Tripathi A, Knight R, et al. A Novel Sparse
740 Compositional Technique Reveals Microbial Perturbations. mSystems. 2019;4:e00016-19.

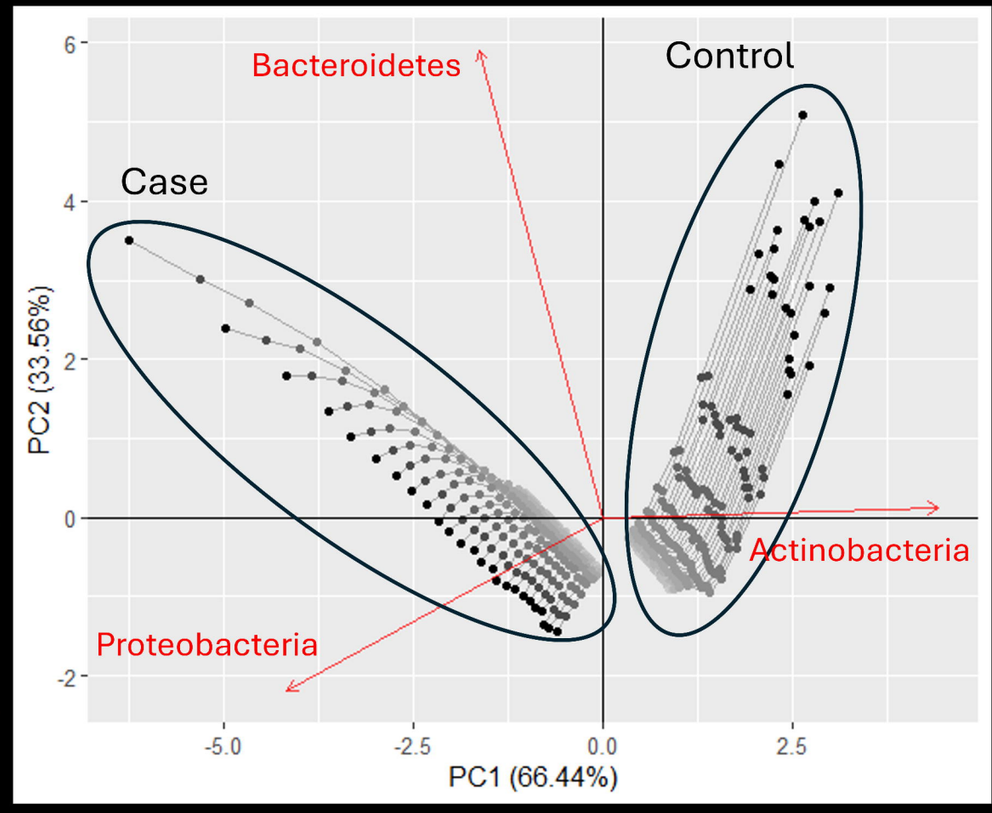
741 62. Mirzayi C, Renson A, Zohra F, Elsafoury S, Geistlinger L, Kasselman LJ, et al. Reporting guidelines for
742 human microbiome research: the STORMS checklist. Nat Med. 2021;27:1885–92.

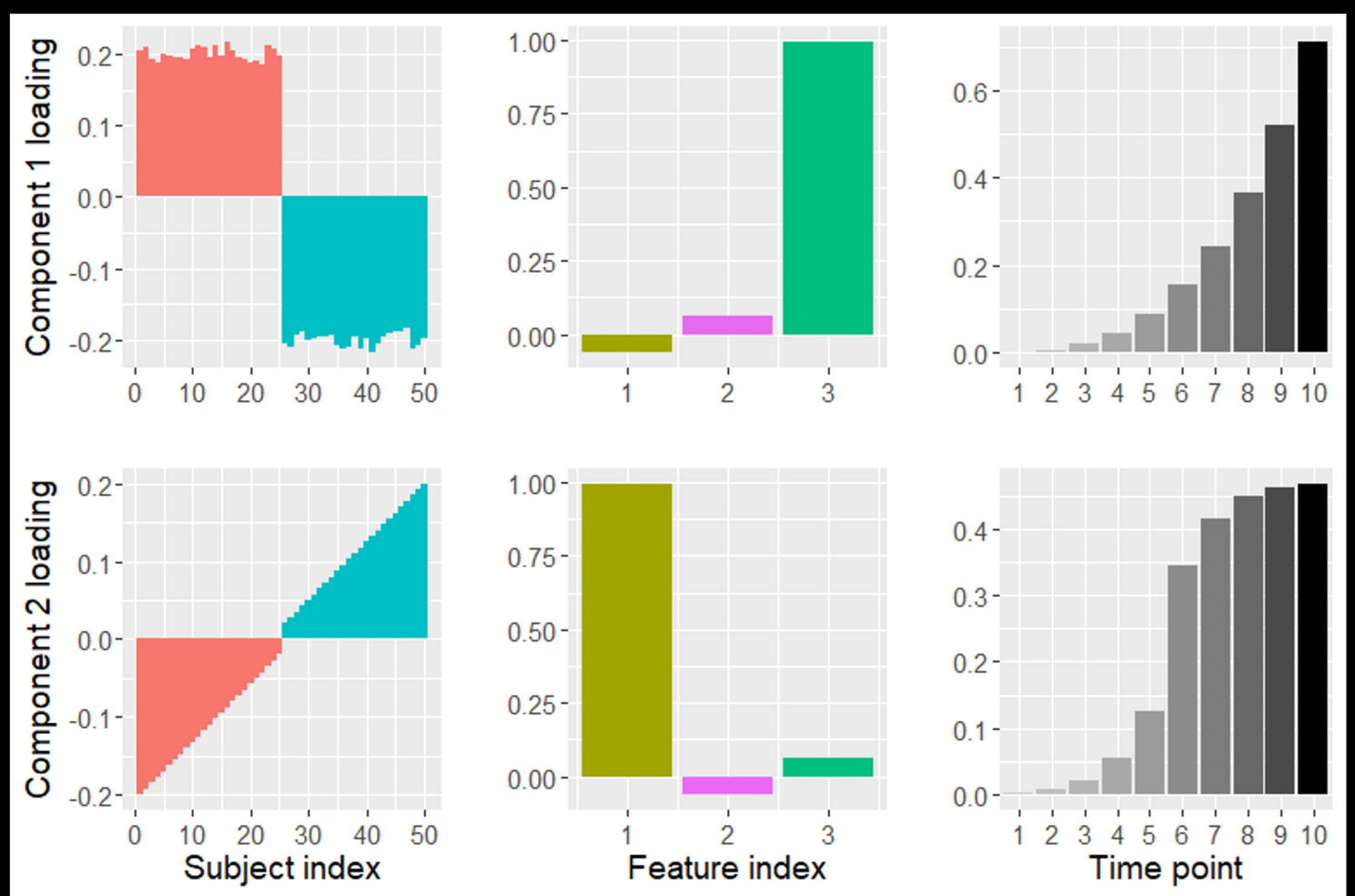
743


Response group
Low
Mid
High
Phylum
Actinobacteriota
Bacteroidota
Firmicutes
Fusobacteriota
Proteobacteria
Cluster
1
2
Not clustered
Component 1
Component 2
Response group
Feature index
Time point [days]
*
***
B
Fusobacterium canifelinum/nucleatum
Porphyromonas pasteri
Lachnoanaerobaculum cf.
Capnocytophaga leadbetteri
Aggregatibacter segnis
Cardiobacterium hominis
Streptococcus cristatus/sanguinis
Streptococcus cristatus/sinensis
Granulicatella elegans
Veillonella parvula/rogosae/tobetsuensis
Veillonella massiliensis
Leptotrichia hongkongensis
Fusobacterium massiliense
Kingella oralis
Veillonella dispar/parvula
Actinomyces naeslundii/oris/viscosus
Prevotella nanceiensis
Rothia dentocariosa
Veillonella atypica/dispar
Feature mode, component 1 (transformed)
Feature mode, component 2 (transformed)
C
Cluster 1
Cluster 2
Relative abundance sum of ASVs per group (mean +/- SEM)
Time point [days]

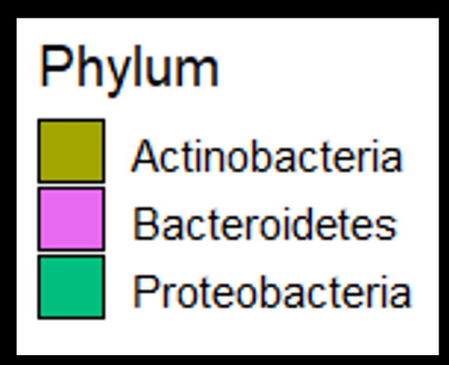

Time point 1
⋮
Time point K
Time point 1
X
E
X
E

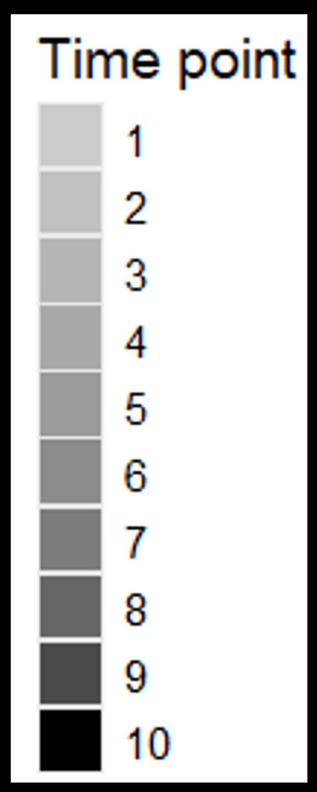

Cases
Controls
Relative abundance
Time point

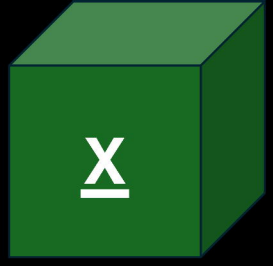

Bacteroidetes
Control
Case
Actinobacteria
Proteobacteria
PC2 (33.56%)
PC1 (66.44%)

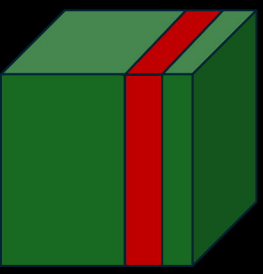

Component 1 loading
Component 2 loading
Subject index
Feature index
Time point

Subject group
Case
Control


Phylum
Actinobacteria
Bacteroidetes
Proteobacteria

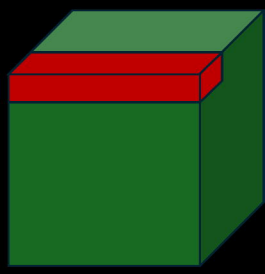

Time point
1
2
3
4
5
6
7
8
9
10



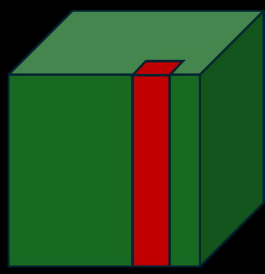


Filtering based on sparsity across all measurements

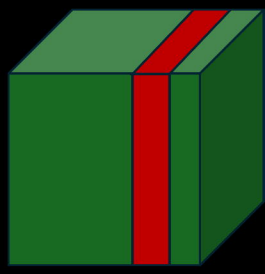
Centred log-ratio transformation per measurement

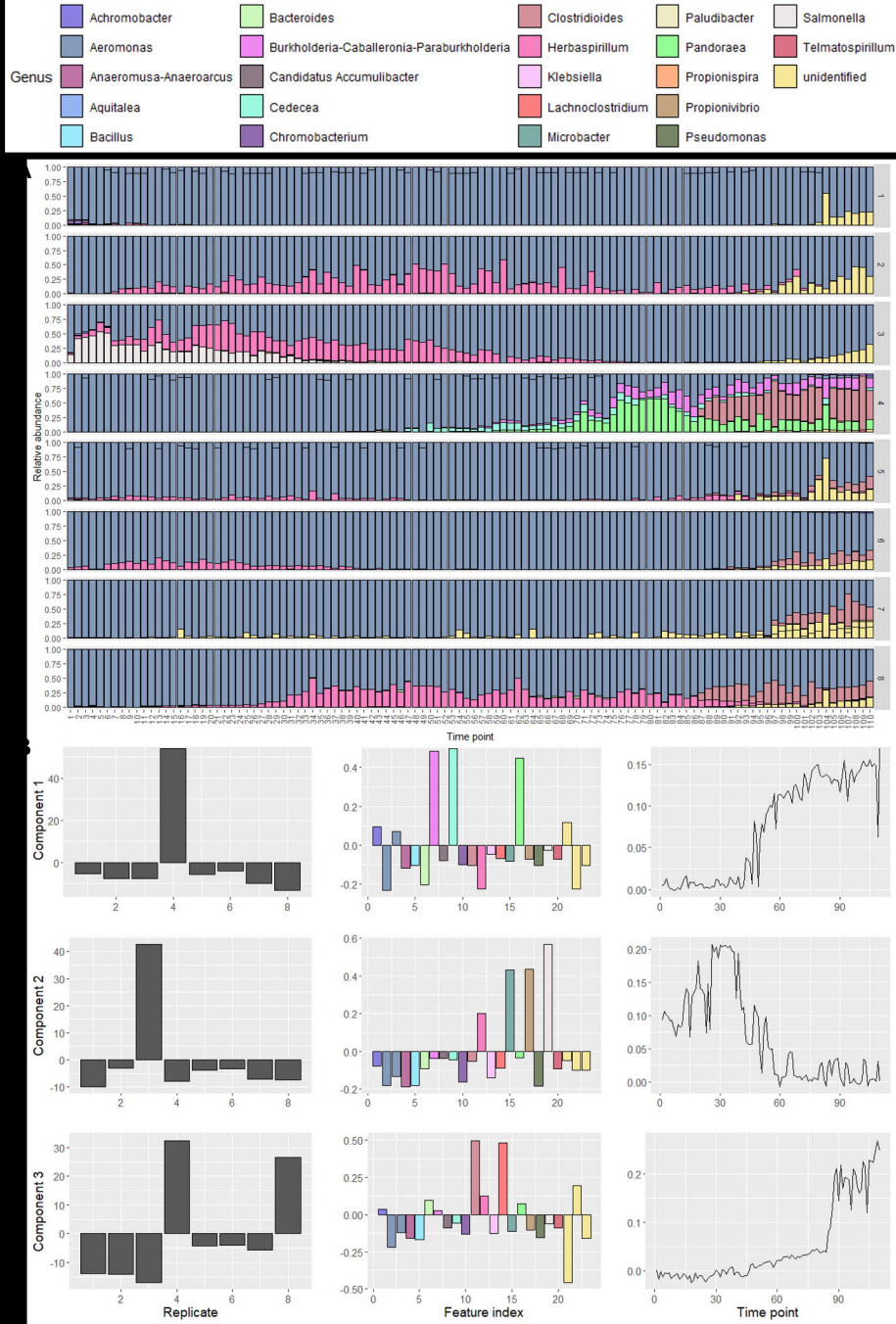


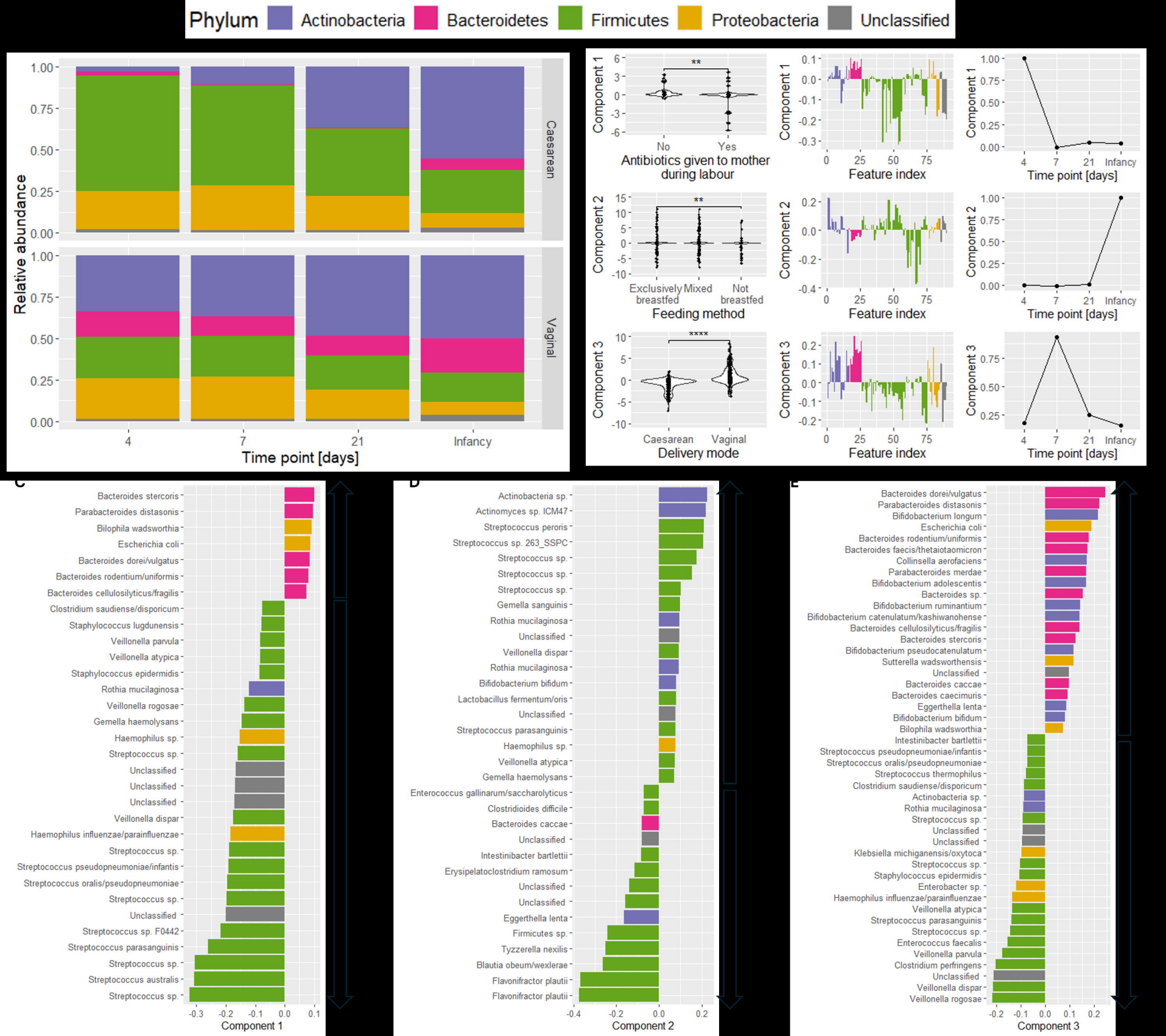
Centring across the subject mode



Scaling within the feature mode




Genus
Achromobacter
Aeromonas
Anaeromusa-Anaeroarcus
Aquitalea
Bacillus
Bacteroides
Burkholderia-Caballeronia-Paraburkholderia
Candidatus Accumulibacter
Cedecea
Chromobacterium
Clostridioides
Herbaspirillum
Klebsiella
Lachnoclostridium
Microbacter
Paludibacter
Pandoraea
Propionispira
Propionivibrio
Pseudomonas
Salmonella
Telmatospirillum
unidentified
Relative abundance
Time point
Component 1
Component 2
Component 3
Replicate
Feature index
Time point


Phylum
Actinobacteria
Bacteroidetes
Firmicutes
Proteobacteria
Unclassified
Relative abundance
Caesarean
Vaginal
Time point [days]
Component 1
Component 2
Component 3
Antibiotics given to mother during labour
No
Yes
Feeding method
Exclusively breastfed
Mixed
Not breastfed
Delivery mode
Caesarean
Vaginal
Feature index
Infancy
C
Bacteroides stercoris
Parabacteroides distasonis
Bilophila wadsworthia
Escherichia coli
Bacteroides dorei/vulgatus
Bacteroides rodentium/uniformis
Bacteroides cellulosilyticus/fragilis
Clostridium saudiense/disporicum
Staphylococcus lugdunensis
Veillonella parvula
Veillonella atypica
Staphylococcus epidermidis
Rothia mucilaginosa
Veillonella rogosae
Gemella haemolysans
Haemophilus sp.
Streptococcus sp.
Unclassified
Veillonella dispar
Haemophilus influenzae/parainfluenzae
Streptococcus pseudopneumoniae/infantis
Streptococcus oralis/pseudopneumoniae
Streptococcus sp. F0442
Streptococcus parasanguinis
Streptococcus australis
D
Actinobacteria sp.
Actinomyces sp. ICM47
Streptococcus peroris
Streptococcus sp. 263_SSPC
Gemella sanguinis
Bifidobacterium bifidum
Lactobacillus fermentum/oris
Streptococcus parasanguinis
Enterococcus gallinarum/saccharolyticus
Clostridioides difficile
Bacteroides caccae
Intestinibacter bartlettii
Erysipelatoclostridium ramosum
Eggerthella lenta
Firmicutes sp.
Tyzzerella nexilis
Blautia obeum/wexlerae
Flavonifractor plautii
E
Bacteroides caccae
Bacteroides caecimuris
Eggerthella lenta
Intestinibacter bartlettii
Streptococcus thermophilus
Klebsiella michiganensis/oxytoca
Enterobacter sp.
Enterococcus faecalis
Veillonella parvula
Clostridium perfringens
Veillonella dispar
Veillonella rogosae