



**Center for Bioinformatics
and Quantitative Biology**

CBQB & Cancer Center Journal Club

Suzi Alvernaz, G1
01/24/2022

CBQB Journal Club



Outline

- Background on Tensor Factorization
- Issues with analyzing repeated microbial measure
- Paper Aims
- Methods employed: normalization & factorization
- CTF Performance & Validation
- Limitations & Conclusions

Matrix rank decomposition & the rotation problem

$$M = AB^T \quad \text{with} \quad M \in \mathbb{R}^{n \times m}, \quad A \in \mathbb{R}^{n \times r}, \quad B^T \in \mathbb{R}^{r \times m} \quad (1)$$

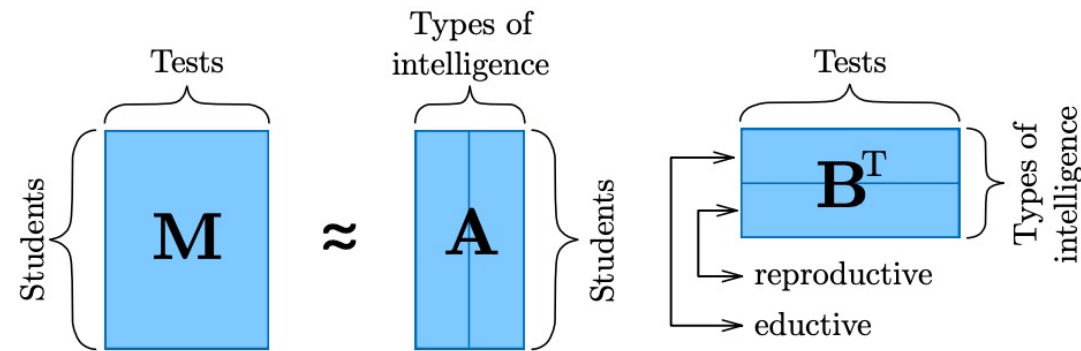


Figure 1: Spearman's hypothesis

Rotation Problem – Non-unique decompositions

Tensor Factorization

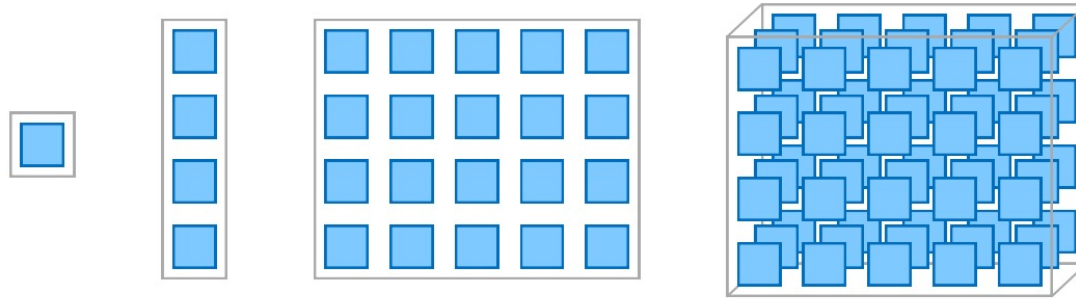


Figure 3: $x \in \mathbb{R}$, $\mathbf{x} \in \mathbb{R}^4$, $\mathbf{X} \in \mathbb{R}^{4 \times 5}$, $\mathcal{X} \in \mathbb{R}^{4 \times 5 \times 3}$

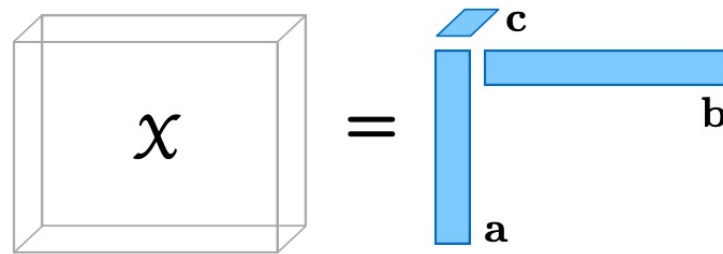


Figure 6: A rank-1 mode-3 tensor

Tensor Factorization

Tensor
Multilinear
form

$$T(M_1, M_2, M_3) \in \mathbb{R}^{d_1} \otimes \mathbb{R}^{d_2} \otimes \mathbb{R}^{d_3}$$
$$T(M_1, M_2, M_3)_{i_1, i_2, i_3} := \sum_{j_1, j_2, j_3 \in [d]} T_{j_1, j_2, j_3} \cdot M_1(j_1, i_1) \cdot M_2(j_2, i_2) \cdot M_3(j_3, i_3) \quad (1)$$

CP
Decomposition

$$T = \sum_{i=1}^r w_i (a_i \otimes b_i \otimes c_i), \quad w_i \in \mathbb{R}, a_i, b_i, c_i \in \mathbb{R}^d \quad (3)$$

Tensor Factorization Application to biological data

Predicting Drug-miRNA Resistance with Layer Attention Graph Convolution Network and Multi Channel Feature Extraction

Year: 2021, Pages: 1083-1089

DOI Bookmark: [10.1109/BIBM52615.2021.9669497](https://doi.org/10.1109/BIBM52615.2021.9669497)

Temporal Phenotyping for Characterisation of Hospital Care Pathways of COVID19 Patients

Mathieu Chambard¹, Thomas Guyet²(✉) , Yê-Lan NGuyen³,
and Etienne Audureau⁴ 

¹ ENS Rennes/IRISA, Rennes, France

² Inria – Centre Grenoble Rhône Alpes, Lyon, France

thomas.guyet@inria.fr

³ AP-HP, Hôpital Cochin, Sorbonne Université, INSERM UMR S 1138, Pierre Louis Institute of Epidemiology and Public Health, Paris, France

⁴ AP-HP, Henri Mondor Hospital, University Paris Est Créteil, Créteil, France

2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)

Graph Attention Mechanism-based Deep Tensor Factorization for Predicting disease-associated miRNA-miRNA pairs

Year: 2021, Pages: 189-196

DOI Bookmark: [10.1109/BIBM52615.2021.9669673](https://doi.org/10.1109/BIBM52615.2021.9669673)

Authors

Deep Tensor Factorization-Based Time Series Remote Sensing Data Imputation for Missing Observations

Fellow, IEEE, Dong Liang, Jianyi Peng, and Chengjun Li

Abstract—Land surface temperature (LST) is widely used in the field of time series remote sensing. However, due to the influence of cloud cover, the large area of LST data observation is missing, which seriously affects the later data analysis. In the past research, various effective interpolation methods have been developed, but they usually cannot effectively interpolate

applications. The National Aeronautics and Space Administration (NASA) provides global MODIS data with high spatial and temporal resolution [6] and these data have been widely used worldwide. However, when cloud pollution factors cause low-quality or missing data on land satellites, which is a very



Context-aware dimensionality reduction deconvolutes gut microbial community dynamics

Cameron Martino^{1,2,3,16}, Liat Shenhav^{4,16}, Clarisse A. Marotz³, George Armstrong ^{2,3}, Daniel McDonald³,
Yoshiki Vázquez-Baeza^{1,5}, James T. Morton ⁶, Lingjing Jiang⁷, Maria Gloria Dominguez-Bello ^{8,9},
Austin D. Swafford¹, Eran Halperin ^{4,10,11,12,13} and Rob Knight ^{1,3,14,15} 

Outline

- Background on Tensor Factorization
- Issues with analyzing repeated microbial measure
- Paper Aims
- Methods employed: normalization & factorization
- CTF Performance & Validation
- Limitations & Conclusions

Compositionality of Microbial Data

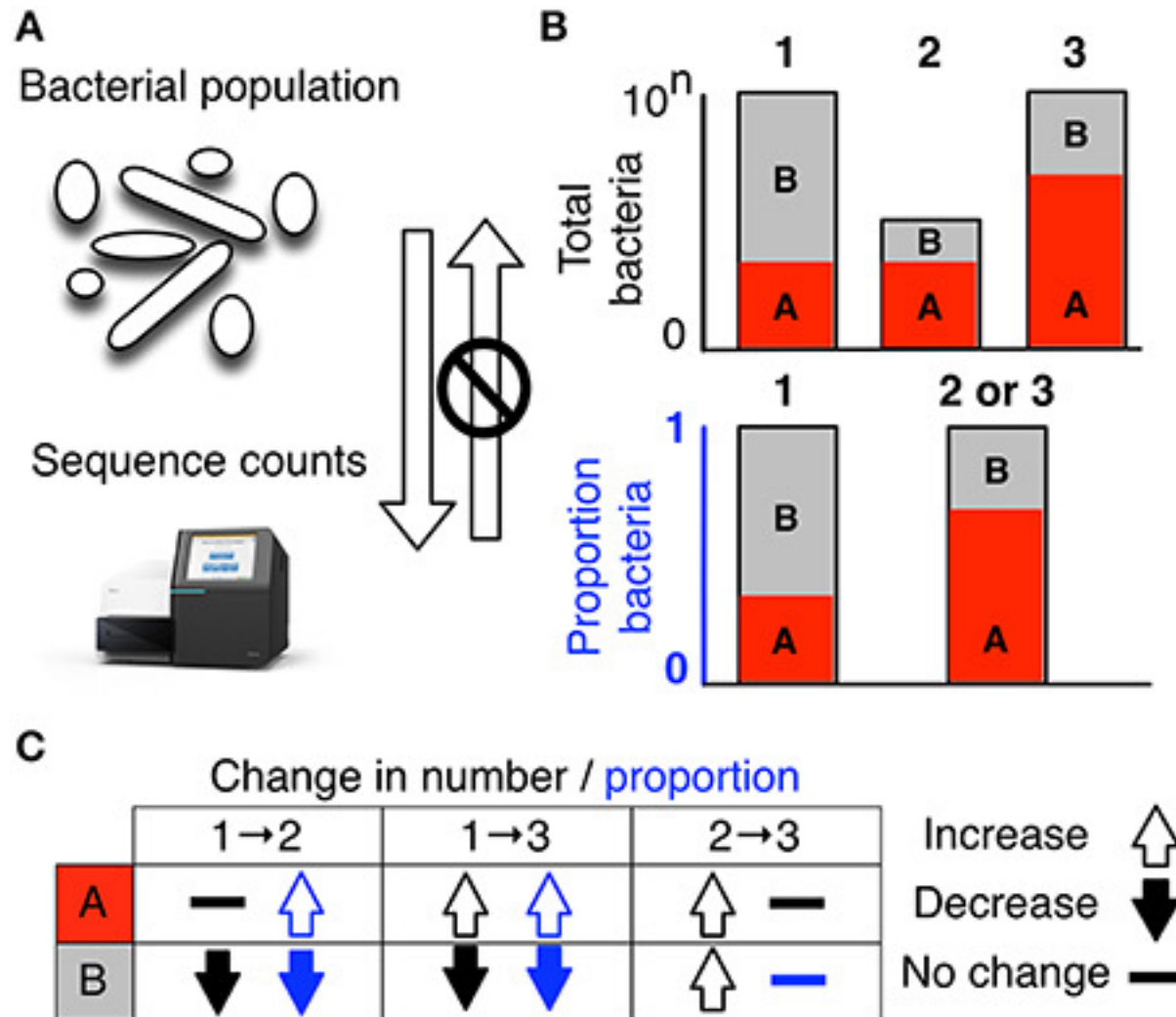
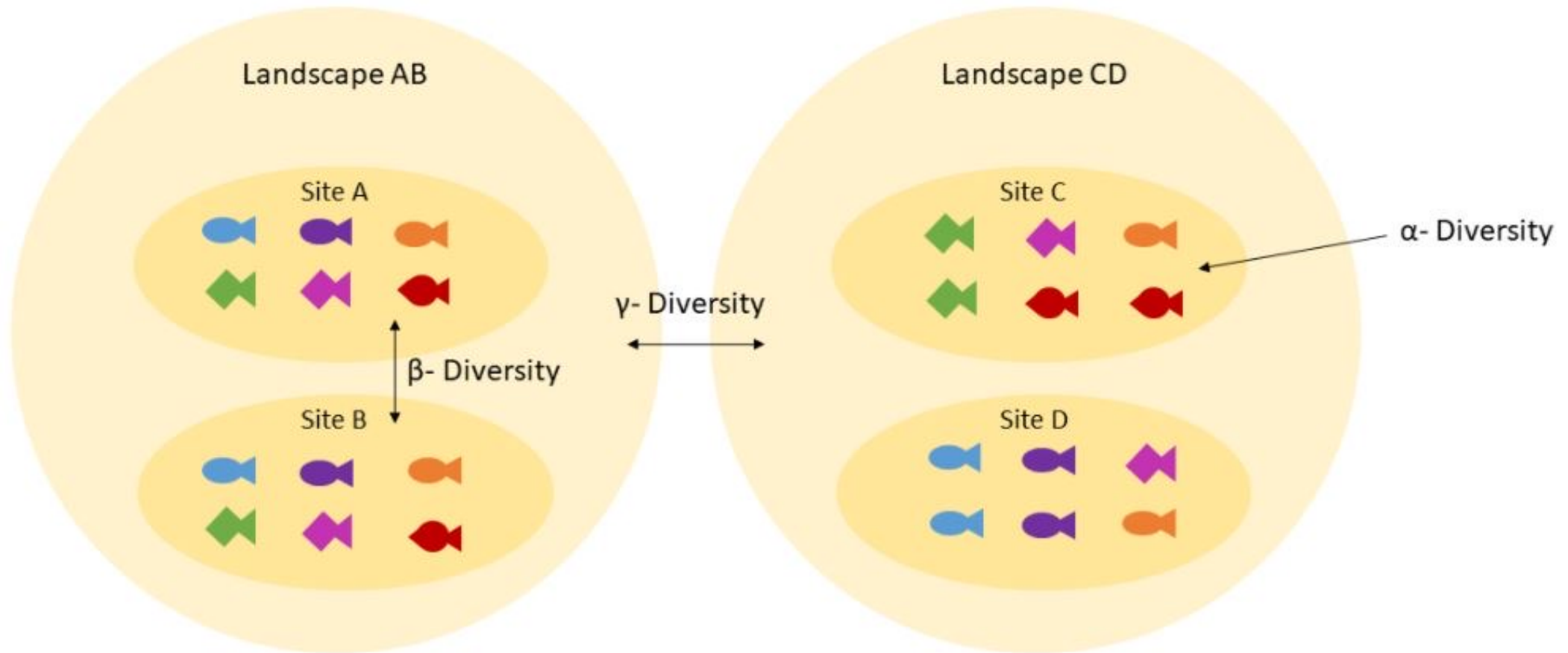


Figure 1 (Gloor, et al., 2017)

Beta diversity measures taxa differences between communities



Issues analyzing repeated microbial measure

- Modeling repeated measures
 - Longitudinal data, multiple body sites etc
- Methods for Microbiome community analysis
 - PCoA – Relationship between sample communities
 - Cannot account for temporal/spatial data
 - Supervised methods
 - Accounts for repeated measure but difficulties clustering phenotypes by microbe communities

Outline

- Background on Tensor Factorization
- Issues with analyzing repeated microbial measure
- **Paper Aims**
- Methods employed: normalization & factorization
- CTF Performance & Validation
- Limitations & Conclusions

Study Aims

- Control for interindividual variation in biological data by combining data using repeated measures
- Account for compositional nature of microbial communities using an unsupervised dimension reduction method
- Generation of diversity and differential abundance analysis based on this deconvolution

Outline

- Background on Tensor Factorization
- Issues with analyzing repeated microbial measure
- Paper Aims
- **Methods employed: normalization & factorization**
- CTF Performance & Validation
- Limitations & Conclusions

CTF Algorithm

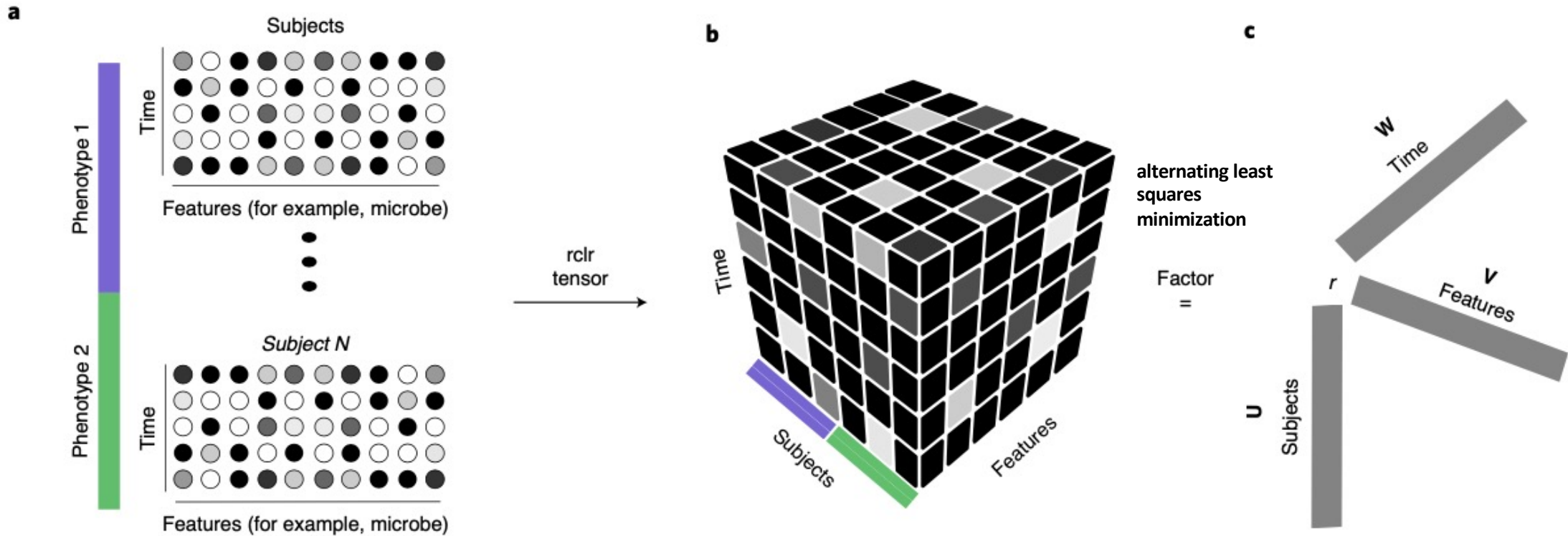


Figure 1 (Martino et al., 2021)

Data Normalization & Tensor Formation via rclr

$$\text{rclr}(\mathbf{x}) = \left[\log \frac{x_1}{g(\mathbf{x})}, \dots, \log \frac{x_D}{g(\mathbf{x})} \right] \quad (1)$$

$$g(\mathbf{x}) = \left(\prod_{i \in \Omega_x} x_i \right)^{1/|\Omega_x|} \quad (2)$$

Tensor Factorization by alternating least squares minimization

$$P_{\Omega}(T)_{ijt} = \begin{cases} T_{ij}, & \text{if } (i, j, t) \in \Omega \\ 0, & \text{Otherwise} \end{cases} \quad (3)$$

$$\min_{\{\sigma_i, \mathbf{a}_i, \mathbf{b}_i, \mathbf{c}_i\}_{i \in [r]}} \left\| P_{\Omega}(T) - P_{\Omega} \left(\sum_{i=1}^r \sigma_i (\mathbf{a}_i \otimes \mathbf{b}_i \otimes \mathbf{c}_i) \right) \right\|_F^2 \quad (4)$$

Subject trajectory = $\mathbf{a} \odot \mathbf{c} = [\mathbf{a}_1 \otimes \mathbf{c}_1, \dots, \mathbf{a}_r \otimes \mathbf{c}_r] \in \mathbb{R}^{d^2 \times r}$

Feature trajectory = $\mathbf{b} \odot \mathbf{c} = [\mathbf{b}_1 \otimes \mathbf{c}_1, \dots, \mathbf{b}_r \otimes \mathbf{c}_r] \in \mathbb{R}^{d^2 \times r}$

CTF Algorithm

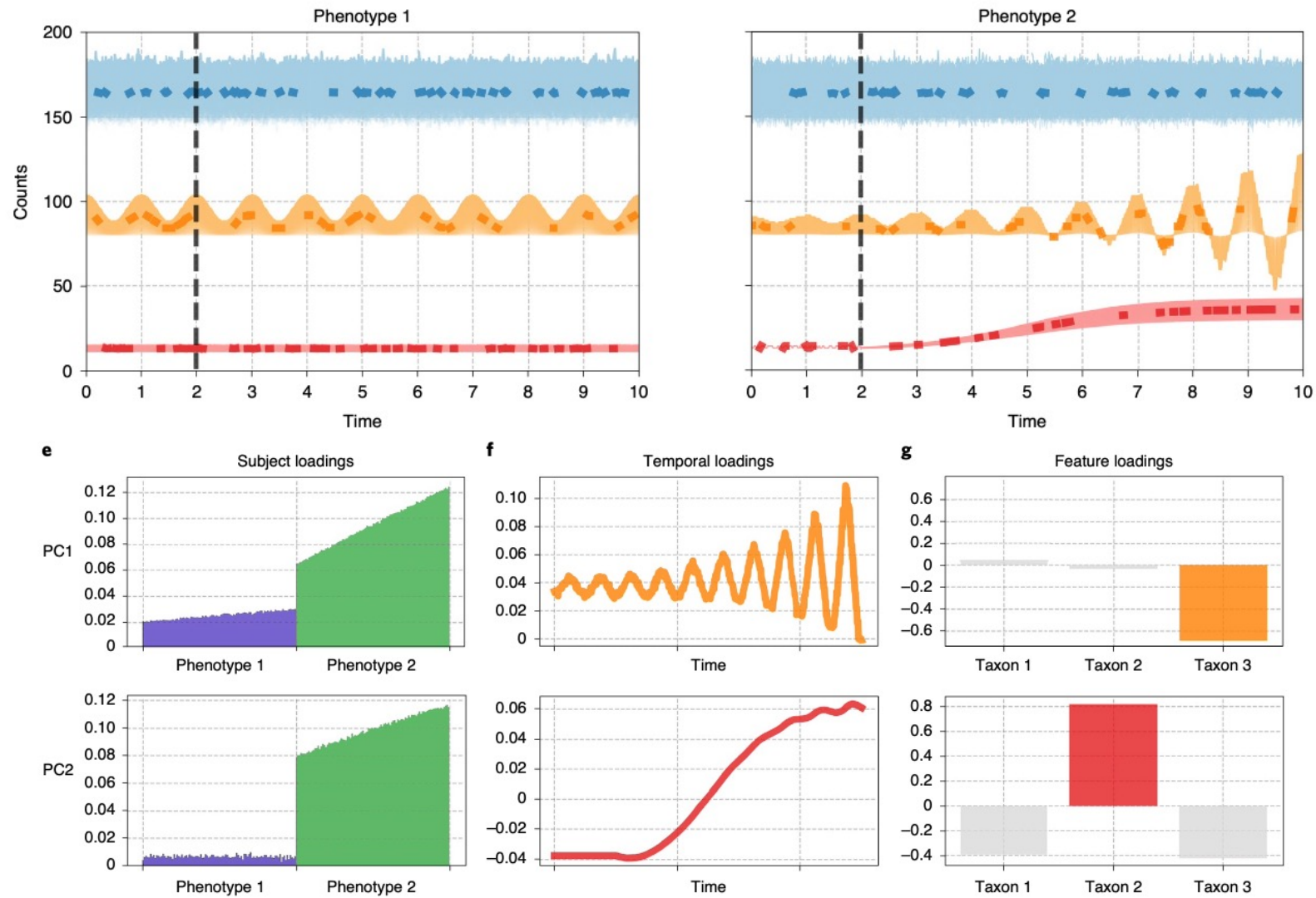


Figure 1 (Martino et al., 2021)

Outline

- Background on Tensor Factorization
- Issues with analyzing repeated microbial measure
- Paper Aims
- Methods employed: normalization & factorization
- **CTF Performance & Validation**
- Limitations & Conclusions

CTF performance against other models

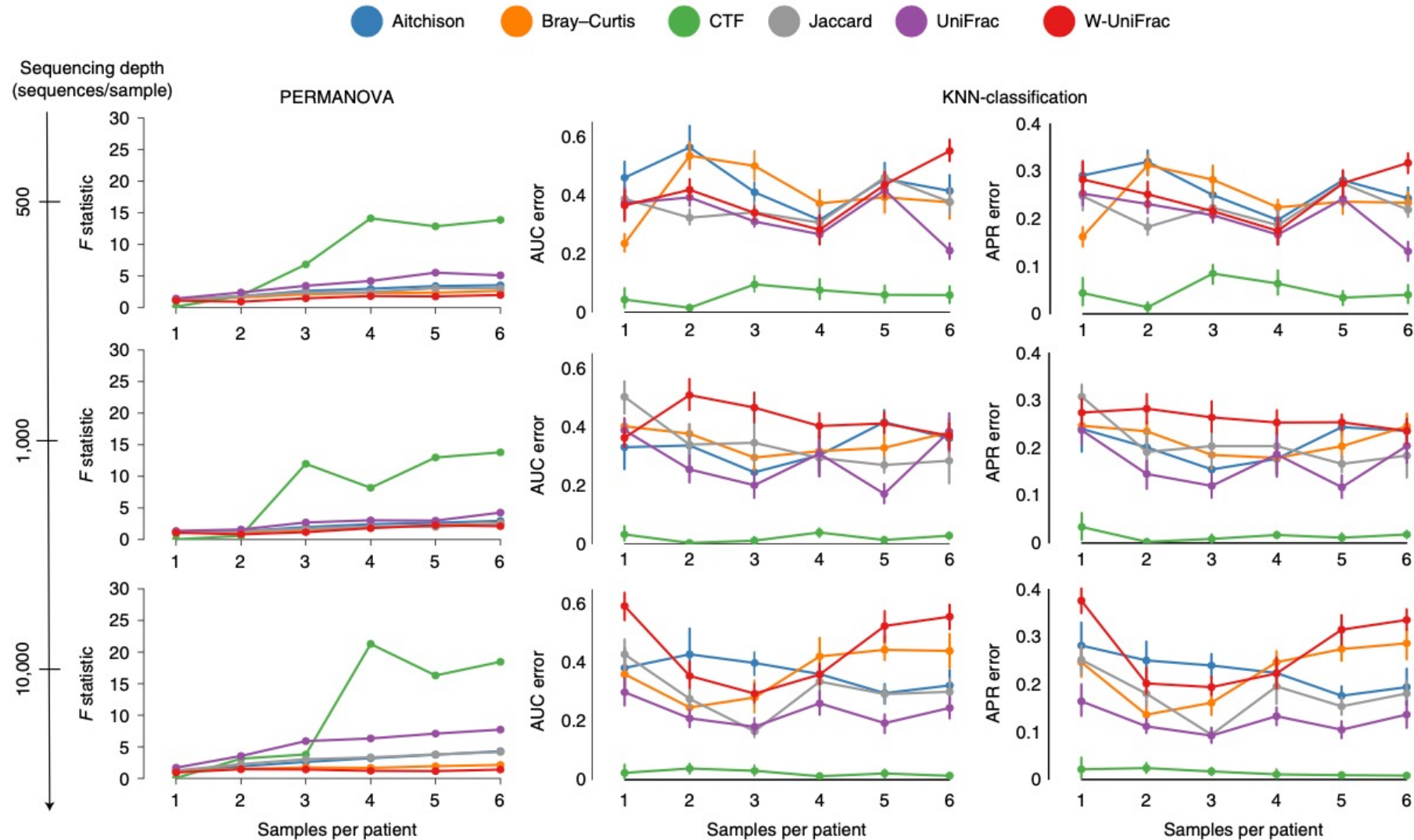


Figure 2 (Martino et al., 2021)

CTF – IBD dataset benchmarking

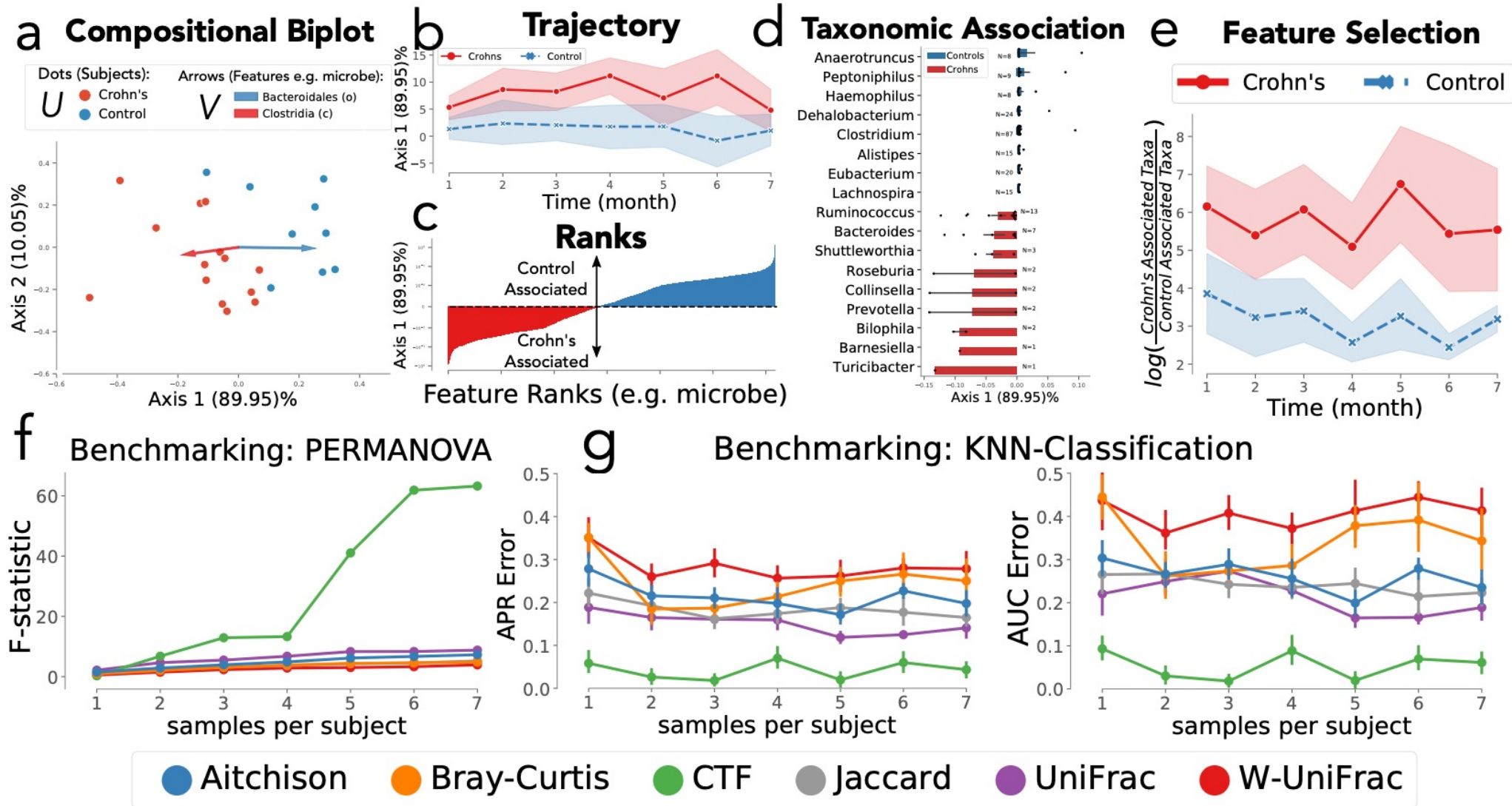


Figure S1 (Martino et al., 2021)

CTF able to predict birth method based on infant microbiome

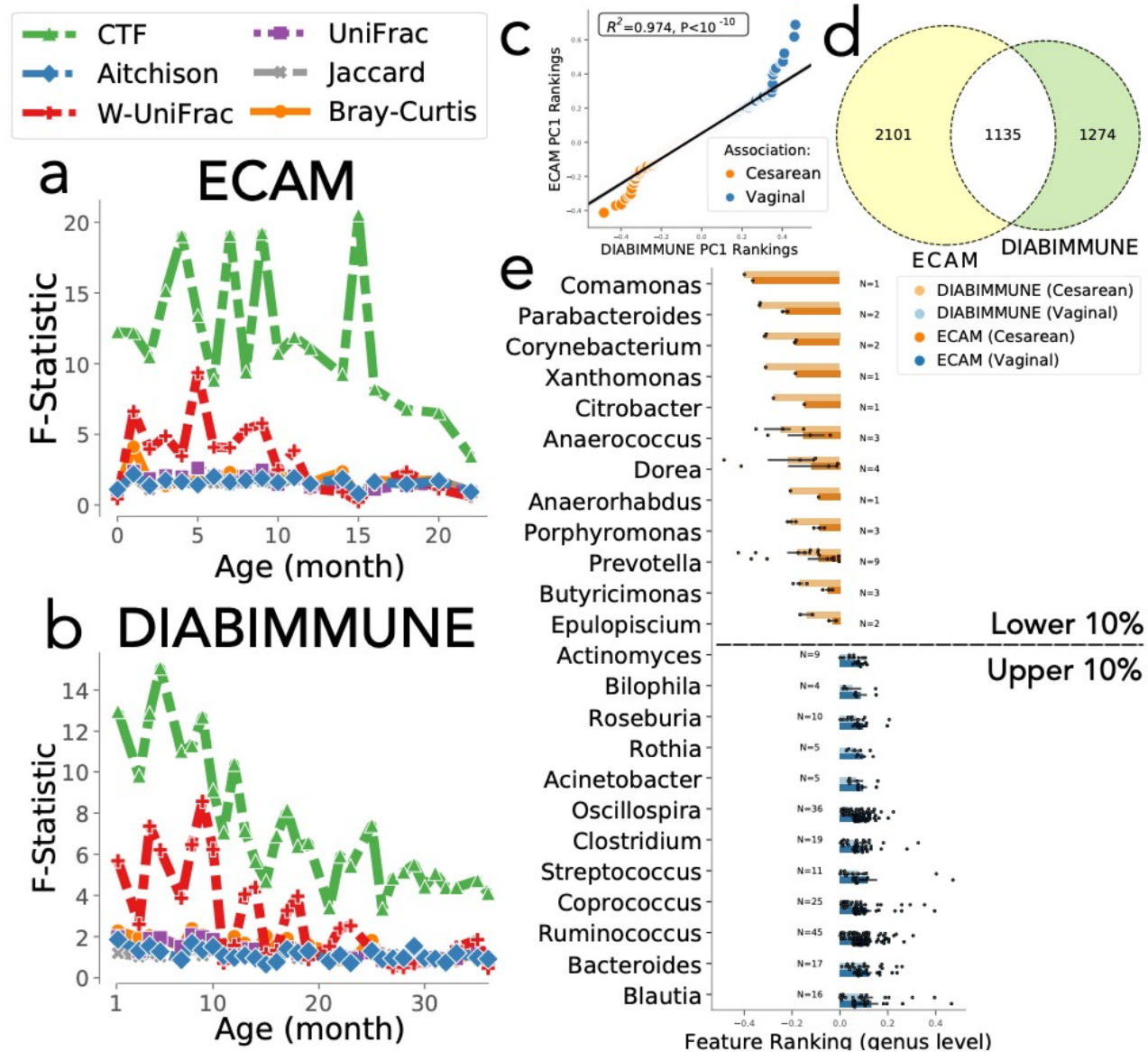


Figure S2 (Martino et al., 2021)

CTF able to predict birth method based on infant microbiome

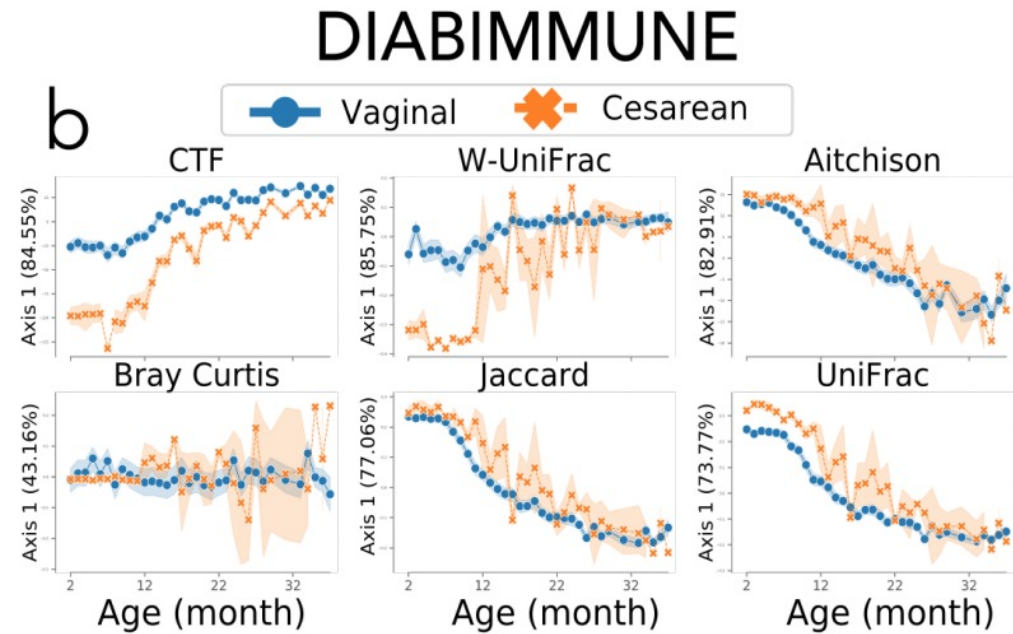
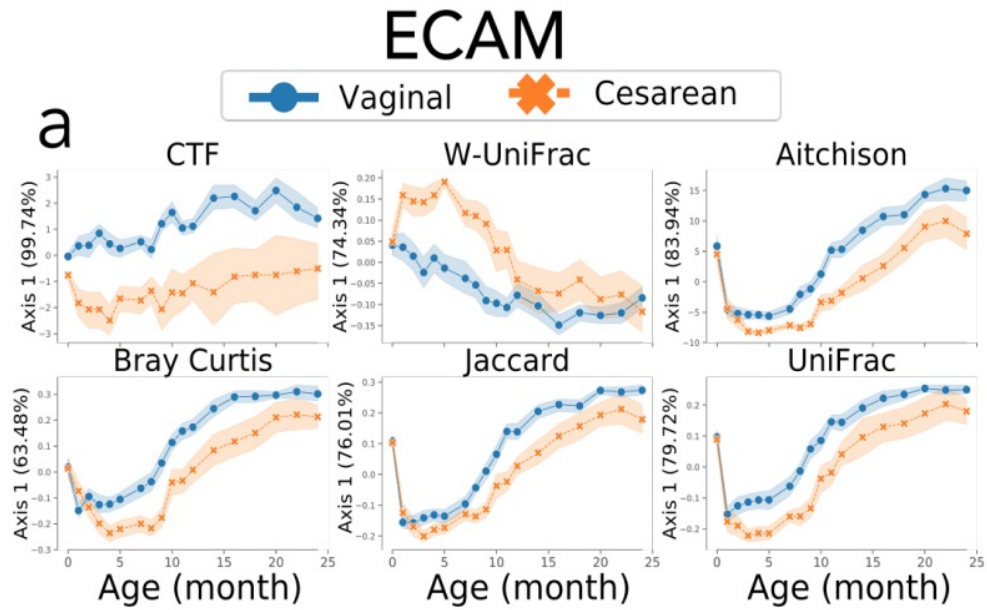


Figure S3 (Martino et al., 2021)

Outline

- Background on Tensor Factorization
- Issues with analyzing repeated microbial measure
- Paper Aims
- Methods employed: normalization & factorization
- CTF Performance & Validation
- Limitations & Conclusions

Limitations & Conclusions

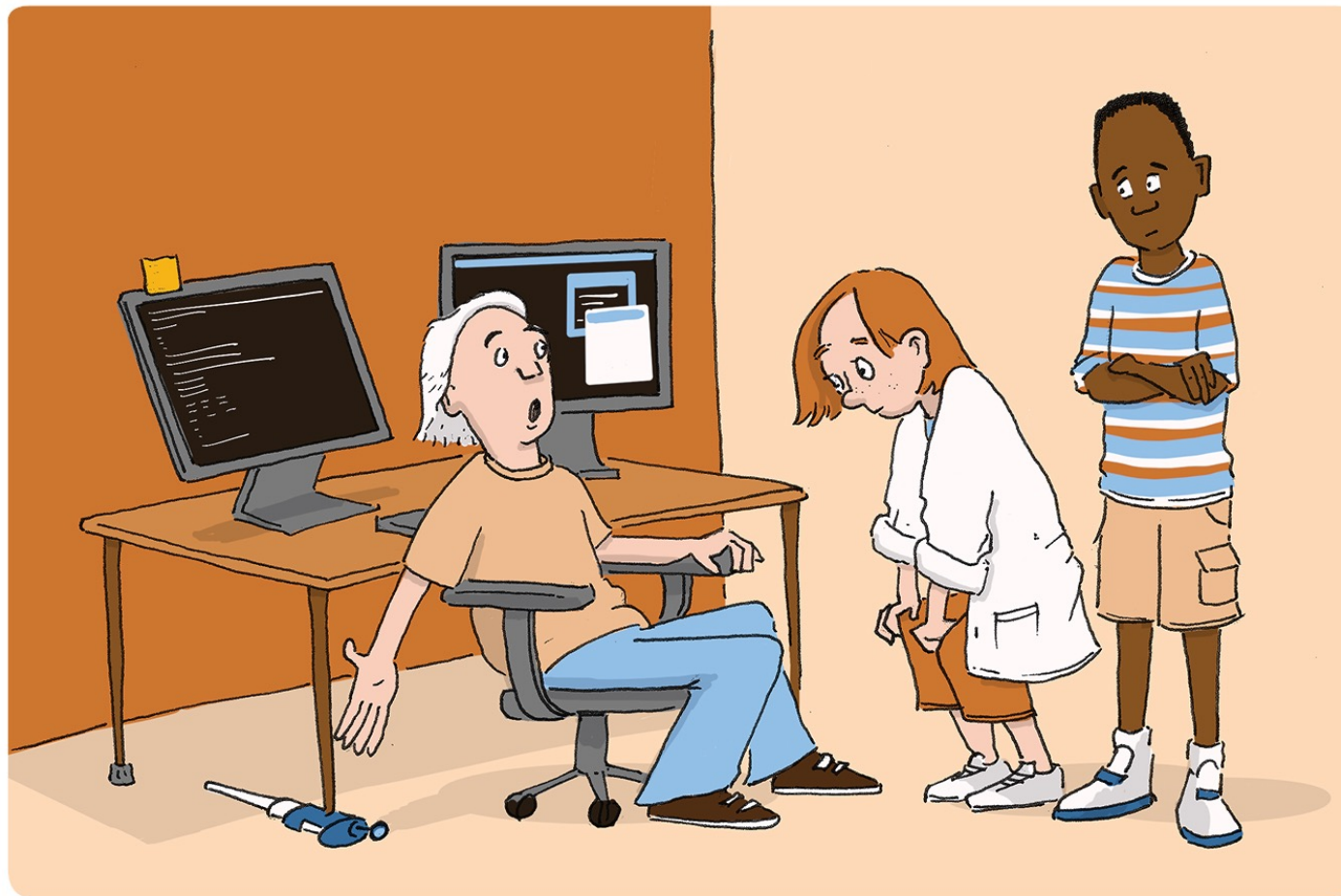
Limitations

- Assumes low ranking underlying data (discrete not gradient groupings)
- Beta-diversity measure does not account for confounders
- Does not account for ordering of longitudinal data (eqn 4)
- Need more applications to various longitudinal datasets

Conclusions

- CTF – supervised method with full use of repeated measures
 - Longitudinal
 - Spatially repeated measures
- Accounts for intraindividual variability
- Provides accurate community identification

Questions?



I use my micropipet every day.
It keeps my workstation from wobbling.