Application of Bayesian Techniques to Multi-omic Longitudinal Data

Daniel Ruiz-Perez[1] ,Jose Lugo-Martinez[2], Ziv Bar-Joseph[2], Giri Narasimhan[1]

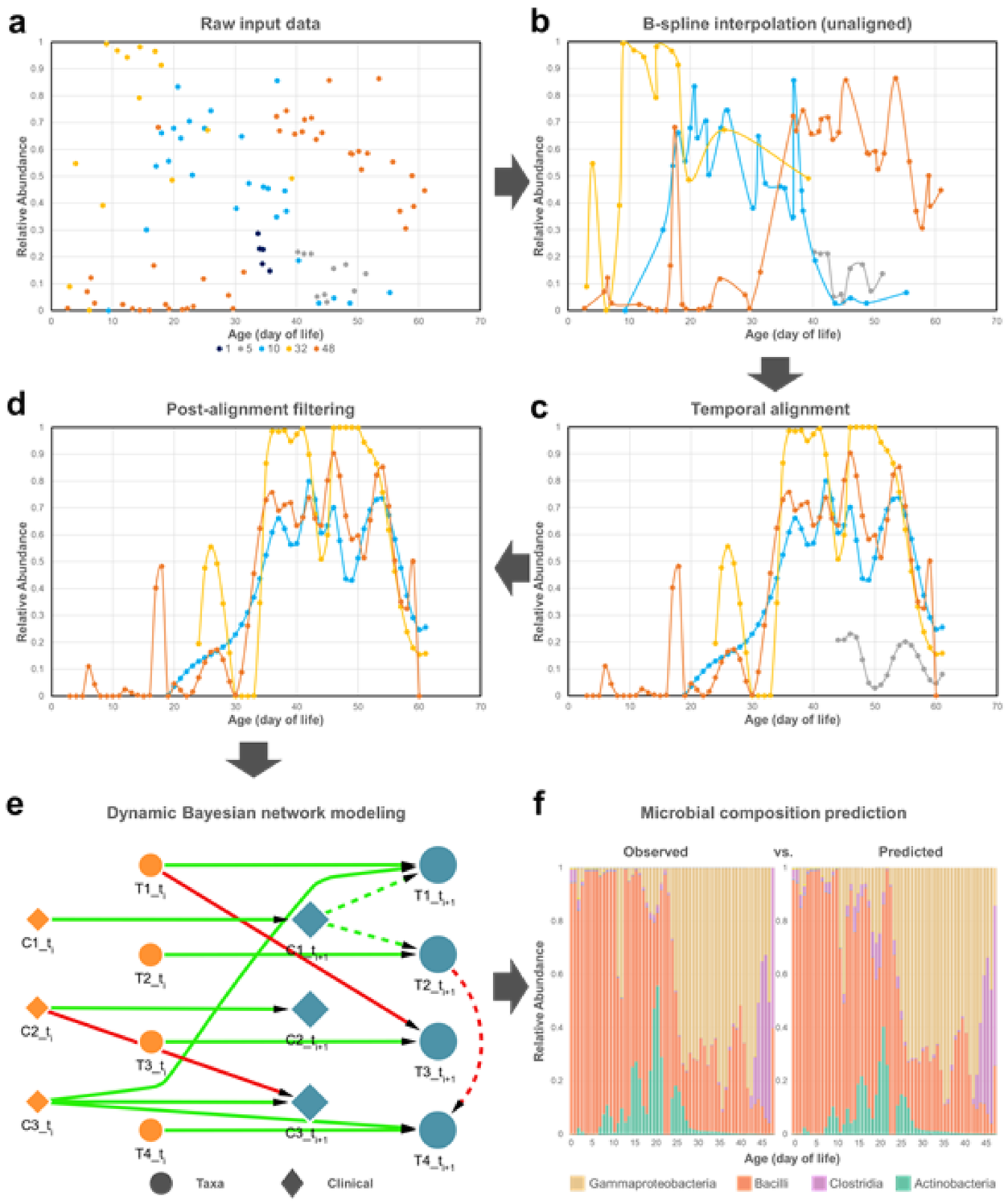
[1]School of Computing and Information Sciences, Florida International University, Miami, United States

[2]Computational Biology Department, Carnegie Mellon University, Pittsburgh, PA, USA.

1. Abstract

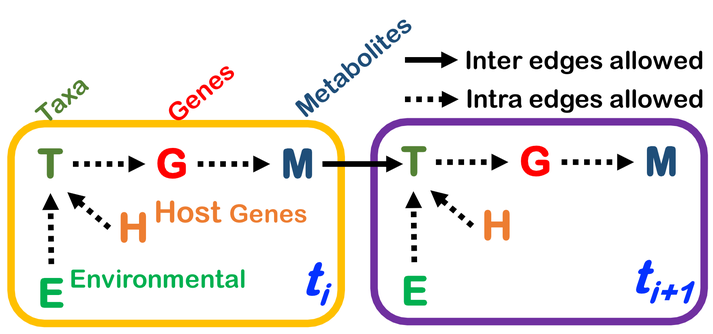
We report updates on the development of a computational pipeline for the inference of heterogeneous interactions from longitudinal studies (PALM), improving on previous work on a pipeline for homogeneous interactions. Metagenomics, metatranscriptomics, and metabolomics data from iHMP IBD were used and modeled using Dynamic Bayesian Networks. The data were interpolated and temporally aligned to account for differential rates of change, missing data, and irregular sampling. A metabolic framework is imposed, and the inferred validations are being validated both computationally and experimentally.

2. Single-omics pipeline



Pipeline of the process. The samples are first interpolated using b-splines to deal with irregular sampling rate and missing timepoint. Then they are temporally aligned to take care of temporal displacements, and the outliers are removed. Then the DBN model is learned using a 2-stage DBN to predict microbe-microbe interactions. Finally, the whole composition of each sample is predicted based on the previous timepoints.

3. Multi-omics incorporation



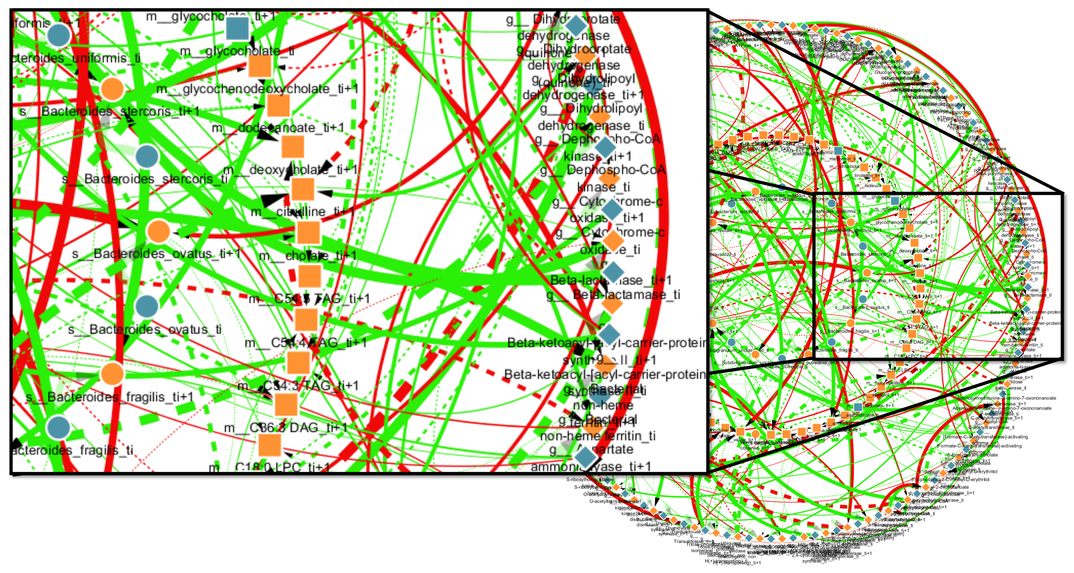
	Clinical	Taxa	Metabolite	Gene
Clinical	0	Intra	0	0
Taxa	0	Inter&Intra	Intra	Intra
Metabolite	0	Inter	Self	0
Gene	0	Inter	Intra	Self

Multi-omics restriction framework, to only allow interactions between specific types of nodes. Also the alignment step now can be done based on different omics.

4. Dataset and validation contributions

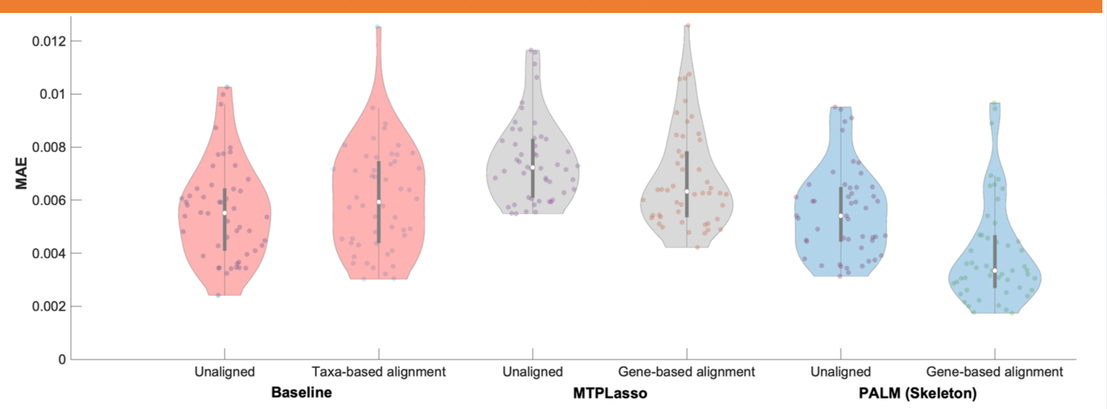
The dataset used was the IBD dataset from the integrative Human Microbiome Project.
Previously reported prediction errors are not biologically grounded, so we generated datasets with “valid” multi-omic interactions to function as the ground truth for our statistical validations. To decide whether an interaction from a gene to a taxon we look at KEGG and see if that gene is in the genome of that taxon, therefore capable of expressing it. To validate interactions between a gene and a metabolite, we look at KEGG pathways and see if that gene is an enzyme capable of metabolizing that compound. Lastly, for interactions from a taxa to a metabolite, we look at the tool MIMOSA and see if the taxa has the potential to produce that metabolite, giving its environment.

5. Learned multi-omics DBN



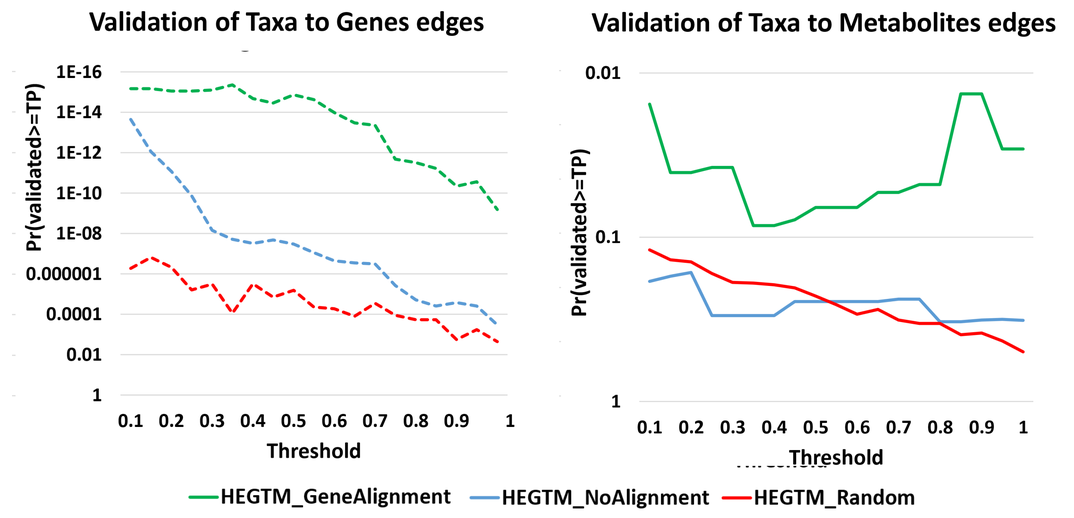
Learned DBN connecting taxa, metabolites, genes, and clinical variables

6. Prediction error validation



Comparison of the Mean Absolute Error (MAE) of our proposed DBN models against a baseline method using only metagenomic data and a previously published approach, MTPLasso, which models longitudinal multi-omics microbial data using a generalizded Lotka-Volterra. Figure also compares the performance of each method on the unaligned and aligned data sets.

7. Learned interactions statistical validation



Comparison of the probability of validating at least as many interactions as were validated by chance. Computed with a Poisson binomial distribution. A baseline that learns a random network is used for comparison in red.

8. In vitro validation of interactions

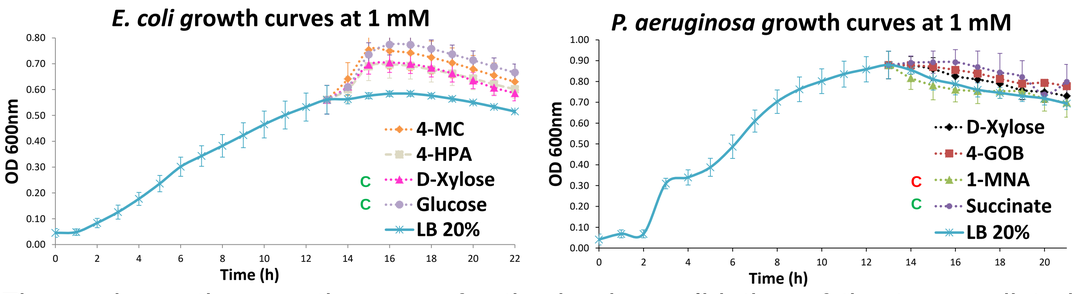


Figure shows the growth curves for the in vitro validation of the top predicted interactions by the DBN. The different metabolites were introduced at the end of the exponential) of the bacteria. Positive and negative controls appear with a C.

9. Conclusions

We present algorithms for the integrated analysis and modeling of longitudinal multi-omic microbiome, host and clinical data. We show that our method, PALM is able to successfully predict future (unseen) taxa abundance using the learned interaction network. We used 3 different in silico approaches to validate different types of predicted multi-omic interactions. Finally, we validate select interactions in the laboratory by performing growth experiments with metabolites for specific taxa, confirming our predictions.