Charting Global Patterns of Gut Microbiome Maturation in Infancy through Microbial Age Modeling

Guilherme Fahur Bottino[1], Kevin S. Bonham[2] , Shelley H. McCann[1] , Fadheela Patel[3] , Kirsty Donald[3] , Curtis Huttenhower[4] , Vanja Klepac-Ceraj[1]

[1]Wellesley College, Wellesley, MA, United States; [2]Tufts University School of Medicine, Boston, MA, United States; [3]University of Cape Town, Cape Town, South Africa; [4]Harvard T.H. Chan School of Public Health, Boston, MA, United States.

GOAL: to develop a gut-microbiome age model to study the effects of the gut microbiome on other age-dependent development features

Multiple environmental factors influence the development of the gut microbiome, which experiences dramatic changes during early infancy [1,2]. There is accumulating evidence that the gut microbiota and its metabolites also regulate various aspects of neurodevelopment, physiology and behavior [3].

Building a model that estimates age from gut-microbial taxonomic profiles can help us understand typical developmental trajectories and capture temporal trends and deviations. Its outputs can be useful to predict other developmental outcomes.

Global metagenomes enable large-scale meta-analysis

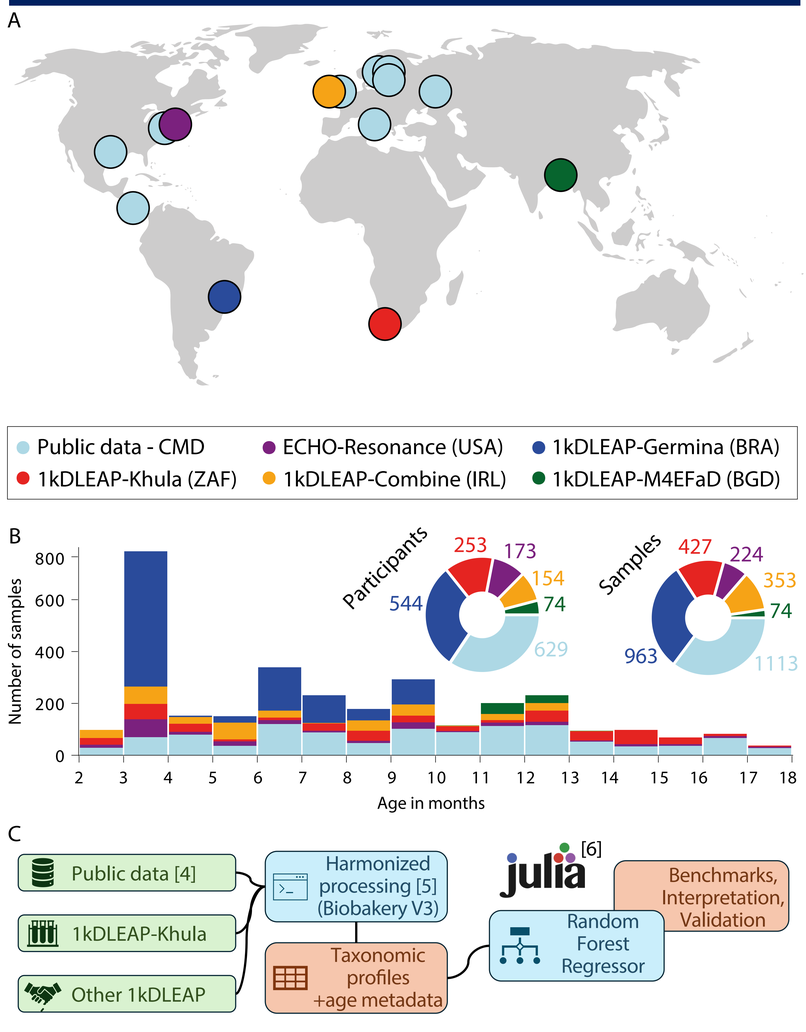

Figure 1. Data origin, description and analysis pipeline. (A) approximate geographical sites of stool sample collection. (B) normalized stacked kernel density plots of age distribution per stool sample, color-coded by data source. (C) simplified illustration of the data analysis, from data curation, sample collection, sequence processing, model training, validation and interpretation.

Harmonized computational processing provides a continuous diversity landscape

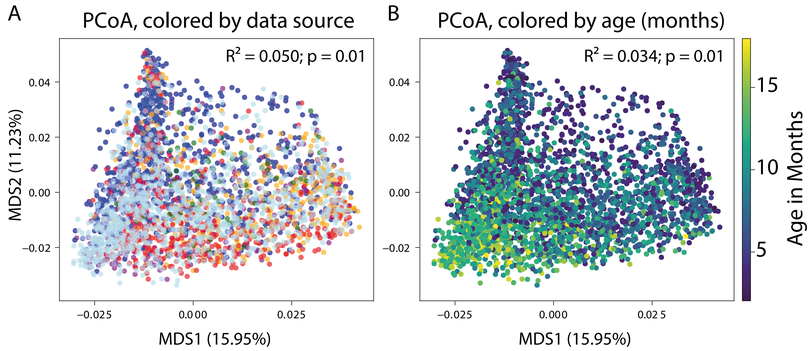

Figure 2. NMDS decomposition of Bray-Curtis distance matrix for all taxonomic profiles analyzed, colored by data source (A) – same key as 1A/B – and age at collection (B). Samples distribute continuously through the diversity ordination, on patterns that strongly reflect data source, but without forming isolated clusters and with a perceivable degree of mixture. First principal component is heavily loaded on the age of sample collection. PERMANOVAS (appended to both plots) show that source and age explain comparable amounts of variance.

Pooled metagenomes predict age with high temporal resolution

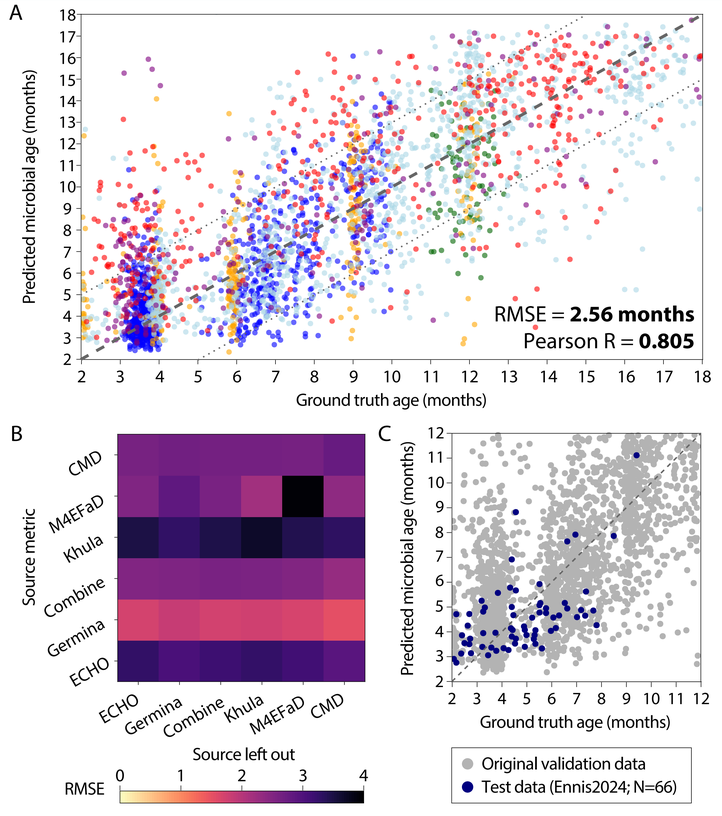

Figure 3. Model benchmark, figures of merit and results from Leave-One-Source-Out-Cross-Validation. (A) scatterplot of predicted ages versus ground truth ages at sample collection (in months) – same color key as 1A/B. Identity line ($y=x$) ±3 months added for reference. (B) cohort metrics for Leave-One-Source-Out-Cross-Validation. (C) Age predictions for an independent external test set (not used during training).

Top changing taxa show feeding transitions and dietary exposures

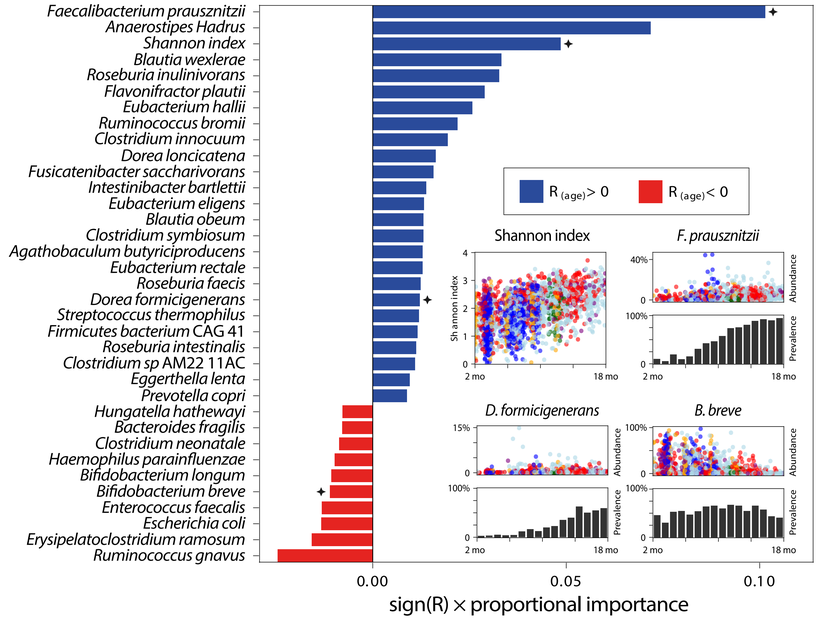

Figure 4. Most important species for age estimation as measured by MDI/GINI importance on Random Forest models, accompanied by scatterplots of prevalences and relative abundances (where present) as a function of age, color-coded by data source – same key as 1A/B –, for selected species.

Learned gut microbial patterns generalize across different sites

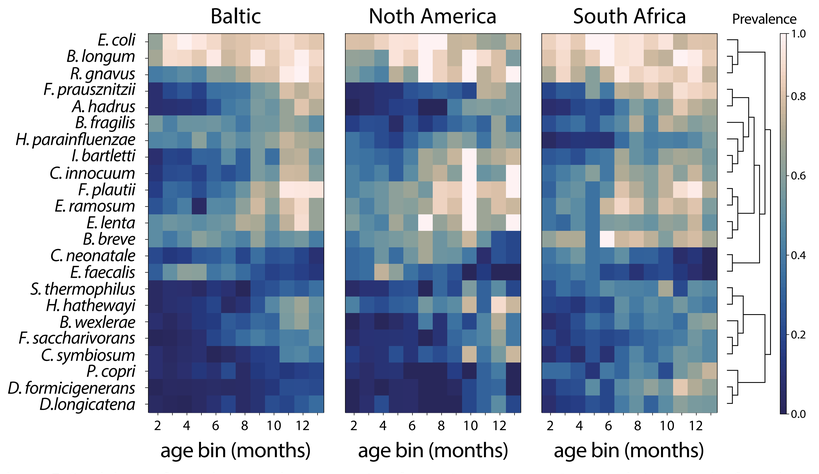

Figure 5. Breakdown of prevalence evolution over time for top important species considering samples from three different data sources (Baltic countries; North America; South Africa). Y-axis was ordered based on hierarchical clustering of mean prevalence vectors. Cluster assignment illustrated on dendrogram to the side.

Microbiome development in the first years of life **follows predictable normative trajectories**, that emerge as underlying **machine-learnable patterns** when a large and diverse dataset of metagenomes is used to train an age prediction model from taxonomic profiles.

References

https://tinyurl.com/2025BottinoHCMPH

Acknowledgements

This research was funded by NIH UG3 OD023313 and Wellcome LEAP 1kD