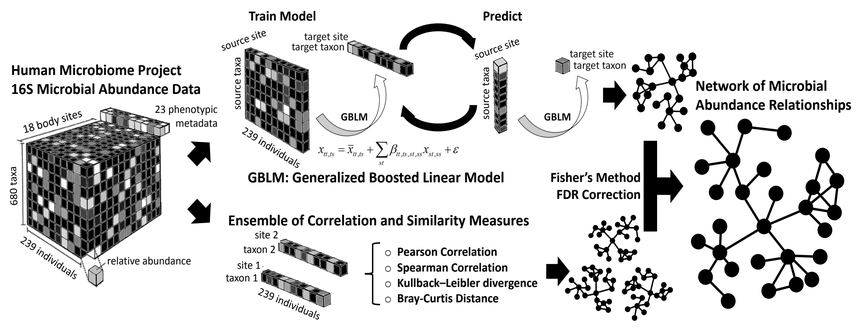
Prediction of bacterial relationships in the human microbiome

[1]Karoline Faust*, [2]J. Fah Sathirapongsasuti*, [3]Jacques Izard, [2]Nicola Segata, [4]Dirk Gevers, [2]Curtis Huttenhower* and [1]Jeroen Raes*

[1]Bioinformatics and (Eco-)Systems Biology, VIB-VUB, Pleinlaan 2, B-1050, Brussels, Belgium
[2]Department of Biostatistics, Harvard School of Public Health, 655 Huntington Avenue, Boston, USA
[3]Department of Molecular Genetics, The Forsyth Institute, 245 First Street, Cambridge, USA
[4]Broad Institue of MIT and Harvard, 7 Cambridge Center, Cambridge, USA
Contact: karoline.faust@vib-vub.be

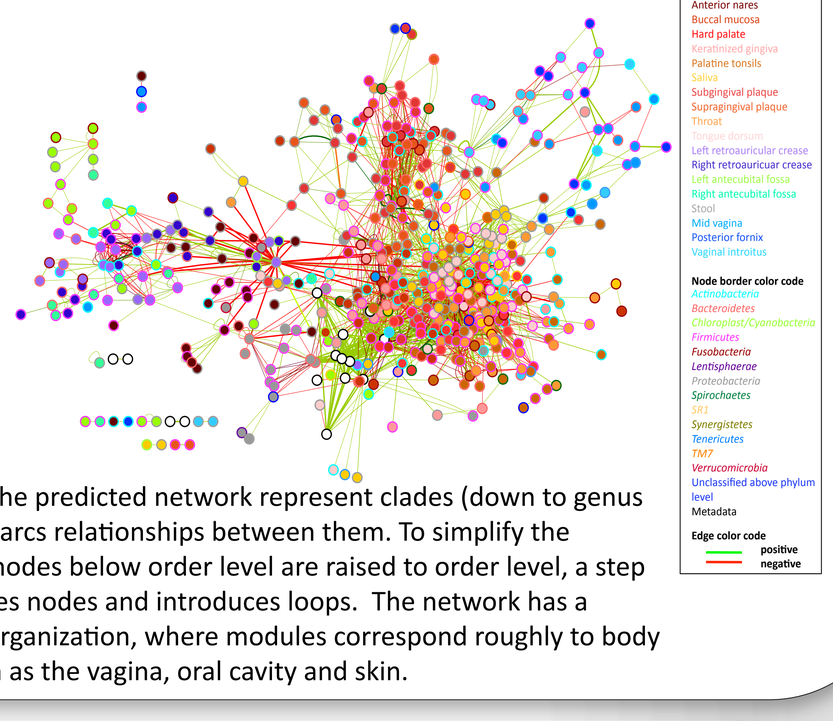
Metagenomic sequencing projects are accumulating abundance data for microbial organisms in a wide variety of environments, including the human body. These data enable ecological studies of microbiota that could not be carried out previously. In macro-ecological data sets, non-random patterns of species distributions were found that reflect ecological relationships, such as the checkerboard pattern, which indicates competition [1]. The analysis of microbial abundance data revealed similar non-random patterns for microorganisms [2]. Recently, the Human Microbiome Consortium [3] has compiled a massive data set of microbial sequences in up to 18 human body sites. Here, we present a network of microbial relationships built from the 16S phylotype data of the Human Microbiome Project.

Prediction of significant co-occurrence and mutual exclusion relationships between human microorganisms



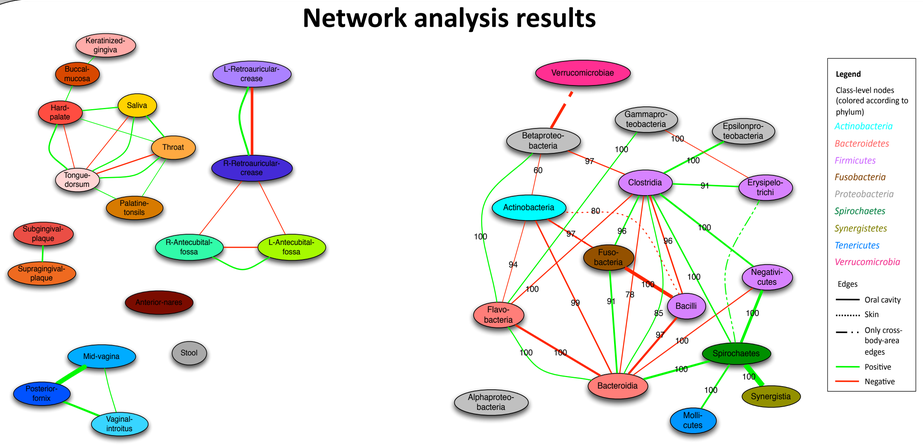
To infer the final network, we merge the results of a sparse linear regression method (GBLM) with those from an ensemble of distance and correlation measures using Fisher's method. We then compute edge-specific p-values from both permutation (for the null value) and bootstrap distributions. For the permutation, we apply a novel re-normalization-based approach that we observed in simulations to reduce the impact of compositional bias [4].

Network of microbial relationships



Nodes in the predicted network represent clades (down to genus level) and arcs relationships between them. To simplify the network, nodes below order level are raised to order level, a step that merges nodes and introduces loops. The network has a modular organization, where modules correspond roughly to body areas such as the vagina, oral cavity and skin.

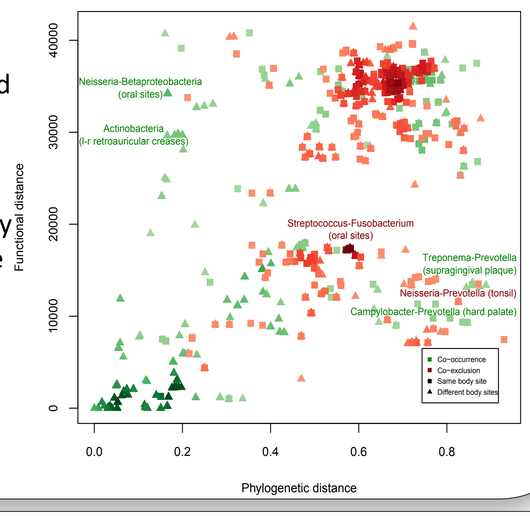
Network analysis results



Body site and class relationship networks. The edge thickness represents the number of positive (green) or negative (red) relationships that member clades of the body sites or classes share. Thus, the body site network can be interpreted to group body sites into different microbial habitat types, whereas the class network highlights negative (e.g. *Bacilli*) and positive (e.g. *Spirochaetes*) hubs.

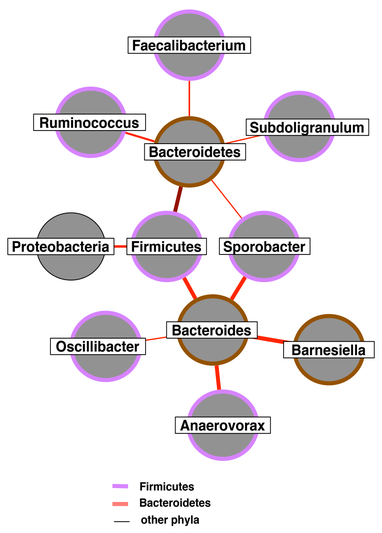
Functional analysis results

For each significant pair of clades we computed phylogenetic and functional distance in terms of 16S dissimilarity and number of shared orthologous gene families respectively.
Clades with small phylogenetic and functional distances tend to be positively associated, whereas the opposite is true for clades with large phylogenetic and functional distances. Interestingly, pairs whose functional distance is larger than expected from their phylogenetic distance tend to form positive relationships.

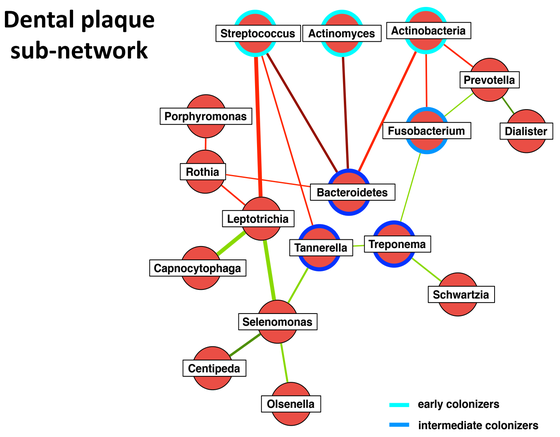


Sub-networks in agreement with known human microbiology

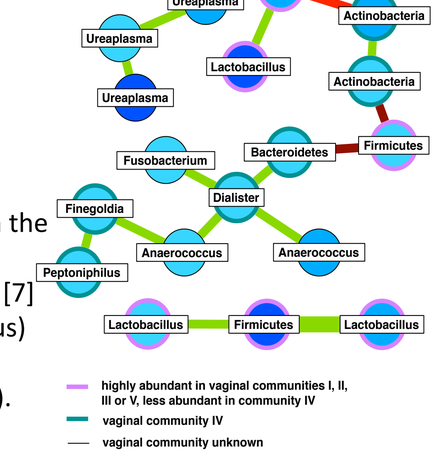
Gut sub-network



Dental plaque sub-network



Vaginal sub-network



Known alternative community configurations in the vagina [5] and successional stages in the subgingival dental plaque [6] were reproduced. In the gut, we found negative relationships between two of the three enterotype [7] drivers (Bacteroides and Ruminococcus) and more generally between Bacteroidetes and Firmicutes (e.g. [8]).

We have inferred and analyzed a network of microbial relationships, where positive edges indicate niche sharing or mutualism and negative edges membership to alternative communities or successional stages or competition. The network is roughly organized into body areas reflecting different microbial niches and reproduces a number of known microbial relationships in the gut, the vagina and the subgingival dental plaque. We anticipate that more detailed community structure will emerge when applying our methodology to data sets that allow the assignment of taxonomy down to species or even strain level.

References
[1] Diamond, J.M. "Assembly of species communities" in Ecology and evolution of communities (eds. Cody, M. & Diamond, J.M.) 342-444 (Harvard University Press, 1975).
[2] Horner-Devine, M.C. et al. "A Comparison Of Taxon Co-Occurrence Patterns For Macro- And Microorganisms" *Ecology* 88, 1345-1353 (2007).
[3] Turnbaugh, P.J. et al. "The Human Microbiome Project" *Nature* 449, 804-810 (2007).
[4] Aitchison, J. "A Concise Guide to Compositional Data Analysis" in CDA Workshop Girona (2003).
[5] Ravel, J. et al. "Vaginal microbiome of reproductive-age women" *PNAS* 108, 4680-4687 (2011).
[6] Dumitrescu, A.L. "Antibiotics and Antiseptics in Periodontal Therapy", (Springer, 2011).
[7] Arumugam, M., Raes J. et al. "Enterotypes of the human gut microbiome" *Nature* 473, 174-180 (2011).
[8] Turnbaugh, P.J. et al. "A core gut microbiome in obese and lean twins" *Nature* 457, 480-485 (2009).