# Dissimilarity measures to characterize compositions of microbial communities

**G. Mateu-Figueras, P. Daunis-i-Estadella, M. Lopez-Siles, J.A. Martín-Fernández**
Universitat de Girona, Spain; gloria.mateu@udg.edu

**Abstract**

The composition of microbial communities play important roles in biology, ecology and human health (Bardgett and others, 2012; Berendsen and others, 2012; Human microbiome Project Consortium, 2012). As indicated in Washburne and others (2016), the analyses of microbial communities aim either (i) to describe community quantifying the relative abundance of individual microbial taxa, (ii) to characterize how the microbial community change across space, time or in a response to a treatment, or (iii) to quantify differences between two communities. Distance and dissimilarities measures like Bray-Curtis or UniFract are commonly used (Coburn and others, 2015; Goslee and Urban, 2007) to compute the differences between microbial communities for non-parametric manova, dimensionality reduction or clustering methods.

Recently the compositional nature of the data has been discussed and the methodology based on logratios is recommended for the analysis of such data sets (Gloor and others, 2016, Tsilimigran and Fodor, 2016). This opens the door to the Aitchison distance as an adequate distance or to other logratio indices, for example, the F-E index to discriminate amongst different intestinal disorders (Lopez-Siles and others, 2014). Following Tsilimigran and Fodor (2016), it remains an open question to study how much the use of the Aitchison distance would beneficially impact results compared with these more commonly used metrics.

In this work we analyse several distances and dissimilarity measures that are popular in the biological sciences. In particular we show how the changes in size and shape of compositions of microbial communities affect to these measures and we explore its subcompositional coherence. To illustrate the performance of these measures real and simulated data sets are analysed.

## References

Bardgett, R.D., Freeman, C. and Ostle, N.J. (2008). Microbial contributions to climate change through carbon cycle feedbacks, *The ISME Journal, 2*(8), 805-814.

Berendsen, R.L., Pieterse, C.M.J. and Bakker, P. (2012). The rhizosphere microbiome and plant health. *Trends in plant science, 17*(8), 478-486.

Corbun, B., Wang, P.W., Caballero, J.D., Clark, S.T., Brahma, V., Donaldson, S., Zhang, Y., Surendra, A., Gong, Y., Tullis, D.E., Yau, Y.C.W., Watwrs, V.J., Hwang, D.M. and Guttman, D.S. (2015). Lung microbiota across age and disease in cystic fibrosis. *Scientific reports* 5:10241, DOI: 10.1038/srep10241

Gloor, G.B., Wu, J.R., Pawlowsky-Glahn, V. and Egozcue J.J. (2016). It's relative: analyzing microbiote data as compositions. *Annals of Epidemiology, 26*(5), 322-329.

Goslee, S.C. and Urban, D.L. (2007). The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software 22*(7), 1-19.

Human Microbiome Project Consortium (2012). Structure, function and diversity of the healthy human microbiome. *Nature, 486*(7402), 207-214.

López Siles, M.; Martinez-Medina, M.; Busquets, D.; Sabat-Mir,M; Duncan, S.H.; Flint, H.J.; Aldeguer, X.; Garcia-Gil, L.J. (2014). Mucosa-associated *Faecalibacterium prausnitzii* and *Escherichia coli* co-abundance can distinguish Irritable Bowel Syndrome and Inflammatory Bowel Disease phenotypes. *International Journal of Medical Microbiology, 304*, 464-475.

Tsilimigran, M.C.B., Fodor, A.A. (2016). Compositional data analysis of the microbiome: fundamentals, tools, and challenges. *Annals of Epidemiology, 26*, 330-335.

Washburne, A.D., Silverman, J.S., Left, J.W., Bennet, D.J., Darcy, J.L., Mukherjee, S., Fierer, N. and David, L.A. (2016) Phylogenetic factorization of compositional data. Preprint available in http://dx.doi.org/10.1101/074112.