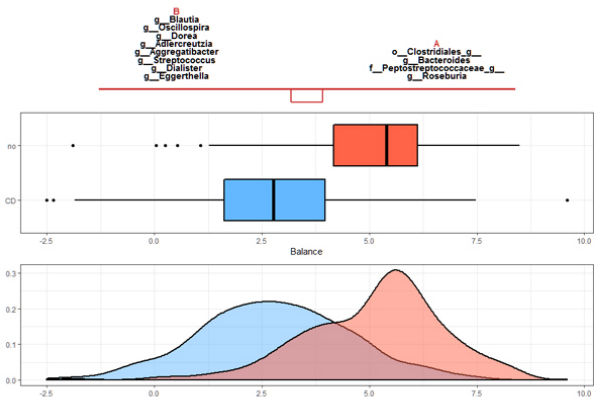
ViRVIG
Visualització, Realitat Virtual i Interacció Gràfica

www.virvig.eu
C. Llorens i Artigas 4-6
Edifici U, 1er pis
E- 08028- Barcelona

Variable selection in microbiome compositional data analysis

Susin, Antonio; Wang, Y.; Le Cao, K.; Calle, M.



Though variable selection is one of the most relevant tasks in microbiome analysis, e.g. for the identification of microbial signatures, many studies still rely on methods that ignore the compositional nature of microbiome data. The applicability of compositional data analysis methods has been hampered by the availability of software and the difficulty in interpreting their results. This work is focused on three methods for variable selection that acknowledge the compositional structure of microbiome data: selbal, a forward selection approach for the identification of compositional balances, and clr-lasso and coda-lasso, two penalized regression models for compositional data analysis. This study highlights the link between these methods and brings out some limitations of the centered log-ratio transformation for variable selection. In particular, the fact that it is not subcompositionally consistent makes the microbial signatures obtained from clr-lasso not readily transferable. Coda-lasso is computationally efficient and suitable when the focus is the identification of the most associated microbial taxa. Selbal stands out when the goal is to obtain a parsimonious model with optimal prediction performance, but it is computationally greedy. We provide a reproducible vignette for the application of these methods that will enable researchers to fully leverage their potential in microbiome studies.

http://dx.doi.org/https://doi.org/10.1093/nargab/lqaa029