How to calculate number of samples in the design of pre/pro-biotics studies (metagenomic studies)

Monleón-Getino A[1*], Rodríguez-Casado CI[2], Méndez-Viera J[3]

[1]amonleong@ub.edu, Section of Statistics, Department of Genetics, Microbiology and Statistics (G.M.S.), University of Barcelona. GRBIO (Research Group in Biostatistics and Bioinformatics).
[2]clara.rodriguez@ub.edu, Section of Statistics, G.M.S, University of Barcelona.
[3]jmendez@ub.edu, Fermentation Service, Faculty of Biology, University of Barcelona.
*Presenting author

Background and objectives:

Nowadays a high amount of pre/probiotic studies use metagenomics to control the effectiveness of the use of pre/probiotic. Many bacterial species are non-cultivable out of their natural environment and, therefore, some sets of species can only be studied all together within their environment (e.g. human gut) using metagenomics, sampling and sequencing DNA from all species. Sample size in metagenomics is an unresolved problem in which statistics and computational methods have much to say, especially from the point of view of Bayesian approach. One possibility is the use of a bietapic method: conduct a pilot study, prior to the final study.

Methodology:

We propose the use of a Bayesian method based on Markov Chain Monte Carlo simulations (MCMC) to calculate the richness, supposing a Dirichlet-multinomial probability distribution for the matrix of richness-abundance. The sample size on the base of sampling effort (90, 95 and 99%) is then calculated assuming saturation or not using parametric and non-parametric nonlinear models.

Results and conclusions:

These calculations have been implemented in the new library **'BDSbiost3'** for R, which has been used in different simulation scenarios of the common practice in metagenomics (low abundance, high richness, oversampling, under sampling) and the results (Sp) are presented with the effort curve and a value proposed of the correct number of samples to reach the extrapolated richness of the 90, 95 and 99%. This function can help researchers in pre/probiotics to better design studies and save on number of samples.

Acknowledgements:

GRBIO (Research Group in Biostatistics and Bioinformatics).