

# A Bayesian Volcano Plot Visualization for Bioinformatics Analysis

Desiré Vázquez Ortiz,  
desire.vazquez@upr.edu  
Department of Mathematics, UPR-RP.

May 11, 2026



THESIS  
DEPARTMENT OF  
MATHEMATICS

## Abstract

Microbiome data analysis is essential for understanding the role of microbial communities in human health, yet statistical interpretation remains challenging due to high dimensionality, sparsity, and multiple testing. Volcano plots are commonly used to visualize differential abundance by combining effect size and statistical significance; however, their reliance on  $p$ -values and the  $-\log_{10}(p)$  transformation can lead to overinterpretation of evidence against the null hypothesis. In this study, microbiome data were analyzed using a pipeline implemented in R, integrating abundance, taxonomy, and metadata through the `phyloseq` framework. Differential abundance between community state types (CST I and CST IV) was assessed using  $\log_2$  fold change and the Wilcoxon rank-sum test, with  $p$ -values adjusted via the Benjamini–Hochberg procedure. A Bayesian-calibrated volcano plot was then constructed using the Sellke–Bayarri–Berger bound to transform  $p$ -values into a lower bound on the posterior probability of the null hypothesis. The classical volcano plot identified numerous statistically significant taxa, whereas the Bayesian-calibrated plot showed that many correspond to non-negligible posterior probabilities of the null hypothesis, indicating weaker evidence. These results demonstrate that conventional thresholds such as  $p = 0.05$  may overstate statistical evidence. The proposed approach provides a more interpretable and conservative framework for visualizing differential abundance in high-dimensional microbiome studies.