

SHAMAN : a shiny application for quantitative metagenomic analysis

Contact : Amine Ghozlane (amine.ghozlane@pasteur.fr)

Background:

Quantitative metagenomics is broadly employed to identify genera or species associated with several diseases. These data are obtained by mapping the reads of each sample against operational taxonomic units (OTU) or a gene catalog. SHAMAN (shaman.pasteur.fr) was one the first web application that allowed to clinician and biologist to perform an interactive analysis of quantitative metagenomics data with a dynamic-interface dedicated to the diagnostic and to the differential analysis. The interface integrates the experimental design (association of sample to one or several conditions), the statistical process for differential analysis and a real-time visualisation system.

SHAMAN is based R, Shiny and DESeq2. The analytical process is divided into four steps : count matrix/annotation submission, normalisation, modelisation and visualisation. The count matrix is normalised at the OTU/gene level using the DESeq2 normalisation method and then, based on the experimental design, a generalised linear model is applied to detect differences in abundance at the considered taxonomic level.

Two years after first release, we can see a great interest from the metagenomics community with 5 publications using SHAMAN (in Nature microbiology, PNAS and Science Advances), 3 publications not involving SHAMAN authors, 74 active users per month (1430 unique visitors since first publication - 70 % are regular users) and 514 downloads of the docker application. Several trainings were also performed to train biologist to use SHAMAN in the Pasteur Network and at ENS.

Project 1:

We want to integrate a full automatized bioinformatic workflow based on Nextflow for targeted metagenomics data. This implementation will follow the current approach already implemented in bash workflow (<https://github.com/aghazlane/masque>). The workflow should also be included in the SHAMAN docker application for local installation on windows/mac/linux.

Project 2:

Several user have requested the inclusion of LEFse analysis and random forest analysis in SHAMAN (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3218848/>). These developments should be perform in R using Shiny API.