

DANCE WITH SAMBA: AN INNOVATIVE APPROACH TO STUDY FISH GUT MICROBIOME USING BAYESIAN NETWORKS.

F. Moroni^{1*}, F. Naya-Català¹, B. Soriano^{1,2,3}, A. I. Hafez², C. Llorens², J. Pérez-Sánchez¹

¹Institute of Aquaculture Torre de la Sal (IATS), Consejo Superior de Investigaciones Científicas (CSIC), Castellón, Spain. ²Biotechvana, Parc Científic Universitat de València, Paterna, Spain. ³Institute for Integrative Systems Biology (I2SysBio), Universitat de Valencia and CSIC (UVEG-CSIC), Spain. E-mail: federico.moroni@csic.es

Introduction

Gut microbiota creates a profound connection with the host, taking part in numerous essential aspects of its physiology (Naya-Català et al., 2022). Moreover, microbial populations build a mazy network of interconnections between species, characterized by cooperation or competition that can affect the whole microbiome (Yajima et al., 2023). It is well documented that several factors can affect the microbial composition, but the interpretation of the probabilistic factors that underlie these interactions are still unrevealed. Indeed, the current approaches in fish microbiota research are mainly focused on the comparison of bacterial abundances between experimental groups or meta-analyses matching different studies. In this context, to make a step forward in the understanding of the fish microbiome dynamics, we implemented SAMBA (Structure-Learning of Aquaculture Microbiomes using a Bayesian Approach), an open-source, web-based platform with a graphical user interface implemented with Shiny (Soriano et al., 2022). This tool uses probabilistic Bayesian Networks (BN) to evaluate the conditional dependencies within a set of experimental variables and taxa. The aim of this study is to show how the implementation of SAMBA works as a model of causal predictions in the gut microbiota of aquaculture species.

Materials and methods

Experimental data for training SAMBA were taken from Spanish national (ThinkInAzul) and H2020 EU projects (AquaIMPACT, AQUAEXCEL²⁰²⁰, AQUAEXCEL3.0, and EATFISH). The experiments were carried out on gilthead sea bream (*Sparus aurata*) under specific experimental conditions such as changes in genetic background, diet composition, or feed additives supplementation, among others. To define how and which biotic and abiotic factors modulate the fish microbiota, and which causal relationships exist within microbial taxa, we set different goals in line with the potential of SAMBA. In a first approach, we combined the taxa abundances within one specific experiment with BN, identifying the positive and negative relationships among the most representative bacteria. In a second approach, we filtered, from the data of already published experiments, those taxa present in both trials that took part of the core microbiota in at least one experiment. The counts of the remaining taxa were introduced in SAMBA, and we built two separated models, aiming to detect common microbes' causal relationships occurring in multiple experiences.

Results and discussion

Both the intra-experiment and the inter-experiment approaches allowed us to obtain the hierarchical disposition of the experimental variables and the taxa within the data populations (Figure 1). Therefore, SAMBA tool can constitute a real step forward microbiomic studies, as it can take advantages from the taxa abundances to deeper understand the role and influence of each OTUs in the gut of livestock species. The directed acyclic graph (DAG) created in the intra-experiment approach (Figure 1a) disclosed the taxa primarily influenced by the experimental variables and the effect of these over other members of the microbial population. Interestingly, our results showed

that the taxa affecting other taxa (i.e., parent taxa) are not always the most abundant, which are the ones usually reported in current 16S analyses. Regarding the inter-experiment approach, we obtained a total of 13 relationships present in both models (Figure 1b). From these results, it is feasible to draw a common structure, identifying which interconnections, shared by the two different experiments, and belonging to the core microbiota taxa, remain unaltered or are strictly linked with the variables that characterize the experiments. In addition, with the SAMBA implementation of the pipeline to infer metagenomes using PICRUSt2 (available with Metacyc and KEGG protocols), it is also possible to correlate the functional metabolic profiles of those taxa to better define their role in the specific relationships and in the total framework of the core microbiota (data not shown).

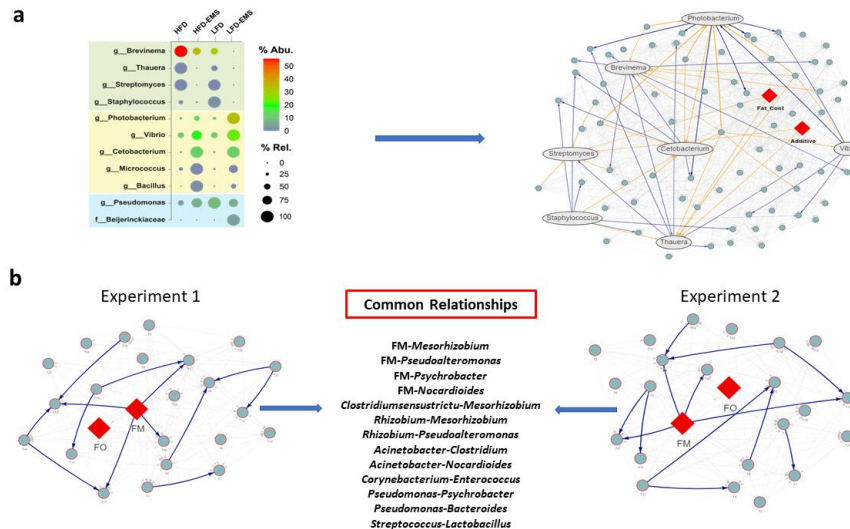


Figure 1. **a** Graphical representation of the relationships of the most abundant OTUs within the experiments (blue arrows= directional relationships from the OTUs, orange arrows= directional relationships to the OTUs, red diamonds=Experimental variables). **b** Graphical representation of the common relationships, shared between two different experiments (blue arrows=shared relationships, red diamonds=Experimental variables, dots=OTUs).

Concluding remarks

The implementation of SAMBA arises as an innovative approach, not yet exploited for aquaculture data (Ruiz-Pérez et al., 2021), to offer researchers relevant information beyond taxa abundances comparison when working with 16S metagenomics datasets. Although the tool is still being trained, when a sufficient amount of shared variables (i.e. mutual taxa, core microbiota) will be available, SAMBA will predict matching information of several experiments, discerning common associations between them that can be related with the experimental variables. In fact, future experimental designs are expected to be adapted to feed SAMBA. Moreover, as this tool is a constantly evolving project, SAMBA will soon include the integration of machine learning algorithms and new interfaces for different omics data, to be able to process complex and integrated results in an easy-to-use package.

Funding. This work was supported by MCIN with funding from European Union NextGenerationEU (PRTR-C17.I1) and by Generalitat Valenciana (THINKINAZUL/2021/024) to JP-S.

References

- Naya-Català, F. et al. (2022); *Biology* 11: 1744.
 Soriano B. et al. (2022); *bioRxiv* doi: 10.1101/2022.12.30.522281.
 Yajima, D. et al. (2023); *Microbiome* 11:53.
 Ruiz-Pérez D. et al. (2021); *MSystems* 30: e01105-20.