

AmpFlow: a containerized pipeline to assist in Reproducible and Replicable Microbiome research

David Aciole Barbosa, Fabiano Menegidio, Yara Natércia Lima Faustino de Maria,
Rafael dos Santos Gonçalves, Marcos Vinicius Yano, Regina Costa de Oliveira,
Daniela L. Jabes, Luiz R. Nunes

UMC

Abstract

The increasing number of studies aimed at evaluating microorganism populations is providing vast and detailed information about a large variety of environments. Such microbiome data represent one of the most promising approaches currently available in biological sciences, with possible applications in industry, agriculture and health. However, technical obstacles still must be overcome, before results from microbiome analyses can be fully incorporated into innovative technologies. In fact, scientists around the world claim that many fields of research are currently affected by a reproducibility challenge, due to difficulties in obtaining all details and resources necessary to reproduce the same experiment/analysis in different laboratories. Microbiome analyses seem to be particularly affected by such reproducibility crisis and the lack of proper standardization for the complex bioinformatics procedures, inherent to such analyses, seems to be one of the main reasons for such problems. A new trend to solve this issue is Docker, a system-independent technology which made possible the packaging of complete software environments by creating additional layers of operating system level virtualization abstraction bundled in the so called containers. In this sense, we present AmpFlow, a containerized pipeline designed to promote reproducible and replicable microbiome analyses. AmpFlow was built in Docker, using a set of simple, yet effective scripts, to deploy tools available from Qiime, which perform: (i) quality checking; (ii) pre-processing and (iii) processing of raw bacterial and fungal sequencing data, creating reproducible OTU tables that are ready to be used in different post-processing platforms, such as the online and R versions of Microbiome Analyst, as well as the Galaxy version of LEfSe.

Funding: FAPESP, CAPES, CNPQ