de RSE CONFERENCE WÜRZBURG 2024

Contribution ID: **122** Type: **Demo**

FuncScan: A pipeline to mine DNA for antimicrobial peptide genes, antibiotic resistant genes, and biosynthetic gene clusters under one umbrella.

Over the past decade, major advances in the field of microbial biotechnology, in particular (eco)genomics have been observed. Because of the increase in genetic data produced at high throughput by next generation sequencing technologies, many bioinformatic tools for their *de novo* and reference-based assembly were and are continually being developed. In order to meet the strong demand for *accurate* and *ideal* assembly into DNA fragments, *i.e.* contigs, many standardized analysis pipelines have been created. This is even more complex in the case where these contigs must be compiled within metagenomes.

However, the diverse functions and biosynthetic capabilities of these ever-growing biomes remain largely elusive. The study of this functional diversity, however, is of utmost importance as it plays key roles in cell signaling, communication, microbial evolution, and defense. Many excellent software tools have been developed to predict antimicrobial resistance genes (ARGs), biosythetic gene clusters (BGCs) and antimicrobial peptide genes (AMPs). However, the use of such standalone tools requires a high minimum level of command line expertise from researchers in the fields of biology and chemistry. Many of these tools are written in different coding languages, lack reproducibility due to differing computing environments required and the results obtained are generated in different formats. This demands a steep learning curve for the researcher, eventually reducing time efficiency prior to downstream processing in the lab.

Here, we address this problem by creating a pipeline called **FuncScan** that can screen nucleotide sequences for functional genes. **FuncScan** is built using a specialized workflow management system called nf-core, which supports FAIR (findable, accessible, interoperable, and reusable) research methods. The pipeline consists of three autonomous workflows that can detect ARGS, BGCs, and AMPs, which can be activated using one command. Each workflow combines different stand-alone tools that are compiled in independent software packages to resolve different execution logics and software dependencies. The standardized result of these workflows is then parsed and further filtered by newly developed parsing tools and visualized using associated Application Programming Interface (API). This allows for the easy comparison of the candidate genes predicted and removes all potential spurious or false positive genes. Therefore, FuncScan exponentially reduces the time and effort required by researchers in the natural product discovery field to identify genes that may be suitable for downstream processing in the lab.

Here, we use ancient DNA extracted from historical dental plaque as case study to explore the presence of *true* antimicrobial peptide genes that can be processed in the lab to generate potentially new AMPs.

Slot length

Workshop (1h)

Primary author: IBRAHIM, Anan (Paleobiotechnology Department, Leibniz-Institut für Naturstoff-Forschung und Infektionsbiologie e. V. Hans-Knöll-Institut, Jena)

Presenter: IBRAHIM, Anan (Paleobiotechnology Department, Leibniz-Institut für Naturstoff-Forschung und Infektionsbiologie e. V. Hans-Knöll-Institut, Jena)

Session Classification: Poster Session

Track Classification: Research Software (legacy): Computational Workflows