

MADAME, a user-friendly and effective data and metadata retrieval tool

Soletta G.¹, Fumagalli S.¹, Agostinetto G.¹, Bruno A.¹, Labra M.¹

E-mail: g.soletta@campus.unimib.it

¹ Department of Biotechnology and Biosciences, University of Milano-Bicocca, Milan (Italy)

Keywords: microbiome data, metadata, ENA, bioinformatics, omics data

Abstract: Publicly available repositories, as the European Nucleotide Archive (ENA), provide huge amounts of microbiome-related data derived from high-throughput DNA sequencing studies. These rich datasets are a fundamental resource for meta-analyses and secondary analyses, leading to novel discoveries, data reuse approaches and data mining applications.

MADAME (MetADAta MicrobiomE) is a currently in development bioinformatic tool designed to automate the process of data and metadata retrieval, enhancing microbiome re-analyses strategies. The tool can be run via command-line interface and will guide users through the pipeline, so it does not require for the user to have specific programming skills. This easy-to-use framework downloads data in a standardized format, according to FAIR principles, for improving the reproducibility of the analyses.

MADAME is entirely written in Python and implements a set of scripts for simplified programmatic access to ENA repository.

It comprehends five main parts: (1) “Accessions”, for performing a free-text query in ENA’s database or, alternatively, accepting a user’s accessions list; (2) “Metadata”, for metadata retrieval, considering both project and sample metadata; (3) “Report”, for providing the user with informative reports about the identified sequencing data, allowing a sensible choice before the actual download of data; (4) “Sequencing data”, for downloading sequencing data in the preferred format; (5) “Publications and annotations”, for finding associated publications and annotations on Europe PMC, since, unfortunately, publications are not often mentioned in the projects’ main page on the repository or in their metadata, and they are crucial for secondary analyses.

The SKIOME Project was used as a case study for testing MADAME, starting from its manually curated list of projects.

Looking ahead, our tool may be included into larger workflows for processing, analyzing, and visualizing microbiome data, making the first steps of data and metadata retrieval accessible for all types of users.