The detection and characterization of emerging infectious agents has been a continuing public health concern. High throughput sequencing (or Next-Generation Sequencing; NGS) technologies have proven promising for unbiased detections of pathogens in complex biological samples. They are efficient and provide comprehensive analyses. However, NGS yields millions of putatively representative reads per sample, such an amount, that efficient data management and visualization resource have become mandatory requirement, through a dedicated Laboratory Information Management System (LIMS), solely to provide perspective regarding the information contained in this huge amount of data. We developed a managing and analytical bioinformatics framework that is engineered to run associated and dedicated Galaxy[1] workflows for the detection and eventually classification of pathogens. In essence, our primary purpose is to assist the biologist in the process of deciding on the most relevant sample-specific sequences in the supplied samples, and determine their relative abundance. To this end, a user-friendly interface is essential. A complete set of specific Galaxy pipelines, producing high quality reads and/or assemblies meaningful for biological interpretation, have been engineered, and serve as the driving engine for a graphical web-interface associating the sample’s metadata and its analysis results. This user-interface has been tailored to associate Galaxy’s restful API providers resources (a Galaxy instance, sufficient storage and grid computing power), with the input data and its metadata. Hence, the web application allows scientists to easily interact with existing a Galaxy metagenomic workflows, facilitates the organization, visualization and aggregation of the most significant and most meaningful bits of information from millions of genomic sequences. In more detail, communication between our Django-based interface [2] and Galaxy uses the Bioblend library[3]. It gives access to a Galaxy primary purpose is the information contained in this huge amount of data.

**Bioinformatics for HTS data**

Systematic bioinformatic approaches are in use to analyse HTS data. High throughput sequencing (or Next-Generation Sequencing; NGS) technologies have proven promising in the detection and characterization of emerging infectious agents. HTS data analyses are: • Mapping • Alignment • Blast • HTS data analyses... • De novo Assembly • Mapping • Blast • Blast

**Architecture**

Interactive visualization systems allow scientists to intuitively navigate and focus on the most pertinent results that might correspond to a family of microorganisms. Through APIs, in the analyse part of MetaGenSense, the KRONA pie chart is available. The scientific will be able to identify specific agents detected through the workflow. If the analysis in terminated, the user can choose to download the results, or share them with other users logged on MetaGenSense.

**References**


**Acknowledgments**

Groupe Exploitation et Infrastructure (E&I), Institut Pasteur, Paris, France.

**LIMS**

A LIMS is an application to manage labaratory data. It is mainly used to track biological samples and associated metadata by recording them in a specific refined database. Data can be recorded using a client interface.

**Conclusion**

MetaGenSense helps to automate workflows from Galaxy, make biologists unfamiliar with designing workflows to use the Galaxy interface, and quickly obtain analysis results from HTS sequencing projects. It uses Galaxy as a workflow management software and bioblend API to remotely manage the data upload, the workflow launching as well as the results analysis. Visualization of end-results is an important component, and is subject to further developments.