**Microbiome profiling on a Galaxy-based framework for microbiology**

Stefano Onano, Gianmauro Cuccuru, Massimiliano Orsini, Andrea Pinna, Andrea Sbardellati, Nicola Soranzo, Paolo Uva, Giorgio Fotia

CRS4, Bioinformatics Research Program
Loc. Piscina Manna, 09010 Pula (CA), Italy

---

**Abstract**

Microbes are essential for many biological processes in the human body, especially in digestion. **Microbiome composition** has been strongly related to different health status or pathologies, like dental caries, psoriasis, metabolic disorders, Crohn’s disease, and colorectal carcinoma.

Currently, NGS technology allows the deep investigation of microbial communities, thus helping the elucidation of the role of each microbiome component. Bioinformatics plays a central role in the NGS data analysis, and this holds particularly true in metagenomics, where millions of sequences are aligned against thousands of genomes, and different algorithms or settings can lead to dissimilar results.

Several tools and pipelines have been recently developed, focused on different aspects of the analysis, from the pre-processing to the reads binning and the community composition reconstruction.

In order to create an environment for metagenomics analysis and to allow data and results sharing among collaborators, we implemented several metagenomics tools into Orione (http://orione.crs4.it), our local platform for genome analysis.

---

**In vitro microbiome**

Orione is a Galaxy-based web server which provides an integrated set of state-of-the-art tools allowing end-to-end NGS data analysis from bacterial resequencing, de novo assembly, scaffolding, bacterial RNA-Seq, gene annotation and metagenomics.

Many of these **wrapper tools** have been developed in-house (BLAT, SSKAE, Edena, SPACE, SOPRA, SEQuel, Mupgy, MUSCLE, Glimmer, Prokka, EDGE-pro, READSCAN and MetaPhlAn-to-Krona). The complete set of Orione new tools will be available soon through the main Galaxy Tool Shed (http://tools.shed.g2.bx.psu.edu).

Orione is backed with the CRSS4 High Performance Computing cluster to support large-scale data processing and storage in a scalable manner, and is fully integrated with a high-throughput NGS facility.

---

**Human microbiome**

The NIH Human Microbiome Project (HMP) collected and sequenced samples from microbial communities found at several different sites on the human body. We downloaded from http://www.hmpdacc.org/HMASM/ the Illumina reads of 26 samples for various body sites.

The heatmaps show the genus- (top) and species-level (bottom) taxonomic profiles of the most abundant groups obtained with different methodologies, hierarchically clustered (average linkage) with the Bray-Curtis similarity.

We only included species and genera with abundance > 1% at the 95th percentile of their distribution.

---

**Metagenomics workflow**

We analyzed two in vitro microbiomes containing 7 species with different abundances. Each sample was processed in Galaxy with the metagenomic workflows illustrated below.

Figures show the rank of the known species obtained with the different approaches and the combined rank.

---

**Conclusion**

We compared several approaches for the analysis of microbiomes using the Orione Galaxy server (http://orione.crs4.it):

- using in vitro microbiomes, most of the approaches reported the 7 species in the top 10 of the most abundant species;
- the analysis of the human microbiome revealed that the taxonomic profiles naturally clusters by body site, but large differences in the species found by the various tools remain.

We took advantage of the Galaxy framework for the management and analysis of metagenomics data to give users the possibility to directly run and compare results from different tools, build complex analysis workflows and share results among colleagues.

---

**References**


---

We thank Porto Conte Ricerche (Alghero, Italy) for providing live-in vitro microbiomes.