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

In vitro microbiome

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.

Orione.crs4.it

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, SSAKE, Edena, SSPACE, SOPRA, SEQuel, Mugsy, MUSCLE, Glimmer, Prokka, EDGE-pro, READSCAN and MetaPhIAn-to-Krona). The complete set of Orione new tools will be available soon through the main Galaxy Tool Shed (http://toolshed.g2.bx.psu.edu).

Orione is backed with the CRS4 High Performance Computing cluster to support large-scale data processing and storage in a scalable manner, and is fully integrated with a highthroughput NGS facility.



Metagenomics workflow



- We analyzed two *in vitro* microbiomes containing 7 species with different
- 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.



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.





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

- . Soranzo, N., Onano, S., Cuccuru, G., Orsini, M., Pinna, A., Sbardellati, A., Uva, P., Fotia, G. Orione: a web-based framework for NGS analysis in microbiology. In preparation, 2013
- 2. Goecks, J., Nekrutenko, A., Taylor, J., The Galaxy Team. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* 11(8), R86, 2010
- 3. The Human Microbiome Project Consortium. A framework for human microbiome research. *Nature* 486(7402), 215-221, 2012
- 4. Naeem, R., Rashid, M., Pain, A. READSCAN: a fast and scalable pathogen discovery program with accurate genome relative abundance estimation. *Bioinformatics* 29(3), 391-392, 2013
- 5. Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., Huttenhower, C. Metagenomic microbial community profiling using unique cladespecific marker genes. Nat. Methods 9(8), 811-814, 2012
- 6. Kosakovsky Pond, S., Wadhawan, S., Chiaromonte, F., Ananda, G., Chung, W. Y., Taylor, J., Nekrutenko, A., The Galaxy Team. Windshield splatter analysis with the Galaxy metagenomic pipeline. Genome Res. 19(11), 2144-2153, 2009

We thank Porto Conte Ricerche (Alghero, Italy) for providing the in vitro microbiomes