InterOmics Tutorial - Tools and methods for the analysis of omics data and biodiversity


Angelica Tulipano✉, **Andreas Gisel**
CNR, Institute for Biomedical Technologies, Bari, Italy

Received 11 March 2014; **Published** 18 March 2014

Tulipano A and Gisel A (2014) *EMBnet.journal* **20**, e759. http://dx.doi.org/10.14806/ej.20.0.759

The CNR Institute for Biomedical Technologies (ITB)[1] in Bari (IT), with support from the Italian *Flagship project InterOmics*[2], organised a *Tutorial-Day*[3] as a satellite event of the *BiP-Day 2013* workshop (see related article in the present volume). The tutorial was organised in three 3-hour events, covering metagenomics, phylogenetics and data analysis of non-coding RNA. The event took place on 6 December 2013 at the Department of Physics *Michelangelo Merlin* of the University of Bari and INFN[4], as their computing infrastructure was used to guarantee the required performance for such data analysis approaches. While the services required for the first two sessions were already hosted on the INFN computer infrastructure, the third session was run on Virtual Machines (VMs). Four VMs with the analysis pipeline pre-installed, each with 16 CPU and 200GB shared memory, were used to serve 40 participants. Giacinto Donvito, from Bari University's Department of Physics continuously monitored the infrastructure to guarantee a flawless service.

The morning session started with a tutorial on the *Classification and quantification of the mi-*

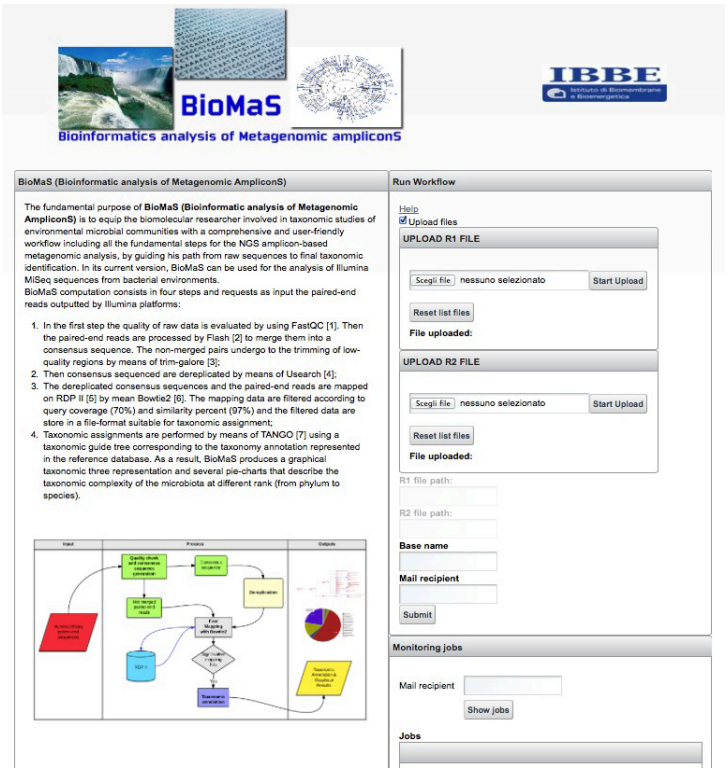

Figure 1. Screenshot from the BioMaS website.

1 www.itb.cnr.it
2 www.interomics.eu
3 www.ba.itb.cnr.it/bip-day/tutorial
4 https://www.ba.infn.it/

crobiome using metagenomic amplicons. Bruno Fosso, from the Department of Biotechnology and Biopharmaceutical Biosciences of the University of Bari (IT), and Monica Santamaria, from the CNR Institute of Biomembranes and Bioenergetics (IBBE)[5], presented a modular pipeline (BioMaS) using third-party tools and ad hoc python and bash scripts. BioMaS is a web-service on the INFN/UNIBA infrastructure (Figure 1). High-level SaaS (Software as a Service) services are applied to facilitate the use of BioMaS components that are already suitably configured and optimised to run on the dedicated infrastructure. BioMaS allows the analysis of both bacterial and fungal environments, and alternative paths can be selected to process data obtained either by Roche 454 or Illumina sequencing technology. The tutorial allowed participants to run a test data-set and understand in detail the pipeline and its functionality.

The second session covered *Instruments for the phylogenetic analysis for studies of biodiversity*. Saverio Vicario, from the CNR – ITB, and Bachir Balech, from the CNR – IBBE, presented the BioVeL[6] infrastructure. This allows users to build customised workflows (Figure 2) by selecting and applying successive 'services', or re-using existing workflows available from BioVeL's library. By giving participants the opportunity to process specially provided test data, the tutorial offered profound insights into BioVeL's significant functionality and performance.

The third session introduced participants to the world of non-coding RNA, in *Mapping and*

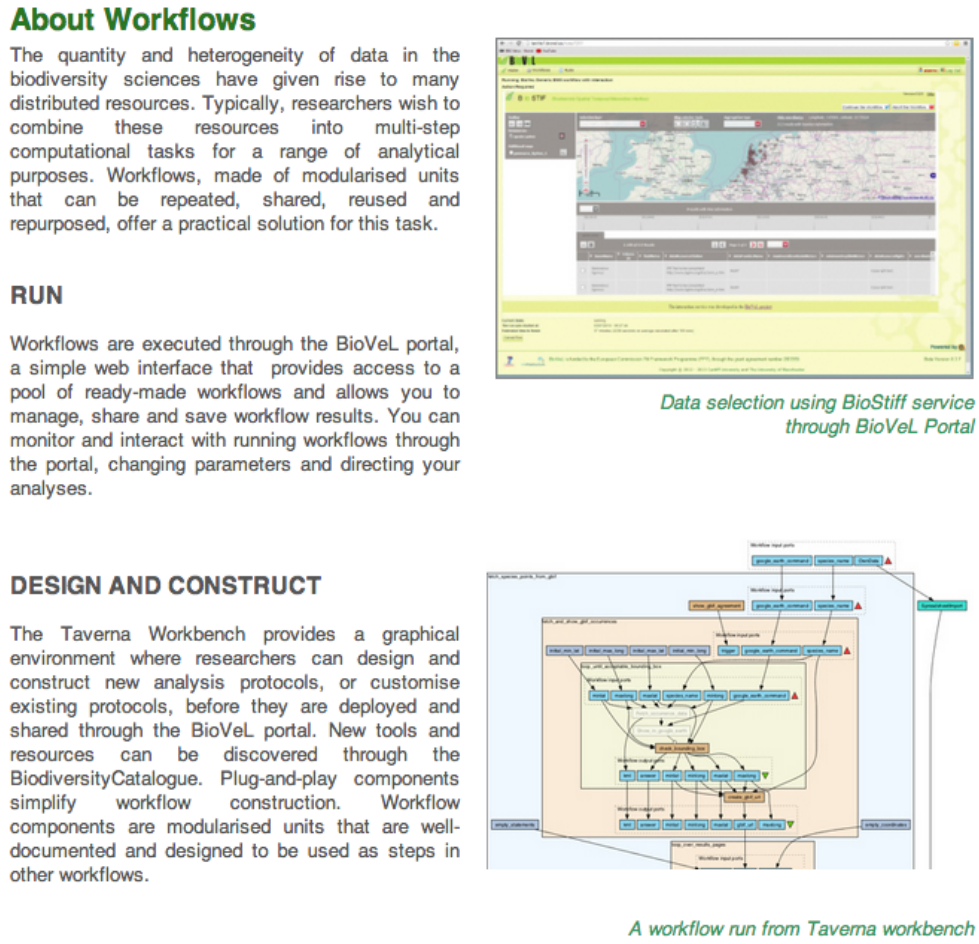

Figure 2. Screenshot from the BioVeL website illustrating its underlying workflows.

5 www.ibbe.cnr.it/

6 www.biovel.org

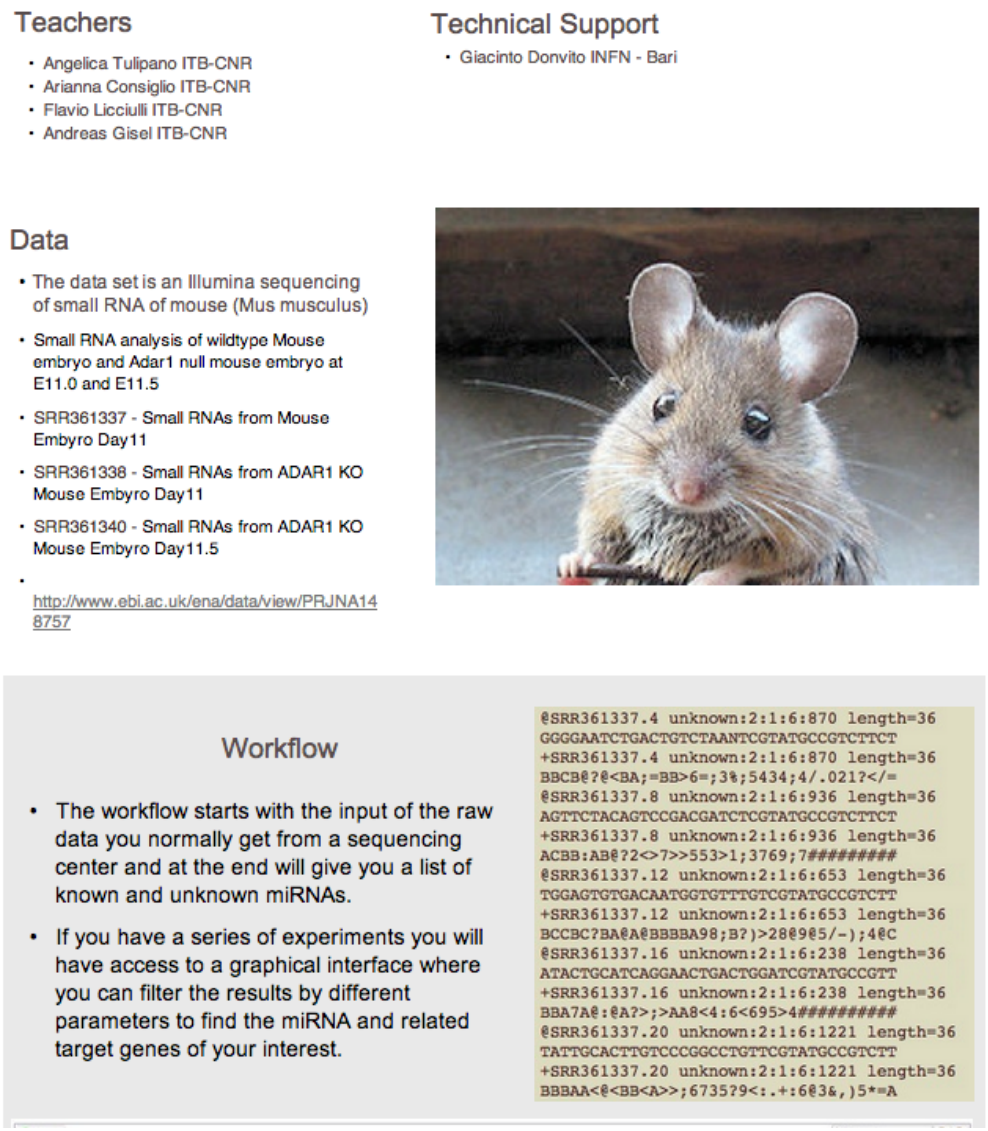

Figure 3. Screenshot from the ncRNA data-analysis website.

analysis of non-coding RNAs and small RNAs from NGS technologies. The specialist team from ITB Bari – Angelica Tulipano, Flavio Licciulli, Arianna Consiglio and Andreas Gisel – demonstrated a simple workflow to get from raw sequencing data to an expression profile of known and unknown miRNA and other non-coding RNAs. The workflow is based on publicly available software and in-house Perl scripts, assembled into a user-friendly pipeline. The results can be uploaded into a MySQL database with a simple graphical interface to visualise, sort and filter the data. A customised data-set of Illumina small RNA sequences, at three time points (Figure 3) , was provided to give the users first-hand experience of the pipeline's functionality.

More than 110 participants (on average 35 per session) attended the tutorials, demonstrating the urgent need for such events to help train life scientists to cope with the large and complex data-sets produced by NGS technologies.