

BITS :: Call for Abstracts 2021 - Poster

<i>Type</i>	Poster
<i>Session</i>	Biological Databases
<i>Title</i>	ITSoneWB: an integrated environment for eukaryotic communities taxonomic profiling on Galaxy.
<i>All Authors</i>	Tangaro M(1), Defazio G(2), Fosso B(1), Licciulli VF(3), Donvito G(4), Pesole G(1,2) and Santamaria M(1)

Affiliation

(1)Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies, National Research Council, Bari

(2)Department of Biosciences, Biotechnology and Biopharmaceutics, University of Bari 'A. Moro', Bari

(3)Institute of Biomedical Technologies, National Research Council, Bari

(4)National Institute for Nuclear Physics (INFN), Bari

Motivation

In amplicon targeted metagenomic analysis (here referred as DNA metabarcoding), taxonomic variants of selected genetic markers from environmental samples are explored based on High-Throughput Sequencing (HTS) technologies. This allows to profile the global composition of biotic communities in a rapid, massive and cheap way. After its great success in prokaryotic microbiomes investigation, a growing number of applications are now focusing on eukaryotic communities. This is producing a great effort, in the framework of the mayor bioinformatics infrastructures, to develop increasingly comprehensive, well-controlled and FAIR-compliant (Wilkinson et al., 2016) reference databases targeted at Eukaryotes and well linked with taxonomy annotation tools. ITSoneWB (ITSone WorkBench) is a Galaxy-based bioinformatic environment targeted at automated and easy analysis of eukaryotic communities biodiversity based on Internal Transcribed Spacer 1 variants HTS integrating comprehensive and high-quality reference data with established pipelines and new tools.

Methods

The ITSoneWB Galaxy virtual instance was deployed on the INFN-Bari ReCaS cloud facility. Finally, a Galaxy virtual cluster was deployed to develop the workbench, exploiting the resources of the INFN-Bari cloud site, consisting in a front-end server running Galaxy and the slurm server with 2 Worker Nodes (WN), each fit out with 8 CPUs and 16 Gb RAM and running CentOS 7. A comprehensive description of Galaxy environment deployment is available in Tangaro M.A. et al. 2020.

The same querying options available in ITSoneDB are available in ITSoneWB, in order to retrieve data by accession number (i.e. by using ENA accession number) and/or by taxon name (e.g. species name). The deployment of metabarcoding data analysis pipelines was obtained through the Galaxy ShedTools (QIIME (Caporaso et al., 2010), QIIME2 (Bolyen et al., 2019, 2) and Mothur (Schloss et al., 2009) or by exploiting an ad-hoc designed wrapper for the BioMaS workflow (Fosso et al., 2015). The Mopo16S (Sambo et al. 2018) tool was deployed and optimized to exploit primer design by using the ITSoneDB content as reference. Finally, a tool for the barcoding-gap inference was de-novo developed and deployed on the ITSoneWB. The whole documentation is available at: <https://github.com/ibiom-cnr/itsonewb/wiki>.

Results

In the framework of ELIXIR ESFRI project, we developed the ITSone WorkBench (ITSoneWB). It is a free and easy-to-use service where the Internal Transcribed Spacer 1 (ITS1) reference database ITSoneDB and DNA metabarcoding tools are integrated in customized and automated workflows addressing the entire eukaryotic domain of life. Actually, ITSoneWB provides the first bioinformatic environment specifically targeted at this promising barcode marker and is centered on ITSoneDB (Santamaria et al., 2018), to our knowledge the first and unique controlled and taxonomically referenced specialized collection of eukaryotic ITS1 sequences. ITSoneDB current version hosts 1,174,761 ITS1 sequences spanning 157,531 eukaryotic species and its reference dataset has been also recently made public in the ENA Browser under accession PRJEB33030. The virtual environment, deployed on the ReCaS-Bari cloud facility thus supplying enough computational power and suitable scalability, hosts established sequence-based taxonomic assignment pipelines, such as BioMaS (Fosso et al., 2015), QIIME (Caporaso et al., 2010), QIIME2 (Bolyen et al., 2019, 2) and Mothur (Schloss et al., 2009) and new services targeting some of the most common and challenging issues of metabarcoding experiments, such as the design of effective universal primers and the evaluation of the barcoding gap in customized taxonomic ranges. Finally, the user is provided with the ITSoneDB connector, to query, cross-referencing and downloading ITS1 data and metadata to feed his own bioinformatic workflow.

Info

<https://github.com/ibiom-cnr/itsnewb/tree/master/galaxy-deployment>

Figure

-

Availability <http://itsnewb.cloud.ba.infn.it/galaxy>

Corresponding Author

Name, Surname Monica, Santamaria

Email m.santameria@ibiom.cnr.it

Submitted on 09.05.2021

Società Italiana di Bioinformatica

C.F. / P.IVA 97319460586

E-mail bits@bioinformatics.it

Sede legale Viale G. Mazzini, 114/B - 00195 Roma

Website bioinformatics.it

message generated by sciencedev.com for bioinformatics.it 20:29:46 09.05.2021
