

Genetics e-services virtual interface

The Genetics virtual laboratory (Geneticsvlab) is now available through the LifeWatch portal. This is an extension to the Genetics e-services, which are available via command line (SSH) access directly to the HCMR PC cluster. This novel virtual environment is designed to be a user friendly graphical interface for accessing QIIME (QIIME Quantitative Insights into Microbial Ecology). QIIME is an open-source bioinformatics pipeline for performing microbiome analysis from raw DNA sequencing data. QIIME is designed to take users from raw sequencing data generated on the Illumina or other platforms to publication quality graphics and statistics. This includes demultiplexing and quality filtering, OTU picking, taxonomic assignment, and phylogenetic reconstruction, and diversity analyses and visualizations. QIIME has been applied to studies based on billions of sequences from tens of thousands of samples. Geneticsvlab provides efficiency and ease in execution of QIIME functions. This interface is a combination of web development languages like HTML and PHP and QIIME scripts, implemented in PYTHON). It is available through the LifeWatch portal, thus maintaining a link to the PC cluster at HCMR. The production version of the Geneticsvlab is now available here (<https://portal.lifewatchgreece.eu/>) and is subject to registration and login to the LifeWarch portal. On the main page users can find links to a comprehensive and self-explainable online QIIME tutorial describing basic QIIME routines and reasoning.

Some basic functionalities as well as some screen shots from the application are shown below.

The screenshot displays the 'Genetic Services' web interface. At the top, a header bar contains the title 'Genetic Services' and a brief description: 'This is a web interface to QIIME (Quantitative insights in Microbial Ecology) which is running on a PC cluster part of the LifeWatch Genetics services infrastructure at the Institute of Marine Biology, Biotechnology and Aquaculture (IMBBC), Hellenic Centre for Marine Research (HCMR), Heraklion, Greece.' Below the header, there are two main sections. On the left, a 'Recent Jobs' table shows a single entry with a 'Submitted At' timestamp of '2015-11-19 11:57:10'. To the right of this table is a 'Submit a new Job' button. Below the 'Recent Jobs' table, there are four service cards: 'Literature Mining', 'Data Services', 'MicroCT Services', and 'Genetic Services'. Each card provides a brief description of the service and its availability status. An arrow points from the 'Genetic Services' card to the 'Submit a new Job' form on the right. The form includes fields for 'Input files' (mapping_file_test.txt, sequences_test.fna), 'Parameters' (Check to demultiplex data, Check to denoise data, split_libraries.py, --min_seq_length, --max_seq_length, --barcode_type), and a 'Submit' button.

Genetic Services

This is a web interface to QIIME (Quantitative insights in Microbial Ecology) which is running on a PC cluster part of the LifeWatch Genetics services infrastructure at the Institute of Marine Biology, Biotechnology and Aquaculture (IMBBC), Hellenic Centre for Marine Research (HCMR), Heraklion, Greece.

Workspace File Management Help

Recent Jobs:

Submitted At
2015-11-19 11:57:10

Submit a new Job

Input files

Select FASTA sequences from loaded files

- mapping_file_test.txt
- sequences_test.fna

Select mapping file

- mapping_file_test.txt
- sequences_test.fna

Parameters

- ☒ Check to demultiplex data
- ☐ Check to denoise data
- split_libraries.py (Demultiplex and quality filter reads)
- min_seq_length 200
- max_seq_length 1000
- barcode_type golay_12
- denoiser.py (Denoise Seqs with Denoiser) & inflate_denoiser_output.py (Inflate denoiser output)

Literature Mining

Biodiversity literature and data constitute a vast public resource open to mining and knowledge extraction. Associating organisms to key features of their life e.g. the environment in which they live, the way they feed, their breeding habits, is cornerstone in explaining biodiversity patterns and informing ecological decisions. Eco-Systems Biology, and in particular network based analysis, can provide holistic pictures of such associations, highlight novel relations and support hypothesis formulation and knowledge discovery. Initial aim of this vLab is to augment species related information based on data available in global biodiversity knowledge and literature aggregators, such as the Encyclopedia of Life (EOL) and the Biodiversity Heritage Library (BHL). Main focus of this virtual lab is the extraction of species - traits associations starting with the environment in which occur. Species and environments associations will be extracted by mining relevant text field clauses of: a) In-house LifewatchGreece data, b) the EOL, and the BHL text collection. Also, interactive web-based visualizations will be developed to summarise the extracted species - environments association and support data exploration and landscape ecology studies.

Available without Sign in

Data Services

Data Services provide the users with tools in order to: a) publish their datasets and make them available to the community by providing information that allows a user to locate and access the resource and its curator/creator, b) import their datasets to the Lifewatch Greece Infrastructure and to GBIF or MedOBIS, c) perform biodiversity data and information quality improvement, and d) search about datasets of interest by providing an efficient way of querying semantic networks. The schema of the data that is provided by the users is mapped to the semantic model of the LVI and the data is transformed to LVI format before it is stored to the Infrastructure. The semantic model is based on CIDOC CRM (<http://www.cidoc-crm.org/>), CRM dig, CRM geo, CRM sci and MarineTLO (<http://www.ics.forth.gr/isl/MarineTLO/>).

Available after Sign in

MicroCT Services

Micro-tomography (micro-computed tomography or microCT) is a method of non-destructive 3D x-ray microscopy, which allows the users to create 3D models of objects from a series of x-ray projection images, similar to the conventional clinical computer tomography. The MicroCT Service will offer a collection of virtual galleries of taxa which will be displayed and disseminated through a web-based framework, and will allow the user to manipulate the 3D models through a series of online tools or to download the datasets for local manipulations.


Available without Sign in

Genetic Services

The online Services of the Genetics working group are focused on 16S Metagenomics, which involve specific 16S ribosomal marker gene amplification by PCR. The 16S analysis is the most common approach for biodiversity studies. The service focuses on taxonomic analysis from data derived from 454 Roche as well as Illumina sequencing technologies and provides the relevant software to allow analysis for both sequencing technologies. Efficient noise removal for 454 data is accomplished using the AmpliconNoise (<http://qiime.org/scripts/ampliconnoise.html>) pipeline and the taxonomic assignment for both de-noised 454 and Illumina data is achieved via QIIME (<http://qiime.org/index.html>).

Available after Sign in

Figure 1: The Geneticsvlab intro Page – link through LifeWatch portal


Genetic Services

This is a web interface to **QIIME (Quantitative Insights in Microbial Ecology)** which is running on a PC cluster part of the **LifeWatch Genetics services** infrastructure at the Institute of Marine Biology, Biotechnology and Aquaculture (IMBBC), Hellenic Centre for Marine Research (HCMR), Heraklion, Greece.

Workspace File Management

Input Files

mapping_file_test.txt

sequences_test.fna

Delete selected files ✖

Upload new input files

Select file(s)...

Add Files

User's Storage Utilization: (156.00 MB)

3.1%

Recent Jobs:

Job ID	Status	Submitted At	
Job1147	Completed	2015-11-19 11:57:10	

Delete selected jobs ✖

Help

Submit a new Job

Input files

Select FASTA sequences from loaded files

mapping_file_test.txt

sequences_test.fna

Select mapping file

mapping_file_test.txt

sequences_test.fna

Parameters

☒ Check to demultiplex data

☐ Check to denoise data

split_libraries.py (Demultiplex and quality filter reads)

-l, --min_seq_length

200

-L, --max_seq_length

1000

-b, --barcode_type

golay_12

denoiser.py (Denoise Seqs with Denoiser) & inflate_denoiser_output.py (Inflate denoiser output)

Figure 2: The GeneticsvLab main Page

The figure above shows the main interface of the Geneticsvlab web application. Similar to Rvlab the Geneticsvlab comprised of 3 main tabs (Workspace tab, Functions area tab and Jobs submission tab). They allows user to: 1) upload of specific files in predefined formats required for successful execution of the QIIME functions 2) Select desired function and respective function parameters and start executions and 3) View the progress of the submitted job and inspect results. Example datasets are available for the users to obtain a clear idea of the type of data that are supported by the Geneticsvlab. These include commonly used formats generated as output (like FASTA) from next generation sequencing technologies. All jobs are submitted to a queuing system on the HCMR PC cluster and upon completion of the submitted job the user can click on the results “Job ID” link which will direct the user to the results display page.

Parameters

☒ Check to demultiplex data
☐ Check to denoise data

split_libraries.py (Demultiplex and quality filter reads)

-l, --min_seq_length

-L, --max_seq_length

-b, --barcode_type

denoiser.py (Denoise Seqs with Denoiser) & inflate_denoiser_output.py (Inflate denoiser output)

--primer

pick_otus.py (Pick Operational Taxonomic Units)

-m, --otu_picking_method

assign_taxonomy.py (Assigning Taxonomy)

-r, --reference_seqs_fp

-m, --assignment_method

-b, --blast_db

Figure 3: Geneticsvlab function parameters

The user can select from a wide variety parameters to that best suit the dataset under analysis. These include demultiplexing and quality filtering parameters, OTU picking strategies, taxonomic assignment algorithms, phylogenetic reconstruction metrics, diversity analyses measures and different visualizations. (see Figure 3 above)

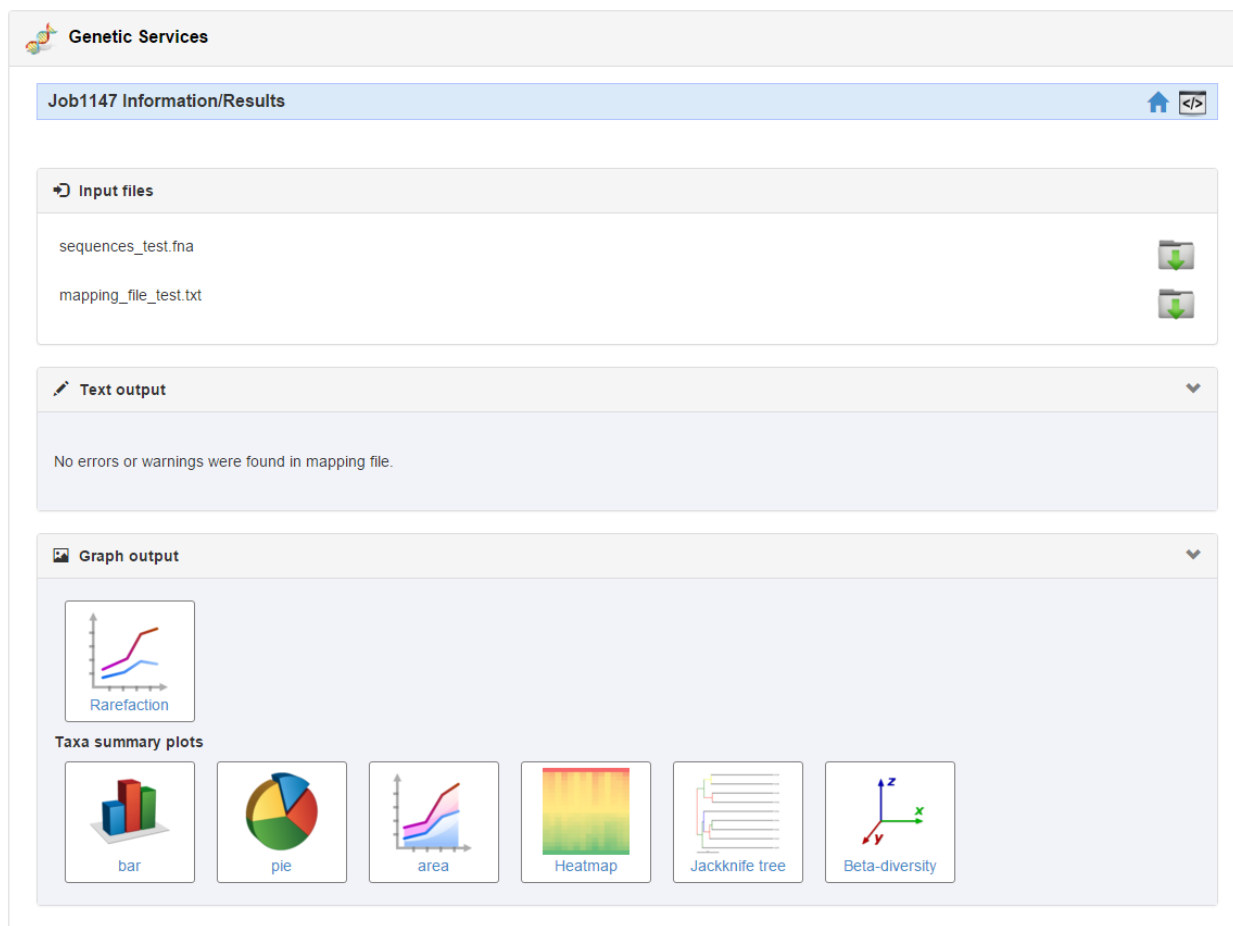


Figure 4: Geneticslab result page

Function results are displayed by using a summary of the results in text form as well as numerous links to charts and plots that provide informative visualizations. As shown above, users may also download results locally to their computer. *Geneticslab* supports multiple interactive web plots where the user can directly interact with the web page for exploring data in greater detail (see below).

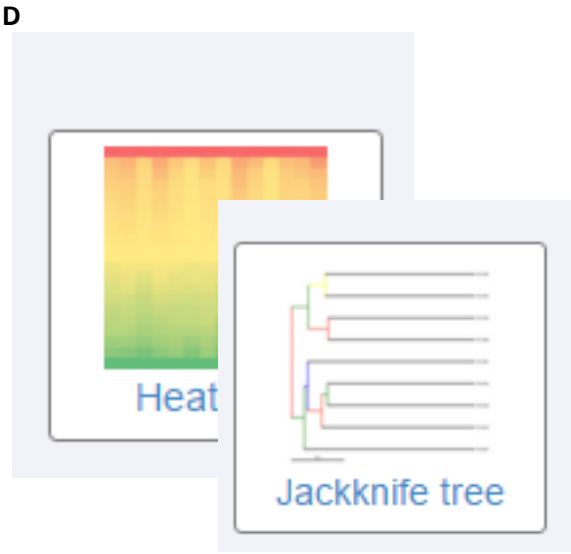
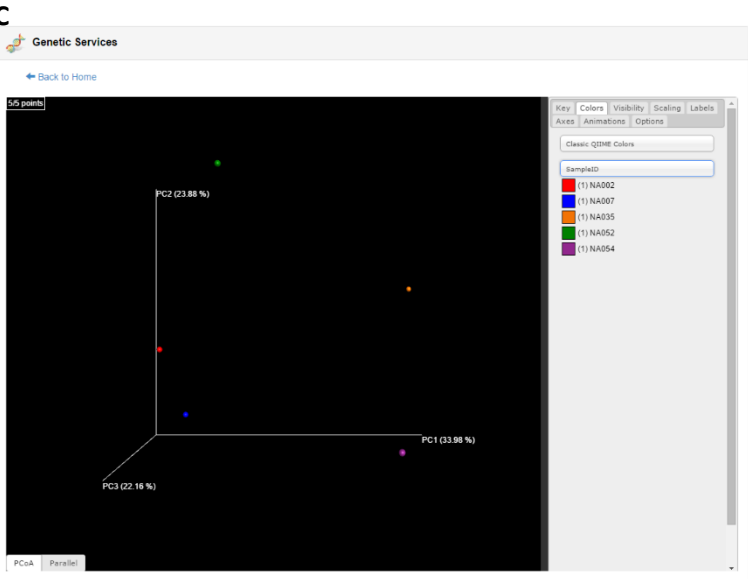
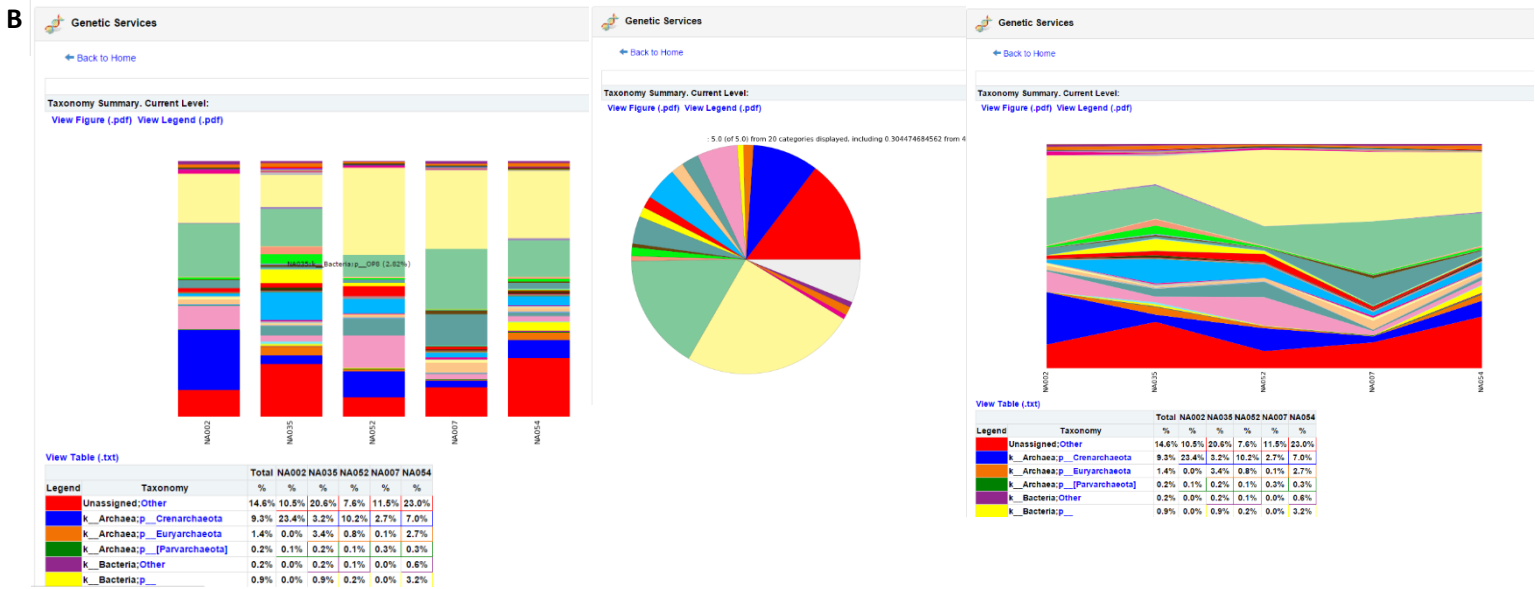
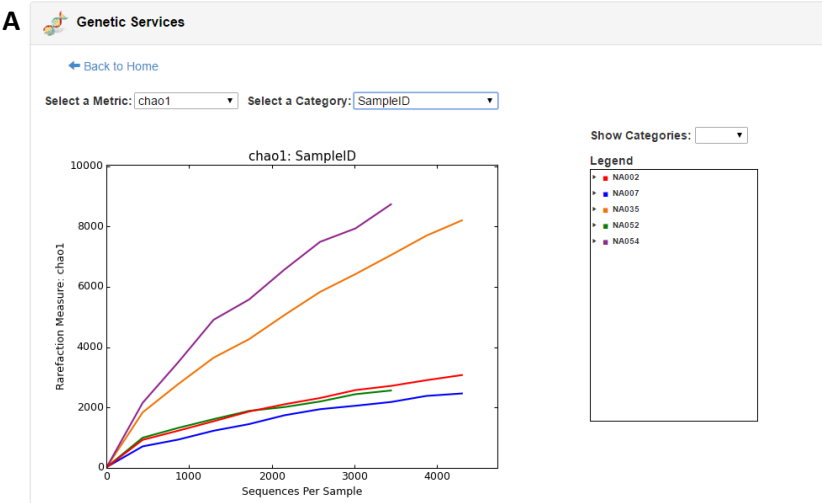


Figure 5: Some Geneticsvlab results. A) Rarefraction plot for evaluation sequencing depth and coverage B) and taxonomic distribution bar, pie and area charts. C) Beta-diversity (Principal Component Analysis), D) Heat map and Phylogenetic Tree visualization

Next Steps

The main emphases regarding our future activities related to the *Geneticsvlab* are centered around two main tasks: the improvement of the web UI and the inclusion of additional functions and parameters for greater user flexibility in performing analyses.

Additionally, the performance evaluation will enable us to understand how the size of input data affects performance, in order to preconfigure resource allocation parameters upon user requests. A variety of use cases with diverse initial parameters have been designed for conducting the evaluation (e.g., number of processors, size of input datasets, size of available primary/secondary memory, number of users).