BioBix
Lab of Bioinformatics and Computational Genomics

PROKAR-SEQ: An analysis and visualization framework for next-generation sequencing based quantification of prokaryotic communities

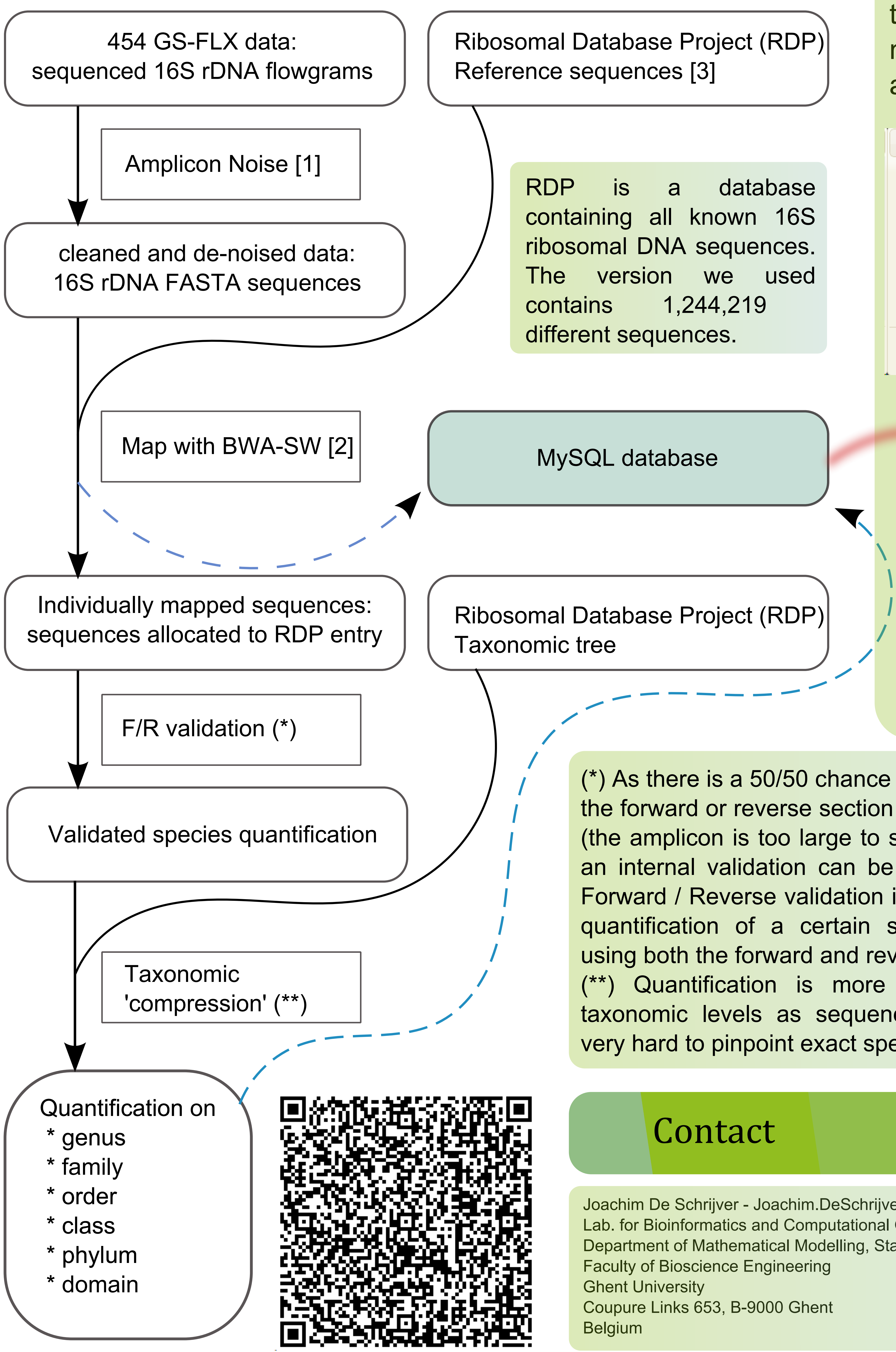
Joachim De Schrijver*[,1], Pieter-Jan Volders[2], Frederiek-Maarten Kerckhof[3], Dagmar Obbels[4], Elie Verleyen[4], Wim Vyverman[4], Tim De Meyer[1] & Wim Van Criekinge[1]

1. Introduction

16S ribosomal (rDNA) base PCR followed by sequencing of these PCR fragments is the preferred method of quantifying prokaryotic microbial communities. Limitations of the sequencing technology (mainly the throughput) were until recently the limiting factor to quantify large amounts of different samples in a high-throughput manner. However, recent development in high-throughput sequencing (454 GS-FLX sequencing) allow large amounts of data to be generated.

PROKAR-SEQ is a modular framework consisting of an analysis module and a visualization module and is backed by a relational MySQL database. The analysis module is written in Perl, the visualization module is written in PHP.

2. Analysis module



454 GS-FLX data:
sequenced 16S rDNA flowgrams

Amplicon Noise [1]

cleaned and de-noised data:
16S rDNA FASTA sequences

Map with BWA-SW [2]

Individually mapped sequences:
sequences allocated to RDP entry

F/R validation (*)

Validated species quantification

Taxonomic 'compression' (**)

Quantification on
* genus
* family
* order
* class
* phylum
* domain

Ribosomal Database Project (RDP)
Reference sequences [3]

RDP is a database containing all known 16S ribosomal DNA sequences. The version we used contains 1,244,219 different sequences.

MySQL database

Ribosomal Database Project (RDP)
Taxonomic tree

(*) As there is a 50/50 chance of sequencing either the forward or reverse section of the 16S amplicon (the amplicon is too large to sequence it entirely), an internal validation can be carried out. F/R or Forward / Reverse validation is a process wherein quantification of a certain species is validated using both the forward and reverse frequency.

(**) Quantification is more reliable on higher taxonomic levels as sequencing errors make it very hard to pinpoint exact species.

3. Visualization module

The analysis results are visualized using an **interactive PHP/ MySQL website**. All data is directly fetched from the database and enables easy visualization independent from the processing step. In this example we also used the Google API to visually locate all the different samples.

Users can select a taxonomic level (genus, family, order, class, phylum or domain) and taxonomic unit (e.g. 'Proteobacteria') using the **taxonomy browser** (Figure 1).

Users can **compare** the taxonomic entry of interest (for example Acidobacteria) across different samples. The abundance is depicted by the size of the green circel (when present) or a red circel (when not present) as shown in Figure 2. This allows a quick interpretation of taxonomic 'clustering'.

The third option is to analyze a certain sample and assess the **taxonomic distribution** of a certain sample (on the desired taxonomic level). Species are ordered by abundance (from the most abundant to the least abundant species) and the relative abundance for each species is given (Figure 3).

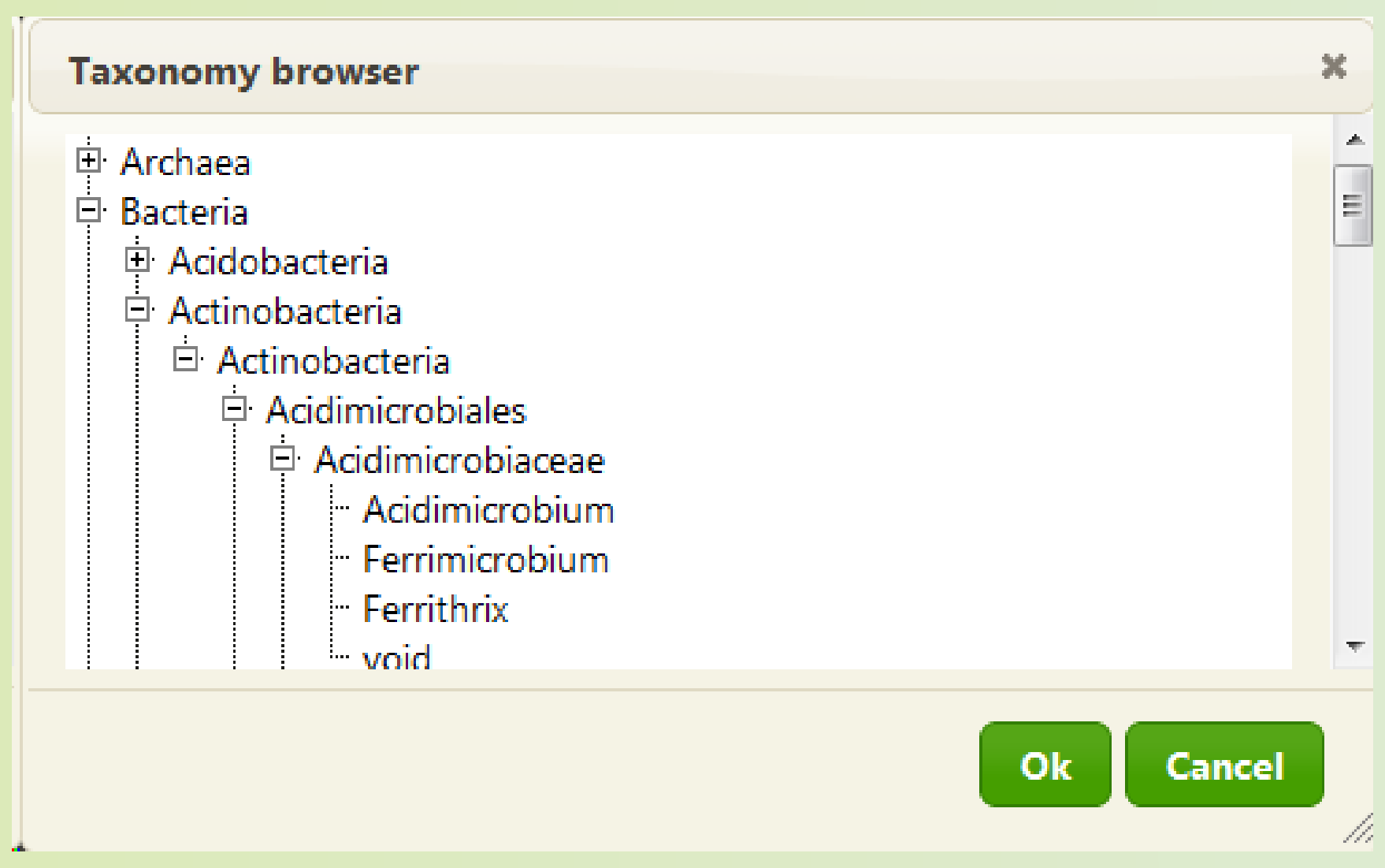


Figure 1: Taxonomy browser

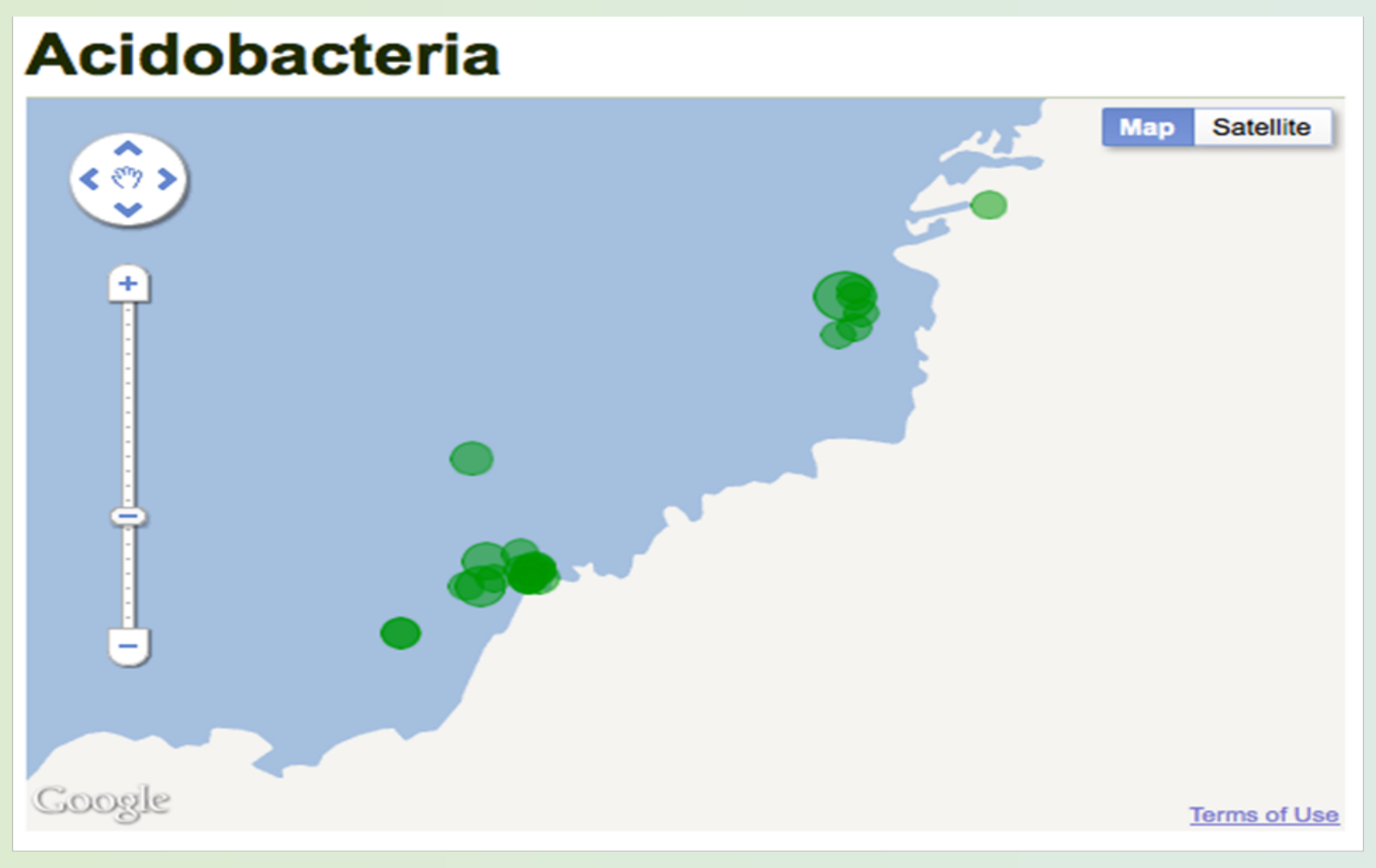
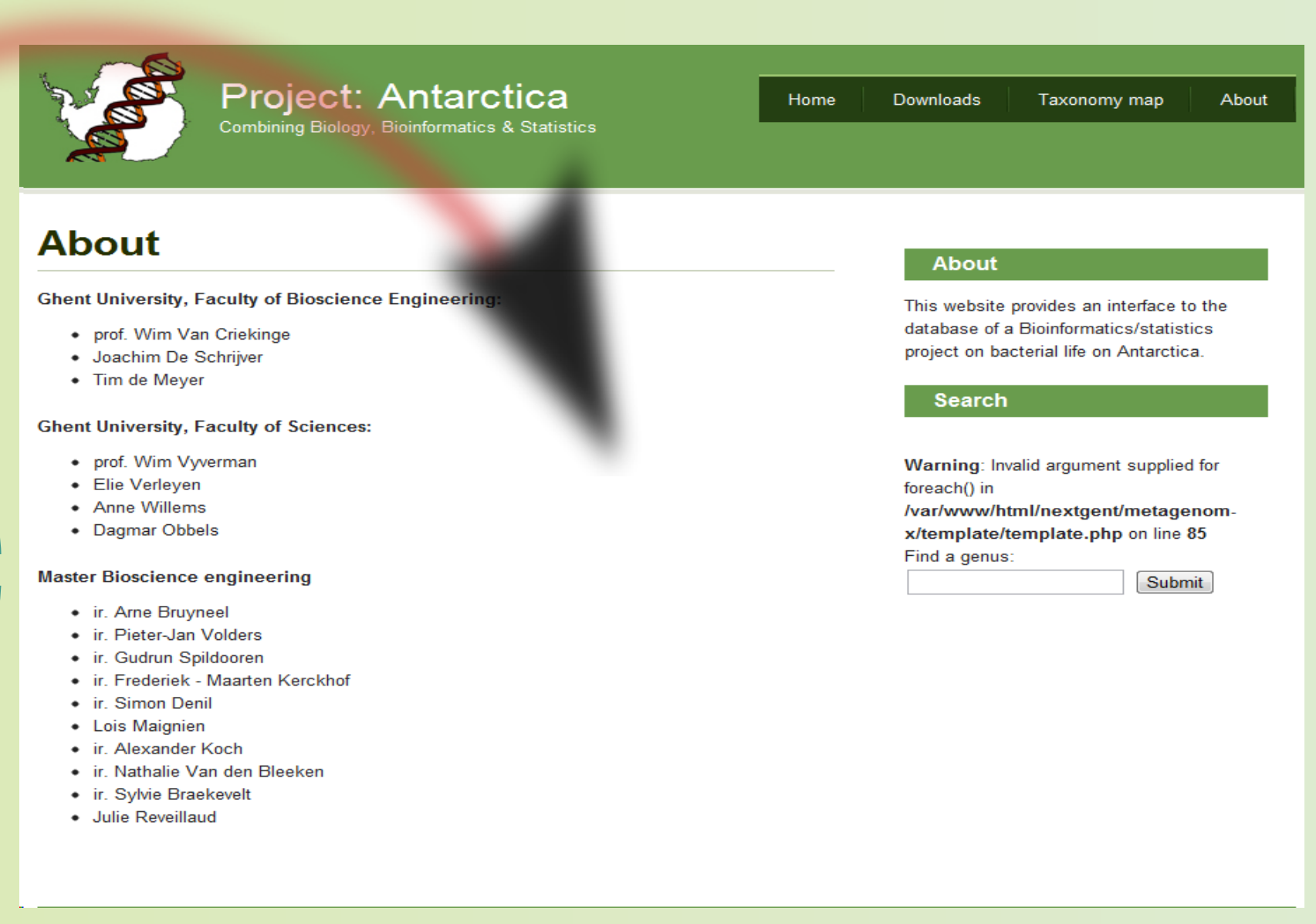


Figure 2: Sample comparison



About

Ghent University, Faculty of Bioscience Engineering:
- prof. Wim Van Criekinge
- Joachim De Schrijver
- Tim de Meyer

Ghent University, Faculty of Sciences:
- prof. Wim Vyverman
- Elie Verleyen
- Anne Willems
- Dagmar Obbels

Master Bioscience engineering
- ir. Arne Bruyneel
- ir. Pieter-Jan Volders
- ir. Gudrun Spildooren
- ir. Frederiek - Maarten Kerckhof
- ir. Simon Denil
- Lois Maignien
- ir. Alexander Koch
- ir. Nathalie Van den Bleeken
- ir. Sylvie Braekevelt
- Julie Reveillaud

About

This website provides an interface to the database of a Bioinformatics/statistics project on bacterial life on Antarctica.

Warning: Invalid argument supplied for foreach() in **/var/www/html/nextgent/metagenom-x/template/template.php** on line **85**
Find a genus:

Website interface

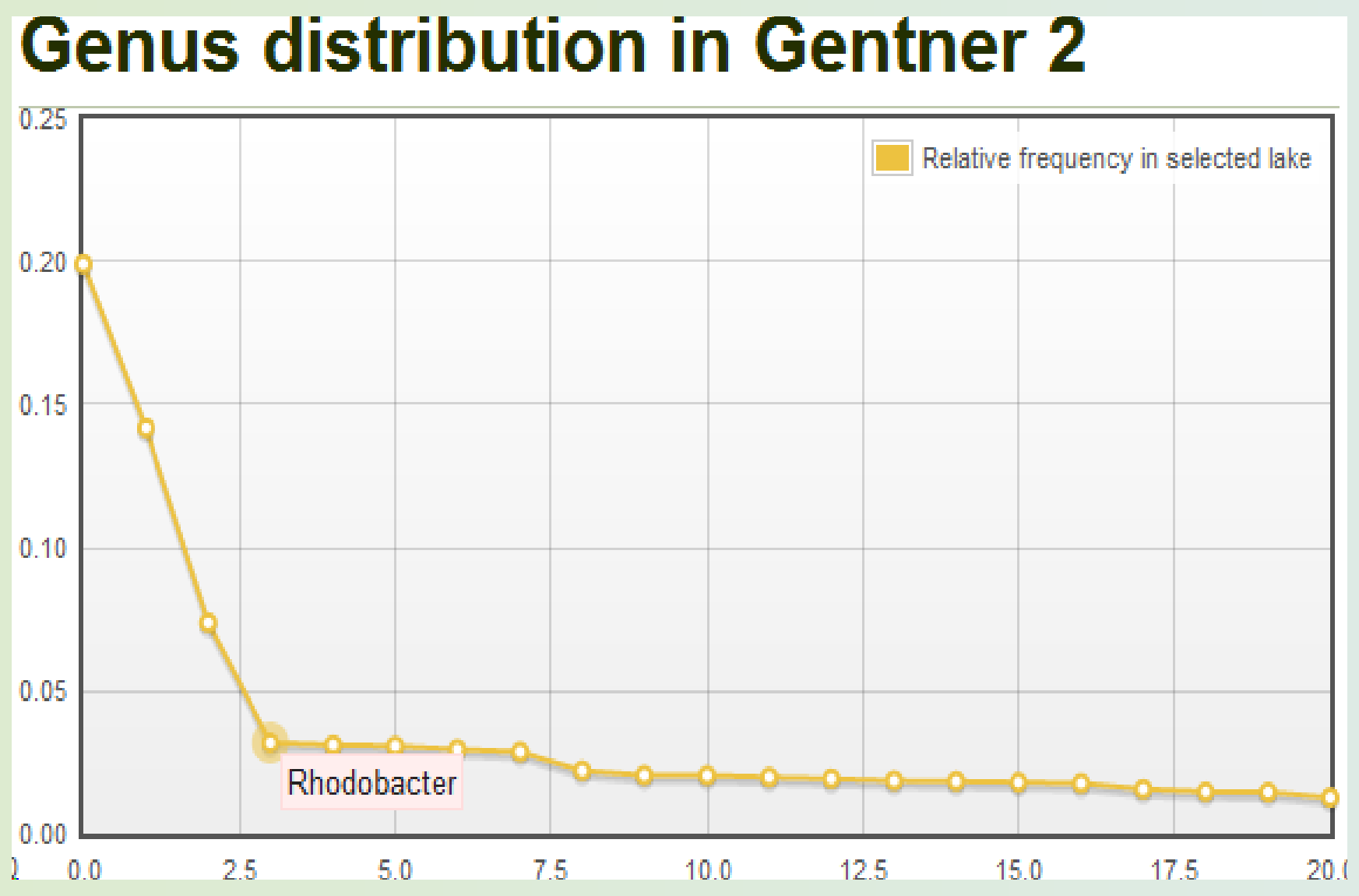


Figure 3: Sample distribution

Availability

The pipeline is completely open-source and available for download.
Download the pipeline at http://athos.ugent.be/metagenom-x/Download/

Author affiliations

1: Lab. for Computational Genomics and Bioinformatics - Ghent Univ.
2: Center for Medical Genetics - Ghent University Hospital
3: Lab. of Microbial Ecology and Technology - Ghent University
4: Laboratory of Prostitology and Aquatic Ecology - Ghent University

Contact

Joachim De Schrijver - Joachim.DeSchrijver@ugent.be
Lab. for Bioinformatics and Computational Genomics (BIOBIX)
Department of Mathematical Modelling, Statistics and Bioinformatics
Faculty of Bioscience Engineering
Ghent University
Coupure Links 653, B-9000 Ghent
Belgium

References

[1] Quince et al. (2011) Removing noise from pyrosequenced amplicons. BMC Bioinformatics
[2] Li et al. (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics
[3] Olsen et al. (1992) The Ribosomal Database Project. Nucleic Acids Res.