

**3rd Course on bioinformatics tools
for Next Generation Sequencing data mining:
use of bioinformatics tools for typing pathogenic *E. coli*
18-19 June, 2018
Istituto Superiore di Sanità, Informatics Training Room**

Content

- **Scientific Report**
- **Course agenda and list of participants (Annex 1)**

Scientific Report

The rapid development of bioinformatics methods for the analysis of Next Generation Sequencing (NGS) data has made the genome sequence-based investigation a realistic alternative to conventional molecular typing of bacterial isolates including PFGE typing for molecular surveillance of VTEC infections.

Discussion groups have already been organized by the European Food Safety Authority (EFSA) and the European Centre for Disease Control (ECDC) to define the most appropriate context for the introduction of NGS as the standard technology for molecular surveillance of foodborne infections. Moreover, an inter-EURLs working group among EURLs responsible of microbial hazards (including also parasites and foodborne viruses) has been established in 2017 and is currently working for facilitating and harmonizing the adoption of NGS among the respective NRLs networks.

In this framework, the EURL-VTEC organized the third edition of a course on the use of bioinformatics tools for assembling the genome, identify the presence of virulence genes and typing Shiga toxin-producing *E. coli* from NGS data, with the aim of facilitating the

adoption of the NGS technology and increasing the skills and level of awareness within the NRLs *E. coli* network.

The event consisted in a two-days course and was held in Rome on 18-19 June 2018, at the IT-training room of the *Istituto Superiore di Sanità* (ISS). Six representatives of as many NRLs and six participants from four Italian Official Laboratories participated in the event. Five additional scientists from ISS structures took part in the course as observers. The travel and accommodation costs of the NRLs participants were supported with a dedicated budget by DG-SANTE, with the only exception of the representative from the NRL of Switzerland. The course was organized in sessions, each composed of interactive oral presentations, and by hands-on training regarding the subjects discussed. Each participant was placed in front of a workstation for carrying out the practical sessions.

The Course was opened by **Dr. Valeria Michelacci**, who welcomed the participants and briefly introduced the workshop's aims and program. She also gave the first presentation about the formats of NGS data and the available servers for NGS data analysis, including ARIES, a shared workspace for intensive data analyses developed within a collaboration among the EURL VTEC and the ISS IT service, available since 2015.

Dr. Arnold Knijn then explained the informatics infrastructure of ARIES, based on Galaxy architecture, including the ARIES cluster organization and giving an overview of ARIES user interface.

Dr. Antonella Maugliani went more in detail with a speech dedicated to practical aspects constituting the basis of the use of ARIES, including how to access information on previous analysis and how to convert the formats of the data.

After this introduction, the course was structured with theoretical lessons followed by specific hands-on exercises, with the aim of giving the participants a direct approach to the data analysis allowing them to be in the condition of using ARIES for data analysis since their return back to the home laboratories. This practical sessions were led by **Dr. Michelacci**, with the support of **Dr. Arnold Knijn** and **Dr. Antonella Maugliani**.

The practical sessions regarded:

- The quality check and trimming of the sequence files.
- The assembly of the reads in contigs and the assembly statistics.
- The use of the EURL-VTEC WGS PT pipeline for the automatic analysis of sequencing data of *E. coli* isolates.
- The Multilocus Sequence Typing schemes and tools, including the classical 7-genes scheme through the tool SRST2 and whole genome MLST (wgMLST) and core genome MLST (cgMLST) through chewBBACA tool.

- The whole genome comparison with a reference-free whole genome SNPs comparison tool (ksnp3)

The course was concluded by **Dr. Michelacci**, who presented a wrap up of the topics discussed during the two days.

Some of the course presentations are available in the EURL-VTEC website (www.iss.it/vtec), section: *E. coli Genomics*.



**3rd Course on bioinformatics tools
for Next Generation Sequencing data mining:
use of bioinformatics tools for typing pathogenic *E. coli***

18-19 June, 2018

Informatics Training Room

(*Building 1, Floor B*)

Istituto Superiore di Sanità

Viale Regina Elena, 299 – Rome, Italy



Organized by:

- The EU Reference Laboratory for *E. coli*
- The ISS IT Service

Funded by the European Commission – DG SANTE

DIRECTOR OF THE COURSE

Stefano MORABITO

EU Reference Laboratory for *E. coli*

Department of Food Safety, Nutrition and Veterinary Public Health
Istituto Superiore di Sanità

SPEAKERS

EU Reference Laboratory for *E. coli*, ISS, Rome, Italy

Antonella MAUGLIANI

Valeria MICHELACCI

Arnold KNIJN

TECHNICAL SECRETARIAT

Susan BABSA, Clarissa FERRERI

EU Reference Laboratory for *E. coli*

Department of Food Safety, Nutrition and Veterinary Public Health
Istituto Superiore di Sanità
Rome, Italy

Fabio GALATI

IT Services

Istituto Superiore di Sanità

Rome, Italy

GENERAL INFORMATION

Venue: Istituto Superiore di Sanità, Informatics Training Room (Building 1, Floor B)

Viale Regina Elena 299, 00161 Rome

This event is part of the scientific and tutorial activities of the EURL-VTEC, funded by the European Commission – DG SANTE

For any information regarding the event, please send an email to crl.vtec@iss.it

PROGRAM

June 18th, Monday

13.00 Registration

13.10 Welcome, housekeeping, and general overview on the training course **Valeria Michelacci**

Session 1

13.20 Introduction to Next Generation Sequencing data formats, basic tools and servers for analysis **Valeria Michelacci**

13.50 IT infrastructure and user interface: The Galaxy architecture and ARIES cluster **Arnold Knijn**

14.10 Practical aspects on the use of ARIES **Antonella Maugliani**

14.30 Preliminary analysis: quality check and trimming **Valeria Michelacci**

15.00 Break

15.30 Hands on: Quality check and trimming of whole genome sequences
Arnold Knijn
Valeria Michelacci
Antonella Maugliani

16.00 Assembly, assembly statistics, virulotyping and serotyping **Valeria Michelacci**

16.30 Hands on: Assembly of the raw reads
Arnold Knijn
Valeria Michelacci
Antonella Maugliani

17.00 Introduction to the EURL-VTEC WGS PT pipeline **Valeria Michelacci**

17.20 Hands on: EURL-VTEC WGS PT pipeline
Arnold Knijn
Valeria Michelacci
Antonella Maugliani

17.40 **End of Session 1**

Session 2

09.00	Hands on: assembly statistics	Arnold Knijn Valeria Michelacci Antonella Maugliani
09.30	Hands on: analysis of EURL-VTEC WGS PT pipeline report	Arnold Knijn Valeria Michelacci Antonella Maugliani
10.00	Multi Locus Sequence Typing from WGS data	Valeria Michelacci
10.30	Hands on: MLST 7 genes	Arnold Knijn Valeria Michelacci Antonella Maugliani
11.00	Hands on: wgMLST	Arnold Knijn Valeria Michelacci Antonella Maugliani
11.40	Whole genome SNPs comparison: Reference-free approach	Valeria Michelacci
12.10	Lunch break	
13.10	Hands on: ref-free wgSNPs	Arnold Knijn Valeria Michelacci Antonella Maugliani
14.00	Final discussion	
15.00	Closure	

Participants

Caterina Flink, **NRL Sweden**
Joerg Hummerjohann, **NRL Switzerland**
Kristine Kreitsman, **NRL Estonia**
Marina Lamparter, **NRL Germany**
Irena Meistere, **NRL Latvia**
Ana Isabel Rocha Roso, **NRL Spain**
Monica Rosaria Molotzu, **IZS Sardegna**
Patrizia Centorame, **IZS Abruzzo and Molise**

Marina Torresi, **IZS Abruzzo and Molise**

Arianna Peruzzo, **IZS Venezie**

Annalisa Paternò, **IZS Lazio e Toscana**

Maria Laura De Marchis, **IZS Lazio and Toscana**

Observers

Alfio D'Angelo, **ISS**

Michele Zuppi, **ISS**

Fiorella Ciaffoni, **ISS**

Manuela Marra, **ISS**

Ornella Moro, **ISS**