



Training Course on WGS data use: bioinformatics tools for aiding STEC outbreak investigation

11-12 July, 2019

**Istituto Superiore di Sanità, Aula Zampieri
Via Giano della Bella, 34 – 00161 Rome, Italy**

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 whole genome sequence-based characterization a consolidate approach for the surveillance and monitoring of STEC infections and for a real time management of STEC outbreaks.

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 training course on the use of bioinformatics tools for aiding STEC outbreak investigation, with the aim to strengthen the competence on the analysis of WGS data within the NRLs network, and improve the capability in using such data for outbreak response and investigation

The event consisted in a two-days course and was held in Rome on 11-12 July 2019, at the Istituto Superiore di Sanità (ISS). Ten representatives of as many NRLs, one from EFSA and one from the European Commission participated in the event. Two additional scientists from ISS structures took part in the course. The travel and accommodation costs of the NRLs participants were supported with a dedicated budget by European Commission, DG-SANTE.

The course was organized in different sessions. The first day was dedicated to interactive oral presentations and hands-on sessions with practical exercises on the subjects discussed, based on genome assembly and typing of STEC isolates. The second day was focused on a simulation of a national outbreak investigation, involving the characterization and cluster analysis of STEC WGS and results interpretation with the aim of examining the correlation between isolates for the source attribution.

The course was based on the bioinformatic tools available on the Galaxy public webserver ARIES, a shared workspace for intensive data analyses developed within a collaboration among the EURL VTEC and the ISS IT service, available since 2015.

The course was opened by **Dr. Stefano Morabito**, who welcomed the participants and briefly introduced the training's aims and program. The first presentation was from **Dr. Valeria Michelacci**, who introduced the NGS data formats and the basic analytical tools, including the quality check of the raw sequences. She also explained the assembly of the reads in contigs in a following presentation. **Dr. Rosangela Tozzoli** presented softwares and servers currently available for the NGS data analysis. **Dr. Antonella Maugliani** gave an overview of ARIES user interface and illustrated the basis of the use of ARIES webserver. **Dr. Federica Gigliucci** explained approaches to retrieve the basic characterization of STEC isolates, based on performing the serotyping, the virulotyping and the 7-genes MLST for the sequence type definition. Moreover, she introduced how to perform cluster analysis on a set of STEC genomes analyzed through cgMLST, in order to examine the correlation between strains especially useful in case of outbreak investigation. Each theoretical lesson was followed by specific hands-on exercises, allowing the participants to be in the condition of using ARIES for data analysis since their return back

to the home laboratories and giving them a direct approach in attributing foodborne disease to specific sources of infection. In detail, a flow chart of the steps needed for *E. coli* outbreak investigation based on WGS analysis was distributed to the participants to serve as a tool for the course and as guidance to aid the investigation during potential real outbreaks in their countries.

The practical sessions regarded:

- The quality check and trimming of the sequence files.
- The 7-genes Multilocus Sequence Typing schemes through the tool SRST2 for analyzing fastq sequences and the tool MLST developed by T. Seemann for the analysis of the contigs.
- The use of dedicated pipelines for the *E. coli* Serotyping and Virulotyping.
- The assembly of the reads in contigs and the assembly statistics.
- Outbreak investigation exercise.

Dr. Stefano Morabito closed the course with a discussion on the outcome of the exercises and then **Dr. Valeria Michelacci** presented the EURL-VTEC WGS PT pipeline for the automatic analysis of sequencing data of *E. coli* isolates.



**Training Course on WGS data use:
bioinformatics tools for aiding STEC outbreak
investigation**

11-12 July, 2019

Venue: Aula Zampieri

**Istituto Superiore di Sanità
Via Giano della Bella, 34 – 00161 Rome, Italy**



Organized by:

The EU Reference Laboratory for *E. coli*

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/TUTORS

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

Federica GIGLIUCCI
Arnold KNIJN
Antonella MAUGLIANI
Valeria MICHELACCI
Stefano MORABITO
Rosangela TOZZOLI

TECHNICAL SECRETARIAT

Guendalina FORNARI LUSWERGH, Clarissa FERRERI

EU Reference Laboratory for *E. coli*
Department of Food Safety, Nutrition and Veterinary Public Health
Istituto Superiore di Sanità
Rome, Italy

GENERAL INFORMATION

Venue: Istituto Superiore di Sanità, Via Giano della Bella, 34 – 00161 Rome, Italy
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

FINAL PROGRAM

July 11th, Thursday

09.00 Registration

09.10 Welcome, housekeeping, and general overview on the training course **Stefano Morabito**

Session 1

09.20 Introduction to Next Generation Sequencing data formats, quality check and basic analytical tools **Valeria Michelacci**

09.50 Bioinformatic analysis of NGS data: approaches and opportunities **Rosangela Tozzoli**

10.10 Introduction to the ARIES webserver user-interface **Antonella Maugliani**

10.30 Hands on: Quality check and Trimming of WGS
Federica Gigliucci
Antonella Maugliani
Valeria Michelacci
Rosangela Tozzoli

11.00 Break

11.30 Basic characterization: 7-genes Multi Locus Sequence Typing (MLST), virulotyping and serotyping **Federica Gigliucci**

12.00 Hands on: 7-genes MLST
Federica Gigliucci
Antonella Maugliani
Valeria Michelacci
Rosangela Tozzoli

12.15 Hands on: virulotyping and serotyping
Federica Gigliucci
Antonella Maugliani
Valeria Michelacci
Rosangela Tozzoli

12.45 End of Session 1 and Lunch break

Session 2

13.45 Assembly and assembly statistics

Valeria Michelacci

14.00 Hands on: Assembly of the raw reads

Federica Gigliucci

Antonella Maugliani

Valeria Michelacci

Rosangela Tozzoli

14.20 Introduction to core genome MLST (cgMLST)

Federica Gigliucci

Federica Gigliucci

Antonella Maugliani

14.40 Hands on: Assembly statistics

Valeria Michelacci

Rosangela Tozzoli

15.10 Introduction to the Outbreak investigation exercise

Stefano Morabito

16.30 End of Session 2

July 12th, Friday

Session 3

09.00	Outbreak investigation exercise	Federica Gigliucci Antonella Maugliani Valeria Michelacci Rosangela Tozzoli
12.00	Discussion on the outcome of outbreak investigation exercise	Stefano Morabito
12.30	Introduction to automatic pipelines for <i>E. coli</i> typing through ARIES	Valeria Michelacci
13.00	Closure	

Participants

Luisa Antunes, **European Commision**

Valentina Rizzi, **EFSA**

Amina Ismail Ahmed, **NRL UK**

Brian Byrne, **NRL Ireland**

Célia Cristina Fialho Leão, **NRL Portugal**

Nuria Gomez-Fernandez, **NRL Spain**

Hans Tomas Jinnerot, **NRL Sweden**

Rasmus Nielsen Klitgaard, **NRL Denmark**

Elżbieta Maćkiw, **NRL Poland**

Bojan Papić, **NRL Slovenia**

Elisabeth Schuh, **NRL Germany**

Gertruda Seskute, **NRL Lithuania**

Gaia Scavia, **ISS**

Eleonora Ventola, **ISS**