



**2<sup>nd</sup> Course on Bioinformatics  
for Next Generation Sequencing Data Mining:  
Use of Tools for Typing Pathogenic *E. coli***

**16-17 June, 2016**

**Istituto Superiore di Sanità, SIDBAE Training Room**

**Content**

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

**Scientific Report**

The rapid advance of next generation sequencing (NGS) platforms and the parallel development of bioinformatics for NGS data management and analyses make the genome sequence-based typing approach a realistic alternative to PFGE and its designated successor for molecular surveillance systems of Shiga toxin-producing *Escherichia coli* (STEC) infections. A recent inventory of the molecular typing methods and IT applications available within the network of the NRLs for *E. coli* showed that some NRLs have already access to NGS facilities.

To consolidate the knowledge on the NGS technology and to increase the level of skill and awareness within the NRL network, the EURL-VTEC organized the *1<sup>st</sup> Course on bioinformatics tools*, held in Rome in June 2015 and focused on the use of bioinformatics for assembling, mining and analyzing NGS data (course program and report available at the EURL-VTEC website, [www.iss.it/vtec](http://www.iss.it/vtec), “*E. coli* genomics” section).

In 2016, the EURL-VTEC organized a second training course on the use of bioinformatics for NGS data mining. The course was focused on the typing of pathogenic *E. coli* strains based on NGS using the tools freely available as web-based services such as those hosted at the Center for Genomic Epidemiology at DTU in Copenhagen

(<https://cge.cbs.dtu.dk/services/>) as well as those available on the ARIES portal. ARIES is a webserver developed by the EURL-VTEC with the aim of being a shared workspace for intensive data analyses, particularly those deriving from NGS, and was presented to the Network of NRLs for *E. coli* during the 1<sup>st</sup> *Course on bioinformatics tools*. ARIES is installed and maintained on the ISS servers and contains tools allowing the analysis of whole genome sequences for the identification and typing of pathogenic *E. coli*.

The event consisted in a two-days course and was held in Rome on 16-17 June 2016, at the IT-training room of the *Istituto Superiore di Sanità* (ISS). Six representatives from as many NRLs for *E. coli*, two from Italian Official Laboratories and four scientists from NRL *E. coli* Italy attended the course. Additionally, representatives from the European Centre for Disease Prevention and Control (ECDC) and the European Food Safety Authority (EFSA) participated in the event. The travel and accommodation costs of the NRLs for *E. coli* participants were supported by DG-SANTE through a dedicated line in the budget of the EURL-VTEC.

The course was organized in sessions, each composed of interactive oral presentations, and hands-on training regarding the subjects discussed. A workstation was made available for each participant for the practical sessions.

The EURL-VTEC and ISS staff (**Dr. Valeria Michelacci**, **Dr. Rosangela Tozzoli**, **Dr. Antonella Maugliani**) managed the presentations and the exercises. **Dr. Massimiliano Orsini**, a Bioinformatician from the *Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise*, Teramo, Italy, who is collaborating with EURL in the development of bioinformatics tools for *E. coli* typing, also participated in the course management, either giving specific presentations or providing help in the hands-on sessions. **Dr. Valentina Rizzi** and **Dr. Saara Kotila** participated as invited speakers, representing EFSA and ECDC, respectively.

**Dr. Stefano Morabito**, Director of the EURL-VTEC, welcomed the participants and briefly introduced the course aims and program.

**Dr. Valentina Rizzi** presented EFSA initiatives on Whole Genome Sequencing (WGS), including the Scientific Colloquium on the use of WGS of food-borne pathogens for public health protection, the collaboration in the ECDC FWD-NEXT Expert Group on introduction of next-generation typing methods for surveillance of food- and water-borne diseases and the contribution for the review of the Technical background Paper 'Applications of Whole Genome Sequencing in food safety management', prepared by Food and Agriculture Organization of the United Nations in collaboration with the World Health Organization. Moreover Dr. Rizzi briefly introduced the projects granted by EFSA for promoting networking on the integration of WGS in microbial food safety: **Innuendo**, led by the

University of Helsinki, and **ENGAGE**, led by the Denmark Technical University. Finally, she introduced the ongoing discussion aimed at expanding the joint EFSA-ECDC molecular typing database to WGS data.

**Dr. Saara Kotila** introduced the surveillance activities of the ECDC FWD network by introducing the model of analysis of WGS data proposed by FWD-NEXT expert group and the decisional flowchart of the already ongoing FWD surveillance in the Epidemic Intelligence information System (EPIS-FWD). Dr. Kotila explained that the possibility of using WGS data to support molecular typing cluster investigation is already available in the EPIS platform and introduced the results of a survey conducted in May 2015 about WGS implementation in EU/EEA Member States (MS). The analysis of the data collected showed that more than 30 % of MS declared the intention to implement WGS in the next two years to complement existing methods, while a few other have already implemented it. Finally, some MS declared not to have plans to do it. Dr. Kotila closed her presentation showing an example of the use of WGS data supporting the epidemiological analysis in EPIS-FWD during a *Salmonella* Enteritidis outbreak investigation, stressing the importance of streamlining data sharing and transfer in the future.

**Dr. Rosangela Tozzoli** gave an overview of the different typing approaches available for VTEC, ranging from serotyping to WGS, including virulotyping, MLVA, PFGE and MLST. She presented the main advantages and drawbacks of each technique, introducing the relevant role, which WGS is expected to play in surveillance and outbreak investigation once the WGS data production and analysis will be streamlined and standardized.

**Dr. Arnold Knijn**, from the ISS IT-service, introduced the ARIES webserver, developed within a collaboration among the EURL-VTEC, the ISS IT service (Dr. Arnold Knijn) and the *Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise* (Dr. M. Orsini). In detail, he presented the informatics infrastructure of ISS, the Galaxy webserver architecture and the advantages deriving from the use of a Galaxy cluster, such as the one currently powering ARIES. Dr. Knijn, also showed the possible drivers for the future improvements of ARIES through the use of a cloud-based infrastructure. He also introduced the user interface of ARIES to the participants.

**Dr. Valeria Michelacci** introduced the basic tools for NGS data mining, from those required for the quality check and trimming of the raw data, to those used for assembling them in contigs. She also revised the main file formats of data deriving from NGS experiments and briefly presented the main strategies used for identifying the presence of genetic features of interest in the sequenced strains, mainly based on the comparison of the data with a database of reference sequences, by mapping the reads or performing *blastn* searches on the contigs.

The rest of the course consisted in practical sessions composed on a brief introduction of specific tools followed by corresponding hands-on sessions, lead by **Dr. Michelacci**, with the support of **Dr. Maugliani**, **Dr. Tozzoli** and **Dr. Orsini**.

The practical sessions regarded:

- The virulotyping and serotyping of the sequenced VTEC strains.
- The comparison of genome sequences through the analysis of the whole genome SNPs (wgSNPs) comparison, using tools available through different webserver (Centre for Genomic Epidemiology and ARIES Galaxy Server).
- The comparison of strains using melting temperatures of PCR products amplified from virulence genes, through the tool HReVAP (High Resolution Virulence Allelic Profiling), developed by the EURL-VTEC as a concept for analyzing the variability of this fraction of the genome in STEC strains.
- The MLST approach for typing *E. coli* strains from NGS data, on the basis of two different schemes: the one based on seven housekeeping genes, through the SRST2 tool, and the accessory genome database, currently based on the analysis of the variability of virulence genes through a dedicated allele-calling tool developed by ARIES group.

The course was concluded by **Dr. Morabito**, who presented a wrap-up of the topics discussed during the two days and proposed a possible approach for a molecular typing NGS data collection system for surveillance, which could be based on MLST typing of extended panels of genes and involve the storage and analysis of data.

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



**2nd course on bioinformatics tools for Next Generation  
Sequencing data mining: use of  
bioinformatics tools for typing pathogenic *E. coli***

**16-17 June, 2016**

**SIDBAE 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 (SIDBAE)

**Funded by the European Commission – DG SANTE**

## DIRECTOR OF THE COURSE

Stefano MORABITO

EU Reference Laboratory for *E. coli*

Dipartimento di Sanità Pubblica Veterinaria e Sicurezza Alimentare  
Istituto Superiore di Sanità

## SPEAKERS

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

Antonella MAUGLIANI

Valeria MICHELACCI

Rosangela TOZZOLI

IT Service (SIDBAE), ISS, Rome, Italy

Arnold KNIJN

Istituto Zooprofilattico Sperimentale Abruzzo e Molise, Italy

Massimiliano ORSINI

European Food Safety Agency, Parma, Italy

Valentina RIZZI

European Center for Disease Prevention and Control, Stockholm, Sweden

Saara KOTILA

## TECHNICAL SECRETARIAT

Susan BABSA, Clarissa FERRERI

EU Reference Laboratory for *E. coli*

Department of Veterinary Public Health and Food Safety

Istituto Superiore di Sanità

Rome, Italy

Fabio GALATI, Arnold KNIJN

Servizio Informatico, documentazione, biblioteca ed attività editoriali (SIDBAE)

Istituto Superiore di Sanità

Rome, Italy

## GENERAL INFORMATION

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

Viale Regina Elena 299, 00161 Rome

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

For any information regarding the event, please send an email to [crl.vtec@iss.it](mailto:crl.vtec@iss.it)

**June 16th, Thursday**

<b>9.00</b>	<b>Registration</b>	
<b>9.15</b>	Welcome, housekeeping, and general overview on the training course	<b>Stefano Morabito</b>
	<b>Session 1</b>	
<b>9.30</b>	Recent and on-going WGS initiatives and activities of EFSA	<b>Valentina Rizzi</b>
<b>10.00</b>	Recent and on-going WGS initiatives and activities of ECDC	<b>Saara Kotila</b>
<b>10.30</b>	<b>Coffee break</b>	
<b>11.00</b>	VTEC strains typing: from traditional methods to NGS	<b>Rosangela Tozzoli</b>
<b>11.30</b>	IT infrastructure and user interface: the Galaxy architecture and ARIES cluster	<b>Arnold Knijn</b>
<b>12.00</b>	Basic tools for NGS data mining: quality check, assembly, annotation, alignment and blast	<b>Valeria Michelacci</b>
<b>12.30</b>	Virulotyping and serotyping	<b>Valeria Michelacci</b>
<b>13.00</b>	<b>Lunch</b>	
<b>14.00</b>	Hands-on exercises: virulotyping and serotyping	<b>Rosangela Tozzoli</b> <b>Valeria Michelacci</b> <b>Antonella Maugliani</b>
<b>15.00</b>	Whole genome SNPs (wgSNPs) comparison: SNPtree, NDtree, CSI Phylogeny and kmer-based analysis	<b>Valeria Michelacci</b>
<b>15.30</b>	Hands on exercises: wgSNPs analysis	<b>Valeria Michelacci</b> <b>Rosangela Tozzoli</b> <b>Antonella Maugliani</b>
<b>17.00</b>	<b>End of Session 1</b>	

**Session 2**

<b>9.30</b>	Hands on exercises: kmer-based wgSNPs analysis	<b>Valeria Michelacci Rosangela Tozzoli Antonella Maugliani</b>
<b>10.15</b>	Introduction to High Resolution Virulence Allelic Profiling (HReVAP)	<b>Valeria Michelacci</b>
<b>10.45</b>	Hands on exercises: HReVAP from melting temperatures	<b>Valeria Michelacci Rosangela Tozzoli Antonella Maugliani</b>
<b>11.15</b>	<b>Coffee break</b>	
<b>11.45</b>	Gene-by-gene approaches for typing through WGS: conventional MLST and <i>E. coli</i> accessory genome MLST (agMLST)	<b>Massimiliano Orsini</b>
<b>12.15</b>	Hands-on exercises: conventional MLST and agMLST	<b>Valeria Michelacci Rosangela Tozzoli Antonella Maugliani Massimiliano Orsini</b>
<b>12.45</b>	Concluding remarks	<b>Stefano Morabito</b>
<b>13.00</b>	<b>Closure of the Course and Lunch</b>	

**Participants**

Giusi AMORE, **EFSA**  
Anna ASPÁN, **NRL Sweden (SVA)**  
Roslen BONDI', **NRL Italy**  
Paola CHIANI, **NRL Italy**  
Federica FLAMINI, **NRL Italy**  
Federica GIGLIUCCI, **NRL Italy**  
Anniina JAAKKONEN, **NRL Finland**  
Juris KĪBILDS, **NRL Latvia**  
Bojan PAPIČ, **NRL Slovenia**  
Rubina PARADISO, **IZS Mezzogiorno – Portici (NA)**

Deirdre PRENDERGAST, **NRL Ireland**

Angelo ROMANO, **IZS Piemonte, Liguria e Valle d'Aosta - Torino**

Ettore SEVERI, **ECDC**

Kinga WIECZOREK, **NRL Poland (NVRI)**

## **Observers**

Arianna ANGELONI, **NRL Italy**

Silvia ARANCIA, **NRL Italy**

Luca DE SABATO, **ISS**

Ilaria DI BARTOLO, **ISS**

Giovanna FRANCIOSA, **ISS**

Fabio MINELLI, **NRL Italy**

Concetta SCALFARO, **ISS**

Gaia SCAVIA, **EURL-VTEC**