



**Basic Course on Bioinformatics tools for Next Generation Sequencing
data mining
11-12 June, 2015
Istituto Superiore di Sanità, SIDBAE Training Room**

Content

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

Scientific Report

The rapid development of next generation sequencing (NGS) platforms and the parallel development of bioinformatics tools for NGS data management and analyses make the genome sequence-based investigation a realistic alternative to conventional molecular typing of bacterial isolates. In particular, the approach promise to become a credible alternative to PFGE and its designated successor for molecular surveillance of VTEC infections.

Discussion groups have been recently 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. In this respect, a recent inventory of the molecular typing methods and IT applications available within the network of the National Reference Laboratories (NRL) for *E. coli* showed that some NRLs have access to NGS facilities.

In this framework, the European Reference Laboratory for *E. coli* (EU-RL VTEC) organized a course on the use of bioinformatics for assembling, mining, and analyzing NGS data,

with the aim of consolidating the knowledge on the NGS technology and to increase the level of skill and awareness within the NRLs *E. coli* network.

The event consisted in a two-days course and was held in Rome on 11-12 June 2015, at the IT-training room of the *Istituto Superiore di Sanità* (ISS). Nine representatives of eight NRLs and five from as many Italian Official Laboratories participated in the event. Two other scientists from NRL Italy, three scientists from other ISS structures, and two EFSA representatives also 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.

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 presentations and the exercises were managed by the EU-RL and ISS staff.

Dr. Massimiliano Orsini, a Bioinformatician from the *Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise*, Teramo, Italy, who is collaborating with EU-RL in the development of bioinformatics tools, also participated in the course management, either giving specific presentations or providing help in the hands-on sessions. An EFSA representative participated as invited speaker.

The Course was opened by **Dr. Stefano Morabito**, Deputy Director of the EU-RL VTEC, who welcomed the participants and briefly introduced the workshop's aims and program. Dr. Morabito also gave an overview of the tools for genomics analysis available as software suites or webservers and introduced the ARIES webserver, a shared workspace for intensive data analyses developed within a collaboration among the EU-RL VTEC, the ISS IT service (Dr. Arnold Knijn), and the *Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise* (Dr. M. Orsini).

Dr. M. Teresa da Silva Felicio, from EFSA, Parma, presented the current Whole Genome Sequencing (WGS) procurement activities in place at EFSA and the conclusions of the discussion groups from EFSA's Scientific Colloquium on the use of WGS of food-borne pathogens for public health protection. The Colloquium was held in Parma on June 2014 and the report is available for consultation on the EFSA website.

Dr. Valeria Michelacci introduced the NGS technology, considering the most widely used platforms *Illumina* and *Ion Torrent*, with particular attention to the methodology underlying the two different systems, their throughput, and the possible problems arising with the two technologies. Moreover, she gave an overview of the first steps in the management of NGS data, introducing the concepts of checking the quality of the raw sequencing reads and assembling them in contigs.

Dr Massimiliano Orsini presented the advantages deriving from the use of servers for NGS data analyses, and underlined the possible problems deriving from the use of local computing capacities, for limits either in the computational power or in the capacity of storage of the resulting data. Other talks provided a deeper explanation of the assembling processes, and an overview of the assembly pipelines available, with particular attention to those recommended for the different NGS systems, and introducing the steps of the assembly project.

Dr. Arnold Knijn, from the ISS IT-service, introduced the informatics infra-structure of ISS, the Galaxy webserver architecture, and the advantages deriving from the use of a Galaxy cluster, increasing the computational activity by portioning the workload among several computing nodes. Dr. Knijn showed the user interface of the Aries Galaxy Cluster at ISS, which will be soon made available for external users.

The rest of the course consisted in hands-on sessions, lead by **Dr. Michelacci**, with the support of **Dr. Antonella Maugliani** and **Dr. Rosangela Tozzoli**, both from EU-RL VTEC, and **Dr. Orsini**.

The practical sessions regarded:

- The quality check of the sequence files.
- The assembly of the reads in contigs by different webserver (i.e. Centre for Genomic Epidemiology and Aries Galaxy Server).
- The possibilities of BLAST searches to identify virulence genes and define the serotype of an *E. coli* strain based on the whole genome sequences.
- The use of the Prokka tool for annotating contigs.
- The progressive alignments among sequences from multiple strains with MAUVE software.
- The possibilities of typing with NGS data.

On the second training day, **Dr Michelacci** gave a short presentation on the current approaches used for typing bacterial strains starting from NGS data, introduced the problems arising in reproducibility and nomenclature assignment, and proposed some novel strategies for designing cross-platform and cross-generational typing strategies.

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

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



Basic Course on Bioinformatics tools for Next Generation Sequencing data mining

11-12 June, 2015

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

Alfredo CAPRIOLI

Antonella MAUGLIANI

Valeria MICHELACCI

Stefano MORABITO

Rosangela TOZZOLI

IT Service (SIDBAE), ISS, Rome, Italy

Arnold KNIJN

Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise, Teramo, Italy

Massimiliano ORSINI

European Food Safety Agency, Parma, Italy

Maria Teresa DA SILVA FELICIO

TECHNICAL SECRETARIAT

Susan BABSA, Clarissa FERRERI

EU Reference Laboratory for *E. coli*

Dipartimento di Sanità Pubblica Veterinaria e Sicurezza Alimentare

Istituto Superiore di Sanità

Rome, Italy

Fabio GALATI

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

Thursday 11 June

9.00 Registration

- 9.15 Welcome, housekeeping, and general overview on the training course
- Alfredo Caprioli**
Stefano Morabito
Fabio Galati

Session 1

- 9.45 Recent and on-going WGS initiatives and activities of EFSA
- M. Teresa da Silva Felicio**

- 10.00 Next generation sequencers: from the bacterial culture to raw data
- Valeria Michelacci**

10.30 Coffee break

- 11.00 Analysis of NGS data: local and remote options, closed and custom platforms (Galaxy)
- Massimiliano Orsini**

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

- 12.00 Assembly and alignment of NGS data
- Valeria Michelacci**
Massimiliano Orsini

13.00 Lunch

- 14.00 Hands-on exercises: DTU Center for Genomic Epidemiology remote closed approach
- Valeria Michelacci**
Rosangela Tozzoli
Antonella Maugliani

- 15.00 Hands-on exercises: remote and custom analysis on ARIES (Assembly of Illumina and IonTorrent reads, QC, Bowtie2 mapping)
- Massimiliano Orsini**
Valeria Michelacci
Arnold Knijn

- 16.00 Blast searches, annotations, visualization of genome comparisons
- Valeria Michelacci**

Session 2

9.30	Hands-on exercises: <i>E. coli</i> Virulotyper and Serotyper, Prokka annotation and Mauve alignments.	Valeria Michelacci
11.00	Coffee break	
11.30	Typing in the NGS era: The way forward!	Valeria Michelacci
12.15	Hands-on exercises: MLST, different tools for phylogenetic trees (SNPs, NDtree)	Valeria Michelacci Rosangela Tozzoli Stefano Morabito
12.45	Concluding remarks	Stefano Morabito Alfredo Caprioli
13.00	Closure of the Course and Lunch	

Participants

Roslen BONDÌ, **NRL Italy**

Maria Grazia DE FALCO, **IZS del Mezzogiorno – Portici (NA)**

Sarah DENAYER, **NRL Belgium (IPH)**

Federica GIGLIUCCI, **NRL Italy**

Sirpa HEINIKAINEN, **NRL Finland**

Ivana KOLÁČKOVÁ, **NRL Czech Republic**

Guerrino MACORI, **IZS Piemonte, Liguria e Valle d'Aosta - Torino**

Eleonora MASTRORILLI, **IZS delle Venezie – Legnaro (PD)**

Antonio PARISI, **IZS Puglia e Basilicata - Putignano (BA)**

Michael POLEMIS, **NRL Greece**

Stefano REALE, **IZS della Sicilia**

Sabine SCHLAGER, **NRL Austria**

Gitte SØRENSEN, **NRL Denmark**

Michael WRIGHT, **NRL UK**

Observers

Paola CHIANI, **NRL Italy**

Paola DE SANTIS, **IZS Lazio e Toscana - Roma**

Ilaria DI BARTOLO, **ISS**

Federica FLAMINI, **NRL Italy**

Beatriz GUERRA, **EFSA**

Francesca LEONI, **IZS Umbria e Marche - Ancona**

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

Gabriele VACCARI, **ISS**