

Update on the preparedness towards the use of NGS for characterization and typing of pathogenic *E. coli*

Valeria Michelacci

12th Annual Workshop of the National Reference Laboratories for *E. coli* in the EU

Rome, October 12th 2017



Istituto Superiore di Sanità, Dep. of Food Safety, Nutrition and Veterinary Public Health
European Union and National Reference Laboratory for *E. coli*, Rome, Italy



EURLS WORKING GROUP ON WGS

AIM:

to promote the use of WGS across the EURLs' networks, build WGS capacity within the EU and ensure liaison with the work of the EURLs and the work of EFSA and ECDC on the WGS mandate sent by the Commission

MEMBERSHIP:

EURL *E. coli* (coordinator)

EURL *Listeria monocytogenes*

EURL CPS

EURL *Salmonella*

EURL *Campylobacter*

EURL Parasites

EURL AR

Observers: SANTE G4, EFSA, ECDC

Kick off:

November 14th 2017



EURLS WORKING GROUP ON WGS

To elaborate **guidance documents, reports and common strategies** for:

- PTs
- DNA extraction and sequencing
- Bioinformatics tools
- WGS data analysis
- Bench marking
- Training on WGS

To Follow-up of ISO activities on WGS

EURL-VTEC nominated as an observer in the **WG 25 “WGS for typing and genomic characterization”** of ISO/TC 34/SC 9 “Food products-Microbiology”



Getting ready for the paradigm shift

EFSA-ECDC:

Set up of an expert working group for «technical support to collect and analyse whole genome sequencing (WGS) data in the joint ECDC-EFSA molecular typing database»

Stefano Morabito from EURL-VTEC will participate

Preparation of a technical report – **deadline April 2019**



ARIES: A Galaxy-based workspace for intensive data analyses

The screenshot displays the Galaxy web interface for ARIES - ISS. The top navigation bar includes 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. The left sidebar lists various tools under categories like 'COMMON TOOLS' and 'HREVP TOOLS'. The main workspace area shows a green header for 'Istituto Superiore di Sanità' and a central graphic featuring a DNA double helix, the word 'ARIES', the logo of the Istituto Superiore di Sanità, and the EU-RL VTEC logo. On the right, a 'History' panel lists 19 datasets, including 'metagenomics biosolids' and various analysis steps like 'Krona chart on data 73' and 'species assigned to Mobio 15 trimmed'.

Galaxy is an open, web-based platform for data intensive biomedical research. The Galaxy team is a part of BX at Penn State, and the Biology and Mathematics and Computer Science departments at Emory University. The Galaxy Project is supported in part by NHGRI, NSF, The Huck Institutes of the Life Sciences, The Institute for CyberScience at Penn State, and Emory University.



Istituto Superiore di Sanità, Dep. of Food Safety, Nutrition and Veterinary Public Health
European Union and National Reference Laboratory for *E. coli*, Rome, Italy



ARIES: Users Policy

Ask for an account: aries@iss.it



Collaborative projects: stefano.morabito@iss.it; valeria.michelacci@iss.it

- ARIES is a collaborative platform. It is open and based on the social science concept
- We protect your data from hacking
- We don't backup your data (50GB workspace allocated to each user)
- We don't see your data (unless youre asking us to do so)
- Read the Disclaimer at: <https://w3.iss.it/site/aries/>

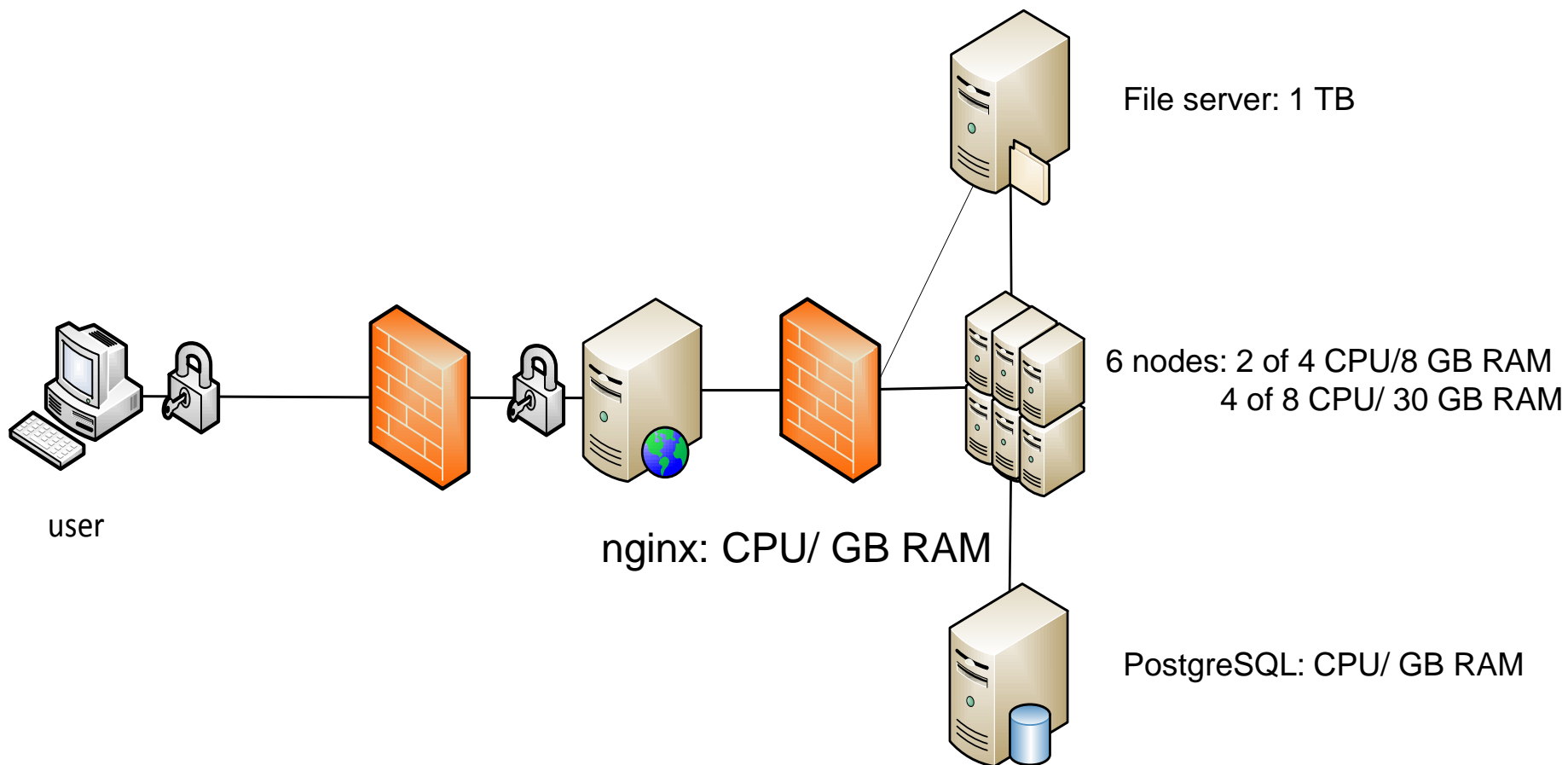
ARIES: Credits

The Galaxy ARIES core group at the EU RL for *E. coli*:

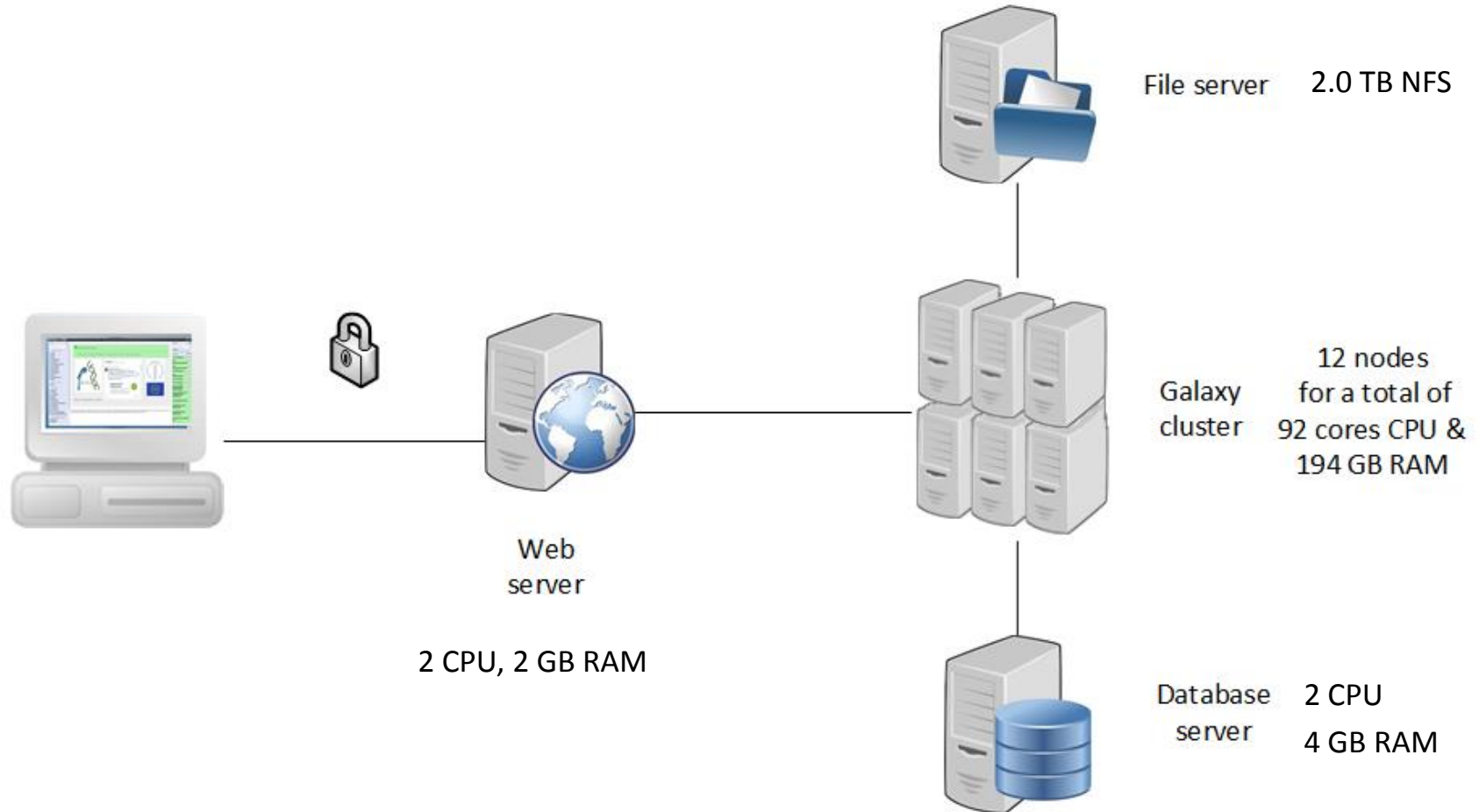
- Stefano Morabito (EU-RL VTEC): stefano.morabito@iss.it
ARIES Scientific coordination, tools design, contact person
- Arnold Knijn (SIDBAE): arnold.knijn@iss.it
ARIES Administrator, tools integration, contact person
- Valeria Michelacci (EU-RL VTEC): valeria.michelacci@iss.it
Tools design, contact person



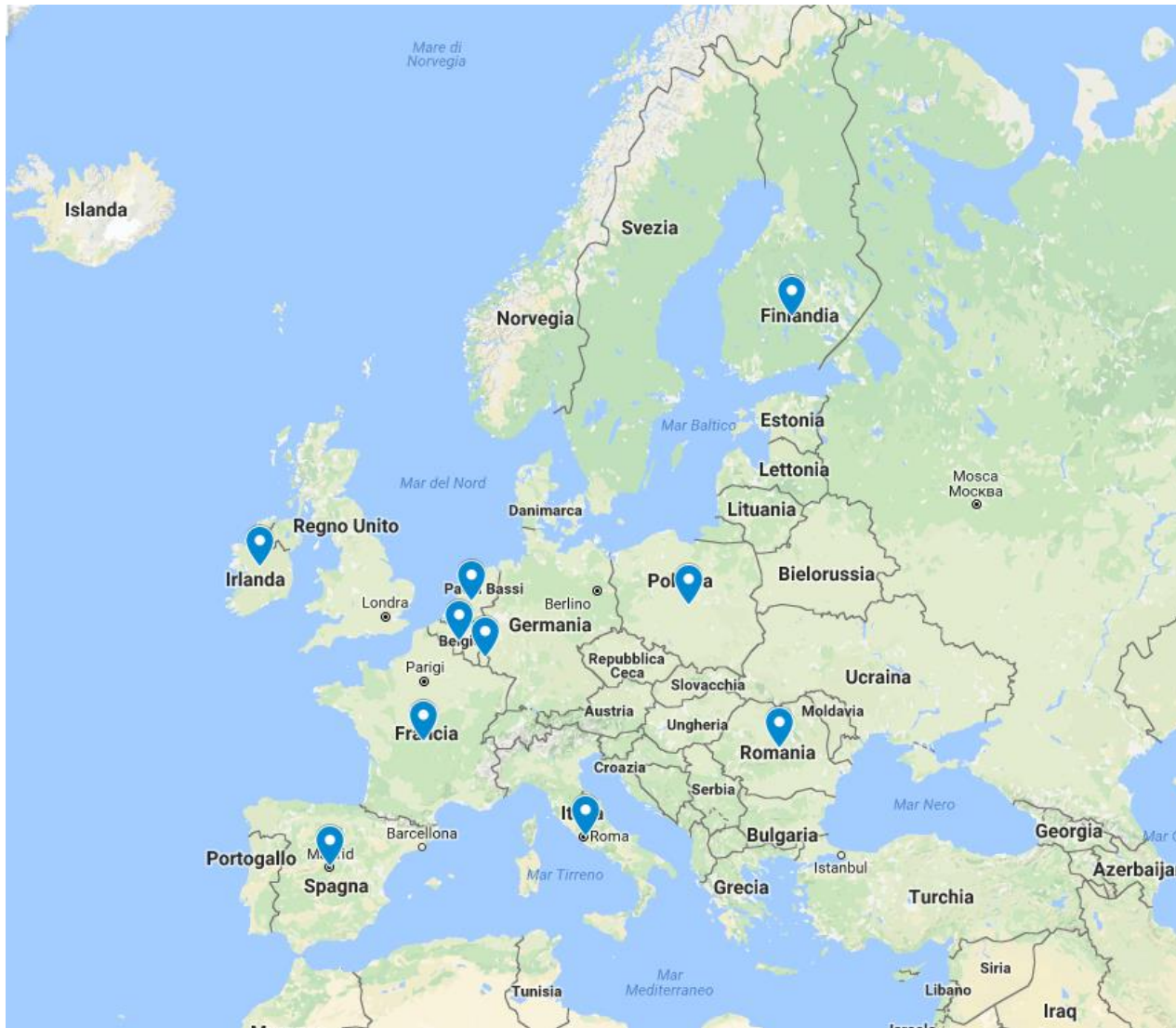
ARIES Under the hood – Workshop 2016



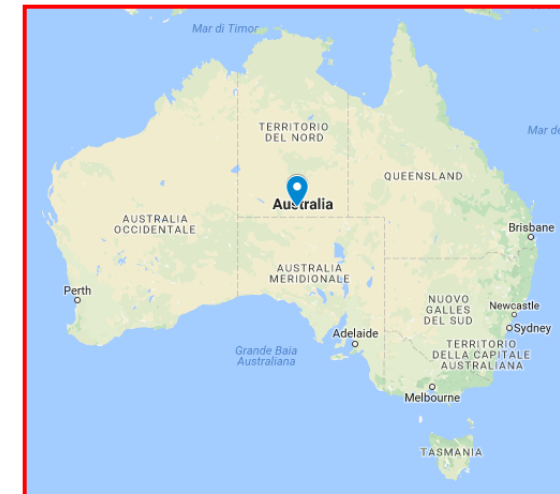
ARIES Under the hood – Workshop 2017



ARIES geographic spread – Workshop 2016



**Total:
50 users**



ARIES geographical spread – Workshop 2017

**Total 76
users**

Including NRLs
from 13 MSs:

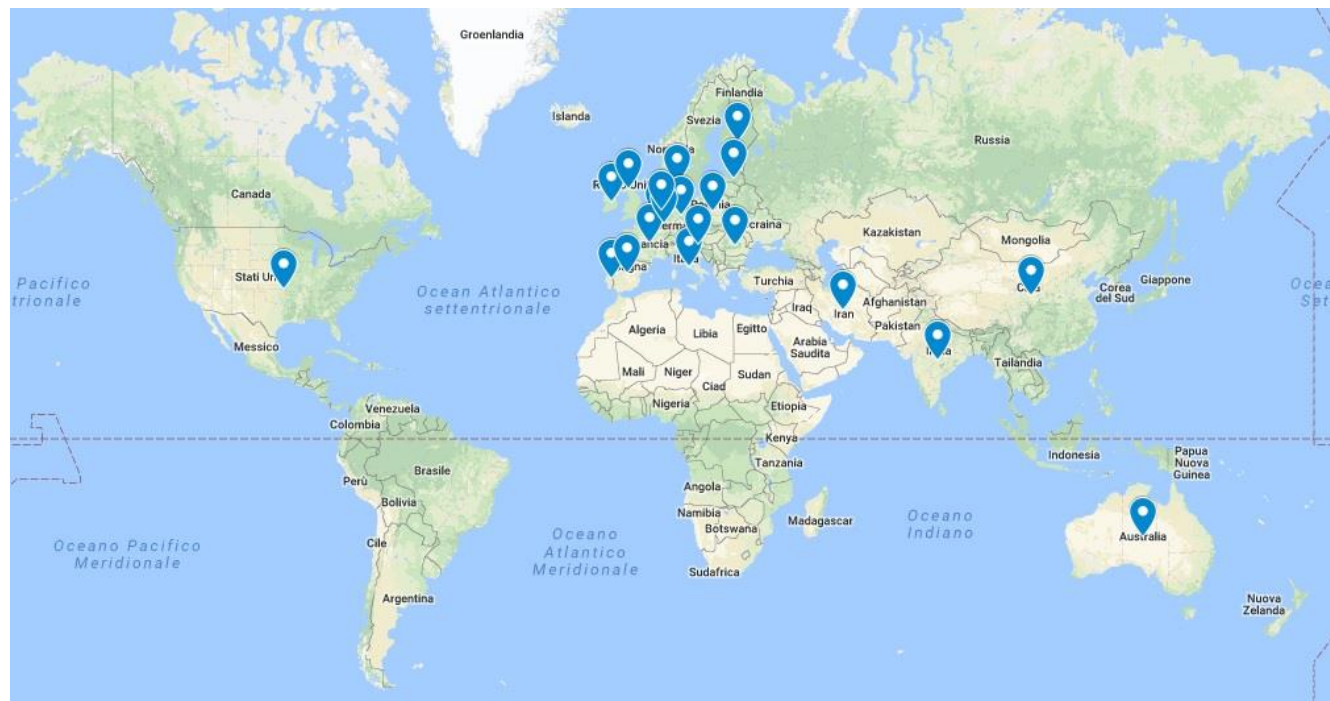
Belgium
Denmark
Finland
Germany
Ireland
Italy

Latvia
The Netherlands
Poland (2 NRLs)
Portugal
Slovenia
United Kingdom

Other users from
Europe:
France, Luxemburgh,
Romania, Spain

Users from
outside Europe:

Australia, Cina,
India, Iran, USA



EURL VTEC WGS PT pipeline

The screenshot shows the Galaxy / ARIES web interface. The top navigation bar includes 'Galaxy / ARIES Analyze Data', 'Workflow', 'Shared Data', 'Visualization', and 'Admin'. The left sidebar contains a 'Tools' section with a search bar and a list of tools under 'In Silico PCR' and 'E coli typing'. The main panel displays the configuration for the 'EURL VTEC WGS PT' tool. The tool title is 'EURL VTEC WGS PT EURL VTEC WGS PT' with a dropdown for 'Options'. Below the title, it states 'performs various typing tools (Galaxy Version 1.0)'. The configuration includes a dropdown for 'Is this single or paired library' set to 'Paired-end'. There are two 'FASTQ file #' fields: 'FASTQ file #1' with value '100: ED180_R1.fastq' and 'FASTQ file #2' with value '101: ED180_R2.fastq'. Both fields have a note: 'Must be of datatype "fastqsanger"'. An 'Execute' button is visible at the bottom of the configuration panel. Below the configuration, there is an 'Overview' section stating 'This tool is in beta test. The tool performs various typing tools:' followed by a list of tools: 'Raw data quality check', 'Trimming', 'Virulotyping', 'Multi Locus Sequence Typing', 'Assembly (A5 for Illumina and SPAdes for IonTorrent data)', 'Assembly statistics', and 'Serotyping'. The list is enclosed in a red box.

Galaxy / ARIES Analyze Data Workflow Shared Data Visualization Admin

Tools

search tools

In Silico PCR

E coli typing

[EURL VTEC WGS PT EURL VTEC WGS PT](#) performs various typing tools

[PHE MPileup](#) PHE SNP and indel caller

[Map with BWA](#) - map short reads (< 100 bp) against reference genome

[Map with BWA-MEM](#) - map medium and long reads (> 100 bp) against reference genome

[VCFs to fasta](#) Takes a set of VCF files and outputs a multi fasta file with only the variant positions.

[Filter VCF](#) filters a VCF file

---MLST---

agMI ST by Assembly

EURL VTEC WGS PT EURL VTEC WGS PT Options

performs various typing tools (Galaxy Version 1.0)

Is this single or paired library

Paired-end

FASTQ file #1

100: ED180_R1.fastq

Must be of datatype "fastqsanger"

FASTQ file #2

101: ED180_R2.fastq

Must be of datatype "fastqsanger"

Execute

EURL VTEC WGS PT Overview This tool is in beta test. The tool performs various typing tools:

- Raw data quality check
- Trimming
- Virulotyping
- Multi Locus Sequence Typing
- Assembly (A5 for Illumina and SPAdes for IonTorrent data)
- Assembly statistics
- Serotyping



Report for ED1182_06062017

2017-06-08 16:10 UTC

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

Report and output

Summary

O76:H19
 STNF
 ehxA, stx1A, stx1B, stx2A, stx2B

Raw data quality check

FASTQC result: [Webpage](#)

Trimming analysis

Maximum length trimming	360
Left-side trimming	10
Right-side trimming	0
Minimum Phred quality score for right-side trimming	25
Average Phred quality score for right-side trimming	27
Minimum length filtering	50

Assembly statistics

Estimated genome size:	5000000 bp
Assembled nucleotides:	5322187 bp
Estimated coverage:	1.06 x
N. contigs:	166
Average contig length:	32061

Serotyping

sseqid	pident	length	positive
wzx_80_Og76_O76	99.60	1239	1234
wzy_66_Ab812031_O76	99.59	1233	1228
fliC_282_AY337479_H19	99.95	1832	1831
fliC_102_AY250002_H19	100.00	1202	1202

Multi Locus Sequence Typing

Sample	ST	adk	fumC	gyrB	icd	mdh	purA	recA	mismatches	uncertainty	depth	maxMAF
dataset_B3309	NF	6	23	32	402	9	8	7	0	-	79.9397142857	0.181818181818

Virulotyping

virulence gene	coverage
cba:7:FJ664719	35.4688
cma:8:FJ664724	11.6373
ehxA:11:EF204925	77.5756
epeA:1:AY258503	0.830392
espl:1:AJ278144	125.905
espP:2:AY258503	1.4125
gad:13:CU928160	53.2741
iha:19:AJ278144	133.536
ireA:1:CU928162	93.2172
lpfA:3:AP010953	57.6702
pet:1:AF056581	0.279064
pic:1:AE014075	96.3387
senB:3:CP000038	44.3197
stx1A:9:AJ314838:c	24.9916

NGS COURSE

388 shown, 100 deleted, 81 hidden

7.85 GB



537: EURL VTEC

WGS PT on data

517: report



528: SPAdes contigs (fasta)

523: EURL VTEC

WGS PT on data

517: trimmed FASTQ



520: SRST2-7loci Alignment File

519: SRST2-7loci Alleles Table

518: virulotyper on data 517: mapping reads

517: ED1182_06062017



WGS - Proficiency test 1

6 samples (the same of PT-PFGE-6)

Shipment scheduled for
13 November 2017
together with PT20

PT-WGS on a
voluntary base

23 participants:
18 NRLs
&
5 Official Labs from Italy

EURL Analysis

WGS data uploaded in the
EURL-PT database



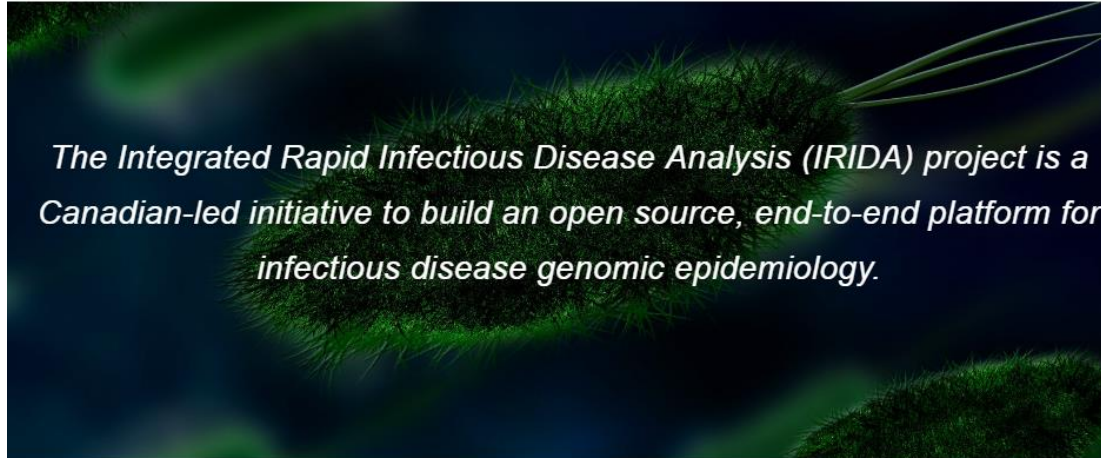
Ref-free & Ref-based
phylogeny

ALL vs ALL

**Identify average
variation in terms
of SNPs between
different labs and
platforms**



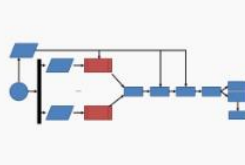
IRIDA database structure



HIGHLIGHTS



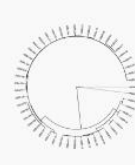
Sequence Data Management



Workflows



Data Integration



Visualizations and Tools

IRIDA instance on
ISS webserver

Possibility to create
**projects shared
with groups of
users**

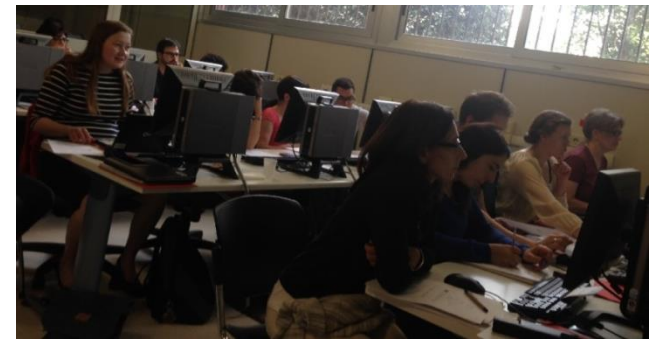
Participants will be
provided with an
account and
instructions to add
sequencing data to a
shared project



Courses on Bioinformatics for NGS data mining

22 attendees in 2015 (7 NRLs, EFSA)

22 attendees in 2016 (6 NRLs, EFSA, ECDC)



EU Reference Laboratory for *E. coli*
Department of Veterinary Public Health and Food Safety
Unit of Foodborne Zoonoses
Istituto Superiore di Sanità



Program for a 3-days training at the EURL-VTEC, Istituto Superiore di Sanità, Rome,
on the use of bioinformatics tools for Next Generation Sequencing data mining
for typing pathogenic *E. coli*

11 attendees (9 NRLs)



Istituto Superiore di Sanità, Dep. of Food Safety, Nutrition and Veterinary Public Health
European Union and National Reference Laboratory for *E. coli*, Rome, Italy



Training at the EURL



European Union Reference Laboratory VTEC

Training at the EURL-VTEC

According to Art. 32 of Regulation (EC) 882/2004 defining the functions and duties of the EU Reference Laboratories, the EURL-VTEC conducts laboratory training for the staff of the NRLs of EU Member States or third countries. The training sessions are based on a hands-on approach and consist in visits to our laboratory of trainees interested in improving their knowledge about methodologies and techniques for pathogenic *E. coli* detection, identification, and typing. A dedicated budget from the European Commission is usually allocated to support visits of scientists from EU NRLs and each year the EURL launches a call to collect applications from the NRLs.

Five different training programs are currently available at EURL-VTEC and are reported below.

Documents

Program for a 5-days training on the detection of VTEC in food matrices according to the ISO TS 13136:2012 and the characterization of the isolated VTEC strains (EU-RL VTEC_Training Program_ISOTS13136_Rev 2) [PDF - 46.69 kbytes]



Program for a 4-days training on the identification and characterization of the different groups of pathogenic *E. coli* by Real Time PCR amplification of their virulence genes (EU-RL VTEC_Training Program_Identification of pathogenic *E. coli*_Rev 1) [PDF - 44.66 kbytes]



Program for a 5-days training on molecular typing of VTEC by PFGE (EU-RL VTEC_Training Program_PFGE typing_Rev 2) [PDF - 123.83 kbytes]



Program for a 5-days training on the design and preparation of PTs on the detection of VTEC in food matrices (EU-RL VTEC_Training Program_PT organization_Rev 0) [PDF - 129.72 kbytes]



Program for a 3-days training on the use of bioinformatics tools for Next Generation Sequencing data mining for typing pathogenic *E. coli* (EURL-VTEC_Training Program_Bioinformatics_Rev 0) [PDF - 120.32 kbytes]



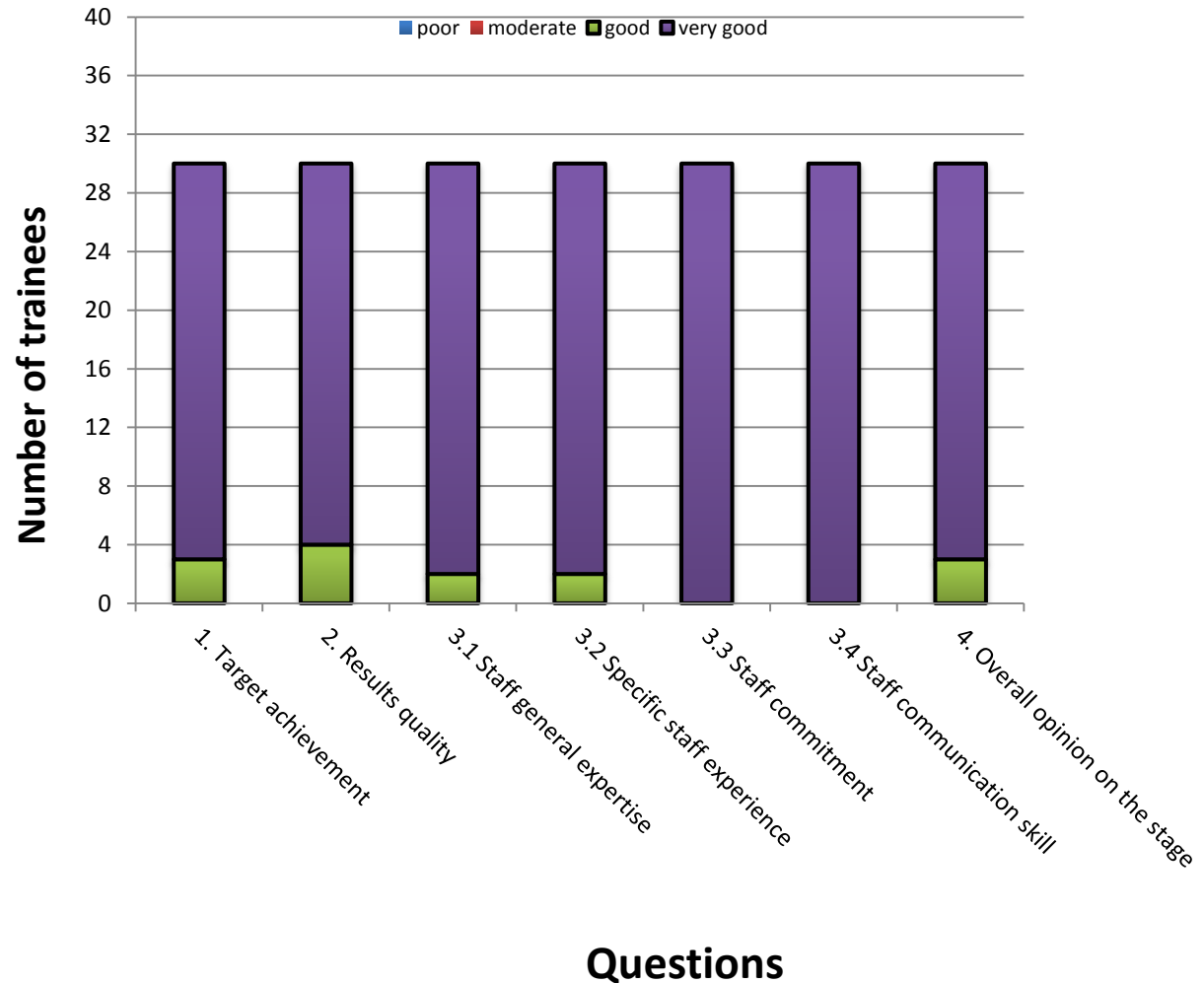
Training satisfaction - 2017

Trainees in all the training modules

30 trainees in 2017
(11 WGS data analysis)

Tot 10 trainees on EU budget

- PTs organization
- ISO TS 13136
- PFGE
- WGS data analysis



Training satisfaction - 2016

Trainees in all the training modules

12 trainees in 2016

Tot 6 trainees on EU budget

- PTs organization
- ISO TS 13136
- PFGE

