

Implementing WGS data collection and analysis for communicable diseases at EU/EEA level

Erik Alm, Senior Expert Applied Molecular Epidemiology Presented by: Saara Kotila, Scientific Officer Molecular Surveillance

STEC activities at ECDC

- Annual case-based data collection and reporting at the EU/EEA-level
- Contribution to EFSA scientific opinions on STEC
- WGS support for possible multi-country events through laboratory contractor
- WGS analysis tools for outbreak investigations



Figure 2. Distribution of confirmed cases of Shiga toxin/verocytotoxin -producing *Escherichia coli* (STEC/VTEC) -infection by month, EU/EEA, 2012–2017



Source: Country reports from Austria, Cyprus, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, the Netherlands, Norway, Poland, Romania, Slovakia, Slovenia, Spain, Sweden and the United Kingdom.

Indicator-based and event-based surveillance: **TESSy and EPIS**



Notified cases

Whole genome sequencing (WGS) at ECDC



ECDC strategic framework for WGS by 2021:

- Cross-border outbreak investigations
- Control-oriented surveillance (cross-border outbreak detection)
- Strategy-oriented surveillance

Extensive list of diseases and pathogens to be implemented

- Food- and waterborne diseases
- Vaccine-preventable diseases
- Multidrug-resistant tuberculosis
- Antimicrobial resistance
- Influenza
- Any emerging infectious disease threat at the EU level



ECDC WGS system upgrade, design principles



- Re-use existing tools and resources for storage, calculation and visualisation
- Integrate data from several different calculation systems in the same Data Warehouse data model
- Encourage open data sharing while also fulfilling all data protection and access control requirements
- Focus on the user experience and service-oriented architecture (data submission and data exploration)



Data flow schematic





Building blocks



Built by ECDC:

- Data upload interface (API and GUI), TESSy
- Data warehouse for derived data, new
- Web service API for data consumption, new
- Cluster detection scripts, existing
- Web page to hold visualisation components, new

Re-used:

- Storage (currently ENA and network drive, cloud storage to be implemented in 2020)
- Calculation systems and operational databases (currently Bionumerics and BIGSdb, more will be added as needed)
- Interactive visualisation (currently MicroReact, we are also looking at NextStrain)
- Charts and tables libraries (charts.js, datatables.js)





ECDC upgraded system expected to be operational Q1 2020

Real-time WGS data collection opened for *Listeria* in 2019

In 2020 we expect to have support for continous surveillance for *Listeria, Salmonella*, STEC, *Neisseria meningitidis*, Influenza and MDR-TB

Outbreak support for STEC

ECDC-EFSA joint database implementation will start Q1 2020



ECDC tool for analysing WGS data for clusters



Visualisation Molecular clusters/datasets My data Settings

These table shows your uploaded data, you can choose to share or not share individual isolates with other users.

Save settings Reload data from source (slow, you must refresh your browser to see the changes)

Column visibility Copy CSV Show 100 T entries

					LISTISO subr	nissions				
Countr y	BatchId 🔶	RecordId 🔶	Created date 🛛 🔻	Modified date	QC (core loci detect ed)	QC (PASS/FAIL)	DateUsedForStatisti 🖨	Cluster code	Event 🔶	Share assembly
Seard	Search Batc	Search RecordId	Search Created date	Search Modified date	Search QC (core loci	Search QC (PASS/FA	Search DateUsedFor!	Search Cluster code	Search Event	Search Share assemi
-			2019-09-19 14:52:02	2019-09-25 08:44:34	97.76	PASS	2019-06-18			
			2019-09-19 14:52:01	2019-09-25 08:44:31	98.56	PASS	2019-05-13			۲
			2019-09-19 14:52:01	2019-09-25 08:44:30	98.45	PASS	2019-05-26	2019-04.LIST.03.CC570		۲
			2019-09-18 18:12:43	2019-09-25 08:44:32	100.0	PASS	2019-08-08			۲
			2019-09-18 18:12:43	2019-09-25 08:44:28	100.0	PASS	2019-08-28	2017-10.LIST.82.CC1.Asc I.0053.ApaI.0044		×
			2019-09-18 17:02:27	2019-09-20 15:13:34	99.94	PASS			UI-490 Cluster B ST32	۲
			2019-09-18 17:02:27	2019-09-20 15:13:39	99.88	PASS			UI-490 Cluster B ST32	
			2019-09-18 17:02:27	2019-09-20 15:13:43	99.94	PASS			UI-490 Cluster B ST32	
			2019-09-18 12:19:37	2019-09-25 08:44:29	99.94	PASS	2018-09-20	2018-10.LIST.18.ST32		
			2019-09-18 12:19:37	2019-09-25 08:44:29	99.94	PASS	2019-07-18	2018-10.LIST.18.ST32	UI-490 Cluster B ST32	
			2019-09-18 12:19:36	2019-09-25 08:44:28	99.82	PASS	2019-06-15	2018-10.LIST.18.ST32	UI-490 Cluster B ST32	
			2019-09-06 10:53:35	2019-09-20 10:16:49	99.88	PASS		2019-09.LIST.01.CC388	<u>UI594</u>	
			2019-08-27 16:23:42	2019-08-28 14:30:00	98.51	PASS		2019-05.LIST.06		
			2019-08-27 16:23:42	2019-09-25 09:23:31	97.99	PASS	2019-04-17	2017-10.LIST.01.CC87		
			2019-08-27 16:23:42	2019-09-25 09:23:41	97.94	PASS	2019-04-19	2017-10.LIST.01.CC87		

Search:



Prototype version 0.17 Powered by <u>MicroReact</u>

LISTISO EBOV IMD ZIKV

Visualisation Molecular dusters/datasets My data Settings

These tables shows molecular clusters detected by ECDC by applying automated algorithms, see respective surveillance protocols for details. Only multi-country clusters and clusters in your country are shown. Included are also generated datasets of particular interest.

Cluster ID 🕴	Event ID 🔶	Cluster status 🛛 🔻	Number of isolate s	Countries 🔶	Updated (7d)	First date 🕴	Last date 🔻 🔻	Last updated 🕴	Extended cluster (7AD)	Neighborhood (50 AD)	
Search Cluster ID	Search Event ID	Search Cluster stat	Search Number of	Search Countries	Search Updated (7	Search First date	Search Last date	Search Last update	Search Extended c	Search Neighborhc	
2017-10.LIST.82.CC1.A scI.0053.ApaI.0044		OPEN	5			2012-07-15	2019-08-28	2019-09-25	2017-10.LIST.82.CC1. AscI.0053.ApsI.0044	2017-10.LIST.82.CC1. AscI.0053.ApaI.0044	
2018-10.LIST.18.ST32	UI-490 Cluster B ST32	OPEN	8			2018-01-27	2019-07-18	2019-09-25	2018-10.LIST.18.ST32	2018-10.LIST.18.ST32	
2019-03.LIST.02.CC1		OPEN	8			2018-02-25	2019-06-12	2019-08-06	2019-03.LIST.02.CC1	2019-03.LIST.02.CC1	
2017-10.LIST.01.CC87		OPEN	37			2011-12-13	2019-06-05	2019-09-25	2017-10.LIST.01.CC87	2017-10.LIST.01.CC87	
2019-06.LIST.10.CC8	<u>UI563</u>	OPEN	5			2019-05-15	2019-05-15	2019-07-09	2019-06.LIST.10.CC8	2019-06.LIST.10.CC8	
2017-10.LIST.69.CC1		OPEN	3			2010-08-27	2019-02-19	2019-04-30	2017-10.LIST.69.CC1	2017-10.LIST.69.CC1	
2016-11.LIST.01	<u>UI452</u>	OPEN	25			2014-07-23	2019-02-18	2019-05-29	2016-11.LIST.01	2016-11.LIST.01	
2017-06.LIST.70.CC1.A scI.0053.ApaI.0040		OPEN	48			2015-08-26	2019-02-18	2019-07-03	2017-06.LIST.70.CC1. AscI.0053.ApsI.0040	2017-06.LIST.70.CC1. AscI.0053.ApsI.0040	
2019-07.LIST.14.CC87		OPEN	2			2018-08-15	2018-11-30	2019-08-07	2019-07.LIST.14.CC87	2019-07.LIST.14.CC87	
2019-07.LIST.12.CC2		OPEN	2			2014-09-26	2018-11-26	2019-08-07	2019-07.LIST.12.CC2	2019-07.LIST.12.CC2	
2019-07.LIST.13.CC6		OPEN	2			2013-10-15	2018-11-21	2019-08-07	2019-07.LIST.13.CC6	2019-07.LIST.13.CC6	
2018-05.LIST.07.CC15 5		OPEN	7			2015-08-30	2018-10-18	2019-08-07	2018-05.LIST.07.CC15 5	2018-05.LIST.07.CC15 5	
2015-04.LIST.04.CC6.A scI.0002.ApsI.0072	UI583_test	OPEN	26			2010-01-27	2015-03-26	2019-09-10	2015-04.LIST.04.CC6. AscI.0002.ApsI.0072	2015-04.LIST.04.CC6. AscI.0002.ApsI.0072	
2019-06.LIST.11.CC6	<u>UI552</u>	OPEN	3					2019-07-09	2019-06.LIST.11.CC6	2019-06.LIST.11.CC6	
2019-08.LIST.01.CC9		CLOSED	2			2018-01-17	2019-06-04	2019-08-06	2019-08.LIST.01.CC9	2019-08.LIST.01.CC9	
2019-04.LIST.03.CC57 0		CLOSED	3			2019-05-26	2019-05-26	2019-09-25	2019-04.LIST.03.CC57	2019-04.LIST.03.CC57	
2017-11.LIST.11.CC87		CLOSED	7			2015-05-30	2019-05-01	2019-07-10	2017-11.LIST.11.CC87	2017-11.LIST.11.CC87	
2019-03.LIST.07.CC6		CLOSED	3			2017-03-24	2019-02-20	2019-07-08	2019-03.LIST.07.CC6	2019-03.LIST.07.CC6	
2019-07.LIST.01.CC2		CLOSED	2			2016-05-24	2019-02-10	2019-07-10	2019-07.LIST.01.CC2	2019-07.LIST.01.CC2	



Powered by MicroReact

Visualisation Molecular clusters/datasets My data Settings



MicroReact Tables and figures Distance matrix Isolate table and download





Prototype version 0.17 Powered by <u>MicroReact</u>

LISTISO EBOV IMD ZIKV

Visualisation Molecular clusters/datasets My data Settings



Pathogen: MicroReact Tables and figures Distance matrix Isolate table and download Listeria T Clustering method: Single linkage Isolate distribution by DateUsedForStatistics Include data from SRA and ENA 5 Include data from all countries Sampled within X days: 4 Uploaded within X days: Search for specific cluster/dataset: 3 2017-10.LIST.01.CC87 Search for specific UI: 2 Search for specific RecordIds: 1 After filtering, also include all matches within X genetic differences: 0 7 2011-12 12 2012 04 2012 02 2012 12 2013 04 2013 02 2013 12 2014 04 2014 02 2014 12 2015 04 2015 02 2015 12 2016 04 2016 02 2015 12 2015 04 2015 02 2015 12 2015 04 2015 02 2015 12 2015 04 2015 02 2015 12 2015 04 2015 02 2015 12 2015 04 2015 02 2015 12 2015 04 2015 02 2015 12 2015 04 2015 02 2015 12 2015 04 2015 02 2015 12 2015 04 201 Update Month

URL for sharing: http://microreact.ecdcnet.e URL for recreating: http://zbiodev1.idmdevdmz









Software options for WGS data upload

17

ENA run

- ECDC WGS upload application
- ECDC Bionumerics client plugin

• Direct TESSy submission

All three options include mandatory submission to TESSy isolate-based subjects

C Dat	and a second sec											
1] 🔸 🌽 🗙	* 🖌	<all entries=""></all>	8	_		_	_	_	_	_	
	Key	Level	Modified date	MLST PubMLST A	ReportingCountry	DateUsedForStatistics	DateOfSampling	DateUploaded	DateReleased	SampleOrigin	InsdcR 💌	
1	SRR5193935		2019-01-30 00:00:09		UK	2015-07-01	2015-07-01		2017-01-24	Homo sapiens		
1	SRR5193925		2019-01-29 20:06:00		UK	2016-11-01	2016-11-01		2017-01-24	Homo sapiens		
1	SRR5193934		2019-01-29 21:25:04		UK	2016-09-01	2016-09-01		2017-01-24	Horno sapiens		
1	SRR5193926		2019-01-30 03:16:06		UK	2016-12-01	2016-12-01		2017-01-24	Horno sapiens		
1	SRR5193931		2019-01-30 00:05:38		UK	2016-10-01	2016-10-01		2017-01-24	Homo sapiens		
1	SRR5193929		2019-01-29 21:41:17		UK	2016-08-01	2016-08-01		2017-01-24	Homo sapiens		
1	SRR5193927		2019-01-29 18:55:37		UK	2016-08-01	2016-08-01		2017-01-24	Horno sapiens		
1	SRR5193922		2019-01-29 17:50:59		UK	2016-04-01	2016-04-01		2017-01-24	Homo sapiens		
1	SRR5193923		2019-01-30 02:31:41		UK	2016-06-01	2016-06-01		2017-01-24	Homo sapiens		
1	SRR5193920		2019-01-30 09:21:15		UK	2016-09-01	2016-09-01		2017-01-24	Homo sapiens		
1	SRR5193932		2019-01-30 01:30:06		UK	2016-09-01	2016-09-01		2017-01-24	Homo sapiens		
1	SRR5193947		2019-01-30 05:04:05		UK	2016-06-01	2016-06-01		2017-01-24	Homo sapiens		
1	SRR5193946		2019-01-30 09:21:06		UK	2016-12-01	2016-12-01		2017-01-24	Homo sapiens		
1	SRR5193945		2019-01-30 09:21:12		UK	2016-05-01	2016-05-01		2017-01-24	Homo sapiens		
1	SRR5193943		2019-01-30 09:21:15		UK	2016-11-01	2016-11-01		2017-01-24	Horno sapiens		



tile Edit Database Analysis Scrints ECDC WGS WGS tools Window Hel

🐙 🐳 🥅 🚙 🐴 V

ECDC WGS upload application



How to use

- Download and install the application
- Read the manual and configure the data upload for each pathogen, personal support is available through the ECDC FWD mailbox
- Configure data import/mapping from Excel/csv, your LIMS system, or enter data directly
- Start uploading in a few clicks according to the reporting protocol
- The epidemiological data are submitted to TESSy only, the WGS reads/assembly is submitted to the selected system(s)



Features

- Can be configured to import data from databases or local files (MySQL, SQL Server, SQLite, Excel, csv)
- Configure only once, single click upload
- Can upload assemblies to TESSy and SFTP, raw reads to SFTP and ENA (configurable)
- Data sharing through SFTP and ENA

