SCIENTIFIC ADVICE

Expert opinion on whole genome sequencing for public health surveillance

Strategy to harness whole genome sequencing to strengthen EU outbreak investigations and public health surveillance

www.ecdc.europa.eu

ECDC SCIENTIFIC ADVICE

Expert opinion on whole genome sequencing for public health surveillance

Strategy to harness whole genome sequencing to strengthen EU outbreak investigations and public health surveillance

This report of the European Centre for Disease Prevention and Control (ECDC) was prepared by Marc Struelens (coordinator), Barbara Albiger, Mike Catchpole, Bruno Ciancio, Denis Coulombier, Laura Espinosa, Karin Johansson, Csaba Ködmön, Daniel Palm, Joana Revez and Ivo Van Walle.

Contributing authors

Frank Aarestrup (Technical University of Denmark, Denmark), Franz Allerberger (Austrian Agency for Health and Food Safety, Austria), João André Carriço (University Lisbon, Portugal), Arta Balode (Paul Stradina University Hospital, Latvia), John Besser (CDC, US), Hans Blystad (National Institute of Public Health, Norway), Vitor Borges (Instituto Nacional de Saúde Dr. Ricardo Jorge, Portugal), Sylvain Brisse (Institut Pasteur Paris, France), Rosa Cano Portero (National Centre of Epidemiology, Spain), Dominique Caugant (National Institute of Public Health, Norway), Bruno Coignard (French Public Health Agency, France), Hajo Grundmann (Groningen University Medical Center, Netherlands), Susan Hahné (National Institute for Public Health and the Environment, Netherlands), Dag Harmsen (University of Munster, Germany), Thea Kølsen Fischer (Statens Serum Institut, Denmark), Marion Koopmans (Erasmus Medical Centre, Netherlands), Pavla Krizova (National Institute of Public Health, Czech Republic), Martin Maiden (University Oxford, UK), Eva Moeller Nielsen (Statens Serum Institut, Denmark), Joël Mossong (National Health Laboratory, Luxembourg), Annalisa Pantosti (Instituto Superiore de Sanità, Italy), Maja Rupnik (Institute of Public Health Maribor, Slovenia), Leo Schouls (National Institute for Public Health and the Environment, Netherlands), Magnus Unemo (Orebro University Medical Centre, Sweden), Steven Van Gucht (Scientific Institute of Public Health, Belgium), Anne-Mieke Vandamme (Katholieke Universiteit Leuven, Belgium), and Guido Werner (Robert Koch Institute, Germany).

A Specific Declaration of Interests was provided by external experts in accordance with the ECDC Independence policy and the Implementing rules on Declarations of Interests and no conflict of interest was identified.

Suggested citation: European Centre for Disease Prevention and Control. Expert opinion on whole genome sequencing for public health surveillance. Stockholm: ECDC; 2016.

Stockholm, August 2016

ISBN 978-92-9193-888-9

doi 10.2900/12442

Catalogue number TQ-01-16-576-EN-N

Contents

Contents iii
Abbreviations iv
Glossary iv
Executive summary 1
1. Background 2
2. Methods 3
3. Results and discussion 4
4. Conclusions and potential implications 10
References 11
Appendix 1. Disease-specific ECDC pilot projects and expert opinions on WGS, as of November 2015 14

Figures

Figure 1. Number of EU/EEA countries applying or planning to apply WGS-based typing for national surveillance within three years, by disease target (categories not mutually exclusive), July 2015 6
Figure 2. Process overview – WGS data production, analysis and integration with epidemiological data for public health surveillance 6

Tables

Table 1. Organisations and initiatives supporting the development and application of WGS-based typing for public health protection 5

Abbreviations

CPE	Carbapenemase-producing *Enterobacteriaceae*
ECDC	European Centre for Disease Prevention and Control
EFSA	European Food Safety Authority
ENA	European Nucleotide Archive
EQA	External Quality Assessment
ERLTB-Net	European Network of Reference Laboratories for Tuberculosis
EU/EEA	European Union/European Economic Area
Euro-GASP	European Gonococcal Antimicrobial Susceptibility Programme
EuSCAPE	European Survey of Carbapenemase-producing *Enterobacteriaceae*
FWD	Food- and waterborne diseases
IBD-LabNet	Invasive Bacterial Disease Laboratory Network
MIRU-VNTR	Mycobacterial Interspersed Repetitive Units – Variable Number Tandem Repeat
MLST	Multi-Locus Sequence Typing
MLVA	Multiple-Locus Variable-number tandem repeat Analysis
NGS	Next-Generation Sequencing
PFGE	Pulsed-Field Gel Electrophoresis
WGS	Whole Genome Sequencing

Glossary

Accessory genome	Subset of genes within a genome that are variably present in the genomes of a given species.
Core genome	Subset of genes within a genome that is common to all strains of a given species.
Federated database	A pool of databases which provide hierarchical reporting to different stakeholders.
k-mers	Very short overlapping sequences typically of around 15 base lengths.
Mobilome	All mobile genetic elements in a genome.
Next generation sequencing	A range of high-throughput technologies that perform massive parallel sequencing of large numbers of sequences.
Raw reads	Sequence data on small random fragments of a genome generated by a DNA sequencer.
Resistome	All the genes and gene variants in a bacterium that confer resistance to antibiotics.
Virulome	The set of genes that contribute to the virulence of a bacterium.
Whole genome sequencing	Sequence determination of the near-complete (typically 95–99% of the total) length of the full genome of a microorganism.

Executive summary

Whole Genome Sequencing (WGS) has become the reference microbial typing method in outbreak studies and is increasingly applied to national surveillance of infectious diseases in EU/EEA countries and beyond. In 2015, ECDC developed an expert opinion on WGS in consultation with multidisciplinary experts. The resulting strategy envisages that within five years ECDC will have contributed to the establishment of standards and systems enabling the EU-wide use of WGS as the method of choice for typing of microbial pathogens, replacing other methods. This will improve the accuracy and effectiveness of disease surveillance, outbreak investigation and evaluation of prevention policies by enhanced assessment of disease and drug resistance transmission dynamics. One of the key elements of ECDC's strategy to realise this vision are the fostering of a multidisciplinary interpretation of the information arising from the combination of epidemiological data and pathogen sequence characterisation to guide public health action. Another key element of the strategy is for ECDC to work together with public health laboratory network partners, EFSA and research and development projects to define the priorities and identify high impact diseases or drug resistance issues, where genome sequence information can make a difference for public health intervention. The strategy will contribute to a global agreement on WGS analytical approaches, epidemiological interpretation criteria and genomic nomenclature by surveillance objective, while retaining flexibility in order to explore improved methods. It will also help with multi-country evaluation of the public health effectiveness of WGS-based typing by measuring outcomes in terms of disease prevention and it will focus training efforts on developing a new, integrative 'genomic epidemiology' discipline, thereby building a common understanding which translates into public health risk assessment.

1. Background

In September 2015, a Joint Strategic consultation with EU/EEA Member State stakeholders advised ECDC to propose a strategy to guide and facilitate the transition across the European Union to the use of Whole Genome Sequencing (WGS) for microbial typing in disease surveillance and outbreak investigations.

This expert opinion addresses the following questions:

- What are the strengths and weaknesses of WGS-based typing for multi-country outbreak detection, investigation and disease surveillance, compared to other molecular typing methods?
- What is the current state and medium-term outlook for the development, harmonisation and inter-laboratory portability of WGS-based typing for epidemiological investigation of the diseases and antimicrobial-resistant pathogens shortlisted for integrated EU surveillance?
- What is the strategy and proposed role of ECDC in collaboration with the EU and global players leading the technical development and public health applications of WGS-based typing?

2. Methods

ECDC experts drafted a first version of ECDC strategy document, examining the advantages and limitations of WGS-based typing for public health applications to provide strategic direction for its EU-wide implementation, based on a narrative review of the scientific literature, previous expert opinions and technical experience gained in national public health institutes and European networks of laboratories. For this purpose, a literature review was performed, based on a search in PubMed for original and review articles published in English between January 2011 and November 2015, using the key words ["whole genome sequencing" OR "next-generation sequencing"] AND ["surveillance" OR "outbreak" OR "survey"]. Based on abstract review of retrieved publications by one author (coordinator), the most relevant articles were selected for a narrative review on the strengths and weaknesses of WGS-based typing for multi-country outbreak detection, investigation and disease surveillance, compared to other molecular typing methods. Grey literature references, including ECDC and EFSA expert meeting reports and expert opinions on public health applications of WGS, were also included in the background information. In October 2015, the draft report version 1 was reviewed by members of ECDC Microbiology and Surveillance Steering committees and revised in accordance with comments provided.

Between June and September 2015, ECDC conducted a questionnaire survey of the National Microbiology Focal Points for the EU/EEA countries, to map current access as of July 2015 to Next-Generation Sequencing (NGS) technology, routine capacity and short-term (within three years) development plans to use WGS-based methods for surveillance and outbreak investigation in the public health laboratories of EU/EEA countries.

In November 2015 the second version of the report was shared for external review by an expert group of multidisciplinary experts and representatives from the Member States. Written comments were elicited and a face-to-face-meeting was held on 19 November 2015 in Stockholm to discuss and revise the draft strategy document, based on presentations of the latest experience on public health use of WGS technology for priority diseases under EU surveillance. The expert group was composed of 17 participants from different EU/EEA countries and the USA, invited on the basis of their scientific and operational expertise in bacteriology, virology, antimicrobial resistance, epidemiology and public health, and 11 members of the Molecular Typing for Surveillance Task Force (MSTF), an ECDC technical advisory group representing National Microbiology Focal Points or National Focal Points for Surveillance. In accordance with ECDC's independence policy, specific declaration of interest were provided by all participants. The presentations from leading translational research groups and public health partners showed that a number of validated WGS analytical tools are now available for public health applications. ECDC was invited to be part of a global discussion on how to achieve common analytical approaches and interpretation criteria for disease surveillance. During the consultation there were further discussions on specific sections of the report: the technical requirements for multi-centre sharing of genomic information for public health use and the proposed strategies for ECDC to support the technology transition in the Member States. These sections were reformulated by consensus agreement with the meeting participants and references to recent original publications were added by the experts consulted.

After the meeting, the final version 3 of the report was shared for comments with the expert meeting participants. In March 2016, it was approved with minor modifications by the ECDC Advisory Forum through a written consultation process.

3. Results and discussion

Whole genome sequencing - a transforming technology

The use of molecular typing information for epidemiological studies and public health investigations of infectious disease is rapidly shifting to Next-Generation Sequencing (NGS) to replace DNA restriction/MLVA typing techniques and classical (Sanger method) single gene sequencing techniques [1-3]. Comparative analysis of Whole Genome Sequencing (WGS) is now the reference typing method used in outbreak studies, thanks to its maximal phylogenetic resolution for disclosing transmission networks and tracing infection sources [1-17]. In many countries WGS-based typing is in the trial implementation phase for use as the routine first-line or second-line typing method for national surveillance of a number of bacterial and viral diseases. This includes use for the purposes of early outbreak detection [18-31] and detection of the emergence and monitoring of the evolution/dynamics of multi-drug resistant pathogen spread [32-39]. It is noteworthy that the available evidence supporting the public health utility of WGS-based typing (as of November 2015) derived mostly from retrospective studies that did not measure the impact of the intervention on population health. Similarly, there have been no cost-effectiveness studies to evaluate the system-wide implementation of the technology for public health application.

The advantages of WGS-based typing over other pathogen typing methods include:

- the optimal resolution of the near-complete genomic sequence comparison for measuring inter-genomic sequence similarity, and inferring the most probable phylogenetic lineages of descent between isolates to infer the direction and route of pathogen transmission, from environmental, animal or human sources and reservoirs;
- the in silico prediction of phenotype and, in particular, acquired antimicrobial resistance mechanisms, pathogenicity and virulence determinants as well as correlates of epidemiological/ecological fitness associated with epidemic spread, also described as 'high-risk clones' [1-3,40].

In addition, recent cost assessments among EU laboratory networks have shown WGS analysis to offer reagent cost-competitive and labour-efficient alternatives to conventional epidemiological typing techniques for foodborne bacterial pathogens and *Neisseria meningitidis* [41]. WGS can produce results with relatively similar turn-around times to other typing techniques, depending on the laboratory throughput and processing workflows.

However, work on how to translate genomic data into meaningful information for public health decision-making is still incomplete. Current technical limitations of WGS-based typing include the complexity and reproducibility of the sequence data produced. Firstly, the sequencing platforms currently used differ in terms of the range of quality control and this may influence the accuracy and inter-laboratory comparability of sequence data [42]. Secondly, standard bioinformatics pipelines have not yet been defined and accepted for the consistent assembly of draft genomes and annotation of the WGS- derived genotypic markers based on a pathogen-specific nomenclature meeting disease-specific surveillance objectives. It is important to note that epidemiological objectives may differ greatly among diseases under surveillance, and therefore the relevant genomic information and the degree of genetic distance resolution required to meet the disease-specific objectives will also differ. For instance, higher levels of resolution and precision are needed for a surveillance system designed to detect and trace-back common source or community cross-infection outbreaks than for monitoring vaccine antigen expression among pathogens circulating in large populations. Another potential limitation of WGS typing for some diseases is the lack of backward compatibility with typing systems such as PFGE and MLVA.

Initiatives to translate WGS into public health practices

A number of organisations, collaborative projects and laboratory consortia are actively working to overcome technical hurdles, develop bioinformatics solutions and pilot studies of WGS-based typing for public health protection in the EU, USA and globally (Table 1).

Table 1. Organisations and initiatives supporting the development and application of WGS-based typing for public health protection

Name (link)	Type of organisation/project	Goal
ECDC (www.ecdc.europa.eu/)	EU health agency	Communicable disease surveillance and risk assessment
EFSA www.efsa.europa.eu/	EU health agency	Food safety monitoring and risk assessment
EMERGE (http://www.emerge.rki.eu/Emerge/EN/h)	European Commission body working in conjunction with EU Member States	Laboratory preparedness and response
PathoNGenTrace (www.patho-ngen-trace.eu/)	EU research project (2012-16)	Bioinformatics solutions for WGS-based diagnostics and surveillance of bacterial pathogens
COMPARE (www.compare-europe.eu/)	EU research project (2014-19)	Bioinformatics solutions and WGS-based data exchange platform for diagnostics, research, surveillance and risk assessment of bacterial and viral pathogens
CDC (www.cdc.gov/)	US health agency	Surveillance and risk assessment; global FWD network (PulseNet International)
GMI (www.globalmicrobialidentifier.org)	Global scientific collaboration initiative	Development of WGS data exchange and analysis tools for diagnostics, research, surveillance and public health response
ESGEM (https://www.escmid.org/research_projects/study_groups/epidemiological_markers/)	ESCMID study group	Collaborative research, evaluation, standardisation and recommendations on microbiological typing systems

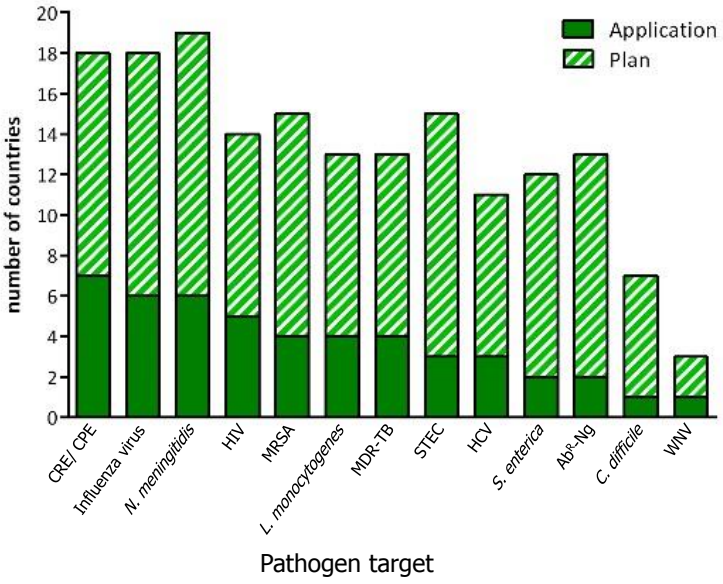
ECDC Roadmap for integration of molecular typing in surveillance

In 2013, ECDC worked with multidisciplinary experts, including National Focal Points for Surveillance and for Microbiology and the Advisory Forum, to develop a five-year strategic roadmap to integrate molecular/genomic typing data into EU surveillance for 12 priority diseases [43]. ECDC also raised awareness of the rapid development and validation of WGS-based typing among the competent bodies for surveillance and microbiology [44]. During the period 2014–15, this roadmap was refined by developing specific surveillance design strategies for each priority disease and underwent a thorough revision in consultation with advisory bodies [45]. Of note, gene sequence or WGS-based typing are recommended in the roadmap for a majority of diseases and multi-drug resistant pathogens. To start implementation, ECDC supported a number of pilot pan-EU studies within disease networks and/or projects for the development and evaluation of WGS typing. These are summarised in Appendix 1.

WGS-based typing in the EU/EEA Member States, July 2015

The survey results based on reports by the National Microbiology Focal Points (NMFP) from 28 countries, indicated a well-advanced implementation of WGS across Europe for surveillance and outbreak investigation. By July 2015, the technology was reportedly accessible to public health laboratories in 19 countries. WGS-based typing was used in 10 countries for national surveillance of one or more diseases, and in 18 countries for outbreak investigations. In addition, 18 countries are planning to implement WGS within the next three years. The use of WGS for surveillance was applied with different sampling schemes, mostly as a supplementary typing method in sentinel surveillance or survey-based samples. WGS is already being used for many diseases and drug-resistant pathogens included in the ECDC roadmap (Figure 1). It is interesting to note that the top three surveillance targets for national WGS-based surveillance are Carbapenem-resistant *Enterobacteriaceae* (CRE), influenza virus and *N. meningitidis*.

Figure 1. Number of EU/EEA countries applying or planning to apply WGS-based typing for national surveillance within three years, by disease target (categories not mutually exclusive), July 2015



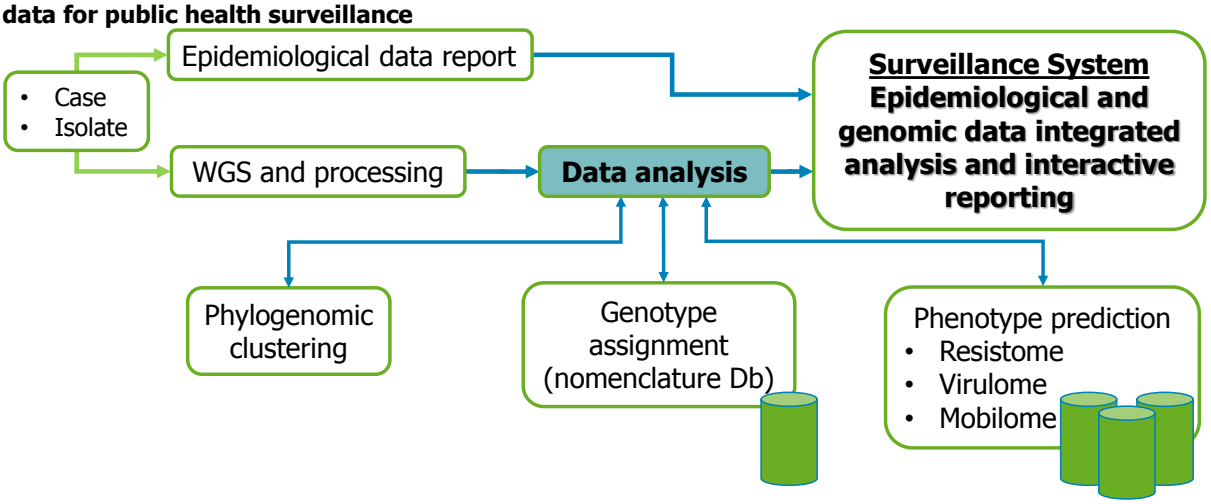
Note: these surveillance systems use diverse sampling frames ranging from survey-based to sentinel and comprehensive longitudinal sampling.

Source: NMFP 2015 survey, (N=28 respondents)

Technical requirements for multi-centre sharing of genomic information for public health use

The main steps involved in the process of WGS data production, bioinformatics processing and analysis, and integration with linked epidemiological data for generic surveillance purposes, are summarised in Figure 2.

Figure 2. Process overview – WGS data production, analysis and integration with epidemiological data for public health surveillance



Within this process, the following technical requirements have been identified as pre-requisites for the implementation of multi-centre genomic information sharing in the context of EU-wide public health surveillance [44,46]:

Quality assurance standards

Sequencing quality control metrics and minimum quality assurance standards must be agreed for each pathogen. They should be validated by means of ring trials to test the equivalence and reproducibility of data generation between laboratories using different sequencing and bioinformatics platforms. Wet/dry laboratory WGS external quality assessment (EQA) services should be provided to public health laboratories participating in surveillance network(s) [44,46].

WGS analysis strategy and genotype nomenclature

Harmonisation of bioinformatics analytical strategies for WGS results visualisation and plain language reporting for epidemiological inference and risk assessment must be agreed for each disease/pathogen with EU surveillance-specific epidemiological objectives. Harmonisation across public health and food safety sectors is required for zoonotic pathogens. There are essentially two main surveillance objectives for WGS-based comparative genome analysis, each of which require distinct data analysis and reporting outputs (Figure 2).

- Phylogenetic analysis with results visualisation as a tree or network graph, based on measurement of the evolutionary distance between genomes and their hypothesised order of descent from their most recent common ancestor. This analysis is used to infer transmission linkage between isolates from different patients or potential infection sources [2,3,44,46,47].
- Prediction of clinically and epidemiologically relevant microbial phenotypes in terms of antigenic profile, drug resistance and virulence, including identification of determinants encoded in the accessory genome and mobile genetic elements such as phages, plasmids and transposons (mobilome). For phenotype prediction, putative function can be assigned by searching gene homologies in public databases that provide experimental evidence of phenotype, such as *in vitro* drug susceptibility or *in vivo* expression of virulence [2,11,32,33,40,44,48].

For cluster detection, several analytical strategies have been used to assess genomic distances and apply cluster or phylogenetic analysis methods to group closely-related strains [46,49]. The three predominant methods currently used in the literature are:

- SNP-calling based phylogenetic reconstruction of assembled draft genomes by alignment to an annotated reference genome;
- Core genome identification for reference scheme construction followed by gene-by-gene assignment to allelic profile (cgMLST) based on draft de-novo assembled genomes;
- K-mer-based grouping of the closest genome matches by comparing across very short sequences.

For surveillance systems designed to detect common source outbreaks caused by closely related/identical strains, it is likely that the standard analysis will consist of two steps. First, using a gene-by-gene based nomenclature, which enables results to be compared across laboratories and a first clustering to be made. This is then followed by SNP analysis to further resolve the phylogenetic structure of identified clusters of isolates with common/closely related cgMLST types [41]. To develop a common nomenclature assignment, an openly accessible database is required for each pathogen species under surveillance. This database should be able to return allele identifiers when provided with a sequence. Minimum database functionality criteria have been defined for the application of WGS in foodborne disease surveillance at EU and global level [41].

WGS data storage and exchange resources

The definition and implementation of international WGS data storage and exchange formats is an important need which is being addressed by several international initiatives (Table 1). Public repositories such as the European Nucleotide Archive (ENA), are available for large-scale storage of genomic data, with the sequence database being mirrored by equivalent repositories across continents. In addition, the technical requirements for storage and sharing of the sequence-based typing datasets depend on the surveillance objectives and study design (continuous, real-time or periodic sentinel surveys, centralised or decentralised analysis). It remains unclear whether a single, global data storage and sharing system can be established in a sustainable manner. There is broad agreement that federated databases designed for food safety and public health monitoring should be interoperable by means of common ontologies and application programming interfaces [46].

Integration of WGS data with epidemiological and clinical data

The combination of genomic with epidemiological and clinical data is critical for meaningful epidemiological surveillance. Therefore, WGS records must be linked with the source patient or environment metadata record using a common source identifier, as part of the notification process to national and EU public health surveillance systems (Figure 2). The protection of sensitive epidemiological and clinical data is guaranteed by only making such data available to the competent authorities, in accordance with EU and national legislation. Sharing of contextual metadata should distinguish between non-confidential minimum data sets that should be deposited into the publicly available databases and sensitive, clinical and epidemiological data that are legally restricted to health authorities. There must

be formal governance of data sharing and use agreements for each disease surveillance scheme, in compliance with the legal framework, as developed for the influenza virus global sequence (GISAID) database.

Epidemiological concordance validation

Empirical interpretation criteria of WGS-based typing data must be established for robust epidemiological inference. Findings from bacterial disease outbreak studies suggest that cases linked to a common source or direct cross-infection event typically exhibit genomes that are either identical or exhibit less than 3–10 SNPs or allelic pairwise differences, depending on the extent of core genome analysed and the species under study. This is because the threshold variation is a function of the mutation and recombination rates of a given species. In contrast, much larger distances in hundreds or thousands of pairwise SNP/allele differences separate them from the genomes of epidemiologically unrelated cases of infection [8,12,16,18,26]. Prospective studies of disease transmission networks should be conducted to validate the epidemiological concordance and pathogen-specific interpretation criteria of WGS relatedness for transmission inference on the various time scales and at the population levels relevant to the surveillance objectives.

Resource investment and expertise diffusion

In order to fully implement WGS-based typing for integrated surveillance in the EU, significant investment will be needed to facilitate access of public health laboratories to technical resources. These resources include the NGS instruments and validated bioinformatics tools, laboratory informatics infrastructure and operational funding. In addition, harnessing the full flexibility and power of NGS for reference diagnostic testing will imply major organisational changes. Appropriate use of WGS-based diagnostics and typing depends on laboratory personnel, public health epidemiologists and risk managers developing new competence.

ECDC strategy

ECDC should act as catalyst to foster wider application of WGS as a frontline pathogen typing method for public health protection in the EU/EEA. ECDC should ensure that WGS is adopted in phases so as not to compromise national or EU-level surveillance, outbreak investigation or risk assessments during the transition from earlier technologies, which will inevitably occur at different speeds across the EU. This entails the following five strategies:

- **Mapping the WGS-based public health initiatives and engage partnerships**
  ECDC will monitor relevant WGS development and validation initiatives within the EU and internationally and engage with research and public health partners in areas of common interest (Table 1). This includes advising on the objectives and design of EU-funded research projects and co-development with its Member State partners of pilot studies to address EU public health surveillance objectives. ECDC will coordinate work with the European Food Safety Authority (EFSA) on joint nomenclature, databases and surveillance systems in a 'one-health' approach to food-borne pathogens. ECDC will also liaise with experts in the clinical microbiology field, and the industry producing genome sequencing technologies and bioinformatics tools to understand and adapt to the changing landscape of infectious disease diagnostics and surveillance.

- **Leading on the integrated analysis of the epidemiological and WGS data**
  ECDC will encourage the integrated analyses of microbiological data and epidemiological data within its public health mandate. This entails both outbreak detection and assessment and long-term surveillance objectives that inform vaccine policy, antimicrobial resistance containment and other public health actions.

  ECDC will work with the Member States to focus on formulating the right risk assessment questions and surveillance objectives, and defining data and analyses to provide reliable answers to the questions. This represents a collective effort on the part of public health institutes, reference laboratories and surveillance experts. ECDC will continue to work with the Member States to ensure that surveillance data collection complies with EU surveillance standards and that case-based linkage of the epidemiological and microbial information is systematically established at a local or national reporting level.

  In its role as coordinator of the disease surveillance networks, ECDC will support the comparative testing of the available WGS analytical platforms for data sharing of sequence information and related epidemiological data. In particular, it will test the inter-operability of open access WGS databases and TESSy/EPIS platform systems for integrated surveillance data collation, analysis and reporting underpinning EU public health goals.

- **Providing guidance on and validation of WGS-based methods for surveillance**
  ECDC will not focus on the WGS and bioinformatics technology development or duplicate existing data repositories, nomenclature databases or analytical platforms. Instead, it will facilitate and contribute to the development of international and cross-sector agreements on WGS quality standards, analytical schemes and genomic type nomenclature for the disease agent/resistance determinants being monitored, in collaboration with the scientific community, EU and international health agencies and national reference laboratories. ECDC will promote the development and maintenance of platform(s) for assigning genomic-derived nomenclature. In addition, to allow for flexibility of analytical strategies, it will stimulate exploratory data mining approaches to uncover unpredicted associations between the epidemiological information on risk factors and nodes of microbial genetic relatedness, following the 'systems biology' paradigm.

- **Supporting the Member States in performing the transition to appropriate use of WGS**
  ECDC will gather further input from its national partners in the public health institutes and reference laboratories who are delivering the public health microbiology services and have first-hand experience of the requirements. ECDC will take into account the diversity of national arrangements and stages of development of the WGS applications across the EU. ECDC will foster capacity for use of WGS technology by organising multidisciplinary workshops on analysis, reporting, interpretation and use of integrated genomic epidemiology data for public health microbiologists, epidemiologists and risk managers. The goal of this 'genomic epidemiology' professional development programme is to reach a common understanding for the use of integrated information on epidemiological and pathogen sequence characteristics to guide public health action.

  In addition, ECDC will organise ring trials to test the multi-centre comparability of WGS data production and analysis within EU disease laboratory networks, as part of pilot studies. If this proves to be effective, it will subsequently organise EQA exercises for routine WGS-based typing of the priority pathogens.

- **Develop, run and evaluate selected pilot implementation studies**
  ECDC will work with its network partners to contribute to the evaluation of operational performance, epidemiological validity and effectiveness of a few disease-specific public health applications of WGS-based typing. These will be selected as first priority surveillance targets, as agreed in consultation with the Member States in accordance with the revision of the EU roadmap on the basis of optimal public health added value, best technical feasibility and broadest Member State capacity [41,45,50].

  Subject to the adoption of the updated roadmap [45], ECDC will start piloting a feasibility study within the food- and waterborne network for *L. monocytogenes* WGS data exchange, in coordination with EFSA. It will develop the business cases for piloting WGS-based surveillance of invasive meningococcal disease, multidrug-resistant gonococci and Carbapenemase-producing *E. coli* and *K. pneumoniae* based on existing nomenclature.

  ECDC will ensure that the public health impact of these 'natural experiments' involving changing the surveillance system to genomic-based typing will be evaluated as an integral part of the pilot trials. Rigorous evaluation of the public health benefits of system implementation will be provided using longitudinal time series analysis, based on agreed surveillance system performance indicators and disease prevention metrics.

4. Conclusions and potential implications

ECDC vision: WGS as frontline typing method for public health protection

In five years' time, ECDC will have contributed to the establishment of standards and systems enabling the EU-wide use of WGS as the method of choice for typing microbial pathogens, replacing other methods. This will improve the accuracy and effectiveness of disease surveillance, outbreak investigation, and evaluation of prevention policies through enhanced assessment of disease and drug resistance transmission dynamics. Furthermore, it will improve pathogen detection, identification, antimicrobial resistance profiling and biological risk prediction. In the process, ECDC will add value at EU and national level by working with public health laboratories and national epidemiology experts to help foster the transition to and wider routine application of WGS-based typing.

Impact on ECDC surveillance operations and systems

The transition to WGS will enhance the added value of EU surveillance, particularly as anticipated new disease-specific surveillance objectives are realised. This, in turn, will have an impact on the overall surveillance organisation and methods required for ECDC's surveillance function. Case definitions may need to be revised. Some of the variables currently collected may become obsolete if more precise information is provided by WGS. Access to additional data sources may be needed, together with information on their characteristics. Different data flows should be envisaged, together with data processing, storage and analysis plans. To ensure the coordinated, harmonious and timely development of surveillance methods and tools and to keep up with the WGS developments, ECDC will:

- Organise joint annual meetings between NFP for Surveillance and for Microbiology where key WGS developments and their impact will be appraised and decisions made on how the EU surveillance system should be adapted (as of 2016).
- Develop and pilot test surveillance process and tool functionalities based on a range of assumptions concerning data source types, envisaged data flows, ECDC proposed strategies and surveillance objectives. This will be done within the framework of the surveillance systems re-engineering project.
- Liaise with the European Commission Directorate General for Health and Food Safety (DG SANTE) to ensure timely integration of new surveillance case definitions into legal texts.
- Liaise with other EU-level and global initiatives as appropriate to ensure coordination and cost-efficient delivery of the strategy.

ECDC strategy implementation

It will be essential for ECDC to effectively prioritise its actions and coordinate the implementation of the present strategy to roll out WGS as the main pathogen typing technique for EU surveillance. To this end, the WGS strategy outlined here is embedded in ECDC's revised strategy and priority roadmap for cross-disease integrated surveillance over the next few years [45]. The roadmap implementation will carefully take into account the operational resources available within Member States and ECDC.

References

1. Struelens MJ, Brisse S. From molecular to genomic epidemiology: transforming surveillance and control of infectious diseases. Euro Surveill. 2013;18(4):20386.
2. Koser CU, Ellington MJ, Cartwright EJ, Gillespie SH, Brown NM, Farrington M, et al. Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. PLoS Pathog. 2012;8(8):e1002824.
3. Sintchenko V, Holmes EC. The role of pathogen genomics in assessing disease transmission. BMJ. 2015;350:h1314.
4. Dallman TJ, Byrne L, Launders N, Glen K, Grant KA, Jenkins C. The utility and public health implications of PCR and whole genome sequencing for the detection and investigation of an outbreak of Shiga toxin-producing *Escherichia coli* serogroup O26:H11. Epidemiol Infect. 2015 Jun;143(8):1672-80.
5. Taylor AJ, Lappi V, Wolfgang WJ, Lapierre P, Palumbo MJ, Medus C, et al. Characterization of Foodborne Outbreaks of *Salmonella enterica* Serovar Enteritidis with Whole-Genome Sequencing Single Nucleotide Polymorphism-Based Analysis for Surveillance and Outbreak Detection. J Clin Microbiol. 2015 Oct;53(10):3334-40.
6. Angelo KM, Chu A, Anand M, Nguyen TA, Bottichio L, Wise M, et al. Outbreak of Salmonella Newport infections linked to cucumbers--United States, 2014. MMWR Morbidity and mortality weekly report. 2015 Feb 20;64(6):144-7.
7. Palmore TN, Henderson DK. Managing transmission of carbapenem-resistant *Enterobacteriaceae* in healthcare settings: a view from the trenches. Clin Infect Dis. 2013 Dec;57(11):1593-9.
8. Epson EE, Pisney LM, Wendt JM, MacCannell DR, Janelle SJ, Kitchel B, et al. Carbapenem-resistant *Klebsiella pneumoniae* producing New Delhi metallo-beta-lactamase at an acute care hospital, Colorado, 2012. Infect Control Hosp Epidemiol. 2014 Apr;35(4):390-7.
9. Espedido BA, Steen JA, Ziochos H, Grimmond SM, Cooper MA, Gosbell IB, et al. Whole genome sequence analysis of the first Australian OXA-48-producing outbreak-associated *Klebsiella pneumoniae* isolates: the resistome and in vivo evolution. PLoS One. 2013;8(3):e59920.
10. Eyre DW, Fawley WN, Best EL, Griffiths D, Stoesser NE, Crook DW, et al. Comparison of multilocus variable-number tandem-repeat analysis and whole-genome sequencing for investigation of *Clostridium difficile* transmission. J Clin Microbiol. 2013 Dec;51(12):4141-9.
11. Lavezzo E, Toppo S, Franchin E, Di Camillo B, Finotello F, Falda M, et al. Genomic comparative analysis and gene function prediction in infectious diseases: application to the investigation of a meningitis outbreak. BMC Infect Dis. 2013;13:554.
12. McAdam PR, Templeton KE, Edwards GF, Holden MT, Feil EJ, Aanensen DM, et al. Molecular tracing of the emergence, adaptation, and transmission of hospital-associated methicillin-resistant *Staphylococcus aureus*. Proc Natl Acad Sci U S A. 2012 Jun 5;109(23):9107-12.
13. Jolley KA, Hill DM, Bratcher HB, Harrison OB, Feavers IM, Parkhill J, et al. Resolution of a meningococcal disease outbreak from whole-genome sequence data with rapid Web-based analysis methods. J Clin Microbiol. 2012 Sep;50(9):3046-53.
14. Walker TM, Monk P, Smith EG, Peto TE. Contact investigations for outbreaks of *Mycobacterium tuberculosis*: advances through whole genome sequencing. Clin Microbiol Infect. 2013 Sep;19(9):796-802.
15. Rebaudet S, Mengel MA, Koivogui L, Moore S, Mutreja A, Kande Y, et al. Deciphering the origin of the 2012 cholera epidemic in Guinea by integrating epidemiological and molecular analyses. PLoS Negl Trop Dis. 2014 Jun;8(6):e2898.
16. Roetzer A, Diel R, Kohl TA, Ruckert C, Nubel U, Blom J, et al. Whole genome sequencing versus traditional genotyping for investigation of a *Mycobacterium tuberculosis* outbreak: a longitudinal molecular epidemiological study. PLoS Med. 2013;10(2):e1001387.
17. Mossong J, Decruyenaere F, Moris G, Ragimbeau C, Olinger CM, Johler S, et al. Investigation of a staphylococcal food poisoning outbreak combining case-control, traditional typing and whole genome sequencing methods, Luxembourg, June 2014. Euro Surveill. 2015 Nov 12;20(45).
18. Dallman TJ, Byrne L, Ashton PM, Cowley LA, Perry NT, Adak G, et al. Whole-Genome Sequencing for National Surveillance of Shiga Toxin-Producing *Escherichia coli* O:157. Clin Infect Dis. 2015 Aug 1;61(3):305-12.
19. Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS, Nielsen EM, et al. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. J Clin Microbiol. 2014 May;52(5):1501-10.
20. Knobloch JK, Niemann S, Kohl TA, Lindner U, Nitschke M, Sayk F, et al. Whole-genome sequencing for risk assessment of long-term Shiga toxin-producing *Escherichia coli*. Emerg Infect Dis. 2014 Apr;20(4):732-3.
21. Harrison OB, Brueggemann AB, Caugant DA, van der Ende A, Frosch M, Gray S, et al. Molecular typing methods for outbreak detection and surveillance of invasive disease caused by *Neisseria meningitidis, Haemophilus influenzae* and *Streptococcus pneumoniae*, a review. Microbiology. 2011 Aug;157(Pt 8):2181-95.

22. Demczuk W, Lynch T, Martin I, Van Domselaar G, Graham M, Bharat A, et al. Whole-genome phylogenomic heterogeneity of *Neisseria gonorrhoeae* isolates with decreased cephalosporin susceptibility collected in Canada between 1989 and 2013. J Clin Microbiol. 2015 Jan;53(1):191-200.
23. den Bakker HC, Allard MW, Bopp D, Brown EW, Fontana J, Iqbal Z, et al. Rapid whole-genome sequencing for surveillance of *Salmonella enterica* serovar Enteritidis. Emerg Infect Dis. 2014 Aug;20(8):1306-14.
24. Leekitcharoenphon P, Nielsen EM, Kaas RS, Lund O, Aarestrup FM. Evaluation of whole genome sequencing for outbreak detection of *Salmonella enterica*. PLoS One. 2014;9(2):e87991.
25. Graham M, Liang B, Van Domselaar G, Bastien N, Beaudoin C, Tyler S, et al. Nationwide molecular surveillance of pandemic H1N1 influenza A virus genomes: Canada, 2009. PLoS One. 2011;6(1):e16087.
26. Kohl TA, Diel R, Harmsen D, Rothganger J, Walter KM, Merker M, et al. Whole-genome-based *Mycobacterium tuberculosis* surveillance: a standardized, portable, and expandable approach. J Clin Microbiol. 2014 Jul;52(7):2479-86.
27. Doro R, Mihalov-Kovacs E, Marton S, Laszlo B, Deak J, Jakab F, et al. Large-scale whole genome sequencing identifies country-wide spread of an emerging G9P[8] rotavirus strain in Hungary, 2012. Infect Genet Evol. 2014 Dec;28:495-512.
28. Cotten M, Watson SJ, Zumla AI, Makhdoom HQ, Palser AL, Ong SH, et al. Spread, circulation, and evolution of the Middle East respiratory syndrome coronavirus. MBio. 2014;5(1).
29. Ravagnan S, Montarsi F, Cazzin S, Porcellato E, Russo F, Palei M, et al. First report outside Eastern Europe of West Nile virus lineage 2 related to the Volgograd 2007 strain, northeastern Italy, 2014. Parasit Vectors. 2015;8:418.
30. Eyre DW, Tracey L, Elliott B, Slimings C, Huntington PG, Stuart RL, et al. Emergence and spread of predominantly community-onset *Clostridium difficile* PCR ribotype 244 infection in Australia, 2010 to 2012. Euro Surveill. 2015;20(10):21059.
31. Lim SK, Stuart RL, Mackin KE, Carter GP, Kotsanas D, Francis MJ, et al. Emergence of a ribotype 244 strain of *Clostridium difficile* associated with severe disease and related to the epidemic ribotype 027 strain. Clin Infect Dis. 2014 Jun;58(12):1723-30.
32. Koser CU, Ellington MJ, Peacock SJ. Whole-genome sequencing to control antimicrobial resistance. Trends Genet. 2014 Sep;30(9):401-7.
33. Zankari E, Hasman H, Kaas RS, Seyfarth AM, Agerso Y, Lund O, et al. Genotyping using whole-genome sequencing is a realistic alternative to surveillance based on phenotypic antimicrobial susceptibility testing. J Antimicrob Chemother. 2013 Apr;68(4):771-7.
34. Pecora ND, Li N, Allard M, Li C, Albano E, Delaney M, et al. Genomically Informed Surveillance for Carbapenem-Resistant *Enterobacteriaceae* in a Health Care System. MBio. 2015;6(4):e01030.
35. Price J, Gordon NC, Crook D, Llewelyn M, Paul J. The usefulness of whole genome sequencing in the management of *Staphylococcus aureus* infections. Clin Microbiol Infect. 2013 Sep;19(9):784-9.
36. Bartels MD, Larner-Svensson H, Meiniche H, Kristoffersen K, Schonning K, Nielsen JB, et al. Monitoring meticillin resistant *Staphylococcus aureus* and its spread in Copenhagen, Denmark, 2013, through routine whole genome sequencing. Euro Surveill. 2015;20(17).
37. Eyre DW, Golubchik T, Gordon NC, Bowden R, Piazza P, Batty EM, et al. A pilot study of rapid benchtop sequencing of *Staphylococcus aureus* and *Clostridium difficile* for outbreak detection and surveillance. BMJ Open. 2012;2(3).
38. Olaitan AO, Diene SM, Kempf M, Berrazeg M, Bakour S, Gupta SK, et al. Worldwide emergence of colistin resistance in *Klebsiella pneumoniae* from healthy humans and patients in Lao PDR, Thailand, Israel, Nigeria and France owing to inactivation of the PhoP/PhoQ regulator *mgrB*: an epidemiological and molecular study. Int J Antimicrob Agents. 2014 Dec;44(6):500-7.
39. Merker M, Blin C, Mona S, Duforet-Frebourg N, Lecher S, Willery E, et al. Evolutionary history and global spread of the *Mycobacterium tuberculosis* Beijing lineage. Nat Genet. 2015 Mar;47(3):242-9.
40. Hill DM, Lucidarme J, Gray SJ, Newbold LS, Ure R, Brehony C, et al. Genomic epidemiology of age-associated meningococcal lineages in national surveillance: an observational cohort study. Lancet Infect Dis. 2015 Dec;15(12):1420-8.
41. European Centre for Disease Prevention and Control. Expert Opinion on the introduction of next-generation typing methods for food- and waterborne diseases in the EU and EEA. Stockholm: ECDC; 2015.
42. European Food Safety Authority EPoBH. Scientific Opinion on the evaluation of molecular typing methods for major food-borne microbiological hazards and their use for attribution modelling, outbreak investigation and scanning surveillance: Part 1 (evaluation of methods and applications). EFSA Journal. 2013;11(12):84.
43. European Centre for Disease Prevention and Control. Roadmap for integration of molecular typing into European-level surveillance and epidemic preparedness – Version 1.2, 2013. Stockholm: ECDC, 2016.
44. European Centre for Disease Prevention and Control. ECDC technical consultation on harnessing genomics for epidemiological surveillance - Meeting report, Paris, 1-2 October 2013. Stockholm: ECDC, 2014.
45. European Centre for Disease Prevention and Control. ECDC roadmap for integration of molecular and genomic typing into European-level surveillance and epidemic preparedness – Version 2.1, 2016-19. Stockholm: ECDC, 2016.

46. European Food Safety Authority. Whole Genome Sequencing of food-borne pathogens for public health protection. EFSA Scientific Colloquium 20. Parma, Italy: EFSA Supporting Publications; 2014. p. 63.
47. European Food Safety Authority EPoBH. Scientific Opinion on the evaluation of molecular typing methods for major food-borne microbiological hazards and their use for attribution modelling, outbreak investigation and scanning surveillance: Part 2 (surveillance and data management activities). EFSA Journal. 2014;12(7):46.
48. Palm D, Johansson K, Ozin A, Friedrich A, Grundmann H, Larsson J, et al. Molecular epidemiology of human pathogens: how to translate breakthroughs into public health practice, Stockholm, November 2011. Euro Surveill. 2012;17(2).
49. Maiden MC, Jansen van Rensburg MJ, Bray JE, Earle SG, Ford SA, Jolley KA, et al. MLST revisited: the gene-by-gene approach to bacterial genomics. Nat Rev Microbiol. 2013 Oct;11(10):728-36.
50. van Belkum A, Tassios PT, Dijkshoorn L, Haeggman S, Cookson B, Fry NK, et al. Guidelines for the validation and application of typing methods for use in bacterial epidemiology. Clin Microbiol Infect. 2007 Oct;13 Suppl 3:1-46.
51. Hasnain SE, O'Toole RF, Grover S, Ehtesham NZ. Whole genome sequencing: a new paradigm in the surveillance and control of human tuberculosis. Tuberculosis (Edinb). 2015 Mar;95(2):91-4.
52. European Centre for Disease Prevention and Control. Carbapenemase-producing bacteria in Europe: interim results from the European Survey on carbapenemase-producing *Enterobacteriaceae* (EuSCAPE) project. Stockholm, Sweden: ECDC, 2013.
53. Bialek-Davenet S, Criscuolo A, Ailloud F, Passet V, Jones L, Delannoy-Vieillard AS, et al. Genomic definition of hypervirulent and multidrug-resistant *Klebsiella pneumoniae* clonal groups. Emerg Infect Dis. 2014 Nov;20(11):1812-20.

Appendix 1. Disease-specific ECDC pilot projects and expert opinions on WGS, as of November 2015

- **Foodborne bacterial diseases:** the WGS technique has been recommended for future foodborne disease surveillance in a recent EFSA scientific colloquium and by EFSA BIOHAZ panel opinion on molecular typing methods [42,46]. ECDC has established an expert group on the introduction of next generation typing methods for the surveillance of FWD diseases composed of microbiologists, epidemiologists and bioinformaticians. The expert group has produced an opinion paper, which aims to support countries that are planning or implementing WGS for surveillance and outbreak investigation [41]. It describes a representative process for routine WGS analysis and outlines minimum requirements for a nomenclature database for WGS of foodborne pathogens. ECDC has started a study with the aim of performing WGS on human isolates of *Listeria monocytogenes* for the period 2010–14. The results will be compared with a complementary project run by EFSA, predominantly for food isolates, so that links between human cases and food can be established.
- **Tuberculosis:** The European Network of Reference Laboratories for Tuberculosis (ERLTB-Net) is assessing WGS technologies for their applicability and usefulness in EU/EEA TB prevention and control. A systematic review on the added value of WGS of *Mycobacterium tuberculosis* for detection of recent transmission and tracing outbreaks concluded that WGS could be a powerful tool in TB epidemiology, due to it having greater discriminatory powers than existing genotyping methods [26,51]. WGS data seem often to be a better match with contact tracing data and geographical distribution of isolates. ERLTB-Net has concluded that no evidence currently exists on the applicability or cost-effectiveness of WGS for resolving TB outbreaks or national/international surveillance.
- **Meningococcal disease:** The Invasive Bacterial Disease Laboratory Network (IBD-LabNet) has performed a pilot WGS-based EU/EEA population survey of invasive disease caused by *Neisseria meningitidis* as an extension of the European meningococcal strain collection. This survey assesses the comparative value and feasibility of WGS versus the current typing scheme. WGS technology has become competitive with PCR-based MLST and antigen gene sequencing in terms of cost, while enhancing resolution for phylogenetic analysis and providing additional information for vaccine antigen typing. WGS-based typing standardisation using the core genome MLST is being developed and nomenclature curated at http://PubMLST.org/Neisseria [49]. ECDC is developing a business case to link these data to TESSy in integrated surveillance.
- **Antibiotic-resistant gonococcal infection:** The ECDC-coordinated European Gonococcal Antimicrobial Susceptibility Programme (Euro-GASP) collects data on antimicrobial susceptibility of gonococcal isolates and associated epidemiological data in the EU/EEA. Characterisation of molecular types of *N. gonorrhoeae* strains circulating in the EU/EEA region in 2009 and 2010 using NG-MAST typing was analysed in the context of epidemiological information and therapeutically-relevant resistance profiles. A second survey of *N. gonorrhoeae* strains in the EU/EEA region since 2013 is ongoing. This study will assess the relevance of WGS in the surveillance of *N. gonorrhoeae* at EU/EEA level. The ECDC Advisory Forum approved a strategy for repeat structured surveys of *N. gonorrhoeae* circulation in the EU. Based on the results, ECDC will produce a business case defining the methodology to be applied in future surveys.
- **Carbapenemase-producing *Enterobacteriaceae* (CPE) infection:** ECDC funded the 'European Survey of Carbapenemase-producing *Enterobacteriaceae*' (EuSCAPE) that collected close to 1400 isolates of CP- *Klebsiella pneumoniae* and *Escherichia coli* together with clinical data from 455 sentinel hospitals in 36 European countries during 2013–14 [52]. CPE isolates were phenotypically characterised for antimicrobial susceptibility and tested for Carbapenemase determinants by PCR. In 2014, together with a group of microbiology, genomics and public health experts, ECDC developed a strategy for an EU-wide surveillance of high-risk CPE clones and mobile genetic vectors to inform risk assessment and control programmes by using WGS-based typing of CPE. This strategy, approved by the ECDC Advisory Forum, follows the model of repeated surveys based on the EuSCAPE project design. A nomenclature for *K. pneumoniae* cgMLST has been described and is publicly available [53]. The EuSCAPE participants have agreed to pilot WGS characterisation of the survey isolates in a centralised manner. Based on the results, ECDC will produce a business case defining the methodology to be applied in future surveys.

European Centre for Disease Prevention and Control (ECDC)

Postal address:
Granits väg 8, SE-171 65 Solna, Sweden

Visiting address:
Tomtebodavägen 11A, SE-171 65 Solna, Sweden

Tel. +46 858601000
Fax +46 858601001
www.ecdc.europa.eu

An agency of the European Union
www.europa.eu

Subscribe to our publications
www.ecdc.europa.eu/en/publications

Contact us
publications@ecdc.europa.eu

Follow us on Twitter
@ECDC_EU

Like our Facebook page
www.facebook.com/ECDC.EU

ECDC is committed to ensuring the transparency and independence of its work

In accordance with the Staff Regulations for Officials and Conditions of Employment of Other Servants of the European Union and the ECDC Independence Policy, ECDC staff members shall not, in the performance of their duties, deal with a matter in which, directly or indirectly, they have any personal interest such as to impair their independence. Declarations of interest must be received from any prospective contractor(s) before any contract can be awarded.
www.ecdc.europa.eu/en/aboutus/transparency

HOW TO OBTAIN EU PUBLICATIONS

Free publications:

- one copy:
  via EU Bookshop (http://bookshop.europa.eu);
- more than one copy or posters/maps:
  from the European Union's representations (http://ec.europa.eu/represent_en.htm);
  from the delegations in non-EU countries (http://eeas.europa.eu/delegations/index_en.htm);
  by contacting the Europe Direct service (http://europa.eu/europedirect/index_en.htm) or
  calling 00 800 6 7 8 9 10 11 (freephone number from anywhere in the EU) (*).

  (*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

Priced publications:

- via EU Bookshop (http://bookshop.europa.eu).

Publications Office