

Molecular epidemiology of human pathogens: how to translate breakthroughs into public health practice, Stockholm, November 2011

D Palm¹, K Johansson¹, A Ozin¹, A W Friedrich², H Grundmann², J T Larsson³, M J Struelens (marc.struelens@ecdc.europa.eu)¹

1. European Centre for Disease Prevention and Control (ECDC), Stockholm, Sweden
2. University Medical Centre Groningen, University of Groningen, Groningen, the Netherlands
3. Statens Serum Institut, Copenhagen, Denmark

Citation style for this article:

Palm D, Johansson K, Ozin A, Friedrich AW, Grundmann H, Larsson JT, Struelens MJ. Molecular epidemiology of human pathogens: how to translate breakthroughs into public health practice, Stockholm, November 2011. Euro Surveill. 2012;17(2):pii=20054. Available online: <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20054>

Article published on 12 January 2012

This report outlines the main conclusions from an expert consultation on molecular epidemiology held on 22–23 November 2011, hosted and organised by the European Centre for Disease Prevention and Control (ECDC) [1]. The consultation brought together researchers, microbiologists and public health experts to discuss how public health can benefit from the recent scientific and technological advances in molecular microbiology, with special focus on the rapidly evolving next-generation sequencing technology.

Interplay of basic and public health microbiological investigations: case study of the *Escherichia coli* O104:H4 outbreak in Germany, 2011

Angelika Fruth (Robert Koch Institute, Germany), Flemming Scheutz (Statens Serum Institut, Denmark) and Martina Bielaszewska (University of Münster, Germany) described how laboratories at different levels contributed to the identification and detailed characterisation of the Shiga toxin-producing *E. coli* (STEC) O104:H4 strain during the 2011 outbreak in Germany [2–4]. The processes applied in the public health microbiology investigations underlined various challenges: obtaining relevant clinical isolates linked to outbreak cases at reference-laboratory level, coping with a high volume of tests and providing rapid feedback on antigenic and virulence profile information of the epidemic strain during the escalation phase of the outbreak.

For the first time in such a large epidemic of food-borne bacterial disease, the whole genome sequence of the pathogen was determined within days of recognition of the outbreak. The relative value of traditional microbiological methods in relation to whole genome sequencing (WGS) was discussed. The general opinion was that WGS approaches are rapidly evolving and will be a powerful tool in future public health emergencies, especially when the infectious agent is less well known. However, in this particular outbreak, traditional

methods were effective for characterising the outbreak strain for epidemiological investigations. It was also considered that other factors related to laboratory investigations need to be strengthened. These include the requirement for stool cultures for pathogen isolation, access to strain collections and typing databases with laboratory information combined with epidemiological data, better communication channels between laboratories and epidemiological institutes and standardisation of traditional characterisation methods between human and food laboratories.

Scientific and technological advances in molecular microbiology: what's in it for public health?

Marc Achtman (Cork University, Ireland) challenged the existing definitions of a bacterial species and the ability of available laboratory methods to investigate population structures of different bacterial species. He argued for core genomic and gene polymorphism analyses, as opposed to serotyping and DNA fingerprinting methods, for the application of phylogenetic analyses to support epidemiological investigations [5]. No single approach can yet handle the wide range of population structures within the group of bacteria causing human disease, due to considerable heterogeneity in genomic diversity and clonality versus recombination between and within species. There are a number of promising developments especially regarding sequencing, but for applications in regional public health laboratories, scientific studies and technical solutions are still needed.

The use of sequencing information to investigate disease dynamics and phylogeographic mapping has a stronger tradition in virology. Anne-Mieke Vandamme (Rega Institute for Medical Research, Belgium) demonstrated how integration of sequencing, epidemiological and spatio-temporal data has been used to understand specific viral epidemics, such as the global migration events and selected national transmission

characteristics of HIV [6,7] and the rapid transmission of influenza A(H1N1) virus during the 2009 pandemic [8].

On behalf of Jacques Schrenzel (Geneva University Hospitals, Switzerland), Marc Struelens (ECDC) presented an overview of state-of-the-art instruments for genomic and proteomic analysis and discussed their possible future application in microbiology laboratories. The main message was that novel mass spectrometry technologies are being introduced into clinical microbiological laboratories for pathogen identification. An open question is whether these technologies will achieve the simplicity combined with higher resolution needed to provide subspecies strain profiling as well as primary diagnostic information in real time. Extrapolating current trends suggests that next-generation proteomic technologies will become routine in clinical laboratories and advanced sequencing applications in reference and research laboratory settings.

Typing as a component of epidemic preparedness and response support

A working group discussed experiences and identified challenges when adopting a validated typing method to operate a quality-assured system for collection and analysis of integrated epidemiological and microbiological data at the European level. Experiences of the European Legionnaires' Disease Surveillance Network (ELDSNET) in setting up the successful international system for sequence-based genotyping of *Legionella* were shared.

A network of engaged laboratories spanning all participating countries was considered key for utilising existing structures and expertise, defining disease-specific objectives, sampling frames and typing techniques, as well as collecting representative isolates.

Human pathogens encompass diverse organisms, but typing techniques should aim at maximum across-species flexibility. Important health events related to these organisms range in scale (endemic, epidemic and emerging) and therefore require different sentinel thresholds. Independent of disease-specific differences, key steps for developing effective typing networks were proposed. First, every pathogen-specific network has to agree on a minimum set of metadata (place, person, time, etc.). Second, a typing approach should be agreed upon, which would preferably be highly versatile, and generate data with suitable portability, reproducibility, timeliness, traceability and biological robustness. Third, a set of rules stipulating a suitable sampling frame should be agreed upon. Fourth, a web-based typing platform should be set up. The manner of submission, access and ownership of contributors must be similar to those of other repositories, e.g. the National Center for Biotechnology Information (NCBI).

The need for regularly conducted structured molecular typing surveys to calibrate the baseline population structure and molecular diversity of the pathogen of interest was identified. Such surveys can also explore the role of potential sources and reservoirs of human pathogens using flexible sampling frames.

A clear threat to sustainability is the unpredictability of funding in the field of European public health. It is therefore important to safeguard against loss of data and isolates by providing the necessary redundancy in terms of biobanking and mirrored databases. Effective collection of enhanced surveillance data is also dependent on centralised capacity-building initiatives, such as regular proficiency testing, training and certification.

Infections pose a potential threat to all European Union (EU) citizens and can only be tackled by joint interventions. Notwithstanding ECDC's clear mandate, EU Member States should provide the necessary support at national level, i.e. maintaining their countries' diagnostic and reference typing services. The need for advanced training and capacity building was also recognised, especially in the field of bioinformatics, and for developing cross-disciplinary competences of public health professionals, as provided by programmes such as the European Programme for Public Health Microbiology (EUPHEM).

Translational research for informing public health action

A second working group appraised if reliable predictions of communicable disease pattern and impact can be made by combining genomic and biological information with epidemiological and clinical data. Work at the Dutch National Institute for Public Health and the Environment (RIVM) was presented, describing efforts to identify risk factors and geographical origin for hepatitis A infection in the Netherlands via continuing collection of partial genome sequence information by a national network of clinical virology laboratories.

Although a large amount of data on pathogenicity has been published in the last 20 years, it was considered that a molecular microbiological risk assessment to foresee disease impact based on pathogen genotyping is not possible at this stage of scientific knowledge. However, data from major epidemiological studies of pathogens such as HIV, influenza virus, enterohaemorrhagic *E. coli* and *Staphylococcus aureus* indicate that genotypic markers can help predict the probability of disease progression or response to therapy. Several participants considered that genome sequence-based disease predictors were better characterised in virology than in bacteriology. It was agreed that much scientific work remains and EU-wide translational research studies should be initiated to increase the accuracy of these predictions. Access to comparative whole genome studies of large samples of bacterial pathogens linked to clinical and epidemiological data

would facilitate the detection of relevant epidemiological–biological associations.

Free sharing of genomic data was advocated, but existing difficulties were also recognised, for example, ownership, confidentiality, national policies, etc. To stimulate scientific progress, future databases hosting typing and epidemiological information combined with innovative bioinformatic solutions should have adjustable access levels, depending on the users and purposes of using the information. Quality assurance and agreed nomenclature for sequence-based typing need to be addressed in parallel to capacity-building initiatives, also in bioinformatics.

Next generation technologies, such as WGS, represent a window of opportunity for EU-wide typing for surveillance in the coming years. Intersectoral collaborations in research networks, combining human and veterinary bacteriology, virology and epidemiology with genomic bioinformatics, molecular biology and population genetics should be facilitated. Guidance and facilitation by ECDC to achieve the goals of common nomenclature as well as intersectoral collaboration and capacity building would be helpful.

Optimising use of novel typing technology for public health

A third working group considered whether novel proteomic and genomic technologies will replace existing methodologies in public health microbiological laboratories within five years. Data were presented showing that WGS applications are already used in the majority of large outbreaks in Europe and should gradually replace currently used microbiological characterisation methodologies or improve them by identification of novel targets.

The discussion was focused on genomic methods: gaps and challenges in routine application of WGS in epidemiological typing were examined. There was no consensus about whether WGS will replace current typing techniques. Arguments against the use of WGS were that currently too much work is required for a small amount of additional useful data, the relative expense and that deriving phenotype from genotype remains an imperfect science.

A number of key hurdles for implementation of WGS for public health purposes were identified. Without either well-utilised local instrumentation or contact with larger sequencing institutes, many public health institutes are excluded from the possibility of using WGS routinely. The time from sampling to results is generally too long to use WGS for daily surveillance. A major unresolved question was how genome data should be analysed for epidemiological characterisation. Existing analysis platforms are still wanting in terms of user friendliness and the inability to produce complete and closed genomes risks fragmenting WGS techniques and analysis methods. Harmonisation of

quality parameters for WGS data production is needed. Routine removal of genetic regions with high variation gave rise to concerns that too much information is left out of analyses – a concern for the analysis of mobile genetic elements that are of particular interest for public health.

To reduce WGS data to epidemiologically useful information is far from trivial. As it was found desirable to work with WGS data in a way similar to a conventional nomenclature of genotypes, several possibilities were examined, including an extended multi-locus sequence typing (MLST)-type classification, single nucleotide polymorphism-based classification or genome grouping into broader ‘natural’ lineages. These possible approaches to derive a ‘type’ are all hampered by considerable feasibility problems and unresolved ambiguities. The issue was thus left open, with the consideration that it was perhaps time to abandon the concept of ‘type’.

Some other unmet needs for public health application of WGS involved the central storage and curation of data. ECDC along with other experts could be involved in setting the design and accreditation standards for genome-based surveillance databases. ECDC could take part in creating a system of unique identifiers and link these to the European Surveillance System (TESSy). In addition, ECDC could support disease-specific studies on the effectiveness of WGS analysis for public health applications such as epidemic investigations.

Updates of ECDC and EU projects related to molecular surveillance and use of new laboratory methods

Harry Vennema (RIVM, the Netherlands) presented the European Commission-funded Global exchange of viral sequences to underpin response to health threats (GESTURE) project [9]. The principles agreed in this project are planned to be discussed for possible implementation in the World Health Organization’s upcoming World Health Assembly. Dag Harmsen (University of Münster, Germany) gave examples of how WGS methodologies have been applied in three outbreak investigations: STEC in Germany (2011) [10], *Klebsiella* OXA-48 in the Netherlands (2011) and *Neisseria meningitidis* in Germany (2001–2006). In two of these, WGS was performed in real time in parallel to traditional investigations, illustrating how sequencing technology can become a method of choice in these situations. Finally, Ivo Van Walle (ECDC) presented the ECDC plan to set up a generic platform for handling microbial typing data. The system, an extension of TESSy, will be piloted for a limited number of pathogens and typing methods and is planned to be released in spring 2012.

Conclusions

The conference clearly illustrated that recent advances in molecular microbial characterisation, primarily by WGS analysis, are opening up tremendous new scientific opportunities for a better understanding of

the pathogenicity, evolution and spread of human pathogens and the epidemiology of the diseases they cause. This technology has considerable potential for improving the resolution and predictive value of microbial typing as applied to public health objectives such as disease surveillance and epidemic investigation. However, many operational hurdles need to be addressed in the coming years, to define the methods to produce, analyse, share genomic information combined with epidemiological data in a meaningful way and build the capacity and fit-for-purpose data management platforms for public health microbiology. ECDC will follow up on the recommendations of the consultation to play the role of a facilitator, bridging the scientific and public health communities to jointly address the generic and disease-specific challenges identified.

References

1. European Centre for Disease Prevention and Control (ECDC). Expert consultation: Breakthroughs in molecular epidemiology of human pathogens - how to translate into public health practice. Stockholm: ECDC. [Accessed 12 Jan 2012]. Available from: http://www.ecdc.europa.eu/en/press/events/Lists/Events/ECDC_Dispatch.aspx?List=43564830%2D6b8a%2D442f%2D84e7%2D2495fa49489b&ID=184&Source=http%3A%2F%2Fstaging%2Eecdc%2Eeuropa%2Eeu%2Fen%2Fpress%2Fevents%2Flists%2Fevents%2Fitems%2Easpx%3FSortField%3DCreated%26SortDir%3DDesc%26View%3D%257bB9767274%252d1B4F%252d4663%252dACD3%252doF569564D9A8%257d
2. Scheutz F, Nielsen EM, Frimodt-Møller J, Boisen N, Morabito S, Tozzoli R, et al. Characteristics of the enteroaggregative Shiga toxin/verotoxin-producing *Escherichia coli* O104:H4 strain causing the outbreak of haemolytic uraemic syndrome in Germany, May to June 2011. *Euro Surveill.* 2011;16(24):pii=19889. Available from: <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19889>
3. Bielaszewska M, Mellmann A, Zhang W, Köck R, Fruth A, Bauwens A, et al. Characterisation of the *Escherichia coli* strain associated with an outbreak of haemolytic uraemic syndrome in Germany, 2011: a microbiological study. *Lancet Infect Dis.* 2011;11(9):671-6.
4. Rasko DA, Webster DR, Sahl JW, Bashir A, Boisen N, Scheutz F, et al. Origins of the *E. coli* strain causing an outbreak of hemolytic-uremic syndrome in Germany. *N Engl J Med.* 2011;365(8):709-17.
5. Didelot X, Bowden R, Street T, Golubchik T, Spencer C, McVean G, et al. Recombination and population structure in *Salmonella enterica*. *PLoS Genet.* 2011;7(7):e1002191.
6. Paraskevis D, Pybus O, Magiorkinis G, Hatzakis A, Wensing AM, van de Vijver DA, et al. Tracing the HIV-1 subtype B mobility in Europe: a phylogeographic approach. *Retrovirology.* 2009;6:49.
7. Faria NR, Suchard MA, Abecasis A, Sousa JD, Ndembi N, Bonfim I, et al. Phylodynamics of the HIV-1 CRF02_AG clade in Cameroon. *Infect Genet Evol.* 2011 May 4. [Epub ahead of print].
8. Lemey P, Suchard M, Rambaut A. Reconstructing the initial global spread of a human influenza pandemic: a Bayesian spatial-temporal model for the global spread of H1N1pdm. *PLoS Curr.* 2009; 1:RRN1031.
9. Executive Agency for Health and Consumers (EAHC). Global exchange of viral sequences to underpin response to health threats (GESTURE). Project No 20084153. EAHC: Luxembourg. [Accessed 12 Jan 2012]. Available from: <http://ec.europa.eu/eahc/projects/database.html?prjno=20084153>
10. Mellmann A, Harmsen D, Cummings CA, Zentz EB, Leopold SR, Rico A, et al. Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. *PLoS One.* 2011;6(7):e22751.