

EFSA Scientific Colloquium n°20

Use of Whole Genome Sequencing (WGS) of food-borne pathogens for public health protection

Parma, Italy, 16-17 June 2014

BRIEFING NOTES FOR DISCUSSION GROUPS

These briefing notes are prepared to provide participants with the relevant background information so as to be prepared for an interactive exchange of views and expertise, during the Colloquium.

Background

Molecular typing methods for food-borne pathogens start to be routinely applied in the European Union (EU) and abroad for public health protection (e.g. investigating food-borne outbreaks, identifying strains of food-borne bacteria with high virulence potential or resistance to antimicrobials) (ECDC, 2007; EFSA, 2009; ECDC, 2013). This follows from continuous advances in the understanding of the molecular characteristics of bacteria and their genetics linked to technological developments, which ultimately have led to the use of available bacterial whole genome sequencing (WGS) methods for food safety applications (EFSA, 2013).

Despite the technological advances, which result in the continuous optimisation of WGS platforms, there are opportunities for improving multiple aspects related to the use of the generated bacterial genome sequencing data. These opportunities which at the same time are true challenges include, among others: (i) the curation and analysis of WGS data and bioinformatics solutions; (ii) the analysis and interpretation of WGS data (e.g. analysis of SNPs, gene by gene comparisons) in the ascertaining of diversity, similarity and relatedness of foodborne bacterial pathogens and prediction of relevant properties (e.g. virulence, AMR) and (iii) the coordination of efforts between the food, veterinary and human health sectors in order to obtain maximum benefits from the use of WGS for food safety and public health protection.

There is currently limited experience in the use of WGS methods in microbial food safety in the EU. Most of that experience comes as a result of retrospective studies that have followed from outbreak investigations (Mellmann et al., 2011; Cao et al., 2013; Mather et al., 2013; Underwood et al., 2013; Zhang et al., 2013). Still, the potential application of WGS to predict pathogenicity of the strain under investigation may provide risk assessors with a powerful tool. However, full integration of routine WGS of food-borne pathogens in food safety will only be possible after

successful collaboration and coordination among scientists paving common pathways to overcome key challenges such as those previously addressed.

Following recent work completed (EFSA, 2013) and still ongoing (mandates from the European Commission to EFSA and ECDC on the building of a molecular typing database and regularly providing scientific analyses in collaboration with each other and EURLs (EFSA-Q-2013-00250, ECDC-CI No 13-292, 22-01-2013); mandate of EFSA to the BIOHAZ Panel 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-Q-2013-00906)), the momentum seems appropriate for EFSA taking the initiative and thus organise a Scientific Colloquium on the use of WGS of food-borne pathogens for public health protection.

This Colloquium will assess the latest scientific information, strengthen alliances with the relevant EU and international bodies to initiate discussions on the use of WGS methods for food safety applications and drive EFSA's ongoing efforts in the collection of molecular typing data by proactively anticipating the specific requirements of WGS data

The objectives of this Colloquium are to:

- Discuss in an open scientific debate the current issues, benefits and future challenges of WGS of foodborne pathogens in public health protection in comparison with current methods.
- Discuss the analysis and the interpretation of WGS data (e.g. analysis of SNPs, gene by gene comparisons) in the ascertaining of diversity, similarity and relatedness of foodborne bacterial pathogens and to predict pathogenicity and other relevant characteristics (e.g. virulence, AMR).
- Discuss the curation and analysis of WGS data, bioinformatics solutions.
- Discuss the coordination of efforts between the food, veterinary and human health sectors in order to obtain maximum benefits from the use of WGS for food safety.

Organizing Committee

Frank Møller Aarestrup, Dorte Lau Baggesen, João André Carriço, Jonathan Green, Benno ter Kuile, Martin C.J. Maiden, John Rossen, Marc Struelens, Johanna Takkinen.

Ernesto Liebana Criado, Maria Teresa da Silva Felício, Andrea Germini, Winy Messens, Valentina Rizzi.

General background documents

- Cao G, Meng J, Strain E, Stones R, Pettengill J, Zhao S, McDermott P, Brown E and Allard M, 2013. Phylogenetics and differentiation of *Salmonella* Newport lineages by whole genome sequencing. PLoS One, 8(2), e55687.

- ECDC (European Centre for Disease Prevention and Control), 2007. Surveillance of communicable diseases in the European Union. A long-term strategy | 2008–2013. (2007) Available online: http://www.ecdc.europa.eu/en/aboutus/key%20documents/08-13_kd_surveillance_of_cd.pdf (last accessed on 13/12/2013).
- ECDC (European Centre for Disease Prevention and Control), 2013. Long-term surveillance strategy 2014-2020. Stockholm: ECDC; 2013 Available online: <http://www.ecdc.europa.eu/en/publications/publications/long-term-surveillance-strategy-2014-2020.pdf>.
- ECDC (European Centre for Disease Prevention and Control), 2013. Surveillance of communicable diseases in Europe - a concept to integrate molecular typing data into EU-level surveillance. Version 2.4. 7 September 2011. Available online at: <http://www.ecdc.europa.eu/en/publications/Publications/surveillance-concept-molecular%20typing-sept2011.pdf> (last accessed on 13/12/2013).
- EFSA (European Food Safety Authority), 2009. Report on the availability of molecular typing methods for *Salmonella*, *Campylobacter*, verotoxigenic *Escherichia coli*, *Listeria monocytogenes* and *Staphylococcus aureus* isolates from food, animals and feedingstuffs in European Union Member States (and in some other reporting countries). EFSA Journal 2009, 272, 1-52. Available online at: <http://www.efsa.europa.eu/en/efsajournal/pub/272r.htm>.
- EFSA (European Food Safety Authority), 2013. Scientific Opinion of the Panel on Biological Hazards 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):3502, 84 pp. doi:10.2903/j.efsa.2013.3502. Available online at: <http://www.efsa.europa.eu/en/efsajournal/pub/3502.htm>
- Mather AE, Reid SW, Maskell DJ, Parkhill J, Fookes MC, Harris SR, Brown DJ, Coia JE, Mulvey MR, Gilmour MW, Petrovska L, de Pinna E, Kuroda M, Akiba M, Izumiya H, Connor TR, Suchard MA, Lemey P, Mellor DJ, Haydon DT and Thomson NR, 2013. Distinguishable epidemics of multidrug-resistant *Salmonella* Typhimurium DT104 in different hosts. Science, 341, 1514-1517.
- Mellmann A, Harmsen D, Cummings CA, Zentz EB, Leopold SR, Rico A, Prior K, Szczepanowski R, Ji Y, Zhang W, McLaughlin SF, Henkhaus JK, Leopold B, Bielaszewska M, Prager R, Brzoska PM, Moore RL, Guenther S, Rothberg JM and Karch H, 2011. Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. PLoS One, 6(7), e22751.
- Underwood AP, Dallman T, Thomson NR, Williams M, Harker K, Perry N, Adak B, Willshaw G, Cheasty T, Green J, Dougan G, Parkhill J and Wain J, 2013. Public health value of next-generation DNA sequencing of enterohemorrhagic *Escherichia coli* isolates from an outbreak. Journal of Clinical Microbiology, 51(1), 232-237.
- Zhang J, Cao G, Xu X, Jin H, Zhang Q, Chen J, Yang X, Pan H, Zhang X, Allard M, Brown E and Meng J, 2013. Whole-Genome Sequences of four *Salmonella* enterica Serotype Newport strains from humans. Genome Announcements, 1(3).

INTRODUCTION

The recent extraordinarily rapid pace of development of technologies in WGS determination for bacteria presents many opportunities for food borne disease surveillance but, as with all rapidly developing areas, also appreciable challenges. It is now clear that WGS is the isolate characterisation technique of the future; indeed, changes in costs for large centres are such that other technologies may soon become too expensive to perform compared to WGS. It is less clear, and certainly less well understood among public health practitioners, which of the multitude of approaches available for WGS determination and analysis should be deployed. In addition, novel means of assessing data and systematising it into 'plain language' reports that can be reported to public health databases and used for public health action need to be developed. Finally, it is important that WGS data are interpreted and made available in the context of what is already known with other approaches.

DISCUSSION POINTS

1. Discuss the WGS methods available for food borne pathogens with respect to: cost; speed; accuracy; convenience; practicality and feasibility.
2. Discuss the means available for the interpretation of WGS data for different applications (research and public health), including the accuracy of different approaches and the ways in which plain language reports can be generated for public health action.
3. Discuss how these data should be curated and stored in such a way as to ensure continuity of existing datasets. How to derive information from WGS data to predict e.g. serotype, phagetype, PFGE type, MLST, resistome, virulome?

BACKGROUND DOCUMENTS

Pallen, M. J., Loman, N. J. and Penn, C. W. (2010). High-throughput sequencing and clinical microbiology: progress, opportunities and challenges. *Current Opinion in Microbiology* 13, 625-631.

Jünemann, S., Sedlazeck, F. J., Prior, K., Albersmeier, A., John, U., Kalinowski, J., Mellmann, A., Goesmann, A., von Haeseler, A., Stoye, J. and Harmsen, D. (2013). Updating benchtop sequencing performance comparison. *Nature Biotechnology*, 31, 294-296.

Aarestrup, F. M., Brown, E. W., Detter, C., Gerner-Smidt, P., Gilmour, M. W., Harmsen, D., Hendriksen, R. S., Hewson, R., Heymann, D. L., Johansson, K., Ijaz, K., Keim, P. S., Koopmans, M., Kroneman, A., Lo Fo Wong, D., Lund, O., Palm, D., Sawanpanyalert, P., Sobel, J. and Schlundt, J. (2012). Integrating genome-based informatics to modernize global disease monitoring, information sharing, and response. *Emerging Infectious Diseases* 18, e1.

Köser, C. U., Ellington, M. J., Cartwright, E. J., Gillespie, S. H., Brown, N. M., Farrington, M., Holden, M. T., Dougan, G., Bentley, S. D., Parkhill, J. and Peacock, S. J. (2012). Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. *Plos Pathogens*, 8, e1002824.

Didelot, X., Bowden, R., Wilson, D. J., Peto, T. E. A. and Crook, D. W. (2012). Transforming clinical microbiology with bacterial genome sequencing. *Nature Reviews Genetics*, 13, 601-12.

Maiden, M. C., van Rensburg, M. J., Bray, J. E., Earle, S. G., Ford, S. A., Jolley, K. A. and McCarthy, N. D. (2013). MLST revisited: the gene-by-gene approach to bacterial genomics. *Nature Reviews Microbiology* 11, 728-736.

DISCUSSION GROUP 2 – *Curation and analysis of WGS data: bioinformatics solutions*

INTRODUCTION

Perhaps the major challenge to the full exploitation of the evolving WGS technologies for trans-national public health microbiological surveillance is how to curate and integrate the data, in order to provide relevant analysis and visualization of those results to public health professionals, so they can act upon them. A key objective has to be the harmonization of data standards and analysis techniques, in order to obtain comparable and portable results from WGS analysis. The aim of this discussion group is to bring together international experts who are leading in all of these areas to identify the most relevant technologies and approaches in order to deliver solutions for Europe.

Whilst some discussion points will be technical around computational infrastructure requirements and database solutions, issues of common data release policies, common quality standards and quality assurance and curation relating to bioinformatics pipelines will also be covered.

DISCUSSION POINTS

1. Discuss the challenges linked to the quality evaluation, annotation, interpretation and storage of the huge amount of information provided by WGS, from raw data to genome assembly and analysis results.
2. Discuss the harmonisation of approaches used for WGS data analysis, including development of genome analysis pipelines and software availability (e.g. open source versus commercial) and including the feasibility of international standards for WGS data analysis
3. Discuss the benefits for public health of specialised online genomic databases for sharing this WGS data and potentially associated metadata and algorithms allowing for real-time data analysis and visualisation.

BACKGROUND DOCUMENTS

Francisco, A.P., Vaz, C., Monteiro, P.T., Melo-Critino, J., Ramirez, M. and Carriço, J.A., 2012. PHYLOViZ: Phylogenetic inference and data visualization for sequence based typing methods. *BMC Bioinformatics*, 13(1), p.87.

ECDC technical consultation on harnessing genomics for epidemiological surveillance. <http://www.ecdc.europa.eu/en/publications/Publications/surveillance-harnessing-genomics-meeting-paris-2013.pdf>.

Carriço, J.A., Sabat, A.J., Friedrich, A.W., Ramirez, M., on behalf of the ESCMID Study Group for Epidemiological Markers (ESGEM), 2013. Bioinformatics in bacterial molecular epidemiology and public health: databases, tools and the next-generation sequencing revolution. *Eurosurveillance*, 18(4),p.20382. Available online: <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20382>

Jolley, K.A. and Maiden, M.C.J., 2010. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics*, 11, p.595.

Gardner, S.N. and Hall, B.G., 2013. When Whole-Genome Alignments Just Won't Work: kSNP v2 Software for Alignment-Free SNP Discovery and Phylogenetics of Hundreds of Microbial Genomes. *PLoS ONE*, 8(12), p.e81760.

DISCUSSION GROUP 3 – WGS of foodborne pathogens: cross-sectorial coordination and international cooperation

INTRODUCTION

The outputs of applying WGS-based typing monitoring in animal and food would be used to carry out further integrated analyses with corresponding surveillance data from cases of infections in humans. The possibility of such integrated analysis improving public health is the main driver for the establishment of the typing-based monitoring within animals, food and feed. This could be achieved by means of enhanced outbreak detection and identification of the sources and transmission pathways for pathogens. The added value of the molecular typing approach to surveillance of food-borne pathogens (ECDC, 2013) was strongly supported by the EU Commission¹, who in 2012 requested both EFSA and ECDC for technical support and collaboration on the collection of data on molecular testing of food/animal and human *Salmonella*, *Listeria monocytogenes* and Shiga toxin-producing *Escherichia coli* isolates.

At the EU level an integrated multidisciplinary approach to surveillance of food-borne pathogens was acknowledged by Directive 99/2003/EC and Decision 1082/2013/EU, which provide criteria for data collection in both human and non-human (i.e. food, animal and feed) sectors. The need for a strong link between data from public health, animal health, and food safety laboratories, including robust epidemiological data as well as close cooperation between MSs, ECDC and EFSA was highlighted by the Commission¹. This requires a high level of interoperability and data integration among all relevant existing monitoring and public health surveillance databases, which is best ensured by addressing harmonisation issues when data is established for integrated analysis across sectors.

Cross-sector and international coordination is essential in order to obtain harmonised surveillance/monitoring including data collection, management and analysis, with the final aim of achieving full integration of efficient and effectively managed WGS databases for food-borne pathogens. The need for harmonisation does also include a need for discussion of and agreement on how to manage questions on legal requirements, intellectual property and data protection issues and confidentiality related to surveillance/monitoring in all sectors (human, food, feed).

The key issue for the usefulness of the data managing system for integrated analysis across sectors will be that those data included from food and feed as well as from humans and environment have a high level of reliability and that they are comparable at the relevant level. A way to obtain this will be through intensive endeavours for harmonisation and standardisation in the different step of the information cycle including focus on the quality of the epidemiological metadata as well as the laboratory results. However, it must be acknowledged that differences between sectors in cultures, legal frameworks and sampling strategies represent a complex challenge for integration of surveillance systems.

DISCUSSION POINTS

¹ See vision paper from the European Commission on the development of data bases for molecular testing of food-borne pathogens in view of outbreak preparedness available at: http://ec.europa.eu/food/food/biosafety/salmonella/docs/vision-paper_en.pdf

1. What are the challenges to integrate WGS analysis into routine monitoring, surveillance and outbreak preparedness as basis for integrated analysis within and across sectors?
2. Discuss the coordination of efforts between the food, veterinary and human health sectors in order to obtain maximum benefits from the use of WGS for food safety and public health protection.
3. Discuss the speed and opportunities for collaboration towards development and validation of WGS applications across sectors (i.e. human, animal and food). Discuss the capacity building and transition management challenges in introducing new WGS technologies versus old typing technologies.

BACKGROUND DOCUMENTS

Aarestrup, F. M., Brown, E. W., Detter, C., Gerner-Smidt, P., Gilmour, M. W., Harmsen, D., Hendriksen, R. S., Hewson, R., Heymann, D. L., Johansson, K., Ijaz, K., Keim, P. S., Koopmans, M., Kroneman, A., Lo Fo Wong, D., Lund, O., Palm, D., Sawanpanyalert, P., Sobel, J. and Schlundt, J. (2012). Integrating genome-based informatics to modernize global disease monitoring, information sharing, and response. *Emerging Infectious Diseases* 18, e1.

ECDC (European Centre for Disease Prevention and Control), 2013. Surveillance of communicable diseases in Europe - a concept to integrate molecular typing data into EU-level surveillance. Version 2.4. 7 September 2011. Available online at: <http://www.ecdc.europa.eu/en/publications/Publications/surveillance-concept-molecular%20typing-sept2011.pdf>.

Vision paper from the European Commission on the development of data bases for molecular testing of food-borne pathogens in view of outbreak preparedness available at: http://ec.europa.eu/food/food/biosafety/salmonella/docs/vision-paper_en.pdf