ISSN 2363-2240

EFSA Scientific Colloquium
Summary Report

20

USE OF WHOLE GENOME SEQUENCING (WGS) OF FOOD-BORNE PATHOGENS FOR PUBLIC HEALTH PROTECTION

16-17 June 2014, Parma, Italy

efsa
European Food Safety Authority


efsa
European Food Safety Authority

EFSA Scientific Colloquium
Summary Report

20

USE OF WHOLE GENOME SEQUENCING (WGS) OF FOOD-BORNE PATHOGENS FOR PUBLIC HEALTH PROTECTION

16-17 June 2014, Parma, Italy

Acknowledgements
EFSA thanks Frank Møller Aarestrup, Dorte Lau Baggesen, João André Carriço, Jonathan Green, Martin C.J. Maiden, John W.A. Rossen, Johanna Takkinen, and the EFSA staff Maria Teresa Da Silva Felicio, Andrea Germini, Ernesto Liebana Criado, Winy Messens and Valentina Rizzi for the preparatory work as the Organising Committee, as well as Marc Struelens for acting as overall chair and Benno Ter Kuile for acting as overall rapporteur to the Colloquium.

ISSN 2363-2240
ISBN 978-92-9199-621-6
doi: 10.2805/66246

About EFSA

The European Food Safety Authority (EFSA) was established and funded by the European Community as an independent agency in 2002 following a series of food scares that caused the European public to voice concerns about food safety and the ability of regulatory authorities to fully protect consumers.

In close collaboration with national authorities and in open consultation with its stakeholders, EFSA provides objective scientific advice on all matters with a direct or indirect impact on food and feed safety, including animal health and welfare and plant protection. EFSA is also consulted on nutrition in relation to Community legislation.

EFSA's work falls into two areas: risk assessment and risk communication. In particular, EFSA's risk assessments provide risk managers (EU institutions with political accountability, i.e. the European Commission, European Parliament and Council) with a sound scientific basis for defining policy-driven legislative or regulatory measures required to ensure a high level of consumer protection with regards to food and feed safety.

EFSA communicates to the public in an open and transparent way on all matters within its remit. Collection and analysis of scientific data, identification of emerging risks and scientific support to the Commission, particularly in case of a food crisis, are also part of EFSA's mandate, as laid down in the founding Regulation (EC) No 178/2002 of 28 January 2002.

For more information about EFSA, please contact:

European Food Safety Authority
Via Carlo Magno 1A
43126 Parma
ITALY

Tel. +39 0521 036 111
Fax +39 0521 036 110
www.efsa.europa.eu

TABLE OF CONTENTS

I. INTRODUCTION 11

II. ABSTRACTS OF SPEAKERS IN THE OPENING PLENARY SESSION 15

III. SUMMARY OF DISCUSSION GROUPS RESULTS 23

1. *Discussion Group 1 – WGS of food-borne pathogens in action* 23
2. *Discussion Group 2 – Curation and analysis of WGS data: bioinformatics solutions* 27
3. *Discussion Group 3 – WGS of food-borne pathogens: cross-sectorial coordination and international cooperation* 34

IV. OVERALL CONCLUSIONS 41

V. ABBREVIATIONS 47

VI. REFERENCES 49

VII. ANNEXES 53

I. INTRODUCTION

The 20th meeting in the EFSA Scientific Colloquium Series was held in Parma, Italy on 16-17 June 2014 and addressed the "*Use of whole genome sequencing (WGS) of food-borne pathogens for public health protection*".

Molecular typing methods for food-borne pathogens are beginning to be routinely applied worldwide for public health protection (e.g. investigating food-borne outbreaks, identifying strains of food-borne bacteria with high virulence potential or resistance to antimicrobials). 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 bacterial WGS methods for food safety applications. The potential of WGS for a large variety of such applications is now actively being considered in several areas including: pathogen characterisation and typing, outbreak detection, risk assessment and high-resolution epidemiology.

While in the USA institutions like the Centers for Disease Control and Prevention (CDC) and the Food and Drug Administration (FDA) have taken initiatives to facilitate a more wide-spread application of WGS, in the European Union (EU) there is currently limited experience in the use of WGS methods in microbial food safety. Most of the limited experience comes from retrospective studies that have followed from outbreak investigations. Still, the potential application of WGS to predict phenotypes of interest (e.g. pathogenicity, antimicrobial resistance (AMR)) of the strain under investigation may provide risk assessors with a powerful tool.

The European Food Safety Authority (EFSA), whose mission is to improve the European Union (EU) food safety and ensure a high level of consumer protection, has shown an increasing interest in WGS over recent years. One of the key recommendations of the EFSA 10th Anniversary Conference[1] was to build a centralized WGS database for food-borne pathogens. The EFSA Advisory Forum further encouraged EFSA to actively promote WGS technology for food safety purposes. In line with these recommendations EFSA has strengthened its collaborations with the European Centre for Disease Prevention and Control (ECDC) and the European Union Reference Laboratories (EURLs) leading to the development of systems for molecular typing of food-borne pathogens. This builds upon the experience of the ECDC molecular surveillance pilot project 2012-14 that links together public health reference laboratories across Europe for real-time sharing of molecular

1 http://www.efsa.europa.eu/en/events/event/121107.htm

typing data on selected pathogens. In addition, the EFSA BIOHAZ Panel recently concluded a self-task mandate on the "Evaluation of molecular typing methods for major food-borne microbiological hazards and their use for attribution modelling, outbreak investigation and scanning surveillance".

Therefore, the moment seemed appropriate for EFSA to organise this Scientific Colloquium to discuss the use of WGS of food-borne pathogens for public health protection, in particular to identify challenges and opportunities and address unresolved issues that impede the implementation of WGS on a large scale for food safety purposes.

The scope of the Colloquium was to bring together leading scientists, representatives of international and European organizations and national food safety authorities to 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 particular requirements relating to WGS data. The specific objectives of the Colloquium were to discuss in an open scientific debate:

- the current issues, benefits and future challenges of WGS of food-borne pathogens in public health protection in comparison with current methods;
- the analysis and the interpretation of WGS data (e.g. analysis of single nucleotide polymorphisms (SNPs), gene by gene comparisons) in order to ascertain diversity, similarity and relatedness of food-borne bacterial pathogens and to predict pathogenicity and other relevant characteristics (e.g. virulence, AMR);
- the curation and analysis of WGS data and bioinformatics solutions;
- 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.

The Colloquium was attended by approximately 90 experts from Europe, USA and Canada and the present publication reports the abstracts of speakers in the opening plenary session and summarizes the discussions and conclusions of the Colloquium.

II. ABSTRACTS OF SPEAKERS IN THE OPENING PLENARY SESSION

Establishing a whole genome sequence-based national network for the detection and traceback of food-borne pathogens

Marc W. Allard, *Food and Drug Administration (FDA), USA*

Additional Authors: Peter Evans, Errol Strain, Ruth Timme, Chris Keys, Steve Musser.

This study outlines how tools will be implemented to create a pathogen detection network called GenomeTrakr, where state, federal and international public health agencies can share data to build a public and transparent reference database with data deposited into a public genomic database (NCBI). Herein we describe the components of the NGS pathogen diagnostic network that includes the generation of a large reference database at UC Davis under the 100K genome project, as well as case studies and current integration among a pilot study consisting of 7 state public health laboratories (AZ, FL, MD, MN, NY, VA and WA) and 11 federal laboratories. Details of the success and failure will be provided concerning communication, coordination, data acquisition, assembly, storage, and analysis. Several recent case studies will be reported in this initial pilot study. The hardware and software implemented allows us to compare and cluster complete genomes of hundreds of taxa at a time, and the software outputs phylogenetic trees for source tracking of food and environmental isolates. Herein, we report enhanced molecular epidemiological insights gained by comparative analysis of *Salmonella* and *Listeria* genomes previously deemed indistinguishable by conventional subtyping methodologies. These results demonstrate an important investigative role for NGS tools within a regulatory environment while highlighting the novel additional insights provided to epidemiological investigations through comparison to a reference database. The GenomeTrakr network of state, federal and international laboratories have released >4,000 unpublished draft genomes for food safety into the SRA database.[2] The FDA also recently opened a public access url describing their whole genome sequencing program[3].

2 http://www.ncbi.nlm.nih.gov/bioproject/183844

3 http://www.fda.gov/Food/FoodScienceResearch/WholeGenomeSequencingProgramWGS/default.htm

Whole genome sequencing of food-borne pathogens: experiences from the reference laboratory

Kathie Grant, *Public Health England (PHE), UK*

WGS is set to revolutionise how microbiology reference laboratories deliver their services. These changes will have a direct impact on the information available for the detection and management of infectious diseases. WGS provides the opportunity to perform bacterial strain identification and detailed finger printing using a single technique. Not only will this reduce the time to results but the increase in information available will improve our understanding of bacterial pathogens and outbreak situations enabling better prevention and control measures to be implemented. WGS has been called the ultimate epidemiological typing method for tracing human cases to the source of the outbreak and innovations in deoxyribonucleic acid (DNA) sequencing technologies mean that we can now deliver the total genetic information of each bacterial strain within a time frame that will directly influence clinical and public health practice.

The Gastrointestinal Bacterial Reference Unit at Public Health England is currently working together with our epidemiological colleagues to implement WGS for national surveillance of food-borne bacterial pathogens and for outbreak investigation. We have used WGS to support the management of a number of recent outbreaks of infectious gastrointestinal illness and it is proving to be an invaluable tool in identifying and clarifying relationships between strains and potential food sources. This presentation will describe the rationale for using WGS to replace traditional reference microbiological methods and share our experience of implementation, using *Salmonella* as an exemplar.

One disrupting technology fits it all – towards standardized bacterial whole genome sequencing for global surveillance

Dag Harmsen, *University of Münster, DE*

Infectious diseases remain a major cause of morbidity and mortality worldwide, representing major threats to human health. They spread through human populations, among animals and humans, in livestock, and via the food chain, making an effective trans-sector 'One Health' approach essential for their control. Whole genome sequencing (WGS) using next generation sequencing (NGS) technology provides comprehensive information for pathogen studies of all types; however, the cross-border and cross-sector application of this approach is hampered by a lack of standardization.

To achieve this goal we propose taking a genome-wide gene by gene (core genome MLST [cgMLST] or MLST+) approach. In contrast to the highly popular single nucleotide polymorphism (SNP) analysis, the MLST+ approach allows for an additive expandable nomenclature that is easily portable, storable and low compute and bandwidth retrievable barcode of every bacterial isolate sequenced by WGS NGS world-wide.

Canada's IRIDA project for genomic epidemiology of food-borne pathogens

Gary Van Domselaar, *Public Health Agency of Canada, CA*

We are in the midst of a technology revolution that is rewriting the way clinicians, public health workers and regulators handle issues of infectious disease. Bioinformatics and genomics technologies are redefining pathogen surveillance, transmission analysis, outbreak response, diagnosis, prevention and control. Implementing these modernized molecular approaches into food safety practice is a Canadian as well as an international priority. This presentation introduces development efforts on Canada's Integrated Rapid Infectious Disease Analysis (IRIDA) platform for infectious disease genomic epidemiology and its role in Canada's current program to modernize food and water safety.

Identification and characterization of food-borne pathogens by whole genome sequencing: a shift in paradigm

Peter Gerner-Smidt, *Centers for Disease Control and Prevention (CDC), USA*

Whole genome sequencing (WGS) is a technology that likely will transform public health microbiology in a few years. It is a technology that may be used to subtype pathogen with unheard precision and most phenotypic traits may be predicted from the sequence of their encoding genes. Therefore, WGS likely will replace most of the numerous workflows used in public health laboratories to characterize food-borne pathogens into one consolidated workflow.

Implementation of WGS will enable faster detection, more targeted investigations and faster control of outbreaks if resources are available to follow upon them. Regarding surveillance of sporadic infections WGS will enable more efficient microbiological source attribution and it will allow efficient targeting scarce public health resources to address the most important problems, e.g. rapidly spreading or emerging pathogens/strains, strains of particular virulence or antimicrobial resistance.

The biggest threat to the successful implementation and use of WGS in public health is no longer technical but rather related to the sharing of data (raw sequence data and associated metadata[4]). If the political, legal and psychological obstacles to free data sharing are not removed, genomics cannot possibly reach its potential for food safety but will be reduced to that of any other typing method.

4 In the context of this scientific colloquium report, metadata refers to epidemiological data (i.e. provenance and phenotype information) associated to the WGS data of a specific bacterial strain.

Genomics Integration in Food Inspection

Catherine Carrillo, *Canadian Food Inspection Agency (CFIA), CA*

Recent developments in the field of pathogen genomics herald a new paradigm for analytical food microbiology in which pathogenic bacteria will be characterized on the basis of their genetic profile rather than traditional approaches relying on morphological behaviours. The ability to identify gene markers associated with virulence and other properties relevant to the identification, risk profiling and typing of food-borne bacterial isolates will play a critical role in informing regulatory decisions and tracing sources of food contamination. In addition, availability of comprehensive databases of pathogen genomic sequences will enhance method development and validation activities, as they will enable robust in silico evaluation prior to wet-lab validation. These applications require assurance of the reliability of the genomic information before it can be incorporated in the regulatory decision making process. The implementation of whole genome sequence information for characterization of food-borne pathogens, molecular method validation and for ad hoc method development will be discussed.

III. SUMMARY OF DISCUSSION GROUPS RESULTS

1. Discussion Group 1 – WGS of food-borne pathogens in action

Chair: Martin C.J. Maiden, *University of Oxford, Department of Zoology, UK*

Rapporteur: John W.A. Rossen, Department of Medical Microbiology University of Groningen, *University Medical Center Groningen, NL*

1. Discuss the WGS methods available for food-borne pathogens with respect to: cost, speed, accuracy, convenience, practicality and feasibility.

A number of methods are currently available for the generation of WGS data, which differ with respect to cost, speed, accuracy, convenience, practicality and feasibility. However, at the time of writing there were two leading benchtop sequencing approaches: the LifeTechnologies PGM Ion torrent and the Illumina MiSeq. Both machines are unsuited for very high throughput sequencing (i.e. for hundreds or thousands of isolates) but are suitable for routine use in clinical microbiology and food safety laboratories. The Illumina platform was widely perceived as the preferred platform, on the basis of data accuracy, practicality of workflow, speed and affordability. It was recognized that WGS-techniques are continuously evolving and it will be exciting to learn what third generation sequencing techniques and concepts used by, e.g. Pacbio and Nanopore, will bring for detection of food-borne pathogens using WGS technologies.

The rapidly changing methods in WGS lead to the question of whether individual laboratories should invest in setting up a sequencing pipeline versus outsourcing the sequencing and/or data analyses to a third party. The latter has the advantage of not requiring major investment in the infrastructure required for WGS and enables non-specialist laboratories to exploit the latest platforms available for this rapidly evolving technology. Moreover, accreditation and validation issues then lie with the laboratory to which the sequencing is outsourced to and not with the clinical microbiology or food safety laboratory/department. However, outsourcing may not meet requirements of flexibility and speed that are key for the tasks of the clinical microbiology and/or food safety laboratories.

Using WGS in clinical and food safety settings has major implications for quality control and assurance for both data generation and analysis. There is an absolute need to define quality

metrics. These may already (partially) exist for creating and handling raw data but need to be more strictly defined for the subsequent data analysis. In addition, the quality of the generated data needs to be externally assessed by, e.g. participation in ring trials. Although such external quality assessment (EQA) programs could be organized by already existing organizations, there may be a need for a new structure. For this, there may be a role for the EURLs. It is clear that such initiatives should not only focus on a specific work field or geographical region but need to be organized in a cross-sector and globally based manner.

Currently, it is not clear if the reliability of the data generated and subsequent analysis techniques and interpretation criteria are sufficiently robust to be used in all applications of public health and food safety as, e.g. in legal proceedings. Therefore, there is an urgent need for establishing guidelines for the use of WGS in epidemiological typing of food-borne pathogens. Such guidelines should include rules on the minimal coverage of generated microorganism genomic data and their reproducibility and accuracy. In addition, WGS methods have a high discriminatory power with respect to comparing genomic data of microorganisms; but global agreement on microorganism-specific calibration and validation of the number of differences (be it SNPs or allele variants) between genomes that will lead to naming them differently needs to be established.

Running costs of current sequence analysis are variable and highly dependent on the platform employed and even more on throughput. For small laboratories it may not be feasible for sequencing to be performed locally, although this is an area of intense research and development and the appearance of very inexpensive 'near patient' sequencing platforms in the near future is likely. In any case, the costs of typing with WGS approaches are already competitive with, if not lower than, costs for conventional typing. In addition to costs incurred in generating the data there are also costs incurred for data analysis, and these may include the provision of new software and/or retraining or recruiting staff with new or different skill sets. To minimise such costs there is a need for capacity building by organizing workshops to share knowledge and to train staff, followed by supporting the trained staff by, for example, e-learning and instruction videos.

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.***

At the time of writing, two distinct and complementary approaches have been employed to analyze and interpret WGS data: SNP calling from a reference sequence and comparisons of allelic variants (gene-by-gene comparison). A plurality of approaches is likely to be an advantage, at least in the short term, while the field is in a period of rapid development and transition, but it is crucial that data are interoperable while methods are still improving. It is not clear which method will in the end be most suitable for food safety and clinical microbiology. The scientific community will ultimately decide which approach will be used as the standard or preferred method, as any scheme will have to have broad acceptability if it is to be widely implemented. The sequence and data analysis approaches employed also depend on the precise questions being addressed and it can be expected that for research questions a more diverse and variable range of analytical approaches will be employed, compared to routine applications where stability of analysis is of primary importance. As SNP calling depends on sequences present in both the bacterial genome to be analyzed and the reference genome with which it is compared, the method does not allow discovery of new genes or insertion elements. Therefore, *de novo* assembly and gene-by-gene analysis will always be required for in-depth analysis of the isolate under investigation. For both approaches, the stability of data and data analysis is key for public health applications.

Once analyses are completed, the outcome has to be reported in a plain language report that is widely understandable and interpretable by a wide range of public health professionals with a diversity of expertise. Consensus needs to be established on what to include and exclude and on how to convey complicated information in a simplified language without trivializing it. The information in the report must retain biological relevance and accuracy, but to communicate with policy makers it may be necessary to categorize the typing results into subtypes or to include a graphical representation of the relationships between the isolates to make it easy to understand. In addition, there may be a need for developing approaches for estimating probabilities of association, e.g. pathogen specific mutation rates and background population diversity. In addition, for gene-by-gene comparison, a common public access reference database is essential. It should be decided/clear who should provide required reference genomic data for this, define a reference genome and write recommendations on which reference genomes should be used. Such agreements should not only be made within the European Union but need to be established on a global level. Moreover, harmonisation across sectors (animal, food, human and environment) is essential, e.g. a stable nomenclature for clones, subclones, and types.

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?*

The establishment of a sustainable means of applying WGS within food safety and clinical microbiology requires the generation, curation, storage, and dissemination of data in appropriately structured databases. The volume of WGS data that will have to be generated to meet food safety, clinical and public health applications will be such that automatic, but supervised, curation will be essential to ensure the integrity of both deposited sequences and sequence-based typing schemes. The uncurated approach employed to date in archival databases such as GenBank will not be suitable for food safety and public health purposes, especially when these data are required as evidence in legal procedures.

WGS data should be publically available in real-time, together with associated epidemiological data[5] (i.e. provenance and phenotype information, including at least country and year of isolation). However, it will be challenging to collect these data and to decide which information should be made available, to whom, and when. Genome sequence databases need to be widely available at all times and therefore mirror databases should be established, so that there is no single point of failure and multiple agencies can share ownership. Interoperability of databases and backwards and forwards compatibility to other sequence based datasets is important; consequently compiled sequences should form the basis of such databases.

These databases must ensure the continuity of future and existing sequence-based datasets i.e. those collected by both sequence-based approaches such as multilocus sequence typing (MLST) and other non-sequence based approaches such as pulsed-field gel electrophoresis (PFGE) fingerprinting. Backwards compatibility with previous typing methods is desirable but not if establishing such backwards compatibility inhibits the implementation of WGS approaches. Therefore, cross-reading with other non-sequence based typing methods may be achieved by WGS of representative historical isolates. In the longer term it is desirable that the majority, if not all historical specimens, are analyzed by WGS approaches. This may become increasingly feasible as WGS costs decline and the technologies become increasingly implemented in the routine environment. More research is needed to link the genomic sequences reliably to phenotypes such as virulence and antimicrobial resistance.

5 In the context of this scientific colloquium report epidemiological data refers e.g. to the provenance and phenotype information associated to the WGS data of a specific bacterial strain.

2. Discussion Group 2 – Curation and analysis of WGS data: bioinformatics solutions

Chair: Jonathan Green, *Public Health England, London, UK*

Rapporteur: João André Carriço, *Faculdade de Medicina de Lisboa, PT*

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 analysed results.

Perhaps the main challenge of using WGS in a global public health setting is the definition of actionable standards for data quality and data analysis and interpretation. Data standards for both genomic and epidemiological data are essential for the effective global sharing of microbial data in such a way that it can be used for international outbreak investigations and surveillance studies. Without data standards, the comparison of data from multiple laboratories and analysis will be very complex, error-prone and unreliable.

At this time, an increasing number of institutes and organizations are intending to adopt genomic approaches and a clear message is that the need for standard protocols and quality metrics is not just limited to data processing and analysis. Sample preparation metrics are needed since they can be used to exclude a sample from being sent to a sequencer, or for validating posterior analysis. These metrics may be dependent on the technology used since different technologies have different requirements on the DNA quantity and quality. The target nucleic acid for the sequencing is important too. For example, DNA preparation kits vary in their ability to provide sequences from plasmids.

Two of the key parameters for successful sequencing are DNA concentration/quality metrics and library preparation metrics. NGS technology companies provide protocols to meet the specific needs of their platforms, and individual laboratories and Consortia such as 'The 100K project', the FP7 'Patho-NGen-Trace' and FDA/'GenomeTrakr' already have protocols developed. The challenge lies in bringing these together, comparing them and making them available to new adopters of NGS approaches.

The outputs from a sequencing run, after initial processing, are the reads stored in FASTQ files. This file format is a "*de facto*" standard that can be easily shared and this is the preferred format to be stored as it allows analysis and later re-analysis by different groups as the algorithms and software evolve. These files are large and the inevitable accumulation of large numbers (in the order of thousands) of genome sequences makes storage and

management of these a challenge. However, it is assumed that data transfer and storage technologies will continue to evolve and cost will continue to decrease. For the FASTQ file processing, defining global read-level quality metrics (such as average read length or depth of coverage against a reference genome) for most or all applications is very challenging at this stage as the metrics have to be chosen carefully on a 'case-by-case' basis, depending on the technology used, subsequent analysis or the organism under investigation.

The most widely emerging data analysis approaches for inferring strain relatedness are K-mer/SNP-based and gene-by-gene approaches, each requiring distinct quality assessment metrics across the whole analysis pipeline being used. The first approaches are based on the analysis of read content (K-mer analysis) or on mapping the obtained reads to a reference genome (SNP-based analysis), while gene-by-gene approaches have an extra step involving the assembly of the reads into contigs or scaffolds that will constitute a draft genome. Applications of both approaches for public health purposes have been published, but these are early models. The definition of appropriate metrics for analytical approaches such as phylogenetic analysis becomes even more complex. At this stage, there is no 'gold standard' for analysis as new approaches continue to emerge but are not yet fully tested. A wide diversity of data analysis approaches across a number of centres is recommended at this stage to allow a 'best practice' to develop.

Forums to allow knowledge sharing between multi-disciplinary groups developing pipelines include dedicated conferences (for example the 'Applied Bioinformatics & Public Health Microbiology Conference') and global consortia, for example Global Microbial Identifier (GMI) initiative, but these need to be supported by more frequent workshops and meetings in order to move forward quickly with their wider adoption. Case studies and ring trials, such as those being undertaken by GMI are useful in providing data on protocols in use for both sample preparation and analysis. Expert groups are then needed to define a consensus on the suitability of the protocols and appropriate quality metrics. This will also need to be supported with training/education on the WGS analysis. It is likely that from these fora, different analysis methodologies will emerge for each different aspect of public health investigations such as long term surveillance, phylogeographic analysis of bacterial spread or outbreak investigation. For analyses that are to be used for public health purposes, the establishment of standard operating procedures (SOPs) will be paramount for accreditation purposes.

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.***

The objective of harmonisation is to ensure that analyses on microbial strains provide data that can be meaningfully compared and, ideally, combined. Harmonisation of analysis is perhaps most simply achieved by use of a single analytical protocol, also commonly called a 'pipeline', by all those wishing to compare data for a similar purpose. The obvious extension of this assumption is that specific data are submitted to a single or small number of 'mirrored sites' where analysis using a defined common pipeline occurs. This is the model underpinning several global microbial gene analysis sites, for example the MLST databases, and this has been very successful for scientific purposes; particularly epidemiological and population biology studies. The alternative is for different groups to host an identical pipeline locally and, while there is no reason why this should not give harmonised analyses, there is the possibility of 'pipeline creep' i.e. local changes to the pipeline made in the different sites asynchronously, leading to divergence of final results. However, we assume an exponential increase in the number of strains being analysed in the future and, therefore, use of a single location for analysis has technical challenges. As strain numbers increase, the data storage requirements escalate proportionately and the dependence on international sharing of the large raw data files requires adequate internet bandwidth.

Effective international surveillance depends on a common nomenclature for the description of related strains at the supra- or sub-species level, providing a means for comparison and communication of the final analysis. Gene-by-gene approaches inherit this from previous MLST schemes and an obvious way forward is to continue the 'sequence type' nomenclature, albeit with extended gene profiles, where these already exist. Many of the current typing methods are not sequence-based schemes, for example serotyping, multiple-locus variable number tandem repeat analysis (MLVA), ribotyping and new nomenclatures will be required. The gene-by-gene approaches would be an obvious way forward, although there are concerns over backwards compatibility, which would require testing of significant collections of strains by the gene-by-gene method.

At this point it is harder to define nomenclature based on K-mer/SNP approaches for several reasons:

- The K-mer method is a fairly simple measure of strain relatedness based on overall similarity. However, nomenclature would be required to translate the sequence similarity distances into something that could be used to describe strains of the same species.

- The K-mer method can be very sensitive to the library preparation protocol. Changes in the commercial kits can lead to slightly different K-mer distributions, compromising the comparison with legacy K-mer data.
- SNP approaches may be required for very fine resolution analysis, for example in local outbreaks, but the establishment of nomenclature is difficult because SNPs will be called against different reference strains and there will be many thousands in number between distant strains of the same species.

An issue that is currently raised is the need for establishing backwards compatibility between WGS methods and currently used molecular typing approaches in order to access the wealth of historical information that is deposited in the available molecular surveillance databases. However, results of some conventional typing methodologies cannot be fully or faithfully reproduced from WGS data at this point due to their reliance on repeat regions (such as MLVA or *spa* typing in *Staphylococcus aureus*) or due to the resulting number of contigs (PFGE). Therefore, establishing equivalence tables between WGS and conventional type nomenclature would require WGS testing of fully characterized representative catalogues of strains to permit the comparison of results. The value in doing so for the purposes of backwards compatibility will be determined by the public health question.

Another current debate is between the use of open source, i.e. freely available, software versus closed source, commercial software. Bioinformaticians generally prefer to use open source solutions as they are relatively transparent, can be easily compared, are free and are supported by communities of people with similar objectives. Commercial solutions are preferred by non-bioinformaticians because they can provide better user interfaces and 'click and go' access to algorithms. Nevertheless, the true cost of analysis is often hidden but can be significant for both open source resources, where the cost is associated with hiring bioinformaticians to develop custom-built solutions, or commercial software, which has costs associated with purchase, maintenance and updates. These costs need to be considered within a sequencing project and at this stage it may be that both open source and commercial solutions are required within an organization to meet the needs of different operators (e.g. bioinformaticians, microbiologists) according to their level of bioinformatical expertise. For widespread, effective analysis of data, easy-to-use solutions are needed for most laboratories. However, it is very important that, for purposes of harmonisation and subsequently accreditation of analyses, commercial software solutions are not a 'black box' in which the details of the analysis performed are opaque to the scientist. The software has to offer users the option to access all the parameters of the algorithms used for the analysis to ensure that the different analytical approaches are comparable. An interim solution currently adopted by some organizations

to provide a user-friendly interface to a set of standardized pre-configured software is the use of workflow managers/schedulers, for example Galaxy, which are software that provide web-based access to individual tools as well as transparent pipelines that can be either commercial or open-source software. Such software has already been used to share pipelines in several bioinformatics projects since they reduce the user complexity for non-bioinformaticians. However these approaches still need to be supported with training of the end users for optimal usage.

An important step in the continuous harmonisation of data analysis is the need for new versions of software/algorithms to be re-evaluated as they are released. For that to occur, changes to versioning of software should be transparent. The analysis results produced by any pipeline should include descriptions of all the parameters and the versions of the software used (i.e. creation of complete log files) so that comparisons of results obtained at different times can be assessed for changes and to ensure some level of backwards compatibility between results. The GMI initiative has established working groups for the evaluation of pipelines and software and is also initiating ring trials for the comparison of methods used and results generated by different centres. A key challenge to these ring trials is that no 'gold standard' exists at this point so it is difficult to be sure of the 'true' result, but trials will be designed based on known results. Blind tests with identical and near identical isolates and also sporadic/outlier strains should be included. Clustering of results by participants' methods should be evaluated to better understand what the different methods could provide us with.

The growing sequencing capacity creates a need for evaluation of current software and algorithm scalability to assess if the huge influx of data will be able to be dealt with the current algorithms/pipelines. Current IT and database approaches are likely to be inadequate. New data compression solutions and 'big data' approaches such as the map-reduce approach should be explored as there is evidence that these have the capacity to meet the needs of other scientific areas dependant on analysis of enormous datasets. Most of the algorithms in current use would need to be revised to work with these approaches.

Finally, the ever-growing need for computing power for the analysis of complex data from thousands of strains needs to be addressed. A possible solution can be computing power on demand via the commercial 'cloud', which could provide expandable standardized infrastructure. However, an assessment of the cost of analysis for the different infrastructure and software approaches is needed and should be pursued, keeping in mind the everyday needs of the different institutions.

3. *Discuss the benefits for public health of specialised online genomic databases for sharing this WGS data and potentially associated metadata[6] and algorithms allowing for real-time data analysis and visualisation.*

The previous success of sequence-based typing methods for surveillance in public health applications such as *spa* typing and MLST for *Staphylococcus aureus* was largely due to the fact that they provided a level of resolution between strains that was useful for public health and provided a defined common language for describing strains. These were closed systems i.e. using an array of defined gene targets. With the Next Generation Sequencing technology we have now access to thousands of loci instead of just a few and the challenge lies on how can we make use of that wealth of information.

Online genomic databases should provide a variety of functions that are powerful tools for public health microbiology including:

- A repository of genomic and epidemiological data from strains of infectious agents (e.g National Center for Biotechnology Information Sequence Read Archive (NCBI SRA), European Nucleotide Archive (EBI ENA)).
- Access to analytical resources and nomenclature for strain 'typing' (e.g. Bacterial Isolate Genome Sequence Database (BIGSdb)).
- Real-time genome sequence comparison and visualization of closely related strain clusters to provide early warning of putative outbreaks (model under development at National Center for Biotechnology Information (NCBI)).

The data submitted to and provided by genomic data repositories should be the 'raw' FASTQ files as these provide the most useful format for re-analysis. These files are large and therefore substantial (petabyte) storage will be required in the coming years. It is unlikely that the current data submission and querying format of systems like NCBI SRA or EBI ENA will have the ability to scale up to the requirements needed for the sequence submission and data query for thousands of microbial genomes per day worldwide. An effort should be made at supranational levels to ensure the necessary continuous funding for these structures in order to update them to deal with the increasing data submission requirements and to keep them free-of-charge for the Scientific Community.

Another important point is how the data should be made available in the databases to provide a coherent and reproducible connection from the raw data to the final results.

6 In the context of this scientific colloquium report, metadata refers to epidemiological data (i.e. provenance and phenotype information) associated to the WGS data of a specific bacterial strain.

From a technical point-of-view, databases need to be interoperable and reachable in a machine-readable way to facilitate data analysis algorithm development. Database interoperability involves the ability to query multiple heterogeneous databases simultaneously from a single interface, and requires that each of the databases is built to support it. This can lead to the establishment of a federated, distributed database, potentially reducing the technical issues around data sharing and privacy, by providing hierarchical reporting to different stakeholders based on data access levels. The conjugated use of domain ontologies, i.e. the definition of a domain-specific terminology and the relationships between terms, and application programming interfaces (APIs) can provide the necessary interoperability and capacity to easily query across multiple databases. For private data in federated systems, encryption is needed for secure data transfer of both WGS data and associated epidemiological data between any authorized parties. This could lead to a greater level of data sharing since data protection could be enforced between authorized parties.

A 'One Health' perspective should drive the sharing of microbial data and analysis approaches. Interoperability between databases in clinical, environmental, food-borne and veterinary institutes would provide rapid identification of links between strains encountered in these different sectors and therefore provide significant public health added value. Nevertheless, a challenge to this concept may be the distinct rules for data ownership and release within the different disciplines. A move to revise and align governmental data ownership and release policy to allow public data sharing may be required to fully exploit the potential of the 'One Health' concept. For example, legislation at national level may not support data access provision to other countries. Isolates and data from industry may be difficult to access due to legal/political or commercial sensitivity issues. The utilisation of federated databases in the different settings with common ontologies may reduce some of these anxieties and lead to greater sharing of data compared with submission to a single global database. On a final note it should be remembered that any database or resource is only as good as the data that it contains and the accessibility of data. Assurance of data quality is therefore essential, as well as query mechanisms to securely find the needed data in the available databases. Whilst different database resources will have their own protocols for acceptance of submissions, standard quality protocols should be encouraged to apply across the different databases. The quality of the data accepted and the analytical workflows provided by any resource should also be transparent to allow any needed re-evaluation, as better sequencing technologies and data analysis algorithms are made available.

3. Discussion Group 3 – WGS of food-borne pathogens: cross-sectorial coordination and international cooperation

Chair: Dorte Lau Baggesen, *National Food Institute, The Technical University of Denmark, DK*

Rapporteur: Johanna Takkinen, *European Centre for Disease Prevention and Control (ECDC), SE*

1. What are the challenges to integrate WGS analysis into routine monitoring, surveillance and outbreak preparedness as a basis for integrated analysis within and across sectors?

Integration of WGS data into routine surveillance and monitoring faces several challenges both within and across food safety/veterinary and public health sectors. The group agreed to exclude industrial data from the scope of the discussion and keep the main focus on human data and official control data from food/feed/animals. A crucial principle, addressed several times in the discussion, was to strive towards an open data sharing policy.

Some felt that "routine monitoring and outbreak response" are too broad as concepts and that the focus should be on source attribution and early detection of emerging clones. On the other hand, routine surveillance covers systematic data collection for action, and thus may include the objectives for source attribution and recognition of emerging clones along with the outbreak detection and investigation. It was recognised that surveillance reporting systems involve stakeholders who are not familiar with this type of data, thus presenting challenges to implementation across teams from different disciplines. In addition, due to financial constraints and lack of expertise in bioinformatics, there are large variations across sectors in Member States' capacity and capability for WGS data production, analyses and interpretation, which hamper the development of a single solution for data integration across Europe.

Two obstacles in data sharing were identified: 1) the technical challenges related to sharing, managing and analysing raw sequence data; and 2) concerns related to sharing of related epidemiological data. The FDA, with the experience from, for example the GenomeTrakr network system, suggested starting with raw data sharing first and, after that, progress with developing routine communication of the results. At present, the FDA puts into their database information on country of origin, date and type of food, including any contaminated food imported from European countries. The specific needs of epidemiological data for different purposes, e.g. for source attribution, differ from those for outbreak detection. This highlights the importance of understanding the different

needs for different purposes as they set up the frame for the requirement to share a minimum set of epidemiological data along with the raw WGS data.

Confidentiality in data sharing was seen as problematic as the food and veterinary sectors both handle commercially sensitive data, which if misused, could potentially have an adverse economic impact on food/animal trade. Furthermore, there are no incentives for the food and veterinary sectors to share data and therefore, this would require identification of advantages and benefits for the sector or legal enforcement. Also, legislation on personal data protection poses certain limitations for data sharing. Centralised data should not include any information that could allow identification of a person but complete block of traceability of information back to the patient level may also create problems. Competent authorities are often legally bound to prevent the identification of individual companies through epidemiological data that may accompany WGS data and may therefore have to limit the amount of information they make available.

Members of the research community are interested in publishing their new findings before making data publicly available, which often conflicts with the public health needs and therefore the data may not be shared before publication. On the other hand, lack of confidence generated by not releasing sensitive background data may prevent authorities from releasing information. These challenges/barriers are, however, not linked specifically to WGS data, and experience on how to overcome these will be obtained through the newly established molecular typing database of food-borne pathogens in the interest of enhancing outbreak preparedness at EU level.

The rapid development of WGS technologies has triggered the launch of many international and national initiatives and study projects. Information arises rapidly from these scientific studies and large international high profile projects, which makes it difficult to have a complete overview of the state of the art and applicability to routine monitoring and surveillance. Along with this, it is apparent that storage of large amounts of data can become a problem.

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.***

As a result of fragmented development and different decision making processes in the US and EU, local governments and authorities develop their own visions, which are divergent. To promote global collaboration and data sharing, a common vision paper should be

developed. The European Commission has published a vision paper on the development of databases for molecular testing of food-borne pathogens with the aim of enhancing surveillance and outbreak preparedness in the EU[7] . This could be used as a starting point to develop a vision for data sharing, not only across food, veterinary and human health sectors, but also between scientific, industrial and regulatory communities as well as between professional disciplines e.g. microbiology, bioinformatics and epidemiology. Such a vision should be developed from an international/global perspective and so be applicable beyond the European Union.

The group discussed widely about a minimum dataset that should accompany the uploaded WGS data to a globally accessible open data source. Reporting of sequence data alone is of limited value for surveillance and scientific purposes, which require some minimum background data to be available. EFSA and ECDC are currently discussing this topic in the context of establishing a common joint molecular typing database for PFGE and MLVA results. As a minimum, the following information has been suggested for inclusion in the common typing database: source of sample (food, animal, feed, human), typing data and date of sampling. Additional epidemiological data will be managed by the respective agencies, i.e. food/veterinary data by EFSA and human data by ECDC. Information related to specific countries is considered potentially sensitive but discussion on this is on-going. There may be a need for further harmonisation of sampling schemes for the food/veterinary sector in the EU.

One way forward that serves the different needs of diverse communities could be to moderate the access and limit this to a relevant part of the data depending on the user group. Allowing open data mining for the subset of data could result in a detection of something new and trigger to cross-border collaboration in research projects and early warning on evolving outbreaks.

The FDA has started with a minimum dataset comprising State and the food type along the WGS data, but now also requests the reason for sampling to be added, e.g. if the sample is outbreak-related. Not all States have agreed with the minimal dataset and thresholds to share data vary among the States. As the Genome Trackr project has developed further, more States are willing to share data. Providing WGS phylogenies with a limited amount of epidemiological data has succeeded in convincing an increasing number of States of the value of the methodology. It is expected that more background data will be reported to the centralized FDA WGS database, e.g. after six months, allowing some time to prepare publications. Coordination on common nomenclature is ongoing in the US.

7 http://ec.europa.eu/food/food/biosafety/salmonella/docs/vision-paper_en.pdf

In addition to aiming for an open data sharing policy, quality and comparability of WGS data across sectors must be ensured. This includes several aspects. First, common protocols across sectors should be developed and shared, where feasible, keeping in mind that valid results may be produced with different technologies. As the protocols are dependent on technology and data analysis application tools, there should be a common record format and repository for sharing information on the specific technology and bioinformatics pipeline used. NBCI collects information on the sequencer used. Along with the rapid development and market release of new sequencers, there is a need to keep track of equipment used to produce WGS data. It is noteworthy that genomes from early days of sequencing are obviously not correct and ongoing improvements in accuracy are to be expected. Software and tools for data management and analyses should also be freely available. It was suggested to focus on determination of quality criteria and to acknowledge that creation of SOPs for the most critical parts of the workflow could be important in the early phases. Specific guidance on how to produce MLST data from WGS data was also requested.

Second, curation of WGS data is necessary, as the use of a common SOP does not guarantee comparable results. Even though this quality check can be done to a large extent automatically, individuals with bioinformatics expertise are needed to check and confirm the quality of data. Of particular importance is the capability to re-assemble the sequences later on. Based on experience at the FDA, the biggest challenge is related to mislabelling of sequences, warranting a centralised curation step at least in the beginning.

Third, EQA schemes should support laboratories performing microbial genome sequencing for public health protection. The FDA has asked the International Organization for Standardization (ISO) molecular typing sub-community to start developing more formal requirements for method validation and approval. The task to organise EQA schemes for national reference laboratories in the veterinary/food sector has been mandated to EURLs through EU funding. The European Commission is also promoting the cooperation across EURLs regarding WGS/NGS. Preferably, a centralised system to offer EQA schemes for WGS methods should be organised, including both sectors. At present, public health laboratories are currently offered EQA schemes for molecular typing through ECDC. The EURL for Verotoxin-producing *Escherichia coli* (VTEC) has successfully organised two EQA schemes for molecular typing jointly with an ECDC funded contractor thus, working efficiently towards comparability of data quality across sectors.

Fourth, reference material is needed for internal quality control and validation of the methodologies and EURLs as well as the Joint Research Centre (JRC) could be optimal sources for this purpose.

Fifth, collaboration across sectors with proof-of-principle prospective epidemiological studies is essential to validate the epidemiological concordance and added value of WGS as well as calibrate interpretation criteria of epidemiologically relevant WGS relatedness versus conventional molecular typing for major surveillance applications (source attribution for sporadic cases, outbreak detection, common source trace-back investigation).

The issues outlined above require global collaboration and harmonisation to the most feasible and reasonable extent possible.

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.***

The group agreed that the first step is to develop a communication strategy about concrete benefits to different stakeholders, including data providers, risk assessors, risk managers, industry, microbiologists and epidemiologists. Scientific research and proof-of-concept studies are needed to support the demonstration of added public health value, e.g. in international food-borne outbreak/epidemic situations. There is a need to have more user-friendly visualization tools available, not requiring complex computer programming/bioinformatics skills for analyses.

The WGS technology represents a significant advance in the field of microbiology and epidemiology and the need for a shift to this technology is unanimously recognised. However, the methodology is not widely accessed or understood yet. This is reflected in laboratories' willingness to start preparations for the change but as they lack the knowhow it is difficult to know how and where to start. There is a huge training need for understanding the methodology and providing visions on how to use it for different purposes e.g. through the Better Training For Safer Food (BTSF)-programme. WGS is foreseen to gradually become a routine analysis as *Salmonella* serotyping and the capacity to at least analyse and interpret it should be available in all EU countries. The actual sequencing can also be purchased from companies although there are still significant delays in receiving the results. As the results are needed faster for outbreak detection and investigation purposes, it was considered that food safety and public health laboratories responsible for such service provision should have access to their own sequencers.

The transition from old techniques to new ones requires additional funding. In the scope of the current EU framework programme for research and innovation Horizon 2020, the EU is about to fund a large research-project for developing a universal microbial WGS data management platform and a suite of bioinformatics tools for various stakeholders in science and risk assessment and management, focusing on food-borne pathogens. This project reaches only a proportion of the research, food safety and public health laboratories in the EU and a strategy to promote the spread of technology to a wider range of service laboratories in other countries must be developed. Transition represents not only a shift in technology but also in philosophy, which requires several discussion rounds in scientific and public health communities. EURLs play a crucial role in supporting the transition from old methods to WGS/NGS in the food sector and they should work in close collaboration with public health laboratory networks.

Countries should commence the necessary investment in national capacities, both in terms of equipment, application tools and competence building without delay. However, the replacement of current techniques with the new methodologies requires comparative analytical and epidemiological validation studies. These studies could be coordinated more across countries so that not every country would need to repeat similar analytical/epidemiological validation studies, which may become very costly. A limiting factor is the availability of appropriate tools and expertise for analyses while the equipment itself is not very expensive.

Finally, the current EU legislation on microbiological criteria for foodstuffs[8] is very specific and determines the methods that are to be used in the food/veterinary control programmes. These methods are often European Committee for Standardization (CEN)/ISO standard methods focusing on isolation of the pathogen. The new NGS technology opens new horizons for diagnostics and food analyses resulting in alternative sequence based methods for pathogen identification and a possible shift to non-culture-based DNA/ribonucleic acid (RNA) detection techniques coupled with sequencing. For example, *Salmonella* serotyping is the basis for EU-wide *Salmonella* control programmes. WGS technology may introduce a completely new way of sub-typing *Salmonella* isolates and consideration should be given to the possibility of updating the legal bases of these programmes. Therefore, it is recommended to open a discussion with policy makers and standardisation organizations on the potential impact of WGS development on EU legislation.

8 Commission Regulation (EC) No 2073/2005 of 15 November 2005 on microbiological criteria for foodstuffs. OJ L 338, 22.12.2005, p.1-26

IV. OVERALL CONCLUSIONS

WGS is a powerful tool that can be applied to a wide range of public health and food safety applications. Both SNP-calling and gene-by-gene comparison may be valid approaches for the different public health aims of WGS, as long as stable and comparable analytical procedures are used. Draft genomes can be assembled, using relatively modest computing power and widely available software, from data generated by commercially available high-throughput next generation sequencing (NGS) platforms. These include relatively inexpensive bench top instruments. Such draft genomes are suitable for the majority of practical applications in food safety. There are a number of ways in which this technology can be implemented in a food safety setting, including (i) installing the instruments in a food safety environment and (ii) outsourcing to commercial or other public health laboratories. The choice between these models will depend on local issues of throughput and available resources.

Further work is required to fully implement WGS in routine application. This includes: (i) the adoption of appropriate quality assurance/quality control (QA/QC) measures; (ii) the development and harmonisation of SOPs; (iii) the establishment of database infrastructure; and (iv) the generation and dissemination of appropriate sets of genomic data. Reference datasets should comprise not only representative complete (i.e. finished) genomes, that can be used for reference mapping if required, but also a larger number of draft genomes with accompanying catalogues of allelic diversity data, required for gene-by-gene approaches. These databases must be open access and widely available, although sensitive epidemiological data will have to remain available only to the competent authorities. Further work is also needed to link these data (i) with previous isolate characterization schemes and nomenclatures, although these may have to be revised in the light of the new insights available, but also (ii) to the phenotypic properties of the isolates from which the genomic data are obtained.

As the standardization of data generation, analysis, and storage is key, transnational organizations should aim to facilitate the implementation and integration of WGS across health sectors, working towards the goal of a 'one health' approach.

It is clear that currently only a few organizations are investing in WGS approaches for real-time analysis of food-borne and other pathogens but many institutes are exploring the utility of these approaches through research collaborations or 'in house' using bench-top machines. There are various factors that will hinder more widespread, routine use of these technologies, including funding for machines, IT infrastructure and impact on current workforce. From a bioinformatics viewpoint, access to necessary bioinformatics knowledge,

bioinformatics staff, skills and infrastructure for data storage and management (both locally and centrally) are key issues. The need to share knowledge of laboratory and bioinformatics approaches was recognised as necessary to expedite use of WGS. Some sequencing protocols are publicly available but knowledge sharing between bioinformatics groups is informal at best and frequent forums/meetings would be greatly helpful.

The greater challenges that lie ahead for the bioinformatics analysis of WGS applied to global public health are perhaps less of a technical IT nature, as technology and computer science is likely to continue to provide the solutions to handle the ever-growing data wave of microbial genomes. The generation and implementation of standards for data storage and sharing are clearly critical needs, since this will drive the data analysis development and will facilitate the harmonisation of WGS data analytical approaches. In terms of data analysis, a few models have emerged but we are still in the early days and, although the current models of data analysis should start to be used and their results compared, novel approaches could still emerge as the sequencing technology keeps evolving.

Models for international data sharing resources were discussed including a single, central resource for sequence storage, management and analysis compared with a system of federated databases underpinned by common ontologies and APIs. Concerns were expressed over the scalability of the current global databases (e.g NCBI and EBI) in their current formats to meet the potential tsunami of data likely to emerge in coming years and also whether the considerable funding required for a single central resource would be found. A network of federated, perhaps national, databases may have advantages in terms of resilience but adoption of common APIs and ontologies would be essential.

There are currently many developments in different countries to establish resources for centralised analysis of microbial genomes for public health and research purposes. It may be that, as the best approaches emerge, they will be most widely adopted but we should endeavour to ensure that there is interoperability between systems as they are designed and not as an afterthought. The discussions highlighted national differences in attitudes to data release and also differences between attitudes in public health and commercial groups.

In the meantime, however, there are urgent needs that can help move global approaches forward:

- Formal structures for multi-disciplinary collaboration, knowledge sharing and training in laboratory sequencing procedures and bioinformatics approaches is likely to expedite emergence and wider use of best practice. GMI has led some important initiatives but is currently unfunded.
- Wider development and adoption of ontologies and APIs will facilitate data sharing and integration across multiple resources. This is essential to ensure best value gained from data generated.
- It is recommended that the major structures leading the sharing of raw data (NCBI SRA and EBI ENA) continue to be adequately funded and supported, as they remain the major driver for this change by providing access to the baseline data for stakeholders in this field, from bioinformaticians to clinicians.

The most effective use of this technology is not yet obvious and this was discussed extensively during the colloquium. A major difference between WGS and the classical typing methods is that WGS allows all genes to be included in the analysis, instead of a well-defined subset of genes or variable intergenic regions. Therefore, the analysis of WGS data will yield new types of insight. As it is conceptually new, the legal and official systems are not yet adapted to the large-scale application of WGS to support food safety policies. At present it is not clear whether new legislation is required for optimal utilization of WGS, or if the existing legal framework is sufficient. Not only the legal power of evidence supplied by WGS needs to be defined, but also issues such as the determination of quality standards.

The need for cooperation is at the same time the Achilles' heel of WGS. There are several impediments for the free sharing of data. While the deposition of the microbial genomic sequence data in databases for public access beyond the control of the owner of the data is common practice, releasing sensitive epidemiological data will not easily become a routine procedure. Legal obstacles are to be expected and a careful balance must be struck between the desirable complete openness from a food safety point of view and the privacy and related concerns that necessitate confidentiality. Possibly a standard for encryption may need to be developed, to allow exchange of data to be limited to authorized parties only. Ignoring these issues is likely to considerably delay the successful large-scale implementation of WGS for public health at international level.

The overall conclusion of the EFSA Colloquium on whole genome sequencing is that the application of WGS for food safety and public health in general requires a paradigm

change and the development of new methodologies for processing these data. Since information obtained by WGS can only be utilized to fully benefit food safety and public health if all parties involved use procedures aimed at generating compatible results, international and inter-sector cooperation are crucial. To exploit the enormous potential of WGS, implementation in the EU must be initiated without delay.

Recommendations

Based on the considerations above the participants of the Colloquium recommended that in general:

- Educational interdisciplinary programmes should be established to foster a more in-depth understanding of WGS allowing epidemiologists, bioinformaticians, microbiologists and food safety experts to process, integrate and correctly interpret WGS and epidemiological data for the benefit of food safety and public health.
- Given that rapid technical developments can be expected, there is no need for stringent standardization of WGS methodologies, but quality metrics need to be defined.
- Databases should be established to contain WGS data for public health, standards should be agreed upon for the curation of these data and minimal demands on the associated epidemiological data should be decided upon.
- The legal consequences of the large-scale application of WGS should be reviewed and proposals should be made for any changes in food safety rules and regulations that may be needed.
- All parties involved should strive to harmonize terminology and data reporting in plain language accounts aimed at policy makers and other non-specialist readers.
- Software and data analysis tools should be developed that are at the same time transparent and easy to use for all staff involved in WGS data processing.
- Reference datasets should comprise both finished complete genomes and draft genomes with catalogues of allelic diversity data.

Specific recommendations for the EU:

- EFSA and ECDC should assume a leading role within the EU framework to stimulate, steer and coordinate efforts for the application of WGS across health sectors to further food safety and protection of public health.
- EU stakeholders in public health should work together with EU Member States' competent bodies in food safety and public health towards a joint strategy and action plan to roll out WGS across sectors for enhanced One-Health surveillance of food-borne diseases.
- EU stakeholders in public health should strengthen their collaboration with international counterparts, especially in the US, to coordinate and harmonize the application of WGS for food safety and health protection.
- EFSA and ECDC should invest in the establishment of databases for molecular typing data and set up procedures to allow data reporting, access and analysis procedures for the safe and optimal use of the information to further public health.
- EU stakeholders in public health should contribute to the advanced training of the people who will produce, analyse and interpret WGS data in conjunction with epidemiological data for informing risk managers in food safety and public health.
- EFSA should stimulate and encourage scientific research on WGS and the implementation of the results to benefit food safety.
- EFSA should initiate and coordinate the development of an EU-wide strategy for communication on food safety issues involving WGS in cooperation with ECDC.
- EFSA and ECDC should instil a sense of urgency in all partners regarding implementation of WGS for food and public health safety across the EU.

V. ABBREVIATIONS

AMR	antimicrobial resistance
API(s)	application programming interface(s)
BIGSdb	Bacterial Isolate Genome Sequence Database
BTSF	Better Training For Safer Food
CDC	Centers for Disease Control and Prevention
CEN	European Committee for Standardization
DNA	deoxyribonucleic acid
EBI ENA	European Bioinformatics Institute European Nucleotide Archive
ECDC	European Centre for Disease Prevention and Control
EFSA	European Food Safety Authority
EQA	external quality assessment
EU	European Union
EURL(s)	European Union Reference Laboratory(ies)
FDA	Food and Drug Administration
GMI	Global Microbial Identifier
ISO	International Organization for Standardization
JRC	Joint Research Centre
MLST	multilocus sequence typing
MLVA	multiple-locus variable number tandem repeat analysis
NCBI	National Center for Biotechnology Information
NCBI SRA	National Center for Biotechnology Information Sequence Read Archive
NGS	next generation sequencing
PFGE	pulsed-field gel electrophoresis
QA/QC	quality assurance/quality control
RNA	ribonucleic acid
SOP(s)	standard operating procedure(s)
SNP(s)	single nucleotide polymorphism(s)
VTEC	Verotoxin-producing *Escherichia coli*
WGS	whole genome sequencing

VI. REFERENCES

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: 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: http://www.efsa.europa.eu/en/efsajournal/pub/272r.htm.

EFSA Panel on Biological Hazards (BIOHAZ), 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).

DG1 (Background documents):

Aarestrup FM, Brown EW, Detter C, Gerner-Smidt P, Gilmour MW, Harmsen D, Hendriksen RS, Hewson R, Heymann DL, Johansson K, Ijaz K, Keim PS, 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.

Didelot X, Bowden R, Wilson DJ, Peto TEA and Crook DW, 2012. Transforming clinical microbiology with bacterial genome sequencing. Nature Reviews Genetics, 13, 601-12.

Jünemann S, Sedlazeck FJ, 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.

Köser CU, Ellington MJ, Cartwright EJ, Gillespie SH, Brown NM, Farrington M, Holden MT, Dougan G, Bentley SD, Parkhill J and Peacock SJ, 2012. Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. Plos Pathogens, 8, e1002824.

Maiden MC, van Rensburg MJ, Bray JE, Earle SG, Ford SA, Jolley KA and McCarthy ND, 2013. MLST revisited: the gene-by-gene approach to bacterial genomics. Nature Reviews Microbiology 11, 728-736.

Pallen MJ, Loman NJ and Penn CW, 2010. High-throughput sequencing and clinical microbiology: progress, opportunities and challenges. Current Opinion in Microbiology 13, 625-631.

DG2 (Background documents):

Carriço JA, Sabat AJ, Friedrich AW and 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.

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

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

Gardner SN and Hall BG, 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.

Jolley KA and Maiden MCJ, 2010. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics, 11, p.595.

DG3 (Background documents):

Aarestrup FM, Brown EW, Detter C, Gerner-Smidt P, Gilmour MW, Harmsen D, Hendriksen RS, Hewson R, Heymann DL, Johansson K, Ijaz K, Keim PS, 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.

VII. Annexes

Annex 1 Programme of the Colloquium

Annex 2 Participants at the Colloquium

ANNEX 1 PROGRAMME OF THE COLLOQUIUM

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

16-17 June 2014, Parma, Italy

PROGRAMME

Overall chair: Marc Struelens, European Centre for Disease prevention and control (ECDC), SE

Overall rapporteur: Benno Ter Kuile, Office for Risk Assessment, Netherlands Food and Consumer Product Safety Authority, Utrecht, NL

MONDAY 16 JUNE

08.30 - 09.00 Registration participants

08.15 - 08.45 Briefing meeting with all chairs and rapporteurs

09.00 - 12.30 SESSION 1: INTRODUCTORY PLENARY SESSION

09.00 - 09.10 Welcome and introduction to EFSA
Marta Hugas, European Food Safety Authority, IT

09.10 - 09.25 Objectives of the Colloquium
Rene Hendriksen, National Food Institute, Technical University of Denmark, Copenhagen, DK
Marc Struelens, European Centre for Disease Prevention and Control (ECDC), SE

09.25 - 09.45 Establishing a Whole Genome Sequence-Based national network for the detection and traceback of foodborne pathogens
Marc Allard, Food and Drug Administration (FDA), USA

09.45 - 09.55 Questions

09.55 - 10.15 Whole genome sequencing of foodborne pathogens: experiences from the reference laboratory
Kathie Grant, Public Health England, UK

10.15 - 10.25 Questions

10:25 - 10:55 COFFEE/TEA BREAK

10:55-11:15 One disrupting technology fits it all – towards standardized bacterial Whole Genome Sequencing for global surveillance
Dag Harmsen, University of Münster, DE

11:15-11:25 Questions

11:25-11:45 Canada's IRIDA project for genomic epidemiology of food-borne pathogens
Gary Van Domselaar, Public Health Agency of Canada, CA

11:45-11:55 Questions

11:55-12:15 Identification and characterization of foodborne pathogens by whole genome sequencing: a shift in paradigm
Peter Gerner-Smidt, Centers for Disease Control and Prevention (CDC), USA

12:15-12:25 Questions

12:25-12:35 General discussion

12:35-12:40 Introduction to Discussion Groups
Andrea Germini, European Food Safety Authority, IT

12:40-14:00 LUNCH

14.00-18.00 SESSION 2: DISCUSSION GROUPS (DG)

DG 1 WGS of foodborne pathogens in action

Chair: *Martin C.J. Maiden, University of Oxford, Department of Zoology, UK*

Rapporteur: *John Rossen, University Medical Centre Groningen, NL*

During the session a presentation will be shared by *Catherine Carrillo, Canadian Food Inspection Agency, CA* on Genomics Integration in Food Inspection

DG 2 Curation and analysis of WGS data: bioinformatics solutions

Chair: *Jonathan Green, Public Health England, London, UK*

Rapporteur: *João André Carriço, Faculdade de Medicina de Lisboa, PT*

DG 3 WGS of foodborne pathogens: cross-sectorial coordination and international cooperation

Chair: *Dorte Lau Baggesen, National Food Institute, The Technical University of Denmark, DK*

Rapporteur: *Johanna Takkinen, European Centre for Disease prevention and control (ECDC), SE*

16:00 COFFEE/TEA BREAK

18:00 End of Discussion Groups

20:00 DINNER

TUESDAY 17 JUNE

09:00-10:00 SESSION 3: CONTINUATION OF DISCUSSION GROUPS
Including discussion on the outcomes of the discussion groups and the production of reports to the plenary session

10:00-10:30 COFFEE/TEA BREAK

10:30-13:30 SESSION 4: FINAL PLENARY SESSION

10:30-10:50 Report back from Discussion Group 1
John Rossen, University Medical Centre Groningen, NL

10:50-11:10 Discussion

11:10-11:30 Report back from Discussion Group 2
João André Carriço, Faculdade de Medicina de Lisboa, PT

11:30-11:50 Discussion

11:50- 12:10	Report back from Discussion Group 3 *Johanna Takkinen, European Centre for Disease prevention and control (ECDC), SE*
12:10-12:30	Discussion
12:30-13:00	General discussion
13:00-13:30	Take-home messages *Benno ter Kuile, Office for Risk Assessment, Netherlands Food and Consumer Product Safety Authority, NL*

13:30 COLLOQUIUM ADJOURNS

Organising Committee

Frank Møller Aarestrup, *National Food Institute, Technical University of Denmark, DK*

Dorte Lau Baggesen, *National Food Institute, The Technical University of Denmark, DK*

João André Carriço, *Faculdade de Medicina de Lisboa, PT*

Maria Teresa Da Silva Felicio, *European Food Safety Authority (EFSA), IT*

Andrea Germini, *European Food Safety Authority (EFSA), IT*

Jonathan Green, *Public Health England, London, UK*

Benno Ter Kuile, *Office for Risk Assessment, Netherlands Food and Consumer Product Safety Authority, NL*

Ernesto Liebana Criado, *European Food Safety Authority (EFSA), IT*

Martin C.J. Maiden, *University of Oxford, Department of Zoology, UK*

Winy Messens, *European Food Safety Authority (EFSA), IT*

Valentina Rizzi, *European Food Safety Authority (EFSA), IT*

John Rossen, *University Medical Centre Groningen, NL*

Marc Struelens, *European Centre for Disease prevention and control (ECDC), SE*

Johanna Takkinen, *European Centre for Disease prevention and control (ECDC), SE*

ANNEX 2 PARTICIPANTS AT THE COLLOQUIUM

Name	Affiliation	Country	Discussion Group (DG)
Henk AARTS	National Institute for Public Health and the Environment (RIVM)	NL	2
Ana AFONSO	European Food Safety Authority (EFSA)	IT	3
Joakim ÅGREN	Swedish National Veterinary Institute	SE	1
Marc ALLARD	FDA CFSAN	US	3
Jose ARNAU	Novozymes A/S	DK	2
Dorte Lau BAGGESEN	National Food Institute, Technical University of Denmark	DK	3 - Chair
Anicet BLANCH	University of Barcelona	ES	1
Yannick BLANCHARD	French Agency for Food, Environmental and Occupational Health & Safety (ANSES)	FR	2
Anne BRISABOIS	French Agency for Food, Environmental and Occupational Health & Safety (ANSES)	FR	3
Rosella BROZZI	European Food Safety Authority (EFSA)	IT	3
Patrick BUTAYE	Veterinary & Agrochemical Research Center (CODA-CERVA)	BE	3
Sabrina CADEL SIX	French Agency for Food, Environmental and Occupational Health & Safety (ANSES)	FR	1
Paolo CALISTRI	IZSAM	IT	3
Cesare CAMMÀ	IZSAM	IT	1
João CARRIÇO	University of Lisbon, Faculty of Medicine	PT	2 - Rapporteur
Catherine CARRILLO	Canadian Food inspection Agency	CA	1

Name	Affiliation	Country	Discussion Group (DG)
John COIA	National Health Service	GB	3
Teresa COQUE	Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS)	ES	3
Maria Teresa DA SILVA FELICIO	European Food Safety Authority (EFSA)	IT	1
Robert DAVIES	AHVLA Weybridge	GB	1
Katrien DE BRUYNE	Applied Maths NV	BE	1
Alessandra DE CESARE	University of Bologna	IT	1
Sigrid DE KEERSMAECKER	Scientific Institute of Public health (WIV-ISP)	BE	1
Klara DE RAUW	National reference centre VTEC/ STEC, UZ Brussel	BE	1
Rikard DRYSELIUS	National Food Agency	SE	1
Stefan EMLER	SmartGene	CH	2
Andrea GERMINI	European Food Safety Authority (EFSA)	IT	Plenary
Peter GERNER-SMIDT	Centers for Disease Control and Prevention (CDC)	US	3
Kathie GRANT	Public Health England	GB	1
Jonathan GREEN	Public Health England	GB	2 - Chair
Beatriz GUERRA ROMÁN	Federal Institute for Risk Assessment (BfR)	DE	1
Tine HALD	National Food Institute, Technical University of Denmark	DK	1
Marja-Liisa HÄNNINEN	University of Helsinki, Department of Food Hygiene and Environmental Health	FI	3
Dag HARMSEN	University Münster (ZMK)	DE	2
Arie HAVELAAR	National Institute for Public Health and the Environment (RIVM)	NL	3
Edward HAYNES	Food & Environment Research Agency (FERA)	GB	3

Name	Affiliation	Country	Discussion Group (DG)
Rene HENDRIKSEN	National Food Institute, Technical University of Denmark	DK	2
Raymond HEYMANS	Netherlands Food and Consumer Product Safety Authority (NVWA)	NL	1
Marta HUGAS	European Food Safety Authority (EFSA)	IT	Plenary
Paul IN 'T VELD	Netherlands Food and Consumer Product Safety Authority (NVWA)	NL	3
Alexander INDRA	Austrian Agency for Health and Food Safety (AGES)	AT	2
Renata KARPISKOVA	Czech Veterinary Research Institute	CZ	3
Sonja KITTL	University of Bern, Institute of Veterinary Bacteriology	CH	1
Rauni KIVISTÖ	University of Helsinki, Department of Food Hygiene and Environmental Health	FI	1
Ivana KOLACKOVA	Czech Veterinary Research Institute	CZ	1
Niels LADEFOGED NIELSEN	Ministry of Food, Agriculture and Fisheries of Denmark	DK	1
Marco LEONI	European Food Safety Authority (EFSA)	IT	2
Ernesto LIEBANA CRIADO	European Food Safety Authority (EFSA)	IT	1
Bjorn-Arne LINDSTEDT	Akershus University Hospital	NO	1
Charlotta LÖFSTRÖM	National Food Institute, Technical University of Denmark	DK	2
Carla MAIA	National Institute of Health, Doutor Ricardo Jorge, INSA I.P.	PT	3
Martin MAIDEN	University of Oxford	GB	1 - Chair
Burkhard MALORNY	Federal Institute for Risk Assessment (BfR)	DE	2
Gerardo MANFREDA	University of Bologna	IT	3

Name	Affiliation	Country	Discussion Group (DG)
Monika MATT	Austrian Agency for Health and Food Safety (AGES)	AT	3
Winy MESSENS	European Food Safety Authority (EFSA)	IT	2
Susanne MOGENSEN	Danish Veterinary and Food Administration	DK	3
Stefano MORABITO	Istituto Superiore di Sanità (ISS)	IT	3
Petra MUELLNER	Epi-interactive	DE	3
Eva Møller NIELSEN	Statens Serum Institut	DK	3
Massimiliano ORSINI	IZSAM	IT	2
Sinikka PELKONEN	Finnish Food Safety Authority (Evira)	FI	3
Liljana PETROVSKA	AHVLA Weybridge	GB	2
Annemarie PIELAAT	National Institute for Public Health and the Environment (RIVM)	NL	1
Stefano PONGOLINI	IZSLER	IT	1
Hannes POUSEELE	Applied Maths NV	BE	2
Zoltan PRAGAI	DSM Nutritional Products Ltd.	CH	1
Alan REILLY	Food Safety Authority of Ireland (FSAI)	IE	1
Valentina RIZZI	European Food Safety Authority (EFSA)	IT	3
John ROSSEN	UMCG	NL	1 - Rapporteur
Mirko ROSSI	University of Helsinki	FI	2
Werner RUPPITSCH	Austrian Agency for Health and Food Safety (AGES)	AT	1
Donal SAMMIN	Department of Agriculture, Food and the Marine	IE	3
Sonia SCARAMAGLI	COOP Italia - Laboratory	IT	1
Gaia SCAVIA	Istituto Superiore Sanità (ISS)	IT	3

Name	Affiliation	Country	Discussion Group (DG)
Annemieke SMET	Ghent University	BE	2
Robert STONES	Food & Environment Research Agency (FERA)	GB	2
Marc STRUELENS	European Centre for Disease Prevention and Control (ECDC)	SE	Overall Chair
Johanna TAKKINEN	European Centre for Disease Prevention and Control (ECDC)	SE	3 - Rapporteur
Benno TER KUILE	Netherlands Food and Consumer Product Safety Authority (NVWA)	NL	Overall Rapporteur
John THRELFALL	EFSA BIOHAZ panel	GB	3
Jean-Claude TWIZERE	University of Liege (ULG)	BE	1
Bernhard URL	European Food Safety Authority (EFSA)	IT	Plenary
Gary VAN DOMSELAAR	PHAC	CA	2
Maria VITALE	Istituto Zooprofilattico della Sicilia "A.Mirri"	IT	2
Pierre WATTIAU	Veterinary & Agrochemical Research Center (CODA-CERVA)	BE	2
Bart WEIMER	University of California Davis	US	2
Ewelina WÓJCIK	Proteon Pharmaceuticals S. A.	PL	2

CD CONTENTS

- PDF of the content of the printed version including annexes 1 (programme of the colloquium) and 2 (participants) with active bookmarks.
- A folder containing all presentations from the event in PDF format
- The programme of the Colloquium in html format with active links from the respective agenda items to the PDFs of the presentations, link to the PDF of the full report and link to the PDF of the participant list.