Next-generation sequencing: opportunities and limitations for human and animal health protection

BfR Communication No. 022/2020 of 14 May 2020

The European Food Safety Authority (EFSA) has published a scientific opinion on the potential utility of next-generation sequencing (NGS) for application in foodborne outbreak investigations and when testing food for microbial pathogens. Entitled “Whole genome sequencing and metagenomics for outbreak investigation, source attribution and risk assessment of food-borne microorganisms”, this opinion document also discusses the value of new molecular laboratory methods for risk assessment in food monitoring. The advantages and disadvantages of the various NGS methods are analysed in the context of the microbiological methods cited in current European Union legislation.

The German Federal Institute for Risk Assessment (BfR), in coordination with the Friedrich Loeffler Institute (FLI), has evaluated the application of the NGS methods concerning food safety and animal health based on the EFSA opinion. This evaluation investigates topics such as the challenges for data exchange and for the harmonisation of new methods among various laboratories. Therefore, staff working in institutions, laboratories and regulatory authorities, who are involved in the typing and characterisation of pathogens in food and animals, are the target audience.

NGS stands for the second and third generation of genetic sequencing, and offers the highest resolution yet available for determining the nucleotide sequence of a DNA molecule or genome.

For pathogens readily available as pure cultures, whole-genome sequencing (WGS) has established itself worldwide as an NGS method. In this process, the genetic material of the causative agent is isolated from the sick individual and compared with isolates of the same pathogen from food or animals. This enables the detection of relationships in the genetic material, allowing cases to be traced back to specific disease outbreaks at various locations. Another NGS method, known as whole-metagenome or shotgun metagenomic sequencing, involves harvesting the genetic material directly from a food or animal sample (for example), which may often contain an array of microbes. This enables the detection of non-culturable or difficult-to-culture microorganisms such as parasites or viruses. Metagenomic sequencing is suitable for use as an initial diagnostic method in cases where a specific pathogen is not yet suspected. Metagenomic sequencing also enables the discovery of new, previously unknown pathogens—such as Schmallenberg virus in 2011.
Joint assessment of the German Federal Institute for Risk Assessment (BfR) and the Friedrich Loeffler Institute (FLI) concerning the conclusions drawn in the scientific opinion published by the European Food Safety Authority (EFSA)

For a number of years now, the EFSA has been involved in various activities to evaluate the potential of next-generation sequencing (NGS) and whole-genome sequencing (WGS) methods for the purposes of pathogen characterisation and typing, investigations of foodborne disease outbreaks, and for targeted risk assessments. Above and beyond the pathogen characterisation and typing described by EFSA, NGS methods are also particularly suitable for the identification of pathogens (known as metagenomics). There is a wide range of overlap here with animal health, particularly in relation to the combating and avoidance of epizootic diseases—even though the two domains are governed by separate legislation. While the focus of the EFSA is, true to its mandate, on food safety and foodborne outbreaks, the potential of NGS and WGS methods must nevertheless be considered across multiple sectors, even while sector-specific differences are naturally taken into account.

Examples of activities in which the EFSA is involved include the preparation of scientific opinions, the co-funding of research projects on WGS, and the technical report prepared for the expansion of the molecular database of the European Centre for Disease Prevention and Control (ECDC) and EFSA in terms of WGS data (EFSA 2019a). The most recent activity in this context was the preparation of a scientific opinion as a self-tasking mandate for the EFSA’s BIOHAZ Panel on the application of WGS and metagenomics to the abovementioned areas, to which reference is made in this document (EFSA BIOHAZ Panel 2019).

This opinion covered the following topics (also referred to as the terms of reference (ToR) in the EFSA opinion):

ToR1: Evaluate NGS for possible use in foodborne outbreak investigations and hazard identification, while accounting for activities and experience from various countries, and underlining the added value for risk assessment.

ToR2: Critically analyse the advantages, disadvantages and limitations of existing NGS methodologies that would yield results comparable to those achieved by the microbiological methods cited in current EU food legislation (e.g. Salmonella serotyping, monitoring of Shiga-toxin producing E. coli (STEC) and antimicrobial resistance (AMR) testing), while taking benchmarking exercises into account.

The EFSA’s commissioning of this scientific opinion as a self-tasking mandate stems to an extent from deliberations about whether to incorporate the results of NGS methods into existing legislative regulations in the future and/or to even expand these to other NGS methods. Since a representative of the BfR also participated in the working group that prepared the opinion, this made it possible to contribute experience in the field of WGS that the BfR had acquired in the typing of food-associated pathogens by means of whole-genome sequencing.

The conclusions for ToR1 (page 50 of the EFSA opinion) summarise the potential and application areas of WGS and metagenomics for the abovementioned topics as follows:

In comparison to conventional typing methodologies, WGS offers a more detailed outcome—which is to say a higher phylogenetic resolution. This opens up new possibilities for the performance of differentiated typing as part of outbreak investigations, source-attribution studies
and hazard identification, as part of risk assessment in relation to issues in the context of food safety law and epizootic legislation.

The conclusions also state that WGS is useful for a variety of applications that require the use of specific kinds of bioinformatic analyses. Sequenced genomes can be used for ongoing outbreak investigations, while the discriminatory power, which is superior to previous methods used (e.g. serotyping, MLST), facilitates high-resolution molecular epidemiology. Accordingly, food contaminated with pathogens can be linked with sporadic human cases, for example, simplifying the epidemiological detection of outbreaks occurring in different geographical regions. The FLI also observes that the same applies to epizootic outbreaks such as brucellosis, salmonellosis in cattle or avian flu. The point is also made that the setting of absolute thresholds for the inclusion or exclusion of isolates within an outbreak is not prudent: other diagnostic or epidemiological data must always be drawn on when defining outbreaks.

For source attribution, i.e. the attribution of sources of specific pathogen subtypes to certain foods or species of animal, WGS offers the possibility of achieving greater precision by the development of refined models that take genome sequencing data into account. Here too, however, datasets should always be complemented with epidemiological data. As a culture-independent NGS methodology, metagenomics also offers major potential for the detection of outbreaks, for source-attribution studies and for risk assessments for non-culturables or slow-growing microorganisms, or for the identification of entirely new pathogens, such as Schmallenberg virus in 2011. As with WGS, the detection of genetic markers for antimicrobial resistance or virulence markers is also possible, for example. The future impact and application of metagenomics to specific application areas will depend on further development and harmonisation as well as the sensitivity achievable by its methods. In the field of detecting new viral pathogens, metagenomics is already taking a leading role. In the future, this role could be extended to the characterisation of genes against antimicrobial resistance (AMR) in samples.

Summing up, one may characterise WGS as having become one of the most significant methodologies in the field of microbiology, thanks to its well-established status in laboratories and its technological maturity. This methodology should therefore be promoted further and data that can be acquired by WGS should be included in risk assessment models in order to fully exploit their potential. The same applies for metagenomics: while this methodology is not yet a common part of the lab toolset—due if nothing else to the high cost of its applications—it nonetheless offers considerable potential for differentiated risk assessments and outbreak investigations in relation to non-culturables or slow-growing microorganisms, including viruses. Targeted metagenomics, i.e. the detection of markers determined beforehand by the sequencing of complex samples, could offer a low-cost alternative to untargeted detection methods. However, targeted metagenomics does not of course permit the detection of previously unknown genetic markers.

The conclusions for ToR2 include the following statements (page 50 f. of the EFSA opinion): WGS enables the collection of data about serotypes for Salmonella and STEC, as well as the presence of genetic resistance determinants. A high level of agreement between conventional microbiological methods and WGS-based methods can be attained. In addition, WGS can also be used to predict the serotypes for Salmonella and STEC isolates not previously serotypeable with conventional methods (agglutination). The authors recommend incorporat
ing WGS-based serotyping into relevant EU regulations requiring the serotyping of *Salmonella* and *E. coli* (e.g. (EU) No 200/2010, 517/2011, 200/2012 and 1190/2012, in agreement with (EC) No 2160/2003). This is because WGS-based serotyping provides a considerably more precise and nuanced picture of types than is possible with contemporary methods. Since the White-Kauffmann-Le Minor (WKL) scheme (Grimont and Weill 2007) for the determination of *Salmonella* serotypes is based on antigen-antibody reactions, this can lead to incorrect or incomplete serotyping results. Genotypic data on serotypes should therefore be integrated into the WKL scheme, so as to document the differences currently underpinning the inaccuracies within the WKL scheme and—in the future—to resolve them.

Differences may exist between the phenotype and the genotype when determining antimicrobial resistance, due to the variable expression of resistance genes (induced resistance). By going beyond this expression, WGS can therefore supply data about whether antimicrobial resistant genes are present in the genome and could potentially disseminate to other bacteria by gene transfer. A gradual integration of WGS data into the harmonised AMR monitoring programmes has been recommended in an earlier technical specification from the EFSA (EFSA 2019b).

During the transition period to NGS-based methods in reference laboratories, new challenges will arise in terms of data exchange and method standardisation. While metagenomics is viewed as a methodology with very high potential, it still exhibits limits in comparison with WGS-based methods (such as sensitivity and a lack of harmonisation). On the view of the BfR and FLI, however, the application of metagenomics is indispensable for the detection of new viral and bacterial pathogens.

The concluding statements recommend viewing WGS methods as the equal of comparable microbiological methodologies. This implies the recognition of serotype predictions using WGS methods for *Salmonella* and *E. coli*, the characterisation of virulence genes with WGS for STEC and the systematic surveying of antimicrobial resistance genes as part of antimicrobial resistance (AMR) monitoring programmes. To further optimise the development, validation and standardisation of such WGS methods between laboratories, projects should therefore be funded to this end, and capacity building should be undertaken within EU member states. The BfR and FLI assume that this will encompass activities such as the promotion of NGS-based investigations for the molecular epidemiology of pathogens in wild animals, pets and livestock, as well as outbreak investigations in primary production. This serves the interests of animal health, animal welfare and the agricultural sector, as well as consumer health protection—since this limits the transmission of zoonotic pathogens via food.

FLI estimates that NGS has seen only limited use in animal husbandry to date, however. Funds for capacity building are urgently needed in all German states and for federal institutes, and in order to keep pace with international methodological developments. These steps include the procurement of hardware (e.g. sequencing equipment, high-performance computers and fast internet connections), software and the hiring of specialist personnel. By mutual agreement and with the direct involvement of the BfR, FLI and the Robert Koch Institute (RKI), bioinformatic methods are being developed for NGS-based molecular epidemiology and outbreak investigations. A validation of the methodology to ISO 17025:2018 is being pursued in order to ensure deployment for animal health controls.
Answers from the BFR and the FLI to related questions:

1. Article 34 of Regulation (EU) 2017/625 stipulates that the methods used for laboratory analyses must satisfy the requirements stated in annex III of the same regulation. Which of those criteria given in annex III of Regulation (EU) 2017/625 are fulfilled by (which) whole-genome sequencing (WGS) methods?

WGS methods generate maximal coverage for DNA sequences within a genome (e.g. the genome of a bacterial isolate), which can then be used for typing, or for the characterisation of specific properties of the isolate or gene. The criteria that a sequence must fulfil in order to satisfy the quality requirements for subsequent analysis (acceptance parameters) are set out in ISO/CD 23418(E) (table A2). This document is not yet an official ISO standard and is currently in the comment phase. Changes are therefore possible. DNA sequences that satisfy these acceptance parameters can then be used for typing or the characterisation of genetic properties. A rule also applies that the alternative typing method (here: WGS-based typing/characterisation) must be validated. The requirements for the validation of an alternative typing method are set out in the draft ISO/FDIS 16140-6, "Microbiology of the food chain – Method validation – Part 6: Protocol for the validation of alternative (proprietary) methods for microbiological confirmation and typing procedures". The reference method is compared here with the alternative method (determination of inclusivity and exclusivity by means of internal and external method quality testing) and a resulting standard is then defined.

For the determination of the serotype of a *Salmonella* isolate based on WGS data, for example, the following criteria are fulfilled from annex III of Regulation (EU) 2017/625: selectivity, reproducibility and repeatability. Since typing methods in general are not qualitative or quantitative detection methods, the criteria listed in annex III of Regulation (EU) 2017/625 are not generally applicable to alternative typing methods. Also applicable are the validation specifications for test methods required as part of accreditation to DIN EN ISO/IEC 17025 by DAkkS, the German national accreditation body (Document 71 SD 4 019, dated 14 January 2015).

Metagenomic approaches to pathogen detection, and viral pathogens in particular, are not only reproducible and repeatable but also fulfil the criterion stated in annex III(3) of the Regulation, namely: “Methods of analysis which are applicable uniformly to various groups of commodities should be given preference over methods which apply only to individual commodities.” FLI has already demonstrated that the methods established for pathogen detection can be applied to a range of sample matrices from a wide variety of sources. Further research is needed to establish which other requirements from the annex cited are fulfilled, such as sensitivity.

2. Regulations (EC) No 2160/2003 and (EC) No 2073/2005 assume the application of certain methods. Which additional and more detailed results for the evaluation of the criteria described in the regulations cited (for Reg. (EC) No 2160/2003, these are the kinds of information that meet and exceed the criteria for the determination of *Salmonella* serotypes, and which supply additional data (for example about pathogenicity,) could be obtained by the application of whole-genome sequencing (WGS) methods—and which are potentially of importance for risk assessment?
By applying WGS methods, a risk assessment can be prepared for certain types of pathogen on the basis of determining the presence or absence of genes for pathogenicity. One typical example here is the identification of stx genotypes in STEC. While the identification of STEC variants of this kind is already performed using PCR methods, these could be supplanted by WGS methods. In addition, the accessible characteristics (that is, the repertoire of pathogenicity genes in an isolate) can be significantly extended, which allows the creation of considerably more nuanced and detailed datasets for risk assessment than previously possible. This is generally applicable to all bacterial species relevant for food monitoring as well as to other microorganisms (such as pathogens relevant for livestock health). Alongside the identification of pathogenicity genes in the genome of a strain, the following genes can also be of interest for a risk assessment: 1.) Identification of genes that promote the formation of biofilms 2.) Resistance genes, which are responsible for insensitivity to antimicrobial substances and disinfectants 3.) Determinants that play a role in gene mobility (plasmids, transposons, IS elements, phages) 4.) Genes that are responsible for certain metabolic properties of a microbe and which promote survival under stress conditions, for example.

3. Are whole-genome sequencing (WGS) methods already suitable for assessing strain variability in pathogenic microorganisms relevant for public health, such as for *Salmonella*, *Campylobacter* spp. and *Listeria*, for example, and therefore appropriate for supplanting the characteristics of microorganisms typically used to date, such as the serotype, phage type, etc. (see concluding remarks to section 3.4.3 of the EFSA opinion cited)?

Predicting the characteristics of microorganisms, such as the serotype and antimicrobial resistance, on the basis of genome sequences is already essentially possible in laboratories specialising in this field. However, the accuracy of such a prediction depends on the scope and content of the sequence databases and knowledge bases (e.g. known genetic markers for antimicrobial resistance) that must be consulted for investigations of this kind. The scope of these databases varies widely from pathogen to pathogen. In addition, accuracy also depends on the availability of the bioinformatics software and IT hardware required for the identification of these characteristics. In the future, the maintenance of sequence databases and the development and validation of bioinformatics software capable of identifying relevant characteristics will be decisive for establishing the reliability of the results obtained from a genomic analysis. In terms of the characteristics to be identified according to EU legislation (serotypes for *Salmonella* and *E. coli*, antimicrobial resistance), WGS methods are already well-established and can replace traditional microbiological identification.

The development of automated analysis software makes it possible to detect all of the additional genetic fixed characteristics of an isolate. This would include MLST types for all genes, for example, or single nucleotide changes throughout the genome. These characteristics supplement traditional characteristics (such as the serotype) to allow higher-resolution analysis. Thanks to the plethora of additional data acquired by means of WGS methods, newly defined characteristics can be routinely added.

The use of WGS data to predict the phage type (of *Salmonella*, for example) is an approach no longer being pursued, since considerably more nuanced sequence-based approaches for determining phylogenetic properties are now available (e.g. MLST, rMLST, cgMLST, SNP analyses). However, further efforts must be made towards international coordination in terms
of the nomenclature of type designations in these new methods, to enable comparisons between laboratories.

4. BfR Opinion No. 047/2019 ("Comparing the genetic material of pathogens to explain disease outbreaks", 28 November 2019), draws attention to QA in the laboratory and the activities of a working group based on section 64 of the German Food and Feed Code (LFGB). Which laboratory standards for quality assurance are needed in order to verify the results acquired with the same method by different users and to confirm agreement between different laboratories (based in different institutions)?

As a rule, the laboratory standards for quality assurance for a method or a test procedure are defined by the DAkkS (Document 71 SD 4 019, dated 14 January 2015) and codified by means of ISO standards (DIN EN ISO/IEC 17025, DIN EN ISO 15189). This applies both to the food sector and to the field of livestock health. Laboratories must satisfy the requirements described in these documents to be accredited for the test procedure. One should distinguish between standardised methods that are deployed as standards within the procedure’s application scope and non-standardised test methods (e.g. in-house methods). A test procedure must always be validated. The scope of the procedure parameters to be determined depends on whether the procedure in question is of a qualitative or quantitative nature. To verify an already validated procedure (e.g. a standardised procedure) that is established in several laboratories, the laboratory should be subjected to routine proficiency testing by an appropriate body (such as a national reference laboratory, NRL).

The LFGB section 64 working group ‘NGS Bacterial Characterisation’ is developing standardised test procedures (Szabo et al. 2019). The validation work being performed by this group demonstrates that the procedure is capable of reliably obtaining correct results, which are then comparable between the laboratories. Validation encompasses a benchmarking study in one laboratory to determine selectivity (inclusivity and exclusivity), and an interlaboratory benchmarking study that will involve at least eight separate laboratories. Reference methods are incorporated into the benchmarking study where available. Interlaboratory benchmarking studies in WGS are very labour-intensive, since the characteristics to be determined must be investigated comparatively for each target organism. This relates both to the quality assurance parameters for the raw sequences to be generated as well as to the subsequent bioinformatic analysis steps up to the final outcome. Currently, the LFGB section 64 working group ‘NGS Bacterial Characterisation’ is preparing a study that aims to determine whether the methods for detecting clusters in Listeria (L.) monocytogenes are suitable for use in the investigation of disease outbreaks. Since no reference procedure exists to date, criteria for reproducibility (potential deviation from lab to lab) are being analysed, as are measurement uncertainties. Since interlaboratory benchmarking studies need to be carried out for the four food-associated pathogens L. monocytogenes, Salmonella, Campylobacter and STEC, standardised procedures will be prepared successively and will require several years to complete.

Corresponding diagnostic procedures are also being developed and validated in relation to livestock health (section 27 of the German Animal Health Act, TierGesG). Pilot studies involving comparative interlaboratory studies with selected German states and international partners are now underway.
Quality assurance measures also require the preparation and establishment of ‘reference genome’ records. Reference strains must also be defined and made available. Laboratories require these reference strains for the routine verification of their internal WGS methods and in order to validate any changes made to a WGS method.

The validation of WGS methods not only impacts laboratory work but is also especially important in terms of the bioinformatic evaluation of WGS data. As a result of the many different algorithms available for the evaluation of WGS data, comparative studies are also important in this context in order to be able to establish robust standards. The LFGB section 64 working group ‘NGS Bacterial Characterisation’ and FLI are helping to account for this aspect of validation.

In terms of both food monitoring and livestock health, however, it will not be possible to make any statements about whether an isolate is involved in a specific outbreak without also consulting individual epidemiological data on isolates.

BfR and FLI routinely participate in interlaboratory testing organised under the United Nations Secretary General’s Mechanism (UNSGM). This testing is organised by the RKI and involves trials of NGS methods for highly virulent bacterial and viral pathogens (BSL-3).

Further information about next-generation sequencing is available on the BfR website

BfR Opinion No. 047/2019, dated 28 November 2019

Application of Whole Genome Sequencing for the Detection of Foodborne Disease Outbreaks (2020). BfR Wissenschaft. German Federal Institute for Risk Assessment, Berlin

Research project: Establishing Next-Generation Sequencing Ability for Genomic Analysis in Europe (ENGAGE)
https://www.bfr.bund.de/en/new_approaches_in_identifying_and_characterising_microbiological_and_chemical_hazards__engage_-202739.html

External next-generation sequencing links

Global Microbial Identifier initiative to build a DNA genome database for the identification and diagnosis of microbial pathogens (the BfR is a participant in this initiative):
https://www.globalmicrobialidentifier.org/

BfR "Opinions app"
5. References


About the BfR

The German Federal Institute for Risk Assessment (BfR) is a scientifically independent institution within the portfolio of the Federal Ministry of Food and Agriculture (BMEL) in Germany. It advises the German federal government and German federal states ("Laender") on questions of food, chemical and product safety. The BfR conducts its own research on topics that are closely linked to its assessment tasks.

This text version is a translation of the original German text which is the only legally binding version.