STATEMENT

APPROVED: 2 March 2021 PUBLISHED: 10 March 2021

EFSA statement on the requirements for whole genome sequence analysis of microorganisms intentionally used in the food chain

European Food Safety Authority (EFSA)

Abstract

Microorganisms, genetically modified or not, may be used in the food chain as such or as production organisms of substances of interest. The placement of such microorganisms or derived substances/products in the European market may be subject to a pre-market authorization process. The authorization process defines the need to perform a risk assessment to establish the safety and/or the efficacy of the microorganisms when used in the food chain as such or as production strains of substances of interest. In order to perform a risk assessment, the microorganism/s subject to the application for authorization need to be characterized. In this regard, data obtained from whole genome sequence analysis can provide information on the unequivocal taxonomic identification of the strains and on the characterization of their potential functional traits of concern which may include virulence factors, resistance to antimicrobials of clinical relevance for humans and animals, production of known toxic metabolites. In fact, in some areas of the regulated products the use of whole genome sequence-based data has been established as a requirement for the risk assessment. This document provides recommendations to applicants on how to describe the process and results which should be provided to the risk assessor in the context of an application for market authorization of a regulated product. Indications are given on how to perform WGS and the quality criteria/thresholds that should be reached as well as the data and relevant information that need to be sent along whenever such kind of data is required.

Keywords

Microorganisms, food chain, whole genome sequence-based data, intentional use

EFSA Question: EFSA-Q-2019-00434

Correspondence: Feedap@efsa.europa.eu
ISSN: 1831-4732

The EFSA Journal is a publication of the European Food Safety Authority, an agency of the European Union.

Table of Contents

1 Introduction

1.1 Background and Terms of Reference

Microorganisms, genetically modified or not, may be used in the food chain as such or as production organisms for substances of interest. The placement of such microorganisms or derived substances/products in the European market may be subject to a pre-market authorization process, according to the relevant Regulatory framework, including:

- Genetically modified microorganisms for deliberate release into the environment, as covered by EFSA's remit under Directive 2001/18/EC,[1]
- Genetically modified food and feed, Regulation (EC) No 1829/2003 of the European Parliament and of the Council of 22 September 2003 on genetically modified food and feed,[2]
- Feed additives, Regulation (EC) No 1831/2003 of the European Parliament and of the Council of 22 September 2003 on additives for use in animal nutrition,[3]
- Foods for which nutrition or health claims are made, subject to Regulation (EC) No 1924/2006 of the European Parliament and of the Council of 20 December 2006 on nutrition and health claims made on foods,[4]
- Food enzymes, Regulation (EC) No 1332/2008 of the European Parliament and of the Council of 16 December 2008 on food enzymes,[5]
- Food additives, Regulation (EC) No 1333/2008 of the European Parliament and of the Council of 16 December 2008 on food additives,[6]
- Food flavourings, Regulation (EC) No 1334/2008 of the European Parliament and of the Council of 16 December 2008 on food flavourings and certain food ingredients with flavouring properties,[7]
- Microorganisms used as plant protection products, Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October 2009 concerning the placing of plant protection products on the market,[8]

[1] Directive 2001/18/EC of the European Parliament and of the Council of 12 March 2001 on the deliberate release into the environment of genetically modified organisms and repealing Council Directive 90/220/EEC - Commission Declaration. Official Journal L 106, 17/04/2001 p. 1.

[2] Regulation (EC) No 1829/2003 of the European Parliament and of the Council of 22 September 2003 on genetically modified food and feed (Text with EEA relevance). Official Journal L 268, 18/10/2003 p. 1.

[3] Regulation (EC) No 1831/2003 of the European Parliament and of the Council of 22 September 2003 on additives for use in animal nutrition. Official Journal L 268 (Text with EEA relevance), 18/10/2003 p. 29.

[4] Regulation (EC) No 1924/2006 of the European Parliament and of the Council of 20 December 2006 on nutrition and health claims made on foods. Official Journal L 404, 30/12/2006, p. 9.

[5] Regulation (EC) No 1332/2008 of the European Parliament and of the Council of 16 December 2008 on food enzymes and amending Council Directive 83/417/EEC, Council Regulation (EC) No 1493/1999, Directive 2000/13/EC, Council Directive 2001/112/EC and Regulation (EC) No 258/97 (Text with EEA relevance). Official Journal L 354, 31/12/2008, p. 7.

[6] Regulation (EC) No 1333/2008 of the European Parliament and of the Council of 16 December 2008 on food additives (Text with EEA relevance). Official Journal L 354, 31/12/2008, p. 16.

[7] Regulation (EC) No 1334/2008 of the European Parliament and of the Council of 16 December 2008 on flavourings and certain food ingredients with flavouring properties for use in and on foods and amending Council Regulation (EEC) No 1601/91, Regulations (EC) No 2232/96 and (EC) No 110/2008 and Directive 2000/13/EC (Text with EEA relevance). Official Journal L 354, 31/12/2008, p. 34.

[8] Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October 2009 concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC and 91/414/EEC. Official Journal L 309, 24/11/2009, p. 1.

- Novel foods, subject to Regulation (EU) 2015/2283 of the European Parliament and of the Council on novel foods.[9]

The authorization process defines the need to perform a risk assessment to establish the safety and/or the efficacy of the microorganisms when used in the food chain as such or as production strains. Therefore, the microorganisms may need to be characterized and the following documents have been developed to support applicants in the preparation and submission of the data required:

- Guideline developed within the Standing Committee on the Food Chain and Animal Health on the taxonomic level of micro-organisms to be included in Annex I to Directive 91/414/EEC (SANCO/10754/2005 rev.5, 15 April 2005)[10],
- Guidance of the GMO Panel on the risk assessment of genetically modified microorganisms and their products intended for food and feed use (EFSA, 2011),
- Guidance Document for the assessment of the equivalence of technical grade active ingredients for identical microbial strains or isolates approved under Regulation (EC) No 1107/2009 (SANCO/12823/2012-rev. 4, 12 December 2014)[11],
- General Scientific guidance for stakeholders on health claim applications (EFSA NDA Panel, 2016a),
- Guidance on the preparations and presentation of the notification and application for authorization of traditional foods from third countries in the context of Regulation (EU) 2015/2283 (EFSA NDA Panel, 2016b),
- Guidance on the preparation and presentation of an application for authorization of a novel food in the context of Regulation (EU) 2015/2283 (EFSA NDA Panel, 2016c),
- Guidance of the FEEDAP Panel on the characterization of microorganisms used as feed additives or as production organisms (EFSA FEEDAP Panel, 2018),
- Statement on the characterization of microorganisms used for the production of food enzymes (EFSA CEP Panel, 2019),

The FEEDAP Guidance document (EFSA FEEDAP Panel, 2018) and the Statement on the characterization of microorganisms for the production of food enzymes (EFSA CEP Panel, 2019) established whole genome sequence (WGS) and WGS-based data analysis as a requirement for the characterization of bacterial and yeast strains intended for use either as products or as production strains. This approach is also recommended for filamentous fungi. Similarly, the Guidance of the NDA Panel (EFSA NDA Panel, 2016a) offers the possibility to use the WGS data to identify/characterize bacteria and yeasts. In the area of plant protection products, Commission Regulation (EU) No 283/2013[12] also recommends the best available technology to be used to characterize the microorganism. The WGS-based data analysis can provide information to assign taxonomic identification of the strains unequivocally, as well as on the characterization of their potential functional traits of concern (e.g., virulence factors, resistance to antimicrobials of clinical relevance for humans and animals, production of known toxic metabolites).

[9] Regulation (EU) 2015/2283 of the European Parliament and of the Council of 25 November 2015 on novel foods, amending Regulation (EU) No 1169/2011 of the European Parliament and of the Council and repealing Regulation (EC) No 258/97 of the European Parliament and of the Council and Commission Regulation (EC) No 1852/2001 (Text with EEA relevance). Official Journal L 327, 11/12/2015, p.1.

[10] Available online: https://ec.europa.eu/food/sites/food/files/plant/docs/pesticides_aas_guidance_taxonomic_level_dir91414.pdf

[11] Available online: https://ec.europa.eu/food/sites/food/files/plant/docs/pesticides_guidance_equivalence-micro-organisms.pdf

[12] Commission Regulation (EU) No 283/2013 of 1 March 2013 setting out the data requirements for active substances, in accordance with Regulation (EC) No 1107/2009 of the European Parliament and of the Council concerning the placing of plant protection products on the market. OJ L93, 3.4.2013, p.1.

The minimum set of information to submit for the risk assessment for the WGS and WGS-based data analysis are indicated in the above-mentioned documents. Currently, the responsibility of the risk assessor (EFSA and/or Member States) is to critically appraise the work done by the applicants and to evaluate the results. In this regard and in the area of regulated products, it is the applicant´s responsibility to perform the sequencing and the analysis for the microorganism under assessment and the information is reviewed by the risk assessment body. Consequently, the reporting of the work performed by the applicants and the data provided, should allow to perform risk assessment in a scientifically sound and harmonized way, to be able to draw conclusions on the identification and characterization of the microorganisms.

Considering the above, EFSA was requested to prepare a document to support applicants on the preparation and submission of the data based on WGS for the characterization of microorganisms intentionally used in the food chain. The document was subject to public consultation and the outcome is available in Annex A.

2 Scope

The scope of the current document is to provide further advice to applicants on how to describe the analysis and results of WGS-based characterization of microorganisms which should be submitted for assessment in the context of an application including, where relevant, indications on how to perform it and any quality criteria/thresholds that should be provided/reached. This document does not define in which cases WGS-based data are necessary, neither aims at establishing assessment criteria to extract conclusions from the WGS-based analyses. For this, applicants should consult the relevant sectoral Regulatory framework and/or Guidance documents, according to the nature and intended use of the product for which authorization is sought.

3 Data requirements

Applicants should prepare the technical dossiers for the risk assessment in accordance with the requirements set in the applicable Regulatory framework and/or relevant guidance documents. The below requirements apply only to those cases in which WGS-based data analysis needs to be submitted for the risk assessment as established in the applicable Regulatory framework and/or Guidance documents.

This document reflects the current state of the art, which is rapidly evolving both in knowledge and technology. Hence, it may not cover future developments and the document will be subject to recurrent updates, when appropriate. Applicants can choose the technologies/methodologies to be used to generate the data and report them accordingly. The report should describe in detail the work done and the results obtained. Therefore, protocols and methodologies followed, software programs (name, version, and parameters), public databases/references used, as well as the outputs of the analysis should be reported. Alternative approaches to those described below may also be followed, provided they allow a proper characterization.

The microorganisms covered in the document include bacteria, yeasts and filamentous fungi. For applications on other taxonomical groups (e.g., viruses, phages, microalgae…), the same principles would apply on a case-by-case basis.

3.1 Microorganism and DNA Extraction

The microorganism/s tested/under analysis should be the one/s subject to the application for authorization. The samples used for DNA extraction, sequencing, WGS-based data analysis and the results reported should correspond to the microorganism/s under assessment, and this needs to be clearly stated.

Each microorganism should be cultivated before DNA extraction as a pure culture (for fungi, where possible, monosporic cultures should be used). An adequate protocol/method for DNA extraction should be applied. Total DNA (*i.e.* including chromosomal and extra-chromosomal elements), should be extracted and subject to analysis.

3.2 Sequencing and data quality control

3.2.1 Library construction

Library construction method, including DNA fragmentation method and selection of fragments should be reported. Any selection of fragments by size should ensure that small plasmids are not lost. The manufacturer's instructions followed, including version number, and any deviations from that method should be provided.

3.2.2 Sequencing strategy and quality control

The applicant should describe the sequencing strategy, instrumentation used, and any base-calling method applied, where applicable.

For short-read sequencing technologies, it is recommended to trim the sequencing reads to avoid assembly or read mapping artifacts, unless the assembler software discourages it. The trimming and adaptor removal criteria applied, including the software, version and parameters, should be reported. A PHRED score threshold of at least 20 should be set for the quality trimming, and the number of reads and total base pairs of sequence data before and after trimming should be reported.

The average read depth achieved should be at least 30-fold with a recommended target of 100-fold.

Contamination in the sequencing reads should be investigated. Assigned reads to an unexpected organism should be less than 5%. If this is not the case then the applicants should provide an explanation. The tool used, the software version and any parameters used for detection of contamination should be provided and accompanying the results. The database, its version (where available) and/or date of accession needs to be indicated.

The sequencing reads can be *de novo* assembled (and annotated), mapped to a reference genome/database, or the two approaches can be used in combination. For bacteria, complete genome sequence should be pursued but draft genome sequence may be accepted.

3.3 *De novo* assembly and annotation

If a *de novo* assembly-based approach is taken, then the following data are required:

- *De novo* assembly including assembler software, version, and parameters. If post-assembly processing is carried out, approach, software, version and parameters should also be reported.
- Contigs:
    - For draft genomes, the total number of contigs produced by the assembler. For bacteria, total contigs should be <500 and for yeasts and filamentous fungi <1,000. If a higher number of contigs is produced, the applicant should provide a justification,
    - The total length of the contigs. Applicants should provide a justification if their assembly size is not within +/- 20% of the expected genome size for the species,
    - N50 metric.
- For genomes of yeast and filamentous fungi, the number of highly conserved genes such as BUSCO genes present in the assembly should be reported since this parameter indicates the completeness and quality of the assembly (https://busco.ezlab.org/). Ideally, >90% complete matches to BUSCO gene set from the most closely related group of yeasts/filamentous fungi should be present in the assembly.

If a genome annotation is carried out to provide any of the required information, the software name, version and parameters used should be reported. The database(s), version (where available) and/or date of accession should be indicated.

3.4 Reference-based read mapping

There is the possibility to use reference-based read mapping as an alternative to *de novo* assembly-based approach or in combination with it, for the characterization of the microorganism. In this case, the sequencing reads need to be mapped against reference genome(s)/database(s).

3.5 Use of whole genome sequence-based data for the characterization of the microorganism

In the next sections it is indicated what should be detailed in the reports. The applicants may report also any other parameter/information considered to be relevant for strain identification/characterization.

3.5.1 Identification of the microorganism

Confirmation of the identity of the microorganism under assessment should be provided. Each microorganism under assessment should be unambiguously identified, where possible, at the species level.

- For bacteria, the identity of the organism under assessment should preferably be established by digital DNA-DNA hybridization (dDDH; Auch et al., 2010a and b), average nucleotide identity (ANI; Goris et al, 2007) or phylogenomic methods. The data from the microorganism under assessment should be compared with the reference genome of the type strain of the expected species. In case the genome of the type strain is not available, genome sequences of another well-identified strain(s) can be used as the reference. For identification at the species level, dDDH should reach >70% identity and ANI should reach >95 % (Chun et al., 2018).
- For yeasts and filamentous fungi, confirmation of identity should be done by phylogenomic analysis (e.g., using a concatenation of several conserved sequences (e.g., AFToL genes including 18S rDNA/ITS)) to produce a phylogeny against available related genomes) or by alignment to a complete reference genome from the same species.

For *de novo* assembly-based approach, a summary of the method and sequence/s used for comparison, and results of the comparison including sequence identity (percent of identity with the compared reference genome) should be indicated. A phylogenetic tree is recommended, in particular for taxa where a high level of identity between related species exists.

If reference-based read mapping approach is used for identification, sequencing reads should be mapped against a suitable reference genome(s) (e.g., type strain or well-known and well-identified strain(s)). The choice of the reference genome(s) needs to be well considered, justified and reported. The software used should be reported, including version number, and all parameters (if the default parameters are used this should be stated). The proportion of reads mapped, proportion of reference genome covered to at least 5× depth and median depth of mapping across the entire genome should be reported.

3.5.2 Genetic modifications

The characterization of the genetic modifications can be done by comparing the WGS data of the genetically modified microorganism (GMM) with that of the non-genetically modified reference genome (parental or recipient strain). *De novo* assembly or read-mapping strategies can be used. For deletions and small modifications (e.g., regulatory elements) a reference-based read mapping approach can be used, for other genetic modifications a *de novo* assembly approach or a combination of the two strategies may be needed.

The sequences and methodology used for analyses and comparison should be described in detail. Based on the alignment between the GMM and the reference, the actual genetic modification should be characterized. These alignments should be provided. A map or graphic presentation should be provided with all insertions, deletions and substitutions found in the genetically modified strain, including coding and non-coding sequences (e.g., promoters, terminators), together with their description.

Certain applications such as GMMs for deliberate release into the environment, as covered by EFSA's remit under Directive 2001/18/EC,[13] may require data supporting the stability of the genetic modification. In those cases, applicants should provide the data following the relevant Regulatory/Guidance document.

[13] Directive 2001/18/EC of the European Parliament and of the Council of 12 March 2001 on the deliberate release into the environment of genetically modified organisms and repealing Council Directive 90/220/EEC - Commission Declaration. Official Journal L 106, 17/04/2001 p. 1.

3.5.3 Identification of genes of potential concern

The WGS data can be interrogated for the presence of genes of potential concern, which may include those coding for or contributing to resistance to antimicrobials relevant to their use in humans and animals, and others coding for virulence, pathogenicity and/or toxigenicity.

A *de novo* assembled sequence can be analyzed with a search/comparison-based approach against maintained databases and the identified hits should be provided in a table. For each reported result the subject sequence (i.e. the sequence in the database) name and accession number, function of the encoded protein, sequence identity, and the percentage length of the subject sequence which is covered should be provided.

If a reference-based read mapping approach is used, the sequencing reads should be compared to maintained reference database/s. The following statistics should be reported along with the subject sequence name and accession and function of the encoded protein: sequence identity, the average depth of mapping and the percentage length of the subject sequence which is covered by reads. A minimum 5× median depth across the entire sequences should be used as a threshold.

The strategy, software and all relevant parameters (including algorithm if specified within the software) used to identify genes of interest should be reported. The database, version (where available) and/or the date when the database was accessed should be provided.

Antimicrobial resistance

When the search for antimicrobial resistance genes (AMR) is required, it is recommended to conduct it against at least two maintained databases. The search should be done applying the minimum available threshold in the database for the length of coverage. In addition, in case of microorganisms for which no or few AMR genes are present in databases, searches with Hidden-Markov model tools are recommended.

In general, query sequence hits with at least 80% identity (at protein level or nucleotide level as reported in the database) and 70% length of the subject sequence should be reported. In the case two or more fragments covering less than 70% length of the subject sequence with at least 80% identity to the same AMR gene are detected, these should be reported, and it should be checked whether the full gene is present.

Toxigenicity and pathogenicity

Depending on the taxon, the assessment may require the search of genes coding for known virulence factors (e.g., toxins, invasion and adhesion factors) or to identify the presence/absence of known metabolic pathways involved in toxigenicity. The search should be done applying the minimum available threshold in the database for the length of coverage.

In general, query sequence hits with at least 80% identity (at protein level or nucleotide level as provided in the database) and 70% length of the subject sequence should be reported.

In the case two or more fragments covering less than 70% length of the subject sequence with at least 80% identity to the same gene are detected, these should be reported, and it should be checked whether the full gene is present.

3.6 Provision of raw data and standard data formats

The WGS raw data should be submitted in the respective standard formats as indicated below.

- The sequencing reads, and trimmed reads where relevant, should be submitted in FASTQ or equivalent formats, compressed or not (e.g. gzip), pair or single end,
- Assembled sequences can be submitted in FASTA format (e.g.,*.fasta),
- Supported formats for annotation[14] are GFF format (*.gff), GenBank format (*.gbff, *.gbk and *.gb), Tabular format (*.csv) and the NCBI's Sequin ASN.1 (*.sqn),

[14]In case the annotation format includes the nucleotide sequence, data in FASTA format is not required.

- For the characterization of the genetic modification, the alignments should be provided in Sequence Alignment/Map format (SAM), or Binary Alignment/Map format (BAM) (Li et al., 2009) or similar file formats.

The list of relevant data and the information that should be reported along these data can be found in Appendix A.

4 Other relevant documents for reference

Other reference documents published by EFSA:

- Final report of ENGAGE - Establishing Next Generation Sequencing Ability for Genomic analysis in Europe (Hendriksen et al., 2018),
- EFSA Scientific Colloquium 24- 'omics in risk assessment: state of the art and next steps (EFSA, 2018),
- Final report of INNUENDO: A cross-sectoral platform for the integration of genomics in the surveillance of food-borne pathogens (Llarena et al., 2018),
- Technical report to provide technical support in the collection and analysis of whole genome sequencing data in the joint ECDC-EFSA molecular typing database (ECDC and EFSA, 2019),
- Technical specifications on harmonised monitoring of antimicrobial resistance in zoonotic and indicator bacteria from food-producing animals and food (EFSA, 2019),
- Self-tasking whole genome sequencing and metagenomics for outbreak investigation, source attribution and risk assessment of foodborne microorganisms (EFSA BIOHAZ Panel, 2019)

Applicants may also wish to consult the EURL-JRC Guideline for the submission of DNA sequences derived from genetically modified organisms and associated annotations within the framework of Directive 2001/18/EC and Regulation (EC) No 1829/2003, European Union, 2016.[15]

Finally, the GMO Panel published in 2018 a Technical Note on the quality of DNA sequencing for the molecular characterisation of genetically modified plants (EFSA GMO Panel, 2018).

[15] Available on-line at: https://gmo-crl.jrc.ec.europa.eu/doc/Guideline-Sequencing-Feb-2016-mod-April-2017.pdf

References

Auch AF, Von Jan M, Klenk H-P, Göker M et al. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. Standards in Genomic Sciences 2010a; 2: 117-134.

Auch AF, Klenk H-P, Göker M. Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs. Standards in Genomic Sciences 2010b; 2: 142-148.

Chun J, Oren A, Ventosa A, Christensen H, Arahal DR, da Costa MS, Rooney AP, Yi H, Xu XW, De Meyer S, Trujillo ME, 2018. Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. International Journal of Systematic and Evolutionary Microbiology 68, 461-466.

ECDC (European Centre for Disease Prevention and Control), EFSA (European Food Safety Authority), Van Walle I, Guerra B, Borges V, Carriço JA, Cochrane G, Dallman T, Franz E, Karpíšková R, Litrup E, Mistou M-Y, Morabito S, Mossong J, Alm E, Barrucci F, Bianchi C, Costa G, Kotila S, Mangone I, Palm D, Pasinato L, Revez J, Struelens M, Thomas-López D and Rizzi V, 2019. EFSA and ECDC technical report on the collection and analysis of whole genome sequencing data from food-borne pathogens and other relevant microorganisms isolated from human, animal, food, feed and food/feed environmental samples in the joint ECDC–EFSA molecular typing database. EFSA supporting publication 2019:EN-1337. 92 pp. doi:10.2903/sp.efsa.2019.EN-1337

EFSA (European Food Safety Authority), 2011. Scientific Opinion on Guidance on the risk assessment of genetically modified microorganisms and their products intended for food and feed use. EFSA Journal 2011;9(6):2193, 54 pp. doi:10.2903/j.efsa.2011.2193.

EFSA (European Food Safety Authority), Aguilera J., Aguilera-Gomez M., Barrucci F., Cocconcelli P.S., Davies H., Denslow N., Dorne J-L., Grohmann L., Herman L., Hogstrand C., Kass G.E.N., Kille P., Kleter G., Nogué F., Plant N. J., Ramon M, Schoonjans R., Waigmann E., Wright M.C. EFSA Scientific Colloquium 24 – 'omics in risk assessment: state of the art and next steps. EFSA supporting publication 2018:EN-1512. 30 pp. doi:10.2903/sp.efsa.2018.EN-1512

EFSA (European Food Safety Authority), Aerts M, Battisti A, Hendriksen R, Kempf I, Teale C, Tenhagen B-A, Veldman K, Wasyl D, Guerra B, Liébana E, Thomas-López D and Belœil P-A, 2019. Scientific report on the technical specifications on harmonised monitoring of antimicrobial resistance in zoonotic and indicator bacteria from food-producing animals and food. EFSA Journal 2019;17(6):5709, 122 pp.https://doi.org/10.2903/j.efsa.2019.5709

EFSA BIOHAZ Panel (EFSA Panel on Biological Hazards), Koutsoumanis K, Allende A, Alvarez-Ordóñez A, Bolton D, Bover-Cid S, Chemaly M, Davies R, De Cesare A, Hilbert F, Lindqvist R, Nauta M, Peixe L, Ru G, Simmons M, Skandamis P, Suffredini E, Jenkins C, Malorny B, Ribeiro Duarte AS, Torpdahl M, da Silva Felício MT, Guerra B, Rossi M and Herman L, 2019. Scientific Opinion on the whole genome sequencing and metagenomics for outbreak investigation, source attribution and risk assessment of food-borne microorganisms. EFSA Journal 2019;17(12):5898, 78 pp. https://doi.org/10.2903/j.efsa.2019.5898

EFSA CEP Panel (EFSA Panel on Food Contact Materials, Enzymes and Processing Aids), Silano V, Barat Baviera JM, Bolognesi C, Bruschweiler BJ, Cocconcelli PS, Crebelli R,Gott DM, Grob K, Lampi E, Mortensen A, Riviere G, Steffensen I-L, Tlustos C, Van Loveren H, Vernis L, Zorn H, Glandorf B, Herman L, Aguilera J and Chesson A, 2019. Statement on the characterisation of microorganisms used for the production of food enzymes. EFSA Journal 2019;17(6):5741, 13 pp.

EFSA FEEDAP Panel (EFSA Panel on Additives and Products or Substances used in Animal Feed), Rychen G, Aquilina G, Azimonti G, Bampidis V, Bastos ML, Bories G, Chesson A, Cocconcelli PS, Flachowsky G, Gropp J, Kolar B, Kouba M, Lopez-Alonso M, Lopez Puente S, Mantovani A, Mayo B, Ramos F, Saarela M, Villa RE, Wallace RJ, Wester P, Glandorf B, Herman L, Kärenlampi S, Aguilera J, Anguita M, Brozzi R and Galobart J, 2018. Guidance on the characterisation of microorganisms used as feed additives or as production organisms. EFSA Journal 2018;16(3):5206, 24 pp. https:/doi:10.2903/j.efsa.2018.5206

EFSA GMO Panel (EFSA Panel on Genetically Modified Organisms), Casacuberta J, Nogué F, Naegeli H, Birch AN, De Schrijver A, Gralak MA, Guerche P, Manachini B, Messéan A, Nielsen EE, Robaglia C, Rostoks N, Sweet J, Tebbe C, Visioli F, Wal J-M, Moxon S, Schneeberger K, Federici S, Ramon M, Papadopoulou N and Jones H, 2018. Scientific Opinion on the technical Note on the quality of DNA sequencing for the molecular characterisation of genetically modified plants. EFSA Journal 2018;16(7):5345, 11 pp.

EFSA NDA Panel (EFSA Panel on Dietetic Products, Nutrition and Allergies), 2016a. General scientific guidance for stakeholders on health claim applications. EFSA Journal 2016;14(1):4367, 38 pp. https:/doi:10.2903/j.efsa.2016.4367

EFSA NDA Panel (EFSA Panel on Dietetic Products, Nutrition and Allergies), Turck D, Bresson J-L, Burlingame B, Dean T, Fairweather-Tait S, Heinonen M, Hirsch-Ernst KI, Mangelsdorf I, McArdle H, Naska A, Neuhäuser-Berthold M, Nowicka G, Pentieva K, Sanz Y, Siani A, Sjödin A, Stern M, Tomé D, Vinceti M, Willatts P, Engel K-H, Marchelli R, Pöting A, Poulsen M, Schlatter J, Gelbmann W, de Sesmaisons-Lecarré A, Verhagen H and van Loveren H, 2016b. Guidance on the preparation and presentation of the notification and application for authorisation of traditional foods from third countries in the context of Regulation (EU) 2015/2283. EFSA Journal 2016;14(11):4590, 16 pp. doi:10.2903/j.efsa.2016.4590

EFSA NDA Panel (EFSA Panel on Dietetic Products, Nutrition and Allergies), Turck D, Bresson J-L, Burlingame B, Dean T, Fairweather-Tait S, Heinonen M, Hirsch-Ernst KI, Mangelsdorf I, McArdle H, Naska A, Neuhäuser-Berthold M, Nowicka G, Pentieva K, Sanz Y, Siani A, Sjödin A, Stern M, Tomé D, Vinceti M, Willatts P, Engel K-H, Marchelli R, Pöting A, Poulsen M, Salminen S, Schlatter J, Arcella D, Gelbmann W, de Sesmaisons-Lecarré A, Verhagen H and van Loveren H, 2016c. Guidance on the preparation and presentation of an application for authorisation of a novel food in the context of Regulation (EU) 2015/2283. EFSA Journal 2016;14(11):4594, 24 pp. doi:10.2903/j.efsa.2016.4594

Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. International Journal of Systematic and Evolutionary Microbiology. 2007;57(Pt 1):81-91. doi:10.1099/ijs.0.64483-0

Hendriksen R., Karlsmose Pedersen S., Leekitcharoenphon P., Malorny B., Borowiak M., Battisti A., Franco A., Alba P., Carfora V., Ricci A., Mastrorilli E., Losasso C., Longo A., Petrin S., Barco L., Wołkowicz T., Gierczyński R., Zacharczuk K., Wolaniuk N., Wasyl D., Zając M, Wieczorek K., Półtorak K., Petrovska-Holmes L., Davies R., Tang Y., Grant K., Underwood A., Dallman T., Painset A., Hartman H., Al-Shabib A., Cowley L., , 2018. Final report of ENGAGE - Establishing Next Generation sequencing Ability for Genomic analysis in Europe. EFSA supporting publication 2018:EN-1431. 252pp. doi:10.2903/sp.efsa.2018.EN-1431

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup, 2009. The sequence alignment/map format and SAMtools. Bioinformatics, 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Llarena A-K, Ribeiro-Gonçalves BF, Nuno Silva D, Halkilahti J, Machado MP, DaSilva MS, Jaakkonen A, Isidro J, Hämäläinen C, Joenperä J, Borges V, Viera L, Gomes JP, Correia C, Lunden J, Laukkanen-Ninios R, Fredriksson-Ahomaa M, Bikandi J, San Millan R, Martinez-Ballesteros I, Laorden L, Mäesaar M, Grantiņa-Ieviņa L, Hilbert F, Garaizar J, Oleastro M, Nevas M, Salmenlinna S, Hakkinen M, Carriço JA and Rossi M, 2018. INNUENDO: A cross-sectoral platform for the integration ofgenomics in the surveillance of food-borne pathogens. EFSA supporting publication 2018:EN-1498.142 pp. doi:10.2903/sp.efsa.2018.EN-1498

Abbreviations

AMR Antimicrobial resistance

ANI Average Nucleotide Identity

dDDH digital DNA-DNA hybridization

GMM Genetically Modified Microorganism

WGS Whole genome Sequence

Glossary

In the context of this document the following definitions apply:

BUSCO genes: data set of genes comprising genes that within a lineage are near-universally present as single-copy orthologs

Complete genome: A complete genome sequence is a product in which the order and accuracy of every base pair have been verified and the number of contigs equals the number of replicons/chromosomes.

Contamination: reads that do not originate from the expected organism (e.g. presence of reads from organisms other than the expected)

Contigs: Assembly of overlapping sequencing reads that make a contiguous consensus region of DNA

***De novo* assembly:** to join sequencing reads into contigs without a reference sequence

Depth: Number of times that a given nucleotide is read in a reconstructed sequence

Draft genome: a collection of independent contigs with unknown order and orientation.

Reference-based read mapping: Placement of sequencing reads against a reference sequence to assess the coverage or differences to the reference

PHRED score: quality score which indicates the likelihood of a correct base assignment

Appendix A – List of information and data to be provided

The below table lists the information and data that should be submitted by the applicants in the technical dossiers in those applications for which WGS-based data analysis is required according to the relevant Regulatory framework or guidance. This form should be duly completed and signed by the applicants at the time of submission.

Section	Item	Provided		Comments
		Yes	NA	
	Reporting of methodologies and outcomes			
3.1	**Microorganism and DNA extraction**			
	Identifier for the microorganism/s subject of the application for authorization (same used in other sections of the dossier)	☐	☐	
	Confirmation of the correspondence of the samples used for DNA extraction, sequencing, WGS-based data analysis and results reported with the microorganism/s subject of the application	☐	☐	
3.2	**Sequencing and data quality control**			
3.2.1	**Library construction**			
	Library construction method (including the DNA fragmentation method and any selection of fragments)	☐	☐	
3.2.2	**Sequencing strategy and quality control**			
	Sequencing strategy and instrumentation used (base-calling method, where applicable)	☐	☐	
	Trimming and adaptor removal criteria applied, software version and parameters used, PHRED score, number of reads and total base pairs of sequence data before and after trimming	☐	☐	
	Average read depth	☐	☐	
	Contamination in the sequencing data – Percent of reads assigned to unexpected organism/s Tool used, the software version and parameters used, and results; the database used, its version and/or date of accession	☐	☐	
3.3	***De novo* assembly and annotation**			
	Assembler software, version and parameters (including those applied in post-assembly processing)	☐	☐	
	Data related to the contigs: total number of contigs produced by the assembler, total length of contigs, N50	☐	☐	
	Number of highly conserved genes present for yeast and filamentous fungi	☐	☐	
	Annotation software name, version and parameters used, databases used, version and/or date of accession	☐	☐	
3.5	**Use of whole genome sequence-based data for the characterization of the microorganism**			
3.5.1	**Identification of the microorganism from the sequencing data**			
	For *de-novo* assembly approach, method used, sequence/s used for comparison, and the results	☐	☐	
	For read-mapping approach, the reference genome used, the software including version number and parameters used, and the results	☐	☐	
3.5.2	**Genetic modifications**[16]			
	Sequences and methodology used for analyses and comparison	☐	☐	
	Alignments between the GMM and the parental/recipient strain	☐	☐	
	Map or graphic presentation, including all insertions, deletions and substitutions found in the genetically modified strain, coding and non-coding sequences (e.g., promoters, terminators)	☐	☐	
3.5.3	**Identification of genes of potential concern**			
	Strategy, software and parameters used to identify genes of interest and database/s used (including version and/or accession date)	☐	☐	
	De novo assembled sequence and search/comparison-based approach	☐	☐	

[16] Certain applications may require data supporting the stability of the genetic modification.

	For relevant hits: ▪ Subject sequence (including name, accession number and function of the encoded protein) ▪ Sequence identity ▪ Percentage length of the subject sequence covered	☐	☐	
	Reference-based read mapping approach For relevant hits: ▪ Subject sequence (including name, accession number and function of the encoded protein) ▪ Sequence identity ▪ Median depth of mapping ▪ Percentage length of the subject sequence covered	☐ ☐	☐ ☐	
3.6	**Provision of raw data and standard data formats**			
	The sequencing reads, and after trimming where relevant, should be submitted in FASTQ or similar formats, compressed or not, pair or single end	☐	☐	
	Assembled sequences can be submitted in FASTA format (e.g.,*.fa)	☐	☐	
	Annotation[17] should be in GFF format (*.gff), GenBank format (*.gbff, *.gbk and *.gb), Tabular format (*.csv) and the NCBI's Sequin ASN.1 (*.sqn)	☐	☐	
	Alignments to characterize the genetic modification should be provided in Sequence Alignment/Map format (SAM), or Binary Alignment/Map format (BAM) or similar file formats	☐	☐	
Name: Date: Signature:				

[17]In case the annotation format includes the nucleotide sequence, data in FASTA format is not required.

Annex A –Outcome of the public consultation on the statement on the requirements for whole genome sequence analysis of microorganisms intentionally used in the food chain

In line with EFSA's policy on openness and transparency and in order for EFSA to receive comments from the scientific community and stakeholders on its work, EFSA engages in public consultations on key issues. The Statement was endorsed by the Scientific Committee of EFSA for public consultation on the 4th of December 2019.[18] The Public consultation period ended on 28th of February. This Annex summarizes the outcome of the public consultation.

[18] https://www.efsa.europa.eu/en/consultations/call/public-consultation-efsa-statement-requirements-whole-genome