Adopted: 28 June 2024

DOI: 10.2903/j.efsa.2024.8912

efsa JOURNAL

STATEMENT

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

European Food Safety Authority (EFSA)

Correspondence: feedap@efsa.europa.eu

Abstract

Microorganisms, genetically modified or not, may be used in the food chain either as active agents, biomasses or as production organisms of substances of interest. The placement of such microorganisms or their derived substances/products in the European market may be subject to a premarket authorisation process. The authorisation process requires a risk assessment in order to establish the safety and/or the efficacy of the microorganism(s) when used in the food chain as such, as biomasses or as production strains. This includes a full molecular characterisation of the microorganism(s) under assessment. For certain regulated products, the use of whole genome sequence (WGS) data of the microorganism is established as a requirement for the risk assessment. In this regard, data obtained from WGS analysis can provide information on the unambiguous taxonomic identification of the strains, on the presence of genes of concern (e.g. those encoding virulence factors, resistance to antimicrobials of clinical relevance for humans and animals, production of harmful metabolites or of clinically relevant antimicrobials) and on the characterisation of genetic modification(s) (where relevant). This document provides recommendations to applicants on how to describe and report the results of WGS analyses in the context of an application for market authorisation of a regulated product. Indications are given on how to perform genome sequencing and the quality criteria/thresholds that should be reached, as well as the data and relevant information that need to be reported, if required. This updated document replaces the EFSA 2021 Statement and reflects the current knowledge in technologies and methodologies to be used to generate and analyse WGS data for the risk assessment of microorganisms.

KEYWORDS

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

CONTENTS

1 | INTRODUCTION

1.1 | Background and Terms of Reference as provided by EFSA

1.1.1 | Background

Regulation (EC) No 1831/2003 establishes the rules governing the Community authorisation of additives for use in animal nutrition. Moreover, Regulation (EC) No 429/2008 provides detailed rules for the implementation of Regulation (EC) No 1831/2003 as regards the preparation and the presentation of applications and the assessment and the authorisation of feed additives.

The Panel on Additives and Products or Substances used in Animal Feed (FEEDAP Panel) has adopted a series of Guidance documents which aim at complementing Regulation (EC) No 429/2008 to help the applicants in the preparation and submission of technical dossiers for the authorisation of additives for use in animal nutrition according to Regulation (EC) No 1831/2003.

At the plenary meeting in September 2021, the FEEDAP Panel identified the following Guidance documents and statement for revision:

- the Guidance on user safety, considering recent scientific developments and the Panel's experience gained during the last years while working under the provisions of Regulation (EC) No 429/2008,
- the Guidance on the assessment of the efficacy of feed additives, making it complementary to the revised Regulation (EC) No 1831/2003 by stimulating innovation and sustainability in particular for additives that are beneficial for the environment and animal welfare, as outlined in the Green Deal,
- the Guidance on the characterisation of microorganisms used as feed additives or as production organisms, harmonising it with related EFSA Guidance documents, and
- the EFSA Statement on the requirements for whole genome sequence analysis of microorganisms intentionally used in the food chain, keeping track of the fast development in this field.

1.1.2 | Terms of Reference

In view of the above, the European Food Safety Authority (EFSA) asks its FEEDAP Panel to:

1. Analyse for the identified Guidance documents which aspects are most relevant to be updated based on the scientific developments and stakeholder perspective;
2. Update the identified Guidance documents, focusing on the most relevant aspects and taking into account the comments received during public and/or targeted consultations.

1.2 | Regulatory context and guidance

Microorganisms, genetically modified or not, may be used in the food chain either as active agents, as biomasses 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 premarket authorisation 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]

[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, 17.4.2001, 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.

- 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]
- Novel foods, subject to Regulation (EU) 2015/2283 of the European Parliament and of the Council of 25 November 2015 on novel foods,[9]
- Commission Regulation (EU) 2022/1438, amending Annex II to Regulation (EC) No 1107/2009 as regards specific criteria for the approval of active substances that are microorganisms,[10]
- Commission Regulation (EU) 2022/1439, amending Regulation (EU) No 283/2013 as regards the information to be submitted for active substances and the specific data requirements for microorganisms,[11]
- Communication from the Commission concerning Part B of the Annex to Commission Regulation (EU) No 283/2013 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,[12]
- Commission Regulation (EU) 2022/1440, amending Regulation (EU) No 284/2013 as regards the information to be submitted for plant protection products and the specific data requirements for plant protection products containing microorganisms,[13]
- Communication from the Commission concerning Part B of the Annex to Commission Regulation (EU) No 284/2013 setting out the data requirements for plant protection products 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,[14]
- Commission Regulation (EU) 2022/1441, amending Regulation (EU) No 546/2011 as regards specific uniform principles for evaluation and authorisation of plant protection products containing microorganisms.[15]

The authorisation process defines the need to conduct a risk assessment in order to establish the safety and/or the efficacy of the microorganism(s) when used in the food chain as such, as biomasses or as production strains. Therefore, the microorganism/s need/s to be characterised 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),[16]
- Opinion of the Scientific Committee on a request from EFSA on the introduction of a qualified presumption of safety (QPS) approach for assessment of selected microorganisms referred to EFSA (EFSA, 2007),
- Guidance document on the assessment of new isolated of baculovirus species already included in Annex I of Council Directive 91/414/EEC (SANCO/0253/2008 rev. 2, 22 January 2008),[17]

[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.

[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]Commission Regulation (EU) 2022/1438 of 31 August 2022 amending Annex II to Regulation (EC) No 1107/2009 of the European Parliament and of the Council as regards specific criteria for the approval of active substances that are micro-organisms (Text with EEA relevance). OJ L 227, 1.9.2022, p. 2–7.

[11]Commission Regulation (EU) 2022/1439 of 31 August 2022 amending Regulation (EU) No 283/2013 as regards the information to be submitted for active substances and the specific data requirements for micro-organisms (Text with EEA relevance). OJ L 227, 1.9.2022, p. 8–37.

[12]Communication from the Commission concerning Part B of the Annex to Commission Regulation (EU) No 283/2013 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 (Text with EEA relevance), C/2023/3552 OJ C 202/03, 9.6.2023, p. 14–24.

[13]Commission Regulation (EU) 2022/1440 of 31 August 2022 amending Regulation (EU) No 284/2013 as regards the information to be submitted for plant protection products and the specific data requirements for plant protection products containing micro-organisms (Text with EEA relevance). OJ L 227, 1.9.2022, p. 38–69.

[14]Communication from the Commission concerning Part B of the Annex to Commission Regulation (EU) No 284/2013 setting out the data requirements for plant protection products 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 (Text with EEA relevance), C/2023/3548 OJ C 202/02, 9.6.2023, p. 2–13.

[15]Commission Regulation (EU) 2022/1441 of 31 August 2022 amending Regulation (EU) No 546/2011 as regards specific uniform principles for evaluation and authorisation of plant protection products containing micro-organisms (Text with EEA relevance). OJ L 227, 1.9.2022, p. 70–116.

[16]https://food.ec.europa.eu/system/files/2016-10/pesticides_aas_guidance_taxonomic_level_dir91414.pdf.

[17]https://food.ec.europa.eu/system/files/2016-10/pesticides_aas_guidance_baculovirus.pdf.

- Guidance of the EFSA Panel on genetically modified microorganisms (GMO) on the risk assessment of genetically modified microorganisms and their products intended for food and feed use (EFSA, 2011),
- Guidance for submission for food additive evaluations (EFSA ANS Panel, 2012),[18]
- 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),[19]
- General scientific guidance for stakeholders on health claim applications (EFSA NDA Panel, 2016),
- Guidance on the scientific requirements for a notification and application for authorisation of traditional foods from third countries in the context of Regulation (EU) 2015/2283 (EFSA NDA Panel, 2024a),[20]
- Guidance on the scientific requirements for an application for authorisation of a novel food in the context of Regulation (EU) 2015/2283 (EFSA NDA Panel, 2024b),[21]
- Guidance of the EFSA Panel on Additives and Products or Substances used in Animal Feed (FEEDAP) on the characterisation of microorganisms used as feed additives or as production organisms (EFSA FEEDAP Panel, 2018),
- Guidance on the approval and low-risk criteria linked to 'antimicrobial resistance' applicable to microorganisms used for plant protection in accordance with Regulation (EC) No 1107/2009 (SANTE/2020/12260, 23 October 2020),[22]
- Guidance on the risk assessment of metabolites produced by microorganisms used as plant protection active substances in accordance with Article 77 of Regulation (EC) No 1107/2009 (SANCO/2020/12258, 23 October 2020),[23]
- Scientific guidance for the submission of dossiers on food enzymes (EFSA CEP Panel, 2021),
- Explanatory notes for the implementation of the data requirements on microorganisms and plant protection products containing them in the framework of Regulation (EC) No 1107/2009,[24]
- Statement on how to interpret the QPS qualification on 'acquired antimicrobial resistance genes' (EFSA BIOHAZ Panel, 2023).

The FEEDAP Guidance document (EFSA FEEDAP Panel, 2018) and the Scientific guidance for the submission of dossiers on food enzymes (EFSA CEP Panel, 2021) establish whole genome sequence (WGS) and WGS-based data analysis as a requirement for the characterisation of bacterial and yeast strains intended for use either as active agents or as production strains. This approach is also recommended for filamentous fungi. Similarly, the Guidance document of the NDA Panel (EFSA NDA Panel, 2024b) also requires the use of WGS data for the taxonomic and hazard identification of microorganisms. In the area of plant protection active substances, Commission Regulation (EU) No 2022/1439 also requires the use of the latest scientific information to identify/characterise the microorganism under assessment. The WGS-based data analysis can provide information to unequivocally establish the taxonomic identification of the strains, as well as information on the characterisation of their genetic modifications (where relevant) and genes of concern (e.g. those encoding virulence factors, resistance to antimicrobials of clinical relevance for humans and animals, production of harmful metabolites or of clinically relevant antimicrobials).

The minimum set of information to submit for the risk assessment for the WGS and WGS-based data analysis is indicated in the above-mentioned documents. The responsibility of the risk assessor (EFSA and/or Member States) is to critically appraise the information provided in the applications and to derive conclusions. In this regard, and in the area of regulated products, it is the applicant's responsibility to perform the sequencing and the analysis of the microorganism/s 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 conduct the risk assessment in a scientifically sound and harmonised way, and ultimately, to draw conclusions on the identification and characterisation of the microorganism(s).

Considering the above, EFSA was requested to prepare a document to support applicants in the preparation and submission of the data based on WGS for the characterisation of microorganisms intentionally used in the food chain.

2 | SCOPE

The scope of the current document is to provide indications to applicants on how to describe the analysis and results of WGS-based characterisation 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; it neither aims at establishing assessment criteria to draw 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 authorisation is sought.

This document reflects the current state of the art, which is rapidly evolving both in knowledge and technology. Therefore, to ensure that the technologies/methodologies to be used to generate and analyse WGS data are in line with

[18] Endorsed by the EFSA FAF Panel (Panel on Food Additives and Flavourings) on 2 July 2020.

[19] https://food.ec.europa.eu/system/files/2016-10/pesticides_ppp_app-proc_guide_phys-chem-ana_equiv_micro-organisms.pdf.

[20] Under Publication.

[21] Under Publication.

[22] https://food.ec.europa.eu/system/files/2020-11/pesticides_ppp_app-proc_guide_180652_microorganism-amr_202011.pdf.

[23] https://food.ec.europa.eu/system/files/2023-06/pesticides_ppp_app-proc_guide_180653_microorganism-metabolites-concern.pdf.

[24] Available at: https://food.ec.europa.eu/system/files/2023-10/pesticides_ppp_app-proc_guide_imp-data-req_micro-organisms-ppp_imp-reg-11072009.pdf.

up-to-date scientific knowledge, the EFSA 2021 Statement is replaced by this updated document. This document will be subject to recurrent updates when appropriate.

Applicants can choose the technologies/methodologies to be used to generate and analyse the WGS data, and report the work done and the results obtained accordingly. 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 characterisation and risk assessment of the microorganism.

The microorganisms covered in the document include bacteria, yeasts, filamentous fungi and viruses (including bacteriophages). For applications on other taxonomical groups, the same principles will apply on a case-by-case basis. For bacteriophages, WGS data of the bacteriophage itself and of the host bacterial strain in which it is replicated should be generated and used for characterisation purposes as described below.

The current update also aims at clarifying requirements for WGS raw data formats and introducing naming conventions for those files.

2.1 | Microorganism and nucleic acid extraction

The samples used for nucleic acid extraction, sequencing, WGS-based data analysis and the results reported should correspond to the strain(s) under assessment and the subject of the application for authorisation.

Before nucleic acid extraction, each microorganism should be cultivated as a pure culture from the master cell bank (for bacteriophages, the phage preparation should be purified from bacterial DNA and should be free of other viruses). The protocol/method for nucleic acid extraction of the strain under assessment should be described in detail. Genomic material (both chromosomal and extra-chromosomal elements) should be extracted and subjected to analysis.

2.2 | Sequencing and data quality control

Approaches using long-read or a combination of short-read and long-read sequencing technologies and hybrid assembly methods are required for bacterial strains, and for viruses that have a genome of 20 kb or larger. This approach is also strongly recommended for yeasts and filamentous fungi. The integration of short-read and long-read sequencing data sets provides the best results in terms of genome completeness (including extra-chromosomal elements) and reliability of correct genome assembly.

2.2.1 | Library construction

The library construction protocol, including, if applied, methods for nucleic acid fragmentation 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.

2.2.2 | Sequencing strategy and quality control

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

The program, software version and parameters used for the quality control and filtering of the sequencing reads and the corresponding values obtained should be reported. In general, the usual quality thresholds for each sequencing technology should be reached (e.g. a per-base PHRED score of at least 20 for short-reads; an average PHRED score of at least 7 for long-reads).

The average read depth achieved should be at least 30-fold with a recommended target of 100-fold. Sufficient genome coverage should be reached to obtain a high-quality assembly or complete/closed genome as described below (Sections 2.3 and 2.4). If the quality of the assembly is not sufficient, different sequencing strategies might be needed.

Contamination of the sequencing reads should be investigated. Assigned reads to an unexpected organism should be less than 5%. Exceedance of this threshold may be acceptable if properly justified. 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 date of accession needs to be indicated.

The sequencing reads can be de novo or reference-based assembled (and annotated), or the two approaches can be used in combination.

2.3 | De novo assembly and annotation

If a de novo assembly-based approach is taken, then the assembly including assembler software, version and parameters should be provided. If post-assembly processing is carried out, the approach, software, version and parameters should also be reported.

For bacteria and viruses, a complete assembled genome should be provided. The methodology used to confirm the completeness of the genome assembly should also be described.

For yeasts and filamentous fungi, if a complete genome assembly cannot be achieved, the following data should be reported:

- Contigs:
  - The total number of contigs produced by the assembler. The total number of contigs should be < 1000; if a higher number is produced, a justification should be provided,
  - The total length of the contigs and N50 metric. Applicants should provide a justification if their assembly size is not within +/– 20% of the expected genome size for the species.

- 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 date of accession should be indicated.

2.4 | Reference-based assembly

There is the possibility to use reference-based assembly as an alternative to de novo assembly-based approach, or in combination with it, for the characterisation of the microorganism. In this case, the sequencing reads need to be mapped against reference genome(s)/database(s). This approach is not suitable for very divergent strains (e.g. heavily genetically modified ones).

2.5 | Use of whole genome sequence-based data for the characterisation of the microorganism

The next sections describe the information to be reported by applicants when using WGS data for the characterisation of the strain under assessment. The applicants should also report any other parameter/information considered to be relevant for the strain identification/characterisation.

2.5.1 | Identification of the microorganism

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

- For bacteria, the identity of the organism under assessment should preferably be established by digital DNA–DNA hybridisation (dDDH) and/or average nucleotide identity (ANI) (Hugenholtz et al., 2021; Meier-Kolthoff et al., 2014; Parks et al., 2020). The data from the microorganism under assessment should be compared with the genome of the type strain of the expected species and with several genomes of type strains of closely related species. In case the genome of a type strain is not available, publicly available genome sequences of other well-identified strain(s) can be used as a reference. The use of the genome of the parental strain and/or of strains from the same lineage is not acceptable. For identification at the species level, dDDH should usually reach > 70% identity and ANI should usually reach > 94% (Chun et al., 2018; Parks et al., 2022; Riesco & Trujillo, 2024). A phylogenomic analysis is recommended when the ANI or dDDH analysis does not unequivocally assign the strain to a specific species.
- For yeasts and filamentous fungi, identification should be done by phylogenomic analysis (e.g. using a concatenation of several conserved sequences to produce a phylogeny against available related genomes), by alignment to a complete reference genome from the same species or by ANI analysis. For phylogenomic analysis, the analysed genes should be chosen according to the genus considered (e.g. AFToL genes including ITS) and/or other more specific markers (Lücking et al., 2020). For ANI-based identification, a similarity of at least 99% should be reached when comparing the strain under assessment against the type material of the expected species. In case the genome of a type material is not available, publicly available genome sequences of another well-characterised strain(s) may be used as a reference. The number of the annotated orthologous genes used in the analysis and their coverage of the reference genome should be provided.
- For viruses, identification should be done by complete genome analysis and comparison of the sequence against maintained and up-to-date databases.

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.

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 justified and reported. The software used should be reported, including version number, and all parameters (if 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.

2.5.2 | Genetic modifications

The characterisation of the genetic modifications should be done by comparing the WGS data of the genetically modified microorganism (GMM) with that of the non-genetically modified reference strain (parental strain). If a different strategy is followed, a justification should be provided.

The sequences and methodology used for analyses and comparison should be described in detail. Based on the alignment between the GMM and the reference strain, any genetic modifications (i.e. intended and unintended) should be reported. The focus of the unintended modifications is on genes of concern and should be assessed on a case-by-case basis. The alignments between the GMM and the reference strain should be provided. A map or graphic presentation should be provided with all insertions, deletions and substitutions found in the genome (chromosome(s) and extra-chromosomal genetic elements) of the genetically modified strain, including coding and non-coding sequences (e.g. promoters, terminators), together with their description (i.e. function) and location. For each inserted, modified or deleted open reading frame (ORF) the amino acid sequence, function and metabolic role should be provided.

Certain applications for the deliberate release of GMMs into the environment under Directive 2001/18/EC[1] may require data supporting the stability of the genetic modification. In those cases, if applicants use WGS, data format requirements recommended in this guidance document shall be considered.

2.5.3 | Identification of genes and/or genetic elements of concern

The WGS data can be interrogated for the presence of genes of concern, which may include those encoding virulence factors, resistance to antimicrobials of clinical relevance for humans and animals, production of harmful metabolites or of clinically relevant antimicrobials or involved in lysogenic activity and transduction.

A de novo assembled sequence can be analysed 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 length percentage of the subject sequence 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, accession number 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 the 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 (AMR) genes is required, it should be conducted against at least two maintained/curated databases. The search should be done applying the minimum available threshold in the database for the length of coverage.

In general, query sequence hits with a minimum of 80% identity (at the protein or nucleotide level as provided by the database) and 70% length coverage of the subject sequence should be reported. In 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, pathogenicity and antimicrobial production

Depending on the taxon, the assessment may require the search of genes coding for known virulence factors (e.g. toxins, invasion and adhesion factors) and/or to identify the presence of known metabolic pathways involved in toxigenicity or production of clinically relevant antimicrobials. For this purpose, comparison against specific up-to-date databases

targeted to detect the above-mentioned relevant genes should be performed. The search should be done applying the minimum available threshold in the database for the length of coverage.

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

Lifecycle and genetic elements for transduction

For bacteriophages, the absence of lysogenic activity and ability to transduce (mobilise) DNA should be assessed. The assessment requires identification of the presence of genetic determinants known to confer lysogeny and of genes involved in genome packaging and essential for the recognition and cleavage of unit length genome (e.g. Ter genes coding for Terminase enzymes/proteins). Regarding the transducing activity, the presence of genetic elements known to indicate the ability to transduce genes needs to be checked. For this purpose, comparison against specific up-to-date databases should be performed. The search should be done applying the minimum available threshold in the database for the length of coverage.

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

2.6 | Provision of WGS raw and processed data and their standard data formats

The WGS raw and processed data should be submitted in their respective standard formats, following specific naming conventions (where relevant) and using the appropriate file extensions as indicated below. In all cases, the naming convention should be applied to the file name before the corresponding file extension and the use of spaces or special characters (e.g. £;!;|) in the file names should be avoided.

1. The sequencing reads, and trimmed reads where relevant, should be submitted in FASTQ format, compressed (using Gzip) or not, paired or single end. Files should use the standard file extensions corresponding to their format (i.e. *.fastq.gz, *.fq.gz, *.fastq or *.fq) and follow a specific name convention: *Species*_deposit_number_sequence (e.g. *Bacillus_subtilis*_XXX12345_sequence.fastq.gz).
2. Assembled sequences can be submitted in FASTA format, compressed (using Gzip) or not. Files should use the standard file extensions corresponding to their format (i.e.*.fasta, *.fna, *.fa, *.fsa_nt, *.fasta.gz, *.fna.gz, *.fa.gz or *.fsa_nt.gz) and follow a specific name convention: *Species*_deposit_number_assembled_genome (e.g. *Bacillus_subtilis*_XXX12345_assembled_genome.fasta.gz).
3. For genetically modified microorganisms, the assembled sequence of the non-genetically modified reference strain used as comparator in the characterisation of the genetic modification should also be submitted in the formats and extensions described in point 2. The following name convention should be used: *Species*_deposit_number_reference_genome (e.g. *Bacillus_subtilis*_XXX12345_reference_genome.fasta.gz).
4. Supported formats for annotation are GFF format (*.gff), GenBank format (*.gb, *.gbk), EMBL format (*.embl) and the ASN.1 format used by NCBI (*.asn).
5. For the characterisation 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 all relevant data and the information that should be reported along these data can be found in Appendix A.

3 | 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),
- Final report of INNUENDO: A cross-sectoral platform for the integration of genomics in the surveillance of food-borne pathogens (Llarena et al., 2018),
- EFSA Scientific Colloquium 24 – 'omics in risk assessment: state of the art and next steps (EFSA, 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),
- EFSA Scientific Colloquium27: Cell Culture-derived Foods and Food Ingredients (EFSA, 2024).

Applicants may also wish to consult the guideline of the European Union Reference Laboratory for Genetically Modified Food and Feed (EURL GMFF) 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.[25]

Finally, the GMO Panel published in 2024 a Technical Note on the quality of DNA sequencing for the molecular characterisation of genetically modified plants (EFSA, 2024).

GLOSSARY

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.
Reference-based assembly	Mapping of sequencing reads against a reference sequence to obtain a consensus sequence.
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.

ABBREVIATIONS

AMR	Antimicrobial resistance
ANI	Average Nucleotide Identity
dDDH	digital DNA–DNA hybridization
GMM	Genetically Modified Microorganism
ORF	Open reading frame
WGS	Whole genome Sequence

ACKNOWLEDGEMENTS

The Panel wishes to thank the following for the support provided to this scientific output (in alphabetical order of the last name): EFSA experts from the FEEDAP Microbiology Working group, including: Margarita Aguilera, Philip Ashton, Stéphane Brétagne, Henrik Christensen, Pier Sandro Cocconcelli, Christina Cuomo, Boet Glandorf, Lieve Herman, Baltasar Mayo, Carmen Peláez, Miguel Prieto Maradona, Amparo Querol, Maria Saarela, José Serrano Sánchez, Lolke Sijtsma, Laurence Vernis, René Van Der Vlugt, and Andrey Yurkov. EFSA also wishes to acknowledge the contribution of EFSA staff from BIOHAZ, FEEDCO, FIP, iDATA, NIF, PLANTS and PREV Units: Reinhard Ackerl, Jaime Aguilera, Montserrat Anguita, Nicole Bozzi Cionci, Rosella Brozzi, Ana Gomes, Varvara Gouliarmou, Beatriz Guerra, Giovanni Iacono, Frédérique Istace, Dafni Kagkli, Christopher Lythgo, Estefanía Noriega Fernández, Nikoletta Papadopoulou, Elisa Pettenati, Adrián César Razquin, Mirko Rossi and Reinhilde Schoonjans.

CONFLICT OF INTEREST

If you wish to access the declaration of interests of any expert contributing to an EFSA scientific assessment, please contact interestmanagement@efsa.europa.eu.

REQUESTOR

European Food Safety Authority

QUESTION NUMBER

EFSA-Q-2022-00250

[25] https://gmo-crl.jrc.ec.europa.eu/doc/Guideline-Sequencing-Feb-2016-mod-April-2017.pdf.

REFERENCES

Chun, J., Oren, A., Ventosa, A., Christensen, H., Arahal, D. R., da Costa, M. S., Rooney, A. P., Yi, H., Xu, X. W., De Meyer, S., & Trujillo, M. E. (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 and EFSA (European Centre for Disease Prevention and Control and European Food Safety Authority), Van Walle, I., Guerra, B., Borges, V., Carriço, J. A., 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., … 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*, EN-1337. https://doi.org/10.2903/sp.efsa.2019.EN-1337

EFSA (European Food Safety Authority). (2007). Opinion of the scientific committee on a request from EFSA on the introduction of a qualified presumption of safety (QPS) approach for assessment of selected microorganisms referred to EFSA. *EFSA Journal*, *5*(12), 587. https://doi.org/10.2903/j.efsa.2007.587

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*, *9*(6), 2193. https://doi.org/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. (2018). EFSA Scientific colloquium 24 – 'Omics in risk assessment: State of the art and next steps. *EFSA Supporting Publication*, EN-1512. https://doi.org/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., & 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*, *17*(6), 5709. https://doi.org/10.2903/j.efsa.2019.5709

EFSA (European Food Safety Authority). (2024). EFSA Scientific Colloquium27: Cell culture-derived foods and food ingredients. *EFSA Supporting Publication*, EN-8664. https://doi.org/10.2903/sp.efsa.2024.EN-8664

EFSA (European Food Safety Authority), César-Razquin, A., Casacuberta, J., Dalmay, T., Federici, S., Jacchia, S., Kagkli, D. M., Moxon, S., & Papadopoulou, N. (2024). Technical Note on the quality of DNA sequencing for the molecular characterisation of genetically modified plants. *EFSA Journal*, *22*(4), e8744. https://doi.org/10.2903/j.efsa.2024.8744

EFSA ANS Panel (EFSA Panel on Food Additives and Nutrient Sources added to Food). (2012). Guidance for submission for food additive evaluations. *EFSA Journal*, *10*(7), 2760. https://doi.org/10.2903/j.efsa.2012.2760

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., … 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*, *17*(12), 5898. https://doi.org/10.2903/j.efsa.2019.5898

EFSA BIOHAZ Panel (EFSA Panel on Biological Hazards), Koutsoumanis, K., Allende, A., Alvarez-Ordoñez, A., Bolton, D., Bover-Cid, S., Chemaly, M., De Cesare, A., Hilbert, F., Lindqvist, R., Nauta, M., Nonno, R., Peixe, L., Ru, G., Simmons, M., Skandamis, P., Suffredini, E., Cocconcelli, P. S., Suarez, J. E., … Herman, L. (2023). Statement on how to interpret the QPS qualification on 'acquired antimicrobial resistance genes'. *EFSA Journal*, *21*(10), 8323. https://doi.org/10.2903/j.efsa.2023.8323

EFSA CEP Panel (EFSA Panel on Food Contact Materials, Enzymes and Processing Aids), Lambré, C., Barat Baviera, J. M., Bolognesi, C., Cocconcelli, P. S., Crebelli, R., Gott, D. M., Grob, K., Lampi, E., Mengelers, M., Mortensen, A., Rivière, G., Steffensen, I.-L., Tlustos, C., Van Loveren, H., Vernis, L., Zorn, H., Glandorf, B., Herman, L., … Chesson, A. (2021). Scientific Guidance for the submission of dossiers on food enzymes. *EFSA Journal*, *19*(10), 6851. https://doi.org/10.2903/j.efsa.2021.6851

EFSA FEEDAP Panel (EFSA Panel on Additives and Products or Substances used in Animal Feed), Rychen, G., Aquilina, G., Azimonti, G., Bampidis, V., Bastos, M. L., Bories, G., Chesson, A., Cocconcelli, P. S., Flachowsky, G., Gropp, J., Kolar, B., Kouba, M., Lopez-Alonso, M., Lopez Puente, S., Mantovani, A., Mayo, B., Ramos, F., Saarela, M., … Galobart, J. (2018). Guidance on the characterisation of microorganisms used as feed additives or as production organisms. *EFSA Journal*, *16*(3), 5206. https://doi.org/10.2903/j.efsa.2018.5206

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

Hendriksen, R., Karlsmose, P. 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., … Cowley, L. (2018). Final report of ENGAGE - establishing next generation sequencing ability for genomic analysis in Europe. *EFSA Supporting Publication*, EN-1431. https://doi.org/10.2903/sp.efsa.2018.EN-1431

Hugenholtz, P., Chuvochina, M., Oren, A., Parks, D. H., & Soo, R. M. (2021). Prokaryotic taxonomy and nomenclature in the age of big sequence data. *The ISME Journal*, *15*(7), 1879–1892. https://doi.org/10.1038/s41396-021-00941-x

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, B. F., Nuno Silva, D., Halkilahti, J., Machado, M. P., DaSilva, M. S., Jaakkonen, A., Isidro, J., Hämäläinen, C., Joenperä, J., Borges, V., Viera, L., Gomes, J. P., Correia, C., Lunden, J., Laukkanen-Ninios, R., Fredriksson-Ahomaa, M., Bikandi, J., San Millan, R., … Rossi, M. (2018). INNUENDO: A cross-sectoral platform for the integration ofgenomics in the surveillance of food-borne pathogens. *EFSA Supporting Publication*, EN-1498. https://doi.org/10.2903/sp.efsa.2018.EN-1498

Lücking, R., Aime, M. C., Robbertse, B., Miller, A. N., Ariyawansa, H. A., Aoki, T., Cardinali, G., Crous, P. W., Druzhinina, I. S., Geiser, D. M., Hawksworth, D. L., Hyde, K. D., Irinyi, L., Jeewon, R., Johnston, P. R., Kirk, P. M., Malosso, E., May, T. W., Meyer, W., … Schoch, C. L. (2020). Unambiguous identification of fungi: Where do we stand and how accurate and precise is fungal DNA barcoding? *IMA Fungus*, *11*, 14. https://doi.org/10.1186/s43008-020-00033-z

Meier-Kolthoff, J. P., Klenk, H.-P., & Göker, M. (2014). Taxonomic use of DNA G+C content and DNA–DNA hybridization in the genomic age. *International Journal of Systematic and Evolutionary Microbiology*, *64*, 352–356.

Parks, D. H., Chuvochina, M., Chaumeil, P. A., Rinke, C., Mussig, A. J., & Hugenholtz, P. (2020). A complete domain-to-species taxonomy for bacteria and archaea. *Nature Biotechnology*, *38*(9), 1079–1086. https://doi.org/10.1038/s41587-020-0501-8

Parks, D. H., Chuvochina, M., Rinke, C., Mussig, A. J., Chaumeil, P. A., & Hugenholtz, P. (2022). GTDB: An ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Research*, *50*(D1), D785–D794. https://doi.org/10.1093/nar/gkab776

Riesco, R., & Trujillo, M. E. (2024). Update on the proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *International Journal of Systematic and Evolutionary Microbiology*, *74*(3), 006300. https://doi.org/10.1099/ijsem.0.006300

How to cite this article: EFSA (European Food Safety Authority), (2024). EFSA statement on the requirements for whole genome sequence analysis of microorganisms intentionally used in the food chain. *EFSA Journal*, *22*(8), e8912. https://doi.org/10.2903/j.efsa.2024.8912

APPENDIX A

List of information and data to be provided

The below table lists the information and data that should be submitted to EFSA 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			
2.1	**Microorganism and nucleic acid extraction**			
	Identifier for the microorganism/s subject of the application for authorisation (same used in other sections of the dossier)	☐	☐	
	Confirmation of the correspondence of the samples used for nucleic acid extraction, sequencing, WGS-based data analysis and results reported with the microorganism/s subject of the application	☐	☐	
2.2	**Sequencing and data quality control**			
2.2.1	**Library construction**			
	Library construction method (including the nucleic acid fragmentation method and any selection of fragments)	☐	☐	
2.2.2	**Sequencing strategy and quality control**			
	Sequencing strategy and instrumentation used (base-calling method, where applicable)	☐	☐	
	Trimming (where applicable), filtering, software version and parameters used, quality thresholds	☐	☐	
	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	☐	☐	
2.3	**De novo assembly and annotation**			
	Assembler software, version and parameters (including those applied in post-assembly processing)	☐	☐	
	Data related to the contigs	☐	☐	
	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	☐	☐	
2.4	**Reference-based assembly**			
	Reference genome(s)/database(s)	☐	☐	
2.5	**Use of whole genome sequence-based data for the characterisation of the microorganism**			
2.5.1	**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	☐	☐	
2.5.2	**Genetic modifications**[26]			
	Sequences and methodology used for analyses and comparison	☐	☐	
	Alignments between the GMM and the parental 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)	☐	☐	
	Amino acid sequence, function and metabolic role of each inserted, modified or deleted open reading frame (ORF)	☐	☐	

(Continues)

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

(Continued)

Section	Item	Provided		Comments
		Yes	NA	
2.5.3	**Identification of genes of 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	☐	☐	
	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 WGS raw and processed data and their standard data formats**			
	The sequencing reads, and after trimming where relevant, should be submitted in FASTQ formats, compressed or not, paired or single end with the corresponding file extension and name convention	☐	☐	
	Assembled sequences can be submitted in FASTA format with the corresponding file extension and name convention[27]	☐	☐	
	Annotation should be in GFF format (*.gff), GenBank format (*.gb, *.gbk), EMBL format (*.embl) and the ASN.1 format used by NCBI (*.asn)	☐	☐	
	Alignments to characterise 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:				

[27]For GMMs, the same information should be also submitted for the non-genetically modified reference strain used as comparator in the characterisation of the genetic modification with the corresponding file extension and name convention.