# SLOVENSKI STANDARD
# oSIST prEN ISO 23418:2020

**01-november-2020**

---

**Mikrobiologija v prehranski verigi - Sekvenciranje celotnega genoma za tipizacijo in genomsko karakterizacijo bakterij v živilih - Splošne zahteve in navodilo (ISO/DIS 23418:2020)**

Microbiology of the food chain - Whole genome sequencing for typing and genomic characterization of foodborne bacteria - General requirements and guidance (ISO/DIS 23418:2020)



Mikrobiologie der Lebensmittelkette - Vollständige Genomsequenzierung zur Typisierung und genomischen Charakterisierung von Bakterien in Lebensmitteln - Allgemeine Anforderungen und Leitfaden (ISO/DIS 23418:2020)



Microbiologie de la chaîne alimentaire - Séquençage de génome complet pour le typage et la caractérisation génomique des bactéries dans les aliments - Exigences générales et recommandations (ISO/DIS 23418:2020)

**Ta slovenski standard je istoveten z:** **prEN ISO 23418**

---

**ICS:**

| | | |
|---|---|---|
| 07.100.30 | Mikrobiologija živil | Food microbiology |

**oSIST prEN ISO 23418:2020** **en,fr,de**

DRAFT INTERNATIONAL STANDARD

# ISO/DIS 23418

ISO/TC **34**/SC **9**

Secretariat: **AFNOR**

Voting begins on:
**2020-09-18**

Voting terminates on:
**2020-12-11**

# Microbiology of the food chain — Whole genome sequencing for typing and genomic characterization of foodborne bacteria — General requirements and guidance

*Microbiologie de la chaîne alimentaire — Séquençage de génome complet pour le typage et la caractérisation génomique des bactéries dans les aliments — Exigences générales et recommandations*

ICS: 07.100.30





This document is circulated as received from the committee secretariat.

**ISO/CEN PARALLEL PROCESSING**

Reference number
ISO/DIS 23418:2020(E)



ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

# Contents

Page

# Foreword

ISO (the International Organizations for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.



This document was prepared by Technical Committee ISO/TC 34, *Food Products*, Subcommittee SC 9, *Microbiology.*



Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

# Introduction

Next generation sequencing (NGS) provides rapid, economical and high-throughput access to microbial whole genome sequences (WGS) and is being applied to an expanding number of problems in food microbiology. WGS are digital representations of the biological potential of the sequenced organism at single base resolution. The digital nature of WGS data is a departure from the continuous nature of phenotypes routinely analyzed in food microbiology. Therefore, WGS offers significant advantages over existing technologies (e.g., serology, pulsed field gel electrophoresis, antibiotic resistance phenotype). WGS-based analyses are used by public health laboratories to detect outbreaks, and to detect mutations, genes and other genetic features to characterize virulence and survival potential. Within the food industry, there is interest in WGS to characterize bacterial isolates from outsourced ingredients and environmental surfaces, to better understand their origin and ecology, and to update procedures to reduce risk. Some companies have developed, or are developing, the capacity to collect and analyze WGS data. Others will turn to third party laboratories to perform these services, as they currently do for other microbiological analyses.

Removed text

This standard is intended to provide guidance for both the laboratory and bioinformatic components of WGS and associated metadata for foodborne microorganisms. This standard is intended to be applicable to all currently available short- and long-read DNA sequencing technologies. It may be applied to analysis of WGS data with proprietary, open-source, and custom software. It is not intended to specify sequencing chemistries, analytical methods, or software. The standard defines laboratory, data, and metadata stewardship practices to ensure that analyses are clearly reported, transparent, open to inquiry, and available for unanticipated uses. This standard is for use by laboratories to develop their management systems for quality and technical operations. Laboratory customers and regulatory authorities may also use it in confirmation or recognizing the competence of laboratories.

# Microbiology of the food chain — Whole genome sequencing for typing and genomic characterization of foodborne bacteria — General requirements and guidance

## 1 Scope

This international standard specifies minimum requirements for generating and analyzing whole-genome sequencing (WGS) data obtained from foodborne bacteria. These requirements are applicable to any sequencing platform or chemistry. This process may include the following stages:

a) Handling of bacterial cultures;

b) Genomic DNA isolation;

c) Library preparation, sequencing, and assessment of raw DNA sequence read quality and storage;

d) Bioinformatics analysis for determining genetic relatedness, genetic content and predicting phenotype, and bioinformatics pipeline validation;

e) Metadata capture and sequence repository deposition; and

f) Validation of the end-to-end WGS workflow (fit for purpose for intended application).



## 2 Normative references



There are no normative references in this document.

## 3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at http://www.electropedia.org/

**3.1**
**adapter sequence**
DNA with a known sequence, which is added to the end of a DNA library fragment, to facilitate the sequencing process (e.g., annealing to a flow cell)

**3.2**
**annotation**
process of identifying genes and other features on genome assemblies

**3.3**
**antibiogram**
summary of antimicrobial susceptibility testing results performed for a specific microorganism, usually represented in tabular form

**3.4**
**assembly**
output from process of aligning and merging sequencing reads into larger contiguous sequences (contigs)

**3.5**
**base calling**
process of assigning nucleotides and quality scores to positions in sequencing reads

**3.6**
**bioinformatics**
collection, storage, and analysis of biological sequence data

**3.7**
**bioinformatics pipeline**
individual programs, scripts, or pieces of software linked together, where output from one program is used as input for the next step in data processing

**3.8**
**carryover-contamination**
samples contaminated with DNA from previously sequenced samples, or substances, including EDTA, phenol-chloroform, protein, excess salts

**3.9**
**Chemical Entities of Biological Interest Ontology**
**ChEBI**
ontology for describing small chemical compounds

**3.10**
**contig**
contiguous stretch of DNA sequence that results from the assembly of smaller, overlapping DNA sequence reads



**3.11**
**controlled vocabulary**
finite set of values that represent the only allowed values for a data item



[SOURCE: ISO 11238:2018(en)]

**3.12**
**coverage**
average number of times each base in a genome is sequenced

**3.13**
**cross-contamination**
contamination of a sample (bacterial isolate or DNA) with other samples

**3.14**
**DNA quality**
indication of DNA purity (free of polysaccharides, contaminants and enzyme inhibitors) and integrity (high molecular weight with little to no evidence of degradation)

**3.15**
**DNA Sample**
portion of DNA extracted from some material

**3.16**
**draft assembly**
de novo genome assembly consisting of contigs with no implied order, typically generated using whole-genome shotgun sequencing with a short-read technology

**3.17**
**Environment Ontology**
**EnvO**
ontology for describing environmental features and habitats

**3.18**
**FoodEx2 Ontology**
**FoodEx2**
standardised food classification and description system developed by the European Food Safety Authority (EFSA)

**3.19**
**Food Ontology**
**FoodOn**
ontology for describing food products, animal feed and food processing

**3.20**
**Gazetteer Ontology**
**GAZ**
ontology for describing geographical locations

**3.21**
**index**
oligonucleotide sequences used in the process of library preparation to tag or barcode DNA from specific samples, so that multiple samples may be combined (multiplexed) in a sequencing reaction

**3.22**
**International Nucleotide Sequence Database Collaboration**
**INSDC**
initiative operated by the DNA Database of Japan (DDBJ), the European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI) and the National Center for Biotechnology Information (NCBI)



**3.23**
**ISO WGS Slim**
ontology Slim containing interoperable fields and terms pertaining to the use of WGS for food microbiology



**3.24**
**isolate**
population of bacterial cells in pure culture derived from a single colony

**3.25**
**kmers**
all possible sequences of length k that are contained in a whole genome sequence

**3.26**
**library**
collection of genomic DNA fragments from a single isolate intended for determining genome sequence

**3.27**
**management system**
quality, administrative and technical systems that govern the operations of an organization

Note 1 to entry: For the purposes of this document organization refers to the laboratory

**3.28**
**mapping**
use of software to align sequencing reads to reference sequences

**3.29**
**metadata**
data that describes and defines other data

[SOURCE: ISO/IEC 11179-1:2015, 3.2.16]

**3.30**
**minimal data for matching**
**MDM**
information required to describe the sample source and provenance of a genomic sequence, as defined by the Global Microbial Identifier[10], and implemented by the International Nucleotide Sequence Database Collaboration

**3.31**
**minimum inhibitory concentration**
**MIC**
lowest concentration that, under defined *in vitro* test conditions, reduces growth by an agreed amount within a defined period of time.

Note 1 to entry: to entry The MIC is expressed in mg/l.

[SOURCE: ISO 16256:2012(en)]

**3.32**
**multi-locus sequence typing**
**MLST**
method of genomic analysis in which nucleotide variants within predefined sets of loci, either core genome loci for cgMLST or whole genome loci for wgMLST, are identified

**3.33**
**N50**
length (N) such that sequence contigs of N or longer include half the bases in the assembly



**3.34**
**NCBITaxon Ontology**
**NCBITaxon**
automatic translation of the NCBI taxonomy database



**3.35**
**NG50**
length (N) of DNA such that sequence contigs of N or longer include half the bases in the genome

**3.36**
**Open Biological and Biomedical Ontology Foundry**
**OBO Foundry**
collection of ontologies created by a collective of ontology developers that are committed to collaboration and adherence to shared principles

**3.37**
**ontology**
controlled vocabulary arranged in a hierarchy, where the terms are connected by logical relationships

**3.38**
**ontology Slim**
set of ontology fields and terms annotated as part of a particular collection, often for a specific purpose, which can be extracted to create a file distinct from the original ontology

**3.39**
**Phred (Q) sequence quality score**
measure of the probability that a base is incorrectly assigned at a given position in the sequence expressed as:

$$Q = -10 \log_{10} P$$

Note 1 to entry: to entry A score of Q30 indicates that there is a 1 in 1000 chance that a base is incorrectly assigned (i.e. the base call is 99.9 % accurate)

**3.40**
**read**
Nucleotide sequence inferred from a fragment of DNA or RNA

**3.41**
**sequence repository**
database in which WGS datasets are stored and managed

Note 1 to entry: to entry A public repository allows unrestricted access to the data, while a private or federated repository restricts access to the data

**3.42**
**sequencing replicate, biological**
sequencing a different colony from the same isolate obtained from the same sample material, to assess biological variation

**3.43**
**sequencing replicate, technical**
resequencing of the same biological sample or library to assess sequence variation due to instrumentation and protocol

**3.44**
**serotype**
classification scheme based on the antigenic detection or sequence-based detection of genes encoding bacteria surface molecules



**3.45**
**Single Nucleotide Polymorphism**
**SNP**
a SNV that passes a particular quality and/or frequency threshold



**3.46**
**Single Nucleotide Variant**
**SNV**
differences between the nucleotide states at the same genomic position of two or more isolates

**3.47**
**strain**
the descendants of a single isolation in pure culture, usually derived from a single initial colony on a solid growth medium[1]

Note 1 to entry: to entry A strain may be considered an isolate or group of isolates that can be distinguished from other isolates of the same genus and species by phenotypic and genotypic characteristics

**3.48**
**validation**
establishment of the performance characteristics of a method and provision of objective evidence that the performance requirements for a specified intended use are fulfilled

[SOURCE: ISO 16140-1:2016(en)]

**3.49**
**validated data entry**
automated process ensuring that data entered into a repository is correct

**3.50**
**verification**
demonstration that a validated method functions in the user's hands according to the method's specifications determined in the *validation* (3.48) study and is fit for its purpose

[SOURCE: ISO 16140-1:2016(en)]

**3.51**
**whole genome sequencing**
**WGS**
process of determining the DNA sequence of an organism's genome using total genomic DNA as input

# 4 Principle

## 4.1 General

Any organization that handles samples, performs sequencing, or performs bioinformatics analyses for WGS analysis shall demonstrate, through provision of evidence, that proper documentation of sample provenance, methods and quality control is collected and maintained for follow-up.

WGS analysis of foodborne bacteria consist of bacterial culture, DNA isolation performed in a microbiological laboratory, sequencing steps conducted at a sequencing facility, and bioinformatics analysis performed in a distinct computational environment.

## 4.2 Laboratory operation: sample preparation and sequencing

Sample preparation and sequencing should include the following steps:

a) Information about the isolates being sequenced, including barcodes for multiplexed samples, is entered into the appropriate record systems, such as a laboratory information management system (LIMS) and/or sample description worksheets.



b) Genomic DNA is extracted from pure cultures and ideally the species identity is confirmed.

c) DNA libraries are prepared from the genomic DNA extraction. This process should include:



   i. DNA fragmentation;

   ii. ligation of indices and adapters;

   iii. quantification, normalization, and quality control of the resulting library and

   iv. pooling of libraries for multiplexed sequencing runs.

d) The libraries are sequenced

e) Quality metrics produced by the sequencing instrument are recorded for each run.

## 4.3 Bioinformatics analysis

Pipelines for bioinformatics analysis may focus on *in silico* predictions of phenotype (e.g. virulence) or detecting clusters of genetically similar isolates (i.e. same strain, sequence type, or serotype). Pipelines based on comparative approaches can be used to detect the presence and states of markers in raw and assembled sequencing data to make *in silico* strain (e.g., sequence type) and phenotype predictions.

Sequence data for multiple isolates can be analyzed using SNP, MLST or kmer distance analysis methods to identify clusters of closely related bacteria. Results from these analyses can be used to infer relationships between isolates which may be illustrated with phylogenetic trees and dendrograms.

a) **SNP Analyses**

For SNP analyses, reads are mapped to a reference sequence or reads are assembled into contigs that are compared. To determine SNPs, SNVs are quality-filtered to identify SNP positions.

b) **MLST Analyses**

 

For MLST analyses, reads are assembled or mapped. Target loci are identified, quality-filtered, and compared to a curated cgMLST or wgMLST database.

c) **Kmer distance analysis**

Sequence data for multiple isolates can be analyzed using kmer distance methods to identify clusters of related bacteria. Kmer analyses have the advantage of being very fast but have some limitations notably in terms of precision. (i.e., they are applicable in species determination, but not recommended for detailed source tracking analysis of closely related strains).

### 4.4 Metadata formats and sequence repository deposition

Metadata records shall be created and safely stored for all sequences. Sequence data and corresponding metadata should be consistently formatted and documented. These metadata can be shared solely at the discretion of the metadata owner. Data and its corresponding metadata shall be subject to security considerations, cost and benefits, legal liability, intellectual property rights, confidential business information, contract restriction or other binding written agreements.

To promote data stewardship best practices[3], this standard provides optional metadata reporting formats which are harmonized to a community data standard (e.g., MDM or OBO Foundry ontologies). These formats and standards facilitate reproducibility and common understanding of terminology. An ISO WGS Slim was created to format and provide values for the recommended metadata fields. WGS and selected metadata can be transferred (uploaded) to a publicly accessible database.

### 4.5 Validation and verification of WGS Workflow



The entire WGS workflow shall be validated to provide assurance that the methods are fit for intended use.



## 5 General laboratory guidance

### 5.1 Bacterial isolation and DNA extraction

Bacterial isolation and DNA extraction should be performed in a general microbiological laboratory adapted to work with the specific bacteria, including pathogens. For sequencing library preparation that involves DNA amplification using polymerase chain reaction (PCR), pre- and post-PCR steps should be carried out in different or segregated areas of the laboratory to avoid carryover-contamination.

### 5.2 Laboratory environment

Air movements, vibration, temperature and humidity can interfere with the performance of many sequencers and should be considered in the placement of the equipment in the laboratory. Laboratories should consult the sequencer manufacturer's site preparation guide for specific guidance.

### 5.3 Standard Operating Procedures (SOPs) and non-conforming work

Laboratories should maintain and adhere to standardized operating procedures (SOPs), workflow documents, reagent inventory controls, and equipment maintenance logs. SOPs should include procedures for using positive and negative controls for the DNA extraction, sequence library preparation and sequencing steps. SOPs should include procedures for monitoring operations for run quality and errors (sample misidentification or cross-contamination).

In the case of sample misidentification or contamination the root cause of errors in sequencing shall be investigated;

i. ensuring that runs containing misidentified samples, or samples contaminated with multiple strains, are not used for bioinformatics analysis or uploaded to databases; and