BS EN ISO 23418:2022

BSI Standards Publication

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

Documents
Document Sharing Hub

Full document details: BS EN ISO 23418:2022

National foreword

This British Standard is the UK implementation of EN ISO 23418:2022. It is identical to ISO 23418:2022.

The UK participation in its preparation was entrusted to Technical Committee AW/9, Microbiology.

A list of organizations represented on this committee can be obtained on request to its committee manager.

Contractual and legal considerations

This publication has been prepared in good faith, however no representation, warranty, assurance or undertaking (express or implied) is or will be made, and no responsibility or liability is or will be accepted by BSI in relation to the adequacy, accuracy, completeness or reasonableness of this publication. All and any such responsibility and liability is expressly disclaimed to the full extent permitted by the law.

This publication is provided as is, and is to be used at the recipient's own risk.

The recipient is advised to consider seeking professional guidance with respect to its use of this publication.

This publication is not intended to constitute a contract. Users are responsible for its correct application.

© The British Standards Institution 2022
Published by BSI Standards Limited 2022

ISBN 978 0 539 05079 0

ICS 07.100.30

Compliance with a British Standard cannot confer immunity from legal obligations.

This British Standard was published under the authority of the Standards Policy and Strategy Committee on 31 July 2022.

Amendments/corrigenda issued since publication

Date	Text affected

Documents
Document Sharing Hub

Full document details: BS EN ISO 23418:2022

EUROPEAN STANDARD

NORME EUROPÉENNE

EUROPÄISCHE NORM

EN ISO 23418

June 2022

ICS 07.100.30

English Version

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

Microbiologie de la chaîne alimentaire - Séquençage de génome entier pour le typage et la caractérisation génomique des bactéries - Exigences générales et recommandations (ISO 23418:2022)

Mikrobiologie der Lebensmittelkette - Vollständige Genomsequenzierung zur Typisierung und genomischen Charakterisierung von Bakterien in Lebensmitteln - Allgemeine Anforderungen und Leitfaden (ISO 23418:2022)

This European Standard was approved by CEN on 20 May 2022.

CEN members are bound to comply with the CEN/CENELEC Internal Regulations which stipulate the conditions for giving this European Standard the status of a national standard without any alteration. Up-to-date lists and bibliographical references concerning such national standards may be obtained on application to the CEN-CENELEC Management Centre or to any CEN member.

This European Standard exists in three official versions (English, French, German). A version in any other language made by translation under the responsibility of a CEN member into its own language and notified to the CEN-CENELEC Management Centre has the same status as the official versions.

CEN members are the national standards bodies of Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Republic of North Macedonia, Romania, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey and United Kingdom.

EUROPEAN COMMITTEE FOR STANDARDIZATION
COMITÉ EUROPÉEN DE NORMALISATION
EUROPÄISCHES KOMITEE FÜR NORMUNG

CEN-CENELEC Management Centre: Rue de la Science 23, B-1040 Brussels

© 2022 CEN All rights of exploitation in any form and by any means reserved worldwide for CEN national Members.

Ref. No. EN ISO 23418:2022 E

Documents
Document Sharing Hub

Full document details: BS EN ISO 23418:2022

Contents

Page

European foreword **3**

Documents
Document Sharing Hub

Full document details: BS EN ISO 23418:2022

European foreword

This document (EN ISO 23418:2022) has been prepared by Technical Committee ISO/TC 34 "Food products" in collaboration with Technical Committee CEN/TC 463 “Microbiology of the food chain” the secretariat of which is held by AFNOR.

This European Standard shall be given the status of a national standard, either by publication of an identical text or by endorsement, at the latest by December 2022, and conflicting national standards shall be withdrawn at the latest by December 2022.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. CEN shall not be held responsible for identifying any or all such patent rights.

This document has been prepared under a Standardization Request given to CEN by the European Commission and the European Free Trade Association.

Any feedback and questions on this document should be directed to the users’ national standards body/national committee. A complete listing of these bodies can be found on the CEN website.

According to the CEN-CENELEC Internal Regulations, the national standards organizations of the following countries are bound to implement this European Standard: Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Republic of North Macedonia, Romania, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey and the United Kingdom.

Endorsement notice

The text of ISO 23418:2022 has been approved by CEN as EN ISO 23418:2022 without any modification.

Documents
Document Sharing Hub

Full document details: BS EN ISO 23418:2022

INTERNATIONAL STANDARD

ISO 23418

First edition
2022-06

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

Microbiologie de la chaîne alimentaire — Séquençage de génome entier pour le typage et la caractérisation génomique des bactéries — Exigences générales et recommandations

Reference number
ISO 23418:2022(E)

Documents
Document Sharing Hub
© ISO 2022

Full document details: BS EN ISO 23418:2022

COPYRIGHT PROTECTED DOCUMENT

© ISO 2022

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

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

© ISO 2022 – All rights reserved

Documents
Document Sharing Hub

Full document details: BS EN ISO 23418:2022

Contents

Page

Foreword ... **v**
Introduction ... **vi**
1 Scope ... **1**
2 Normative references ... **1**
3 Terms and definitions ... **1**
4 Principle ... **6**
4.1 General ... 6
4.2 Laboratory operation: sample preparation and sequencing ... 6
4.3 Bioinformatics analysis ... 7
4.3.1 General ... 7
4.3.2 SNP analyses ... 7
4.3.3 MLST analyses ... 7
4.3.4 Kmer distance analysis ... 7
4.4 Metadata formats and sequence repository deposition ... 7
4.5 Validation and verification of WGS workflow ... 8
5 General laboratory guidance ... **8**
5.1 Bacterial isolation and DNA extraction ... 8
5.2 Laboratory environment ... 8
5.3 Standard operating procedures and nonconforming work ... 8
5.4 Laboratory information management system ... 8
5.5 Laboratory competence ... 8
6 Laboratory operations ... **9**
6.1 Sample preparation and storage ... 9
6.2 Bacterial isolates ... 9
6.3 DNA isolation ... 9
6.4 Library preparation and sequencing ... 9
6.4.1 Library preparation ... 9
6.4.2 DNA sequencing ... 10
6.4.3 Use of controls ... 10
6.4.4 Assessing raw read data quality ... 10
6.4.5 Sample and data storage and retention ... 10
7 Bioinformatic data analysis ... **11**
7.1 Requirements for software and bioinformatic pipelines used for data analysis ... 11
7.2 Logging and documentation ... 11
7.3 Quality assessments ... 11
7.4 SNP analyses ... 12
7.5 MLST analyses (cgMLST and wgMLST) ... 12
7.6 Target gene detection ... 13
7.7 Phylogenetic tree or dendrogram generation ... 13
7.8 Metrics and log files ... 13
7.9 Interpreting and reporting the results of bioinformatics analyses ... 13
7.9.1 Interpreting results from bioinformatics pipelines ... 13
7.9.2 Reporting genomic analysis results ... 14
8 Metadata ... **14**
8.1 General ... 14
8.2 Metadata interoperability and future-proofing ... 14
8.2.1 General ... 14
8.2.2 Ontologies ... 14
8.2.3 ISO WGS Slim ... 14
8.3 Formatting metadata using this document ... 15
8.4 Metadata associated with sample collection ... 15

© ISO 2022 – All rights reserved

AVA Documents
Document Sharing Hub
Full document details: BS EN ISO 23418:2022

8.5 Metadata associated with the isolate ... 16
8.6 Metadata associated with the sequence ... 17

9 Sequence repositories ... 19

10 Validation and verification ... 20
10.1 Validation ... 20
10.1.1 General ... 20
10.1.2 Validation of laboratory operations ... 21
10.1.3 Validation of the bioinformatics pipeline ... 21
10.1.4 Validation of the end-to-end workflow ... 22
10.2 Verification ... 22
10.2.1 General ... 22
10.2.2 Verification of laboratory operations ... 22
10.2.3 Verification of the bioinformatics pipeline ... 22

Annex A (informative) **Development of quality metrics and use of controls ... 24**

Annex B (informative) **Laboratory contact information fields ... 29**

Annex C (informative) **Geographic location of sample collection fields ... 31**

Annex D (informative) **Isolate passage history fields ... 32**

Annex E (informative) **Antibiogram results and methods fields ... 33**

Annex F (informative) **Virulence factor detection and methods fields ... 35**

Annex G (informative) **Sequence quality control metrics ... 36**

Annex H (informative) **Metadata specification ... 37**

Annex I (informative) **Instructions for ontology slim integration by software developers ... 40**

Bibliography ... 44

© ISO 2022 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS EN ISO 23418:2022

Foreword

ISO (the International Organization 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*, in collaboration with the European Committee for Standardization (CEN) Technical Committee CEN/TC 463, *Microbiology of the food chain,* in accordance with the Agreement on technical cooperation between ISO and CEN (Vienna Agreement).

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.

© ISO 2022 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS EN ISO 23418:2022

Introduction

Next generation sequencing (NGS) provides rapid, economical and high-throughput access to microbial whole genome sequences and is being applied to an expanding number of problems in food microbiology. Whole genome sequences are representations of the biological potential of the sequenced organism at single base resolution. Whole genome sequencing (WGS) offers significant advantages over existing technologies (e.g. serotyping, pulsed field gel electrophoresis, antibiotic resistance phenotype) for many applications. 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 using whole genome sequences to characterize bacterial isolates from 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 analyse whole genome sequence data. Others are turning to third-party laboratories to perform these services, as they have done for other microbiological analyses.

This document provides guidance for both the laboratory and bioinformatic components of whole genome sequences and associated metadata for bacterial foodborne microorganisms sampled along the food chain (e.g. ingredients, food, feed, production environment). Although microbiology of the food chain includes viruses and fungi, this document is only intended for bacteria. This document is intended to be applicable to all currently available next generation DNA sequencing technologies. It may be applied to analysis of whole genome sequence data with proprietary, open-source or custom software. It is not intended to specify sequencing chemistries, analytical methods or software. This document defines laboratory, data and metadata stewardship practices to ensure that analyses are clearly reported, transparent and open to inquiry. This document is for use by laboratories to develop their management systems for quality and technical operations. Laboratory customers and regulatory authorities can also use it in confirming or recognizing the competence of laboratories. This document can also be applied in other domains (e.g. environment, human health, animal health).

© ISO 2022 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS EN ISO 23418:2022

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

WARNING — In order to safeguard the health of laboratory personnel, it is essential that handling of bacterial cultures is only undertaken in properly equipped laboratories, under the control of a skilled microbiologist, and that great care is taken in the disposal of all incubated materials. Persons using this document should be familiar with normal laboratory practice. This document does not purport to address all safety aspects, if any, associated with its use. It is the responsibility of the user to establish appropriate safety and health practices.

1 Scope

This document specifies the minimum requirements for generating and analysing whole genome sequencing (WGS) data of bacteria obtained from the food chain. This process can include the following stages:

a) handling of bacterial cultures;

b) axenic 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;

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

This document is applicable to bacteria isolated from:

— products intended for human consumption;

— products intended for animal feed;

— environmental samples from food and feed handling and production areas;

— samples from the primary production stage.

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 terminology databases for use in standardization at the following addresses:

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

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

© ISO 2022 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS EN ISO 23418:2022

3.1
adapter sequence
DNA with a known sequence that 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* (3.4)

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* (3.38) into larger contiguous sequences (*contigs* (3.10))

3.5
base calling
process of assigning nucleotides and quality scores to positions in sequencing *reads* (3.38)

3.6
bioinformatics
collection, storage and analysis of biological data including sequences

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
sample contamination linked to previous experiments, transferred to the current analysis (e.g. carryover-contamination from amplification products in prior polymerase chain reaction (PCR) experiments to the current PCR analysis, or carryover-contamination of previously sequenced samples from one sequencing run to another)

3.9
Chemical Entities of Biological Interest Ontology
ChEBI
ontology (3.35) for describing small chemical compounds

3.10
contig
contiguous stretch of DNA sequence that results from the *assembly* (3.4) of smaller, overlapping DNA sequence *reads* (3.38)

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

[SOURCE: ISO 11238:2018, 3.18, modified — Note 1 to entry deleted.]

3.12
coverage
number of times that a given base position is read in a sequencing run

Note 1 to entry: The number of *reads* (3.38) that cover a particular position.

© ISO 2022 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS EN ISO 23418:2022

[SOURCE: ISO 20397-2:2021, 3.6, modified — Admitted term "coverage depth" deleted.]

3.13
cross-contamination
contamination of a sample (bacterial *isolate* (3.23) or DNA) with other samples during the preparation of a sequencing run

3.14
DNA sample
portion of DNA extracted from the processed sample

3.15
draft assembly
de novo genome *assembly* (3.4) consisting of *contigs* (3.10) with no implied order, typically generated using whole genome shotgun sequencing with a short-read technology

3.16
Environment Ontology
EnvO
ontology (3.35) for describing environmental features and habitats

3.17
FoodEx2 Ontology
FoodEx2
standardized food classification and description system developed by the European Food Safety Authority (EFSA)

3.18
Food Ontology
FoodOn
ontology (3.35) for describing food products, animal feed and food processing

3.19
Gazetteer Ontology
GAZ
ontology (3.35) for describing geographical locations

3.20
index
oligonucleotide sequences used in the process of library preparation to tag or barcode DNA from specific samples, so that multiple samples (i.e. multiple *libraries* (3.25)) can be combined (multiplexed) in a pool of libraries and analysed in a single sequencing reaction

3.21
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)

3.22
International Organization for Standardization whole genome sequencing slim
ISO WGS Slim
ontology (3.35) slim containing interoperable fields and terms pertaining to the use of *WGS* (3.49) for microbiology of the food chain

3.23
isolate
population of bacterial cells in pure culture derived from a single *strain* (3.45)

© ISO 2022 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS EN ISO 23418:2022

3.24
kmer
possible sequence of length *k* that is contained in a whole genome sequence

3.25
library
collection of genomic DNA fragments from a single *isolate* (3.23) intended for determining genome sequence(s)

Note 1 to entry: A collection of libraries, each of a single isolate, is called a "pool of libraries" and is loaded on a sequencer to be analysed. This multiplexing of libraries would still provide the result for a single isolate if unique indices are used for each individual single isolate's library preparation.

Note 2 to entry: A library of mixed DNA, i.e. originating from a mixture of multiple species, can be made. However, this is not within the scope of this document as this refers to metagenomics sequencing.

3.26
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.

3.27
mapping
use of software to align sequencing *reads* (3.38) to reference sequences

3.28
metadata
data that defines and describes other data

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

3.29
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[1], and implemented by the *International Nucleotide Sequence Database Collaboration* (3.21)

3.30
multi-locus sequence typing
MLST
method of genomic analysis that identifies nucleotide variants within predefined sets of loci

Note 1 to entry: Originally used for seven loci, it is now also applied to either core genome loci for cgMLST or whole genome loci for wgMLST.

3.31
N50
length (N) such that sequence *contigs* (3.10) of N or longer include half the bases in the *assembly* (3.4)

3.32
NCBITaxon
automatic translation of the National Center for Biotechnology Information (NCBI) taxonomy database into obo/owl

3.33
NG50
length (N) of DNA such that sequence *contigs* (3.10) of N or longer include half the bases in the genome

© ISO 2022 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS EN ISO 23418:2022

3.34
Open Biological and Biomedical Ontology Foundry
OBO Foundry
collection of *ontologies* (3.35) created by a collective of ontology developers that are committed to collaboration and adherence to shared principles

3.35
ontology
controlled vocabulary (3.11) arranged in a hierarchy, where the terms are connected by logical relationships

3.36
ontology slim
set of ontology fields and terms annotated as part of a particular collection, often for a specific purpose, which may be extracted to create a file distinct from the original *ontology* (3.35)

3.37
Phred sequence quality score
Q
measure of the probability (*P*) that a base is incorrectly assigned at a given position in the sequence expressed as:

$$Q = -10 \lg P$$

Note 1 to entry: A score of Q30 indicates that there is a 1 in 1 000 chance that a base is incorrectly assigned (i.e. the base call is 99,9 % accurate).

3.38
read
nucleotide sequence inferred from a fragment of DNA or RNA

3.39
sequence repository
database in which *whole genome sequencing* (3.49) datasets are stored and managed

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

3.40
sequencing replicate
<biological> sequencing a different colony from the same *isolate* (3.23) obtained from the same sample material, to assess biological variation

3.41
sequencing replicate
<technical> resequencing of the same biological sample or *library* (3.25) to assess sequence variation due to instrumentation and protocol

3.42
serotype
classification scheme based on the antigenic protein detection or sequence-based detection of genes encoding bacterial surface molecules

3.43
single nucleotide polymorphism
SNP
single nucleotide variant (3.44) that passes a particular quality or frequency threshold

© ISO 2022 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS EN ISO 23418:2022

3.44
single nucleotide variant
SNV
differences between the nucleotides at the same genomic position of two or more *isolates* (3.23)

3.45
strain
descendants of a single isolation in pure culture, usually derived from a single initial colony on a solid growth medium

Note 1 to entry: A strain may be considered an *isolate* (3.23) or group of isolates that may be distinguished from other isolates of the same genus and species by phenotypic and genotypic characteristics.

Note 2 to entry: See Reference [2].

3.46
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, 2.81]

3.47
validated data entry
automated process ensuring that data entered into a repository are correct

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

[SOURCE: ISO 16140-3:2021, 3.21, modified — Note 1 to entry deleted.]

3.49
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

WGS analyses of bacteria along the food and feed chain consists of culturing the pure bacterial isolate, DNA isolation performed in a microbiological laboratory, sequencing steps conducted in an appropriate sequencing environment and bioinformatics analysis performed in a distinct computational environment.

NOTE The microbiology laboratory, the sequencing facility and the bioinformatics facility can be the same organization.

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) or sample description worksheets, or both.

b) Pure isolates (identified at least to the genus level and ideally to the species level) are cultured and genomic DNA is extracted.

© ISO 2022 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS EN ISO 23418:2022

c) DNA sequencing libraries are prepared from quality controlled genomic DNA (see Table A.1 for guidance on DNA quantity and quality metrics). This process should include:

   1) DNA fragmentation, if required for the applied sequencing technology;

   2) ligation of indices and adapters, consistent with the applied sequencing technology's protocols;

   3) quantification, normalization and quality control of the resulting library;

   4) pooling of libraries in the case of multiplexed sequencing runs.

d) Libraries (i.e. pool of libraries) are sequenced.

e) Quality metrics produced by the sequencing instrument are ideally recorded for each run to allow monitoring of the performance.

4.3 Bioinformatics analysis

4.3.1 General

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

4.3.2 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.

4.3.3 MLST analyses

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

4.3.4 Kmer distance analysis

Sequence data for multiple isolates may be analysed 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 may be shared solely at the discretion of the metadata owner. Sequence data and its corresponding metadata shall be subject to security considerations, cost and benefits, intellectual property rights, confidential business information, contract restriction or other binding written agreements.

NOTE Licensing or a privacy policy, or both, can be applied to metadata or sequence data, or both, to protect private or proprietary information.

To promote data stewardship best practices[3], this document 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 [illegible]

© ISO 2022 – All rights reserved

AVA Documents
Document Sharing Hub
Full document details: BS EN ISO 23418:2022