BS ISO 24480:2024

BSI Standards Publication

Biotechnology — Validation of database used for nucleotide sequence evaluation

Documents
Document Sharing Hub

Full document details: BS ISO 24480:2024

National foreword

This British Standard is the UK implementation of ISO 24480:2024.

The UK participation in its preparation was entrusted to Technical Committee BTI/1, Biotechnologies.

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 2024
Published by BSI Standards Limited 2024

ISBN 978 0 539 22786 4

ICS 07.080

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 30 November 2024.

Amendments/corrigenda issued since publication

Date	Text affected

Documents
Document Sharing Hub

Full document details: BS ISO 24480:2024

International Standard

ISO 24480

First edition
2024-11

Biotechnology — Validation of database used for nucleotide sequence evaluation

Biotechnologie — Validation de la base de données utilisée pour l'évaluation de la séquence nucléotidique

Reference number
ISO 24480:2024(en)

© ISO 2024

Documents
Document Sharing Hub

Full document details: BS ISO 24480:2024

COPYRIGHT PROTECTED DOCUMENT

© ISO 2024

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

Documents
Document Sharing Hub
Full document details: BS ISO 24480:2024

Contents

Page

Foreword ... **iv**
Introduction ... **v**
1 Scope ... **1**
2 Normative references ... **1**
3 Terms and definitions ... **1**
4 General ... **3**
5 Common requirements of the database ... **4**
6 Inclusivity database: database for inclusivity evaluation ... **4**
6.1 Quality criteria ... 4
6.2 Requirements of an inclusivity database ... 4
6.3 Individual data quality indicators ... 5
6.3.1 Data provenance and updates ... 5
6.3.2 Length of the entries ... 5
6.3.3 Number of unidentified nucleotides (N) ... 5
6.4 Validation of the inclusivity database ... 5
7 Exclusivity database: Database for the exclusivity evaluation ... **6**
7.1 Quality criteria ... 6
7.2 Requirements of the exclusivity database ... 6
7.3 Validation of the exclusivity database ... 6
8 Validation report ... **7**
Annex A (informative) **Example of a data entry format** ... **9**
Annex B (informative) **Example for the validation of an inclusivity database (Inclusivity)** ... **10**
Annex C (informative) **Example on validation of an exclusivity database (Exclusivity)** ... **13**
Annex D (informative) **Example of use of an inclusivity database and an exclusivity database** ... **17**
Annex E (informative) **Example on dataset verification command for an inclusivity database** ... **23**
Bibliography ... **24**

© ISO 2024 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 24480:2024

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 document 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).

ISO draws attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO takes no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents. ISO shall not be held responsible for identifying any or all such patent rights.

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 276, *Biotechnology*.

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 2024 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 24480:2024

Introduction

A valid database is important for nucleotide sequence evaluation. The development of inclusivity and exclusivity panels for diagnostics and surveillance using community genomic databases, e.g., Genbank, has been evaluated[1],[2],[3]. However, a specific validation procedure for the databases has yet to be provided.

Considering the current database quality, inclusivity and exclusivity are almost impossible to be validated with ideal accuracy. Therefore, in this document a practical procedure for evaluating the quality of nucleotide sequence database to be used for the development of inclusivity and exclusivity panels is comprehensively described. The degree of data accuracy to be used is determined according to the user's intended test purpose. This evaluation can become a part of the validated diagnostic or surveillance method. Ensuring the quality of the database improves its sufficiency for validating the whole measuring system.

In polymerase chain reaction (PCR) and DNA microarray technologies, nucleotide sequence is used as primers or probes to detect the target nucleic acids. Those technologies utilize initially the hybridization of two single strand DNA molecules with complementary sequences. During the design process of the primers or probes, nucleotide sequence database is used for evaluating specificity and exclusivity of probes or primers. In general, target DNA sequences can be confirmed to match the intended sequences but not others by similarity (homology) search on nucleotide databases with computer tools, for example BLAST.

The validated databases can be used for evaluating specificity of probe or primer sequences and ensuring the selectivity of the qualification and quantification measurement system.

Validation of the entire nucleotide sequence database is not appropriate for the database providers because there are wide varieties of purpose of uses by users. It is almost impossible for the users, however, to evaluate the quality of each data entry especially in huge sequence databases. The database can reflect the fitness for the intended test purpose of users.

This document provides the minimum requirements of a practical procedure for the validation of database used for nucleotide sequence evaluation.

© ISO 2024 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 24480:2024

Documents
Document Sharing Hub

Biotechnology — Validation of database used for nucleotide sequence evaluation

1 Scope

This document describes a practical procedure for nucleotide sequence database evaluation and validation. This document describes minimum requirements for the validation of a nucleotide sequence database. This document is applicable only for databases consisting of entries of nucleotide sequences.

This document is not applicable to the general evaluation of the entire database quality including the quality of each data entry.

EXAMPLE The use of the validated database is for confirming a representative sequence specificity including primers or probes for qualification and quantification of target nucleic acids by conventional polymerase chain reaction (PCR), quantitative polymerase chain reaction (qPCR), digital polymerase chain reaction (dPCR) and microarray technologies.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 20691, *Biotechnology — Requirements for data formatting and description in the life sciences*

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/

3.1
nucleotide sequence specificity
capacity to exclusively recognize a specific nucleic acid target sequence, distinguishing it from other nucleic acids and contaminants

Note 1 to entry: It describes the degree of similarity to specifically match to the nucleotide sequence to be searched by distinguishing it from other nucleotide sequences, and the tendency for a primer or probe with the matched nucleotide sequence to hybridize with its intended target and not hybridize with other non-target sequences.

Note 2 to entry: “sequence specificity” can be considered to be the combination of *inclusivity* (3.4) and *exclusivity* (3.5)

3.2
selectivity
extent to which a method can determine particular analyte(s) in a mixture(s) or matrice(s) without interferences from other components of similar behaviour

Note 1 to entry: Selectivity is the recommended term in analytical chemistry to express the extent to which a particular method can determine analyte(s) in the presence other components. Selectivity can be graded. The use of the term “specificity” for the same concept is to be discouraged as this often leads to confusion.

© ISO 2024 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 24480:2024

Note 2 to entry: Sequence specificity in molecular biomarker analysis is differentiated from chemical analyte selectivity.

[SOURCE: ISO 16577:2022, 3.3.73]

3.3
sequence similarity
proportion of matched number of units, including nucleotides and amino acids, to the number of units in specified regions between two nucleic acids or proteins

Note 1 to entry: gap or deletion can be considered to compare the unit sequence of the regions between two nucleic acids or proteins.

Note 2 to entry: "sequence similarity" can be evaluated simply based on a proportion of matched number of units, whereas the term "homology" contains biological meaning in the comparison of two nucleic acids or proteins.

3.4
inclusivity
property of a nucleotide sequence to show high *sequence similarity* (3.3) specifically with intended target nucleotide sequence

Note 1 to entry: The term "inclusivity" is used as same meaning of "sensitivity" in some cases [4].

3.5
exclusivity
property of a nucleotide sequence to show low *sequence similarity* (3.3) with those excluding intended target nucleotide sequences

3.6
nucleic acid test
NAT
technique used to detect or quantify a target nucleic acid with specific sequence, by using of oligonucleotide as a primer or probe

3.7
representative sequence
group of nucleotide sequence data containing one or more target sequences in a complete or partial sequence intended for detection or quantification

3.8
undesirable sequence
group of nucleotide sequence data containing one or more nucleotide sequences, which are potentially either influencing or intentionally excluded, or both, for detection or quantification

3.9
intended test purpose
purpose of nucleic acid detection or quantification using oligonucleotide, e.g., primers or probes, whose design is evaluated by using validated nucleotide sequence databases containing representative and *undesirable sequences* (3.8)

3.10
exploring key
data used for examining data entries stored in database

EXAMPLE Key words, sequence data, taxon data, tissue name etc.

3.11
inclusivity database
database used for evaluating *inclusivity* (3.4) of a specified nucleotide sequence

3.12
exclusivity database
database used for evaluating *exclusivity* (3.5) of a specified nucleotide sequence

© ISO 2024 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 24480:2024

3.13
provenance information
information that documents the history of a described object and related described activities, and that contains information about the origin or source of the described object, any changes that can have taken place since it was originated, and who has had custody of it since it was originated

[SOURCE: ISO/TS 23494-1:2023, 3.13]

3.14
finalized provenance information
provenance information transformed into a representation specified by the common provenance model, and which is prepared to be conserved or archived and which is considered as being immutable

Note 1 to entry: Finalized provenance information is a subset of provenance information.

[SOURCE: ISO/TS 23494-1:2023, 3.5]

3.15
basic local alignment search tool
BLAST
sequence comparison algorithm optimized for speed that is used to search sequence databases for optimal local alignments to a query

[SOURCE: ISO 20813:2019, 3.1 modified — Notes to entry have been deleted.]

3.16
massively parallel nucleotide sequencing
next generation sequencing
NGS
high throughput nucleotide sequencing method capable of determining multiple DNA sequences simultaneously and in parallel

Note 1 to entry: The data from a single massively parallel sequencing analysis comprises of millions of sequences and the output is a file containing all sequences.

[SOURCE: ISO 16577:2022, 3.7.10, modified —"whole genome sequencing" and "WGS" have been deleted.]

4 General

There are significant differences between the inclusivity and exclusivity confirmation roles. A database used for the inclusivity analysis shall cover all intended sequence entries. In addition, other recognized unintended sequence entries should be contained to show the sequence similarity is specific to the intended sequence entries in a recognized extent, although exclusivity cannot be confirmed only with the recognized undesirable sequences. Entries with high and reliable quality of both intended sequences and recognized undesirable sequences should be included in the database[3]. Quality of the sequence entries for inclusivity analysis is described in the validation of inclusivity database.

On the other hand, a database used for exclusivity analysis should include as many sequence entries as possible including related and not likely related ones. Even though the entries are not likely related sequences, the database should present those sequences because they can contain sequences that can be unintentionally hybridized by the primers or probes in non-specific manner and rise the amplification background. The quality of the sequence entries for the exclusivity analysis is described in the validation of the exclusivity database (for details, see 7.3). One example for uses of a validated database is for an *in silico* analysis evaluating nucleotide sequence specificity in designing primers or probes for nucleic acid measurement including various PCR-based methods and microarray analysis, i.e., inclusivity and exclusivity analyses. Databases for the inclusivity analysis are used for evaluation on how the designed primers and probes can work to distinguish a target in the nucleic acid measurement methods. A database for the exclusivity analysis is used for evaluating the possibility of the designed primers and probes showing non-specific detection or quantification in the nucleic acid measurement methods. Therefore, when users validate a database, i.e., confirm that the requirements of the database for a specific intended use have been fulfilled,

© ISO 2024 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 24480:2024