PD ISO/TS 24420:2023



BSI Standards Publication

Biotechnology — Massively parallel DNA sequencing — General requirements for data processing of shotgun metagenomic sequences

bsi.

National foreword

This Published Document is the UK implementation of ISO/TS 24420:2023.

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.

This publication is not to be regarded as a British Standard.

© The British Standards Institution 2023
Published by BSI Standards Limited 2023

ISBN 978 0 539 14415 4

ICS 07.080

Compliance with a Published Document cannot confer immunity from legal obligations.

This Published Document was published under the authority of the Standards Policy and Strategy Committee on 31 May 2023.

Amendments/corrigenda issued since publication

Date	Text affected

PD ISO/TS 24420:2023

TECHNICAL SPECIFICATION

ISO/TS 24420

First edition
2023-05-19

Biotechnology — Massively parallel DNA sequencing — General requirements for data processing of shotgun metagenomic sequences

Biotechnologie — Séquençage d'ADN massivement parallèle — Exigences générales pour le traitement des données des séquences métagénomiques "Shotgun"

Reference number
ISO/TS 24420:2023(E)

© ISO 2023



COPYRIGHT PROTECTED DOCUMENT

© ISO 2023, Published in Switzerland

All rights reserved. Unless otherwise specified, 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
Ch. de Blandonnet 8 • CP 401
CH-1214 Vernier, Geneva, Switzerland
Tel. +41 22 749 01 11
Fax +41 22 749 09 47
copyright@iso.org
www.iso.org

Contents

Page

Foreword ..... **iv**
Introduction ..... **v**
1 Scope **1**
2 Normative references **1**
3 Terms and definitions **1**
4 Processing workflow **4**
5 Data processing **5**
5.1 Facilities and software requirements 5
5.2 Sequence quality control and error determination 5
5.3 Sequence assembly 6
6 Data analysis **6**
6.1 Annotation 6
6.2 Calculation of species relative abundance 7
6.2.1 Species analysis 7
6.2.2 Gene analysis 7
7 Data archive and metadata **7**
7.1 Original data 7
7.2 Sequencing analytical data 7
7.3 Data directory and archive 8
7.3.1 General 8
7.3.2 Directory of data elements 8
7.3.3 Data archiving 8
7.4 Metadata 8
Annex A (informative) **Examples of data format** **10**
Annex B (informative) **Directory of data elements** **16**
Bibliography **18**

© ISO 2023 – All rights reserved

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 2023 – All rights reserved

Introduction

Shotgun metagenomic sequencing of organisms' genomes from a complex sample is widely used in life science and clinical applications (e.g. human complex disease associated analysis, environmental microecology and other fields) in order to gain knowledge of their composition and function. It has potential to provide significant scientific data for life science research.

The utility of this technique is its ability to reveal the microbial diversity and abundance found in microbial populations from multiple environments and to determine sequence information (taxonomic characterization, functional annotation, and comparative analysis/metagenomics) for individual organisms in these populations. The resulting data can be subjected to comparative analytics. Massively parallel shotgun metagenomic sequencing generates a large amount of data containing a high complexity of microbial genomes and a large number of unknown species. It is important to use effective processing procedures and address quality control for shotgun metagenomic sequencing data. A standardised data format is essential to promote data sharing.

As with any advanced technology, massively parallel sequencing technologies is error prone. Overcoming these shortcomings to ensure a reliable sequencing and analytical outcome is important. This document provides a uniform standard for the collation, storage and subsequent analysis of metagenomic data, and guidelines. It provides requirements and recommendations for the workflow and process of shotgun metagenomic analyses including quality control of sequencing data and metadata, and the compositional and functional analysis of microbial community. These requirements and recommendations can ensure accuracy of data generated from metagenomic analysis, address potential errors and facilitate downstream applications.

© ISO 2023 – All rights reserved

Biotechnology — Massively parallel DNA sequencing — General requirements for data processing of shotgun metagenomic sequences

1 Scope

This document illustrates the workflow of shotgun metagenomic sequence data processing of host-derived microbiome and environmental metagenomes.

This document specifies the requirements for quality control of shotgun metagenomic sequence data processing for massively parallel DNA sequencing.

This document provides guidelines for data directory, data archive and metadata for shotgun metagenomic sequence data.

This document applies to data storage, sharing and interoperability of shotgun metagenomic sequence data.

This document applies to shotgun metagenomic sequence data processing and analyses, but excludes functional analysis.

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 20397-1:2022, *Biotechnology — Massively parallel sequencing — Part 1: Nucleic acid and library preparation*

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
attribute value
value associated with an attribute instance

[SOURCE: ISO 21962:2003, 1.5.2.3]

3.2
category
set of items or concepts that share a common attribute or feature

© ISO 2023 – All rights reserved