© ISO 2022 – All rights reserved

ISO/DTS~~.2~~ 24420:~~2022~~2023(E)

Date: 2023-01-09

ISO/TC 276

Secretariat: DIN

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"

iTeh STANDARD PREVIEW
(standards.iteh.ai)

~~DTS.2 stage~~

ISO/DTS 24420
https://standards.iteh.ai/catalog/standards/sist/4816721b-9c79-4124-8496-

~~**Warning for WDs and CDs**~~

~~This document is not an ISO International Standard. It is distributed for review and comment. It is subject to change without notice and may not be referred to as an International Standard.~~

~~Recipients of this draft are invited to submit, with their comments, notification of any relevant patent rights of which they are aware and to provide supporting documentation.~~

~~*To help you, this guide on writing standards was produced by the ISO/TMB and is available at http://www.iso.org/iso/how-to-write-standards.pdf*~~

~~*A model manuscript of a draft International Standard (known as "The Rice Model") is available at http://www.iso.org/iso/model-document-rice-model.pdf*~~

iTeh STANDARD PREVIEW
(standards.iteh.ai)
ISO/DTS 24420
https://standards.iteh.ai/catalog/standards/sist/4816721b-9c79-4124-8496-35a2a2a367d8/iso-dts-24420

© ISO 2022

iTeh STANDARD PREVIEW
(standards.iteh.ai)
ISO/DTS 24420
https://standards.iteh.ai/catalog/standards/sist/4816721b-9c79-4124-8496-35a2a2a367d8/iso-dts-24420

© ISO 2023

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

Fax: +41 22 749 09 47

Email: copyright@iso.org

Website: www.iso.orgwww.iso.org

Published in Switzerland

iTeh STANDARD PREVIEW
(standards.iteh.ai)
ISO/DTS 24420
https://standards.iteh.ai/catalog/standards/sist/4816721b-9c79-4124-8496-35a2a2a367d8/iso-dts-24420

Formatted: Font: 11 pt

Formatted: Space Before: 0 pt, Line spacing: Exactly 11 pt, Tab stops: Not at 487.6 pt

Contents

iTeh STANDARD PREVIEW
(standards.iteh.ai)
ISO/DTS 24420
https://standards.iteh.ai/catalog/standards/sist/4816721b-8c79-4124-8496-35a2a2a367d8/iso-dts-24420

Formatted: Font: 11 pt

Formatted: Space Before: 0 pt, Line spacing: Exactly 11 pt, Tab stops: Not at 487.6 pt

© ISO 2022 – All rights reserved 3

© ISO 2023 – All rights reserved

iTeh STANDARD PREVIEW
(standards.iteh.ai)
ISO/DTS 24420
https://standards.iteh.ai/catalog/standards/sist/4816721b-9c79-4124-8496-35a2a2a367d8/iso-dts-24420

Formatted: Font: 11 pt

Formatted: Space Before: 0 pt, Line spacing: Exactly 11 pt, Tab stops: Not at 487.6 pt

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 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/patentswww.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.htmlwww.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.htmlwww.iso.org/members.html.

iTeh STANDARD PREVIEW
(standards.iteh.ai)
ISO/DTS 24420
https://standards.iteh.ai/catalog/standards/sist/4816721b-9c79-4124-8496-35e2e2e267d8/iso-dts-24420

Formatted: Font: 11 pt

Formatted: Space Before: 0 pt, Line spacing: Exactly 11 pt, Tab stops: Not at 487.6 pt

© ISO 2022 – All rights reserved 5

© ISO 2023 – All rights reserved

Introduction

Shotgun metagenomic sequencing genomes of organisms in a complex sample in a community to gain knowledge of its composition and function is widely used in life science and clinical applications, such as human complex disease associated analysis, environmental microecology and other fields. 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 (Taxonomictaxonomic 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.

iTeh STANDARD PREVIEW
(standards.iteh.ai)
ISO/DTS 24420
https://standards.iteh.ai/catalog/standards/sist/4816721b-9c79-4124-8496-35a2a2a367d8/iso-dts-24420

Formatted: Font: 11 pt

Formatted: Space Before: 0 pt, Line spacing: Exactly 11 pt, Tab stops: Not at 487.6 pt

iTeh STANDARD PREVIEW
(standards.iteh.ai)
ISO/DTS 24420
https://standards.iteh.ai/catalog/standards/sist/4816721b-9c79-4124-8496-35a2a2a367d8/iso-dts-24420

Formatted: Font: 11 pt

Formatted: Space Before: 0 pt, Line spacing: Exactly 11 pt, Tab stops: Not at 487.6 pt

© ISO 2022 – All rights reserved 7

© ISO 2023 – All rights reserved

iTeh STANDARD PREVIEW
(standards.iteh.ai)
ISO/DTS 24420
https://standards.iteh.ai/catalog/standards/sist/4816721b-9c79-4124-8496-35a2a2a367d8/iso-dts-24420

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.

iTeh STANDARD PREVIEW
(standards.iteh.ai)

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 — ~~General requirements for massively~~Massively parallel sequencing — Part 1: Nucleic acid and library preparation*

ISO/DTS 24420
https://standards.iteh.ai/catalog/standards/sist/c79-4124-8496-35a2a2a367d8/iso-dts-24420

~~ISO 20397-2:2021, *Biotechnology — Massively parallel sequencing — Part 2: Methods to evaluate the quality of sequencing data*~~

43 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 ~~things~~items or concepts that share a common attribute or feature

© ISO 2023 – All rights reserved

3.3
classification
exhaustive set of mutually exclusive categories to aggregate data at a pre-prescribed level of specialization for a specific purpose

[SOURCE: ISO 17115:2007, 2.7.1]

3.4
clean data
sequencing data obtained after a pre-processing procedure which usually includes multiple trimming and filtering steps to ensure specific quality levels (e.g., per-base quality, host/contaminant sequences removed, linkers/adaptors removed)

3.5
code
system of rule(s) to convert information such as text, images, sounds or electric, photonic or magnetic signals into another form or representation to facilitate analysis, communication or storage in a storage medium

[SOURCE: ISO 20691:2022, 3.6]

3.6
coding
encoding
process of assigning code to things or concepts

iTeh STANDARD PREVIEW
(standards.iteh.ai)

3.7
contig
contiguous sequence of DNA created by assembling overlapping sequenced fragments of a chromosome or plasmid

ISO/DTS 24420:2023
https://standards.iteh.ai/catalog/standards/sist/4816721b-9c79-4124-8496-35a2a2a367d8/iso-dts-24420

3.8
data format
arrangement of data according to preset specifications

Note 1 to entry: Preset specifications are usually made for computer processing.

3.9
data element
single unit of data that in a certain context is considered indivisible

[SOURCE: ISO/TS 21089:2018, 3.44]

3.10
directory
list of data items, which gives itemized information enabling traceability, identification and findability of related data

Note 1 to entry: A directory can be arranged in alphabetical, chronological or systematic order.

3.11
directory identifier

Formatted: Font: 11 pt

Formatted: Space Before: 0 pt, Line spacing: Exactly 11 pt, Tab stops: Not at 487.6 pt