

DIN EN ISO 23418



ICS 07.100.30

Einsprüche bis 2020-11-11

Entwurf

**Mikrobiologie der Lebensmittelkette –
Vollständige Genomsequenzierung zur Typisierung und genomischen
Charakterisierung von Bakterien in Lebensmitteln –
Allgemeine Anforderungen und Leitfaden (ISO/DIS 23418:2020);
Deutsche und Englische Fassung prEN ISO 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);
German and English version prEN ISO 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);
Version allemande et anglaise prEN ISO 23418:2020

Anwendungswarnvermerk

Dieser Norm-Entwurf mit Erscheinungsdatum 2020-09-11 wird der Öffentlichkeit zur Prüfung und
Stellungnahme vorgelegt.

Weil die beabsichtigte Norm von der vorliegenden Fassung abweichen kann, ist die Anwendung dieses Entwurfs
besonders zu vereinbaren.

Stellungnahmen werden erbeten

- vorzugsweise online im Norm-Entwurfs-Portal von DIN unter www.din.de/go/entwuerfe bzw. für Norm-Entwürfe der DKE auch im Norm-Entwurfs-Portal der DKE unter www.entwuerfe.normenbibliothek.de, sofern dort wiedergegeben;
- oder als Datei per E-Mail an nal@din.de möglichst in Form einer Tabelle. Die Vorlage dieser Tabelle kann im Internet unter www.din.de/go/stellungnahmen-norm-entwuerfe oder für Stellungnahmen zu Norm-Entwürfen der DKE unter www.dke.de/stellungnahme abgerufen werden;
- oder in Papierform an den DIN-Normenausschuss Lebensmittel und landwirtschaftliche Produkte (NAL), 10772 Berlin oder Saatwinkler Damm 42/43, 13627 Berlin.

Die Empfänger dieses Norm-Entwurfs werden gebeten, mit ihren Kommentaren jegliche relevanten
Patentrechte, die sie kennen, mitzuteilen und unterstützende Dokumentationen zur Verfügung zu stellen.

Gesamtumfang 104 Seiten

DIN-Normenausschuss Lebensmittel und landwirtschaftliche Produkte (NAL)

Nationales Vorwort

Dieses Dokument (prEN ISO 23418:2020) wurde vom Technischen Komitee ISO/TC 34 „Food products“ in Zusammenarbeit mit dem Technischen Komitee CEN/TC 463 „Mikrobiologie der Lebensmittelkette“ erarbeitet, dessen Sekretariat von AFNOR (Frankreich) gehalten wird.

Das zuständige deutsche Normungsgremium ist der Arbeitsausschuss NA 057-01-06 AA „Mikrobiologie der Lebensmittelkette“ im DIN-Normenausschuss Lebensmittel und landwirtschaftliche Produkte (NAL).

Um Zweifelsfälle in der Übersetzung auszuschließen, ist die englische Originalfassung beigelegt. Die Nutzungsbedingungen für den deutschen Text des Norm-Entwurfes gelten gleichermaßen auch für den englischen Text.

Aktuelle Informationen zu diesem Dokument können über die Internetseiten von DIN (www.din.de) durch eine Suche nach der Dokumentennummer aufgerufen werden.

<i>Titel de:</i>	Mikrobiologie der Lebensmittelkette — Vollständige Genomsequenzierung zur Typisierung und genomischen Charakterisierung von Bakterien in Lebensmitteln — Allgemeine Anforderungen und Leitfaden (ISO/DIS 23418:2020)
<i>Titel en:</i>	Microbiology of the food chain — Whole genome sequencing for typing and genomic characterization of foodborne bacteria — General requirements and guidance (ISO/DIS 23418:2020)
<i>Titel fr:</i>	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)

Inhalt

	Seite
Europäisches Vorwort	4
Vorwort	5
Einleitung	6
1 Anwendungsbereich	7
2 Normative Verweisungen	7
3 Begriffe	7
4 Kurzbeschreibung	12
4.1 Allgemeines	12
4.2 Laborbetrieb: Probenvorbereitung und Sequenzierung	12
4.3 Bioinformatische Analyse	13
4.4 Metadatenformate und Hinterlegung von Sequenzdaten in Repositorien	13
4.5 Validierung und Verifizierung des WGS-Workflows	14
5 Allgemeiner Laborleitfaden	14
5.1 Bakterienisolierung und DNA-Extraktion	14
5.2 Laborumgebung	14
5.3 Standardverfahrensanweisungen (SOP, en: standard operating procedure) und nicht konforme Arbeit	14
5.4 Labor-Informations- und Management-System (LIMS)	15
5.5 Kompetenz im Labor	15
6 Laborbetrieb	15
6.1 Herstellung und Lagerung der Proben	15
6.2 Bakterienisolate	15
6.3 Isolierung der DNA	16
6.4 Vorbereitung der Bibliothek	16
6.4.1 DNA-Sequenzierung	16
6.4.2 Verwendung von Kontrollen	17
6.4.3 Beurteilung der Qualität von Read-Rohdaten	17
6.4.4 Speicherung und Aufbewahrung von Proben und Daten	17
7 Bioinformatische Datenanalyse	17
7.1 Anforderungen an Software und/oder bioinformatische Pipelines, die für die Datenanalyse verwendet werden	17
7.2 Protokollierung und Dokumentation	18
7.3 Qualitätsbeurteilungen	18
7.4 SNP-Analysen	19
7.5 MLST-Analysen (cgMLST und wgMLST)	19
7.6 Detektion von Zielgenen	20
7.7 Phylogenetische Baum- oder Dendrogramm-Generierung	20
7.8 Metriken und Protokolldateien	20
7.9 Interpretation und Meldung der Ergebnisse bioinformatischer Analysen	21
7.9.1 Interpretation der Ergebnisse aus bioinformatischen Pipelines	21
7.9.2 Angabe über die Ergebnisse der Genomanalyse	21
8 Metadaten	21
8.1 Allgemeines	21

8.2	Interoperabilität und Zukunftssicherheit von Metadaten.....	21
8.2.1	Ontologien.....	22
8.2.2	ISO WGS-Slim.....	22
8.3	Formatierung von Metadaten mit der Norm.....	22
8.4	Metadaten im Zusammenhang mit der Probenahme.....	22
8.5	Metadaten im Zusammenhang mit dem Isolat.....	23
8.6	Mit der Sequenz verbundene Metadaten.....	23
9	Sequenz-Repositorien.....	23
10	Validierung und Verifizierung.....	23
10.1	Validierung.....	23
10.1.1	Validierung des Laborbetriebs.....	23
10.1.2	Validierung der bioinformatischen Pipeline.....	23
10.1.3	Validierung des durchgehenden Workflows.....	24
10.2	Verifizierung.....	25
10.2.1	Verifizierung des Laborbetriebs.....	25
10.2.2	Verifizierung der bioinformatischen Pipeline.....	26
Anhang A (informativ)		33
A.1	Leitfaden für die Entwicklung von Qualitätsmetriken für Sequenzierungstechnologien mit kurzer und langer Read-Länge.....	33
A.2	Empfehlung für die Qualitätsbeurteilung von Short-Read-Daten.....	35
A.3	Empfohlene Verwendung von Kontrollen.....	36
Anhang B (informativ)	Felder für Labor-Kontaktinformationen.....	38
Anhang C (informativ)	Geographischer Standort der Probenahmefelder.....	40
Anhang D (informativ)	Verlaufsfelder der Isolat-Passage.....	41
Anhang E (informativ)	Ergebnisse des Antibiotogramms und Verfahrensfelder.....	42
Anhang F (informativ)	Detektion von Virulenzfaktoren und Verfahrensfelder.....	44
Anhang G (informativ)	Metriken zur Qualitätskontrolle der Sequenz.....	45
Anhang H (informativ)	Spezifikation der Metadaten.....	46
H.1	Spezifikation der Metadaten für Einreichungen bei NCBI/DDBJ.....	46
H.2	Spezifikation der Metadaten für Einreichungen bei EBI.....	48
Anhang I (informativ)	Anweisungen für die Integration der Ontologie Slim durch Softwareentwickler.....	50
I.1	Einleitung.....	50
I.2	Ratschläge für die Implementierung der ISO WGS Slim.....	51
I.2.1	Datentabellen für die Integration von Ontologien.....	51
I.2.2	Integration von Ontologien in einer SQL-Datenbank.....	52
I.3	Ansätze zur Integration von Ontologien in Systeme zur Erfassung und Verwaltung von Metadaten.....	52
I.3.1	Datentabellen.....	52
I.3.2	Integration von Ontologien in einer SQL-Datenbank.....	53
Literaturhinweise.....		55

Europäisches Vorwort

Dieses Dokument (prEN ISO 23418:2020) wurde vom Technischen Komitee ISO/TC 34 „Food products“ in Zusammenarbeit mit dem Technischen Komitee CEN/TC 463 „Mikrobiologie der Lebensmittelkette“ erarbeitet, dessen Sekretariat von AFNOR gehalten wird.

Dieses Dokument ist derzeit zur parallelen Umfrage vorgelegt.

Anerkennungsnotiz

Der Text von ISO/DIS 23418:2020 wurde von CEN als prEN ISO 23418:2020 ohne irgendeine Abänderung genehmigt.

Vorwort

ISO (die Internationale Organisation für Normung) ist eine weltweite Vereinigung nationaler Normungsinstitute (ISO-Mitgliedsorganisationen). Die Erstellung von Internationalen Normen wird üblicherweise von Technischen Komitees von ISO durchgeführt. Jede Mitgliedsorganisation, die Interesse an einem Thema hat, für welches ein Technisches Komitee gegründet wurde, hat das Recht, in diesem Komitee vertreten zu sein. Internationale staatliche und nichtstaatliche Organisationen, die in engem Kontakt mit ISO stehen, nehmen ebenfalls an der Arbeit teil. ISO arbeitet bei allen elektrotechnischen Normungsthemen eng mit der Internationalen Elektrotechnischen Kommission (IEC) zusammen.

Die Verfahren, die bei der Entwicklung dieses Dokuments angewendet wurden und die für die weitere Pflege vorgesehen sind, werden in den ISO/IEC-Direktiven, Teil 1 beschrieben. Es sollten insbesondere die unterschiedlichen Annahmekriterien für die verschiedenen ISO-Dokumentenarten beachtet werden. Dieses Dokument wurde in Übereinstimmung mit den Gestaltungsregeln der ISO/IEC-Direktiven, Teil 2 erarbeitet (siehe www.iso.org/directives).

Es wird auf die Möglichkeit hingewiesen, dass einige Elemente dieses Dokuments Patentrechte berühren können. ISO ist nicht dafür verantwortlich, einige oder alle diesbezüglichen Patentrechte zu identifizieren. Details zu allen während der Entwicklung des Dokuments identifizierten Patentrechten finden sich in der Einleitung und/oder in der ISO-Liste der erhaltenen Patenterklärungen (siehe www.iso.org/patents).

Jeder in diesem Dokument verwendete Handelsname dient nur zur Unterrichtung der Anwender und bedeutet keine Anerkennung.

Für eine Erläuterung des freiwilligen Charakters von Normen, der Bedeutung ISO-spezifischer Begriffe und Ausdrücke in Bezug auf Konformitätsbewertungen sowie Informationen darüber, wie ISO die Grundsätze der Welthandelsorganisation (WTO, en: World Trade Organization) hinsichtlich technischer Handelshemmnisse (TBT, en: Technical Barriers to Trade) berücksichtigt, siehe www.iso.org/iso/foreword.html.

Dieses Dokument wurde vom Technischen Komitee ISO/TC 34, *Food Products*, Unterkomitee SC 9, *Microbiology* erarbeitet.

Rückmeldungen oder Fragen zu diesem Dokument sollten an das jeweilige nationale Normungsinstitut des Anwenders gerichtet werden. Eine vollständige Auflistung dieser Institute ist unter www.iso.org/members.html zu finden.

Einleitung

Next generation sequencing (NGS) bietet schnellen, kostengünstigen Zugang zu mikrobiellen Gesamtgenomsequenzen (WGS, en: Whole Genome Sequencing) im Hochdurchsatzverfahren und wird bei einer wachsenden Zahl von Problemen in der Mikrobiologie von Lebensmitteln angewandt. WGS sind digitale Darstellungen des biologischen Potentials zur Sequenzierung des Organismus bei Einzelbasenauflösung. Die digitale Beschaffenheit der WGS-Daten ist eine Abkehr von der kontinuierlichen Variation der Phänotypen, die routinemäßig in der Mikrobiologie von Lebensmitteln analysiert werden. Daher bietet WGS erhebliche Vorteile gegenüber bestehenden Technologien (z. B. Serologie, Pulsfeld-Gelelektrophorese, phänotypische Antibiotikaresistenz). WGS-basierte Analysen werden von Laboren für öffentliche Gesundheit verwendet, um Ausbrüche zu erkennen und Mutationen, Gene und andere genetische Merkmale zur Charakterisierung der Virulenz und des Überlebenspotentials nachzuweisen. Innerhalb der Lebensmittelindustrie besteht Interesse an WGS, um bakterielle Isolate aus ausgegliederten Bestandteilen und umgebenden Flächen zu charakterisieren, ihre Herkunft und Ökologie besser zu verstehen und die Verfahren zur Risikominderung zu aktualisieren. Einige Unternehmen haben die Kapazitäten zur Erfassung und Analyse von WGS-Daten geschaffen oder schaffen diese gerade. Andere werden sich an Drittlabore wenden, um diese Dienstleistungen zu erbringen, wie sie es derzeit für andere mikrobiologische Analysen tun.

Gestrichener Text

Diese Norm soll sowohl für das Labor als auch für die bioinformatischen Komponenten der WGS und die zugehörigen Metadaten für lebensmittelbedingte Mikroorganismen als Leitfaden dienen. Diese Norm soll auf alle derzeit verfügbaren Technologien zur DNA-Sequenzierung mit kurzen und langen Reads anwendbar sein. Sie darf zur Analyse von WGS-Daten mit proprietärer, Open Source- und kundenspezifischer Software eingesetzt werden. Sie ist nicht zur Festlegung der Sequenzierungsschemikalien, analytischen Verfahren oder Software vorgesehen. Die Norm definiert Labor-, Daten- und Metadaten-Verwaltungspraktiken, um sicherzustellen, dass Analysen eindeutig angegeben werden, transparent und für Anfragen zugänglich sind sowie für unvorhergesehene Zwecke zur Verfügung stehen. Diese Norm soll Labore in die Lage versetzen, ihre Managementsysteme für Qualität und technische Abläufe zu entwickeln. Kunden von Laboren und Aufsichtsbehörden dürfen sie auch zur Anerkennung und Bestätigung der Kompetenz von Laboren nutzen.

1 Anwendungsbereich

Diese Internationale Norm legt Mindestanforderungen für die Generierung und Analyse von Daten der Gesamtgenomsequenzierung (WGS) fest, die aus pathogenen Mikroorganismen aus Lebensmitteln gewonnen werden. Diese Anforderungen gelten für jede Sequenzierplattform oder -chemie. Dieser Prozess kann die folgenden Phasen umfassen:

- a) Umgang mit Bakterienkulturen;
- b) Isolierung der genomischen DNA;
- c) Vorbereitung der genomischen DNA-Bibliothek, Sequenzierung und Beurteilung der Read-Qualität von Rohdaten und deren Speicherung;
- d) bioinformatische Analyse zur Bestimmung der genetischen Verwandtschaft, des genetischen Gehalts und zur Vorhersage des Phänotyps sowie Validierung der bioinformatischen Pipeline;
- e) die Erfassung von Metadaten und die Hinterlegung von Sequenzdaten in Repositorien und
- f) Validierung des durchgehenden WGS-Workflows (gebrauchstauglich für die beabsichtigte Anwendung).

2 Normative Verweisungen

Es gibt keine normativen Verweisungen in diesem Dokument.

3 Begriffe

Für die Anwendung dieses Dokuments gelten die folgenden Begriffe.

ISO und IEC stellen terminologische Datenbanken für die Verwendung in der Normung unter den folgenden Adressen bereit:

- ISO Online Browsing Platform: verfügbar unter <https://www.iso.org/obp>
- IEC Electropedia: verfügbar unter <http://www.electropedia.org/>

3.1

Adaptersequenz

DNA mit einer bekannten Sequenz, die an das Ende eines DNA-Bibliotheksfragments angefügt wird, um den Sequenzierungsprozess zu erleichtern (z. B. Annealing an eine Flusszelle)

3.2

Annotation

Prozess der Identifizierung von Genen und anderen Merkmalen der Genomassemblierung

3.3

Antibiogramm

Zusammenfassung der Ergebnisse antimikrobieller Anfälligkeitsprüfungen, die für einen spezifischen Mikroorganismus durchgeführt wurden, üblicherweise in tabellarischer Form

3.4

Assembly

Ergebnis aus dem Prozess des Alignments und der Zusammenführung von Sequenzabschnitten (Reads) zu größeren zusammenhängenden Sequenzen (Contigs)

3.5

Basenaufruf

Prozess, bei dem Nukleotide und Qualitätswerte Positionen in den Sequenzabschnitten zugewiesen werden

3.6

Bioinformatik

Erfassung, Speicherung und Analyse biologischer Sequenzdaten

3.7

bioinformatische Pipeline

einzelne Programme, Skripte oder miteinander verknüpfte Softwarekomponenten, wobei die Ausgabe eines Programms als Eingabe für den nächsten Schritt der Datenverarbeitung verwendet wird

3.8

carry-over-Kontamination

Proben, die mit DNA aus zuvor sequenzierten Proben oder Stoffen, einschließlich EDTA, Phenol-Chloroform, Protein, überschüssigen Salzen, kontaminiert sind

3.9

Chemical Entities of Biological Interest Ontology

ChEBI (en: Chemical Entities of Biological Interest Ontology)

Ontologie zur Beschreibung kleiner chemischer Verbindungen

3.10

Contig

zusammenhängendes Stück einer DNA-Sequenz, die sich aus der Assemblierung kleinerer, überlappender DNA-Reads ergibt

3.11

kontrolliertes Vokabular

endlicher Vorrat von Werten, die die einzig zulässigen Werte für ein Datenelement darstellen

[QUELLE: ISO 11238:2018 (en)]

3.12

Abdeckung

durchschnittliche Häufigkeit, die jede Base in einem Genom sequenziert wird

3.13

Kreuzkontamination

Kontamination einer Probe (Bakterienisolat oder DNA) mit anderen Proben

3.14

DNA-Qualität

Angabe der DNA-Reinheit (frei von Polysacchariden, Kontaminanten und Enzyminhibitoren) und Integrität (hohes Molekulargewicht mit geringen bis gar keinen Anzeichen auf Abbau)

3.15

DNA-Probe

Teil einer DNA, die aus einem Material extrahiert wird

3.16

Draft-Assemblierung

de novo-Assemblierung von Genomen, bestehend aus Contigs ohne implizite Reihenfolge, die üblicherweise mit Hilfe der Gesamtgenom-Schrotschuss-Sequenzierung mit einer Short-Read-Technologie erzeugt werden

3.17

Umweltontologie

EnvO (en: Environment Ontology)

Ontologie zur Beschreibung von Umweltmerkmalen und Lebensräumen

3.18

FoodEx2-Ontologie

FoodEx2

genormtes System zur Klassifizierung und Beschreibung von Lebensmitteln, entwickelt von der Europäischen Behörde für Lebensmittelsicherheit (EFSA)

3.19

Lebensmittel-Ontologie

FoodOn (en: Food Ontology)

Ontologie zur Beschreibung von Lebensmittelprodukten, Tierfutter und Lebensmittelverarbeitung

3.20

ontologisches Gazetteer (Ortsverzeichnis)

GAZ

Ontologie zur Beschreibung geographischer Orte

3.21

Index

Oligonukleotidsequenzen, die im Prozess der Bibliotheksvorbereitung verwendet werden, um DNA aus spezifischen Proben zu markieren oder mit einem Barcode zu versehen, so dass mehrere Proben in einer Sequenzierungsreaktion kombiniert (gemultiplext) werden können

3.22

Internationale Nukleotidsequenz-Datenbank-Zusammenarbeit

INSDC (en: International Nucleotide Sequence Database Collaboration)

Initiative, die von der DNA-Datenbank Japans (DDBJ), dem Europäischen Laboratorium für Molekularbiologie, dem Europäischen Institut für Bioinformatik (EMBL-EBI) und dem Nationalen Zentrum für Biotechnologieinformation (NCBI) betrieben wird

3.23

ISO WGS-Slim

zusammenfassende Ontologie mit interoperablen Feldern und Begriffen bezüglich der Verwendung von WGS für die Mikrobiologie von Lebensmitteln

3.24

Isolat

Population von Bakterienzellen in Reinkultur, die aus einer einzigen Kolonie stammen

3.25

k-mere

alle möglichen Sequenzen der Länge k, die in einer ganzen Genomsequenz enthalten sind

3.26

Bibliothek

Sammlung von genomischen DNA-Fragmenten aus einem einzigen Isolat zur Bestimmung der Genomsequenz

3.27

Managementsystem

Qualitäts-, Verwaltungs- und technische Systeme, die den Betrieb einer Organisation regeln

Anmerkung 1 zum Begriff: In diesem Dokument bezieht sich „Organisation“ auf das „Labor“.

3.28

Mapping

Verwendung von Software für das Alignment von Sequenzabschnitten mit Referenzsequenzen

3.29

Metadaten

Daten, die andere Daten beschreiben und definieren

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

3.30

minimale Daten zum Abgleich

MDM (en: minimal data for matching)

Informationen, die zur Beschreibung der Probenquelle und -herkunft einer genomischen Sequenz erforderlich sind, wie durch den Global Microbial Identifier [10] definiert und durch die Internationale Nukleotidsequenz-Datenbank-Zusammenarbeit implementiert

3.31

minimale Hemmkonzentration

MHK

geringste Konzentration, die unter definierten *In-vitro*-Prüfbedingungen in einem vorgegebenen Zeitintervall das Wachstum um einen vereinbarten Betrag verringert

Anmerkung 1 zum Begriff: Die MHK wird in mg/l angegeben.

[QUELLE: ISO 16256:2012 (en)]

3.32

Multilocus-Sequenztypisierung

MLST (en: multi-locus sequence typing)

Verfahren der Genomanalyse, bei der Nukleotidvarianten innerhalb vordefinierter Sätze von Loci, entweder Kerngenomloci für cgMLST oder ganze Genomloci für wgMLST, identifiziert werden

3.33

N50

Länge (N), so dass Contigs von N oder länger die Hälfte der Basen in der Assembly umfassen

3.34

NCBITaxon Ontology

NCBITaxon

automatische Übersetzung der NCBI-Taxonomiedatenbank

3.35

NG50

Länge (N) der DNA, so dass Contigs von N oder länger die Hälfte der Basen im Genom umfassen

3.36

Open Biological and Biomedical Ontology Foundry

OBO Foundry

Sammlung von Ontologien, die von einer Gemeinschaft von Ontologie-Entwicklern erstellt wurden und die sich zur Zusammenarbeit und zur Einhaltung gemeinsamer Grundsätze verpflichtet haben

3.37

Ontologie

kontrolliertes Vokabular, das in einer Hierarchie angeordnet ist, wobei die Begriffe durch logische Beziehungen miteinander verbunden sind

3.38

Ontologie-Slim

Reihe von Ontologiefeldern und -begriffen, die als Teil einer bestimmten Sammlung kommentiert werden, oft zu einem bestimmten Zweck, und die extrahiert werden können, um eine Datei zu erstellen, die sich von der ursprünglichen Ontologie unterscheidet

3.39

Phred-Qualitätswert (Q) der Sequenz

Maß für die Wahrscheinlichkeit, dass eine Base an einer bestimmten Position falsch zugeordnet wird in der Sequenz, die ausgedrückt wird als:

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

Anmerkung 1 zum Begriff: Ein Wert von Q30 zeigt an, dass eine Wahrscheinlichkeit von 1 zu 1 000 besteht, dass eine Base falsch zugeordnet ist (d. h. der Basenaufbau ist zu 99,9 % richtig).

3.40

Read

aus einem DNA- oder RNA-Fragment abgeleitete Nukleotidsequenz

3.41

Sequenz-Repository

Datenbank, in der WGS-Datensätze gespeichert und verwaltet werden

Anmerkung 1 zum Begriff: Ein öffentliches Repository ermöglicht den uneingeschränkten Zugang zu den Daten, während ein privates Repository oder eines im Verbund den Zugang zu den Daten einschränkt.

3.42

Parallelproben der Sequenzierung, biologisch

Sequenzierung einer anderen Kolonie aus demselben Isolat, das aus demselben Probenmaterial gewonnen wurde, um die biologische Variation zu bewerten

3.43

Parallelproben der Sequenzierung, technisch

Resequenzierung derselben biologischen Probe oder Bibliothek zur Bewertung der Sequenzvariation aufgrund der Instrumentierung und des Protokolls

3.44

Serotyp

Klassifizierungsschema auf der Grundlage des antigenen Nachweises oder des sequenzbasierten Nachweises von Genen, die Bakterienoberflächenmoleküle kodieren

3.45

Einzelnukleotid-Polymorphismus

SNP (en: Single Nucleotide Polymorphism)

SNV, die eine bestimmte Qualitäts- und/oder Frequenzschwelle überschreitet

3.46

Einzelnukleotid-Variante

SNV (en: Single Nucleotide Variant)

Unterschiede zwischen den Nukleotidzuständen an der gleichen genomischen Position von zwei oder mehr Isolaten

3.47

Stamm

Nachkommen einer einzelnen Isolierung in Reinkultur, die in der Regel von einer einzigen Ausgangskolonie auf einem festen Wachstumsmedium abstammen [1]

Anmerkung 1 zum Begriff: Ein Stamm darf als ein Isolat oder eine Gruppe von Isolaten angesehen werden, die sich von anderen Isolaten derselben Gattung und Art durch phänotypische und genotypische Merkmale unterscheiden lassen.

3.48

Validierung

Ermittlung der Leistungskenngrößen eines Verfahrens und Bereitstellung eines objektiven Nachweises, dass die Leistungsanforderungen für eine festgelegte vorgesehene Anwendung erfüllt sind

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

3.49

validierte Dateneingabe

automatisierter Prozess, der sicherstellt, dass die in ein Repository eingegebenen Daten korrekt sind

3.50

Verifizierung

Nachweis, dass ein validiertes Verfahren bei der Anwendung durch den Anwender nach den in der *Validierungsuntersuchung* (3.48) bestimmten Verfahrensspezifikationen funktioniert und gebrauchstauglich ist

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

3.51

Gesamtgenomsequenzierung

WGS (en: Whole Genome Sequencing)

Prozess zur Bestimmung der DNA-Sequenz des Genoms eines Organismus unter Verwendung einer genomischen Gesamt-DNA als Eingabe

4 Kurzbeschreibung

4.1 Allgemeines

Jede Organisation, die Proben handhabt, Sequenzierungen oder bioinformatische Analysen für die WGS-Analyse durchführt, muss durch die Bereitstellung von Nachweisen belegen, dass eine ordnungsgemäße Dokumentation der Probenherkunft, der Verfahren und der Qualitätskontrolle erfasst und zur Nachverfolgung aufbewahrt wird.

Die WGS-Analyse pathogener Mikroorganismen aus Lebensmitteln besteht aus einer Bakterienkultur, einer DNA-Isolierung in einem mikrobiologischen Labor, Sequenzierungsschritten in einer Sequenzierungseinrichtung und bioinformatischen Analysen, die in einer speziellen Computerumgebung durchgeführt werden.

4.2 Laborbetrieb: Probenvorbereitung und Sequenzierung

Die Probenvorbereitung und Sequenzierung sollten die folgenden Schritte umfassen:

- a) Informationen über die zu sequenzierenden Isolate, einschließlich Barcodes für Multiplex-Proben, werden in die entsprechenden Aufzeichnungssysteme, wie z. B. ein Laborinformationsmanagementsystem (LIMS) und/oder Arbeitsblätter zur Probenbeschreibung, eingegeben.

- b) Genomische DNA wird aus Reinkulturen extrahiert, und idealerweise wird die Artidentität bestätigt.
- c) Aus der genomischen DNA-Extraktion werden DNA-Bibliotheken hergestellt. Dieser Prozess sollte Folgendes umfassen:
 - i) DNA-Fragmentierung;
 - ii) Ligation von Indizes und Adaptern;
 - iii) Quantifizierung, Normalisierung und Qualitätskontrolle der resultierenden Bibliothek; und
 - iv) Pooling von Bibliotheken für Multiplex-Sequenzierungsläufe.
- d) Die Bibliotheken sind sequenziert.
- e) Die vom Sequenzierungsinstrument erzeugten Qualitätsmetriken werden für jeden Lauf aufgezeichnet.

4.3 Bioinformatische Analyse

Pipelines für bioinformatische Analysen können sich auf *In-silico*-Vorhersagen des Phänotyps (z. B. Virulenz) oder auf den Nachweis von Clustern genetisch ähnlicher Isolate (d. h. gleicher Stamm, Sequenztyp oder Serotyp) konzentrieren. Pipelines, die auf vergleichenden Ansätzen basieren, können verwendet werden, um das Vorhandensein und die Zustände von Markern in rohen und assemblierten Sequenzierungsdaten zu erkennen, um *In-silico*-Stamm- (z. B. Sequenztyp) und Phänotypvorhersagen zu machen.

Sequenzdaten für mehrere Isolate können mit SNP-, MLST- oder k-mer-Abstands-Analyseverfahren analysiert werden, um Cluster eng verwandter Bakterien zu identifizieren. Die Ergebnisse dieser Analysen können verwendet werden, um auf Beziehungen zwischen Isolaten zu schließen, die mit phylogenetischen Bäumen und Dendrogrammen veranschaulicht werden können.

a) SNP-Analysen

Bei SNP-Analysen werden Reads auf eine Referenzsequenz abgebildet oder Reads zu Contigs assembliert, die verglichen werden. Zur Bestimmung von SNPs werden SNVs qualitätsgefiltert, um SNP-Positionen zu identifizieren.

b) MLST-Analysen

Für MLST-Analysen werden die Reads assembliert oder abgebildet. Die Zielloci werden identifiziert, qualitätsgefiltert und mit einer kuratierten cgMLST- oder wgMLST-Datenbank verglichen.

c) Analyse des k-mer-Abstandes

Sequenzdaten für mehrere Isolate können mit k-mer-Abstandsverfahren analysiert werden, um Cluster verwandter Bakterien zu identifizieren. K-mer-Analysen haben den Vorteil, dass sie sehr schnell sind, weisen aber einige Einschränkungen auf, insbesondere in Bezug auf die Präzision. (D. h. sie sind bei der Artbestimmung anwendbar, werden aber nicht für detaillierte Quellennachweisanalysen von eng verwandten Stämmen empfohlen.)

4.4 Metadatenformate und Hinterlegung von Sequenzdaten in Repositorien

Für alle Sequenzen müssen Metadatenätze erstellt und sicher gespeichert werden. Sequenzdaten und entsprechende Metadaten sollten konsistent formatiert und dokumentiert werden. Diese Metadaten können nach alleinigem Ermessen des Eigentümers der Metadaten freigegeben werden. Daten und die entsprechenden Metadaten müssen Sicherheitserwägungen, Kosten und Nutzen, gesetzlicher Haftung,

geistigen Eigentumsrechten, vertraulichen Geschäftsinformationen, Vertragsbeschränkungen oder anderen verbindlichen schriftlichen Vereinbarungen unterliegen.

Zur Förderung von bewährten Verfahren für die Datenverwaltung [3] bietet diese Norm optionale Metadaten-Berichtsformate, die auf einen gemeinschaftlichen Datenstandard abgestimmt sind (z. B. Ontologien für MDM oder OBO Foundry). Diese Formate und Normen erleichtern die Reproduzierbarkeit und das gemeinsame Verständnis der Terminologie. Zur Formatierung und Bereitstellung von Werten für die empfohlenen Metadatenfelder wurde eine ISO WGS Slim erstellt. WGS und ausgewählte Metadaten können in eine öffentlich zugängliche Datenbank übertragen (hochgeladen) werden.

4.5 Validierung und Verifizierung des WGS-Workflows

Der gesamte WGS-Workflow muss validiert werden, um sicherzustellen, dass die Verfahren für den bestimmungsgemäßen Gebrauch geeignet sind.

5 Allgemeiner Laborleitfaden

5.1 Bakterienisolierung und DNA-Extraktion

Die Bakterienisolierung und DNA-Extraktion sollte in einem allgemeinen mikrobiologischen Labor durchgeführt werden, das für die Arbeit mit den spezifischen Bakterien, einschließlich Krankheitserregern, geeignet ist. Bei der Vorbereitung einer Sequenzierungsbibliothek, die eine DNA-Amplifikation mittels Polymerase-Kettenreaktion (PCR) beinhaltet, sollten Vor- und Nach-PCR-Schritte in verschiedenen oder getrennten Bereichen des Labors durchgeführt werden, um eine carry-over-Kontamination zu vermeiden.

5.2 Laborumgebung

Luftbewegungen, Vibrationen, Temperatur und Feuchtigkeit können die Leistung vieler Sequenzer beeinträchtigen und sollten bei der Platzierung der Geräte im Labor berücksichtigt werden. Labore sollten den Leitfaden des Sequenziergeräteherstellers zur Vorbereitung des Standorts als Orientierungshilfe konsultieren.

5.3 Standardverfahrensanweisungen (SOP, en: standard operating procedure) und nicht konforme Arbeit

Labore sollten Standardverfahrensanweisungen (SOP), Arbeitsablaufdokumente, Bestandskontrollen von Reagenzien und Gerätewartungsprotokolle pflegen und einhalten. SOPs sollten Verfahren für die Verwendung von Positiv- und Negativkontrollen für die DNA-Extraktion, die Vorbereitung der Sequenzbibliothek und die Sequenzierungsschritte enthalten. SOPs sollten Verfahren zur Überwachung von Arbeitsvorgängen auf Ausführungsqualität und Fehler (falsche Kennzeichnung von Proben oder Kreuzkontamination) enthalten.

Im Falle einer falschen Kennzeichnung oder Kontamination von Proben muss die Grundursache für Fehler in der Sequenzierung untersucht werden:

- i) Sicherstellung, dass Läufe, die falsch gekennzeichnete Proben oder mit mehreren Stämmen kontaminierte Proben enthalten, nicht für bioinformatische Analysen verwendet oder in Datenbanken hochgeladen werden; und
- ii) Umsetzung von Maßnahmen zur Aufrechterhaltung der Qualität und zur Verhinderung des erneuten Auftretens von Fehlern.

ANMERKUNG Falsche Kennzeichnung und/oder Kontamination von Proben oberhalb von Schwellenwerten, die bestimmt wurden, um die Ergebnisse bioinformatischer Analysen zu beeinflussen, sind kritische Fehler, die eine weitere Verwendung der betroffenen Daten untersagen sollten.

5.4 Labor-Informations- und Management-System (LIMS)

Die Informationen zu Proben müssen mit einem LIMS oder einem ähnlichen System zur Dokumentation und Verfolgung von Informationen erfasst werden.

5.5 Kompetenz im Labor

Labore sollten Aufzeichnungen führen, die die Aus- und Fortbildung sowie die Eignung der Personen dokumentieren, die Sequenzierungs- und bioinformatische Analysen durchführen, sowie Richtlinien zur Aufbewahrung von Proben.

Das Labor sollte seine Leistung durch Vergleich mit den Ergebnissen anderer Labore überwachen, sofern diese verfügbar und angemessen sind. Diese Überwachung sollte geplant und überprüft werden und u. a. eine der folgenden Maßnahmen umfassen:

- i) jährliche Teilnahme an einem Eignungsprüfungsprogramm;
- ii) jährliche Teilnahme an Vergleichsprüfungen zwischen Laboren mit Ausnahme von Eignungsprüfungen;
- iii) jährliche Verifizierung des Analyseprozesses durch Einführung von Blindproben oder Proben, deren Merkmale dem Bearbeiter nicht bekannt sind.

Die Daten aus diesen Überwachungstätigkeiten sollten analysiert, zur Kontrolle und gegebenenfalls zur Verbesserung der Tätigkeit des Labors verwendet werden. Wenn sich herausstellt, dass die Ergebnisse der Analyse von Daten aus diesen Überwachungstätigkeiten außerhalb der vordefinierten Kriterien liegen, sollten geeignete Maßnahmen ergriffen werden, um zu verhindern, dass falsche Ergebnisse verwendet werden.

6 Laborbetrieb

6.1 Herstellung und Lagerung der Proben

Jegliches zu sequenzierende Material (bakterielles Isolat oder extrahierte genomische DNA) sollte so gehandhabt und gelagert werden, dass das Risiko des Probenabbaus, der falschen Kennzeichnung und Kreuzkontamination minimiert wird.

6.2 Bakterienisolate

Bakterienisolate sollten mit Verfahren gelagert und kultiviert werden, die das Potential für die Einführung genetischer Veränderungen (z. B. Verlust von Plasmiden oder Polymorphismen, die durch die Kultur und Passage eingeführt werden) minimieren. Bakterienisolate sollten wegen der Möglichkeit der Einführung genetischer Veränderungen nicht länger als zwei Wochen auf Agarplatten oder im Stabformat (bei 4 °C bis 25 °C) gelagert werden. Wenn Proben in dieser Form länger als zwei Wochen aufbewahrt werden, sollte dies in den Metadaten der Probe vermerkt werden. Wenn das Labor ein Bakterienisolat erhält, muss das Labor die Reinheit des Isolats sicherstellen und idealerweise die Spezies bestätigen, bevor weitere Schritte durchgeführt werden. Jede Probe sollte nicht mehr als fünf einzelnen Koloniepässagen unterzogen werden, um die Einführung genetischer Veränderungen und Engpässe zu vermeiden. Wenn die Sorge besteht, dass potentiell instabile Elemente (z. B. Plasmide) während der Passage aus einer Probe verloren gehen könnten, sollten Sequenzen von zwei oder drei biologischen Parallelproben gesammelt werden. Die Anzahl der einzelnen Koloniepässagen sollte vermerkt werden, wenn sie in den Metadaten der Probe größer als fünf ist. Bakterienisolate sollten mit Verfahren wie dem Einfrieren als Glycerinvorrat bei -80 °C archiviert werden.

6.3 Isolierung der DNA

Wenn eine Isolierung bakterieller DNA erforderlich ist, sollte das Extraktionsverfahren für die Herstellung von DNA geeignet sein, die für die Organismen und die verwendete Sequenzierplattform geeignet ist. Die Isolierung bakterieller DNA wird durch eine Reihe von Faktoren beeinflusst, darunter Zellart (grampositiv oder -negativ), Wachstumsphase (früh-, mittel-, spät-logarithmisch oder -stationär) und Nährmedium. Die Quantität und Qualität der DNA sollte beurteilt und dokumentiert werden.

ANMERKUNG Die Lagerungsbedingungen werden die DNA-Integrität und die Vorbereitung der Bibliothek für bestimmte Sequenziertechnologien beeinflussen. Verschiedene Sequenzierplattformen haben unterschiedliche Anforderungen an DNA-Quantität, -Qualität und -Integrität. Wenn ein Isolierungskit verwendet wird, kann die DNA bis zum Abschluss der Analyse nach den Anweisungen des Herstellers aufbewahrt werden. Die Vorbereitung von Einweg-DNA-Aliquoten kann dazu beitragen, Kreuzkontaminationen und wiederholtes Auftauen der Probe zu verhindern.

6.4 Vorbereitung der Bibliothek

Das Labor sollte das vom Hersteller empfohlene Protokoll befolgen. Die Verfahren dürfen für spezifische Bedürfnisse angepasst werden, aber alle Änderungen müssen vollständig dokumentiert und validiert werden.

ANMERKUNG 1 Größenauswahlverfahren, die bei einigen Verfahren zur Vorbereitung der Bibliothek verwendet werden (z. B. beim Aufbau von Einzelmolekül-Echtzeitbibliotheken mit großformatigem Einsatz (> 2 kb)), können zum Verlust kleiner Plasmide führen.

ANMERKUNG 2 Bei der PCR-Anreicherung von Bibliotheken ist Vorsicht geboten, da dies unter Umständen zu einer Verringerung der Komplexität der Bibliothek und zu einer Verringerung der Anzahl der verschiedenen DNA-Moleküle in der Vorbereitung führen kann. Die Bibliothekskomplexität kann auch durch die Menge an DNA-Ausgangsmaterial oder die Menge an DNA, die während der Aufreinigungsschritte der Bibliotheksvorbereitung verloren geht, beeinflusst werden. Die Komplexität von Bibliotheken kann mit dem Verfahren von Daley und Smith geschätzt werden [2].

Besteht die Möglichkeit, dass Bibliotheken wieder verwendet werden, so sind die Bibliotheken nach den Empfehlungen des Herstellers aufzubewahren. Das Labor muss das verwendete Rückverfolgbarkeitsverfahren (d. h. Barcode oder gleichwertig), die verwendete Sequenzierungsplattform, den Betreiber, der die Bibliothek hergestellt hat, das Datum der Herstellung der Bibliothek und die Chargeninformationen für den/die verwendeten Kit(s) dokumentieren. Das Multiplexen von Proben erfordert die Auswahl und Zuweisung von Barcodes zur Identifizierung einzelner Proben und wird üblicherweise in einem Arbeitsblatt dokumentiert, um die Zuordnung von Sequenzdaten zu den richtigen Metadaten zu ermöglichen. Wenn alle gemultiplexten Proben derselben Bakteriengattung angehören (z. B. alle Salmonellen), sollten Schritte unternommen werden, um sicherzustellen, dass äquimolare DNA-Eingaben verwendet werden (d. h. Bibliotheksnormalisierung) und dass die korrekte Sequenz mit den entsprechenden Metadaten verknüpft ist. Wenn die gemultiplexten Proben mehrere Gattungen repräsentieren, müssen die geschätzte Erfassung, die Genomgröße und die Größe der Bibliotheksfragmente bei der Schätzung der für jede Probe aufzunehmenden DNA-Menge berücksichtigt werden.

6.4.1 DNA-Sequenzierung

Sequenzierinstrumente müssen nach den Empfehlungen des Herstellers oder anderer Experten betrieben und gewartet werden, und die Dokumentation der Instandhaltungsverfahren muss gepflegt werden. Plattformspezifische Sequenzierungsmetriken (z. B. Clusterdichte, Anzahl der Reads, durchschnittliche Basenqualität usw.) müssen für jeden Sequenzierungslauf aufgezeichnet und überwacht werden.

ANMERKUNG Plattformspezifische Empfehlungen zur Minimierung der carry-over-Kontamination sind in Anhang A1 enthalten.

6.4.2 Verwendung von Kontrollen

Beim Umgang mit einem Bakterienisolat und einem DNA-Extrakt sollte das Labor eine Wasserleerprobe oder eine unbeimpfte Bouillon als Negativkontrolle während der DNA-Extraktion verwenden, um eine mögliche Kreuzkontamination zu beurteilen. Eine positive Extraktionskontrolle zur Beurteilung der Effizienz des Verfahrens kann nach Bedarf einbezogen werden. Wenn die Vorbereitung der Bibliothek Multiplexing und PCR-Amplifikationsschritte umfasst, wird dringend empfohlen, sowohl Positiv- als auch Negativkontrollen einzuschließen. Es wird auch empfohlen, stets denselben DNA-Extrakt für die Positivkontrolle zu verwenden, um Vergleiche der Sequenzierungsqualität von Lauf zu Lauf zu ermöglichen.

ANMERKUNG Empfehlungen für die Verwendung von Positiv- und Negativkontrollen sind in Anhang A3 enthalten.

6.4.3 Beurteilung der Qualität von Read-Rohdaten

Der Basenaufwurf sollte mit einer für das Instrument und die Sequenzierchemie spezifischen Software durchgeführt werden. Metriken können auf Lafebene und auf Probenebene definiert werden. Metriken müssen dokumentiert werden, um die Qualität der Sequenzrohdaten zu bewerten. Dazu könnten das Einsatzformat, die Verteilung der Sequenzlänge, die Anzahl der Reads und eine Beurteilung der Basenzusammensetzung (d. h. AT/GC-Bilanz oder TAGC (taxon annotated GC-coverage)-Plot oder -Äquivalent) gehören. Qualitätswerte, Read-Länge und Taxonomieprüfung sollten als erste Überprüfung der Sequenzierungsleistung verwendet werden (siehe auch Teil 7.3). Die durch Mapping errechnete Erfassung, die auf eine de novo-Assemblierung oder ein geeignetes Referenzgenom zurückgelesen wird, sollte ebenfalls bewertet werden.

Die Qualität und Quantität der DNA-Sequenz-Lesequalität und -quantität wirkt sich auf die nachgelagerte Assemblierung, die Lesekartierung und die Fähigkeit zur Nutzung von WGS-Daten für die bakterielle Herkunftsverfolgung und Genomcharakterisierung aus. Zu den Sequenzierungsartefakten, die sich auf nachgelagerte Analysen auswirken können, gehören unter anderem plattformspezifische Fehlerprofile, Schwankungen in den Qualitätswerten über die Sequenzabschnitte, Verzerrungen der Sequenzdaten aufgrund der Basenzusammensetzung, Abweichungen von der optimalen Größe der Bibliotheksfragmente und Kontamination durch bekannte und unbekannte Spezies, die nicht das Sequenzierungsziel darstellen.

ANMERKUNG Allgemeiner Leitfaden für die Entwicklung von Qualitätsmetriken sind in den Anhängen A1 und A2 enthalten.

6.4.4 Speicherung und Aufbewahrung von Proben und Daten

Das Labor muss eine Richtlinie für die Lagerung und Aufbewahrung von Prüfkörpern, DNA-Proben, DNA-Bibliotheken und Sequenzierungsdaten dokumentieren.

7 Bioinformatische Datenanalyse

7.1 Anforderungen an Software und/oder bioinformatische Pipelines, die für die Datenanalyse verwendet werden

Bioinformatische Pipelines sollten vor der Verwendung für die Datenanalyse validiert werden (siehe Abschnitt 10.1.2). Pipeline-Entwickler sollten Prüfdatensätze mit ihrer Software vertreiben. Anwender sollten sicherstellen, dass die Pipelines korrekt installiert sind, indem sie die Prüfdatensätze analysieren und prüfen, ob die erwarteten Ergebnisse erzeugt werden. Die Hauptkomponenten von bioinformatischen Pipelines (z. B. Aligner, Variantenaufreifer, Open-Reading-Frame-Nachweis) sollten vorzugsweise in Arbeiten beschrieben werden, die in wissenschaftlichen Fachzeitschriften veröffentlicht werden. Die intern entwickelten Verfahren sollten vollständig validiert werden. Software sollte unter Anwendung von Verfahren der Software-Qualitätskontrolle und -Qualitätssicherung entwickelt und gewartet werden.

7.2 Protokollierung und Dokumentation

Alle datenanalytischen Schritte und Analysen sollten protokolliert und dokumentiert werden. Es sollte ein Plan zur Aktualisierung der bioinformatischen Pipeline entwickelt und umgesetzt werden, sobald Aktualisierungen der Software-Komponenten verfügbar werden. Die Auswirkungen der Aktualisierung der Softwarekomponenten sollten evaluiert und dokumentiert werden. Möglicherweise ist eine erneute Validierung erforderlich (siehe Abschnitt 10.1.2). Wenn Datensätze übertragen werden, sollte die Datenintegrität vor und nach der Übertragung überprüft werden (z. B. anhand der MD5-Prüfsumme). Ausnahmeprotokolle sollten verwendet werden, um Abweichungen von SOPs während einzelner bioinformatischer Analysen zu dokumentieren, z. B. dass die SOP nicht wie beschrieben befolgt wurde.

7.3 Qualitätsbeurteilungen

Die Qualität der Sequenzdaten sollte nach Abschluss des Sequenzierungslaufs beurteilt und dokumentiert werden. Qualitätsmetriken sollten plattformspezifisch sein (siehe Anhang A1 und A2). Die Anwender sollten ihre Spezifikationen für die Qualitätsbeurteilungsparameter festlegen und aufzeichnen. Zu den Kriterien, die zur Beurteilung der Sequenzqualität für ein Isolat verwendet werden, können gehören:

- durchschnittlicher Qualitätswert und Anzahl der Basen, die über einem bestimmten Qualitätsschwellenwert liegen;
- Anzahl der Reads und deren durchschnittlicher Phred-Wert;
- Es sollten Kontaminationsprüfungen durchgeführt und annehmbare Grenzwerte für Kontaminationen (z. B. Sequenzierung carry-over- oder Kreuzkontamination aus der Probenvorbereitung) festgelegt werden, die für bioinformatische Analysen geeignet sind.

Bei bioinformatischen Pipelines zur Assemblierung sollte die Qualität der Assemblierung vor Beginn der Analysen bewertet werden. Als allgemeine Indikatoren für die Qualität der Assemblierung werden die folgenden Maßnahmen empfohlen:

- Die Lesetiefe muss ausreichend sein, um sicherzustellen, dass Varianten in der Assembly zuverlässig nachgewiesen werden können.
- Anzahl der Contigs. Bei Draft-Assemblierungen kann es unter Umständen erforderlich sein, eine niedrige Abdeckung und/oder kleine Contigs zu entfernen, bevor die Anzahl der Contigs angegeben wird.
- N50 und/oder NG50 und Länge des längsten Contigs.
- Die Gesamtlänge aller Contigs oder Scaffolds sollte in etwa der bekannten Genomgröße des Zielorganismus entsprechen.
- Vorhandene artspezifische konservierte Elemente (z. B. Kerngenom).

Labore sollten in Sequenzierungsdaten auf Kontamination prüfen und Grenzwerte festlegen, die für spezifische Anwendungen geeignet sind. Diese Kontaminationen können von einer anderen Spezies/Gattung oder von der gleichen Art stammen. Zu den empfohlenen Verfahren gehören unter anderem eine oder mehrere der folgenden Verfahren:

- k-mer-Hashing im Vergleich zu einer Referenzsequenz-Datenbank;
- Berechnung der durchschnittlichen Nukleotidentität (ANI, en: average nucleotide identity) von Sequenzdaten;
- Überprüfung der Anzahl von rDNA-Allelen in Reads oder Assemblys;

- Verifizierung von Serotypen mit bioinformatischen Serotyp-Vorhersagetools und
- Vergleich von Assemblys mit Referenzdatenbanken.

Wenn Ergebnisse von nicht-WGS-phänotypischen oder molekularen Prüfungen für ein bakterielles Isolat vorliegen, dürfen sie mit WGS-Befunden verglichen werden, um die Konsistenz der genomischen Vorhersagen zu bewerten. Beispiele hierfür sind u. a.

- vorhandene oder fehlende bekannte Resistenzelemente bei Isolaten mit antimikrobiellen Suszeptibilitätsprofilen,
- Serotyp,
- antigenische Loci und
- vorhandene oder fehlende Virulenz oder pathogene Elemente.

7.4 SNP-Analysen

Für SNP-Analysen kann entweder eine genetisch ähnliche Draft-Assemblierung oder eine fertige Genomsequenz als Referenz verwendet werden [3]. Referenzsequenzen sollten vor der Analyse kuratiert werden (z. B. durch Entfernen kleiner Contigs oder Contigs mit geringer Abdeckungstiefe), wie es für bestimmte Anwendungen erforderlich ist. SNVs sollten mit Hilfe von Qualitätswerten, Abdeckungstiefe, Dichte und Maskierung hochvariabler Regionen gefiltert werden, je nach Anwendung, um Fehler zu reduzieren, die durch Sequenzierungs- und Alignment-Artefakte, InDels (Insertionen/Deletionen), Strukturvarianten, Rekombination und mobile genetische Elemente verursacht werden. Die zur Identifizierung der SNP-Positionen verwendeten Filterbedingungen sollten dokumentiert werden.

ANMERKUNG Die Analyse von Vergleichs- und simulierten Datensätzen kann dazu beitragen, falsch-positive und falsch-negative Ergebnisse für SNP-Pipelines zu ermitteln.

Das Referenzgenom sollte den Sequenzen der Prüfpersonen genetisch ähnlich sein, da die Falsch-Positiv-Rate für die SNP-Identifizierung mit zunehmendem genetischen Abstand zur Referenz tendenziell zunimmt [13]. Anwender sollten angeben, was als „genetisch ähnlich“ betrachtet wird.

7.5 MLST-Analysen (cgMLST und wgMLST)

Die Kriterien für das Hinzufügen oder Entfernen von Allelen oder Loci aus cgMLST- oder wgMLST-Datenbanken sollten klar definiert werden. Wenn MLST-Allelen-Bestimmungen mit assemblierten Genomen durchgeführt werden:

- dasselbe Genom-Annotationsverfahren sollte bei der Generierung von Datenbanken und bei nachfolgenden Analysen verwendet werden; und
- es sollten Mindestqualitätsstandards für die Assemblierung festgelegt werden, wie z. B. der Prozentsatz der detektierten Kernloci.

Wenn MLST-Allelen-Bestimmungen mit Read-Mapping durchgeführt werden:

- Kriterien für die SNP- und InDel-Identifizierung (z. B. Alignment-Tiefe, minimale prozentuale Abdeckung der Loci, Anzahl der zulässigen Diskrepanzen usw.) sollten dokumentiert werden; und
- es sollten Mindestqualitätsstandards für WGS-Datensätze festgelegt werden, wie z. B. der Prozentsatz der detektierten Kernloci.

- Die Qualitätskriterien hängen von der Gattung ab. Bei der Erstellung eines proprietären Schemas ist Vorsicht geboten. Loci mit einzelnen Allelen und solche mit hoher Längenvariabilität sollten entfernt werden.

7.6 Detektion von Zielgenen

Datenbanken, die zur Detektion von Zielgenen verwendet werden, z. B. Virulenzgen, Antibiotikaresistenzgen, Serotyp usw., sollten unter Angabe der Versionsnummer dokumentiert werden. Die Kriterien, anhand derer entschieden wird, ob das Zielgen vorhanden ist oder nicht, sollten klar definiert werden (z. B. prozentuale Abdeckung und prozentuale Identität).

7.7 Phylogenetische Baum- oder Dendrogramm-Generierung

Distanz-, Parsimonie- und Maximal-Kompatibilitäts-Analyseverfahren können zum schnellen Screening von WGS-Datensätzen und zur Identifizierung von Clustern eng verwandter Isolate verwendet werden. Die Ergebnisse werden üblicherweise als Dendrogramm oder paarweise Distanzmatrix dargestellt. Bäume oder Dendrogramme können je nach Anwendung aus der paarweisen Distanzmatrix, gemeinsamen Loci, Genen oder Varianten aufgebaut werden. Bayes'sche und Maximum-Likelihood-Verfahren der phylogenetischen Baumkonstruktion werden den distanzbasierten Verfahren vorgezogen, da sie statistisch konsistent sind (d. h. sie konvergieren auf die richtige Topologie, je mehr Daten erfasst werden). Bootstrapping sollte verwendet werden, um die statistische Unterstützung für Topologien unter distanzbasierten, parsimonischen, maximal kompatiblen und maximal wahrscheinlichen Analyseverfahren zu schätzen. „*A posteriori*“-Wahrscheinlichkeiten sollten verwendet werden, wenn Bäume durch Bayes'sche Verfahren erzeugt werden.

7.8 Metriken und Protokolldateien

Metriken und Protokolldateien sollten aufbewahrt werden und Text enthalten, der Folgendes beschreibt:

- a) die Identitäten der analysierten Isolate;
- b) die Identitäten von beliebigen Referenzsequenzen (für SNP);
- c) Version der Datenbank (für MLST);
- d) die Version der verwendeten bioinformatischen Pipeline, Parametereinstellungen und Benutzeridentifikation;
- e) Zeitstempel; und
- f) beliebige Filter- oder Maskierungsbedingungen.

Gegebenenfalls sollten Metriken angegeben werden, die Folgendes beschreiben:

- a) Anzahl der Reads;
- b) Prozent der abgebildeten Reads;
- c) durchschnittliche Insertionsgröße und
- d) durchschnittliche Alignmenttiefe.

7.9 Interpretation und Meldung der Ergebnisse bioinformatischer Analysen

7.9.1 Interpretation der Ergebnisse aus bioinformatischen Pipelines

Ergebnisse aus bioinformatischen Pipelines sollten im Zusammenhang mit Informationen zu Metadaten über die Herkunft von Isolaten und zur Epidemiologie (d. h. Rückverfolgungsinformationen) interpretiert werden. Schwellenwerte, die für einen Zweck (z. B. Klonausbrüche) festgelegt wurden, sollten nicht für die Interpretation verschiedener Zwecke (z. B. persistente oder residente Krankheitserreger) verwendet werden [14].

Die Diversität der Isolate sollte bei der Interpretation von Dendrogrammen oder Allel-/SNP-Unterschieden berücksichtigt werden. Einige Abstammungslinien dürfen geklont sein. In diesen Fällen können wenige Unterschiede Ausbrüche/Cluster führen. Andere Abstammungslinien haben eine größere Vielfalt. Im Allgemeinen ist bei Kontaminationsereignissen mit einer einzigen Punktquelle, die über einen kurzen Zeitraum auftreten, nur mit sehr wenigen genetischen Veränderungen zu rechnen. Bei großflächigen Kontaminationsereignissen können größere Unterschiede beobachtet werden.

7.9.2 Angabe über die Ergebnisse der Genomanalyse

Die folgenden Informationen sollten in den Berichten enthalten oder auf Anfrage erhältlich sein:

- a) Version der Pipeline;
- b) Identität der Eingabedaten;
- c) verwendete Referenzgenom- oder MLST-Datenbank und ggf. Version;
- d) analytische Einstellungen, falls Optionen verfügbar sind (z. B. Mindestabdeckungseinstellungen für Aufrufe, Filterung oder Maskierung) und
- e) Interpretation und Schlussfolgerungen der Ergebnisse des genomischen Vergleichs, falls Teil des Antrags.

8 Metadaten

8.1 Allgemeines

Die Organisation muss eine Richtlinie für die Erfassung von Metadaten verabschieden. Metadaten in privaten Repositorien sollten so detailliert wie möglich sein, aber der Grad der Detailliertheit liegt im Ermessen des Anwenders. Wenn es gestattet ist, dürfen Metadaten mit Partnern geteilt werden und bis zu einer Granularitätsebene abstrahiert werden, die den Richtlinien der Organisation für den Datenaustausch entspricht.

8.2 Interoperabilität und Zukunftssicherheit von Metadaten

Die Empfehlungen zu Struktur und Inhalt der Metadaten in diesem Standard sollen die Interoperabilität und den Nutzen von Metadaten bei Vergleichen zwischen verschiedenen Datenbanken gewährleisten. Die Empfehlungen zu Metadaten sollen auch dazu dienen, die Daten so zu strukturieren, dass sie für unvorhergesehene Nutzungen zugänglich sind (Zukunftssicherheit). Der hier definierte Ansatz zur Normung von Metadaten erfasst Informationen über Labore und Laborprozesse, Produktionsumgebungen, Lebensmittelerzeugnisse und Lebensmittelverarbeitung sowie bioinformatische Prozesse und Qualitätskontrollmetriken. Die Norm ist flexibel gestaltet, da einige interessierte Parteien mehr, andere weniger spezifische Informationen erfassen werden.

Genomsequenz-Metadaten, die in privaten Repositorien gespeichert sind, können Informationen zur Beschreibung der Probe, des Isolats und der Sequenz enthalten. Metadatenfelder und -werte können in den Tabellen 2, 3 und 4 und den Anhängen B bis H beschriebenen Format geliefert werden. Metadaten, die nach dieser Norm erfasst wurden, können je nach Probenart geliefert werden (siehe Abschnitt 8 und Anhänge). Ein Nullwert, z. B. „fehlt“, „nicht erhoben“, „nicht bereitgestellt“ oder „eingeschränkter Zugang“, ist jedoch auch annehmbar, wenn Informationen nicht bekannt oder verfügbar sind. Die ISO WGS-Slim und andere Ontologien können zur Formatierung und Bereitstellung von Werten für beschriebene Metadatenfelder verwendet werden.

8.2.1 Ontologien

Ontologien kodieren die Berechnungslogik, die von Softwaresystemen verwendet werden kann, um die Automatisierung und komplexere Abfragen zu verbessern [4], [5]. Die hierarchische Struktur der Ontologien ermöglicht bessere Vergleiche von Informationen auf verschiedenen Granularitätsebenen [4], [5].

8.2.2 ISO WGS-Slim

Die ISO WGS-Slim enthält genormte Felder und Begriffe, die von bestehenden Ontologien und anderen Gemeinschaftsstandards abgeleitet sind (z. B. INSDC Minimal Data for Matching (MDM) und Antibiotogramm-Standards) [6]. Die Felder geben die für die Erfassung empfohlenen Informationsarten an, während die Begriffe als mögliche Werte dienen, die zum Befüllen der Felder verwendet werden können. Die ISO WGS-Slim enthält auch synonyme Begriffskennzeichnungen verschiedener Organisationen (z. B. FoodEx2 [7]), um eine bevorzugte Verwendung von Vokabular zu vermeiden und die Interoperabilität und Datenharmonisierung zu erleichtern.

Die ISO WGS-Slim kann zur Formatierung und Bereitstellung von Werten für die in den Tabellen 2, 3 und 4 und den Anhängen B bis H beschriebenen Metadatenfelder verwendet werden. Geographische, taxonomische, Lebensmittel- und Verarbeitungs-, Umwelt- und Arzneimittelfelder in dieser Norm, die ein umfangreicheres Vokabular erfordern, können mit FoodEx2 und den Ontologien GAZ, NCBITaxon, FoodOn, EnvO und CheBI [5], [6], [7] bereitgestellt werden.

Die ISO WGS Slim kann in Tabellenkalkulationen mit Metadaten und Informationsmanagementsystemen implementiert werden. Weitere Anweisungen sind in Anhang I enthalten.

8.3 Formatierung von Metadaten mit der Norm

Die Metadatenfelder in den Tabellen 2, 3 und 4 sind allgemein, und einige erfordern zusätzliche detaillierte Felder, um die Informationen zu strukturieren und die Verwendung von Freitext zu minimieren. Diese zusätzlichen Felder werden in den Anhängen B bis H beschrieben.

Jedes Feld in den Tabellen und Anhängen enthält eine Definition, eine bestimmte Rolle bei der Datenanalyse oder -harmonisierung und spezifische Formatierungsempfehlungen. Werte für einige Felder (z. B. Lebensmittelerzeugnis) können durch die Implementierung der ISO WGS-Slim bereitgestellt werden, Werte für andere Felder (z. B. Datum) können durch eine validierte Eingabe eingeschränkt werden, während andere Felder als Freitext bereitgestellt werden können.

Die mit dieser Norm formatierten Metadaten sind mit den INSDC-Datenstandards kompatibel. Ein Leitfaden zur Vorbereitung von Metadaten für die Einreichung bei öffentlichen INSDC-Repositorien ist in Anhang H (siehe Abschnitt 9) enthalten.

8.4 Metadaten im Zusammenhang mit der Probenahme

Die Felder in Tabelle 2 können zur Erfassung von Metadaten im Zusammenhang mit der Probenahme verwendet werden. Zusätzliche Felder sind in Anhang B und C enthalten.

8.5 Metadaten im Zusammenhang mit dem Isolat

Die Felder in Tabelle 3 können zur Erfassung von Metadaten im Zusammenhang mit dem Isolat verwendet werden. Zusätzliche Felder sind in Anhang B, D, E und F enthalten.

8.6 Mit der Sequenz verbundene Metadaten

Die Felder in Tabelle 4 können zur Erfassung von Metadaten im Zusammenhang mit der Sequenz verwendet werden. Zusätzliche Felder sind in Anhang B und G enthalten.

9 Sequenz-Repositorien

Genomische Sequenzdaten müssen in einem genormten bioinformatischen Dateiformat zur Verwendung in bioinformatischen Pipelines verfügbar sein. Die Betreiber müssen Verfahren implementieren, um zu überprüfen, ob die Metadaten und die Sequenz korrekt zugeordnet sind, um die referentielle Integrität zu erhalten. Die Betreiber privater Repositorien müssen Fehler korrigieren, wenn sie erkannt werden, die Datensätze mit diesen Fehlern in öffentlichen Repositorien aktualisieren und WGS-Datensätze entfernen, wenn die referentielle Integrität nicht überprüft werden kann.

WGS-Daten und ausgewählte Metadaten dürfen in eine öffentlich zugängliche Datenbank übertragen (hochgeladen) werden. Organisationen müssen unter Umständen Metadaten vor der Einreichung bei öffentlichen Repositorien transformieren, damit keine Details oder identifizierbare Informationen preisgegeben werden. Metadaten, die nach den Tabellen und Anhängen dieser Norm bereitgestellt werden, können so formatiert werden, dass sie die MDM-Anforderungen für die Übermittlung mikrobieller Sequenzen an öffentliche INSDC-Repositorien erfüllen. Weitere Anweisungen sind in Anhang H enthalten.

10 Validierung und Verifizierung

10.1 Validierung

Die Leistungsmerkmale von WGS-basierten Verfahren sind für den bestimmungsgemäßen Gebrauch festzulegen. Die Validierung des WGS-Workflows kann für die verschiedenen Komponenten getrennt durchgeführt werden (siehe Tabelle 1). Letztendlich muss jedoch der gesamte Workflow validiert worden sein. Die Validierung wird den Nachweis erbringen, dass das Verfahren wiederholbar, reproduzierbar und genau ist.

10.1.1 Validierung des Laborbetriebs

Die Validierung des Laborbetriebs kann je nach Arbeitsablauf im Labor von der Kultur bis zur DNA-Sequenzierung und allen Phasen dazwischen reichen. Validierungsparameter und Annahmekriterien für verschiedene Phasen sind in Tabelle 1 beschrieben.

10.1.2 Validierung der bioinformatischen Pipeline

Die Leistung von bioinformatischen Pipelines sollte auf jeder geeigneten Ebene der Analyse beurteilt werden (Tabelle 1). Bioinformatische Pipelines dürfen unter Verwendung verschiedener Arten von Datensätzen evaluiert und/oder validiert werden. Die Validierung kann Probanden umfassen, die im Ursprungslabor unter Verwendung eines spezifischen WGS-Workflows zusammen mit Standard-(Benchmark)-Datensätzen und/oder simulierten Datensätzen erzeugt wurden.

- a) Standarddatensätze (oder Benchmark-Sätze) sind Fälle, in denen die Herkunft, der Phänotyp und/oder die epidemiologische Beziehung der Isolate bekannt sind und die Sequenzdaten öffentlich zugänglich gemacht wurden. Standarddatensätze können für den Vergleich der Ergebnisse verschiedener bioinformatischer Pipelines nützlich sein. Der Nutzen von Standarddatensätzen bei der Erstellung gebrauchstauglicher Arbeitsabläufe kann begrenzt sein, es sei denn, sie wurden mit demselben

Verfahren der Laborvorbereitung einschließlich der Sequenzierertechnologie erzeugt. Beispiele für Standarddatensätze umfassen:

- i) Gen-FS Project - *L. monocytogenes*, *S. enterica* ser. Bareilly, *E. coli*, und *C. jejuni* sind verfügbar unter <https://github.com/WGS-standards-and-analysis/datasets>
 - ii) BioProject PRJNA252728 (NCBI) vom National Institute of Standards and Technology (NIST)
 - iii) FDA-ARGOS project - BioProject PRJNA231221 (NCBI)
- b) Simulierte Daten – Anwendungen können verwendet werden, um synthetische Sequenz-Lesedaten aus realen Genomsequenzdaten zu erzeugen. Simulierte Datensätze können einen größeren Bereich von Parameterwerten und Fehlern prüfen, als dies in der Regel bei realen Sequenzdaten der Fall ist. Simulierte Daten sind äußerst nützlich, da bekannte Unterschiede (z. B. Nukleotid-Polymorphismen, InDels und Strukturvarianten) eingeführt werden können, was Vertrauen in die endgültigen Messungen schafft. Aufgrund von Schwierigkeiten bei der Modellierung der molekularen Evolution können Simulationen jedoch zu unrealistischen Ergebnissen führen; Pipelines schneiden mit simulierten Daten oft besser ab als mit echten Datensätzen.
- c) Probandendatensätze (d.h. echte Sequenzierungsdaten) sollen die in einem bestimmten Labor verwendeten Organismenarten und mikrobiologischen Verfahren widerspiegeln. Probandendatensätze werden unter Verwendung eines spezifischen durchgehenden Workflows generiert, der DNA-Isolierung, Bibliotheksvorbereitung, Sequenzierung und bioinformatische Analyse umfasst. Probandendatensätze werden in der Regel aus Standard- oder Referenzsammlungen mit bekannten Merkmalen oder aus Sammlungen von Isolaten im Zusammenhang mit einem Ausbruch mit bekannten epidemiologischen Informationen abgeleitet, je nach dem zu validierenden Workflow der Anwendung.

Validierungsdatsätze sollten aus Daten von Zielbakterienarten bestehen, die die Komplexität und Fehler repräsentieren, die üblicherweise bei bestimmungsgemäßem Gebrauch auftreten. Validierungsdatsätze sollten potentiell konfundierende Isolate enthalten, wie z. B. Genomsequenzen, die entweder sehr eng oder entfernt mit dem Zielbakterium verwandt sind. Die Datensätze dürfen auch mehrere Arten umfassen. Zusätzlich können Daten mehrerer Arten oder Stämme in einer einzigen Datei verwendet werden, um die Fähigkeit zur Detektion von Kreuzkontaminationen zu validieren. Annehmbare Bedingungen müssen auf der Grundlage von Leistungszielen festgelegt und je nach Anwendung dokumentiert werden, z. B:

- Genauigkeit der Annotation und Merkmalsvorhersage und/oder
- Genauigkeit der Stamm- oder Artvorhersagen, Bewertung der Verwandtschaft in Übereinstimmung mit bekannten epidemiologischen Informationen.

Validierungsdatsätze müssen mit der bioinformatischen Pipeline analysiert und die Ergebnisse anhand der festgelegten Leistungsziele und Annahmekriterien beurteilt werden (siehe Tabelle 1). Berichte, die die Validierungsergebnisse beschreiben, sollten ausreichen, um die Analysen zu replizieren. Jede größere Änderung in den bioinformatischen Pipelines muss evaluiert und dokumentiert werden. Wenn eine größere Auswirkung beobachtet wird, muss möglicherweise eine erneute Validierung durchgeführt werden.

10.1.3 Validierung des durchgehenden Workflows

Für jede WGS-Anwendung muss eine durchgehende Validierung erfolgen, wenn die Validierung eines der Schritte innerhalb des WGS-Workflows (Tabelle 1) für die beabsichtigte Anwendung fehlt oder wenn die Validierung des Laborbetriebs oder der bioinformatischen Analyse keine Probanden umfasste (siehe 10.1.2). Die Validierung des durchgehenden WGS-Workflows hilft bei der Festlegung von Schwellenwerten für biologisch relevante Unterschiede gegenüber Unterschieden, die mit dem Kultur- und Sequenzierungsprozess zusammenhängen. Die Validierung von WGS-Workflows durch Vergleich mit historischen Goldstandards (Pulsfeld-Gelelektrophorese, 7-Gen-MLST, Phagentypisierung usw.) stellt eine Herausforderung dar, da WGS eine höhere Auflösung der Daten bietet. Bakterienisolate, die früher identisch

oder nicht unterscheidbar waren, können jetzt messbare Unterschiede aufweisen. Je nach Anwendung sollten geeignete Probengenomsequenz-Datensätze erstellt werden, d.h. es sollten Isolate ausgewählt werden, die die Variabilität der Organismen repräsentieren, die für spezifische Anwendungen analysiert werden sollen.

ANMERKUNG Metriken, die mit der Methodik verbunden sind und von Van Belkum *et al.* (2007) beschrieben werden, können bei der Charakterisierung von Unterschieden zwischen eng verwandten Genomen nützlich sein [1]. Ein Beispiel für die Validierung der Quellennachverfolgung auf der Grundlage dieser Metriken ist in Portmann *et al.* (2018) [8] dargestellt. Ein Beispiel für die Validierungsstrategie, die sich speziell auf die erschöpfende Charakterisierung der bioinformatischen Analyse eines WGS-Workflows konzentriert, wird in Bogaerts *et al.* 2019 [9] dargestellt. Jede Stufe des Arbeitsablaufs sollte wie in Tabelle 1 beschrieben validiert werden.

10.2 Verifizierung

Die Verifizierung muss nachweisen, dass das ausführende Labor in der Lage ist, das validierte Verfahren für eine bestimmte WGS-Anwendung korrekt einzusetzen. Die Verifizierung muss für den gesamten Arbeitsablauf oder einen der Schritte innerhalb des Arbeitsablaufs durchgeführt werden (das Labor, das die Laborvorgänge durchführt und/oder die Einheit, die die bioinformatische Analyse durchführt).

10.2.1 Verifizierung des Laborbetriebs

Das ausführende Labor muss innerhalb des Anwendungsbereichs objektiv nachweisen, dass das validierte Verfahren in seinem Anwendungsbereich eingesetzt wird und dass die festgelegten Anforderungen erfüllt sind. Bei anerkannten/normalisierten Verfahren (z. B. Verfahren, die in maßgeblichen Büchern, wissenschaftlichen Fachzeitschriften, publizierten Normen, Anweisungen des internationalen Konsens oder Vorschriften veröffentlicht wurden) ist keine Validierung, sondern nur eine Verifizierung erforderlich.

Tabelle 1 — Validierung der Workflow-Stufen

Validierungsstufe	Wiederholpräzision (Genauigkeit/Präzision)	Vergleichpräzision (Genauigkeit/Präzision)	Übereinstimmung mit anderen Verfahren (Genauigkeit/Richtigkeit)
1. Reinkultur	Einbeziehen verschiedener Subkulturen am selben Tag durch denselben Bearbeiter	Einbeziehen verschiedener Subkulturen an verschiedenen Tagen durch verschiedene Bearbeiter	Einbeziehen verwandter und nicht verwandter Stämme (z. B. Ausbruch und Nichtausbruch) oder von Stämmen, die den/die zu betrachtenden Marker nicht enthalten
2. DNA-Extraktion	Einbeziehen verschiedener DNA-Extraktionen aus derselben Subkultur am selben Tag durch denselben Bearbeiter und unter Verwendung derselben Reagenzienchargen	Einbeziehen verschiedener DNA-Extraktionen aus derselben Subkultur durch verschiedene Bearbeiter an verschiedenen Tagen unter Verwendung verschiedener Reagenzienchargen	Einbeziehen von DNA verwandter und nicht verwandter Stämme (z. B. zugehöriger Ausbruch und Nichtausbruch) oder von Stämmen, die den/die zu betrachtenden Marker nicht enthalten
3. DNA-Sequenzierung	Einbeziehen von Bibliotheken desselben Stammes (z. B. in dreifacher Ausfertigung), die von demselben Bearbeiter am selben Tag und im selben Lauf (innerhalb der Laufgenauigkeit) erzeugt wurden	Einbeziehen von Bibliotheken, die von verschiedenen Bearbeitern an verschiedenen Tagen (zwischen Laufgenauigkeit) auf verschiedenen Instrumenten erzeugt wurden	Einbeziehen von Bibliotheken verwandter und nicht verwandter Stämme (z. B. zugehöriger Nichtausbruch) oder von Stämmen, die den/die zu betrachtenden Marker nicht enthalten

Validierungs- stufe	Wiederholpräzision (Genauigkeit/Präzision)	Vergleichpräzision (Genauigkeit/Präzision)	Übereinstimmung mit anderen Verfahren (Genauigkeit/Richtigkeit)
4. Bioinfor- matische Pipeline	Nachweis identischer Ergebnisse aus demselben Datensatz mindestens zweimal auf demselben Computer/derselben IT- Infrastruktur unter Verwendung derselben Version der Software mit denselben Optionen/Parametern	Nachweis vergleichbarer Ergebnisse aus demselben Datensatz mindestens zweimal auf verschiedenen Computern/unterschied- licher IT-Infrastruktur unter Verwendung derselben Version der Software mit denselben Optionen/Parametern	Nachweisen, dass Ergebnisse mit anderen Pipelines für dieselbe Anwendung vergleichbar sind und Festlegen aller bekannten Unterschiede zwischen Pipelines, die das Ergebnis beeinflussen können, z. B. eingebaute Referenzdatenbanken. Wenn keine solche Pipeline zur Verfügung steht, sollten simulierte Daten, bei denen die evolutionären Beziehungen der Isolate bekannt sind und die die in realen Daten erwartete Variabilität widerspiegeln, verwendet werden, um die Fähigkeit der Pipeline, die richtige Antwort zu geben, nachzuweisen.
Annahme- kriterien	Die Interpretation der Ergebnisse sollte sich nicht ändern, d. h. es sollten keine signifikanten Unterschiede beobachtet werden, wenn der WGS- Workflow im gleichen Labor mit den gleichen Bearbeitern, die das gleiche Instrument verwenden, wiederholt wird.	Die Interpretation der Ergebnisse sollte sich nicht ändern, d. h. es sollten keine signifikanten Unterschiede bei der Reproduktion des WGS-Workflows in verschie- denen Laboren, mit verschie- denen Bearbeitern oder verschiedenen Instrumenten erzielt werden. Es werden geringfügige Unterschiede erwartet, die durch metho- dische Schritte bedingt sind. Auch Genominhalte wie Plas- mide können verloren gehen. Sie sollten nicht als signifi- kant angesehen werden.	Der WGS-Workflow muss in der Lage sein, dieselben Schlussfolgerungen zu ziehen wie andere Goldstandard- Typisierungs- und/oder Charakterisierungsverfahren (z. B. epidemiologische Inferenz/Konkordanz, Differenzierung nicht verwandter Stämme bei der Gruppierung eng verwandter Isolate, andere genotypische Verfahren, vergleichbare Daten mit geschlossenen Referenzgenomen).

10.2.2 Verifizierung der bioinformatischen Pipeline

Wenn handelsübliche oder bioinformatische Open-Source-Pipelines verwendet werden, die von ihren Entwicklern validiert wurden, die Validierungsprüfungen veröffentlicht werden und die Validierungsdatensätze öffentlich zugänglich sind, muss die Validierungsprüfung möglicherweise erst nach der Installation der Software (teilweise) wiederholt werden. Hier können Prüfdatensätze verwendet werden, die von den Pipeline-Entwicklern vertrieben werden. Die erfolgreiche Ausführung eines Prüfdatensatzes bedeutet jedoch nicht notwendigerweise, dass eine bioinformatische Pipeline validiert oder einsatzbereit ist. Mit Prüfdatensätzen kann überprüft werden, ob bioinformatische Pipelines und die damit verbundenen Abhängigkeiten korrekt installiert sind und erwartungsgemäß funktionieren. Der Anwender muss die Funktionalität der Pipeline nach den vorher festgelegten Parametern aufzeigen. Die Prüfdaten werden als Eingabe für eine bioinformatische Pipeline verwendet, und die Ausgabe wird mit den erwarteten Ergebnissen verglichen. Prüfdatensätze sind in der Regel klein (z. B. Lambda-Phagengenom) und werden mit der Software oder Pipeline vertrieben. Wenn die Daten zu groß sind, um mit der Software gebündelt zu werden, können die Zugangsnummern von Datenspeichern angegeben werden.

Tabelle 2 — Empfohlene Metadatenfelder und mit der Probenahme verbundene Werte

Metadatenfelder und Definition	Rolle bei Datenanalysen/ Harmonisierung	ISO-Empfehlungen
Kontaktinformationen des Labors zur Probenahme Der Name des Labors, das die zu analysierende Probe genommen hat, sowie der Name und die Kontaktinformationen einer Person, die weitere Einzelheiten zu dem Projekt oder der Probe angeben kann, sollten ebenfalls bereitgestellt werden.	Einrichtung der Überwachungskette und Bereitstellung von Kontaktinformationen für Folgeanalysen.	Kontaktinformationen können durch die Informationsfelder in Anhang B angegeben werden.
Geographischer Standort der Probenahme Die geographische Herkunft der Probe.	INSDC-Datenstandard (erfüllt das MDM-Feld „geo_loc“).	Beispiele für geographische Standortinformationen können durch die Informationsfelder in Anhang C festgelegt werden.
Probenahmedatum Datum, an dem die Probe genommen wurde.	INSDC-Datenstandard (erfüllt das MDM-Feld „collection_date“).	Das Probenahmedatum kann als JJJJ-MM-TT nach ISO 8601 unter Verwendung einer validierten Dateneingabe aufgezeichnet werden.
Probenart Die Art des Materials, aus dem das Isolat gewonnen wurde. Proben werden in der Regel als Lebensmittelerzeugnisse, Körperprodukte oder Gewebe kategorisiert, oder Umweltproben, die aus dem Bereich der Lebensmittelproduktion und Lebensmittelhandhabung entnommen werden.	Rückverfolgung und andere Analysen.	Probearten können aus der ISO WGS-Slim ausgewählt werden.
Lebensmittelerzeugnis Produkte, die für den menschlichen Verzehr und als Futtermittel bestimmt sind	INSDC-Datenstandard (erfüllt das MDM-Feld „isolation_source“).	Lebensmittelerzeugnisse und Inhaltsstoffe können sowohl für menschliche Nahrung als auch für Tiernahrung gelten. Wenn Lebensmittelerzeugnisse zutreffen, können Deskriptoren aus der ISO WGS-Slim ausgewählt werden.
Lebensmittelverarbeitung Verarbeitung, die für ein Lebensmittel gilt, z. B. Entbeinen, Enthäuten, Pasteurisieren.	INSDC-Datenstandard (erfüllt das MDM-Feld „isolation_source“).	Wenn Lebensmittelverarbeitung zutrifft, können Deskriptoren aus der ISO WGS Slim ausgewählt werden.
Umweltmaterial Ein Stoff, der aus der natürlichen oder vom Menschen geschaffenen Umwelt gewonnen wird, z. B. Boden, Wasser, Stalldung.	INSDC-Datenstandard (erfüllt das MDM-Feld „isolation_source“).	Lebensmittel werden als ein vom Umweltmaterial getrennter Bereich betrachtet. Wenn Umweltmaterialien zutreffen, können Deskriptoren aus der ISO WGS Slim ausgewählt werden.

Metadatenfelder und Definition	Rolle bei Datenanalysen/ Harmonisierung	ISO-Empfehlungen
Umweltstandort Ein Umweltstandort kann einen Standort in der natürlichen oder bebauten Umgebung beschreiben, z. B. Schlachthof, Einzelhandel, Futtermittelbetrieb, Nahrungsmittelmaschinen, Flächen, die zur Verarbeitung und Zubereitung von Lebensmitteln verwendet werden.	INSDC-Datenstandard (erfüllt das MDM-Feld „isolation_source“).	Lebensmittelbezogene Umweltstandorte können unter anderem, aber nicht ausschließlich, Lebensmittelproduktion, -verarbeitung, -verteilung und Einzelhandelsumgebungen umfassen, die beprobt wurden. Wenn Umweltstandorte zutreffen, können Deskriptoren aus der ISO WGS Slim ausgewählt werden.
Sammelbehälter Das Instrument oder der Behälter, das/der zur Probenahme verwendet wird, z. B. steriler Plastikbeutel, Plastikbehälter, Tupfer (mit oder ohne Transportmedium), Schlauch.	INSDC-Datenstandard (erfüllt das MDM-Feld „isolation_source“).	Die Sammelbehälter sind nicht immer bekannt, aber wenn sie angegeben werden, können sie aus der ISO WGS Slim ausgewählt werden.
Probenahmeverfahren Das zur Probenahme verwendete Verfahren	INSDC-Datenstandard MDM, wenn das für die Probenahme verwendete Verfahren bekannt ist (erfüllt das MDM-Feld „isolation_source“).	Die Probenahmeverfahren sind nicht immer bekannt, aber wenn sie angegeben werden, können sie aus der ISO WGS Slim ausgewählt werden.

Tabelle 3 — Empfohlene Metadatenfelder und mit dem Isolat verbundene Werte

Metadatenfelder und Definition	Rolle bei Datenanalysen/ Harmonisierung	ISO-Empfehlungen
Kontaktinformationen des mikrobiologischen Labors Der Name des Labors, das den zu sequenzierenden Organismus isoliert hat, sowie der Name und die Kontaktinformationen einer Person, die weitere Einzelheiten zu dem Projekt oder Isolat angeben kann, sollten ebenfalls bereitgestellt werden.	Einrichtung der Überwachungskette und Bereitstellung von Kontaktinformationen für Folgeanalysen.	Kontaktinformationen können durch die Informationsfelder in Anhang B angegeben werden.
Organismus Die Spezies des zu sequenzierenden Isolats.	INSDC-Datenstandard (erfüllt das MDM-Feld „organism“).	Der wissenschaftliche Name für die Art kann unter Verwendung genormter taxonomischer Bezeichnungen von NCBI Taxon angegeben werden.
Stamm Der Name oder Identifikator des Stammes.	INSDC-Datenstandard (erfüllt das MDM-Feld „strain or isolate“).	Der Stammidentifikator kann als Freitext zur Verfügung gestellt werden.
Isolat Der Name oder Identifikator des Isolats.	INSDC-Datenstandard (erfüllt das MDM-Feld „strain or isolate“).	Der Isolatidentifikator kann als Freitext zur Verfügung gestellt werden.

Metadatenfelder und Definition	Rolle bei Datenanalysen/ Harmonisierung	ISO-Empfehlungen
Serotyp Der Serotyp des Isolats oder Stammes, wie er durch <i>In-vitro</i> - oder <i>In-silico</i> -Verfahren (z. B. WGS, PCR oder immunologische Verfahren) bestimmt wird.	MDM des öffentlichen Repositorys (erforderlich für das EBI-Feld „Serotyp“).	Die Ergebnisse des Serotyps können, falls verfügbar, als Freitext zur Verfügung gestellt werden.
Isolierungsmedium Die Nährmedien, die zur Isolierung des zu sequenzierenden Organismus von anderen Organismen in der Probe verwendet werden.	Berechenbare Vergleiche von Methodiken.	Deskriptoren dieses Materials können aus der ISO WGS Slim ausgewählt werden.
Verlauf der Isolatpassage Die Anzahl, wie oft ein Isolat nach einem bestimmten Verfahren seriell subkultiviert wird.	Berechenbare Vergleiche von Methodiken. Eine Erhöhung der Anzahl der Passierungen eines Isolats kann unter Umständen zur Anhäufung zusätzlicher Mutationen führen.	Details zur Isolatpassage können durch die Informationsfelder in Anhang D angegeben werden.
Ergebnisse des Antibiotogramms Die minimalen Hemmkonzentrationen (Wert, Einheit, Zeichen (<, >, =)) und die Resistenzphänotypen (resistent, sensitives Zwischenprodukt oder unbestimmt) des sequenzierten Isolats im Vergleich zu verschiedene geprüfte Antibiotika.	Berechenbare Vergleiche von Antibiotogrammen.	Wenn die Ergebnisse der Antibiotogramme vorliegen, können die Informationen entsprechend den Feldern in Anhang E festgelegt werden. Die Quelle der Breakpoints (und die Version), die für die Interpretation/Klassifizierung der MHK-Werte verwendet werden, kann angegeben werden.
Verfahren der Antibiotogramme Das Laborprotokoll zur Bestimmung von Resistenzphänotypen und minimalen Hemmkonzentrationen von Antibiotika, die im Vergleich zu ein Isolat geprüft werden. Das Protokoll sollte die geprüften Antibiotika, die Laborprüfverfahren und die Prüfnorm sowie die für die Prüfung verwendeten Kontroll-/Referenzstämme enthalten.	Berechenbare Vergleiche von Antibiotogrammen.	Antibiotogramm-Verfahren (falls zutreffend) können entsprechend den Feldern in Anhang E festgelegt werden.
Ergebnisse des Virulenzfaktors Die Virulenzfaktoren, deren Vorhandensein im sequenzierten Isolat durch phänotypische oder Zielamplifikationsverfahren bestimmt wird, z. B. Shiga-Toxine, Hämolysine.	Berechenbare Vergleiche der Virulenz.	Wenn die Ergebnisse der Virulenz vorliegen, können die Informationen nach den Feldern in Anhang F festgelegt werden.
Verfahren zur Prüfung des Virulenzfaktors Das Laborprotokoll zur Bestimmung von Virulenzphänotypen und Markern.	Berechenbare Vergleiche der Virulenz.	Verfahren zur Virulenzprüfung (falls zutreffend) können entsprechend den Feldern in Anhang F festgelegt werden.

Tabelle 4 — Empfohlene Metadatenfelder und mit der Sequenz verbundene Werte

Metadatenfelder und Definition	Rolle bei Datenanalysen/ Harmonisierung	ISO-Empfehlungen
Kontaktinformationen der Sequenzierungseinrichtung Der Name der Einrichtung, die den isolierten Organismus sequenziert hat, sowie der Name und die Kontaktinformationen einer Person, die weitere Einzelheiten zum Projekt und zur Sequenzierung angeben kann, sollten bereitgestellt werden.	Einrichtung der Überwachungskette und Bereitstellung von Kontaktinformationen für Folgeanalysen.	Kontaktinformationen können durch die Informationsfelder in Anhang B angegeben werden.
Sequenzierungsdatum Das Datum, an dem der Sequenzierungslauf initiiert wurde.	Verfolgen von Sequenzierungsläufen.	Die Sequenzierung kann als JJJJ-MM-TT nach ISO 8601 unter Verwendung einer validierten Dateneingabe aufgezeichnet werden.
Nährmedien Formulierung von Stoffen in flüssiger, halbfester oder fester Form, die natürliche und/oder synthetische Bestandteile enthalten, die die Vermehrung (mit oder ohne Hemmung bestimmter Mikroorganismen), die Identifizierung oder die Erhaltung der Lebensfähigkeit von Mikroorganismen unterstützen sollen	Berechenbare Vergleiche von Methodiken.	Deskriptoren dieses Materials können aus der ISO WGS Slim ausgewählt werden.
DNA-Extraktionsverfahren Das Verfahren, mit dem genomische DNA aus einer Probe durch chemische, physikalische oder mechanische Mittel gewonnen wird.	Berechenbare Vergleiche von Methodiken und Qualitätslenkung.	Der Name des handelsüblichen Kits und die Versionsnummer oder das Laborprotokoll sind einzufügen, das zur Extraktion der genomischen DNA des isolierten Organismus unter Verwendung von Freitext verwendet wurde.
Sequenzierung von Parallelproben Eine technische Sequenzierungswiederholung stellt die Resequenzierung der gleichen biologischen Probe dar, um die experimentelle Variation zu beurteilen. Eine biologische Sequenzierung stellt biologisch ein Sequenzierungsexperiment dar, das auf einer anderen Kolonie desselben Isolats basiert, die aus demselben Probenmaterial gewonnen wurde, um die biologische Variation zu beurteilen.	Verfolgen von Sequenzierungsläufen und Analysieren der Variabilität von Reads und Sequenzen.	Parallelproben innerhalb einer Reihe von Sequenzierungsläufen können entweder als technisch oder biologisch beschrieben werden. Wenn Sequenzierungswiederholungen zutreffen, können Deskriptoren aus der ISO WGS Slim ausgewählt werden.
Verfahren zur Vorbereitung der Sequenzbibliothek Das Verfahren zur Erstellung einer Bibliothek aus DNA-Fragmenten unter Verwendung von Oligonukleotiden mit der Rolle von Adaptern.	Berechenbare Vergleiche von Methodiken und Qualitätslenkung.	Der Name des handelsüblichen Kits und die Versionsnummer oder das Laborprotokoll sind einzufügen, das für die Vorbereitung der Bibliotheken zur Sequenzierung als Freitext verwendet wurde.