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

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

[ISO 23418:2022](https://standards.iteh.ai/catalog/standards/sist/3a188c8c-8076-4bc3-a45b-48845fda28d1/iso-23418-2022)

<https://standards.iteh.ai/catalog/standards/sist/3a188c8c-8076-4bc3-a45b-48845fda28d1/iso-23418-2022>



iTeh STANDARD PREVIEW
(standards.iteh.ai)

ISO 23418:2022

<https://standards.iteh.ai/catalog/standards/sist/3a188c8c-8076-4bc3-a45b-48845fda28d1/iso-23418-2022>



DOCUMENT PROTÉGÉ PAR COPYRIGHT

© ISO 2022

Tous droits réservés. Sauf prescription différente ou nécessité dans le contexte de sa mise en œuvre, aucune partie de cette publication ne peut être reproduite ni utilisée sous quelque forme que ce soit et par aucun procédé, électronique ou mécanique, y compris la photocopie, ou la diffusion sur l'internet ou sur un intranet, sans autorisation écrite préalable. Une autorisation peut être demandée à l'ISO à l'adresse ci-après ou au comité membre de l'ISO dans le pays du demandeur.

ISO copyright office
Case postale 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Genève
Tél.: +41 22 749 01 11
E-mail: copyright@iso.org
Web: www.iso.org

Publié en Suisse

Sommaire

Page

Avant-propos	v
Introduction	vi
1 Domaine d'application	1
2 Références normatives	1
3 Termes et définitions	1
4 Principe	7
4.1 Généralités	7
4.2 Opérations réalisées en laboratoire: préparation et séquençage de l'échantillon	7
4.3 Analyse bioinformatique	7
4.3.1 Généralités	7
4.3.2 Analyses des SNP	8
4.3.3 Analyses des MLST	8
4.3.4 Analyse de la distance k-mer	8
4.4 Formats de métadonnées et dépôt dans la base de données de séquences	8
4.5 Validation et vérification du processus de WGS	8
5 Recommandations générales pour le laboratoire	9
5.1 Isolement bactérien et extraction de l'ADN	9
5.2 Environnement du laboratoire	9
5.3 Procédure opérationnelles normalisées et travaux non conformes	9
5.4 Système de management de l'information des laboratoires	9
5.5 Compétence du laboratoire	9
6 Opérations du laboratoire	10
6.1 Préparation et stockage des échantillons	10
6.2 Isolats bactériens	10
6.3 Isolement de l'ADN	10
6.4 Préparation de la librairie et séquençage	10
6.4.1 Préparation de la librairie	10
6.4.2 Séquençage de l'ADN	11
6.4.3 Utilisation de contrôles	11
6.4.4 Évaluation de la qualité des données de lectures brutes	11
6.4.5 Stockage et rétention des échantillons et des données	12
7 Analyse bioinformatique des données	12
7.1 Exigences relatives aux pipelines et logiciels bioinformatiques utilisés pour l'analyse des données	12
7.2 Connexion et documentation	12
7.3 Évaluations de qualité	12
7.4 Analyses des SNP	14
7.5 Analyses des MLST (cgMLST et wgMLST)	14
7.6 Détection de gènes cibles	14
7.7 Génération d'arbre phylogénétique ou de dendrogramme	14
7.8 Métriques et fichiers journaux	15
7.9 Interprétation et consignation des résultats des analyses bioinformatiques	15
7.9.1 Interprétation des résultats des pipelines bioinformatiques	15
7.9.2 Consignation des résultats des analyses génomiques	15
8 Métadonnées	16
8.1 Généralités	16
8.2 Interopérabilité et pérennité des métadonnées	16
8.2.1 Généralités	16
8.2.2 Ontologies	16
8.2.3 ISO WGS Slim	16
8.3 Formatage des métadonnées au moyen du présent document	17

8.4	Métadonnées associées à la collecte d'échantillons	17
8.5	Métadonnées associées à l'isolat	18
8.6	Métadonnées associées à la séquence.....	20
9	Bases de données de séquences.....	22
10	Validation et vérification	22
10.1	Validation	22
10.1.1	Généralités.....	22
10.1.2	Validation des opérations de laboratoire	24
10.1.3	Validation du pipeline bioinformatique	24
10.1.4	Validation du processus de bout en bout.....	25
10.2	Vérification.....	26
10.2.1	Généralités.....	26
10.2.2	Vérification des opérations de laboratoire	26
10.2.3	Vérification du pipeline bioinformatique	26
Annexe A	(informative) Élaboration de métriques de qualité et utilisation de contrôles.....	27
Annexe B	(informative) Champs de coordonnées du laboratoire	33
Annexe C	(informative) Champs de localisation géographique de la collecte d'échantillon.....	35
Annexe D	(informative) Champs d'historique de repiquage de l'isolat.....	36
Annexe E	(informative) Champs de méthodes et de résultats pour l'antibiogramme.....	37
Annexe F	(informative) Champs de détection des facteurs de virulence et de méthodes.....	39
Annexe G	(informative) Métriques de contrôle qualité des séquences.....	40
Annexe H	(informative) Spécifications des métadonnées.....	41
Annexe I	(informative) Instructions pour l'intégration de l'ontologie Slim par les développeurs de logiciels.....	44
Bibliographie	49

Avant-propos

L'ISO (Organisation internationale de normalisation) est une fédération mondiale d'organismes nationaux de normalisation (comités membres de l'ISO). L'élaboration des Normes internationales est en général confiée aux comités techniques de l'ISO. Chaque comité membre intéressé par une étude a le droit de faire partie du comité technique créé à cet effet. Les organisations internationales, gouvernementales et non gouvernementales, en liaison avec l'ISO participent également aux travaux. L'ISO collabore étroitement avec la Commission électrotechnique internationale (IEC) en ce qui concerne la normalisation électrotechnique.

Les procédures utilisées pour élaborer le présent document et celles destinées à sa mise à jour sont décrites dans les Directives ISO/IEC, Partie 1. Il convient, en particulier, de prendre note des différents critères d'approbation requis pour les différents types de documents ISO. Le présent document a été rédigé conformément aux règles de rédaction données dans les Directives ISO/IEC, Partie 2 (voir www.iso.org/directives).

L'attention est attirée sur le fait que certains des éléments du présent document peuvent faire l'objet de droits de propriété intellectuelle ou de droits analogues. L'ISO ne saurait être tenue pour responsable de ne pas avoir identifié de tels droits de propriété et averti de leur existence. Les détails concernant les références aux droits de propriété intellectuelle ou autres droits analogues identifiés lors de l'élaboration du document sont indiqués dans l'Introduction et/ou dans la liste des déclarations de brevets reçues par l'ISO (voir www.iso.org/brevets).

Les appellations commerciales éventuellement mentionnées dans le présent document sont données pour information, par souci de commodité, à l'intention des utilisateurs et ne sauraient constituer un engagement.

Pour une explication de la nature volontaire des normes, la signification des termes et expressions spécifiques de l'ISO liés à l'évaluation de la conformité, ou pour toute information au sujet de l'adhésion de l'ISO aux principes de l'Organisation mondiale du commerce (OMC) concernant les obstacles techniques au commerce (OTC), voir www.iso.org/avant-propos.

Le présent document a été élaboré par le comité technique ISO/TC 34, *Produits alimentaires*, sous-comité SC 9, *Microbiologie*, en collaboration avec le comité technique CEN/TC 463, *Microbiologie de la chaîne alimentaire*, du Comité européen de normalisation (CEN) conformément à l'Accord de coopération technique entre l'ISO et le CEN (Accord de Vienne).

Il convient que l'utilisateur adresse tout retour d'information ou toute question concernant le présent document à l'organisme national de normalisation de son pays. Une liste exhaustive desdits organismes se trouve à l'adresse www.iso.org/fr/members.html.

Introduction

Le séquençage à haut débit (NGS, *next generation sequencing*) permet un accès rapide, économique, et à haut débit à des séquences de génomes microbiens entiers et est appliqué en réponse à un nombre croissant de problèmes dans le secteur de la microbiologie des aliments. Les séquences de génomes entiers sont des représentations du potentiel biologique de l'organisme séquencé avec une résolution à la base. Le séquençage de génomes entiers (WGS) offre des avantages significatifs par rapport aux technologies existantes (par exemple, sérotypage, électrophorèse sur gel en champ pulsé, phénotype de résistance aux antibiotiques) dans de nombreuses applications. Les analyses basées sur le WGS sont utilisées par les laboratoires de santé publique pour détecter les épidémies, ainsi que pour détecter les mutations, gènes et autres éléments génétiques caractérisant la virulence et le potentiel de survie. L'industrie alimentaire s'intéresse à l'utilisation de séquences de génomes entiers pour caractériser des isolats bactériens provenant d'ingrédients et de surfaces environnementales, afin de mieux comprendre leur origine et leur écologie, et d'actualiser les modes opératoires dans le but de réduire le risque. Des sociétés ont développé ou développent actuellement leur capacité de collecte et d'analyse de données de séquences de génomes entiers. D'autres confient ces opérations techniques à des laboratoires tiers, comme elles le font pour d'autres analyses microbiologiques.

Le présent document fournit des recommandations pour les parties à la fois de laboratoire et de bioinformatique des séquences de génomes entiers, ainsi que pour les métadonnées associées relatives aux micro-organismes bactériens d'origine alimentaire échantillonnés tout au long de la chaîne alimentaire (par exemple, ingrédients, produits alimentaires, aliments pour animaux, environnement de production). Bien que la microbiologie de la chaîne alimentaire comprenne les virus et les champignons, le présent document concerne uniquement les bactéries. Le présent document est destiné à être applicable à toutes les technologies disponibles actuellement de séquençage d'ADN. Il peut être appliqué à l'analyse des données de séquences de génomes entiers au moyen d'un logiciel commercial, libre de droits ou personnalisé. Il n'a pas vocation à spécifier les chimies de séquençage, les méthodes analytiques, ni le logiciel d'analyse. Le présent document définit les bonnes pratiques de laboratoire, de gestion des données et des métadonnées, afin de s'assurer que les analyses sont transparentes, clairement consignées dans un rapport et utilisables dans des investigations. Le présent document est destiné à être utilisé par les laboratoires pour le développement de leurs systèmes de management de la qualité et de leurs opérations techniques. Les clients des laboratoires et les autorités réglementaires peuvent également l'utiliser pour confirmer ou reconnaître la compétence des laboratoires. Le présent document peut aussi être appliqué à d'autres domaines (par exemple, environnement, santé humaine, santé animale).

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

AVERTISSEMENT — Afin de protéger la santé du personnel de laboratoire, il est essentiel que la manipulation des cultures bactériennes soit effectuée uniquement dans des laboratoires dotés d'un équipement approprié, sous le contrôle d'un microbiologiste expérimenté, et qu'un grand soin soit apporté à l'élimination de l'ensemble des matériaux ayant servi à l'incubation. Il convient que les utilisateurs du présent document connaissent les pratiques courantes de laboratoire. Le présent document ne prétend pas couvrir tous les aspects de sécurité liés, le cas échéant, à son utilisation. Il incombe à l'utilisateur de mettre en place des pratiques de santé et de sécurité appropriées.

1 Domaine d'application

Le présent document spécifie les exigences minimales pour générer et analyser des données de séquençage de génome entier (WGS) de bactéries provenant de la chaîne alimentaire. Ce processus peut comprendre les étapes suivantes:

- a) manipulation des cultures bactériennes;
- b) isolement de l'ADN génomique axène;
- c) préparation de la librairie, séquençage et évaluation de la qualité et du stockage des lectures de séquences brutes d'ADN;
- d) analyse bioinformatique visant à déterminer la parenté génétique et le contenu génétique, à prédire le phénotype et à valider le pipeline bioinformatique;
- e) capture des métadonnées et dépôt dans des bases de données de séquences;
- f) validation du processus de WGS de bout en bout (adapté à l'application prévue).

Le présent document est applicable aux bactéries isolées à partir de ce qui suit:

- des produits destinés à la consommation humaine;
- des produits destinés à l'alimentation animale;
- des échantillons environnementaux prélevés dans des zones de production et de manipulation de produits alimentaires et d'aliments pour animaux;
- des échantillons de production primaire.

2 Références normatives

Le présent document ne contient aucune référence normative.

3 Termes et définitions

Pour les besoins du présent document, les termes et définitions suivants s'appliquent.

L'ISO et l'IEC tiennent à jour des bases de données terminologiques destinées à être utilisées en normalisation, consultables aux adresses suivantes:

- ISO Online browsing platform: disponible à l'adresse <https://www.iso.org/obp>
- IEC Electropedia: disponible à l'adresse <https://www.electropedia.org/>

3.1 adaptateur

ADN possédant une séquence connue, qui est ajouté à l'extrémité d'un fragment de la librairie d'ADN afin de faciliter le procédé de séquençage (par exemple, appariement sur la cellule en flux de séquençage)

3.2 annotation

procédé d'identification des gènes et autres caractéristiques sur des *assemblages* (3.4) génomiques

3.3 antibiogramme

synthèse des résultats des essais de sensibilité aux agents antimicrobiens réalisés pour un micro-organisme spécifique, généralement représenté sous forme de tableau

3.4 assemblage

produit du procédé d'alignement et de fusion des *séquences nucléotidiques lues* ou *lectures* (3.38) en séquences contiguës plus longues (*contigs* (3.10))

3.5 attribution des bases base calling

procédé consistant à affecter des nucléotides et des scores de qualité à des emplacements dans les *lectures* (3.38)

3.6 bioinformatique

collecte, stockage et analyse de données biologiques, y compris des séquences

3.7 pipeline bioinformatique

programmes individuels, scripts ou éléments de logiciels liés ensemble, dans lesquels le produit d'un programme est utilisé comme entrée pour l'étape suivante de traitement des données

3.8 intercontamination

contamination des échantillons due à de précédentes analyses, transférée à l'analyse en cours (par exemple, intercontamination de produits d'amplification de précédentes analyses de réaction de polymérisation en chaîne (PCR) à l'analyse PCR en cours, ou intercontamination d'échantillons séquencés précédemment d'un cycle de séquençage à un autre)

3.9 Chemical Entities of Biological Interest (ontologie) ChEBI

ontologie (3.35) utilisée pour la description de petits composés chimiques

3.10 contig

fragment contigu de séquence d'ADN résultant de l'*assemblage* (3.4) de *lectures* (3.38) de séquences d'ADN plus petites se chevauchant

3.11**vocabulaire contrôlé**

jeu fini de valeurs qui correspondent aux seules valeurs admises pour un élément de données

[SOURCE: ISO 11238:2018, 3.18, modifié — La Note 1 à l'article a été supprimée.]

3.12**couverture**

nombre de fois qu'une position de base donnée est lue dans un cycle de séquençage

Note 1 à l'article: Nombre de *lectures* (3.38) qui couvrent une position particulière.

[SOURCE: ISO 20397-2:2021, 3.6, modifié — Le terme admis «profondeur de couverture» a été supprimé.]

3.13**contamination croisée**

contamination d'un échantillon (*isolat* (3.23) bactérien ou ADN) avec d'autres échantillons au cours de la préparation d'une série d'échantillons d'ADN en vue de leur séquençage

3.14**échantillon d'ADN**

portion d'ADN extraite de l'échantillon traité

3.15**assemblage préliminaire**

assemblage (3.4) de génome *de novo* composé de *contigs* (3.10) sans ordre implicite, généralement obtenu par séquençage tronqué de génome entier avec une technologie de séquençage générant des lectures courtes

3.16**Environment Ontology****EnvO**

ontologie (3.35) utilisée pour la description des caractéristiques environnementales et des habitats

3.17**FoodEx2 (ontologie)****FoodEx2**

classification alimentaire normalisée et système de description développé par l'Autorité européenne de sécurité des aliments (EFSA)

3.18**Food Ontology****FoodOn**

ontologie (3.35) utilisée pour la description des produits alimentaires, des aliments pour animaux et de la transformation des aliments

3.19**Gazetteer (ontologie)****GAZ**

ontologie (3.35) utilisée pour la description des localisations géographiques

3.20**index**

séquences oligonucléotidiques utilisées dans le procédé de préparation de bibliothèques pour étiqueter ou marquer par code-barres l'ADN d'échantillons spécifiques afin de pouvoir combiner (multiplexer) plusieurs échantillons (c'est-à-dire plusieurs *bibliothèques* (3.25)) dans un groupe de bibliothèques et de les analyser au cours d'une seule réaction de séquençage

3.21

International Nucleotide Sequence Database Collaboration

INSDC

initiative conduite par la Base de données d'acides nucléiques du Japon (DDBJ), l'Institut européen de bioinformatique du Laboratoire européen de biologie moléculaire (EMBL-EBI) et le National Center for Biotechnology Information (NCBI)

3.22

séquençage de génome entier slim de l'Organisation internationale de normalisation

ISO WGS Slim

ontologie (3.35) slim contenant des champs interopérables et des termes relatifs à l'utilisation du *WGS* (3.49) dans le secteur de la microbiologie de la chaîne alimentaire

3.23

isolat

population de cellules bactériennes en culture pure dérivée d'une *souche* (3.45) unique

3.24

k-mer

séquence possible de longueur *k* contenue dans une séquence de génome entier

3.25

librairie

collection de fragments d'ADN génomique provenant d'un *isolat* (3.23) unique, destinée à déterminer la ou les séquences du génome

Note 1 à l'article: Une collection de librairies, chacune d'elle provenant d'un isolat unique, est appelée «groupe de librairies» et est transférée dans un séquenceur en vue d'être analysée. Ce multiplexage de librairies permettrait tout de même d'obtenir le résultat correspondant à un seul isolat si des index uniques étaient utilisés pour la préparation de la librairie de chaque isolat unique.

Note 2 à l'article: Une librairie d'ADN mixte, c'est-à-dire provenant d'un mélange de plusieurs espèces, peut être constituée. Cependant, ce type de librairie fait référence au séquençage métagénomique et est donc exclu du domaine d'application du présent document.

3.26

système de management

systèmes qualité, administratifs et techniques qui gouvernent les opérations d'un organisme

Note 1 à l'article: Pour les besoins du présent document, l'« organisme » désigne le laboratoire.

3.27

alignement

utilisation d'un logiciel pour aligner les *lectures* (3.38) sur des séquences de référence

3.28

métadonnées

données qui définissent et décrivent d'autres données

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

3.29

données minimales pour appariement

MDM

informations nécessaires pour décrire la source de l'échantillon et la provenance d'une séquence génomique, comme défini par le Global Microbial Identifier,^[1] mises en œuvre par l'*International Nucleotide Sequence Database Collaboration* (3.21)

3.30**typage par séquençage multilocus
MLST**

méthode d'analyse génomique visant à identifier les variants nucléotidiques au sein d'ensembles prédéfinis de loci

Note 1 à l'article: Utilisé à l'origine pour sept loci, il est désormais appliqué à d'autres loci de la partie commune du génome des souches pour la cgMLST ou aux loci du génome entier pour le wgMLST.

3.31**N50**

longueur (N) telle que les *contigs* (3.10) de séquences de longueur N ou plus incluent la moitié des bases de l'assemblage (3.4)

3.32**NCBITaxon**

traduction automatique de la base de données taxonomiques du National Center for Biotechnology Information (NCBI) en obo/owl

3.33**NG50**

longueur (N) d'ADN telle que les *contigs* (3.10) de séquences de longueur N ou plus incluent la moitié des bases du génome

3.34**Open Biological and Biomedical Ontology Foundry
OBO Foundry**

collection d'*ontologies* (3.35) créée par un collectif de développeurs d'ontologie déterminés à collaborer et à adhérer à des principes partagés

3.35**ontologie**

vocabulaire contrôlé (3.11) organisé de manière hiérarchique, dans lequel les termes sont connectés par des relations logiques

3.36**slim (ontologie)**

ensemble de champs d'ontologie et termes annotés dans le cadre d'une collection donnée, souvent dans un but spécifique, pouvant être extrait pour créer un fichier distinct de l'*ontologie* (3.35) d'origine

3.37**score Phred de qualité de séquence**

Q

mesure de la probabilité (*P*) qu'une base soit attribuée de manière incorrecte à un emplacement donné dans la séquence, exprimée comme:

$$Q = -10 \lg P$$

Note 1 à l'article: Un score Q30 indique une probabilité de 1 sur 1 000 qu'une base soit affectée de manière incorrecte (c'est-à-dire que l'attribution des bases est exacte à 99,9 %).

3.38**lecture****séquence nucléotidique lue**

séquence nucléotidique déduite d'un fragment d'ADN ou d'ARN

3.39

base de données de séquences

base de données dans laquelle des jeux de données de *séquençage de génome entier* (3.49) sont stockés et gérés

Note 1 à l'article: Une base de données publique autorise le libre accès aux données, tandis qu'une base de données privée ou fédérée limite l'accès aux données.

3.40

réplicat de séquençage

<biologique> séquençage d'une colonie différente à partir du même *isolat* (3.23) obtenu à partir du même échantillon, pour évaluer la variation biologique

3.41

réplicat de séquençage

<technique> reséquençage du même échantillon biologique ou de la même *librairie* (3.25) pour évaluer la variation de séquence due aux équipements et au protocole

3.42

sérotype

schéma de classification basé sur la détection de protéines antigéniques ou la détection de séquences de gènes codant les molécules situées à la surface des bactéries

3.43

polymorphisme d'un seul nucléotide

SNP

variant mononucléotidique (3.44) qui passe un seuil donné de qualité ou de fréquence

3.44

variant mononucléotidique

SNV

différences entre les nucléotides au même emplacement génomique de deux *isolats* (3.23) ou plus

3.45

souche

descendants d'un isolement unique en culture pure, généralement dérivés d'une seule colonie initiale sur un milieu de croissance solide

Note 1 à l'article: Une souche peut être considérée comme un *isolat* (3.23) ou un groupe d'isolats pouvant être distingué des autres isolats du même genre et de la même espèce par ses caractéristiques phénotypiques et génotypiques.

Note 2 à l'article: Voir la Référence [2].

3.46

validation

étude des caractéristiques de performance d'une méthode et démonstration objective que les exigences en termes de performance correspondent à l'utilisation prévue et spécifiée

[SOURCE: ISO 16140-1:2016, 2.81]

3.47

entrée de données validées

procédé automatisé garantissant que les données saisies dans une base de données sont correctes

3.48

vérification

démonstration apportant la preuve que la méthode validée lorsqu'elle est appliquée par un laborantin est conforme aux spécifications de la méthode déterminées lors de l'étude de validation et qu'elle est en adéquation avec l'utilisation prévue

[SOURCE: ISO 16140-3:2021, 3.21, modifié — La Note 1 à l'article a été supprimée.]

3.49 séquençage de génome entier WGS

processus permettant la détermination de la séquence d'ADN du génome d'un organisme à partir de l'ADN génomique total

4 Principe

4.1 Généralités

Les analyses de WGS des bactéries présentes dans la chaîne alimentaire humaine et animale consistent à cultiver l'isolat bactérien pur, à purifier l'ADN dans un laboratoire de microbiologie, à conduire des étapes de séquençage dans un environnement de séquençage approprié et à effectuer une analyse bioinformatique dans un environnement informatique distinct.

NOTE Le laboratoire de microbiologie, le service de séquençage et le service de bioinformatique peuvent appartenir à la même organisation.

4.2 Opérations réalisées en laboratoire: préparation et séquençage de l'échantillon

Il convient que la préparation et le séquençage de l'échantillon incluent les étapes suivantes:

- a) des informations relatives aux isolats séquencés, incluant des codes-barres pour les échantillons multiplexés, sont saisies dans les systèmes d'enregistrement appropriés, comme un système de management de l'information des laboratoires (LIMS) et/ou des fiches de description d'échantillon;
- b) des isolats purs (identifiés au moins au niveau du genre et dans l'idéal au niveau de l'espèce) sont mis en culture et leur ADN génomique est extrait;
- c) des librairies de séquençage d'ADN sont préparées à partir de l'ADN génomique, après avoir été soumis à un contrôle qualité (voir le [Tableau A.1](#) pour obtenir des recommandations relatives aux métriques de qualité et de quantité de l'ADN). Il convient que ce procédé inclue:
 - 1) la fragmentation de l'ADN, si nécessaire selon la technologie de séquençage appliquée;
 - 2) la ligature des index et adaptateurs, selon les protocoles de la technologie de séquençage appliquée;
 - 3) la quantification, la normalisation et le contrôle qualité de la librairie résultante;
 - 4) le regroupement des librairies en cas de cycles de séquençage multiplexé;
- d) les librairies (c'est-à-dire les groupes de librairies) sont séquencées;
- e) dans l'idéal, les métriques de qualité produites par l'équipement de séquençage sont enregistrées à chaque série d'échantillons d'ADN à des fins de surveillance des performances.

4.3 Analyse bioinformatique

4.3.1 Généralités

Les pipelines pour l'analyse bioinformatique peuvent être axés sur des prédictions *in silico* de phénotype (par exemple, virulence) ou la détection de groupes d'isolats génétiquement similaires (c'est-à-dire mêmes souche, type de séquence ou sérotype). Des pipelines basés sur des approches comparatives peuvent être utilisés pour détecter la présence et les états de marqueurs dans des données de séquençage brutes et assemblées pour constituer une souche *in silico* (par exemple, type de séquence) et des prédictions de phénotypes.

Les données de séquences d'isolats multiples peuvent être analysées en utilisant des méthodes d'analyse SNP, MLST ou de la distance k-mer afin d'identifier des groupes de bactéries étroitement apparentées. Les résultats de ces analyses peuvent être utilisés pour déduire les relations entre les isolats, qui peuvent être illustrées au moyen d'arbres phylogénétiques et de dendrogrammes.

4.3.2 Analyses des SNP

Pour les analyses des SNP, les lectures sont alignées sur une séquence de référence ou les lectures sont assemblées en contigs qui sont comparés. Pour déterminer les SNP, un filtre qualité est appliqué aux SNV afin d'identifier les emplacements des SNP.

4.3.3 Analyses des MLST

Pour les analyses des MLST, les lectures sont assemblées ou alignées. Les allèles sont identifiés, un filtre qualité leur est appliqué et ils sont comparés à une base de données de cgMLST ou de wgMLST.

4.3.4 Analyse de la distance k-mer

Les données de séquence d'isolats multiples peuvent être analysées en utilisant les méthodes de distances k-mers afin d'identifier des groupes de bactéries apparentées. Les analyses de k-mers ont l'avantage d'être très rapides, mais présentent certaines limites notamment en matière de précision (c'est-à-dire qu'elles sont applicables à la détermination des espèces, mais ne sont pas recommandées pour l'analyse détaillée de suivi des sources de contamination portant sur des souches étroitement apparentées).

iTeh STANDARD PREVIEW
(standards.itih.ai)

4.4 Formats de métadonnées et dépôt dans la base de données de séquences

Les enregistrements des métadonnées doivent être créés et stockés de manière sécurisée pour toutes les séquences. Il convient que les données de séquence et les métadonnées correspondantes soient formatées et documentées de manière uniforme. Ces métadonnées peuvent être partagées uniquement à la discrétion du propriétaire des métadonnées. Les données de séquence et les métadonnées correspondantes doivent être soumises à des considérations liées à la sécurité, aux coûts et aux bénéfices, aux droits de propriété intellectuelle, aux informations commerciales à caractère confidentiel, aux restrictions par contrat et/ou autres accords écrits contraignants.

NOTE Les données de séquence et/ou les métadonnées peuvent faire l'objet d'une licence et/ou d'une politique de confidentialité visant à protéger les informations privées ou exclusives.

Afin de favoriser les meilleures pratiques de gestion des données,^[3] le présent document propose des formats facultatifs pour les rapports de métadonnées, qui sont harmonisés selon une norme communautaire (par exemple, ontologies MDM ou OBO Foundry). Ces formats et ces normes facilitent la reproductibilité et la compréhension commune de la terminologie. Un ISO WGS Slim a été créé pour formater et fournir des valeurs pour les champs de métadonnées recommandés. Le WGS et les métadonnées sélectionnées peuvent être transférés (téléchargés) dans une base de données publique.

4.5 Validation et vérification du processus de WGS

L'ensemble du processus de WGS doit être validé pour fournir l'assurance que les méthodes sont adaptées à l'application prévue.

NOTE [L'Article 10](#) et le [Tableau 4](#) fournissent davantage de détails concernant la validation et la vérification du processus de WGS.

5 Recommandations générales pour le laboratoire

5.1 Isolement bactérien et extraction de l'ADN

Il convient que l'isolement bactérien et l'extraction de l'ADN soient réalisés dans un laboratoire de microbiologie générale adapté à la manipulation des bactéries spécifiques, y compris pathogènes. Pour la préparation d'une librairie de séquençage impliquant une amplification de l'ADN par réaction de polymérisation en chaîne (PCR), il convient que les étapes pré- et post-PCR soient réalisées dans des zones différentes ou séparées du laboratoire, afin d'éviter les intercontaminations.

5.2 Environnement du laboratoire

Les mouvements, les vibrations, la température et l'humidité de l'air peuvent interférer avec les performances de nombreux séquenceurs et il convient d'en tenir compte dans le positionnement de l'équipement dans le laboratoire. Il convient que les laboratoires consultent le guide de préparation du site du fabricant du séquenceur, afin d'obtenir des recommandations spécifiques.

5.3 Procédure opérationnelles normalisées et travaux non conformes

Il convient que les laboratoires tiennent à jour et respectent des procédures opérationnelles normalisées (PON), des documents de processus, des contrôles d'inventaire de réactifs et des dossiers de maintenance des équipements. Il convient que les PON incluent des procédures concernant l'utilisation de contrôles positifs et négatifs pour les étapes d'extraction de l'ADN, de préparation des librairies de séquences et de séquençage. Il convient que les PON incluent des procédures pour les opérations de surveillance portant sur la qualité des séries d'échantillons d'ADN et les erreurs (mauvaise identification d'échantillon ou contamination croisée).

En cas de mauvaise identification d'échantillon ou de contamination, l'analyse des causes premières des erreurs dans le séquençage doit être réalisée:

- a) en s'assurant que les séries d'échantillons d'ADN contenant des échantillons mal identifiés, ou des échantillons contaminés par des souches multiples, ne sont pas utilisées pour l'analyse bioinformatique à des fins d'interprétation des échantillons ni téléchargées dans les bases de données;
- b) en mettant en œuvre des mesures pour maintenir la qualité et éviter la répétition des erreurs.

5.4 Système de management de l'information des laboratoires

Les informations concernant les échantillons doivent être enregistrées dans un LIMS ou un système similaire de documentation et de suivi des informations.

5.5 Compétence du laboratoire

Il convient que les laboratoires tiennent à jour des enregistrements documentant la formation, l'éducation et les aptitudes des personnes réalisant le séquençage et l'analyse bioinformatique, ainsi que la politique de conservation des échantillons.

Il convient que le laboratoire surveille ses performances dans le cadre de l'analyse WGS par comparaison avec les résultats d'autres laboratoires, lorsqu'ils sont disponibles et s'il y a lieu. Il convient que cette surveillance soit planifiée et examinée chaque année dans l'idéal, et qu'elle inclue, sans toutefois s'y limiter, l'un des éléments suivants:

- a) participation à un programme d'essai d'aptitude;
- b) participation à des comparaisons interlaboratoires autres que des essais d'aptitude;
- c) vérification du procédé d'analyse en introduisant des échantillons «aveugles», ou des échantillons dont les caractéristiques ne sont pas connues de l'opérateur.