

## Annexe 2 Glossaire

### Sources

1. EFSA BIOHAZ panel. (2019). Whole genome sequencing and metagenomics for outbreak investigation, source attribution and risk assessment of food-borne microorganisms. *EFSA Journal*, 17, e05898.
2. Bertucci M. (2020) Bio-prospecting for new carbohydrate active enzymes from the microbiomes of the termite gut and anaerobic digester: an omics mediated approach. Ph.D. Thesis. UCLouvain, 250 pp.

### **Amplification en chaîne par polymérase**

Technologie permettant de réaliser rapidement et précisément de nombreuses copies d'un fragment spécifique d'ADN.

### **Amplification isotherme induite par boucle**

Technique d'amplification des acides nucléiques réalisée à température constante et reposant sur une structure tige-boucle.

### **Analyse 16S**

Étude des gènes de l'ARN ribosomique 16S conduisant à la structure de la communauté bactérienne de l'environnement étudié (microbiome).

### **Analyses « omiques »**

Analyse à haut débit de l'ADN (génomique), de l'ARN (transcriptomique), des protéines (protéomique) et/ou des métabolites (métabolomique) d'un environnement spécifique.

### **Approche « Une seule santé »**

Elle a été définie comme l'effort de collaboration de multiples disciplines (au niveau local, national et mondial) pour atteindre une santé optimale pour les personnes, les animaux et l'environnement.

### **Assemblage**

Résultat du processus d'alignement et de fusion des lectures de séquençage en séquences contiguës plus grandes (contigs).

### **Banque génomique (*library*)**

Ensemble de fragments de séquence amplifiables obtenus à partir de l'ADN cible original à séquencer.

### **Bioinformatique**

Collecte, stockage et analyse de séquences génomiques et d'expression génétique, à l'aide d'outils analytiques et de modélisation/prédiction.

### **Clade**

Monophylétique d'organismes constitué d'un ancêtre commun et de tous ses descendants directs.

### **Complexes clonaux (CC)**

Un complexe clonal est un groupe d'organismes apparentés basé sur la similarité de séquence d'une cible choisie.

**Contig**

Ensemble assemblé de séquences d'ADN se chevauchant et codant pour un ou plusieurs gènes.

**Couverture**

Nombre moyen de lectures de séquençage représentant un nucléotide donné, c'est-à-dire le nombre moyen de fois qu'une base nucléotidique spécifique est lue pendant le séquençage. Elle est calculée à partir de la longueur du génome original, du nombre de lectures, et de la longueur moyenne des lectures. Une couverture élevée peut diminuer les erreurs d'assemblage.

**Diversité génétique/génomique**

La diversité génétique est le nombre total de différences génétiques entre les organismes, tant entre les populations qu'au sein de celles-ci.

**Données épidémiologiques**

Ensemble de données décrivant l'unité d'échantillonnage (la date et le lieu de l'échantillonnage, le type et l'origine de l'échantillon, par exemple animal/alimentation/alimentation), qui doit être couplé aux données de typage moléculaire lorsqu'un isolat bactérien peut être obtenu à partir de l'échantillon.

**Électrophorèse en champ pulsé (ECP)**

Il s'agit d'une variante de l'analyse par endonucléase de restriction, une technique permettant de séparer de longs brins d'ADN à travers une matrice de gel d'agarose et de les visualiser sous forme de bandes. Le pouvoir discriminatoire de l'ECP dépend du nombre et de la répartition des sites de restriction dans le génome, y compris l'ADN extra-chromosomique, qui définissent le nombre et la taille des bandes dans le profil, et peut être augmenté en utilisant des endonucléases de restriction différentes ou combinées.

**Élément génétique mobile (EGM)**

Un morceau de matériel génétique capable de se déplacer à l'intérieur d'un génome ou de se transférer d'une cellule à une autre. Différents types d'EGM sont connus, par exemple les séquences d'insertion (IS), les transposons (Tn), les éléments intégratifs et conjugatifs (ICE), les intégrons, les introns, les plasmides ou les bactériophages.

**Génome de base**

Les parties du génome (gènes ou groupes de gènes) partagées par tous les membres d'un sous-ensemble défini (par exemple, l'espèce ou le genre) de bactéries.

**Génome entier (y compris le génome accessoire)**

Séquence(s) génomique(s) et leurs métadonnées associées.

**Identification du sérotype**

La classification des bactéries basée sur la détection antigénique ou séquentielle des molécules de surface des bactéries, en ce qui concerne *E. coli* se réfère spécifiquement à l'antigène somatique O du LPS.

**Laboratoire humide (*Wet laboratory*)**

Laboratoires où des produits chimiques, des médicaments ou d'autres matières biologiques sont testés et analysés, par opposition à un laboratoire sec où des analyses informatiques ou de mathématiques appliquées sont effectuées à l'aide de modèles générés par ordinateur.

**Laboratoires de référence européens (EURL)**

Laboratoires pour les aliments pour animaux et les denrées alimentaires, qui, entre autres : (i) sont chargés de fournir aux laboratoires nationaux de référence (LNR) des détails sur les méthodes d'analyse, y compris les méthodes de référence et les matériaux de référence et (ii) coordonnent, dans leur domaine de compétence, les dispositions pratiques nécessaires à l'application de nouvelles méthodes d'analyse et informent les LNR des avancées dans ce domaine. Les activités des laboratoires de référence doivent couvrir tous les domaines de la législation sur les aliments pour animaux et les denrées alimentaires et de la santé animale, en particulier ceux pour lesquels il est nécessaire de disposer de résultats d'analyse normalisés et harmonisés. Ces laboratoires sont soutenus dans le cadre du règlement (CE) n° 882/2004.

**Lignée**

Un groupe de bactéries qui partagent toutes un ancêtre, généralement utilisé pour définir les sous-groupes clonaux au sein des populations bactériennes.

**Marqueur génétique**

Gène ou séquence d'ADN qui peut être utilisé pour identifier une espèce ou un sous-type microbien particulier ou pour prédire un phénotype particulier (par exemple, potentiel de croissance, potentiel de virulence, résistance antimicrobienne).

**Métadonnées**

Données qui définissent et décrivent d'autres données. Les métadonnées peuvent être associées à la collection d'échantillons, à l'isolat ou à la séquence. Les métadonnées doivent être fournies en fonction du type d'échantillon (données épidémiologiques), des tests effectués ou des informations sur les opérations réalisées (données techniques) qui sont conservées en tant que description des données de séquençage stockées.

**Microbiome**

Communauté d'espèces microbiennes présentes dans un environnement spécifique.

**Multilocus variable-number tandem-repeat analysis (MLVA)**

Méthode utilisée pour effectuer un typage moléculaire en utilisant la variation naturelle du nombre de séquences d'ADN répétées en tandem que l'on trouve dans de nombreux loci différents du génome d'une variété d'organismes.

**Normalisation**

Processus de mise en œuvre et d'élaboration de normes techniques sur la base du consensus de différentes parties, notamment des entreprises, des utilisateurs, des groupes d'intérêt, des organismes de normalisation et des gouvernements.

**Phylogénie**

Fait référence aux relations évolutives entre les organismes.

**Pipeline**

Algorithmes informatiques pour la détection et l'interprétation de variants à partir de l'alignement de séquences génomiques.

**Polymorphismes de nucléotide unique (PNU)**

Un polymorphisme de nucléotide unique est une substitution d'un seul nucléotide qui se produit à une position spécifique dans le génome.

**Pouvoir discriminatoire**

Capacité à distinguer des souches qui devraient être considérées comme non apparentées dans le contexte épidémiologique de l'objectif de l'application.

**RAM acquise**

Capacité des bactéries à résister à l'activité d'un agent antimicrobien auquel elles étaient auparavant sensibles, c'est-à-dire des bactéries qui survivent à des concentrations antimicrobiennes plus élevées par rapport à la population de type sauvage. La résistance acquise résulte de la variation et/ou de l'échange de gènes par transfert horizontal de gènes.

**RAM intrinsèque (ou insensibilité)**

Tolérance innée à un agent antimicrobien ou à une classe d'antimicrobiens spécifique partagée par tous les membres d'un groupe bactérien (au niveau de l'espèce ou au-dessus), c'est-à-dire la population de type sauvage, en raison de caractéristiques structurelles ou fonctionnelles inhérentes.

**Reads**

Séquences d'acides nucléiques obtenues par séquençage à haut débit.

**Schéma de nomenclature des allèles**

Un ensemble de dénominations pour les variations alléliques (c'est-à-dire les séquences d'allèles) de chaque locus d'un ensemble de loci (c'est-à-dire le schéma) défini pour une espèce ou un genre.

**Schéma cgMLST**

Un nombre fixe et convenu de gènes pour chaque espèce ou groupe d'espèces étroitement apparentées, qui convient parfaitement pour normaliser le génotypage bactérien basé sur le *whole genome sequencing* (WGS).

**Séquençage à haut débit**

Désigne les techniques automatisées permettant le séquençage de centaines de millions de molécules d'ADN.

**Séquençage de nouvelle génération (NGS ou Next Generation Sequencing)**

Méthode à haut débit utilisée pour déterminer la séquence nucléotidique d'un génome ou d'une partie de celui-ci. Cette technique utilise les technologies de séquençage de l'ADN qui sont capables de traiter plusieurs séquences d'ADN en parallèle. Également appelé séquençage massivement parallèle.

**Sérotypage**

Schéma de classification basé sur la détection antigénique ou séquentielle des molécules de surface des bactéries. Pour les Enterobacteriaceae, il s'agit plus particulièrement des antigènes somatiques O du lipopolysaccharide (LPS), H flagellaire et K (ou Vir) capsulaire.

**Souche**

Une souche est considérée comme une culture pure, et une population uniforme de bactéries qui est génétiquement différente des autres populations de la même espèce, possédant un ensemble de caractéristiques définies. Une souche est souvent utilisée comme référence de laboratoire, ou maintenue par sous-culture.

**Souches persistantes**

Souches responsables de contaminations récurrentes observées le long de la chaîne alimentaire.

**Sous-espèces**

En taxonomie bactérienne, des sous-groupes d'une espèce qui diffèrent par leurs caractéristiques phénotypiques ou génotypiques.

**Sous-type**

Groupement de bactéries au sein d'une espèce qui partagent certaines caractéristiques, généralement obtenues par typage moléculaire (sous-type moléculaire ou génotypique). Les protéines, telles que les toxines, peuvent également être divisées en sous-types.

**Surveillance**

Conformément à la directive 2003/99/CE, le terme « surveillance » s'applique à un système de collecte, d'analyse et de diffusion de données sur l'apparition de zoonoses, d'agents zoonotiques et de la résistance antimicrobienne qui y est associée.

**Transcription du gène**

Séquence d'ARNm.

**Transcriptomique**

Étude de l'ensemble des transcriptions d'ARN qui sont produites par le génome.

**Transfert horizontal de gène (TH)**

Échange d'informations génétiques entre des cellules qui ne partagent pas nécessairement un parent commun par des processus autres que la descendance.

**Typage PNU**

Le génotypage PNU est la mesure des variations génétiques des polymorphismes de nucléotides uniques entre les membres d'une espèce.

**Type de séquence (TS)**

Désignation numérique d'un profil de séquence d'ADN allélique particulier. À l'origine, sept loci sont indexés pour lesquels chaque séquence unique pour chaque locus se voit attribuer un numéro d'allèle arbitraire et unique, qui est incorporé dans le profil allélique. Les TS sont utilisés dans les schémas de typage séquentiel multilocus comme unité de comparaison basée sur l'enregistrement des variants alléliques. Les isolats qui possèdent des allèles identiques pour toutes les séquences sont affectés à un type de séquence commun.

**Typage par analyse de séquences multilocus (MLST ou Multilocus sequence typing)**

Désigne le séquençage de plusieurs gènes ou d'un locus génétique, présentant un polymorphisme suffisant pour être utilisé dans un schéma de typage. Il s'agit idéalement de gènes « domestiques », c'est-à-dire de gènes codant pour des enzymes qui interviennent dans le métabolisme primaire de l'organisme en question et qui sont donc présents dans tous les isolats.

- Le rMLST (Ribosomal Multilocus Sequence Typing) est une approche similaire au MLST qui indexe la variation des gènes codant les sous-unités protéiques du ribosome bactérien (gènes *rps*) comme moyen d'intégrer la taxonomie et le typage microbiens.
- Le wgMLST (Whole genome MLST) est défini comme un ensemble non redondant de gènes présents dans un ensemble de génomes représentant une espèce, un peu comme un pangénome. Par conséquent, un schéma wgMLST inclut un plus grand nombre de gènes et peut également inclure des éléments hautement variables tels que des gènes répétitifs et des pseudogènes, s'ils sont présents dans tout le génome inclus (Pearce *et al.*, 2018).
- Les schémas du cgMLST (Core genome MLST) équilibrent le nombre de loci utilisés dans un schéma avec la résolution maximale possible, en incluant les loci présents dans la majorité des isolats (entre 95 % et 99 %) dans un groupe donné de bactéries. Idéalement, ces gènes reflètent la véritable généalogie au sein de l'espèce et ne changent pas de présence au fil du temps. Les éléments qui ne sont pas soumis à des pressions de sélection strictes, tels que les gènes répétitifs et les pseudogènes, doivent être exclus (Pearce *et al.*, 2018).

**Validation**

Établissement des caractéristiques de performance d'une méthode et fourniture d'une preuve objective que les exigences particulières pour une utilisation prévue spécifiée sont remplies. Les résultats obtenus par une méthode alternative doivent démontrer qu'ils sont comparables à ceux obtenus par la méthode de référence.

**Whole Genome Sequencing (WGS)**

Processus de détermination de la séquence d'ADN du génome d'un organisme en utilisant l'ADN génomique total comme entrée.