Appendix 2 Glossary

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.

Acquired AMR

Ability of bacteria to resist the activity of an antimicrobial agent to which it was previously susceptible, i.e. bacteria that survive at higher antimicrobial concentrations compared to the wild-type population. Acquired resistance results from gene variation and/or exchange by horizontal gene transfer.

Allele nomenclature schema

A collection of name designations for allelic variations (i.e. allele sequences) of each locus of a set of loci (i.e. schema) defined for a species or genus.

Assembly

Output from process of aligning and merging sequencing reads into larger contiguous sequences (contigs)

Bioinformatics

Collection, storage, and analysis of genome sequences and gene expression, using analytical and modeling/predictive tools.

cgMLST scheme

A fixed and agreed upon number of genes for each species or group of closely related species that is ideally suited to standardize whole genome sequencing (WGS) based bacterial genotyping.

Clade

Monophyletic of organisms that consisting of a common ancestor and all its lineal descendants.

Clonal complexes (CC)

A clonal complex is a group of related organisms based on the sequence similarity of a chosen target.

Contig

Assembled set of overlapping DNA sequences coding for one or more genes

Core genome

Those parts of the genome (genes or gene clusters) shared by all members of a defined subset (e.g. species or genus) of bacteria.

Coverage

Is the average number of sequencing reads representing a given nucleotide, i.e. the average number of times a specific nucleotide base is read during sequencing. It is calculated from the length of the original genome, the number of reads, and the average read length. A high coverage may decrease errors in assembly.

Discriminatory power

The ability to distinguish between strains that should be considered unrelated in the epidemiological context of the application purpose.

Epidemiological data

Dataset describing the sample unit (e.g. date and place of sampling, type and origin of sample, for example animal/food/feed), which needs to be coupled with molecular typing data when a bacterial isolate can be obtained from the sample.

EU Reference Laboratories (EURLs)

Laboratories for feed and food, which, among others: (i) shall be responsible for providing national reference laboratories (NRLs) with details of analytical methods, including reference methods and reference materials and (ii) coordinating, within their area of competence, practical arrangements needed to apply new analytical methods and informing NRLs of advances in this field. The activities of reference laboratories should cover all the areas of feed and food law and animal health, in particular those areas where there is a need for standardized and harmonized analytical results. These laboratories are supported in the scope of Regulation (EC) No 882/2004.

Gene transcript

mRNA sequence.

Genomic/genetic diversity

Genetic diversity is the total number of genetic differences between organisms both between and within populations.

Genetic marker

Gene or DNA sequence which can be used to identify a particular microbial species or subtype or to predict a particular phenotype (e.g. growth potential, virulence potential, antimicrobial resistance).

High throughput sequencing

Refers to automated techniques allowing the sequencing of hundreds of millions of DNA molecules.

Horizontal gene transfer (HGT)

Exchange of genetic information among cells that do not necessarily share a common parent by processes other than descent.

Intrinsic AMR (or insensitivity)

Innate tolerance to a specific antimicrobial agent/class shared by all members of a bacterial group (species level or above), i.e. the wild-type population, due to inherent structural or functional characteristics.

Library

Set of amplifiable sequence fragments obtained from the original target DNA to be sequenced.

Lineage

A group of bacteria all of which share an ancestor, usually used to define clonal subgroups within bacterial populations.

Loop-mediated isothermal AMPlification

Nucleic acid amplification technique performed at constant temperature relying on a stem-loop structure.

Metadata

Data that defines and describes other data. Metadata can be associated with the sample collection, with the isolate or with the sequence. Metadata should be supplied according to the sample type (epidemiological data), the testing performed or the operations performed information (technical data) that is held as a description of stored sequencing data.

Microbiome

Community of microbial species present in a specific environment.

Mobile genetic element (MGE)

A piece of genetic material capable of moving its location within a genome or transfer from one cell to another cell. Different types MGE are known, e.g. Insertion Sequences (IS), transposons (Tn), Integrative conjugative elements (ICE), integrons, introns, plasmids or bacteriophage.

Monitoring

In agreement with the Directive 2003/99/EC, the term 'monitoring' will be applied to a system of collecting, analyzing and disseminating data on the occurrence of zoonoses, zoonotic agents and antimicrobial resistance related thereto.

Multilocus sequence typing (MLST)

Refers to the sequencing of multiple genes or a genetic locus, displaying enough polymorphism to be used in a typing scheme. These are ideally 'house-keeping' genes, i.e. genes encoding enzymes that are involved in primary metabolism of the organism in question and which are therefore present in all isolates.

- Ribosomal Multilocus Sequence Typing (rMLST) is a similar approach to MLST that indexes variation of the genes encoding the bacterial ribosome protein subunits (*rps* genes) as a means of integrating microbial taxonomy and typing.
- Whole genome MLST (wgMLST) is defined as a non-redundant set of genes that are present across a set of genomes representing a species, akin to a pan-genome. Consequently, a wgMLST scheme includes a greater number of genes and may also include highly variable elements such as repetitive genes and pseudogenes, if they are present in all included genome (Pearce *et al.*, 2018).
- Core genome MLST (cgMLST) schemes balance the number of loci used in a scheme with the maximum possible resolution, by including those loci present in the majority of isolates (ranging from 95% to 99%) in a given grouping of bacteria. Ideally these genes reflect the true genealogy within the species and do not change presence over time; and elements not under strict selection pressures, such as repetitive genes and pseudogenes should be excluded (Pearce *et al.*, 2018).

Multilocus variable-number tandem-repeat analysis (MLVA)

Method used to perform molecular typing utilizing the naturally occurring variation in the number of tandem repeated DNA sequences found in many different loci in the genome of a variety of organisms.

Next generation sequencing (NGS)

A high-throughput method used to determine the nucleotide sequence of a genome or of a portion of it. This technique utilizes DNA sequencing technologies that are capable of processing multiple DNA sequences in parallel. Also called massively parallel sequencing.

OMICS analyses

High throughput analysis of DNA (genomics), RNA (transcriptomics), proteins (proteomics) and/or metabolites (metabolomics) of a specific environment.

One Health

Has been defined as the collaborative effort of multiple disciplines - working locally, nationally, and globally - to attain optimal health for people, animals and the environment.

Persistent strains

Strains responsible for recurrent contaminations observed along the food chain level.

Polymerase Chain Reaction

Technology to make numerous copies of a specific fragment of DNA quickly and accurately.

Phylogeny

Refers to the evolutionary relationships between organisms.

Pipeline

Computational algorithms for detecting and interpreting variants from alignment of genomic sequences.

Pulsed-field gel electrophoresis (PFGE)

Is a variant of the restriction endonuclease analysis (REA); a technique to separate long strands of DNA though an agarose gel matrix and visualized as bands. The discriminatory power of PFGE depends on the number and distribution of restriction sites throughout the genome, including extra-chromosomal DNA, which define the number and sizes of bands in the profile, and can be increased by using different or combinations of restriction endonucleases.

Reads

Nucleic acid sequences obtained from high throughput sequencing.

Sequence type (ST)

Numerical designation for a particular allelic DNA sequence profile. Originally, seven loci are indexed for which each unique sequence for each locus is assigned an arbitrary and unique allele number, which is incorporated into the allelic profile. STs are used in multilocus sequence typing schemes as the unit of comparison based on the record of allelic variants. Isolates that possess identical alleles for all sequences are assigned to a common Sequence Type.

Serogroup identification

Classification of bacteria based on the antigenic or sequence-based detection of bacteria surface molecules, with respect to *E. coli* refers specifically to the LPS somatic O antigen.

Serotyping

Classification scheme based on the antigenic or sequence-based detection of bacteria surface molecules, for the Enterobacteriaceae refers specifically the lipopolysaccharide (LPS) somatic O, the flagellar H and the capsular K (or Vir) antigens.

Single Nucleotide Polymorphism (SNP)

A single-nucleotide polymorphism is a substitution of a single nucleotide that occurs at a specific position in the genome.

16S analysis

Study of the bacterial 16S rRNA genes leading to the bacterial community structure of the studied environment (microbiome).

SNP typing

SNP genotyping is the measurement of genetic variations of single nucleotide polymorphisms between members of a species.

Standardization

Process of implementing and developing technical standards based on the consensus of different parties that include firms, users, interest groups, standards organizations and governments.

Strain

A strain is considered a pure culture, and a uniform population of bacteria that is genetically different from other populations of the same species, possessing a set of defined characteristics. A strain is often used as a laboratory reference, or maintained by subculture.

Subspecies

In bacterial taxonomy, subgroups of a species that differ in their phenotypic or genotypic features.

Subtype

A grouping of bacteria within a species that share certain characteristics, usually derived by molecular typing (molecular or genotypic subtype). Proteins, such as toxins, may also be divided into subtypes.

Transcriptomics

Study of the complete set of RNA transcripts that are produced by the genome.

Wet laboratory

Laboratories where chemicals, drugs or other biological matter are tested and analyzed, in contrast to a dry laboratory where computational or applied mathematical analyses are done with assistance of computer-generated models.

Validation

Establishment of the performance characteristics of a method and provision of objective evidence that the particular requirements for a specified intended use are fulfilled. Results obtained by an alternative method should demonstrate that they are comparable to those obtained by the reference method.

Whole genome (including accessory genome)

Genomic sequence(s) and their associated metadata.

Whole genome sequencing (WGS)

Process of determining the DNA sequence of an organism's genome using total genomic DNA as input.