Chapter 12
Genome-Based Prediction of Bacterial Antibiotic Resistance

Deepali Kalambhe, Lokesh K. M., Gourab Basak, Abhilash Jadhao, and Sumeet Singh

Abstract Genome-based techniques, especially whole-genome sequencing (WGS), present a huge potential to predict antimicrobial resistance in the microbes. The advancement in the inexpensive DNA sequencing technology, bioinformatics tools and handy online databases on nucleotide sequence has transformed the entire diagnostic microbiology and bacterial investigation. Genome sequencing in conjugation with the online bioinformatics tools helps in predicting real-time AMR determinants. This approach allows establishing global pathogen surveillance and AMR tracking based on genomics which is essential to combat, control and prevent the increasing threat of AMR. Tools genome-based surveillance tools are either available at public genome data domains or can be operated locally. Public database centres such as NCBI and European Nucleotide Archive (ENA) allow online submission of nucleotide sequencing data along with phenotypic antimicrobial susceptibility data. However, there is a need for optimization of databanks as well as phenotypic predictions based on the genomic data. This chapter discusses the latest genome-based techniques, bioinformatics tools and genomic databases for predicting antimicrobial resistance (AMR).

Keywords Whole-genome sequencing · Bioinformatics · Antimicrobial resistance · Next-generation sequencing · In silico analysis

D. Kalambhe (✉) · L. K. M. · G. Basak
Centre for One Health, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India

A. Jadhao
Department of Veterinary Pathology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India

S. Singh
Department of Veterinary Microbiology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India

C. S. Mukhopadhyay et al. (eds.), *Biotechnological Interventions Augmenting Livestock Health and Production*, Livestock Diseases and Management,
https://doi.org/10.1007/978-981-99-2209-3_12

12.1 Introduction

With the discovery of penicillin as the world's first antibiotic in the late 1920s, it has made a renaissance in the medical therapeutics and science; subsequently discovery of various antibiotics from time to time had saved millions of lives across the globe. But alike every coin has two opposite sides, antibiotic resistance also came up along with it in the path within a couple of years. Antimicrobial resistance (AMR) is an emerging and rapidly growing public health concern. According to Food and Agriculture Organization of the United States, AMR is the microbe's ability to persist and continuously grow in presence of the antimicrobial compounds that are intended to inhibit or to kill them (https://www.fao.org/antimicrobial-resistance/background/what-is-it/en/). The transmission of AMR strains of microbes (bacteria) from animals to humans is well documented which are termed as zoonotic AMR strains (Economou and Gousia 2015). Thus, AMR threatens the practice of medicine inferring animal and/or human health risks. Besides, it also possesses implications in food safety and security. Reports highlight that the AMR organisms in United States are responsible for more than two million infective cases along with 23,000 mortalities per annum (CDC 2013). Similarly, Europe is no behind with 25,000 live-loses annually keeping in pace with the former nation (Gelband et al. 2015). Globally, 0.7 million annual deaths occur presently which is expected to touch ten million by 2050 ceasing 66 trillion financial wealth. Moreover, it is been three long decades since the introduction of any new antibiotic.

12.2 Why Antimicrobials Are Used in Animal Production and How AMR Is Knocking?

Diseases are the main reasons for the use of antimicrobials. Though if modified environmental hygiene, proper balanced nutrition and more importantly husbandry and management practices are espoused unanimously, the animals can be preven

from any infectious or metabolic diseases in turn exempting them from any antimicrobials. Besides, the major concern these days is about the use of such chemicals as growth promoters and production enhancers, which gradually keep on bioaccumulation producing antimicrobial resistance (Fig. 12.1). In this regard, limited access to health professionals, oversight and regulation of their use and incomplete completion of drug regiment amplify the condition many folds whereas restricted training provisions for these experts time to time add a cheery on the top. As a result, owners fill the gap with over-the-counter drugs, which really pose a risk impose in the health of animals, man as well as in the environment. These may sometimes also because of substandard and/or falsified drugs which fail to fulfil its target in place encourage the microbes to get acclimatize with the condition provided. Besides, lack of knowledge, awareness and proper use help in increasing height of the graph.

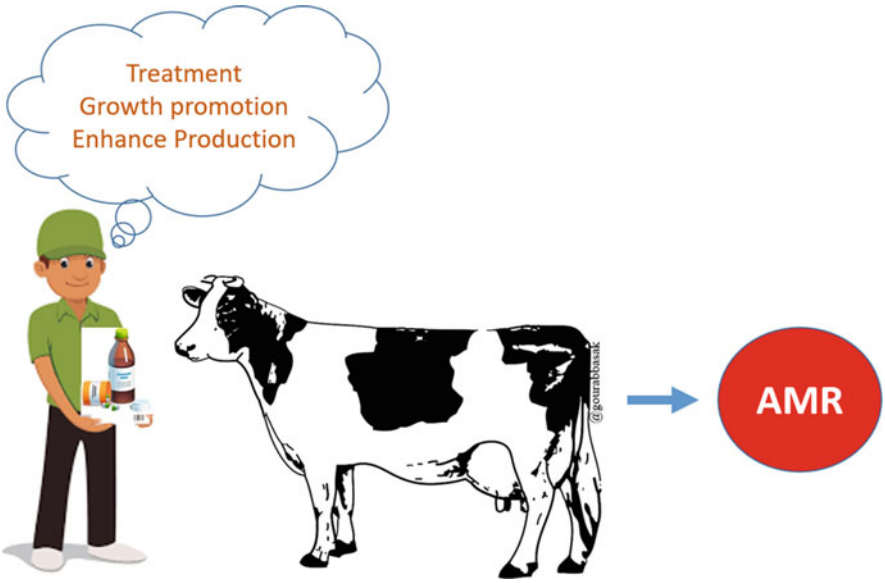

Fig. 12.1 Schematic representation of the drivers behind antimicrobials usage in animal production and their impact

As a result, detection of AMR organisms and the responsible genes became mandatory to arrest the proficiency of AMR contractions. Monitoring of antimicrobial resistance in foodborne pathogens isolated from clinical, food and environmental samples again becomes very important. As because monitoring aid in recognizing and mitigating resistant strains spread from animals to humans inferring public health risk. As the morbidity and economic burden are increasing with the rise in resistance rates, therefore, to guide treatment decisions, precise detection of antibiotic resistance is required. Currently, approaches including phenotypic detection and rapid genomic detection methods like PCR for resistance determinants are in the limelight. Clinical laboratories usually use culture-based antimicrobial susceptibility testing (AST) as their principal approach. But, only or solely phenotypic detection would not help in the long run as more precision can be drawn efficiently through the genotypic methods. In this context, the genome-based detection of Antimicrobial Resistance/Susceptibility Testing offers the potential benefit for rapid, reliable and precise predictions of every known resistance phenotype for a strain.

12.3 Whole-Genome Sequencing for Antimicrobial Susceptibility Testing (WGS-AST)

Whole-genome sequencing for antimicrobial susceptibility testing enabled to assess genes accounting AMR, their location along with their potentiality for multidrug resistance and rapid dissemination (Karp et al. 2017). Single-nucleotide variants,

insertions/deletions, copy number alterations and significant structural variants can easily be detected by whole-genome sequencing. In silico examination for the presence of antimicrobial genes can be done using software and databases including Resistance Gene Identifier in Comprehensive Antimicrobial Resistance Database (CARD), Antibiotic Resistance Database and Resistance Gene Finder (ResFinder). CDC, FDA, FSIS and ARS actively collaborate on nationwide AMR surveillance for near 20 years in the National Antimicrobial Resistance Monitoring System in the United States (Karp et al. 2017).

Previously used Sanger sequencing was highly accurate for relatively shorter DNA fragments, for the longer DNA stretches, the process was time-consuming and involved multiple reactions (Sanger et al. 1977). It took several years to sequence the bacterial genome and employed millions of dollars. With the advent of next-generation sequencing in the early 2000, the revolutionary approach linked DNA sequencing to food safety and public health surveillance on a routine basis. Whole-genome sequencing is an advanced and comprehensive method for determining the complete DNA sequence of an organism. Using next-generation sequencing (NGS) techniques, like Illumina and Nanopore, the sequence of complete chromosomal, plasmid and mitochondrial DNA can be determined in a single reaction and much less time. The entire bacterial genome can be sequenced in small random fragments (1000 bp) multiple times in a single reaction (Vincent et al. 2013). The complete DNA sequence is determined with the use of state-of-the-art bioinformatics tools. WGS provides a very high-resolution base to base view of the genome. Continuous advancement in biotechnology, bioinformatics and information technology enhances capability of using NGS to augment food safety and public health (Allard et al. 2019). This technique can determine a large amount of data in a short time on a routine basis and facilitates to maintain a database for further reference.

- *Base to base (Single-Nucleotide Polymorphism, SNPs)*
- *Gene to gene (Multilocus Sequence Typing, MLST)*

Base to base comparison of the test strain with the reference strain provides the nucleotide difference at specific positions owing to the genetic mutations. SNPs occur throughout the genome in both coding and non-coding regions. The SNP identification reference strain choice is quite significant and can be customized to any given situation (i.e. closely related to the outbreak strain) providing a more precise SNP difference-assessment in an outbreak setting. SNPs profiles of all the isolates are compared pairwise and displayed in the phylogenetic tree. This approach uses all the information of the genome (coding and non-coding regions) and provides greater accuracy for the reconstruction of a strain phylogeny (Pettengill et al. 2014).

Gene to gene approach works by assessing the sequence variation in the coding region of the genes. The accessibility of number and nature of the genes to be assessed to any given situation makes this approach more flexible. In this, genes identified in test strain are compared against the reference databases of genes with all known gene variants of the species. Each unique allele sequence is given a number and the genome is compared based on allele numbers. This approach cannot assess

the variation in non-coding regions. It is more popular in clinical settings as fewer bioinformatics skills are required in this analysis approach.

SNPs provide precise information and the flexibility to choose more closely associated reference strains for precise assessment of relatedness. It is more discriminatory as it assesses non-coding regions also. But for practical purposes, both the approaches are equally discriminatory and epidemiologically concordant (Brown et al. 2018).

Various DNA genome databases are available to share information regarding outbreaks identification and diagnostics of causative agents, which facilitates the early recognition and investigation of international foodborne outbreaks. PulseNet by Centre for Disease Control and Prevention, Food Safety Inspection System (FSIS) by United States Department of Agriculture and GenomeTrakr by U.S. Food and Drug Administration are the networks to utilize whole-genome sequencing for pathogen identification. These databases collect and share genomic and geographic data from foodborne pathogens which can be accessed by researchers and public health officials for real-time comparison and analysis of an outbreak (Stevens et al. 2017). These networks promise speed for foodborne illness outbreak investigations and reduce foodborne illnesses and deaths. Data are stored in the standardized method so that the large volume is greatly reduced to exchange with any part of the world, and least post-processing is required ensuring fast comparison of data from databases in different regions of the world.

WGS has been used in routine public health surveillance since 2014 for *Listeria monocytogenes*, surveillance for *Campylobacter* outbreaks incorporated in 2018, followed by Shiga-toxin producing *E. coli* and *Salmonella* in 2019. PublicNet, FSIS, GenomeTrakr work along with other public health partners to make improvements in hazard detection characterization methods. These networks use WGS for isolation, characterization and surveillance of outbreaks to detect and prevent contamination events and follow foodborne illness outbreaks (Brown et al. 2018).

Genome sequence possesses many advantages as well as has the potential symbiotic interaction between genomics and phenotypic-based AST. Following selective cultivation of the bacterium of interest from a clinical sample, WGS-AST is performed. It is also possible to perform AST after direct shotgun sequencing of clinical samples (Meta genomic-AST). Due to the presence of potentially low amount of pathogen of interest relative to the host DNA, metagenomics is more difficult, expensive and prone to false-negative results. As DNA sequencing is easier and faster than acquiring enough culture growth for phenotypic assessment, slow-growing or difficult-to-culture bacteria (such as *Mycobacterium tuberculosis*) are the key early targets for metagenomic-AST (Doyle et al. 2018; Votintseva et al. 2017). WGS-AST can determine the antibiotic resistance phenotypes of the entire genome simultaneously and phenotypes where multiple loci contribute can be easily screened, unlike culture-based AST or nucleic acid amplification tests (NAATs). The later are often limited by the number of resistant phenotypes that can be determined from one test (except for multiplex PCRs). The genome sequence data are digitally saved and can be queried for additional purposes once it is obtained (Feijao et al. 2018) (Fig. 12.2).

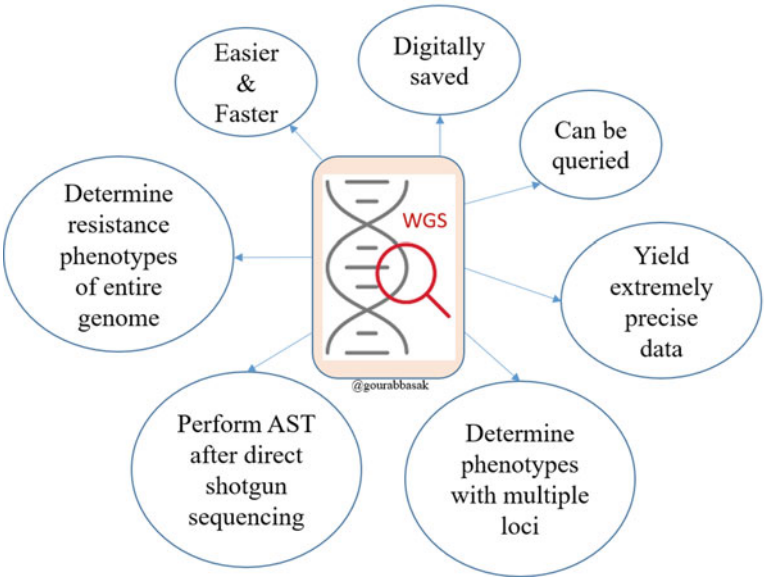

Fig. 12.2 Whole-genome sequencing as tool for antimicrobial susceptibility testing

Genomes can be sequenced to extremely high depths, yielding extremely precise sequence data. Unlike NAATs, template amplification does not rely on primer specificity, lowering the risk of false-negative results. The collection of genomes in clinical laboratories has resulted in a data source that may be utilized to track disease evolution (Gardy and Loman 2018). If new antibiotic resistance loci are discovered, these databases may be searched right away to see how long these genes have been circulating and how they got into the clinical use.

12.4 Next-Generation Sequencing (NGS) Technologies Driving WGS-AST

Next-generation sequencing (NGS) is a high-throughput, low-cost and quick second-generation sequencing technology while whole-genome sequencing (WGS) is a comprehensive method of analyzing the entire genomic DNA of a cell at a single time by using sequencing techniques such as Sanger sequencing, shotgun approach or high-throughput NGS sequencing. Second-generation devices, such as the Illumina sequencing-by-synthesis technology, drastically lowered the cost of data generation allowing for large-scale sequencing of thousands of pathogen

genomes and the application of shotgun metagenomics for clinical diagnosis. Illumina sequencing reads are short (300 bp), paired-end and have a low per-base error rate (usually 0.1%). De novo assembly generally results in genomes fragmented into many contigs and collapsed repeat regions, despite the fact that Illumina sequencing provides for extensive shotgun coverage with high consensus accuracy.

Longer reads are produced using third-generation single-molecule sequencing, as demonstrated by Oxford Nanopore Technology (ONT) and Pacific Biosciences (PacBio) technologies. However, extensive read genome assemblies include fewer gaps and generally span long repeat regions, allowing complicated structural features like tandem repeats and nested insertions to be resolved (Giordano et al. 2017). Third-generation technologies have a greater cost per base and higher per-base error rates (5 to 15%) than Illumina, despite improvements in chemistry and base calling algorithms narrowing the gap (Rhoads and Au 2015; Lu et al. 2016). As existing technologies mature and become more cost-effective, and new approaches emerge, the future of clinical sequencing is continuously shifting. Illumina is currently the most popular WGS-AST platform. However, the estimated minimum cost of $80 per genome is still too high for clinical laboratories to use on a regular basis.

12.5 WGS-AST Based on Searching Catalogues of Resistance Loci

Using a "rules-based" classification based on the presence of one or more known antimicrobial resistance (AMR) genes or mutations is the easiest technique. Cross-referencing the genomic sequence against databases of antibiotic resistance determinants is required for this. The majority of the databases were created via curation of the literature on molecular genetic studies that link antibiotic resistance phenotypes to genes (Xavier et al. 2016).

12.5.1 Multispecies Database

CARD	McArthur et al. (2013), Jia et al. (2017)
Resfinder	Zankari et al. (2012)
Pointfinder	Zankari et al. (2017)
ARG-ANNOT	Gupta et al. (2014)
MEGARes	Liu and Pop (2009)
Resfams	Lakin et al. (2017)
RAST	Gibson et al. (2015)
Bacterial antimicrobial resistance reference gene database	BARRGD, https://www.ncbi.nlm.nih.gov/bioproject/313047

There are also databases created specifically for single species, such as Dream TB (Sandgren et al. 2009) and MUBII-TB-DB (Flandrois et al. 2014) for Mycobacterium tuberculosis. The data produced at two phases in the next-generation sequencing method, raw sequence data and assembled contigs, are used by software tools for rules-based antibiotic resistance catalogue matching. In terms of speed and precision, each has its own set of trade-offs. There is good agreement between what is known about the genetic basis of resistance and the resistance phenotype for many species and antibiotic resistance phenotypes. For numerous characteristics across several pathogen species, rules-based WGS-AST has been proven to have high sensitivity and specificity (95 per cent) (Bradley et al. 2015; Clausen et al. 2016; Mason et al. 2018). Despite the fact that the number of strains examined and the within-species genetic diversity of the test set varied greatly between investigations (Kos et al. 2015). Miotto et al. (2017) divided the predictive power of M. tuberculosis mutations into high, moderate and minimal confidence in an exhaustive investigation of the genetic basis of resistance in M. tuberculosis. As a result, rule-based approaches may not always be enough for accurate WGS-AST.

12.6 Advantages and Disadvantages of Whole-Genome Sequencing

SN	Advantages	Disadvantages	Authors
1	Provides huge genomic data in a single assay	High cost and resources	Ellington et al. (2017), Meienberg et al. (2016)
2	AMR bacteria can be typed and tracked using unique allele profiles	Processing and storing a large amount of data	Hendriksen et al. (2019), Meienberg et al. (2016), Gonzaga-Jauregui et al. (2012)
3	Help to investigate a drug-resistant foodborne and inconsistent resistance patterns among indistinguishable PFGE types of *Salmonella* serovar	Sanger sequencing is required to validate genetic variations	Gong et al. (2018), Gonzaga-Jauregui et al. (2012)
4	Reveal the co-carriage of individual genes generating diverse MDR patterns, allelic trends over time, horizontal transfer and distribution	Large number of variants can be detected in non-coding areas, which may or may not be relevant	Ellington et al. (2017), Gong et al. (2018), Gonzaga-Jauregui et al. (2012)
5	Defines MDR as resistance to three or more drug classes	Physicians are not familiar in interpreting genomic data	Ng and Kirkness (2010)
6	Sequence-based surveillance allows more precise definition of multidrug resistance (MDR) when compared phenotypic method	The function of the majority of gene in the human genome is unclear hence, much of the "knowledge" found in a human genome sequence is currently useless	Hendriksen et al. (2019), Ng and Kirkness (2010)

(continued)

SN	Advantages	Disadvantages	Authors
7	Sequencing allows identification of a single-nucleotide variants, insertions/deletions, copy number alterations and significant structural variants and monogenic disease	Enormous amount of data are generated. Policies and security procedures to protect the privacy and security of this data are still being developed	Abdelbary et al. (2017), Ng and Kirkness (2010)
8	Used to diagnose genetic mutations and to identify genetic carriers of recessive disorders like cystic fibrosis	Requires high informatics capacity and special software	Katsanis and Katsanis (2013), Ng and Kirkness (2010)
9	WGS in cancer research allow the identification of genetic drivers of tumours and new biological therapies	Large incidental findings may increase the risk of overdiagnosis	Zhao et al. (2019), Mazzarotto et al. (2020)
10	Both coding and non-coding variations are detected	Potential inclusion of non-validated genes in genetic testing	Meienberg et al. (2016), Mazzarotto et al. (2020)
11	Detects structural variants	False-positive findings	Gong et al. (2018), Gonzaga-Jauregui et al. (2012), Mazzarotto et al. (2020)

WGS in cancer research could lead to the discovery of new biological therapies and genetic causes of tumours.

12.7 Antibiotic Resistance Techniques Based on Bioinformatics

Bioinformatics uses computer software tool such as sequence and structural alignment application that develop extrapolations and process sequence data as reads or assemblies for finding novel biology from plethora of biological data generated from gene sequences, cell populations, or protein samples (Luscombe et al. 2001). Bioinformatics has become an essential approach for the consolidation of knowledge on antibiotic resistance. Bioinformatics approaches such as molecular docking are commonly used to evaluate ligand–protein interactions and to quantify binding energy during the docking process. It can be utilized to investigate the links between traditional mathematical modelling and omic scope predictions, as well as specific features of the immune system (Rapin et al. 2010). Swiss-model (online homology modelling), Autodock Vina (ligand-protein docking), Avogadro (ligand energy minimization) and Chimera (3D docked complexes) are a few examples of regularly used bioinformatics tools for understanding the antimicrobial profile of bacteria. Bioinformatics method for reducing antibiotic resistance has expanded in recent years, and it now includes bioinformatics techniques based on whole-genome sequencing (Ndagi et al. 2020).

Now a days the whole-genome sequencing (WGS) of pathogens is in vogue due to its easy accessibility, rapid increase in output, rapid analysis and reduced cost (Quainoo et al. 2017). With the advancement in sequencing technologies and analysis tools, the genome sequencing offers a suitable framework for scientific advancement, notably in biomolecular modelling and medication creation, with a focus on antibiotic resistance (Gwinn et al. 2019). Genotyping technologies enable better understanding of disease transmission hence helpful for epidemic management (Quainoo et al. 2017). The whole-genome analysis of bacteria by sequencing provides better insight of related lineages and outbreaks in hospitals (Biswas et al. 2008; Quainoo et al. 2017).

Deoxyribonucleic acid (DNA) sequencing is an excellent platform for protein modelling and drug development (Blundell et al. 2006). Advancements in genome sequencing, protein expression, high-throughput crystallography and nuclear magnetic resonance (NMR) have revolutionized the possibilities for using protein three-dimensional structures to speed drug development. Structural biology and bioinformatics have well-established functions in target identification of the remarkable bacteria resistance in our environment (Blundell et al. 2006). High-throughput structure determination tools offer effective strategies for combating bacterial resistance as it permits screening of complicated bacterium proteins for the identification of lineages, sequence alignment and three-dimensional modelling structures.

12.8 In Silico Analysis of Serovar, Serogroup and Antigenic Profile

With the progression of genetic studies, the inclination of microbiologists have increased for the whole-genome sequencing for the purpose of genotyping just like the molecular serotyping which has replaced the traditional serotyping method. Presently, the multilocus sequence typing (MLST) and serovar-specific gene markers or DNA fragments are more popular in silico serovar prediction techniques for identifying the genes expressing surface antigen. However, these serovar-specific gene markers or DNA fragments can distinguish only a few serovars. This shortcoming was overcome by Zhang et al. (2019) who developed in silico serovar prediction technique that compares 1089 genomes covering 106 serovars to a set of 131 serovar-specific gene markers. According to available literature, the method best suits as a good diagnostic tool for culture-independent and metagenomics methodologies, and also as an alternative for confirming other genome-based investigations. This set of bioinformatics procedures is beneficial for identifying a certain type of gene marker and may help in the development of more cost-effective molecular assays for detecting specific gene markers of all major serovars.

12.9 In Silico Plasmid Identification

Plasmids are the mobile genetic element that are placed external to chromosomal DNA and are capable of autonomous replication. Plasmids are either circular or linear in nature and are abundantly found in microorganisms especially in the bacterial and archaeal domains (Jesus et al. 2019). Plasmids often carry genes for virulence and resistance and by virtue of its mobility it serves as "vehicle" for the transport of genetic information between the bacterial species and genus (Frost et al. 2005). Thus, plasmids are important for the acquisition of virulence traits and spread of antibiotic resistance (Jesus et al. 2019). The growing incidence of plasmid-mediated microbial resistance against the commonly used therapeutic drugs is a big concern for the spread of resistance across the human and veterinary healthcare settings. Therefore, it is absolutely necessary to investigate the molecular epidemiology of plasmids along with the molecular epidemiology of different bacterial strains. There are few online available tools such as cBar, PLACNET, plasmidSPAdes and Recycler (Zhou and Xu 2010; Lanza et al. 2018; Antipov et al. 2016; Rozov et al. 2017) that may be used to extract and assemble plasmids data for specific markers or unique characteristic sequences from high-throughput sequencing (HTS) data. Some other software tools such as plasmidFinder and MOB-suite Plasmid Profiler are also available to reconstruct or detect plasmids in HTS data; however, these are quite difficult for the users to understand the list of hits and evaluate the impact of these alleged plasmids on the host bacteria (Carattoli et al. 2017; Robertson and Nash 2018; Zetner et al. 2017).

The National Centre for Biotechnology Information (NCBI) has approximately 13,924 reference plasmid sequence (RefSeq) entries stored in its data bank though, with a dearth of essential tools for retrieving these massive amounts of plasmid sequence data (Jesus et al. 2019; O'Leary et al. 2016). Plasmid Atlas (pAT-LAS) 102, on the other hand provides an easily accessible visual analytics tool for users to explore the NCBI database for RefSeq plasmids for plasmid identification from HTS data (Jesus et al. 2019). The de novo annotations-based CARD, ResFinder, Virulence Factors Database (VFDB) and PlasmidFinder, pATLAS allows users to envisage and reconnoitre the metadata associated with all plasmids available in NCBI's RefSeq database, as well as their putative antibiotic resistance and virulence genes and plasmid families (Wang et al. 2015; Zankari et al. 2012; Chen et al. 2005; Carattoli et al. 2017 and Jesus et al. 2019).

12.10 Metagenomics for Antimicrobial Surveillance

Surveillance of antimicrobial resistance (AMR) mainly relies on the passive reporting of laboratory generated data on phenotypes of microorganisms. The Danish Monitoring System (DANMAP) (https://www.danmap.org.) is one such surveillance type that provides data on antimicrobial resistance gene pattern based

on the molecular studies generated from the laboratory. However, this type of antimicrobial gene surveillance does not cover all the relevant information as it is confined to a selected spectrum of microorganism. Microbial culture-based techniques can provide a good insight to antimicrobial resistance organism by allowing the whole-genome sequencing of resistance strains of organisms. However, these techniques are tedious and the applicability is limited to only a few genes of interest from the easily culturable organisms. Culture-based methods are not useful to study the antimicrobial resistance profile of unculturable microorganisms. The metagenomics overcome this hurdle as it targets the sequence analysis of genomic material directly extracted from a sample without any culture isolation of microorganisms. The metagenomic approach is relatively quick that gives high-quality information in comparison to culture-based techniques (Hendriksen et al. 2019).

The metagenomic can be performed in two ways namely; 16S metagenomic sequencing and whole-metagenome sequencing (WMS). 16S metagenomic targets an amplicon of a small variable segment of a highly conserved 16S RNA gene present in all bacterial community. 16S metagenomic gives an insight on the possible microbial taxa and its relative abundance in a sample. On the other hand, in whole-metagenome sequencing approach, the entire genomic DNA is fragmented and sequenced without any amplification. The relative abundance of taxa and known functional and resistance genes in a sample can be determined by comparing the fragmented shotgun reads to available databases of known functional and resistance genes.

Other than the above two approaches, metagenomics has a longitudinal metagenomic approaches appropriate for studying the issues of burden and build-up of antibiotic resistance. Longitudinal metagenomic directly focuses a change in microbial resistance in a sample taken from a patient during a treatment course thus helps alleviating the emergence and transmission of resistance.

Therefore, the metagenomic approaches are useful for monitoring the antimicrobial resistance organisms and resistance genes using short-read sequencing that quantifies thousands of transmissible resistance genes in a single sample without microbial culture (Sukhum et al. 2019). It provides more accurate information on the presence of microbial taxa, pathogenesis and virulence. The metagenomic data would be useful for analyses of novel genes of interest. Owing to its direct application on samples from healthy or clinical cases, and on samples from the potential reservoir, metagenomics outstands as a tool for a single-point surveillance of antimicrobial resistance allowing identification of all resistance genes and their context in all reservoirs.

12.11 Use of Comprehensive Antibiotic Resistance Database (CARD)

CARD is basically a data organizing software system that provides high-quality reference records and achieved genomic sequence data within a defined vocabulary. For the research in resistome and genome-based antimicrobial resistance prediction,

the CARD biocuration team has created and included the Resistance Gene Identifier (RGI) software in the Antibiotic Resistance Ontology (ARO) system for the hindrance-free interaction with software development initiatives (Brian et al. 2020). The use of CARD was popularized in 2017 as a consequence of ease of curating exhaustive reference sequence, modifying ontological framework, ability to curate over 500 extra microbial resistance models, to facilitate the development of innovative classification paradigm and the expansion of analytical tools.

Recently there is an addition of a new module called "Resistomes and Variations" in CARD system which helps to analyze the in silico prediction of resistance variants from over 82 pathogens and one lakh genomes from the database. The inclusion of module on resistance variations has enabled the summarization of the expected resistance using the data in CARD. It has allowed identifying the trends in AMR mobility, understanding of the previously unexplained and novel resistance variants in microbes.

References

Abdelbary M, Basset P, Blanc D et al (2017) In: Tibayrenc M (ed) Genetics and evolution of infectious diseases: the evolution and dynamics of methicillin-resistant *Staphylococcus aureus*, 2nd edn. Elsevier, pp 553–572. ISBN: 9780127999425

Allard MW, Stevens EL, Brown EW (2019) All for one and one for all the true potential of whole-genome sequencing. Lancet Infect Dis 19:683–684

Antipov D, Hartwick N, Shen M et al (2016) Plasmid spades: assembling plasmids from whole-genome sequencing data. Bioinformatics 32:3380–3387

Biswas S, Raoult D, Rolain JM (2008) A bioinformatics approach to understanding antibiotic resistance in intracellular bacteria through whole-genome analysis. Int J Antimicrob Agents 32: 207–220

Blundell TL, Sibanda BL, Montalvao RW et al (2006) Structural biology and bioinformatics in drug design: opportunities and challenges for target identification and lead discovery. Philos Trans R Soc B 361:413–423

Bradley P, Gordon NC, Walker TM et al (2015) Rapid antibiotic-resistance predictions from genome sequence data for *Staphylococcus aureus* and *mycobacterium tuberculosis*. Nat Commun 6:10063. https://doi.org/10.1038/ncomms10063

Brian PA, Amogelang RR, Tammy TYL et al (2020) Antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res 2019(48):D1

Brown NA, Aisner DL, Oxnard GR (2018) Precision medicine in non-small cell lung cancer: current standards in pathology and biomarker interpretation. Am Soc Clin Oncol Educ book 38: 708–715

Carattoli A, Villa L, Feudi C et al (2017) Novel plasmid-mediated colistin resistance mcr-4 gene in Salmonella and Escherichia coli. Italy 2013, Spain and Belgium, 2015 to 2016. Eur Secur 22(31):30589. (Epub ahead of print)

Centers for Disease Control and Prevention (2013) Antibiotic resistance threats in the United States 2013. http://www.cdc.gov/drugresistance/threat-report-2013/. Accessed Apr 2022

Chen L, Yang J, Yu J et al (2005) VFDB: a reference database for bacterial virulence factors. Nucleic Acids Res 33:325–328

Clausen PT, Zankari E, Aarestrup FM (2016) Benchmarking of methods for identification of antimicrobial resistance genes in bacterial whole-genome data. J Antimicrob Chemother 71: 2484–2488. https://doi.org/10.1093/jac/dkw184

Doyle RM, Burgess C, Williams R et al (2018) Direct whole-genome sequencing of sputum accurately identifies drug resistant mycobacterium tuberculosis faster than MGIT culture sequencing. J Clin Microbiol 56:e00666–e00618

Economou V, Gousia P (2015) Agriculture and food animals as a source of antimicrobial-resistant bacteria. Infect Drug Resist 8:49. https://doi.org/10.2147/IDR.S55778

Ellington MJ, Ekelund O, Aarestrup FM et al (2017) The role of whole-genome sequencing in antimicrobial susceptibility testing of bacteria: report from the EUCAST subcommittee. Clin Microbiol Infect 23:2–22. https://doi.org/10.1016/j.cmi.11.012

Feijao P, Yao HT, Fornika D et al (2018) MentaLiST–a fast MLST caller for large MLST schemes. Microbial Genom 4:e000146. https://doi.org/10.1099/mgen.0.000146

Flandrois JP, Lina G, Dumitrescu O (2014) MUBII-TB-DB: a database of mutations associated with antibiotic resistance in mycobacterium tuberculosis. BMC Bioinformatics 15:107. https://doi.org/10.1186/1471-2105-15-107

Frost LS, Leplae R, Summers AO et al (2005) Mobile genetic elements: the agents of open source evolution. Nat Rev Microbiol 3:722–732

Gardy JL, Loman NJ (2018) Towards a genomics-informed, real-time, global pathogen surveillance system. Nat Rev Genet 19:9–20. https://doi.org/10.1038/nrg.2017.88

Gelband H, Miller-Petrie M, Pant S et al (2015) State of the World's antibiotics, 2015. Center for Disease Dynamics, Economics, and Policy, Washington, DC; http://cddep.org/publications/state_worlds_antibiotics_2015#sthash.l18BFUu2.dpbs

Gibson MK, Forsberg KJ, Dantas G (2015) Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. ISME J 9:207–216. https://doi.org/10.1038/ismej.2014.106

Giordano F, Aigrain L, Quail MA et al (2017) De novo yeast genome assemblies from MinION, PacBio and MiSeq platforms. Sci Rep 7:3935. https://doi.org/10.1038/s41598-017-03996-z

Gong J, Pan K, Fakih M et al (2018) Value-based genomics. Oncotarget 9(21):15792–15,815

Gonzaga-Jauregui C, Lupski JR, Gibbs RA (2012) Human genome sequencing in health and disease. Annu Rev Med 63:35–61

Gupta SK, Padmanabhan BR, Diene SM et al (2014) ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents Chemother 58:212–220. https://doi.org/10.1128/AAC.01310-13

Gwinn M, Maccannell D, Armstrong GL (2019) Next generation sequencing of infectious pathogens. JAMA 321:893–894

Hendriksen RS, Munk P, Njage P et al (2019) Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. Nat Commun 10:1124–08853

Jesus TF, Ribeiro-Gonçalves B, Silva DN et al (2019) Plasmid ATLAS: plasmid visual analytics and identification in high-throughput sequencing data. Nucleic Acids Res 47:D188–D194

Jia B, Raphenya AR, Alcock B et al (2017) CARD. 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic Acids Res 45:D566–D573. https://doi.org/10.1093/nar/gkw1004

Karp BE, Tate H, Plumblee JR et al (2017) National antimicrobial resistance monitoring system: two decades of advancing public health through integrated surveillance of antimicrobial resistance. Foodborne Pathog Dis 14:545–557. https://doi.org/10.1089/fpd.2017.2283

Katsanis SH, Katsanis N (2013) Molecular genetic testing and the future of clinical genomics. Nat Rev Genet 14(6):415–426

Kos VN, Déraspe M, McLaughlin RE et al (2015) The resistome of Pseudomonas aeruginosa in relationship to phenotypic susceptibility. Antimicrob Agents Chemother 59:427–436. https://doi.org/10.1128/AAC.03954

Lakin SM, Dean C, Noyes NR et al (2017) MEGARes: an antimicrobial resistance database for high-throughput sequencing. Nucleic Acids Res 45:574–580. https://doi.org/10.1093/nar/gkw1009

Lanza VF, Baquero F, Martinez JL et al (2018) In-depth resistome analysis by targeted metagenomics. Microbiome 6(11):0387

Liu B, Pop MB (2009) Antibiotic resistance genes database. Nucleic Acids Res 37:D443–D447. https://doi.org/10.1093/nar/gkn656

Lu H, Giordano F, Ning Z (2016) Oxford Nanopore MinION sequencing and genome assembly. Genomics Proteomics Bioinform 14:265–279. https://doi.org/10.1016/j.gpb.2016.05.004

Luscombe NM, Greenbaum D, Gerstein M (2001) What is bioinformatics? A proposed definition and overview of the field. Methods Inf Med 40:346–358

Mason A, Foster D, Bradley P et al (2018) Accuracy of different bioinformatics methods in detecting antibiotic resistance and virulence factors from Staphylococcus aureus whole-genome sequences. J Clin Microbiol 56:e01815–e01817

Mazzarotto F, Olivotto I, Walsh R (2020) Advantages and perils of clinical whole-exome and whole-genome sequencing in cardiomyopathy. Cardiovasc Drugs Ther 34(2):241–253

McArthur AG, Waglechne RN, Nizam F et al (2013) The comprehensive antibiotic resistance database. Antimicrob Agents Chemother 57:3348–3357. https://doi.org/10.1128/AAC.00419-13

Meienberg J, Bruggmann R, Oexle K et al (2016) Clinical sequencing: is WGS the better WES? Hum Genet 135(3):359–362

Miotto P, Tessema B, Tagliani E et al (2017) A standardised method for interpreting the association between mutations and phenotypic drug resistance in mycobacterium tuberculosis. Eur Respir J 50:1701354. https://doi.org/10.1183/13993003.01354-2017

Ndagi U, Falaki AA, Abdullahi M et al (2020) Antibiotic resistance: bioinformatics-based understanding as a functional strategy for drug design. RSC Adv 10(31):18451–18468

Ng PC, Kirkness EF (2010) Whole-genome sequencing. Methods Mol Biol 628:215–226

O'Leary NA, Wright MW, Brister JR et al (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion and functional annotation. Nucleic Acids Res 44:D733–D745

Pettengill EA, Pettengill JB, Binet R (2014) Phylogenetic analyses of Shigella and Enteroinvasive Escherichia coli for the identification of molecular epidemiological markers: whole-genome comparative analysis does not support distinct genera designation. Front Microbiol 6:1573. https://doi.org/10.3389/fmicb.2015.01573

Quainoo S, Coolen JPM, van Hijum SAFT et al (2017) Whole-genome sequencing of bacterial pathogens: the future of nosocomial outbreak analysis. Clin Microbiol Rev 4:1015–1063

Rapin N, Lund O, Bernaschi M (2010) Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system. PLoS One 4:e9862

Rhoads A, Au KF (2015) PacBio sequencing and its applications. Genomics, Proteomics Bioinformatics 13:278–289. https://doi.org/10.1016/j.gpb.2015.08.002

Robertson J, Nash JHE (2018) MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies. Microb Genom 4:1–7

Rozov R, Kav AB, Bogumil D (2017) Recycler: an algorithm for detecting plasmids from de novo assembly graphs. Bioinformatics 33:475–482

Sandgren A, Strong M, Muthukrishnan P et al (2009) Tuberculosis drug resistance mutation database. PLoS Med 6:e2. https://doi.org/10.1371/journal.pmed.1000002

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci 74:5463–5467

Stevens EL, Timme R, Brown EW et al (2017) The public health impact of a publically available, environmental database of microbial genomes. Front Microbiol 8:808

Sukhum KV, Diorio-Toth L, Dantas G (2019) Genomic and metagenomic approaches for predictive surveillance of emerging pathogens and antibiotic resistance. Clin Pharmacol Ther 106(3):512–524. https://doi.org/10.1002/cpt.1535

Vincent BM, Lancaster AK, Scherz-Shouval R et al (2013) Fitness trade-offs restrict the evolution of resistance to amphotericin B. PLoS Biol 11(10):e1001692

Votintseva AA, Bradley P, Pankhurst L et al (2017) Same-day diagnostic and surveillance data for tuberculosis via whole-genome sequencing of direct respiratory samples. J Clin Microbiol 55: 1285–1298. https://doi.org/10.1128/JCM.02483-16

Wang F, He X, Huang B, Chen P et al (2015) Detection and analysis of resistance mutations of hepatitis B virus. Int J Clin Exp Med 8:9630–9639

Xavier BB, Lammens C, Ruhal R et al (2016) Identification of a novel plasmid-mediated colistin-resistance gene, mcr-2, in Escherichia coli, Belgium, June 2016. Euro Surveill 21(27):30280. https://doi.org/10.2807/1560-7917.ES.2016.21.27.30280

Zankari E, Hasman H, Cosentino S et al (2012) Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother 67:2640–2644. https://doi.org/10.1093/jac/dks261

Zankari E, Allesøe R, Joensen KG et al (2017) PointFinder: a novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. J Antimicrob Chemother 72:2764–2768. https://doi.org/10.1093/jac/dkx217

Zetner A, Cabral J, Mataseje L et al (2017) Plasmid Proler: comparative analysis of plasmid content in WGS data. BioRxiv 121:350

Zhang X, Payne M, Lan R (2019) In silico identification of Serovar-specific genes for salmonella serotyping. Front Microbiol 10:835

Zhao EY, Jones M, Jones SJM (2019) Whole-genome sequencing in cancer. Cold Spring Harb Perspect Med 9(3):a03457

Zhou F, Xu Y (2010) cBar: a computer program to distinguish plasmid-derived from chromosome-derived sequence fragments in metagenomics data. Bioinformatics 26(16):2051–2052. https://doi.org/10.1093/bioinformatics/btq299