www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277-7695
ISSN (P): 2349-8242
NAAS Rating: 5.23
TPI 2022; SP-11(11): 585-593

www.thepharmajournal.com
Received: 29-09-2022
Accepted: 30-10-2022

Krishna J Patel
M.V.Sc Scholar,
Department of Veterinary Microbiology, College of Veterinary Science and Animal Husbandry, Kamdhenu University, Anand, Gujarat, India

Mayank J Patel
M.V.Sc Scholar,
Department of Animal Reproduction, Gynaecology and Obstetrics, College of Veterinary Science and Animal Husbandry, Kamdhenu University, Anand, Gujarat, India.

Esha Sinha
Ph.D. Scholar, Department of Veterinary Microbiology, ICAR-IVRI, Bareilly, Uttar Pradesh, India

Corresponding Author:
Krishna J Patel
M.V.Sc Scholar,
Department of Veterinary Microbiology, College of Veterinary Science and Animal Husbandry, Kamdhenu University, Anand, Gujarat, India

Profiling of antimicrobial resistance genes using next generation sequencing: An overview

Krishna J Patel, Mayank J Patel and Esha Sinha

Abstract
Antimicrobial resistance (AMR) is a global problem because there is a paucity of antibiotics available now a day to treat multidrug-resistant bacterial infections in humans and animals. Detection of AMR present in bacteria that may pose a threat to veterinary and public health and routinely detected by using standardized phenotypic methods. Technologies such as next-generation sequencing are expanding abilities to detect and study antimicrobial resistance. This Review provides a detailed overview of antimicrobial resistance identification and characterization methods from traditional antimicrobial susceptibility testing to recent deep-learning methods and focusing on sequencing-based resistance discovery and various tools and databases used in antimicrobial resistance studies.

Keywords: Antimicrobial resistance, polymerase chain reaction, whole genome sequencing, whole metagenomics sequencing, next generation sequencing

Introduction
Antibiotic resistance is one of the major global threat now a day. Many earlier manageable bacterial infections are turning into increasingly more difficult to treat (Su *et al*., 2019) [11]. Antimicrobials are small molecules that can inhibit bacteria. These small molecules are commonly used as therapeutics for bacterial infections, but some bacteria can develop and survive despite antimicrobial pressures, a property known as antimicrobial resistance (AMR) (Boolchandan *et al*., 2019) [3]. Accurate detection of antimicrobial resistance is necessary to guide treatment decisions (Su *et al*., 2019) [11]. Recently phenotypic and genotypic methods are used for the detection of AMR in bacteria. Culture-based antimicrobial susceptibility testing (AST) is still the primary method used by clinical laboratories (Su *et al*., 2019) [11]. However, the genotypic methods are used for screening of AMR genes in bacteria *viz*. Polymerase Chain Reaction (PCR), Whole Genome Sequencing (WGS) and Metagenomics. Minimum three days are required for cultural isolation and susceptibility test in phenotypic method, so it is a very time-consuming and some antibiotic genes could be silent, but it could act as a source for gene spread. In genotypic methods, no need for cultural isolation and it is a modern molecular biotechnological method to detect and characterized genes encoding antibiotic resistance (Su *et al*., 2019) [11].

Polymerase Chain Reaction (PCR)
PCR is a technique that was developed by Kary Mullis and has revolutionized molecular biology, enabling rapid and exponential amplification of target DNA sequences using a forward and reverse PCR primer and an enzyme known as DNA polymerase in the presence of deoxyribonucleotides.
Conventional PCR comprises three steps:
1. Denaturing of the double-stranded DNA at 95 °C,
2. Annealing of the PCR primers at 50 to 60 °C and
3. Extension of the DNA at 72 °C.

PCR is used routinely in microbiology laboratories for detecting individual genes with complementary sequences (Anjum *et al*., 2018) [2].

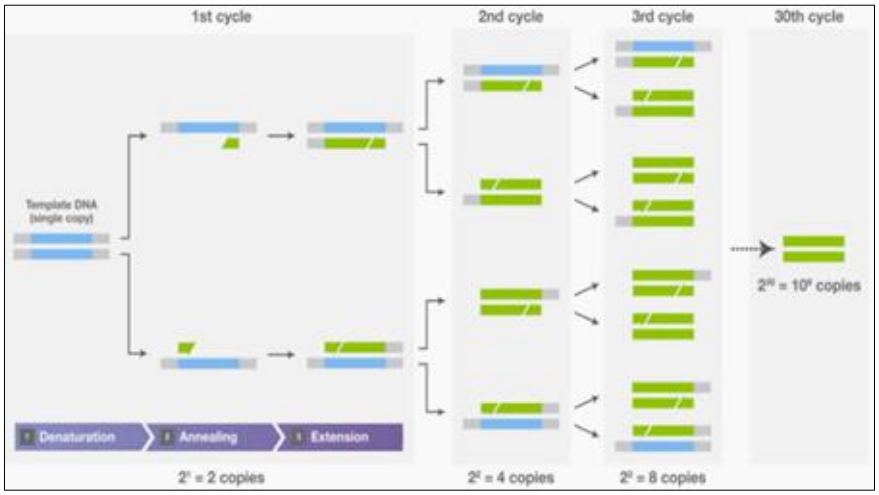

Fig 1: Polymerase Chain Reaction steps

Whole Genome Sequencing (WGS)

Sequencing is performed either using long-read third-generation technologies that have higher error rates but more complete assemblies. WGS-AST algorithms operate on the raw reads and/or assembled contigs (Su *et al*., 2019) [11].

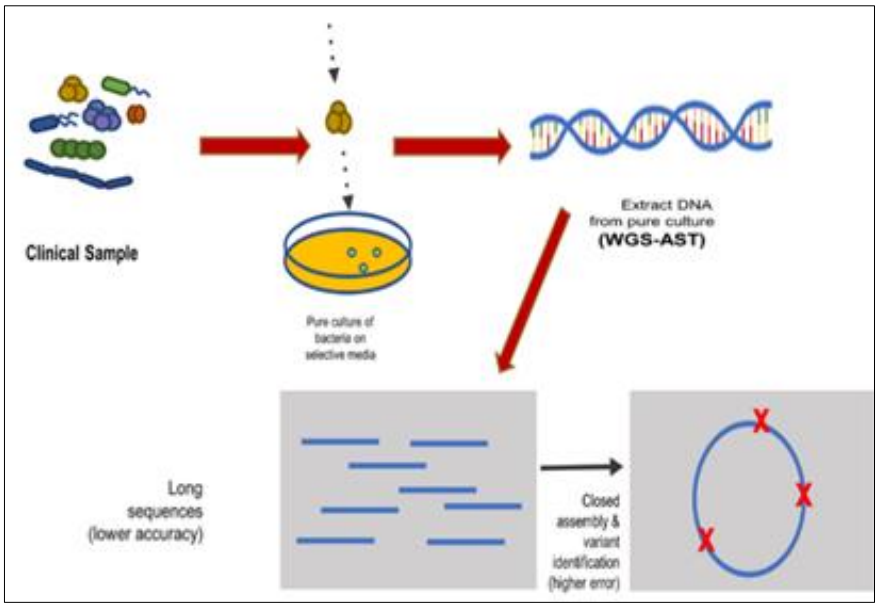

Fig 2: WGS-AST algorithms

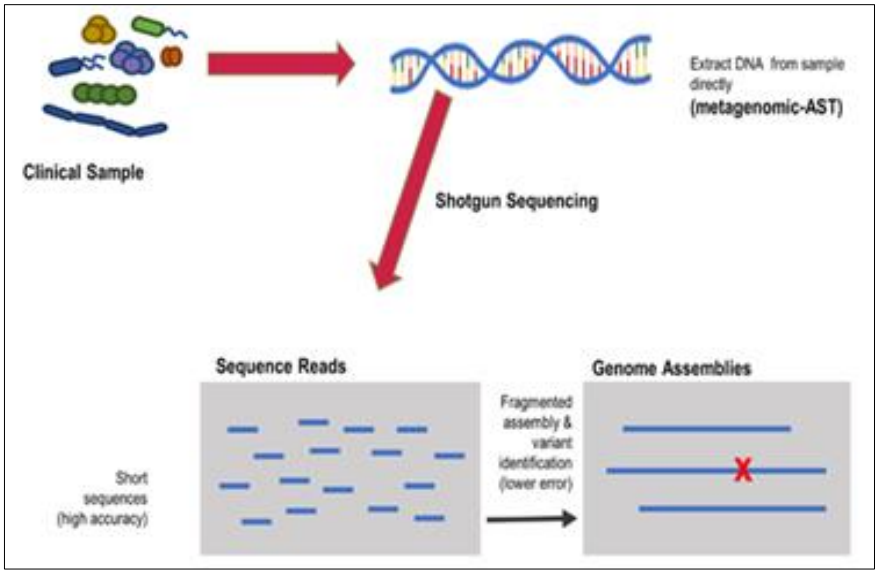

Fig 3: Shortgun Sequencing

Metagenomics

Sequencing is performed either using short-read second-generation technology, which tends to produce fragmented whole-genome assemblies of high accuracy (Su *et al*., 2019) [11].

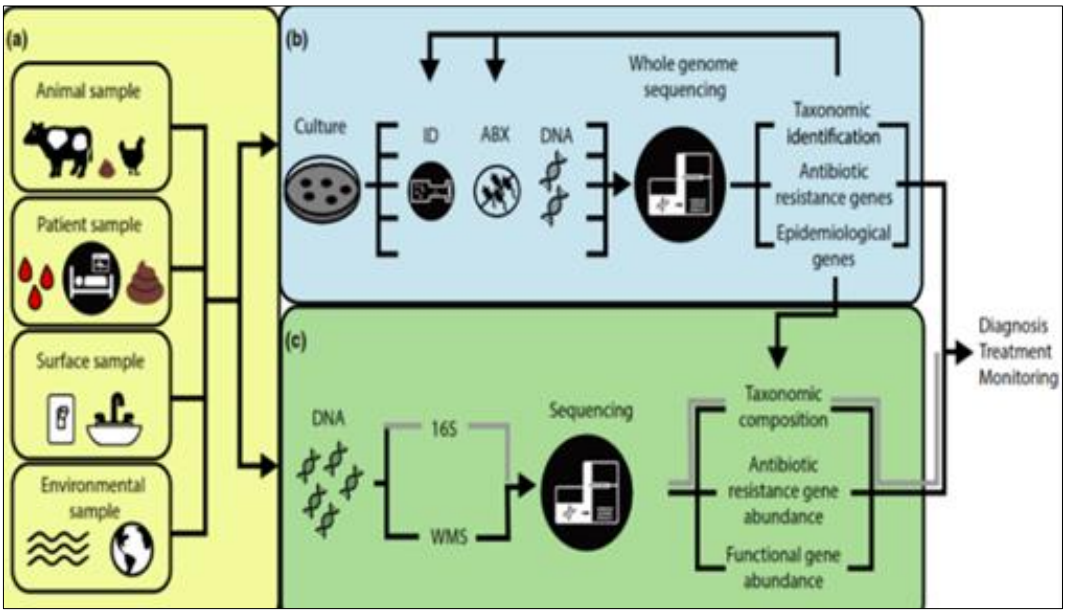

Fig 4: Taxonomic identification/composition, Antibiotic resistance gene/gene abundance, Epidemiological genes/functional gene abundance detection through WGS and Metagenomic techniques from biological and environmental samples.

Sequencing technologies

Culture-based methods, isolates are extracted from samples and processed individually to determine species identification and antibiotic susceptibility. After DNA extraction each isolate can be sequenced using WGS. WGS analyses can determine taxonomic identification, antibiotic resistance genotype, and epidemiological genes. These characteristics can be compared with culture-based results, creating genotype-phenotype characterization of isolates and improved characterization of species and resistome. Metagenomics analyses, genomic DNA is extracted directly from a sample, then 16S or whole metagenomics sequencing (WMS) is done. After 16S amplification and sequencing, taxonomic composition is determined. After WMS, taxonomic composition, antibiotic resistance gene abundance, and functional gene abundance can be analyzed (Sukhum *et al*., 2019) [12].

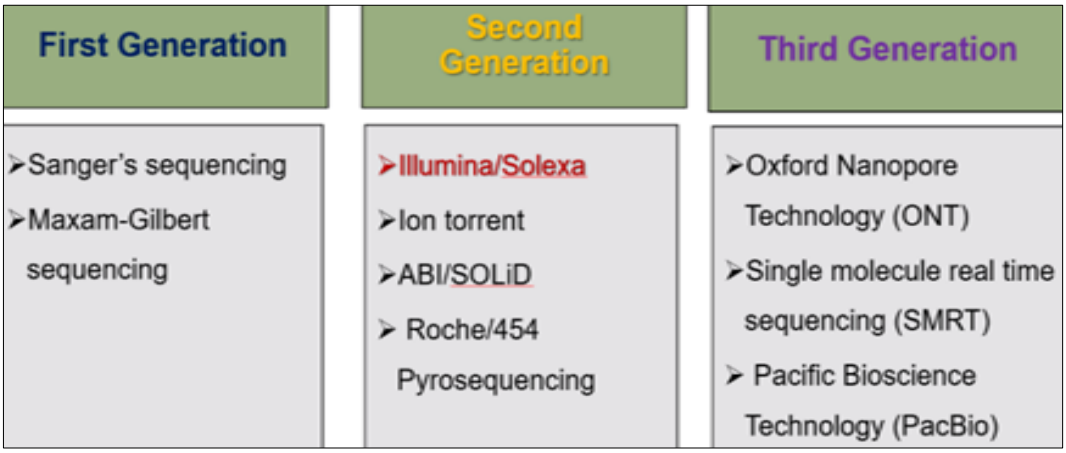

Fig 5: Types of sequencing technologies

First Generation

Sanger Sequencing is known as the chain termination method or dideoxynucleotide method and Maxam-Gilbert is known as the chemical degradation method (Tyagi *et al*., 2020) [13]. It works on whole genome shotgun sequencing approach. DNA is cloned into a high-copy-number plasmid. PCR amplification is carried out with primers that flank the target. Primer is complementary to known sequence immediately flanking the region of interest. Each round of primer extension is terminated by the incorporation of fluorescently labelled dideoxynucleotides (ddNTPs). Sequence is determined by high-resolution electrophoretic separation of the single-stranded, end-labeled extension products in a capillary-based polymer gel. Laser excitation of fluorescent labels as fragments of discreet lengths exit the capillary, coupled to four-color detection of emission spectra, provides the readout that is represented in a Sanger sequencing trace (Shendure & Ji *et al*., 2008) [10].

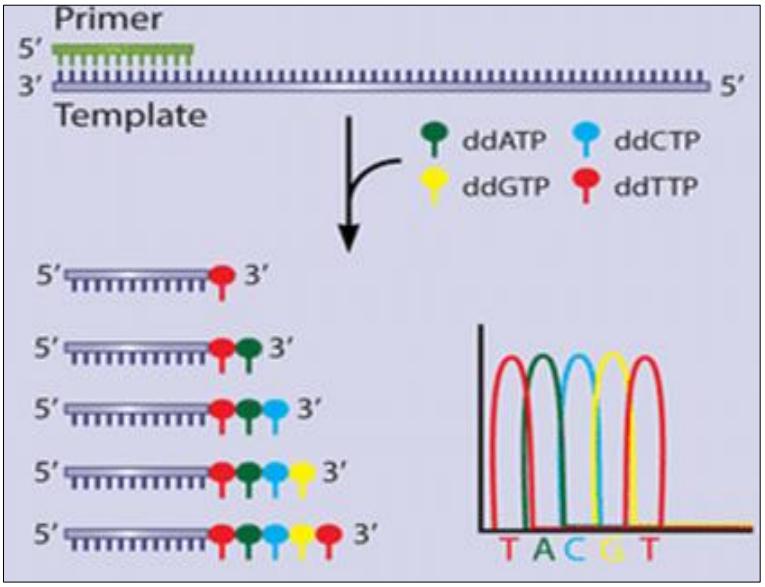

Fig 6: First generation (Short gun sequencing)

Second Generation

Second-generation sequencing is based on massive parallel and clonal amplification of molecules by polymerase chain reaction (PCR). It can also be termed as a short-read sequencing approach. The fluorescent dye is identified through laser excitation and imaging, and subsequently, it is enzymatically cleaved to allow the next round of incorporation (Pereira *et al*., 2020; Tyagi *et al*., 2020) [13]

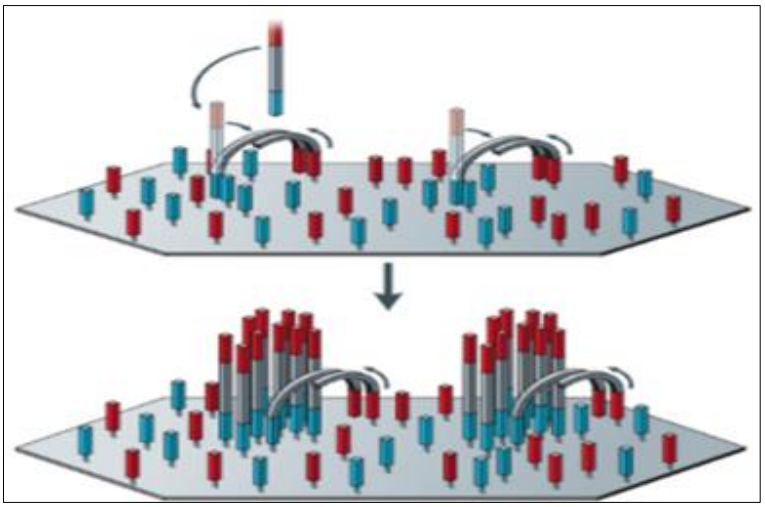

Fig 7: Second generation **(**Massive parallel sequencing)

Third Generation

Third-generation sequencing relies on single-molecule sequencing without a prior clonal amplification (Pereira *et al*., 2020; Tyagi *et al*., 2020) [9, 13].

Next-generation sequencing (NGS)

Next Generation Sequencing (NGS) is also known as high-throughput sequencing or second generation sequencing technology. These technologies allow for sequencing of DNA and RNA much more quickly and cheaply than previously used Sanger Sequencing method (Mardis, 2008) [8].

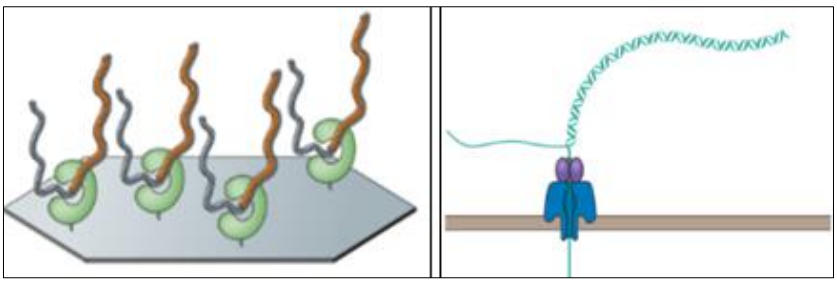

Fig 8: Third generation (Single molecule sequencing

Characteristics of Second Generation Sequencing Technology

- Generation of many millions of short reads in parallel.
- Speed up of sequencing the process compared to the first generation.
- Low cost of sequencing.
- Sequencing output is directly detected without the need for Electrophoresis. (Kchouk *et al.*, 2017) [7]

Next generation sequencing techniques is the targeted enrichment step where gene panels focus on a limited number of genes. Whole genome sequencing does not require targeted enrichment. Gene panels target a set number of genes at a higher sequencing depth and lower cost when compared to whole exome and Whole genome sequencing can provide more coverage Whole genome sequencing can provide more even coverage of the genome and protein coding regions when compared to whole exome sequencing and gene panels, however there is a lower sequencing depth at a higher cost per sample (Dunn *et al*., 2018) [4].

Next Generation DNA Sequencing Technology

1. Sequencing by synthesis (Illumina)
2. Ion Torrent or PGM or Proton
3. Sequencing by ligation (SOLiD technology)
4. 454 life sciences or pyrosequencing

Sequencing technology	First generation	Second Generation	Third Generation
Principle	Chain termination or dideoxy and Chemical method	Sequencing-by-ligation and Sequencing-by-synthesis	single-molecule sequencing
Resolution	Averaged across many copies of the DNA molecule	Averaged across many copies of the DNA molecule	Single DNA molecule
Current raw read accuracy	High	High	Low
Current read length	Moderate (800- 1000bp	Short	> 1000bp
Current throughput	Low	High	High
Current cost	High cost per base Low cost per run	Low cost per base High cost per run	Low cost per base High cost per run
Time of result	Hours	Days	<1 day

Fig 9: Comparison of first, second and third generation sequencing

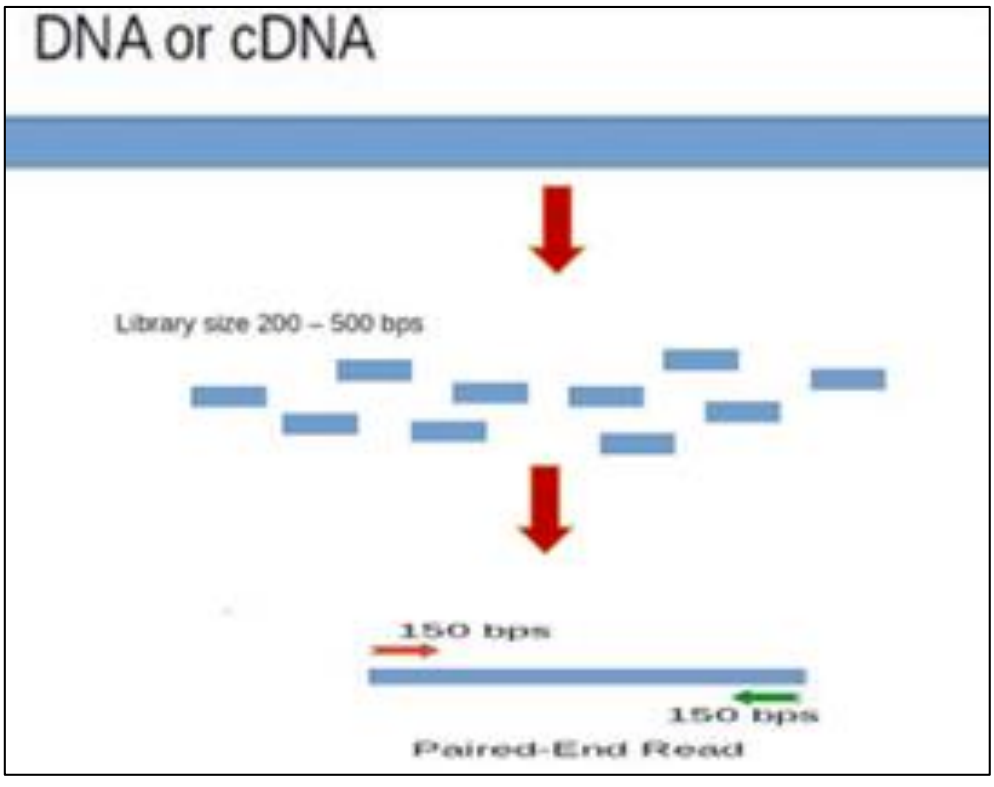

Fig 10: Paired End Read

Method	Pacific Bioscience	Solexa/Illumina (Sequencing by synthesis)	Ion Torrent or PGM	Nanopore
Average read length (bp)	Up to 50 kb	50 to 250 bp	200 bp	Up to 100 kb
Average accuracy	86 %	98 %	98 %	95 %
Average reads per run	500 k	Up to 3 billion	Up to 5 million	100 k
Time per run	30 minutes to 6 hours	Depend upon sequencer and specific read length	2 hours	Vary
Cost per 1 million bases	$2	$ 0.05 to 0.15	$ 1	$3
Advantages	Longest read length , Fast .	Potential for high sequence yield, depending upon sequencer model	Less expensive equipment, Fast	Longest read, ready to use
Disadvantages	Low yield , cost and error	Equipment can be very expensive	Homopolymer errors.	Low yield , cost ,error And stability

Fig 11: Comparison of sequencing technologies

Illumina/solexa sequencing

Illumina technology is sequencing by synthesis approach and is currently the most used technology in the NGS market (Kchouk *et al.*, 2017) [7].

Illumina Sequencing Platform

Illumina sequencing technology basically work on four basic steps which include library preparation, cluster generation, sequencing, and analysis. Illumina sequencing is termed as bridge amplification (amplification part) (Kchouk *et al.*, 2017) [7].

Fig 12: Illumina sequencing platforms

Library Preparation

Genomics DNA is randomly fragmented into shorter fragments of ~200 bps followed by 5' and 3' end adapter ligations. These short fragments are then PCR amplified and purified. Paired end sequencing allows sequencing of both ends of a fragment and generates high-quality reads. As the distance between the 2 reads of a pair is known, pair end sequencing helps generate quality alignments to a genome to detect genomic rearrangements and repetitive sequence elements, as well as gene fusions and novel transcripts. **(**1010 genome.com/illumina-sequencing)

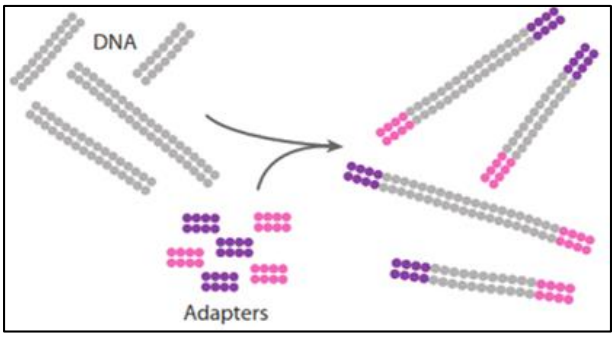

Fig 13: Genomic DNA Sample

Preparation of Genomic DNA

DNA samples are randomly fragmented into sequences and adapters are ligated to both ends of each sequence (Mardis, 2008) [8].

Flow Cell

Several samples can be loaded onto the eight-lane flow cell for simultaneous analysis on an Illumina Sequencing System (Mardis, 2008) [8]. Bind single stranded fragments randomly to

the inside surface of the flow cell channels. Unlabelled nucleotides and polymerase enzyme are added to initiate the solid phase bridge amplification (Mardis, 2008) [8].

Bridge PCR Amplification

Fragment becomes double stranded (Mardis, 2008) [8].

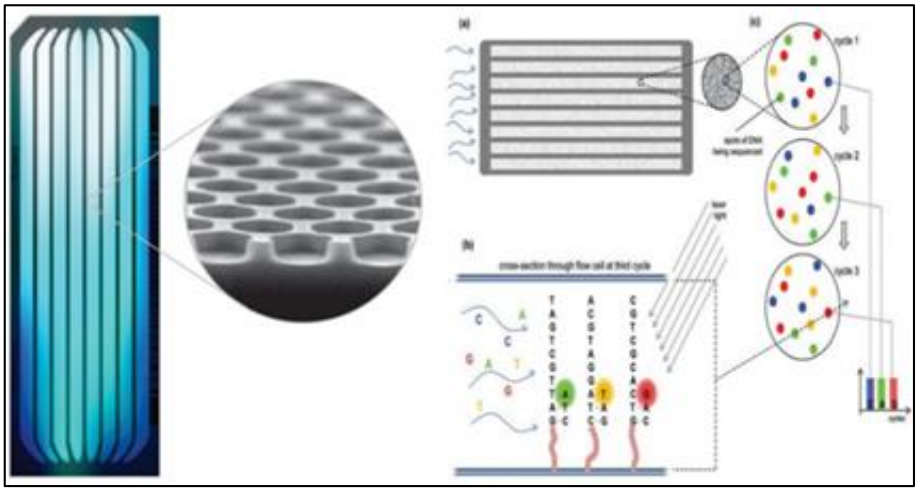

Fig 14: Flow cell

Cluster Generation

Each attached sequence to the solid plate is amplified by PCR bridge amplification that creates several identical copies of each sequence, a set of sequences made from the same original sequence is called a cluster. Each cluster contains approximately one million copies of the same original sequence (Mardis, 2008) [8].

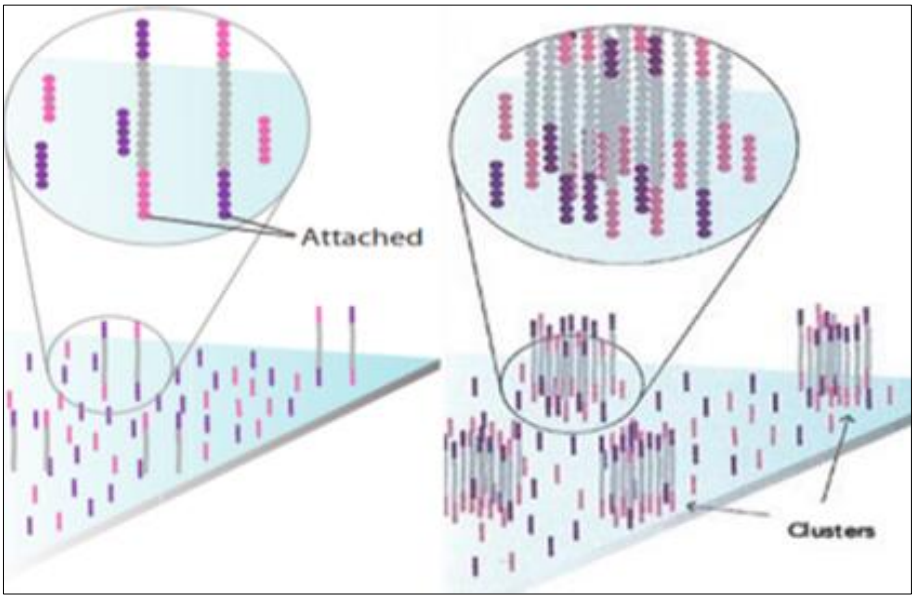

Fig 15: Cluster amplification

Sequencing by Synthesis

Each type of nucleotide is labeled with a fluorescent specific for each type to be unique. The nucleotides have an inactive 3'-hydroxyl group which ensures that only one nucleotide is incorporated. Clusters are excited by laser for emitting a light signal specific to each nucleotide, which will be detected by a coupled-charge device (CCD) camera and Computer programs will translate these signals into a nucleotide sequence. The process continues with the elimination of the terminator with the fluorescent label and the starting of a new cycle with a new incorporation (Mardis, 2008) [8].

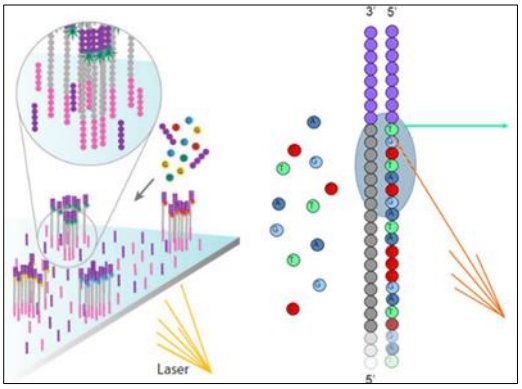

Fig 16: Sequencing

Data Analysis

Reads are aligned to a reference genome with various bioinformatics software (Mardis, 2008) [8].

Sequencing based approaches for antimicrobial resistance genes detection

Next-generation sequencing are expanding our abilities to detect and study antimicrobial resistance. The various bioinformatics software can process sequence data either as reads or as assemblies (Boolchandani *et al*., 2019) [3].

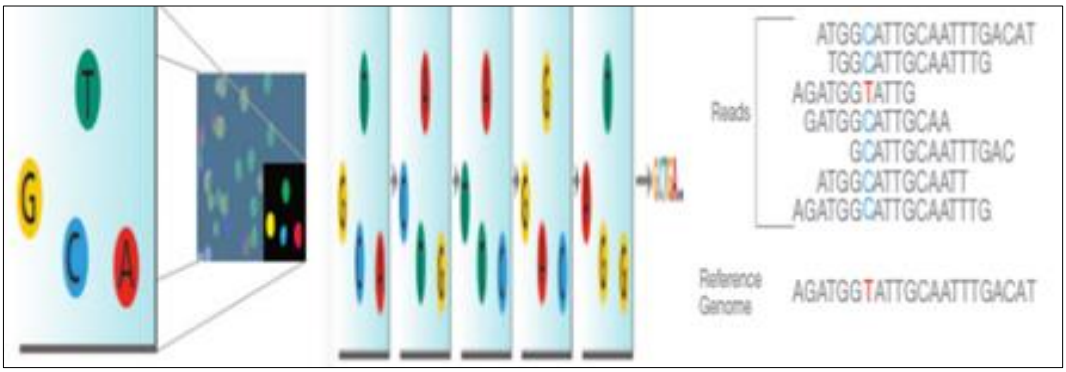

Fig 17: Alignment and data analysis

Assembly Based Approach

Computationally expensive and time consuming, particularly in resistome profiling of large complex communities. Identification of both known and novel resistance genes that share low similarity with reference database; however, requires high genome coverage. Genomic context such as regulatory and mobile element sequences can be captured Read-based approach (Boolchandani *et al*., 2019) [3].

Read Based Approach

Fast and less computationally demanding, enabling resistome analysis of large data sets identification of resistance genes is dependent on completeness of the reference database of organisms under analysis. Nearby genes and genomic context are missing; may lead to false positives (Boolchandani *et al*., 2019) [3].

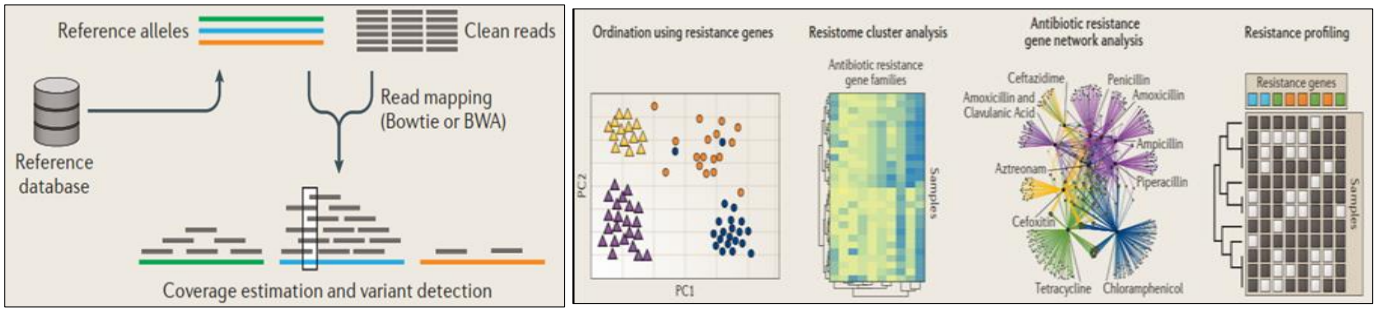

Fig 18: Downstream analysis

AMR gene databases/tools

CARD

The Comprehensive Antibiotic Resistance Database (CARD) is a web service and database in FASTA formate. CARD has two analysis options: BLAST and RGI (Resistance Gene Identifier)The BLAST option performs standard BLAST searches on smaller sequences uploaded by the user (but not whole genomes) against the CARD reference sequences. The RGI supports two detection model types: (i) Protein homologue models to detect AMR genes (ii) Protein variant models to detection of mutations in AMR. Currently, the RGI only analyzes protein sequences,and if assembled contigs are submitted to the tool, the RGI first predicts open reading frames and then analyzes the predicted protein sequences. The RGI is also developed as a command-line tool for bulk analysis of many genomes simultaneously (Anjum *et al*., 2018) [2].

ResFinder

ResFinder is a web server composed of a BLAST based alignment for detection of acquired AMR genes in assembled WGS data and a curated database in FASTA format containing the resistance genes. It can possible to analyze assembled data or raw reads (Anjum *et al*., 2018) [2].

ARG-ANNOT

Antibiotic Resistance Gene-Annotation (ARG-ANNOT) uses a local BLAST in Bio-Edit software. It provides the user with three databases in FASTA format (i) Nucleotide sequences (ii) Protein and (iii) Mutational gene. Analyzes assembled data without Web interface (Anjum *et al*., 2018) [2].

Pathogen watch

Fast predictions of resistant genotypes and clustering. Facilitates processing, clustering and exploration of microbial genome assemblies. Processes of whole genome assemblies and epicollect for generic metadata gathering (Pathogen. watch).

Sr. No.	List of Web databases used for AMR gene depositories / references	List of tools used for AMR gene detection
1.	CARD	ResFinder
2.	Resfams	ARG-ANNOT
3.	ARDB	ResfinderFG
4.	MEGARes	RGI
5.	NDARO	ARGs-OAP (v2)
6.	Mustard	ARIBA
7.	FARME database	NCBI-AMRFinder
8.	SARD (v2)	SRST2
9.	Lahey list of beta-lactamases	Pathogen Watch
10.	BLDB	PATRIC
11.	LacED	SSTAR
12.	CBMAR	KmerResistance
13.	MUBII-TB-DB	DeepARG
14.	u-CARE	

Fig 19: Bioinformatics tools for AMR gene detection

Conclusions

Antibiotic resistance genes can be detected by two approaches phenotypic method and genotypic method, out of these two methods genotypic method is successful method in which the silent genes can detected which is not possible through phenotypic method. AMR genes can successfully be screened by NGS platforms more rapidly as compared to detection of individual gene by PCR. The genome sequencing involves three generations. First generation sequencing the Sanger sequencing involves chain termination by ddNTPs; and second generation sequencing involves Illumina/Solexa, Ion torrent; and third generation includes single molecule real time sequencing approach (SMRT), Pacific Bioscience Technology (PacBio) etc. Illumina sequencing platform involves four steps library preparation, cluster generation, sequencing by synthesis, and analysis either by read based approach or assembly based approach.NGS technologies are widely used and help in screening of many genes simultaneously with low cost, less time and more output. Illumina gene panel can be used to detect AMR in combination with bioinformatics databases/softwares like CARD, ResFinder, ARG-ANNOT etc. A number of accessible bioinformatics tools available for the prediction of AMR determinants to support and expand the global pathogen surveillance and AMR tracking based on genomics.

References

1. Alves LDF, Westmann CA, Lovate GL, de Siqueira GMV, Borelli TC, Guazzaroni ME, *et al*. Metagenomic approaches for understanding new concepts in microbial science. International Journal of Genomics; c2018. Article ID 2312987.
2. Anjum MF, Zankari E, Hasman H. Molecular methods for detection of antimicrobial resistance. Antimicrobial Resistance in Bacteria from Livestock and Companion Animals. 2018;6(4):33-50.
3. Boolchandani M, D'Souza AW, Dantas G. Sequencing based methods and resources to study antimicrobial resistance. Nature Reviews Genetics. 2019;20(6):356-370.
4. Dunn P, Albury CL, Maksemous N, Benton MC, Sutherland HG, Smith RA, *et al*. Next generation sequencing methods for diagnosis of epilepsy syndromes. Frontiers in Genetics. 2018;9:20.
5. https://pathogen.watch.
6. Illumina Sequencing by Synthesis 1010Genome.http:/www.1010genome.com.
7. Kchouk M, Gibrat JF, Elloumi M. Generations of sequencing technologies: from first to next generation. Biology and Medicine. 2017;9(3):8.
8. Mardis ER. Next-generation DNA sequencing methods. Annual Review Genomics and Human Genetics. 2008;9(1):387-402.
9. Pereira R, Oliveira J, Sousa M. Bioinformatics and computational tools for next generation sequencing analysis in clinical genetics. Journal of Clinical Medicine. 2020;9(1):132.
10. Shendure J, Ji H. Next-generation DNA sequencing. Nature biotechnology. 2008;26(10):1135-1145.
11. Su M, Satola SW, Read TD. Genome-based prediction of bacterial antibiotic resistance. Journal of Clinical Microbiology. 2019;57(3):e01405-18.
12. Sukhum KV, Diorio Toth L, Dantas G. Genomic and metagenomic approaches for predictive surveillance of emerging pathogens and antibiotic resistance. Clinical Pharmacology and Therapeutics. 2019;106(3):512-524.
13. Tyagi P, Bhide M. History of DNA Sequencing. Folia Veterinaria. 2020;64(2):66-73.