International Journal of Scientific Research and Engineering Development-– Volume 3 Issue 6, Nov-Dec 2020

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Genome Sequencing to Access Food Borne Illness

Lisha Deb, Anoushka Das*, Purvi Patawri*, Abhilipsa Dash*, Tiyasa Guha**

*(Department of Biotechnology, VIT University, Vellore)

** (Department of Biotechnology, VIT University, Vellore)

* (Department of Biotechnology, VIT University, Vellore)

** (Department of Biotechnology, VIT University, Vellore)

*** (Department of Biotechnology, VIT University, Vellore)

----------**********----------

Abstract:

Food borne ailments are the result of consuming contaminated food or water that contains pathogens. These remain a chief public health issue around the globe that substantially impacts the economy. Suitable health catastrophe management is principally reliant on the early exposure of probable public health hazards that is laden by constantly changing tendencies in illness outbreaks. A cohesive multidisciplinary method that is a union of sustained pathogen syndrome and diagnostic surveillance, genomics-based, and standardized global analytical networks gathering clinical, epidemiological and genetic data alike is needed to comprehend the dynamic forces of food borne illness and to alleviate budding effects of imminent threats. In this paper, we review various sequencing techniques used for surveillance of food borne diseases.

.*Keywords* **—Food borne diseases, whole genome sequencing, outbreak and surveillance, next-generation sequencing, high-throughput sequencing**

----------**********----------

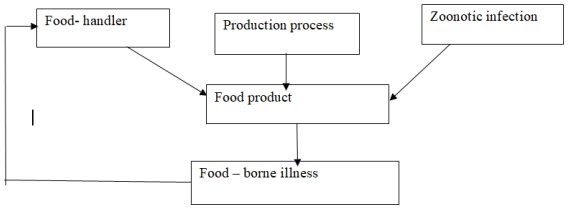
I. INTRODUCTION

In the recent years, numerous genome sequencing methods have dominated the public health labs and food regulatory agencies, replacing the traditional methods of diagnosis. In 1995, the first ever complete sequence of any bacterial organism was generated. *Haemophilus influenzae* was used to reveal the entire set of genetic information used to encode a free-living organism. This project also is notable for making the genome sequence data and the bioinformatics software used to assemble it freely available to the scientific community. Such sharing aimed to be (and was) consistent with the policies initially set out by the ongoing Human Genome Project [46] later codified in the 1996 Bermuda Principles [47]. The policies of open sharing of genomic data and open source release of bioinformatics software tools set out by these seminal sequencing efforts were instrumental in cementing openness into the scientific culture [48].

There has been a continuous rise in vector-borne diseases over decades, with some ending in a pandemic and others in small outbreaks. While the most common and deadly vector-borne diseases Āhave been associated with virus carrying mosquitoes such as Aedes aegypti, also known as “yellow fever mosquito”, which can carry a number of disease agents along with dengue fever virus (DENV) [49]. Similar to dengue, malaria is also considered as one of the most common and fatal vector-borne disease which is caused by bite of an infected Anopheles mosquito. While these diseases have been widely covered in the global news media

to monitor their respective outbreaks there is another category that has been escalating every year, it is the food-borne disease [49]. According to the WHO, food-borne diseases (FBD) are caused by 'unsafe' food which carries harmful bacteria, viruses, parasites or chemical substances that culminate in hundreds of diseases. Almost 1 in every 10 people in the world falls ill because of FBD, which estimates to around 600 million lives and about 420,000 lives are lost. Food contamination can occur at any stage of production [50].

Based on recent studies conducted across the world, the four main species involved in food-borne diseases are : *Salmonella* spp, *Campylobacter* spp, *Listeria monocytogenes* and Shiga toxin-producing *Escherichia* coli [54]. According to surveys conducted across various countries of the world, 82% of all hospitalisation for food borne diseases was caused by theses four species in the United States [51]. In Iran though, this rate was as low as 16.5% [52] and about 15% in China, where the main spot for food poisoning is occupied by two strains of *V. vulnificus [53].*



The evolution of genome sequencing started in 1975 with the first generation sequencing by Sanger and Coulson. This method used radioactive isotope labelling as its base, followed by gel electrophoresis and autoradiography [1].

2005 saw the commercial availability of a new, high throughput sequencing technology, commonly known as Next-Generation Sequencing (NGS). Various sequencing techniques like real-time sequencing technology (real-time PCR, microarray), parallel sequencing, Real-Time Whole-Genome Sequencing (in silico and in vitro end-point PCR, multiplex PCR), etc. have been mapped into second generation sequencing. This review paper has shed light on these genome sequencing methods in association with food-borne illnesses, after accessing the accomplishments achieved by the first generation.

Bioinformatics has come a long way in the diagnosis and surveillance of food-borne illness programs. In the wake of the first generation of sequencing, it was proposed that there should be four mechanisms to a well-designed surveillance system. First, early alert of the disease should result in a swift action for preclusion of any further spread by the concerned agencies. Second, test conducted on affected individuals can help isolate and characterize the magnitude and intensity of the disease. Third, immediately after attaining the results, it should be reported to provisional or national levels for record and awareness. Fourth, case-control studies and lab examination of stool and food trials may be carried out to attain detailed statistics on the link between the food and the disease [55]. The current genome sequencing technologies, encompassing both short and long read sequencing techniques have ensured the implementation of these programs.

II. FIRST GENERATION

The first generation of genome sequencing follows the principles of chain termination method developed by Sanger and Coulson in 1975, also called Sanger Chain Termination Method. The first bacterial genome sequence, for Haemophilus influenza, was done by this method in 1995. The sequencing method was named as "random shotgun" sequencing as it was done from both ends.

The core idea of Sanger method is that dideoxyribonucleotide (ddNTP) cannot form phosphodiester bond during the synthesis of the DNA due to lack of hydroxyl group, so adding a certain proportion of ddNTP with radioactive isotope label into four DNA synthesis reaction to

inhibit the process. This process is followed by gel electrophoresis and autoradiography which leads to the determination of DNA sequence according to the position of electrophoretic bands [1].

This has been being used and studied for over 30 years now, but its limitations have made it a secondary tool to sequencing. Sanger sequencing technique was able to sequence only a limited number of nucleotides, generally fewer than 1000 base pairs and this consumed time and hefty amount of money along with specific skilled expertise.

III. MOLECULAR CHARACTERIZATION

During the time of modification and development of new genome sequencing methods, scientists started merging Sanger's method with Next Generation Sequencing for molecular characterization to acquire more efficiency in getting desired results.

By this time, more emphasis was given upon closing or finishing genomes and the first assembly programs such as Allora, Celera Assembler, and TIGR Assembler were used as genome assembly algorithm [2].

The first H. Influenza bacterial genome was closed after initial assembly with TIGR Assembler and various algoritms were used to map and fill physical gaps, while sequence gaps were closed by primer walking [2].

3.1.Viruses

3.1.1.Norovirus (NoV)

NoVis is one of the human pathogens associated with non-bacterial gastroenteritis outbreaks. Epidemiological surveys have shown that this is mostly transmitted by contaminated food and water and affects noth adults and children [3]. Norovirus isolated from clinical samples were genotyped using RT-PCR with primers to amplify specific region. Purified products underwent Sanger's sequencing, which was done by using ABI Prism Genetic Analyzer [4].

3.1.2.Hepatitis A (HAV)

HAV associated foodbourne outbreaks are frequently reported worldwide and has been classified into six genotypes [5]. Viral WGS was used to link the relationship between different sources followed by PCR and Sanger's sequencing approaches [6]. HCV genotype patient was first determined by NS5B Sanger sequencing [7].

3.1.3.Human adenoviruses (HAdVs)

They are of major public health importance with a variety of clinical manifestations, including gastroenteritis [8]. This may vary from absence of symptoms in healthy carriers to death in immune-compromised individuals [9]. In the past PCR and Sanger sequencing approaches were employed to detect HAdV contamination in food and water. However these methods identified only a limited number of predominant species [10]. In an experiment conducted over a period of one year in 2013, fifty-nine samples, out of 141, were characterized by conventional Sanger sequencing as belonging to four HAdV species and four types: A, B C, and F. The remaining samples could not be characterized [8].

IV. SEQUENCING TECHNIQUES

Fast and accurate identification and typing of pathogens are essential for operative surveillance and outbreak detection. The current routine

procedure is based on a variety of techniques, making the procedure laborious, time-consuming, and expensive .For proper surveillance and outbreak detection, it is important to detect and estimate pathogen pathogenicity on food items to understand different food borne diseases. Traditional techniques which were used for detection of pathogenicity facilitate detection of pathogensof low concentration in produce, in theory. It is a fact that culture based techniques although offer valuable information, they can only detect organisms that are culturable, and therefore, many others are left out. This type of conventional approach is very time consuming and labour intensive and allows identification of single species at a time wherein large commercial research is the requirement.

4.1 Next generation sequencing (NGS)

NGS approaches have become popular in the recent times due to their advantages over conventional methods of detection.

NGS theoretically allows for the non-targeted detection of multiple spoilage agents and pathogens and enables both culturable and non-culturable taxa to be characterized. Although methods such as liquid chromatography–mass spectrometry (LC-MS) also allow for the non-targeted detection of bacterial species[12], they suffer from a lack of standardisation and data on public databases. Additionally, these methods do not give information on the strains and genomes of the bacteria detected as NGS does[12].NGS is increasingly being utilised in research in food microbiology, with most studies focusing on whole genome sequencing (WGS) of bacterial isolates and the exploration of the micro biome associated with specific commodities [12].These studies highlight the diversity of microbes associated with foodstuffs, but few studies have, as yet, explored the use of NGS to screen fresh produce for the presence of human pathogens, while also delivering information about the associated micro biome [16].

4.2 Real-time sequencing technology

(Evaluation of bacterial contamination in raw milk, ultra-high temperature milk and infant formula using single molecule)

Several studies have been performed by incorporating various molecular techniques, such as pulsed-field gel electrophoresis, real-time PCR, and microarray, to investigate the prevalence of contaminated bacteria and the level of the contamination in dairy products [15].Though a few upgrades have been indicated utilizing these atomic based discovery strategies, one regular constraint shared by these procedures is the disappointment in giving a general microbial profile in the examples, as a large portion of these methodologies require the recognition of specific target microorganisms.

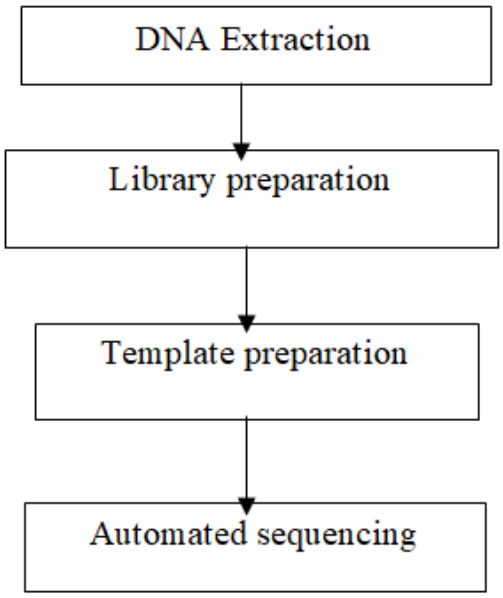
Their study reports for the first time the feasibility of using the latest PacBio SMRT sequencing technology to define the bacterial contamination profile of dairy products based on the full-length 16S rRNA gene[15]. The PacBio SMRT technology is beneficial over other sequencing approaches as it has previously been shown to offer a higher taxonomical resolution in profiling the bacterial communities in the environmental samples due to its high capacity in producing long sequence reads [15].

Clear bacterial structural differences were revealed across 16 samples of raw milk, UHT milk, and infant formula, by employing the SMRT sequencing technology. Their results have for starters shown this sequencing method can give an exact taxonomic grouping to the accuracy of the variety and species levels, and consequently generally increment our ability in separating between bacterial contaminants. However, because of the low number of tests utilized in this investigation, further top to bottom work will be expected to set up a more complete arrangement of rules for its future application in the dairy quality control during modern creation.

4.3 Parallel sequencing

The short read technologies currently in use are collectively known as massively parallel sequencing and are often also referred to as second-generation sequencing billions of nucleotide sequences are produced during each run, where each genome is sequenced multiple times in small random pieces to generate very large data sets. [12]

Even though platforms have different biochemistry and arrays, the workflows include similar steps: (i) DNA extraction; (ii) library preparation, which usually includes shearing the DNA either mechanically or enzymatically, adding adaptors and barcodes/indexes and amplification; (iii) template preparation, either by bridge amplification or emulsion PCR; and (iv) automated sequencing .The short read sequencing platforms differ substantially in terms of their engineering, sequencing chemistry, output (length of reads, number of sequences), accuracy and cost .[12]



4.4 Real-Time Whole-Genome Sequencing

(Study conducted for Routine Typing, Surveillance, and Outbreak Detection of Verotoxigenic Escherichia coli)

In the current study we developed a novel method for FBP detection based on parallel sequencing of multiple amplicons, an approach which is clearly different from the 16S rRNA-based microbial profiling method as the latter relies on sequencing of a single gene. [14].This novel FBP detection approach is based on a primer collection called "foodborne pathogen-panel" (FPP), which generates species- specific amplicons designed for sequencing on an Illumina platform. FPP was evaluated by means of in silico and in vitro end-point PCR using DNA from selected FBPs, while specificity, sensitivity and the use of multiplex PCR were validated and optimized employing single- or mixed- DNA template experiments.[14]

In total, sequence reads ranging in number from 95,059 to 422,836 were generated by Illumina MiSeq DNA sequencing. Primer sets targeting the FBPs yielded specific PCR products, as displayed by the presence of sequencing reads that mapped onto FBPs ,reaching more than 66.1 % of mapped reads of the total sequencing output in the cheese sample and perfectly aligned with the targeted gene sequences. These findings corroborate the high specificity of the primer pairs selected for the FBPs used in this experiment. [14]

An advantage of the FPP method is the simultaneous identification of multiple FBPs, thus allowing direct detection in complex matrices without enrichment steps. Moreover, assessment of the sensitivity and specificity of FPP indicate that it is an accurate method, which can be compared with other DNA-based techniques such as qPCR and multiplex qPCR (Velusamy et al., 2010; Villamizar-Rodriguez et al., 2015).

4.5 Limit of detection using ScriptSeq

In the study conducted the most sensitive NGS approach employed ScriptSeq (in conjunction with

enrichment by ribosomal depletion), with bioinformatic analysis using Kraken in conjunction with the mini database [12]there were huge contrasts in the LoD saw between bioinformatics strategies for task of Salmonella, featuring the importance of normalizing pipelines and information bases.

When NGS is compared to both culturing and rapid methods, it can give us a bounty of additional information, such as that regarding other members of the microbiome, allowing for potential source tracking of contamination, and changes in the microbiome associated with contamination or spoilage[12].

4.6 UltraStrain for sensitive strain typing

UltraStrain is a highly sensitive, rapid and efficient method for metagenomic taxonomic classification at strain level. In UltraStrain pipeline, the reads filtering step uses a synthetic reference genome consisting of differentiating regions from known S. enterica strains to filter out the reads that are not specific to S. enterica species, greatly improving the efficacy as well as efficiency of the process. Strain identification through the proposed statistical learning provides a fast and accurate solution for metagenome sample data analysis[20].

Experiments on both computer-generated data sets and real sample demonstrate that UltraStrain achieves high accuracy even at very low abundance level. It can achieve both shorter run time and higher sensitivity, which indicates its usability as a standalone pathogen identification pipeline[20]. In addition, their trials showed that the sensitivity of UltraStrain can be improved by using deeper sequencing of the sample, which could be particularly useful when it is necessary to perform strain typing on sample with extremely low abundance of target strains[20].

The proposed algorithm can be further improved in many characteristics. For example, although it is established with the target of high-sensitivity S. enterica in mind, the proposed agenda can be easily extended to taxonomic profiling and analyze other bacteria strains by adapting its filter and reference library designs. Furthermore, the capacity of current calculation in managing test with more than one target strains from similar species actually needs further examination. Significantly, the current methodology, as its essential objective is for ultra sensitive strain typing, comes up short on the capacity to precisely recognize the relative abundance of numerous microscopic organisms species/strains present in an example as given by other comparable devices. In this way, it is foreseen that it could be utilized related to other metagenomic pipelines when essential.

4.7 Information portability between WGS platforms.

Henceforth, WGS-based genotype information portability between various NGS platforms and logical pipelines seems feasible temporarily. Distinctive computational methodologies can be utilized to foresee antibiotic resistance from WGS information, the least complex by mapping of the sequence peruses against a reference data set of resistance genes or mutations, scoring the absence or presence of these elements, and anticipating an opposition profile in like manner. Nonetheless, the foundation of curated information bases on drug resistance genes will be important to defeat the quality gaps in distributed pheno genotype connections that are right now hampering the precision of susceptibility phenotype expectations from WGS information.

V. INVESTIGATING OUTBREAKS, SURVEILLANCE SYSTEM AND MODERNIZING FOOD SAFETY WITH GENOMICS, BIOINFORMATICS, AND OPEN DATA ACCESS

With rapid population growth we are witnessing several changes in life style, diets, import and export system, modifications of food

and drinks and frequency of emergence of diseases. To keep a record of these changes in different fields of work, a single surveillance system will not be enough. This necessity of monitoring billions of people spread over hundreds of places has called for experimenting with new tools and combination of systems. This has resulted in different types of surveillance like, active surveillance system, passive surveillance system and integrated system. While different governments of different countries might not have the same approach towards surveilling FDS but their approaches have the foundational base of the above mentioned surveillance systems.

A few enormous scope pilot ventures have been implemented that apply WGS and current bioinformatics examinations to existing foodborne diseases reconnaissance programs. All the more explicitly, these projects are pointed toward supplanting current subtyping approaches that support a great part of the advanced sanitation lab tasks, with WGS information for ongoing molecular observation. These modernization endeavours speak to one of the most urgent changes throughout the entire existence of sanitation, with advantages and in general effect just beginning to be figured it out. To get a sense of the scale required for this move, it is critical to survey the part of sub-atomic subtyping in irresistible infection observation and control.

4.1 WGS as a swap for current subtyping techniques

At its most basic level, subtyping is used to discriminate strains from the same species and to infer genetic relatedness, linking clinical cases representing a possible outbreak and further linking them to potential sources of infection [22]. More often, this goal is challenging to achieve amongst a background of sporadic cases in the absence of clear epidemiological links [23]. Use of standardized laboratory protocols, standardized approaches for analysis and interpretation of data, and a common convention for naming molecular subtypes, collectively have been critical to large-scale deployment of subtyping for routine surveillance.

In considering WGS as a swap for current subtyping techniques, it is important that in atomic the study of disease transmission the key supposition that will be that subtyping information is an intermediary for the fundamental genomic data from which it is inferred. Existing composing strategies would thus be able to be seen as transitory arrangements in a period when fast and modest WGS was impractical; rise of NGS and appropriation of WGS are settling this constraint for general wellbeing and sanitation examinations. Despite the fact that WGS information can be dissected utilizing a conventional phylogenetic structure, the use of NGS in epidemiological observation requires approaches for WGS-based subtyping and furthermore for relating WGS information to a subtype through a classification plot. WGS-subtyping encourages effective examination of WGS information and is fundamental given the exponential increment in accessible information. A terminology is indispensable to the correspondence of results to general wellbeing or sanitation experts, permitting the observing of epidemiological patterns and encouraging a quick reaction focused on sickness anticipation and control.

Of the two main strategies proposed for WGS-based subtyping, the first is based on the analysis of single nucleotide variant [SNV; also called single nucleotide polymorphism (SNP)] and small insertions/deletions (indels) between strains. Although this type of analysis can be performed on draft genome assemblies, several tools have been developed that directly compare raw sequence reads to a related reference genome sequence [24].

This process, which is referred to as variant detection by reference mapping, relies on algorithms that align each read to a reference genome and index the variation between them, also assigning confidence levels to each variant position

based upon the sequence coverage and level of agreement between reads supporting the SNV [25]. Reference mapping methodology has been used extensively in studies that have successfully used WGS in outbreak investigations [26]. Reference mapping also is the approach that has been employed in analyzing *S. enterica* data within the large-scale, international GenomeTrakr project [27].

The second major strategy for WGS-based subtyping is the 'gene-by-gene' approach, based on the original MLST concept [28] but extended to the whole-genome level (wgMLST) [29]. MLST is based on indexing variation where each locus, a gene or gene fragment, is used as the basic unit of comparison. 53 ribosomal protein subunit genes suitable for resolving bacterial isolates at all taxonomic levels; and core genome MLST (cgMLST) [30], which targets the genes shared by all or most members of a species (i.e., core genes). Genome-wide approaches to MLST have been applied to *Campylobacter jejuni* [31] and several other pathogens [32]. The approach recently has been validated in a PulseNet International pilot project performing real-time NGS-based typing of *L. monocytogenes*. PulseNet International also recently has committed to the wgMLST approach for their routine surveillance of foodborne disease [33].

A drawback of cgMLST is that the numbers of genes in the core for any group of strains are dramatically lower than the total number available in a species 'pan-genome,' which is comprised of both the core and any accessory genes present in only some strains [34]. It is possible, however, to design *ad hoc* MLST schemes based on the expanded number of genes shared by a smaller subset of genomes, thus providing additional discriminatory power when a low level of genetic variability is expected, as is the case in a rapidly expanding outbreak [35].

4.2 Quality control measures for WGS usage

To allow rigid utilization of WGS information as standard general wellbeing practice, quality control measurements, (for example, arrangement inclusion) and understanding rules are required. Remorsefully, such measurements and standards are as yet being characterized for the field and stay a "moving objective." Additionally they do shift with bacterial species, the time period of an examination, and the philosophy attempted for the investigations; consequently, no simple "one size fits all" approach exists. In spite of the fact that it stays untimely to depict quality measurements and understanding models in explicit terms, key variables affecting grouping information age have been uncovered (in any event for the develop sequencing advancements) and there are continuous worldwide endeavours to formalize how to produce solid information and how to vigorously decipher the information with certainty. The undertaking is a troublesome one since the sequencing boundaries, time span of examination, and transformative elements of the creature all impact the connection of genomic variety and epidemiological understanding in a perplexing manner. Attributable to the significance of information age and understanding in foodborne episode examinations, they stay high needs and will get impressive consideration for years to come.

VI. CURRENT BIOINFORMATICS TOOLS

Several large-scale pilot projects have been implemented that apply NGS and modern bioinformatics analyses to existing foodborne disease surveillance programs. Whole genome sequencing (WGS) is now making it feasible to sequence food-borne isolates in real-time, such as during food-borne illness outbreak investigations [36]

The utilization of bioinformatics in the microbiological risk examination is conveying clear advantages both for the examination of food borne microorganisms and for decision making on clinically significant treatments.

4.1Regulatory bioinformatics

"Regulatory bioinformatics" strives to develop and implement a standardized and transparent bioinformatic framework to support the implementation of existing and emerging technologies in regulatory decision-making. For example, GSRS2015 explored a range of applications of regulatory bioinformatics in the development and use of medical products and as tools to manage the safety of the food supply [37].

4.1.1Current WGS technologies can be subdivided into two categories based on the length of the sequence reads they produce; (i) short read sequencing technologies, producing reads up to 600 base-pairs long (e.g., Illumina, Ion Torrent), and (ii) long read technologies, capable of producing reads longer than 1000 base-pairs and often longer than 70,000 base-pairs (e.g., Pacific Biosciences, Oxford Nanopore).

5.2 Short read sequencing technologies:

5.2.Illumina sequencing

It is currently the prevailing HTS technology and also offers the highest fidelity.[40] The HiSeq2000 and MiSeqIllumina platforms to community amplicon sequencing generates the information that on comparison, these differ in scale and thus support totally different applications. The HiSeq platform permits massively parallel sequencing at the lowest cost. Comparable information will be generated on the MiSeq for smaller projects where it's vital to process samples quickly.[45]

5.2.2 Ion Torrent sequencing

It generates comparable number of sequence reads and displays similar microbial community structure to that obtained in a pyrosequencing pilot study. Ion Torrent sequencing also proves to be accurate based on sequences generated in controls. [39]

5.3 Long read technologies:

PacBio and ONT are similar: long read length, high error rate and relatively low throughput. However, they have distinct aspects, such as homopolymer error in PacBio and context-specific mismatches in ONT. PacBio sequences a molecule multiple times to generate high-quality consensus data, while ONT can only sequence a molecule twice. Together with the higher quality of the raw data, PacBio can generate extremely-low-error-rate data for high-resolution studies, which is not feasible for ONT. However, ONT has a few advantages: in addition to slightly longer mappable length, ONT MinION provides very high throughput as the nanopores can sequence multiple molecules. [38]

5.4 Metagenomics:

Microbial community analysis of metagenomic samples using 16S rRNA sequencing has been applied to diverse studies like the study of bacterial communities associated with fresh vegetables etc. It has also been used for determining the sensitivity of metagenomic shotgun sequencing for detection of a pathogen contaminant associated with the spinach.[41] and *Salmonella* contamination among cilantro samples[43] The database-based approach for the strain-level analysis of metagenomic data permits to achieve bioinformatic separation of individual bacterial strains present in an enriched food sample.[42] Metagenomics identifies the entire gene content and when coupled to transcriptomics or proteomics, allows the identification of functional capacity and biochemical activity of microbial populations.[44]

Although the application of NGS has brought great advances to epidemiological research over traditional methodologies for strain characterization, the whole process from the sample processing to sequencing and data analysis is more complex, leading to new challenges: multiple protocols for sample and library preparations are available and each sequencing run can use different versions of sequencing units and consumable

reagents; moreover, in terms of data analysis, there are potentially hundreds of different software and respective versions that can be used and need to be tracked. Therefore, to facilitate comparative analyses the entire process from sample to analysis of results should be annotated.[36]

VII. FUTURE PROSPECTS AND DISCUSSION

The first generation of genomic sequencing methods and analysis pipelines are now in the final stages of translation into routine application in modern food safety labs, and may soon replace many conventional laboratory tests. These advanced genomic and bioinformatics systems have proven their worth in reducing response times to emerging foodborne disease outbreaks, with substantial socioeconomic benefits in terms of improved public health, reduced health care costs, and avoidance of lost productivity due to illness (Scharff et al., 2016). The ongoing global efforts to modernize our food safety systems with genomics and bioinformatics have been impressive, but there remain many challenges and opportunities.[56]

Our present investigative limit actually requires the refined of bacterial disengages, which can take a few days. Culture-autonomous indicative testing utilizing metagenomic advancements vows to get rid of essential refined of detaches, hence shortening our reaction times significantly more.

Culture free metagenomics procedures have their own issues, notwithstanding, for example, the huge measure of non-target information produced, tainting from ecological sources, and a current failure to recognize arrangements got from live or dead microorganisms. The immense and ever-developing size of the microbe genome information bases requires considerable superior processing assets and novel algorithmic ways to deal with investigate such huge informational collections on a valuable timescale.. Metadata normalization is gaining acceptable ground, however will require extensive continued exertion over several years to arrive at development.

VIII. CONCLUSION

While the developed countries have undertaken to use WGS for better surveillance and have already started sharing the data in different open-to-all databases, it is the low- and middle-income countries that are short on enthusiasm of taking up new tools and technologies for the betterment of public health. This is a result of lack of human resources, communication and limited surveying.

There are many ways for reaching out to these countries and different fields of research so that there will be a significant boost in the use of this new technology and also, monitoring and controlling of outbreaks of FBD would become uncomplicated when it comes to working together on a global scale. Writing and accessing substantial amount of literatures has been proven to show a positive shift in the willing nature of organizations to work towards it. Peer-reviewed literatures hold importance in this aspect. A surge of attention through press and commercial media platform is needed to support this system for its efficient working. Implementing various policies and regulations, such as asking people to contribute in sequencing studies, along with giving then the guarantee of confidentiality and ethical rights.

More online databases should be available for free to reach out organizations that work towards this issue and also, to educational institutes so that the students can get an insight on the current progress of this system and its future aspects. There are several of these online platforms under development and many already available.

Working with new technologies along with the traditional methods will give them a more transparent outlook on it. Translating this system globally would require the involvement of everyone

from clinicians to health experts, laboratories workers, researchers and the staff of assessment of these outbreaks. Workshops, collaborations, honest exchange of information, and response can lead to a acquiring major benefits for local, national and global public health.

ACKNOWLEDGMENT

This paper would not have been possible without the kind support and help of many individuals and organizations. We would like to extend our sincere and foremost thanks to VIT University, Vellore and VIT School of Biosciences and Technology for enabling us to have this wonderful opportunity.
We are highly indebted to Dr. Ramalingam C.for his guidance and constant supervision as well as for providing necessary information regarding the project & also for his support in completing the project.
We would like to express our gratitude towards our parents & PhD scholars of this project for their kind co-operation and encouragement which helped us in completion of this project.
Our thanks and appreciations also go to our colleague in developing the project and people who have willingly helped us out with their abilities.

REFERENCES

[1] Yang, X., Noyes, N. R., Doster, E., Martin, J. N., Linke, L. M., Magnuson, R. J., et al. (2016). Use of metagenomic shotgun sequencing technology to detect foodborne pathogens within the microbiome of the beef production chain. Appl. Environ. Microbiol. 82, 2433–2443. doi: 10.1128/AEM.00078-16

[2] KorenS ,Phillippy AM . 2015. One chromosome, one contig: complete microbial genomes from long-read sequencing and assembly. CurrOpinMicrobiol 23:110–120. doi:10.1016/j.mib.2014.11.014.

[3] MORILLO, Simone Guadagnucci et al . Identification and Molecular Characterization of Norovirus in São Paulo State, Brazil. Braz. J. Microbiol., São Paulo , v. 39, n. 4, p. 619-622, Dec. 2008 .

[4] Ferreira, M.S.R.; Miagostovich, M.P.; Xavier, M.P.T.P.; Fumian, T.M.; Fialho, A.M.; Leite, J.P.G. (2005). Norovirus associated with outbreaks of acute gastroenteritis in middle Paraiba region of the state of Rio de Janeiro. In: National Meeting of Virology.

[5] Chi H , Haagsma EB , Riezebos-Brilman A , van den Berg AP , Metselaar HJ , de Knegt RJ . 2014. Hepatitis A related acute liver failure by consumption of contaminated food. J ClinVirol 61:456–458. doi:10.1016/j.jcv.2014.08.014.

[6] Chiapponi C , Pavoni E , Bertasi B , Baioni L , Scaltriti E , Chiesa E , Cianti L , Losio MN , Pongolini S . 2014. Isolation and genomic sequence of hepatitis A virus from mixed frozen berries in Italy. Food Environ Virol 6:202–206. doi:10.1007/s12560-014-9149-1.

[7] K. Sandres-Sauné, P. Deny, C. Pasquier, V. Thibaut, G. Duverlie, J. Izopet Determining hepatitis C genotype by analyzing the sequence of the NS5b region J Virol Methods, 109 (2003), pp. 187-193

[8] Iaconelli M, Valdazo-González B, Equestre M, et al. Molecular characterization of human adenoviruses in urban wastewaters using next generation and Sanger sequencing. Water Research. 2017 Sep;121:240-247. DOI: 10.1016/j.watres.2017.05.039.

[9] ArnbergN . 2012. Adenovirus receptors: implications for targeting of viral vectors. Trends PharmacolSci 33:442–448. doi:10.1016/j.tips.2012.04.005.

[10] Mena KD ,Gerba CP . 2009. Waterborne adenovirus. Rev Environ ContamToxicol 198:133–167. doi:10.1007/978-0-387-09647-6_4.

[11] J. Ronholm, NedaNasheri, Nicholas Petronella, Franco Pagotto Clinical Microbiology Reviews Aug 2016, 29 (4) 837-857

[12] Lewis, E., J. A. Hudson, N. Cook, J. D. Barnes, and E. Haynes. 2020. "Next-Generation Sequencing as a Screening Tool for Foodborne Pathogens in Fresh Produce." *Journal of Microbiological Methods* 171 (April). Elsevier B.V. doi:10.1016/j.mimet.2020.105840.

[13] Besser J, Carleton HA, Gerner-Smidt P, Lindsey RL, Trees E. Next-generation sequencing technologies and their application to the study and control of bacterial infections. *ClinMicrobiol Infect*. 2018;24(4):335-341. doi:10.1016/j.cmi.2017.10.013

[14] Joensen, K. G., Scheutz, F., Lund, O., Hasman, H., Kaas, R. S., Nielsen, E. M., &Aarestrup, F. M. (2014). Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic Escherichia coli. *Journal of Clinical Microbiology*, *52*(5), 1501–1510.https://doi.org/10.1128/JCM.03617-13

[15] Hou, Q., Xu, H., Zheng, Y., Xi, X., Kwok, L. Y., Sun, Z., … Zhang, W. (2015). Evaluation of bacterial contamination in raw milk, ultra-high temperature milk and infant formula using single molecule, real-time sequencing technology. *Journal of Dairy Science*, *98*(12), 8464–8472.https://doi.org/10.3168/jds.2015-9886

[16] Ferrario, C., Lugli, G. A., Ossiprandi, M. C., Turroni, F., Milani, C., Duranti, S., … Ventura, M. (2017). Next generation sequencing-based multigene panel for high throughput detection of food-borne pathogens. *International Journal of Food Microbiology*, *256*, 20–29.https://doi.org/10.1016/j.ijfoodmicro.2017.05.001

[17] Miralles, M. M., Maestre-Carballa, L., Lluesma-Gomez, M., & Martinez-Garcia, M. (2019). High-throughput 16S rRNA sequencing to assess potentially active bacteria and foodborne pathogens: A case example in ready-to-eat food. *Foods*, *8*(10).https://doi.org/10.3390/foods8100480

[18] Miralles, M. M., Maestre-Carballa, L., Lluesma-Gomez, M., & Martinez-Garcia, M. (2019). High-throughput 16S rRNA sequencing to assess potentially active bacteria and foodborne pathogens: A case example in ready-to-eat food. *Foods*, *8*(10).https://doi.org/10.3390/foods8100480

[19] Li, Y., Zhang, S., Li, J., Chen, M., He, M., Wang, Y., … Pang, B. (2019). Application of digital PCR and next generation sequencing in the etiology investigation of a foodborne disease outbreak caused by Vibrio parahaemolyticus. *Food Microbiology*, *84*.https://doi.org/10.1016/j.fm.2019.05.017

[20] Yang, W., Huang, L., Shi, C., Wang, L., & Yu, R. (2019). UltraStrain: An NGS-based ultra sensitive strain typing method for salmonella enterica. *Frontiers in Genetics*, *10*(APR).https://doi.org/10.3389/fgene.2019.00276

[21] Desdouits, M., de Graaf, M., Strubbia, S., Oude Munnink, B. B., Kroneman, A., Le Guyader, F. S., & Koopmans, M. P. G. (2020). "Novel opportunities for NGS-based one health surveillance of foodborne viruses." *One Health Outlook*, *2*(1). https://doi.org/10.1186/s42522-020-00015-6

[22] Sabat AJ, Budimir A, Nashev D, Sá-Leão R, van DijlJm, Laurent F, Grundmann H, Friedrich AW; ESCMID Study Group of Epidemiological Markers (ESGEM).Euro Surveill. 2013 Jan 24;18(4):20380. doi: 10.2807/ese.18.04.20380-en.

[23] WeiweiLi[a] ShuyuWu[b] PingFu[a] JikaiLiu[a] HaihongHan[a] LiBai[a] XiaoyanPei[a] NingLi[a] XiumeiLiu[a] YunchangGuo

[24] Reinert T, Schøler LV, Thomsen R, Tobiasen H, Vang S, Nordentoft I, Lamy P, Kannerup AS, Mortensen FV, Stribolt K, Hamilton-Dutoit S, Nielsen HJ, Laurberg S, Pallisgaard N, Pedersen JS, Ørntoft TF, Andersen CL.Gut. 2016 Apr;65(4):625-34. doi: 10.1136/gutjnl-2014-308859. Epub 2015 Feb 4.

[25] J Appl Genet. 2016 Feb;57(1):71-9. doi: 10.1007/s13353-015-0292-7. Epub 2015 Jun 9.PMID: 26055432

[26] Taboada EN, Graham MR, Carriço JA, Van Domselaar G. Food Safety in the Age of Next Generation Sequencing, Bioinformatics, and Open Data Access. *Front Microbiol*. 2017;8:909. Published 2017 May 23. doi:10.3389/fmicb.2017.00909

[27] Allard MW, Strain E, Melka D, Bunning K, Musser SM, Brown EW, Timme R.J ClinMicrobiol. 2016 Aug;54(8):1975-83. doi: 10.1128/JCM.00081-16. Epub 2016 Mar 23.PMID: 27008877

[28] Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG.ProcNatlAcadSci U S A. 1998 Mar 17;95(6):3140-5. doi: 10.1073/pnas.95.6.3140.PMID: 9501229

[29] Sheppard SK, Maiden MC. The evolution of Campylobacter jejuni and Campylobacter coli. *Cold Spring HarbPerspect Biol*. 2015;7(8):a018119. Published 2015 Jun 22. doi:10.1101/cshperspect.a018119

[30] Jolley KA, Maiden MC.BMC Bioinformatics. 2010 Dec 10;11:595. doi: 10.1186/1471-2105-11-595.PMID: 21143983

[31] Shepherd, A., Ivins, E. R., A, G., Barletta, V. R., Bentley, M. J., Bettadpur, S., Briggs, K. H., Bromwich, D. H., Forsberg, R., Galin, N., Horwath, M., Jacobs, S., Joughin, I., King, M. A., Lenaerts, J. T. M., Li, J., Ligtenberg, S. R. M., Luckman, A., Luthcke, S. B. et al., 2012, 'A reconciled estimate of ice-sheet mass balance',Science338(6111), 1183–1189 (DOI: 10.1126/science.1228102)

[32] Pightling AW, Pettengill JB, Luo Y, Baugher JD, Rand H, Strain E. Interpreting Whole-Genome Sequence Analyses of Foodborne Bacteria for Regulatory Applications and Outbreak Investigations. *Front Microbiol*. 2018;9:1482. Published 2018 Jul 10. doi:10.3389/fmicb.2018.01482

[33] Jackson BR, Tarr C, Strain E, Jackson KA, Conrad A, Carleton H, Katz LS, Stroika S, Gould LH, Mody RK, Silk BJ, Beal J, Chen Y, Timme R, Doyle M, Fields A, Wise M, Tillman G, Defibaugh-Chavez S, Kucerova Z, Sabol A, Roache K, Trees E, Simmons M, Wasilenko J, Kubota K, Pouseele H, Klimke W, Besser J, Brown E, Allard M, Gerner-SmidtP.Clin Infect Dis. 2016 Aug 1;63(3):380-6. doi: 10.1093/cid/ciw242. Epub 2016 Apr 18.PMID: 27090985

[34] Tettelin H, Riley D, Cattuto C, MediniD.CurrOpinMicrobiol. 2008 Oct;11(5):472-7. doi: 10.1016/j.mib.2008.09.006.PMID: 19086349

[35] Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG.ProcNatlAcadSci U S A. 1998 Mar 17;95(6):3140-5. doi: 10.1073/pnas.95.6.3140.PMID: 9501229

[36] Taboada EN, Graham MR, Carriço JA and Van Domselaar G (2017) Food Safety in the Age of Next Generation Sequencing, Bioinformatics, and Open Data Access. Front. Microbiol. 8:909. doi: 10.3389/fmicb.2017.00909

[37] Healy, M. J., Tong, W., Ostroff, S., Eichler, H.-G., Patak, A., Neuspiel, M., …Slikker, W. (2016). *Regulatory bioinformatics for food and drug safety. Regulatory Toxicology and Pharmacology, 80, 342–347*.doi:10.1016/j.yrtph.2016.05.021

[38] Chen Z, Erickson DL, Meng J. Benchmarking hybrid assembly approaches for genomic analyses of bacterial pathogens using Illumina and Oxford Nanoporesequencing. *BMC Genomics*. 2020;21(1):631. Published 2020 Sep 14. doi:10.1186/s12864-020-07041-8

[39] Fujimoto M, Moyerbrailean GA, Noman S, et al. Application of ion torrent sequencing to the assessment of the effect of alkali ballast water treatment on microbial community diversity. *PLoS One*. 2014;9(9):e107534. Published 2014 Sep 15. doi:10.1371/journal.pone.0107534

[40] Sekse C, Holst-Jensen A, Dobrindt U, et al. High Throughput Sequencing for Detection of Foodborne Pathogens. *Front Microbiol*. 2017;8:2029. Published 2017 Oct 20. doi:10.3389/fmicb.2017.02029

[41] Leonard SR, Mammel MK, Lacher DW, Elkins CA. Application of metagenomic sequencing to food safety: detection of Shiga Toxin-producing Escherichia coli on fresh bagged spinach. *Appl Environ Microbiol*. 2015;81(23):8183-8191. doi:10.1128/AEM.02601-15

[42] Saltykova A, Buytaers FE, Denayer S, et al. Strain-Level Metagenomic Data Analysis of Enriched In Vitro and In Silico Spiked Food Samples: Paving the Way towards a Culture-Free Foodborne Outbreak Investigation Using STEC as a Case Study. *Int J Mol Sci*. 2020;21(16):5688. Published 2020 Aug 8. doi:10.3390/ijms21165688

[43] Jarvis KG, White JR, Grim CJ, et al. Cilantro microbiome before and after nonselective pre-enrichment for Salmonella using 16S rRNA and metagenomic sequencing. *BMC Microbiol*. 2015;15:160. Published 2015 Aug 12. doi:10.1186/s12866-015-0497-2

[44] Jagadeesan B, Gerner-Smidt P, Allard MW, et al. The use of next generation sequencing for improving food safety: Translation into practice. *Food Microbiol*. 2019;79:96-115. doi:10.1016/j.fm.2018.11.005

[45] Caporaso JG, Lauber CL, Walters WA, et al. Ultra-high-throughput microbial community analysis on the IlluminaHiSeq and MiSeq platforms. *ISME J*. 2012;6(8):1621-1624. doi:10.1038/ismej.2012.8

[46] Pereira S, Gibbs RA, McGuire AL. Open access data sharing in genomic research. *Genes (Basel)*. 2014;5(3):739-747. Published 2014 Aug 29. doi:10.3390/genes5030739

[47] Marshall, E. (2001). Bermuda rules: community spirit, with teeth. *Science* 291, 1192. doi: 10.1126/science.291.5507.1192

[48] Lord, P., Macdonald, A., Sinnott, R., Ecklund, D., Westhead, M., and Jones, A. (2005). *Large-scale Data Sharing in the Life Sciences: Data Standards, Incentives, Barriers and Funding Models ("The Joint Data Standards Study").* Technical Report, No. UKeS-2006-02. Edinburgh: National e-Science Centre.

[49] Institute of Medicine (US) Forum on Microbial Threats. Vector-Borne Diseases: Understanding the Environmental, Human Health, and Ecological Connections, Workshop Summary. Washington (DC): National Academies Press (US); 2008. Summary and Assessment. Available from:https://www.ncbi.nlm.nih.gov/books/NBK52939/

[50] Saskia L. Smits, Marion P. G. Koopmans(2017). Genomics and Foodborne

Viral Infections https://link.springer.com/chapter/10.1007/978-3-319-43751-4_9

[51] Daniel Dewey-Mattia, MPH, KarunyaManikonda, MPH, Aron J. Hall, DVM,Matthew E. Wise, PhD, and Samuel J. Crowe, PhD(2018). Surveillance for Foodborne Disease Outbreaks — United States, 2009–2015.MMWR SurveillSumm. 2018 Jul 27; 67(10): 1–11.doi: 10.15585/mmwr.ss6710a1

[52] Hossein Masoumi Asl,Mohammad Mehdi Gouya, Mohammad Mehdi Soltan-dallal,and Nooshin Aghili(2015). Surveillance for foodborne disease outbreaks in Iran, 2006-2011.Med J Islam Repub Iran. 2015; 29: 285. PMID: 26913248https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4764277/

[53] Shi GUO, Dan LIN, Li-li WANG, and Hong HU (2018). Monitoring the Results of Foodborne Diseases in Sentinel Hospitals in Wenzhou City, China from 2014 to 2015.Iran J Public Health. 2018 May; 47(5): 674–681. PMID: 29922609https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6005988/

[54] Celine Nadon, Ivo Van Walle, Peter Gerner-Smidt, Josefina Campos, Isabel Chinen, Jeniffer Concepcion-Acevedo, Brent Gilpin, Anthony M. Smith, Kai Man Kam, Enrique Perez, Eija Trees, Kristy Kubota, Johanna Takkinen, Eva Møller Nielsen, Heather Carleton, FWD-NEXT Expert Panel(2017). PulseNet International: Vision for the implementation of whole genome sequencing (WGS) for global food-borne disease surveillance.https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2017.22.23.30544#html_fulltext

[55] Ewen C. D. Todd(1996). Worldwide Surveillance of Foodborne Disease: the Need to Improve. *J Food Prot* (1996) 59 (1): 82–92.https://doi.org/10.4315/0362-028X-59.1.82

[56] Taboada, E. N., Graham, M. R., Carriço, J. A., & Van Domselaar, G. (2017, May 23). Food safety in the age of next generation sequencing, bioinformatics, and open data access. *Frontiers in Microbiology*. Frontiers Media S.A. https://doi.org/10.3389/fmicb.2017.00909

