VETERINARY RESEARCH INTERNATIONAL
Journal homepage: www.jakraya.com/journal/vri

REVIEW ARTICLE

Viral Metagenomics: A Review

Mohammed Fawaz*, Ashwin Ashok Raut, Anamika Mishra, Nitin Kambla and Vijay

Indian Veterinary Research Institute, Izatnagar, Bareilly-243 122 (UP), India.

*Corresponding Author:

Mohammed Fawaz

Email: mohdfawaz@gmail.com

Received: 11/09/2014

Revised: 27/10/2014

Accepted: 30/10/2014

Abstract

Sequencing technology has shown significant improvement since its first installment in 1977 by Frederick Sanger and colleagues. The Human Genome Project resulted in great refinement of sequencing technology with many promising discoveries. These newer sequencing technologies are called the Next Generation Sequencing (NGS) technologies. Public Health and Veterinary scientists started picking up NGS technology in a number of animal disease diagnosis including the discovery of Schmallenberg viruses. A review on the use of NGS technology in animal disease diagnosis including discovery of novel microbes, differentiation of various pathotypes, evolutionary aspect of microbes and host-pathogen interaction, transcriptomics (RNA-sequencing), disease surveillance etc. is dealt with in here. With the continuous decent in the cost of DNA sequencing it is estimated that NGS would become the gold standard in the molecular diagnosis of animal diseases. What has to be undermined is that most of the efforts in the area of animal disease diagnosis related to NGS stops once a novel microbe or evolutionary pattern has been identified without further studies on the pathogenicity and related effects on the host.

Keywords: Next generation sequencing, Molecular diagnostics, Sequencing platforms.

Metagenomics Vs Viral Metagenomics

Metagenomics as the name implies is different from genomics in that it takes into account all the genomes of all the microorganisms present in the study environment whether it be a water body like sea, human body or the intestine of an animal – studying the microorganisms at a community level. 16 S rRNA gene of the metagenome is sequenced to ascertain the microbial diversity since this gene is highly conserved across species. Since viruses do not have conserved genes, the viral metagenome is sequenced as a whole lot. Metagenomics is an alternative culture-independent and sequence independent approach that does not rely on the presence of any particular gene in all the subjects being studied (Riesenfeld *et al*., 2004). Metagenomic detection of the viral pathogens is becoming now a powerful diagnostic tool in veterinary diagnostic virology with the advent of next generation sequencing technologies.

The first report of a pathogenic agent smaller than any known bacterium appeared in 1892 with the observation by Russian scientist Dimitrii Ivanowsky that causative agent of tobacco mosaic disease was not retained by the unglazed porcelain filters used at that time to remove bacteria from extracts. Six years later, Martinus Beijerinck made the same observation independently in Holland but more importantly made the conceptual leap that the pathogen responsible for tobacco mosaic virus must be a distinctive agent. The same year (1898), the German scientists Friedrich Loeffler and Paul Frosch, both former students and assistant of Koch, observed that the causative agent of foot and mouth disease, a wide spread devastating infection of cattle and other livestock, was also filterable. Beijerinck termed these agent *Contagium vivium fluidum* to emphasize its infectious nature and distinctive reproductive (replicating only inside host cell) and physical properties. The term eventually evolved simply to viruses, appropriating the term virus from the Latin for “poison”. Culture techniques have been the gold standard for the detection of viruses for over a century. The consensus is that we have barely begun to chart the viral world, which is the ‘dark matter’ of biological universe and rich source of future discoveries (Ng *et al*., 2011).

Clinical virology field shows a gradual substitution of the traditional virus discovery methods with novel molecular biology technology. Sequence

dependent and sequence independent methods are being used for the detection of viruses. Sequence dependent methods, including PCR using consensus primers and hybridization methods such as microarray, require the knowledge of the nucleic acids for the detection of novel viruses. Sequence independent viral metagenomic approach does not rely on prior knowledge of viruses in the study sample. Sequence – independent single-primer amplification (SISPA) method has been used successfully for the discovery of viruses (Reyes *et al*., 1991). Another sequence-independent technique, the viral metagenomics provides superior capability to detect known and unknown viruses than the traditional and molecular sequence-dependent and sequence-independent methods.

Compared to other viral discovery methods, viral metagenomics is less biased. Potentially, any viruses in the samples, culturable or unculturable, known or novel can be readily detected with the viral metagnomic approach (Mokili *et al*., 2012). Viral metagenomic methods have evolved significantly since they were first developed. In earlier viral metagenomic study, preliminary sample preparation involved shearing of DNA and cloning. The process of sample preparation has since been streamlined and the sequencing speed increase with the advent of high throughput sequencing technologies. The replacement of cloning with high throughput methods has revolutionized metagenomics (Mokili *et al*., 2012). There is no specific treatment for viral diseases; hence prevention is better than cure and of special importance here, the details of which would be mentioned in the coming sections.

Sample Preparation

Because viral genomes are relatively short, bacterial and eukaryotic nucleic acids can severely interfere with the isolation and detection of viral DNA or RNA. Thus, removal of non-viral nucleic acid is necessary (Thurber *et al*., 2009). Depending on the nature of sample i.e., tissue, blood, body fluid, have to homogenize, fractionize, or dilute a representative amount of original sample. This can be performed using the protocol used for other applications. It should be kept in mind that the choice of protocol affects the output depending on the host cell disruption and preservation of intact pathogen. After this, most common method is filtration, centrifugation and nuclease treatment (Thurber *et al.*, 2009). Knowledge of the size of the microbes is essential for judging the kind of filters to be used for viral enrichment of the sample. Bacteria normally range in size from 0.5µm to 5µm, Mycoplasma normally has size in the range of 0.1µm and viruses range in size from 0.02µm to 0.3µm. 0.45µm and 0.22µm are normally used by researchers working in the area of viral metagenomics. 0.22µm filter are commonly used to remove bacteria in the search for viruses. However, in the recent years viruses have been discovered which are as large as bacteria. These large viruses are often called giruses (Van Etten *et al*., 2010).

Ultracentrifugation can be used to concentrate the viruses. Most of the small laboratories might not be having this facility. Smaller laboratories might still consider a smaller bench top one, if they are dealing with larger volume of samples. Nuclease treatment is an enrichment method for viral nucleic acid which can be applied directly to the filtered material or could be used as a complement to ultracentrifugation. The capsid protects the viral genome while degrading all the external nucleic acids (Allander *et al*., 2001). A cocktail of enzymes are used including DNase, RNase, Benzonase and Turbo DNase etc. The application of these technique in a clinical setting will require that any virus enrichment methods are simple to perform, fast, robust, effective, standardized and do not require significant capital expenditure (Hall *et al*., 2014). The material extracted after nuclease treatment are normally much smaller that amplification is normally required before sequencing. Two most common amplification strategies are phi 29 polymerase based multiple displacement amplification and random PCR using modified version of sequence independent single-primer amplification (SISPA). The amplification step can introduce bias in the final result which is of little concern in pathogen detection but would influence the result of other application. Random hexamers, adapters and barcodes are added and PCR amplified. The PCR product is the library construct which could be used for sequencing directly (amplicon sequencing). In the case of viruses next generation sequencing is done for the entire nucleotides.

Library Preparation for Viral Metagenomics

Early metagenomic study involved the generation of shot gun libraries and direct sequencing of the total DNA content using the Sanger dideoxy chain termination method. The Sanger technology has been the standard method for sequencing since it was first described. With the advent of next generation sequencing, metagenomics have taken a greater cause. Library construction usually starts with double stranded DNA which is fragmented, end repaired, addition of non-template dA tails, adapter and barcodes added (if

Table 1: Comparison of NGS platforms

Machine	Chemistry	Read length	Gb/run time	Advantages	Disadvantages
High-end instrument					
454 GS FLX+ (Roche)	Pyrosequencing	700b	0.7/23 hours	•Long read lengths	• High reagent costs • High error rate in homopolymers •Appreciable hands on time
HiSeq X ten (Illumina)	Reversible terminator	1Kb		• 1000 dollar genome sequencing	• Only useful for large scale use
HiSeq 2000/2500 (Illumina)	Reversible terminator	2 x 100 b	600 /11 days (regular mode) or 120/ 2 days (rapid run mode)	•Cost effectiveness •Massive throughput •Minimal hands-on time	• Long run time •Short read lengths
5500xl SOLiD (Life Technologies)	Ligation	2 x 60 b	150/ 8 days	• Low error rate •Massive throughput	• Very short read lengths •Long run times
PacBio RS (Pacific Bioscience)	Real-time sequencing	3000 b	3 per day	•Simple sample preparation •Low reagent costs •very long read length	• High error rate •Expensive system •Difficult installation
Bench-top instruments					
454 GS junior (Roche)	Pyrosequencing	500 b	0.035/8 hours	• Long read lengths	• Appreciable hands on time •High reagent costs • High error rate in homopolymers
Ion Torrent (Life Technologies)	Proton detection	100 b	0.01-0.1(314chip) 0.1-0.5 (316 chip) up to 1 (318 chip) per 3 hours	• Short run times • Appropriate throughput for microbial application	• Appreciable hands-on time •High error rate in homopolymers
MiSeq (Illumina)	Reversible terminator	2 x 150 b	1.5/27 hours	•Cost-effectiveness •Short run times •Appropriate throughput for microbial application • Minimal hands-on time	• Read lengths too short for efficient assembly

multiplexing the sample), size selected and PCR performed to generate final library for sequencing. Different sequencing strategies are now available through various companies.

Bioinformatic Analysis

Analysis of the copious data generated by high-throughput sequencing is the most challenging aspect of metagenomics (Mokili *et al*., 2012). An inherent difficulty in assigning taxonomic designations to viral sequences is that there is no universally homologous nucleic acid component present in all viruses that can be used to build phylogenetic tree of life (Rohwer *et al*., 2002; Brussow *et al*., 2009). In most metagenomic studies, sequences generated by high throughput sequencing are queried by homology search tools to previously documented sequences stored either in a local database or in public databases such as Genbank (Mokili *et al*., 2012). Unfortunately, homology searches against known sequences in Genbank cannot characterize unknown viruses (Mokili *et al*., 2012). Bioinformatics analyses should be performed on high quality data. For this reads are typically processed through a software pipeline to remove any background sequence including host and bacterial DNA that had not been removed by viral enrichment methods like filtration, centrifugation and nuclease treatment. The resulting sequence reads are assembled with strict parameters to generate contigs, each made of sequence generated from the same organism quasi-species. Using a stringent assembly parameter is critical to avoid sequence chimerization (Mokili *et al*., 2012). The contigs sequences are then compared to the Genbank

non-redundant nucleotide database using BLAST (Benson *et al*., 2011).

Most intriguing aspect of viral metagenomes is that majority of the sequence has no significant similarity to anything known. A typical viral metagenome can contain 60% and 99% unknown sequence. Factors contributing to this variation include sample type, length of sequence reads, the homology search method (BLASTn, tBLASTx, etc.), the similarity threshold (E-value cutoff), the database and version of the database used for the homology search (Vieites *et al*., 2009). Depending on how they are viewed, the unknown can represent either a formidable challenge or a treasure trove for virus discovery. Although researcher often tend to consider the unknown as 'Junks', these sequence could be valuable blueprint for the discovery of novel viruses. There is a dearth of suitable bioinformatics methods to characterize the unknown sequence.

Veterinary Related Viral Metagenomics

Viral metagenomics started with Mya Breitbart studying the virus community in ocean and later in the human feces. From there on the journey had begun, testing several environment and lately has gained the interest of veterinary scientists and the OIE regional and reference laboratories worldwide, the focal point being Sweden.

Study on the Wild Animals

Most of the emerging infectious diseases (EIDs) originate especially from wildlife reservoirs (Woolhouse *et al*., 2005). EID events are dominated by zoonoses (60.3 % of EIDs): the majority of these (71.8 %) originate in wildlife (Jones *et al*., 2008). Changes in human living patterns, along with environmental and climate changes, pose unprecedented challenges to the global health of people, animals and ecosystems (wood *et al*., 2012). With the destruction of habitat and movement of wild animals into human contact area resulted in wider occurrence of EIDs. Some of the most threatening emerging pathogens are RNA viruses due to their unparallel ability to adapt to new host and environment (Grenfell *et al*., 2004).

Bats are an important reservoirs and vectors for spread of EIDs. Bats are associated with zoonoses of potentially great global public health impact and are the source of Lyssa virus (Warrel et al., 2004), Hendra virus (Murray *et al*., 1995), Nipah virus (Luby *et al*., 2009), Severe Acute Respiratory Syndrome (SARS)-like Corona viruses (Ksiazek *et al*., 2003; Wang *et al*., 2006) and Ebola and Marburg viruses (Towner *et al*., 2009; Towner *et al*., 2007; Leroy *et al*., (2009); Leroy *et al*., 2005). A comprehensive list of reservoir status of bats up to the year 2006 has been provided by Calisher *et al*. (2006).

A preliminary study of viral metagenomics of French bats in contact with humans identified new mammalian virus (Dacheux *et al*., 2014). They detected several new mammalian viruses including rotavirs, gammaretroviruses, bornaviruses and bunyaviruses with the identification of the first nairoviruses. Bats are the major reservoir for Hepaciviruses and Pegiviruses (Quan *et al*., 2013). Influenza virus, of the new H18N11 subtype was discovered in a flat faced fruit bat (Tong *et al*., 2013). Virome profiling of bats from Myanmar by metagenomic analysis of tissue sample revealed more novel mammalian virus (He *et al*., 2013). Novel bat viruses of the genera mamastrovirus, Bocavirus, Circovirus, Iflavirus and Orthohepdavirus were discovered. In the metagenomic study of the viruses of African straw-coloured fruit bat by Baker *et al*. (2013), they detected a chiropteran poxvirus and isolated a novel adenovirus. Jamaican fruit bat (*Artibeus jamaicensis*) is thought to be a potential reservoir host of Tacaribe virus, an arena virus closely related to the South American hemorrhagic fever viruses. Shaw *et al*. (2012) performed transcriptome sequencing and annotation from lung, kidney and spleen tissues using 454 and Illumina platforms to develop this species as an animal model. Metagenomic analysis of viruses from bat fecal sample revealed many novel viruses in insectivorous bats in China (Ge *et al*., 2012). This study provided the first preliminary understanding of the viruses of bat population in China. In a viral metagenomic study of old world frugivorous bats (*Pteropus giganteus*) from Bangladesh by Epstein *et al*. (2010), resulted in the identification of GBV-D, a novel GB-like Flavivirus.

Apart from bats, Phan *et al*. (2013) study provided an initial assessment of the enteric virome in the dropping of pigeon, a feral urban species with frequent human contact. Metagenomic analysis of the Ferret fecal viral flora by Smits *et al*. (2013) discovered novel viruses from the families Picornaviridae, Papillomaviridae and Anelloviridae as well as known viruses from the families Astroviridae, Coronaviridae, Parvoviridae and Hepeviridae. Metagemomic analysis of the viral flora of Pinemarten and European Badger feces conducted by Van de brand *et al*. (2012), revealed a new Anellovirus and Bocavirus species in Pinemarten and a new Circovirus-like virus and Gemini virus-related DNA virus in European badger. The fecal viral flora of wild rodents was ascertained by viral metagenomic study by Phan *et al*. (2011). Rodents have been a constant source of zoonotic infections.

Study on Domestic Animals

Viral metagenomic studies have also been conducted on domestic animals which come in contact with wild animals and also on domestic animals having diseases of unknown etiologies. Belák *et al*. (2013) has published a paper on the new viruses in veterinary medicine, detected by metagenomic approach. Here the suspected viral etiological agents investigated were pigs with post weaning multisystemic wasting syndrome (PMWS), shaking mink syndrome; Honeybees with unspecified symptoms; Nursery and weaned pigs with diarrhea; Broilers with severe interstitial nephritis (Table 2).

Viral metagenomics demonstrated that domestic pigs were a potential reservoir for Nudumu virus (Masembe *et al*., 2012). Pigs can act as reservoir for emerging infectious disease. In the study, during a routine general surveillance for African swine fever domestic pigs in Uganda were screened for DNA and RNA viruses using NGS. Metagenomic analysis of turkey gut RNA virus community has been conducted by Day *et al*. (2010). They conducted this with regard to enteric disease syndrome of suspected viral origin.

Table 2: Viral Metagenomics of domestic animals

Syndromes/conditions	Identified viral genome sequence
Pigs with postweaning multisystemic wasting syndrome	Porcine circo virus type 2, Torque Teno virus, Porcine boca-like virus
Shaking mink syndrome	Astrovirus
Honeybees with unspecified symptoms	Aphid lethal paralysis virus, Israeli acute paralysis virus, Lake Sinai virus
Nursery and Weaned pigs with diarrhea	Porcine astrovirus, kobuvirus, calicivirus, Rotavirus A
Broilers with severe interstitial nephritis	Newcastle disease virus

(Data has been borrowed from Belak et al. (2013).

Next Generation Sequencing Platforms

Knowledge of the encrypted information in bio-molecules is indispensable for basic biological research as well as applied research including in the area of molecular diagnostics. DNA sequencing method is used to identify the infectious agent, study the evolutionary pattern of infectious agent and also comparative DNA sequence study of the healthy and sick cell genome, exome and transcriptome. DNA sequencing also provides the frame work for classifying and selecting the treatment protocol for infectious disease. The year of 2013 has seen U.S Food and Drug Administration (FDA) giving its first approval for next generation sequencing based diagnostic test. DNA sequence represents a single format onto which a broad range of biological phenomena can be projected for high throughput data collection (Shendure *et al*., 2008).

The encrypted codes of DNA started to yield with Frederick Sanger's chain termination sequencing method. Based on the dideoxy chain termination approach of Sanger *et al*. (1977), applied bio-systems marketed the first fully automated sequencing machine ABI PRISM 3700 in 1987 and later upgraded (ABI 3730). Since the inception of the first version of Sanger's method of sequencing several improvements have been made and one of these is the introduction of automated multi capillary-based instrument using flurophore labeling with multi spectral imaging-later referred to as 'first generation sequencing' platform. The first generation sequencing encountered bottleneck in throughput, scalability, speed and resolution which spurred the discovery of newer sequencing technologies. Human genome project have also played a decisive role in the improvement of sequencing technology. The newer technological advances in the cycle-array sequencing gave rise to what is known as 'second generation sequencing' or next generation sequencing. The technology involve repetitive cycle were the DNA sequence is immobilized on a solid support; is determined one base position at a time using enzymatic manipulation and image based data collection. As of today there are a number of NGS platforms based on different chemistry and protocol of template preparation that determine the type and character of the data produced. The efficiency and throughput of the sequencing technique is improving at a very fast rate and the cost of sequencing is reducing even greater than that of Moore's law. Pacific bioscience has commercialized a new generation of sequencing technology - the single molecule sequencing in real time and is called 'Third generation sequencing' technology. Third generation sequencers have read length many times higher than that of next generation sequencer but comes with the cost of higher error rate. The 3rd generation sequencers have failed to be in the limelight; market being the deciders of the best platform.

The merits of next generation sequencing over Sanger biochemistry of sequencing are several including high degree of parallelism, lowering the reagent volume used to picoliter or femtoliter. The advantage of next generation sequencing is often offset by several drawbacks. The most important of these include read length (read length being much smaller than the conventional sequencing) and raw accuracy

(base-calls generated by the new platform being at least ten fold less accurate than base-calls generated by Sanger sequencing). It is hoped that the next generation sequencing would improve in the days to come. Taking into account the bottlenecks of next generation sequencing, it could be considered at three point first, sample preparation second being next generation sequencing which has been improving steadily, and the third that is data collection and analysis. There is significant improvement in the second, but the first and the third has to be looked forth.

Next generation sequencing has shown great potential in the diagnosis of animal disease, detection of novel pathogens as demonstrated in several peer reviewed articles. Next generation sequencing reports all the nucleotides present in given sample. The application of next generation sequencing for routine use in diagnostics and reference laboratories to aid in the diagnosis of animal disease has been considered (Belak *et al*., 2013). The lower limit of detection is determined by the abundance of pathogen in relation to the background material. By deeper sequencing of the sample and the continued development in the next generation sequencing technology it would be able to detect low copy number pathogens. Sample preparation and enrichment protocol have dramatic effect on the next generation sequencing based diagnosis. The Illumina/Solexa and Roche 454 next generation sequencing platform are the most used platforms for viral metagenomics.

The Illumina/Solexa method is based on sequencing-by-synthesis chemistry using fragments of the sample DNA ligated to oligonucleotide adapters. The adapters on a solid support act as primers for DNA polymerase to incorporate reversible terminator nucleotides each labeled with a different fluorescent dye (Mokili *et al*., 2012). Novel bat viruses of the genera mamastro, boca, circo, ifla and orthohepadna in bats from Myanmar (He *et al*., 2013), three novel group 1 coronaviruses from three North American bats (Donaldson *et al*., 2010), are examples of viruses discovered using Illumina/ Solexa sequencing platform. The 454 FLX titanium pyro-sequencing commercialized by Roche has been the one most used for the discovery and characterization of novel viruses. Roche has been used in the metagenomic analysis of turkey gut RNA virus community (Day *et al*., 2010), identification of GBV-D; a Novel GB-like Flavivirus from old world frugivorous Bats in Bangladesh (Li *et al*., 2011) etc. comparison of platforms given in Table 1.

454 Pyrosequencing (Roche)

The pyro-sequencing technique was developed by Mostafa Ronagi and Pal Nyren at the Royal Institute of Technology in Stockholm in 1996 (Ronaghi *et al*., 1996). Pyro-sequencing version of next generation sequencing was licensed to 454 Life sciences which was later acquired by Roche Diagnostics. The method is based on emulsion PCR where DNA templates are amplified in water droplet in oil solution. In water droplets there is single DNA template attached to a single primer coated bead, that then form a colony clone inside the droplet. The beads are deposited in pico-liter volume wells and the sequencing machines contain many of these wells. The wells also contain the sequencing enzymes. Smaller beads bearing immobilized enzymes (ATP sulfurylase and luciferase) are also added in each pico-liter volume wells. Pyro-sequencing detects the released PPi (pyrophosphate) during the DNA synthesis (Shendure *et al*., 2008). At each of several hundred cycles, a single species of unlabeled nucleotide is introduced. On templates where this results in an incorporation event, pyrophosphate is released. Incorporation events immediately drive the generation of a burst of light, which is detected by charge coupled device. Currently there are two platform under this-GS FLX system and GS junior system. The GS FLX is associated with higher running cost and commonly used by larger facilities. The GS junior is a smaller bench top version with lower setup and running cost aimed at research laboratories. 454 pyro-sequencing is commonly used among NGS but it has higher error rate in homo-polymer region (i.e., three or more consecutive identical based) caused by accumulated light intensity variance and per base cost is higher than other next generations sequencing platforms.

AB SOLiD (Life Technologies)

This platform has it origin in the works of Shendure *et al*. (2005). Sequencing by oligonucleotide ligation and detection (SOLiD) also uses emulsion PCR to generate clonal DNA fragments on beads. One fragment of the library would be present in each of the magnetic bead which is clonally amplified. After breaking the emulsions, beads bearing amplification products are selectively recovered and then immobilized to a solid planar substrate to generate a dense disordered array. The starting sequence of the entire fragments attached to the bead would be a universal adapter sequence so that the starting sequence is both known and identical. Primers hybridize to the adapter sequence and a set of four fluorescently labeled di-basic probes compete for ligation to the sequencing primer. Sequencing by synthesis is driven by a DNA ligase instead of a DNA polymerase. During rounds of

ligation extension every base is scored at least twice with fluorescently labeled probes with data translated into color space (two base encoding). This platform has higher accuracy but the read length is short, so not normally used for applications requiring *denovo* assembly (transcriptomics, epigenomics, resequencing). Currently two platforms are available 5500 and 5500xl.

Illumina (Illumina Inc.)

Commonly referred to as 'the Solexa', this platform has its origin in the work by Turcatti and colleagues (Turcatti *et al*., 2008). Uses process known as bridge PCR to generate clonal colony of library DNA fragments on the surface of glass flow cell (single molecule array). In the solid phase bridge amplification of the single DNA template, one end of the DNA molecule attached to a solid support using an adapter; the molecule subsequently bends over and hybridizes to a complementary adapter (creating a bridge), thereby forming a template for the synthesis of complementary strand. The amplicons arising from a single template remain immobilized and clustered to a single physical location on an array. The resulting clusters each consist of ~1,000 clonal amplicons. Millions of such 'clusters' are located within each of the eight 'lanes' that are on a single flow-cell. Here the sequencing by synthesis uses proprietary reversible terminator technology that enables the detection of single bases as they are incorporated. Fluorescently labeled terminator is imaged as each de-oxyribonucleotide is incorporated and then cleaved to allow addition of the next nucleotide. The most recently introduced platform by Illumina is HiSeq 2500 either new or upgrade for the HiSeq 2000 and MiSeq. HiSeq design and capacity is most suitable for sequencing centers while MiSeq bench top equipped with smaller flow cell is aimed at wider range of laboratories and clinical diagnostic market.

Ion Torrent (life technologies)

The ion torrent platform uses silicon chips (transistor based sensors) to measure the hydrogen ion released during DNA polymerase activity. The lack of reliance on imaging makes this platform the first 'post-light' sequencing instrument (Rothberg *et al*., 2011). The most widely used technology for construction of integrated circuits, the complementary metal-oxide semiconductors (CMOS) process along with the electrochemical detection method – ion-sensitive field-effect transistors (ISFET) is used in this platform. The sensor and underlying electronics provide a direct transduction from the incorporation event to an electrical signal. Ion chips are manufactured on wafers, cut into individual die and robotically packed with a disposable polycarbonate flow cell that isolates fluids to a region above sensory array and away from the electronics. It is possible to increase the capacity of the platform by constructing chips of higher density of sensors and micro wells since the sequencing chips are produced in the same way as semiconductors. There are currently three different chips available with capacity from 10 Mb to 1 Gb. Ion torrent represent an affordable bench top system designed for small projects, such as sequencing microbial pathogens.

Single Molecule Sequencers

Heliscope is the first single molecule sequencer to hit the market. Here the templates are immobilized on glass flow cells. Fluorescent nucleotides are added one by one and a reversible terminating nucleotide is used to stop the process until the image has been captured by cameras. This platform has been hampered by high price and poor instrument sales and following the delisting form the stock market, it has plummeted down and there is a serious doubt on the future of this platform. Pacific bioscience introduced a third generation sequencing platform that is able to sequence single DNA molecule in real time without the need for amplification as well as onerous sample and library preparation. This is achieved by Zero Mode Wave guide (ZMW). Dye labelled nucleotides is continuously incorporated into a growing DNA strand by highly processive, strand-displacing φ29-derived DNA polymerase molecule. Each DNA polymerase has a ZMV detector, which continuously image labeled nucleotide as they enter the strand. The sequencing chip contains 1000's of ZMWs, so can handle large number of sequences simultaneously. The most important peculiarity is that read length is the highest among all the currently available NGS platforms. However, Pacific bioscience third generation sequencing also has the highest error rate of about 7 to 13 percentage.

Conclusion

Scientists all over the world have seen the power of next generation sequencing technology with peer reviewed articles that are published and the output and success from these research articles. More over new technologies and innovation are coming up in the research related to sequencing chemistry. As said earlier, the sequencing cost and time is plummeting down at a rate even greater than that of Moore's law. The sequencing commercial providers are on a race for the $ 1000 genome.

So, the only thing to look forward is to apply this next generation sequencing technology to the disease of veterinary and medical importance including disease surveillance and diagnostics. Next generation sequencing technology is used in amplicon sequencing, *de novo* sequencing, re-sequencing, RNA seq and meta-genomics. Viral meta-genomics is another area which has opened the black box of viruses. What is happening in the field of next generation sequencing is to be underscored. The work is stopped once a new virus is discovered by meta-genomic approach. The work has to be continued to the next step like trying to isolate the viral agent, study their role in the development of disease. This would advance science of disease and pathogens to its near comprehension.

References

Allander T, Emerson SU, Engle RE, Purcell RH and Bukh J (2001). A virus discovery method incorporating DNase treatment and its application to the identification of two bovine parvovirus species. *Proceedings of the National Academy of Sciences*, 98 (20): 11609-11614.

Baker KS, Leggett RM, Bexfield NH, Alston M, Daly G, Todd S and Murcia PR (2013). Metagenomic study of the viruses of African straw-coloured fruit bats: detection of a chiropteran poxvirus and isolation of a novel adenovirus. *Virology*, 441(2): 95-106.

Belák S, Karlsson OE, Blomström AL, Berg M and Granberg F (2013). New viruses in veterinary medicine, detected by metagenomic approaches. *Veterinary Microbiology*, 165(1): 95-101.

Benson DA, Karsch-Mizarchi I, Lipman DH, Ostell J and Sayers EW (2011). GenBank. *Nucleic Acids Research*, 38: D46-D51.

Brüssow H (2009). The not so universal tree of life or the place of viruses in the living world. *Philosophical Transactions of the Royal Society Biological Sciences*, 364(1527): 2263-2274.

Calisher CH, Childs JE, Field HE, Holmes KV and Schountz T (2006). Bats: important reservoir hosts of emerging viruses. *Clinical Microbiology Review*, 19(3): 531-545.

Dacheux L, Cervantes-Gonzalez M, Guigon G, Thiberge J M, Vandenbogaert M, Maufrais C and Bourhy H (2014). A Preliminary Study of Viral Metagenomics of French Bat Species in Contact with Humans: Identification of New Mammalian Viruses. *PLoS One*, 9(1): e87194.

Day JM, Ballard LL, Duke MV, Scheffler BE and Zsak L (2010). Metagenomic analysis of the turkey gut RNA virus community. *Virology Journal*, 7(313): 422X-7.

Donaldson EF, Haskew AN, Gates JE, Huynh J, Moore, CJ and Frieman MB (2010). Metagenomic analysis of the viromes of three North American bat species: viral diversity among different bat species that share a common habitat. *Journal of Virology*, 84(24): 13004-13018.

Epstein JH, Quan PL, Briese T, Street C, Jabado O, Conlan S and Lipkin WI (2010). Identification of GBV-D, a novel GB-like flavivirus from old world frugivorous bats (Pteropus giganteus) in Bangladesh. *Plos Pathogen*, 6(7): e1000972.

Ge X, Li Y, Yang X, Zhang H, Zhou P, Zhang Y and Shi Z (2012). Meta-genomic analysis of viruses from bat fecal samples reveals many novel viruses in insectivorous bats in China. *Journal of Virology*, 86(8): 4620-4630.

Grenfell BT, Pybus OG, Gog JR, Wood JL, Daly JM, Mumford JA and Holmes EC (2004). Unifying the epidemiological and evolutionary dynamics of pathogens. *Science*, 303(5656): 327-332.

Hall RJ, Wang J, Todd AK, Bissielo AB, Yen S, Strydom H and Peacey M (2014). Evaluation of rapid and simple techniques for the enrichment of viruses prior to metagenomic virus discovery. *Journal of Virological Methods*, 195:194-204.

He B, Li Z, Yang F, Zheng J, Feng Y, Guo H and Tu C (2013). Virome profiling of bats from Myanmar by metagenomic analysis of tissue samples reveals more novel mammalian viruses. *PloS One*, 8(4): e61950.

Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL and Daszak P (2008). Global trends in emerging infectious diseases. *Nature*, 451(7181): 990-993.

Ksiazek TG, Erdman D, Goldsmith CS, Zaki SR, Peret T, Emery S and Anderson LJ (2003). A novel coronavirus associated with severe acute respiratory syndrome. *New England Journal of Medicine*, 348(20): 1953-1966.

Leroy EM, Epelboin A, Mondonge V, Pourrut X, Gonzalez JP, Muyembe-Tamfum JJ and Formenty P (2009). Human Ebola outbreak resulting from direct exposure to fruit bats in Luebo, Democratic Republic of Congo, 2007. *Vector Borne Zoonotic Diseases*, 9(6): 723-728.

Leroy EM, Kumulungui B, Pourrut X, Rouquet P, Hassanin A, Yaba P and Swanepoel R (2005). Fruit bats as reservoirs of Ebola virus. *Nature*, 438(7068): 575-576.

Li L, Shan T, Wang C, Côté C, Kolman J, Onions D and Delwart E (2011). The fecal viral flora of California sea lions. *Journal of Virology*, 85(19): 9909-9917.

Luby SP, Hossain MJ, Gurley ES, Ahmed BN, Banu S, Khan SU and Rahman M (2009). Recurrent zoonotic transmission of Nipah virus into humans,

Bangladesh, 2001-007. *Emerging Infectious Diseases*, 15(8): 1229.

Masembe C, Michuki G, Onyango M, Rumberia C, Norling M, Bishop RP and Fischer A (2012). Viral metagenomics demonstrates that domestic pigs are a potential reservoir for Ndumu virus. *Virology Journal*, 9(1): 218.

Mokili JL, Rohwer F and Dutilh BE (2012). Metagenomics and future perspectives in virus discovery. *Current Opinion in Virology*, 2(1): 63-77.

Murray K, Selleck P, Hooper P, Hyatt A, Gould A, Gleeson L and Rodwell B (1995). A morbillivirus that caused fatal disease in horses and humans. *Science*, 268(5207): 94-97.

Ng TFF, Willner DL, Lim YW, Schmieder R., Chau B, Nilsson C and Breitbart M (2011). Broad surveys of DNA viral diversity obtained through viral metagenomics of mosquitoes. *PLoS One*, 6:e20579-e20579.

Phan TG, Kapusinszky B, Wang C, Rose RK, Lipton HL and Delwart EL (2011). The fecal viral flora of wild rodents. *PLoS Pathogen*, 7(9): e1002218.

Phan TG, Vo NP, Boros Á, Pankovics P, Reuter G, Li OT and Delwart E (2013). The viruses of wild pigeon droppings. *PloS One*, 8(9): e72787.

Quan PL, Firth C, Conte JM, Williams SH, Zambrana-Torrelio CM, Anthony SJ and Lipkin WI (2013). Bats are a major natural reservoir for hepaciviruses and pegiviruses. *Proceedings of the National Academy of Sciences*, 110(20): 8194-8199.

Reyes GR and Kim JP (1991). Sequence-independent, single-primer amplification (SISPA) of complex DNA populations. *Molecular and cellular probes*, 5(6): 473-481.

Riesenfeld CS, Schloss PD and Handelsman J (2004). Metagenomics: genomic analysis of microbial communities. *Annual Review of Genetics*, 38:525-552.

Rohwer F and Edwards R (2002). The Phage Proteomic Tree: a genome-based taxonomy for phage. *Journal of Bacteriology*, 184(16): 4529-4535.

Ronaghi M, Karamohamed S, Pettersson B, Uhlén M and Nyrén P (1996). Real-time DNA sequencing using detection of pyrophosphate release. *Analytical Biochemistry*, 242(1): 84-89.

Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M and Bustillo J (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature*, 475(7356): 348-352.

Sanger F, Nicklen S and Coulson AR (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12): 5463-5467.

Shaw TI, Srivastava A, Chou WC, Liu L, Hawkinson A, Glenn TC and Schountz T (2012). Transcriptome sequencing and annotation for the Jamaican fruit bat (Artibeus jamaicensis). *PloS One*, 7(11): e48472.

Shendure J and Ji H (2008). Next-generation DNA sequencing. *Nature Biotechnology*, 26(10): 1135-1145.

Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon, JP, Rosenbaum AM, Church GM (2005). Accurate multiplex polony sequencing of an evolved bacterial genome. *Science*, 309(5741): 1728-1732.

Smits SL, Raj VS, Oduber MD, Schapendonk CM, Bodewes R, Provacia L and Haagmans BL (2013). Metagenomic analysis of the ferret fecal viral flora. *PloS One*, 8(8): e71595.

Thurber RV, Haynes M, Breitbart M, Wegley L and Rohwer F (2009). Laboratory procedures to generate viral metagenomes. *Nature Protocols*, 4(4): 470-483.

Tong S, Zhu X, Li Y, Shi M, Zhang J, Bourgeois M and Donis RO (2013). New world bats harbor diverse influenza A viruses. *PLoS Pathogen*, 9(10): e1003657.

Towner JS, Amman BR, Sealy TK, Carroll SAR, Comer JA, Kemp A and Rollin PE (2009). Isolation of genetically diverse Marburg viruses from Egyptian fruit bats. *PLoS Pathogen*, 5(7):e1000536.

Towner JS, Pourrut X, Albariño CG, Nkogue CN, Bird B H, Grard G and Leroy EM (2007). Marburg virus infection detected in a common African bat. *PLoS One*, 2(8): e764.

Turcatti G, Romieu A, Fedurco M and Tairi AP (2008). A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. *Nucleic Acids Research*, 36(4): e25.

Van den Brand JM, Van Leeuwen M, Schapendonk C M, Simon JH, Haagmans BL, Osterhaus AD and Smits SL (2012). Metagenomic analysis of the viral flora of pine marten and European badger feces. *Journal of Virology*, 86(4): 2360-2365.

Van Etten JL, Lane LC and Dunigan DD (2010). DNA viruses: the really big ones (giruses). *Annual Review of Microbiology*, 64: 83.

Vieites JM, Guazzaroni ME, Beloqui A, Golyshin PN and Ferrer M (2009). Metagenomics approaches in systems microbiology. *FEMS Microbiology Reviews*, 33(1): 236-255.

Wang LF, Shi Z, Zhang S, Field H, Daszak P and Eaton BT (2006). Review of bats and SARS. *Emerging Infectious Diseases*, 12(12): 1834.

Warrell MJ and Warrell DA (2004). Rabies and other lyssavirus diseases. *The Lancet*, 363(9413): 959-969.

Wood JL, Leach M, Waldman L, MacGregor H, Fooks A R, Jones KE and Cunningham AA (2012). A framework for the study of zoonotic disease emergence and its drivers: spillover of bat pathogens as a case study. *Philosophical Transactions of the Royal Society: Biological Sciences*, 367(1604): 2881-2892.

Woolhouse ME, Haydon DT and Antia R (2005). Emerging pathogens: the epidemiology and evolution of species jumps. *Trends in Ecology and Evolution*, 20(5): 238-244.