



Bioinformatics and Microarray-Based Technologies to Viral Genome Sequence Analysis

6

Mayank Pokhriyal, Barkha Ratta, and Brijesh S. Yadav

Abstract

Identification of microbial pathogen is an important event which lead to diagnosis, treatment, and control of infections produce by them. The high-throughput technology like microarray and new-generation sequencing machine are able to generate huge amount of nucleotide sequences of viral and bacterial genome of both known and unknown pathogens. Few years ago it was the DNA microarrays which had great potential to screen all the known pathogens and yet to be identified pathogen simultaneously. But after the development of a new generation sequencing, technologies and advance computational approach researchers are looking forward for a complete understanding of microbes and host interactions. The powerful sequencing platform is rapidly transforming the landscape of microbial identification and characterization. As bioinformatics analysis tools and databases are easily available to researchers, the enormous amount of data generated can be meaningfully handled for better understanding of the microbial world. Here in this chapter, we present commentary on how the computational method incorporated with sequencing technique made easy for microbial detection and characterization.

M. Pokhriyal

Viral Research and Diagnostic Laboratory (VRDL), Government Medical College, Haldwani, Uttarakhand, India

B. Ratta

Division of Biochemistry, Indian Veterinary Research Institute, Bareilly, Uttar Pradesh, India

B. S. Yadav (✉)

Center of Bioengineering Research, University of Information Science and Technology (UIST), St. Paul, Republic of North Macedonia

e-mail: brijeshbioinfo@gmail.com

6.1 Introduction

6.1.1 Importance of Microbes

Microorganisms like viruses, bacteria, and fungus have evolved to survive in every type of conditions on our planet including human and animal bodies. Although many are not harmful, a few cause life-threatening diseases. Traditionally these are identified by culturing in appropriate media, biochemical analysis, and serological testing. However, more numbers of microbes are yet to be characterized than are known. Through the human body, microbes play vital roles in all of their ecosystems (Hentges 1993). In spite of being extremely small, the sheer numbers of microbes living on the planet have large effects on the cycling of nutrients and compounds, essential for the survival of all organisms.

Microbes are encountered in all walks of human life and there prevails a constant interaction. The vast majority of the bacteria in the body are rendered harmless by the protective effects of the immune system, and a few are beneficial. In fact, the relationship between microbes and humans is delicate and complex. Ten times as many microbes live on or inside your body. The microbes living in our digestive system break down food and produce useful vitamins. The millions of microbes that coat our skin and intestinal lumen form a protective barrier against more dangerous microbes. Without them, our bodies would be open to microbial attack. In spite of these benefits, a relatively small number of microbes are harmful to humans. Many diseases and epidemics are caused by microbes: the plague during the Middle Ages, smallpox, AIDS, influenza, food poisoning, and anthrax. These diseases result in severe illness, or even death. As scientists learn more about bacteria, fungi, and viruses, they are better able to treat and prevent these diseases (Yadav et al. 2013).

6.1.2 Viral Diseases in Humans and Livestock

Viral diseases of livestock can be devastating both to farmers and the wider community. Rinderpest, a disease of cattle caused by rinderpest virus, spread rapidly across Africa by 1892 and led to death of nearly 95% of the cattle in East Africa. In the early years of the twentieth century, rinderpest was common in Asia and parts of Europe, and its prevalence increased in Asia. In 1957, Thailand had to appeal for aid because many buffaloes had died, due to which paddy fields could not be prepared for rice. In India, spread of rinderpest was controlled by numerous eradication programs throughout the twentieth century (Barrett et al. 2006). As a result rinderpest was eradicated from India in the year 1995 (Barrett et al. 2006). In human and animals, viral diseases keep emerging very frequently.

Research on various aspects of animal diseases particularly diagnosis is carried out regularly to develop effective measures to guarantee animal health through containment of emerging diseases. Many viral diseases have been eradicated from globe by taking effective control measures. Smallpox was the first viral disease to be eradicated globally followed by Rinderpest or cattle plague. In the last 30 years or so,

about 40 new viral diseases have been identified (Zappa et al. 2009). A recent study has shown that by 2020, about 10 new human diseases would emerge compared to the past, and mobility of humans and animals have increased; with this, rate of spread of disease has also increased (Morens and Fauci 2013). It has become vital to have an arsenal of diagnostic techniques which can identify a pathogen whether new or old in shortest possible time to combat effective control measures. New technologies are being developed in molecular biology at a very rapid rate and many of them are being applied in diagnosis. PCR (polymerase chain reaction) is once such technique which was developed for amplification of a fragment of DNA in the early 1980s (Mullis 1990). The PCR and related techniques like real-time PCR are now the most widely used molecular method for the detection and identification of viruses (Espy et al. 2006). Multiplex PCR allows detection of more than one species of virus in a single assay (Elnifro et al. 2000). However, these techniques have limitations of multiplexibility and versatility and require extensive prior knowledge of the sequences to be amplified (Miller et al. 2009).

Another technique, microarray, developed in mid-1990s (Kostrzynska and Bachand 2006) for studying gene expression, is now being increasingly applied for diagnostics. Microarray consists of thousands to millions of oligonucleotide probes representing different genes of known or unknown pathogens deposited or directly synthesized on a surface in an ordered fashion (Anderson et al. 2008; Manoj 2009). These numbers give huge capability to microarray to detect and quantitate hundreds and thousands of genes simultaneously. The capability of microarray to detect a large number of genes was used for diagnosis at the beginning of this century when the SARS corona virus was identified using a diagnostic microarray chip (Wang et al. 2002). Since then several microarray chips have been developed and tested (Wang et al. 2002, 2003; Martín et al. 2006; Chiu et al. 2006; Quan et al. 2007; Gardner et al. 2010; Yadav et al. 2015).

6.2 Virus Identification Methods

A virus is identified by showing presence of its proteins or nucleic acids or via an immunological response to the host. Each method has its own advantages and drawbacks with respect to sensitivity, specificity, efficiency, and feasibility. Depending on the virus type, concentration, and circumstances under which the viral sample was collected, certain methods may be more effective than others. The most commonly used methods to detect and identify different viruses can be broadly divided into four categories, that is:

- A. Electron microscopy: A direct method for detecting virus requires purified and high concentration of virus, but only experienced technician then discerns the virus by its physical structure features. Viruses such as poxviruses and herpes viruses can be easily identified using this technology (Nii 1971).
- B. Virus isolation in cell culture: Another method is to grow the virus in cell cultures (Leland and Christine 2007) and observe for virus-induced changes such as cell

rounding, disorientation, swelling, shrinking, or death in the cells. This method is time-consuming, has low sensitivity, and also cannot be used for viruses like hepatitis B, parvovirus, and papillomavirus, which cannot grow in cell culture (Goldsmith and Miller 2009).

- C. Immunofluorescence-based assays: ELISA (enzyme-linked immunosorbent assay) (Voller et al. 1978), one of the widely used methods for detecting viruses, relies on the presence of antigens or antibodies in bodily fluids.
- D. Molecular techniques: Molecular biology techniques based on nucleic acid sequences are advanced and much faster than any other techniques and are now mostly used for virus diagnosis (Kreuze et al. 2009).

6.2.1 Nucleic Acid Sequence-Based Diagnosis

Large-scale availability of genomic and nucleotide sequences of disease-causing agents in public databases (GenBank of the USA, European molecular biology laboratory of Europe, and DNA databank of Japan) and progress in nucleic acid amplification techniques have enabled application of nucleic acid-based techniques for pathogen identification. These techniques include polymerase chain reaction (PCR) (Mullis 1990), loop-mediated isothermal amplification (LAMP) (Tomita et al. 2008), ligase chain reaction, nucleic acid sequence-based amplification/isothermal amplification (NASBA, (Liu et al. 2006), strand displacement amplification (Su et al. 2013), Qb replicase method (An et al. 2003), and branched DNA probes (Baumeister et al. 2012). The PCR is one of the most commonly used methods for virus detection (Cunningham 2004). Real-time PCR, an innovation of PCR, is one of the most sensitive methods of detecting pathogens (Mackay et al. 2002). Both PCR and real-time PCR have limited multiplexing capability.

Like PCR and real-time PCR, DNA microarray is a genome-based method which was originally developed for studying variation in gene expression but has since been adapted for pathogen detection. It has huge multiplex capability and can screen for all the known pathogen in one experiment (Wilson et al. 2002; Bryant et al. 2004). The DNA microarray also called DNA chip has high sensitivity and specificity (Kostrzynska and Bachand 2006). The microarray chip used for diagnostic purpose contains thousands of different oligonucleotide probes specific to respective pathogens. The probes designed for diagnostic assays are unique to a specific pathogen with respect to all the other pathogen genomes and also to host other nonspecific genome sequences present in the clinical samples. Specially designed software are used for designing probes and also for data analysis.

6.2.2 DNA Microarray-Based Virus Identification

DNA microarray has been used efficiently in clinical diagnostics for identifying disease-related genes with the help of its biomarkers (Loy and Bodrossy 2006) and also for disease diagnosis. The microarray-based virus diagnosis started at the

beginning of this century (Skena et al. 1995). Microarrays based on oligonucleotide probes representing nucleic acid sequences conserved between members of a taxonomic group were first used for detection of the then unknown SARS coronavirus (Wang et al. 2002, 2003) and since then has been used for detection of many viruses (Chiu et al. 2007; Chou et al. 2006; Quan et al. 2007; Martín et al. 2006, Chen et al. 2010). The detection of multiple rhinovirus serotypes in cell culture and clinical specimen (Wang et al. 2002), papillomavirus in cervical lesions (An et al. 2003), parainfluenza virus 4 in nasopharyngeal aspirates (Chiu et al. 2006), influenza virus from nasal wash and throat swabs (Lin et al. 2006), gammaretrovirus in prostate tumors (Urisman et al. 2006), foot and mouth disease virus from animal tissue (Martín et al. 2006), coronaviruses and rhinoviruses from nasal lavage (Kistler et al. 2007), metapneumovirus from bronchoalveolar lavage (Chiu et al. 2007), different respiratory pathogens including influenza virus and non-influenza agents in nasal swabs and lung tissue (Quan et al. 2007), and common food born viruses such as coxsackievirus, hepatitis A virus, norovirus, and rotavirus identified using tiling microarray (Chen et al. 2010). Grubaugh et al. (2013) identified 13 of 14 flaviviruses (*Culex* flavivirus, dengue-3, and Japanese encephalitis viruses) using microarray platform.

6.2.3 Animal Virus Diagnosis Using Microarray

Microarray technique is one of the most recent diagnostics in veterinary field (Feilletter 2004). Jack et al. (2009) developed a microarray assay for identifying viruses that cause vesicular or vesicular-like lesions in livestock animals. They were able to differentiate foot and mouth disease virus (FMDV), vesicular stomatitis virus (VSV), swine vesicular disease virus, vesicular exanthema of swine virus (VESV), BHV-1, orf virus, pseudocowpox virus, bluetongue virus serotype 1, and bovine viral diarrhea virus 1 (BVDV1). Leblanc and co-researcher (2009) used magnetic bead microarray for the rapid detection and identification of the four recognized species in the *Pestivirus* genus of the *Flaviviridae* family, i.e., classical swine fever virus, border disease virus, and BVDV1 and 2, which allowed specific and sensitive virus detection. They concluded that based on the simplicity of the assay, the protocols for hybridization and magnetic bead detection offer an emerging application for molecular diagnosis in virology that is amenable for use in a modestly equipped laboratory. Porcine reproductive and respiratory syndrome virus (PRRSV) and foot and mouth disease virus (FMDV) were detected in a cDNA microarray (Liu et al. 2006). GreenChip array facilitated the discovery of Ebola virus, in the porcine respiratory illness outbreak in the Philippines (Barrette et al. 2009). These chips have also been used for screening veterinary clinical samples (Mihindukulasuriya et al. 2008). Canine coronavirus (CCoV), feline infectious peritonitis virus (FIPV), feline coronavirus (FCoV), bovine coronavirus (BCoV), porcine respiratory coronavirus (PRCoV), turkey enteritis coronavirus (TCoV), transmissible gastroenteritis virus (TGEV), and human respiratory coronavirus (HRCoV) are identified based on microarray hybridization (Chen et al. 2010). Sharma et al. (2012) have designed

microarray chip for identification of animal viruses; the chip successfully identified the new castle disease virus in sheep and mixed infection of bovine viral diarrhoea 2 and bovine herpes virus 1 (Ratta et al. 2013) in cattle.

6.3 Basic Concept of DNA Microarray

Since its inception DNA microarray has advanced biological sciences more profoundly than any other technique (Benedetti et al. 2000; Rockett and Dix 2000; Staudt and Brown 2000; Brown and Botstein 1999). The basic principle involved in DNA microarrays is the reverse of Southern blot (Southern 1975). Unlike Southern blot where targets are immobilized and a probe is labeled, in DNA microarray, probes are immobilized on a membrane and then hybridized against the labeled target population (Kurian et al. 1999). Since unlike Southern blot, in microarray, target is labeled not probes, target can be hybridized to a large number of probes and consequently microarray chips have the capacity to simultaneously detect tens or hundreds or thousands of specific nucleic acid targets present in biological samples in a single experiment (Schna et al. 1995). Microarrays have been used in a large number of applications such as genome-wide genotyping, expression profiling, RNA detection, protein arrays, and pathogen nucleic acid detections (Petricoin et al. 2002).

6.3.1 Probe Designing for Pathogen Identification

Probes, which are short DNA sequences that are similar to parts of the sequence of target, are the most important constituent of a microarray chip. Probe has to be specific to a pathogen only, and then only it can identify a pathogen unambiguously. Besides specificity probe designing requires consideration of many other factors, which influence hybridization processes such as guanine/cytosine (G/C) content, melting point, secondary structure, sequence specificity, polynucleotide tract, and probe length. Probe length affects the sensitivity and specificity of hybridization, while other factors contribute to nonspecific hybridization. Maximizing the specificity and the sensitivity are often conflicting goals in terms of achieving probe design. Some well-known examples for microarray chips are as follows.

The probes for rotavirus were selected by using the following criteria: lengths of about 20 nucleotides, melting temperatures between 65°C and 75°C, and two or more mismatches with homologous sequences in other genotypes (Chizhikov et al. 2002). For orthopoxviruses, C23L/B29R gene sequences of different orthopoxvirus species were aligned using clustalx software to find variable regions suitable to design species-specific oligoprobes. Criteria for oligoprobe design were as follows: one or more mismatches with other orthopoxvirus species, length of 13–21 nucleotides, and predicted melting temperatures between 36°C and 58°C (Majid et al. 2003).

One of the very first microarray chips to utilize a generalized computational algorithm to find conserved regions among viral pathogens was by Wang et al. (2002). The oligonucleotide probe length used in this chip was 70-mer. The chip based on this conserved sequence was used to identify a coronavirus from the then unknown SARS-coronavirus samples. Chou et al. (2006) developed a comprehensive algorithm for designing conserved probes. They assumed that a virus genus (G) is a collection of n viruses, in which each virus v_i ($i = 1, \dots, n$) is associated with a subset of G . Comparison of this virus with another virus in the genus identifies similar sequences. The similarity was defined as having either (i) more than 75% local sequence similarity in a 50-bp window with any virus or (ii) >15 consecutive bases pairing. From these set of conserved sequences, those sets are picked up which would identify the entire virus in the genus alone or in combination. It was not necessary that a single conserved sequence would pick up the entire virus in the genus. From the conserved sequences, 70-mer conserved probe were selected based on the following criteria: (i) GC content between 40% and 60%, (ii) <5 continuous mononucleotide repeats, (iii) <25-bp BLASTN sequence identity matches, and (iv) ≤ 15 consecutive bases pairing with other viral sequences in the noncognate viral genus (Chou et al. 2006).

Jabado et al. (2008) used protein families database (Pfam) as the basis for designing conserved oligonucleotide probes. Their strategy for finding conserved region entails identifying short conserved region and corresponding nucleic acid region. First they identified most conserved nonoverlapping 20 amino acids sequences then extracted corresponding nucleotide sequences. If the gaps were found, the flanking regions were taken into account. The sequences which were not part of Pfam were extracted and homologous gene clustered. All clustered sequence was searched for common motifs with software such as MEME. Three motifs were selected for each sequence cluster. The nucleic acid sequence extracted for each protein motif was used for probe design. The conditions for probe design were ≤ 5 mismatches to the template, a $T_m > 60^\circ\text{C}$, no repeats exceeding a length of 10 nt, no hairpins with stem lengths exceeding 11 nt, and <33% overall sequence identity to non-viral genomes (Jabado et al. 2008).

Pan-Microbial Detection Array (MDA) is the most comprehensive array designed for virus diagnosis. This array includes probes for all the virus sequences present in the database at the time of designing. The designing strategy used for this array was family based. First, the sequence of all the viruses reported for virus families was grouped, and from the group, all the sequences similar to human and nonfamily virus sequences were removed. From the resultant viral sequences, probes were designed irrespective of the location or the gene using primer 3. The primers were selected on the basis of conservation within family and the number of probes per target sequence which was 50 probes per target sequence (Gardner et al. 2010).

6.3.2 Clinical Sample Processing and Hybridization

The clinical samples for identification of a virus are collected from different places, from different animals, and from different sources. The clinical samples could be blood, a swab from the nose, vagina, or mouth, tissue and stool, or any other. Most of these samples have been tested in microarrays. Foot and mouth disease virus (FMDV) was identified in ticks collected from a livestock market in Nairobi, Kenya (Sang et al. 2006). Vincent (2009) identified clinical porcine respiratory and reproductive syndrome virus in a nasal swab from pig samples. The clinical samples for microarray analysis are collected in solutions (like RNAlater) which makes RNA stable or directly in trizol for RNA isolation (Wang et al. 2002; Chiu et al. 2007). Samples for microarrays are generally processed by the methods adopted by Wang et al. (2002). This method employs anchored random nonamer primers for cDNA synthesis, nonspecific amplification, and introduction of amino-allyl nucleotide into the amplification product. Labeling is done by covalent binding of fluorescent dye (cyanine with aminoallyl group of nucleotide (Jabado et al. 2008; Wang et al. 2002; Gardner et al. 2010). Hybridization is done overnight, and the temperature of hybridization is defined by the probe sequences which vary between 65°C and 70°C (Tan et al. 2003; Wang et al. 2003; Jayaraman et al. 2006; Gardner et al. 2010). Hybridization signals are generated by using light of a predefined wavelength to stimulate the emission of the fluorescent signal. The amount of emission is determined by the amount of fluorescent dyes bound, which is correlated with the amount of targets-probe hybrids at the spot.

6.3.3 Data Processing on the Basis of Signal Intensity

Virus prediction from the signal intensity data has been the subject of intense study. Unlike gene expression microarrays, diagnostic microarrays have no up or down-regulated gene. In diagnostic microarray, a signal cannot be low, high, or unchanged; it has to be defined in binary numbers—present or absent, on or off. A signal is defined as present if it is above a predefined cutoff. This cutoff signal has arrived differently as the microarray technology for diagnosis developed. In the first broad-spectrum microarray chip, the control hybridization was carried out with the RNA from uninfected cell culture. In this particular chip, each spot was spiked with a known probe in a fixed ratio to normalize the expression of all the probes. The complementary sequence of the spike probe was labeled with Cy3 dye. Two color hybridization experiments were done for both infected and uninfected cell culture. The Cy5 signal for each probe was normalized against Cy3 signal. Cy5 labeled-infected and uninfected signal for each probe were calculated; from these values, a value of 1500 was arbitrarily defined as cutoff value, and the prediction was based on aggregate hybridization (Wang et al. 2003). In diagnostic arrays, it is not always possible to have controls for identifying cutoff value like Wang et al. (2003) where they used mock-infected cell culture as control. To overcome this control problem, random probes which have no sequence similarity with NR database have been

introduced (Gardner et al. 2010). If control samples are available, they are used for calculating the cutoff value; otherwise signal intensity of random probes is used for calculating cutoff value. The cutoff value is kept at median + 2 SD of the random probe signal intensity, 95–99% percentile of random signal intensity (Gardner et al. 2010). The virus identification from the signal intensity data makes use of different approaches. The simplest approach for predicting presence of virus in the sample is based on averaging signal intensity of all the probes of a virus or virus genus/family: if it is above a cutoff value, virus is predicted to be present; the other approach is based on the number/percentage of probes giving positive signal above a preset threshold (Sharma et al. 2012; Yadav et al. 2015). The first broad-spectrum virus-detecting microarray chip, ViroChip, made prediction based on the aggregate hybridization pattern (Wang et al. 2003). Chou et al. (2006) adopted similar criteria but with some modifications. Their method makes use of both signal intensity and a number of probes for making prediction. First, sum of all the signal intensity of a probe set is calculated, and then this is divided by the maximum intensity obtained for any group in the probe set. Based on the percentage, prediction is made. For the GreenChip (Palacios et al. 2007), virus prediction was done by a specially developed software called Green-LAMP. This software subtracts background values from the probe intensities, calculate Z-score from log intensities by dividing with standard deviation and compute tail probabilities (p-value) assuming log normality. Presence of a positive or negative signal is computed from a fixed P-value threshold, 0.1 for arrays with matched controls and 0.023 otherwise. The background levels are derived from matched control samples or from random 60-mer control probes if matched control samples are not present. The following assumptions were made for making predictions: (1) spot intensities are normally distributed, (2) spots represent independent observations, and (3) there are relatively few (<100) positive probes for any given virus. DetectiV (Watson et al. 2007) tool was developed for handling array data by selecting groups of probes comprising a species, genus, or family and computing a one-sample t-test with null hypothesis that log intensity ratio for each group is zero. PhyloDetect (Rehrauer et al. 2008), another tool for making virus prediction, converts the probe intensities to binary indicators (e.g., by thresholding against the median + 2 SD of the background intensities). In Pan-Microbial Detection Array (MDA) (Gardner et al. 2010), the threshold for positive signal was kept at 99th percentile of negative controls which were randomly generated probes. Prediction for presence or absence of a pathogen was based upon two conditional probabilities: the probability of observing a signal in the presence of a specific microbial target in sample and the probability of observing signal in the absence of microbial target in the sample. Liu et al. (2006) compared five different methods of virus prediction and concluded that hypergeometric distribution and log transform ranking method give good prediction, but other methods like the number of probes with a threshold or ratio method also give suitable prediction.

6.4 Recent Advances for Identification and Characterization of Microbes

Over the last two decades, sequence analysis of conserved genes has become a reliable, accurate, inexpensive, and scalable method of microbial identification in health and environmental sciences. These advantages have resulted in the routine use of sequencing methods to complement, and sometimes replace, traditional phenotypic methods of identification.

Various molecular techniques have emerged in the recent decades offering speed combined with specific and sensitive detection. They are simple, rapid and reliable, and dependent on the presence of nucleic acids (DNA and RNA) that code for the proteins. These methods include polymerase chain reaction (PCR), microarrays, metagenomics, next-generation sequencing, and many others. Detection of DNA is now possible on a single molecule, and high-throughput analysis allows thousands of detection reactions to be performed at once, thus allowing a range of characteristics to be rapidly and simultaneously determined. Some of the recent molecular detection methods can be performed in the laboratory or clinical settings and also at the farm site. Although some of these techniques provide immediate result, many require extensive computational approaches for analysis and interpretation of the data.

Metagenomics is recently introduced where we study the genomic content of an environmental sample of microbes. It is a derivation of conventional microbial genomics, with the key difference being that it bypasses the requirement for obtaining pure cultures for sequencing. Metagenomics holds the promise of revealing the genomes of the majority of microorganisms that cannot be readily obtained in pure form. Since the samples are obtained from communities rather than isolated populations, metagenomics may serve to establish hypotheses concerning interactions between community members. This process begins with sample and metadata collection and proceeds with DNA extraction, library construction, sequencing, read preprocessing, and assembly. Community composition analysis is employed at several stages of this workflow, and databases and computational tools are used to facilitate the analysis. Advances in the throughput and cost-efficiency of sequencing technology are fueling a rapid increase in the number and size of metagenomics datasets being generated. However, bioinformaticists are faced with the problem of how to handle and analyze these datasets in an efficient and useful way (Tringe and Rubin 2005). The goal of metagenomics studies is to get a basic understanding of the microbial world both surrounding us and within us.

Information from metagenomics studies will be fully exploited only if appropriate data-management and data-analysis methods are in place. One was that the data were immediately accessible in a form suitable for computer analysis; another was that the data were freely available, without impediment to all researchers, be they in academia or industry. The three nucleic acid sequence archives GenBank, EMBL-Bank, and DDBJ have spearheaded the cause of free availability of sequence information. In the process sequences of a large number of fragments have been registered in the international DNA databanks. However, the details of function of the sequence are

not available and are of limited use. Analysis and comparison of complex metagenomic data is driving the development of a new class of bioinformatics and visualization software. The field is moving forward rapidly, driven by enormous improvements in sequencing technology and the availability of many complementary technologies. Analysis and clustering of metagenomic sequences with the help of bioinformatics tools according to phenotypes and genomes might in future help in environmental preservation (Kunin et al. 2008).

6.4.1 New-Generation Sequencing

Sequencing is one technique that transformed biology from qualitative to a quantitative science and leads to the emergence of bioinformatics as an important discipline. Initially, sequencing started with radioisotope-labeled sequencing products analyzed on slab gels. This slow process was overtaken by fluorescent labeling and capillary electrophoresis that improved speed and data quality of sequencing. Recently the next-generation sequencing platforms have made possible massive parallel sequencing without the need for lengthy electrophoresis. There are various approaches for next-generation sequencing like sequencing by hybridization, microelectrophoresis, cyclic array sequencing and real-time observation of single molecules. These diverse approaches and sophistication of next-generation sequencing have brought great challenges for bioinformaticists to tackle alignment, sequence scoring, data assembly, storage, and release of huge amounts of data (Kunin et al. 2008).

The ability to simultaneously acquire huge amount of sequence data when applied to clinical and environmental samples helps in the identification of pathogenic microbes. Moreover, genome variability and evolution within the host can be tracked over short periods of time. These approaches were already being used in diagnostic virology for detection of novel pathogenic viruses and for mapping of resistance to antiviral drugs (Barzon et al. 2011).

6.5 Conclusion and Future Perspective

DNA microarray is a rapid method for virus identification. Proof of concept has already been shown in at least two cases where DNA microarray identified pathogen when all the other method for diagnosis failed. Currently the biggest problems in designing probes for diagnostic arrays are in designing conserved probes at genus level. In viral genera, getting conserved probes which would identify all the virus in the genus is next to impossible for almost all genera; as a consequence genus has to be arbitrarily subdivided based upon sequence conservation within the genus. This strategy has not been adopted though in MDA array the probes have been selected based upon conservation but not on uniqueness within a subgroup. Another problem with using microarray is the sequence heterogeneity within the virus species. All the unique probes of a virus species would not bind to all the isolates so a threshold has to be set for making prediction. The capacity of chips is going up, the

new-generation chips can incorporate a million probes, while the number of viruses reported in ICTV is just above 2000; however the number of virus sequence reported and stored in NCBI database is in hundreds of thousands. Thus it is possible to make probes for all the sequences and incorporate them in a chip, but doing that would create problems in interpretation of results because of cross-hybridization signals. The one way to avoid cross-hybridization signal is to reduce the size of probes which is currently set at about 70-mer. Microarray is costly as camper to PCR, and so it is generally restricted to the few commercial laboratories that can possess the capital, or those laboratories developing expertise in this field. This is a long procedure for the numbers of genes involved in this technique. ViroChip pan-viral microarray recently used deep sequencing technology to 17 respiratory samples collected from individuals infected with the 2009 H1N1 influenza virus early during the pandemic and deep sequencing which can test for thousands of potential pathogens simultaneously. Consequently diagnostic strategy of rapid ViroChip-based testing followed by deep sequencing could show to be a useful public health response to infectious disease outbreaks in the future (Yadav et al. 2014). Thus, identifying viral species using the previously reported viral microarray probe design strategy with new approaches is very impressive. The use of microarray in pathogen identification is still an intensive area of research. The new design strategies are constantly coming up. It is hoped that in near future a very precise and cost-effective chip would be developed, but to increase its practical usage in clinical microbiology laboratories, it has to become more affordable, be convenient to handle, and be accurate.

References

- An HJ, Cho NH, Lee SY, Kim IH, Lee C, Kim SJ, Mun MS, Kim SH, Jeong JK (2003) Correlation of cervical carcinoma and precancerous lesions with human papillomavirus (HPV) genotypes detected with the HPV DNA chip microarray method. *Cancer* 97:1672–1680
- Anderson BT, Ruane AC, Roads JO, Kanamitsu M, Salvucci G (2008) A new metric for estimating local moisture cycling and its influence upon seasonal precipitation rates. *J Hydrometeorol* 9:576–588
- Barrett T, Pastoret P, Taylor W (2006) Rinderpest and peste des petits ruminants: virus plagues of large and small ruminants. Elsevier Academic Press, Amsterdam. ISBN 0-12-088385-6
- Barrette RW, Metwally SA, Rowland JM, Xu L, Zaki SR, Nichol ST (2009) Discovery of swine as a host for the reston ebolavirus. *Science* 325:204–206
- Barzon L, Lavezzo E, Militello V, Toppo S, Palù G (2011) Applications of next-generation sequencing technologies to diagnostic virology. *Int J Mol Sci* 12(11):7861–7884
- Baumeister MA, Zhang N, Beas H, Brooks JR, Canchola JA, Cosenza C, Kleshik F, Rampersad V, Surtihadi J, Battersby TR (2012) A sensitive branched DNA HIV-1 signal amplification viral load assay with single day turnaround. *PLoS One* 7(3):e33295
- Benedetti VMD, Biglia N, Sismondi P, De Bortoli M (2000) DNA chips: the future of biomarkers. *Int J Biol Markers* 15(1):1–9
- Brown PO, Botstein D (1999) Exploring the new world of the genome with DNA microarrays. *Nat Genet* 21(1):33–37
- Bryant PA, Venter D, Robins-Browne R, Curtis N (2004) Chips with everything: DNA microarrays in infectious diseases. *Lancet Infect Dis* 4(2):100–111

- Chen Q, Li J, Deng Z, Xiong W, Wang Q, Hu YQ (2010) Comprehensive detection and identification of seven animal coronaviruses and human respiratory coronavirus 229 E with a microarray hybridization assay. *Intervirology* 53(2):95–104
- Chiu CY, Rouskin S, Koshy A, Urisman A, Fischer K, Yagi S, Schnurr D, Eckburg PB, Tompkins LS, Blackburn BG, Merker JD, Patterson BK, Ganem D, DeRisi JL (2006) Microarray detection of human parainfluenzavirus 4 infection associated with respiratory failure in an immunocompetent adult. *Clin Infect Dis* 43(8):e71–e76
- Chiu CY, Alizadeh AA, Rouskin S, Merker JD, Yeh E, Yagi S, Schnurr D, Patterson BK, Ganem D (2007) Diagnosis of a critical respiratory illness caused by human metapneumovirus by use of a pan-virus microarray. *J Clin Microbiol* 45:2340–2343
- Chizhikov V, Wagner M, Ivshina A, Hoshino Y, Kapikian AZ, Chumakov K (2002) Detection and genotyping of human group A rotaviruses by oligonucleotide microarray hybridization. *J Clin Microbiol* 40:2398–2407
- Chou CC, Lee TT, Chen CH, Hsiao HY, Lin YL, Ho MS, Yang PC, Peck K (2006) Design of microarray probes for virus identification and detection of emerging viruses at the genus level. *BMC Bioinform* 28(7):232–240
- Cunningham CO (2004) Use of molecular diagnostic tests in disease control: making the leap from laboratory to field application. In: *Current trends in the study of bacterial and viral fish and shrimp diseases*. World Scientific Publishing Co., Singapore, pp 292–312
- Elnifro EM, Ashshi AM, Cooper RJ, Klapper PE (2000) Multiplex PCR: optimization and application in diagnostic virology. *Clin Microbiol Rev* 13:559–570
- Espy MJ, Uhl JR, Sloan LM, Buckwalter SP, Jones MF, Velter EA, Yao JDC, Wengenach NL, Rosenblatt JE, Cockerill FR, Smith TF (2006) Real-time PCR in clinical microbiology: applications for routine laboratory testing. *Clin Microbiol Rev* 19(1):165–256
- Feilotter HE (2004) Microarrays in veterinary diagnostics. *Anim Health Res Rev* 5(2):249–255
- Gardner SN, Crystal JJ, Kevin SM, Tom RS (2010) A microbial detection array (MDA) for viral and bacterial detection. *BMC Genomics* 11:668
- Goldsmith CS, Miller SE (2009) Modern uses of electron microscopy for detection of viruses. *Clin Microbiol Rev* 22(4):552–563
- Grubaugh ND, Petz LN, Melanson VR, McMenamy SS, Turell MJ, Long LS, Pisarcik SE, Kengluetcha A, Jaichapor B, O’Guinn ML, Lee JS (2013) Evaluation of a field-portable DNA microarray platform and nucleic acid amplification strategies for the detection of arboviruses, arthropods, and bloodmeals. *Am J Trop Med Hyg* 88(2):245–253
- Hentges DJ (1993) The anaerobic microflora of the human body. *Clin Infect Dis* 16(Suppl 4):S175
- Jabado OJ, Liu Y, Conlan S, Quan PL, Hegyi H, Lussier Y, Briese T, Palacios G, Lipkin WI (2008) Comprehensive viral oligonucleotide probe design using conserved protein regions. *Nucleic Acids Res* 36(1):e3
- Jack PJ, Amos-Ritchie RN, Reverter A, Palacios G, Quan PL (2009) Microarray based detection of viruses causing vesicular or vesicular-like lesions in livestock animals. *Vet Microbiol* 133:145–153
- Jayaraman A, Hall CK, Genzer J (2006) Computer simulation study of molecular recognition in model DNA microarrays. *Biophys J* 91:2227–2236
- Kistler A, Avila PC, Rouskin S, Wang D, Ward T, Yagi S, Schnurr D, Ganem D, DeRisi JL, Boushey HA (2007) Pan-viral screening of respiratory tract infections in adults with and without asthma reveals unexpected human coronavirus and human rhinovirus diversity. *J Infect Dis* 196(6):817–825
- Kostrzynska M, Bachand A (2006) Application of DNA microarray technology for detection, identification, and characterization of food-borne pathogens. *Can J Microbiol* 52(1):1–8
- Kreuzer J, Perez A, Untiveros M, Quispe D, Fuentes S, Barker I (2009) Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: a generic method for diagnosis, discovery and sequencing of viruses. *Virology* 388:1–7
- Kunin V, Copeland A, Lapidus A, Mavromatis K, Hugenholtz P (2008) A bioinformatician’s guide to metagenomics. *Microbiol Mol Biol Rev* 72(4):557–578

- Kurian KM, Watson CJ, Wyllie AH (1999) DNA chip technology. *J Pathol* 187(3):267–271
- Leblanc N, Gantelius J, Schwenk JM, Stahl K, Blomberg J, Andersson- Svahn H, Belak S (2009) Development of a magnetic bead microarray for simultaneous and simple detection of four pestiviruses. *J Virol Methods* 155:1–9
- Leland DS, Christine CG (2007) Role of cell culture for virus detection in the age of technology. *Microbiol Rev* 20(1):49–78
- Lin B, Wang Z, Vora GJ, Thornton JA, Schnur JM, Thach DC, Blaney KM, Ligler AG, Malanoski AP (2006) Broad-spectrum respiratory tract pathogen identification using resequencing DNA microarrays. *Genome Res* 16:527–535
- Liu YC, Huang GS, Wu MC, Hong MY, Hsiung KP (2006) Detection of foot and mouth disease and porcine reproductive and respiratory syndrome viral genes using microarray chip. *Vet Res Commun* 30(2):191–204
- Loy A, Bodrossy L (2006) Highly parallel microbial diagnostics using oligonucleotide microarrays. *Clin Chim Acta* 363:106–119
- Mackay IM, Arden KE, Nitsche A (2002) Real-time PCR in virology. *Nucleic Acids Res* 30(6):1292–1305
- Majid L, Vladimir C, Maxim M, Sergei S, Konstantin C (2003) Detection and discrimination of orthopoxviruses using microarrays of immobilized oligonucleotides. *J Virol Methods* 112(1–2):67–78
- Manoj KR (2009) The widely used diagnostics “DNA microarray”- a review. *Ame Jourof Inf Dis* 5(3):207–218
- Martin V, Perales C, Abia D, Ortíz AR, Domingo E, Briones C (2006) Microarray-based identification of antigenic variants of foot-and-mouth disease virus: a bioinformatics quality assessment. *BMC Genomics* 7:117
- Mihindukulasuriya KA, Wu G, St. Leger J, Nordhausen RW, Wang D (2008) Identification of a novel coronavirus from a beluga whale by using a panviral microarray. *J Virol* 82:5084–5088
- Miller SW, Avidor-Reiss T, Polyanovsky A, Posakony JW (2009) Complex interplay of three transcription factors in controlling the tormogen differentiation program of *Drosophila* mechanoreceptors. *Dev Biol* 329(2):386–399
- Morens DM, Fauci AS (2013) Emerging infectious diseases: threats to human health and global stability. *PLoS Pathog* 9(7):e1003467
- Mullis K (1990) The unusual origin of the polymerase chain reaction. *Sci Am* 4:56–65
- Nii S (1971) Electron microscopic observations on FL cells infected with herpes simplex virus. I. Viral forms. *Biken J* 14(2):177–190
- Palacios G, Quan PL, Jabado OJ, Conlan S (2007) Panmicrobial oligonucleotide array for diagnosis of infectious diseases. *Emerg Infect Dis* 13(1):73–81
- Petricoin EF, Zoon KC, Kohn EC, Barrett JC, Liotta LA (2002) Clinical proteomics: translating benchside promise into bedside reality. *Nat Rev Drug Discov* 1:683–695
- Quan PL, Palacios G, Jabado OJ, Conlan S (2007) Detection of respiratory viruses and subtype identification of influenza A viruses by GreeneChipResp oligonucleotide microarray. *J Clin Microbiol* 45(8):2359–2364
- Ratta B, Yadav BS, Pokhriyal M, Saxena M, Sharma B (2013) Microarray chip based identification of a mixed infection of bovine herpesvirus 1 and bovine viral diarrhea 2 from Indian cattle. *Curr Microbiol*. <https://doi.org/10.1007/s00284-013-0448-9>
- Rehrauer H, Schönmann S, Eberl L (2008) PhyloDetect: a likelihood-based strategy for detecting microorganisms with diagnostic microarrays. *Bioinformatics* 24:83–89
- Rockett JC, Dix DJ (2000) DNA arrays: technology, options and toxicological applications. *Xenobiotica* 30(2):155–177
- Sang R, Onyango C, Gachoya J, Mabinda E, Konongoi S, Ofula V, Dunster L, Okoth F, Coldren R, Tesh R, da Rossa AT, Finkbeiner S, Wang D, Crabtree M, Miller B (2006) Tickborne arbovirus surveillance in market livestock, Nairobi, Kenya. *Emerg Infect Dis* 12(7):1074–1080
- Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270(5235):467–470

- Sharma B, Pokhriyal M, Rai GK, Saxena M, Ratta B, Chaurasia M, Liu BS, Sen A, Mondal B (2012) Isolation of Newcastle disease virus from a non-avian host (sheep) and its implications. *Arch Virol* 157(8):1565–1567
- Staudt LM, Brown PO (2000) Genomic views of the immune system. *Annu Rev Immunol* 18:829–859
- Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* 98(3):503–517
- Su H, Long J, Guo Q, Meng X, Tan Y, Cai Q, Chen Z, Meng X (2013) Long-tail probe-mediated cycled strand displacement amplification: Label-free, isothermal and sensitive detection of nucleic acids. *Talanta* 15(116):330–334
- Tan PK, Downey TJ, Spitznagel EL Jr, Xu P, Fu D, Dimitrov DS, Lempicki RA, Raaka BM, Cam MC (2003) Evaluation of gene expression measurements from commercial microarray platforms. *Nucleic Acids Res* 31:5676–5684
- Tomita N, Mori Y, Kanda H, Notomi T (2008 May) Loop-mediated isothermal amplification (LAMP) of gene sequences and simple visual detection of products. *Nat Protoc* 3(5):877
- Tringe SG, Rubin EM (2005) Metagenomics: DNA sequencing of environmental samples. *Nat Rev Gen* 6:805–814
- Urisman A, Molinaro RJ, Fischer N, Plummer SJ, Casey G (2006) Identification of a novel Gammaretrovirus in prostate tumors of patients homozygous for R462Q RNASEL variant. *PLoS Pathol* 2(3):e25
- Vincent AL (2009) Characterization of influenza A virus isolated from pigs during an outbreak of respiratory disease in swine and people during a county fair in the United States. *Vet Microbiol* 137:51–59
- Voller A, Bartlett A, Bidwell DE (1978) Enzyme immunoassays with special reference to ELISA techniques. *J Clin Pathol* 31:507–520
- Wang D, Coscoy L, Zylberberg M, Avila PC, Boushey HA, Ganem D, DeRisi JL (2002) Microarray-based detection and genotyping of viral pathogens. *Proc Natl Acad Sci* 99(24):15687–15692
- Wang D, Urisman A, Liu Y-T, Springer M, Ksiazek TG (2003) Viral discovery and sequence recovery using DNA microarrays. *PLoS Biol* 1(2):e2
- Watson M, Dukes J, Abu-Median AB, King DP, Britton P (2007) DetectiV: visualization, normalization and significance testing for pathogen-detection microarray data. *Genome Biol* 8(9):R190
- Wilson WJ, Strout CL, DeSantis TZ, Stilwell JL, Carrano AV, Andersen GL (2002) Sequence-specific identification of 18 pathogenic microorganisms using microarray technology. *Mol Cell Probes* 216(2):119–127
- Yadav BS, Ronda V, Vashista DP, Sharma B (2013) Sequencing and computational approach to identification and characterization of microbial organisms. *Biomed Eng Comput Biol* 5:43–49
- Yadav BS, Pokhriyal M, Vasishtha DP, Sharma B (2014) Animal viruses probe dataset (AVPDS) for microarray-based diagnosis and identification of viruses. *Curr Microbiol*. <https://doi.org/10.1007/s00284-013-0477-4>
- Yadav BS, Pokhriyal M, Ratta B, Kumar A, Saxena M, Sharma B (2015) Viral diagnosis in Indian livestock using customized microarray chips. *Bioinformatics* 11(11):489–492
- Zappa A, Antonella A, Zanetti A (2009) Emerging and re-emerging viruses in the era of globalization. *Blood Transfus* 7(3):167–171