Author Manuscript Author Manuscript Author Manuscript Author Manuscript

HHS Public Access
Author manuscript
Virus Res. Author manuscript; available in PMC 2018 July 15.

Published in final edited form as:
Virus Res. 2017 July 15; 239: 172–179. doi:10.1016/j.virusres.2017.06.005.

Evolution Of Selective-Sequencing Approaches For Virus Discovery And Virome Analysis

Arvind Kumar[1], **Satyapramod Murthy**[1], and **Amit Kapoor**[1,2,*]

[1]Center for Vaccines and Immunity, The Research Institute at Nationwide Children's Hospital, Columbus, OH 43205, USA

[2]Department of Pediatrics, College of Medicine and Public Health, Ohio State University, Columbus, OH 43210

Abstract

Recent advances in sequencing technologies have transformed the field of virus discovery and virome analysis. Once, mostly confined to the traditional Sanger sequencing based individual virus discovery, is now entirely replaced by high throughput sequencing (HTS) based virus metagenomics that can be used to characterize the nature and composition of entire viromes. To better harness the potential of HTS for study of viromes, sample preparation methodologies used different refinements to exclude amplification of non-viral components that can overshadow low-titer viruses. These virus-sequence enrichment approaches mostly focused on the sample preparation methods, like enzymatic digestion of non-viral nucleic acids and size exclusion of non-viral constituents by column filtration, ultrafiltration or density centrifugation. However, recently an approach of virus-sequence enrichment called virome-capture sequencing, focused on the amplification or HTS library preparation stage, was shown to increase the ability of virome characterization. This new approach has the potential to further transform the field of virus discovery and virome analysis, but its technical complexity and sequence-dependence warrants further improvements. In this review we have listed the different methods, their applications and evolution, for selective sequencing based virome analysis and also the major refinements needed to harness the full potential of HTS for virome analysis.

1. Introduction

Despite being the simplest of biological entity, viruses play an enormously important and complex role in human and animal health, environmental ecology and are known to shape the evolution of their hosts (Greenbaum and Ghedin, 2015; Koonin and Dolja, 2013; Koonin et al., 2015; Koonin and Wolf, 2012; Mager and Stoye, 2015). Virus discovery during most

*Corresponding author. Amit Kapoor, Ph.D., Associate Professor, Department of Pediatrics, College of Medicine, The Ohio State University, Center for Vaccines and Immunity, The Research Institute at Nationwide Children's Hospital, 700 Children's Drive Columbus, Ohio 43205, Office: 614 7224064, Lab: 614 7243722, Fax: 614 7223680, Kapoor.102@osu.edu, Amit.Kapoor@Nationwidechildrens.org.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

of the previous century followed a traditional approach comprising isolation of viruses in cell culture or animal models (Leland and Ginocchio, 2007; Palacios and Oberste, 2005). These virus isolates were then classified based on their morphological and serological properties (Gelderblom, 1996; Muir et al., 1998). Importantly, the widely used cell culture isolation of viruses failed to identify viruses that were refractory to grow in vitro, like hepatitis C virus (HCV) (Choo et al., 1989). However, most of these traditional approaches were very specific or focused on a group of viruses. The inherent dependence of these approaches on the biological properties or sequence of known viruses resulted in our limited knowledge of the virus world, as it’s known now.

Use of sequence dependent (*i.e*; PCR and microarray) and sequence independent (*i.e*; sequence independent single primer amplification (SISPA) and random priming) approaches for nucleic acid amplification integrated with Sanger or HTS allowed new means to identification of new viruses and their profile after 1980 (Bishop-Lilly et al., 2010; Chang et al., 1994; Day et al., 2010; Grard et al., 2012; Kapoor et al., 2015; Ladner et al., 2016; Linnen et al., 1996; Matsui et al., 1991; Mokili et al., 2012; Muerhoff et al., 1997; Nichol et al., 1993; Qin et al., 2014; Quan et al., 2010; Simons et al., 1995b) (Figure 1). PCR and microarray based approach require the knowledge of viral sequence for generation of consensus primers and probes while SISPA and random priming methods can be used without sequence information. In SISPA method, a primer-binding sequence is ligated to both ends of a cDNA fragment which can be amplified in PCR with a single primer. Random PCR uses a slightly different approach in which the primer-binding sequence is introduced during cDNA synthesis by using a primer binding extension sequence on random primers. Viruses can live everywhere including within the cells or their host and environment. Comprehensive identification of viruses poses unique challenges due to their tiny and complex nature of genomes and no correlate of bacteria’s or eukaryotic conserved ribosomal sequences in viruses that can facilitate their pan-amplification and sequencing (Iwen et al., 2002; Ju and Zhang, 2015). Considering these challenges, an ideal sequencing based virus identification approach should use a selection method to increase the virus sequences or reduce non-viral sequences while remaining unbiased towards the nature and composition of the virus genome.

The term metagenomics can be defined as characterization of genetic information directly from clinical or environmental samples without culturing them (Handelsman et al., 1998).The culture-independent nature of this approach allowed discovery of unprecedented microbial diversity that remained underestimated by culture-dependent approaches. This vastly useful information gained by bacterial metagenomics studies ignited interest of marine virologists to adapt this approach for virus metagenomics (Breitbart et al., 2002). Soon after, several groups used this approach, or its modified forms, to identify human and animal viruses. The early use of the metagenomics approach for human and animal clinical samples was very successful. New viruses were identified from in vitro cultures supernatants, respiratory secretions, stool suspensions, urine and other body fluids (Allander et al., 2001; Allander et al., 2005; Jones et al., 2005; Kapoor et al., 2008; Victoria et al., 2008). Metagenomics remains the most widely used approach for new virus identification and virome analysis to date. Here we discuss the use and limitations of virus metagenomics;

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

recently used selective sequencing based virome analysis and provides a perspective on future refinements of virus metagenomics.

2. Evolution and use of sequencing technologies for virus metagenomics

Simple dideoxy or Sanger sequencing was not very applicable for virus metagenomics studies but was the only option available to virologists before 2005. The process involved plasmid cloning of fragmented or amplified nucleic acids from environmental or clinical samples followed by their sequencing. Technical and cost constrains allowed sequencing of only a few hundred clones in most of these early studies but even this small scale sequencing revealed presence of vast diversity among viral communities and also led to identification of several human and animal viruses in clinical samples (Allander et al., 2001; Allander et al., 2005; Chang et al., 1994; Choo et al., 1989; Kapoor et al., 2009; Kapoor et al., 2008; Linnen et al., 1996; Matsui et al., 1991; Muerhoff et al., 1997; Nichol et al., 1993; Nishizawa et al., 1997; van der Hoek et al., 2004; Victoria et al., 2008). These studies highlighted the importance of adapting more efficient sequencing technologies for virus metagenomics. The section below briefly describes different sequencing platforms, their advantages and disadvantages for virus metagenomics based virus discoveries and virome analysis.

454 or pyroseqeuncing

This is the first non-Sanger deep sequencing technology used for metagenomics based virus discovery. The principle of 454 (later purchased by Roche) technology is sequencing-by-synthesis chemistry in which DNA molecules are amplified through an emulsion PCR, generating multiple clones of DNA using a single template. Pyrophosphates released during base incorporation are enzymatically converted to a light signal which can be detected using a charged couple devices camera (Droege and Hill, 2008). 454 sequencing was widely used to identify several new viruses and virome profiles from human and animal samples (Day et al., 2010) including arboviruses (Bishop-Lilly et al., 2010), orbiviruses (Li et al., 2014), arenaviruses (Palacios et al., 2008), Lujo virus (Briese et al., 2009), astrovirus (Quan et al., 2010), gyroviruses (Phan et al., 2012), porcine bocaviruses (Yu et al., 2013), picornaviruses (Boros et al., 2012), rhabdoviruses (Grard et al., 2012), coronaviruses (Honkavuori et al., 2014), gammapapillomavirus (Phan et al., 2013) and seadornavirus (Reuter et al., 2013). Most of these viruses were identified from samples like serum, respiratory, and fecal samples. Some studies used this technology to identify viruses from tissue and organ samples- canine adenovirus 1 (van der Heijden et al., 2012), bat hepacivirus and pegivirus (Quan et al., 2013). This method generally allows sequencing of 500,000 to 5 million individual reads that can be between 100–450 nucleotides long. Although, this technology offered a higher yield than Sanger sequencing at a lower cost, this technology has been supplanted by other NGS technologies due to high cost, error rate in homopolymeric regions and artificial amplification.

Ion torrent

Life Technologies/ Thermo Fisher Scientific released pH-mediated or semiconductor sequencing technology based Ion personal genomics machine (PGM) sequencer platform in 2010 followed by Ion Proton (2012) and Ion S5 series (2015). These platforms are

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

conceptually similar to the 454 pyrosequencing platform in template preparation and sequencing steps. Adapter-ligated DNA fragments are amplified by emulsion-PCR on the surface of beads which are distributed into microwells where a sequencing-by-synthesis reaction occurs. Protons released during nucleotide incorporation are detected using an ion sensor, which can measure slight shifts in pH (Merriman et al., 2012; Reuter et al., 2015). Rapid sequencing run make this sequencers particularly useful for targeted detection of viruses in clinical samples like HIV (Archer et al., 2012; Chang et al., 2013; Gibson et al., 2014), hepatitis B virus (Yan et al., 2015), HCV (Gaspareto et al., 2016; Marascio et al., 2016), and rapid genome sequencing of several viruses including toscana virus (Nougairede et al., 2013), polyomavirus (Anthony et al., 2013), porcine reproductive and respiratory syndrome virus (Kvisgaard et al., 2013), orthoreovirus (Steyer et al., 2013), bluetongue virus (Lorusso et al., 2014), rotavirus (Ndze et al., 2014; Nyaga et al., 2014), influenza virus (Van den Hoecke et al., 2015) etc. Although, some studies used this technologies to study virome in skin (Bzhalava et al., 2013), ticks (Tokarz et al., 2014; Xia et al., 2015), gut virome in piglets (Karlsson et al., 2016) and seals (Kluge et al., 2016), this platform is not the ideal choice for virome study in human clinical samples due to lower outputs.

Illumina

Solexa/Illumina introduces high throughput platform and less expensive in cost Genome analyzer II platform in 2006 followed by MiSeq (2011), NextSeq 500 (2014) and the HiSeq series (2012–2014). The principle of Illumina sequencing involves reversible-termination sequencing by synthesis (SBS) with fluorescently labelled nucleotides (Liu et al., 2012). Introduction of these platforms accelerated the rate of virus discovery in humans and animals like severe fever with thrombocytopenia virus (Xu et al., 2011), bas-congo virus (Grard et al., 2012), titi monkey adenovirus (Chen et al., 2011), canine bocavirus 3 (Li et al., 2013), snake arenaviruses (Stenglein et al., 2012), human polyomavirus 9 (Sauvage et al., 2011), simian adenovirus c (Chiu et al., 2013), theiler’s disease associated virus (Chandriani et al., 2013), human hepegivirus 1 (Kapoor et al., 2015), jingmen tick virus (Qin et al., 2014), guaico culex virus (Ladner et al., 2016), protoparvovirus (Phan et al., 2016), marmota himalayana hepatovirus (Yu et al., 2016) etc. High throughput and low error rates (below 1% and mainly substitutions) are the main reasons that Illumina technologies have dominated the viral discovery field in the past several years.

PacBio sequencing

Pacific Biosciences commercialized single-molecule real-time (SMRT) sequencing platform known as RS II in 2010 followed by Sequel (2015) for generation of long read (average 10–50 kb) without clonal amplification bias. Hairpin adapters are ligated on the end of template DNA molecule for generation of capped template (SMRT-bell). Zero-mode waveguides (ZMW) with a single molecule of DNA and fluorescent nucleotides is affixed with a single DNA polymerase enzyme at the bottom. During polymerization fluorescent tag of nucleotides is cleaved off and diffuses out of the observation area of the ZMW and detected by detector in real time (Reuter et al., 2015; Rhoads and Au, 2015). Single molecule sequencing is extremely well suited for virus metagenomics studies, because of the very long reads. It is otherwise nearly impossible to obtain long contigs in a biologically diverse sample (Brinzevich et al., 2014; Tombacz et al., 2015) (Archer et al., 2012; Bergfors et al.,

2016), (Schleiss et al., 2014; Tombacz et al., 2014; Wittmann et al., 2014) but lower throughput, higher costs per base sequencing and higher error rates currently limits the scope for viral metagenomics study of low titer viruses.

Nanopore sequencing

Oxford Nanopore Technologies released a nanopore based portable sequencer MinION in 2014. The principle of this technology is measuring information about the characteristic changes that are induced as the biological molecules (DNA/RNA) passing through the nanopore by a molecular motor protein. This is the first portable sequencer with capability of RNA/DNA sequencing, longer read length (approximate 300 kb) in few hours, and real time sequence analysis. Nanopore platform has been used for viral detection in human clinical samples (Greninger et al., 2015; Hoenen et al., 2016) and genome sequencing (Karamitros et al., 2016; Kilianski et al., 2016). Although this platform looks promising for epidemiological investigation during an outbreak but low output and high error rate (around 10%) (McGinn et al., 2016) is only a concern for virus discovery.

3. Selective sequencing approaches

Currently high throughput sequencing technologies are capable of producing of billions of sequence reads but detection of viruses in clinical samples is still challenging due to the presence of extremely small quantity of viral nucleic acids in combination with a relatively high background of host, bacterial and other contaminating genetic material. Therefore efficient and sensitive metagenomics study of viruses in clinical or environmental samples require removal of non-virus nucleic acids and/or enrichment of virus-derived nucleic acids (Capobianchi et al., 2013; Conceicao-Neto et al., 2015; Hall et al., 2014). These virus enrichment methods can be broadly classified as sample preparation methods and sequencing library preparation methods (Figure-2), or alternatively can be also called pre-extraction (Figure-2A) and post-extraction virus enrichment methods (Figure-2B), respectively.

Choice of pre-extraction virus enrichment methods largely depend on the nature of sample to be analyzed. Generally the virus metagenomics samples are contaminated with host (bacteria or eukaryote) and environmental nucleic acids. For virome analysis of environmental samples, several methods including ultra-filtration, iron chloride precipitation and density gradient centrifugation (polyethylene glycol, sucrose cushion) are available for pre extraction virus enrichment (Andrews-Pfannkoch et al., 2010; John et al., 2011). For virome analysis of clinical samples with low abundance of host cells, like cerebrospinal fluid, respiratory samples, serum, urine or stools, filtration and nuclease digestion or density centrifugation can be used as a method of pre-extraction virus enrichment (Batty et al., 2013; Conceicao-Neto et al., 2015; Daly et al., 2011; Hall et al., 2014; Kohl et al., 2015; Rosseel et al., 2015). For virome analysis of clinical samples with abundance of host cells, like blood or tissues, pre-extraction based enrichment is not appropriate as the virus genome itself can be present in its non-capsidated or transcribed form.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Virus particle enrichment methods

Most of the pre-extraction virus enrichment methods are based on the physical properties of virions and their differences from the other living forms. Virus particles or virions are encapsidated forms of virus genomes. The three virus properties used in pre-extraction virus enrichment methods are the size of virus particles, their density, and the presence of capsid that protects the virus genome. Even the first recognition of virus as an entity in early 19th century, then recognized as liquid poison, was done on their ability to pass through the Cahmberland filters that can retain most of bacteria (Lecoq, 2001). Expect for a few large recently identified viruses (Halary et al., 2016), most animal viruses are less than 200–300 nm in diameter and therefore can pass through most filters with 0.2 to 0.45 micron pore size. Filtration is therefore the most commonly used method for selectively sequencing the viruses from environmental or clinical samples. Notably, several metagenomic based virus identification studies used the 0.45 micron filter and not the 0.2 micron filter because the pore diameters of commercial filters are quite variable and using a filter with pore diameter closer to most viruses can reduce the amount of virus in the filtrate (Conceicao-Neto et al., 2015).

The compact nature of virions allows use of density gradient centrifugation (sucrose, cesium chloride, poly ethylene glycol) as a method to enrich samples for virus derived nucleic acids. Density centrifugation is commonly used in molecular virology labs and was also used in the first virus metagenomics study for enriching viruses in the sea water samples. This method was also used to study viromes of lakes, soils and from extreme environments (Brum et al., 2013; Kleiner et al., 2015; Thurber et al., 2009). The major limitation of density centrifugation method is its applicability for analysis of clinical samples. Moreover, the differences in density of enveloped and non-enveloped viruses and other virus variables may require analysis of several gradient fractions. The presence of contaminating nucleic acids in the density gradient solution can further contaminate the samples with environmental nucleic acids.

Most virus capsid protects viral genome from degradation by endonuclease (Benzonase and Dnase I), endo-exo nucleases (Micrococcal nuclease) and endoribonuclease (Rnase A). Benzonase and microcccal nuclease digests all forms of nucleic acid including DNA and RNA (single stranded, double stranded, linear and circular). RNase A specifically degrades single-stranded RNA at C and U residues while DNase I nonspecifically cleaves single and double stranded DNA to release di-, tri- and oligonucleotide products with 5´-phosphorylated and 3´-hydroxylated ends. This property of virions was frequently used to study viruses in molecular biology research by plant and animal virologists (Rosseel et al., 2015; Thurber et al., 2009). However, the nuclease treatment or nuclease mediated removal of non-capsidated nucleic acids to enrich a sample for virus nucleic acids was first attempted by Allander et.al. (Allander et al., 2001). The study showed that DNAse treatment of virus containing serum samples can help enrich the metagenomic libraries from virus derived sequences. This work described the identification of two novel parvovirus species in bovine serum. Soon after, this group and then several others used a similar approach to identify a plethora of human and animal viruses. Although very helpful, nuclease treatment is not very effective in degrading unprotected nucleic acids and often fails to enrich samples enough for

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

identification of low-titer viruses. Moreover, nuclease treatment cannot be used to find viruses in cells or tissue samples where all virus nucleic acids are not encapsidated.

Positive selection based virus or virome capture methods

Viral pathogens usually have much smaller genome relative to their host, so even with pre extraction virus enrichment; the percentage of viral reads in NGS data is relatively low. In our experience, we generally get 70–90% host nucleic acid in clinical samples besides pre extraction virus enrichment. In positive selection method, samples are enriched for viruses using PCR assays, microarray or virus capture (in solution based hybridization) approaches. The simplest examples of the positive selection approach are generic PCR assays which uses degenerate primers to target several related viruses or their variants (Irving et al., 2014; Kwok and Chiang, 2016; Simons et al., 1995a). PCR based approach is restricted by its ability to detect only a limited number of viruses due to issue of multiplexity. To avoid this, several groups developed a positive selection approach for enriching samples for viruses of a defined taxonomic group (family, genus and species) based on DNA microarray (Gardner et al., 2010; Palacios et al., 2007; Wang et al., 2002). DNA microarray have been employed to characterize and discovery for a number of novel or variant viruses including human cardioviruses, porcine circovirus, rhinovirus and adenovirus respectively (Chen et al., 2011; Chiu et al., 2008; Kistler et al., 2007; Palacios et al., 2007; Rota et al., 2003). However, due to their limited specificity, none of these methods were suitable for comprehensive characterization of vertebrate viruses or analysis of viromes. Most recently two very comprehensive probe sets were developed for positive selection based sequencing of all known vertebrate viruses and their variants (Briese et al., 2015; Wylie et al., 2015). These methods use virus-specific probes in a liquid-phase hybridization to capture viral sequences from metagenomics libraries. Human exon capture sequencing is an ideal example of positive selection method (Kozarewa et al., 2015) based on these approach. One of these approaches was termed Virome Capture Sequencing (VirCapSeq) (Briese et al., 2015). Basically, VirCapSeq is similar to any other DNA hybridization based enrichment methods, including the well-known exon capture method used for transcriptomics. In principle, VirCapSeq is made of set of specific oligonucleotides (or probes), like used on viral microarrays, that when mixed with samples can hybridize or capture complementary viral nucleic acids (Figure 2 B–b). VirCapSeq include probes for all viruses known to infect vertebrates by targeting their protein coding regions. However, the design of VirCapSeq was challenging due to the constant influx of genome data from newly identified viruses and the biased abundance of virus genomes of pathogenic viruses, like HIV and HCV in sequence databases. The design of VirCapSeq, therefore required use of a unique probe set containing 2 million probes or oligonucleotides that differed by at least 10% to reduce sequence redundancy and widen the coverage of virus sequences. Stringent hybridization conditions of VirCapSeq allow specific enrichment of virus-like sequences in libraries generated from different sample types. Compared to common metagenomics and other virus enrichment approaches (filtration and nuclease digestion prior to total nucleic acid extraction and RiboZero rRNA depletion after extraction), the use of VirCapSeq gained 100–10,000-fold more virus sequences in the metagenomics libraries. The sensitivity of VirCapSeq was also compared with real time PCR for detection of viral sequence in blood and serum samples spiked with live enterovirus D68. VirCapSeq was able to detect 10 copies/ml of virus in both

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

sample types which is comparable to the sensitivity of PCR based methods. Interestingly, use of VirCapSeq allows the identification of viruses whose genomes are vastly different from those used to design the probes, as all viruses of a taxonomic group share some highly conserved stretches of nucleotides. Another similar positive selection approach was termed ViroCap and was simultaneously developed by Wylie et.al.(Wylie et al., 2015). Like VirCapSeq, ViroCap also targeted most virus species (of 34 virus families) that are known to infect vertebrates and excluded known endogenous retroviruses. To design ViroCap, approximately 1 billion bases of sequence data was condensed to ~200 million bases of probe sequences. Use of ViroCap allowed 296 to 674 – fold increases in the number of virus reads compared to simple metagenomics sequencing.

Although both these positive selection approaches described above are biased towards sequences of known viruses, they can be efficiently used to study mixed infections and provide more sensitive characterization of all viruses present in clinical samples. As the appreciation of the role that viromes play in the human health is increasing, these platforms can provide useful tools to study the dynamics of human virome in longitudinal samples collected before and after appearance of diseases.

Negative selection based virus metagenomics

An ideal negative subtraction strategy will allow removal of non-viral nucleic acids including the nucleic acids derived from the host, the reagents and the environment. A negative selection approach can use the principle of suppression subtractive hybridization (SSH) or representational difference analysis (RDA) which allows comparison of two DNA populations (tester and driver) and enrichment of differentially distributed molecules (Diatchenko et al., 1996; Lisitsyn et al., 1993). The genomic DNA sample that contains the sequences of interest is referred to as tester and the reference sample is referred to as driver. Tester and driver DNAs are hybridized in excess of driver and the hybrid sequences are then removed. Consequently, the remaining unhybridized DNAs represent DNA molecules that are present in the tester yet absent from the driver DNA. Although these approaches have been usually employed to identify genes with a differential expression pattern (Yin et al., 2013), but successfully have been used to identify human herpes virus-8 (Chang et al., 1994), GB virus (Simons et al., 1995b), TT virus (Nishizawa et al., 1997), Walrus Calicivirus (Ganova-Raeva et al., 2004) and a new strain of murine hepatitis virus (Islam et al., 2015). Another simpler approach can be based on the simple subtraction of non-target nucleic acid using biotinylated probes (Figure-2B–a). One example of a negative selection approach is subtraction of ribosomal RNA (rRNA) from extracted ribonucleic acids. This method is very efficient in finding RNA viruses in samples, like serum and cerebrospinal fluid, where a majority of host RNA is rRNA (Matranga et al., 2014; Rosseel et al., 2015). Several commercial kit including Illumina's Ribo-Zero rRna removal kit, Ambion's GLOBINclear kit provide sequence specific (human/mouse/ rat/bacteria) biotinylated oligos designed to hybridize to the rRNA large. The rRNA:DNA hybrid can be captured on a streptavidin magnetic bead and removed from the sample, depleting the rRNA and thereby enriching these samples for virus RNA. New England Biolab's NEBNext rRNA depletion kit use slight approach which is based on solution hybridization of DNA oligo with rRNA and then degradation by RNase H, which digest DNA/RNA hybrid molecules. Recently a

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

new approach known as DASH (depletion of abundant sequences by hybridization) was described by Gu et al for depletion of rRNA with Cas9 protein complexed with a library of rRNA (Gu et al., 2016). However a major limitation is the unavailability of rRNA subtraction reagents for other vertebrate and invertebrate species. Moreover, rRNA subtraction is not promising when looking for viruses in cells and tissue samples that usually contain an abundance of genomic DNA and transcripts (Li et al., 2016).

Although theoretically straightforward, experimental optimization of the unbiased amplification of complex mixture of nucleic acids followed by their stringent hybridization and appropriate washing to elute unique sequences requires rigorous experimental design and validation. Kim et al used multiple displacement amplification approach for amplification of single stranded DNA and double stranded DNA (Kim et al., 2011). Although multiple displacement amplification approach is good but it should be avoided for metagenomic investigations that require quantitative estimates of microbial taxa and gene functional groups due to generation of amplification bias (Marine et al., 2014). Duhaime et al develop linker-amplified shotgun libraries (LASLs) method for generation of HTS sequencing library from 1 pg amount of DNA without amplification bias (Duhaime et al., 2012). Besides library preparation, library sequencing platform can also impact viral metagenomes (Solonenko et al., 2013). The impact of an ideal negative selection approach will be far reaching as it will enable not only the sensitive analysis of viromes but also the comparative analysis of changes in viromes within a host or between, or to compare viromes in samples obtained at different time points.

4. Conclusions and perspectives

Recent advances in sequencing technologies were successfully exploited for virus discovery and virome analysis yielding significant new information about viruses that live with and around us. However, existing approaches are not sensitive enough to analyze the virome of many sample types where the host and environmental nucleic acids are typically present in orders of magnitude higher than viral nucleic acids. Viruses residing within cells and tissues represent such a scenario where conventional viral metagenomics fail to identify low-titer viruses; however this problem can be overcome with increased depth. Moreover, the vast amount of sequence data generated using a metagenomics approach poses bioinformatics challenges for comparative analysis of viromes. Use of appropriate selective sequencing approaches, therefore can help in overcoming these challenges by reducing the amount of useless data.

Although in positive selection approaches, like VirCapSeq or ViroCap, a very broad selection of probes for all known vertebrate viruses allow comprehensive analysis of related viruses in clinical samples. The major limitation of this technique is its inherent bias of enriching samples for only known and targeted viruses. Even the most basic virus enrichment approaches like centrifugation and filtration were biased and therefore giant viruses remained unidentified (Halary et al., 2016). Some constituents of viromes are either fluctuating or continuously evolving and therefore cannot be studied using a constant probe set. In conclusion, these methods are more appropriate for a selected group of viruses and specially to study the relative presence and dynamic changes in their populations. In the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

future, a more focused VirCapSeq assay to study viromes and its fluctuations in specific diseases or body sites like blood, respiratory system, liver or gastrointestinatl tract, holds more enormous potential to gain new insights into the role of viromes in human health and development of chronic diseases. However, we believe that only a simple and efficient negative selection approach will become an ideal selective sequencing approach for the discovery of new viruses and also for meaningful analysis of viromes.

Acknowledgments

Research in the Kapoor lab is supported by National Institutes of Health award (HL119485), the US Department of Agriculture award, National Science Foundation award (FAIN- 1619072) and The Research Institute at Nationwide Children's Hospital.

References

Allander T, Emerson SU, Engle RE, Purcell RH, Bukh J. 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 of the United States of America. 2001; 98:11609–11614. [PubMed: 11562506]

Allander T, Tammi MT, Eriksson M, Bjerkner A, Tiveljung-Lindell A, Andersson B. Cloning of a human parvovirus by molecular screening of respiratory tract samples. Proceedings of the National Academy of Sciences of the United States of America. 2005; 102:12891–12896. [PubMed: 16118271]

Andrews-Pfannkoch C, Fadrosh DW, Thorpe J, Williamson SJ. Hydroxyapatite-mediated separation of double-stranded DNA, single-stranded DNA, and RNA genomes from natural viral assemblages. Applied and environmental microbiology. 2010; 76:5039–5045. [PubMed: 20543058]

Anthony SJ, St Leger JA, Navarrete-Macias I, Nilson E, Sanchez-Leon M, Liang E, Seimon T, Jain K, Karesh W, Daszak P, Briese T, Lipkin WI. Identification of a novel cetacean polyomavirus from a common dolphin (Delphinus delphis) with Tracheobronchitis. PloS one. 2013; 8:e68239. [PubMed: 23874559]

Archer J, Weber J, Henry K, Winner D, Gibson R, Lee L, Paxinos E, Arts EJ, Robertson DL, Mimms L, Quinones-Mateu ME. Use of four next-generation sequencing platforms to determine HIV-1 coreceptor tropism. PloS one. 2012; 7:e49602. [PubMed: 23166726]

Batty EM, Wong TH, Trebes A, Argoud K, Attar M, Buck D, Ip CL, Golubchik T, Cule M, Bowden R, Manganis C, Klenerman P, Barnes E, Walker AS, Wyllie DH, Wilson DJ, Dingle KE, Peto TE, Crook DW, Piazza P. A modified RNA-Seq approach for whole genome sequencing of RNA viruses from faecal and blood samples. PloS one. 2013; 8:e66129. [PubMed: 23762474]

Bergfors A, Leenheer D, Bergqvist A, Ameur A, Lennerstrand J. Analysis of hepatitis C NS5A resistance associated polymorphisms using ultra deep single molecule real time (SMRT) sequencing. Antiviral Res. 2016; 126:81–89. [PubMed: 26707078]

Bishop-Lilly KA, Turell MJ, Willner KM, Butani A, Nolan NM, Lentz SM, Akmal A, Mateczun A, Brahmbhatt TN, Sozhamannan S, Whitehouse CA, Read TD. Arbovirus detection in insect vectors by rapid, high-throughput pyrosequencing. PLoS neglected tropical diseases. 2010; 4:e878. [PubMed: 21085471]

Boros A, Nemes C, Pankovics P, Kapusinszky B, Delwart E, Reuter G. Identification and complete genome characterization of a novel picornavirus in turkey (Meleagris gallopavo). J Gen Virol. 2012; 93:2171–2182. [PubMed: 22875254]

Breitbart M, Salamon P, Andresen B, Mahaffy JM, Segall AM, Mead D, Azam F, Rohwer F. Genomic analysis of uncultured marine viral communities. Proceedings of the National Academy of Sciences of the United States of America. 2002; 99:14250–14255. [PubMed: 12384570]

Briese T, Kapoor A, Mishra N, Jain K, Kumar A, Jabado OJ, Lipkin WI. Virome Capture Sequencing Enables Sensitive Viral Diagnosis and Comprehensive Virome Analysis. mBio. 2015; 6:e01491–01415. [PubMed: 26396248]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Briese T, Paweska JT, McMullan LK, Hutchison SK, Street C, Palacios G, Khristova ML, Weyer J, Swanepoel R, Egholm M, Nichol ST, Lipkin WI. Genetic detection and characterization of Lujo virus, a new hemorrhagic fever-associated arenavirus from southern Africa. PLoS pathogens. 2009; 5:e1000455. [PubMed: 19478873]

Brinzevich D, Young GR, Sebra R, Ayllon J, Maio SM, Deikus G, Chen BK, Fernandez-Sesma A, Simon V, Mulder LC. HIV-1 interacts with human endogenous retrovirus K (HML-2) envelopes derived from human primary lymphocytes. J Virol. 2014; 88:6213–6223. [PubMed: 24648457]

Brum JR, Culley AI, Steward GF. Assembly of a marine viral metagenome after physical fractionation. PloS one. 2013; 8:e60604. [PubMed: 23580170]

Bzhalava D, Johansson H, Ekstrom J, Faust H, Moller B, Eklund C, Nordin P, Stenquist B, Paoli J, Persson B, Forslund O, Dillner J. Unbiased approach for virus detection in skin lesions. Plos One. 2013; 8:e65953. [PubMed: 23840382]

Capobianchi MR, Giombini E, Rozera G. Next-generation sequencing technology in clinical virology. Clin Microbiol Infect. 2013; 19:15–22. [PubMed: 23279287]

Chandriani S, Skewes-Cox P, Zhong W, Ganem DE, Divers TJ, Van Blaricum AJ, Tennant BC, Kistler AL. Identification of a previously undescribed divergent virus from the Flaviviridae family in an outbreak of equine serum hepatitis. Proceedings of the National Academy of Sciences of the United States of America. 2013; 110:E1407–1415. [PubMed: 23509292]

Chang MW, Oliveira G, Yuan J, Okulicz JF, Levy S, Torbett BE. Rapid deep sequencing of patient-derived HIV with ion semiconductor technology. J Virol Methods. 2013; 189:232–234. [PubMed: 23384677]

Chang Y, Cesarman E, Pessin MS, Lee F, Culpepper J, Knowles DM, Moore PS. Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. Science. 1994; 266:1865–1869. [PubMed: 7997879]

Chen EC, Yagi S, Kelly KR, Mendoza SP, Tarara RP, Canfield DR, Maninger N, Rosenthal A, Spinner A, Bales KL, Schnurr DP, Lerche NW, Chiu CY. Cross-species transmission of a novel adenovirus associated with a fulminant pneumonia outbreak in a new world monkey colony. PLoS pathogens. 2011; 7:e1002155. [PubMed: 21779173]

Chiu CY, Greninger AL, Kanada K, Kwok T, Fischer KF, Runckel C, Louie JK, Glaser CA, Yagi S, Schnurr DP, Haggerty TD, Parsonnet J, Ganem D, DeRisi JL. Identification of cardioviruses related to Theiler's murine encephalomyelitis virus in human infections. Proceedings of the National Academy of Sciences of the United States of America. 2008; 105:14124–14129. [PubMed: 18768820]

Chiu CY, Yagi S, Lu X, Yu G, Chen EC, Liu M, Dick EJ Jr, Carey KD, Erdman DD, Leland MM, Patterson JL. A novel adenovirus species associated with an acute respiratory outbreak in a baboon colony and evidence of coincident human infection. mBio. 2013; 4:e00084. [PubMed: 23592261]

Choo QL, Kuo G, Weiner AJ, Overby LR, Bradley DW, Houghton M. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. Science. 1989; 244:359–362. [PubMed: 2523562]

Conceicao-Neto N, Zeller M, Lefrere H, De Bruyn P, Beller L, Deboutte W, Yinda CK, Lavigne R, Maes P, Van Ranst M, Heylen E, Matthijnssens J. Modular approach to customise sample preparation procedures for viral metagenomics: a reproducible protocol for virome analysis. Scientific reports. 2015; 5:16532. [PubMed: 26559140]

Daly GM, Bexfield N, Heaney J, Stubbs S, Mayer AP, Palser A, Kellam P, Drou N, Caccamo M, Tiley L, Alexander GJ, Bernal W, Heeney JL. A viral discovery methodology for clinical biopsy samples utilising massively parallel next generation sequencing. PloS one. 2011; 6:e28879. [PubMed: 22216131]

Day JM, Ballard LL, Duke MV, Scheffler BE, Zsak L. Metagenomic analysis of the turkey gut RNA virus community. Virology journal. 2010; 7:313. [PubMed: 21073719]

Diatchenko L, Lau YF, Campbell AP, Chenchik A, Moqadam F, Huang B, Lukyanov S, Lukyanov K, Gurskaya N, Sverdlov ED, Siebert PD. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. Proceedings of the National Academy of Sciences of the United States of America. 1996; 93:6025–6030. [PubMed: 8650213]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Droege M, Hill B. The Genome Sequencer FLX System--longer reads, more applications, straight forward bioinformatics and more complete data sets. Journal of biotechnology. 2008; 136:3–10. [PubMed: 18616967]

Duhaime MB, Deng L, Poulos BT, Sullivan MB. Towards quantitative metagenomics of wild viruses and other ultra-low concentration DNA samples: a rigorous assessment and optimization of the linker amplification method. Environmental microbiology. 2012; 14:2526–2537. [PubMed: 22713159]

Ganova-Raeva L, Smith AW, Fields H, Khudyakov Y. New Calicivirus isolated from walrus. Virus research. 2004; 102:207–213. [PubMed: 15084403]

Gardner SN, Jaing CJ, McLoughlin KS, Slezak TR. A microbial detection array (MDA) for viral and bacterial detection. BMC genomics. 2010; 11:668. [PubMed: 21108826]

Gaspareto KV, Ribeiro RM, Malta FM, Gomes-Gouvea MS, Muto NH, Romano CM, Mendes-Correa MC, Carrilho FJ, Sabino EC, Pinho JR. Resistance-associated variants in HCV subtypes 1a and 1b detected by Ion Torrent sequencing platform. Antivir Ther. 2016

Gelderblom, HR. Structure and Classification of Viruses. In: Baron, S., editor. Medical Microbiology. 4. Galveston (TX): 1996.

Gibson RM, Meyer AM, Winner D, Archer J, Feyertag F, Ruiz-Mateos E, Leal M, Robertson DL, Schmotzer CL, Quinones-Mateu ME. Sensitive deep-sequencing-based HIV-1 genotyping assay to simultaneously determine susceptibility to protease, reverse transcriptase, integrase, and maturation inhibitors, as well as HIV-1 coreceptor tropism. Antimicrob Agents Chemother. 2014; 58:2167–2185. [PubMed: 24468782]

Grard G, Fair JN, Lee D, Slikas E, Steffen I, Muyembe JJ, Sittler T, Veeraraghavan N, Ruby JG, Wang C, Makuwa M, Mulembakani P, Tesh RB, Mazet J, Rimoin AW, Taylor T, Schneider BS, Simmons G, Delwart E, Wolfe ND, Chiu CY, Leroy EM. A novel rhabdovirus associated with acute hemorrhagic fever in central Africa. PLoS pathogens. 2012; 8:e1002924. [PubMed: 23028323]

Greenbaum BD, Ghedin E. Viral evolution: beyond drift and shift. Curr Opin Microbiol. 2015; 26:109–115. [PubMed: 26189048]

Greninger AL, Naccache SN, Federman S, Yu G, Mbala P, Bres V, Stryke D, Bouquet J, Somasekar S, Linnen JM, Dodd R, Mulembakani P, Schneider BS, Muyembe-Tamfum JJ, Stramer SL, Chiu CY. Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. Genome Med. 2015; 7:99. [PubMed: 26416663]

Gu W, Crawford ED, O'Donovan BD, Wilson MR, Chow ED, Retallack H, DeRisi JL. Depletion of Abundant Sequences by Hybridization (DASH): using Cas9 to remove unwanted high-abundance species in sequencing libraries and molecular counting applications. Genome biology. 2016; 17:41. [PubMed: 26944702]

Halary S, Temmam S, Raoult D, Desnues C. Viral metagenomics: are we missing the giants? Current opinion in microbiology. 2016; 31:34–43. [PubMed: 26851442]

Hall RJ, Wang J, Todd AK, Bissielo AB, Yen S, Strydom H, Moore NE, Ren X, Huang QS, Carter PE, Peacey M. Evaluation of rapid and simple techniques for the enrichment of viruses prior to metagenomic virus discovery. Journal of virological methods. 2014; 195:194–204. [PubMed: 24036074]

Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem Biol. 1998; 5:R245–249. [PubMed: 9818143]

Hoenen T, Groseth A, Rosenke K, Fischer RJ, Hoenen A, Judson SD, Martellaro C, Falzarano D, Marzi A, Squires RB, Wollenberg KR, de Wit E, Prescott J, Safronetz D, van Doremalen N, Bushmaker T, Feldmann F, McNally K, Bolay FK, Fields B, Sealy T, Rayfield M, Nichol ST, Zoon KC, Massaquoi M, Munster VJ, Feldmann H. Nanopore Sequencing as a Rapidly Deployable Ebola Outbreak Tool. Emerging infectious diseases. 2016; 22:331–334. [PubMed: 26812583]

Honkavuori KS, Briese T, Krauss S, Sanchez MD, Jain K, Hutchison SK, Webster RG, Lipkin WI. Novel coronavirus and astrovirus in Delaware Bay shorebirds. PLoS One. 2014; 9:e93395. [PubMed: 24699424]

Irving WL, Rupp D, McClure CP, Than LM, Titman A, Ball JK, Steinmann E, Bartenschlager R, Pietschmann T, Brown RJ. Development of a high-throughput pyrosequencing assay for

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

monitoring temporal evolution and resistance associated variant emergence in the Hepatitis C virus protease coding-region. Antiviral research. 2014; 110:52–59. [PubMed: 25057759]

Islam MM, Toohey B, Purcell DF, Kannourakis G. Suppression subtractive hybridization method for the identification of a new strain of murine hepatitis virus from xenografted SCID mice. Archives of virology. 2015; 160:2945–2955. [PubMed: 26347284]

Iwen PC, Hinrichs SH, Rupp ME. Utilization of the internal transcribed spacer regions as molecular targets to detect and identify human fungal pathogens. Med Mycol. 2002; 40:87–109. [PubMed: 11860017]

John SG, Mendez CB, Deng L, Poulos B, Kauffman AK, Kern S, Brum J, Polz MF, Boyle EA, Sullivan MB. A simple and efficient method for concentration of ocean viruses by chemical flocculation. Environmental microbiology reports. 2011; 3:195–202. [PubMed: 21572525]

Jones MS, Kapoor A, Lukashov VV, Simmonds P, Hecht F, Delwart E. New DNA viruses identified in patients with acute viral infection syndrome. J Virol. 2005; 79:8230–8236. [PubMed: 15956568]

Ju F, Zhang T. 16S rRNA gene high-throughput sequencing data mining of microbial diversity and interactions. Appl Microbiol Biotechnol. 2015; 99:4119–4129. [PubMed: 25808518]

Kapoor A, Kumar A, Simmonds P, Bhuva N, Singh Chauhan L, Lee B, Sall AA, Jin Z, Morse SS, Shaz B, Burbelo PD, Lipkin WI. Virome Analysis of Transfusion Recipients Reveals a Novel Human Virus That Shares Genomic Features with Hepaciviruses and Pegiviruses. mBio. 2015; 6:e01466–01415. [PubMed: 26396247]

Kapoor A, Slikas E, Simmonds P, Chieochansin T, Naeem A, Shaukat S, Alam MM, Sharif S, Angez M, Zaidi S, Delwart E. A newly identified bocavirus species in human stool. J Infect Dis. 2009; 199:196–200. [PubMed: 19072716]

Kapoor A, Victoria J, Simmonds P, Slikas E, Chieochansin T, Naeem A, Shaukat S, Sharif S, Alam MM, Angez M, Wang C, Shafer RW, Zaidi S, Delwart E. A highly prevalent and genetically diversified Picornaviridae genus in South Asian children. Proceedings of the National Academy of Sciences of the United States of America. 2008; 105:20482–20487. [PubMed: 19033469]

Karamitros T, Harrison I, Piorkowska R, Katzourakis A, Magiorkinis G, Mbisa JL. De Novo Assembly of Human Herpes Virus Type 1 (HHV-1) Genome, Mining of Non-Canonical Structures and Detection of Novel Drug-Resistance Mutations Using Short- and Long-Read Next Generation Sequencing Technologies. Plos One. 2016; 11:e0157600. [PubMed: 27309375]

Karlsson OE, Larsson J, Hayer J, Berg M, Jacobson M. The Intestinal Eukaryotic Virome in Healthy and Diarrhoeic Neonatal Piglets. Plos One. 2016; 11:e0151481. [PubMed: 26982708]

Kilianski A, Roth PA, Liem AT, Hill JM, Willis KL, Rossmaier RD, Marinich AV, Maughan MN, Karavis MA, Kuhn JH, Honko AN, Rosenzweig CN. Use of Unamplified RNA/cDNA-Hybrid Nanopore Sequencing for Rapid Detection and Characterization of RNA Viruses. Emerging infectious diseases. 2016; 22:1448–1451. [PubMed: 27191483]

Kim MS, Park EJ, Roh SW, Bae JW. Diversity and abundance of single-stranded DNA viruses in human feces. Applied and environmental microbiology. 2011; 77:8062–8070. [PubMed: 21948823]

Kistler A, Avila PC, Rouskin S, Wang D, Ward T, Yagi S, Schnurr D, Ganem D, DeRisi JL, Boushey HA. Pan-viral screening of respiratory tract infections in adults with and without asthma reveals unexpected human coronavirus and human rhinovirus diversity. The Journal of infectious diseases. 2007; 196:817–825. [PubMed: 17703411]

Kleiner M, Hooper LV, Duerkop BA. Evaluation of methods to purify virus-like particles for metagenomic sequencing of intestinal viromes. BMC genomics. 2015; 16:7. [PubMed: 25608871]

Kluge M, Campos FS, Tavares M, de Amorim DB, Valdez FP, Giongo A, Roehe PM, Franco AC. Metagenomic Survey of Viral Diversity Obtained from Feces of Subantarctic and South American Fur Seals. Plos One. 2016; 11:e0151921. [PubMed: 26986573]

Kohl C, Brinkmann A, Dabrowski PW, Radonic A, Nitsche A, Kurth A. Protocol for metagenomic virus detection in clinical specimens. Emerging infectious diseases. 2015; 21:48–57. [PubMed: 25532973]

Koonin EV, Dolja VV. A virocentric perspective on the evolution of life. Curr Opin Virol. 2013; 3:546–557. [PubMed: 23850169]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Koonin EV, Dolja VV, Krupovic M. Origins and evolution of viruses of eukaryotes: The ultimate modularity. Virology. 2015; 479–480:2–25.

Koonin EV, Wolf YI. Evolution of microbes and viruses: a paradigm shift in evolutionary biology? Front Cell Infect Microbiol. 2012; 2:119. [PubMed: 22993722]

Kozarewa, I., Armisen, J., Gardner, AF., Slatko, BE., Hendrickson, CL. Overview of Target Enrichment Strategies. In: Ausubel, Frederick M., et al., editors. Current protocols in molecular biology. Vol. 112. 2015. p. 7 21 21-23.

Kvisgaard LK, Hjulsager CK, Fahnoe U, Breum SO, Ait-Ali T, Larsen LE. A fast and robust method for full genome sequencing of Porcine Reproductive and Respiratory Syndrome Virus (PRRSV) Type 1 and Type 2. J Virol Methods. 2013; 193:697–705. [PubMed: 23891870]

Kwok H, Chiang AK. From Conventional to Next Generation Sequencing of Epstein-Barr Virus Genomes. Viruses. 2016; 8:60. [PubMed: 26927157]

Ladner JT, Wiley MR, Beitzel B, Auguste AJ, Dupuis AP 2nd, Lindquist ME, Sibley SD, Kota KP, Fetterer D, Eastwood G, Kimmel D, Prieto K, Guzman H, Aliota MT, Reyes D, Brueggemann EE, St John L, Hyeroba D, Lauck M, Friedrich TC, O'Connor DH, Gestole MC, Cazares LH, Popov VL, Castro-Llanos F, Kochel TJ, Kenny T, White B, Ward MD, Loaiza JR, Goldberg TL, Weaver SC, Kramer LD, Tesh RB, Palacios G. A Multicomponent Animal Virus Isolated from Mosquitoes. Cell host & microbe. 2016; 20:357–367. [PubMed: 27569558]

Lecoq H. Discovery of the first virus, the tobacco mosaic virus: 1892 or 1898? Comptes rendus de l'Academie des sciences. Serie III, Sciences de la vie. 2001; 324:929–933.

Leland DS, Ginocchio CC. Role of cell culture for virus detection in the age of technology. Clin Microbiol Rev. 2007; 20:49–78. [PubMed: 17223623]

Li D, Li Z, Zhou Z, Li Z, Qu X, Xu P, Zhou P, Bo X, Ni M. Direct next-generation sequencing of virus-human mixed samples without pretreatment is favorable to recover virus genome. Biology direct. 2016; 11:3. [PubMed: 26754142]

Li L, Pesavento PA, Leutenegger CM, Estrada M, Coffey LL, Naccache SN, Samayoa E, Chiu C, Qiu J, Wang C, Deng X, Delwart E. A novel bocavirus in canine liver. Virol J. 2013; 10:54. [PubMed: 23402347]

Li M, Zheng Y, Zhao G, Fu S, Wang D, Wang Z, Liang G. Tibet Orbivirus, a novel Orbivirus species isolated from Anopheles maculatus mosquitoes in Tibet, China. Plos One. 2014; 9:e88738. [PubMed: 24533145]

Linnen J, Wages J Jr, Zhang-Keck ZY, Fry KE, Krawczynski KZ, Alter H, Koonin E, Gallagher M, Alter M, Hadziyannis S, Karayiannis P, Fung K, Nakatsuji Y, Shih JW, Young L, Piatak M Jr, Hoover C, Fernandez J, Chen S, Zou JC, Morris T, Hyams KC, Ismay S, Lifson JD, Hess G, Foung SK, Thomas H, Bradley D, Margolis H, Kim JP. Molecular cloning and disease association of hepatitis G virus: a transfusion-transmissible agent. Science. 1996; 271:505–508. [PubMed: 8560265]

Lisitsyn N, Lisitsyn N, Wigler M. Cloning the differences between two complex genomes. Science. 1993; 259:946–951. [PubMed: 8438152]

Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M. Comparison of next-generation sequencing systems. J Biomed Biotechnol. 2012; 2012:251364. [PubMed: 22829749]

Lorusso A, Marcacci M, Ancora M, Mangone I, Leone A, Marini V, Camma C, Savini G. Complete Genome Sequence of Bluetongue Virus Serotype 1 Circulating in Italy, Obtained through a Fast Next-Generation Sequencing Protocol. Genome Announc. 2014:2.

Mager DL, Stoye JP. Mammalian Endogenous Retroviruses. Microbiol Spectr. 2015; 3 MDNA3-0009-2014.

Marascio N, Pavia G, Strazzulla A, Dierckx T, Cuypers L, Vrancken B, Barreca GS, Mirante T, Malanga D, Oliveira DM, Vandamme AM, Torti C, Liberto MC, Foca A. The Sinergie-Umg Study, G. Detection of Natural Resistance-Associated Substitutions by Ion Semiconductor Technology in HCV1b Positive, Direct-Acting Antiviral Agents-Naive Patients. Int J Mol Sci. 2016:17.

Marine R, McCarren C, Vorrasane V, Nasko D, Crowgey E, Polson SW, Wommack KE. Caught in the middle with multiple displacement amplification: the myth of pooling for avoiding multiple displacement amplification bias in a metagenome. Microbiome. 2014; 2:3. [PubMed: 24475755]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Matranga CB, Andersen KG, Winnicki S, Busby M, Gladden AD, Tewhey R, Stremlau M, Berlin A, Gire SK, England E, Moses LM, Mikkelsen TS, Odia I, Ehiane PE, Folarin O, Goba A, Kahn SH, Grant DS, Honko A, Hensley L, Happi C, Garry RF, Malboeuf CM, Birren BW, Gnirke A, Levin JZ, Sabeti PC. Enhanced methods for unbiased deep sequencing of Lassa and Ebola RNA viruses from clinical and biological samples. Genome biology. 2014; 15:519. [PubMed: 25403361]

Matsui SM, Kim JP, Greenberg HB, Su W, Sun Q, Johnson PC, DuPont HL, Oshiro LS, Reyes GR. The isolation and characterization of a Norwalk virus-specific cDNA. J Clin Invest. 1991; 87:1456–1461. [PubMed: 2010555]

McGinn S, Bauer D, Brefort T, Dong L, El-Sagheer A, Elsharawy A, Evans G, Falk-Sorqvist E, Forster M, Fredriksson S, Freeman P, Freitag C, Fritzsche J, Gibson S, Gullberg M, Gut M, Heath S, Heath-Brun I, Heron AJ, Hohlbein J, Ke R, Lancaster O, Le Reste L, Maglia G, Marie R, Mauger F, Mertes F, Mignardi M, Moens L, Oostmeijer J, Out R, Pedersen JN, Persson F, Picaud V, Rotem D, Schracke N, Sengenes J, Stahler PF, Stade B, Stoddart D, Teng X, Veal CD, Zahra N, Bayley H, Beier M, Brown T, Dekker C, Ekstrom B, Flyvbjerg H, Franke A, Guenther S, Kapanidis AN, Kaye J, Kristensen A, Lehrach H, Mangion J, Sauer S, Schyns E, Tost J, van Helvoort JM, van der Zaag PJ, Tegenfeldt JO, Brookes AJ, Mir K, Nilsson M, Willcocks JP, Gut IG. New technologies for DNA analysis - a review of the READNA Project. N Biotechnol. 2016; 33:311–330. [PubMed: 26514324]

Merriman B, Ion Torrent R, Team D, Rothberg JM. Progress in ion torrent semiconductor chip based sequencing. Electrophoresis. 2012; 33:3397–3417. [PubMed: 23208921]

Mokili JL, Rohwer F, Dutilh BE. Metagenomics and future perspectives in virus discovery. Current opinion in virology. 2012; 2:63–77. [PubMed: 22440968]

Muerhoff AS, Leary TP, Desai SM, Mushahwar IK. Amplification and subtraction methods and their application to the discovery of novel human viruses. J Med Virol. 1997; 53:96–103. [PubMed: 9298739]

Muir P, Kammerer U, Korn K, Mulders MN, Poyry T, Weissbrich B, Kandolf R, Cleator GM, van Loon AM. Molecular typing of enteroviruses: current status and future requirements. The European Union Concerted Action on Virus Meningitis and Encephalitis. Clin Microbiol Rev. 1998; 11:202–227. [PubMed: 9457433]

Ndze VN, Esona MD, Achidi EA, Gonsu KH, Doro R, Marton S, Farkas S, Ngeng MB, Ngu AF, Obama-Abena MT, Banyai K. Full genome characterization of human Rotavirus A strains isolated in Cameroon, 2010–2011: diverse combinations of the G and P genes and lack of reassortment of the backbone genes. Infect Genet Evol. 2014; 28:537–560. [PubMed: 25460824]

Nichol ST, Spiropoulou CF, Morzunov S, Rollin PE, Ksiazek TG, Feldmann H, Sanchez A, Childs J, Zaki S, Peters CJ. Genetic identification of a hantavirus associated with an outbreak of acute respiratory illness. Science. 1993; 262:914–917. [PubMed: 8235615]

Nishizawa T, Okamoto H, Konishi K, Yoshizawa H, Miyakawa Y, Mayumi M. A novel DNA virus (TTV) associated with elevated transaminase levels in posttransfusion hepatitis of unknown etiology. Biochemical and biophysical research communications. 1997; 241:92–97. [PubMed: 9405239]

Nougairede A, Bichaud L, Thiberville SD, Ninove L, Zandotti C, de Lamballerie X, Brouqui P, Charrel RN. Isolation of Toscana virus from the cerebrospinal fluid of a man with meningitis in Marseille, France, 2010. Vector Borne Zoonotic Dis. 2013; 13:685–688. [PubMed: 23808972]

Nyaga MM, Stucker KM, Esona MD, Jere KC, Mwinyi B, Shonhai A, Tsolenyanu E, Mulindwa A, Chibumbya JN, Adolfine H, Halpin RA, Roy S, Stockwell TB, Berejena C, Seheri ML, Mwenda JM, Steele AD, Wentworth DE, Mphahlele MJ. Whole-genome analyses of DS-1-like human G2P[4] and G8P[4] rotavirus strains from Eastern, Western and Southern Africa. Virus Genes. 2014; 49:196–207. [PubMed: 24952422]

Palacios G, Druce J, Du L, Tran T, Birch C, Briese T, Conlan S, Quan PL, Hui J, Marshall J, Simons JF, Egholm M, Paddock CD, Shieh WJ, Goldsmith CS, Zaki SR, Catton M, Lipkin WI. A new arenavirus in a cluster of fatal transplant-associated diseases. N Engl J Med. 2008; 358:991–998. [PubMed: 18256387]

Palacios G, Oberste MS. Enteroviruses as agents of emerging infectious diseases. J Neurovirol. 2005; 11:424–433. [PubMed: 16287683]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Palacios G, Quan PL, Jabado OJ, Conlan S, Hirschberg DL, Liu Y, Zhai J, Renwick N, Hui J, Hegyi H, Grolla A, Strong JE, Towner JS, Geisbert TW, Jahrling PB, Buchen-Osmond C, Ellerbrok H, Sanchez-Seco MP, Lussier Y, Formenty P, Nichol MS, Feldmann H, Briese T, Lipkin WI. Panmicrobial oligonucleotide array for diagnosis of infectious diseases. Emerging infectious diseases. 2007; 13:73–81. [PubMed: 17370518]

Phan TG, Dreno B, da Costa AC, Li L, Orlandi P, Deng X, Kapusinszky B, Siqueira J, Knol AC, Halary F, Dantal J, Alexander KA, Pesavento PA, Delwart E. A new protoparvovirus in human fecal samples and cutaneous T cell lymphomas (mycosis fungoides). Virology. 2016; 496:299–305. [PubMed: 27393975]

Phan TG, Li L, O'Ryan MG, Cortes H, Mamani N, Bonkoungou IJ, Wang C, Leutenegger CM, Delwart E. A third gyrovirus species in human faeces. J Gen Virol. 2012; 93:1356–1361. [PubMed: 22422066]

Phan TG, Vo NP, Aronen M, Jartti L, Jartti T, Delwart E. Novel human gammapapillomavirus species in a nasal swab. Genome Announc. 2013; 1:e0002213. [PubMed: 23516180]

Qin XC, Shi M, Tian JH, Lin XD, Gao DY, He JR, Wang JB, Li CX, Kang YJ, Yu B, Zhou DJ, Xu J, Plyusnin A, Holmes EC, Zhang YZ. A tick-borne segmented RNA virus contains genome segments derived from unsegmented viral ancestors. Proceedings of the National Academy of Sciences of the United States of America. 2014; 111:6744–6749. [PubMed: 24753611]

Quan PL, Firth C, Conte JM, Williams SH, Zambrana-Torrelio CM, Anthony SJ, Ellison JA, Gilbert AT, Kuzmin IV, Niezgoda M, Osinubi MO, Recuenco S, Markotter W, Breiman RF, Kalemba L, Malekani J, Lindblade KA, Rostal MK, Ojeda-Flores R, Suzan G, Davis LB, Blau DM, Ogunkoya AB, Alvarez Castillo DA, Moran D, Ngam S, Akaibe D, Agwanda B, Briese T, Epstein JH, Daszak P, Rupprecht CE, Holmes EC, Lipkin WI. Bats are a major natural reservoir for hepaciviruses and pegiviruses. Proc Natl Acad Sci U S A. 2013; 110:8194–8199. [PubMed: 23610427]

Quan PL, Wagner TA, Briese T, Torgerson TR, Hornig M, Tashmukhamedova A, Firth C, Palacios G, Baisre-De-Leon A, Paddock CD, Hutchison SK, Egholm M, Zaki SR, Goldman JE, Ochs HD, Lipkin WI. Astrovirus encephalitis in boy with X-linked agammaglobulinemia. Emerging infectious diseases. 2010; 16:918–925. [PubMed: 20507741]

Reuter G, Boros A, Delwart E, Pankovics P. Novel seadornavirus (family Reoviridae) related to Banna virus in Europe. Archives of virology. 2013; 158:2163–2167. [PubMed: 23624680]

Reuter JA, Spacek DV, Snyder MP. High-throughput sequencing technologies. Mol Cell. 2015; 58:586–597. [PubMed: 26000844]

Rhoads A, Au KF. PacBio Sequencing and Its Applications. Genomics Proteomics Bioinformatics. 2015; 13:278–289. [PubMed: 26542840]

Rosseel T, Ozhelvaci O, Freimanis G, Van Borm S. Evaluation of convenient pretreatment protocols for RNA virus metagenomics in serum and tissue samples. Journal of virological methods. 2015; 222:72–80. [PubMed: 26025457]

Rota PA, Oberste MS, Monroe SS, Nix WA, Campagnoli R, Icenogle JP, Penaranda S, Bankamp B, Maher K, Chen MH, Tong S, Tamin A, Lowe L, Frace M, DeRisi JL, Chen Q, Wang D, Erdman DD, Peret TC, Burns C, Ksiazek TG, Rollin PE, Sanchez A, Liffick S, Holloway B, Limor J, McCaustland K, Olsen-Rasmussen M, Fouchier R, Gunther S, Osterhaus AD, Drosten C, Pallansch MA, Anderson LJ, Bellini WJ. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science. 2003; 300:1394–1399. [PubMed: 12730500]

Sauvage V, Foulongne V, Cheval J, Ar Gouilh M, Pariente K, Dereure O, Manuguerra JC, Richardson J, Lecuit M, Burguiere A, Caro V, Eloit M. Human polyomavirus related to African green monkey lymphotropic polyomavirus. Emerging infectious diseases. 2011; 17:1364–1370. [PubMed: 21801611]

Schleiss MR, McAllister S, Armien AG, Hernandez-Alvarado N, Fernandez-Alarcon C, Zabeli JC, Ramaraj T, Crow JA, McVoy MA. Molecular and biological characterization of a new isolate of guinea pig cytomegalovirus. Viruses. 2014; 6:448–475. [PubMed: 24473341]

Simons JN, Leary TP, Dawson GJ, Pilot-Matias TJ, Muerhoff AS, Schlauder GG, Desai SM, Mushahwar IK. Isolation of novel virus-like sequences associated with human hepatitis. Nat Med. 1995a; 1:564–569. [PubMed: 7585124]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Simons JN, Pilot-Matias TJ, Leary TP, Dawson GJ, Desai SM, Schlauder GG, Muerhoff AS, Erker JC, Buijk SL, Chalmers ML, et al. Identification of two flavivirus-like genomes in the GB hepatitis agent. Proceedings of the National Academy of Sciences of the United States of America. 1995b; 92:3401–3405. [PubMed: 7724574]

Solonenko SA, Ignacio-Espinoza JC, Alberti A, Cruaud C, Hallam S, Konstantinidis K, Tyson G, Wincker P, Sullivan MB. Sequencing platform and library preparation choices impact viral metagenomes. BMC genomics. 2013; 14:320. [PubMed: 23663384]

Stenglein MD, Sanders C, Kistler AL, Ruby JG, Franco JY, Reavill DR, Dunker F, Derisi JL. Identification, characterization, and in vitro culture of highly divergent arenaviruses from boa constrictors and annulated tree boas: candidate etiological agents for snake inclusion body disease. mBio. 2012; 3:e00180–00112. [PubMed: 22893382]

Steyer A, Gutierrez-Aguire I, Kolenc M, Koren S, Kutnjak D, Pokorn M, Poljsak-Prijatelj M, Racki N, Ravnikar M, Sagadin M, Fratnik Steyer A, Toplak N. High similarity of novel orthoreovirus detected in a child hospitalized with acute gastroenteritis to mammalian orthoreoviruses found in bats in Europe. J Clin Microbiol. 2013; 51:3818–3825. [PubMed: 24025904]

Thurber RV, Haynes M, Breitbart M, Wegley L, Rohwer F. Laboratory procedures to generate viral metagenomes. Nature protocols. 2009; 4:470–483. [PubMed: 19300441]

Tokarz R, Williams SH, Sameroff S, Sanchez Leon M, Jain K, Lipkin WI. Virome analysis of Amblyomma americanum, Dermacentor variabilis, and Ixodes scapularis ticks reveals novel highly divergent vertebrate and invertebrate viruses. J Virol. 2014; 88:11480–11492. [PubMed: 25056893]

Tombacz D, Csabai Z, Olah P, Havelda Z, Sharon D, Snyder M, Boldogkoi Z. Characterization of novel transcripts in pseudorabies virus. Viruses. 2015; 7:2727–2744. [PubMed: 26008709]

Tombacz D, Sharon D, Olah P, Csabai Z, Snyder M, Boldogkoi Z. Strain Kaplan of Pseudorabies Virus Genome Sequenced by PacBio Single-Molecule Real-Time Sequencing Technology. Genome Announc. 2014:2.

Van den Hoecke S, Verhelst J, Vuylsteke M, Saelens X. Analysis of the genetic diversity of influenza A viruses using next-generation DNA sequencing. BMC genomics. 2015; 16:79. [PubMed: 25758772]

van der Heijden M, de Vries M, van Steenbeek FG, Favier RP, Deijs M, Brinkhof B, Rothuizen J, van der Hoek L, Penning LC. Sequence-independent VIDISCA-454 technique to discover new viruses in canine livers. J Virol Methods. 2012; 185:152–155. [PubMed: 22664180]

van der Hoek L, Pyrc K, Jebbink MF, Vermeulen-Oost W, Berkhout RJ, Wolthers KC, Wertheim-van Dillen PM, Kaandorp J, Spaargaren J, Berkhout B. Identification of a new human coronavirus. Nat Med. 2004; 10:368–373. [PubMed: 15034574]

Victoria JG, Kapoor A, Dupuis K, Schnurr DP, Delwart EL. Rapid identification of known and new RNA viruses from animal tissues. PLoS pathogens. 2008; 4:e1000163. [PubMed: 18818738]

Wang D, Coscoy L, Zylberberg M, Avila PC, Boushey HA, Ganem D, DeRisi JL. Microarray-based detection and genotyping of viral pathogens. Proceedings of the National Academy of Sciences of the United States of America. 2002; 99:15687–15692. [PubMed: 12429852]

Wittmann J, Dreiseikelmann B, Rohde M, Meier-Kolthoff JP, Bunk B, Rohde C. First genome sequences of Achromobacter phages reveal new members of the N4 family. Virol J. 2014; 11:14. [PubMed: 24468270]

Wylie TN, Wylie KM, Herter BN, Storch GA. Enhanced virome sequencing using targeted sequence capture. Genome Res. 2015; 25:1910–1920. [PubMed: 26395152]

Xia H, Hu C, Zhang D, Tang S, Zhang Z, Kou Z, Fan Z, Bente D, Zeng C, Li T. Metagenomic profile of the viral communities in Rhipicephalus spp. ticks from Yunnan, China. Plos One. 2015; 10:e0121609. [PubMed: 25799057]

Xu B, Liu L, Huang X, Ma H, Zhang Y, Du Y, Wang P, Tang X, Wang H, Kang K, Zhang S, Zhao G, Wu W, Yang Y, Chen H, Mu F, Chen W. Metagenomic analysis of fever, thrombocytopenia and leukopenia syndrome (FTLS) in Henan Province, China: discovery of a new bunyavirus. PLoS Pathog. 2011; 7:e1002369. [PubMed: 22114553]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Yan L, Zhang H, Ma H, Liu D, Li W, Kang Y, Yang R, Wang J, He G, Xie X, Wang H, Wei L, Lu Z, Shao Q, Chen H. Deep sequencing of hepatitis B virus basal core promoter and precore mutants in HBeAg-positive chronic hepatitis B patients. Sci Rep. 2015; 5:17950. [PubMed: 26647737]

Yin Y, Jiang L, Fang D, Jiang L, Zhou J. Differentially expressed genes of human microvascular endothelial cells in response to anti-dengue virus NS1 antibodies by suppression subtractive hybridization. Viral immunology. 2013; 26:185–191. [PubMed: 23697814]

Yu JM, Li JS, Ao YY, Duan ZJ. Detection of novel viruses in porcine fecal samples from China. Virol J. 2013; 10:39. [PubMed: 23363449]

Yu JM, Li LL, Zhang CY, Lu S, Ao YY, Gao HC, Xie ZP, Xie GC, Sun XM, Pang LL, Xu JG, Lipkin WI, Duan ZJ. A novel hepatovirus identified in wild woodchuck Marmota himalayana. Sci Rep. 2016; 6:22361. [PubMed: 26924426]

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

HIGHLIGHTS

- Description of virus enrichment techniques for metagenomics based virome analysis.
- Usefulness of recently developed virome capture sequencing techniques.
- Perspective on negative and positive selection approaches for virome analysis.

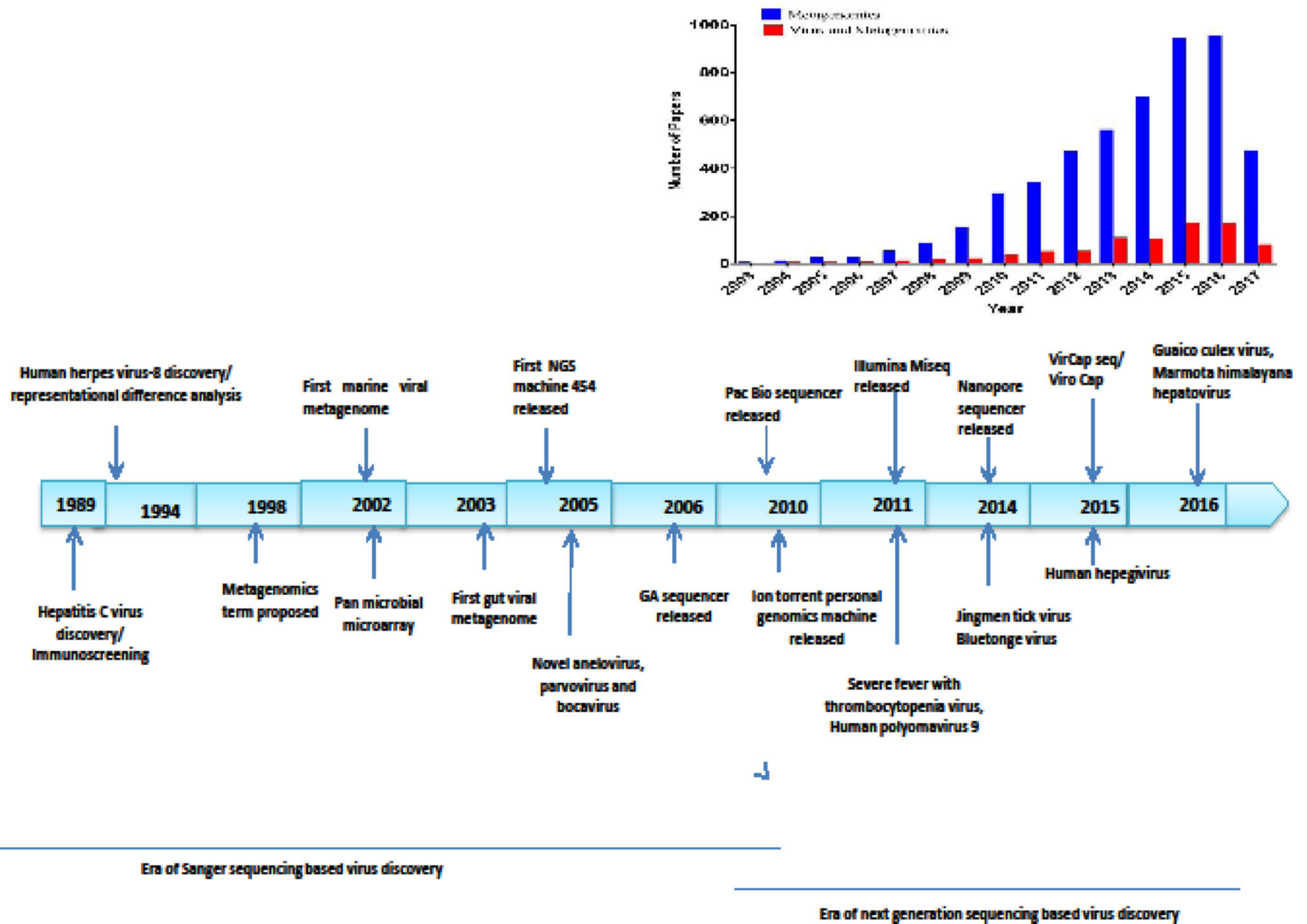

Figure 1.

Overview of important landmark virus discoveries using selective sequencing approach, next generation sequencers released year and the number of published papers on metagenomics, virus and metagenomics.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

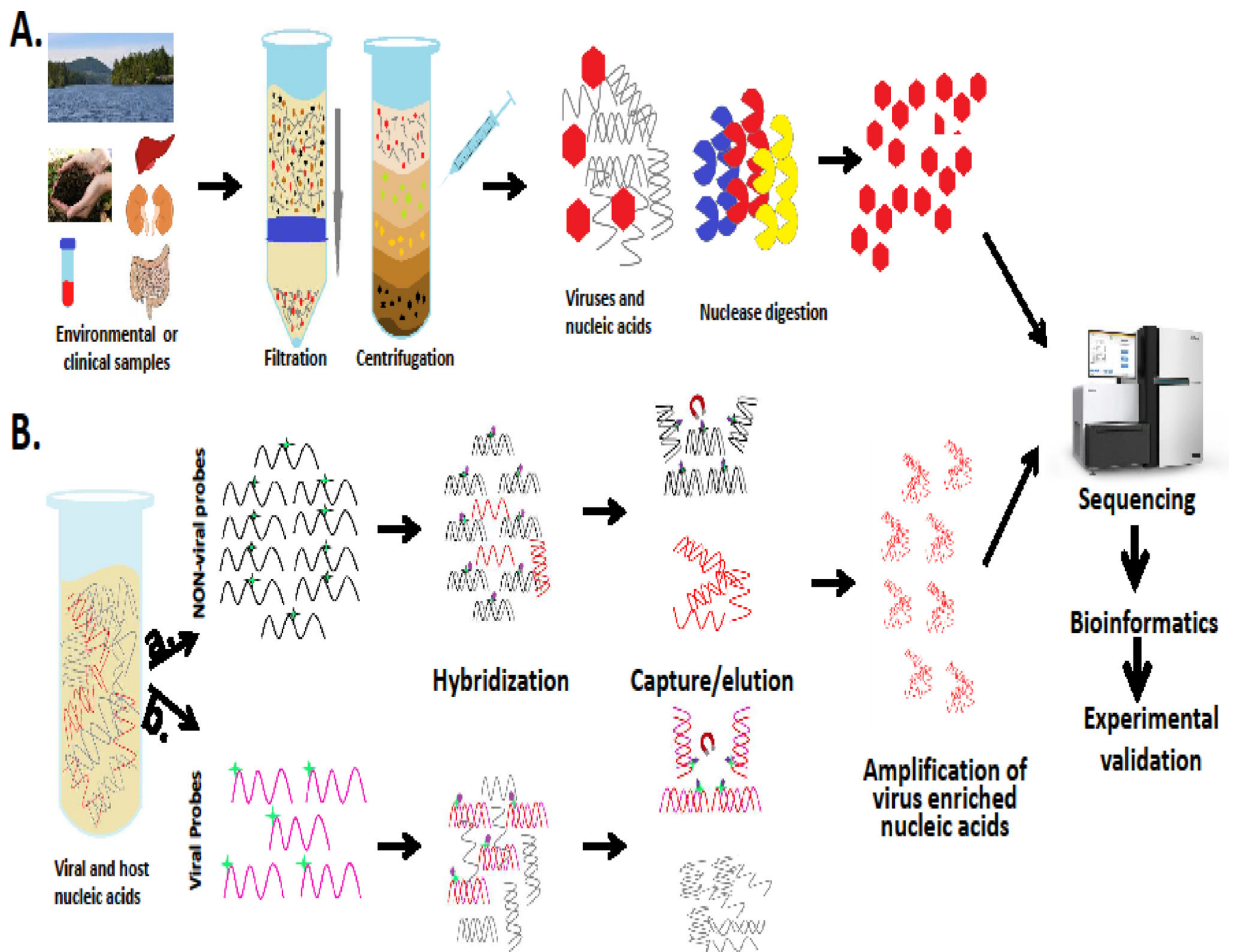

Figure-2.

Experimental outline of virus enrichment and selective sequencing approaches. (A). Details of the pre-extraction virus enrichment based selective sequencing. Briefly, samples are filtered and/or centrifuged before treatment with nucleases (DNAse, RNAse, Benzonase etc.), followed by their amplification and sequencing. (B). Outline of negative (non-viral probes) and positive (viral probes) selection method using hybridization based capture and selective enrichment of virus nucleic acids, followed by amplification and sequencing. Sequencing data analysis requires bioinformatics based virus discovery or virome analysis and experimental validation of results.