MINI-REVIEW

Potential role of viral metagenomics as a surveillance tool for the early
detection of emerging novel pathogens

Hareem Mohsin1 · Azka Asif2 · Minhaj Fatima3 · Yasir Rehman4

Received: 4 August 2020 / Revised: 1 October 2020 / Accepted: 19 October 2020 / Published online: 11 November 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract
Since the early times, human beings have always been faced with deadly microbial infections, both bacterial and viral. Pathogens such as viruses are always evolving owing to the processes of antigenic shift and drift. Such viral evolution results in the emergence of new types and serovars that prove deadly for humans-like influenza pandemics, severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS). The pandemic of novel coronavirus SARS-CoV-2 is the recent example. It has resulted in a great loss of human lives and a serious burden on economy across the globe. To counter such situations, a system should exist for the early detection of emerging viral pathogens. This will help prevent possible outbreaks and save human lives. Most of such deadly novel viruses and viral outbreaks are known to be originated from animal hosts. Regular monitoring of potential hot spots of such emerging microbes, such as zoos and animal markets, through metagenomics could help assess the presence of new viruses and pathogens. In this review, we focus on the potential of viral metagenomics and propose a surveillance system based on it for the early detection and hence prevention of such emerging viral infections.

Keywords  Viral outbreaks · Metagenomics · Surveillance · Public health · Infections · Pandemics

Introduction
Human race has faced microbial infections from its onset. Both viral and bacterial pathogens are always ready to initiate infections that sometimes spread like wildfire resulting in horrifying outbreaks (Christou 2011). Such outbreaks, endemic or pandemic, have resulted in human death toll over thousand folds. Though the immune system of human beings, supported with vaccination, is good enough to fight off the invading pathogens; however, pathogens continue to evolve (Read and Mackinnon 2008). The antigenic drift and antigenic shift happening due to mutations change the pathogens in a way that they are rendered unrecognizable by the host immune system (Both et al. 1983). This helps them evade the host immune system and cause infections. This is most important in case of viral infections such as flu and at times can be very difficult to control and cure. Early detection of such new viruses is of utmost importance so that their spread could be controlled. The recent outbreak of a new coronavirus, severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2), is one such example. Most of the deadly viral outbreaks are known to have been originated from animals, especially bats, and the same is true for SARS-CoV-2 (Andersen et al. 2020). The microbiome of an organism in general, and virome in particular, thus holds great importance, and it is where the new viruses originate. Routine viral analysis includes cell culturing, polymerase chain reaction-based in vitro DNA amplification, serological and immunological testing, hybridization, histology and electron microscopy (Rabenau et al. 2007). Though these methods can be very specific and at times very robust, however, their detection range is limited since many of
these methods can be biased and are based on previous knowledge. Microbiome and virome can be best studied via metagenomics (Wommack et al. 2012), which has grown by leaps and bounds, courtesy of the next-generation sequencing technologies. This review discusses the potential of metagenomics for the early detection of such new viruses. This could help in prevention, and possible cure, of such deadly viral outbreaks.

Genetic basis of viral antigenic shift and drift

As discussed before, antigenic drift and antigenic shift cause changes in the antigenic properties of a pathogen, rendering them unrecognizable by the host immune. Thus, the pathogens survive the immune system of the host and are able to cause infection. Antigenic drift is a small change in the antigenic properties due to minor mutations, whereas antigenic shift is a change owing to the re-assortment process which results in the emergence of anew viral subtypes having a combination of antigenic properties of the two different parental viral genomes co-infecting a host cell. Influenza virus is the classical example of antigenic shifts which had led to several influenza pandemics. The change in the antigenic type of hemagglutinin (HA) and neuraminidase (NA) contributes in the escape of the influenza virus from host’s immunological memory. This not only causes infection in an individual but also vanquishes the herd immunity (Cann 2001).

Emergence of viral outbreaks

The extensive studies of emergence of viral pandemics started by the work of Andrewes (1967), Kilbourne (1968) and Webster and Laver (1971). They studied the antigenic diversity of Influenza A virus strains and proposed that the viral strains enter the nonresistant human population via human–animal interaction in close proximity. Same phenomenon was also observed in the pandemics caused by Asian influenza virus A in China (1957) (McDonald 1958) and Hong Kong (1968) (Ha et al. 2003; Laver et al. 1980). Such changes did not arise due to mutational events like in antigenic drift but due to the antigenic shift resulting in the production of antigenic hybrids (re-assortants) followed by the transmission of the new viral subtypes to humans via animals. On the contrary, the strain B of influenza virus has never been involved in causing pandemics because of having no host other than humans; thus, they thrive via antigenic drift in humans rather than the antigenic shift (Gibbs et al. 2005).

Viral outbreaks in history

Emergence of viral species, owing to antigenic shift and drift, continues globally with increasing human–animal interactions. The deadliest pandemics that appeared in antiquity include the plague of Justinian (bubonic plague) in sixth century and the Black Death (1347–1351) spreading across Europe and Asia. Both pandemics erupted from zoonotic sources, i.e. rats/fleas and causative agent was bacteria *Yersinia pestis* (Sohysiak 2008; Wagner et al. 2014). Affected individuals were treated by self-isolation and by the use of certain chemicals such as alkalooids (Horgan 2014). Other horrific pandemics include third (1852–1860) and sixth (1910–1911) cholera pandemic (a total of seven cholera pandemics occurred) in Asia, North America, Europe, Russia and Africa. The flu pandemic (1918) targeted more strikingly the young individuals with healthy immune system as compared to the juvenile and older ones with weak immune system (Cunha 2004). The notorious influenza virus and its deadly strains also hit the world with deadly outbreaks such as Russian flu (1889–1890), Spanish flu (1918–1919), Asian flu (1957–1958), Hong Kong flu (1968–1970), Swine flu (2009–2010), severe acute respiratory syndrome (SARS) (2002–2003) and Middle East respiratory syndrome (MERS) (2015–Present). SARS and MERS are caused by coronaviruses, via bats as vectors. Moreover, bats are also considered as potential vectors for the spread of COVID-19 (Banerjee et al. 2019). All these outbreaks, especially SARS and MERS, were dealt by improving sanitation and quarantining methods. Other non-pharmaceutical methods like self-isolation, maintenance of good hygiene, disinfection and restricting public gatherings were also practiced (Hossain et al. 2020). SARS along with above-mentioned methods was also treated with oxygen therapies, i.e. invasive and noninvasive ventilation (Hung et al. 2004). Table 1 shows some of the world’s worst outbreaks, the death toll and fatality/mortality rate.

The novel coronavirus SARS-CoV-2 attack

Humans, with increasing human–animal interaction, have unknowingly headed toward viral outbreaks that can easily turn into pandemics due to increased globalization. Not every outbreak reached the pandemic level as the novel coronavirus SARS-CoV-2 has. This “mass killer” has sent chills down the spine of the populace causing COVID-19. The outbreak started in December 2019 from Wuhan City, Hubei province in China, and spread like wildfire engulfing more than 210 countries. A study was conducted by Chen et al. (2019) in which pair-end sequencing with
Illumina MiSeq platform revealed that the new coronavirus is highly related to the bat-associated coronavirus. The virus was isolated from *Rhinolophus* spp. (horse-shoe bats) which are considered as exotic food. The possible origin of this virus is considered to be bats which were also related with the origin of SARS and MERS as well (Cui et al. 2019). Potential vectors for coronavirus include bats, pigs, rats and pangolins; however, the exact zoonotic vector is still not clear (LePan 2020). Another study suggests arthropod (cockroaches) vectors for coronavirus due to unhygienic conditions (Dehghani and Kassiri 2019). The pandemic is still in its full gear, and drugs like hydroxychloroquine, azithromycin and the antiviral drug remdesivir are being tested for treatment. The combination of hydroxychloroquine and azithromycin being used at the start has now proven to provide no benefit. World Health Organization (WHO) has suspended this combination as it caused an increase in cardiovascular cases in a study conducted by Mehra et al. (2020). Another research conducted by Sun (2020) revealed that remdesivir showed efficiency in vitro and in vivo in laboratory animals, but when administered by IV method, the efficacy was limited (Sun 2020). A recent drug named as dexamethasone is also considered as a potential treatment, but certain limitations are still there. The drug is effective in case of patients which require mechanical ventilation, but The National Institute of Health (NIH) is against the use of dexamethasone when the patient is not in need of supplemental oxygen (Horby et al. 2020; NIH 2020).

**Role of animals in viral outbreaks**

The excessive interference of anthropogenic activities in rich diverse ecosystems have provided more chances of human–animal interactions. This leads to the increased transmission of infectious and contagious diseases from animals to humans and then among humans as well (Fauci 2005). The animals, in most cases, act as reservoirs for viral species and become a major reason for the viral outbreaks. As discussed before, viruses can recombine in these animal hosts and can result in new viral subtypes.

As discussed above, animals can serve as reservoirs for many viruses. Birds also act as a reservoir for many viruses and spread diseases while migrating annually over long distances. Examples include West Nile virus in America (Rizzoli et al. 2015) and Avian influenza virus (Kalthoff et al. 2010). Primates, i.e. monkeys are also involved in viral disease spread as they are known to be reservoirs for dengue virus and HIV. Rodents, also being notorious in spreading diseases, are involved in the transmission of Hantaviruses (Strauss and Strauss 2007).

**Emergence and re-emergence of viral zoonosis**

The pandemics usually occur due to the emergence of new infectious agents. As discussed above, antigenic shifts play a major role in this process. The result of such antigenic shifts have been witnessed by the world in form of 1918 (H1N1), 1957 (H2N2), 1968 (H3N2) and 1977 (H1N1) pandemics while re-emergence in the form of epidemics owes to the antigenic drift events (Alexander and Brown 2000). Considering SARS, MERS and the current outbreak of COVID-19, all share common zoonotic source, i.e. bats (Cunningham et al. 2017; Cupertino et al. 2020). These viruses perhaps have also originated due to antigenic shifts and drifts. Thus, it can be said that the re-emergence of these viral outbreaks occurred with evolved antigenic properties.

**Viral metagenomics**

The use of metagenomics analysis for the detection of pathogens has been proposed before as well (Miller et al. 2013). With the advent of high-throughput DNA sequencing technologies, there has been a staggering increase in
metagenomic studies. The culture-independent high-throughput techniques, being holistic in nature, are able to sequence the complete or near-complete genomes of the viral pathogens (Miller et al. 2013). Total DNA is extracted from diverse sites and is sequenced followed by assembly and annotation of the genomes. This can be done for a specific enzyme class, pathogens, novel metabolites or organisms, etc. Target specific amplification and sequencing can also be achieved via this process (Robinson et al. 2013). Other approaches such as microarray techniques and deep sequencing strategies have been promising methods as well in recent years for the detection of novel viral strains. It has been reported by Greninger et al. (2010) that despite the genetic mechanisms of antigenic shift and drift in influenza virus, metagenomics has showed promising results in detection of viral species. Influenza virus, composed of segmented genome, can be typed using a pan-viral microarray technique known as Virochip, to track the production of viral re-assortants and to identify other co-infections evolved with influenza. On the other hand, NGS or re-sequencing microarrays (RMA) (Dacheux et al. 2010) can play an important role in determining the viral mutations conferring virulence and resistance against the antivirals. Kuroda et al. (2010) have reported the identification of quasispecies of 2009 pandemic of H1N1 influenza A virus by employing deep sequencing strategies (Greninger et al. 2010).

Metagenomic analysis of animal viromes

The recent epidemics having devastating impacts on the human health are some of the most typical examples of outbreaks in which zoonotic sources are involved. Therefore, for outbreak surveillance, it is necessary to deeply understand and comprehend metagenomes of animals for any viral segment which can be a source of viral outbreak. Literature reports that by metagenomic approaches, avian viruses have been identified in Australian Muscovy ducks and Pacific black ducks. Study reveals some unrelated viruses as well (Vibin et al. 2018). Pangolins in China are reported to be associated with the transmission of SARS-CoV-2 working as an intermediate host. The coronavirus was identified in the pangolin specie Manis javanica in Southern China via metagenome sequencing (high-throughput sequencing of RNA) (Lam et al. 2020). Novel sequences from taxonomic groups of picornaviruses, bocaviruses, sapoviruses and single-stranded DNA viruses were isolated from rat viromes. These novel sequences showed less homology with already present sequences in the databases (Sachsenröder et al. 2014). As bats have been reported to be the reservoir for many deadly viruses, our next discussion is based on bat viromes only.

Metagenomics of bat viromes

Coronavirus (SARS-CoV-2), belonging to a new evolutionary branch, is reported to share a huge percentage of similar gene sequences and the behavior pattern with SARS-CoV (causative agent of SARS) (Chan et al. 2020). Another study reports the identification of novel coronavirus from a bat specie Rhinolophus affinis (Lam et al. 2020). Even before these deadly outbreaks, Marburg and Nipah viruses were also of zoonotic origin and transmitted via bats directly to humans (Amman et al. 2012; Rahman et al. 2012). Bat species from the family Vespertilionidae, in the region of South Africa, is reported to inhabit the roofs of houses thus having direct interaction with humans (Geldenhuys et al. 2018).

Novel viruses reported form metagenomic analysis of bats

Genus Neoromicia of bats was studied by Geldenhuys et al. (2018), and metagenomic analysis was performed using Illumina sequencing (NextSeq500). This showed the presence of novel species of family Circoviridae whose members are largely unknown and are considered to be transmitted via zoonosis. A novel Cyclovirus (NeoCycloV-1) was also identified from the bat genome. Sequences of other viral families were also identified like Adenoviridae, Coronaviridae, Herpesviridae, Parvoviridae, Papillomaviridae, Phenuiviridae and Picornaviridae.

Another study reported the presence of new species of alphacoronavirus and betacoronavirus from the bats inhabiting Myanmar which had no similarities with already-known coronaviruses (Valitutto et al. 2020). Not only coronaviruses, but other novel strains like Sapovirus genus (SaVs) (family Calciviridae) were also discovered to be associated with pigs. In 2017, metagenomes of two bats Eidolon helvum and Epomophorous gambianus were sequenced. Upon analysis, genomes sequences of six SaVs were recognized in Eidolon helvum. On the phylogenetic and taxonomic basis, SaVs obtained from the bat virome were recognized as novel genotypic strains (Yinda et al. 2017). Another pathogenic novel papillomavirus, named as Miniopterus schreibersii papillomavirus type 1 (MscPV1), was identified from bat virome (Miniopterus schreibersii) via high-throughput sequencing method. This novel genome showed identity with Eredhizon dorsatum papillomavirus (EdPV1). The pattern of similarities hinted at divergent evolution (Tse et al. 2012). Metagenomic analysis of bat viromes has also been helpful in identifying novel sequences of viruses belonging to genus Alphapapillomavirus, Betaretrovirus, Alpharetrovirus, Varicellovirus, Cyprindivirus, Chlorovirus and Cucumovirus. These novel sequences were not previously present in any existing database (Zheng et al. 2018). The novel species
of Rotavirus J was also identified from bat virome from Serbia through the metagenomics approach (Bányai et al. 2017).

**Viral metagenomics-a potential tool for surveillance**

Viral metagenomics holds a lot of potential as a surveillance tool for the detection and identification of new viruses (Dacheux et al. 2014; Wu et al. 2016). A few research reports have also confirmed this potential of viral metagenomics (Brinkmann et al. 2016). In order to identify novel viral pathogens, the surveillance systems must evolve continuously for detection and identification as the pathogens are ever evolving. The surveillance system should be strong enough to detect such emerging infectious agents at potential hot spots such as animal markets, zoos, etc., where there are enhanced chances of the emergence of new viral species leading to potential outbreaks. Continuous monitoring of animal metagenomes is of utmost importance because more than 60% of novel viruses are of zoonotic origin (Brucker 2020; Lloyd-Smith et al. 2009). As is the case of coronavirus SARS-CoV-2, the animal market serving bats as exotic food item was the epicenter of COVID-19 (Jalava 2020).

Gardy and Loman (2018) proposed an amalgam of surveillance and early outbreak response system that was based on the phylodynamic approach. The proposed system was limited to the surveillance of potential hot spots already identified by Jones et al. (2008). They proposed that any place among the spots with animal habitats interacting with humans would be checked on routine basis by sample collection of animal tissues and in case of bats, the bats guano. An early outbreak response could be generated with such routine screenings, which would be followed by a signal on mobile applications using satellite data by geographic information system (GIS) analysis. For instance, a particular area where people are using a particular medicine for treatment will be checked by the health care team after an alert is sent. Sequences would be obtained via metagenomic analysis and uploaded to the public repository and compared with already-sequenced reference genomes and metadata to identify the presence of any novel sequence.

**Proposed surveillance system**

We propose that a combination of metagenomics, and epidemiology can be an ideal surveillance system for advanced and digital detection of emerging viral pathogens. NGS has made possible the robust and high-throughput sequencing of metagenomes/genomes. Third-generation sequencing techniques such as single-molecule real time sequencing and nanopore sequencing has made possible better gap closure due to long reads along with minimum amplification-based errors. The nanopore MinION sequencing technology is even portable and has been used in space shuttles as well (Castro-Wallace et al. 2017). We suggest that the surveillance system should not be limited to already-identified hot spots. All kinds of potential areas like zoos, unchecked animal breeding centers, including places where wild animals notorious for outbreaks in the past like bats, pigs, etc., are bred or kept, animal markets, and areas with high animal–human interactions should be included in routine screening. In addition to metagenome sequencing, the epidemiological data about such hot spots should also be recorded such as population density, climate conditions and other related aspects. Once any novel pathogenic sequence is found in any place, all the nearby places with similar parameters should be rigorously scrutinized immediately in order to keep a strong check on any possible spread outbreak. This high-throughput and robust sequencing coupled with epidemiology will form a potentially omnifarious system. This system can proactively detect the emerging pathogens.

In order for any such surveillance system to be applicable, we first need to identify signature sequences of different classes of viruses and develop a comprehensive database. This database should be upgraded and expanded on regular basis. In addition to such signature sequences, other salient features of viral genome sequences that can help detect any possible divergent strain should also be present in this database.

The first step should involve the isolation and sequencing of metagenomic DNA from the potential hot spots. Samples of animal tissues and excreta should be used for metagenomic DNA isolation. Animals from controlled breeding centers can be used as controls. Subsequent sequencing, assembly and annotation of the samples will unveil some salient features such as mutations, variations (insertions and deletions) or any transposable elements. Presence of any viral signature sequences and other sequences hallmarks should be assessed. The metagenome sequences should also be compared with existing databases. This will result in phylogenetic lineages that are cognizant of the divergent or convergent evolution. This will help in the early detection of novel and evolving viruses, and thus, preventive measures can be taken in time to prevent any possible outbreak.

In addition, we also suggest using population genetic markers and population dynamics to ascertain the least and most vulnerable regions for a particular family of viruses. For this purpose, information from past outbreaks can be a useful tool. This will be an extension to the surveillance system because genetic markers are from the population and will be used to study which regions are likely to be most affected by a certain viral outbreak. For instance, a SNP prevalence in a certain population renders them more vulnerable to a particular type of disease. So, we suggest...
the identification of genetic markers particular for attacks from a certain class of viruses as well.

For successful targeting of viral genomes residing in animal hosts, tools must be robust and high-throughput enough to produce data with accuracy and efficiency. Table 2 represents some of the tools for the analysis of viral sequences and some of the viral databases.

### Conclusion

In the wake of emergence and re-emergence of viral outbreaks, human beings have suffered a lot. Antigenic shifts and drifts are the key players in viral outbreaks. There is a dire need of a surveillance system for the early detection of emerging viral pathogens as well as regulation of potential hot spots. This will help take necessary actions in time to prevent outbreaks that can turn into epidemics and pandemics. This will save a lot of lives and the negative impacts on the economy due to such outbreaks.

### References

Alexander D, Brown I (2000) Recent zoonoses caused by influenza A viruses. Rev Sci Tech 19:197–225 ((Office International des Epizooties))

Amman BR et al (2012) Seasonal pulses of Marburg virus circulation in juvenile *Roussettus aegyptiacus* bats coincide with periods of increased risk of human infection. PLoS Pathog 8:e1002877

Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF (2020) The proximal origin of SARS-CoV-2. Nat Med 26(4):450–452. https://doi.org/10.1038/s41591-020-0820-9

Andrewes CH (1967) The natural history of viruses. Nat Hist Viruses

Banerjee A, Kulcsar K, Misra V, Frieman M, Mossman K (2019) Bats and coronaviruses. Viruses 11:41. https://doi.org/10.3390/v11010041

Bányai K et al (2017) Candidate new rotavirus species in Schreiber’s bats, Serbia. Infect Genet Evol 48:19–26

Both GW, Sleigh M, Cox N, Kendal A (1983) Antigenic drift in influenza virus H3 hemagglutinin from 1968 to 1980: multiple evolutionary pathways and sequential amino acid changes at key antigenic sites. J Virol 48:52–60

Brinkmann A, Nitsche A, Kohli C (2016) Viral metagenomics on blood-feeding arthropods as a tool for human disease surveillance. Int J Mol Sci 17:1743
Webster R, Laver W (1971) Antigenic variation in influenza virus. Biology and chemistry. Prog Med Virol 13:271
Wommack KE et al (2012) VIROME: a standard operating procedure for analysis of viral metagenome sequences. Stand Genomic Sci 6:421–433
Wu Z et al (2016) Deciphering the bat virome catalog to better understand the ecological diversity of bat viruses and the bat origin of emerging infectious diseases. ISME J 10:609–620
Yinda CK et al (2017) Novel highly divergent sapoviruses detected by metagenomics analysis in straw-colored fruit bats in Cameroon: divergent bat sapoviruses. Emerg Microbes Infect 6:1–7
Zheng X-y et al (2018) Viral metagenomics of six bat species in close contact with humans in southern China. Arch Virol 163:73–88

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.