# New pathogen discovery

## Balachandran Ravindran*

Institute of Life Sciences, Bhubaneswar 751 023, India

Epidemics and pandemics caused by infectious pathogens have been a major cause of concern from the public health and economic point of view. The general approach in the past has been to investigate outbreaks in the shortest possible time using microbiological tools and initiate intervention strategies. However significant progress made in nucleic acid sequencing technologies and development of bioinformatics tools have enabled rapid diagnosis of infectious pathogens in human and animal communities. These tools have also enabled ambitious programme the world over to discover newer pathogens and more significantly predict future outbreaks – something the global science community had not succeeded in the past.

**Keywords:** Bioinformatics tools, pathogens, viral infections.

JOSHUA Lederberg famously said 'The single biggest threat to man's continued dominance on this planet is the virus'. The statement possibly was made far ahead of its time, since it is more true in contemporary times than ever before. Recorded history suggests empires and civilizations have been obliterated due to outbreak of infectious diseases – the inflenza pandemic in Europe less than a century ago is a testament to this statement when nearly 1.5 million people died in a short span of time changing the course of history of Europe in 20th century. Global economy, trade and human movement between countries crash easily by sudden outbreak of infectious diseases largely of viral origin. Investigating and diagnosing identity of pathogens that cause epidemics and pandemics is an expensive exercise and practiced largely in resource-rich countries. Before the advent of currently available molecular tools, conventional microbiological and/or virological tools were the mainstay for pathogen detection in infected human and animal populations. The century-old Koch's postulates has been the bedrock to declare a microbe as a pathogen. In short, it stated isolation of a microbe from infected individuals, absence of the microbe in healthy individuals, its successful growth *in vitro* and the ability of the isolated microbe to cause disease in experimental animal models. Although not originally included by Robert Koch, the ability to isolate the microbe from infected experimental animal model was included as an additional requirement to attribute the

microbe to be the causative agent of an infectious disease. There have been several exceptions to this basic rule including the microbe *Mycobacterium leprae*[1]. Molecular and computational tools developed in recent years have made identification of epidemics caused by infectious pathogens in animal and human communities very rapid and accurate. Thus a SARS or Ebola outbreak gets tracked rapidly by global efforts and gets contained in spite of increased movement of infected population due to globally connected network of people. Extensive global traffic to some extent tends to circumvent the advantages of development of tools for rapid detection of an epidemic. It also needs to be recognized however that no pandemic of infectious disease has ever been predicted so far in spite of significant progress made in recent years using molecular and computational tools[2]. The future challenge in this sphere of activity will be in not only identifying novel pathogens but also in predicting possible outbreak of an epidemic or pandemic in human and animal communities. This can be expected to be more successful in comparison to predicting earthquakes. This short review summarizes recent developments of the last few years which have transformed our ability to discover novel pathogens. The review is 'virus centric' given the high frequency and possibilities of mutations of viruses to cause novel diseases in comparison to bacteria, fungi or higher organisms.

Contrary to the common perception, viruses constitute the largest and most genetically diverse source of biomass in the globe. It has been estimated that viruses far outnumber archea and bacteria in the ocean alone and viral pathogens constitute some of them[3]. Many of the recent outbreaks of viral infections in human communities, viz. SARS, MARS, Ebola, HIV/AIDS, etc. have their origin in wildlife or domestic animals[4,5]. Ian Lipkin and colleagues have been in the forefront of discovering novel viruses over the last few years – as opposed to conventional microbiological methods high-throughput sequencing has revolutionized viral discovery programme[3,6]. Generation of millions/billions of nucleic acid sequences from biological samples coupled with development of algorithms for computational analysis of these sequences has enabled discovery of several novel viral sequences. Repeated sampling of wildlife species such as bats that are known to harbour emerging zoonotic pathogens has resulted in the discovery of as many as 58 viruses. Extrapolation of this to about 5486 described mammalian species suggests that more than 320,000

viruses are waiting to be discovered[3,6]. As these discoveries using molecular and computational tools cannot be verified by conventional methods of *in vitro* culture techniques to satisfy Koch's postulates to qualify them as pathogenic viruses, Ian Lipkin and colleagues have developed a hierarchy of confidence in the relevance of virus discovery, viz. possible causal relationship, probable causal relationship and at the final level confirmed causal relationship. The next step, less expensive and more rapid than NGS sequencing, has been the recent development of chips for viral discovery. A virome capture sequencing platform for vertebrate viruses (VirCapSeq-VERT) has increased the sensitivity of sequence-based virus detection and characterization[7]. About 2 million probes covering members of 207 viral taxa known to infect vertebrates including humans have been recently developed by Ian Lipkin and his associates. This technology has led to 100–10,000 fold increase in viral reads generated using biological samples such as blood, serum and tissue extracts. Even considering that only a small number of such sequences eventually get identified as potential pathogenic viruses that cause diseases in animals and humans, the discovery of absolute number of viruses is very high. Development of web-based protocols has further opened up several opportunities for integrating large sequence database for viral identification and discovery. A programme for monitoring Emerging Infectious Disease (ProMED-mail) has been updating continuously submissions from several investigators. The data are distributed to several thousand subscribers in 185 countries after screening by a panel of experts[3,6]. It is now possible for several investigators to have access to such database for viral identification and discovery. A more recent development is a cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples – designated as SURPI for 'sequence-based ultrarapid pathogen identifications'[8]. Developed by Charles Chiu and colleagues, it can be deployed on both cloud-based and standalone servers. It leverages two aligners for accelerated analyses, SNAP and RAPSearch but several-fold faster in performance. In *fast* mode SURPI detects viruses and bacteria by scanning datasets of 7–500 million reads in 11 min to 5 h. A cloud-based open sourced community tool is a new development that has been validated[9] – Pathosphere.org allows communication, collaboration and sharing of NGS data amongst investigators in academia, industry and government. Although high-end sequence technology has been a major development in the area, issues such as collection of appropriate biological samples, purification of RNA/DNA, storage, transportation, etc. were not given enough emphasis in the past. This is a crucial issue since sequencing platforms need to be in a centralized facility while collection and storage of biological samples for analysis need to be undertaken in human and animal communities – the operational issues while collecting

samples from wildlife will be even more critical. Fortunately these technical issues have begun to be addressed recently, including training personnel for biosafety – long-term storage of RNA is not any more a rate-limiting step in viral discovery programme[10,11].

The key to viral discovery is not as much as anaylsis of biological samples from disease-affected population, which is needed for diagnosis and to initiate intervention. The target population for making viral discovery is essentially asymptomatic carriers of pathogens, who function as reservoirs. Viruses being obligate parasites, often can be found in hosts that do not display symptoms of the disease in animal or human communities. A discovery programme thus needs to target analysis of biological samples from hosts that do not necessarily display clinical manifestations. Several animal species such as bats, camel, hares, birds, pigs, rodents and primates have been reported to be the most common hosts for human viral diseases. Influenza, Nepah, Chikungunya, MARS, Hepatitis E, SARS, Rabies, West Nile, Rift Valley and Hanta virus have originated from one of these animal species[12–15].

Considering the close evolutionary proximity of humans and other primates and their receptors, analysis of very large sequences of viruses from wild primates could lead to predictions about future outbreak in humans by a novel infective virus of primate origin. In a comment in the *Lancet* on anatomy of a pandemic, Peter Daszak[16] borrows an expression made popular during Iraq war – known unknowns are more likely to be discovered than unknown unknowns – predicting an outbreak caused by microbes that do have known close relatives will be far more challenging. Stephen Morse and colleagues and others[17,18] have discussed extensively in a series on zoonoses on prediction of next zoonotic pandemics. Several pandemics such as HIV/AIDS, influenza, severe acute respiratory syndrome have all been caused by viruses that emerged from animals often due to ecological, behavioural and socioeconomic changes. The authors have proposed that new mathematical modelling, communications and informatics technologies have the potential to identify hitherto unreported microbes in other species which could allow prediction of future pandemics. By using database of all known infectious agents during the last several decades they identified global hotspots for emerging infectious diseases – presence or absence of infections in wildlife, static and changing human population density and wildlife diversity were factored-in to predict regions in the globe that are most likely to experience pandemics. The mapping reveals a disturbing scenario of India being one of such global hotspots. Currently the strategy in India like in most other countries has been to essentially study an epidemic or pandemic outbreak after the event has already occurred. The approach is to identify an already known and identified pathogen and take remedial intervention methods. This is clearly insufficient for a geographical area which has been classified as a global

hotspot for future outbreak of pandemics. The Department of Biotechnology is at a threshold of addressing this issue seriously in the North-East region considering the biodiversity of the region sharing borders with several neighbouring countries – wild animals do not recognize borders drawn by countries. The strategy is to use several of the above described technologies to screen animal and human populations with a view to map novel pathogen sequences in biological and environmental samples and analyse them using new emerging computational tools.

1. Byrd, A. L. and Segre, J. L., Adapting Koch's postulates. *Science*, 2016, **351**, 224–226.
2. Lane, R., Andrew Dobson: taking a macroscopic view of zoonoses. *Lancet*, 2012, **380**, 1899.
3. Lipkin, W. I. and Anthony, S. J., Virus hunting. *Virology*, 2015, **479–480**, 194–199.
4. Delwart, E., Animal virus discovery: improving animal health, understanding zoonoses, and opportunities for vaccine development. *Curr. Op. Virol.*, 2012, **2**, 344–352.
5. Karesh, W. B. *et al.*, Ecology of zoonoses: natural and unnatural histories. *Lancet*, 2012, **380**, 1936–1945.
6. Lipkin, W. I. and Firth, C., Viral surveillance and discovery. *Curr. Op. Virol.*, 2013, **3**, 199–204.
7. Briese, T. *et al.*, Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. *mBio*, 2015, **6**, e01491–01415.
8. Naccache, S. N. *et al.*, A cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples. *Genome Res.*, 2014, **24**, 1180–1192.
9. Kilianski, A. *et al.*, Pathosphere.org: pathogen detection and characterization through a web-based, open source informatics platform. *BMC Bioinf.*, 2015, **16**, 416.
10. Duale, N. *et al.*, Long-term storage of blood RNA collected in RNA stabilizing Tempus tubes in a large biobank – evaluation of RNA quality and stability. *BMC Res. Notes*, 2014, **7**, 633.
11. DiEuliis, D., Johnson, K. R., Morse, S. S. and Schindel, D. E., Opinion: specimen collections should have a much bigger role in infectious disease research and response. *Proc. Natl. Acad. Sci. USA*, 2016, **113**, 4–7.
12. Kosoy, M., Khlyap, L., Cosson, J. F. and Morand, S., Aboriginal and invasive rats of genus Rattus as hosts of infectious agents. *Vector Borne Zoonotic Dis.*, 2015, **15**, 3–12.
13. Brook, C. E. and Dobson, A. P., Bats as 'special' reservoirs for emerging zoonotic pathogens. *Trends Microbiol.*, 2015, **23**, 172–180.
14. Mandl, J. N. *et al.*, Reservoir host immune responses to emerging zoonotic viruses. *Cell*, 2015, **160**, 20–35.
15. Quan, P. L. *et al.*, Bats are a major natural reservoir for hepaciviruses and pegiviruses. *Proc. Natl. Acad. Sci. USA*, 2013, **110**, 8194–8199.
16. Daszak, P., Anatomy of a pandemic. *Lancet*, 2012, **380**, 1883–1884.
17. Morse, S. S. *et al.*, Prediction and prevention of the next pandemic zoonosis. *Lancet*, 2012, **380**, 1956–1965.
18. Kilpatrick, A. M. and Randolph, S. E., Drivers, dynamics, and control of emerging vector-borne zoonotic diseases. *Lancet*, 2012, **380**, 1946–1955.