Fruits of Virus Discovery: New Pathogens and New Experimental Models

David Wang

Departments of Molecular Microbiology and Pathology & Immunology
Washington University in St. Louis, School of Medicine
St. Louis, MO 63110

Email: davewang@wustl.edu
ABSTRACT

The application of high-throughput sequencing has lead to a tremendous increase in the rate of discovery of viral sequences. In some instances, novel pathogens have been identified. What has been less well appreciated is that novel virus discoveries in distinct hosts have led to the establishment of unique experimental systems to define host-virus interactions. These new systems have opened new frontiers in the study of fundamental virology and infectious disease.

Next Generation Sequencing (NGS) Catalyzes the Rate of Virus Discovery

In the past decade, new developments in sequencing technology, along with commensurate bioinformatic capability and tools, and their application to interrogate samples for the presence of virus-derived nucleic acids, have lead to a dramatic increase in the rate at which novel virus sequences have been discovered. In single studies of diverse specimen types evidence for literally tens to hundreds of novel viruses can now be generated. This has resulted in a nearly quadrupling of the number of unique virus species in Genbank since 2005 (Figure 1). These observations have demonstrated that the world of viruses is vastly greater than previously recognized. For the past half-century, the predominant viruses that were known and studied were those that could be grown in cell culture systems, and it was only on rare occasions that viruses were identified that could not be cultured (e.g. human norovirus, or hepatitis C). Thus, our perspective on the diversity of viruses has been heavily biased towards those that could be propagated in a very limited set of experimental culture systems. In many regards, this situation parallels the 1990s observations in the bacterial world when consensus 16S ribosomal RNA gene sequencing demonstrated a much greater diversity of bacterial taxa than had been previously identified by culturing.

The goal of this GEM is to describe the most substantive fruits arising from the NGS revolution. Technical aspects of viral discovery by NGS will not be discussed herein. Rather, the
focus will be upon providing exemplars of novel pathogens and novel model systems that have emerged from the first decade of NGS-based viral discovery efforts.

Identification of Novel Pathogens

Many novel viruses have been identified by application of NGS to diseased specimens, thereby generating candidate agents that may play causal roles in the disease in question. Two examples that illustrate new paradigms emerging from viral discovery efforts are presented.

**Merkel Polyomavirus and human cancer.** For decades, there has been controversy as to whether polyomaviruses, which have well defined transforming properties in animal models, play causal roles in any human tumors. In 2008, NGS analysis of specimens collected from Merkel cell carcinoma, a rare tumor that develops more frequently in immunocompromised individuals identified a novel polyomavirus. Merkel polyomavirus was clonally integrated in multiple tumors and subsequent analyses demonstrated the presence of this virus in ~80% of Merkel cell carcinomas tested (9). The accumulated data to date provide the strongest data implicating a polyomavirus in the etiology of a human cancer. Furthermore, detailed analysis of the Merkel genome lead to the identification of a novel open reading frame, termed ALTO (3). While its function is still cryptic, comparative genomic analyses suggest that there is an entire clade of polyomaviruses that encode ALTO. Thus, from this discovery, not only was a novel pathogen identified, but new insights were gleaned into the fundamental virologic and evolutionary properties of polyomaviruses that were not previously recognized.

**Astrovirus VA1/PS.** Astroviruses have traditionally been associated with gastrointestinal disease in humans and animals. Therefore, it was not surprising when application of NGS to stool samples from an unexplained outbreak of gastroenteritis in a Virginia day care center lead to the discovery of a novel astrovirus, Astrovirus VA1, that was present in multiple samples from the outbreak (10). More surprising was the identification of Astrovirus PS, a virus that shares 95% nucleotide identity with Astrovirus VA1, in brain tissue of an
immunocompromised child with unexplained encephalitis (17). The detection of this novel astrovirus in brain tissue strongly implicates this virus as the cause of the patient's symptoms and represents the first example of astrovirus associated with neurologic disease in humans. Interestingly, additional NGS studies have also identified other novel astroviruses associated with neurologic diseases of cows and mink (2, 14). Taken together, these NGS based studies have expanded our understanding of the tissue tropism and potential disease spectrum of astroviruses.

Establishment of New Animal Experimental Models.

Aside from identifying novel pathogens, a key benefit from the flurry of virus discovery activities has been the establishment of new animal model systems for understanding viral pathogenesis and virus-host interactions. This includes model systems for specific pathogens and more generalized models to examine broader questions of fundamental host-virus interactions.

Animal hepatitis C models. For many years, studies of hepatitis C were limited by the absence of a cell culture system and the lack of small animal models. The recent discoveries of rodent (6, 13) and canine (12) homologs of hepatitis C provide avenues to develop new small animal models for the study of these relatives of hepatitis C. While these studies are only in their infancy, paradigms for the effectiveness of such parallel rodent models include studies of murine gammaherpesvirus 68 as a surrogate for understanding Epstein-Barr virus and Kaposi's sarcoma herpesvirus and the use of murine norovirus as a model for human norovirus.

Bats as reservoirs of viral diversity. A major finding of the past decade has been the identification of bats as a source of tremendous virus diversity. Catalyzed in a large part by the discovery of SARS-like coronavirus in bats, subsequent NGS and taxon-specific consensus PCR based viral discovery efforts have demonstrated that bats harbor novel sequences derived from many known viral families (5, 15, 19). However, efforts to progress beyond discovery to
experimentally characterize these viruses have been hampered by the general inability to cultivate the majority of the novel bat viruses in common mammalian cell lines. Thus significant effort in recent years has focused on developing bat reagents— including genomics, transcriptomics, and new immortalized bat cell lines and primary bat cultures (4, 7, 16, 18). Such tools are critically important for defining the fundamental aspects of viral infection in bats, which in turn will help to answer question such as “What makes bats such permissive hosts to viral infection” and “What factors govern cross species transmission of viruses from bats”. It seems likely that with development of appropriate bat cell culture systems, the experimental tractability of bat viruses will mature much as the development of insect cell cultures in the 1960s revolutionized the culture of arboviruses.

**Virus infection in the model organism C. elegans.** C. elegans is a highly genetically tractable model organism that has played key roles in the discoveries of broadly conserved fundamental processes such as the caspase cell death pathways and RNA interference. However, its utility in the study of virus-host interactions has been limited due to the absence of any viruses known to naturally infect C. elegans. In fact, although nematodes are among the most abundant and diverse animals on earth, there were no viruses naturally infecting any nematode that had been molecularly characterized until 2011 when Orsay virus, the first and to date only known virus capable of infecting C. elegans, was discovered (8). Orsay virus is a positive sense RNA virus most similar to nodaviruses. Experimental conditions for infection of the laboratory reference C. elegans strain, as well as a viral reverse genetic system for Orsay virus (11), have now been established providing a robust new model system to define proviral and antiviral factors against this RNA virus. Such studies may provide new insights into not only invertebrate virology, but because ~40% of genes in C. elegans are conserved in mammals, the system also has potential to identify host pathways that are broadly evolutionarily conserved. The discovery of Toll in *Drosophila* and the subsequent characterization of Toll-like receptors in
vertebrates provide a paradigm for the power of model organism studies of infection and immunity.

**Future Directions and Challenges**

It is clear that the advent of NGS has triggered a new age of discovery of novel virus sequences. Continued application to humans, animals, plants and environmental niches will continue to uncover novel virus diversity for some time. Recent estimates for mammals suggest on the order of $10^5$ novel viruses that remain to be discovered (1), which would provide much fodder for future studies. These include traditional efforts to define pathogenic roles of novel viruses in diseases as well as the exploitation of new genomes to identify by comparative analyses novel genes, properties or host interactions absent in their previously described relatives. Furthermore, additional novel experimental models will undoubtedly arise from the multitude of new viruses identified. One limitation is that current NGS based viral discovery methods still rely largely upon alignment to known viruses, and thus truly novel viruses from previously unrecognized taxa will continue to evade detection until robust alignment independent bioinformatic tools are developed. In addition, the rate of discovery of new viral sequences has rapidly outpaced the rate at which culture systems for viruses have been developed, thereby limiting experimental studies of these new viruses. To maximally benefit from the wealth of viruses being discovered, there must be commensurate effort and resources dedicated to building fundamental tools and reagents for propagating these novel viruses. Despite these challenges, there will clearly be more low-lying fruit to be picked and exploited to advance our understanding of fundamental virology and viral pathogenesis.

**Acknowledgements**
Work in the author's laboratory relevant to this topic is supported in part by NIH grants U54 AI057160 and R21 AI097865 and by an Investigator in the Pathogenesis of Infectious Diseases Award from the Burroughs Wellcome Fund.
FIGURE LEGEND

Figure 1. Novel virus sequences deposited in Genbank based upon NCBI taxonomy criteria. The cumulative number of unique virus species nodes is plotted.
REFERENCES


