Available online at www.sciencedirect.com

SciVerse ScienceDirect

Current Opinion in Microbiology

Virome genomics: a tool for defining the human virome

Kristine M Wylie[1,2], George M Weinstock[1] and Gregory A Storch[2]

High throughput, deep sequencing assays are powerful tools for gaining insights into virus–host interactions. Sequencing assays can discover novel viruses and describe the genomes of novel and known viruses. Genomic information can predict viral proteins that can be characterized, describe important genes in the host that control infections, and evaluate gene expression of viruses and hosts during infection. Sequencing can also describe variation and evolution of viruses during replication and transmission. This review recounts some of the major advances in the studies of virus–host interactions from the last two years, and discusses the uses of sequencing technologies relating to these studies.

Addresses
[1] The Genome Institute, Washington University School of Medicine, Campus Box 8501, 4444 Forest Park Avenue, St. Louis, MO 63108, United States
[2] The Department of Pediatrics, Washington University School of Medicine, Campus Box 8116, 660 S. Euclid Avenue, St. Louis, MO 63110, United States

Corresponding author: Wylie, Kristine M (kwylie@genome.wustl.edu)

Current Opinion in Microbiology 2013, **16**:479–484

This review comes from a themed issue on **Host–microbe interactions: viruses**

Edited by **Carlos F Arias**

For a complete overview see the Issue and the Editorial

Available online 23rd May 2013

1369-5274

http://dx.doi.org/10.1016/j.mib.2013.04.006

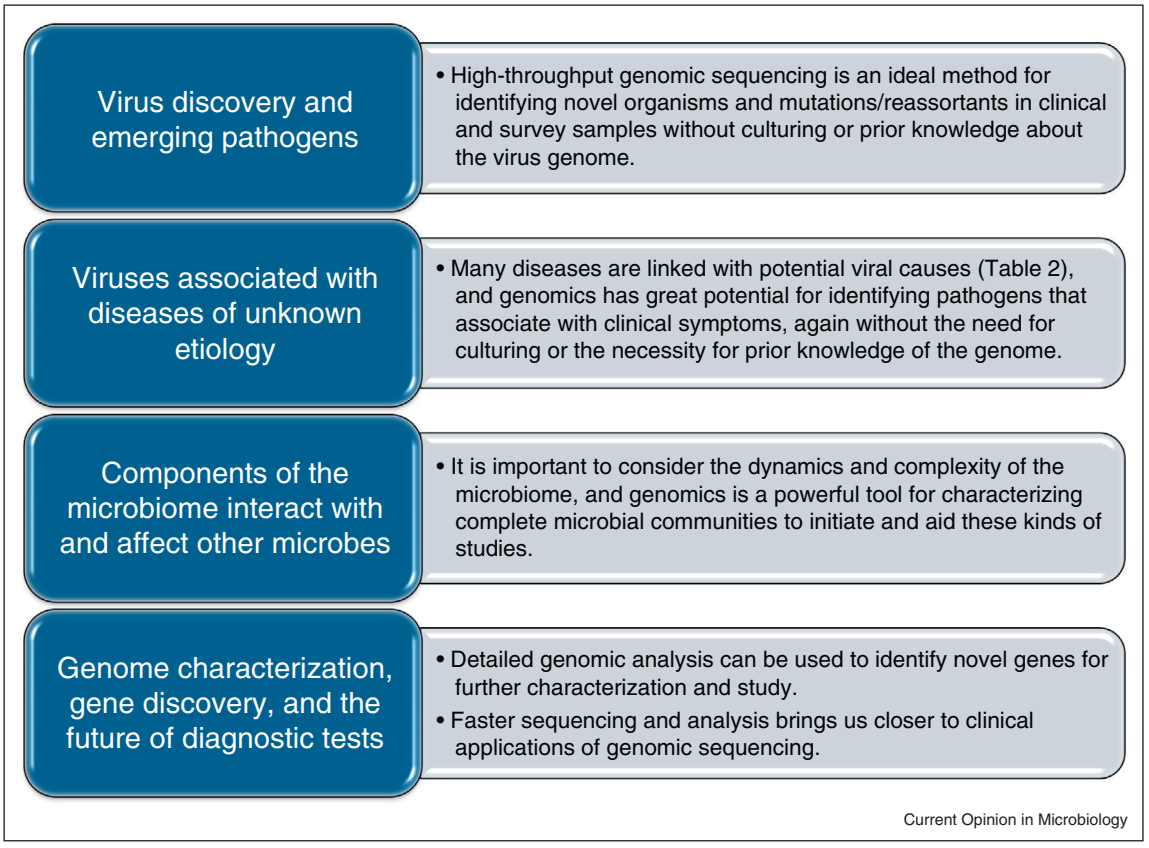
Introduction

The dynamic interactions between viruses and their hosts during infection are complex. Viruses manipulate the cellular environment and subvert host immune responses in order to replicate, and the host counters the virus's maneuvers in order to control the infection. The interplay between virus and host can be studied in many ways: natural infections; model systems (either animal or cell culture); manipulation of virus–host interactions; identification of the proteins involved in virus–host interactions; and studies of protein functions. High throughput, deep sequencing is a powerful tool for gaining insights into virus–host interactions. Sequencing assays can predict novel viruses and describe the genomes of novel and known viruses. Genomic information can be used to discover viral proteins that can then be characterized, describe genes in the host that are important in controlling infections, and evaluate gene expression of viruses and hosts during infection. Sequencing can also assess variation and evolution of viruses during replication and transmission. This review recounts some of the major advances in the studies of virus–host interactions from the last two years, and discusses the uses (or potential uses) of sequencing technologies relating to these studies (Figure 1).

Virus discovery and emerging pathogens

In order to understand how viruses interact with their hosts and how they affect human health, we must understand the scope of viral diversity and be able to detect the viruses present in clinical samples. High-throughput, deep sequencing has proven to be an effective tool for this purpose. The relatively unbiased approach it offers for screening clinical samples enables virus discovery without preconceptions about which viruses might be present in the samples. After 10 years of applying this technology to virus detection, eukaryotic virus discovery continues to be robust. A recent example of this is the novel rhabdovirus, Bas-Congo virus, which is an emerging pathogen associated with acute hemorrhagic fever [1], notable for being the first instance of a rhabdovirus being implicated as a cause of hemorrhagic fever. This virus was characterized in the context of a small outbreak, and the presence of antibodies in an asymptomatic caregiver suggested that person-to-person transmission had occurred. Another emerging pathogen, human coronavirus EMC (HCoV-EMC), was recently identified and characterized following an outbreak in the Middle East [2]. This betacoronavirus causes symptoms resembling those of its sister species, SARS coronavirus, including respiratory symptoms and acute renal failure, although HCoV-EMC is most closely related taxonomically to bat coronaviruses. Using modern technologies, the genome of HCoV-EMC was completely sequenced, and assays have been developed to monitor its presence. This virus is particularly interesting because coronaviruses are typically highly restricted to a specific host, but HCoV-EMC can infect cells from primates (human and monkey), swine, and bats (four families) in culture, suggesting that this virus may utilize a receptor shared among these host groups and may be readily transmitted between hosts [3]. These and similar studies demonstrate that continued viral discovery is needed in order to identify and prepare for the effects of emerging viral pathogens on human health. The techniques and technologies are in place for identifying pathogens with similarities to known viruses

Figure 1



The major topics of this review are summarized. Areas of virology research are noted in the blue boxes, and the text in the gray boxes describes how genomics can be used as a tool in that area of research.

(even remote similarities, see Table 1). The availability of samples and the funding required for the experiments currently bottleneck virus discovery efforts.

Testing samples from affected individuals during outbreaks of diseases of unknown etiology is important for surveillance of pathogens, but it is also important to identify viruses producing symptoms that are mild or subclinical because infection with these viruses may nevertheless have long-term implications for human health. For example, polyomaviruses and papillomaviruses may establish chronic infections. Some of these viruses, including many alpha papillomaviruses and Merkel cell polyoma virus, are associated with cancer [4–5]. In light of their capacity to transform cells, identifying the full range of polyomaviruses and monitoring their presence could ultimately provide insight into the development of some cancers.

Potentially emerging pathogens are of great concern, and influenza pandemics are of particular interest because transmissions between animal reservoirs and humans are observed and the emergence of pandemic strains is expected. Current molecular technologies allow us to screen for transmission of influenza between animal species and between humans, and to evaluate mutations and quasispecies. In two highly publicized studies, researchers identified mutations that correlated with airborne mammal-to-mammal transmission of the virus, a trait critical for the development of a pandemic [6•,7•]. Researchers used either an H5N1 influenza that originated in birds but had infected humans or a reassortant virus with the avian subtype H5 hemagglutinin and the other seven segments from a 2009 pandemic H1N1 virus. In these studies, viruses were passaged in ferrets, and isolates that had acquired mutations allowing for airborne transmission between ferrets were identified. The strains were sequenced and the mutations that correlated with ferret-to-ferret transmission were identified. The mutations were in the host receptor binding protein hemagglutinin and in the polymerase complex protein basic polymerase 2, and the studies showed that surprisingly few mutations would be necessary for the virus to achieve the capacity for airborne transmission between ferrets. Furthermore, a second study showed that two of the mutations were already circulating among H5N1 strains in birds [8]. These studies provide valuable insight into the plausibility of emergence of an H5N1 pandemic