Annual Review of Virology

Expanding the RNA Virosphere by Unbiased Metagenomics

Yong-Zhen Zhang,[1,2] Yan-Mei Chen,[1,2] Wen Wang,[2] Xin-Chen Qin,[2] and Edward C. Holmes[1,2,3]

[1]Shanghai Public Health Clinical Center and School of Public Health, Fudan University, Shanghai 200433, China; email: zhangyongzhen@shphc.org.cn

[2]Department of Zoonosis, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Changping, Beijing 102206, China

[3]Marie Bashir Institute for Infectious Diseases and Biosecurity, Charles Perkins Centre, School of Life and Environmental Sciences and Sydney Medical School, The University of Sydney, Sydney, New South Wales 2006, Australia

Annu. Rev. Virol. 2019. 6:119–39

First published as a Review in Advance on May 17, 2019

The *Annual Review of Virology* is online at virology.annualreviews.org

https://doi.org/10.1146/annurev-virology-092818-015851

Keywords

metatranscriptomics, virosphere, diversity, evolution, genomics, taxonomy

Abstract

Although viruses comprise the most abundant genetic material in the biosphere, to date only several thousand virus species have been formally defined. Such a limited perspective on virus diversity has in part arisen because viruses were traditionally considered only as etiologic agents of overt disease in humans or economically important species and were often difficult to identify using cell culture. This view has dramatically changed with the rise of metagenomics, which is transforming virus discovery and revealing a remarkable diversity of viruses sampled from diverse cellular organisms. These newly discovered viruses help fill major gaps in the evolutionary history of viruses, revealing a near continuum of diversity among genera, families, and even orders of RNA viruses. Herein, we review some of the recent advances in our understanding of the RNA virosphere that have stemmed from metagenomics, note future directions, and highlight some of the remaining challenges to this rapidly developing field.

INTRODUCTION

The past decade has seen a revolution in our understanding of the viral universe. For many years it has been known that bacterial viruses (i.e., bacteriophage) are some of the most abundant biological entities on Earth, with a famous estimate of their global population size of 10^{31} based on the assumption that each of the approximately 10^{30} bacteria on Earth carries some 10 phage (1). This view is supported by intensive marine sampling over the last 15 years that has revealed an impressive array of diverse phage (2–5). Until recently, however, there was no understanding of the diversity and abundance of eukaryote-infecting viruses, particularly those with RNA genomes (**Figure 1**). This has radically changed with metagenomic sequencing, in turn transforming our understanding of the total diversity, abundance, and structure of the virus world—the so-called virosphere (6–8). Metagenomics has led to a seismic shift in virology, opening up new research pathways and dismantling long-established dogmas, while at the same time presenting a variety of new challenges. In the case of RNA viruses this deluge of data has led to the discovery of a multitude of new viruses, genera, and families (6, 7). The flip side of this revolution is that it is now clear that we

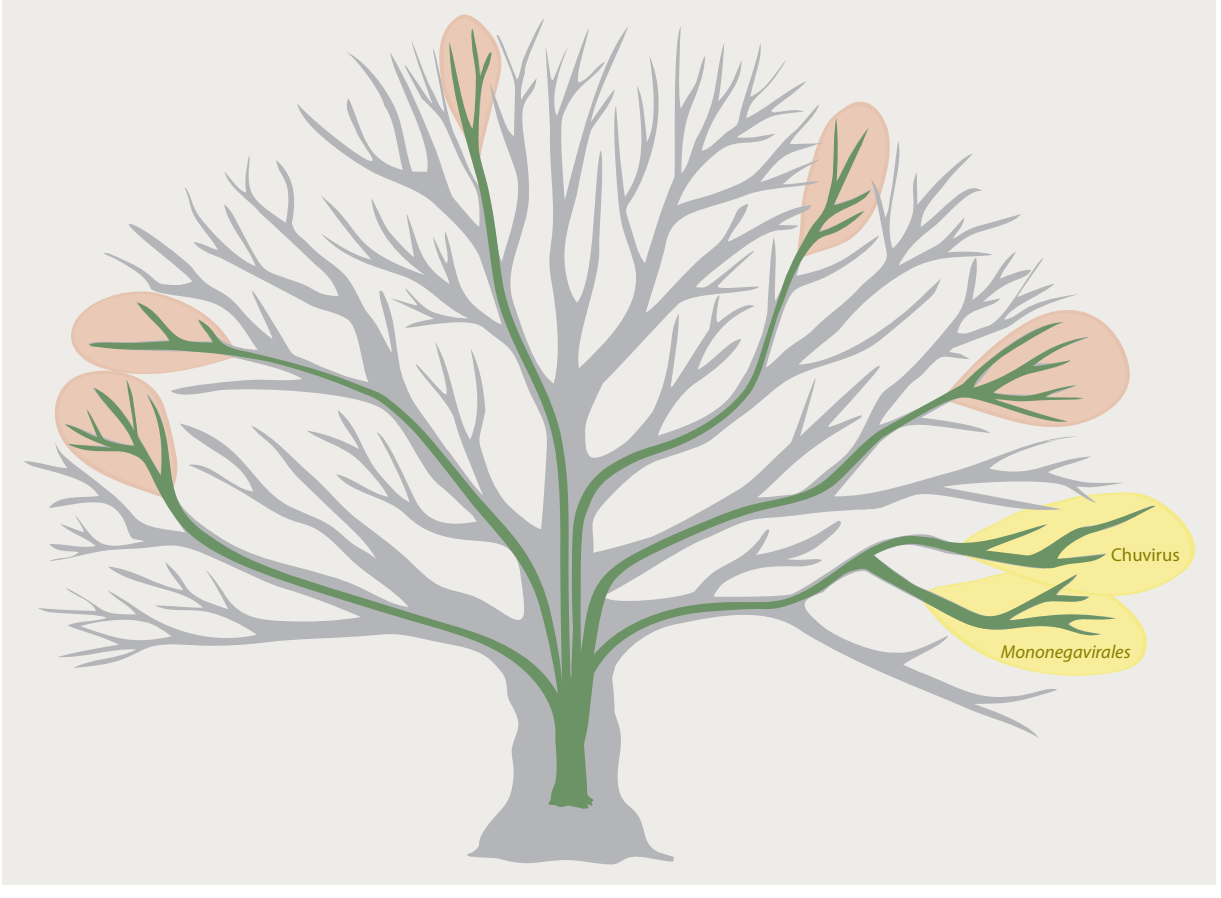


Figure 1

Schematic representation of the diversity and continuity of the RNA virosphere. The hypothetical tree in gray represents the totality of the RNA viruses that exist in nature, while those branches shown in green represent the viruses that have been described to date. The recently identified chuviruses and their relationship to the known order *Mononegavirales* are highlighted as examples.

have sampled only a miniscule fraction (likely significantly less than 1%) of the known virosphere (9). Although estimates of the number of viruses in nature are perhaps little more than arm waving, it seems reasonable to conclude that the total number of eukaryotic virus species in existence will be in excess of 100 million (9), and new RNA viruses are being discovered on a daily basis. Virology has plainly entered an exciting new discovery phase. Categorizing this amazing array of viruses, let alone analyzing all their phenotypic properties through experimental assays, poses a gargantuan challenge.

Since the first virus (tobacco mosaic virus) was identified at the end of the nineteenth century (10), to date only several thousand virus species have been defined by the International Committee on Taxonomy of Viruses (ICTV, **https://talk.ictvonline.org/taxonomy/**), although many more are waiting for classification and the proposal of new taxonomic groups is now a routine activity. Not only is our sample of the virosphere small, but it is also highly biased, and many of the best-described viruses are agents of infectious disease in humans or economically important animals and plants. Although these viruses are significant for public health and agriculture, their hosts obviously represent only a tiny proportion of the cellular organisms on Earth. Arguably the first sign that the RNA virosphere was far more expansive than previously realized was the work by Suttle and colleagues (11–13) that reported a multitude of RNA viruses in marine environments. In the past five years the focus has shifted toward terrestrial organisms, where sampling is necessarily more complex because it requires the capture and analysis of individual organisms but which has led to the discovery of staggering levels of viral diversity (14–17). Combined, this body of work has told us that viruses likely can infect all cellular organisms and are present in all environments, that the origin and evolutionary history of viruses remain unclear yet likely involve complex inter-virus interactions, and that our sampling of the virosphere is so small that any attempt to predict the evolution of key phenotypic traits based on this limited sampling of diversity, such as predictors of disease emergence (18), is hopelessly biased.

The overemphasis on viruses as agents of disease in part stems from the fact that they are often very difficult to identify and characterize using the traditional methods available to virologists, particularly isolation followed by visualization using electron microscopy and/or characterization by immunologic assays or consensus polymerase chain reaction (PCR) based on the known viral sequences. This is especially true for divergent RNA viruses that share little or no sequence similarity to existing viruses and that are still difficult to describe. Fortunately, this has begun to change with the application of metagenomic methods of virus discovery, particularly what has become known as metatranscriptomics (i.e., bulk RNA sequencing), which has become a common vehicle for virus discovery over the last decade (7). Through metagenomics it is now possible not only to discover novel viruses far more rapidly than before but also to characterize the entire virome present in single or even groups of host organisms (7, 8, 19). Consequently, a great number of divergent viruses have been identified in diverse host organisms sampled from disparate localities that are very difficult to characterize using traditional methods, with illustrative examples provided by the chuviruses (negative-sense RNA viruses) and the jingmenviruses (positive-sense viruses) (15, 16, 20). Because it simply records transcribed RNA, metatranscriptomics is a powerful tool to determine any type of etiologic agent infecting humans, animals, or plants (21–26).

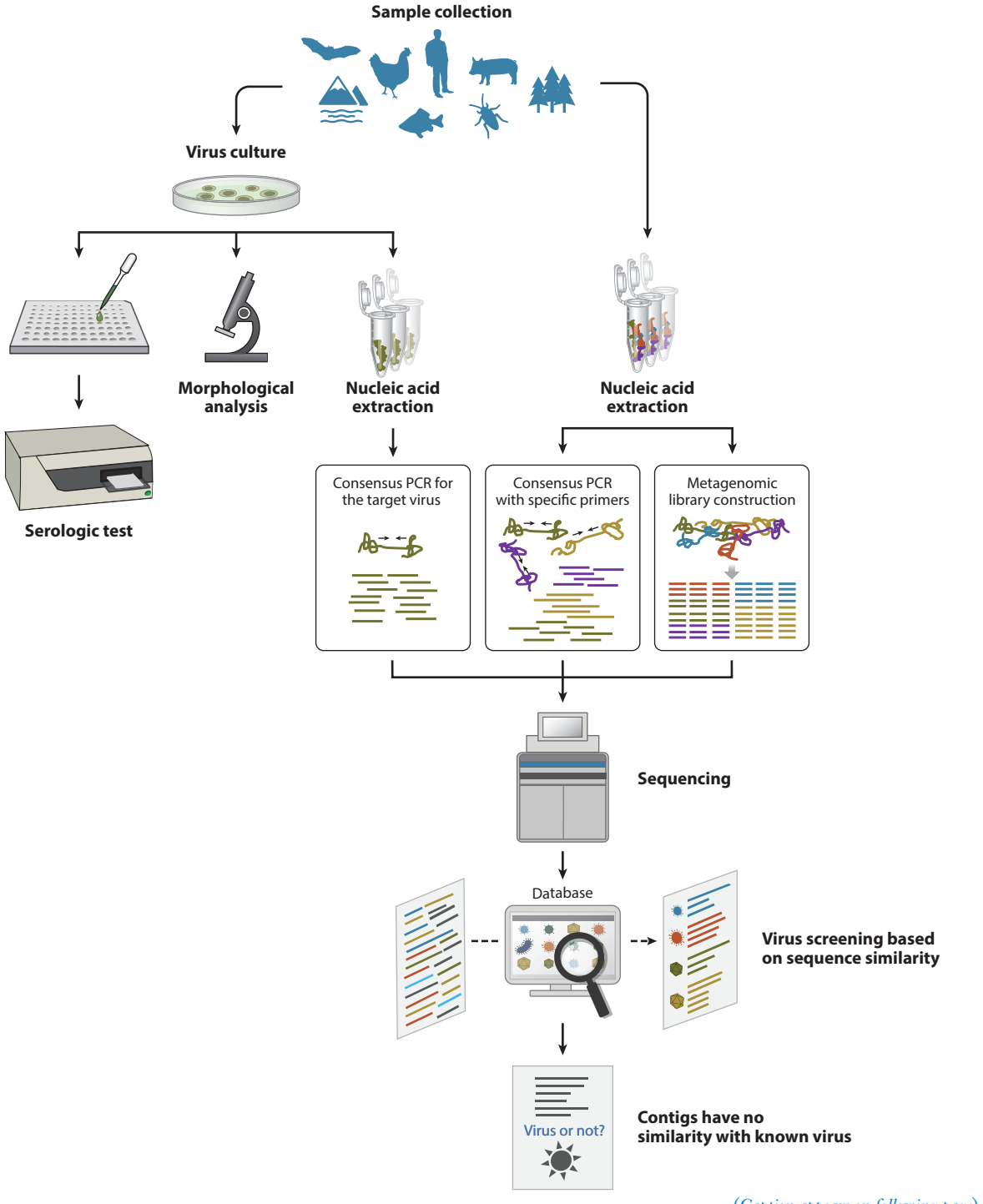
The explosive increase in the number of recognized novel and divergent viruses has therefore led to a new understanding of nature and the origin of virological diversity and the evolutionary patterns and processes that have given rise to it. Herein, we outline the main insights into the nature of the virosphere provided by this metagenomic revolution.

METHODS OF VIRUS DISCOVERY

Before the advent of molecular-based methods (27), a variety of techniques were commonly used to identify the viral agents causing disease in humans, animals, and plants. These included filtration, cell culture, electron microscopy, immunologic assays, and even the use of animal models (**Figure 2**). Indeed, these traditional methods have played an enormous role in exploring the virus world, establishing much of what we regard as modern virology. However, to characterize and define an unrecognized virus, these methods necessarily require the successful isolation and sometimes the availability of specific antibodies against known viruses. Substantial time and effort are therefore needed to complete the discovery process. In addition, due to the extraordinary genetic diversity of viruses that is now apparent, most viruses may not be culturable using current cell lines, nor are there the available person-hours to perform such huge amounts of laboratory work. It is therefore no surprise that many viruses already defined by the ICTV have not yet been isolated successfully. Hence, although the propagation of viruses in cells or laboratory animals has been the gold standard for virus discovery for over a century, these methods are clearly unsuitable and unfeasible for much of the virosphere. Culture-independent methods will therefore inevitably come to the fore (28).

As their name implies, molecular approaches to virus discovery involve only the determination and comparison of viral nucleotide sequences within a sample. Since the end of the 1980s, PCR methods have played an important role in the identification of novel viruses, and a number of important human agents have been determined in this manner, including hepatitis E virus (29), Nipah virus (30, 31), severe acute respiratory syndrome coronavirus (32), and Huaiyangshan virus (33). An even greater number of novel viruses have been identified and characterized by genomic amplification and sequencing, such as novel coronaviruses (34), hantaviruses (35), hepaciviruses (36), influenza viruses (37), paramyxoviruses (38), and picornaviruses (23). However, as the primers used in PCR-based methods are designed based on the available sequences of known viruses, it is obviously a challenge to identify novel viruses that share low or no genomic similarity with known viruses; hence, it is impossible to characterize the entire virome present in single or groups of host organisms.

One of the great advances of metagenomic approaches to virus discovery is that they do not rely on the availability of known closely related viruses (28, 39, 40). Indeed, in theory, a combination of metagenomics and high-throughput sequencing technology could determine all known and novel viruses present in a wide variety of samples (7). Due to the constraints of sample preparation and sequencing techniques, early metagenomic studies were much more successful in the detection of DNA than RNA viruses (2). Although the haul of RNA viruses could be improved by virus enrichment approaches (41) and sequence-independent amplification prior to sequencing (42), these also seem to be subject to a variety of biases and may still preferentially amplify DNA viruses. For these reasons unbiased RNA shotgun sequencing (metatranscriptomics) has recently come to the fore (43). The power of this technique is that it requires only sequencing of the expressed RNA within a host, followed by large-scale bioinformatic analysis to distinguish microbe from host. To increase, without bias, the relative abundance of virus compared to host, a common preparation step involves removing the ribosomal RNA (rRNA) during library preparation. Importantly, metatranscriptomics is simple yet unbiased, and it is highly efficient for virus discovery (7). From metatranscriptomic data it is possible to detect any infecting RNA viruses should they be present in a sufficient number of reads and have recognizable similarity to other RNA virus genes. The most common sequence tag in this case is the RNA-dependent RNA polymerase (RdRp), which is the most conserved sequence among RNA viruses and hence a popular phylogenetic marker (44). Indeed, a remarkable diversity of novel RNA viruses has been identified in both invertebrate



(*Caption appears on following page*)

Figure 2 (*Figure appears on preceding page*)

A schematic pipeline for virus discovery using different techniques. A variety of viruses are shown in different colors. Lines represent fragments or contigs of viral genomes generated by metagenomic library construction and sequencing, consensus PCR, and genome assembly. Abbreviation: PCR, polymerase chain reaction.

and vertebrate animals sampled from land and ocean environments using this method (15–17), including novel negative-sense RNA viruses (**Figure 3**).

Despite their transformative power, all metagenomic approaches have limitations and areas in which improvements in technology will greatly enhance pathogen discovery. Crucially, the RNA sequence may come from the host itself, as well as any microbe (virus, bacteria, fungus, or parasite) infecting that host, or be found in its diet, in the surrounding environment, or associated with another organism within that host. For example, the possession of variant genetic code in a virus sampled from an invertebrate suggested that the true host was in fact a protist (16). One simple way to help assess whether a specific virus is actively replicating in the host from which it is sampled—that it is a true infection rather than acquired from the surrounding environment—is to consider its abundance. This can be readily achieved in metatranscriptomics by measuring the relative abundance of specific transcripts (16). The analysis of transcript abundance also provides important insights into virus genome composition and structure, although confirmation by PCR and/or Sanger sequencing is often still merited. For example, comparing the abundance of RNA fragments generated by metatranscriptomics helped identify the two segments encoding the structural proteins in Jingmen tick virus that shared no similarity to any known viruses (20). In addition, clues as to the true host of an RNA virus can be obtained by placing the sequences in phylogenetic context. Hence, for example, a virus that actively replicates in a vertebrate will most likely be related to other vertebrate viruses, whereas a closer relationship to an invertebrate or other more divergent host species may be indicative of infection of a different host (45). As noted below, the viromes of vertebrate species commonly contain viruses that seem legitimately associated with vertebrates as well as those that are more likely to infect co-sampled parasitic species (16).

The second limitation of metagenomics is that for a virus to be detected it must be present in a sufficient number of reads within the coverage of a specific sequencing run, which is itself dependent on the RNA quality within a sample. This will obviously present challenges in the case of viruses associated with transient or latent infections (46). Fortunately, the greater number of reads being produced by increasingly powerful sequencing platforms will make the detection of low-frequency viruses even easier, and it is likely that new CRISPR-guided protocols will be developed to remove more of the host genome, in turn increasing the proportion that is microbial. A complicating factor is that the reagents commonly used in metagenomics are frequently contaminated with a range of viruses (47), so that great care must be taken to distinguish bona fide viruses. Important discriminatory criteria are virus frequency, whether they are found in all libraries, and whether they are most closely related to marine viruses, as these are often the source of contamination. Ultimately, for metagenomic studies to advance, it may be necessary to include a blank control library in every study.

Finally, all metagenomic approaches are entirely dependent on identifying homology (i.e., significant sequence similarity) between the sequences detected and those present in the databases. While this works for viruses that are clearly related to those from families identified to date, it will be less effective for RNA viruses that are so divergent in sequence that they cannot be readily detected. This is especially true for basal eukaryotes, bacteria, and particularly Archaea, for which no RNA viruses have been clearly identified to date apart from one disputed case (48), perhaps because they are too divergent in sequence to identify. Given that it will be impossible to depict

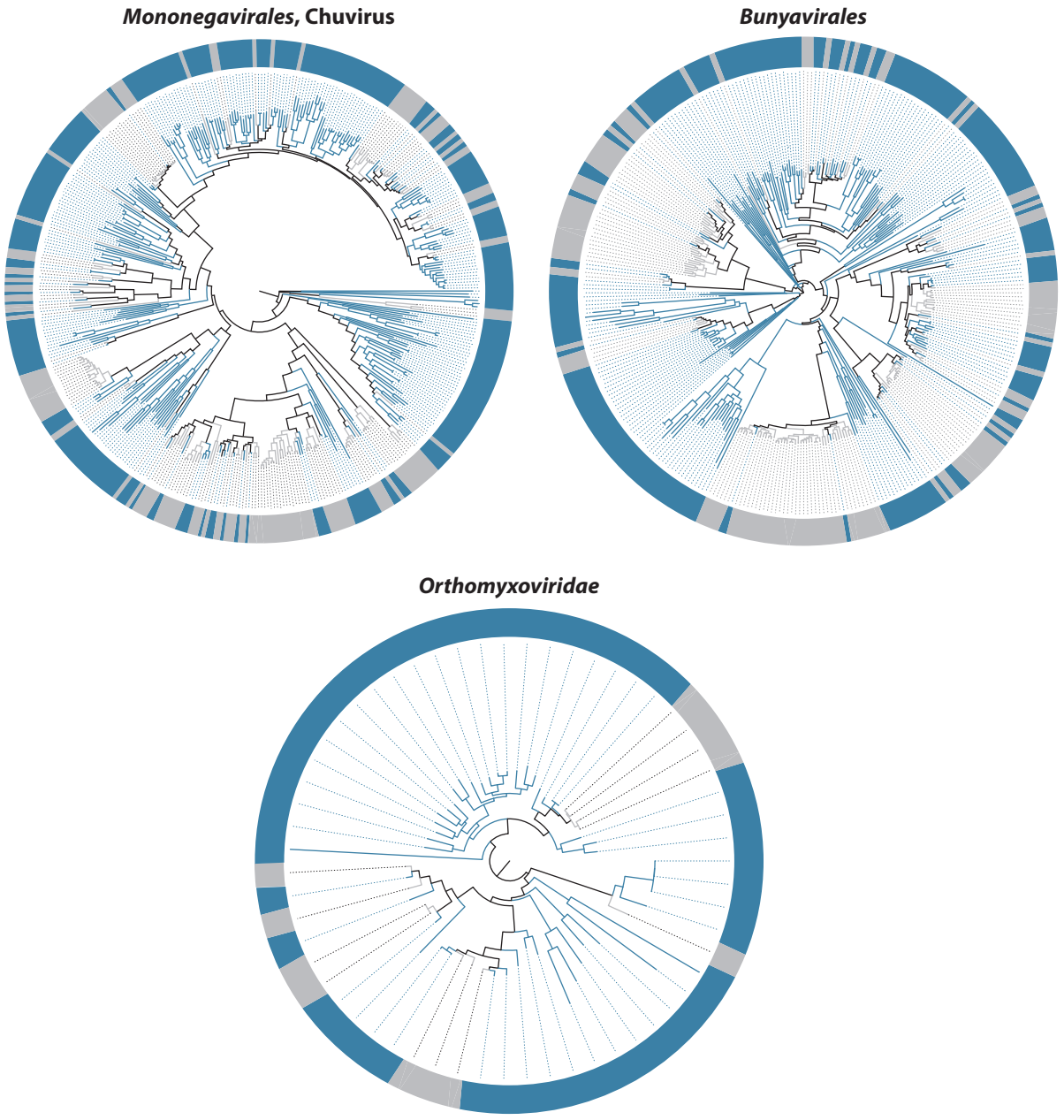


Figure 3

The extensive virus diversity within different groups of negative-sense RNA viruses identified using metatranscriptomics. Within each circular phylogeny, the virus species discovered by metagenomics are shaded blue, while those discovered by traditional and PCR methods are shaded gray. Abbreviation: PCR, polymerase chain reaction.

significant relationships between primary sequences that exceed a certain level of diversity (see the next section), the most likely way forward is to search for viruses based on elements of protein structure (49, 50).

THE INVERTEBRATE RNA VIROME

If there is one group of organisms in which metatranscriptomics has been truly transformational, it is the invertebrates (7). Before the rise of metagenomics, relatively little was known about the true diversity of invertebrate RNA viruses. Only a relatively small number of viruses had been described among the millions of invertebrate species, and many of the viruses identified were done so in the context of human and animal disease, such as the *Bunyavirales* (**Figure 4**). Hence, there was, and still is, a strong focus on arthropods (phylum Arthropoda), particularly mosquitoes and ticks, which are well known as vectors for specific human viruses. This picture has radically changed with metagenomics (**Figure 3**). In particular, we now know that the natural virome of some invertebrate species can be immense, with mosquitoes and ticks again representing high-profile examples, and that the viruses that are disease causing in vertebrates (and hence that are able to productively replicate in vertebrates) are present at much lower levels (15, 16, 51–58). There is also mounting evidence that some of the natural benign viruses found in invertebrates can effectively block their pathogenic relatives, which opens up new opportunities for disease control (59, 60). It is also possible that by encoding PIWI-interacting RNAs, the endogenous virus elements (EVEs), which are portions of viral genomes that have integrated into the host genome and that are particularly common in arthropod (61), present in mosquito genomes could in part explain their immunity to specific viruses (62).

The diversity of RNA viruses in invertebrates is immense and covers not only essentially all families of animal RNA virus identified to date but also families that were once thought to be restricted to plants only, such as the *Luteoviridae* and the *Tombusviridae*, as well as a number of novel virus groups of which the jingmenviruses and chuviruses are important exemplars (15, 16, 20) (**Figure 1**). From an ecological perspective arthropods may be particularly important because they interact with both vertebrates and plants and therefore provide a conduit for viruses to move between these disparate host types (15, 63), helping to unify the virosphere. Indeed, arthropods harbor viruses that are closely related to those found in plants and vertebrates and that act as vectors for the infection of both (15, 16). Phylogenetic analysis reveals that many of these arthropod viruses fall in basal positions on expansive evolutionary trees that also contain viruses from vertebrates, suggesting that at least some have been in existence for the entire evolutionary history of the Metazoa (see the next section) (16, 17). Given the many millions of invertebrate species in existence, of which only a negligible fraction has had their viromes characterized to date (and some phyla barely sampled at all), it is obvious that many millions of invertebrate viruses remain to be discovered. Not only will these new viruses continue to fill the gaps in virus diversity, highlighting the continuity of the virosphere, but also it is certain that new viral families, and perhaps orders, will also be discovered (**Figure 1**). We strongly suspect that a similar increase in the diversity of invertebrate DNA viruses will also be forthcoming when equivalent large-scale sampling regimes take place.

Not only are the RNA viruses found in invertebrates extremely diverse, but they are also highly abundant and in many cases possess a far wider range of genome structures than their vertebrate counterpoints (see the next section). In some instances the level of abundance of RNA viruses in invertebrates is staggering, representing over 50% of the number of transcripts in the sample (excluding rRNA) (16). Abundance values over 1% of the total number are commonplace in invertebrates yet are rare in vertebrates, even in the case of viruses responsible for overt disease

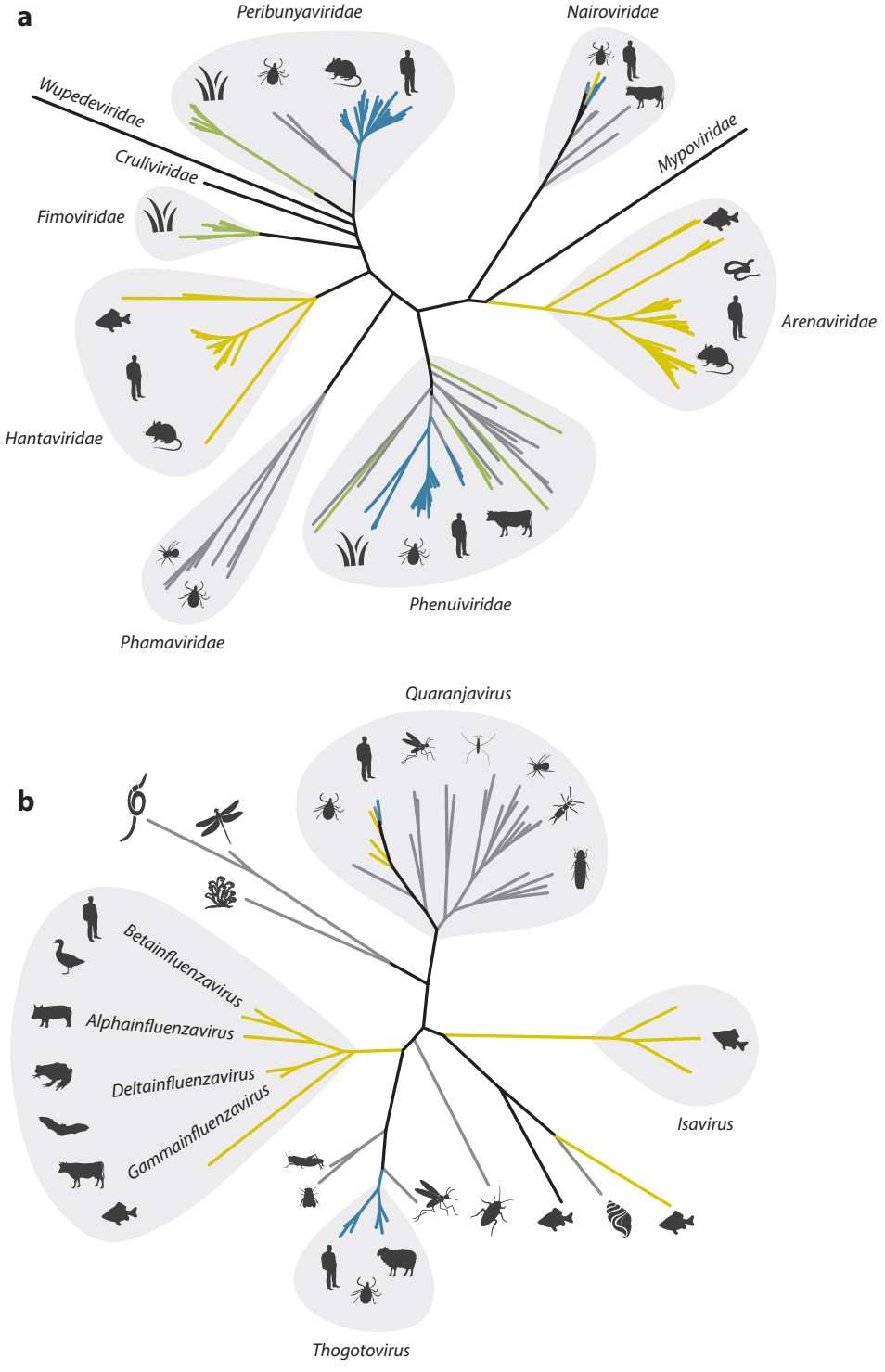


Figure 4

Phylogenetic trees of the RNA virus order *Bunyavirales* (*a*) and the family *Orthomyxoviridae* (*b*) illustrating the wide diversity of hosts infected. Hosts are represented by different branch colors: plants (*green*), arthropods only (*gray*), vertebrates only (including humans, *yellow*), and both arthropods and vertebrates (including humans, *blue*).

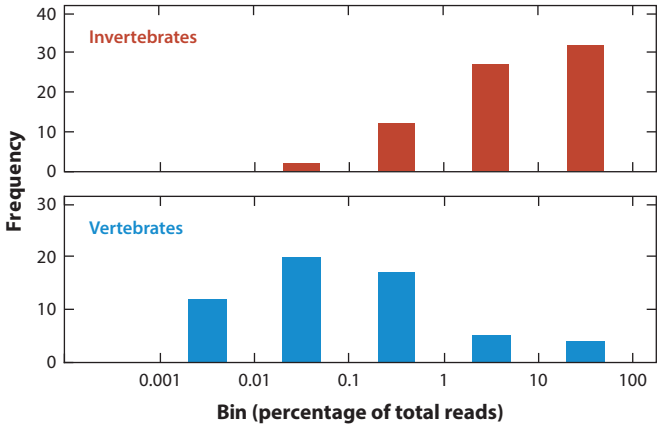


Figure 5

Contrasting abundance of invertebrate and vertebrate RNA viruses identified through metatranscriptomics and categorized by frequency (percentage) categories. Data are from Reference 16 for invertebrates and Reference 17 for vertebrates.

(**Figure 5**). The outstanding question, of course, is how invertebrate species are able to harbor such an enormous load of viruses, which must impose some cost in terms of replicative burden. At the very least this observation suggests that many of these viruses are unlikely to cause overt and serious pathology in their invertebrate hosts and may simply represent benign, or perhaps mutualistic, passengers. This is clearly an area in which future study will be highly profitable and may go further to change the dogma that viruses are consistently pathogenic (64).

THE DIVERSITY AND EVOLUTION OF VERTEBRATE RNA VIRUSES

Describing the range of viruses present in vertebrates in nature as well as their origins is central to understanding emerging disease and human health (65). The most basic insights from metagenomics in this respect are that vertebrate RNA viruses are more diverse and older than previously realized and that even seemingly healthy animals can carry a wide range of RNA viruses (6, 19). A telling fact is that every virus family that had previously been discovered only in mammals has now been identified in a more divergent class of vertebrates, particularly fish (17) (**Figure 4**). Indeed, bony fish carry a staggering range of RNA viruses including relatives of those that comprise important pathogens in humans, such as hantaviruses, arenaviruses, filoviruses, and influenza viruses (17, 45), as well as important disease-causing DNA viruses such as the hepadnaviruses (45, 66, 67). The most obvious conclusion from this is that these families of RNA viruses are at least as old as the vertebrates (that is, approximately 400 million years), as also appears to be the case for the hepadnaviruses (67, 68). Further, because these vertebrate viruses are then most closely related to those found in invertebrates, albeit often connected by a relatively long branch, it seems probable that these viruses are as old as the Metazoa, at over 600 million years (17).

Illustrative examples of this ancient evolutionary history are the orthomyxoviruses, a group of segmented negative-sense RNA viruses that includes the influenza viruses. Until recently, the orthomyxoviruses comprised a relatively small family of RNA viruses sampled from a small number of mammals, birds, and ticks (69). A plausible scenario from this was that the orthomyxoviruses perhaps originated in birds and may have been primarily tick-borne before losing the vector in the case of the directly transmitted influenza viruses. Metagenomics paints a radically

different picture (**Figure 4*b***). Studies of more diverse vertebrates show that influenza viruses are present in amphibians, bony fish, and, remarkably, jawless vertebrates such as the hagfish (17). This strongly suggests that this lineage is likely as old as vertebrates themselves and again exhibits a macroevolutionary pattern that reflects a combination of virus-host codivergence and cross-species transmission (70). It is inevitable that a wider sampling of vertebrate taxa will reveal the presence of even more divergent influenza viruses. Wider studies of the orthomyxoviruses as a whole reveal that this virus family is present in diverse arthropods including mosquitoes, flies, spiders, and earthworms (15, 16). Such a taxonomic distribution places the origin of the orthomyxoviruses perhaps coincident with the origin of the Metazoa. In addition, although all orthomyxoviruses sampled to date possess segmented genomes, this long evolutionary history has also witnessed changes in segment numbers within the orthomyxoviruses, from 6 to 10 (15, 17, 71). Similarly, until recently the standard view on the origin of hepatitis delta virus was that it was derived from the human genome, particularly as it was always associated with hepatitis B virus (HBV) (72). However, the metagenomic discovery of this virus in birds in the absence of HBV suggests a very different and older evolutionary history (73).

Although metagenomics has led to a radical new view of the diversity and perhaps disease-causing propensity of vertebrate viruses, it has arguably had less of an impact on our understanding of the human virome. As expected given its intense focus, the rate of discovery of human viruses has slowed dramatically, with a little over 200 human viruses documented to date (74), and metagenomics analysis has told us that most healthy humans carry a very limited load of viruses (bacteriophage excluded). Although bacteriophages are commonplace in the human microbiome, what role they play in modulating human health and disease is unclear (75). However, the power of pathogen discovery offered by metagenomics will unquestionably be of immense importance in responding to future outbreaks of infectious disease in humans, and the regular screening of those working at the human-animal interface may provide a means for the highly effective surveillance to help mitigate future outbreaks (76). Indeed, metagenomics promises to transform the way in which microbial infections are diagnosed, opening up the possibility of a one-stop shop for the identification of any infectious agent (bacteria, virus, fungus, parasite) within a host and from a variety of tissue types (26, 46, 77). This promises a genomic revolution on a par with that currently going on with the use of human genomic sequencing in cancer.

As noted above, vertebrate RNA virus genomes are normally far less abundant parts of the total cellular transcriptome than those in invertebrates (17) (**Figure 5**). Although this may to some extent again reflect the vagaries of sampling, it is tempting to speculate that the dramatic reduction in the frequency of vertebrate virus RNA genome copies per cell is due to the evolution of adaptive immunity that has dramatically reduced the viral burden. An increased sampling of taxa will go a long way toward answering this question. Similarly, it is also notable that some of the vertebrate viruses characterized to date are shorter and possess less complex genome structures than their invertebrate relatives (78), perhaps because shorter genomes will generate fewer targets for a fully active vertebrate immune response. A good example of the differing genome organizations of invertebrate and vertebrate viruses is provided by the *Flaviviridae* and the closely related flavi-like viruses. Prior to metagenomics the flaviruses appeared to be a group of small (~10 kb), unsegmented viruses that infected vertebrates with arthropods (particularly mosquitoes and ticks) acting as vectors. Post metagenomics they now appear to be group of largely invertebrate viruses that have secondarily infected vertebrates; contain groups that are insect specific (60); have genomes that can be in excess of 25 Kb; and can possess both segmented (comprising between four and five) and unsegmented genomes (20, 78), and in some cases can be multicomponent such that their genome segments are located in different virus particles (79), which was previously observed only in some plant RNA viruses.

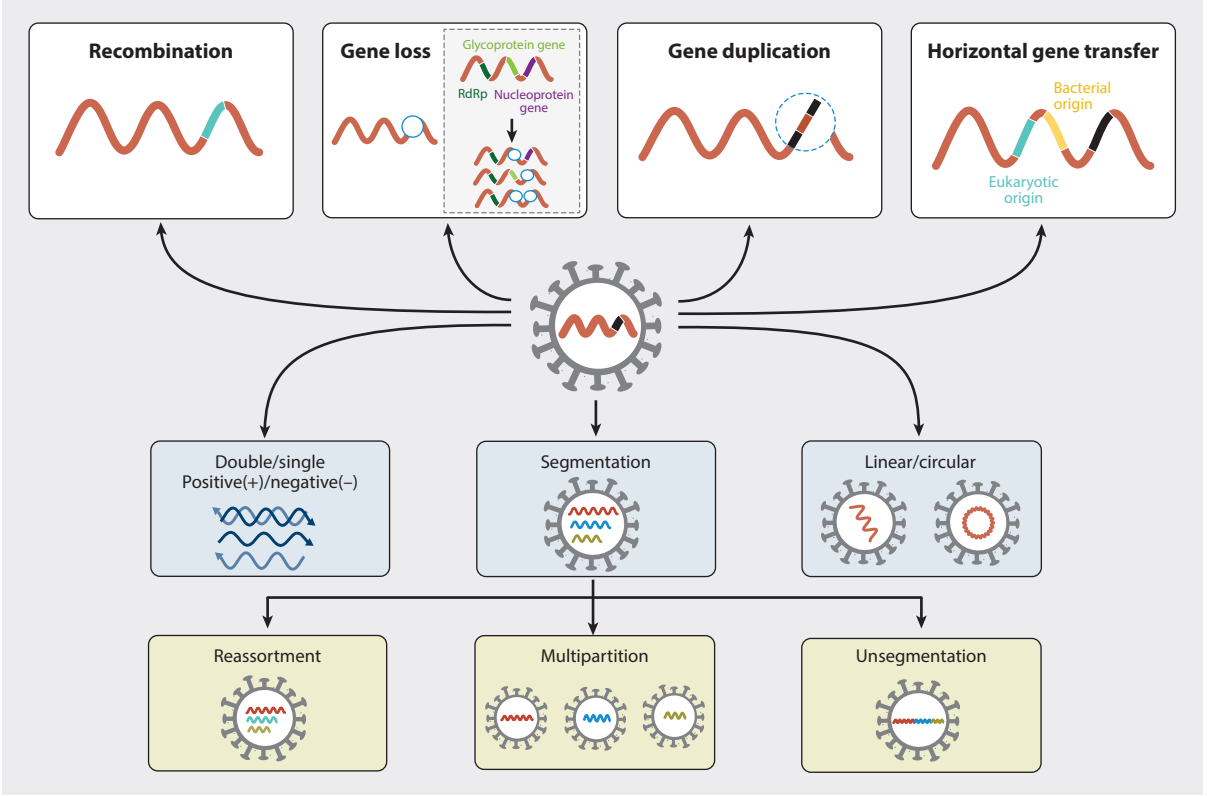


Figure 6

Genome diversity and evolutionary mechanisms that generate this diversity in RNA viruses. The viral genome is represented by a wavy line in each case. Different genome segments (for segmented viruses) or genes of different evolutionary origins are indicated by different line colors. Abbreviation: RdRp, RNA-dependent RNA polymerase.

THE EVOLUTION OF RNA VIRUS GENOME STRUCTURES

The change in our understanding of the diversity and flexibility of RNA virus genome structures following the metagenomic revolution has arguably been as profound as that of their phylogenetic diversity (**Figure 6**). As noted previously, RNA virus genomes are more diverse, have more complex structures, and have a wider range of lengths than previously anticipated (19). For example, while the presence of genome segmentation, and the number of genome segments, was traditionally considered to be a relatively stable taxonomic marker, metagenomics has shown that segment patterns and numbers vary dramatically among RNA viruses and can even vary within a single family. Hence, segmentation no longer appears to be a strong taxon-defining trait. As noted above, one of the first hints that might be the case were the *Flaviviridae* and their close relatives (20, 78), although it now appears to be true of many virus families, including members of the *Picornavirales*, the *Tombusviridae-Nodaviridae* cluster, and the *Hepeviridae-Virgaviridae* cluster (16). Segment numbers can also vary substantially within individual families; for example, a range of one to six is seen in the *Partitiviridae*-like viruses. Obviously, that different patterns of genome segmentation can evolve so frequently has important implications for understanding the selective factors that led to its evolution in the first place. For example, the discovery of segmented orthomyxoviruses in arthropods means that the process of genomic reassortment, which naturally co-occurs with

segmentation, is an ancient innovation that long predates the evolution of acquired immunity in vertebrates. More generally, other than the fact that RNA viruses are seemingly universally small, with the size cap currently sitting at less than 45 Kb, there are few other generalities that can be imposed on the evolution of their genome structures. It is, of course, highly likely that the length profiles of RNA viruses will increase with greater sampling, and the length of the longest RNA virus has increased continually with the discovery of novel nidoviruses (79–81). An RNA virus with the length and complexity of a large double-strand DNA virus represents something of a virological holy grail.

The new diversity of RNA viruses has revealed a variety of other changes in genome structure including genome size, the number of genes, and their orientation (**Figure 6**), and metatranscriptomics has demonstrated the existence of negative-strand RNA viruses that lack the typical nucleoprotein or glycoprotein genes, and even both (15, 16). For example, two arena-like viruses discovered in marine fish have three RNA segments instead of the two seen in mammals or reptiles (17). Interestingly, arena-like viruses with three RNA segments have also been found in arthropods (16), indicative of a complex evolutionary history. As noted above, invertebrate RNA viruses are a particularly rich source of genomic diversity (15, 16) and exhibit greater variation in architectures than their counterparts found in vertebrates. For example, metagenomics suggests that the chuviruses—a diverse group of negative-sense RNA viruses that are related to the *Mononegavirales* and that infect a range of invertebrate species—contain a wide array of genome structures, including both segmented and nonsegmented forms as well as those with likely circular genomes (15). That such a diversity of structures was found in the relatively small number of eukaryotes sampled to date suggests that many more genomic surprises lie in wait.

The diversity of genome structures harbored by even relatively closely related viruses also provides important insights into the basic processes of genome evolution. Although clear-cut cases of gene duplication still appear to be relatively rare among RNA viruses (82), which may in part be a function of the difficulty in detecting sequence homology in the face of high levels of sequence divergence, metagenomics has told us that lateral (horizontal) gene transfer is a relatively common evolutionary process, particularly in invertebrate RNA viruses (14, 16, 19). Lateral gene transfer is readily manifest in the markedly incongruent phylogenetic trees inferred using different virus proteins, such as those encoding the RdRp and those encoding viral capsids (16). Most striking is that there is even evidence for lateral gene transfer involving host genes inserted into viral genomes, although there were few recorded cases until the rise of metagenomics (83). For example, two viruses associated with the sea slater (an invertebrate) contain exoribonucleases of eukaryotic origin (16). The occurrence of lateral gene transfer is also important from the perspective of evolutionary theory, as this process will usually result in an increase in genome size that in turn is thought to result in an increased burden of deleterious mutations and hence a cost to overall fitness (84). As the metagenomic revolution continues, more cases of this important evolutionary process will undoubtedly be identified. Indeed, we expect that as data accumulate, the evolutionary processes that shape the evolution of RNA viruses will more closely resemble those in bacteria, in which lateral gene transfer is an important means of generating evolutionary novelty (85). To better determine the evolutionary processes that shape viral genome structures, and hence how new viruses are created, it is important to use the new wealth of metatranscriptomic data to generate denser and more informative phylogenetic trees, as these will make it easier to determine the frequency, pattern, and direction of gene duplications and losses, lateral gene transfers, and genomic rearrangements.

METAGENOMICS AND VIRUS TAXONOMY

The identification and characterization of a rapidly increasing database of viruses, some of which are highly divergent, undoubtedly represent a major challenge, and perhaps unique opportunity,

to the current taxonomy of viruses (86) and to our understanding of the processes that generate viral species (87). As there has already been considerable discussion on the potential impact that metagenomics will have on virus taxonomy and evolution (4, 7–9, 14, 19, 40, 88–91), this issue is considered only briefly here.

Because viruses were traditionally described based on morphological rather than molecular data, the rise of metagenomics necessarily means that rules of engagement for virus taxonomy have changed, with sequence-only studies coming to fore (14, 86). While the sheer scale of the virosphere means that genome sequence–based studies are the only practical way to proceed, a key concern is that because virus classification is based on the underlying phylogenetic tree linking RNA viruses, and only a tiny fraction of that tree has been sampled, it may be the case that established hard taxonomic boundaries will collapse as more viruses are sampled (19). As already noted, one of the major impacts of metagenomics is to fill gaps in the tree of RNA viruses, and these gaps are often used as important taxonomic boundaries. It will therefore be interesting to see whether the currently established taxonomic boundaries within viruses can withstand the onslaught of new data.

A related concern is that many of the phylogenies used as the basis for classification schemes are based on only a single virus gene—the RdRp—as this is the most conserved component of the RNA virus genome and is common to all RNA viruses. However, metagenomic studies have shown that RdRp trees do not necessarily reflect the evolutionary history of the entire viral genome, which can be shaped by relatively frequent interspecific recombination and lateral gene transfer (6). Hence, gene trees do not necessarily match species trees, which is particularly evident in comparisons of virus capsids and those based on the RdRp (16). Although RdRp-based trees are the only reasonable basis for phylogenetic approaches to virus classification, care should be taken to recall that this then does not accurately depict the evolutionary processes that have shaped this evolutionary pattern. A phylogenetic tree is a model of the evolutionary process, and one that may look increasingly out of date for viruses as more taxa are sampled (6).

As noted above, another well-documented challenge for metagenomic-based studies of virus evolution is detecting viruses that are too divergent to be identified using the currently available sequence similarity searching methods such as BLAST and HMMER. Highly divergent reads are a common occurrence in any metagenomic data set, and some of these may comprise the so-called dark matter of the virome: viral sequences that are too divergent to prevent accurate characterization (7). Indeed, in many cases the putative protein sequences identified in metagenomic surveys are so divergent in sequence that it is impossible to accurately assign their origin and function. The reality of the matter is that no analytical method based on primary sequence similarity will be able to robustly identify divergent RNA viruses beyond a specific cutoff point of sequence similarity. Beyond this there is no more sequence similarity than predicted by chance alone preventing any meaningful subsequent analyses (92). Hence, the lack of recognizable sequence similarity in metagenomic data sets does not mean that more divergent viruses do not exist but rather that they are too divergent in sequence to be detected. Pragmatically, the only way to proceed in these circumstances is likely through elements of protein structure and particularly conserved domains, although this will impose a major computational burden. Similar and long-standing reservations also apply to instances of trying to infer the phylogenetic history of viruses, and hence their associated genome structures, from the analysis of RdRp genes alone. Although very short sequence motifs exist that comprise a powerful argument for the common ancestry of RNA viruses as a whole (44), these are in no way of sufficient length to be able to construct reliable sequence alignments and form reliable phylogenetic trees (92). Any attempts to infer expansive phylogenies of RNA viruses from sequence data alone should be treated with great caution (93). Metagenomics therefore raises a number of challenges that will need to be met by advances in computational

biology and a fusing of primary sequence and protein structure approaches to categorizing biological diversity.

METAGENOMICS AND THE NEW VIROLOGY

Much of virology, particularly in the context of human disease, has implicitly assumed that viral infections are rare and that any particular disease can be assigned to an individual pathogen. This view is in accord with the Koch's postulates that have dominated work in infectious diseases since the nineteenth century. Metagenomics is challenging this view, leading to new approaches to define pathogens (94) and to what might be considered a new virology. It is now clear that many species, particularly wildlife, carry an abundance of viruses at any one time, that these are often not associated with any disease or impact on fitness, and that the presence of one virus may have a profound impact on the presence and abundance of another. For example, a recent study of the viruses associated with wild birds in Australia revealed that they are commonly infected with multiple viruses, that most of these do not result in observable symptoms, and that those birds infected by (low-pathogenicity) influenza A virus had a higher prevalence of other RNA viruses (95). The precise explanations for these intermicrobial interactions are unclear. Hence, it may sometimes be more profitable to think of viruses as members of a wider community of microbes that may interact in a complex manner rather than as independently evolving and isolated entities. Intermicrobial interactions of this sort are also being documented in the viruses involved in human disease. Most famously, *Wolbachia* bacteria are able to block the replication of a number of important human viruses transmitted by *Aedes aegypti* mosquitoes including dengue, Chikungunya, and Zika (96, 97). Interestingly, this blocking effect seems to be strongest in viruses that are newly emerged in these invertebrate hosts. For example, while *Wolbachia* bacteria are able to block the replication of a number of *Drosophila* viruses in a laboratory setting, this does not appear to be the case in natural *Drosophila* viromes, with the metatranscriptomic analyses of individual flies revealing no association between the *Wolbachia* and the presence or abundance of viruses (98). Future metatranscriptomics studies of individual animals are likely to provide even more insights into the nature of these intermicrobial interactions. More fundamentally, we still do not know why host species, even those that are closely related, differ in the diversity and abundance of RNA viruses they carry. For example, *Culex* and *Aedes* genus mosquitoes sampled at the same location differed profoundly in the number and abundance of viruses they carry (55).

Analyses of the genetic diversity produced by metagenomics have helped provide a better understanding of the relative roles played by virus-host codivergence and virus cross-species transmission in the long-term evolution of RNA viruses (84). Cross-species transmission should not be regarded as an unusual form of viral evolution that is associated only with disease; the reality of the matter is that virus host jumping is a normal mode of virus evolution and perhaps most often occurs in the absence of any overt disease. Indeed, cross-species transmission occurs in every family of RNA viruses studied and often at very high frequency (64, 69, 99). As noted above, this pervasive cross-species transmission also occurs on a background of virus-host codivergence over periods of millions, and perhaps hundreds of millions, of years. What is less clear, however, is whether there are distinct breaks in the expansive diversity of RNA viruses, perhaps manifest as gaps in phylogenetic trees. This leads to the question of whether such breaks represent specific biological features that inhibit the appearance of intermediate viruses, such as the inability to infect certain cell types, or whether the presence of long branches on deep virus phylogenies simply reflects inadequate sampling. As a simple case in point, does the relatively long branch that often exists between some invertebrate and vertebrate viruses mean that only a subset of viral families of specific phenotypes are able to replicate in such different host types, or does the

gap just reflect an absence of data? Greatly enhanced sampling will go a long way to provide an answer.

Because of metagenomics we now know that extant families of RNA viruses may be many millions of years old, have been associated with particular types of hosts over geological timescales, and have evolutionary histories strongly coupled with those hosts. Hence, as eukaryotic populations experienced mass extinction events through evolutionary time, these would also have greatly impacted the diversity of their associated viruses (100), although there are currently insufficient data to determine how. Indeed, it is obvious that the currently available sample of viruses necessarily reflects only a small subset of viruses that have been able to survive to the present day, with many past viral lineages likely experiencing extinction (84). Key to understanding these long-term virus-host interactions will be the establishment of robust timescales of virus evolution, although this will be complicated by the high frequency of cross-species transmission. Dating exercises for viruses have been undertaken in a variety of ways, although they all have both benefits and limitations. Because of the huge levels of sequence divergence observed between RNA viruses, particularly those assigned to different families, molecular clock dating through comparisons of recently sampled sequences is applicable only to relatively shallow comparisons (i.e., a few hundred or thousand years) where there has been measurable evolution over the time period of sampling (101, 102) and is greatly complicated by the time-dependent nature of virus evolution (103). However, the broad-scale match between host and virus phylogenies generated by metagenomic data (7), backed-up by the phylogenetic distribution of EVEs (104, 105), strongly suggests that these associations have been established over many millions of years, which constitutes a valuable calibration point to infer evolutionary timescales. While current data indicate that many families of RNA viruses span the entire evolutionary history of the vertebrates and likely the animals, whether these will extend for the time span of the eukaryotes as a whole remains to be seen, particularly as only a limited number of metagenomic studies of plant RNA viruses have been undertaken to date (106, 107). Similarly, finding a continuity of viral evolution among bacteria and viruses would be a major achievement.

Finally, another of the most interesting and important generalities stemming from recent metagenomic studies of the virosphere is that some virus families are particularly commonplace in metagenomic screens, notably astroviruses, caliciviruses, and picornaviruses, with the latter especially diverse and abundant (16, 17, 22, 45, 108). That these viruses are common in marine environments, either in seawater, freshwater, or marine animals such as fish, suggests that they are especially robust in harsh environments (45). There are a number of reasons why this could be. First, it is possible that specific phenotypic features of these viruses, particularly their small and compact icosahedral capsids, make them better able to survive in relatively tough conditions such as marine environments. Second, and related, viruses of this type may simply be of greater antiquity, which explains their high levels of diversity and abundance. Indeed, it seems plausible that their simple positive-sense RNA genomes that encode a single polyprotein were one of the earliest virus structures to evolve (84), and icosahedral capsids, which are very common in viruses, are based on a simple, stable, and hence likely ancient pattern of folding symmetry (109). It will be interesting to see if this generality is upheld with additional sampling.

CONCLUSIONS

Metagenomics has transformed our understanding of virology and will continue to do so as more hosts are sampled and both sequencing and computational technologies continually improve. Not only is the virosphere immense and continually changing, with the high mutation rate of RNA viruses ensuring that new variants are generated every day, but also much of our understanding of

virus diversity, evolution, and function is based on a relatively small number of exemplar cases in a limited number of hosts and may not withstand the vast increase in sampling that will occur in the near future. The challenge for those working in this area is not only to continue to describe this rich biodiversity, which likely requires the development of new techniques that are able to detect those sequences too divergent to be assayed by currently available techniques, but more importantly to determine the evolutionary and ecological rules that shape this diversity, as well as the function of the myriad viral gene products. Addressing these issues will require not only virus and other microbial metagenomic data but also a clear understanding of the host genes that are up- or downregulated upon infection. Conducting combined analyses of genomic sequence data of different sources will also be central to understanding how viruses interact with each other and with the other microbes that infect a host and how their diversity and evolution are structured by aspects of host immunity. For example, in the case of human influenza virus there is growing evidence that prior immunity (immune imprinting) plays a major role in mediating disease emergence and abundance (110). A key question for the future is determining how broadly effects of this sort impact RNA viruses. Although much of virology has necessarily considered the RNA virome as a distinct and self-contained entity, the reality will be very different, and it is critical to consider hosts and microbes in a more wholistic manner. In short, despite over 100 years of research, studies of RNA viruses are very much still in their infancy.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We apologize to those authors we could not cite due to space restrictions. This study was supported by the National Natural Science Foundation of China (grants 81861138003 and 81672057) and the Special National Project on Research and Development of Key Biosafety Technologies (2016YFC1201900 and 2016YFC1200101). E.C.H. is funded by an ARC Australian Laureate Fellowship (FL170100022).

LITERATURE CITED

1. Hendrix RW, Smith MCM, Burns RN, Ford ME, Hatfull GF. 1999. Evolutionary relationships among diverse bacteriophages and prophages: All the world's a phage. *PNAS* 96:2192–97
2. Angly FE, Felts B, Breitbart M, Salamon P, Edwards RA, et al. 2006. The marine viromes of four oceanic regions. *PLOS Biol.* 4:e368
3. Desnues C, Rodriguez-Brito B, Rayhawk S, Kelley S, Tran T, et al. 2008. Biodiversity and biogeography of phages in modern stromatolites and thrombolites. *Nature* 452:340–43
4. Edwards RA, Rohwer F. 2005. Viral metagenomics. *Nat. Rev. Microbiol.* 3:504–10
5. Paez-Espino D, Eloe-Fadrosh EA, Pavlopoulos GA, Thomas AD, Huntemann M, et al. 2016. Uncovering Earth's virome. *Nature* 536:425–430
6. Koonin EV, Dolja VV. 2018. Metaviromics: a tectonic shift in understanding virus evolution. *Virus Res.* 246:A1–3
7. Zhang YZ, Shi M, Holmes EC. 2018. Using metagenomics to characterize an expanding virosphere. *Cell* 172:1168–72
8. Zhang YZ, Wu WC, Shi M, Holmes EC. 2018. The diversity, evolution and origins of vertebrate RNA viruses. *Curr. Opin. Virol.* 31:9–16

9. Geoghegan JL, Holmes EC. 2017. Predicting virus emergence amidst evolutionary noise. *Open Biol.* 7:170189
10. Lecoq H. 2001. Discovery of the first virus, the tobacco mosaic virus: 1892 or 1898? *C. R. Acad. Sci. III* 324:929–33
11. Culley AI, Lang AS, Suttle CA. 2003. High diversity of unknown picorna-like viruses in the sea. *Nature* 424:1054–57
12. Culley AI, Lang AS, Suttle CA. 2006. Metagenomic analysis of coastal RNA virus communities. *Science* 312:1795–98
13. Suttle CA. 2005. Viruses in the sea. *Nature* 437:356–61
14. Dolja VV, Koonin EV. 2018. Metagenomics reshapes the concepts of RNA virus evolution by revealing extensive horizontal virus transfer. *Virus Res.* 244:36–52
15. Li CX, Shi M, Tian JH, Lin XD, Kang YJ, et al. 2015. Unprecedented RNA virus diversity in arthropods reveals the ancestry of negative-sense RNA viruses. *eLife* 4:e05378
16. Shi M, Lin XD, Tian JH, Chen LJ, Chen X, et al. 2016. Redefining the invertebrate virosphere. *Nature* 540:539–43
17. Shi M, Lin XD, Chen X, Tian JH, Chen LJ, et al. 2018. The evolutionary history of vertebrate RNA viruses. *Nature* 556:197–202
18. Babayan SA, Orton RJ, Streicker DG. 2018. Predicting reservoir hosts and arthropod vectors from evolutionary signatures in RNA virus genomes. *Science* 362:577–80
19. Shi M, Zhang YZ, Holmes EC. 2018. Meta-transcriptomics and the evolutionary biology of RNA viruses. *Virus Res.* 243:83–90
20. Qin XC, Shi M, Tian JH, Lin XD, Gao DY, et al. 2014. A tick-borne segmented RNA virus contains genome segments derived from unsegmented viral ancestors. *PNAS* 111:6744–49
21. Delwart E. 2012. Animal virus discovery: improving animal health, understanding zoonoses, and opportunities for vaccine development. *Curr. Opin. Virol.* 2:344–52
22. Eden J-S, Chisholm R-H, Bull RA, White PA, Holmes EC, Tanaka MM. 2017. Persistent infections in immunocompromised hosts are rarely sources of new pathogen variants. *Virus Evol.* 3:vex018
23. Kapoor A, Victoria J, Simmonds P, Slikas E, Chieochansin T, et al. 2008. A highly prevalent and genetically diversified *Picornaviridae* genus in South Asian children. *PNAS* 105:20482–87
24. Li L, Victoria J, Kapoor A, Naeem A, Shaukat S, et al. 2009. Genomic characterization of novel human parechovirus type. *Emerg. Infect. Dis.* 15:288–91
25. Victoria JG, Kapoor A, Dupuis K, Schnurr DP, Delwart EL. 2008. Rapid identification of known and new RNA viruses from animal tissues. *PLOS Pathog.* 4:e1000163
26. Wilson MR, Naccache SN, Samayoa E, Biagtan M, Bashir H, et al. 2014. Actionable diagnosis of neuroleptospirosis by next-generation sequencing. *N. Engl. J. Med.* 370:2408–17
27. Lipkin WI, Travis GH, Carbone KM, Wilson MC. 1990. Isolation and characterization of Borna disease agent cDNA clones. *PNAS* 87:4184–88
28. Mokili JL, Rohwer F, Dutilh BE. 2012. Metagenomics and future perspectives in virus discovery. *Curr. Opin. Virol.* 2:63–77
29. Reyes GR, Purdy MA, Kim JP, Luk KC, Young LM, et al. 1990. Isolation of a cDNA from the virus responsible for enterically transmitted non-A, non-B hepatitis. *Science* 247:1335–39
30. CDC (Cent. Dis. Control Prev.). 1999. Outbreak of Hendra-like virus—Malaysia and Singapore, 1998–1999. *MMWR Morb. Mortal. Wkly. Rep.* 48:265–69
31. Chua KB, Bellini WJ, Rota PA, Harcourt BH, Tamin A, et al. 2000. Nipah virus: a recently emergent deadly paramyxovirus. *Science* 288:1432–35
32. Drosten C, Günther S, Preiser W, van der Werf S, Brodt HR, et al. 2003. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N. Engl. J. Med.* 348:1967–76
33. Zhang Y-Z, Zhou DJ, Xiong Y, Chen XP, He YW, et al. 2011. Hemorrhagic fever caused by a novel tick-borne Bunyavirus in Huaiyangshan, China. *Zhonghua Liu Xing Bing Xue Za Zhi* 32:209–20
34. Woo PC, Lau SK, Lam CS, Lau CC, Tsang AK, et al. 2012. Discovery of seven novel mammalian and avian coronaviruses in the genus *Deltacoronavirus* supports bat coronaviruses as the gene source of *Alphacoronavirus* and *Betacoronavirus* and avian coronaviruses as the gene source of *Gammacoronavirus* and *Deltacoronavirus*. *J. Virol.* 86:3995–4008

35. Guo WP, Lin XD, Wang W, Tian JH, Cong ML, et al. 2013. Phylogeny and origins of hantaviruses harbored by bats, insectivores, and rodents. *PLOS Pathog.* 9:e1003159
36. Drexler JF, Corman VM, Müller MA, Lukashev AN, Gmyl A, et al. 2013. Evidence for novel hepaciviruses in rodents. *PLOS Pathog.* 9:e1003438
37. Tong S, Zhu X, Li Y, Shi M, Bourgeois M, et al. 2013. New World bats harbor diverse influenza A viruses. *PLOS Pathog.* 9:e1003657
38. Drexler JF, Corman VM, Müller MA, Maganga GD, Vallo P, et al. 2012. Bats host major mammalian paramyxoviruses. *Nat. Commun.* 3:796
39. Firth C, Lipkin WI. 2013. The genomics of emerging pathogens. *Annu. Rev. Genom. Hum. Genet.* 14:281–300
40. Greningera AL. 2018. A decade of RNA virus metagenomics is (not) enough. *Virus Res.* 244:218–29
41. Conceicao-Neto N, Zeller M, Lefrere H, De Bruyn P, Beller L, et al. 2015. Modular approach to customise sample preparation procedures for viral metagenomics: a reproducible protocol for virome analysis. *Sci. Rep.* 5:16532
42. Djikeng A, Halpin R, Kuzmickas R, DePasse J, Feldblyum J, et al. 2008. Viral genome sequencing by random priming methods. *BMC Genom.* 9:5
43. Lim YW, Schmieder R, Haynes M, Willner D, Furlan M, et al. 2013. Metagenomics and metatranscriptomics: windows on CF-associated viral and microbial communities. *J. Cyst. Fibros.* 12:154–64
44. Koonin EV. 1991. The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. *J. Gen. Virol.* 72:2197–206
45. Geoghegan JL, Di Giallonardo F, Cousins K, Shi M, Williamson JE, Holmes EC. 2018. Hidden diversity and evolution of viruses in market fish. *Virus Evol.* 4:vey031
46. Wilson MR, O'Donovan BD, Gelfand JM, Sample HA, Chow FC, et al. 2018. Chronic meningitis investigated via metagenomic next-generation sequencing. *JAMA Neurol.* 75:947–55
47. Naccache SN, Greninger AL, Lee D, Coffey LL, Phan T, et al. 2013. The perils of pathogen discovery: origin of a novel parvovirus-like hybrid genome traced to nucleic acid extraction spin columns. *J. Virol.* 87:11966–77
48. Bolduc B, Shaughnessy DP, Wolf YI, Koonin EV, Roberto FF, Young M. 2012. Identification of novel positive-strand RNA viruses by metagenomic analysis of archaea-dominated Yellowstone hot springs. *J. Virol.* 86:5562–73
49. Bamford DH, Grimes JM, Stuart DI. 2005. What does structure tell us about viral evolution? *Curr. Opin. Struct. Biol.* 15:655–63
50. Fédry J, Liu Y, Péhau-Arnaudet G, Pei J, Li W, et al. 2017. The ancient gamete fusogen HAP2 is a eukaryotic class II fusion protein. *Cell* 168:904–15
51. Colmant AMG, Hobson-Peters J, Bielefeldt-Ohmann H, van den Hurk AF, Hall-Mendelin S, et al. 2017. A new clade of insect-specific flaviviruses from Australian *Anopheles* mosquitoes displays species-specific host restriction. *mSphere* 2:e00262-17
52. Junglen S, Drosten C. 2013. Virus discovery and recent insights into virus diversity in arthropods. *Curr. Opin. Microbiol.* 16:507–13
53. Marklewitz M, Zirkel F, Kurth A, Drosten C, Junglen S. 2015. Evolutionary and phenotypic analysis of live virus isolates suggests arthropod origin of a pathogenic RNA virus family. *PNAS* 112:7536–41
54. Remnant EJ, Shi M, Buchmann G, Blacquière T, Holmes EC, et al. 2017. A diverse range of novel RNA viruses in geographically distinct honey bee populations. *J. Virol.* 91:e00158-17
55. Shi M, Neville P, Nicholson J, Eden JS, Imrie A, Holmes EC. 2017. High-resolution metatranscriptomics reveals the ecological dynamics of mosquito-associated RNA viruses in Western Australia. *J. Virol.* 91:e00680-17
56. Tokarz R, Sameroff S, Tagliafierro T, Jain K, Williams SH, et al. 2018. Identification of novel viruses in *Amblyomma americanum*, *Dermacentor variabilis*, and *Ixodes scapularis* ticks. *mSphere* 3:e00614-17
57. Tokarz R, Williams SH, Sameroff S, Sanchez Leon M, Jain K, Lipkin WI. 2014. Virome analysis of *Amblyomma americanum*, *Dermacentor variabilis*, and *Ixodes scapularis* ticks reveals novel highly divergent vertebrate and invertebrate viruses. *J. Virol.* 88:11480–92
58. Webster CL, Waldron FM, Robertson S, Crowson D, Ferrari G, et al. 2015. The discovery, distribution, and evolution of viruses associated with *Drosophila melanogaster*. *PLOS Biol.* 13:e1002210

59. Colmant AMG, Hall-Mendelin S, Ritchie SA, Bielefeldt-Ohmann H, Harrison JJ, et al. 2018. The recently identified flavivirus Bamaga virus is transmitted horizontally by *Culex* mosquitoes and interferes with West Nile virus replication *in vitro* and transmission *in vivo*. *PLOS Negl. Trop. Dis.* 12:e0006886
60. Hall-Mendelin S, McLean BJ, Bielefeldt-Ohmann H, Hobson-Peters J, Hall RA, et al. 2016. The insect-specific Palm Creek virus modulates West Nile virus infection in and transmission by Australian mosquitoes. *Parasit. Vectors* 9:414
61. Holmes EC. 2011. The evolution of endogenous viral elements. *Cell Host Microbe* 10:368–77
62. Whitfield ZJ, Dolan PT, Kunitomi M, Tassetto M, Seetin MG, et al. 2017. The diversity, structure, and function of heritable adaptive immunity sequences in the *Aedes aegypti* genome. *Curr. Biol.* 27:3511–19
63. Bennett AJ, Bushmaker T, Cameron K, Ondzie A, Niama FR, et al. 2018. Diverse RNA viruses of arthropod origin in the blood of fruit bats suggest a link between bat and arthropod viromes. *Virology* 528:64–72
64. Roossinck MJ. 2015. Plants, viruses and the environment: ecology and mutualism. *Virology* 479–80:271–77
65. Olival KJ, Hosseini PR, Zambrana-Torrelio C, Ross N, Bogich TL, Daszak P. 2017. Host and viral traits predict zoonotic spillover from mammals. *Nature* 546:646–50
66. Dill JA, Camus AC, Leary JH, Di Giallonardo F, Holmes EC, Ng TFF. 2016. Distinct viral lineages of hepadnavirus from fish and amphibians reveal the complex evolutionary history of hepadnaviruses. *J. Virol.* 90:7920–33
67. Lauber C, Seitz S, Mattei S, Suh A, Beck J, et al. 2017. Deciphering the origin and evolution of hepatitis B viruses by means of a family of non-enveloped fish viruses. *Cell Host Microbe* 22:387–99
68. Suh A, Weber CC, Kehlmaier C, Braun EL, Green RE, et al. 2014. Early Mesozoic coexistence of amniotes and Hepadnaviridae. *PLOS Genet.* 10:e1004559
69. Allison AB, Ballard JR, Tesh RB, Brown JD, Ruder MG, et al. 2015. Cyclic avian mass mortality in the northeastern United States is associated with a novel orthomyxovirus. *J. Virol.* 89:1389–403
70. Geoghegan JL, Duchêne S, Holmes EC. 2017. Comparative analysis estimates the relative frequencies of co-divergence and cross-species transmission within viral families. *PLOS Pathog.* 13:e1006215
71. Bacharach E, Mishra N, Briese T, Zody MC, Kembou Tsofack JE, et al. 2016. Characterization of a novel orthomyxo-like virus causing mass die-offs of tilapia. *mBio* 7:e00431-16
72. Taylor J, Pelchat M. 2010. Origin of hepatitis Delta virus. *Future Microbiol.* 5:393–402
73. Wille M, Netter HJ, Littlejohn M, Yuen L, Shi M, et al. 2018. A divergent hepatitis D-like virus in birds. *Viruses* 10:720
74. Woolhouse MEJ, Brierley L. 2018. Epidemiological characteristics of human-infective RNA viruses. *Sci. Data* 5:180017
75. Tan SK, Relman DA, Pinsky BA. 2017. The human virome: implications for clinical practice in transplantation medicine. *J. Clin. Microbiol.* 55:2884–93
76. Holmes EC, Rambaut A, Andersen KG. 2018. Pandemics: spend on surveillance, not prediction. *Nature* 558:180–82
77. Wilson MR, Suan D, Duggins A, Schubert RD, Khan LM, et al. 2017. A novel cause of chronic viral meningoencephalitis: Cache Valley virus. *Ann. Neurol.* 82:105–14
78. Shi M, Lin XD, Vasilakis N, Tian JH, Li CX, et al. 2016. Divergent viruses discovered in arthropods and vertebrates revise the evolutionary history of the *Flaviviridae* and related viruses. *J. Virol.* 90:659–69
79. Ladner JT, Wiley MR, Beitzel B, Auguste AJ, Dupuis AP II, et al. 2016. A multicomponent animal virus isolated from mosquitoes. *Cell Host Microbe* 20:357–67
80. Gorbalenya AE, Enjuanes L, Ziebuhr J, Snijder EJ. 2006. *Nidovirales*: evolving the largest RNA virus genome. *Virus Res.* 117:17–37
81. Saberi A, Gulyaeva AA, Brubacher JL, Newmark PA, Gorbalenya AE. 2018. A planarian nidovirus expands the limits of RNA genome size. *PLOS Pathog.* 14:e1007314
82. Simon-Loriere E, Holmes EC. 2013. Gene duplication is infrequent in the recent evolutionary history of RNA viruses. *Mol. Biol. Evol.* 30:1263–69
83. Meyers G, Rumenapf T, Thiel HJ. 1989. Ubiquitin in a togavirus. *Nature* 341:491
84. Holmes EC. 2009. *The Evolution and Emergence of RNA Viruses*. New York: Oxford Univ. Press

85. Ochman H, Lawrence JG, Groisman EA. 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature* 405:299–304
86. Simmonds P, Adams MJ, Benkő M, Breitbart M, Brister JR, et al. 2017. Consensus statement: virus taxonomy in the age of metagenomics. *Nat. Rev. Microbiol.* 15:161–68
87. Bobay LM, Ochman H. 2018. Biological species in the viral world. *PNAS* 115:6040–45
88. Bexfield N, Kellam P. 2011. Metagenomics and the molecular identification of novel viruses. *Vet. J.* 190:191–98
89. Delwart EL. 2007. Viral metagenomics. *Rev. Med. Virol.* 17:115–31
90. Rosario K, Breitbart M. 2011. Exploring the viral world through metagenomics. *Curr. Opin. Virol.* 1:289–97
91. Tang P, Chiu C. 2010. Metagenomics for the discovery of novel human viruses. *Future Microbiol.* 5:177–89
92. Zanotto PM, Gibbs MJ, Gould EA, Holmes EC. 1996. A reevaluation of the higher taxonomy of viruses based on RNA polymerases. *J. Virol.* 70:6083–96
93. Wolf YI, Kazlauskas D, Iranzo J, Lucía-Sanz A, Kuhn JH, et al. 2018. Origins and evolution of the global RNA virome. *mBio* 9:e02329-18
94. Fredricks DN, Relman DA. 1996. Sequence-based identification of microbial pathogens: a reconsideration of Koch's postulates. *Clin. Microbiol. Rev.* 9:18–33
95. Wille M, Eden J-S, Shi M, Klaassen M, Hurt AC, Holmes EC. 2018. Virus–virus interactions and host ecology are associated with RNA virome structure in wild birds. *Mol. Ecol.* 27:5263–78
96. Moreira LA, Iturbe-Ormaetxe I, Jeffery JA, Lu G, Pyke AT, et al. 2009. A *Wolbachia* symbiont in *Aedes aegypti* limits infection with dengue, Chikungunya, and *Plasmodium*. *Cell* 139:1268–78
97. Tan CH, Wong PJ, Li MI, Yang H, Ng LC, O'Neill SL. 2017. *w*Mel limits zika and chikungunya virus infection in a Singapore *Wolbachia*-introgressed *Ae. aegypti* strain, *w*Mel-Sg. *PLOS Negl. Trop. Dis.* 11:e0005496
98. Shi M, White VL, Schlub T, Eden J-S, Hoffmann AA, Holmes EC. 2018. No detectable effect of *Wolbachia w*Mel on the prevalence and abundance of the RNA virome of *Drosophila melanogaster*. *Proc. R. Soc. B* 285:20181165
99. Nambulli S, Sharp CR, Acciardo AS, Drexler JF, Duprex WP. 2016. Mapping the evolutionary trajectories of morbilliviruses: what, where and whither. *Curr. Opin. Virol.* 16:95–105
100. Wang LF, Walker PJ, Poon LL. 2011. Mass extinctions, biodiversity and mitochondrial function: Are bats 'special' as reservoirs for emerging viruses? *Curr. Opin. Virol.* 1:649–57
101. Drummond AJ, Pybus OG, Rambaut A, Forsberg R, Rodrigo AG. 2003. Measurably evolving populations. *Trends Ecol. Evol.* 18:481–88
102. Biek R, Pybus OG, Lloyd-Smith JO, Didelot X. 2015. Measurably evolving pathogens in the genomic era. *Trends Ecol. Evol.* 30:306–13
103. Duchêne S, Holmes EC, Ho SYW. 2014. Analyses of evolutionary dynamics in viruses are hindered by a time-dependent bias in rate estimates. *Proc. R. Soc. B* 281:20140732
104. Katzourakis A, Gifford RJ. 2010. Endogenous viral elements in animal genomes. *PLOS Genet.* 6:e1001191
105. Aiewsakun P, Katzourakis A. 2015. Endogenous viruses: connecting recent and ancient viral evolution. *Virology* 479–480:26–37
106. Roossinck MJ, Martin DP, Roumagnac P. 2015. Plant virus metagenomics: advances in virus discovery. *Phytopathology* 105:716–27
107. Mushegian A, Shipunov A, Elena SF. 2016. Changes in the composition of the RNA virome mark evolutionary transitions in green plants. *BMC Biol.* 14:68
108. Miranda JA, Culley AI, Schvarcz CR, Steward GF. 2016. RNA viruses as major contributors to Antarctic virioplankton. *Environ. Microbiol.* 18:3714–27
109. Krupovic M, Koonin EV. 2017. Multiple origins of viral capsid proteins from cellular ancestors. *PNAS* 114:E2401–10
110. Gostic KM, Ambrose M, Worobey M, Lloyd-Smith JO. 2016. Potent protection against H5N1 and H7N9 influenza via childhood hemagglutinin imprinting. *Science* 354:722–26

Annual Review
of Virology

Volume 6, 2019

Contents

Genome Replication, Regulation of Gene Expression, and Biosynthesis

Virus Cell Biology

Transformation and Oncogenesis

Pathogenesis