Deep Recombination: RNA and ssDNA Virus Genes in DNA Virus and Host Genomes

Kenneth M. Stedman

Biology Department and Center for Life in Extreme Environments, Portland State University, Portland, Oregon 97207; email: kstedman@pdx.edu

Annu. Rev. Virol. 2015. 2:203–17

First published online as a Review in Advance on September 2, 2015

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

This article's doi: 10.1146/annurev-virology-100114-055127

Keywords

chimeric viruses, hybrid viruses, RNA-DNA recombination, virus evolution

Abstract

Viruses are notorious for rapidly exchanging genetic information between close relatives and with the host cells they infect. This exchange has profound effects on the nature and rapidity of virus and host evolution. Recombination between dsDNA viruses is common, as is genetic exchange between dsDNA viruses or retroviruses and host genomes. Recombination between RNA virus genomes is also well known. In contrast, genetic exchange across viral kingdoms, for instance between nonretroviral RNA viruses or ssDNA viruses and host genomes or between RNA and DNA viruses, was previously thought to be practically nonexistent. However, there is now growing evidence for both RNA and ssDNA viruses recombining with host dsDNA genomes and, more surprisingly, RNA virus genes recombining with ssDNA virus genomes. Mechanisms are still unclear, but this deep recombination greatly expands the breadth of virus evolution and confounds virus taxonomy.

INTRODUCTION

Virus Classification

In 1971, David Baltimore presciently proposed six classes of viruses based on the form of their packaged genomes and the mechanism they use to make mRNA. This classification is widely used to this day (10) (**Figure 1**). Negative-strand ssRNA viruses, like Ebola virus, are Class V; dsDNA viruses, like camelpox virus, are Class I. The remaining four classes are ssDNA viruses (Class II), positive-strand ssRNA viruses (Class IV), dsRNA viruses (Class III), and retroviruses, which replicate using reverse transcription (Class VI). There appears to be considerable recombination, both homologous and illegitimate, between closely related viruses, much less between viruses of different families, and practically none between different classes of viruses (11).

Chimeric Viruses

A chimeric or hybrid virus refers to a virus that is composed of genetic parts from two different parent viruses (see sidebar, Chimera Versus Hybrid). The term chimeric or chimera virus has been widely used to describe the cause of the next fictional pandemic (1–4) as well as some very promising vaccine candidates (5–7). Chimeric viruses tend to appear in fiction as a force that will

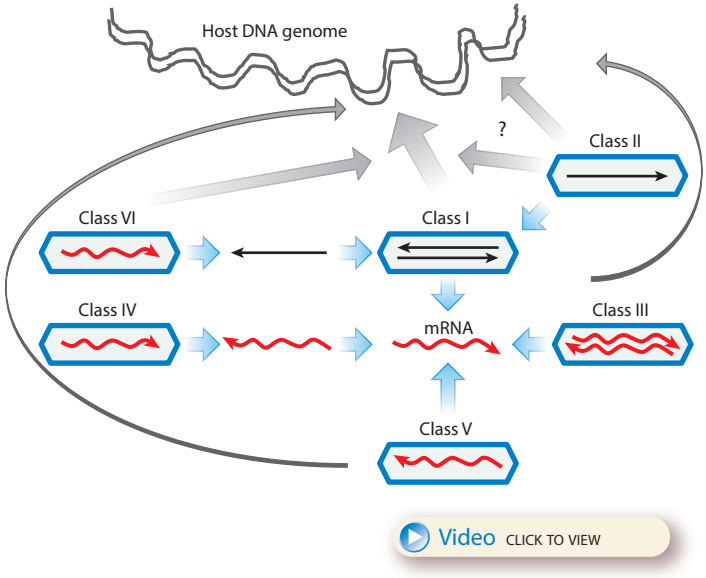

Figure 1

Schematic of different virus classes [after Baltimore 1971 (10)] and the integration of some viruses into cellular genomes. Black arrows represent DNA. Red arrows represent RNA. Blue hexagons represent virus capsids. Blue arrows represent virus replication mechanisms to produce mRNA for protein translation. Gray arrows represent integration into the host genome, with the thickness of the arrow representing the gross frequency of recombination. The recombination arrow from Class VI viruses, the retroviruses, proceeds through a dsDNA intermediate. Recombination between ssDNA viruses (Class II) and dsDNA genomes may proceed through dsDNA (*lower arrow*) or directly as ssDNA (*upper arrow*) (55). Recombination between Class V and Class III viruses is probably mediated by cellular retrotransposons (see **Figure 2**). For an animation of the schematic, see **Video 1** or download a PowerPoint slideshow.

CHIMERA VERSUS HYBRID

The Chimera is a fire-breathing creature of Greek legend with the body of a goat between the head of a lion and the tail of a snake. In biology, a chimera is an organism made up of cells derived from two different zygotes. For example, in the generation of a transgenic mouse, an embryonic stem cell derived from one mouse is injected into the blastocyst of another mouse with a different genetic background, and a chimeric mouse then develops. Molecular chimeras can be formed by polymerase chain reaction (PCR) as a product derived from two different sources. This is a significant problem in PCR-based surveys of molecular diversity. Apparent diversity could have been generated by the PCR reaction itself and not by actual underlying diversity. For this reason, standard procedures require a check for molecular chimeras.

A hybrid, on the other hand, generally describes the product of mating between two genetically different individuals of the same or similar species. These hybrids may or may not be reproductively viable. The classic example of a hybrid animal is the mule.

kill off most humans and possibly all life on this planet unless some charismatic hero or heroine saves us all. The chimeric viruses used in vaccines are recombinant combinations of genes from closely related viruses—in the case of ChimeriVax, from different flaviviruses (6). A virus is typically described as chimeric when it is derived from viruses that otherwise infect different hosts. The most famous, or infamous, chimeric viruses are the influenza viruses, which often undergo genome reassortment, whereby genome segments from different influenza viruses are incorporated into a new influenza virus with a different host range, immune reactivity, pathogenicity, or all of the above (8). The genome reassortment process in influenza is well understood and bears superficial conceptual resemblance to independent assortment of eukaryotic chromosomes. Some more far-fetched fictional examples are chimeras of Ebola virus and camelpox virus (2). This combination seems particularly unlikely given that Ebola virus is a negative-strand ssRNA virus that replicates in the absence of a DNA intermediate (Class V) and camelpox (like all poxviruses) is a dsDNA virus that encodes its own DNA replicase and does not replicate its genome using an RNA intermediate (Class I). Until the past few years, it was thought that RNA viruses did not recombine at all with DNA viruses (9).

Recombination of Virus Genomes with Host Genomes

Most retroviruses must integrate into cellular genomes in order to replicate, so there is extensive evidence of retroviruses and retroviral elements in many genomes. For instance, up to 8% of the human genome is clearly derived from retrovirus insertion (12). On the flip side, many, if not most, of the oncogenic retroviruses—also known as RNA tumor viruses—have incorporated host oncogenes into their genomes. It is also not surprising that there is extensive evidence of dsDNA viruses or their genes in cellular dsDNA genomes; almost all known genomes contain such proviruses (13). Moreover, there is extensive recombination, both homologous and illegitimate, between different dsDNA viruses, particularly the tailed bacteriophages (14–17). Until recently, there was little to no evidence for insertion of either nonretroviral RNA viruses or ssDNA viruses into cellular DNA.

However, evidence is now emerging for nonretroviral RNA and ssDNA virus genomes that have integrated at least in part into cellular dsDNA genomes. Moreover, a number of recently discovered ssDNA viruses appear to have acquired RNA genes by recombination with coinfecting RNA viruses (9, 18–20). This review examines these examples, explores potential mechanisms for

this deep recombination, and proposes future work to determine the prevalence and provenance of these chimeric viruses.

UNEXPECTED VIRUS GENES IN CELLULAR GENOMES

RNA Virus Genomes in Cellular Genomes

In the mid-1970s Zhdanov and coworkers (21–23) reported the presence of measles, tick-borne encephalitis, and Sindbis viruses—viruses that use only RNA for replication—incorporated into cellular dsDNA genomes. All were in long-term [as long as 15 years (23)] cell culture systems. The presence of genomic DNA copies of RNA viruses was confirmed by DNA-RNA hybridizations and CsCl gradients. The authors predicted that the genomes of these RNA viruses had been incorporated by the activity of reverse transcriptases from coinfecting retroviruses (21). Unfortunately, their experiments were apparently difficult to reproduce and were widely ignored (24, 25). However, these early reports were corroborated, at least in part, almost 25 years later with the advent of polymerase chain reaction (PCR) amplification and whole-genome sequencing. First, DNA complementary to the genome of lymphocytic choriomeningitis virus (LCMV), a segmented ssRNA virus, was found in mouse spleen DNA by PCR after infection, but no RNA could be detected (26). As anticipated by Zhdanov, the incorporation of RNA into DNA was inhibited by the presence of the reverse transcriptase inhibitor azidothymidine (26). For LCMV the process appears to include recombination with a retrotransposon (27). About 50% of the genome of a flavivirus-like (positive-strand ssRNA) virus was found in mosquito dsDNA genomes also using PCR (25).

With the advent of whole-genome sequencing, a number of extensive bioinformatics searches have been undertaken to determine whether RNA virus sequences are present in cellular genomes. Importantly, given the fragmented nature of many host genome sequences and the ease of contamination, the best studies have also confirmed junctions between RNA virus–like sequences and the cellular genome. For instance, Taylor & Bruenn (28) found a number of totivirus-like sequences in diverse fungal genomes (28). Totiviruses are very simple dsRNA viruses that encode an RNA-dependent RNA polymerase and a capsid protein. Similarly, many partitivirus-like sequences were found in plant genomes, including in the model plant *Arabidopsis thaliana*; partitiviruses are also dsRNA viruses (29). In 2010, three independent bioinformatics studies found multiple nonretrovirus RNA virus–like sequences in animal, particularly vertebrate, genomes (30–32).

Bornavirus-like sequences in cellular genomes. The most common RNA virus–like sequences found in mammalian genomes are similar to bornavirus genes. Bornaviruses are negative-strand enveloped ssRNA viruses that, unusually for RNA viruses, replicate exclusively in the nucleus of host cells (reviewed in 33). The mRNA from the N protein gene and, less frequently, the P protein gene from bornaviruses seems to have been incorporated into diverse animal genomes, from primates to insects (30–33). By phylogenetic analysis, the integration event seems to have taken place independently in each lineage. One insertion, in the thirteen-lined ground squirrel, seems to have occurred quite recently. By contrast, some of the insertions seem to have taken place in common ancestors of current genomes more than 65 million years ago. Horie et al. (30) also showed that in cell cultures stably infected with bornaviruses, the N protein gene incorporates into different places in the host genome. However, they were not able to detect integration during infection (33).

Filovirus-like sequences in cellular genomes. In addition to bornavirus-like sequences, filovirus-like sequences were also detected in host genomes by bioinformatics surveys of vertebrate genomes (31, 34). Filoviridae is infamous for being the virus family that includes Ebola

virus. Five vertebrate genomes have been found to contain sequences similar to the filovirus N protein gene, three with sequences similar to the filovirus GP35 gene, and one with a sequence similar to the filovirus L protein gene (31). Like bornaviruses, filoviruses are negative-strand enveloped ssRNA viruses, but unlike bornaviruses, they replicate in the cytoplasm of infected host cells. Filovirus-like sequences have been found in a broad range of mammals, including many that are not known to be filovirus hosts (32, 34).

How Do RNA Virus Sequences Get Into Cellular DNA Genomes?

The insertion of partial RNA virus genomes into host genomes is thought to proceed through the activity of endogenous retroelements, similar to integration of processed pseudogenes (35, 36). This model is based on the presence of poly(A) sequences adjacent to RNA virus genes in host genomes as well as terminal direct repeat (TDR) sequences flanking them, as would be expected if integration were mediated by LINE-1 (33, 35) (**Figure 2**). Bornaviruses replicate in the nucleus, potentially giving them ample opportunity to integrate into the genome or associate with endogenous retroelements. That single, highly expressed genes are incorporated supports the LINE-1-mediated integration model. The mechanisms for the integration of other RNA

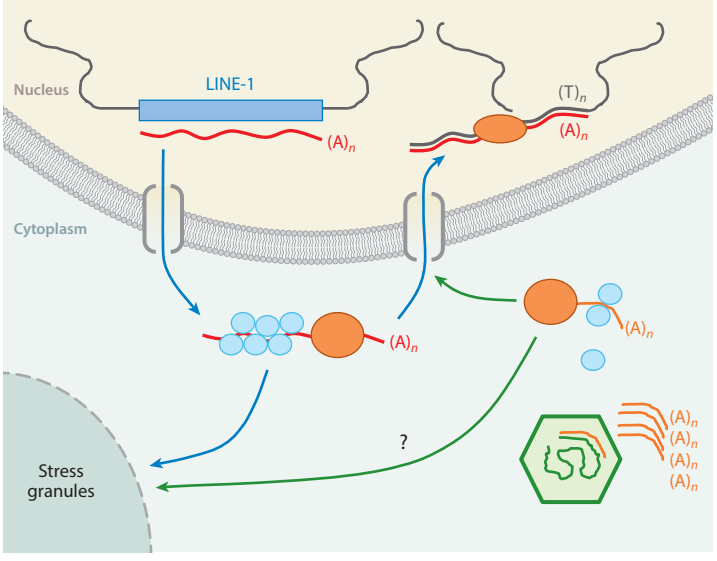

Figure 2

Proposed LINE-1-mediated mechanism of integration of RNA virus genes into host genomes. LINE-1 elements in the host genome (*blue rectangle*) are transcribed into mRNA (*red line*). The resulting mRNAs are exported to the cytoplasm, where they combine with ORF1 (*light blue circles*) and ORF2 (*orange ovals*), the products of the two LINE-1 open reading frames. These complexes can be either relocated to stress granules or reimported back into the nucleus (*blue arrows*), where they can recombine with the host genome (*dark gray lines*) by reverse transcription. RNA viruses (*green hexagon*) infect the cell and produce mRNA (*orange lines*) that can bind to the ORF products of LINE-1 and be imported to the nucleus for reverse transcription or, possibly, relocated to stress granules (*green arrows*). For an animation of the mechanism, see **Video 2** or download a PowerPoint slideshow.

virus genomes or parts thereof into host cellular genomes are less clear but must involve a reverse transcriptase. Curiously, most of the exogenous viral elements (EVEs) found by Katzourakis & Gifford (32) in insects did not have TDRs or poly(A) sequences. However, this lack may be due to different retrotransposon activities in the host genomes, which may not generate TDRs or produce flanking poly(A) sequences (37).

The prevailing model for LINE-1 retrotransposition, target-primed reverse transcription, has multiple steps (38) (**Figure 2**). First, the LINE-1 element is transcribed into mRNA; the mRNA is then transported to the cytoplasm, where both open reading frames (ORFs), ORF1 and ORF2, are translated. Next, the gene products assemble on the LINE-1 mRNA and are imported back into the nucleus. There, the product of LINE-1 ORF2, an endonuclease/reverse transcriptase, makes a nick in the genome at a poorly conserved target site (5′-TTTT/AAA-3′) (39, 40). The thymidine residues in the DNA then pair with the poly(A) tail of the LINE-1 mRNA. The reverse transcriptase activity of LINE-1 ORF2 makes a DNA copy of the LINE-1 mRNA using the 3′-OH generated by the nick as a primer. After that, the second strand is made, presumably by the product of LINE-1 ORF2. The dsDNA is subsequently inserted into the host genome. Finally, any gaps generated by the nicking process are repaired by cellular DNA repair machinery. If the nick is downstream of the site used for reverse transcription, then a target site duplication or TDR is created, but if the nick is upstream, the target site could be deleted (37).

Subcellular compartments or foci containing RNA and RNA-binding proteins have been implicated as being extremely important for RNA processing and translation, as well as for LINE-1-mediated retrotransposition and the control thereof (41). The best known of these foci are the stress granules and perinuclear (P) granules. An excellent review on the positive and negative interplay between these granules and viruses was published recently (42). It is possible that bornaviruses and filoviruses disrupt stress and/or P granules in a manner that leads to the mRNAs of these viruses being integrated into the genome as EVEs more frequently than are those of other viruses. Curiously, there seems to be more LINE-1 activity in aging cells (43). Thus, there could be more RNA virus insertion in aging somatic cells than in younger ones.

An alternative putative mechanism for integration of RNA viruses into host genomes is that RNA recombination occurs between viral RNA and the mRNA of a retrotransposon before the retrotransposon is converted to DNA. This process seems to have occurred in grapevines (44) and in cells infected with LCMV (27).

Why Are RNA Virus Sequences Maintained in Host Genomes?

Curiously, of the nonretroviral RNA viruses, sequences similar only to bornaviruses and filoviruses are found in vertebrate animal genomes, despite the wide net cast in bioinformatics studies (31). This immediately raises the question of why bornavirus-like and filovirus-like RNAs are incorporated or maintained in host genomes to the exclusion of sequences from other viruses. The proposed mechanism, LINE-1-mediated insertion (30, 31, 33) (**Figure 2**), predicts that many different RNAs, both cellular and viral, would be incorporated. It may be that all virus sequences that are not positively selected are lost. The presence of presumably unselected cellular pseudogenes and the absence of viral pseudogenes, however, argue against this. A more likely hypothesis is that the integrated RNA virus gene provides defense against future infection by the same or a similar virus. A clear example of this is bees harboring parts of the Israeli acute paralysis virus (a positive-strand ssRNA virus) in their genomes. Bees that had incorporated some part of the structural genes of the viral genome were resistant to future infection (45). Whether other EVEs also confer resistance is less clear (33, 46). Given the severe outcomes of diseases caused by filoviruses and bornaviruses, there could be considerable selective pressure for maintenance of genes that protect from

infection. In *Drosophila*, RNA virus infection leads to production and genome insertion of small cDNA fragments that can modulate virus infection through RNA interference (47). Thus it appears that reverse transcription of RNA viruses, even those that are not retroviruses, is quite prevalent.

It is possible, if not probable, that there are many other RNA virus genomes, or fragments thereof, that were inserted into host genomes and have since mutated beyond recognition. In addition, only a very small proportion of the viruses in the biosphere have been characterized (48–50). All of the studies published to date on EVEs used database searches of known virus genomes to find these integrated viruses. If there is no annotated sequence in the database, then an integrated virus will escape detection. It is unclear how many integrated viruses are missed in this way, but the number could be vast.

ssDNA Virus Genomes in Host Genomes

Similar to RNA virus genomes, there are a few examples of either complete or partial ssDNA virus genomes in host genomes. The best characterized of these are the parvoviruses (51). Due to its use in gene therapy, probably the most studied of the parvoviruses is adeno-associated virus 2 (AAV2) (52). AAV2 integrates its complete genome into human chromosome 19 (53). Similarly, the ssDNA bacteriophage that encodes cholera toxin in *Vibrio cholerae* (CTXφ) also inserts into the host genome (54, 55). In addition to genes from parvoviruses (56, 57), bioinformatics surveys of animal genomes for ssDNA virus–like sequences have found a gene similar to the circovirus rolling circle replication initiation protein (Rep) gene (32) (**Figure 3**). Circoviruses are small circular ssDNA viruses common in numerous and diverse environments (58, 59). A broad survey of other eukaryotic genomes (60) also found many ssDNA Rep gene homologs, but very few capsid protein gene homologs, in eukaryote genomes. One major caveat of this and other bioinformatics studies is that capsid protein genes are notoriously difficult to identify by sequence similarity searches (61).

Mechanism of ssDNA Virus Insertion into Host Genomes

Parvoviruses are proposed to integrate at host genome sequences similar to their viral replication origins (62). The proposed mechanism for incorporation of circovirus *rep* sequences is analogous: There is a DNA sequence in the host genome that is similar to the origin of replication of the circovirus genome (hairpin loop), and this feature is recognized by the Rep protein (60) (**Figure 3**). In contrast, microviruses, ssDNA viruses of bacteria, have been proposed to integrate in a manner similar to the ssDNA cholera phage, by harnessing the cellular XerC/D mechanism normally used for resolution of chromosome dimers after replication (55, 63).

HYBRID OR CHIMERIC VIRUSES

In addition to these examples of ssDNA and nonretroviral RNA virus genes and genomes that have been incorporated into cellular genomes, there are a few examples of ssDNA and RNA virus genes that have recombined with other virus genomes.

ssDNA Virus Genes in dsDNA Virus Genomes

Much like in cellular dsDNA genomes, there are a few examples of ssDNA virus–like sequences in dsDNA virus genomes. A parvovirus Rep gene homolog was found in the sequence of human herpesvirus 6 (HHV-6), but not in the very closely related HHV-7 (64). The HHV-6 gene, *U94*,

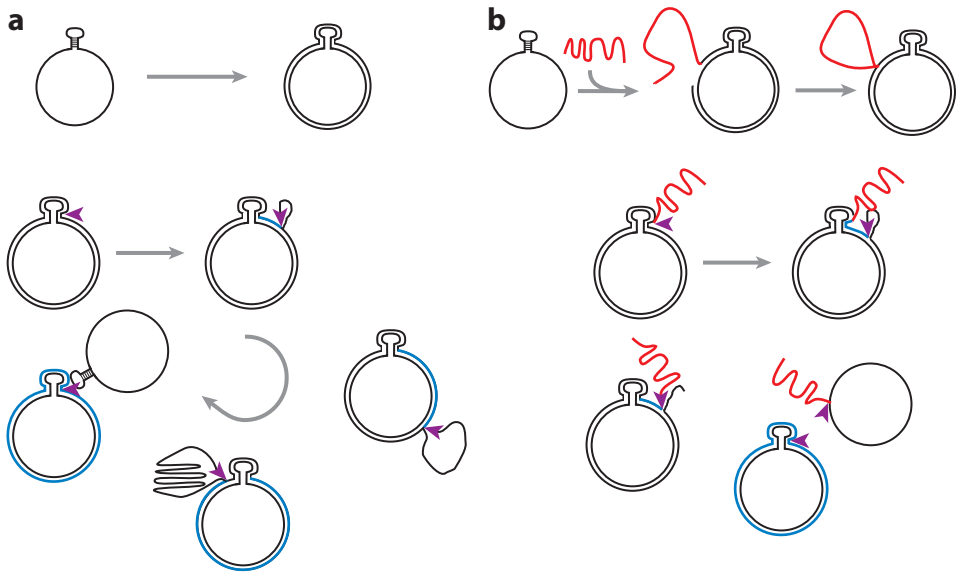

Figure 3

Replication of circular ssDNA viruses and possible insertion of RNA. Blue represents newly replicated DNA. (*a*) Replication of ssDNA viruses. After infection, the ssDNA virus genome is made into dsDNA by cellular DNA polymerases. The dsDNA is cut by the Rep protein (*purple arrowhead*), and the 5′ end of the nontemplate strand is covalently attached to a conserved tyrosine in Rep. The 3′-OH generated by Rep provides a primer for cellular DNA polymerases that replicate around the genome until a complete genome has been made. At this point Rep reverses its activity, a covalently closed DNA genome is released, and the process repeats. (*b*) RNA capture. RNA (*red*) could be incorporated into the ssDNA genome at the first step of replication by serving as a primer for second-strand synthesis. RNA could also be used as the primer for ssDNA synthesis and either become covalently attached to Rep or be incorporated at the end of replication. Figure modified with permission from Stedman (75).

can complement AAV2 *rep* mutants (65). The U94 gene product is thought to act in the latent phase of infection and may protect against infection by other herpesviruses (66). It may also be involved in integration of the HHV-6 genome into the host genome (67). The sequence of rat cytomegalovirus, a betaherpesvirus, also contains a homolog of the parvovirus *rep* gene (68). Similar to HHV-6, only the rat cytomegalovirus contains this gene; other cytomegaloviruses do not. Last but not least, canarypox virus was found to contain a *rep*-like gene similar to that found in circoviruses (69), but this gene is not present in other closely related poxviruses.

The presence of integrated ssDNA virus *rep* genes in poxvirus and herpesvirus genomes confounds virus metagenome analyses (see sidebar, Metagenomes and Metaviromes). Many of these analyses report the presence or absence of different virus families in environmental virus assemblies by sequence similarity of metagenomic reads to annotated database sequences (70, 71). Hits to poxvirus or herpesvirus sequences may in fact be due to sequences similar to the integrated circovirus *rep* homolog in canarypox and parvovirus *rep* homologs in herpesviruses. Thus metagenome analyses indicating the presence of poxviruses and herpesviruses in a virus assemblage need to be checked carefully to determine whether the sequences are similar to multiple viruses from each of these families or just to canarypox, HHV-6, and rat cytomegalovirus. If the latter, then the sequences in the metagenome are more likely derived from an ssDNA virus than from a dsDNA virus.

METAGENOMES AND METAVIROMES

A metagenome is nucleic acid sequence data from an environmental sample. A metavirome or viral metagenome, terms that are used interchangeably, is nucleic acid sequence data collected from a virus-enriched sample (89). Generally, viral metagenomes contain so little biomass that the nucleic acid must be amplified before sequence data can be obtained. Amplification clearly generates some biases, such as the overrepresentation of circular ssDNA virus genomes (90) when using multiple displacement amplification. Metagenomes provide a partial snapshot of a community under certain conditions. Most viral metagenomes have extremely high diversity, such that complete genomes can only rarely be assembled (48, 49).

RNA Virus Genes in ssDNA Virus Genomes

More surprising than the presence of ssDNA virus sequences in dsDNA virus genomes was the discovery of RNA virus sequences in putative ssDNA virus genomes (9). RNA virus–like sequences in DNA virus metagenomes were reported in two separate instances (72, 73), but these sequences either were not discussed (72) or were explicitly stated to be contamination (73). The lack of previous assignment of these sequences is probably due to the massive diversity present in both virus metagenomes (49), making assembly of larger sequences problematic. However, a clear instance of an ssDNA virus genome containing an RNA virus–derived gene was the result of assembly of a low-diversity metavirome from an acidic hot lake. This ssDNA virus genome appeared to have recombined with a capsid protein gene from an RNA virus (9). The authors called this genome Boiling Springs Lake RNA-DNA hybrid virus (BSL-RDHV) and noted that similar partial sequences, including sequence resembling genes for both capsid protein and ssDNA Rep, were present in three marine metagenomic data sets (9). Since then, two more complete genomes of similar viruses, both from metaviromes, have been reported (18, 19). Whether another ssDNA virus found in a dragonfly metagenome also has an RNA virus capsid protein gene homolog is controversial (20, 74, 75). Unfortunately, to date, the host or hosts for the viruses represented by these genomes have not been conclusively determined.

A bioinformatics study of approximately 100 viral metagenomes was undertaken using sequences similar to the capsid protein gene from BSL-RDHV to discover similar chimeric virus genomes (20). Thirteen new chimeric virus genomes were found in both viral metagenomes and whole-genome shotgun sequence analysis. Whereas all of the capsid protein genes were found to be monophyletic, or clearly related to one another, the associated Rep protein genes were, surprisingly, related to those from three different groups of ssDNA viruses. This result implies that the capsid protein gene was acquired once in the common ancestor of all of these viruses, whereas the Rep protein genes have undergone rampant exchange since then.

A new family of ssDNA viruses from silkworms, the bidnaviruses (76), appears on the basis of phylogenetic analysis to contain sequences derived from both dsRNA and dsDNA viruses (77). The mechanisms and reasons for this recombination are not clear but may involve recombination or insertion with the politon family of transposons. These transposons may also replicate as viruses (77, 78).

Mechanism of RNA Incorporation by ssDNA Viruses

Recently, a number of potential mechanisms were proposed whereby ssDNA viruses could capture RNA virus sequences in their genomes (75). One set of these mechanisms is based on relaxed substrate specificity of the Rep protein. Rep could ligate RNA instead of DNA at a

number of steps during rolling circle replication of the ssDNA virus genome (**Figure 3*b***). Another potential mechanism is that an mRNA from a coinfecting virus is used as a primer for either first- or second-strand replication (**Figure 3*a*,*b***). Any RNA-DNA recombination event would have to be followed by reverse transcription, presumably by cellular reverse transcriptases (38). In theory, insertion of an ssDNA virus Rep protein mRNA into an RNA virus followed by reverse transcription of the whole virus is possible (79, 80), but this seems less likely than RNA acquisition by ssDNA viruses, given the diversity of Rep protein genes and lack of diversity in capsid protein genes in these chimeric virus genomes (20). A retroelement-mediated insertion of mRNA into the dsDNA form of an ssDNA virus, similar to that proposed for RNA virus genes in host genomes (**Figure 2**), is also possible, but repeat sequences and poly(A) tails have not been detected flanking transferred genes in ssDNA viruses, nor have transposable elements been detected. DNA recombination rates in ssDNA viruses are known to be very high (81), also probably due to the activity of Rep.

Why Are RNA Virus Sequences Maintained in ssDNA Viruses?

All the mRNA capture mechanisms described above predict that any mRNA could be incorporated; thus, it is surprising that only genes resembling the BSL-RDHV capsid protein gene have been found to date in chimeric ssDNA viruses (20, 75). This suggests that selection for an RNA virus capsid protein gene that can incorporate ssDNA might be very strong; possibly, the RDHV-like capsid protein gene is one of very few that can incorporate a different kind of nucleic acid genome, DNA as well as RNA (9, 20). Moreover, a capsid protein gene would be the most likely gene to be stably maintained, because capsids are critical for the determination of host range and the avoidance of host defenses. There are also many ssDNA virus genomes that appear to contain capsid protein genes that are not similar to other known capsid protein genes; these could be derived from as-yet-uncharacterized RNA viruses (58). Environmental RNA viruses are extremely poorly studied (82), so this possibility is not unlikely. The presence of only a few RNA virus gene types is similar to the case with bornavirus and filovirus sequences in animal genomes, where the presence of these genes may also provide strong positive selection (see above). A recent study examining artificial recombination in ssDNA viruses found that very few were viable (83).

Other Evidence for RNA-DNA Recombination in Viruses

Recently, an exhaustive bioinformatics survey examined virus genomes for chimeric virus genes. Chimeric virus genes were defined as single virus genes that are derived from different sources. In more than 3,000 genomes comprising 120,000 genes, about 10% of the genes were found to be chimeras. Cluster analysis showed that 20 of these sequences were chimeric between DNA and RNA viruses (84). Generally, positive-strand ssDNA and ssRNA viruses were much more prone to having chimeric genes than were negative-strand ssRNA viruses. Again, this bioinformatics study was constrained by the limited number of known virus genome sequences.

FUTURE DIRECTIONS

Clearly chimeric viruses exist, even between RNA and ssDNA viruses. As more and more virus genomes are completed, there will undoubtedly be RNA virus sequences found in dsDNA virus genomes—maybe not Ebola virus in a poxvirus genome, as in fiction (2), but deep recombinants nonetheless. The frequency of this deep recombination is unknown and is a critical area for future research.

There is much more recombination than expected between RNA viruses, ssDNA viruses, and host genomes. This deep recombination leads not only to host genome diversity but also to an increase in the already astronomical diversity contained in virus genomes, and it offers multiple new trajectories for virus evolution (48, 75).

One way to address the frequency of occurrence of deep recombination would be to determine the mechanisms used (75). Progress has been made in understanding the mechanisms of deep recombination in host genomes, but not in viruses. Frequencies of recombination may be exceedingly small, but given the colossal number of viruses predicted on Earth, estimated at 10^{31}, and the sheer number of infections, even highly infrequent events will occur with a finite frequency and, if beneficial, are likely to be fixed by natural selection (14).

Reverse transcription of infecting RNA viruses by endogenous reverse transcriptases appears to be much more common than was previously appreciated (31–34, 47). Nonetheless, stable incorporation of viral mRNAs in host genomes seems to depend much more on selection than on opportunity. However, the discrepancy between the lack of viral pseudogenes and the presence of other pseudogenes needs to be resolved.

Another complication of deep recombination is for virus taxonomy. Does one classify a chimeric virus together with one or the other of its progenitors or create a new classification (20)? Should new classifications be made for each new chimeric virus? This subject has been discussed in depth for bacteriophages and somewhat for other viruses (14, 17, 85). Deep recombination in viruses appears to be an extreme case of the mosaic nature of viruses and their evolution (14–16, 86–88).

Finally, it is critical to find and characterize more environmental RNA viruses and ssDNA viruses, as these seem to be the epicenter of such recombination events. Unfortunately the diversity and prevalence of environmental RNA viruses are still very poorly studied (82); the study of ssDNA viruses is only slightly better in this regard (18). As all of the chimeric viruses are currently known only as genomes, probably the most critical goal is to find a host-virus pair so that they can be studied in more detail, possibly allowing elucidation of their obscure origins.

SUMMARY POINTS

1. There are considerable numbers of nonretroviral RNA virus genes in cellular genomes, probably inserted by endogenous retrotransposons.
2. RNA and ssDNA virus genes in host genomes appear to be maintained by positive selection.
3. ssDNA viruses recombine with large dsDNA viruses.
4. Some ssDNA viruses appear to be able to capture RNA virus capsid protein genes.
5. RNA viruses are often reverse transcribed by endogenous reverse transcriptases on infection.

FUTURE ISSUES

1. The frequency of deep recombination should be determined.
2. Mechanisms of recombination between RNA and DNA should be elucidated, particularly for ssDNA viruses and RNA viruses.
3. Hosts must be found for chimeric viruses found in metagenomes.

DISCLOSURE STATEMENT

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

ACKNOWLEDGMENTS

I apologize to all the researchers whose work was inadvertently insufficiently cited. Thanks to members of the Stedman laboratory for their comments on this manuscript. Thanks to the Gordon and Betty Moore Foundation, the National Science Foundation, Portland State University, and NASA for support.

LITERATURE CITED

1. Woo J. 2000. *Mission: Impossible II*. Hollywood, CA: Paramount
2. Jennings C. 2004–2008. *ReGenesis*. Toronto: Movie Cent. Netw.
3. Preston R. 1998. *The Cobra Event*. New York: Ballantine
4. Cussler C, Cussler D. 2004. *Black Wind* (*A Dirk Pitt Novel*). New York: Putnam
5. Guy B, Guirakhoo F, Barban V, Higgs S, Monath TP, Lang J. 2010. Preclinical and clinical development of YFV 17D-based chimeric vaccines against dengue, West Nile and Japanese encephalitis viruses. *Vaccine* 28:632–49
6. Lai CJ, Monath TP. 2003. Chimeric flaviviruses: novel vaccines against dengue fever, tick-borne encephalitis, and Japanese encephalitis. *Adv. Virus Res.* 61:469–509
7. Villar L, Dayan GH, Arredondo-Garcia JL, Rivera DM, Cunha R, et al. 2015. Efficacy of a tetravalent dengue vaccine in children in Latin America. *N. Engl. J. Med.* 372:113–23
8. Webster RG, Bean WJ, Gorman OT, Chambers TM, Kawaoka Y. 1992. Evolution and ecology of influenza A viruses. *Microbiol. Rev.* 56:152–79
9. **Diemer GS, Stedman KM. 2012. A novel virus genome discovered in an extreme environment suggests recombination between unrelated groups of RNA and DNA viruses. *Biol. Dir.* 7:13**
10. Baltimore D. 1971. Expression of animal virus genomes. *Bacteriol. Rev.* 35:235–41
11. King AMQ, Adams MJ, Carstens EB, Lefkowitz E, eds. 2011. *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*. Amsterdam: Elsevier
12. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860–921
13. Casjens S. 2003. Prophages and bacterial genomics: What have we learned so far? *Mol. Microbiol.* 49:277–300
14. 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
15. Pedulla ML, Ford ME, Houtz JM, Karthikeyan T, Wadsworth C, et al. 2003. Origins of highly mosaic mycobacteriophage genomes. *Cell* 113:171–82
16. Hatfull GF, Hendrix RW. 2011. Bacteriophages and their genomes. *Curr. Opin. Virol.* 1:298–303
17. Krupovic M, Prangishvili D, Hendrix RW, Bamford DH. 2011. Genomics of bacterial and archaeal viruses: dynamics within the prokaryotic virosphere. *Microbiol. Mol. Biol. Rev.* 75:610–35
18. McDaniel LD, Rosario K, Breitbart M, Paul JH. 2014. Comparative metagenomics: natural populations of induced prophages demonstrate highly unique, lower diversity viral sequences. *Environ. Microbiol.* 16:570–85
19. Hewson I, Ng G, Li W, LaBarre BA, Aguirre I, et al. 2013. Metagenomic identification, seasonal dynamics, and potential transmission mechanisms of a *Daphnia*-associated single-stranded DNA virus in two temperate lakes. *Limnol. Oceanogr.* 58:1605–20
20. **Roux S, Enault F, Bronner G, Vaulot D, Forterre P, Krupovic M. 2013. Chimeric viruses blur the borders between the major groups of eukaryotic single-stranded DNA viruses. *Nat. Commun.* 4:2700**

9. The first report of RNA virus genes in ssDNA virus genomes.

20. A comprehensive bioinformatics study of metagenomes and identification of multiple chimeric RNA-ssDNA viruses.

21. Zhdanov VM. 1975. Integration of viral genomes. *Nature* 256:471–73

22. Zhdanov VM, Bogomolova NN, Gavrilov VI, Andzhaparidze OG, Deryabin PG, Astakhov AN. 1974. Infectious DNA of tick-borne encephalitis virus. *Arch. Gesamte Virusforsch.* 45:215–24
23. Zhdanov VM, Parfanovich MI. 1974. Integration of measles virus nucleic acid into the cell genome. *Arch. Gesamte Virusforsch.* 45:225–34
24. Weiss RA, Kellam P. 1997. Virology: illicit viral DNA. *Nature* 390:235–36
25. Crochu S, Cook S, Attoui H, Charrel RN, De Chesse R, et al. 2004. Sequences of flavivirus-related RNA viruses persist in DNA form integrated in the genome of *Aedes* spp. mosquitoes. *J. Gen. Virol.* 85:1971–80

26. Klenerman P, Hengartner H, Zinkernagel RM. 1997. A non-retroviral RNA virus persists in DNA form. *Nature* 390:298–301

27. Geuking MB, Weber J, Dewannieux M, Gorelik E, Heidmann T, et al. 2009. Recombination of retrotransposon and exogenous RNA virus results in nonretroviral cDNA integration. *Science* 323:393–96
28. Taylor DJ, Bruenn J. 2009. The evolution of novel fungal genes from non-retroviral RNA viruses. *BMC Biol.* 7:88
29. Chiba S, Kondo H, Tani A, Saisho D, Sakamoto W, et al. 2011. Widespread endogenization of genome sequences of non-retroviral RNA viruses into plant genomes. *PLOS Pathog.* 7:e1002146

30. Horie M, Honda T, Suzuki Y, Kobayashi Y, Daito T, et al. 2010. Endogenous non-retroviral RNA virus elements in mammalian genomes. *Nature* 463:84–87

31. Belyi VA, Levine AJ, Skalka AM. 2010. Unexpected inheritance: multiple integrations of ancient Bornavirus and Ebolavirus/Marburgvirus sequences in vertebrate genomes. *PLOS Pathog.* 6:e1001030

32. Katzourakis A, Gifford RJ. 2010. Endogenous viral elements in animal genomes. *PLOS Genet.* 6:e1001191

33. Horie M, Kobayashi Y, Suzuki Y, Tomonaga K. 2013. Comprehensive analysis of endogenous bornavirus-like elements in eukaryote genomes. *Philos. Trans. R. Soc. B.* 368:1–11
34. Taylor DJ, Leach RW, Bruenn J. 2010. Filoviruses are ancient and integrated into mammalian genomes. *BMC Evol. Biol.* 10:193
35. Kazazian HH Jr. 2014. Processed pseudogene insertions in somatic cells. *Mob. DNA* 5:20
36. Esnault C, Maestre J, Heidmann T. 2000. Human LINE retrotransposons generate processed pseudogenes. *Nat. Genet.* 24:363–67
37. Han JS. 2010. Non-long terminal repeat (non-LTR) retrotransposons: mechanisms, recent developments, and unanswered questions. *Mob. DNA* 1:15
38. Prak ETL, Kazazian HH. 2000. Mobile elements and the human genome. *Nat. Rev. Genet.* 1:134–44
39. Cost GJ, Feng Q, Jacquier A, Boeke JD. 2002. Human L1 element target-primed reverse transcription in vitro. *EMBO J.* 21:5899–910
40. Cost GJ, Boeke JD. 1998. Targeting of human retrotransposon integration is directed by the specificity of the L1 endonuclease for regions of unusual DNA structure. *Biochemistry* 37:18081–93
41. Mandal PK, Ewing AD, Hancks DC, Kazazian HH Jr. 2013. Enrichment of processed pseudogene transcripts in L1-ribonucleoprotein particles. *Hum. Mol. Genet.* 22:3730–48
42. Tsai WC, Lloyd RE. 2014. Cytoplasmic RNA granules and viral infection. *Annu. Rev. Virol.* 1:147–70
43. Gorbunova V, Boeke JD, Helfand SL, Sedivy JM. 2014. Sleeping dogs of the genome. *Science* 346:1187–88
44. Tanne E, Sela I. 2005. Occurrence of a DNA sequence of a non-retro RNA virus in a host plant genome and its expression: evidence for recombination between viral and host RNAs. *Virology* 332:614–22
45. Maori E, Tanne E, Sela I. 2007. Reciprocal sequence exchange between non-retro viruses and hosts leading to the appearance of new host phenotypes. *Virology* 362:342–49
46. Fujino K, Horie M, Honda T, Merriman DK, Tomonaga K. 2014. Inhibition of Borna disease virus replication by an endogenous bornavirus-like element in the ground squirrel genome. *PNAS* 111:13175–80

47. Goic B, Vodovar N, Mondotte JA, Monot C, Frangeul L, et al. 2013. RNA-mediated interference and reverse transcription control the persistence of RNA viruses in the insect model *Drosophila*. *Nat. Immunol.* 14:396–403

21. First report of RNA virus genes in cellular genomes; widely ignored.

26. First widely accepted report of RNA virus genes in cellular genomes.

30–32. Concurrent bioinformatics studies that showed multiple RNA virus genes in animal genomes.

47. Indication that reverse transcription of infecting RNA viruses is widespread.

48. Mokili JL, Rohwer F, Dutilh BE. 2012. Metagenomics and future perspectives in virus discovery. *Curr. Opin. Virol.* 2:63–77
49. Angly FE, Felts B, Breitbart M, Salamon P, Edwards RA, et al. 2006. The marine viromes of four oceanic regions. *PLOS Biol.* 4:e368
50. Anthony SJ, Epstein JH, Murray KA, Navarrete-Macias I, Zambrana-Torrelio CM, et al. 2013. A strategy to estimate unknown viral diversity in mammals. *mBio* 4:e00598–13
51. Cotmore SF, Tattersall P. 2013. Parvovirus diversity and DNA damage responses. *Cold Spring Harb. Perspect. Biol.* 5:a012989
52. Asokan A, Schaffer DV, Samulski RJ. 2012. The AAV vector toolkit: poised at the clinical crossroads. *Mol. Ther.* 20:699–708
53. Ward P, Walsh CE. 2012. Targeted integration of a rAAV vector into the AAVS1 region. *Virology* 433:356–66
54. Faruque SM, Mekalanos JJ. 2012. Phage-bacterial interactions in the evolution of toxigenic *Vibrio cholerae*. *Virulence* 3:556–65
55. Das B. 2014. Mechanistic insights into filamentous phage integration in *Vibrio cholerae*. *Front. Microbiol.* 5:650
56. Kapoor A, Simmonds P, Lipkin WI. 2010. Discovery and characterization of mammalian endogenous parvoviruses. *J. Virol.* 84:12628–35
57. Liu H, Fu Y, Xie J, Cheng J, Ghabrial SA, et al. 2011. Widespread endogenization of densoviruses and parvoviruses in animal and human genomes. *J. Virol.* 85:9863–76

58. Excellent review article on environmental ssDNA viruses.

58. Rosario K, Duffy S, Breitbart M. 2012. A field guide to eukaryotic circular single-stranded DNA viruses: insights gained from metagenomics. *Arch. Virol.* 157:1851–71

59. Li L, Kapoor A, Slikas B, Bamidele OS, Wang C, et al. 2010. Multiple diverse circoviruses infect farm animals and are commonly found in human and chimpanzee feces. *J. Virol.* 84:1674–82
60. Liu H, Fu Y, Li B, Yu X, Xie J, et al. 2011. Widespread horizontal gene transfer from circular single-stranded DNA viruses to eukaryotic genomes. *BMC Evol. Biol.* 11:276
61. Meng G, Zhang X, Plevka P, Yu Q, Tijssen P, Rossmann MG. 2013. The structure and host entry of an invertebrate parvovirus. *J. Virol.* 87:12523–30
62. Belyi VA, Levine AJ, Skalka AM. 2010. Sequences from ancestral single-stranded DNA viruses in vertebrate genomes: the Parvoviridae and Circoviridae are more than 40 to 50 million years old. *J. Virol.* 84:12458–62
63. Krupovic M, Forterre P. 2011. Microviridae goes temperate: Microvirus-related proviruses reside in the genomes of Bacteroidetes. *PLOS ONE* 6:e19883

64. First report of ssDNA virus genes in host genomes.

64. Thomson BJ, Efstathiou S, Honess RW. 1991. Acquisition of the human adeno-associated virus type-2 *rep* gene by human herpesvirus type-6. *Nature* 351:78–80

65. Thomson BJ, Weindler FW, Gray D, Schwaab V, Heilbronn R. 1994. Human herpesvirus-6 (HHV-6) is a helper virus for adeno-associated virus type-2 (AAV-2) and the AAV-2 *rep* gene homolog in HHV-6 can mediate AAV-2 DNA-replication and regulate gene-expression. *Virology* 204:304–11
66. Caselli E, Bracci A, Galvan M, Boni M, Rotola A, et al. 2006. Human herpesvirus 6 (HHV-6) U94/REP protein inhibits betaherpesvirus replication. *Virology* 346:402–14
67. Kaufer BB, Flamand L. 2014. Chromosomally integrated HHV-6: impact on virus, cell and organismal biology. *Curr. Opin. Virol.* 9:111–18
68. Vink C, Beuken E, Bruggeman CA. 2000. Complete DNA sequence of the rat cytomegalovirus genome. *J. Virol.* 74:7656–65
69. Tulman ER, Afonso CL, Lu Z, Zsak L, Kutish GF, Rock DL. 2004. The genome of canarypox virus. *J. Virol.* 78:353–66
70. Roux S, Tournayre J, Mahul A, Debroas D, Enault F. 2014. Metavir 2: new tools for viral metagenome comparison and assembled virome analysis. *BMC Bioinform.* 15:76
71. Roux S, Faubladier M, Mahul A, Paulhe N, Bernard A, et al. 2011. Metavir: a web server dedicated to virome analysis. *Bioinformatics* 27:3074–75
72. López-Bueno A, Tamames J, Velázquez D, Moya A, Quesada A, Alcami A. 2009. High diversity of the viral community from an Antarctic lake. *Science* 326:858–61

73. Whon TW, Kim MS, Roh SW, Shin NR, Lee HW, Bae JW. 2012. Metagenomic characterization of airborne viral DNA diversity in the near-surface atmosphere. *J. Virol.* 86:8221–31
74. Rosario K, Dayaram A, Marinov M, Ware J, Kraberger S, et al. 2012. Diverse circular ssDNA viruses discovered in dragonflies (Odonata: Epiprocta). *J. Gen. Virol.* 93:2668–81
75. Stedman K. 2013. Mechanisms for RNA capture by ssDNA viruses: grand theft RNA. *J. Mol. Evol.* 76:359–64
76. Hu Z, Li G, Li G, Yao Q, Chen K. 2013. *Bombyx mori bidensovirus*: the type species of the new genus *Bidensovirus* in the new family *Bidnaviridae*. *Chin. Sci. Bull.* 58:4528–32
77. Krupovic M, Koonin EV. 2014. Evolution of eukaryotic single-stranded DNA viruses of the *Bidnaviridae* family from genes of four other groups of widely different viruses. *Sci. Rep.* 4:5347
78. Krupovic M, Bamford DH, Koonin EV. 2014. Conservation of major and minor jelly-roll capsid proteins in Polinton (Maverick) transposons suggests that they are bona fide viruses. *Biol. Dir.* 9:6
79. Krupovic M. 2012. Recombination between RNA viruses and plasmids might have played a central role in the origin and evolution of small DNA viruses. *BioEssays* 34:867–70
80. Krupovic M, Ravantti JJ, Bamford DH. 2009. Geminiviruses: a tale of a plasmid becoming a virus. *BMC Evol. Biol.* 9:112
81. Martin DP, Biagini P, Lefeuvre P, Golden M, Roumagnac P, Varsani A. 2011. Recombination in eukaryotic single stranded DNA viruses. *Viruses* 3:1699–738
82. Steward GF, Culley AI, Mueller JA, Wood-Charlson EM, Belcaid M, Poisson G. 2013. Are we missing half of the viruses in the ocean? *ISME J.* 7:672–79
83. Monjane AL, Martin DP, Lakay F, Muhire BM, Pande D, et al. 2014. Extensive recombination-induced disruption of genetic interactions is highly deleterious but can be partially reversed by small numbers of secondary recombination events. *J. Virol.* 88:7843–51
84. Jachiet PA, Colson P, Lopez P, Bapteste E. 2014. Extensive gene remodeling in the viral world: new evidence for nongradual evolution in the mobilome network. *Genome Biol. Evol.* 6:2195–205
85. Abrescia NGA, Bamford DH, Grimes JM, Stuart DI. 2012. Structure unifies the viral universe. *Annu. Rev. Biochem.* 81:795–822
86. Hendrix RW. 2009. Jumbo bacteriophages. *Curr. Top. Microbiol. Immunol.* 328:229–240
87. Botstein D. 1980. A theory of modular evolution for bacteriophages. *Ann. N.Y. Acad. Sci.* 354:484–90
88. Duffy S, Shackelton LA, Holmes EC. 2008. Rates of evolutionary change in viruses: patterns and determinants. *Nat. Rev. Genet.* 9:267–76
89. Thurber RV, Haynes M, Breitbart M, Wegley L, Rohwer F. 2009. Laboratory procedures to generate viral metagenomes. *Nat. Protoc.* 4:470–83
90. Kim KH, Bae JW. 2011. Amplification methods bias metagenomic libraries of uncultured single-stranded and double-stranded DNA viruses. *Appl. Env. Microbiol.* 77:7663–68

75. Hypothetical mechanisms for RNA virus gene acquisition by ssDNA viruses.

Annual Review of Virology

Volume 2, 2015

Contents