JVI Accepts, published online ahead of print on 30 March 2011
J. Virol. doi:10.1128/JVI.02203-10

Minireview

What does virus evolution tell us about virus origins?

Edward C. Holmes[1,2]

[1]*Center for Infectious Disease Dynamics, Department of Biology, The Pennsylvania State University, Mueller Laboratory, University Park, PA 16802, USA.*

[2]*Fogarty International Center, National Institutes of Health, Bethesda, MD 20892. USA.*

Correspondence to:

Dr. Edward C. Holmes

Department of Biology,

The Pennsylvania State University,

University Park,

PA 16802, USA.

Tel: 1 814 863 4689; Fax: 1 814 865 9131; E-mail: ech15@psu.edu

Main text = 3582 words, Abstract = 169 words, Figures = 1

Despite recent advances in our understanding of diverse aspects of virus evolution, particularly at the epidemiological scale, revealing the ultimate origins of viruses has proven to be a more intractable problem. Herein I review some current ideas on the evolutionary origins of viruses, and assess how well these theories accord with what we know about the evolution of contemporary viruses. I note the growing evidence for the theory that viruses arose before the Last Universal Cellular Ancestor (LUCA). This ancient origin theory is supported by the presence of capsid architectures that are conserved among diverse viral taxa, including among RNA and DNA viruses, and the strongly inverse relationship between genome size and mutation rate across all replication systems, such that pre-LUCA genomes were probably both small and highly error prone and hence RNA virus-like. I also highlight the advances that are needed to come to a better understanding of virus origins, most notably the ability to accurately infer deep evolutionary history from the phylogenetic analysis of conserved protein structures.

As has been true for many years, the central debating point in discussions of the origin of viruses is whether they are ancient, first appearing before the Last Universal Cellular Ancestor (LUCA), or that they evolved more recently, such that their ancestry lies with genes that 'escaped' from the genomes of their cellular host organisms and subsequently evolved independent replication. Although the escaped gene theory has traditionally dominated thinking on viral origins (reviewed in ref. 37), in large part because as viruses are parasitic on cells now it has been argued that this must have always have been the case, and because there is no gene shared by all viruses, recent data is providing increasingly strong support for a far more ancient origin. Herein I briefly review some contemporary ideas on the origins of viruses and assess how well they accord with available data. Although there have been a number of important reviews of virus origins published in recent years (14, 15, 24, 26), which interested readers should consult for a more detailed discussion of individual theories, I will take a rather different perspective. First, while most research on viral origins has focused on DNA viruses, in which the phylogenetic links between viral and cellular genes are rather easier to discern, I will direct most of my attention toward RNA viruses. Second, although a frequent theme in discussions of viral origins has been to list the phenotypic similarities, and presumably homologies, between diverse types of virus, it is my strong contention that an understanding of the fundamental mechanisms of viral evolution, particularly the error-prone nature of RNA-based replication and what this means for the evolution of genome size and complexity, is also able to shed light on the ancestry of viruses. Indeed, most studies of viral origins have deemphasized the processes that govern the evolution of contemporary viruses. Finally, I will outline a number of the research themes that might reasonably provide important new data on the complex issue of virus origins.

RECENT DATA ON VIRAL ORIGINS

Studies of viral origins have been re-energized by two remarkable observations made in the last dozen years: the discovery and genome sequencing of the giant amoebal mimivirus (32, 42), and the growing number of reports of apparent homology between the capsid architectures of viruses that possess no primary sequence similarity (2, 4, 29).

The discovery of mimivirus has undoubtedly had a major impact on theories of viral origins, including our notion of how a virus might be defined (7). While phylogenetic analysis indicates that a small proportion (<1%) of the gene content of mimivirus is of host origin, and which has been used to bolster theories that viruses primarily exist as 'gene robbers' that evolved after cellular species (35, 36), many more genes (at least 25%) clearly link mimivirus to other large dsDNA viruses (22, 23), and particularly those of the Nucleo-Cytoplasmic Large DNA Virus (NCLDV) lineage that comprises asfarviruses, ascoviruses, iridoviruses, phycodnaviruses, poxviruses as well as the recently discovered Marseillevirus that infects the same amoebal host as mimivirus (22, 51). More striking is that most (~70% at the time of writing) mimivirus genes have no known homologs, in either virus or cellular genomes, so that their origins are unknown (12), although the data currently available suggests that they are unlikely to come from the amoebal host genome (42). More importantly, the discovery of mimivirus highlights our profound ignorance of the virosphere. It is therefore a truism that a wider sampling of viruses in nature is likely to tell us a great deal more about viral origins.

Although perhaps less lauded, the discovery of conserved protein structures among diverse viruses with little if any primary sequence similarity has even grander implications for our understanding of viral origins. This deep structural similarity is beautifully illustrated by the jelly-roll capsid, a tightly structured protein barrel that represents the major capsid subunit of virions with an icosahedral structure (8, 43). Not only is the jelly-roll capsid highly conserved, but this conservation extends to both RNA and DNA viruses, including such viruses as picornaviruses (ssRNA+), birnaviruses (dsRNA), herpesviruses (dsDNA), and some DNA phages, and hence strongly arguing for their ancient common ancestry. Other highly conserved

capsid architectures include the 'PRD1-adenovirus lineage', characterized by a double β-barrel fold which is found in dsDNA viruses as diverse as phage PRD1, human adenovirus, mimivirus, as well as a variety of archaean viruses (3, 4, 29), the HK97-like lineage, which encompasses tailed dsDNA viruses that infect bacteria, archaea and eukaryotes, and the BTV-like lineage which is found in a number of dsRNA including members of the *Reoviridae* and *Totiviridae* (2). More recently, a common virion architecture has been proposed for some viruses that do not possess an iscosahedral capsid, including the archaean virus *Halorubrum* pleomorphic virus 1 (HRPV-1) (38).

Because of their remarkable conservation, it has been claimed that these conserved structures signify the existence of distinct 'lineages' of virion architectures with ancestries dating back to a pre-cellular world (1, 30), although the evolutionary relationships between these lineages is far less clear. While the deep common ancestry of viruses infecting hosts from the different domains of life is not in itself conclusive proof of a pre-LUCA origin, particularly as cross-species transmission is a very common mode of virus evolution, it at least greatly reduces the number of possible gene escape events required to explain the diversity of extant viruses and pushes any such escape events far back into evolutionary time. This uncertainty notwithstanding, it is clear that analyses of similarities in virion structure should be extended to as many different types of virus as possible. Outside of the virion, it is notable that a palm subdomain protein structure, which is comprised of a four-stranded antiparallel β-sheet and two α-helices, is conserved among some RNA-dependent and DNA-dependent polymerases, again suggesting that it is of ancient origin (17), while the presence of a superfamily 3 helicase also links diverse RNA and DNA viruses (26).

Despite the growing evidence for highly conserved protein structures and its indications of ancient common ancestry, proponents of the escaped gene theory counter that these similarities could have arisen more recently due to either strong convergent evolution and/or

lateral gene transfer (LGT) (36). It is right to think that convergent evolution may be commonplace in viral capsids that are likely subject to strong selection pressure to be small. Indeed, convergent evolution between divergent protein structures has previously been noted in viruses (19), and convergence is rampant in some other systems, with C_4 photosynthesis a notable case in point (44). Although the lack of a definitive phylogenetic tree of all viruses makes it impossible to conclusively rule out convergent evolution as an explanation for the similarity between the capsid structures of highly divergent viruses, two further observations strongly argue against this process; first, that these structures occur across such a very range of viral taxa, thereby necessitating multiple convergent events, and more convergence needs to be invoked, the less likely it becomes, and second that virion architectures form a variety of different structures (the 'lineages' noted above), whereas selectively driven convergence might be expected to result in a single favorable capsid structure.

I believe that frequent LGT is similarly unlikely. In particular, LGT appears to be rare among RNA viruses, with only a examples documented to date (21). This is to be expected given the major selective constraints against large genome size in these organisms; increasing genome size through LGT would in turn result in an elevated number of deleterious mutations per replication and hence major fitness losses. Indeed, while large dsDNA organisms utilize both gene duplication (common in eukaryotes) and/or LGT (common in bacteria) to create evolutionary novelty (46), both seem to occur only sporadically in RNA viruses (21). Although LGT would not result in an increased genome size if there was a direct gene replacement, any such replacement event would have to occur precisely at a gene boundary otherwise it would likely result in a deleterious genotype. Given that the earliest replicating RNA molecules almost certainly possessed higher error rates than those of contemporary RNA viruses, and which would have imposed major constraints on their genome size (see below), it seems unlikely that LGT was so widespread as to disperse common protein structures among RNA viruses, or between RNA and DNA viruses. As such, the most plausible scenario from the available data is

that the deep similarities in capsid structure among viruses are indeed indicative of an ancient common ancestry.

Quite what the world where these ancient virus-like replicators resided looked like is open to debate, and there are a number of rather different versions of the pre-LUCA theory. One important idea is that there was an 'ancient virus world' of primordial replicators that existed before any cellular organisms, and that both RNA (first) and DNA (later) viruses originated at this time, donating some features to the first cellular organisms (24, 26). The obligatory parasitic behavior we see in contemporary viruses therefore represents a more recent adaptation. A competing theory is that RNA cells existed before the LUCA and that RNA viruses were parasites on these RNA cells that later evolved DNA has a way of escaping host cell responses (13, 14). As such, viruses were responsible for one of the major innovations in evolutionary history. Given that we are attempting to reconstruct events that happened billions of years ago, such that the trace of common ancestry has all but disappeared, it is always going to be extremely challenging to choose between theories of pre-LUCA life. Indeed, it is patently easier to create theories for viral origins than to test them. These fundamental limitations notwithstanding, I believe Koonin's argument that a 'pre-cellular stage of evolution must have involved genetic elements of virus-like size and complexity' is a compelling one (27). Indeed, as I will argue below, a consideration of how RNA viruses evolve today strongly suggests that the earliest replicating molecules shared some clear similarities with viruses.

Despite the mounting evidence for an ancestry of viruses that predates the LUCA, it is important to keep in mind that this does mean that, on occasion, new viruses can be created through gene escape events that must have happened far more recently. This point is dramatically illustrated by human hepatitis delta virus (HDV) which has been shown to contain a ribozyme sequence that is closely related to the CPEB3 ribozyme present in a human intron (45). As HDV is only found in humans and requires human hepatitis B virus to replicate, this discovery represents powerful evidence that origin of HDV lies with the human transcriptome

rather than with a pre-LUCA world. I doubt that this will be the last documentation of viral origin through host gene escape.

ERROR RATES AND VIRAL ORIGINS

One of the most profound observations made in evolutionary genetics in recent years is that there is a strongly inverse relationship between mutation rate per genome replication and genome size (16; Fig. 1). Hence, the highest error rates per nucleotide of any system are reported in the tiny viroids (< 400 nt in length) that possess hammerhead ribozymes (16), while mutation rates that are orders of magnitude lower are observed in bacteria and eukaryotes (10, 16). This association between error rate and genome size is remarkable for two reasons. First, it covers mutation rates and genome sizes that vary over some eight orders of magnitude. Aside from the allometric relationship between body size and metabolic rate (20), associations of this scale are few and far between in nature. Second, there is a marked absence of data points in which mutation rates are overly high or abnormally low for a specific genome size, strongly suggesting that mutation rate is a trait optimized by opposing selection pressures (Fig. 1). Mutation rates that are too high are likely to be selected against because they produce an excessively high number of deleterious mutations per replication and therefore result in fitness losses, while mutation rates that are too low either reduce the rate of adaptive evolution (5), or are subject to a physiological cost on increased fidelity that prevents the evolution of a zero mutation rate (47).

Because the first replicating systems were likely composed of RNA – an hypothesis greatly strengthened by the recent demonstration of how RNA might be effectively synthesized in a pre-biotic atmosphere (40) – they would have been both very small and highly error-prone. Therefore, any increase in genome size and complexity must have required either a reduction in error rate, or a buffering against the effect of deleterious mutations (i.e. mutational robustness),

perhaps in the form of complex secondary structures that increase neutral space (31). Crucially, that RNA viruses are still very much at the mercy of their mutation rates, because artificially increasing error rates through the application of chemical mutagens frequently induces fitness losses (9), also suggests that they evolved from primitive RNA replicators that never possessed error-correction, rather than from higher fidelity cellular polymerases that then evolved to become more error-prone. To put it another way, because of the huge fitness costs that are associated with producing genomes that are overly long (i.e. an increased mutational load), it seems untenable that a high-fidelity DNA replication system in which a wide array of genome sizes are permitted could give rise to an RNA-replicating organism that is strongly genome size limited and so susceptible to major fitness losses. Indeed, the trend depicted in Fig. 1 suggests that error rates have been progressive reduced over evolutionary time. In this case, simplicity really does seem to imply antiquity.

That DNA genomes are usually far larger than those of RNA viruses is also commonly cited as the reason underlying the evolution of DNA from RNA; DNA has an intrinsically higher replication fidelity, which in turn allows genomes to increase in size and hence complexity (33). However, as Forterre has pointed out, an increase in complexity/stability is unlikely to result in a sufficiently large individual fitness benefit to favor the evolution of DNA over RNA (13). In addition, analysis of the relationship between error rate and genome size also reveals that it is only double-strand (ds) DNA organisms that have markedly reduced error rates (and larger genomes) compared to RNA-based organisms (Fig. 1). Indeed, one of the most important conclusions arising from studies of viral evolution in recent years is that many single-strand (ss) DNA viruses evolve at broadly similar rates to RNA viruses, and similarly possess very small genomes (11). Hence, it was not simply the invention of DNA that facilitated the evolution of complexity, but the invention of dsDNA. Here, again, mimivirus may be of great importance. Because mimivirus possesses a genome that is far larger than those of other dsDNA viruses

(and similar to those of some bacterial species), it is also predicted to have a lowest mutation rate yet recorded for a virus.

HOW TO IMPROVE OUR UNDERSTANDING OF VIRAL ORIGINS?

Despite the sea-change in our views of viral origins, with a pre-LUCA ancestry looking increasing likely, it is clear that we are still a long way from understanding this critical moment in the history of life on earth. I believe that two major research themes will have a major effect on studies of virus origins. First, and most obviously, it is clear that we need far more studies of viral biodiversity, with a particular focus on environments and potential hosts that have been only poorly sampled to date. As viruses are the most abundant source of nucleic acid on earth, with every cellular organism likely to be infected by multiple viruses, our sample of current viral biodiversity is by definition miniscule. Despite the remarkable advances in metagenomic surveys of viral biodiversity (48) and what this might mean for viral origins (28), a more detailed exploration of the virosphere should undoubtedly be a research priority. As the discovery of mimivirus fundamentally changed our understanding of virus definitions and origins, so it is the case that the discovery of new viruses will continue to do much the same in future. As a specific case in point, despite the growing catalog of DNA viruses from Archaea (41), including those with ssDNA genomes (38), to date no RNA viruses has been described from this major domain of life. Determining whether the current absence of RNA viruses from the archaea is due to (i) insufficiently intensive sampling, (ii) that RNA viruses have never existed in these organisms, or (iii) that the Archaea have evolved mechanisms that are strongly efficient at eliminating RNA viruses, is therefore central to studies of viral origins. Only a massively increased sampling will tell.

The second major advance needed is in the area of phylogenetics, particularly with respect to RNA viruses in which evolutionary history has been especially difficult to resolve. For

a while, the phylogenetic analysis of specific virus proteins reasonably appeared to hold the key to revealing the deep evolutionary relationships of RNA viruses (25, 39). Indeed, it might seem a relatively straightforward task to take a set of sequences from a gene of known homology, such as the RNA-dependent RNA polymerase that characterizes all RNA viruses, align them and then infer an evolutionary tree, or even a more complex network-like structure, using the suite of phylogenetic methods now available. However, the reality of the matter is that the amino acid sequences of RNA viruses assigned to different families are often so divergent that the standard methods of multiple sequence alignment followed by phylogenetic inference are unable to recover a reliable panoramic phylogeny encompassing all RNA viruses. More starkly, viruses assigned to different families of RNA viruses often possess no more sequence similarity than expected by chance alone (52). Inferring robust phylogenetic trees on these sequence data alone is evidently a fruitless exercise. A lack of sequence similarity at the inter-family level will also make it difficult to distinguish a specific mode of evolutionary change, such as the explosive radiation of lineages leading to different viral families, from a lack of phylogenetic resolution at the root of a viral tree that is an inevitable outcome of extreme levels of sequence divergence (28).

Although it likely that all studies of deep virus phylogeny are likely to be highly challenging at best, a number of specific improvements are possible. One idea is to use aspects of genome organization, such as gene content and/or gene order, as a phylogenetic trait. However, while these traits may be useful in identifying clusters of related RNA viruses such as the picorna-like viruses (28), or provide insights into the evolution of some groups of large dsDNA viruses where there are a sufficient number of changes to undertake a meaningful phylogenetic analysis (34), the diverse array of genome organizations used by viruses make it untenable on a large scale. A more practical approach may therefore be to undertake 'alignment free' analyses of evolutionary history. A variety of methods have been developed in this area (6, 50), often making use of phylogenetic profiles, in which each entry in a vector

quantifies the alignment between a specific target sequence and a knowledge-base Position Specific Scoring Matrix (PSSM) (18). To date, the results of analyses using these methods have been encouraging, and do at least as good a job as standard phylogenetic methods based on multiple sequence alignment in revealing key aspects of evolutionary history (6). However, whether they can provide new insights into systems as diverse as different families of RNA viruses, where multiple sequence alignments fail completely, is another question entirely. Indeed, it is notable that all alignment-free methods currently deal with data sets where multiple sequence alignment is still viable to some extent.

An additional, and potentially even more powerful approach to reconstructing deep evolutionary history is to use features of protein structure, particularly in cases where primary sequence similarity is absent altogether. Indeed, this may be the only practical way to glean new information on the origins of viruses in the face of extreme diversity in primary sequence data and genome organization. In its simplest guise, this can simply mean using protein structures as a guide for amino acid sequence alignment, as has been attempted for some analyses of diverse RNA viruses (49). However, although useful, this approach will clearly be unable to remove all the phylogenetic noise caused by multiple substitutions at single amino acid sites that plague comparisons between very highly divergent sequences.

A more profitable approach would therefore be to code aspects of protein structure as phylogenetic characters. Although there has been some attempt to infer phylogenies using elements of protein structure (2), these methods are still in their infancy and hence provide little phylogenetic precision at present. Simple methods could be based on clustering metrics employing some measure of structural distance or scoring binary differences between structures and then inferring their relationships using a parsimony procedure. However, to make more robust insights it is clear that we will ultimately require far more advanced approaches, ideally incorporating a fully probabilistic model of protein structure evolution, although this represents a major technical challenge and may first require the ability to accurately infer protein structure

271 from primary sequence. Despite the scale of this problem I believe that the time to invest in this
272 project is now. Not only will the development of phylogenetic methods of this kind greatly assist
273 in studies of viral origins, but it will directly benefit any research program that is based on
274 characterizing the deep relationships among organisms or proteins, and where primary
275 sequence similarity has been lost in evolutionary time.

REFERENCES

1. **Bamford, D.H.** 2003. Do viruses form lineages across different domains of life? Res. Microbiol. **154:**231-236.
2. **Bamford, D.H., J.M. Grimes, and D.I. Stuart.** 2005. What does structure tell us about viral evolution? Curr. Op. Struct. Biol. **15:**1-9.
3. **Benson, S.D., J.K.H. Bamford, D.H. Bamford, and R.M. Burnett.** 1999. Viral evolution revealed by bacteriophage PRD1 and human adenovirus coat protein structures. Cell **98:**825-833.
4. **Benson, S.D., J.K.H. Bamford, D.H. Bamford, and R.M. Burnett.** 2004. Does common architecture reveal a viral lineage spanning all three domains of life? Mol. Cell **16:**673-685.
5. **Bonhoeffer, S. and P. Sniegowski.** 2002. Virus evolution: the importance of being erroneous. Nature **420:**367.
6. **Chang, G.S., Y. Hoon, K. D. Ko, G. Bhardwaj, E.C. Holmes, R.L. Patterson, and van D.B. Rossum,** 2008. Phylogenetic profiles reveal evolutionary relationships within the 'twilight zone' of sequence similarity. Proc. Natl. Acad. Sci. USA **105:**13474-13479.
7. **Claverie, J.-M.** 2006. Viruses take center stage in cellular evolution. Genome Biol. **7**:110.
8. **Coulibaly, F., C. Chevalier, I. Gutsche, J. Pous, J. Navaza, S. Bressanelli, B. Delmas, and F.A. Rey,** 2005. The birnavirus crystal structure reveals structural relationships among icosahedral viruses. Cell **120:**761-772.
9. **Crotty, S., C.E. Cameron, and R. Andino.** 2001. RNA virus error catastrophe: direct test by using ribavirin. Proc. Natl. Acad. Sci. USA **98:**6895-6900.
10. **Drake, J.W., B. Charlesworth, D. Charlesworth, and J.F. Crow.** 1998. Rates of spontaneous mutation. Genetics **148:**1667-1686.
11. **Duffy, S., L.A. Shackelton, and E.C. Holmes.** 2008. Rates of evolutionary change in viruses: Patterns and determinants. Nat. Rev. Genet. **9:**267-276.

12. **Filée, J and M. Chandler.** 2010. Gene exchange and the origin of giant viruses. Intervirology **53:**354-361.

13. **Forterre, P.** 2005. The two ages of the RNA world, and the transition to the DNA world: A story of viruses and cells. Biochimie **87:**793-803.

14. **Forterre, P**. 2006. The origin of viruses and their possible roles in major evolutionary transitions. Virus Res. **117:**5-16.

15. **Forterre, P. and D. Prangishvili.** 2009. The origin of viruses. Res. Microbiol. **160:**466-472.

16. **Gago, S., S.F. Elena, R. Flores, and R. Sanjuán**. 2009. Extremely high mutation rate of a hammerhead viroid. Science **323:**1308.

17. **Gorbalenya, A.E., F.M. Pringle, J.L. Zeddam, B.T. Luke, C.E. Cameron, J. Kalmakoff, T.N. Hanzlik, K.H Gordon, and V.K. Ward.** 2002. The palm subdomain-based active site is internally permuted in viral RNA-dependent RNA polymerases of an ancient lineage. J. Mol. Biol. **324:**47-62.

18. **Gribskov, M., A.D. McLachlan, and D. Eisenberg.** 1987. Profile analysis: detection of distantly related proteins. Proc. Natl. Acad. Sci. USA **84:**4355-4358.

19. **Heldwein, E.E., H. Lou, F.C. Bender, G.H. Cohen, R.J. Eisenberg, and S.C. Harrison,** 2006. Crystal structure of glycoprotein B from herpes simplex virus 1. Science **313:**217-220.

20. **Hemmingsen, A. M.** 1960. Energy metabolism as related to body size and respiratory surfaces, and its evolution. Reports of the Steno Memorial Hospital and Nordisk Insulin Laboratorium **9:**6-110.

21. **Holmes, E.C**. 2009. The evolution and emergence of RNA viruses. Oxford Series in Ecology and Evolution (OSEE). Oxford University Press, Oxford, UK.

22. **Iyer, L.M., S. Balaji, E.V. Koonin, and L. Aravind.** 2006. Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. Virus Res. **117:**156-184.

23. **Koonin, E.V.** 2005. Virology: Gulliver among the Lilliputians. Curr. Biol. **15:**R167-169.

24. **Koonin, E.V.** 2009. On the origin of cells and viruses: primordial virus world scenario. Ann. N. Y. Acad. Sci. **1178:**47-64.

25. **Koonin, E.V. and V.V. Dolja.** 1993. Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences. Crit. Rev. Biochem. Mol. Biol. **28:**375-430.

26. **Koonin, E.V., T.G. Senkevich, and V.V. Dolja.** 2006. The ancient virus world and evolution of cells. Biol. Direct **1:**29.

27. **Koonin, E.V., T.G. Senkevich, and V.V. Dolja.** 2009. Compelling reasons why viruses are relevant for the origin of cells. Nat. Rev. Micro. **7**:615.

28. **Koonin, E.V., Y.I. Wolf, K. Nagasaki, and V.V. Dolja.** 2008. The Big Bang of picorna-like virus evolution antedates the radiation of eukaryotic supergroups. Nat. Rev. Microbiol. **6:**925-939.

29. **Krupovic, M. and D. H. Bamford.** 2008. Virus evolution: how far does the double beta-barrel viral lineage extend? Nat. Rev. Microbiol. **6:**941-948.

30. **Krupovic, M. and D. H. Bamford.** 2010. Order to the viral universe. J. Virol. **84:**12476-12479.

31. **Kun, A., Santos, M. and Szathmáry, E.** 2005. Real ribozymes suggest a relaxed error threshold. Nat. Genet. **37:**1008-1011.

32. **La Scola, B., S. Audic, C. Robert, L. Jungang, X. de Lamballerie, M. Drancourt, R. Birtles, J.-M. Claverie, and D. Raoult.** 2003. A giant virus in amoebae. Science **299:**2033.

33. **Lazcano, A., R. Guerrero, L. Margulis, and J. Oro.** 1988. The evolutionary transition from RNA to DNA in early cells. J. Mol. Evol. **27:**283-290.

34. **McLysaght, A., P.F. Baldi, and B.S. Gaut.** 2003. Extensive gene gain associated with adaptive evolution of poxviruses. Proc. Natl. Acad. Sci. USA **100:**14960-14965.

35. **Moreira, D. and P. López-Garcia**. 2005. Comment on 'The 1.2-megabase genome sequence of mimivirus'. Science **308:**1114a.

36. **Moreira, D. and P. López-Garcia**. 2009. Ten reasons to exclude viruses from the tree of life. Nat. Rev. Micro. **7:**306-311.

37. **Morse, S.S.** 1994. Toward an evolutionary biology of viruses. pp. 1-28. *In* S.S. Morse (ed.), The evolutionary biology of viruses. Raven Press, New York.

38. **Pietilä, M.K., E. Roine, L. Paulin, N. Kalkkinen, and D.H. Bamford.** 2009. An ssDNA virus infecting archaea: a new lineage of viruses with a membrane envelope. Mol. Microbiol. **72:**307-319.

39. **Poch, O., I. Sauvaget, M. Delarue, and N. Tordo.** 1989. Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. EMBO J. **8:**3867-3874.

40. **Powner, M.W., B. Gerland, and J.D. Sutherland.** 2009. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. Nature **459:**239-422.

41. **Prangishvili, D., P. Forterre, and R.A. Garrett.** 2006. Viruses of the Archaea: a unifying view. Nat. Rev. Microbiol. **4:**837-848.

42. **Raoult, D., S. Audic, C. Robert, C. Abergel, P. Renesto, H. Ogata, B. La Scola, M. Suzan, and J.-M. Claverie.** 2004. The 1.2-megabase genome sequence of mimivirus. Science **306:**1344-1350.

43. **Rossmann, M.G., E. Arnold, J.W. Erickson, E.A. Frankenberger, J.P. Griffith, H.-J. Hecht, J.E. Johnson, G. Kamer, M. Luo, A.G. Mosser, R.R. Rueckert, B. Sherry, and G. Vriend.** 1985. Structure of a human common cold virus and functional relationship to other picornaviruses. Nature **317:**145-153.

44. **Sage, R.F.** 2004. The evolution of C_4 photosynthesis. The New Phytologist **161:**341–370.

45. **Salehi-Ashtiani, K., A. Lupták, A. Litovchick, and J.W. Szostak.** 2006. A genomewide search for ribozymes reveals an HDV-like sequence in the human CPEB3 gene. Science **313:**1788-1792.

46. **Shackelton, L.A. and E. C. Holmes.** 2004. The evolution of large DNA viruses: Combining genomic information of viruses and their hosts. Trends Microbiol. **12:**458-465.

47. **Sniegowski, P.D., P.J. Gerrish, T, Johnson, and A. Shaver.** 2000. The evolution of mutation rates: separating causes from consequences. BioEssays **22:**1057-1066.

48. **Suttle, C.A.** 2007. Marine viruses – major players in the global ecosystem. Nat. Rev. Micro. **5:**801-812.

49. **Vieth, S., A.E. Torda, M. Asper, H. Schmitz, and S. Günther.** 2004. Sequence analysis of L RNA of Lassa virus. Virology **318:**153-168.

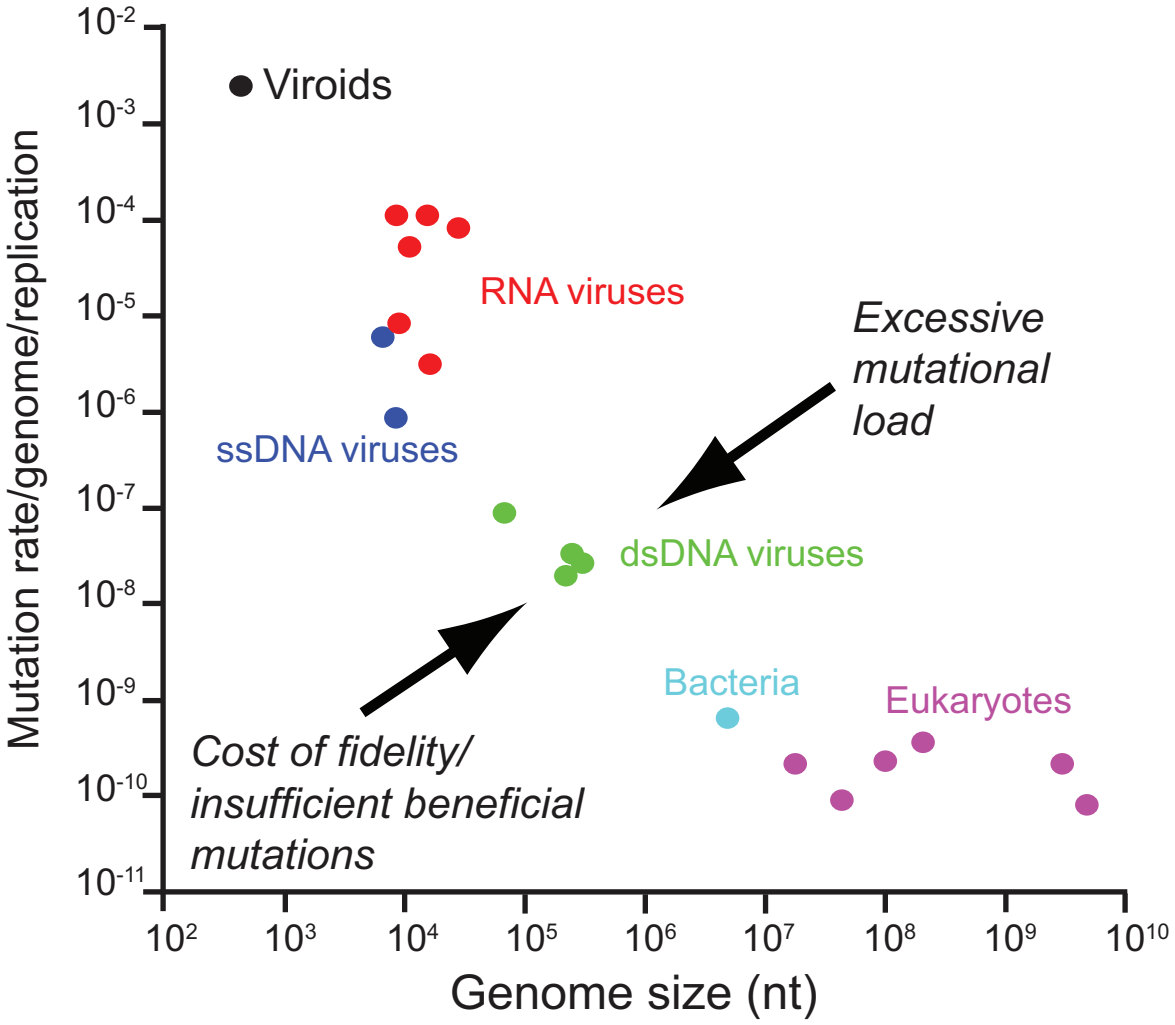
50. **Wu, G.A., S.R. Jun, G.E. Sims, and S.H. Kim.** 2009. Whole-proteome phylogeny of large dsDNA virus families by an alignment-free method. Proc. Natl. Acad. Sci. USA **106:**12826-12831.

51. **Yutin. N., Y.I. Wolf, D. Raoult, and E.V. Koonin.** 2009. Eukaryotic large nucleo-cytoplasmic DNA viruses: clusters of orthologous genes and reconstruction of viral genome evolution. Virol. J. **6:**223.

52. **Zanotto, P.M. de A., M.J. Gibbs, E.A. Gould, and E.C. Holmes.** 1996. A reassessment of the higher taxonomy of viruses based on RNA polymerases. J. Virol. **70:**6083-6096.

FIGURE LEGENDS

FIG. 1. The relationship between error rate and genome size for different genetic systems including viruses. The competing evolutionary forces that might be responsible for the narrow band of viable error rates and genome sizes are also shown. Adapted from ref. 15.



JOURNAL OF VIROLOGY, Sept. 2011, p. 9655 Vol. 85, No. 18
0022-538X/11/$12.00 doi:10.1128/JVI.05689-11

AUTHOR'S CORRECTION

What Does Virus Evolution Tell Us About Virus Origins?

Edward C. Holmes

Center for Infectious Disease Dynamics, Department of Biology, The Pennsylvania State University, Mueller Laboratory, University Park, Pennsylvania 16802, and Fogarty International Center, National Institutes of Health, Bethesda, Maryland 20892

Volume 85, number 11, p. 5247–5251, 2011. Page 5249, Fig. 1, *y* axis: "Mutation rate/genome/replication" should read "Mutations/site/replication."