Adaptation by copy number variation in monopartite viruses

Avraham Bayer[1], Greg Brennan[2] and Adam P Geballe[1,3]

Viruses evolve rapidly in response to host defenses and to exploit new niches. Gene amplification, a common adaptive mechanism in prokaryotes, archaea, and eukaryotes, has also contributed to viral evolution, especially of large DNA viruses. In experimental systems, gene amplification is one mechanism for rapidly overcoming selective pressures. Because the amplification generally incurs a fitness cost, emergence of adaptive point mutations within the amplified locus or elsewhere in the genome can enable collapse of the locus back to a single copy. Evidence of gene amplification followed by subfunctionalization or neofunctionalization of the copies is apparent by the presence of families of paralogous genes in many DNA viruses. These observations suggest that copy number variation has contributed broadly to virus evolution.

Addresses
[1] Divisions of Human Biology and Clinical Research, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
[2] Department of Medical Microbiology and Immunology, University of California, Davis, Sacramento, CA, USA
[3] Departments of Medicine and Microbiology, University of Washington, Seattle, WA, USA

Corresponding author: Geballe, Adam P (ageballe@fredhutch.org)

Introduction

Viruses are extraordinarily diverse in their genomic architecture, gene content, and replication strategies. However, they share common challenges, including the need to exploit the host cell biosynthetic machinery and to evade host defense systems. Over time, evolutionary pressures select for viruses that are better able to replicate and spread to new hosts.

The genetic adaptations that confer fitness improvements can occur by changes as simple as point mutations in single genes. For most RNA viruses, point mutations are generated at a high frequency during genome replication because the replicases lack proof-reading function [1,2]. DNA virus polymerases have much higher fidelity so that point mutations arise less often [3]. This slower mutation rate is likely sufficiently rapid to keep pace with the slow evolution of host genes. However, there are situations, such as drug selection or infections of new and more resistant host species, where a more rapid mechanism of adaptation is beneficial.

In contrast to viruses with segmented genomes, this paper focuses on those in which the genome consists of one polymer of single or double-stranded nucleic acid. Acquisition of new genes by reassortment as occurs with segmented viruses is not possible for these monopartite viruses. However, they can adapt by acquiring large genomic segments by horizontal gene transfer from the host cell, recombination with other viruses, or amplification of endogenous regions of their own genomes. Because viral genome size is often strictly limited by packaging constraints, mechanisms that expand the genome size seem surprising. Nonetheless, considerable evidence exists for viruses having assimilated host genes [4–6] and, as we discuss below, experimental and observational data provide compelling support that copy number variation can and does occur and may play a prominent role in viral evolution.

Gene amplification in experimental systems

The first demonstration of experimental selection for viral gene amplification was described in bacteriophage T4 with an amber nonsense mutation in the essential gene 17 [7]. These phage replicated inefficiently in *Escherichia coli* containing an ochre suppressor. However, some progeny replicated well due to overexpression of gene 17, resulting from tandem amplification of up to 6 copies of a ~4 kb genomic segment containing genes 17 and 18.

Shortly after demonstration of gene amplification as an adaptive mechanism in phage, it was observed in the eukaryotic virus, vaccinia virus (VACV) [8]. Selection for resistance to hydroxyurea, an inhibitor of ribonucleotide reductase, yielded VACV mutants with multiple tandem copies of the viral ribonucleotide reductase gene. More recently, VACV selected for resistance to rifampin, which inhibits VACV virion assembly, yielded a virus with a duplication of a ~2.4 kb segment of the genome [9]. Among the genes in this segment, a truncated variant of A17 was found to account for the rifampin-resistant phenotype. A17 was considered the likely candidate

among the amplified genes since it was known to interact with the scaffold protein D13 during virion assembly. Prior to this study, all known rifampin-resistant mutations were in the D13 gene. Proof of A17's role came from finding that insertion of one extra copy of either the truncated or full length gene alone into VACV conferred rifampin-resistance. The observation that a truncation of A17 conferred rifampin-resistance underscores the fact that amplification does not need to involve a complete gene to provide a replication benefit. Intriguingly, this observation also suggests one mechanism for subfunctionalization, wherein a fragment of a gene encoding a single function may serve as the initiating event to permit separation of multiple functions.

In addition to examples of selection for drug resistant mutants, evolution by gene amplification has been shown to enable adaptation to overcome host cellular defenses. For example, VACV encodes two proteins, E3L and K3L, that block the host protein kinase R (PKR) antiviral pathway [10]. E3L is a double-stranded RNA-binding protein that is a potent inhibitor of PKR in many primate cells, while K3L acts as a pseudosubstrate inhibitor of PKR in some rodent cells but has limited efficacy in antagonizing primate PKRs. As a result, VACV containing K3L but lacking E3L (VACVΔE3L) replicates very inefficiently in human cells. Serial passage of VACVΔE3L through human cells selected for viruses with amplification of K3L and portions of the flanking genes [11[••]]. Consistent with a simple gene dosage mechanism, the viruses with K3L amplification expressed high levels of the K3L protein and knocking down K3L expression reversed the replication phenotype.

In a similarly designed set of experiments, Brennan *et al.* engineered a VACV recombinant lacking VACV PKR antagonists (E3L and K3L) and containing a cytomegalovirus PKR antagonist, rTRS1, that is active in some but not all Old World monkey cells [12[•]]. This virus replicated poorly in one African green monkey cell line and not at all in human and rhesus cells. However, after a few passages, mutant viruses emerged that replicated to ~10-fold higher titers as a result of amplification of the rTRS1 locus. Notably, these viruses replicated better than the parent virus not only in the African green monkey cells in which they were selected, but also in cells from other species. This result highlights the potential role of gene amplification in conferring rapid adaptation that could facilitate cross-species transmission.

A common feature in these experiments is the rapidity with which the amplifications arose. Erlandson, *et al.* isolated a virus with the A17 amplification after a single plaque purification in the presence of rifampin [9]. The emergence of adapted VACVs with amplifications of PKR antagonists took only 4–6 passages in multiple independent experiments [11[••],12[•]]. It is unclear when the

ribonucleotide reductase duplication arose following exposure to hydroxyurea, but resistance was noted between 3 and 6 passages, suggesting a similar time course [13]. One possible explanation for these kinetics is that random gene amplifications arise often during VACV replication. Though rare at any particular locus in the starting pool, a virus with an adaptive amplification will be enriched with each passage until it emerges as the dominant viral species after relatively few passages. Consistent with this model, Elde *et al.* detected a low frequency of gene duplications at several other loci in the VACV genome distant from the K3L amplification [11[••]]. Sequencing methods with higher fidelity than those used previously should help clarify the frequency, location, and size distribution of amplifications in unselected viral populations.

These studies reveal that the gene amplification mechanism is quite versatile. In the VACV experiments, the amplifications occurred at multiple loci throughout the genome, including both natural VACV sequences and those from heterologous sequences [8,9,11[••],12[•]]. Multiple different 'breakpoints' were created by the amplification in different VACV experiments. Although no sequence similarities were recognized at these sites in the VACV experiments, most of the T4 phage with amplification of genes 17 and 18 shared the same breakpoint within a short region of homology (20 of 24 bases) in genes 16 and 19 [14]. Even though the mechanism accounting for the initial duplication is unclear and might vary in phage vs. eukaryotic viruses, once duplicated, expansion and contractions by a recombination mechanism appear to occur in a highly dynamic manner. For example, after plaque purification of VACVs with the amplified loci, the progeny virus genomes consistently contain a variable number of copies [11[••],15]. In some cases, the amplifications increase the genome size by 10–20% [8,11[••]]. However, it is not known if these large genome variants are replication competent.

Although the amplification increases viral replication under selective conditions, it likely incurs a fitness cost. In support of this idea, the gene 17 and 18 locus amplification in the T4 phage with the amber mutation disappeared immediately when the selective pressure was removed by propagation in *E. coli* with an amber suppressor [14] (Figure 1a). Similarly, the amplification of the VACV A17 locus was lost when the virus was propagated in the absence of rifampin [9]. The K3L average gene copy number decreased when the VACVΔE3L having multiple K3L copies was passed in rodent cells, in which a single copy of K3L is sufficient for efficient replication [11[••]]. In the absence of pressure for high level expression of the amplified gene, viruses having the amplifications might be less fit because overexpression of the gene might be toxic to cellular processes or perturb the normal expression pattern that may otherwise be finely tuned for

Figure 1
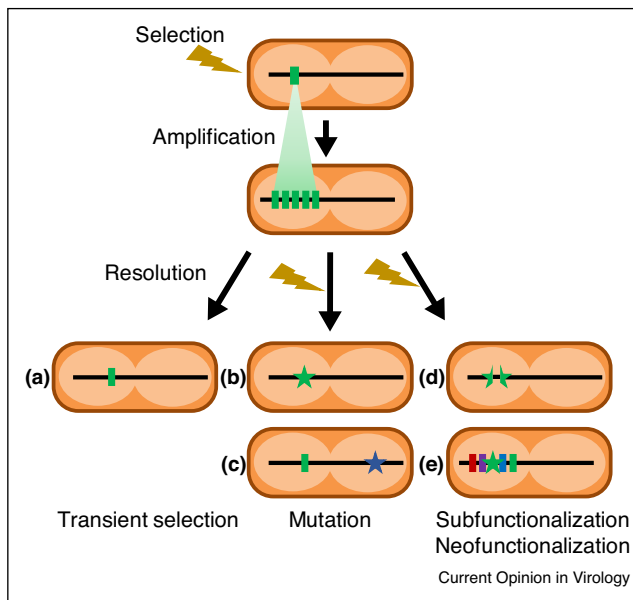
Initiation and resolution of gene amplification in viruses. In the presence of selective pressure (lightning bolt), a gene with incomplete activity in counteracting that selective pressure (green bar) may undergo amplification (center), providing a replication benefit and enabling the virus to sample more evolutionary space per round of replication. If selective pressure is eliminated **(a)** (e.g. by removal of drug selection or entry into a new nonimmune host), the amplification may collapse without changes. In other cases, adaptive point mutations (stars) may evolve in either **(b)** the amplified gene or **(c)** another viral gene, that relieve selective pressure and enable the amplification to collapse. Alternatively, amplified genes may also undergo **(d)** subfunctionalization, dividing the activity of one multifunctional gene into two or more new genes, or **(e)** neofunctionalization, wherein a mutation in one or more of amplified genes acquire novel functions.

optimal replication. The overexpressed protein might also lead to enhanced targeting of the infected cells by the host immune system. Alternatively or in addition, the unnecessary expenditure of energy and resources needed for replication, transcription and translation of the extra gene copies might aid in selection for the collapse of the amplified locus. However, there are not compelling data to distinguish among these and others possible explanations for the costs of the amplification.

Mutation allowing collapse of the amplified region — the genetic 'accordion'

In cases where the selective pressure persists, new adaptive mutations can arise, eliminating the need for the amplification and allowing collapse back to a single copy. Along with selection of K3L amplification, Elde *et al.* detected viruses with a missense mutation (H47R) in K3L, which enabled the virus to replicate well even with only a single copy of the mutant gene [11[••]]. Remarkably this exact same mutation had been identified in an

entirely independent screen of random K3L mutants to identify ones with an improved ability to block human PKR function in a yeast-based assay [16]. Thus, it is possible that the extra K3L genes in the amplified virus facilitated sampling of many mutations at this locus. The sequence of gene amplification followed by selection of an adaptive mutations and subsequent collapse of the amplified locus, referred to as a 'genetic accordion,' is depicted in Figure 1b.

Pressure conferred by the fitness costs of the amplification can also select for mutations in genes outside of the amplified locus (Figure 1c). For instance, Brennan *et al.* identified mutations in A24R and A35R, and the emergence of these mutations correlated with loss of rTRS1 amplification after extended passage under selection in the African green monkey cells [15]. Cone *et al.* also identified mutations in A24R along with K3L amplification, after passage of VACVΔE3L in human fibroblasts [17]. A24R encodes a subunit of the viral RNA polymerase, raising the hypothesis that some A24R mutations may lessen the processivity of the RNA polymerase, resulting in less accumulation of dsRNA, which activates PKR, and thus less dependence on potent inhibition of the PKR pathway. Indeed, Cone *et al.* found that two A24R mutant viruses each produced lower amounts of dsRNA than the parent virus. However, paradoxically, one of these mutants activates the PKR pathway, leaving unanswered the question of how it improves VACV fitness. Regardless, these studies illuminated roles for several genes not previously implicated in the PKR pathway.

Logically, one might expect that the amplified gene would be the most likely to accumulate adaptive mutations; nonetheless, these experiments demonstrate that mutations can arise at distant loci as well. Thus, in the case of these large viruses, the replication benefit accruing from gene amplification may simply be a readily accessible adaptive mechanism that can support sufficient rounds of replication to enable sampling of mutations at any point in the genome. Conceptually, this mechanism may be particularly advantageous in cases where multiple mutations are required to adapt the amplified gene while only a single mutation is required in another, distant gene.

Evidence of gene amplification in nature

Gene duplications occur in all three domains of life. The studies described above demonstrate that gene amplification can occur in viruses as well, at least under well-defined laboratory conditions. However, many genes in large viruses are dispensable for replication in cell culture. Thus, if amplification at one locus generates compensatory deletions elsewhere, the resulting virus might be able to replicate in cell culture but not in nature. Although sequencing performed during the VACV studies did not reveal any such deletions [11[••],12[•],15], VACV may be a

special case as its unconventional structure might be unusually tolerant of genome expansion [18]. Regardless, there is compelling evidence from viral genome sequence analyses that gene duplication has occurred broadly during the evolutionary history of viruses [19]. Because there is a correlation between the genomic complexity (size and number of genes) and the number of duplicated genes, most of the examples come from large DNA viruses [19, 20•,21,22].

One of the insights emerging from the experimental evolution studies is that gene amplifications in viruses may be transient. If a selective pressure is present for only a brief time or adaptive point mutations arise that obviate the need for the extra copies, the amplified segment can collapse back to a single copy (Figure 1a). Thus, detection of amplification in nature could easily be missed. However, an alternative outcome following amplification is the evolution of modified or new functions in the gene copies [22].

Gene amplification followed by subfunctionalization (Figure 1d) or neofunctionalization (Figure 1e) likely explains the presence of paralogous genes within a virus. For example, the ~230 kb genomes of cytomegaloviruses (CMV) encode hundreds of genes. A surprising finding when the human CMV was sequenced was the presence of multiple families of related genes [23,24]. The US12 family, arranged as a series of 10 tandemly repeated genes in the unique short region of the genome, encode a superfamily of proteins characterized by 7 transmembrane segments, several of which contribute to evasion of natural killer cells [24,25••,26]. The fact that these genes are found in primate CMVs but not rodent CMVs and that they are present in direct tandem copies suggests they emerged during a relatively recent amplification event.

In contrast, the US22 gene family, which consists of ~12 members in most CMVs, appears to have arisen earlier, as homologs are found in rodent and primate CMVs, and the genes are now dispersed throughout the genomes [23,27]. Several of the gene products have dsRNA-binding activity, which maps to their most conserved regions [28,29]. Two of the twelve MCMV US22 family genes, *m142* and *m143*, are the only family members that are essential for viral replication [27]. Their protein products work together in a complex that antagonizes PKR, suggesting amplified copies evolved specialized subfunctions of the PKR antagonism mechanism [29–32]. Whether PKR antagonism, which is a highly conserved and critical function in many viruses [33], was the driver of the initial US22 gene family duplication is unknown. PKR has been evolving rapidly, presumably in order to evade viral antagonists encoded by many viruses [34,35]. The subfunctionalization and neofunctionalization of viral genes within the CMV US12 and US22 families adds support to

the hypothesis that amplification may be a critical mechanism for evading host defense systems.

Analyses of many other DNA virus genomes reveal evidence of past gene amplification events, often clustered towards the viral termini [19,36]. Fowlpox virus contains 10 gene families, including a remarkable one with 31 members having ankyrin repeats, clustered at both ends of the genome [37]. Adenoviruses have smaller genomes, but have undergone genus-specific or species-specific duplication events followed by subsequent divergence [38]. The γ-herpesvirus genomes contain many gene copy number expansions, both in protein coding genes as well as tandemly repeated copies of RNA polymerase III transcripts, including EBERs, microRNAs, and tRNAs [39–41].

One of the most striking examples of gene duplication can be found in one of the largest viruses known to date, mimivirus that infects *Acanthamoeba polyphaga*. Within the 1.2 million base pair genome, ~38% of the genes are thought to have originated from duplication events [42,43]. Many of the duplicated genes exist in tandem arrays and appear to be involved in virus host-interactions. For example, the protein kinase family (N232) and the F-box containing cluster (N165), interfere with host signaling or protein degradation respectively. The huge expansion of the mimivirus genome resulting from the amplifications may be adaptive, as the virus may have enlarged to mimic 'prey' to the amoeban predator in order to enter via phagocytosis. Thus mimivirus might have evolved its large genome and physical size because the beneficial effects on uptake by its host exceeded the potential costs of the genomic expansion that are evident in more conventional viruses [20•,42,44].

While DNA viruses, and in particular large dsDNA viruses, exhibit numerous examples of gene duplication, this is not the case with small dsRNA viruses, ssDNA viruses or ssRNA viruses [21]. Among the unusual gene duplications found in RNA viruses, the primate lentiviruses Vpr and Vpx accessory proteins likely arose by gene duplication [45,46]. They remain ~30% similar in amino acid sequence and both antagonize the host restriction factor SAMHD1 [47]. The smaller size of RNA and ssDNA viruses may preclude exploitation of amplifications as a general mechanism for adaptation. Interestingly, the Coronaviridae contain one of the largest genome sizes among RNA viruses, ~30 kb in length. Unlike other RNA viruses which typically have high point mutation rates due to low fidelity replicases [3], coronavirus replication employs exonuclease activity that improves replication fidelity [48]. Limited evidence suggests that a genomic duplication occurred to generate proteases of human coronavirus, PL1pro and PL2pro [49]. Together, these observations suggest genomic expansion

is restricted primarily by genomic size constraints, rather than by type of nucleic acid.

Conclusion

For large DNA viruses in particular, duplication and subsequent dynamic copy number variation of genomic segments can enable rapid adaptation to various challenges such as inhibitory drugs and host antiviral defense systems. In the laboratory, experimental evolution has illuminated a role for transient gene amplification as a rapid adaptive mechanism, but one that often incurs a fitness cost. This fitness trade-off may in principle fill the same adaptive niche for DNA viruses as error-prone replication and rapid mutation rates in RNA viruses. Amplification events may occur more frequently in nature than is currently recognized due to their often-transient nature. However, viruses from multiple different families contain paralogous genes that seem to have arisen by gene duplication events. In these cases, the copied genes accumulated mutations that conferred modified or new functions. These observations support the hypothesis that transient gene amplification events may play a prominent, albeit largely invisible, role in virus evolution. Duplicated genes also can provide new genetic substrates for the evolution of genes with specialized or new functions just as it does in every other branch of life.

Conflict of interest statement

None declared.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Roberts JD, Bebenek K, Kunkel TA: **The accuracy of reverse transcriptase from HIV-1**. *Science* 1988, **242**:1171-1173.

2. Steinhauer DA, Domingo E, Holland JJ: **Lack of evidence for proofreading mechanisms associated with an RNA virus polymerase**. *Gene* 1992, **122**:281-288.

3. Sanjuan R *et al.*: **Viral mutation rates**. *J Virol* 2010, **84**:9733-9748.

4. Gilbert C, Cordaux R: **Viruses as vectors of horizontal transfer of genetic material in eukaryotes**. *Curr Opin Virol* 2017, **25**:16-22.

5. Lai MM: **RNA recombination in animal and plant viruses**. *Microbiol Rev* 1992, **56**:61-79.

6. Renner DW, Szpara ML: **Impacts of genome-wide analyses on our understanding of human herpesvirus diversity and evolution**. *J Virol* 2018:92.

7. Wu DG, Black LW: **Gene amplification mechanism for the hyperproduction of T4 bacteriophage gene 17 and 18 proteins**. *J Mol Biol* 1987, **195**:769-783.

8. Slabaugh M *et al.*: **Vaccinia virus-encoded ribonucleotide reductase: sequence conservation of the gene for the small subunit and its amplification in hydroxyurea-resistant mutants**. *J Virol* 1988, **62**:519-527.

9. Erlandson KJ *et al.*: **Duplication of the A17L locus of vaccinia virus provides an alternate route to rifampin resistance**. *J Virol* 2014, **88**:11576-11585.

10. Langland JO, Jacobs BL: **The role of the PKR-inhibitory genes, E3L and K3L, in determining vaccinia virus host range**. *Virology* 2002, **299**:133-141.

11. Elde NC *et al.*: **Poxviruses deploy genomic accordions to adapt
•• rapidly against host antiviral defenses**. *Cell* 2012, **150**:831-841.
This paper describes the' accordion model' in which gene amplification provides extra templates that can facilitate emergence of adaptive point mutations, in turn allowing collapse of the transiently amplified locus.

12. Brennan G *et al.*: **Adaptive gene amplification as an
• intermediate step in the expansion of virus host range**. *PLoS Pathog* 2014, **10**:e1004002.
This paper illustrates how viral gene amplification in response to a restriction factor in one host might facilitate successful transmission into a new host species.

13. Slabaugh MB, Mathews CK: **Hydroxyurea-resistant vaccinia virus: overproduction of ribonucleotide reductase**. *J Virol* 1986, **60**:506-514.

14. Wu DG, Wu CH, Black LW: **Reiterated gene amplifications at specific short homology sequences in phage T4 produce Hp17 mutants**. *J Mol Biol* 1991, **218**:705-721.

15. Brennan G *et al.*: **Experimental evolution identifies vaccinia virus mutations in A24R and A35R that antagonize the protein kinase R pathway and accompany collapse of an extragenic gene amplification**. *J Virol* 2015, **89**:9986-9997.

16. Kawagishi-Kobayashi M *et al.*: **Regulation of the protein kinase PKR by the vaccinia virus pseudosubstrate inhibitor K3L is dependent on residues conserved between the K3L protein and the PKR substrate eIF2alpha**. *Mol Cell Biol* 1997, **17**:4146-4158.

17. Cone KR *et al.*: **Emergence of a viral RNA polymerase variant during gene copy number amplification promotes rapid evolution of vaccinia virus**. *J Virol* 2017:91.

18. Condit RC, Moussatche N, Traktman P: **In a nutshell: structure and assembly of the vaccinia virion**. *Adv Virus Res* 2006, **66**:31-124.

19. Gao Y *et al.*: **Extent and evolution of gene duplication in DNA viruses**. *Virus Res* 2017, **240**:161-165.

20. Shukla A, Chatterjee A, Kondabagil K: **The number of genes
• encoding repeat domain-containing proteins positively correlates with genome size in amoebal giant viruses**. *Virus Evol* 2018, **4**:vex039.
This paper explores the role of 'accordion-like' evolution of amoeba giant virus genomes, and the hypothesis that the large viral size arose by duplications that helped these virus enter into the host cells.

21. Simon-Loriere E, Holmes EC: **Gene duplication is infrequent in the recent evolutionary history of RNA viruses**. *Mol Biol Evol* 2013, **30**:1263-1269.

22. Espinosa-Cantu A *et al.*: **Gene duplication and the evolution of moonlighting proteins**. *Front Genet* 2015, **6**:227.

23. Chee MS *et al.*: **Analysis of the protein-coding content of the sequence of human cytomegalovirus strain AD169**. *Curr Top Microbiol Immunol* 1990, **154**:125-169.

24. Weston K, Barrell BG: **Sequence of the short unique region, short repeats, and part of the long repeats of human cytomegalovirus**. *J Mol Biol* 1986, **192**:177-208.

25. Fielding CA *et al.*: **Control of immune ligands by members of a
•• cytomegalovirus gene expansion suppresses natural killer cell activation**. *Elife* 2017, **6**.
This paper reveals a new function in natural killer cell evasion for a tandemly amplified family of paralogous genes in primate cytomegaloviruses.

26. Lesniewski M *et al.*: **Primate cytomegalovirus US12 gene family: a distinct and diverse clade of seven-transmembrane proteins**. *Virology* 2006, **354**:286-298.

27. Menard C *et al.*: **Role of murine cytomegalovirus US22 gene family members in replication in macrophages**. *J Virol* 2003, **77**:5557-5570.

28. Hakki M, Geballe AP: **Double-stranded RNA binding by human cytomegalovirus pTRS1**. *J Virol* 2005, **79**:7311-7318.

29. Child SJ *et al.*: **Double-stranded RNA binding by a heterodimeric complex of murine cytomegalovirus m142 and m143 proteins**. *J Virol* 2006, **80**:10173-10180.

30. Budt M *et al.*: **Specific inhibition of the PKR-mediated antiviral response by the murine cytomegalovirus proteins m142 and m143**. *J Virol* 2009, **83**:1260-1270.

31. Child SJ, Geballe AP: **Binding and relocalization of protein kinase R by murine cytomegalovirus**. *J Virol* 2009, **83**:1790-1799.

32. Valchanova RS *et al.*: **Murine cytomegalovirus m142 and m143 are both required to block protein kinase R-mediated shutdown of protein synthesis**. *J Virol* 2006, **80**:10181-10190.

33. Langland JO *et al.*: **Inhibition of PKR by RNA and DNA viruses**. *Virus Res* 2006, **119**:100-110.

34. Elde NC *et al.*: **Protein kinase R reveals an evolutionary model for defeating viral mimicry**. *Nature* 2009, **457**:485-489.

35. Rothenburg S *et al.*: **Rapid evolution of protein kinase PKR alters sensitivity to viral inhibitors**. *Nat Struct Mol Biol* 2009, **16**:63-70.

36. Shackelton LA, Holmes EC: **The evolution of large DNA viruses: combining genomic information of viruses and their hosts**. *Trends Microbiol* 2004, **12**:458-465.

37. Afonso CL *et al.*: **The genome of fowlpox virus**. *J Virol* 2000, **74**:3815-3831.

38. Davison AJ, Benko M, Harrach B: **Genetic content and evolution of adenoviruses**. *J Gen Virol* 2003, **84(Pt 11)**:2895-2908.

39. Arias C *et al.*: **KSHV 2.0: a comprehensive annotation of the Kaposi's sarcoma-associated herpesvirus genome using next-generation sequencing reveals novel genomic and functional features**. *PLoS Pathog* 2014, **10(1)**:e1003847.

40. Moss WN *et al.*: **RNA families in Epstein-Barr virus**. *RNA Biol* 2014, **11**:10-17.

41. Nicholas J: **Evolutionary aspects of oncogenic herpesviruses**. *Mol Pathol* 2000, **53**:222-237.

42. Suhre K: **Gene and genome duplication in Acanthamoeba polyphaga Mimivirus**. *J Virol* 2005, **79**:14095-14101.

43. Filee J: **Genomic comparison of closely related Giant Viruses supports an accordion-like model of evolution**. *Front Microbiol* 2015, **6**:593.

44. Rodrigues RAL *et al.*: **Giants among larges: how gigantism impacts giant virus entry into amoebae**. *Curr Opin Microbiol* 2016, **31**:88-93.

45. Tristem M *et al.*: **Evolution of the primate lentiviruses: evidence from vpx and vpr**. *EMBO J* 1992, **11**:3405-3412.

46. Tristem M *et al.*: **Origin of vpx in lentiviruses**. *Nature* 1990, **347**:341-342.

47. Lim ES *et al.*: **The ability of primate lentiviruses to degrade the monocyte restriction factor SAMHD1 preceded the birth of the viral accessory protein Vpx**. *Cell Host Microbe* 2012, **11**:194-204.

48. Minskaia E *et al.*: **Discovery of an RNA virus 3′-&5′ exoribonuclease that is critically involved in coronavirus RNA synthesis**. *Proc Natl Acad Sci U S A* 2006, **103(13)**:5108-5113.

49. Ziebuhr J, Thiel V, Gorbalenya AE: **The autocatalytic release of a putative RNA virus transcription factor from its polyprotein precursor involves two paralogous papain-like proteases that cleave the same peptide bond**. *J Biol Chem* 2001, **276**:33220-33232.