BMC Evolutionary Biology

BioMed Central

Research article

Whole genome molecular phylogeny of large dsDNA viruses using composition vector method

Lei Gao*†[1,2] and Ji Qi†[3]

Address: [1]The Institute of Theoretical Physics, Academia Sinica, Beijing 100080, China, [2]The T-Life Research Center, c/o Department of Physics, Fudan University, Shanghai 200433, China and [3]Center for Comparative Genomics and Bioinformatics, Penn State University, University Park, PA 16802, USA

Email: Lei Gao* - highlei@gmail.com; Ji Qi - jxq11@psu.edu

* Corresponding author †Equal contributors

Published: 15 March 2007

BMC Evolutionary Biology 2007, **7**:41 doi:10.1186/1471-2148-7-41

Received: 19 September 2006
Accepted: 15 March 2007

This article is available from: http://www.biomedcentral.com/1471-2148/7/41

Abstract

Background: One important mechanism by which large DNA viruses increase their genome size is the addition of modules acquired from other viruses, host genomes or gene duplications. Phylogenetic analysis of large DNA viruses, especially using methods based on alignment, is often difficult due to the presence of horizontal gene transfer events. The recent composition vector approach, not sensitive to such events, is applied here to reconstruct the phylogeny of 124 large DNA viruses.

Results: The results are mostly consistent with the biologist's systematics with only a few outliers and can also provide some information for those unclassified viruses and cladistic relationships of several families.

Conclusion: With composition vector approach we obtained the phylogenetic tree of large DNA viruses, which not only give results comparable to biologist's systematics but also provide a new way for recovering the phylogeny of viruses.

Background

Viruses are small, infectious, obligate intracellular parasites that are capable of replicating themselves within their host cells. They are even smaller than the smallest elementary biosystem, yet still possess some properties of living systems such as having a genome and the ability to adapt to changing environments. However, viruses cannot capture and store free energy and they are not functionally active outside their host cell [1].

Traditionally, viruses were characterized by morphological features including capsid size, shape, structure, etc., as well as physicochemical and antigenic properties. As more and more viral genomes are being sequenced, the evolutionary relationship of a great many families and genera is being explored [2,3] by sequence analysis on single gene or gene families, such as polymerase, capsid and movement genes [4-10]. The virus taxonomy system is approved and updated by the International Committee on the Taxonomy of Viruses (ICTV).

However, it is full of ambiguity for phylogenetic analysis based on single gene when using conserved or similar genes since horizontal gene transfer (HGT) between viruses, along with gene duplication, gene capture from host appears to have been frequent in large DNA viruses [11-14]. Genetic mosaicism of phages has been known for a long time. Homologous morons (coding region and

transcription control sequences) are found in many lineages of phages [15], and these kinds of genetic acquirements have also been considered the main sources of increasing genome size in large DNA viruses [9,16]. The high substitution rate of viruses also limits the sequence-based methods from revealing the distant evolutionary relationships [11,17]. For example, the herpes simplex virus type 1 mutates 10 times greater than mammalian genes (nearly 3.5×10^8 substitutions per site per year) [18,19]. Some attempts have been made to combine viral structure and function characteristics and genomic information of their hosts into sequence information [11,17] although quantifying such structural similarity has proved to be extremely challenging.

In the meantime there are several other attempts to infer viral phylogeny from their whole genomes [12,20-25] to avoid the problem of gene rearrangement, gene loss, gene duplication and lateral gene transfer. However, some of them infer the majority consensus tree of the many trees of individual genes or use the combined sequences of many shared genes [12,21,22]. Some of them employ gene content [12,22,23] and gene order [12,22] method, but the former has to correct for the genome size effect [26] and the latter can be hindered by a lack of synteny conservation or the variation of the evolving rate of synteny between taxa [12,26]. Above methods are partly or completely based on alignment of conserved or similar sequences which is hard to infer more distant evolutionary relationships. Gao and Stuart [20,25] apply new alignment-free methods to resolve virus relationships respectively, which appear to be sufficiently powerful to explore the phylogeny of viruses at large evolutionary distance.

Since viruses have no universal common genes just like SSU rRNA in cellular life, it is difficult to reconstruct the phylogenetic tree for distinct type of virus with most of former methods. We present here a phylogenetic analysis of large DNA viruses with the Composition Vector (CV) method [20,27,28] and discuss their relationships at a deep level. The CV method does not require extended alignments, predefined operational orthologs, or even predefined homologs (for details see material and methods). We show that the results are mostly consistent with the biologist's systematics with only a few outliers and also provide some information for those unclassified viruses and cladistic relationships of several families.

Results and discussion

A phylogenetic tree including 124 dsDNA viruses is shown in Figure 1. Apparently, despite numerous horizontal gene transfer among large DNA viruses [13], our analysis is able to divide the 124 dsDNA viruses into 10 families with only 4 outliers, CuniNPV, IIV-6, IcHV-1 and OsHV-1 (see Additional file 1: 124 large dsDNA virus names, abbreviations, and NCBI accession numbers for viruses names, abbreviations and accession numbers).

Phylogenetic relationships of all 124 dsDNA viruses coming from 33 genera, 10 families are well consistent with the taxonomy by ICTV [1] and other phylogenetic studies [9] with few exceptions.

Adenoviridae

Fig. 1 and Fig. 2a supports the division of this family into four genera. It is notable that the two genera, *Atadenovirus* and *Siadenovirus*, which both comprise viruses from a variety of hosts locate between another two genera, *Mastadenovirus* whose hosts are mammals and *Aviadenovirus* whose hosts are birds. This variety of host origin supports the hypothesis that interspecies transmission, i.e. host switches of adenoviruses, may have occurred [29].

Baculoviridae

According to the classification in ICTV database, one of the largest families in dsDNA viruses, *Baculoviridae*, is composed of two genera, *Granulovirus* and *Nucleopolyhedrovirus*. However, the 4-subbranch of this family shown in Fig. 1 and Fig. 2b complies with the classification of their hosts. *Dipteran*-infecting baculovirus, CuniNPV, locates the most deep [30] and stay outside the whole large clade, followed by the *Hymenoptera* baculovirus (NeleNPV and NeseNPV) and *Lepidoptera* baculovirus (the rest of them) [12,31]. There are three hypotheses on the origin of *Baculoviruses*: originated within *Lepidoptera* with subsequent horizontal transmissions to other insect orders [32]; originated with the cocladogenesis of the viruses and their hosts [33]; originated from the ancestral baculoviruses that were probably able to infect the hosts of different orders, with ancient coevolution between the hosts and pathogens then leading to the different order host specialization [12]. Our analysis apparently provides equal support to the last two hypotheses as postulated by Herniou [30], without comparing the division time of viruses and their hosts.

Poxviridae

The division of *Poxviridae* into two subfamilies *Chordopoxvirinae* (ChPV) and *Entomopoxvirinae* (EnPV) shown in Fig. 1 and Fig. 2d, is the same as in the systematics of ICTV. Within the cluster *Orthopoxviruses* in ChPV, ECTV and CPXV are the most divergent, which is similar to McLysaght's study [34] and is also supported by another analysis based on multiple genes alignment [3]. *Capripoxvirus*, *Leporipoxvirus*, *Suipoxvirus* and *Yatapoxvirus* form another cluster, in which *Capripoxvirus* and *Suipoxvirus* are much closer to each other, and this supports the hypothesis that they might have evolved from a common ancestor [3]. In Figure 1, DPV, an unclassified *Poxvirus*, is

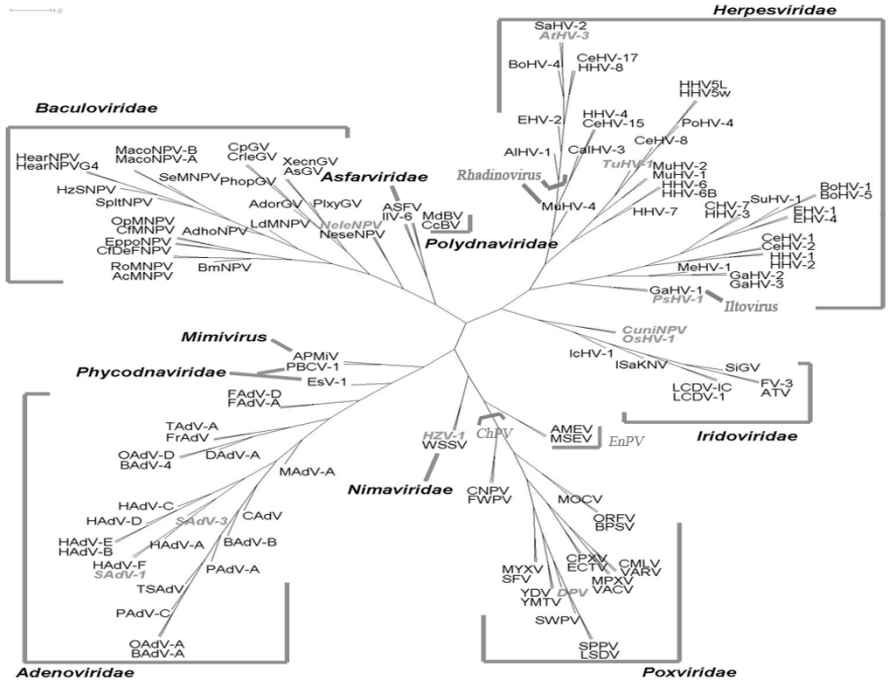

Figure 1

Tree of 124 large DNA viruses. Phylogenetic analysis of large DNA viruses based on CV method with K-string length *K* = 5. Altogether 33 genera from 10 families are presented. The nine unclassified dsDNA viruses are indicated by grey. Family names are placed close to the corresponding branches. Note that this is an unrooted tree and the branches are not to scale.

assigned to ChPV subfamily, which agrees with Afonso's result [35].

Herpesviridae

Within the *Herpesviridae* family, Fig. 1 and Fig. 2c also supports the observations of an early split of the *Beta-* and *Gammaherpesviruses* from the *Alphaherpesviruses* [36]. It is worth mentioning that TuHV-1, previously known only to belong to the *Beta*-subfamily, now is found to cluster with *Cytomegaloviruses* in Fig. 1, which follows Bahr's analysis [37]. According to the taxonomy system of ICTV, the *Gamma*-subfamily consists of two genera, *Lymphocryptovirus* and *Rhadinovirus*. MuHV-4, whose position was previously unresolved and various [36,38] is now assigned to *Rhadinoviruses* according to ICTV, and is the most divergent in fig. 1. The rest two ungulate herpesviruses within *Rhadinoviruses*, AIHV-1 and EHV-2, are divergent from others which is in accordance with other analyses [36,38]. However, another ungulate herpesvirus, BoHV-4, clusters closely to HHV-8 and CeHV-17, which is incompatible with the hypothesis that herpesviruses have coevolved with their hosts [39].

All *Iridoviruses* except IIV-6 fall into one cluster in fig. 1. IIV-6 and ASFV from *Asfarviridae* group together, which partly supports the theory that *Iridoviridae* and *Asfarviridae* are monophyletic [40]. It is interesting to note that ISaKNV, which was still an unclassified *Iridovirus* at the time we fixed our data sets, has been placed in a new genus *Megalocytivirus*, which supports both our analysis and Do's [41]. The same is true of PsHV-1, which is assigned to *Alphaherpesvirinae* in fig. 1 just as ICTV has done not long ago, and it should belong to *Iltovirus* for it clusters to GaHV-1. Similarly, WSSV, an unclassified marine invertebrate virus [42], has also been classified into a new virus family, *Nimaviridae*, which is again supported by our results.

Our results could also provide some clues to these hypotheses about origins and evolution of viruses of several families.

Several unclassified viruses are analyzed for obtaining some hints for their possible taxonomic statuses. As shown in fig. 1, PsHV-1 may belong to *Iltovirus*, which was supported by Thureen's result [43]; AtHV-3 to *Rhadinovirus* the same as McGeoch's analysis [44]; TuHV-1 to Cytomegalovirus just as Bahr's result [37]. SAdV-3 and SAdV-1 are close to HAdV-(A, B, C, D, E, F), and they may be two new species of *Mastadenovirus* [45,46]. NeleNPV and NeseNPV group together, and they may belong to a new genus according to Herniou's results [30]. In fig. 1 DPV locates between *Suipoxvirus* and *Yatapoxvirus* but not very closes to each of them, further supporting the idea

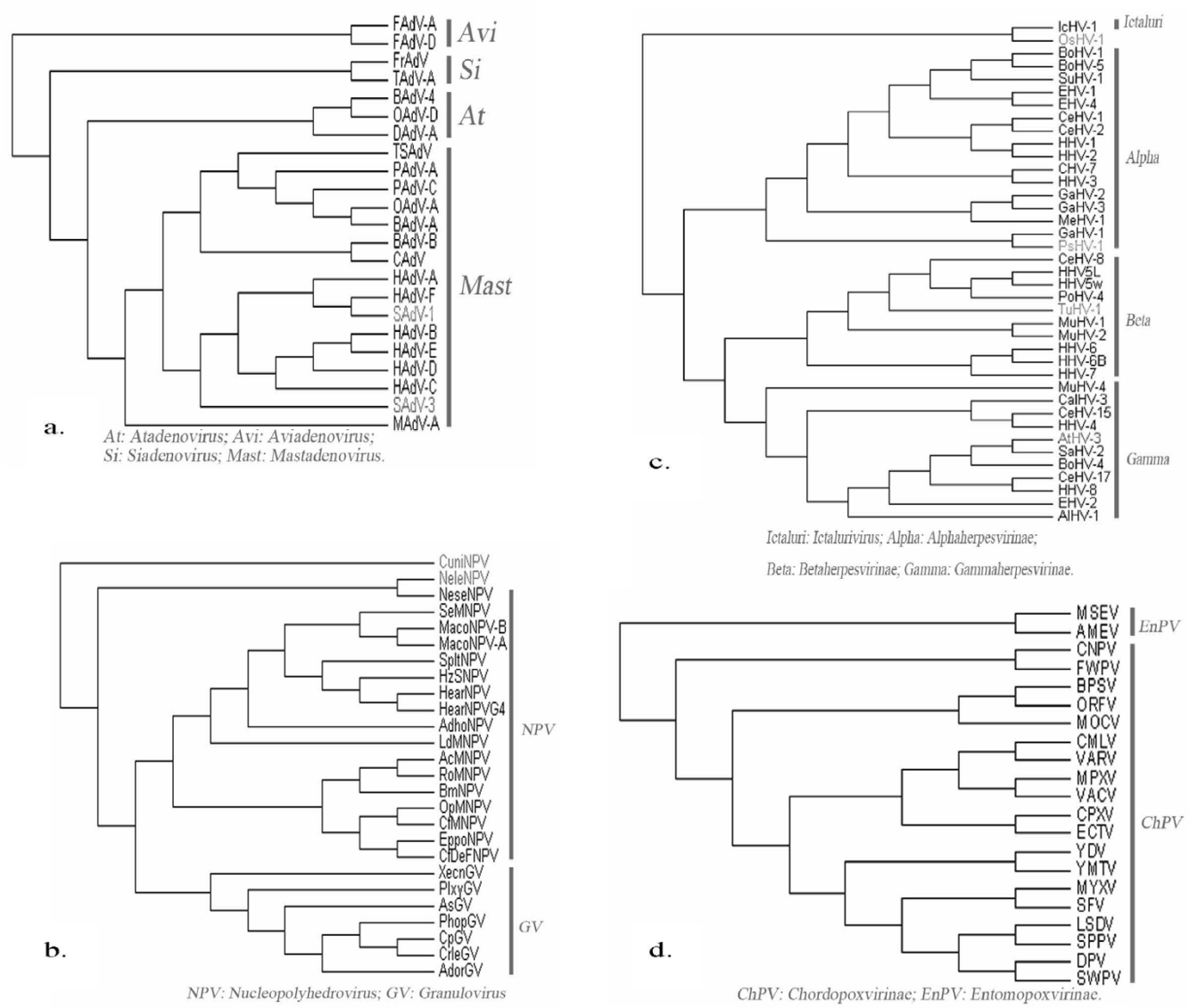

Figure 2

Trees of 4 virus families. Phylogenetic analysis of 4 virus families based on CV method with K-string length *K* = 5. a: Adenoviridae; b: Baculoviridae; c: Herpesviridae; d: Poxviridae. The unclassified dsDNA viruses are indicated by grey. Genus or subfamily names are placed close to the corresponding branches.

that it appears to be assigned to a new genus *Cervidpoxvirus* [35,47]. HZV-1, originally defined as *Baculovirus* but currently as an unclassified dsDNA virus [48], clusters with WSSV whin *Nimaviridae* in our results.

However, there are some outliers in fig. 1. IcHV-1 and OsHV-1 group closely but jump out of the branch of *Herpesviridae*, which is consistent with dissimilarities in sequence comparisons between OsHV-1 and the three vertebrate herpesvirus subfamilies [49] and up-to-date classification to two new families [50]. CuniNPV stay outside the whole large clade of *Baculoviridae* [51], and it also should belong to a new genus according to Herniou's proposal [30]. IIV-6 closes with ASFV [40] but stays outside all other *Iridoviruses*, which may support partly the theory that *Iridoviridae* and *Asfarviridae* are monophyletic [40].

Conclusion

We present here a phylogenetic analysis of large DNA viruses with the CV method and discuss their relationships at a deep level. The results support the biologist's systematics in overall structure and in many details and provide some clues to these hypotheses about origins and evolution of viruses of several families.

It should be pointed out that although baculoviruses and their hosts are obviously subject to coevolution [30], the phylogenetic relationships of many families and the lower taxonomic levels cannot be fully explained by only the hypothesis of coevolution, e.g. the variety of host origin of *Atadenovirus* and *Siadenovirus* and the location of BoHV-4.

Some traditional methods, e.g. the measures by concatenating aligned sequences, are efficient and powerful to recover the phylogeny of virus with closely evolutionary relationship. However, definition and selection of orthologs may limit their application to distance evolutionary viruses. Furthermore, these methods, in some cases, need adjustment or fine turning.

The CV method could circumvents the ambiguity of choosing orthologs especially for viruses since substitution rate of viruses is high and only a few number of universal common genes could be found (another paper about the stable analysis of the CV method will be submitted subsequently), it may suggest a new angle to Large DNA viruses evolution. Furthermore, the CV method is robust to HGT events. It has been observed that combining many genes could reduce sampling error and converge phylogenies on correct solution with good support [12,52]. Herniou obtained 32 different tree topologies by using 63 individual genes and one tree based on the combined alignment of the 63 genes, while the latter was consistent with most individual gene trees [12]. The CV method could use the information from all coding proteins so that it may still construct stable trees even dashing with a few horizontal transferred genes. We used two sets of data in our previous analysis on bacteria: one is based on whole genomes, and the other is a set of ribosome proteins. Both the results lead to reasonable phylogenetic trees but the first one is better, this shows that these orthologs only appeared in a subset species would also help to stabilize the tree topology. In this way, the method could be a well supplement to the traditional methods. The CV method may provide a quick reference in viruses phylogeneny and a fast analysis of co-evolution of viruses and their hosts whenever their proteomes are available [26].

Methods

All viral genomes were downloaded from NCBI before May. 24th, 2005. There are two available data sets of virus complete genomes. Those in GenBank [53] are the original data submitted by their authors. Those at the National Center for Biotechnological Information (NCBI) [54] are reference genomes curated by NCBI staff. Since the latter represents the approach of one and the same group using the same set of tools, it may provide a more consistent background for comparison. Therefore, we used all the translated amino acid sequences (the .faa files with NC_accession numbers) from NCBI. There are 1489 viral genomes, including 248 phages. Under the assumption that small DNA viruses (genome size < 10 k) probably have a different evolutionary history than large DNA viruses [9,11], and their mutation rate approaches that of RNA viruses (the order of substitutions per site per year, [55]), only large DNA viruses (total length of all coding proteins > 4 k) were used in the phylogenetic analysis, which included 124 viruses (phages have been excluded). Among the 124 dsDNA viruses there are seven viruses that are classified to certain families but their lower taxonomy states remain unknown, and two viruses are tentative species, and one virus that is only recognized as a dsDNA virus. The Additional file 1 lists the dsDNA viruses used, their abbreviations, and the NCBI accession numbers [see Additional file 1: 124 large dsDNA virus names, abbreviations, and NCBI accession numbers].

The main steps of the method are (see [28] for details): First, collect all amino acid sequences of a species. Second, calculate the frequency of appearance of overlapping oligopeptides of length K. A random background needs to be subtracted from these frequencies by using a Markov model of order (K - 2) in order to diminish the influence of random neutral mutations at the molecular level and to highlight the shaping role of selective evolution. Some strings that contribute mostly to apomorphic characters become more significant after the subtraction. The subtraction procedure is an essential step in our method. Third, putting these "normalized" frequencies in a fixed order, a composition vector of dimension 20^K is obtained for each species. Fourth, the correlation $C(A, B)$ between two species A and B is determined by taking projection of one normalized vector on another, i.e., taking the cosine of the angle between them. Lastly, the normalized distance between the two species is defined to be $D = (1 - C)/2$. Once a distance matrix has been calculated it is straightforward to construct phylogenetic trees by following the standard neighbor-joining method in the Phylip package [56].

The best choice of K is related to the uniqueness of sequence reconstruction from its K-word components and is determined basically by the length of the sequence at hand. According to so-called "sequencing by hybridization" [57], for dsDNA viral genomes with length around 4,000 a.a., the minimal K is estimated to be 5.

Only large genome viruses are used in our analysis to avoid the problem of small sample size when using CV method whose subtraction procedure is based on statistics. The CV method avoids the problems caused by HGT on the application of prokaryotic phylogeny by using whole genome sequences, because the extent of lateral transfer has been increasingly restricted to smaller and

smaller gene pools of closer and closer related species as time goes by [58]. However, its application on classification of small DNA viruses may be affected by HGT because of relative shorter genome length, that's one of the reasons only large DNA viruses are used.

Authors' contributions

LG carried out the molecular phylogenetic studies, participated in the design of program, and drafted the manuscript. JQ carried out the design of program and algorithm, participated in molecular phylogenetic studies, and helped to draft the manuscript.

Additional material

Additional File 1

124 large dsDNA virus names, abbreviations, and NCBI accession numbers.

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2148-7-41-S1.doc]

Acknowledgements

The authors thank professor Bailin Hao for discussion and comments. The manuscript has greatly benefited from stimulating comments and suggestions of the three anonymous reviewers. The computation of this project was performed on the HP-SC45 Sigma-X parallel computer of ITP and ICTS, CAS. This paper was partailly based on the PhD thesis work of the two authors.

References

1. van Regenmortel MHV, Fauquet CM, Bishop DHL, Carstens EB, Estes MK, Lemon SM, Maniloff J, Mayo MA, McGeoch DJ, Pringle CR, Wickner RB: *Virus Taxonomy: Seventh Report of the International Committee on Taxonomy of Viruses* Academic Press, San Diego; 2000.
2. Davison AJ, Benko M, Harrach B: **Genetic content and evolution of adenoviruses.** *J Gen Virol* 2003, **84:**2895-2908.
3. Gubser C, Hue S, Kellam P, Smith GL: **Poxvirus genomes: a phylogenetic analysis.** *J Gen Virol* 2004, **85:**105-117.
4. Bulach DM, Kumar CA, Zaia A, Liang B, Tribe DE: **Group II nucleopolyhedrovirus subgroups revealed by phylogenetic analysis of polyhedrin and DNA polymerase gene sequences.** *J Invertebr Pathol* 1999, **73:**59-73.
5. Chen X, Ijkel WFJ, Dominy C, Zanotto P, Hashimoto Y, Faktor O, Hayakawa T, Wang CH, Prekumar A, Mathavan S, Krell PJ, Hu Z, Vlak JM: **Identification, sequence analysis and phylogeny of the lef-2 gene of Helicoverpa armigera single-nucleocapsid baculovirus.** *Virus Res* 1999, **65:**21-32.
6. Koonin EV: **The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses.** *Journal of General Virology* 1991, **72:**2197-2206.
7. Melcher U: **The '30K' superfamily of viral movement proteins.** *Journal of General Virology* 2000, **81:**257-266.
8. Tetart F, Desplats C, Kutateladze M, Monod C, Ackermann HW, Krisch HM: **Phylogeny of the major head and tail genes of the wide-ranging T4-Type bacteriophage.** *J Bacteriol* 2001, **183:**358-366.
9. Tidona CA, Darai G: **Iridovirus homologues of cellular genes: implications for the molecular evolution of large DNA viruses.** *Virus Genes* 2000, **21:**77-81.
10. Tidona CA, Schnitzler P, Kehm R, Darai G: **Is the major capsid protein of Iridoviruses a suitable target for the study of viral evolution?** *Virus genes* 1998, **16:**59-66.
11. Shackelton LA, Holmes EC: **The evolution of large DNA viruses: combining genomic information of viruses and their hosts.** *Trends in Microbiology* 2004, **12:**458-465.
12. Herniou EA, Luque T, Chen X, Vlak JM, Winstanley D, Cory JS, O'Reilly DR: **Use of whole genome sequence data to infer baculovirus phylogeny.** *Journal of Virology* 2001, **75:**8117-8126.
13. Filee J, Forterre P, Laurent J: **The role played by viruses in the evolution of their hosts: a view based on informational protein phylogenies.** *Res Microbiol* 2003, **154:**237-243.
14. Hughes AL: **Origin and evolution of viral interleukin-10 and other DNA virus genes with vertebrate homologues.** *J Mol Evol* 2002, **54:**90-101.
15. Hendrix RW, Lawrence JG, Hatfull GF, Casjens S: **The origins and ongoing evolution of viruses.** *Trends in Microbiology* 2000, **8:**504-508.
16. Bugert JJ, Darai G: **Poxvirus homologues of cellular genes.** *Virus Genes* 2000, **21:**111-133.
17. Bamford DH, Burnett RM, Stuart D: **Evolution of viral structure.** *Theoretical Population Biology* 2002, **61:**461-470.
18. Li W: *Molecular Evolution* Sinauer Associates, Inc; 1997.
19. Sakaoka H, Kurita K, Iida Y, Takada S, Umene K, Kim YT, Ren CS, Nahmias AJ: **Quantitative analysis of genomic polymorphism of herpes simplex virus type 1 strains from six countries: studies of molecular evolution and molecular epidemiology of the virus.** *J Gen Virol* 1994, **75:**513-527.
20. Gao L, Qi J, Wei H, Sun Y, Hao B: **Molecular phylogeny of coronaviruses including human SARS-CoV.** *Chinese Science Bulletin* 2003, **48:**1170-1174.
21. Harrison RL, Bonning BC: **Comparative analysis of the genomes of Rachiplusiaou and Autographa californica multiple nucleopolyhedroviruses.** *J Gen Virol* 2003, **84:**1827-1842.
22. Hyink O, Dellow RA, Olsen MJ, Caradoc-Davies KMB, Drake K, Herniou EA, Cory JS, O'Reilly DR, Ward VK: **Whole genome analysis of the Epiphyas postvittana nucleopolyhedrovirus.** *J Gen Virol* 2002, **83:**957-971.
23. Montague MG, Hutchison CA: **Gene content phylogeny of herpesviruses.** *Proc Natl Acad Sci USA* 2000, **97:**5334-5339.
24. Rohwer F, Edwards R: **The phage proteomic tree: a genome-based taxonomy for phage.** *Journal of Bacteriology* 2002, **184:**4529-4535.
25. Stuart G, Moffett K, Bozarth RF: **A whole genome perspective on the phylogeny of the plant virus family Tombusviridae.** *Archives of Virology* 2004, **149:**1595-1610.
26. Snel B, Huynen MA, Dutilh BE: **Genome trees and the nature of genome evolution.** *Annu Rev Microbiol* 2005, **59:**191-209.
27. Chu K, Qi J, Yu Z, Anh V: **Origin and Phylogeny of Chloroplasts Revealed by a Simple Correlation Analysis of Complete Genomes.** *Mol Biol Evol* 2004, **21:**200-206.
28. Qi J, Wang B, Hao B: **Whole Proteome Prokaryote Phylogeny without Sequence Alignment: A K-String Composition Approach.** *Journal of Molecular Evolution* 2004, **58:**1-11.
29. Farkas SL, Benkö M, Élö P, Ursu K, Dán Á, Ahne W, Harrach B: **Genomic and phylogenetic analyses of an adenovirus isolated from a corn snake (Elaphe guttata) imply common origin with the members of the proposed new genus Atadenovirus.** *J Gen Virol* 2002, **83:**2403-2410.
30. Herniou EA, Olszewski JA, O'Reilly DR, Cory JS: **Ancient coevolution of baculoviruses and their insect hosts.** *Journal of Virology* 2004, **78:**3244-3251.
31. Zanotto PM, Kessing BD, Maruniak JE: **Phylogenetic interrelationships among baculoviruses: evolutionary rates and host associations.** *Journal of Invertebrate Phathology* 1993, **62:**147-164.
32. Rohrmann GF: **Evolution of occluded baculoviruses.** In *The biology of baculoviruses* Edited by: Granados R, Federici B. Boca Raton, Fla: CRC Press, Inc; 1986:203-215.
33. Federici BA: **Baculovirus pathogenesis.** In *The baculoviruses* Edited by: Miller LK. New York: Plenum Press; 1997:33-56.
34. McLysaght A, Baldi PF, Gaut BS: **Extensive gene gain associated with adaptive evolution of poxviruses.** *Proc Natl Acad Sci USA* 2003, **100:**15655-15660.
35. Afonso CL, Delhon G, Tulman ER, Lu Z, Zsak A, Becerra VM, Zsak L, Kutish GF, Roch DL: **Genome of Deerpox Virus.** *J Virol* 2005, **79:**966-977.
36. Albà MM, Das R, Orengo CA, Kellam P: **Genomewide function conservation and phylogeny in the Herpesviridae.** *Genome Res* 2001, **11:**43-54.

37. Bahr U, Darai G: **Analysis and characterization of the complete genome of tupaia (tree shrew) herpesvirus.** *J Virol* 2001, **75:**4854-4870.
38. McGeoch DJ, Davison AJ: **The descent of human herpesvirus 8.** *Seminars in Cancer Biology* 1999, **9:**201-209.
39. McGeoch DJ, Cook S: **Molecular phylogeny of the Alphaherpesvirinae subfamily and a proposed evolutionary timescale.** *J Mol Biol* 1994, **238:**9-22.
40. Iyer LM, Aravind L, Koonin EV: **Common origin of four diverse families of large eukaryotic DNA viruses.** *J Virol* 2001, **75:**11720-11734.
41. Do JW, Moon CH, Kim HJ, Ko MS, Kim SB, Son JH, Kim JS, An EJ, Kim MK, Lee SK, Han MS, Cha SJ, Park MS, Park MA, Kim YC, Kim JW, Park JW: **Complete genomic DNA sequence of rock bream iridovirus.** *Virology* 2004, **325:**351-363.
42. Yang F, He J, Lin X, Pan D, Zhang X, Xu X: **Complete genome sequence of the shrimp white spot bacilliform virus.** *J Virol* 2001, **75:**11811-11820.
43. Thureen DR, Keeler CL Jr: **Psittacid Herpesvirus 1 and Infectious Laryngotracheitis Virus: Comparative Genome Sequence Analysis of Two Avian Alphaherpesviruses.** *J Virol* 2006, **80:**7863-7872.
44. McGeoch DJ, Gatherer D, Dolan A: **On phylogenetic relationships among major lineages of the Gammaherpesvirinae.** *J Gen Virol* 2005, **86:**307-316.
45. Kovács GM, Davison AJ, Zakhartchouk AN, Harrach B: **Analysis of the first complete genome sequence of an Old World monkey adenovirus reveals a lineage distinct from the six human adenovirus species.** *J Gen Virol* 2004, **85:**2799-2807.
46. Kovács GM, Harrach B, Zakhartchouk AN, Davison AJ: **Complete genome sequence of simian adenovirus 1: an Old World monkey adenovirus with two fiber genes.** *J Gen Virol* 2005, **86:**1681-1686.
47. Lefkowitz EJ, Wang C, Upton C: **Poxviruses: past, present and future.** *Virus Research* 2006, **117:**105-118.
48. Cheng CH, Liu SM, Chow TY, Hsiao YY, Wang DP, Huang JJ, Chen HH: **Analysis of the complete genome sequence of the Hz-1 virus suggests that it is related to members of the Baculoviridae.** *J Virol* 2002, **76:**9024-9034.
49. Davison AJ, Trus BL, Cheng N, Steven AC, Watson MS, Cunningham C, Le Deuff RM, Renault T: **A novel class of herpesvirus with bivalve hosts.** *J Gen Virol* 2005, **86:**41-53.
50. McGeoch DJ, Rixon FJ, Davison AJ: **Topics in herpesvirus genomics and evolution.** *Virus Research* 2006, **117:**90-104.
51. Moser BA, Becnel JJ, White SE, Afonso C, Kutish G, Shanker S, Almira E: **Morphological and molecular evidence that Culex nigripalpus baculovirus is an unusual member of the family Baculoviridae.** *J Gen Virol* 2001, **82:**283-297.
52. Mitchell A, Mitter C, Regier JC: **More taxa or more characters revisited: combining data from nuclear protein-encoding genes for phylogenetic analyses of Noctuoidea (Insecta: Lepidoptera).** *Syst Biol* 2000, **49:**202-224.
53. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL: **GenBank.** *Nucleic Acids Res* 2006, **34:**D16-D20.
54. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Edgar R, Federhen S, Geer LY, Helmberg W, Kapustin Y, Kenton DL, Khovayko O, Lipman DJ, Madden TL, Maglott DR, Ostell J, Pruitt KD, Schuler GD, Schriml LM, Sequeira E, Sherry ST, Sirotkin K, Souvorov A, Starchenko G, Suzek TO, Tatusov R, Tatusova TA, Wagner L, Yaschenko E: **Database resources of the National Center for Biotechnology Information.** *Nucleic Acids Res* 2006, **34:**D173-D180.
55. Truyen U, Gruenberg A, Chang S, Obermaier B, Veijalainen P, Parrish C: **Evolution of the feline-subgroup parvoviruses and the control of canine host-range in vivo.** *J Virol* 1995, **69:**4702-4710.
56. Felsenstein J: **PHYLIP (phylogeny inference package) version 3.5c.** [http://evolution.genetics.washington.edu/phylip.html].
57. Pevzner PA: *Computational Molecular Biology: An Algorithmic Approach Volume 75*. MIT Press; 2000.
58. Woese CR: **The universal ancester.** *Proc Natl Acad Sci USA* 1998, **95:**6854-6859.