

Use of Whole Genome Sequence Data To Infer Baculovirus Phylogeny

ELISABETH A. HERNIOU,^{1,2} TERESA LUQUE,¹ XINWEN CHEN,³ JUST M. VLAK,³
DOREEN WINSTANLEY,⁴ JENNIFER S. CORY,² AND DAVID R. O'REILLY^{1*}

Department of Biology, Imperial College, London SW7 2AZ,¹ Ecology and Biocontrol Group, Centre for Ecology and Hydrology–Oxford, Oxford OX1 3SR,² and Horticulture Research International, Wellesbourne,⁴ United Kingdom, and Laboratory of Virology, Wageningen University, Wageningen, The Netherlands³

Received 21 March 2001/Accepted 18 May 2001

Several phylogenetic methods based on whole genome sequence data were evaluated using data from nine complete baculovirus genomes. The utility of three independent character sets was assessed. The first data set comprised the sequences of the 63 genes common to these viruses. The second set of characters was based on gene order, and phylogenies were inferred using both breakpoint distance analysis and a novel method developed here, termed neighbor pair analysis. The third set recorded gene content by scoring gene presence or absence in each genome. All three data sets yielded phylogenies supporting the separation of the *Nucleopolyhedrovirus* (NPV) and *Granulovirus* (GV) genera, the division of the NPVs into groups I and II, and species relationships within group I NPVs. Generation of phylogenies based on the combined sequences of all 63 shared genes proved to be the most effective approach to resolving the relationships among the group II NPVs and the GVs. The history of gene acquisitions and losses that have accompanied baculovirus diversification was visualized by mapping the gene content data onto the phylogenetic tree. This analysis highlighted the fluid nature of baculovirus genomes, with evidence of frequent genome rearrangements and multiple gene content changes during their evolution. Of more than 416 genes identified in the genomes analyzed, only 63 are present in all nine genomes, and 200 genes are found only in a single genome. Despite this fluidity, the whole genome-based methods we describe are sufficiently powerful to recover the underlying phylogeny of the viruses.

Members of the *Baculoviridae* are circular double-stranded DNA viruses with a genome size ranging from 90 to 180 kb (26). They are pathogenic for arthropods, with most having been isolated from *Lepidoptera* species. Traditionally, baculovirus classification has been based on the morphology of the occlusion bodies they form in infected cells (5). Viruses in the genus *Nucleopolyhedrovirus* (NPV) form polyhedral occlusion bodies, each containing many virions (45), whereas viruses in the genus *Granulovirus* (GV) form ovoid occlusion bodies usually containing a single virion (57). The lepidopteran NPVs have been subdivided into groups I and II based on molecular phylogenies (6, 59).

Most analysis of baculovirus phylogeny has been based on the *polyhedrin/granulin* gene, which encodes the major occlusion body protein (3, 59), but other genes have been used recently (6, 7, 9, 10, 31). Comparison of these analyses reveals that conflicts are often observed between phylogenies based on different genes. In particular, polyhedrin phylogenies often disagree with other gene phylogenies (10, 31). These conflicts could be due to erroneous phylogenetic inferences caused by unequal rates of evolution or to lack of an unambiguous phylogenetic signal in the sequences. Alternatively, they could reflect real differences in the phylogeny of individual genes due to recombination. Accumulating evidence of frequent horizontal transfers in some prokaryotic lineages has led researchers to question whether phylogenetic trees are the most appropriate way to represent the evolutionary history of such organisms (13). Horizontal transfer is a particular issue for some viruses

in which recombination is a known evolutionary driver (27, 41). Exchange of genetic material is known to occur between coinfecting baculoviruses or between baculoviruses and their hosts (12, 18, 33, 56). There is also evidence of gene exchange between baculoviruses and other infectious agents of their hosts (24, 38, 42, 46). However, the extent to which such gene exchanges have shaped baculovirus evolution is unclear. A key question is whether it is possible and appropriate to construct a single phylogenetic tree representing their evolutionary history or whether such a “backbone” tree is obscured by frequent horizontal transfers.

The availability of complete genome sequence data for several organisms has led to an interest in the use of such data for phylogenetic reconstruction. Complete genome sequences contain phylogenetic information at several levels (34). In addition to the nucleotide sequence and amino acid sequences of the encoded proteins, the gene content and the order of genes on a genome may be phylogenetically informative (47, 50). Gene content or gene order data sets are independent of the sequences of individual genes and should complement phylogenies based on nucleotide or amino acid sequences. Complete genome approaches have recently been employed to infer the phylogeny of the herpesviruses (22, 40).

Several baculovirus genome sequences have now been published (1, 2, 8, 20, 23, 25, 30, 35), permitting the use of whole genome approaches to infer their phylogeny. Baculovirus gene arrangements have previously been compared using gene parity plot analysis (8, 23, 28, 30). These studies confirmed that gene order comparisons between baculoviruses could be phylogenetically informative; more closely related genomes clearly had a more similar gene order. However, parity plot analysis does not give quantitative information on relatedness, making

* Corresponding author. Mailing address: Department of Biology, Imperial College of Science, Technology, and Medicine, Imperial College Rd., London SW7 2AZ, United Kingdom. Phone: 44 (0)20 75945376. Fax: 44 (0)20 75842056. E-mail: d.o'reilly@ic.ac.uk.

TABLE 1. Baculovirus genomes

Baculovirus genome	Classification	Size (bp)	No. of genes ^a	Accession no.	Reference
AcMNPV	NPV Group I	133,894	155 (1)	L22858	2
BmNPV	NPV Group I	128,413	140 (5)	L33180	20
OpMNPV	NPV Group I	131,990	153 (1)	U75930	1
LdMNPV	NPV Group II	161,046	164 (1)	AF081810	35
SeMNPV	NPV Group II	135,611	140 (1)	AF169823	30
HaSNPV	NPV Group II	131,403	135	AF271059	8
XcGV	GV	178,733	181	AF162221	25
PxGV	GV	100,999	120	AF270937	23
CpGV	GV	123,500	143	U53466	Luque et al. (submitted)

^a BmNPV, *Bombyx mori* NPV; OpMNPV, *Orgyia pseudotsugata* multicapsid NPV; PxGV, *Plutella xylostella* GV.

^b The numbers in parantheses indicate genes not recognized when the original sequence paper was published.

it difficult to use this method to build trees. Here we present a comprehensive analysis of the relationships between nine lepidopteran baculoviruses whose genomes have been completely sequenced, comprising three group I NPVs, three group II NPVs, and three GVs. Phylogenies were generated based on three independent character sets: the individual sequences of genes shared by all nine viruses, gene order, and gene content. The utility of these data sets for the reconstruction of baculovirus phylogenies was assessed. Methods based on both gene content and gene order successfully resolved the three major groups and further resolved the species of the group I NPVs. However, the genomic data available to date are not strong enough to allow the generation of well-supported phylogenies that resolve the species among the group II NPVs and the GVs. The relationships between the viruses in these groups were only resolved with strong support in a phylogeny based on the combined sequences of all 63 genes shared between these nine viruses.

MATERIALS AND METHODS

Baculovirus sequences. The genomes used are listed in Table 1. The gene identity and gene order data for each genome were taken from the sequence annotations in the literature.

Phylogenetic inference based on gene sequences. The nine baculoviruses included in this study share 63 genes (Table 2). For each gene, amino acid sequences were aligned with ClustalW (54) using default parameters and the Blossum matrix. The alignments were checked and refined by eye using MacClade 4 (37) prior to being compiled in a single file of 25,788 characters, of which 15,907 were parsimoniously informative. Each gene represents a defined subset of this file. Gaps were treated as missing data. Maximum parsimony analyses were performed in PAUP* (phylogenetic analysis using parsimony [*and other meth-

ods]) (51). Phylogenies, either of the entire data set or of individual subsets, were estimated by exhaustive searches using a PAM-weighted amino acid step matrix (53). Branch support was evaluated by bootstrap analysis. For each gene, the most parsimonious tree was retained to calculate a majority rule consensus tree. Topologies of the most parsimonious trees were compared against each subset using the Shimodaira-Hasegawa (SH) test (49) implemented in the software package PAML (58).

All data sets and trees are deposited in TreeBase under the accession numbers S625, M964, and M965 (<http://www.herbaria.harvard.edu/treebase>).

Phylogenetic inference based on gene order. Phylogenetic analysis based on gene order was carried out in two ways. The first was a modification of the breakpoint distance analysis method of Blanchette et al. (4), originally described for the analysis of mitochondrial gene order. A breakpoint between two genomes is where two genes that are adjacent in one genome are separated in the other. The method makes no assumptions about the mechanisms involved in genome rearrangements. The number of breakpoints was counted between a pair of genomes. This was then divided by the number of genes in common between those genomes to yield a relative breakpoint distance. This modification was implemented to compensate for bias in the calculated distances due to differences in genome length. Without correction, comparisons between small genomes would give shorter distances simply because they have fewer genes. The *bro* gene family was omitted from this analysis because of the difficulty of establishing orthology between *bro* genes of different genomes. Calculation of the relative breakpoint distances from pairwise comparisons of all nine genomes resulted in a distance matrix which was then used for phylogenetic reconstruction with the Neighbor program from PHYLIP (16). The resulting phylogenetic tree was visualized in TreeView (43).

We have also developed a new approach to inferring phylogeny from gene order data, which we term neighbor pair analysis. Only the 63 shared genes were considered in this approach. A matrix recording the presence or absence of each possible neighboring gene pair in each genome was compiled. Neighbor gene pairs resulting in constant characters (present in all genomes or absent from all genomes) were not taken into account. This resulted in a data matrix containing 103 characters, of which 73 were parsimoniously informative. Similar to breakpoint analysis, neighbor pair analysis is independent of the mechanism of gene rearrangement. It has the advantage that it allows the binary encoding of conservation of gene order, which can then be analyzed by maximum parsimony. Branch support was evaluated by bootstrap analysis, and alternative topologies were assessed using the Kishino/Hasegawa test (KH test) (32) implemented in PAUP.

Phylogenetic inference based on gene content. A matrix was generated recording the presence or absence of each baculovirus gene in each genome. The *bro* gene family was omitted as before. A total of 409 distinct genes were recorded in this matrix. Of these, 145 were parsimoniously informative, i.e., present in more than one genome but not in all. Phylogenetic analyses were performed using maximum parsimony in PAUP. Branch support was assessed by bootstrap analysis, and alternative topologies were evaluated using the KH test. Character changes (i.e., gene acquisition or loss) were mapped onto the trees using MacClade 4.

RESULTS

Gene sequence phylogenies. Comparison of the baculovirus genomes included in this study revealed that 63 genes are common to these nine genomes (Table 2). This number is lower than that previously reported by Chen et al. (8), because

TABLE 2. Functions of the 63 shared baculovirus genes

Gene function or type ^b			
Replication or expression	Structure	Auxiliary ^a	Unknown
39K (<i>ac36</i>), <i>dbp1</i> (<i>ac25</i>), <i>dnapol</i> (<i>ac65</i>), <i>helicase</i> (<i>ac95</i>), <i>lef1</i> (<i>ac14</i>), <i>lef2</i> (<i>ac6</i>), <i>lef3</i> (<i>ac67</i>), <i>lef4</i> (<i>ac90</i>), <i>lef5</i> (<i>ac99</i>), <i>lef6</i> (<i>ac28</i>), <i>lef8</i> (<i>ac50</i>), <i>lef9</i> (<i>ac62</i>), <i>lef11</i> (<i>ac37</i>), <i>ie1</i> (<i>ac147</i>), <i>me53</i> (<i>ac139</i>), <i>p47</i> (<i>ac40</i>), <i>vlf1</i> (<i>ac77</i>)	<i>ac23</i> , <i>fp25K</i> (<i>ac61</i>), <i>gp41</i> (<i>ac80</i>), <i>odv-e18</i> (<i>ac143</i>), <i>odv-e25</i> (<i>ac94</i>), <i>odv-ec27</i> (<i>ac144</i>), <i>odv-e56</i> (<i>ac148</i>), <i>adv-e66</i> (<i>ac46</i>), <i>p6.9</i> (<i>ac100</i>), <i>p74</i> (<i>ac138</i>), <i>p95</i> (<i>ac83</i>), <i>pk1</i> (<i>ac10</i>), <i>polh</i> (<i>ac8</i>), <i>vp39</i> (<i>ac89</i>), <i>vp1054</i> (<i>ac54</i>)	<i>alk-exo</i> (<i>ac133</i>), <i>fgf</i> (<i>ac32</i>), <i>sod</i> (<i>ac31</i>), <i>ubiquitin</i> (<i>ac35</i>)	38K (<i>ac98</i>), 38.7K (<i>ac13</i>), <i>ac22</i> , <i>ac29</i> , <i>ac38</i> , <i>ac53</i> , <i>ac66</i> , <i>ac68</i> , <i>ac75</i> , <i>ac76</i> , <i>ac78</i> , <i>ac81</i> , <i>ac82</i> , <i>ac92</i> , <i>ac93</i> , <i>ac96</i> , <i>ac106</i> , <i>ac109</i> , <i>ac110</i> , <i>ac115</i> , <i>ac119</i> , <i>ac142</i> , <i>ac145</i> , <i>ac146</i> , <i>p40</i> (<i>ac101</i>), <i>p12</i> (<i>ac102</i>), <i>p45</i> (<i>ac103</i>)

^a Auxiliary genes are genes that are not directly involved in viral gene expression or genome replication or in the formation of progeny virus particles.

^b Where a common gene name is given, the open reading frame number in AcMNPV is also given in parentheses.

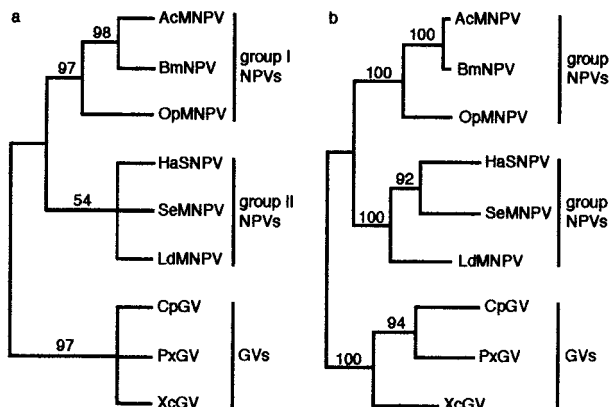


FIG. 1. Gene sequence phylogenies. (a) Majority rule consensus tree of the most parsimonious trees obtained for each of the 63 genes shared by all nine baculoviruses. The numbers indicate the percentages of individual gene trees supporting each branch. (b) Most parsimonious tree based on the combined sequences of the 63 shared genes. Numbers indicate the percentages of bootstrap support from 1,000 replicates. Trees are rooted using the GVs as a sister group to the NPVs.

ie0 (*ac141*) and *p10* (*ac137*) are not present in the *Cydia pomonella* GV (CpGV) genome (T. Luque, R. Finch, N. Crook, D. R. O'Reilly, and D. Winstanley, submitted for publication). It is likely to decrease as more baculovirus genomes become available. Phylogenetic trees were generated for each of these 63 shared genes, resulting in 32 different tree topologies (see Fig. A1). Most of the topological variation was in the arrangement of the GVs and in the monophyly and arrangement of the group II NPVs. The majority rule consensus tree of the most parsimonious tree for each gene (Fig. 1a) shows that most gene phylogenies support the NPV-GV division and the subdivision of the NPVs into two groups.

The alignments of the 63 conserved genes were also combined, and phylogenies were reconstructed based on this combined alignment. This analysis yielded a single most parsimonious tree with high bootstrap support (Fig. 1b). Seven individual gene phylogenies (*ac22*, *ac81*, *ac119*, *ac142*, *ac145*, *lef8*, and *lef9*) had this topology. Furthermore, SH tests showed that most individual gene phylogenies are compatible with this topology (see Table A1), the only exception being *adv-e66*.

Gene order phylogeny. Two approaches were used to provide a measure of the difference in synteny (i.e., gene order) between baculovirus genomes. First, a matrix of relative breakpoint distances was compiled based on pairwise comparisons of all the genomes (the distance matrix is available at <http://www.bio.ic.ac.uk/staff/dor/oreilly.htm>). The distance tree generated from this matrix (Fig. 2a) differs from the combined gene tree (Fig. 1b) in the relationships among the group II NPVs but is consistent with the majority rule consensus tree shown in Fig. 1a. Second, a binary matrix recording the presence of conserved neighboring gene pairs in each genome was compiled (available at <http://www.bio.ic.ac.uk/staff/dor/oreilly.htm>). This matrix was analyzed by maximum parsimony. The most parsimonious tree (Fig. 2b) has a different topology again, differing from the relative breakpoint distance tree (Fig. 2a) in the relationships within the group II NPVs and the GVs but differing from the combined gene tree (Fig. 1b) only in the relationships within the GVs. Furthermore, KH tests of this neigh-

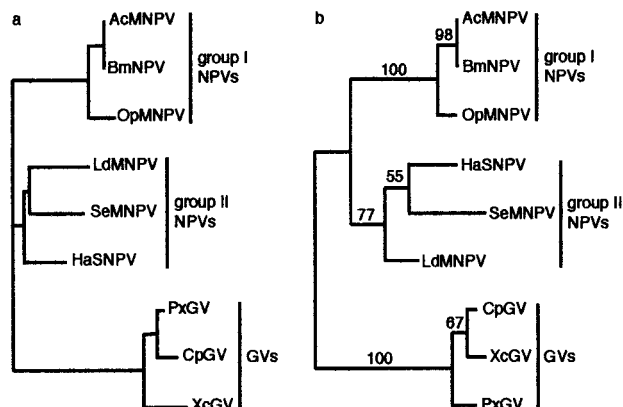


FIG. 2. Gene order phylogenies. (a) Neighbor-joining tree based on relative breakpoint distances. (b) Most parsimonious tree based on the neighboring gene pair analysis. Numbers indicate the percentages of bootstrap support from 1,000 replicates. Trees are rooted using the GVs as a sister group to the NPVs.

bor pair data set demonstrated that it is compatible with a total of 26 single gene tree topologies, including both tree topologies shown in Fig. 1b and 2a (see Table A1).

Gene content phylogeny. A matrix recording the presence or absence of all baculovirus genes in each genome was compiled (available at <http://www.bio.ic.ac.uk/staff/dor/oreilly.htm>). Maximum parsimony analysis of this data set gave a single most parsimonious tree (Fig. 3). Again, this tree separates the NPVs and GVs and resolves the NPVs into two subgroups. It differs from previous trees in the relationships among the group II NPVs and the GVs. The tree is consistent with the majority-rule consensus tree (Fig. 1a). KH tests demonstrated that it is also compatible with 24 of the single gene trees, including all tree topologies shown in Fig. 1 and 2 (see Table A1).

DISCUSSION

Comparative genomics will become an increasingly powerful tool for inferring biological function as more genome sequences become available. However, to exploit this approach

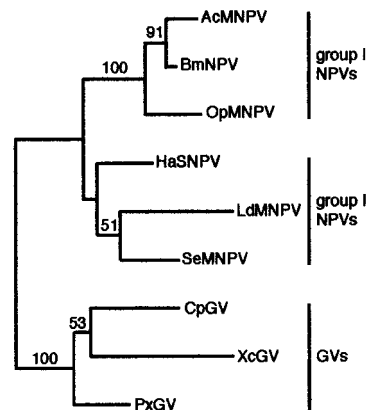


FIG. 3. Gene content phylogeny. Most parsimonious tree based on the gene content data set. Percentages of bootstrap support (1,000 replicates) greater than 50% are shown. The tree is rooted using the GVs as a sister group to the NPVs.

fully it will be critical to develop methods that place the data in an appropriate evolutionary context. Baculoviruses provide a case in point. Several complete sequences have been published, and it is likely that many additional sequences will be available soon (1, 2, 8, 20, 23, 25, 30, 35). It will be essential to establish relationships among baculoviruses reliably in order to effectively interpret the wealth of information about the biology and evolutionary history of these viruses contained within these data. The rapidly increasing availability of complete genome sequences has prompted an interest in using information other than nucleotide or amino acid sequence data for the generation of molecular phylogenies. Gene content has already been used in phylogenetic analyses of herpesviruses, prokaryotes, and eukaryotes (17, 40, 50, 52). Gene order has also been used to reconstruct phylogenies of herpesviruses, animal mitochondrial genomes, and bacteria. However, its use can be hindered by a lack of synteny conservation or a lack of synteny variation (4, 22, 36, 55). Here we evaluated methods based on gene order, gene content, and conserved gene sequences for the analysis of the relationships between nine lepidopteran baculoviruses. This represents the most comprehensive analysis to date of baculovirus phylogeny.

All the approaches used agreed on the separation of the NPVs and GVs and the division of the NPVs into groups I and II, as postulated by Zanotto et al. (59) and Bulach et al. (6). They all also resolved the relationships between the group I NPVs. Relationships between viruses in the other groups were only clearly resolved by the combined gene sequence analysis (Fig. 1b). Several lines of evidence support this tree as the most plausible representation of the relationships between these viruses. First, it is very strongly supported by bootstrap analysis (>90% support for all nodes). Second, it is based on a very large data set. It has been observed previously that with a consistent method, combining genes reduces sampling error and causes the phylogenies to converge toward the correct solution with good support (39). We believe this effect is observed here. Third, although the gene order and gene content-based analyses yielded different optimal topologies, the combined gene topology was always present among suboptimal trees that were compatible with the data. Furthermore, partition homogeneity testing demonstrated that the phylogenetic signal yielded by these approaches is congruent with that of the gene sequences ($P = 0.01$). Finally, this tree was also found to be the best tree when SH tests were performed for the whole data set under maximum likelihood criteria in PAML.

The fact that methods based on gene order and gene content could resolve the viruses into the three major groups demonstrates that such approaches do permit phylogenetic reconstruction. However, the data available from the baculoviruses that have been sequenced to date are relatively weak and prone to homoplasy. (Homoplasy is defined as similarity due to independent evolutionary change. This can either be due to convergent evolution [e.g., two genomes appear similar because they have independently acquired the same gene] or reversal to an ancestral state [e.g., two genomes appear similar because a gene was acquired and subsequently lost during the evolution of one but was never present in the lineage of the other].) The weakness of the data was reflected by the large number of suboptimal trees compatible with the data sets, as shown by KH tests. Gene order and gene content have the

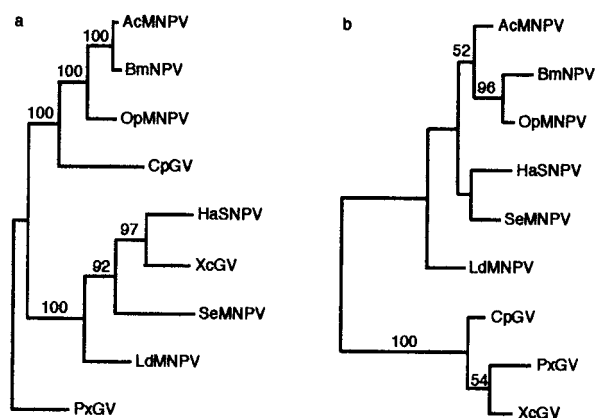


FIG. 4. *odv-e66* (a) and *polyhedrin* (b) gene phylogenies. The single most parsimonious tree is shown in each case. Percentages of bootstrap support (1,000 replicates) greater than 50% are shown. Trees are rooted using the GVs as a sister group to the NPVs.

advantage of providing independent data sets from the gene sequences, with independent dynamics and rates of evolution. This is particularly true for gene content analysis, as the parsimoniously informative genes used (present in more than one genome but not present in all) do not overlap with the genes used for sequence-based analyses (present in all genomes). We anticipate that the future addition of more species from the group II NPVs and the GVs will improve species sampling and reduce homoplastic noise and thus provide better phylogenetic resolution using whole genome-based approaches. The approaches we have developed here will also prove valuable for the phylogenetic analysis of other organisms, including other large DNA viruses.

It is worth noting that, although the gene sequence data contain the strongest phylogenetic signal, only a few individual genes actually gave the best tree (*ac22*, *ac81*, *ac119*, *ac142*, *ac145*, *lef8*, and *lef9*). This underlines the danger of using phylogenies based on one gene or a small number of genes to infer the evolution of whole genomes or species. Thus, we recommend that reconstruction of baculovirus phylogenies should ideally be based on a combined analysis of all genes conserved among all baculoviruses. Whole genome approaches based on gene content and gene order should be used to complement this analysis, as it is clear that both are phylogenetically informative and can provide additional support for the combined gene tree. As more genomes are sequenced, they will become increasingly powerful tools.

Most of the topological variation between the data sets resided within the GVs and group II NPVs. A number of factors contribute to this. First, each group contains one genome (*Xestia c-nigrum* GV [XcGV] and *Lymantria dispar* multicapsid NPV [LdMNPV]) that is markedly larger than the others, creating an imbalance in the character distribution for gene content and gene order, which results in a long branch attraction effect (15). This is most noticeable for the gene content phylogeny (Fig. 3), where smaller genomes are attracted to each other at the base of their respective groups. Second, species within the groups are either too similar or too different to provide appropriate characters to resolve their relationships. Gene order data are not very informative for the GVs

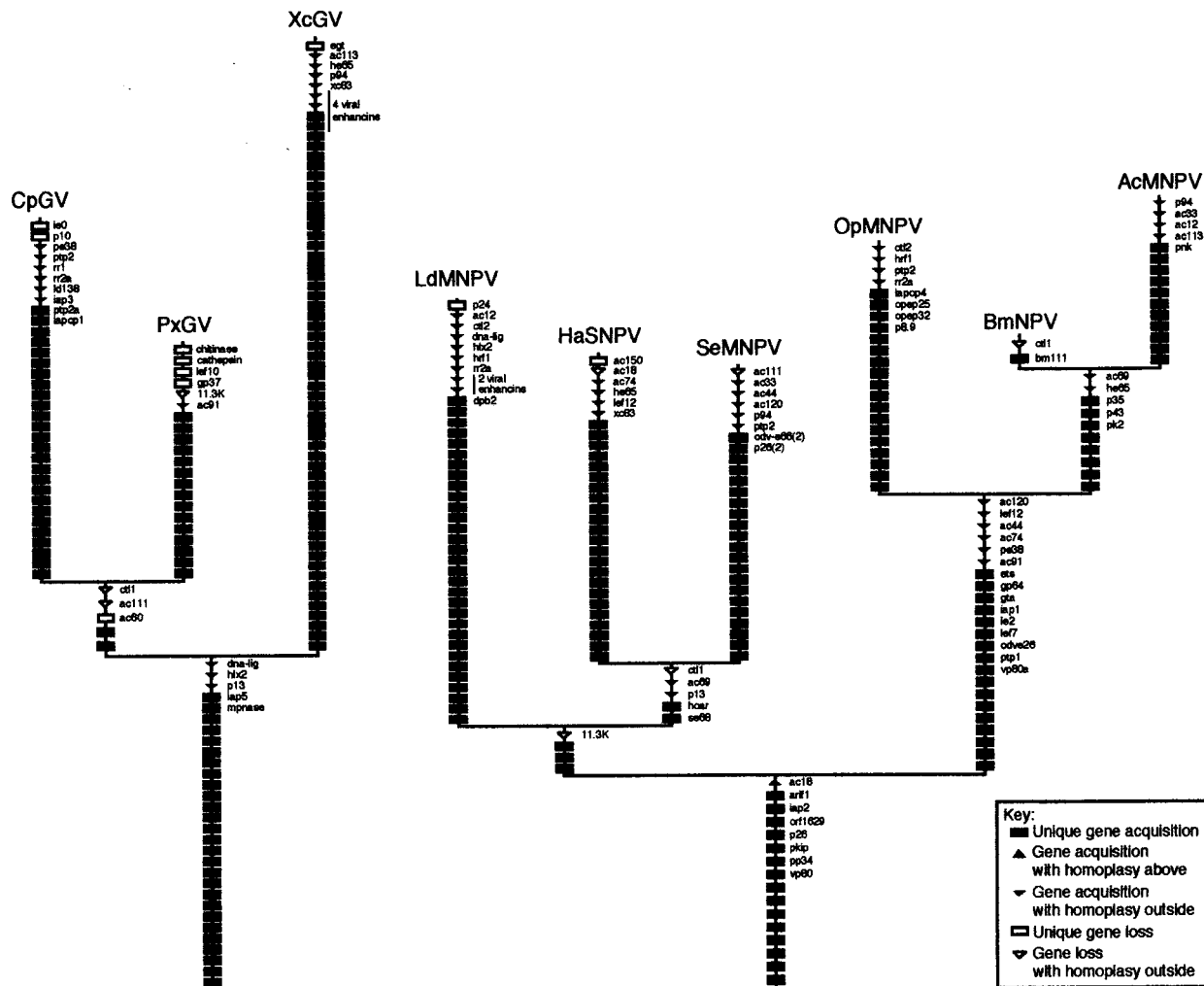


FIG. 5. Gene content data mapped onto the most parsimonious tree based on the combined sequences of the 63 common genes. Shown are gene content changes predicted to have taken place during baculovirus evolution. Gene acquisitions and losses are represented by solid and open symbols, respectively. Where the state of a gene is predicted to have changed only once, a rectangle is used, whereas triangles are used to denote genes whose state has changed multiple times. An upward-pointing triangle is used to illustrate where the additional change of state occurs further up in the same lineage. For example, *ac18* is acquired at the base of the NPV lineage but is subsequently lost from the HaSNPV lineage. Downward-pointing triangles illustrate where the character state of a gene changes independently in different parts of the lineage. For example, DNA ligase, helicase 2, and p13 are represented by downward-pointing triangles at the base of the GV lineage because all three genes are also independently acquired by some NPVs. Only gene content changes that can be unambiguously assigned to a particular branch are shown, with the exception of gene content changes leading to the separation of NPVs and GVs (see the text for details). The tree is rooted using the GVs as a sister group to the NPVs.

because of the almost identical order of the 63 conserved genes among these viruses (Fig. 2b). Conversely, relationships within the group II NPVs are obscured by their extensive differences in genome arrangements.

An additional pattern emerging from these data is that the monophyly of the group II NPVs is far less well supported than that for the group I NPVs or the GVs. This could indicate a sampling artifact whereby the species representing the other two groups are much more closely related than the group II species. Alternatively, it might indicate that the group II NPVs are an older group than the other two. Similarly, our understanding of baculovirus evolution might change when nonlepidopteran NPVs become available for phylogenetic analysis.

The *odv-e66* gene yielded a tree that was incompatible with

all individual trees and genome trees (Table A1). The phylogeny of *odv-e66* (Fig. 4a) agrees only with that of the consensus tree (Fig. 1a) in the arrangement of the group I NPVs. The rest of the tree suggests a complex history, possibly including several duplications, horizontal transfers, and gene losses. The presence of a second copy of *odv-e66* in *Spodoptera exigua* multicapsid NPV (SeMNPV) provides independent evidence for duplication (30). This gene codes for a structural protein present in the envelopes of occluded virions. Understanding its complex evolutionary history might provide clues to its precise role in the virus life cycle. Of the 63 common genes, the only other gene whose phylogenetic tree disagrees (with strong bootstrap support) with the consensus tree (Fig. 1a) is the *polyhedrin* gene. The *polh*-based phylogeny consistently and strong-

ly places *Autographa californica* multicapsid NPV (AcMNPV) at the base of the group I NPVs (Fig. 4b), suggesting a horizontal transfer of the *polh* gene in the AcMNPV lineage, as previously noted (10). The otherwise low bootstrap scores for this tree reflect the weak phylogenetic signal in *polh* amino acid sequence alignments. Great caution should therefore be taken when interpreting phylogenies based solely on this gene.

A highly informative way to visualize the gene content data is to map them onto the optimal phylogenetic tree, revealing where gene content changes are likely to have occurred during the evolution of these viruses. Figure 5 presents all the gene content changes that can be unambiguously assigned to a particular branch on the basis of the existing data. The exceptions to this are the genes prior to the GV-NPV division. Because the gene content of the most recent common ancestor of NPVs and GVs is not known, we cannot say whether a given gene has been lost by one group or acquired by the other. For presentation purposes only, all of these genes have been coded as acquisitions, as this indicates more clearly the genes unique to each group of viruses.

The tree in Fig. 5 gives a unique view of the gene content changes that define the different baculovirus groups. For example, at the base of the tree it can be seen that 43 genes distinguish these NPVs from these GVs. Sixteen of these are unique to NPVs (although the *ac18* homologue was subsequently lost in the *Helicoverpa armigera* single-nucleocapsid NPV [HaSNPV] lineage) and 27 are unique to GVs. Potential functions can be ascribed to six of the NPV-specific genes but to only two of the GV-specific genes. Three NPV-specific genes (*vp80*, *pp34*, and *orf1629*) code for structural proteins (19). This may be associated with the structural differences between NPVs and GVs. ARIF 1 is implicated in rearrangement of the cytoskeleton during NPV infection (48). This may be relevant to the differences in subcellular architecture during NPV and GV infections (57). The relationship of the other genes to differences between NPVs and GVs is uncertain. The functions of p26 and PKIP are not clear (14, 29). The *iap* genes are implicated in the inhibition of apoptosis (11). It is intriguing that individual members of this gene family appear to be unique to both GVs (*iap5*) and NPVs (*iap2*). The only other GV unique gene with a potential function encodes a metallo-proteinase which is thought to contribute to the proteolysis of infected tissue (21).

Twenty genes distinguish the group I and group II NPVs. Pearson et al. (44) have previously noted that *gp64* is unique to the group I NPVs and suggested that acquisition of this gene promoted the diversification of these viruses. Morse et al. (42) have noted that *gp64* is related to a Thogoto virus (a tick-borne orthomyxo-like virus) glycoprotein, further supporting the idea that acquisition of *gp64* may have promoted baculovirus diversification. Our analysis shows that *gp64* is only 1 of 17 genes unique to the group I NPVs. Intriguingly, four of these genes, including *gp64* (*gp64*, *odve26*, *ptp1*, and *vp80a*), code for structural proteins. It is tempting to speculate that acquisition of novel structural proteins may contribute to baculovirus speciation by causing alterations in host range. Of the other group I NPV-specific genes, two (*ie2* and *lef7*) are implicated in the regulation of viral gene expression and one is another *iap* (*iap1*). For all of these genes it is possible to postulate an association with virus host range. The functions of the remain-

ing group I NPV-specific genes are not known. Only three genes, whose functions are also unknown, appear to be unique to the group II NPVs based on present data.

A striking feature of the tree in Fig. 5 is the number of homoplastic changes predicted. Of particular note is the number of genes that appear to have been acquired independently in different parts of the lineage (indicated by downward solid triangles in the figure). The analysis predicts that 25 genes have been acquired independently at least twice, and 4 genes (*he65*, *p94*, *ptp2*, and *rr2a*) appear to have been acquired three times. Further study will be required to determine whether these represent independent acquisitions from the host or other genome or horizontal transfers between baculoviruses. It is important to bear in mind that these predictions should be interpreted with caution. They represent the most parsimonious interpretation of the presently available data, but it is possible that, as further data become available, the mapping of the tree may change. Nonetheless, the picture that emerges is one of baculoviruses continuously sampling their genomic environment (either the host genome or the genomes of coinfecting agents) for beneficial genes during the course of their evolution.

There is abundant other evidence in the data analyzed here of the fluid nature of baculovirus genomes. Of more than 416 genes identified in these nine genomes, only 63 are present in all genomes, and 200 are present in only one genome (although it is conceivable that some of these might represent highly diverged homologues not recognized by present comparison methods). Similarly, analysis of the gene order data points to frequent gene rearrangements in the course of baculovirus evolution. For example, the patristic distance between AcMNPV and CpGV is 61, implying a minimum of 61 rearrangements between the 63 conserved genes since their last common ancestor. As noted above, comparison of individual gene phylogenies to the whole genome phylogeny provides further support for horizontal transfer of genes between genomes. Despite this fluidity, we show that it is possible to recover a single, well-supported tree that describes the evolution of these viruses. The challenge now will be to relate biological differences to the evolutionary groups that have been highlighted so that we can begin to understand what features of baculovirus biology and ecology have driven the diversification of this group of viruses.

ACKNOWLEDGMENTS

We thank M. Tristem, A. Burt, C. Lopez Vaamonde, J. Olszewski (Imperial College), D. T. J. Littlewood, and M. Wilkinson (Natural History Museum, London) for critical reading of the manuscript and Z. Yang for advice on the utilization of PAML.

This research was supported by Natural Environment Research Council CASE studentship award GT04/99/TS/142 to E.A.H.

APPENDIX

Shimodaira-Hasegawa and Khisino-Hasegawa tests. In this study, phylogenetic trees were generated based on individual gene alignments, a combined alignment of the 63 shared genes, gene order, and gene content data sets. The strength of the phylogenetic signal inherent in the different data sets was assessed by performing Shimodaira-Hasegawa (SH) or Khisino-Hasegawa (KH) tests (Table A1). These tests compare several

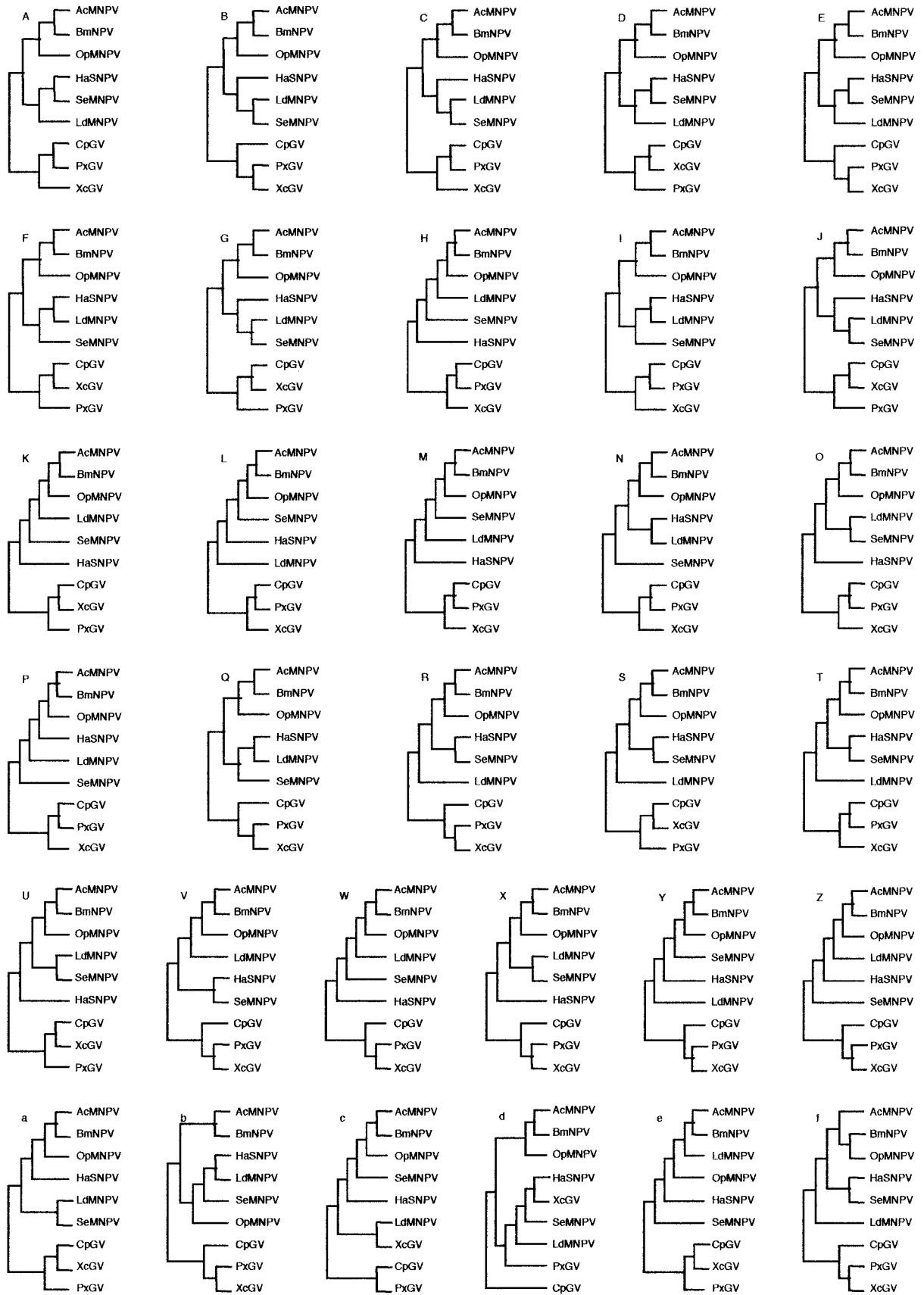


FIG. A1. Most-parsimonious tree topologies obtained for the individual phylogenetic analyses of the 63 shared genes and for the combined gene alignment, gene order, and gene content data sets. Table A1 shows which data set(s) gave rise to each tree.

TABLE A1. SH and KH test results

Data set	Size ^a	% Informative	Tree topologies ^b																															
			A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	a	b	c	d	e	f
<i>38.7K</i>	422	63.7												+																				**
<i>38K</i>	382	59.2			+																											**	**	**
<i>39K/pp31</i>	392	59.2			+																											**	**	**
<i>ac22</i>	439	51.7		+																												**	**	**
<i>ac23</i>	770	77.9																													**	**	**	**
<i>ac29</i>	142	42.3			+																										**	**	**	
<i>ac38</i>	272	52.9			+																										**	**	**	
<i>ac53</i>	160	72.5			+																										**	**	**	**
<i>ac66</i>	1,069	68.6														+															**	**	**	**
<i>ac68</i>	201	53.7																					+								**	**	**	
<i>ac75</i>	156	76.3				+																									**	**	**	
<i>ac76</i>	88	76.1								+																					**	**	**	
<i>ac78</i>	169	50.3																										+			**	**	**	
<i>ac81</i>	267	54.3		+																											**	**	**	**
<i>ac82/tlp20</i>	265	66.4				+																									**	**	**	
<i>ac92</i>	285	63.9									+																				**	**	**	
<i>ac93</i>	174	66.1				+						+																			**	**	**	
<i>ac96</i>	193	65.8									+																				**	**	**	
<i>ac106</i>	380	47.9																						+							**	**	**	
<i>ac109</i>	468	56.2																													**	**	**	
<i>ac110</i>	61	75.4																													**	**	**	
<i>ac115</i>	227	65.6																													**	**	**	
<i>ac119</i>	583	67.6		+																											**	**	**	
<i>ac142</i>	520	65.2		+																											**	**	**	**
<i>ac145</i>	102	65.7		+																											**	**	**	
<i>ac146</i>	217	81.1																													**	**	**	
<i>alk-exo</i>	517	59.0																													**	**	**	**
<i>dbp1</i>	365	72.6						+																							**	**	**	**
<i>dnapol</i>	1,200	63.4						+																							**	**	**	**
<i>fgf</i>	456	70.8																													+	**	**	**
<i>fp/25K</i>	249	50.6																													**	**	**	**
<i>gp41</i>	421	55.1																													**	**	**	**
<i>helicase</i>	1,349	73.4																													**	**	**	**
<i>ie1</i>	753	71.8									+																				**	**	**	**
<i>lef1</i>	314	58.0																													**	**	**	**
<i>lef2</i>	317	53.3																													**	**	**	**
<i>lef3</i>	463	75.4																													**	**	**	**
<i>lef4</i>	553	63.1																													**	**	**	**
<i>lef5</i>	320	59.1																													**	**	**	**
<i>lef6</i>	197	71.6																													**	**	**	**
<i>lef8</i>	942	53.0																													**	**	**	**
<i>lef9</i>	532	50.6																													**	**	**	**
<i>lef11</i>	193	50.3																													**	**	**	**
<i>me53</i>	523	62.1																													**	**	**	**
<i>odv-e18</i>	111	56.8																													**	**	**	**
<i>odv-e25</i>	242	64.5																													**	**	**	**
<i>odv-e56</i>	393	64.9																													**	**	**	**
<i>odv-e66</i>	800	59.3																													**	**	**	**
<i>odv-ec27</i>	312	73.7																													**	**	**	**
<i>p12</i>	134	82.1																													**	**	**	**
<i>p40</i>	424	77.4																													**	**	**	**
<i>p45</i>	421	67.9																													**	**	**	**
<i>p47</i>	488	51.0																													**	**	**	**
<i>p6.9</i>	117	29.9																													**	**	**	**
<i>p74</i>	785	53.6																													**	**	**	**
<i>p95</i>	1,096	55.5																													**	**	**	**
<i>pk1</i>	329	65.7																													**	**	**	**
<i>polh</i>	248	47.6																													**	**	**	**
<i>sod</i>	160	46.3																													**	**	**	**
<i>ubi</i>	172	20.3																													**	**	**	**
<i>vlf1</i>	438	63.2																													**	**	**	**
<i>vp1054</i>	632	40.8																													**	**	**	**
<i>vp39</i>	416	61.3																													**	**	**	**
All 63 genes	25,788	61.7		+																											**	**	**	**
Gene content	417	32.6									*	+																		*	**	**	**	**
Gene order	104	70.2																													*	**	**	**

^a Size of the aligned sequences.

^b +, most parsimonious tree topology; *, tree topology significantly different from the best topology as determined by the KH test performed on the binary data sets in PAUP; **, tree topology significantly different from the best topology as determined by the SH test performed on the amino acid sequence data sets in PAML.

tree topologies in relation to individual data sets to assess the compatibility of suboptimal trees. SH tests were performed for the molecular data sets, whereas KH tests were performed for the gene order and gene content data binary matrices. Individual phylogenetic analyses of the 63 common genes gave rise to 32 different tree topologies, which are presented in Fig. A1. The combined gene alignment, gene order, and gene content data sets yielded tree topologies that were included in this set of 32 trees. The trees of topologies c, d, and e (Fig. A1) are incompatible with most of the individual gene data sets as well as with the combined gene alignment, gene order, and gene content data sets (Table A1). In contrast, the tree of topology A is compatible with all data sets except that of the *odv-e66* gene (Table A1).

REFERENCES

- Ahrens, C. H., R. L. Q. Russell, C. J. Funk, J. T. Evans, S. H. Harwood, and G. F. Rohrmann. 1997. The sequence of the *Orgyia pseudotsugata* multicapsid nuclear polyhedrosis virus genome. *Virology* **229**:381–399.
- Ayres, M. D., S. C. Howard, J. Kuzio, M. Lopez-Ferber, and R. D. Possee. 1994. The complete DNA sequence of *Autographa californica* nuclear polyhedrosis virus. *Virology* **202**:586–605.
- Bideshi, D. K., Y. Bigot, and B. A. Federici. 2000. Molecular characterization and phylogenetic analysis of the *Harrisina brillians* granulovirus granulin gene. *Arch. Virol.* **145**:1933–1945.
- Blanchette, M., T. Kunisawa, and D. Sankoff. 1999. Gene order breakpoint evidence in animal mitochondrial phylogeny. *J. Mol. Evol.* **49**:193–203.
- Blissard, G. W., B. Black, N. Crook, B. A. Keddle, R. Possee, G. Rohrmann, D. A. Theilmann, and L. Volkman. 2000. Seventh report of the international committee on taxonomy of viruses, p. 195–202. In H. V. Van Regenmortel, D. H. L. Bishop, M. H. Van Regenmortel, and Claude M. Fauquet (ed.), *Virus taxonomy*. Academic Press, San Diego, Calif.
- Bulach, D. M., C. A. Kumar, A. Zaia, B. Liang, and D. E. Tribe. 1999. Group II nucleopolyhedrovirus subgroups revealed by phylogenetic analysis of polyhedrin and DNA polymerase gene sequences. *J. Invertebr. Pathol.* **73**:59–73.
- Chen, X., W. F. J. Ijkel, C. Dominy, P. Zanotto, Y. Hashimoto, O. Faktor, T. Hayakawa, C.-H. Wang, A. Prekumar, S. Mathavan, P. J. Krell, Z. Hu, and J. M. Vlak. 1999. Identification, sequence analysis and phylogeny of the *lef-2* gene of *Helicoverpa armigera* single-nucleocapsid baculovirus. *Virus Res.* **65**:21–32.
- Chen, X., W. F. J. Ijkel, R. Tarchini, X. Sun, H. Sandbrink, H. Wang, S. Peters, D. Zuidema, R. K. Lankhorst, J. M. Vlak, and Z. Hu. 2001. The sequence of the *Helicoverpa armigera* single nucleocapsid nucleopolyhedrovirus genome. *J. Gen. Virol.* **82**:241–257.
- Chen, X. W., Z. H. Hu, J. A. Jehle, Y. Q. Zhang, and J. M. Vlak. 1997. Analysis of the ecdysteroid UDP-glucotransferase gene of *Heliothis armigera* single nucleocapsid baculovirus. *Virus Res.* **15**:219–225.
- Clarke, E. E., M. Tristem, J. S. Cory, and D. R. O'Reilly. 1996. Characterization of the ecdysteroid UDP-glucosyltransferase gene from *Mamestra brassicae* nucleopolyhedrovirus. *J. Gen. Virol.* **77**:2865–2871.
- Clem, R. J., and L. K. Miller. 1994. Control of programmed cell death by the baculovirus genes *p35* and *iap*. *Mol. Cell. Biol.* **14**:5212–5222.
- Croizier, G., and H. C. T. Ribeiro. 1992. Recombination as a possible major cause of genetic heterogeneity in *Anticarsia gemmatalis* nuclear polyhedrosis virus populations. *Virus Res.* **26**:183–196.
- Doolittle, W. F. 1999. Phylogenetic classification and the universal tree. *Science* **284**:2124–2128.
- Fan, X., J. R. McLachlin, and R. F. Weaver. 1998. Identification and characterization of a protein kinase-interacting protein encoded by the *Autographa californica* nuclear polyhedrosis virus. *Virology* **240**:175–183.
- Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Biol.* **27**:401–410.
- Felsenstein, J. 2000. PHYLIP (phylogeny inference package), version 3.6. Department of Genetics, University of Washington, Seattle.
- Fitz-Gibbon, S. T., and C. H. House. 1999. Whole genome-based phylogenetic analysis of free living microorganisms. *Nucleic Acids Res.* **27**:4218–4222.
- Fraser, M. J., L. Cary, K. Boonvisudhi, and H.-G. H. Wang. 1995. Assay for movement of lepidopteran transposon IFP2 in insect cells using a baculovirus genome as a target DNA. *Virology* **211**:397–407.
- Funk, C. J., S. C. Braunagel, and G. F. Rohrmann. 1998. Baculovirus structure, p. 7–27. In L. K. Miller (ed.), *The baculoviruses*. Plenum Press, New York, N.Y.
- Gomi, S., K. Majima, and S. Maeda. 1999. Sequence analysis of the genome of *Bombyx mori* nucleopolyhedrovirus. *J. Gen. Virol.* **80**:1323–1337.
- Goto, C., T. Hayakawa, and S. Maeda. 1998. Genome organization of *Xestia c-nigrum* granulovirus. *Virus Genes* **16**:199–210.
- Hannenhalli, S., C. Chappey, E. V. Koonin, and P. A. Pevzner. 1995. Genome sequence comparison and scenarios for gene rearrangements—a test-case. *Genomics* **30**:299–311.
- Hashimoto, Y., T. Hayakawa, Y. Ueno, T. Fujita, Y. Sano, and T. Matsumoto. 2000. Sequence analysis of the *Plutella xylostella* granulovirus genome. *Virology* **275**:358–372.
- Hawtin, R. E., K. Arnold, M. D. Ayres, P. M. Zanotto, S. C. Howard, G. W. Gooday, L. H. Chappell, P. A. Kitts, L. A. King, and R. D. Possee. 1995. Identifying and preliminary characterization of a *chitinase* gene in the *Autographa californica* nuclear polyhedrosis virus genome. *Virology* **212**:673–685.
- Hayakawa, T., R. Ko, K. Okano, S.-I. Seong, C. Goto, and S. Maeda. 1999. Sequence analysis of the *Xestia c-nigrum* granulovirus genome. *Virology* **262**:277–297.
- Hayakawa, T., G.-F. Rohrmann, and Y. Hashimoto. 2000. Patterns of genome organization and content in lepidopteran baculoviruses. *Virology* **278**:1–12.
- Holmes, E. C., M. Worobey, and A. Rambaut. 1999. Phylogenetic evidence for recombination in dengue virus. *Mol. Biol. Evol.* **16**:405–409.
- Hu, Z. H., B. M. Arif, F. Jin, J. W. M. Martens, X. W. Chen, J. S. Sun, D. Zuidema, R. W. Goldbach, and J. M. Vlak. 1998. Distinct gene arrangement in the *Buzura suppressaria* single-nucleocapsid nucleopolyhedrovirus genome. *J. Gen. Virol.* **79**:2841–2851.
- Huh, N. E., and R. F. Weaver. 1990. Categorizing some early and late transcripts directed by the *Autographa californica* nuclear polyhedrosis virus. *J. Gen. Virol.* **71**:2195–2200.
- Ijkel, W., E. A. van Strien, J. G. Heldens, R. Broer, D. Zuidema, R. W. Goldbach, and J. M. Vlak. 1999. Sequence and organization of the *Spodoptera exigua* multicapsid nucleopolyhedrovirus genome. *J. Gen. Virol.* **80**:3289–33604.
- Kang, W., M. Tristem, S. Maeda, N. E. Crook, and D. R. O'Reilly. 1998. Identification and characterization of the *Cydia pomonella* granulovirus *cathepsin* and *chitinase* genes. *J. Gen. Virol.* **79**:2283–2292.
- Kishino, H., and M. Hasegawa. 1989. Evaluation of maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *J. Mol. Evol.* **29**:170–179.
- Kondo, A., and S. Maeda. 1991. Host range expansion by recombination of the baculoviruses *Bombyx mori* nuclear polyhedrosis virus and *Autographa californica* nuclear polyhedrosis virus. *J. Virol.* **65**:3625–3632.
- Koonin, E. V., L. Aravind, and A. S. Kondrashov. 2000. The impact of comparative genomics on our understanding of evolution. *Cell* **101**:573–576.
- Kuzio, J., M. N. Pearson, S. H. Harwood, C. J. Funk, J. T. Evans, J. M. Slavicek, and G. F. Rohrmann. 1999. Sequence and analysis of the genome of a baculovirus pathogenic for *Lymantria dispar*. *Virology* **253**:17–34.
- Le, T. H., D. Blair, T. Agatsuma, P. F. Humair, N. J. H. Campbell, M. Iwagami, D. T. J. Littlewood, B. Peacock, D. A. Johnston, J. D. Bartley, Rollinson, E. A. Herniou, D. S. Zarlenga, and D. P. McManus. 2000. Phylogenies inferred from mitochondrial gene orders—a cautionary tale from the parasitic flatworms. *Mol. Biol. Evol.* **17**:1123–1125.
- Maddison, D. R., and W. R. Maddison. 2000. MacClade 4. Sinauer Associates, Sunderland, Mass.
- Malik, N. S., S. Henikoff, and T. H. Eickbush. 2000. Poised for contagion: evolutionary origins of the infectious abilities of invertebrate retroviruses. *Genome Res.* **10**:1307–1318.
- Mitchell, A., C. Mitter, and J. C. Regier. 2000. More taxa or more characters revisited: combining data from nuclear protein-encoding genes for phylogenetic analyses of Noctuoidea (Insecta: Lepidoptera). *Syst. Biol.* **49**:202–224.
- Montague, M. G., and C. A. Hutchison. 2000. Gene content phylogeny of herpesviruses. *Proc. Natl. Acad. Sci. USA* **97**:5334–5339.
- Morris, A., M. Marsden, K. Halcrow, E. S. Hughes, R. P. Brettell, J. E. Bell, and P. Simmonds. 1999. Mosaic structure of the human immunodeficiency virus type 1 genome infecting lymphoid cells and the brain: evidence for frequent in vivo recombination events in the evolution of regional populations. *J. Virol.* **73**:8720–8731.
- Morse, M. A., A. C. Marriott, and P. A. Nuttall. 1992. The glycoprotein of Thogoto virus (a tick-borne orthomyxo-like virus) is related to the baculovirus glycoprotein gp64. *Virology* **186**:640–646.
- Page, R. D. M. 1996. TreeView: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* **12**:357–358.
- Pearson, M. N., C. Groten, and G. F. Rohrmann. 2000. Identification of the *Lymantria dispar* nucleopolyhedrovirus envelope fusion protein provides evidence for a phylogenetic division of the *Baculoviridae*. *J. Virol.* **74**:6126–6131.
- Rohrmann, G.-F. 1999. Nuclear polyhedrosis viruses. In R. G. Webster and A. Granoff (ed.), *Encyclopedia of virology*, 2nd ed. Academic Press, London, United Kingdom.
- Rohrmann, G. F., and P. A. Karplus. 2001. Relatedness of baculovirus and gypsy retrotransposon envelope proteins. *BMC Evol. Biol.* **1**:1.
- Rokas, A., and P. W. H. Holland. 2000. Rare genomic changes as a tool for phylogenetics. *TREE* **15**:454–459.
- Roncarati, R., and D. Knebel-Mörsdorf. 1997. Identification of the early actin-rearrangement-inducing factor gene, *arif-1*, from *Autographa californica*

- nica* multicapsid nuclear polyhedrosis virus. *J. Virol.* **71**:7933–7941.
49. **Shimodaira, H., and M. Hasegawa.** 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* **16**:1114–1116.
 50. **Snel, B., P. Bork, and M. A. Huynen.** 1999. Genome phylogeny based on gene content. *Nat. Genet.* **21**:108–110.
 51. **Swofford, D. L.** 2001. PAUP*. Phylogenetic analysis using parsimony (*and other methods), 4th ed. Sinauer Associates, Sunderland, Mass.
 52. **Tekaia, F., A. Lazcano, and B. Dujon.** 1999. The genomic tree as revealed from whole proteome comparisons. *Genome Res.* **9**:550–557.
 53. **Telford, M. J.** 2000. Evidence for the derivation of the *Drosophila* fushi tarazu gene from a Hox gene orthologous to lophotrochozoan Lox5. *Curr. Biol.* **10**:349–352.
 54. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**:4673–4680.
 55. **Tillier, E. R. M., and R. A. Collins.** 2000. Genome rearrangement by replication-directed translocation. *Nat. Genet.* **26**:195–197.
 56. **Wang, H. H., M. J. Fraser, and L. C. Cary.** 1989. Transposon mutagenesis of baculoviruses—analysis of Tfp3 lepidopteran transposon insertions at the fp locus of nuclear polyhedrosis. *Virus Genes.* **81**:97–108.
 57. **Winstanley, D., and D. O'Reilly.** 1999. Granuloviruses. In R. G. Webster and A. Granoff (ed.), *Encyclopedia of virology*, 2nd ed. Academic Press, London, United Kingdom.
 58. **Yang, Z. H.** 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**:555–556.
 59. **Zanotto, P. M. D., B. D. Kessing, and J. E. Maruniak.** 1993. Phylogenetic interrelationships among baculoviruses: evolutionary rates and host associations. *J. Invertebr. Pathol.* **62**:147–164.