A Cophylogenetic Perspective of RNA–Virus Evolution

Andrew P. Jackson and Michael A. Charleston

Department of Zoology, University of Oxford, Oxford, United Kingdom

The extent to which viruses and their hosts codiverge remains an open question, given that numerous cases of both ''cospeciation'' and horizontal switching have recently been documented. DNA viruses that form persistent infections are thought to be the most likely candidates for phylogenetic congruence. Phylogenetic reconciliation analysis was used to compare established phylogenies for four RNA viruses and their hosts. The analysis employs a cophylogeny mapping technique, implemented in TreeMap v2.0, to find the most parsimonious combinations of evolutionary events able to reconcile any incongruence. This technique is guaranteed to recover all potentially optimal solutions to the reconciled tree and specifically tests the null hypothesis that an associate phylogeny is no more congruent with a host phylogeny than would be a random tree with the same taxon set. Phylogenies for *Hantavirus*, *Spumavirus*, and avian sarcoma leukosis virus were found to be significantly similar to their host trees, whereas *Lyssavirus* and *Arenavirus* displayed no significant congruence. These results demonstrate that RNA viruses are able to form stable associations with their hosts over evolutionary time scales and that the details of such associations are consistent with persistent infection being a necessary but not sufficient precondition.

Introduction

RNA viruses are known for their roles in disease. They cause acute and highly virulent infections such as Mosquito-borne *Flavivirus*, causative agent of Yellow and Dengue fevers, and *Lyssavirus*, responsible for rabies worldwide. Other RNA viruses cause persistent infections, which may cause chronic illness or be largely asymptomatic. Retroviruses have become infamous as the human immunodeficiency virus (HIV) epidemic has spread across the world. Many details regarding the long-term association of viruses with their hosts remain uncertain, as do their origins generally. This study is concerned with the correspondence of viral and host phylogenies, and the phenomenon of host-virus codivergence, in five RNA viruses for which sufficient sequence data have been acquired:

Arenavirus *(Arenaviridae)*

Sigmodontine and murine rodents comprise the natural hosts for these viruses, where they cause persistent and largely asymptomatic infections (Southern 1996). Casual infection of humans by *Arenavirus* can lead to acute and potentially fatal haemorrhagic fever, for example, Lassa fever in Africa (McCormick et al. 1987). Macrophages are the predominant target for infection, a feature that facilitates its persistence in rodents (King et al. 1990).

Hantavirus *(Bunyaviridae)*

The hantaviruses have a worldwide distribution. Old World viruses infect arvicorine and murine rodents naturally; rodent infections are persistent and benign but, upon infecting humans, hantaviruses cause hemorrhagic fever and renal disorders (Schmaljohn and Dalrymple 1983). New World viruses are hosted by sigmodontine rodents and also cause disease in humans. The Sin Nombre virus (SNV) was found to be responsible for certain pulmonary disorders in the Southwestern United States (Nichol et al. 1993; Childs et al. 1994).

Lyssavirus *(Rhabdoviridae)*

Lyssavirus is found naturally among chiropteran and carnivoran mammals and is spread by the exchange of body fluids. There are seven phylogroups of rabies; this study concerns group I, the classical acute rabies in carnivores (Badrane et al. 2001).

Spumavirus *(Retroviridae)*

Spumavirus is a retrovirus that is widespread amongst primates. Routes of transmission are controversial, and both sexual (Broussard et al. 1997) and non-sexual (Blewett et al. 2000) means have been implicated. Infections are typically benign, asymptomatic, and lifelong (Coffin 1996). Increasing or dynamic antigenic variation is not a feature of this persistence (Broussard et al. 1997). However, *Spumavirus* may assist other retroviruses, such as HIV, in concomitant infections through the provision of a transcriptional transactivator (Blewett et al. 2000).

Avian Sarcoma Leucosis Virus (ASLV; Retroviridae)

The retroviruses of the avian sarcoma leucosis line come in both endogenous and exogenous forms. They cause a persistent, largely asymptomatic but oncogenic infection in galliform birds, frequently resulting in lymphoma. Host ranges are poorly known and transmission is by fluid exchange (Payne et al. 1991).

Host-Virus Codivergence

Ancient association of host and viral evolutionary lineages has been indicated by a number of studies, for example, in *Arenavirus* (Bowen, Peters, and Nichol 1997). This pattern is most consistent with phylogenetic estimates of Papovaviridae (Chan et al. 1992; Shadan and Villarreal 1993) and Herpesviridae (McGeoch et al. 1995; McGeoch, Dolan, and Ralph 2000). A recent review suggested that such ''cospeciation'' would follow from persistent viral infection over an evolutionary time scale, whereas acute

Key words: RNA–virus, phylogenetic reconciliation, cophylogeny, coevolution.

E-mail: andrew.jackson@zoo.ox.ac.uk.

Mol. Biol. Evol. 21(1):45–57. 2004
DOI: 10.1093/molbev/msg232
Molecular Biology and Evolution vol. 21 no. 1

infections would not (Villarreal, DeFilippis, and Gottlieb 2000). We use the term *codivergence* to refer to the concomitant divergence of host and viral lineages to prevent confusion with the general term *coevolution*. Codivergence is also distinct from phylogenetic *congruence*, which is a topological phenomenon and has many explanations, one of which is codivergence.

The positive relationship between virulence and transmission rate is widely known in viruses (Messenger, Molineux, and Bull 1999) and other parasites (Clayton and Tompkins 1994; Lipsitch, Siller, and Nowak 1996; Day 2001), and persistent and acute infections can be distinguished on the basis of virulence and transmission. Villarreal, DeFilippis, and Gottlieb (2000) proposed character combinations for idealized persistent and acute infections. From these ideal states it follows that those viruses eventually codiverging with their hosts will be those that (1) select their hosts specifically through vertical transmission or conspecific horizontal transmission, thereby limiting their transmission rate; (2) have few detrimental effects on the health of their host, i.e., their sole means of transmission; (3) form persistent, latent or lifelong infections due to their avirulence. Villarreal, DeFilippis, and Gottlieb (2000) have suggested that DNA viruses are far more likely to conform than RNA viruses. Because of the lack of any proof-reading functions in the viral polymerase, the latter have elevated substitution rates (Domingo et al. 1996). This elevation of substitution rates introduces instability and results in the formation of distinct lineages, even within a single host (Peeters and Sharp 2000). This should effectively uncouple host and virus evolutionary time scales. Hence, RNA viruses, it is supposed, will not retain the genetic integrity required to produce a codivergent cladogenic pattern with their hosts. To date, we have sought to identify phylogenetic congruence between four genera of RNA viruses and their hosts through the analysis of existing phylogenies with reconciliation analysis.

Cophylogenetic Reconciliation

Analyzing the similarity between phylogenies of coevolving taxa is part of a general comparative method (Brooks and McLennan 1993). Two associated phylogenies are almost invariably incongruent in places; while this immediately presents a puzzle, it is appreciated that this disagreement also represents a source of information about coevolutionary events (Nelson and Platnick 1981; Page 1990).

Phylogenetic reconciliation is the term given for the positing of evolutionary events to explain differences between trees that share a common history. The need for reconciliation analysis has been demonstrated at several conceptual levels, including species biogeography (Page 1988, 1990), the comparison of host and associate organismal phylogenies (Humphries, Cox, and Nielsen 1986; Page and Charleston 1998), and the resolution of orthologous and paralogous gene lineages (Goodman et al. 1979; Page and Charleston 1997). The present study is concerned with the second application and approached the problem through cophylogenetic mapping. In this procedure, nodes in the associate tree are mapped to positions in the host tree, based on the known associations, and any of four evolutionary events—codivergence, duplication, loss or switching—are posited for each associate lineage in a parsimonious way (Page 1990, 1994). Figuratively, the associate tree is laid over the host tree and embedded within it (Page and Charleston 1998).

Initial applications of this technique in the program Component (Page 1989), and subsequently TreeMap v1.0 (Page 1995), were unable to infer horizontal switching correctly or at all. A solution came in the form of the ''Jungle,'' which is a directed graph (digraph) containing all the ''potentially optimal'' reconstructions, given a certain weighting structure for the four event types given above (Charleston 1998). Jungles, as incorporated into TreeMap v2.0 (http://evolve.zoo.ox.ac.uk/software; Charleston and Page 2002), solve the problem of ''weakly incompatible host switches'' (i.e., switches that require subsequent loss events to make the source and destination contemporary), which previous methods failed to do (Charleston 1998) and permit the relative cost of event types to be customized, when appropriate.

Reconciliation using cophylogeny mapping and jungles guarantees to obtain all solutions potentially optimal under some set of costs (a set known as POpt). We believe that it is more exhaustive than methods that only evaluate the degree of tree similarity using maximum likelihood (Huelsenbeck, Rannala, and Yang 1997; Legendre, Desdevises, and Bazin 2002) or parsimony (Johnson, Drown, and Clayton 2000). Cophylogeny mapping produces reconciled trees with minimal assumptions about the relative likelihood of evolutionary events. The method assumes that the component phylogenies are robust and that codivergence is the default pathway during cladogenesis. These are conservative compared to likelihood-based methods (Huelsenbeck, Rannala, and Large 2000), which use explicit macroevolutionary models that cannot be verified in practice. In contrast, generalized parsimony assigns costs to particular evolutionary events, in order to evaluate solutions according to minimal total cost (Ronquist 1995). By mapping associates, cophylogeny mapping produces solutions that are easily interpreted. This is preferable to treating associates as characters on a host tree (Brooks 1981); the fundamental difficulty with this approach is that non-independent associates are treated as independent characters. This results in solutions that defy interpretation and incorrectly estimate the required number of events for reconciliation (Page 1990, 1994). Dowling's (2002) comparison of reconciliation methods indicated that these inaccuracies persist in revised versions (Page and Charleston 2002). Finally, unlike many other methods that are metrics, this method recognizes the general asymmetry in an association (the associate tracks the host phylogeny) during reconciliation and significance testing (see *Methods*).

In the present study, the existing phylogenies of the five RNA viruses described above were reconciled with those of their hosts using TreeMap v2.0. Randomized trees were used to evaluate the null hypothesis that the observed congruence was no greater than that expected between random trees. It is possible to constrain the jungle

construction to restrict the number of events of a particular type, or total number of non-codivergence events, or even the total cost. However, given that little information was available a priori to discriminate between any particular evolutionary scenario, the strategy throughout the analysis was to make each jungle analysis unconstrained, thereby recovering all the maps in POpt and accepting each solution as an equally valid working hypothesis.

Methods

Source Phylogenies

TreeMap requires fully resolved source phylogenetic trees; the 10 phylogenies used in the five reconciliation analyses, described in figure 1(*a–e*), were inferred both from published estimates and through phylogenetic reconstruction of sequences deposited in GenBank (accession records are available online as Supplementary Material). The results of reconciliation analysis are clearly dependent on the quality of the original trees, and sampling strategy should reflect both the hypothesis being tested and the diversity of the symbiosis. The following taxon sets contain as many of the available associations as possible, although size was limited because of complexity. Some taxa were excluded to provide resolution or to prevent artificial inflation of the congruence; for example, cat and marmoset spumaviruses are not included because these are outgroups to the main catarrhine clade, thereby guaranteeing two extra codivergence events:

Arenavirus

Sigmodontine host relationships were inferred from published phylogenies (Myers, Lundrigan, and Tucker 1995; Engel et al. 1998); viral phylogeny was taken from Bowen et al. (1997). The taxon set was limited to New World viruses and constrained to 11 in order to reduce computational complexity during cophylogeny mapping. The Flexal virus, which infects two hosts, *Oryzomys buccinatus* and *Oryzomys albigularis*, was excluded for this reason also.

Hantavirus

Host relationships for hantaviruses were a consensus of published estimates (Myers, Lundrigan, and Tucker 1995; Engel et al. 1998). Viral relationships were inferred by comparing several concordant phylogenies employing the glycoprotein locus (M segment) (Levis et al. 1998; Morzunov et al. 1998; Monroe et al. 1999; Sanchez et al. 2001).

Lyssavirus

Canid relationships were inferred from published phylogenies (Wayne et al. 1997; Bininda-Emonds, Gittleman, and Purvis 1999). Viral phylogeny was reconstructed using the phosphoprotein locus, with sequences (920 characters) deposited by Nadin-Davies et al. (1997). Given that the original study included a variety of mammalian hosts, it was prudent to reconstruct the phylogeny with purely Canid viruses. Sequences were aligned in Clustal X (Thompson et al. 1997) and then adjusted by eye in SeAl v1.0 (A. Rambaut, University of Oxford). Phylogenies were reconstructed in PAUP* v4.0 (Swofford 1998) using maximum parsimony (MP), maximum likelihood (ML), and Neighbor-Joining (NJ) methods. An MP bootstrap consensus tree was built with 100 nonparametric, bootstrapped replicates. A heuristic search using a TBR swapping algorithm was employed, after an initial tree was obtained by ''as-is'' stepwise addition (the maximum number of trees held was automatically increased). An Acctran mode of character state optimization was used, and all multiple states were considered as uncertainty. Maximum likelihood trees were built with a GTR + I model, the optimal model as designated by Modeltest v3.06 (Posada and Crandall 1998), and 100 bootstrap replicates. Neighbor-Joining trees were built with a minimum evolution criterion, maximum likelihood genetic distances, and 100 bootstrap replicates.

Spumavirus

Primate relationships were inferred from published accounts (Page and Goodman 2001). Viral phylogeny was estimated using two loci: the long terminal repeat (LTR—237 characters) locus and of the RNA Polymerase (*Pol*—283 characters). A robust phylogeny was possible for 10 viral strains using a combined data set. Spumaviruses from *C. hamlyni* and *P. pygmaeus* could not be placed robustly in this combined tree and were removed to ensure complete resolution. Phylogenetic reconstruction was carried out as described above, except that for LTR two short regions that showed no sequence homology were removed. Weighting by successive approximation (Swofford et al. 1996) was used to exclude homoplasious characters in MP searches, and a transversion model (TVM + Γ) was recommended by Modeltest for both ML and NJ searches. The tree was midpoint rooted.

ASLV

Dimcheff et al. (2000) published estimates of both host and parasite relationships in ASLV, using mtDNA and *gag*, respectively. The trees used in this study are taken from these estimates but include fewer taxa. The original taxonomic selection was sampled to provide a manageable taxon set that was representative of all galliform hosts. Where a host possessed multiple viral isolates, these were considered a single associate where all isolates formed a monophyletic clade. Where isolates from the same host were paraphyletic, for example in the cases of *Colinus virginianus* and *Phasianus colchicus*, these were recorded as multiple associates in the jungle analysis.

Cophylogeny Mapping

Phylogenies for each of five associations are shown in figure 1(*a–e*). Likelihood scores for host and associate topologies were compared using a Shimodaira-Hasegawa (SH) test (Shimodaira and Hasegawa 1999), when using the associate data set for associations c and d, where associate trees were estimated. These tanglegrams were subjected to reconciliation analysis through cophylogeny mapping. In figure 1*a* and *b* (*Arenavirus* and *Lyssavirus*)

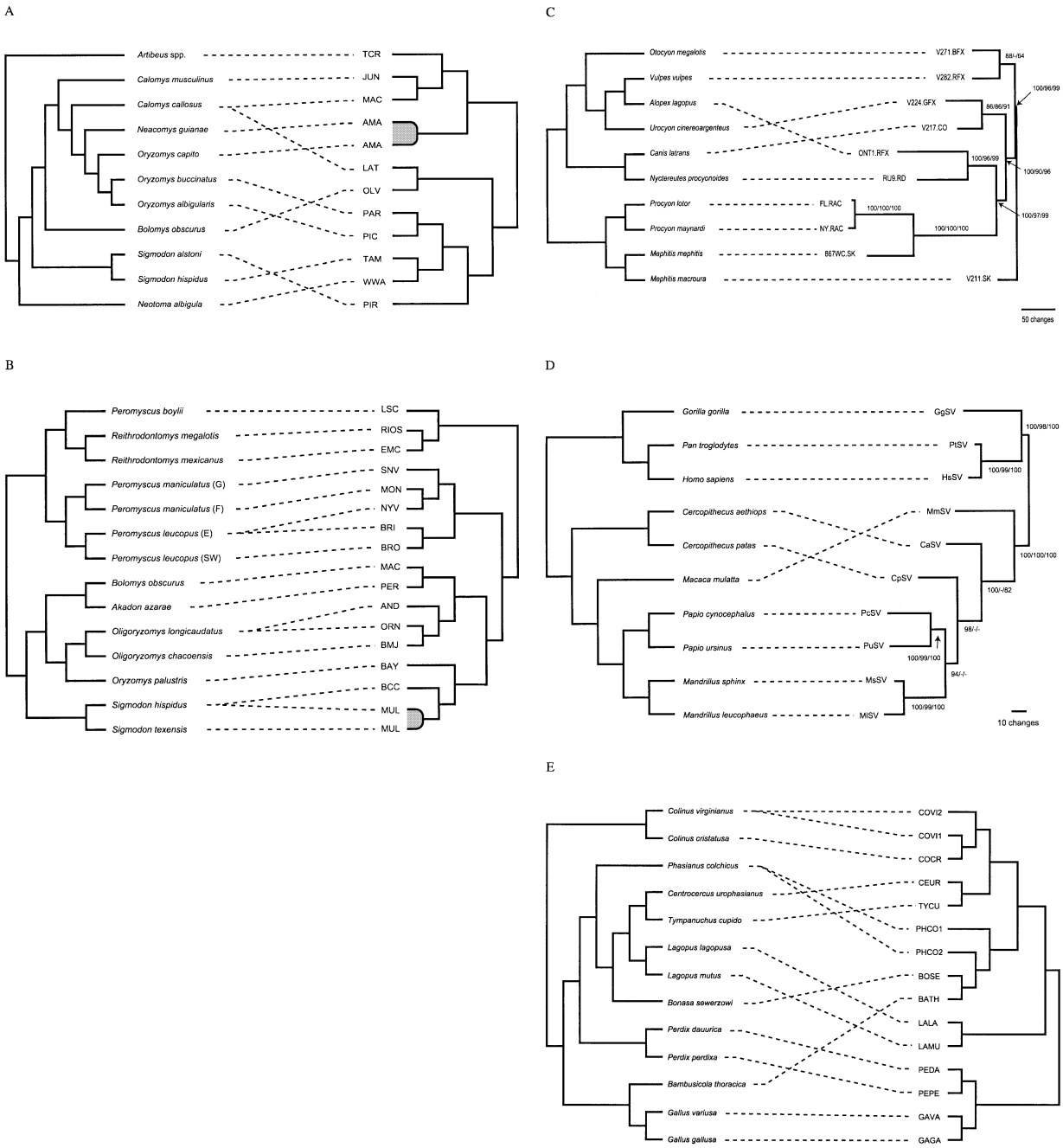

FIG. 1.—Tanglegrams describing the source phylogenies used in reconciliation analysis for *A*, *Arenavirus*; *B*, *Hantavirus*; *C*, *Lyssavirus*; *D*, *Spumavirus*; and *E*, ASLV. In all cases, hosts are shown on the left, viruses on the right; associated taxa are joined by dotted lines. Virus labels are described in table 1. In the case of *A* and *B*, a pair of shaded terminal nodes in the virus tree denotes a single viral species present in multiple hosts. In the cases of *C* and *D*, phylograms are drawn to a scale as shown, and bootstrap proportions are given in the format MP/ML/NJ.

shaded terminal nodes indicate a widespread taxon, that is, a virus present on two hosts. To accommodate this, the virus is assumed to form two isolated populations in each; this is the simplest interpretation and one that accommodates most scenarios. Outgroups were removed from input trees for *Lyssavirus* and *Spumavirus*, because these artificially inflate the number of codivergence events (CEs). A practical limit to the size of tanglegrams manageable by TreeMap is ≈20, although this varies according to the level of congruence between the trees. For highly incongruent tanglegrams, smaller numbers may have unreasonable calculation times or memory requirements. Hence, it is often prudent not to attempt a full jungle analysis (i.e., with all event bounds unconstrained) in the first instance. Rather, the analysis is repeated with decreasing constraints, becoming increasingly inclusive. Thus, the number of CEs required for a solution to be accepted is initially high and is gradually reduced; i.e., the bounds become more permissive, until all solutions are found or the analysis cannot be completed. In most cases, it is possible to find all potentially optimal solutions.

Significance Testing

The significance of any congruence between two trees can be evaluated in TreeMap through randomization, using a Markov model to reconstruct random associate trees. Because of the inherent asymmetry in the relationship between associate and host and the hypothesis being tested (i.e., that the associate tracks the host through evolutionary time), it is appropriate to randomize only the associate tree during significance testing. The null hypothesis is that the genuine associate tree has no more congruence with the host tree than does a randomized associate tree. This is rejected if the genuine level of congruence is seen in no more than 5% of randomizations. Congruence can be measured in terms of the maximum number of CEs or the minimum number of non-codivergence events (NCEs—i.e., duplication, loss, and switching events). All solutions were considered equally valid hypotheses and tested using both CE and NCE frequency. The number of NCEs was considered to be a more sensitive indicator of significant correspondence because this quantity has a wider distribution than the number of CEs (M.A. Charleston, unpublished manuscript).

''Cherry-Picking'' Test

Significant phylogenetic congruence can often derive solely from extant associations, reflecting a difference between contemporary and past dynamics. A jungle analysis can be constrained to test the specific effect of terminal nodes, or associations involving congruent pairs of sister taxa (''cherries''), through the removal of one association from each of these couplets, followed by reanalysis to identify any change in significance. This test is introduced as the ''cherry-picking test'' (CPT). In the case of *Hantavirus* and ASLV, the contribution of cherries to the overall result was evaluated using this test. Significance, as measured by minimum NCEs, was calculated after removing five terminal nodes from the jungle analysis. For *Hantavirus*, three cherries involved multiple interactions; eight permutations resulted from removing associations, and each was tested separately. The potential for reduced sample size to have an effect on significance would ordinarily require a control test, in which the same number of associations would be removed from the analysis at random over a number of trials.

Results

Source Phylogenies

The source phylogenies used in the reconciliation analysis are described by figure 1(*a–e*). *Lyssavirus* and *Spumavirus* MP phylogenies are represented as phylograms as these were reconstructed from sequence data. Other trees, each a consensus of published estimates, are represented as cladograms. *Lyssavirus* and *Spumavirus* phylogenies were fully resolved, although at a cost of removing unresolved taxa, and they were relatively robust, as evidenced by bootstrap values in excess of 75%. Most nodes are supported by MP, ML, and NJ methods. The *Lyssavirus* MP tree had a length of 366.0 steps with 286 parsimony informative (p.i.) characters, RI = 0.870, and RC = 0.774. The *Spumavirus* (LTR) MP tree had a length of 157.6 steps, 44 p.i. characters, RI = 0.853, and RC = 0.820. *Spumavirus* (*Pol*) had a length of 155.1 steps, 64 p.i. characters, RI = 0.865, and RC = 0.778. LTR and *Pol* loci differed only in the placement of the *Cercopithecus hamlyni* virus and were not significantly different (Kishino-Hasegawa test, one-tailed, full optimization, $P = 0.286$ using *pol* data and 0.302 using LTR data). The combined *Spumavirus* tree had a length of 251.3 steps, 155 p.i. characters, RI = 0.905, and RC = 0.850.

Reconciliation Analysis

The results of reconciliation analysis are shown in table 1 and summarized in figure 2. POpt for *Arenavirus* and *Lyssavirus* included 25 and 74 solutions, respectively. *Arenavirus* and *Lyssavirus* phylogenies show no significant congruence with those of their respective hosts. The *Lyssavirus* result is corroborated by the SH test ($\Delta = 164.0$, $P < 0.001$). *Arenavirus* becomes less significant if the *Flexal* virus is included, although not all analyses could be completed. Their reconciled trees possess no more CEs and no fewer NCEs than an average random associate tree when reconciled. Adjusted incongruence is the number of NCEs required to reconcile two trees, divided by the number of lineages ($2N - 2$, where N is the number of taxa). Adjusted incongruence for these analyses suggests that *Arenavirus* shows less congruence with its hosts than *Lyssavirus*, the former requiring a maximum of 1.90 NCEs per taxon to be reconciled, whereas the latter required only 1.28.

Conversely, all nine solutions generated for *Hantavirus* showed a significant degree of congruence with their host phylogenies ($P < 0.01$), as measured by NCE frequency. Significance values using CE frequencies for *Hantavirus* could not be obtained because of computational complexity introduced by the large number of taxa and multiple interactions involved. The maximum adjusted incongruence (1.00) for these taxa is lower than for the previous two. The effect of removing associated terminal sister taxa in the CPT was to slightly reduce the significance of congruence between *Hantavirus* and Sigmodontine phylogenies. Although seven of the eight permutations of the edited tree showed no reduction in significance ($P < 0.01$) for the minimum number of NCEs, evaluation with the maximum CEs was usually nonsignificant. However, the overall result remained that these two trees were significantly similar under NCEs, even when the tipward associations were removed in the CPT. Given this result, no control test, in which random terminal nodes are removed, was required.

The *Spumavirus* phylogeny was not significantly different from the host phylogeny; this was supported by the SH test ($\Delta = 5.28$, $P = 0.228$). Expanded phylogenies for *Pol* and LTR, which contained viruses from *C. hamlyni* and *P. pygmaeus*, produced single parsimonious trees. Although not consistently resolved, the position of these two viruses was consistent with codivergence. Two further viruses from *M. cyclopsi* and *P. panicus*, which were placed using a third locus (*int*), also enhanced the congruence when included ($P < 0.01$ when testing the

Table 1
Results of Reconciliation Analysis

				CE				NCE					
	Taxa			Min.		Max.		Min.		Max.		Adjusted Incongruence	
Association	Host	Associate	|POpt|	#	p	#	p	#	p	#	p	Lower	Upper
Sigmodontinae—*Arenavirus*	11	12	25	8	1.00	14	0.05	19	0.24	38	1.00	0.95	1.90
Sigmodontinae—*Hantavirus*	14	17	9	20	0.06	22	**<0.01**	16	**<0.01**	26	**<0.01**	0.62	1.00
Canidae—*Lyssavirus*	10	10	74	4	1.00	12	0.19	14	0.14	23	1.00	0.78	1.28
Primates—*Spumavirus*	10	10	8	12	0.09	16	**<0.01**	4	**<0.01**	10	**<0.01**	0.22	0.56
Galliformes—ASLV	13	15	66	12	0.52	16	0.18	22	**<0.01**	38	1.00	0.92	1.58

NOTE.—Five jungles were created using TreeMap for *Arenavirus*, *Hantavirus*, *Lyssavirus*, *Spumavirus*, and avian sarcoma leukosis virus (ASLV). The number of taxa involved and the number of solutions comprising POpt are given, alongside the minimum and maximum numbers of CEs and NCEs, with their significance values; values in bold denote significant congruence. Lower and upper adjusted incongruence describes the minimum and maximum number of NCEs required to reconcile the two trees, divided by the number of lineages.

maximum CEs for both *pol* and LTR topologies). ASLV was significantly congruent using minimum NCE ($|POpt| = 66$; $P = 0.02$). However, a solution for the ASLV jungle is shown in figure 3. This clearly demonstrates that this similarity is derived from five matched cherries at the tips of the reconciled tree. The remainder shows no indication of congruence. A CPT for this jungle shows the tree is nonsignificant when these matching cherries are removed ($P = 0.55$ for minimum NCE), and a control test showed that the change in significance was not caused by reduced sample size. Randomly removing the same number of associations over 30 trials produced fewer NCEs and greater significance in all cases.

Discussion

This study has identified significant phylogenetic congruence between RNA viruses and their hosts. Charleston and Robertson (2002) analyzed the cophylogenetic patterns between a set of Lentiviruses, the group comprising HIV and its Simian counterparts, and their primate hosts and found significant congruence. An additional simulation indicated that this was also consistent with preferential host switching, rather than the immediately intuitive codivergence of host and pathogen lineages during cladogenesis. This observation is a salutary reminder that phylogenetic congruence can indicate mechanisms other than codivergence. *Phylogenetic congruence* refers simply to the recapitulation of a first cladogenic pattern in a second, associated tree. Crucially, congruence contains no temporal assumption, and so three possible explanations exist: (1) preemptive cladogenesis by the pathogen; (2) codivergence (or ''cospeciation''), i.e., synchronous codivergence; and (3) delayed cladogenesis by the pathogen (Hafner and Nadler 1990; Page 1990). Codivergence (the common conclusion) therefore differs fundamentally from congruence (the observation) by assuming biological reality. Studies of host-RNA virus associations have frequently brushed over other potential explanations of phylogenetic congruence, attributing it to cospeciation (Bowen, Peters, and Nichol 1997; Charrel, de Micco, and de Lamballerie 1999; Monroe et al. 1999; Dimcheff et al. 2000). Where there is a great disparity between the evolutionary rates of host and pathogen, the diversification of a pathogen with preferential host switching toward related hosts is a feasible mechanism and would produce a degree of phylogenetic congruence. Congruence near the base of the reconciled tree is a more effective indicator that codivergence is the characteristic dynamic, because this signal has clearly not been obliterated by subsequent switching events; such dynamics have been identified in DNA viruses such as Herpesvirus (McGeoch et al. 1995) and are indicated by RNA viruses such as *Hantavirus* and *Spumavirus* (see below). However, the incidences of ''cospeciation'' in *Arenavirus* (Bowen, Peters, and Nichol 1997), ASLV (Dimcheff et al. 2000), T-cell lymphoma virus (TLV) (Meertens et al. 2001), or GB virus (Charrel, de Micco, and de Lamballerie 1999) cannot be said to characterize the long-term host-virus associations in the absence of a significant level of congruence across the tree.

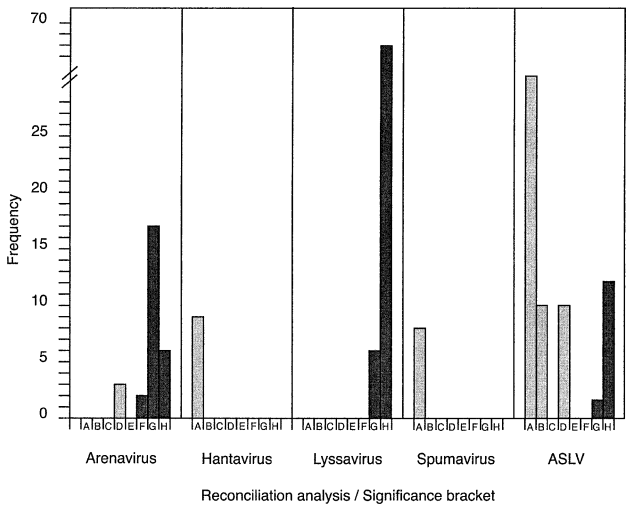

FIG. 2.—A summary diagram showing the frequency distribution of significance values of all potentially optimal reconstructions for six jungles. For each analysis, the significance value, as measured by NCE frequency (see text for explanation) for each solution is put into one of eight significance brackets (*A–H*): $P < 0.01$ (*A*), $0.01 < P < 0.02$ (*B*), $0.02 < P < 0.03$ (*C*), $0.03 < P < 0.04$ (*D*), $0.04 < P < 0.05$ (*E*), $0.05 < P < 0.1$ (*F*), $0.1 < P < 0.5$ (*G*), $0.5 < P < 1.0$ (*H*). Significant solutions are represented with light shading; nonsignificant solutions, with dark shading.

Phylogenetic Incongruence: *Arenavirus* and *Lyssavirus*

Arenaviruses are thought to be half of an ancient association (Johnson, Webb, and Justines 1973; Bowen,

Peters, and Nichol 1997) and this is consistent with their forming benign, persistent infections in rodents (Southern 1996). The hypothesis has been based on isolated instances of phylogenetic congruence, but no temporal analysis exists to identify such topological congruence as cospeciation. The essential distribution of Lassa-like viruses in Old World murine hosts and Tacaribe strains in New World sigmodontine hosts suggests some historical delimitation of virus specificity, although this is also explicable on biogeographical grounds. The seminal study by Bowen, Peters, and Nichol (1997) emphasized the instances of phylogenetic congruence apparent between host and *Arenavirus* phylogenies. Certain related viral strains shared congeneric hosts, for example, JUN and MAC were found in *Calomys* spp., and FLE, PAR, and PIC strains were found in *Oryzomys*. However, in each case of putative cospeciation raised by Bowen, Peters, and Nichol (1997) the qualification of an associated incongruence is required. Thus, *Callomys* also hosts the LAT strain, unrelated to JUN and MAC; *Oryzomys* also hosts the AMA strain, unrelated to either PAR or PIC. The conclusion here is that codivergence occurs no more frequently than expected by chance, and so it is incongruence, and the processes that cause it, that have characterized murid-*Arenavirus* coevolution. If the flexal virus is included, most significance measures cannot be obtained by reason of computational complexity; however, the value for minimum NCE becomes even less significant ($P = 0.1$) than previously.

Admittedly, this analysis has not included all viral strains, and the Old World arenaviruses are notable absentees. Hugot, Gonzalez, and Denys (2001) used TreeMap v1.0 to analyze Old World arenaviruses alone and raised a number of methodological issues. They incorporated secondary host affiliations hitherto ignored by phylogenetic studies. This may turn out to be an important distinction, but before introducing another level of complexity into the analysis, it should be established that these secondary hosts are not simply casual and transient spillover infections from sympatric primary hosts. With only five viral strains, the analysis was unlikely to produce a significant result because it is often impossible to obtain a level of congruence that is distinguishable from a random assemblage from such a small data set. In evaluating their reconciled trees, Hugot, Gonzalez, and Denys (2001) used a general parsimony principle. TreeMap uses a parsimony principle in minimizing the number of events posited as a pragmatic measure to limit POpt. Despite this, it is ill-advised to discriminate between solutions using the number of events because, first, NCEs are not necessarily comparable and second, not all events are observable, akin to the problem of ''multiple hits'' in phylogenetic reconstruction, but with no known correction. A single host switch may not be comparable to a single loss because the loss may necessitate subsequent losses toward the tips, especially if the first event is near the root. The issue of ''cophylogeny without ''cospeciation'' raised by Hugot, Gonzalez, and Denys (2001) leads to a single associate being present on several hosts and has long been recognized within cophylogeny theory as the problem of

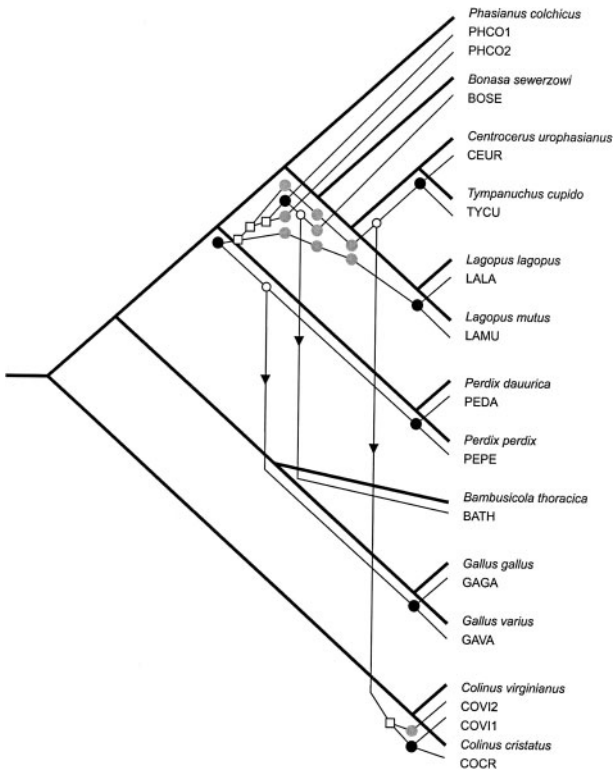

FIG. 3.—A representative reconciled tree for Avian sarcoma leucosis virus (ASLV, Retroviridae) infecting galliform birds. This solution is one of 66 potentially optimal reconstructions. This reconciled tree showed significant congruence when using minimum NCEs ($P = 0.02$), but significance was lost following a CPT ($P = 0.55$). Evolutionary events are symbolized thus: codivergence (black circle), duplication (square), loss (gray circle), or switch (arrow).

''widespread taxa'' (Platnick and Nelson 1978; Page 1988). This phenomenon is not currently included within cophylogeny mapping, because there is no way of reliably inferring the point of entry into the infected host clade or the sequence in which each host was colonized. Without knowledge of whatever substructure exists within the associate population (e.g., host-related strains), one cannot map the associate into the host tree.

Lyssavirus also demonstrated no significant phylogenetic congruence with its hosts. This is of little surprise given what is known about Rhabdovirus biology; wide host ranges are common and the virus causes an acute illness, resulting in high rates of horizontal transmission. Furthermore, a lack of proof-reading function in the RNA polymerase results in a high synonymous mutation rate (Wagner and Rose 1996) and *Lyssavirus* evolves at a much greater rate than its hosts (Kissi, Tordo, and Bourhy 1995). This mutability may contribute to the ability of *Lyssavirus* to readily infect novel hosts. Indeed, previous studies have documented historical host switches of *Lyssavirus* from bats to carnivores (Bourhy, Kissi, and Tordo 1999; Badrane and Tordo 2001), a transfer that has apparently produced a more acute disease in the form of classical rabies. In this study, domestic dogs were removed from the phylogeny because this cosmopolitan species harbors a plethora of *Lyssavirus* strains that distribute throughout

the phylogeny, demonstrating the ease with which the virus spreads either to or from dogs as they have spread across the world. However, despite the recent origin and ease of horizontal transfer of *Lyssavirus* in carnivores, which ensures that codivergence cannot possibly be a characteristic pattern, the reconciliation of host and viral phylogenies produces occasional incidences of codivergence. From two to six nodes, from a total of nine, were codivergent in the *Lyssavirus* analysis, compared with between four and seven of 10 in the *Arenavirus* analysis. As table 1 shows, the *Lyssavirus* phylogeny required fewer non-codivergence events than the *Arenavirus* phylogeny to be reconciled (0.78–1.28 compared with 0.95–1.9). This illustrates the need for a test of nonsignificant incongruence in cophylogenetic analysis, rather than simply identifying apparent codivergence events.

Phylogenetic Congruence: *Hantavirus* and *Spumavirus*

Various aspects of *Hantavirus* biology have indicated that there may be congruence between murid host and viral phylogenies (Hjelle et al. 1995; Plyusnin, Vapalahti, and Vaheri 1996; Monroe et al. 1999). The resolution of viral clades infecting murine (Wang et al. 2000), sigmodontine (Levis et al. 1998), and arvicoline (Monroe et al. 1999) rodents, and within host genera and species, suggests that biological boundaries exist to prevent switching between these hosts. Evidence also indicates that viral genetic structure and diversity mirrors that of host populations. In North America, BRI and BRO viruses are sibling species associated with distinct allopatric populations of *Peromyscus leucopus* (Morzunov et al. 1998), whereas SNV and MON viruses are similarly related but found in distinct, sympatric ecotypes of *Peromyscus maniculatus* (Monroe et al. 1999). This phylogenetic correspondence is consistent with a model of host-virus codivergence through evolutionary time, but also, to a large degree, with host distribution, suggesting a model of host tracking through ecological time.

Putative cases of host switching by *Hantavirus* have been identified, suggesting that any biological boundaries can be overcome under some circumstances. Monroe et al. (1999) note the cross-family transmission of *Hantavirus* from *Microtus* to *Peromyscus*, although there is no suggestion that this has led to a permanent association. Morzunov et al. (1998) discuss possible explanations for the disjunctive relationships of SN-like viruses infecting *P. leucopus*. Figure 4 illustrates that the consensus of current solutions favors a host switch event involving NYV. As figure 1 suggests, this is from the lineage leading to the common ancestor of *P. maniculatus* and *P. leucopus*. South American hantaviruses also provide potential cases of host switching; the AND strain was identified as an exception to the general pattern of codivergence (Levis et al. 1998) and host switching is favored as an explanation here. Sanchez, Abbott, and Nichol (2001) suggested that Limestone Canyon Virus (LSC) represents a case of host switching, because it is hosted by *Peromyscus boylii* yet groups with viruses from the host genus *Reithrodontomys*. However, the rodent mitochondrial phylogeny suggests that *Peromyscus* is paraphyletic and that *P. boylii* groups with *Reithrodontomys* (Engel et al. 1998). In this situation, the position of LSC is consistent with codivergence.

Previous studies of *Hantavirus* diversity in North and South America have concluded that the recognized strains are genetically distinct, indicative of a long period of isolated and stable coevolution (Levis et al. 1998; Morzunov et al. 1998; Monroe et al. 1999); the result of the CPT also shows that codivergence is associated with deep nodes as well as recent cladogenesis. This result contrasts well with ASLV, for which the original jungle proves significantly congruent but a CPT shows that this significance derives from five matching ''cherries.'' Clearly, host switching and/or lineage loss has obscured any trace of ancient codivergence events, recent CEs being visible only until future NCEs obliterate their mark. Regarding *Hantavirus*, if there is a low level of interstrain genetic divergence (Bohlman et al. 2002), this may be the result of the recent radiation of rodent hosts (Monroe et al. 1999). As described above, congruence suggests, but does not demand, codivergence, and so, to substantiate a mechanism of codivergence, a comparison of genetic distances will be required to identify proportional interspecific divergence in the host, and thereby demonstrate synchronous cladogenesis. If we are to accept a model of codivergence for *Hantavirus*, we must explain the lack of host switching, the apparent host specificity, and local adaptation, despite opportunities for interspecific transmission; this requires a hypothesis of poor dispersal and low viability outside the host, perhaps related to the loss of Arthropod vectors in this genus of the Bunyaviridae.

Virally encoded reverse transcriptase gives retroviruses the ability to integrate with the host genome and be inherited across generations (Coffin 1996). Phylogenies of endogenous retroviruses, as these inherited proviruses are called, have been shown to recapitulate those of their primate hosts (Johnson and Coffin 1999). Many of the endogenous retroviruses, which number in the thousands, present in a primate genome are of ancient origin and cannot express virus due corruption by point mutations (Boeke and Stoye 1997). Hence, host-viral codivergence is unsurprising in this case, because these proviruses effectively became host markers long ago (Johnson and Coffin 1999). However, endogenous retroviruses may represent our ultimate expectation for all retroelements. The important distinction for exogenous retroviruses is that they retain the ability to express and transmit virus, often virulently, and therefore, the possibility of transcending a purely codivergent pattern. The apparent congruence between *Spumavirus* and primate phylogenies identified previously (Schweizer and Neumann-Haefelin 1995; Bieniasz et al. 1995; Broussard et al. 1997) has been shown here to be significant—indeed, consistent with strict cospeciation. Moreover, it is improved further if other, tentatively placed viruses are added to those shown in figure 1. This is in keeping with the lifelong and apparently asymptomatic infections (Coffin 1996) and the low levels of genetic variation observed within and between individuals (Schweizer et al. 1999). Moreover, although no CPT was performed (as this would make the taxon number insignificant), CEs within the *Spumavirus* reconciled tree are common among the deeper nodes (fig. 4); this suggests

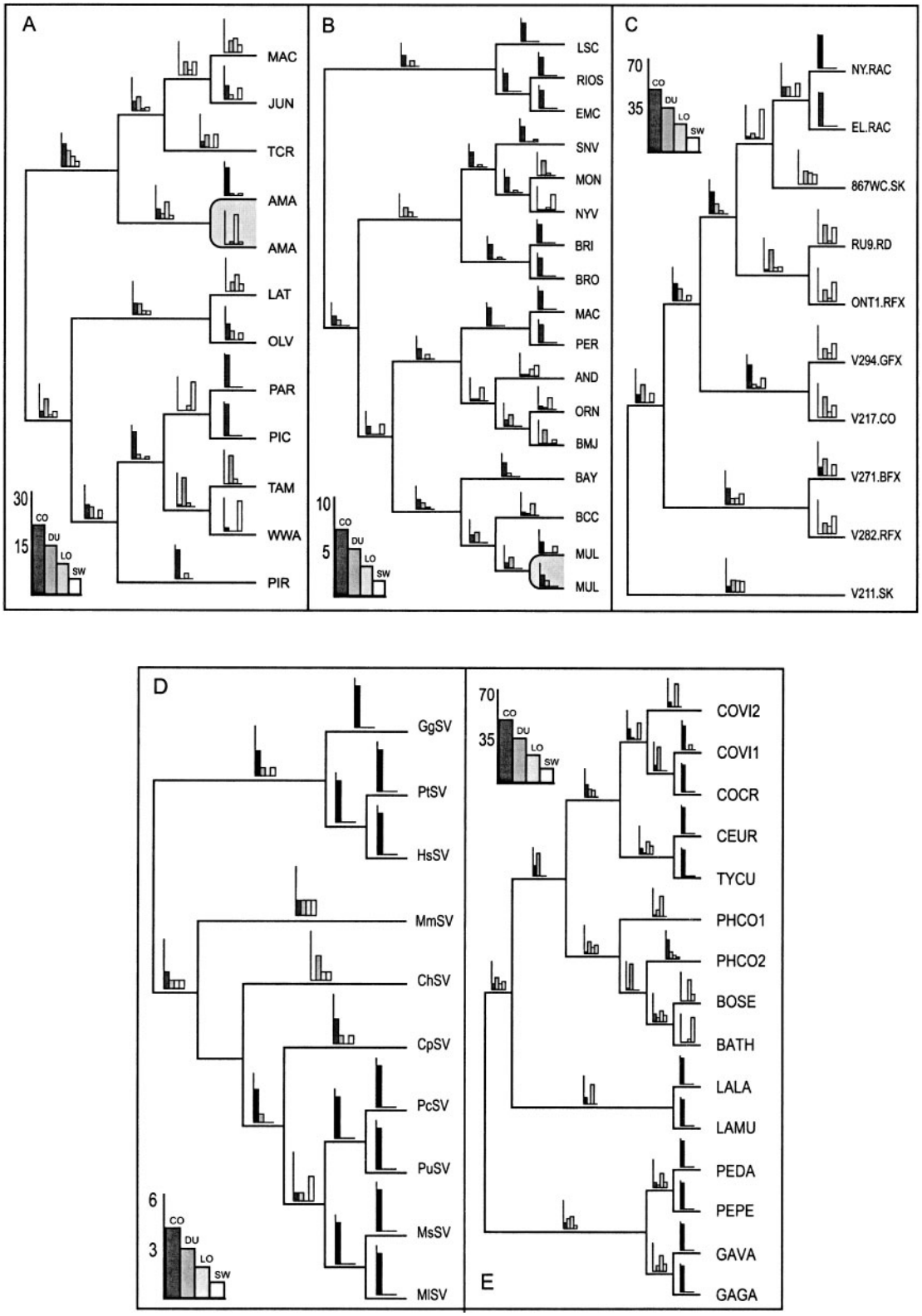

FIG. 4.—Consensus diagrams showing the origin of associate lineages across all solutions for *A*, *Arenavirus*, *B*, *Hantavirus*, *C*, *Lyssavirus*, *D*, *Spumavirus*, and *E*, ASLV. For each associate lineage, up to four bars are shown representing each of the four evolutionary events posited by TreeMap. These are shown in order from left to right (and darkest to lightest): codivergence, duplication, loss, and switching. The scale varies as indicated. Shaded arcs indicate viral species that infect multiple hosts.

that codivergence has characterized *Spumavirus* diversification since before recent times.

The *Spumavirus* result is consistent with the expectations of Villarreal, DeFilippis, and Gottlieb (2000), but it contrasts with two other retroviruses: *Lentivirus* and TLV. *Lentivirus* forms persistent infections and displays species-specific diversity in its primate hosts (Allan et al. 1991). Although host switching has been demonstrated (Jin et al. 1994), most remarkably from two primates into humans (Chen et al. 1997; Gao et al. 1999), host-virus codivergence has been suggested (Beer et al. 1999). Correspondence between *Lentivirus* and primate phylogenies could in large part be due to preferential host switching (Charleston and Robertson 2002), and this is consistent with the extraordinarily high replication rate in Lentiviruses and the consequent intrahost genomic variability (variability rate of between 10^{-2} and 10^{-3} per site per year) (Hahn et al. 1986; Coffin 1992; Wain-Hobson 1996). In stark contrast, *Spumavirus* displays genomic stability within its hosts, which is probably maintained by a low replication rate, and does not rely on antigenic variation for its persistence (variability rate of $\approx 3 \times 10^{-4}$ per site per year); a low ratio of nonsynonymous to synonymous mutations further suggests that purifying selection effects a conservative rate of amino acid substitution in surface antigen (Schweizer et al. 1999). As noted for *Hantavirus* above, special circumstances must be required to suppress the normally high rates of substitution and diversification in RNA viruses; clearly, in the case of *Spumavirus*, but not *Lentivirus*, a life strategy has evolved that has brought about such suppression of replication. Any satisfaction at having made this comparison is tempered by TLV, a virus that shares its life strategy with *Spumavirus* and has a variability rate 2–4 orders of magnitude lower than *Lentivirus* (Pecon-Slattery, Franchini, and Gessian 1999). Human TLV is the etiological agent of various human diseases including adult T-cell leukemia and chronic progressive myelopathy. It replicates through clonal expansion of host cells rather than the normal cycle involving reverse transcription; this results in first, a very low replication rate and thus, low intrahost variability, and second, its oncogenic properties (Pecon-Slattery, Franchini, and Gessian 1999). Despite these features, the seminal TLV phylogeny dismissed the possibility of host-viral codivergence (Koralnik et al. 1994), and rampant host switching has since been identified as the norm (Crandall 1996; Meertens et al. 2001).

Conclusion

An ancient association of viral lineages with their hosts has been thought to be the preserve of DNA viruses such as Herpesvirus (Bowen, Peters, and Nichol 1997) and Polyomavirus (Shadan and Villarreal 1993). These results have indicated that RNA viruses, specifically *Hantavirus* and *Spumavirus*, show greater phylogenetic congruence with their hosts than predicted by chance and so can also be codivergent. The observation that these viruses cause persistent and largely benign infections and form distinct genetic entities within host populations is consistent with the theory that such characters are the basis of host specificity and the long-term association of phyletic lineages (Villarreal, DeFilippis, and Gottlieb 2000). However, the converse view, that the elevated base substitution rate of RNA viruses relative to their hosts should decouple their cladogenic patterns, persists and is consistent with other cases, such as *Lyssavirus*, where the null hypothesis cannot be rejected. Hence, it may be said that persistent and benign infections are a necessary, but not sufficient, precondition to codivergence in both RNA and DNA viruses. *Arenavirus*, as well as HTLV and ASLV, demonstrate that infections of this type can produce alternative patterns. Furthermore, a complication is that even where this type of infection is evident, codivergence does not necessarily follow from congruence; apparent codivergence involving *Lentivirus* (Charleston and Robertson 2002) and perhaps *Arenavirus* can be explained through preferential host switching to related hosts long after host diversification.

Few general diagnostic indicators seem to exist to predict when a virus will codiverge with its hosts. Particular life history traits such as low virulence will result in poor dispersal and vertical transmission that may predispose codivergence. Conversely, the use of a hematophagous vector would intuitively make host specificity less likely. However, even where conditions appear optimal, as in the case of *Arenavirus*, the result may not be consistent codivergence; this uncertainty may indicate that special circumstances must be required to overcome the inherent disparity between virus and host substitution rates and to couple viral and host diversification.

Acknowledgments

We are grateful to Eddie Holmes and to anonymous reviewers for helpful comments during the development of this study. A.P.J. is supported by a Medical Research Council Research Studentship, and M.A.C. is supported by a Royal Society University Research Fellowship.

Literature Cited

Allan, J. S., M. Short, M. E. Taylor, S. Su, V. M. Hirsch, P. R. Johnson, G. M. Shaw, and B. H. Hahn. 1991. Species-specific diversity among Simian immunodeficiency viruses from African green monkeys. J. Virol. **65**:2816–2828.

Badrane, H., and N. Tordo. 2001. Host switching in *Lyssavirus* history from the Chiroptera to the Carnivora orders. J. Virol. **75**:8096–8104.

Badrane, H., C. Bahloul, P. Perrin, and N. Tordo. 2001. Evidence of two *Lyssavirus* phylogroups with distinct pathogenicity and immunogenicity. J. Virol. **75**:3268–3276.

Beer, B. E., E. Bailes, R. Goeken, G. Dapolito, S. G. Norley, R. Kurtii, J. P. Gautier, A. Gautier-Hion, D. Vallet, P. M. Sharp, and V. M. Hirsch. 1999. Simian immunodeficiency virus (SIV) from sun-tailed monkeys: evidence for host-dependent evolution of SIV within the *C. lhoesti* superspecies. J. Virol. **73**:7734–7744.

Bieniasz, P. D., A. Rethwilm, R. Pitman, M. D. Daniel, I. Chrystie, and M. O. McClure. 1995. A comparative study of higher primate foamy viruses, including a new virus from a Gorilla. Virology **207**:217–228.

Bininda-Emonds, O. R. P., J. L. Gittleman, and A. Purvis. 1999. Building large trees by combining phylogenetic information:

a complete phylogeny of the extant Carnivora (Mammalia). Biol. Rev. **74**:143–175.

Blewett, E. L., D. H. Black, N. W. Lerche, G. White, and R. Eberle. 2000. Simian foamy virus infections in a Baboon breeding colony. Virology **278**:183–193.

Boeke, J. D., and J. P. Stoye. 1997. Retrotransposons, endogenous retroviruses and the evolution of retroelements. Pp 343–436 in J. M. Coffin, S. H. Hughes, and H. E. Varmus, eds. Retroviruses. Cold Spring Harbor Laboratory Press, Plainview, N.Y.

Bohlman, M. C., S. P. Morzunov, J. Meissner, M. B. Taylor, K. Ishibashi, J. Rowe, S. Levis, D. Enria, and S. C. St Jeor. 2002. Analysis of *Hantavirus* genetic diversity in Argentina: S segment-derived phylogeny. J. Virol. **76**:3765–3773.

Bourhy, H., B. Kissi, and N. Tordo. 1999. Molecular diversity of the *Lyssavirus* genus. Virology **194**:70–81.

Bowen, M. D., C. J. Peters, and S. T. Nichol. 1997. Phylogenetic analysis of the Arenaviridae: patterns of virus evolution and evidence for cospeciation between arenaviruses and their rodent hosts. Mol. Phylogenet. Evol. **8**:301–316.

Brooks, D. R. 1981. Hennig's parasitological method—a proposed solution. Syst. Zool. **30**:229–249.

Brooks, D. R., and D. A. McLennan. 1993. Comparative-study of adaptive radiations with an example using parasitic flatworms (Platyhelminthes, Cercomeria). Am. Nat. **142**:755–778.

Broussard, S. R., A. G. Comuzzie, K. L. Leighton, M. Leland, E. M. Whitehead, and J. S. Allan. 1997. Characterisation of new Simian foamy viruses from African nonhuman Primates. Virology **237**:349–359.

Chan, S. Y., L. Ho, C. K. Ong, V. Chow, B. Drescher, M. Durst, J. Termeulen, L. Villa, J. Luande, and H. N. Mgaya. 1992. Molecular variants of human papillomavirus type 16 from four continents suggest ancient pandemic spread of the virus and its coevolution with humankind. J. Virol. **66**:2057–2066.

Charleston, M. A. 1998. Jungles: a new solution to the host/parasite phylogeny reconciliation problem. Math. Biosci. **149**:191–123.

Charleston, M. A., and R. D. M. Page. 2002. TreeMap v2.0. Application for Apple Macintosh.

Charleston, M. A., and D. L. Robertson. 2002. Preferential host switching by primate lentiviruses can account for phylogenetic similarity with the primate phylogeny. Syst. Biol. **51**:528–535.

Charrel, R. N., P. de Micco, and X. de Lamballerie. 1999. Phylogenetic analysis of GB viruses A and C: evidence for cospeciation between virus isolates and their primate hosts. J. Gen. Virol. **80**:2239–2335.

Chen, Z. L. A., D. L. Sodora, P. Telfer, P. Reed, A. Gettie, J. M. Kanu, R. F. Sadek, J. Yee, D. D. Ho, L. Zhang, and P. A. Marx. 1997. Human immunodeficiency virus type-2 (HIV-2) seroprevalence and characterisation of a distinct HIV-2 genetic subtype from the natural range of simian immunodeficiency virus-infected sooty mangabeys. J. Virol. **71**:3953–3960.

Childs, J. E., T. G. Ksiazek, C. F. Spiropoulou et al. (13 coauthors). 1994. Serologic and genetic identification of *Peromyscus maniculatus* as the primary rodent reservoir for a new *Hantavirus* in the southwestern United States. J. Infect. Dis. **169**:127–180.

Clayton, D. H., and Tompkins, D. M. 1994. Ectoparasite virulence is linked to mode of transmission. Proc. R. Soc. Lond. Ser. B **256**:211–217.

Coffin, J. M. 1992. Genetic diversity and evolution of retroviruses. Curr. Top. Microbiol. Immunol. **176**:143–164.

Coffin, J. M. 1996. Retroviridae: the viruses and their replication. Pp 763–841 *in* B. N. Fields, D. M. Knipe, P. M. Howley et al., eds. Fundamental virology, 3rd edition. Lippincott-Raven, Philadelphia.

Crandall, K. A. 1996. Multiple interspecies transmission of human and simian T-cell leukaemia/lymphotropic virus type I sequences. Mol. Biol. Evol. **13**:115–131.

Day, T. 2001. Parasite transmission modes and the evolution of virulence. Evolution **55**:2389–2400.

Dimcheff, D. E., S. V. Drovetski, M. Krishnan, and D. P. Mindell. 2000. Cospeciation and horizontal transmission of avian sarcoma and leukosis virus *gag* genes in Galliform birds. J. Virol. **74**:3984–3995.

Domingo, E., C. Escarmis, N. Sevilla, A. Moya, S. F. Elena, J. Quer, I. S. Novella, and J. J. Holland. 1996. Basic concepts in RNA virus evolution. FASEB J. **10**:859–864.

Dowling, A. P. G. 2002. Testing the accuracy of TreeMap and Brooks parsimony analysis of coevolutionary patterns using artificial associations. Cladistics **18**:416–435.

Engel, S. R., K. M. Hogan, J. F. Taylor, and S. K. Davis. 1998. Molecular systematics and paleobiogeography of the South American sigmodontine rodents. Mol. Biol. Evol. **15**:35–49.

Gao, F., E. Bailes, D. L. Robertson, Y. Chen, C. M. Rodenburg, S. F. Michael, L. B. Cummins, L. O. Arthur, M. Peeters, G. M. Shaw, P. M. Sharp, and B. H. Hahn. 1999. Origin of HIV-1 in the chimpanzee, *Pan troglodytes troglodytes*. Nature **397**:436–441.

Goodman, M., J. Czelusniak, G. W. Moore, A. E. Romero-Herrera, and G. Matsuda. 1979. Fitting the gene lineage to its species lineage: a parsimony strategy illustrated by cladograms constructed from globin sequences. Syst. Zool. **28**:132–168.

Hafner, M. S., and S. A. Nadler. 1990. Cospeciation in host-parasite assemblages—comparative-analysis of rates of evolution and timing of cospeciation events. Syst. Zool. **39**:192–204.

Hahn, B. H., G. M. Shaw, M. E. Taylor, R. R. Redfield, P. D. Markham, S. Z. Salahuddin, F. Wong-Staal, R. C. Gallo, E. S. Parks, and W. P. Parks. 1986. Genetic variation in HTLV-III/LAV over time with AIDS or at risk of AIDS. Science **232**:1548–1553.

Hjelle, B., S. W. Lee, W. M. Song, N. Torrez-Martinez, J. W. Song, R. Yanagihara, I. Gavrilovskaya, and E. R. Mackow. 1995. Molecular linkage of *Hantavirus* pulmonary syndrome to the white-footed mouse *Peromyscus leucopus*: genetic characteristics of the M genome of New York virus. J. Virol. **69**:8137–8141.

Huelsenbeck, J. P., B. Rannala, and Z. H. Yang. 1997. Statistical tests of host-parasite cospeciation. Evolution **51**:410–419.

Huelsenbeck, J. P., B. Rannala, and B. Larget. 2000. A Bayesian framework for the analysis of cospeciation. Evolution **54**:352–364.

Hugot, J. P., J. P. Gonzalez, and C. Denys. 2001. Evolution of the Old World Arenaviridae and their rodent hosts: generalised host transfer or association by descent? Infect. Genet. Evol. **1**:13–20.

Humphries, C. J., J. M. Cox, and E. S. Nielsen. 1986. *Notofagus* and its parasites: a cladistic approach to coevolution. Pp 55–76 *in* A. R. Stone and D. L. Hawksworth, eds. Coevolution and systematics. Clarendon Press, Oxford.

Jin, M. J., J. Rogers, J. E. Phillips-Conroy, J. S. Allan, R. C. Desrosiers, G. M. Shaw, P. M. Sharp, and B. H. Hahn. 1994. Infection of a yellow baboon with a simian immunodeficiency virus from African green monkeys: evidence for cross-species transmission in the wild. J. Virol. **68**:8454–8460.

Johnson, K. P., D. M. Drown, and D. H. Clayton. 2000. A data based parsimony method of cophylogenetic analysis. Zoological Scripta **30**:79–87.

Johnson, W. E., and J. M. Coffin. 1999. Constructing primate phylogenies from ancient retrovirus sequences. Proc. Natl. Acad. Sci. USA **96**:10254–10260.

Johnson, W. E., P. A. Webb, and G. Justines. 1973. Biology of Tacaribe-complex viruses. Pp 241–258 *in* F. Lehmann-Grube, ed. Lymphocytic choriomeningitis virus and other arenaviruses. Springer-Verlag, Berlin.

King, C. C., R. de Fries, S. R. Kolhekar, and R. Ahmed. 1990. In vivo selection of lymphocyte-tropic and macrophage-tropic variants of lymphocytic choriomeningitis virus during persistent infection. J. Virol. **64**:5611–5616.

Kissi, B., N. Tordo, and H. Bourhy. 1995. Genetic polymorphism in the Rabies virus nucleoprotein gene. Virology **209**:526–537.

Koralnik, I. J., E. Boeri, W. C. Saxinger, A. Lomonico, J. Fullen, A. Gessain, H. G., Guo, R. C. Gallo, P. Markham, V. Kalyanaraman et al. (17 co-authors). 1994. Phylogenetic associations of human and simian t-cell leukemia/lymphotropic virus type-I strains—evidence for interspecies transmission. J. Virol. **68**:2693–2707.

Legendre, P., Y. Desdevises, and E. Bazin. 2002. A statistical test for host-parasite coevolution. Syst. Biol. **51**:217–234.

Levis, S., S. P. Morzunov, J. E. Rowe, D. Enria, N. Pini, G. Calderon, M. Sabattini, and S. C. St Jeor. 1998. Genetic diversity and epidemiology of hantaviruses in Argentina. Infect. Genet. Evol. **177**:529–538.

Lipsitch, M., S. Siller, and M. A. Nowak. 1996. The evolution of virulence in pathogens with vertical and horizontal transmission. Evolution **50**:1729–1741.

McCormick, J. B., P. A. Webb, J. W. Krebs, K. M. Johnson, and E. S. Smith. 1987. A prospective study of the epidemiology and ecology of Lassa fever. J. Infect. Dis. **155**:437–444.

McGeoch, D. J., A. Dolan, and A. C. Ralph. 2000. Toward a comprehensive phylogeny for mammalian and avian herpesviruses. J. Virol. **74**:10401–10406.

McGeoch, D. J., S. Cook, A. Dolan, F. E. Jamieson, and E. A. R. Telford. 1995. Molecular phylogeny and evolutionary timescale for the family of mammalian herpesviruses. J. Mol. Biol. **247**:443–458.

Meertens, L., J. Rigoulet, P. Mauclere, M. van Beveren, G. M. Chen, O. Diop, G. Dubreuil, M-C. Georges-Goubot, J-L. Berthier, J. Lewis, and A. Gessain. 2001. Molecular and phylogenetic analyses of 16 novel Simian T-cell leukaemia virus type 1 from Africa: close relationship of STLY-1 from *Allenopithecus nigroviridis* to HTLV-1 subtype B strains. Virology **287**:275–285.

Messenger, S. L., I. J. Molineux, and J. J. Bull. 1999. Virulence evolution in a virus obeys a trade-off. Proc. R. Soc. Lond. Ser. B **266**:397–404.

Monroe, M. C., S. P. Morzunov, A. M. Johnson, M. D. Bowen, H. Artsob, T. Yates, C. J. Peters, P. E. Rollin, T. G. Ksiazek, and S. T. Nichol. 1999. Genetic diversity and distribution of *Peromyscus*-borne hantaviruses in North America. Emerg. Infect. Dis. **5**:75–86.

Morzunov, S. P., J. E. Rowe, T. G. Ksiazek, C. J. Peters, S. C. St. Jeor, and S. T. Nichol. 1998. Genetic analysis of the diversity and origin of hantaviruses in *Peromyscus leucopus* mice in North America. J. Virol. **72**:57–64.

Myers, P., B. Lundrigan, and P. K. Tucker. 1995. Molecular phylogenetics of the Oryzomyine rodents: the genus *Oligoryzomys*. Mol. Phylogenet. Evol. **4**:372–382.

Nadin-Davies, S. A., W. Huang, J. Armstrong, G. Allen Casey, C. Bahloul, N. Tordo, and A. I. Wandeler. 1997. Antigenic and genetic divergence of rabies viruses from bat species indigenous to Canada. Virus Res. **74**:139–156.

Nelson, G., and N. I. Platnick. 1981. Systematics and biogeography. Cladistics and vicariance. Columbia University Press, New York.

Nichol, S. T., C. F. Spiropoulou, S. P. Morzunov, P. E. Rollin, T. G. Ksiazek, and H. Feldman. 1993. Genetic identification of a *Hantavirus* associated with an outbreak of acute respiratory illness. Science **262**:914–917.

Page, R. D. M. 1988. Quantitative cladistic biogeography: constructing and comparing area cladograms. Syst. Zool. **37**:254–270.

Page, R. D. M. 1989. COMPONENT. Application for Apple Macintosh. (http://taxonomy.zoology.gla.ac.uk/rod/cpw.html).

Page, R. D. M. 1990. Component analysis—a valiant failure? Cladistics **6**:119–136.

Page, R. D. M. 1994. Parallel phylogenies—reconstructing the history of host-parasite assemblages. Cladistics **10**:155–173.

Page, R. D. M. 1995. TreeMap v1.0. Application for Apple Macintosh. (http://taxonomy.zoology.gla.ac.uk/rod/treemap.html).

Page, R. D. M., and M. A. Charleston. 1997. From gene to organismal phylogeny: reconciled trees and the gene tree species tree problem. Mol. Phylogenet. Evol. **7**:231–240.

Page, R. D. M., and M. A. Charleston. 1998. Trees within trees: phylogeny and historical associations. Trends Ecol. Evol. **13**:356–359.

Page, R. D. M., and M. A. Charleston. 2002. TreeMap versus BPA (again): a response to Dowling. Technical Reports in Taxonomy, 02–02. http://taxonomy.zoology.gla.ac.uk/publications/tech-reports/

Page, S. L., and M. Goodman. 2001. Catarrhine phylogeny: noncoding DNA evidence for a diphyletic origin of the mangabeys and for a human-chimpanzee clade. Mol. Phylogenet. Evol. **18**:14–25.

Payne, L. N., S. R. Brown, N. Bumstead, J. A. Howes, J. A. Frazier, and M. E. Thouless. 1991. A novel subgroup of exogenous avian leucosis virus in chickens. J. Gen. Virol. **72**:801–807.

Pecon-Slattery, J., G. Franchini, and A. Gessian. 1999. Genomic evolution, patterns of global dissemination, and interspecies transmission of human and simian T-cell leukaemia/lymphotropic viruses. Genome Res. **9**:525–540.

Peeters, M., and P. M. Sharp. 2000. Genetic diversity of HIV-1: the moving target. AIDS **14**:S129–S140.

Platnick, N. I., and G. Nelson. 1978. A method of analysis for historical biogeography. Syst. Zool. **27**:1–16.

Plyusnin, A., O. Vapalahti, and A. Vaheri. 1996. Hantaviruses: genome structure, expression and evolution. J. Gen. Virol. **77**:2677–2687.

Posada, D., and K. A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics **14**:817–818.

Ronquist, F. 1995. Reconstructing the history of host-parasite associations using generalised parsimony. Cladistics **11**:73–89.

Sanchez, A. J., K. D. Abbott, and S. T. Nichol. 2001. Genetic identification and characterisation of Limestone Canyon Virus, a unique *Peromyscus*-borne *Hantavirus*. Virology **286**:345–353.

Schmaljohn, C. S., and J. M. Dalrymple. 1983. Analysis of Hantaan virus RNA: evidence for a new genus of Bunyaviridae. Virology **131**:482–491.

Schweizer, M., and D. Neumann-Haefelin. 1995. Phylogenetic analysis of primate foamy viruses by comparison of *pol* sequences. Virology **207**:577–582.

Schweizer, M., H. Schleer, M. Pietrek, J. Liegibel, V. Falcone, and D. Neumann-Haefelin. 1999. Genetic stability of foamy

viruses: long term study in an African green monkey population. J. Virol, **73**:9256–9265.
Shadan, F. F., and L. P. Villarreal. 1993. Coevolution of persistently infecting small DNA viruses and their hosts linked to host-interactive regulatory domains. Proc. Natl. Acad. Sci. USA **90**:4117–4121.
Shimodaira, H., and M. Hasegawa. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. **16**:114–116.
Southern, P. J. 1996. Arenaviridae: the viruses and their replication. Pp 675–689 *in* B. N. Fields, D. M. Knipe, P. M. Howley et al., eds. Fundamental virology, 3rd edition. Lippincott-Raven, Philadelphia.
Swofford, D. W. 1998. PAUP* Phylogenetic Reconstruction Using Parsimony and other methods. Sinauer Associates, Sunderland, Mass.
Swofford, D. L., G. J. Olsen, P. J. Waddell, and D. M. Hillis. 1996. Phylogenetic inference. *in* D. M. Hillis, C. Morris, and B. K. Mable, eds. Molecular systematics. Sinauer Associates, Sunderland, Mass.
Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **24**:4876–4882.
Villarreal, L. P., V. R. DeFilippis, and K. A. Gottlieb. 2000. Acute and persistent life strategies and their relationship to emerging diseases. Virology **272**:1–6.
Wagner, R. R., and J. K. Rose. 1996. Rhabdoviridae: the viruses and their replication. Pp. 561–575 *in* B. N. Fields, D. M. Knipe, P. M. Howley et al., eds. Fundamental virology, 3rd edition. Lippincott-Raven, Philadelphia.
Wain-Hobson, S. 1996. Running the gamut of retroviral variation. Trends Microbiol. **4**:135–141.
Wang, H., K. Yoshimatsu, H. Ebihara, M. Ogino, K. Araki, H. Kariwa, Z. Wang, Z. Luo, D. Li, C. Hang, and J. Arikawa. 2000. Genetic diversity of hantaviruses isolated in China and characterisation of novel hantaviruses isolated from *Niviventer confucianus* and *Rattus rattus*. Virology **278**:332–345.
Wayne, R. K., E. Geffin, D. J. Girman, K. P. Koepfli, L. M. Lau, and C. R. Marshall. 1997. Molecular systematics of the Canidae. Syst. Biol. **46**:622–653.

Keith Crandall, Associate Editor

Accepted August 1, 2003