Hantavirus Evolution in Relation to Its Rodent and Insectivore Hosts: No Evidence for Codivergence

Cadhla Ramsden,* Edward C. Holmes,† and Michael A. Charleston‡

*Center for Infectious Disease Dynamics, Department of Biology, Mueller Laboratory, The Pennsylvania State University; †Fogarty International Center, National Institutes of Health, Bethesda, MD 20892, USA; and ‡School of Information Technologies, Sydney Bioinformatics, and Centre for Mathematical Biology, University of Sydney, New South Wales 2006, Australia

Hantaviruses are considered one of the best examples of a long-term association between RNA viruses and their hosts. Based on the appearance of strong host specificity, it has been suggested that hantaviruses cospeciated with the rodents and insectivores they infect since these mammals last shared a common ancestor, approximately 100 million years ago.

We tested this hypothesis of host–virus codivergence in two ways: 1) we used cophylogenetic reconciliation analysis to assess the fit of the virus tree onto that of the host and 2) we estimated the evolutionary rates and divergence times for the Hantavirus genus using a Bayesian Markov Chain Monte Carlo method and similarly compared these with those of their hosts. Our reconciliation analysis provided no evidence for a history of codivergence between hantaviruses and their hosts. Further, the divergence times for the Hantavirus genus were many orders of magnitude too recent to correspond with the timescale of their hosts’ speciation. We therefore propose that apparent similarities between the phylogenies of hantaviruses and their mammalian hosts are the result of a more recent history of preferential host switching and local adaptation. Based on the presence of clade-defining amino acids in all genomic segments, we propose that the patterns of amino acid replacement in these viruses are also compatible with a history of host-specific adaptation.

Introduction

Hantaviruses are single-stranded negative-sense RNA viruses whose genome consists of three segments. These segments are designated Small (S), Medium (M), and Large (L) and code for the nucleocapsid, the two envelope glycoproteins (G1 and G2), and the RNA-dependent RNA polymerase, respectively (Schmaljohn 1996). Unlike the rest of the family Bunyaviridae, hantaviruses are not vector borne but rather are transmitted directly or indirectly between hosts during aggressive interactions or through the inhalation of infectious aerosols released in urine and feces (Plyusnin and Morzunov 2001). Hantaviruses have historically been classified primarily as agents of disease in rodents belonging to the family Muridae, where they appear to cause persistent, asymptomatic infections (Johnson 2001). However, recent studies have found evidence that infected rodents may suffer decreased survivorship, particularly during winter (Calisher et al. 2005; Kallio et al. 2007). In contrast, many hantaviruses are highly pathogenic in humans. For example, infection with Old World hantaviruses can cause hemorrhagic fever with renal syndrome (HFRS) with mortality rates as high as 15%, whereas infection with the New World viruses can result in hantavirus pulmonary syndrome (HPS) with epidemic fatality rates as high as 50% (Peters et al. 1999; Zeier et al. 2005). Human cases of HPS or HFRS almost exclusively result from increased contact between humans and rodent excrement that can occur during periods of high rodent density. Notably, recent South American HPS outbreaks have included two confirmed and one suspected epidemic with human-to-human transmission, highlighting the potential for hantaviruses to evolve from a spillover zoonotic disease into an emerging human pathogen (Padula et al. 1998; Toro et al. 1998; Martinez et al. 2005).

Phylogenetic inference of the relationships between members of the genus Hantavirus has revealed three consistently well-defined clades, each associated with only one of the three subfamilies of Muroid rodents: Arvicolinae, Murinae, and Sigmodontinae (Plyusnin et al. 1996; Hughes and Friedman 2000; Plyusnin and Morzunov 2001; Jackson and Charleston 2004). Comparisons of the hantavirus phylogeny with that of the members of the Muridae family they infect have revealed strong topological correspondence between them (Hughes and Friedman 2000; Plyusnin and Morzunov 2001; Jackson and Charleston 2004). These topological similarities have been used repeatedly to assert that hantaviruses are one of the most robust examples of a long-standing history of codivergence between a host and a pathogen, potentially dating back to the split between rodents and insectivores (Hjelle, Lee et al. 1995; Plyusnin et al. 1996; Morzunov et al. 1998; Monroe et al. 1999; Vapalahti et al. 1999; Hughes and Friedman 2000; Plyusnin and Morzunov 2001; Jackson and Charleston 2004; Nemirov et al. 2004). However, cospeciation is not the only evolutionary process that can produce congruence between host and pathogen phylogenies. In particular, host switching followed by pathogen speciation can also generate congruence between trees, particularly when pathogens preferentially switch among closely related hosts (Page 1996; Charleston and Robertson 2002; de Vienne et al. 2007).

There are substantial inconsistencies between what we know of hantavirus biology and a history of codivergence with Muroid rodents. Numerous hantaviruses are capable of infecting multiple rodent species (e.g., Andes virus, Gonzalez et al. 2002; Dobrava virus, Sibold et al. 2001; Puumala virus, Dekonenko et al. 2003; Tula virus, Plyusnin et al. 1994), and cross-species transmission (host switching) events between rodents have occurred repeatedly (Levis et al. 1998; Morzunov et al. 1998; Monroe et al. 1999; Bohlm et al. 2002; Nemirov et al. 2002). In addition, hantaviruses have recently been isolated in a variety of insectivore species from a range of genera: (e.g., Anourusorex, Song, Baek, et al. 2007; Blarina, Arai et al. 2007; Sorex, Arai et al. 2008; Suncus, Song, Kang, et al. 2007), and initial phylogenetic analyses of these viruses have indicated...
that neither the shrew- nor the rodent-associated viruses are monophyletic (Arai et al. 2008). Critically, the presence of mixed rodent and insectivore hantavirus clades greatly undermines the argument that rodents and hantaviruses have codiverged, especially given that insectivores are more closely related to the mammalian order Carnivora (among others) than they are to rodents (Nikaido et al. 2001).

Finally, temporal congruence must exist between the divergences of hantaviruses and their hosts for the hypothesis of cospeciation between these groups to be validated (Page 1996). Previous inspections of the correspondence between the phylogenies of hantaviruses and rodents have used topology as the main criterion by which codivergence was determined, and temporal congruence was then assumed rather than established independently (Hughes and Friedman 2000; Sironen et al. 2001). Following the assumption of codivergence, rates of nucleotide substitution for hantaviruses were estimated to be in the range of $10^{-7}$ nucleotide substitutions per site per year (Hughes and Friedman 2000; Sironen et al. 2001), several orders of magnitude lower than those normally observed in RNA viruses that replicate with RNA-dependent RNA polymerase (Drake 1999; Jenkins et al. 2002; Hanada 2003). Further, under the extreme low long-term rates, nucleotide substitutions per site per year (subs/site/year), several polytomies (Hughes and Friedman 2000; Sironen et al. 2001) were observed rather than established independently. Following the assumption of both recent evolutionary change are in stark contrast to both recent estimates of the short-term rate of nucleotide substitution in rodent-associated hantaviruses ($10^{-2}$ to $10^{-3}$ subs/site/year) (Hughes and Friedman 2000; Sironen et al. 2001), and experimentally measured mutation frequencies ($10^{-3}$ mutations per kb, Sironen et al. 2008). Moreover, extensive genetic variation has been found within individual hosts infected with hantaviruses (Plyusnin et al. 1995, 1996a; Feuer et al. 1999) and hence similar to that observed within rapidly evolving RNA viruses. Taken together, these observations strongly suggest that hantaviruses are evolving substantially faster than predicted based on shared divergence times with their rodent hosts, highlighting the inconsistencies between the hantavirus biology and the hypothesis of codivergence.

In sum, there is great uncertainty regarding the history and timescale of the evolution of hantaviruses, which impacts our ability to predict the likelihood and scale of future host jumps. Here, we explicitly examine the evolutionary history of hantaviruses and their rodent/insectivore hosts for the first time, using rigorous statistical methods to stringently test the hypothesis of codivergence.

Materials and Methods
Phylogenetic Analyses
To test the level of congruence between hantavirus and host phylogenies, two sets of phylogenetic trees were inferred. All available hantavirus sequences corresponding to each of the three genomic segments were downloaded from GenBank. Sequences were aligned manually using Se-Al (v2.0a11 Carbon, http://tree.bio.ed.ac.uk/software/seal/) and examined for evidence of recombination using the RDP3 program (Martin et al. 2005). As each species group formed a monophyletic cluster (data not shown), a maximum of three randomly chosen representative sequences (supplementary table S1, Supplementary Material online) were used from each hantavirus species for phylogenetic analysis. Whenever possible, only full-length sequences were included, with untranslated regions excluded from the analysis (nucleotides 37–1322, 52–3465, and 37–6507 were included for the S, M, and L segments, respectively). For each data set, the uncorrected pairwise distances among sequences were calculated to determine the level of saturation at each codon position (Lewontin 1989) using PAUP* v4.0b10 (Swofford 2003). Saturation was observed (pairwise distances of 75–100%) at the third position of the S segment and the first and third codon positions of the M segment, whereas site saturation was not detected within the L segment. Saturated codon positions were removed from the S and M sequence alignments, resulting in final data sets of 65 taxa (886 nt, first and second codon positions) for the S segment, 57 taxa (1153 nt, second codon positions) for the M segment, and 42 taxa (6477 nt, all codon positions) for the L segment.

Phylogenetic relationships among the hantaviruses in each data set were estimated using a Bayesian Markov Chain Monte Carlo (MCMC) method implemented in BEAST v1.4.7 (Drummond and Rambaut 2007), using the general time reversible (GTR) + $\Gamma$ + I model of nucleotide substitution as determined by Modeltest v3.7 (Posada and Crandall 1998; we also attempted to use the CP + GTR + I model but did not achieve statistical convergence). This method of phylogenetic analysis was chosen as it returns rooted trees (essential for codivergence analyses) and provides an internal statistical measure of nodal support (posterior probabilities) through the analysis of millions of plausible trees. The Bayesian skyline coalescent model assuming an uncorrelated lognormal relaxed molecular clock was used in all cases. A minimum of two independent runs was performed for each data set with sampling every 1,000 generations. Each run was continued until the effective sample size of all parameters was greater than 200. All runs were combined using LogCombiner v1.4.7, and maximum clade credibility trees were determined using TreeAnnotator with a burn-in of 10% of the sampled trees (http://beast.bio.ed.ac.uk/Programs). As sequences for all hantaviruses were not available for each segment, the resultant topologies for each segment were manually combined into a single composite phylogeny for the analysis. Prior to creating this composite phylogeny, the topological congruence of the three segment trees was compared using the Shimodaira–Hasegawa test (Shimodaira and Hasegawa 1999) available in PAUP* (Swofford 2003). To directly compare the evolutionary history of each segment, only those taxa that were present in all three phylogenies were retained. Importantly, the three segment trees were not statistically different in topology, with $P$ values ranging from 0.2 to 1.0.

The rodent and insectivore species included in this phylogenetic analysis were those considered to be the reservoir hosts for all hantaviruses found in any of the data sets (Nemirov et al. 2004; supplementary table S2, Supplementary Material online). All available full-length cytochrome
b (cty b) gene sequences (the most widely available mitochondrial gene) corresponding to these hosts were downloaded from GenBank and manually aligned (n = 95, 1143 bp). A relaxed clock Bayesian phylogeny was reconstructed as above using a single representative taxon for each species (as each species included formed a monophyletic group, data not shown). However, in this case, both the Bayesian skyline and Yule (birth–death) models were used as coalescent priors. Cyt b sequences were not available for two species: Necromys benefactus and Bandicota indica, hosts of Maciel and Thailand viruses, respectively. Therefore, these taxa were manually added to the rodent phylogeny based on their position in the Tree of Life (N. benefactus, http://tolweb.org/tree/phylogeny.html) or from the literature (B. indica, Castle et al. 2004). As the orders Rodentia and Insectivora are not closely related within Eutheria, representatives of several other mammals were included to aid in the resolution of the tree (Nannospalax ehrenbergi, Brachyphylla cavernarum, and Canis lupus).

Phylogenetic Reconciliation

To determine the level of congruence between the host and hantavirus phylogenies, we used the program TreeMap (v2.0, http://www.it.usyd.edu.au/~mcharles/), the most intuitive and statistically rigorous of the methods that have been developed for this purpose (the differences between methods have been reviewed elsewhere, see Dowling 2002; Jackson and Charleston 2004; Ricklefs et al. 2004; Stevens 2004 as examples). TreeMap provides multiple solutions, whereby the nodes of the virus phylogeny are mapped onto the associated nodes of the host tree (assumed to be fixed) while allowing for different potentially optimal (POpt) combinations of host switches, duplications, losses, and codivergence events (CEs). These maps attempt to explain both the observed virus phylogeny and the distribution of viruses on the host lineages in the most parsimonious manner (Charleston 1998; Jackson and Charleston 2004). For our analysis, the POpt solutions were found by weighing both the numbers of CEs and non-codivergence events (NCEs) required to reconcile the host and virus trees whenever possible. All reconciliation analyses were repeated with decreasing constraints on the acceptable solutions until all solutions had been found, or the analyses could not be completed.

Given that no information was available regarding the true value of each event cost (e.g., duplication and codivergence), we chose not to consider these values when calculating the total cost of each solution found (Charleston 1998). Instead, we considered all POpt solutions as hypothetical scenarios that may explain the observed phylogenetic patterns of the virus as well as the host–virus associations. Significance testing was performed in TreeMap by creating 1,000 viral trees with randomized branches and mapping these random trees onto the fixed host tree. The proportion of these reconciliations with equal or fewer numbers of CEs or the same or greater numbers of CEs compared with the “real” viral tree was then calculated. The null hypothesis was that the level of congruence that exists is no greater than that expected between randomly generated trees, with a 5% level of rejection.

Due to the computational intractability of cophylogenetic mapping (Libeskind-Hadas and Charleston 2008) and corresponding limitations inherent in TreeMap, the complexity of the host and virus phylogenies was reduced as much as possible prior to completing the full phylogenetic reconciliation analysis. We condensed all individuals from the same host and virus species to a single representative taxon and further reduced any pairs of sister taxa that mapped to the same associated or paired-associated taxa on the host/virus phylogeny to a single representative branch. This yields a considerable computational gain while creating a slight bias in the direction of nonsignificance. We then analyzed complete clade-to-clade matches separately (Arvicolinae rodents and their viruses, Murinae rodents and their viruses) and reduced these clades to a single branch in both the host and the virus phylogenies for the overall analysis (supplementary fig. S1, Supplementary Material online). The Sigmodontinae clade was considered to be a special case, as two alternate host topologies must be considered (see Results). To be as conservative as possible, we first performed the phylogenetic reconciliation analysis on the topology most likely to result in significant congruence between the host and the virus trees (the Sigmodontinae subfamily as monophyletic). As this analysis did not result in significance congruence, the less-optimistic version (paraphyletic Sigmodontinae) of the host topology also cannot result in congruence. Due to the complexity of the Sigmodontinae data set, we were only able to complete the significance test with 100 trials using the number of CEs (see Results); however, the results of these random trials were so far from significant that further tests were unnecessary. Once the Arvicolinae, Murinae, and Sigmodontinae (monophyletic) clades were reduced to single branches, we performed phylogenetic reconciliation using all remaining host–virus associations (supplementary fig. S1, Supplemental Material online). All POpt solutions contained in any reconciliation analysis of the complete, unmodified phylogenies would contain maps that are also found in the condensed virus/host clade analysis. Therefore, analyzing these complete paired clades separately will have no impact on the overall result.

Evolutionary Rates and Dates

We also used the Bayesian MCMC approach available in BEAST v1.4.7 to estimate the long-term rates of nucleotide substitution and divergence times (i.e., time to the most recent common ancestor—TMRCA) in the genus Hantavirus for both the S and the M segments. New data sets were constructed of sequences containing a minimum of 275 nt and for which the year-of-sampling was known (dates ranged from 1983 to 2006). Only rodent-associated viruses were included in the analysis, resulting in data sets of 164 taxa spanning 28 years for the S segment (868 nt, first and second codon positions only) and 38 taxa spanning 23 years for the M segment (1146 nt, second codon position only). The BEAST analysis was undertaken in the manner outlined above for the complete hantavirus phylogenies, with uncertainty in all estimates reflected in the 95% high-probability density (HPD) intervals. To determine whether these
tip-dated hantavirus sequences contained sufficient signal to estimate evolutionary rates and divergence times, we randomized the date–sequence relationships in our data sets and repeated the BEAST analysis to test the strength of the deeper phylogenetic signal in the data.

Pattern of Amino Acid Replacements

To look for evidence of host-specific adaptation, we identified amino acids that uniquely defined the major hantavirus lineages by tracing unambiguous changes along the branches leading to each of the following virus clades: Arvicolinae, Murinae, Sigmodontinae, ingroup shrew viruses, and Thottapalayam (outgroup shrew virus) of our previously constructed phylogenies using MacClade 4.08 (Maddison DR and Maddison WP 2005). This was done separately for the S, M, and L segment phylogenies. Each amino acid identified as potentially clade defining was then verified by examining the amino acid present at that position in each sequence of additional alignments for the S, M, and L segments, constructed using all available hantavirus sequences of length at least 400 nt from GenBank. We then classified these amino acid changes as conservative or radical according to Hanada et al. (2007). To ensure a conservative approach, we considered only those amino acids that were invariant within each clade, yet different between clades. No amino acid positions were considered in regions containing low levels of sequence similarity.

Results

Phylogenetic Analysis

Hantaviruses

The rooted phylogenetic trees reconstructed using the Bayesian MCMC method in BEAST (Drummond and Rambaut 2007) resulted in nearly identical topologies between all three hantavirus genomic segments (fig. 1). In particular, the three primary clades of the rodent-associated viruses corresponding to the three rodent subfamilies were recovered in all phylogenies with high nodal support in all cases (Bayesian posterior probabilities, BPP = 1.0). Critically, neither the shrew- nor the rodent-associated viruses were monophyletic in any of the recovered phylogenies (fig. 1). Instead, the shrew viruses formed either 2 or 3 distinct clades, depending on the segment and the taxa included (fig. 1). Thottapalayam virus has been used as an outgroup to the rodent-associated hantaviruses in previous studies, and the rooted phylogenies returned by BEAST in our analyses support its position as basal to the rest of our sampled taxa (BPP = 1.0). Cao Bang, Jemez Springs, and Ash River consistently grouped together (BPP = 0.98); however, the position of Tanganya alternated between grouping with these three hantaviruses (S segment, BPP = 1.0) or as the sister clade to Thottapalayam virus (L segment, BPP = 0.50; supplementary fig. S1, Supplementary Material online). Camp Ripley sequences were only present in the L segment phylogeny, where they fell outside the primary shrew hantavirus clade (BPP = 0.93). Arai et al. (2008) found strong support for the presence of the ingroup shrews as a sister clade to the Murinae-associated rodent hantaviruses, a position also recovered in all our phylogenetic trees (BPP = 0.51, 0.68, and 0.99). The position of all major groups was consistent between the L segment trees created using either the complete segment or only the sequence region that was present in the majority of taxa (data not shown).

Hosts

The rooted phylogenetic trees recovered from both the Yule and Bayesian Skyline models were identical with...
respect to topology, with the shrew and rodent species forming distinct monophyletic groups (BPP = 0.94 and 0.88, respectively). These clades were separated by lineages corresponding to the nonrodent, nonshrew mammals included as background taxa in the analysis, which were then removed from the phylogeny prior to the cophylogenetic reconciliation analysis. Both the Murinae and Arvicolinae subfamilies of rodents formed monophyletic groups (BPP = 1.0), with the Murinae clade occupying the basal position among rodents. However, the Sigmodontinae subfamily formed two distinct clades: the larger clade contains the predominantly South American sigmodontids (BPP = 1.0), whereas the smaller and more basal clade included the exclusively North American Sigmodontinae (Reithrodontomy and Peromyscus, BPP = 1.0). Interestingly, these two clades do not occupy a sister relationship to each other; rather, the Arvicolinae subfamily forms the sister clade to the South American Sigmodontinae (fig. 2). The relationship between the South American Sigmodontinae and the Arvicolinae and between these clades and the North American Sigmodontinae, both topologies were considered in our reconciliation analysis.

Phylogenetic Reconciliation

To test the extent of congruence between the host and the virus trees, composite trees were created by combining the three hantavirus segment trees and adding those taxa manually to the host phylogeny for which cyt *b* sequences were not available (fig. 2). We first analyzed the congruence between each of the three rodent subfamily clades and their associated virus clades separately. Strikingly, none of these clades were any more congruent than could be expected given a randomly generated virus tree (table 1). Two POpt solutions were found for both the Arvicolinae and the Murinae clades, containing no more than 8 and 6 CEs (4 and 3 codivergent nodes), respectively. For both clades, the genuine level of congruence found in the POpt solutions was seen in greater than 5% of the 1,000 random viral topologies mapped on to the corresponding host clades, indicating a lack of congruence between these groups (table 1). Similarly, nonsignificant *P* values were returned when the number of NCEs was tested, although these *P* values were substantially lower in both cases (table 1). The cophylogenetic reconciliation of the Sigmodontinae clade proved to be more difficult, due to the computational complexity of this large group of viruses. To reduce the computational burden, solutions could only be found by increasing the
constraints required for a random viral to be included as an “acceptable” solution tree (an additional constraint of a maximum of three host switches), which yielded a more conservative test of significance. This highly constrained test did not yield significantly small numbers of acceptable randomized viral trees ($P = 0.3 \pm 0.05$); therefore, further random trials were not necessary.

Finally, we tested the deeper levels of congruence between our sampled hantaviruses and their rodent and shrew hosts by collapsing the previously analyzed clades to single branches. This resulted in a topology where the shrew viruses and hosts remained expanded while the Sigmodontinae, Arvicoline, and Murinae were collapsed (supplementary fig. S1, Supplementary Material online). The reconciliation analysis of this composite tree resulted in 13 POpt solutions, none of which contained more than five codivergent nodes (table 1). When significance was measured using the maximum number of CEs, all 1,000 random viral trees could be mapped onto the host tree with at least 10 CE trees (table 1). Similarly, a $P$ value of 0.987 was obtained when the minimum number of NCEs was tested (table 1). Because the L segment phylogeny returned slightly different groupings of shrew-associated viruses from those of the M and S segments (fig. 1), the cophylogenetic reconciliation analysis was also done using the L segment virus topology. The results of this analysis also showed no congruence between host and virus phylogenies ($P = 0.998 \pm 0.002$ and $0.997 \pm 0.003$, using CEs and NCEs, respectively).

## Discussion

If hantaviruses have been codiverging with their rodent and insectivore hosts, there should be strong topological similarities between the host and the virus phylogenies. However, our cophylogenetic reconciliation analyses revealed that these topologies are no more similar than would be expected by chance alone, at both large and small (clade specific) scales. Moreover, age estimates for the sampled hantavirus genetic diversity must be broadly similar to those estimated for their rodent and insectivore hosts in order for the divergence times of these clades to be congruent (Page 1996; de Vienne et al. 2007). Yet, based on our estimates of TMRCA for the rodent viruses alone, this does not seem plausible.

The case for a history of codivergence between hantaviruses and their hosts has rested primarily on the observation that those viruses associated with rodents fall into three main clades that correspond to the rodent subfamilies they infect: Arvicoline, Murinae, and Sigmodontinae (Plyusnin et al. 1996; Morzunov et al. 1998; Hughes and Friedman 2000; Plyusnin and Morzunov 2001; Nemirov...
et al. 2004). However, the recent discovery of distinct, shrew-associated hantaviruses (Song, Baek, et al. 2007; Yadav et al. 2007; Arai et al. 2008) reveals that rodent hantaviruses are not monophyletic but instead share a mixed evolutionary history with those infecting shrews (Arai et al. 2008, fig. 2). Furthermore, at least one host-switching event between rodents and shrews is needed to explain the current host–virus associations in context of the estimated phylogenetic relationships (fig. 2). A history of codivergence between hosts and hantaviruses since the split between rodents and shrews has previously been hypothesized (Hughes and Friedman 2000). However, this scenario cannot be validated even if the rodent and shrew viruses are forced to become reciprocally monophyletic, given that insectivores share a more recent common ancestor with cats, dogs, and whales than they do with rodents (Nikaido et al. 2001). It is therefore extremely unlikely that the common ancestor of hantaviruses infected the shared common ancestor of rodents and shrews, which would have existed more than 100 Ma (Nikaido et al. 2001).

Although there is no evidence for a long-term history of codivergence between hantaviruses and their hosts, there is some congruence (although not significant) between host and viral topologies within the major clades (table 2).

### Table 2

Clade-Defining and Unique Amino Acids for the Five Major Clades of the Rodent- and Shrew-Borne Hantaviruses and for the Small, Medium, and Large Genomic Segments

<table>
<thead>
<tr>
<th>Segment</th>
<th>Position</th>
<th>Arvicolinae</th>
<th>Murinae</th>
<th>Sigmodontinae</th>
<th>Ingroup Shrews</th>
<th>Thottapalayam</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small</td>
<td>90</td>
<td>D</td>
<td>G&lt;sup&gt;a&lt;/sup&gt;</td>
<td>D</td>
<td>D</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>112</td>
<td>I</td>
<td>I</td>
<td>L</td>
<td>L</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>242</td>
<td>F</td>
<td>W</td>
<td>F</td>
<td>M</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>248</td>
<td>P</td>
<td>K</td>
<td>P</td>
<td>E</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>E</td>
<td>D</td>
<td>E</td>
<td>Gap</td>
<td>—</td>
</tr>
<tr>
<td>Medium</td>
<td>164</td>
<td>K</td>
<td>K</td>
<td>R</td>
<td>K</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>167</td>
<td>L</td>
<td>L</td>
<td>M</td>
<td>I</td>
<td>L</td>
</tr>
<tr>
<td></td>
<td>214</td>
<td>T</td>
<td>S</td>
<td>T</td>
<td>T</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>337</td>
<td>T</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>415</td>
<td>S</td>
<td>T</td>
<td>S</td>
<td>S</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>550</td>
<td>Y</td>
<td>F&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Y</td>
<td>Y</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>588</td>
<td>L</td>
<td>F</td>
<td>M</td>
<td>F</td>
<td>L</td>
</tr>
<tr>
<td></td>
<td>857</td>
<td>Q</td>
<td>S</td>
<td>Q</td>
<td>Q</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>897</td>
<td>G</td>
<td>K&lt;sup&gt;a&lt;/sup&gt;</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>955</td>
<td>L</td>
<td>F&lt;sup&gt;c&lt;/sup&gt;</td>
<td>L</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td></td>
<td>1006</td>
<td>N</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>1081</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>—</td>
</tr>
<tr>
<td>Large</td>
<td>22</td>
<td>V</td>
<td>V</td>
<td>L</td>
<td>—</td>
<td>I</td>
</tr>
<tr>
<td></td>
<td>187</td>
<td>T&lt;sup&gt;a&lt;/sup&gt;</td>
<td>V</td>
<td>V</td>
<td>—</td>
<td>V</td>
</tr>
<tr>
<td></td>
<td>206</td>
<td>A</td>
<td>A</td>
<td>S&lt;sup&gt;c&lt;/sup&gt;</td>
<td>—</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>308</td>
<td>A</td>
<td>L</td>
<td>S</td>
<td>—</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>374</td>
<td>N</td>
<td>T</td>
<td>G</td>
<td>—</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>379</td>
<td>D</td>
<td>R</td>
<td>N</td>
<td>—</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>385</td>
<td>L</td>
<td>M</td>
<td>I</td>
<td>—</td>
<td>I</td>
</tr>
<tr>
<td></td>
<td>472</td>
<td>T</td>
<td>T</td>
<td>A</td>
<td>—</td>
<td>V</td>
</tr>
<tr>
<td></td>
<td>522</td>
<td>A</td>
<td>A</td>
<td>V</td>
<td>—</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>528</td>
<td>F</td>
<td>F</td>
<td>X&lt;sup&gt;c&lt;/sup&gt;</td>
<td>—</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>607</td>
<td>A</td>
<td>A</td>
<td>X&lt;sup&gt;c&lt;/sup&gt;</td>
<td>—</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>693</td>
<td>Q</td>
<td>Q</td>
<td>H&lt;sup&gt;c&lt;/sup&gt;</td>
<td>—</td>
<td>Q</td>
</tr>
<tr>
<td></td>
<td>867</td>
<td>L</td>
<td>L</td>
<td>F&lt;sup&gt;c&lt;/sup&gt;</td>
<td>—</td>
<td>L</td>
</tr>
<tr>
<td></td>
<td>930</td>
<td>E</td>
<td>E</td>
<td>D&lt;sup&gt;a&lt;/sup&gt;</td>
<td>—</td>
<td>E</td>
</tr>
<tr>
<td></td>
<td>963</td>
<td>L</td>
<td>M</td>
<td>F</td>
<td>—</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>1048</td>
<td>K</td>
<td>G</td>
<td>G</td>
<td>—</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>1055</td>
<td>I</td>
<td>V</td>
<td>V</td>
<td>—</td>
<td>V</td>
</tr>
<tr>
<td></td>
<td>1261</td>
<td>R</td>
<td>H</td>
<td>K</td>
<td>—</td>
<td>R</td>
</tr>
<tr>
<td></td>
<td>1468</td>
<td>R</td>
<td>K</td>
<td>K</td>
<td>—</td>
<td>K</td>
</tr>
<tr>
<td></td>
<td>1483</td>
<td>Q</td>
<td>Q</td>
<td>H</td>
<td>—</td>
<td>K</td>
</tr>
<tr>
<td></td>
<td>1530</td>
<td>H</td>
<td>F</td>
<td>F</td>
<td>—</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>1569</td>
<td>S</td>
<td>S</td>
<td>A</td>
<td>—</td>
<td>T</td>
</tr>
<tr>
<td></td>
<td>1578</td>
<td>Q</td>
<td>Q</td>
<td>E</td>
<td>—</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>1775</td>
<td>K</td>
<td>K</td>
<td>Q</td>
<td>—</td>
<td>T</td>
</tr>
<tr>
<td></td>
<td>1847</td>
<td>M&lt;sup&gt;c&lt;/sup&gt;</td>
<td>I</td>
<td>I</td>
<td>—</td>
<td>I</td>
</tr>
<tr>
<td></td>
<td>1854</td>
<td>N&lt;sup&gt;c&lt;/sup&gt;</td>
<td>S</td>
<td>S</td>
<td>—</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>1993</td>
<td>D</td>
<td>D</td>
<td>A&lt;sup&gt;a&lt;/sup&gt;&lt;sup&gt;b&lt;/sup&gt;</td>
<td>—</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>2114</td>
<td>I</td>
<td>V</td>
<td>V</td>
<td>—</td>
<td>L</td>
</tr>
</tbody>
</table>

<sup>a</sup> An amino acid difference in only one major clade, all other clades have the same amino acid.

<sup>b</sup> A radical amino acid change that involves a change in both charge and size.

<sup>c</sup> A conservative amino acid change (same charge and size).

<sup>d</sup> A radical amino acid change involving charge only.
**Sigmodontinae** clade but were unable to differentiate between codivergence and preferential host switching. Indeed, viruses within the **Sigmodontinae** clade do exhibit some patterns consistent with codivergence. For example, *Peromyscus maniculatus* and *Peromyscus leucopus* both exhibit metapopulation structure (with corresponding mitochondrial DNA divergence) across the United States (Morzunov et al. 1998; Plyusnin and Morzunov 2001). Their associated viruses, Sin Nombre and Blue River viruses, respectively, mimic this genetic clustering (Plyusnin and Morzunov 2001). However, the genetic differences between viruses corresponding to different host populations of the same species could also be explained through preferential host switching (Nemirov et al. 2004; de Vienne et al. 2007).

The evolutionary rates and TMRCA estimated for the Hantavirus genus are also incompatible with a history of codivergence between these viruses and their hosts. When the divergence times of rodent species have been used to calibrate the evolutionary rates of the rodent hantaviruses, a rate that is anomalously low for RNA viruses was obtained ($\sim 10^{-7}$ subs/site/year; Hughes and Friedman 2000; Sironen et al. 2001). Assuming a history of codivergence, we would expect that the evolutionary rates estimated in this manner would broadly agree with those estimated independently from nucleotide sequence data. However, both the short-term substitution rates estimated previously ($10^{-2}$ to $10^{-3}$ subs/site/year, Ramsden et al. 2008) and the longer term rates estimated here ($6.76 \times 10^{-4}$ subs/site/year) diverge substantially from host-dependent estimations. Similar discordance exists between the divergence times estimated for the rodent subfamilies and our estimates for the associated viruses: the Murinae subfamily of rodents is thought to have diverged $\sim 12$ Ma, the Arvicolinae $\sim 18$ Ma, and the Sigmodontinae $\sim 19$ Ma (Steppan et al. 2004). These estimates are in stark contrast to our much more recent divergence estimates for the associated hantaviruses. However, caution must be used when interpreting our TMRCA results. Using a Bayesian MCMC method to estimate the divergence dates of viral lineages results only in estimates of the age of the sampled genetic diversity. Therefore, this method is extremely sensitive to the exclusion of extant or extinct basal viral lineages in the data set, a factor that is certainly present with hantaviruses due to the limited data available. The estimation of divergence dates from sequence data for other viruses has resulted in the estimation of TMRCA values that appear too recent given their often high prevalence in animal populations (Sharp et al. 2000; Holmes 2003). However, even if we assume that the TMRCAs we estimated for the major rodent-associated virus clades are an order of magnitude too recent, these viruses must have diverged millions of years too late to have coevolved with their hosts.

Our cophylogenetic reconciliation analysis suggests that hantaviruses have not codiverged with rodents or insectivores. However, it is clear that hantaviruses do segregate into distinct, well-supported clades that are associated with specific host types (e.g., rodent subfamilies, fig. 2). Curiously, geographical associations alone cannot explain these associations, as the majority of clades contain viruses from disparate geographical regions (fig. 2; Bohlman et al. 2002; Nemirov et al. 2004). Without evidence for a history of cospeciation between hantaviruses and their hosts, an alternative hypothesis is needed to explain the observed pattern of hantavirus infection worldwide. Preferential host switching, governed by a combination of geographical proximity and adaptation to specific host types, is a good candidate explanation.

Switching between closely related hosts is widely known to occur in hantaviruses and could contribute to a cophylogenetic pattern that resembles codivergence (Charleston and Robertson 2002; Jackson and Charleston 2004; de Vienne et al. 2007). Not only has host switching been documented between closely related rodents (e.g., Dobrava infects both Apodemus agrarius and Apodemus flavicollis, Sibold et al. 2001), but it also occurs between hosts from different subfamilies. In addition to their reservoir hosts, Sin Nombre virus has been detected in *Mus musculus* (Nichol 1999), Four Corners virus in *Reithrodontomys mexicanus* (Hjelle, Anderson, et al. 1995), and El Moro Canyon virus in *P. maniculatus* (Rawlings et al. 1996). Furthermore, hantaviruses that are not phylogenetically related have been found to infect the same or similar host species, a phenomenon most likely related to host switching. For example, Khabarovsk virus infects *Microtus fortis* but is more closely related to viruses infecting *Clethrionomys* sp. than to other *Microtus* viruses (Horling et al. 1996). Similarly, Saarema virus is more closely related to Dobrava than to Hantaan, although both Saarema and Hantaan are endemic to *A. agrarius* (Nemirov et al. 2002). Finally, Laguna Negra, Rio Mearim, Anajatuba, and Rio Mamore viruses infect a range of hosts (*Calomys laucha*, *Holochilus sciureus*, *Oligoryzomys meseki*, and *Oligoryzomys microtus*, respectively); however, they cluster into a Brazilian clade (Anajatuba and Rio Mearim) and a Paraguay/Bolivia clade (Laguna Negra and Rio Mamore), representing multiple host switches (Bohlman et al. 2002).

Neither geographical proximity nor phylogenetic relatedness appears sufficient to predict the likelihood of a successful host switch. Therefore, additional factors must contribute to the successful establishment of a hantavirus in a new host. Once contact with a new host species has been made, a range of both host and viral features may contribute to the likelihood of a cross-species transmission event, including variations in host susceptibility and the ability of a virus to infect and reproduce in a new cell type. To this end, we identified unique, clade-defining amino acids that may be evidence of virus adaptation to a specific host type, indicative of a more recent evolutionary history than suggested under codivergence. The Murinae subclade contained more unique amino acids in the S and M segments than any other group (table 2). Interestingly, the Murinae-associated viruses form a sister clade to the ingroup shrew viruses and are most likely to have been involved in the host-switching event with these shrew viruses required to generate concordance between the virus topology and the host associations (fig. 2). The S segment codes for the nucleocapsid protein, which is known to contain the majority of the hantavirus epitopes, as well as to be involved in antibody binding (Yamada et al. 1995; Yoshimatsu et al. 1996; Tischler et al. 2008), making it a likely candidate for clade-specific adaptation. The M segment encodes the two glycoproteins and is expected to contain epitope regions...
that are highly variable both within and between clades (Plyusnin et al. 1996b; Horling and Lundkvist 1997). As these two glycoproteins are cleaved during translation by host proteases, the regions involved in complexing with the host cell machinery would need to rapidly adapt to a new host (Shi and Elliott 2002). In addition, regions of the glycoproteins are thought to be involved in binding to the cell receptors used for host cell entry (Larson et al. 2005). Therefore, these regions may evolve to become more efficient mediators of cell entry once exposed to a new host cell type. The L segment contained many more clade-defining amino acids than either the S or the M segment, and in contrast to the other two segments, most of these were unique to the Sigmodontinae-associated viruses (table 2). The L protein contains a single open reading frame and has transcriptase, replicase, and endonuclease functions (Kukkonen et al. 2005). This protein (along with the nucleocapsid) is critical for all stages of viral RNA synthesis and was expected to be under strong functional constraints. What is behind the high proportion of unique amino acids among the Sigmodontinae viruses is therefore unclear, although one possible explanation is that the jump from Old to New World rodents required a period of adaptive evolution in the polymerase so that the virus could replicate efficiently within the cells of the new host. A more detailed examination of the amino acid variation and the functions performed by regions within the L protein would help to illuminate the adaptation of hantaviruses to these New World rodents.

This study demonstrates that the biology and evolutionary history of the Hantavirus genus is incompatible with the hypothesis of codivergence with its hosts. However, a full understanding of hantavirus evolution is hampered by lack of a understanding of the number and range of host taxa infected by these viruses. Preliminary work has suggested that hantaviruses may infect a wide variety of mammals, including moose, cats, muskrat, and bats; however, these (and other) groups have not been rigorously sampled (Zier et al. 2005). Therefore, the impact of an ascertainment bias must be carefully considered when weighing any hypothesis concerning our understanding of the evolutionary history or future emergences of these viruses. Hantaviruses are a rapidly emerging zoonotic disease, and an understanding of their evolutionary history or future emergences of these viruses is therefore unclear, although one possible explanation is that the jump from Old to New World rodents required a period of adaptive evolution in the polymerase so that the virus could replicate efficiently within the cells of the new host. A more detailed examination of the amino acid variation and the functions performed by regions within the L protein would help to illuminate the adaptation of hantaviruses to these New World rodents.

This study demonstrates that the biology and evolutionary history of the Hantavirus genus is incompatible with the hypothesis of codivergence with its hosts. However, a full understanding of hantavirus evolution is hampered by lack of a understanding of the number and range of host taxa infected by these viruses. Preliminary work has suggested that hantaviruses may infect a wide variety of mammals, including moose, cats, muskrat, and bats; however, these (and other) groups have not been rigorously sampled (Zier et al. 2005). Therefore, the impact of an ascertainment bias must be carefully considered when weighing any hypothesis concerning our understanding of the evolutionary history or future emergences of these viruses. Hantaviruses are a rapidly emerging zoonotic disease, and future research must focus on elucidating those factors critical for effective cross-species transmission.

Supplementary Material

Supplementary tables and figure are available at Molecular Biology and Evolution online (http://www.mbe. oxfordjournals.org/).

Funding

Natural Sciences and Engineering Research Council (Canada) to C.R.

Acknowledgments

The authors would like to thank T. Avšič-Županc for providing date-of-sampling information for her sequences, as well as Richard Burhans for the use of his computer. We also thank F. L. Melo, P. M. de A. Zanotto, and S. Duffy for the invaluable insights and discussions on hantavirus evolution.

Literature Cited


Lewontin RC. 1989. Inferring the number of evolutionary events from DNA coding sequence differences. Mol Biol Evol. 6:15–32.


Nichol ST. 1999. Factors in the emergence and control of rodent-borne viral diseases (hantaviral and arenaviral diseases). Elsevier SAS.


Aoife McLysaght, Associate Editor

Accepted October 9, 2008