2010

Family level phylogenies reveal modes of macroevolution in RNA viruses

Andrew Kitchen
Pennsylvania State University, University Park

Laura Shackelton
Pennsylvania State University, University Park

Edward Holmes
Pennsylvania State University, University Park, edward.holmes@sydney.edu.au

Family level phylogenies reveal modes of macroevolution in RNA viruses

Andrew Kitchen[a,1], Laura A. Shackelton[a,b,1], and Edward C. Holmes[a,c,2]

[a]Center for Infectious Disease Dynamics, Department of Biology, Mueller Laboratory, Pennsylvania State University, University Park, PA 16802; [b]Global Health Discovery, Bill and Melinda Gates Foundation, Seattle, WA 98105; and [c]Fogarty International Center, National Institutes of Health, Bethesda, MD 20892

Edited by Peter Palese, Mount Sinai School of Medicine, New York, NY, and approved November 22, 2010 (received for review July 27, 2010)

Despite advances in understanding the patterns and processes of microevolution in RNA viruses, little is known about the determinants of viral diversification at the macroevolutionary scale. In particular, the processes by which viral lineages assigned as different "species" are generated remain largely uncharacterized. To address this issue, we use a robust phylogenetic approach to analyze patterns of lineage diversification in five representative families of RNA viruses. We ask whether the process of lineage diversification primarily occurs when viruses infect new host species, either through cross-species transmission or codivergence, and which are defined here as analogous to allopatric speciation in animals, or by acquiring new niches within the same host species, analogous to sympatric speciation. By mapping probable primary host species onto family level viral phylogenies, we reveal a strong clustering among viral lineages that infect groups of closely related host species. Although this is consistent with lineage diversification within individual hosts, we argue that this pattern more likely represents strong biases in our knowledge of viral biodiversity, because we also find that better-sampled human viruses rarely cluster together. Hence, although closely related viruses tend to infect related host species, it is unlikely that they often infect the same host species, such that evolutionary constraints hinder lineage diversification within individual host species. We conclude that the colonization of new but related host species may represent the principle mode of macroevolution in RNA viruses.

emergence | molecular evolution | host jumping

Although there is a large body of work considering the microevolution of RNA viruses, especially in the guise of molecular epidemiology and studies of how various evolutionary processes shape intraspecies and intrahost genetic diversity, far less is known about how these infectious agents change at a macroevolutionary scale (1–3). In particular, there is no clear understanding of the mechanisms that determine the appearance and maintenance of phylogenetically discrete viral lineages that are often assigned as different virus "species" (4). Although providing a strict definition of a virus species has necessarily proven difficult, and clearly has a large arbitrary component (5), it is important to examine the evolutionary processes that result in the appearance of phylogenetically and often phenotypically distinct lineages of RNA viruses.

We use a phylogenetic approach to explore the modes of macroevolution in RNA viruses. To help focus this study, we use, as informative analogies, terms borrowed from speciation theory as applied to animals. Hence, we ask whether viral lineages primarily differentiate through sympatric processes, which we define here as virus adaptation to different niches within the same host species, or allopatric processes, which we take to mean virus adaptation to different host species (6). In the case of sympatric divergence, viruses would remain pathogens of the same host species but evolve different niches within that host, such as using different cell types, inducing immune responses that are not cross-protective, or establishing different seasonalities. Alternatively, in the case of allopatric divergence, viruses would infect different host species, either by cross-species transmission (i.e., host jumping) or codivergence with hosts over extended time periods. We take these processes to be conceptually (although not mechanistically) analogous to the separation of animal populations by physical barriers. Importantly, determining the relative frequencies of allopatric vs. sympatric divergence will assist in predicting which virus groups are most likely to jump species boundaries and emerge in new hosts.

To determine which mode of macroevolution may play the dominant role in RNA viruses, we performed a series of phylogenetic analyses using well-characterized and representative viral families and genera for which topologically robust phylogenetic trees could be inferred. By identifying the probable primary reservoir host of each virus, we determined the extent of sympatric lineage diversification (indicated by the phylogenetic clustering of viruses that share the same host species) vs. allopatric diversification (in which closely related viruses infect different host species). We focus on viruses of humans, because these are more thoroughly sampled than those of other host species.

Results

Five families or genera of RNA viruses for which it is possible to estimate robust large-scale phylogenetic trees were analyzed. Genome regions for each of the virus groups were selected based on sequence availability and a review of the literature for regions that have previously been shown to be indicative of evolutionary relationships. The known or probable host species of each virus was then mapped onto the virus tree estimated here (Figs. 1–5), and the degree of host-virus association was assessed for each virus data set using three phylogeny-trait association tests (Table 1 and *Materials and Methods*). Due to uncertainties in host species identification, broadly defined taxonomic units and multiple host species were used when assessing host-virus associations, which make such tests conservative and biased toward fewer host shifts across virus trees.

Genus *Alphavirus*. Alphavirus genomes are unsegmented, single-strand, positive-sense RNA molecules. Transmission is typically via an arthropod vector, with closely related viruses using closely related vector species. Unfortunately, vector-borne transmission makes it difficult to identify primary reservoir hosts with certainty. Further, alphaviruses have been observed to infect many vertebrate hosts, which complicates the identification of the specific reservoir host species primarily responsible for virus amplification and transmission.

It is apparent from the alphavirus phylogeny that viral lineages with divergent reservoir hosts may be each other's closest relatives

Author contributions: A.K., L.A.S., and E.C.H. designed research; A.K. and L.A.S. performed research; A.K. and L.A.S. analyzed data; and A.K., L.A.S., and E.C.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]A.K. and L.A.S. contributed equally to this work.

[2]To whom correspondence should be addressed. E-mail: ech15@psu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1011090108/-/DCSupplemental.

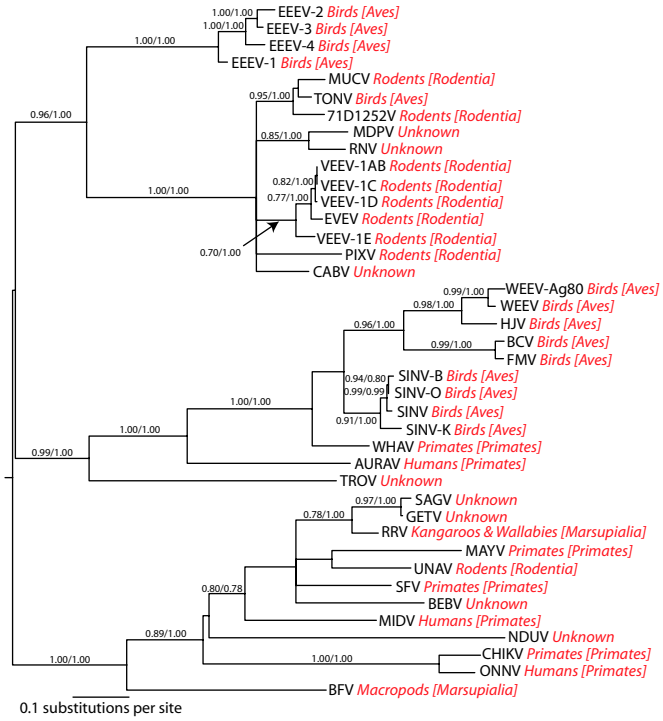

Fig. 1. Maximum-likelihood tree of the genus *Alphavirus*. Viral taxa are denoted in black, alongside their probable reservoir host range (in red) and categories used in the association analysis (in red and bracketed). The tree is midpoint rooted with quartet puzzling, and Bayesian posterior support values ≥70% are given above branches (quartet puzzling/Bayesian posterior).

(Fig. 1). However, there are also several monophyletic groups that share broadly defined reservoir hosts, such as rodents, primates, and birds, and that produce strong signals of reservoir host structure in the tree [association index (AI) and parsimony score (PS): $P < 0.001$] (Table 1). Of these three broadly defined host taxa groups, there is statistical support [maximum monophyletic clade (MC): $P < 0.001$] that alphaviruses with avian and rodent reservoir hosts are more likely to be closely related to other viruses infecting avian and rodent hosts, respectively, than to viruses using other reservoir hosts.

Family Caliciviridae. Calicivirus genomes are unsegmented, single-strand, positive-sense RNA molecules. Transmission is direct and does not involve vectors. Our phylogeny shows that species criteria are quite broadly defined in this family [e.g., International Committee on Taxonomy of Viruses (ICTV) definitions], with single species corresponding to monophyletic groups of lineages that display substantial sequence and reservoir host diversity (Fig. 2).

We find well-supported clusters of viruses sometimes infecting very divergent hosts. For example, the human Lordsdale and Desert Shield noroviruses are most closely related to swine and bovine noroviruses, respectively (Fig. 2). Additionally, viruses of dogs, cats, rabbits, walruses, reptiles, and apes cluster together (i.e., the genus *Vesivirus*), as do viruses of pigs, humans, and mink (i.e., the genus *Sapovirus*). Although there is significant support for the clustering of viruses with the same reservoir hosts across the calicivirus phylogeny (AI and PS: $P < 0.005$; Table 1), only viruses of two broadly defined reservoir hosts (primates and lagomorphs; MC: $P < 0.05$) fall into this category (of four with sample sizes >1). This suggests there is substantial host switching within the Caliciviridae. Indeed, our estimate for the scaled PS statistic (0.661; Table 1), which accounts for both sample size and the number of character states (i.e., reservoir host categories), suggests that caliciviruses exhibit higher levels of host switching than the other four virus groups analyzed here.

Genus *Flavivirus*. Viruses of the genus *Flavivirus* are the most well-sampled virus group analyzed here, with 53 species officially recognized by the ICTV and 56 taxa included in this study. Flaviviruses, which have unsegmented, single-strand, positive-sense RNA genomes, can be phylogenetically distinguished by their mode of transmission: mosquito-borne, tick-borne, and those with no known vector (7). These divisions are well supported in our phylogenetic analysis (Fig. 3).

Importantly, flaviviruses infect a wide range of hosts, from birds to marsupials and primates, as well as insects alone (although these viruses were too divergent to be included in this analysis). Though viruses known to infect the same reservoir host taxa are dispersed throughout the tree, there is significant clustering of viruses that use the same reservoir hosts (PS and AI: $P < 0.001$; Table 1), particularly for viruses of primates and birds (MC: $P < 0.001$; Table 1). For example, Usutu, Murray Valley encephalitis, Japanese encephalitis, West Nile, Kunjin, and St. Louis encephalitis viruses are each other's closest relatives, and all have birds as their natural reservoir hosts. However, when accounting for the sample size and number of reservoir host categories, the flaviviruses displayed intermediate levels of host switching relative to the other viruses studied here (scaled PS = 0.420; Table 1).

Family Paramyxoviridae. The paramyxoviruses are a family of viruses within the order Mononegavirales, which have unsegmented, negative-sense, single-stranded RNA genomes. They infect a wide range of animal hosts and are transmitted directly via the respiratory tract. The two subfamilies, Paramyxovirinae and Pneumovirinae, are further subdivided into five and two genera, respectively, which are well-supported in our phylogenetic analysis (Fig. 4). As is the case with many viruses described here, a number of paramyxoviruses (such as Nipah virus and Hendra virus) have a primary reservoir host but can productively infect a wide range of animals, whereas others, such as measles virus, clearly infect a single host species (humans).

As shown in our paramyxovirus phylogeny, important human pathogens have been identified from five of the seven established genera and rarely cluster with each other. For example, human and avian metapneumoviruses are sister taxa within the genus *Metapneumovirus*, whereas human respiratory syncytial virus (RSV) clusters only with its fellow pneumoviruses—bovine RSV, ovine RSV, and murine pneumonia virus. Despite this, the association indexes indicate that the phylogeny is significantly structured by reservoir host (PS and AI: $P < 0.001$; Table 1). This is likely driven by the clustering of viruses infecting birds (*Aves*), carnivores (*Carnivora*), and marine mammals (*Cetacea*), all of which are significant (MC: $P < 0.05$; Table 1). Unsurprisingly, the scaled PS estimates (Table 1) indicate that the paramyxoviruses switch hosts at a higher rate than all virus groups studied here except the Caliciviridae.

Family Rhabdoviridae. Like paramyxoviruses, members of the family Rhabdoviridae belong to the order Mononegavirales. The family is composed of six ICTV-recognized genera comprised of 46 species, including at least five species unassigned to any genus. The Rhabdoviridae are notable for their diverse host ranges, infecting vertebrates, invertebrates, and even plants (Fig. 5). Interestingly, unlike the other virus groups studied here, though some rhabdoviruses are transmitted by vectors (i.e., the vesiculoviruses, ephemeroviruses, and unassigned viruses; ref. 8), others are directly transmitted, and the broad host ranges of individual viruses means that primary hosts are often not known with certainty. For example, rabies and vesicular stomatitis viruses, which are the best-studied rhabdoviruses causing human

Table 1. Statistical analyses of virus-host associations

Taxonomic group	Observed mean (95% HPD)	Null mean (95% HPD)	Significance
Alphavirus			
AI	0.74 (0.58–0.93)	2.21 (1.54–2.78)	0.001*
PS	7.45 (7.00–8.00)	14.2 (12.0–16.0)	0.001*
Scaled PS[†]	0.397 (NA)	NA	NA
Aves[‡]	7.08 (5.00–9.00)	2.52 (1.12–4.47)	0.001*
Marsupialia	1.00 (1.00–1.00)	1.02(1.00–1.00)	1.000
Primates	2.00 (2.00–2.00)	1.41 (1.00–2.00)	0.202
Rodentia	4.59 (4.00–5.00)	1.40 (1.00–2.00)	0.001*
Caliciviridae			
AI	1.58 (1.41–1.78)	2.52 (1.96–3.07)	0.005*
PS	13.2 (13.0–14.0)	17.4 (15.1–19.4)	0.003*
Scaled PS	0.661 (NA)	NA	NA
Artiodactyla	2.00 (2.00–2.00)	1.54 (1.00–2.77)	0.350
Carnivora	2.00 (2.00–2.00)	1.40 (1.00–2.00)	0.249
Lagomorpha	3.00 (3.00–3.00)	1.12 (1.00–2.00)	0.001*
Primates	3.75 (3.00–4.00)	1.71 (1.00–3.00)	0.013*
Reptilia[§]	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.000
Rodentia[§]	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.000
Flavivirus			
AI	1.21 (1.21–1.21)	4.92 (4.05–5.73)	0.001*
PS	15.8 (15.8–15.8)	30.7 (27.8–33.3)	0.001*
Scaled PS	0.420 (NA)	NA	NA
Artiodactyla	1.25 (1.25–1.25)	1.08 (1.00–1.69)	0.778
Aves	9.00 (9.00–9.00)	1.77 (1.00–3.00)	0.001*
Chiroptera	2.00 (2.00–2.00)	1.19 (1.00–2.00)	0.104
Marsupialia[§]	1.00 (1.00–1.00)	1.01 (1.00–1.00)	1.000
Primates	4.00 (4.00–4.00)	1.58 (1.00–2.81)	0.005*
Rodentia	3.75 (3.75–3.75)	2.45 (1.73–3.94)	0.316
Paramyxoviridae			
AI	1.62 (1.31–1.87)	4.32 (3.64–4.91)	0.001*
PS	17.1 (16.5–17.5)	27.1 (24.6–29.4)	0.001*
Scaled PS	0.476 (NA)	NA	NA
Artiodactyla	1.93 (1.00–2.00)	1.27 (1.00–2.00)	0.162
Aves	7.73 (5.00–9.00)	1.68 (1.00–2.51)	0.001*
Carnivora	2.00 (2.00–2.00)	1.03 (1.00–1.25)	0.020*
Cetacea	2.00 (2.00–2.00)	1.02 (1.00–1.00)	0.011*
Chiroptera	2.00 (2.00–2.00)	1.10 (1.00–2.00)	0.052
Primates	2.75 (2.50–3.00)	1.72 (1.00–2.69)	0.230
Reptilia[§]	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.000
Rodentia	2.01 (2.00–2.00)	1.17 (1.00–2.00)	0.099
Scandentia[§]	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.000
Rhabdoviridae			
AI	0.39 (0.07–0.86)	3.31 (2.69–3.85)	0.001*
PS	9.24 (8.50–9.50)	23.4 (20.9–25.7)	0.001*
Scaled PS	0.162 (NA)	NA	NA
Artiodactyla	4.07 (4.00–5.00)	1.52 (1.00–2.45)	0.004*
Carnivora[§]	1.00 (1.00–1.00)	1.01 (1.00–1.00)	1.000
Chiroptera	3.06 (2.00–3.50)	1.48 (1.00–2.49)	0.012*
Fish	4.00 (4.00–4.00)	1.30 (1.00–2.01)	0.002*
Plants	8.77 (5.00–9.00)	1.51 (1.00–2.42)	0.001*
Primates[§]	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.000
Reptilia	2.00 (2.00–2.00)	1.01 (1.00–1.01)	0.006*
Scandentia[§]	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.000

HPD, highest probability density interval.

*Statistically significant at the $P < 0.05$ level.

[†]The scaled PS is the observed mean PS minus the minimum PS (i.e., the number of taxa categories minus 1) divided by the expected mean PS minus the minimum PS.

[‡]The host-virus association of each individual taxa group (in italics) was determined using the maximum MC statistic.

[§]Taxonomic units represented only once as a reservoir host in the virus phylogenies.

disease, infect many species in addition to their respective bat and artiodactyl reservoir hosts. Rabies virus is now stably maintained in dog populations, which serve as additional reservoir hosts.

Similar to the other virus families and genera studied here, the Rhabdoviridae display significant phylogenetic structure correlated with their probable reservoir hosts (AI and PS: $P < 0.001$). Notably, of the reservoir host taxa with sufficient sample sizes, all but the carnivores (i.e., five of six) clustered together significantly (MC: $P < 0.05$). However, of these host taxa, three are extremely broadly defined: plants, fish, and reptiles (artiodactyls and bats are the other two). It is likely that dividing these most broad-scale categories into more narrowly defined reservoir host taxa, along with greater sampling from nonmammalian hosts, might reduce the significance of these groups. Furthermore, the scaled PS estimate for the Rhabdoviridae is by far the lowest estimated here (0.162), indicating much lower relative rates of host-jumping despite the great diversity of rhabdovirus hosts.

Host Switching of Human Viruses. Because human viruses are the most thoroughly studied viruses with the most accurately known reservoir hosts, analyzing the host species associations as human/nonhuman dichotomies provides a more powerful test of allopatric vs. sympatric processes. Importantly, this test is biased toward sympatric divergence, and thus is a conservative test of allopatric divergence. Under this analysis, only the human Caliciviridae, specifically the human sapoviruses, cluster together in a significant manner (i.e., the AI, PS, and MC statistics have P values <0.05). Hence, the majority of human viruses studied here are the product of host jumping.

Discussion

It might be assumed a priori that the barriers to viral emergence in novel host species are relatively difficult to overcome. Indeed, most cases of viral emergence in reality represent transient "spillover" infections, in which only a few individuals of a novel host species acquire the new virus and without establishment of a sustained chain of transmission (9). In combination with the obviously higher rate of intra- vs. interhost species contact, this may lead to the expectation that the majority of lineage diversification events that give rise to different virus species take place within single host species. Indeed, the phylogenies of well-studied viral families and genera analyzed here suggest that at least half of all virus divergence events (54.5–75.0%; Table S1) occur from niche diversification within single or closely related host species (e.g., species of carnivores, primates, or rodents). However, these figures are also clearly overestimates produced by the broad host taxa categories necessarily used here, and reflect current uncertainty about the host range of many viral lineages, itself a function of a small and biased sample of overall viral biodiversity. In particular, a limited coverage of taxa means that both host-jumping and virus-host codivergence among a group of related host-species would appear as sympatric divergence events in our study. Hence, overestimation of the number of sympatric events will be most pronounced in groups of viruses infecting hosts that belong to the most broadly defined (e.g., plants and fish) or relatively speciose (e.g., rodents) groups, and least pronounced among viruses that infect the least speciose host categories (e.g., primates and carnivores). Indeed, this is what we observe with respect to the Rhabdoviridae, which have both the most broadly defined hosts and the lowest scaled PS estimates in our study. Because observed rates of sympatric divergence will likely decrease with an increased understanding of both viral biodiversity and reservoir host specificity, our findings are therefore likely to underestimate the important and possibly dominant role that host-switching plays in RNA virus macroevolution.

The respective roles of sympatric vs. allopatric viral diversification are likely to be functions of the intrinsic evolutionary constraints faced by RNA viruses. RNA viruses have extremely

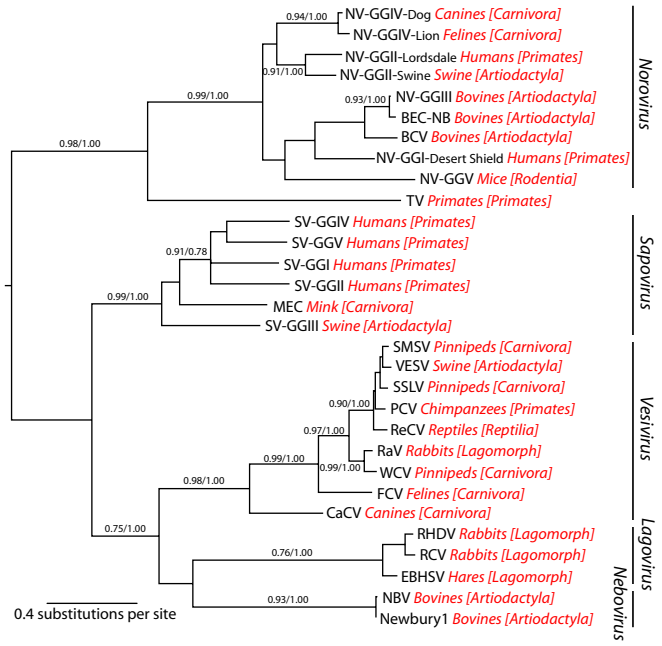

Fig. 2. Maximum-likelihood tree of the family Caliciviridae. Tree labels and description of rooting and node support values as in Fig. 1. Genera within the family are indicated by black bars to the right of the tree.

small genomes, which effectively constrain their evolution through a combination of pleiotropy, epistasis, and fitness tradeoffs resulting from the need to maximize the functionality encoded in small genomes (e.g., the use of overlapping reading frames; ref. 11). We propose that the stronger these evolutionary constraints, the more difficult it becomes to occupy new niches within the current host species. For example, both the phylogenetic data presented here, as well as recent evidence of epidemic human respiratory viruses circulating among chimpanzees (12, 13), suggest that the Paramyxoviridae seem particularly prone to cross-species transfers. This is consistent with results from an earlier study of paramyxovirus F proteins, which found that viruses infecting different host species but with similar tissue tropism and transmission routes (i.e., through the respiratory epithelium) were closely related (14). Additionally, several paramyxoviruses infect more than one species, such as canine distemper virus (canids, raccoons, and mink), and rinderpest virus and bovine parainfluenza virus 3, both of which infect several ungulate species. Furthermore, several paramyxoviruses that infect related hosts, such as the human measles, mumps, and parainfluenza 1 and 3 viruses, are not each other's closest relative (Fig. 4; ref. 14). Combined, these lines of evidence strongly suggest that there are substantial constraints on the ability of paramyxoviruses to infect different cell types within individual host species, yet relatively weaker constraints on their ability to infect different host species. Indeed, the Paramyxoviridae are characterized here as having relatively high rates of allopatric divergence (Table 1 and Table S1).

Lineage diversification within hosts may also require extensive changes in cell tropism, immunogenic peptides, and/or seasonality, all of which may pose a substantial adaptive challenge. In addition, genome-scale epistatic interactions (15), changes in immunogenic regions, and overlaps between immunogenic and receptor-binding site (e.g., as observed in small, single-stranded DNA viruses; ref. 16) will likely necessitate complex compensatory changes in other viral proteins, in turn increasing the difficulty of achieving the level of phenotypic diversification that may be necessary for divergence within a single host species. Notably,

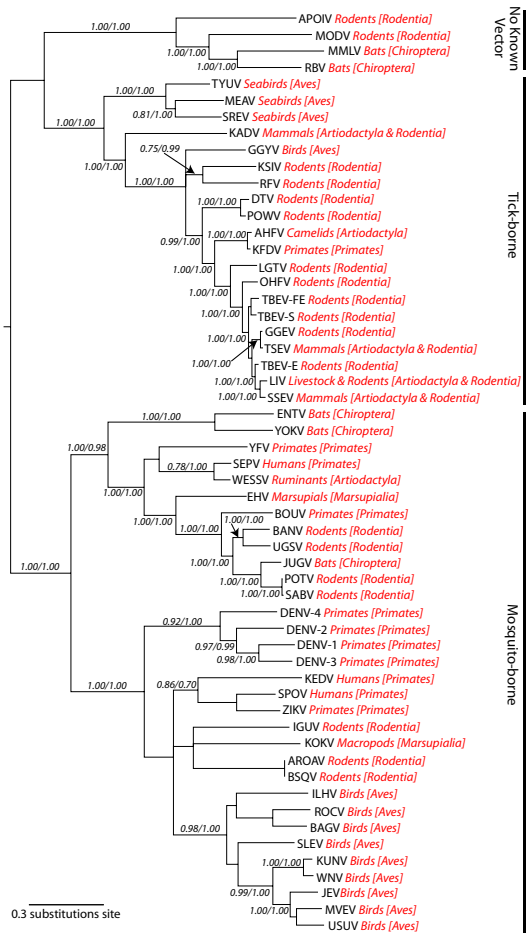

Fig. 3. Maximum-likelihood tree of the genus *Flavivirus*. Tree labels and description of rooting and node support values as in Fig. 1. Vector class (mosquito, tick, or unknown) are indicated by black bars to the right of the tree.

by isolating viral lineages in time, changes in seasonality may lead to sympatric divergence. However, as virus seasonality is evidently a dynamic process, the mechanics of which are not well known (17), it is currently unclear how viral genotype or phenotype contribute to observed differences in seasonality among viruses.

Although some adaptive changes may be necessary for the successful infection of a new host species, such as those at receptor-binding sites (18), these may be relatively minor compared with those required to occupy new niches within a host. In addition, the closer the donor and recipient host species are in phylogenetic space then, on average, the fewer the number of mutations likely required for adaptation to the new host (4). For example, foot-and-mouth disease virus from pigs requires one amino acid replacement to replicate in guinea pigs (19), whereas avian influenza viruses may require as many as 13 mutations to productively replicate in mammals (20). The relatively frequent transfer of viral lineages between humans and other primates further illustrates this concept (12, 13, 21, 22). It therefore seems reasonable to speculate that cross-reactive immune responses, coupled with the slim possibility that diverging viruses may acquire diverse and nonoverlapping niches within the same host, may represent an evolutionary barrier less easily traversed by the virus than that of infection of a novel host species, especially if the two hosts are closely related.

It is clear that the most frequent type of allopatric divergence —virus-jumping between hosts that are close phylogenetically— is systematically underestimated in our study due to a poor un-

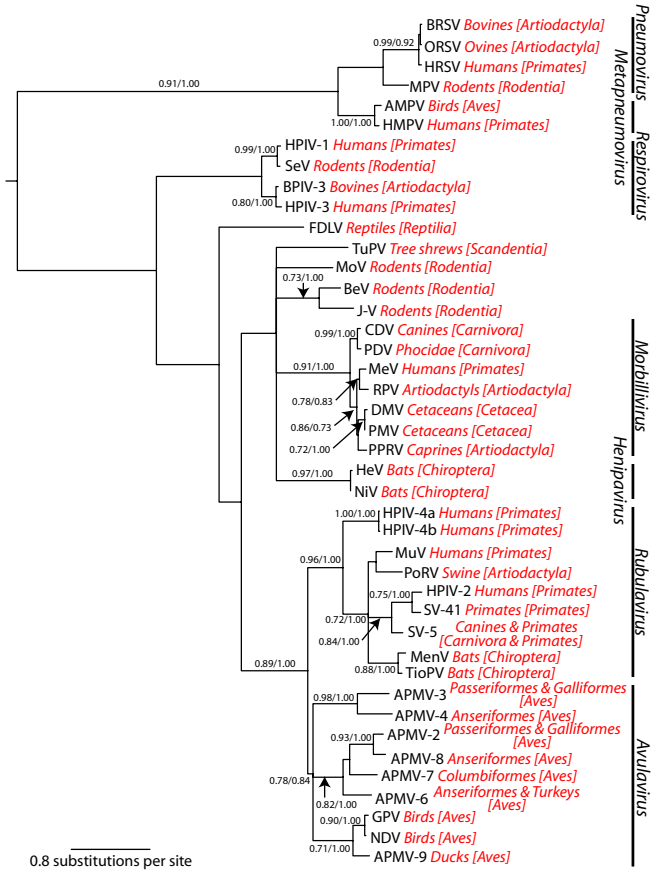

Fig. 4. Maximum-likelihood tree of the family Paramyxoviridae. Tree labels and description of rooting and node support values as in Fig. 1. Genera within the family are indicated by black bars to the right of the tree.

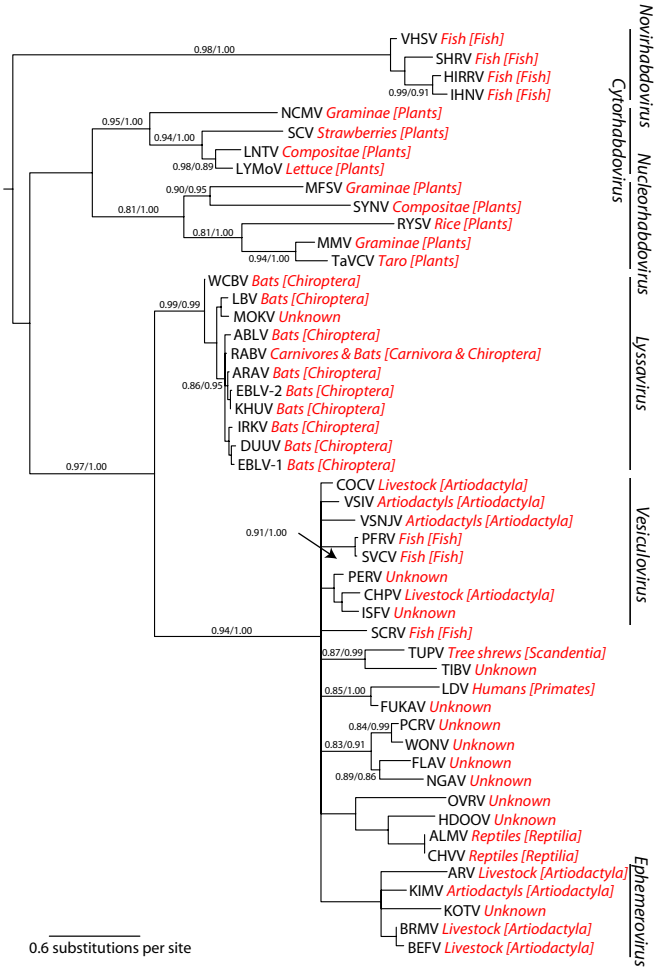

Fig. 5. Maximum-likelihood tree of the family Rhabdoviridae. Tree labels and description of rooting and node support values as in Fig. 1. Genera within the family are indicated by black bars to the right of the tree.

derstanding of reservoir host specificity, particularly with respect to viruses with nonhuman reservoirs. Indeed, our analysis of human vs. nonhuman viruses shows that within the broadly defined host taxa categories used here, allopatric divergence is more common than sympatric divergence. Sampling bias will also have a major effect on estimates of the extent of virus-host codivergence which, by definition, occurs in closely related host species. However, this caveat notwithstanding, visual inspection of the phylogenies presented here provides little support for virus-host codivergence in these representative RNA virus families. For example, the vesiviruses in the family Calciviridae contain viruses that infect reptiles, swine, rabbits, and primates interspersed with viruses of carnivores. It is also possible that there are intermediaries in viral phylogenies that have not yet been sampled from the environment or went extinct due to stochastic processes, such that observed sister lineages may be separated by a third, unsampled, viral lineage (23). Combined, these findings further suggest a key role for allopatric processes in RNA virus macroevolution.

Overall, our findings suggest that despite the evident evolutionary barriers to switching hosts, it is more likely that viruses will successfully escape niche overlap through allopatric rather than sympatric processes, although the precise mechanisms underlying this form of divergence are generally unknown and evidently require further study. Hence, a host switch may require both less net movement through sequence space and cost less in terms of fitness than the substantial changes that may be necessary to acquire a new niche within the same host species. When ecological conditions allow for frequent host-jumping, such as communities rich in biodiversity, we would expect allopatric divergence to be particularly frequent. Hence, although cross-species transmission and emergence is normally regarded as an unusual mode of viral evolution, our analyses in fact suggest that may be a common form of RNA virus macroevolution. More generally, this study establishes a model for studying macroevolutionary processes in RNA viruses. Besides the obvious necessity for more accurate determination of primary host reservoir species, additional research is needed to identify modes of transmission and target tissues for many of the viruses studied here.

Materials and Methods

Sequence Alignments. RNA virus genera or families were selected based on the availability of amino acid sequences with sufficient sequence conservation to estimate reliable phylogenies. Although only a small number of RNA virus families meet these criteria, they are sufficient to demonstrate the general utility of the method and provide an initial insight into modes of virus macroevolution. Sequences of viruses from the families Paramyxoviridae and Caliciviridae, and the genus *Flavivirus*, were collected from GenBank, aligned using MUSCLE (24), and translated/visualized with Se-Al (http://tree.bio.ed.ac.uk/software/seal). The 373-aa Paramyxoviridae alignment consisted of conserved N-protein regions from 42 taxa; the 329-aa Caliciviridae alignment was comprised of conserved major capsid protein regions from 30 taxa; and the 3,522-aa *Flavivirus* alignment was comprised of complete polyprotein sequences from 56 taxa, excluding Tamana bat virus, which was too divergent to be included. Scott Weaver (University of Texas Medical Branch, Galveston, TX) kindly provided an alignment of partial E1 glycoproteins from the genus

Alphavirus, to which additional sequences, retrieved from GenBank, were added. The final 349-aa alignment contained 40 alphavirus sequences. Southern elephant seal louse virus was excluded due to inadequate sequence data, and salmon pancreas disease virus because of its extreme divergence. The family Rhabdoviridae L protein (polymerase) sequences (partial, 158 aa) were obtained and aligned as described in Bourhy et al. (8). Additional sequences were gathered from GenBank, resulting in a 158-aa alignment of 50 virus taxa. Full lists of all viruses analyzed with GenBank accession numbers are provided in Tables S2–S6.

Phylogenetic Trees. Phylogenetic trees were estimated using the maximum-likelihood (ML) method available in TREE-PUZZLE (25), in each case incorporating the WAG+Γ model of amino acid substitution with relevant parameters estimated from the data (parameter values available on request). All ML analyses were run for at least 100,000 tree-puzzling steps. Phylogenetic trees were also estimated using the Bayesian criteria implemented in MrBayes v3.1.2 (26, 27), which uses Markov chain Monte Carlo (MCMC) simulation to estimate posterior distributions. The WAG+Γ model was again used in all cases, and all other priors had default values. All MCMC analyses were run for 10 million generations (samples taken every 1,000 generations, and the first 10% discarded as burn-in), with multiple heated chains, and in duplicate to ensure convergence. Both sets of trees were midpoint rooted with nodes $\geq$70% posterior support labeled.

Host Species Association. The known or probable primary host species were determined, as far as is possible, for every virus in the analysis. Sources used to identify host species included, but were not limited to, GenBank host and reference entries accompanying sequence data, the ICTVdB (http://www.ictvonline.org/), Grard et al. (28), Kuzmin et al. (29), and Tidona and Darai (30). Reservoir hosts for some viruses included in this study could not be identified with confidence. Because there is often limited knowledge of the primary reservoir for a given virus, future research may reveal that some host species assignments are inaccurate, or may be more precisely determined than reported here. Complete lists of references used to identify probable primary reservoir hosts are provided in Tables S2–S6.

The degree of association between hosts and viruses was assessed in the five virus data sets using PS (31), AI (32), and MC (33) statistics. The first two statistics indicate the association of hosts and viruses (i.e., phylogenetic clustering of viruses that infect the same hosts) across entire trees and among all hosts and viruses using the frequency of host categories among descendent viruses at each node and by counting the number of host category shifts throughout the tree, respectively. The MC statistic assesses the association between specific hosts and viruses by estimating the size of the largest cluster of viruses using the same reservoir hosts. These analyses were performed using the program BaTS v0.90 beta (33), which calculates empirical distributions of these statistics from the credible sample (posterior distribution) of trees provided by the Bayesian phylogenetic analyses (described previously). The significance of association statistics was assessed by randomizing the host-virus associations 1,000 times and recalculating the statistics across the distribution of trees to estimate null distributions. To allow for direct comparisons between phylogenies, a scaled PS was calculated for each virus data set. This is calculated by dividing the observed mean PS by the mean PS expected under the null distribution (which accounts for both the number of virus and host taxa in each data set), and offsetting both observed and expected mean PS values by the number of host categories (which determines the minimum PS). This statistic produces a scaled value in which 0 indicates the greatest amount of intrahost lineage diversification (i.e., the minimum observable PS given the number of host categories), $<$1 more intra- than interhost lineage diversification (i.e., more clustering of viruses and hosts) than expected if random, and $>$1 more inter- than intrahost lineage diversification (i.e., an overdispersion of viruses and hosts) than expected if random.

ACKNOWLEDGMENTS. We thank Dr. Scott Weaver (University of Texas Medical Branch, Galveston, TX) for kindly providing an alignment of alphavirus proteins, and Dr. Hervé Bourhy for advice concerning rhabdoviruses. Support for this work was provided in part by National Institutes of Health Grant R01 GM080533 (to E.C.H.).

1. Fargette D, et al. (2008) Diversification of rice yellow mottle virus and related viruses spans the history of agriculture from the neolithic to the present. *PLoS Pathog* 4: e1000125.
2. Gibbs AJ, Ohshima K, Phillips MJ, Gibbs MJ (2008) The prehistory of potyviruses: Their initial radiation was during the dawn of agriculture. *PLoS ONE* 3:e2523.
3. Pagán I, Holmes EC (2010) Long-term evolution of the Luteoviridae: Time scale and mode of virus speciation. *J Virol* 84:6177–6187.
4. Holmes EC (2009) *The Evolution and Emergence of RNA Viruses* (Oxford Univ Press, Oxford).
5. Gibbs AJ, Gibbs MJ (2006) A broader definition of 'the virus species'. *Arch Virol* 151: 1419–1422.
6. Schluter D (2001) Ecology and the origin of species. *Trends Ecol Evol* 16:372–380.
7. Gaunt MW, et al. (2001) Phylogenetic relationships of flaviviruses correlate with their epidemiology, disease association and biogeography. *J Gen Virol* 82:1867–1876.
8. Bourhy H, Cowley JA, Larrous F, Holmes EC, Walker PJ (2005) Phylogenetic relationships among rhabdoviruses inferred using the L polymerase gene. *J Gen Virol* 86: 2849–2858.
9. Parrish CR, et al. (2008) Cross-species virus transmission and the emergence of new epidemic diseases. *Microbiol Mol Biol Rev* 72:457–470.
10. Mallet J, Meyer A, Nosil P, Feder JL (2009) Space, sympatry and speciation. *J Evol Biol* 22:2332–2341.
11. Holmes EC (2003) Error thresholds and the constraints to RNA virus evolution. *Trends Microbiol* 11:543–546.
12. Köndgen S, et al. (2008) Pandemic human viruses cause decline of endangered great apes. *Curr Biol* 18:260–264.
13. Kaur T, et al. (2008) Descriptive epidemiology of fatal respiratory outbreaks and detection of a human-related metapneumovirus in wild chimpanzees (*Pan troglodytes*) at Mahale Mountains National Park, Western Tanzania. *Am J Primatol* 70:755–765.
14. Taber SW, Pease CM (1990) Paramyxovirus phylogeny: Tissue tropism evolves slower than host specificity. *Evolution* 44:435–438.
15. Rambaut A, et al. (2008) The genomic and epidemiological dynamics of human influenza A virus. *Nature* 453:615–619.
16. Nelson CDS, Palermo LM, Hafenstein SL, Parrish CR (2007) Different mechanisms of antibody-mediated neutralization of parvoviruses revealed using the Fab fragments of monoclonal antibodies. *Virology* 361:283–293.
17. Altizer S, et al. (2006) Seasonality and the dynamics of infectious diseases. *Ecol Lett* 9: 467–484.
18. Anishchenko M, et al. (2006) Venezuelan encephalitis emergence mediated by a phylogenetically predicted viral mutation. *Proc Natl Acad Sci USA* 103:4994–4999.
19. Núñez JI, et al. (2001) A single amino acid substitution in nonstructural protein 3A can mediate adaptation of foot-and-mouth disease virus to the guinea pig. *J Virol* 75: 3977–3983.
20. Finkelstein DB, et al. (2007) Persistent host markers in pandemic and H5N1 influenza viruses. *J Virol* 81:10292–10299.
21. Gao F, et al. (1999) Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. *Nature* 397:436–441.
22. Leroy EM, et al. (2004) Multiple Ebola virus transmission events and rapid decline of central African wildlife. *Science* 303:387–390.
23. Pybus OG, Rambaut A, Holmes EC, Harvey PH (2002) New inferences from tree shape: Numbers of missing taxa and population growth rates. *Syst Biol* 51:881–888.
24. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797.
25. Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: Maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 18:502–504.
26. Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–755.
27. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.
28. Grard G, et al. (2010) Genomics and evolution of *Aedes*-borne flaviviruses. *J Gen Virol* 91:87–94.
29. Kuzmin IV, Novella IS, Dietzgen RG, Padhi A, Rupprecht CE (2009) The rhabdoviruses: Biodiversity, phylogenetics, and evolution. *Infect Genet Evol* 9:541–553.
30. Tidona CA, Darai G, eds (2002) *The Springer Index of Viruses* (Springer, Berlin).
31. Slatkin M, Maddison WP (1989) A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics* 123:603–613.
32. Wang TH, Donaldson YK, Brettle RP, Bell JE, Simmonds P (2001) Identification of shared populations of human immunodeficiency virus type 1 infecting microglia and tissue macrophages outside the central nervous system. *J Virol* 75:11686–11699.
33. Parker J, Rambaut A, Pybus OG (2008) Correlating viral phenotypes with phylogeny: Accounting for phylogenetic uncertainty. *Infect Genet Evol* 8:239–246.

Supporting Information

Kitchen et al. 10.1073/pnas.1011090108

Table S1. Scaled parsimony estimates of relative allopatric and sympatric divergences

	Minimum*	Observed mean	Expected mean	(Obs − min)/(exp − min)	Obs/total divergences†
Alphavirus (*n* = 30)‡	3.0	7.45	14.2	0.397	0.257
Caliciviridae (*n* = 30)	5.0	13.2	17.4	0.661	0.455
Flavivirus (*n* = 56)	5.0	15.8	30.7	0.420	0.287
Paramyxoviridae (*n* = 42)	8.0	17.1	27.1	0.476	0.417
Rhabdoviridae (*n* = 38)	6.5	9.24	23.4	0.162	0.250

*The minimum parsimony score is the number of character states (i.e., reservoir host categories) minus 1.
†The total divergences in a rooted tree is the number of taxa minus 1, which is the sum total of both allopatric and sympatric speciation events. Thus, the observed mean parsimony score (Obs) divided by the total divergences is an estimate of the mean proportion of allopatric speciation events in the phylogeny.
‡The number of virus taxa for which probable host reservoirs have been assigned and used in the analyses of virus-host associations.

Table S2. Genus *Alphavirus*

Species	Abbreviation	Accession no.	Probable host	Ref.
Aura virus	AURAV	NP_819019	Humans (*Primates*)	1
Barmah Forest virus	BFV	NP_819002	Macropods (*Marsupialia*)	2
Bebaru virus	BEBV	AAL35779	Unknown	3
Buggy Creek virus	BCV	AAL35789	Birds (*Aves*)	4
Cabassou virus	CABV	AAD14567	Unknown	3
Chikungunya virus	CHIKV	NP_690589	Primates (*Primates*)	1, 5
Eastern equine encephalitis virus 1	EEEV-1	ABQ63086	Birds (*Aves*)	1, 6
Eastern equine encephalitis virus 2	EEEV-2	AAF04801	Birds (*Aves*)	1, 6
Eastern equine encephalitis virus 3	EEEV-3	AAF04802	Birds (*Aves*)	1, 6
Eastern equine encephalitis virus 4	EEEV-4	AAF04803	Birds (*Aves*)	1, 6
Everglades virus	EVEV	P36330	Rodents (*Rodentia*)	1, 3
Fort Morgan virus	FMV	AAL35788	Birds (*Aves*)	1
Getah virus	GETV	AAO33339	Unknown	1
Highlands J virus	HJV	AAB02205	Birds (*Aves*)	1, 3
Mayaro virus	MAYV	AAL35780	Primates (*Primates*)	1, 7
Middelburg virus	MIDV	AAL35777	Humans (*Primates*)	1
Mosso das Pedras virus	MDPV	AAD14563	Unknown	
Mucambo virus	MUCV	ACV42452	Rodents (*Rodentia*)	1, 3
Ndumu virus	NDUV	AAL35778	Unknown	1
O'nyong-nyong virus	ONNV	NP_740711	Humans (*Primates*)	1
Pixuna virus	PIXV	AAD14561	Rodents (*Rodentia*)	1
Rio Negro virus	RNV	AAL35787	Unknown	
Ross River virus	RRV	NP_740686	Macropods (*Masupialia*)	1, 3
Sagiyama virus	SAGV	AAL35781	Unknown	
Salmon pancreas disease virus	SPDV	NP_647497	Salmon (*Fish*)	1
Semliki Forest virus	SFV	NP_819008	Primates (*Primates*)	1, 3
Sindbis virus	SINV	NP_640677	Birds (*Aves*)	1, 6
Sindbis virus Babanki	SINV-B	AA033325	Birds (*Aves*)	1, 6
Sindbis virus Kyzylagach	SINV-K	AAL35791	Birds (*Aves*)	1, 6
Sindbis virus Ockelbo	SINV-O	P27285	Birds (*Aves*)	1, 6
Tonate virus	TONV	AAL35785	Birds (*Aves*)	1, 3
Trocara virus	TROV	AAL55092	Unknown	
Una virus	UNAV	AAL35783	Rodents (*Rodentia*)	1, 7
Venezuelan equine encephalitis virus IAB	VEEV-IAB	AAD37000	Rodents (*Rodentia*)	1, 8
Venezuelan equine encephalitis virus IC	VEEV-IC	AAK66990	Rodents (*Rodentia*)	1, 8
Venezuelan equine encephalitis virus ID	VEEV-ID	P36329	Rodents (*Rodentia*)	1, 8
Venezuelan equine encephalitis virus IE	VEEV-IE	AAW30006	Rodents (*Rodentia*)	1,8
Venezuelan equine encephalitis virus 71D1252V	71D1252V	AAL35786	Rodents (*Rodentia*)	1,8
Western equine encephalitis virus	WEEV	P13897	Birds (*Aves*)	1, 6
Western equine encephalitis virus Ag80	WEEV-Ag80	AAL35792	Birds (*Aves*)	1, 6
Whataroa virus	WHAV	AA033329	Primates (*Primates*)	1, 9

1. Stollar V (2002) *The Springer Index of Viruses*, eds Tidona CA, Darai G (Springer, Berlin), pp 1148–1155.
2. Kay BH, Boyd AM, Ryan PA, Hall RA (2007) Mosquito feeding patterns and natural infection of vertebrates with Ross River and Barmah Forest viruses in Brisbane, Australia. *Am J Trop Med Hyg* 76:417–423.
3. Hanley KA, Weaver SC (2008) *Origin and Evolution of Viruses*, eds Domingo E, Parrish CR, Holland JJ (Academic, New York), pp 351–391.
4. O'Brien VA, Meteyer CU, Ip HS, Long RR, Brown CR (2010) Pathology and virus detection in tissues of nestling house sparrows naturally infected with Buggy Creek virus (Togaviridae). *J Wildl Dis* 46:23–32.
5. Powers AM, Logue CH (2007) Changing patterns of chikungunya virus: Re-emergence of a zoonotic arbovirus. *J Gen Virol* 88:2363–2377.
6. Yuill TM (1986) The ecology of tropical arthropod-borne viruses. *Annu Rev Ecol Syst* 17:189–219.
7. Powers AM, et al. (2006) Genetic relationships among Mayaro and Una viruses suggest distinct patterns of transmission. *Am J Trop Med Hyg* 75:461–469.
8. Weaver SC, Ferro C, Barrera R, Boshell J, Navarro J-C (2004) Venezuelan equine encephalitis. *Annu Rev Entomol* 49:141–174.
9. Lavergne A, et al. (2006) Mayaro virus: Complete nucleotide sequence and phylogenetic relationships with other *alphaviruses*. *Virus Res* 117:283–290.

Table S3. Family Caliciviridae

Species	Abbreviation	Accession no.	Probable host	Ref.
Bovine calicivirus virus	BCV	CAA09481	Bovines (*Artiodactyla*)	1
Bovine enteric calicivirus virus	BEC-NB	AAP83353	Bovines (*Artiodactyla*)	2
Canine calicivirus	CaCV	NP_786912	Canines (*Carnivora*)	3
European brown hare syndrome virus	EBHSV	NP_786903	Hares (*Lagomorphs*)	4
Feline calicivirus	FCV	Q66915	Felines (*Carnivora*)	5
Mink enteric sapovirus	MEC	AAN64326	Mink (*Carnivora*)	6
Nebraska-like virus	NBV	NP_663315	Bovines (*Artiodactyla*)	7
Newbury agent 1 virus	Newbury1	AAY60849	Bovines (*Artiodactyla*)	8
Norovirus genogroup I (Desert Shield)	NV-GGI	U04469	Humans (*Primates*)	9
Norovirus genogroup II (Lordsdale)	NV-GGII-Lord	X86557	Humans (*Primates*)	10
Norovirus genogroup II (swine calicivirus)	NV-GGII-Swine	BAB83516	Swine (*Artiodactyla*)	11
Norovirus genogroup III (Newbury Agent 2)	NV-GGIII	AAN76437	Bovines (*Artiodactyla*)	12
Norovirus genogroup IV (dog)	NV-GGIV-Dog	EU224456	Canines (*Carnivora*)	13
Norovirus genogroup IV (lion)	NV-GGIV-Lion	ABR15783	Felines (*Carnivora*)	14
Norovirue genorgoup V (murine norovirus 1)	NV-GGV	ABU55565	Mice (*Rodentia*)	15
Primate calicivirus virus/VESV-like/Pan-1	PCV	AAC61759	Chimpanzees (*Primates*)	5
Rabbit calicivirus (RCV Australia)	RCV	NC_011704	Rabbits (*Lagomorphs*)	16
Rabbit hemorrhagic disease virus	RHDV	ACF57788	Rabbits (*Lagomorphs*)	4
Rabbit vesivirus	RaV	YP_873923	Rabbits (*Lagomorphs*)	17
Reptile calicivirus	ReCV	AAX48222	Reptiles (*Reptiles*)	18
San Miguel sea lion virus	SMSV	AAA16220	Pinnipeds (*Artiodactyla*)	5
Sapovirus genogroup I (Sapporo)	SV-GGI	AAB60927	Humans (*Primates*)	19
Sapovirus genogroup II (MC10)	SV-GGII	YP_052971	Humans (*Primates*)	19
Sapovirus genogroup III (porcine enteric virus)	SV-GGIII	Q9QEJ5	Swine (*Artiodactyla*)	20
Sapovirus genogroup IV (Chiba/000671/1999/JP)	SV-GGIV	CAH10754	Humans (*Primates*)	19
Sapovirus genogroup V (Arg39/1995/ARG)	SV-GGV	AAP48604	Humans (*Primates*)	19
Steller sea lion vesivirus	SSLV	YP_002004565	Pinnipeds (*Carnivora*)	21
Tulane virus	TV	ACB38131	Primates (*Primates*)	22
Vesicular exanthema of swine virus	VESV	AAC13889	Swine (*Artiodactyla*)	5
Walrus calicivirus virus	WCV	AAG42492	Pinnipeds (*Carnivora*)	23

1. Liu BL, et al. (1999) Molecular characterization of a bovine enteric calicivirus: Relationship to the Norwalk-like viruses. *J Virol* 73:819–825.
2. Clarke IN, Lambden PR (2002) *The Springer Index of Viruses*, eds Tidona CA, Darai G (Springer, Berlin), pp 180–183.
3. Matsuura Y, et al. (2002) Complete nucleotide sequence, genome organization and phylogenic analysis of the canine calicivirus. *Virus Genes* 25:67–73.
4. Rossi C, Meyers G (2002) *The Springer Index of Viruses*, eds Tidona CA, Darai G (Springer, Berlin), pp 176–179.
5. Neill JD (2002) *The Springer Index of Viruses*, eds Tidona CA, Darai G (Springer, Berlin), pp 188–191.
6. Guo M, Evermann JF, Saif LJ (2001) Detection and molecular characterization of cultivable caliciviruses from clinically normal mink and enteric caliciviruses associated with diarrhea in mink. *Arch Virol* 146:479–493.
7. Smiley JR, Chang KO, Hayes J, Vinjé J, Saif LJ (2002) Characterization of an enteropathogenic bovine calicivirus representing a potentially new calicivirus genus. *J Virol* 76:10089–10098.
8. Bridger JC, Hall GA, Brown JF (1984) Characterization of a calici-like virus (Newbury agent) found in association with astrovirus in bovine diarrhea. *Infect Immun* 43:133–138.
9. Hyams KC, et al. (1993) Norwalk virus infection among Desert Storm troops. *J Infect Dis* 167:986–987.
10. Dingle KE, Lambden PR, Caul EO, Clarke IN (1995) Human enteric Caliciviridae: The complete genome sequence and expression of virus-like particles from a genetic group II small round structured virus. *J Gen Virol* 76:2349–2355.
11. Sugieda M, et al. (1998) Detection of Norwalk-like virus genes in the caecum contents of pigs. *Arch Virol* 143:1215–1221.
12. Woode GN, Bridger JC (1978) Isolation of small viruses resembling astroviruses and caliciviruses from acute enteritis of calves. *J Med Microbiol* 11:441–452.
13. Martella V, et al. (2008) Detection and molecular characterization of a canine norovirus. *Emerg Infect Dis* 14:1306–1308.
14. Martella V, et al. (2007) Norovirus in captive lion cub (Panthera leo). *Emerg Infect Dis* 13:1071–1073.
15. Wobus CE, Thackray LB, Virgin HW, 4th (2006) Murine norovirus: A model system to study norovirus biology and pathogenesis. *J Virol* 80:5104–5112.
16. Capucci L, Fusi P, Lavazza A, Pacciarini ML, Rossi C (1996) Detection and preliminary characterization of a new rabbit calicivirus related to rabbit hemorrhagic disease virus but nonpathogenic. *J Virol* 70:8614–8623.
17. Martín-Alonso JM, et al. (2005) Isolation and characterization of a new Vesivirus from rabbits. *Virology* 337:373–383.
18. Smith AW, Anderson MP, Skilling DE, Barlough JE, Ensley PK (1986) First isolation of calicivirus from reptiles and amphibians. *Am J Vet Res* 47:1718–1721.
19. Chiba S, Nakata S (2002) *The Springer Index of Viruses*, eds Tidona CA, Darai G (Springer, Berlin), pp 184–187.
20. Guo M, et al. (1999) Molecular characterization of a porcine enteric calicivirus genetically related to Sapporo-like human *caliciviruses*. *J Virol* 73:9625–9631.
21. McClenahan SD, et al. (2008) Genomic characterization of novel marine vesiviruses from Steller sea lions (Eumetopias jubatus) from Alaska. *Virus Res* 138:26–35.
22. Farkas T, Sestak K, Wei C, Jiang X (2008) Characterization of a rhesus monkey calicivirus representing a new genus of Caliciviridae. *J Virol* 82:5408–5416.
23. Ganova-Raeva L, Smith AW, Fields H, Khudyakov Y (2004) New Calicivirus isolated from walrus. *Virus Res* 102:207–213.

Table S4. Genus *Flavivirus*

Species	Abbreviation	Accession no.	Probable host	Ref.
Alkhurma hemorrhagic fever virus	AHFV	NP_722551	Camelids (*Artiodactyla*)	1
Apoi virus	APOIV	NP_620045	Rodents (*Rodentia*)	2
Aroa virus	AROAV	YP_001040004	Rodents (*Rodentia*)	2
Bagaza virus	BAGV	YP_002790883	Birds (*Aves*)	3
Banzi virus	BANV	ABI54472	Rodents (*Rodentia*)	2
Bouboui virus	BOUV	ABI54473	Primates (*Primates*)	2, 4
Bussuquara virus	BSQV	AAV34152	Rodents (*Rodentia*)	2
Deer tick virus	DTV	AAL32169	Rodents (*Rodentia*)	5
Dengue virus 1	DENV-1	NP_059433	Primates (*Primates*)	2
Dengue virus 2	DENV-2	AAC59275	Primates (*Primates*)	2
Dengue virus 3	DENV-3	P27915	Primates (*Primates*)	2
Dengue virus 4	DENV-4	GNWVDF	Primates (*Primates*)	2
Edge Hill virus	EHV	ABI54476	Marsupials (*Marsupialia*)	2
Entebbe bat virus	ENTV	YP_950477	Bats (*Chiroptera*)	2
Gadgets Gully virus	GGYV	ABB90669	Birds (*Aves*)	2, 6
Greek goat encephalitis virus	GGEV	ABB90677	Rodents (*Rodentia*)	7
Iguape virus	IGUV	AAV34154	Rodents (*Rodentia*)	2
Ihleus virus	ILHV	YP_001040006	Birds (*Aves*)	2, 8
Japanese encephalitis virus	JEV	NP_059434	Birds (*Aves*)	2, 9
Jugra virus	JUGV	ABI54482	Bats (Chiroptera)	2
Kadam virus	KADV	ABB90670	Mammals (*Artiodactyla* and *Rodentia*)	2
Karshi virus	KSIV	ABB90671	Rodents (*Rodentia*)	10
Kedougou virus	KEDV	ABI54477	Humans (*Primates*)	2
Kokobera virus	KOKV	YP_001040007	Macropods (*Marsupialia*)	2, 11
Kunjin virus	KUNV	P14335	Birds (*Aves*)	12
Kyasanur Forest disease virus	KFDV	AAQ91607	Primates (*Primates*)	2, 13
Langat virus	LGTV	NP_620108	Rodents (*Rodentia*)	2
Louping ill virus	LIV	NP_044677	Livestock and rodents (*Artiodactyla* and *Rodentia*)	2, 14, 15
Meaban virus	MEAV	ABB90668	Seabirds (*Aves*)	2
Modoc virus	MODV	NP_619758	Rodents (*Rodentia*)	2
Montana myotis leukoencephalitis virus	MMLV	NP_689391	Bats (*Chiroptera*)	2
Murray Valley encephalitis virus	MVEV	NP_051124	Birds (*Aves*)	2, 13
Omsk hemorrhagic fever virus	OHFV	AAR98531	Rodents (*Rodentia*)	2, 13
Potiskum virus	POTV	ABI54483	Rodents (*Rodentia*)	2, 16
Powassan virus	POWV	NP_620099	Rodents (*Rodentia*)	2, 17
Rio Bravo virus	RBV	NP_620044	Bats (*Chiroptera*)	2
Rocio virus	ROCV	AAV34158	Birds (*Aves*)	2, 13
Royal Farm virus	RFV	ABB90673	Rodents (*Rodentia*)	2, 18
Saboya virus	SABV	ABI54478	Rodents (*Rodentia*)	2,4
Saumarez Reef virus	SREV	ABB90674	Seabirds (*Aves*)	2
Sepik virus	SEPV	ABI54479	Humans (*Primates*)	2
Spanish sheep encephalitis virus	SSEV	ABB90676	Mammals (*Artiodactyla* and *Rodentia*)	2
Spondweni virus	SPOV	ABI54480	Humans (*Primates*)	19
St. Louis encephalitis virus	SLEV	AAV34160	Birds (*Aves*)	2, 20
Tick-borne encephalitis virus European	TBEV-E	NP_043135	Rodents (*Rodentia*)	2, 21
Tick-borne encephalitis virus Far Eastern	TBEV-FE	BAB72162	Rodents (*Rodentia*)	2, 21
Tick-borne encephalitis virus Siberian	TBEV-S	AAD34205	Rodents (*Rodentia*)	2, 21
Turkish sheep encephalitis virus	TSEV	ABB90675	Mammals (*Artiodactyla* and *Rodentia*)	2
Tyuleniy virus	TYUV	ABB90672	Seabirds (*Aves*)	2, 22
Uganda S virus	UGSV	ABI54481	Rodents (*Rodentia*)	2, 23
Usutu virus	USUV	YP_164264	Birds (*Aves*)	2, 13
Wesselsbron virus	WESSV	ABI54474	Ruminants (*Artiodactyla*)	2, 24
West Nile virus	WNV	NP_041724	Birds (*Aves*)	2, 13
Yellow fever virus	YFV	Q9YRV3	Primates (*Primates*)	2
Yokose virus	YOKV	NP_872627	Bats (*Chiroptera*)	2, 25
Zika virus	ZIKV	ABI54475	Primates (*Primates*)	2

1. Charrel RN, et al. (2007) Alkhurma hemorrhagic fever virus in Ornithodoros savignyi ticks. *Emerg Infect Dis* 13:153–155.
2. Westaway EG (2002) *The Springer Index of Viruses*, eds Tidona CA, Darai G (Springer, Berlin), pp 306–319.
3. Bondre VP, et al. (2009) Genetic characterization of Bagaza virus (BAGV) isolated in India and evidence of anti-BAGV antibodies in sera collected from encephalitis patients. *J Gen Virol* 90:2644–2649.
4. Grard G, et al. (2010) Genomics and evolution of *Aedes*-borne flaviviruses. *J Gen Virol* 91:87–94.
5. Telford SR, 3rd, et al. (1997) A new tick-borne encephalitis-like virus infecting New England deer ticks, *Ixodes dammini*. *Emerg Infect Dis* 3:165–170.
6. St George TD, et al. (1985) The isolation of arboviruses including a new flavivirus and a new Bunyavirus from *Ixodes* (*Ceratixodes*) *uriae* (*Ixodoidea*: *Ixodidae*) collected at Macquarie Island, Australia, 1975–1979. *Am J Trop Med Hyg* 34:406–412.
7. Papa A, Pavlidou V, Antoniadis A (2008) Greek goat encephalitis virus strain isolated from *Ixodes ricinus*, Greece. *Emerg Infect Dis* 14:330–332.

8. Wong S, Lau S, Woo P, Yuen K-Y (2007) Bats as a continuing source of emerging infections in humans. *Rev Med Virol* 17:67–91.
9. Vaughn DW, Hoke CH Jr. (1992) The epidemiology of Japanese encephalitis: Prospects for prevention. *Epidemiol Rev* 14:197–221.
10. Turell MJ, et al. (2008) Assay for and replication of Karshi (mammalian tick-borne flavivirus group) virus in mice. *Am J Trop Med Hyg* 78:344–347.
11. Doherty RL, et al. (1971) Studies of the epidemiology of arthropod-borne virus infections at Mitchell River Mission, Cape York Peninsula, North Queensland. IV. Arbovirus infections of mosquitoes and mammals, 1967–1969. *Trans R Soc Trop Med Hyg* 65:504–513.
12. Marshal ID (1988) *The Arboviruses: Epidemiology and Ecology*, ed Monath TP (CRC, Boca Raton, FL), Vol 3, pp 151–189.
13. Calisher CH, Gould EA (2003) Taxonomy of the virus family *Flaviviridae*. *Adv Virus Res* 59:1–19.
14. McGuire K, Holmes EC, Gao GF, Reid HW, Gould EA (1998) Tracing the origins of louping ill virus by molecular phylogenetic analysis. *J Gen Virol* 79:981–988.
15. Pfeffer M, Dobler G (2010) Emergence of zoonotic arboviruses by animal trade and migration. *Parasit Vectors* 3:35.
16. Omilabu SA, Fagbami AH, Olaleye OD (1989) Susceptibility of laboratory and domestic animals to experimental infection with Potiskum virus. *Microbios* 60:53–58.
17. Artsob H (1989) *The Arboviruses: Epidemiology and Ecology*, ed Monath TP (CRC, Boca Raton, FL), Vol 4, pp 29–49.
18. Darwish MA, Hoogstraal H, Roberts TJ, Ahmed IP, Omar F (1983) A sero-epidemiological survey for certain arboviruses (Togaviridae) in Pakistan. *Trans R Soc Trop Med Hyg* 77:442–445.
19. Weissenböck H, Hubálek Z, Bakonyi T, Nowotny N (2010) Zoonotic mosquito-borne flaviviruses: Worldwide presence of agents with proven pathogenicity and potential candidates of future emerging diseases. *Vet Microbiol* 140:271–280.
20. Reisen WK (2003) Epidemiology of St. Louis encephalitis virus. *Adv Virus Res* 61:139–183.
21. Randolph SE, Miklisova D, Lysy J, Rogers DJ, Labuda M (1999) Incidence from coincidence: Patterns of tick infestation on rodents facilitat transmission of tick-borne encephalitis virus. *Parasitol* 118:177–186.
22. Clifford CM, Yunker CE, Thomas LA, Easton ER, Corwin D (1971) Isolation of a Group B arbovirus from Ixodes uriae collected on Three Arch Rocks National Wildlife Refuge, Oregon. *Am J Trop Med Hyg* 20:461–468.
23. Kemp GE, Causey OR, Setzer HW, Moore DL (1974) Isolation of viruses from wild mammals in West Africa, 1966–1970. *J Wildl Dis* 10:279–293.
24. Baba SS, Fagbami AH, Ojeh CK, Olaleye OD, Omilabu SA (1995) Wesselsbron virus antibody in domestic animals in Nigeria: Retrospective and prospective studies. *New Microbiol* 18: 151–162.
25. Tajima S, Takasaki T, Matsuno S, Nakayama M, Kurane I (2004) Genetic characterization of Yokose virus, a *flavivirus* isolated from the bat in Japan. *Virol* 332:28–44.

Table S5. Family Paramyxoviridae

Species	Abbreviation	Accession no.	Probable host	Ref.
Avian metapneumovirus virus	AMPV	YP_443837	Birds (*Aves*)	1
Avian paramyxovirus virus 2	APMV-2	ACA49104	Passeriformes and galliformes (*Aves*)	2
Avian paramyxovirus virus 3	APMV-3	ACB46865	Passeriformes and galliformes (*Aves*)	2
Avian paramyxovirus virus 4	APMV-4	ACJ06712	Anseriformes (*Aves*)	2
Avian paramyxovirus virus 6	APMV-6	NP_150057	Anseriformes and turkeys (*Aves*)	2
Avian paramyxovirus virus 7	APMV-7	ACN72640	Columbiformes (*Aves*)	2
Avian paramyxovirus virus 8	APMV-8	ACN88139	Anseriformes (*Aves*)	2
Avian paramyxovirus virus 9	APMV-9	ACJ82939	Ducks (*Aves*)	2
Beilong virus	BeV	YP_512244	Rodents (*Rodentia*)	3
Bovine parainfluenza virus 3	BPIV-3	NP_037641	Bovines (*Artiodactyla*)	4
Bovine respiratory syncytial virus	BRSV	NP_048050	Bovines (*Artiodactyla*)	5, 6
Canine distemper virus	CDV	NP_047201	Carnivores (*Carnivora*)	7
Dolphin morbillivirus virus	DMV	NP_945024	Cetaceans (*Cetacea*)	7
Fer-de-lance virus	FDLV	NP_899654	Reptiles (*Reptilia*)	8
Goose paramyxovirus SF02	GPV	NP_872273	Birds (*Aves*)	9
Hendra virus	HeV	NP_047106	Bats (*Chiroptera*)	10
Human metapneumovirus virus	HMPV	YP_012605	Humans (*Primates*)	11
Human parainfluenza virus 1	HPIV-1	NP_604433	Humans (*Primates*)	4
Human parainfluenza virus 2	HPIV-2	NP_598401	Humans (*Primates*)	12
Human parainfluenza virus 3	HPIV-3	NP_067148	Humans (*Primates*)	4
Human parainfluenza virus 4a	HPIV-4a	P17240	Humans (*Primates*)	12
Human parainfluenza virus 4b	HPIV-4b	P17241	Humans (*Primates*)	12
Human respiratory syncytial virus	HRSV	NP_056858	Humans (*Primates*)	5
J-virus	J-V	YP_338075	Rodents (*Rodentia*)	13
Measles virus	MeV	AF504047	Humans (*Primates*)	7
Menangle virus	MenV	YP_415508	Bats (*Chiroptera*)	14
Mossman virus	MoV	NP_958048	Rodents (*Rodentia*)	15
Mumps virus	MuV	NP_054707	Humans (*Primates*)	12
Murine pneumonia virus	MPV	AAW79176	Rodents (*Rodentia*)	5
Newcastle disease virus (avian paramyxovirus 1)	NDV	NP_071466	Birds (*Aves*)	2
Nipah virus	NiV	NP_112021	Bats (*Chiroptera*)	10
Ovine respiratory syncytial virus	ORSV	Q83957	Ovines (*Artiodactyla*)	5, 16
Peste-des-petits-ruminants virus	PPRV	YP_133821	Caprines (*Artiodactyla*)	7
Phocine distemper virus	PDV	P35944	Phocidae (*Carnivora*)	7, 17
Porcine rubulavirus virus	PoRV	YP_001331027	Swine (*Artiodactyla*)	12
Porpoise morbillivirus virus	PMV	AAX47056	Cetaceans (Cetacea)	7
Rinderpest virus	RPV	YP_087120	Artiodactyls (*Artiodactyla*)	7
Sendai virus	SeV	NP_056871	Rodents (*Rodentia*)	4
Simian parainfluenza virus 5	SV-5	YP_138511	Canines and primates (*Carnivora* and *Primates*)	12
Simian virus 41	SV-41	YP_138504	Primates (*Primates*)	12
Tioman virus	TioPV	NP_665864	Bats (*Chiroptera*)	18
Tupaia virus	TuPV	NP_054690	Tree shrews (*Standentia*)	19

1. Easton AJ (2002) *The Springer Index of Viruses*, eds Tidona CA, Darai G (Springer, Berlin), pp 663–666.
2. Peeters BPH, Koch G (2002) *The Springer Index of Viruses*, eds Tidona CA, Darai G (Springer, Berlin), pp 636–644.
3. Li Z, et al. (2006) Beilong virus, a novel *paramyxovirus* with the largest genome of non-segmented negative-stranded RNA viruses. *Virology* 346:219–228.
4. Banerjee AK, De BP (2002) *The Springer Index of Viruses*, eds Tidona CA, Darai G (Springer, Berlin), pp 651–655.
5. Collins PL (2002) *The Springer Index of Viruses*, eds Tidona CA, Darai G (Springer, Berlin), pp 667–673.
6. Valarcher J-F, Schelcher F, Bourhy H (2000) Evolution of bovine respiratory syncytial virus. *J Virol* 74:10714–10728.
7. Barrett T, Rima BK (2002) *The Springer Index of Viruses*, eds Tidona CA, Darai G (Springer, Berlin), pp 645–650.
8. Kurath G, Batts WN, Ahne W, Winton JR (2004) Complete genome sequence of Fer-de-Lance virus reveals a novel gene in reptilian paramyxoviruses. *J Virol* 78:2045–2056.
9. Zou J, Shan S, Yao N, Gong Z (2005) Complete genome sequence and biological characterizations of a novel goose paramyxovirus-SF02 isolated in China. *Virus Genes* 20:13–21.
10. Eaton BT, Broder CC, Middleton D, Wang L-F (2006) Hendra and Nipah viruses: Different and dangerous. *Nat Rev Microbiol* 4:23–35.
11. Kahn JS (2006) Epidemiology of human metapneumovirus. *Clin Microbiol Rev* 19:546–557.
12. Ito Y, Tsurudome M (2002) *The Springer Index of Viruses*, eds Tidona CA, Darai G (Springer, Berlin), pp 656–659.
13. Jun MH, Karabatsos N, Johnson RH (1977) A new mouse *paramyxovirus* (J virus). *Aust J Exp Biol Med Sci* 55:645–647.
14. Bowden TR, Westenberg M, Wang LF, Eaton BT, Doyle DB (2001) Molecular characterization of Menangle virus, a novel *paramyxovirus* which infects pigs, fruit bags, and humans. *Virology* 283:358–373.
15. Miller PJ, Boyle DB, Eaton BT, Wang L-F (2003) Full-length genome sequence of Mossman virus, a novel paramyxovirus isolated from rodents in Australia. *Virology* 317:330–344.
16. Grubbs ST, Kania SA, Potgieter LND (2001) Prevalence of ovine and bovine respiratory syncytial virus infections in cattle determined with a synthetic peptide-based immunoassay. *J Vet Diagn Invest* 13:128–132.
17. Nielsen L, et al. (2009) Genetic diversity and phylogenetic analysis of the attachment glycoprotein of phocine distemper viruses of the 2002 and 1988 epizootics. *Virus Res* 144:323–328.
18. Chua KB, et al. (2001) Tioman virus, a novel *paramyxovirus* isolated from fruit bats in Malaysia. *Arch Virol* 147:1323–1348.
19. Tidona CA, Darai G (2002) *The Springer Index of Viruses*, eds Tidona CA, Darai G (Springer, Berlin), pp 660–662.

Table S6. Family Rhabdoviridae

Species	Abbreviation	Accession no.	Probable host	Ref.
Adelaide river virus	ARV	AY854635	Livestock (*Artiodactyla*)	1
Almpiwar virus	ALMV	AY854645	Reptiles (*Reptilia*)	1
Aravan virus	ARAV	EF614259	Bats (*Chiroptera*)	1
Australian bat lyssavirus	ABLV	AF081020	Bats (*Chiroptera*)	1
Berrimah virus	BRMV	AY854636	Livestock (*Artiodactyla*)	1
Bovine ephemeral fever virus	BEFV	NC_0052526	Livestock (*Artiodactyla*)	1
Chandipura virus	CHPV	AJ810083	Livestock (*Artiodactyla*)	1
Charleville virus	CHVV	AY854644	Reptiles (*Reptilia*)	1
Cocal virus	COCV	EU373657	Livestock (*Artiodactyla*)	1
Duvenhage virus	DUVV	AY854659	Bats (*Chiroptera*)	1
European bat lyssavirus 1	EBLV-1	AY854656	Bats (*Chiroptera*)	1
European bat lyssavirus 2	EBLV-2	AY854658	Bats (*Chiroptera*)	1
Flanders virus	FLAV	AF523199	Unknown	1
Fukuoka virus	FU.K.AV	AY854651	Unknown	1
Hirame rhabdovirus	HIRRV	AF104985	Fish (*Fish*)	1
Humpty Doo virus	HDOOV	AY854643	Unknown	1
Infectious hematopoetic necrosis virus	IHNV	X89213	Fish (*Fish*)	1
Irkut virus	IRKV	EF614260	Bats (*Chiroptera*)	1
Isfahan virus	ISFV	AJ810084	Unknown	1
Khujand virus	KHUV	EF614261	Bats (*Chiroptera*)	1
Kimberley virus	KIMV	AY854637	Artiodactyls (*Artiodactyla*)	1
Kotonkan virus	KOTV	AY854638	Unknown	1
Lagos bat virus	LBV	AY854654	Bats (*Chiroptera*)	1
Le Dantec virus	LDV	AY854650	Humans (*Primates*)	1
Lettuce necrotic yellows virus	LNYV	NC_007642	Compositae (*Plants*)	1
Lettuce yellow mottle virus	LYMoV	NC_011532	Lettuce (*Plants*)	1
Maize mosaic virus	MMV	AY618418	Graminae (*Plants*)	1
Maize fine streak virus	MFSV	NC_005974	Graminae (*Plants*)	1
Mokola virus	MOKV	AY854653	Unknown	1
Ngaingan virus	NGAV	AY854649	Unknown	1
Northern cereal mosaic virus	NCMV	NC_002251	Graminae (*Plants*)	1
Oak-Vale virus	OVRV	AY854670	Unknown	1
Parry Creek virus	PCRV	AY854647	Unknown	1
Perinet virus	PERV	AY854652	Unknown	1
Pike fry rhabdovirus	PFRV	FJ872827	Fish (*Fish*)	1
Rabies virus	RABV	AY854669	Carnivores and bats (*Carnivora* and *Chiroptera*)	1
Rice yellow stunt virus	RYSV	AB011257	Rice (*Plants*)	1
Siniperca chuatsi rhabdovirus	SCRV	NC_008514	Fish (*Fish*)	1
Snakehead rhabdovirus virus	SHRV	AF147498	Fish (*Fish*)	1
Sonchus yellow net virus	SYNV	L32603	Compositae (*Plants*)	1
Spring viremia of carp virus	SVCV	U18101	Fish (*Fish*)	1
Strawberry crinkle virus	SCV	AY331385	Strawberries (*Plants*)	1
Taro vein chlorosis virus	TaVCV	AY674964	Taro (*Plants*)	1
Tibrogargan virus	TIBV	AY854646	Unknown	1
Tupaia virus	TUPV	NC_007020	Tree shrews (*Scandentia*)	1
Vesicular stomatis virus Indiana	VSIV	J02428	Artiodactyls (*Artiodactyla*)	1
Vesicular stomatis virus New Jersey	VSNJV	AY074804	Artiodactyls (*Artiodactyla*)	1
Viral hemorrhagic septicaema virus	VHSV	Y18263	Fish (*Fish*)	1
West Caucasian bat virus	WCBV	EF614258	Bats (*Chiroptera*)	1
Wongabel virus	WONV	AY854648	Unknown	1

1. Kuzmin IV, Novella IS, Dietzgen RG, Padhi A, Rupprecht CE (2009) The rhabdoviruses: Biodiversity, phylogenetics, and evolution. *Infect Genet Evol* 9:541–553.