Implication of phylogenetic systematics of rodent-borne hantaviruses allows understanding of their distribution

Vincent Herbreteau, Jean-Paul Gonzalez, Jean-Pierre Hugot

Implication of Phylogenetic Systematics of Rodent-Borne Hantaviruses Allows Understanding of their Distribution

Vincent HERBRETEAU,[a] Jean-Paul GONZALEZ,[b] and Jean-Pierre HUGOT [a,b]

[a] *Institut de Recherche pour le Développement (IRD), 75231 Paris Cedex 05, France*

[b] *Muséum National d'Histoire Naturelle, Département Systématique et Évolution, UMR Origin, Structure and Evolution of Biodiversity, 75231, Paris cedex 05, France.*

KEYWORDS: rodent-borne hantaviruses, *Thottapalayam*, S gene, phylogeny, Bayesian analysis, biogeography, coevolution

ABSTRACT

Hantaviruses' distribution is reassessed after performing a cladistic analysis on ninety-three strains isolated from rodents, and one used as outgroup: *Thottapalayam* isolated from a shrew. While most hantaviruses found in wild animals were collected in Northern Asia, Europe, North America and South America, only *Thottapalayam* and *Thailand* were found in South and Southeastern Asia. *Thottapalayam* is highly divergent from the other known hantaviruses and may represent the emerging tip of a different lineage. Serological surveys carried out to detect evidence of Hantavirus in human populations revealed positive samples in West and Central Africa but also in Thailand with a first case recently confirmed. This suggests that Hantaan-related viruses may infect humans out of their well-documented range. Thus, if rodents are probably the primary reservoir, other mammals may be involved in the cycle of hantaviruses. Additional work is needed out of the traditional areas where hantaviruses have been recorded. New viruses, different hosts and different human syndromes, may be discovered in the future mainly in Southeastern Asia and in Africa where murine rodents are present and highly diversified.

INTRODUCTION

Wild rodents are the usual reservoirs of the zoonotically important hantaviruses (genus *Hantavirus*, family *Bunyaviridae*). Several serologically distinct groups have been associated with different syndromes. In the Old World, *Hantaan*, *Dobrava*, *Seoul* and *Puumala* cause the clinical forms of hemorrhagic fever with renal syndrome) (HFRS)[1.] In the New World, *Sin Nombre* and *Andes* are responsible for hantavirus pulmonary syndrome (HPS)[2.] A last group, *Tula*, widely distributed in Russia and Eastern Europe has never been associated with a human disease.

The hantavirus genome has 3 segments; large (L 6,5 kb), medium (M 3,7 kb) and small (S 1,8 kb) encoding: viral transcriptase-replicase, surface glycoprotein precursor (G1 and G2) and nucleocapsid protein respectively.[3,4] Different analyses based on the alignment of the M or S sequences.[1,5,6,7,8,9,10,11,12,13,14] have been performed and used to discuss the distribution of these viruses, relative to the biogeography and evolutionary history of their hosts. Murine rodents are the primary hosts and, because each virus group seems to be associated with a particular rodent group, the hypothesis of coevolution has been suggested.[1,3,5,6] Since the methodology to establish such coevolution has been questioned and no firm conclusions have been reached, this paper revisits the coevolution of the virus by examining the evolutionary relationship of the S genes of various hantaviruses to their respective murine hosts.

MATERIALS AND METHODS

Sequences and alignment

The data set includes *Hantavirus* S sequences, from 94 taxa, found in GenBank (Table 1): 92 isolated from different rodent hosts, one isolated in Korea from a bat (Kim, direct submission 1995) and *Thottapalayam* detected in India from a shrew *(Suncus murinus)* by Carey et al.,[16] identified by Xiao et al.[17] and recently introduced (complete S sequence) in GenBank by Schmaljohn and Toney (direct submission, 2004). Only complete CDS, whose wild host was certainly identified, were considered. The S sequence of Dugbe virus previously used as outgroup by Hughes and Friedman,[11] was first considered but finally excluded as explained in the outgroup paragraph. A virus described in Thailand from a *Bandicota indica*[1] could not be included in the data set because the S sequence of this virus remains unknown. Selected sequences range between 1,130 and 2,082 nucleotides from which: first 42 correspond to the primer and nucleotides; 1,342-2,082 to the codon stop and non-coding region. Alignment of the coding part (nucleotides 43 to 1,341) was performed using CLUSTAL-X automatic procedure, [18] then improved manually using SE-AL v2.0a11 [19] and validated using the amino acid translation. When applied to the non-coding part, the same procedure made visible the impossibility to detect real homologies between most of the sequences. Thus, only the coding part of the gene was used for phylogenetic analyses. Sequences were analyzed at nucleotide level.

Aligning and coding indels

Sequence alignment made necessary to postulate several gaps, particularly between nucleotides 766 and 813. Various approaches have been employed to deal with insertion-deletions (indels), ranging from their total exclusion to their treatment as missing data or as a fifth character state. Gaps are considered reliable characters by many systematists and the first approach means the loss of potential phylogenetic information.[20] In addition, in our data set, indels are observed within the hypervariable region where the percentage of parsimony informative characters is superior to 90% (less than 60% for the rest of the matrix). Therefore, we kept this region for analysis. Standard procedures for coding gaps suffer of several weaknesses: either the different sites are analyzed independently (gap = new state) and each gap is artificially over weighted relatively to the number of sites, or each site is coded "?" (gap = missing data) and optimization procedure makes the whole zone devoid of phylogenetic information. To express potential phylogenetic information contained in zones with internested insertions/deletions and substitutions, 8 characters coding the presence/absence of deletions between nucleotides 766 and 813 were added. Finally, the matrix includes 1,323 RNA characters and 8 presence/absence characters.

Outgroup rooting

An outgroup sequence must be closely related to the rest of the sequences, but comparatively more different than the others are between themselves. The

introduction of the *Thottapalayam* sequence within the previously aligned rodent-borne sequences: makes necessary the addition of several deletions; shows that, *Thottapalayam* possesses several conservative parts of the rodent-borne sequences. Thus, if *Thottapalayam* may certainly be considered a Hantavirus, it is highly divergent from other members of the genus. This is confirmed by the values of the total-character distances calculated using PAUP: within the rodent-borne group, distances vary from 2 to 516; between *Thottapalayam* and the others, distances range between 765 and 859. Thus *Thottapalayam* may be considered a valuable outgroup and was included in the data set.

Sequence analyses

Two methods likely to give results interpretable in an evolutionary context were used: maximum parsimony analysis (MP) and Bayesian analysis (MB). MACCLADE 4.0 [21] and TREEVIEW 1.3 [22] were used for data and tree handling and for computation of statistics. MP analysis was computed using PAUP* 4.0b10.[23] Robustness of nodes was assessed using bootstrap method [24] computed after 10,000 replicates of heuristic search with closest stepwise addition of taxa. MODELTEST 3.0 [25] was used to determine the best fitting likelihood settings: the general time reversible model [26] with among-site substitution rate heterogeneity described by a gamma distribution with eight categories [27] and a fraction of sites (INV) constrained to be invariable (GTR+I+G, selected by AIC). MB analysis using these settings was performed

using MrBayes v3.0B4[28] Bayesian approach allows defining an explicit probability model of character evolution and obtaining a rapid approximation of posterior probabilities of trees, through the use of the Markov Chain Monte Carlo (MCMC) approach. MrBayes also allows performing phylogenetic analyses of data sets combining information from different subsets, evolving under different evolutionary models. Two partitions were distinguished in our original data set: partition 1=nucleotide (characters 1-1,323) for which the likelihood model chosen was the GTR+I+G; partition 2=indels (characters 1,324-1,331) treated as presence/absence. Analysis was conducted with four independent Markov chains, run for 2,000,000 metropolis-coupled MCMC generations, with tree sampling every 10 generations and burn-in after 3,300 trees. Consensus tree was computed using the "halfcompat" option, equivalent of 50% majority rule. Proportion values of posterior probability of bipartition, considered equivalent to bootstrap values [29,30] were used for evaluation of robustness of the nodes.

Virus taxonomy

In the following and in the figures:

- Virus species listed in the Eighth Report of the International Committee on Taxonomy of Viruses [31] are in italic script.
- Strain names are in roman script, or are represented using their abbreviation in caps when an abbreviation has been proposed.
- When different strains of a same virus species are included, a number

or an adjective (generally dealing with the geographic origin) is added.

The correspondence between the virus species, strain names and abbreviations is given in Table 1.

RESULTS

General

MP or MB analyses yield consistent results. All bipartitions found by MP analysis with a bootstrap value superior or equal to 95% were also found by MB analysis with a posterior probability equal or superior to 95%. In addition, MB analysis gave a resolution and a support superior or equal to 50% for several nodes, which were unresolved, or resolved with a bootstrap inferior to 50%, in the MP analysis. Even if MB analysis is likely to favor higher values when compared to bootstrap analysis,[28,29,30] the results are fully congruent and are represented on Figure 1. Figures 2 and 3 detail the composition of the 3 main clades.

The cladogram is rooted between a basal branch corresponding with *Thottapalayam* and a monophyletic group including all the rodent-borne parasites, distributed following 3 main clades: *CLADE-1* includes *"Seoul, Hantaan, Dobrava"; CLADE-2* and *CLADE-3* are sister clades including: *"Bayou, Sinnombre, Andes",* and *"Islavista, Tula, Puumala",* respectively. Each clade and the sister-grouping of *CLADE-2* and *CLADE-3*, have a support superior or equal to 78%. *CLADE-1* groups 22 taxa: all the viruses hosted by

Murinae rodents, and the single strain found on a bat; *CLADE-2* groups 23 taxa: all the viruses hosted by Sigmodontinae rodents; *CLADE-3*, groups 48 taxa: all the viruses hosted by Arvicolinae rodents. Regarding the geographic distribution: *CLADE-1* is exclusively Palearctic, except Tchoupitoulas collected in the Nearctic (Louisiana); *CLADE-2* is found exclusively in the New World and associates strains from the Nearctic and Neotropics; *CLADE-3* may be divided into one Nearctic subclade (*Islavista*) and the sister-grouping of 2 Palearctic subclades *(Tula + Puumala).*

CLADE-1: "Seoul, Hantaan, Dobrava" (Figure 2).

Viruses hosted by *Rattus* spp. are distinguished from those hosted by *Niviventer confucianus* and *Apodemus* spp. With the exception of the parasite of *Niviventer* (considered by taxonomists closer from *Rattus*) this distribution matches the taxonomy of the rodents at genus level. However, different virus strains hosted by the same rodent species are not grouped together. The bat virus is included in *Hantaan*; its closer relative is HT.76118. Regarding the geographic distribution: *Seoul* is found in eastern China, with the exception of Sapporo (Japan) and Tchoupitoulas (Louisiana), which are sister-taxa. *Hantaan* also is restricted to the eastern part of the Palearctic region, but with a wider distribution including several provinces in China, Korea and the Amur area (Northeastern Siberia). *Dobrava* has a European distribution extending from Estonia toward Greece, through Western Russia, Slovakia and Bosnia.

The arrangement of *Dobrava* viruses on the cladogram generally fit with a North to South distribution.

CLADE-2: "Bayou, Sinnombre, Andes" (Figure 2).

From the three subclades two are hosted by Sigmondontini rodents (*Bayou, Andes*), while Neotomini rodents host *Sinnombre*. *Bayou*, found in 3 states of Southeastern North America (Florida, Louisiana and Texas) is hosted by two different genera, *Oryzomys* and *Sigmodon*. *Sinnombre* is subdivided into: a group of three taxa found in Arizona, New Mexico, Costa Rica and hosted by *Peromyscus* sp. and *Reithrodontomys* spp.; a group hosted by *Peromyscus* spp. ranging from Northeastern to Southwestern and Central United States. *Andes*, is entirely found in the Neotropics and hosted by Sigmodontini rodents: *Oligoryzomys* is the most frequent, together with several other genera (*Akodon, Bolomys, Calomys, Sigmodon*). The most divergent species in this group is *Caño Delgadito* from Venezuela; the other species are arranged following their geographic origin: Laguna Negra and Rio Marmore (Bolivia and Paraguay); *Andes*-Chile 1 and 2, (Chile); the last seven ones from Northern Argentina. Distribution of virus taxa within *CLADE-2* generally fits with the taxonomy of rodents at host tribe level and a dominant genus may be recognized for each of the main subgroups. However: the Sigmodontini parasites are not monophyletic; as in *CLADE-1*, no congruence is observed at host species level (closely related viruses hosted by different host species, viruses hosted by a same host species not closely related on the cladogram).

CLADE-3: "Prairie, Tula, Puumala" (Figure 3).

CLADE-3 is the sister group of CLADE-2 and is hosted by Arvicolinae rodents. *Tula* and *Puumala* are strictly Palearctic, *Islavista* is strictly Nearctic. *Microtus* spp. is the dominant hosts for *Islavista* and *Tula*. *Islavista* may be subdivided in two groups: Isla Vista 1, 2, 3 are Californian, Prairie Vole and Prospect Hill 1 and 2 are from South Central United States. *Tula* has a European distribution extending North to South, from Poland, Germany, Moravia, Western Russia and Slovakia. In *Puumala*: *Microtus,* associated with *Lemnus*, is present in a small basal group including three virus species found in the extreme East of Russian Siberia (Vladivostok, *Khabarovsk* and *Topografov*), the other species are hosted by *Clethrionomys rufocanus* or *C. glareolus.* Parasites of *C. rufocanus* are Japanese. Parasites of *C. glareolus* have a distribution extending from Northwestern Europe (Denmark, Belgium) to Scandinavia, Finland and South Central Russia. In *Puumala* a dominant host species may be recognized for each of the main subgroups. But, in *Islavista* and *Tula* there is no general congruence between the virus and host classifications at the species level: closely related viruses hosted by different host species; viruses hosted by a same host species, not closely related on the cladogram.

DISCUSSION

Clades, groups, robustness of the nodes and molecular data.

Our analysis confirms the 3 main clades previously described within the hantaviruses [6,11] and supports the subdivision of each clade into 3 subclades. *"Seoul", "Hantaan", "Dobrava", "Andes", "Tula", "Puumala*," already have been named. We propose new names for several new groups: "*Bayou*", including *Bayou, Black Creek, Muleshoe*; "*Sinnombre*", including: *Sin Nombre*, Convict, *Monongahela, New York* and *Limestone*, *El Moro, Rio Segundo*; *"IslaVista*", including: Prairie Vole, *Prospect Hill, Isla Vista*. The support for corresponding nodes of the cladogram is generally between 80 and 100. The alignment shows that main clades and subclades are supported by amino acid changes caused by synonymous or non-synonymous nucleotide differences. Most changes occur in a hypervariable (HV) region identified by several previous studies [32,33]. Hughes and Friedman [11] defined the HV region as residues 242–281. In our alignment, HV region corresponds to amino acid residues 249-317 and includes 92% of informative sites (62% in whole matrix). This region also includes several regular indels corresponding with the main subdivisions of the cladogram.

Host specificity and correspondence with host taxonomy

The topology of the 3 main clades matches the phylogeny of the 3 host subfamilies to which they are respectively devolved. Dominant host genera

(pointed by arrows on Figure 1) may be recognized by mapping the host genus, as a character, on the cladogram of the virus. Within *CLADE-1*, *Rattus* is dominant for *Seoul*, *Apodemus* for the association *Hantaan-Dobrava*. Within *CLADE-2*, each subclade is linked to a particular host tribe (Neotomini or Sigmodontini); *Sigmodon* appears as a potential primitive host; *Peromyscus* and *Oligorizomys* are the dominant host genera for *Sinnombre* and *Andes*, respectively. Within *CLADE-3*, *Microtus* appears as a primitive host genus, and *Clethrionomys* is the dominant host genus for the *Puumala* subclade. The good correspondence of the phylogenies at their highest level is consistent with the hypothesis of coevolution: the Hantavirus and the Muridae may have evolved and dispersed in parallel. But, whatever the clade considered, there is a mismatch of the host and parasite distributions at species level. This looks like if host specificity disappeared somewhere between the species and/or genus level. Depending which clade is considered, this limit is variable: host switching at genus level appears difficult and unlikely within *CLADE-1* and *CLADE-3*, and easier in *CLADE-2*; within *CLADE-2* the highest diversity, thus weakest specificity at genus level, is observed in *Andes*.

Biogeography of rodent-borne hantaviruses

CLADE-1 is Palearctic except *Tchoupitoulas*, reported from a wild *Rattus norvegicus* in New Orleans. *R. norvegicus* is a cosmopolitan species, which dependence for human quarters is well known and the presence of this Hantavirus in the New World may probably be, interpreted a case of

dispersion by humans. *CLADE-2* is exclusively found in the New World: Figure 2 shows that, unexpectedly following the hypothesis of coevolution, the parasites of the Nearctic Sigmodontini (*Bayou*) are not closely related to the parasites of Neotropical Sigmodontini (*Andes*). Most of the Sigmodontini biodiversity is found in the Neotropics while their sister group, the Neotomini, is dominant in North America. *Bayou* seems limited to Southeastern United States, and may perhaps be interpreted resulting of an ancient isolation of its hosts in a remote part of their range. *CLADE-3* has a mixed distribution with one small Nearctic subclade (*Isla Vista*) and 2 Palearctic subclades (*Tula, Puumala*). *Prairie* and *Tula* are hosted by different species of genus *Microtus,* as are Vladivostok and Khabarovsk which are the sister group of *Puumala* (hosted by *Clethrionomys* spp.). This distribution is consistent with a Palearctic origin, a passage into the New World probably transported by the Arvicolinae (*Microtus* looks a good candidate), a later dispersion in North and South America following the migrations of the Sigmodontinae. This scenario mimics the scenario generally accepted for the radiation of the Muridae starting from their South Asian center of origin and is compatible with the hypothesis of a parallel evolution. Within the subclades a different pattern is suggested, because transmission between different rodent species in a same genus (and between different genera in the Neotropics) looks possible.

CONCLUSION

Two different patterns of dispersion are explaining the evolution of the hantaviruses: the first one, characterized by a strong specificity for a particular group of hosts, explains the ancient history of this group and its coevolution with the Muridae; the second one, characterized by a slack specificity, is corresponding with the recent and current history of the viruses and their opportunistic circulation by using contacts between closely related rodent genera, species and/or populations. This second pattern explains why, when the distribution of the hosts and parasites is enough documented *(Dobrava, Tula, Puumala, Andes*), a geographic gradient become visible. Different pattern, following different specificity is in agreement with what is known about Hantavirus survival outside their hosts. Sauvage et al.[34] considering the role of indirect transmission on virus persistence, suggest that viruses remain active outside the host, which could permit transmission without physical contact with the infectious rodents. This explains how hantaviruses may switch when the specific barrier is low and when different hosts have overlapping territories.

While most Bunyaviridae are hosted by arthropods, hantaviruses have rodents as principal hosts. However, 2 strains have been isolated from non-rodent mammals: *Thottapalayam*, isolated from a shrew; the *Hantaan* virus isolated from a bat. Considering its strong differences with the other hantaviruses, *Thottapalayam* cannot be interpreted resulting from a recent host switching

between a rodent and a shrew. Further investigations are needed to decide if this adaptation to a different group of mammals is an exception, or may represent the emerging tip of a different lineage.

The bat virus is included in *Hantaan*; its closer relative is HT.7611. No significant difference of branch length is observed between the two strains and their Total-character distance equals 4; this suggests that the two sequences are almost identical: thus, the presence of a Hantavirus in *R. ferrumequinum* must probably be interpreted as the result of an horizontal transfer.

Most of hantaviruses found in wild animals were collected in the Holarctic, or the Neotropics (Northern Asia, Europe, North America, and South America). But, *Thottapalayam* comes from South Asia, where is also found *Thailand*, hosted by *Bandicota indica*, a Muridae rodent. Serological surveys carried out to detect evidence of Hantavirus infection in human populations revealed that: in Thailand, in different provinces and/or in different environments, 1,2% to 31,4% of individual tested had Hantavirus antibody[35,36,37]; the recent publication of the first human case in Thailand[38] confirms the presence of Hantavirus in South Asia; also, screenings performed in West and Central Africa where a human case has not yet been reported, show that humans may have been infected by Hantaan-related virus[39.] All this suggests that, if rodents are probably the primary reservoir, other mammals may be involved in the cycle of hantaviruses; new viruses, different hosts and different human syndromes may be expected to be discovered in the future. Additional work is

needed out of the traditional areas where Hantaviruses have been recorded, mainly in Southeastern Asia and in Africa where Muridae rodents are present and highly diversified.

REFERENCES

01. Schmaljohn, C. A., A. Schmaljohn, and J. Dalrymple. 1987. Hantaan virus M RNA: coding strategy, nucleotide sequence, and gene order. Virology 157:31–39.

02. Childs, J. E., T. G. Ksiazek, C. F. Spiropoulou, J. W. Krebs, S. Morzunov, G. O. Maupin, K. L. Gage, P. E. Rollin, J. Sarisky, R. E. Enscore, J. K. Frey, C. J. Peters, and S. T. Nichol. 1994. Serologic and genetic identification of *Peromyscus maniculatus* as the primary rodent reservoir for a new hantavirus in the southwestern United States. J. Infect. Dis. 169:1271–1280.

03. Schmaljohn, C., G. Jennings, J. Hay, and J. M. Dalrymple. 1986. Coding strategy of the S-genome segment of Hantaan virus. Virology 155:633–643.

04. Elliot, R. M., C. S. Schmaljohn, and M. S. Collett. 1991. Bunyaviridae genome structure and gene expression. Curr. Top. Microbiol. Immunol. 69:91–141.

05. Dekonenko, A., V. Yakimenko, A. Ivanov, V. Morozov, P. Nikitin, S. Khasanova, T. Dzagurov, E. Tkachenko, and C. Schmaljohn. 2003.

Genetic similarity of Puumala viruses found in Finland and western Siberia and of the mitochondrial DNA of their rodent hosts suggests a common evolutionary origin. Infect. Gen. Evol. 3:245–257.

06. Nichol, S.T. 1999. Genetic analysis of Hantaviruses and their host relationships. In: Saluzzo JF, Dodet B, eds. Factors in the emergence and control of rodent-borne viral diseases. Paris, France: Elsevier SAS:99-109.

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

08. Heisk,e A., B. Anheier, J. Pilaski, V. E. Volchkov, and H. Feldmann. 1999. A new *Clethrionomys*-derived Hantavirus from Germany: evidence for distinct genetic sublineages of Puumala viruses in Western Europe. Virus Res. 61:101–112.

09. Hjelle, B., F. Chavez-Giles, N. Torrez-Martinez, T. Yates, J. Sarlisky, J. Webb, and M. Ascher. 1994. Genetic identification of a novel Hantavirus of the harvest mouse *Reithrodontomys megalotis.* J. Virol. 68:6751–6754.

10. Horling, J., V. Chizhikov, A. Lundkvist, M. Jonsson, L. Ivanon L, A. Dekonenko, B. Niklasson, T. Dzagurova, C. J. Peters, E. Tkachenko, and S. Nichol. 1996. Khabarovsk virus: a phylogenetically and serologically

distinct Hantavirus isolated from *Microtus fortis* trapped in far-east Russia. J. Gen. Virol. 77:687–694.

11. Hughes, A. L., and R. Friedman. 2000. Evolutionary diversification of protein-coding genes of Hantaviruses. Mol. Biol. Evol. 17:1558–1568.

12. Kariwa, H., K. Yoshimatsu, and J. Sawabe. 1999. Genetic diversities of Hantaviruses among rodents in Hokkaido, Japan and Far East Russia. Virus. Res. 59:219– 228.

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

14. Lopez, N., P. Padula, C. Rossi, S. Miguel, A. Edelstein, E. Ramirez E, and M. T. France-Fernandez. 1997. Genetic characterization and phylogeny of Andes virus and variants from Argentina and Chile. Virus Res. 50:77–84.

15. Li, W.-H. 1997. Molecular Evolution. Sinauer As., Sunderland, MA.

16. Carey, D,, R. Reuben, K. Panicker, R. Shope, and R. Myers. 1971. Thottapalayam virus: a presumptive arbovirus isolated from a shrew in India. Indian J. Med. Res. 59:1758-60.

17. Xiao, S.-Y., J. W. Leduc, Y. K. Chu, and C. S. Schmaljohn. 1994. Phylogenetic analyses of virus isolates in the genus Hantavirus, family Bunyaviridae. Virology 198:205–217.

18. Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position- specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673–4680.

19. Rambaut, A. Se-Al. Sequence Alignment Editor. Ver. 1.0 alpha 1. 1996. University of Oxford, Oxford, U.K.

20. Barriel, V. 1994. Phylogénies moléculaires et insertions-délétions de nucléotides. CR Acad Sci Sér III 317:693–701.

21. Maddison, D. R. W., and P. Maddison. MacClade 4: Analysis of phylogeny and character evolution. Version 4.0. 2000 Sinauer Associates, Sunderland, Massachusetts.

22. Page, R. D. M. 1996. TreeView: An application to display phylogenetic trees on personal computers. Comp. Appl. Bios. 12:357-358.

23. Swofford, D. L. PAUP*. Phylogenetic Analysis Using Parsimony (and Other Methods). Version 4. 0b10 2001 Sinauer Associates, Sunderland, Massachusetts.

24. Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791.

25. Posada, D., and K. A. Crandall. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14:817-818.

26. Yang Z. 1994. Estimating the pattern of nucleotide substitution. J. Mol. Evol. 39:105–111.

27. Yang Z. 1996. Among-site rate variation and its impact on phylogenetic analyses. Trends Ecol. Evol. 11:367–372.

28. Huelsenbeck, J. P., and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17:754–755.

29. Cummings, M. P.,S. A. Handley, D. S. Myers, D. L. Reed, A. Rokas, and K. Winka. 2003. Comparing Bootstrap and Posterior Probability Values in the Four-Taxon Case. Syst. Biol. 52:477–487.

30. Zhaxybayeva, O., and J. P. Gogarten. 2002. Bootstrap, Bayesian probability and maximum likelihood mapping: exploring new tools for comparative genome analyses. BMC Genomics 3:1-15.

31. Nichols S.T., Beaty B.J., Elliott R.M., Goldbach R., Plyusnin A., Schmaljohn C.S. and Tesh R.B. (2006). Family Buyaviridae. In: Eight Report of the International Committee on Taxonomy of Viruses. Eds C. Fauquet, M. Mayo, J. Maniloff, U. Desselberger, L. A. Ball. Elsevier/Academic Press.

32. Lundkvist, A., H. Kallio-Kokko, K. B. Sjölander, H. Lankinen, B. Niklasson, A. Vaheri, and O. Vapalahti. 1996. Characterization of Puumala virus nucleocapsid protein : identification of B-cell epitopes. Virology 216:397-406.

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

34. Sauvage, F., M. Langlais, N. G. Yoccoz, and D. Pontier. 2003. Modelling Hantavirus in fluctuating populations of bank voles: the role of indirect transmission on virus persistence. Journal of Animal Ecology 72:1–13.

35. Elwell, M. R., G. S. Ward, M. Tingpalapong, J. W. Leduc. 1985. Serologic evidence of Hantaan-like virus in rodents and man in Thailand. Southeast Asian J. trop. Med. Public Health 16:349-354.

36. Nitatpattana, N., G. Chauvency, J. Dardaine, T. Poblap, K. Jumronsawat, W. Tangkanakul, D. Poonsuksombat, S. Yoksan, and J.-P. Gonzalez. 2000. Serological study of Hantavirus in the rodent population of Nakhon Pathom and Nakhon Ratchasima provinces in Thailand. Southeast Asian J. Trop. Med. Public Health 31:277-282.

37. Sawasdikol, S., M. Tamura, and P. Jamjit. 1989. Antibody to hemoragic fever with renal syndrome in man and rat in Thailand. Bull. Dept. Med. Sci. 31:125-130.

38. Suputtamongkol Y, Nitatpattana N., Chyakulkeree M, Palabodeewat S., S Yoksan, & JP Gonzalez, 2005. Hantavirus infection in Thailand: first clinical case report. Southern Asian J Trop. Med. & Pub. Health, 36,1:217-220.

39. Gonzalez, J.-P., J. B. McCormick, D. Baudon, J. P. Gautun, D. Y. Meunier, E. Dournon, and A. J. Georges. 1984. Serological evidence for hantaan-related virus in Africa. The Lancet 324:1036-1037.

FIGURE CAPTIONS

FIGURE 1. Cladogram resulting of Bayesian analysis using GTR+I+G model. Names given to the main groups according to previous publications, except *"Bayou"* and *"Islavista"* which are proposed as new names. Different color patterns are attributed to different geographic areas. Arrows point out optimization of the host genera on the cladogram.

FIGURE 2. Detail of *CLADE-1* and *CLADE-2* of Figure 1. Posterior probability numbered when inferior to 95% (probability of no numbered nodes between 95 and 100). For each virus strain the scientific name of the host is given; different color patterns are attributed to different host groups and to different biogeographic areas. Abbreviation: *Reith.* = *Reithrodontomys. Oligo.* = *Oligorizomys.*

FIGURE 3. Detail of *CLADE-3* of Figure 1. See legend of Figure 2.

TABLE 1. List of Hantavirus strains included in present study. First two columns: name of strain and acronym used in text and figures. Columns 3 and 4: scientific and Family names of principal host. Columns 5 and 6: accession number of sequence and number of nucleotides given in Genbank. Region refers to biogeographic areas: Palearctic (PAL), Nearctic (NEA), Neotropical (NEO), Oriental (ORIENT). Column 7: in which country, and when possible in which province or locality, virus strain has been collected.

	Virus species & strain	**Abbreviation**	**Host species**	**Family**	**Accession nb**	**Nucl**	**Region**	**Distribution**
1	*Dobrava*-Estonia	DOBV-Estonia1	*Apodemus agrarius*	Murinae	AJ009773	1671	PAL	Estonia (Saaremaa)
2	*Dobrava*-Estonia	DOBV-Estonia2	*Apodemus agrarius*	Murinae	AJ009775	1671	PAL	Estonia (Saaremaa)
3	*Dobrava*-Slovakia	DOBV-Slovakia1	*Apodemus agrarius*	Murinae	AJ269549	1704	PAL	Slovakia (Kosice)
4	*Dobrava*-Bosnia	DOBV-Bosnia	*Apodemus flavicollis*	Murinae	L41916	1670	PAL	Bosnia
5	*Dobrava*-Greece	DOBV-Greece1	*Apodemus flavicollis*	Murinae	AJ410615	1290	PAL	Greece (Northeast)
6	*Dobrava*-Greece	DOBV-Greece2	*Apodemus flavicollis*	Murinae	AJ410619	1290	PAL	Greece (Northeast)
7	*Dobrava*-Russia	DOBV-Russia1	*Apodemus sylvaticus*	Murinae	AF442623	1637	PAL	Russia (Krasnodar)
8	*Dobrava*-Russia	DOBV-Russia2	*Apodemus sylvaticus*	Murinae	AF442622	1196	PAL	Russia (Goryachiy)
9	*Dobrava*-Slovakia	DOBV-Slovakia2	*Apodemus sylvaticus*	Murinae	AJ269550	1704	PAL	Slovakia (Kosice)
10	*Hantaan*-76118	HTNV-76118	*Apodemus sylvaticus*	Murinae	M14626	1696	PAL	South Korea
11	*Hantaan*-Maaji	HTNV-Maaji	*Apodemus agrarius*	Murinae	AF321094	1700	PAL	Korea (Maaji)
12	*Hantaan*-Amur AP61	AMRV.AP61	*Apodemus peninsulae*	Murinae	AB071183	1290	PAL	Russia FE (Solovey)
13	*Hantaan*-Amur AP63	AMRV.AP63	*Apodemus peninsulae*	Murinae	AB071184	1696	PAL	Russia FE (Solovey)
14	*Hantaan*-Guizhou	HTNV-Guizhou	*Apodemus sylvaticus*	Murinae	AB027097	1635	PAL	China (Guizhou)
15	*Hantaan*-Anhui	HTVN-Anhui	*Niviventer confucianus*	Murinae	AB027523	1654	PAL	China (Anhui)
16	***Hantaan*-Bat**	**HTNV-Bat**	***Rhinolophus ferrumequinum***	**Rinolophidae**	**U37768**	**1696**	**PAL**	**Korea**

17	*Seoul*-L99	SEOV-L99	*Rattus losea*	Murinae	AF288299	1764	PAL	China (Jiangxi)
18	*Seoul*-Sapporo	SEOV-Sapporo	*Rattus norvegicus*	Murinae	M34881	1769	PAL	Japan (Sapporo)
19	*Seoul*-Shanxi	SEOV-Shanxi	*Rattus rattus*	Murinae	AF288643	1772	PAL	China (Shanxi)
20	*Seoul*-Tchoupitoulas	SEOV-Tchoupi	*Rattus rattus*	Murinae	AF329389	1785	NEA	USA (Louisiana)
21	*Seoul*-Zhejiang	SEOV-lZhejiang1	*Rattus rattus*	Murinae	AB027522	1692	PAL	China (Zhejiang)
22	*Seoul*-Zhejiang	SEOV-Zhejiang2	*Rattus rattus*	Murinae	AF288653	1772	PAL	China (Zhejiang)
23	*Sin Nombre*	SNV	*Peromyscus maniculatus*	Neotominae	L25784	2059	NEA	USA (S-West & Central)
24	*SinNombre*-Convict Creek	SNV-Conv.74	*Peromyscus maniculatus*	Neotominae	L33683	1287	NEA	USA (California)
25	*SinNombre*-Convict Creek	SNV-Conv.107	*Peromyscus maniculatus*	Neotominae	L33816	1287	NEA	USA (California)
26	*SinNombre*-Monongahela	MGLV	*Peromyscus maniculatus*	Neotominae	U32591	2082	NEA	USA (Appalachian)
27	*NewYork*-RI1	*NYork*-RI1	*Peromyscus leucopus*	Neotominae	U09488	2078	NEA	USA (North East)
28	Limestone Canyon	*Limestone Canyon*	*Peromyscus boylii*	Neotominae	AF307322	1209	NEA	USA (Arizona)
29	*El Moro Canyon*	*El Moro Canyon*	*Reithrodontomys megalotis*	Neotominae	U11427	1896	NEA	USA (New Mexico)
30	Rio Segundo	RioSegundo	*Reithrodontomys mexicanus*	Neotominae	U18100	1749	NEO	Costa Rica
31	*Andes*-AH1	ANDV-AH1	*Oligoryzomys longicaudatus*	Sigmodontinae	AF004660	1876	NEO	Argentina
32	*Andes*-Bermejo	BMJV	*Oligoryzomys chacoensis*	Sigmodontinae	AF482713	1933	NEO	Argentina (Oran)
33	*Andes*-Chile	ANDV-Chile1	*Oligoryzomys longicaudatus*	Sigmodontinae	AF291702	1871	NEO	Chile (Aysen)
34	*Andes*-Chile	ANDV-Chile2	*Oligoryzomys longicaudatus*	Sigmodontinae	NC003466	1871	NEO	Chile (Aysen)
35	*Andes*-Lechiguanas	LECV	*Oligoryzomys flavescens*	Sigmodontinae	AF482714	1938	NEO	Argentina (Lechiguana)
36	*Andes*-Norte	ANDV-Norte	*Oligoryzomys chacoensis*	Sigmodontinae	AF325966	1921	NEO	Argentina Norte
37	*Andes*-Oran	ORNV	*Oligoryzomys longicaudatus*	Sigmodontinae	AF482715	1919	NEO	Argentina (Oran)
38	*Andes*-Pergamino	PRGV	*Akodon azarae*	Sigmodontinae	AF482717	1860	NEO	Argentina
39	Maciel	Maciel	*Bolomys benefactus*	Sigmodontinae	AF482716	1869	NEO	Argentina (Maciel)
40	*Laguna Negra*	*Laguna Negra*	*Calomys laucha*	Sigmodontinae	AF005727	1904	NEO	Paraguay, Bolivia
41	Rio Mamore	*RioMamore*	*Oryzomys microtis*	Sigmodontinae	U52136	1975	NEO	Bolivia
42	*Bayou*	*Bayou*	*Oryzomys palustris*	Sigmodontinae	L36929	1958	NEA	USA (Louisiana)
43	*Black Creek Canal*	*Black Creek*	*Sigmodon hispidus*	Sigmodontinae	L39949	1989	NEA	USA (Florida)
44	*Muleshoe*	*Muleshoe*	*Sigmodon hispidus*	Sigmodontinae	U54575	1989	NEA	USA (Texas)
45	*Caño Delgadito*	*CanoDelgadito*	*Sigmodon alstoni*	Sigmodontinae	AF000140	1130	NEO	Venezuela (Portuguesa)

46	*Isla Vista*	*Isla Vista 1*	*Microtus californicus*	Arvicolinae	U19302	1720	NEA	USA (California)
47	*Isla Vista*	*Isla Vista 2*	*Microtus californicus*	Arvicolinae	U31534	1720	NEA	USA (California)
48	*Isla Vista*	*Isla Vista 3*	*Microtus californicus*	Arvicolinae	U31535	1302	NEA	USA (California)
49	*Prospect Hill*	*ProspectHill1*	*Microtus montanus*	Arvicolinae	M34011	1675	NEA	USA
50	*Prospect Hill*	*ProspectHill2*	*Microtus montanus*	Arvicolinae	Z49098	1675	NEA	USA
51	Prairie Vole	PrairieVole	*Microtus ochrogaster*	Arvicolinae	U19303	1722	NEA	USA (?)
52	*Topografov*	*Topografov*	*Lemmus sibiricus*	Arvicolinae	AJ011646	1951	PAL	Russia FE (Taymyr)
53	*Khabarovsk*	*Khabarovsk*	*Microtus fortis*	Arvicolinae	U35255	1845	PAL	Russia FE (Khabarovsk)
54	Vladivostock	Vladivostock	*Microtus fortis*	Arvicolinae	AB011630	1228	PAL	Russia FE (Vladivostok)
55	*Tula*-Germany1	TULV-Germany1	*Microtus arvalis*	Arvicolinae	AF164093	1832	PAL	Germany
56	*Tula*-Germany2	TULV-Germany2	*Microtus arvalis*	Arvicolinae	AF289821	1828	PAL	Germany
57	*Tula*-Lodz	TULV-Lodz1	*Microtus arvalis*	Arvicolinae	AF063892	1852	PAL	Poland
58	*Tula*-Lodz	TULV Lodz2	*Microtus arvalis*	Arvicolinae	AF063897	1852	PAL	Poland
59	*Tula*-Moravia	TULV-Moravia	*Microtus arvalis*	Arvicolinae	Z69991	1831	PAL	Moravia
60	*Tula*-Slovakia	TULV-Slovakia1	*Microtus arvalis*	Arvicolinae	AJ223601	1831	PAL	Slovakia (Koziky)
61	*Tula*-Slovakia	TULV-Slovakia2	*Microtus arvalis*	Arvicolinae	AJ223600	1831	PAL	Slovakia (Koziky)
62	*Tula*-Slovakia	TULV-Slovakia3	*Microtus arvalis*	Arvicolinae	Z48235	1831	PAL	Slovakia (Malacky)
63	*Tula*-Slovakia	TULV-Slovakia4	*Microtus arvalis*	Arvicolinae	Y13979	1833	PAL	Slovakia (Kosice)
64	*Tula*-Slovakia	TULV-Slovakia5	*Microtus arvalis*	Arvicolinae	Y13980	1832	PAL	Slovakia (Kosice)
65	*Tula*-Slovakia	TULV-Slovakia6	*Microtus arvalis*	Arvicolinae	Z68191	1831	PAL	Slovakia (Malacky)
66	*Tula*-Russia	TULV-Russia	*Microtus gregalis*	Arvicolinae	Z30941	1847	PAL	Russia (Tula)
67	*Tula*-Serbia	TULV-Serbia	*Microtus subterraneus*	Arvicolinae	AF017659	1834	PAL	Serbia (Cacac)
68	*Puumala*-Bashkortostan	PUUV-Bashkor	*Clethrionomys glareolus*	Arvicolinae	AF442613	1733	PAL	Russia (Bashkortostan)
69	*Puumala*-Belgium	PUUV-Belgium	*Clethrionomys glareolus*	Arvicolinae	AJ277030	1837	PAL	Belgium (Thuin)
70	*Puumala*-CG1820	PUUV-CG1820	*Clethrionomys glareolus*	Arvicolinae	M32750	1784	PAL	?
71	*Puumala*-Denmark	PUUV-Denmark	*Clethrionomys glareolus*	Arvicolinae	AJ238791	1831	PAL	Denmark
72	*Puumala*-Evo	PUUV-Evo	*Clethrionomys glareolus*	Arvicolinae	Z30703	1832	PAL	Finland
73	*Puumala*-Kamiiso	HOKV-Kamiiso	*Clethrionomys rufocanus*	Arvicolinae	AB010730	1833	PAL	Japan (Hokkaido)
74	*Puumala*-Japan	HOKV-Japan	*Clethrionomys rufocanus*	Arvicolinae	AB010731	1833	PAL	Japan (Tobetsu)

75	*Puumala*-Karelia	PUUV-Karelia1	*Clethrionomys glareolus*	Arvicolinae	AJ238790	1832	PAL	Russia (Karelia, Gomselga)
76	*Puumala*-Karelia	PUUV-Karelia2	*Clethrionomys glareolus*	Arvicolinae	AJ238788	1828	PAL	Russia (Karelia, Karhumaki)
77	*Puumala*-Karelia	PUUV-Karelia3	*Clethrionomys glareolus*	Arvicolinae	AJ238789	1830	PAL	Russia (Karelia, Kolodozero)
78	*Puumala*-Kazan	PUUV-Kazan	*Clethrionomys glareolus*	Arvicolinae	Z84204	1826	PAL	Sweden?
79	*Puumala*-Norway	PUUV-Norway1	*Clethrionomys glareolus*	Arvicolinae	AJ223369	1849	PAL	Norway (Eidsvoll)
80	*Puumala*-Norway	PUUV-Norway3	*Clethrionomys glareolus*	Arvicolinae	AJ223374	1828	PAL	Norway (Mellansel)
81	*Puumala*-Norway	PUUV-Norway4	*Clethrionomys glareolus*	Arvicolinae	AJ223375	1829	PAL	Norway (Mellansel)
82	*Puumala*-Norway	PUUV-Norway5	*Clethrionomys glareolus*	Arvicolinae	AJ223376	1871	PAL	Norway (Solleftea)
83	*Puumala*-Norway	PUUV-Norway6	*Clethrionomys glareolus*	Arvicolinae	AJ223377	1882	PAL	Norway (Solleftea)
84	*Puumala*-Norway	PUUV-Norway7	*Clethrionomys glareolus*	Arvicolinae	AJ223380	1827	PAL	Norway (Tavelsjo)
85	*Puumala*-Omsk	PUUV-Omsk1	*Clethrionomys glareolus*	Arvicolinae	AF367067	1732	PAL	Omsk-Russia (W Siberia)
86	*Puumala*-Omsk	PUUV-Omsk2	*Clethrionomys glareolus*	Arvicolinae	AF367068	1732	PAL	Omsk-Russia (W Siberia)
87	*Puumala*-Omsk	PUUV-Omsk3	*Clethrionomys glareolus*	Arvicolinae	AF367069	1732	PAL	Omsk-Russia (W Siberia)
88	*Puumala*-Omsk	PUUV-Omsk4	*Clethrionomys glareolus*	Arvicolinae	AF367070	1732	PAL	Omsk-Russia (W Siberia)
89	*Puumala*-Slovakia	PUUV-Slovakia	*Clethrionomys glareolus*	Arvicolinae	AF294652	1809	PAL	Slovakia
90	*Puumala*-Sotkamo	PUUV-Sotkamo	*Clethrionomys glareolus*	Arvicolinae	X61035	1830	PAL	Finland (Sotkamo)
91	*Puumala*-Udmurtia	PUUV-Udmurtia	*Clethrionomys glareolus*	Arvicolinae	Z21497	1827	PAL	Finland (Udmurtia)
92	*Puumala*-Vranica	PUUV-Vranica	*Clethrionomys glareolus*	Arvicolinae	U14137	1828	PAL	Bosnia (Vranica)
93	***Thottapalayam***	**Thottalayam**	***Suncus murinus***	**Soricidae**	**AY526097**	**1530**	**ORIENT**	**India (Thottalayam)**