JOURNAL OF VIROLOGY, Dec. 1996, p. 9060–9063 Vol. 70, No. 12
0022-538X/96/$04.00+0

Quasispecies in Wild-Type Tula Hantavirus Populations

ALEXANDER PLYUSNIN,[1]* YING CHENG,[1] HEIKKI LEHVÄSLAIHO,[2] AND ANTTI VAHERI[1]

Haartman Institute, Department of Virology, FIN-00014 University of Helsinki, Helsinki,[1] and Center for Scientific Computing, FIN-02101 Espoo,[2] Finland

Received 25 March 1996/Accepted 5 September 1996

Tula virus (TUL) is a recently detected hantavirus carried by European common voles. Reverse transcriptase PCR cloning was used to study TUL S segment/N protein quasispecies. Both the distribution and character of mutations observed in three mutant spectra indicated limited selection at the protein level. At least 8% of the mutations were neutral or well tolerated; fixation of such mutations may play a role in TUL evolution in its natural host.

Hantaviruses (family *Bunyaviridae*) are worldwide-distributed agents carried by persistently infected rodent or insectivore hosts. Some hantaviruses are human pathogens, and others are thought to be apathogenic. Their RNA genome is of negative polarity and consists of three segments, large (L), medium (M), and small (S), which encode a viral polymerase, surface glycoproteins (G1 and G2), and a nucleocapsid protein (N), respectively (for a review, see reference 19). Both genetic drift (8, 15–18, 22, 23) and genetic shift (5, 7, 11) seem to be involved in the evolution of hantaviruses, similar to that of other members of the *Bunyaviridae* family (10). The evolution of hantaviruses follows that of their natural carriers (1, 15, 16, 24), which should be regarded as the evolutionary scene for these agents.

Tula hantavirus (TUL) was recently detected in European common voles (*Microtus arvalis*) (16). Populations of wild-type TUL and Puumala hantavirus (PUU) from individual animals are represented by complex mixtures of different but closely related variants (16, 17). Thus, hantaviruses resemble other RNA viruses existing in the form of quasispecies that allows for rapid evolution, via selection from preexisting variants, resulting in the establishing of altered mutant spectra with a higher fitness to a new environment (for reviews, see references 4 and 6). The aim of this study was to investigate in detail the nature of nucleotide (nt) and deduced amino acid (aa) substitutions in the S segment encoding the N protein (S/N reference) of three closely related TUL strains. These strains, which were in circulation within the same local rodent population, are designated TUL/23Ma/87 (or T23, for short), TUL/53Ma/87 (T53), and TUL/175Ma/87 (T175) (16).

Reverse transcriptase PCR of total RNA isolated from lung tissue was performed as described earlier (16, 18). PCR amplicons corresponding to the entire S segment (about 1,800 bp) were cloned by using the pGEM-T cloning system (Promega). Altogether, 117 clones were selected as follows: 19 from strain T23, 40 from T53, and 58 from T175. Sequencing (Sequenase version 2.0; United States Biochemical) was carried out to obtain partial sequences of the S segment (nt 801 to 1530). The program suite Genetic Data Environment (21) was used to analyze the sequence data.

Mutation frequency and distribution of substitutions in TUL S/N quasispecies. The part of the S segment selected for our analysis (Fig. 1A) included nt 801 to 1335 from the coding region and 195 bases of the 3′ noncoding region. TUL strains T23, T53, and T175 differ by 6 to 19 nt within this region, giving an opportunity to compare closely related RNA sequences. As for the N protein, its C-terminal part (aa 249 to 430) is even more conserved and contains only two sites of substitutions. In the quasispecies analysis, a total of 81.9 kb were sequenced, and 74 nucleotide mutations that corresponded to 45 deduced amino acid substitutions were found (Fig. 1B). The distribution of nucleotide substitutions varied from strain to strain, and no obvious correlation to the distribution of the S segment genetic variability calculated for TUL master sequences (Fig. 1A) was observed. No identical mutations were found in the virus populations from different animals, suggesting rarity of the events. Nevertheless, there were four sites where substitutions in more than one mutant spectrum were found. The first site was formed by nt 898 to 900 (strain T53, clone 33 [T53/33] and T175/48). The second site consisted of nt 913 to 915 (T175/40, T175/21, and T23/14). Variants with mutated nt 1312 to 1313 were found in mutant spectra of strains T23 and T53 and those with mutated nt 1092 were found in mutant spectra of strains T23 and T175. These "hot spots" might be related to peculiarities of the template RNA (G+C content or secondary structure, etc.).

The mutation frequency in TUL quasispecies ($\sim 1 \times 10^{-3}$) (Table 1) is comparable to the corresponding values calculated for the PUU S nucleotide quasispecies (3×10^{-3}) (17) and also for TUL and PUU M nucleotide quasispecies (our unpublished observations). These frequencies are well within the range estimated for other RNA viruses, i.e., 10^{-3} to 10^{-5} (4), reflecting similar misincorporation capacities of viral RNA polymerases responsible for the appearance of mutated RNA molecules in the absence of proofreading and repair mechanisms. The frequencies of nucleotide substitutions calculated for the coding and noncoding regions were equal, and the frequency of deduced amino acid substitutions was approximately twice as high as that of nucleotide substitutions. These data, taken together with an even distribution of mutations between codon positions and an absence of any specific pattern in deduced amino acid substitutions (see below), suggest that selection operating at the protein level in the mutant spectra is limited.

The frequencies of nucleotide substitutions calculated for three studied virus populations (Table 1) were four- to fivefold higher than the level of Ampli*Taq* polymerase misincorporation, 0.2×10^{-3}, previously determined for our system by using the PUU S segment (17). Control experiments with reampli-

* Corresponding author. Mailing address: Haartman Institute, Department of Virology, POB 21, FIN-00014 University of Helsinki, Helsinki, Finland. Phone: 358-0-434 6486. Fax: 358-0-434 6491. Electronic mail address: Alexander.Plyusnin@helsinki.fi.

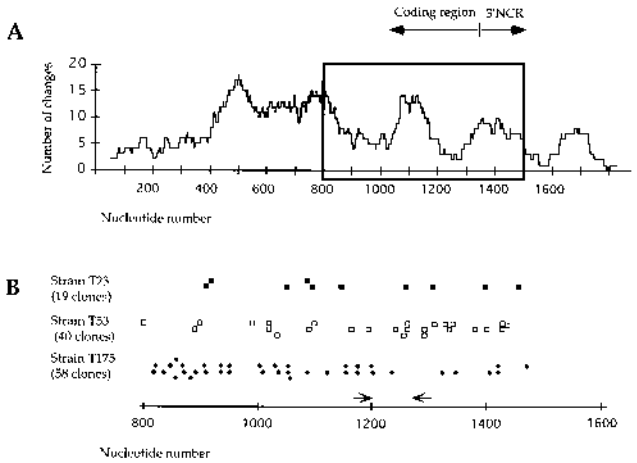

FIG. 1. (A) Distribution of nucleotide replacements in the S segment of TUL (strains T23, T53, T76, T175, and T249; the window size was 101). The region selected for sequencing is framed. 3′NCR, 3′ noncoding region. (B) Nucleotide substitutions that were found in quasispecies of strains T23 (10 substitutions), T53 (26 substitutions), and T175 (38 substitutions). Arrows indicate region (nt 1205 to 1267) where high frequency of *Taq* polymerase misincorporation was registered.

fication and recloning of the TUL S segment (strain T175) showed the same average level of misincorporation for nt 800 to 1530, with the exception of a short region, nt 1205 to 1267, (Fig. 1B) where mutations appeared with an approximately 20-fold higher frequency. Eight of 74 (11%) mutations observed were located within this region: one in T23, four in T53, and three in T175 mutant spectrum. These particular mutations seem most likely to originate from PCR errors. Otherwise, the level of PCR error in our experiments should not be considered crucial.

Character of mutations. Most nucleotide mutations (Table 2) were transitions (51 of 74, or 69%). Within transitions, substitutions A→G and U→C occurred more often than others (G→A and C→U), correlating with the data reported for some other RNA viruses (2, 13). Although the total numbers of clones obtained from the three populations of wild-type TUL were different, the dominance of a master sequence was observed in all of them: about half of the clones carried a master S nucleotide sequence, and master N protein sequences could be deduced for an even higher proportion of the clones. Master sequences might often represent just a minority in the population of viral molecules (3). However, as here for TUL, a similar dominance of the master sequence was reported for an NS3 quasispecies of hepatitis C virus (12). Most TUL mutants were represented by a single clone, but several were represented by two clones each (e.g., clones T53/32 and T53/37). The mutants of the second type were not found in strain T23 (most probably because of the smaller number of clones) but occupied substantial proportions of mutant spectra from strains T53 and T175 (16 and 19%, respectively). Notably, no mutations were observed more than once in our control reamplification and recloning experiments.

TABLE 1. Mutation frequencies in the S/N quasispecies

Strain	No. of nucleotides sequenced (kb)	Frequency of substitutions (10^{-3})[a]		
		Coding region		3′NCR
		nt	aa	
T23	13.3	0.79	1.5	0.78
T53	28.0	0.89	2.2	1.23
T175	40.6	1.06	2.3	0.67

[a] The mean frequencies ± standard errors of substitutions for the nt and aa coding region and the 3′ noncoding region (3′NCR) were 0.91 ± 0.08, 2.00 ± 0.25, and 0.89 ± 0.17, respectively. The level of *Taq* polymerase misincorporation was 0.2×10^{-3} nt (with the exception of nt 1205 to 1267).

Six nucleotide substitutions (identified in the footnotes to Table 2) are of special interest. They were found both within the mutant spectra from T23, T53, and T175 strains, where they were in minority, and in master sequences of other TUL strains. Therefore, these substitutions should be considered as quasineutral or well tolerated. Five of these six nucleotide substitutions were located within the coding region; three of them were silent, and the other two led to amino acid substitutions which were conservative in both cases. These mutations appeared at a frequency of 7.3×10^{-5} and represented a substantial part (8%) of all registered nucleotide substitutions. It seems highly unlikely that such mutations might represent PCR artifacts.

Random fixation of quasineutral mutations in TUL evolution? Within the coding region nucleotide substitutions were distributed evenly between the 1st, 2nd, and 3rd codon positions, inducing all possible variants of amino acid changes and an even appearance of defective genomes (Table 2). Thus, some highly unfit mutants which are expected to be subjected to negative selection are nevertheless present in the mutant spectrum. Such a presence might be favored by dominance of the TUL master S/N sequences: mutants could survive behind the master virus that has the highest fitness to a defined environment. Alternatively, the unfit mutants could belong to a young population, i.e., only a few replication cycles from the initial infection. However, this scenario looks less probable because all three animals showed old immunity to TUL (our unpublished observations).

It is impossible to trace directly the fate of any particular mutant by studying populations of wild-type hantaviruses from trapped animals. Nevertheless, by comparing sequences belonging to the mutant spectrum of one TUL strain with master sequences of other strains, one could approach the question of whether newly appeared nucleotide and amino acid substitutions are tolerable for the virus. More than half (25 of 45) of the amino acid substitutions deduced for mutant spectra of three TUL strains (Table 2) are nonconservative ones. On the other hand, two amino acid substitutions, which should be regarded as neutral or well tolerated (Leu-279→Pro, Ile-286→Val) (Table 2), are conservative. Thus, amino acid mutations, which have higher chances to be fixed in a master sequence, are preferably (or even exclusively) conservative. Such a conclusion correlates with the neutralist view on RNA virus evolution (9) concerning the rarity of advantageous mutations and the deleterious nature of most of the nonneutral ones. The fact that as a result of such evolution, hantavirus variants which differ within antigenically important regions might appear (7, 15, 17) does not seem to be unique for hantaviruses. This correlates with the data observed for other RNA viruses suggesting antigenic variation without immune selection (for a review, see reference 3).

Taken together, our data on S/N quasispecies propose that the random fixation of quasineutral mutations may play a role in TUL evolution. On the other hand, as for other RNA viruses (4), there is no evidence to suggest that this mechanism contributes disproportionally more than selection. Limited negative selection at the protein level in TUL quasispecies does not exclude the possibility that negative selection may operate at the RNA level, for example, by preserving functionally impor-

TABLE 2. Mutations in the TUL S/N quasispecies (nt 801 to 1500/aa 249 to 430)

Strain and clone(s)	Position[a]		nt substitution (aa)	
	nt	aa	Master sequence	Mutant sequence[b]
T23 (19 clones)				
14	914	291	TCA (Ser)	T**T**A (Leu)
18	924	294	CC**G** (Pro)	CC**G** (Pro)
22	1058	339	AC**G** (Thr)	AC**G** (Thr)
14	1092	350	GG**G** (Gly)	GG**G** (Gly)
15	1101	353	GA**G** (Glu)[c]	GA**G** (Glu)[c]
4	1150	370	ACT (Thr)	**G**CT (Ala)
10	1264	408	AGA (Arg)	**G**GA (Gly)
18	1312	424	AAC (Asn)	**G**AC (Asp)
10	1405	3′NCR	A	**G**
12	1462	3′NCR	C	**A**
T53 (40 clones)				
41	801	253	CC**T** (Pro)[d]	CC**T** (Pro)[d]
5	895	285	TTG (Leu)	**A**TG (Met)
33	900	286	ATT (Ile)	AT**C** (Ile)
41	991	317	CCT (Pro)	**T**CT (Ser)
32, 37	1022	327	GCA (Ala)	G**T**A (Val)
1	1038	332	TTT (Phe)	TT**C** (Phe)[e]
43	1096	352	GCA (Ala)	**A**CA (Thr)
20	1105	355	AAG (Lys)	**G**AG (Glu)
28	1169	376	CAA (Gln)	C**G**A (Arg)
5	1200	386	ATG (Met)	AT**A** (Ile)
8	1245	401	GAT (Asp)	GA**A** (Glu)
26	1262	407	CTC (Leu)	C**C**C (Pro)
16, 19	1267	409	ACA (Thr)	**T**CA (Ser)
17, 18	1295	418	AAG (Lys)	A**T**G (Met)
41	1313	424	AAC (Asn)	A**G**C (Ser)
38	1325	428	TTA (Leu)	T–A[f]
41	1341	3′NCR	A	**G**
29	1347	3′NCR	T	**C**
10	1386	3′NCR	G	**T**
38	1406	3′NCR	A	**G**
14,21	1432	3′NCR	C	**T**
22	1441	3′NCR	T	**C**[g]
T175 (58 clones)				
58	825	261	GTG (Val)	GT**A** (Val)
44	826	262	AGC (Ser)	**G**GC (Gly)
36	842	267	CTA (Leu)	C**C**A (Pro)
39	852	270	AGA (Arg)	AG**G** (Arg)
37	863	274	CTC (Leu)	C**C**C (Pro)
58	866	275	CGA (Arg)	C**A**A (Gln)
45	873	277	TCA (Ser)	TC**T** (Ser)
29	878	279	CTG (Leu)	C**C**G (Pro)[h]
24	888	282	ATT (Ile)	AT**G** (Met)
48	898	286	ATT (Ile)	**G**TT (Val)[i]
40	913	291	TCA (Ser)	**C**CA (Pro)
21	915	291	TCA (Ser)	TC**G** (Ser)
34, 38	943	301	AAG (Lys)	**G**AG (Glu)
31, 18	957	305	GCA (Ala)	GC**G** (Ala)
48	1009	323	ATT (Ile)	**G**TT (Val)
39	1016	325	GGG (Gly)	G**A**G (Glu)
36	1035	331	GCA (Ala)	GC**G** (Ala)
54	1042	334	ACT (Ser)	**C**CT (Pro)
5, 14	1060	430	AGA (Arg)	**G**GA (Gly)
44	1063	341	AAC (Asn)	**G**AC (Asp)
59	1092	350	GGA (Gly)	GG**T** (Gly)
3	1129	363	TAT (Tyr)	**C**AT (His)
7, 53	1160	373	ATG (Met)	A**A**G (Lys)
17, 25	1180	380	AGA (Arg)	**G**GA (Gly)
3	1205	388	GAG (Glu)	G**G**G (Gly)
50	1209	389	TGG (Trp)	TG**A**[j]
12	1241	400	GGT (Gly)	G**A**T (Asp)
42	1328	429	AAA (Lys)	A**G**A (Arg)
34	1351	3′NCR	T	**C**
48	1412	3′NCR	G	**A**
30, 54	1425	3′NCR	A	**G**
18	1476	3′NCR	G	**A**

[a] 3′NCR, 3′ noncoding region.
[b] Mutated nucleotides are in boldface type and those from the coding sequence are underlined.
[c] GAG (Glu) was found in master sequence of TUL strains M5286, M5292, M5293, M5294, and M5302 (15).
[d] CCT (Pro) was found in master sequence of TUL strains T23, T76, T175, T249, M5286, M5293, M5294, and M5302 (15, 16).
[e] TTC (Phe) was found in master sequence of TUL strains T76 and T175 (16).
[f] Deletion, appearance of TAA (termination codon).
[g] C was found in master sequence of TUL strain T76 (16).
[h] CCG (Pro) was found in master sequence of TUL strain T249 (16).
[i] GTT (Val) was found in master sequence of TUL strain T76 (16).
[j] Termination codon.

tant stem-loop structures, similar to what has been reported for picornaviruses (14, 20). Also, selection should be involved at the stage of virus transmission from one rodent to another, when randomly sampled aliquots of virus population are distributed into another environment. However, this selection might operate preferably against changes in master geno- and phenotypes, with the result that nucleotide substitutions between master genotypes are mostly silent and deduced amino acid substitutions are mostly synonymous.

We thank Anja Virtanen and Leena Kostamovaara for expert technical assistance.

This work was supported by grants from the Medical Research Council of the Academy of Finland and The Sigrid Jusélius Foundation, Helsinki.

REFERENCES

1. **Antic, D., Y. C. Kang, K. Spik, C. S. Schmaljohn, O. Vapalahti, and A. Vaheri.** 1992. Comparison of the deduced gene products of the L, M and S genome segments of hantaviruses. Virus Res. **24:**35–46.
2. **Cattaneo, R.** 1994. Biased (A→I) hypermutation of animal RNA virus genomes. Curr. Opin. Genet. Dev. **4:**895–900.
3. **Domingo, E., J. Diez, M. A. Martinez, J. Hernandez, A. Holguin, B. Borrego, and M. G. Mateu.** 1993. New observations on antigenic diversification of RNA viruses. Antigenic variation is not dependent on immune selection. J. Gen. Virol. **74:**2039–2045.
4. **Domingo, E., and J. J. Holland.** 1994. Mutation rates and rapid evolution of RNA viruses, p. 161–184. *In* S. S. Morse (ed.), The evolutionary biology of viruses. Raven Press, New York.
5. **Henderson, W. W., M. C. Monroe, S. C. St. Jeor, W. P. Thayer, J. E. Rowe, C. J. Peters, and S. Nichol.** 1995. Naturally occurring Sin Nombre virus genetic reassortants. Virology **214:**602–610.
6. **Holland, J. J., J. C. De La Torre, and D. A. Steinhauer.** 1992. RNA virus populations as quasispecies. Curr. Top. Microbiol. Immunol. **176:**1–20.
7. **Hörling, J., Y. Cheng, A. Plyusnin, K. Persson, H. Lehväslaiho, A. Vaheri, B. Niklasson, and Å. Lundkvist.** 1995. Nucleotide and deduced amino acid sequences of the M and S genome segments of a Swedish Puumala virus isolate. Virus Res. **39:**321–330.
8. **Kariwa, H., Y. Isegawa, J. Arikawa, I. Takashima, S. Ueda, K. Yamanishi, and N. Hashimoto.** 1994. Comparison of nucleotide sequences of M genome segments among Seoul virus strains isolated from Eastern Asia. Virus Res. **33:**27–38.
9. **Kimura, M.** 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge.
10. **Kingsford, L.** 1991. Antigenic variance. Curr. Top. Microbiol. Immunol. **169:**181–216.
11. **Li, D., A. L. Schmaljohn, K. Anderson, and C. S. Schmaljohn.** 1995. Complete nucleotide sequences of the M and S segments of two hantavirus isolates from California: evidence for reassortment in nature among viruses related to hantavirus pulmonary syndrome. Virology **206:**973–983.

12. **Martell, M., J. I. Esteban, J. Quer, J. Genesca, A. Weiner, R. Esteban, J. Guardia, and J. Gomez.** 1992. Hepatitis C virus (HCV) as a population of different but closely related genomes: quasispecies nature of HCV genome distribution. J. Virol. **66:**3225–3229.
13. **O'Hara, P. J., S. T. Nichol, F. M. Horodyski, and J. J. Holland.** 1984. Vesicular stomatitis virus defective interfering particles can contain extensive genomic sequence rearrangements and base substitutions. Cell **36:**915–924.
14. **Pilipenko, E. V., V. M. Blinov, B. K. Chernov, T. M. Dmitrieva, and V. I. Agol.** 1989. Conservation of the secondary structure elements of the 5′-untranslated region of cardio- and aphthovirus RNAs. Nucleic Acids Res. **17:**5701–5711.
15. **Plyusnin, A., Y. Cheng, O. Vapalahti, M. Pejcoch, J. Unar, Z. Jelinkova, H. Lehväslaiho, and A. Vaheri.** 1995. Genetic variation in Tula hantaviruses: sequence analysis of the S and M segments of strains from Central Europe. Virus Res. **39:**237–250.
16. **Plyusnin, A., O. Vapalahti, H. Lankinen, H. Lehväslaiho, N. Apekina, Y. Myasnikov, H. Kallio-Kokko, H. Henttonen, Å. Lundkvist, M. Brummer-Korvenkontio, I. Gavrilovskaya, and A. Vaheri.** 1994. Tula virus: a newly detected hantavirus carried by European common voles. J. Virol. **68:**7833–7838.
17. **Plyusnin, A., O. Vapalahti, H. Lehväslaiho, N. Apekina, T. Mikhailova, I. Gavrilovskaya, J. Laakkonen, J. Niemimaa, H. Henttonen, M. Brummer-Korvenkontio, and A. Vaheri.** 1995. Genetic variation of wild Puumala viruses within the serotype, local rodent population and individual animals. Virus Res. **38:**25–41.
18. **Plyusnin, A., O. Vapalahti, K. Ulfves, H. Lehväslaiho, N. Apekina, I. Gavrilovskaya, V. Blinov, and A. Vaheri.** 1994. Sequences of wild Puumala virus genes show a correlation of genetic variation with geographic origin of the strains. J. Gen. Virol. **75:**405–409.
19. **Plyusnin, A., O. Vapalahti, and A. Vaheri.** 1996. Hantaviruses: genome structure, expression and evolution. J. Gen. Virol. **77:**2677–2687.
20. **Pöyry, T., L. Kinnunen, and T. Hovi.** 1992. Genetic variation in vivo and proposed functional domains of the 5′ noncoding region of poliovirus RNA. J. Virol. **66:**5313–5319.
21. **Smith, S.** 1992. Genetic data environment, version 2.0. Harvard Genome Laboratory, Harvard University, Cambridge, Mass.
22. **Spiropoulou, C. F., S. Morzunov, H. Feldmann, A. Sanchez, C. J. Peters, and S. T. Nichol.** 1994. Genome structure and variability of a virus causing hantavirus pulmonary syndrome. Virology **200:**715–723.
23. **Vapalahti, O., H. Kallio-Kokko, E.-M. Salonen, M. Brummer-Korvenkontio, and A. Vaheri.** 1992. Cloning and sequencing of Puumala virus Sotkamo strain S and M RNA segments: evidence for strain variation in hantaviruses and expression of the nucleocapsid protein. J. Gen. Virol. **73:**829–838.
24. **Xiao, S.-Y., J. W. LeDuc, Y. K. Chu, and C. S. Schmaljohn.** 1994. Phylogenetic analysis of virus isolates in the genus Hantavirus, family Bunyaviridae. Virology **198:**205–217.