

CORRESPONDENCE

Open Access

# Identification of tell-tale patterns in the 3' non-coding region of hantaviruses that distinguish HCPS-causing hantaviruses from HFRS-causing hantaviruses

Sathish Sankar<sup>1</sup>, Jayanta Borkakoti<sup>2</sup>, Mageshbabu Ramamurthy<sup>1</sup>, Balaji Nandagopal<sup>1</sup>, Perumal Vivekanandan<sup>2</sup> and Sridharan Gopalan<sup>1</sup>

Hantaviruses are negative-sense, single-stranded RNA viruses with small (S)-, middle (M)-, and large (L)-segments. Each segment has a coding region and non-coding regions (NCRs)<sup>1</sup>. Hantaviruses associated with human diseases specifically cause either hemorrhagic fever with renal syndrome (HFRS) or hantavirus cardiopulmonary syndrome (HCPS)<sup>2</sup>. A few hantaviruses have been associated with both syndromes. Efforts to identify differences in amino-acid sequences and nucleotide sequences between hantaviruses causing HFRS and HCPS have failed. Clinically, the need to distinguish hantaviruses based on their disease-causing ability is being increasingly recognized<sup>3</sup> because of the discovery of several new hantaviruses in the last decade. In addition, differences between hantavirus genotypes in susceptibility to antiviral agents are well-documented. For example, ribavirin is useful for treating HFRS, but not HCPS<sup>4</sup>. Favipiravir may be useful for treating HCPS<sup>5</sup>. However, specific genomic differences, if any, between HFRS-causing hantaviruses and HCPS-causing hantaviruses have yet to be elucidated.

Previous studies on other RNA viruses indicate important roles for the 3' NCRs in virus replication and infectivity<sup>6</sup>. Among hantaviruses, the nucleotide sequence of the 3' NCR of the S-segment is highly variable<sup>1</sup>, making meaningful sequence comparisons of the 3' NCR extremely challenging<sup>7</sup>. Therefore, there is a paucity of reports

that systematically analyze the 3' NCR of hantaviruses. In this study, we investigated the association, if any, between the length, CpG content, and RNA-folding free energy of hantavirus NCRs and the ability of hantaviruses to cause HFRS, HCPS, or both. Our results highlight an important and yet unknown link between the length and CpG content of hantavirus 3' NCR and human disease.

The S-, M-, and L-segments of hantavirus sequences with complete coding DNA sequences along with the 3' NCR and 5' NCR were obtained from GenBank. The accession numbers of all 896 sequences used for analysis are given in Supplementary Table S1. Viruses reported in the International Committee on Taxonomy of Viruses classification and fulfilling the above criteria were selected and used for further analysis. The reservoir hosts for Old World hantaviruses are primarily restricted to Eurasia, while the reservoirs for New World hantaviruses are found in the Americas<sup>2</sup>. The sequences analyzed were classified into four categories (HCPS only, HFRS only, both HCPS and HFRS, and unknown) based on their ability to cause human disease (Supplementary Fig. S2).

We analyzed the length of the 3' NCR ( $n = 896$ ) and 5' NCR ( $n = 631$ ) of hantavirus sequences using BioEdit v7.2.5 (<http://www.mbio.ncsu.edu/bioedit/page2.html>). Data were analyzed using Student's *t*-test. Box plots were made using MS-Excel. The results were considered statistically significant at a *P*-value of  $<0.05$ .

The lengths of the 3' NCRs of the S- and L-segments were significantly longer in New World hantaviruses than Old World hantaviruses (Fig. 1a, b;  $P < 0.0001$ ). The existence of such striking differences in the 3' NCR lengths between New world hantaviruses and Old World

Correspondence: Perumal Vivekanandan ([vperumal@bioschool.iitd.ac.in](mailto:vperumal@bioschool.iitd.ac.in))

<sup>1</sup>Sri Sakthi Amma Institute of Biomedical Research, Sri Narayani Hospital and Research Centre, Sripuram, Vellore 632055 Tamil Nadu, India

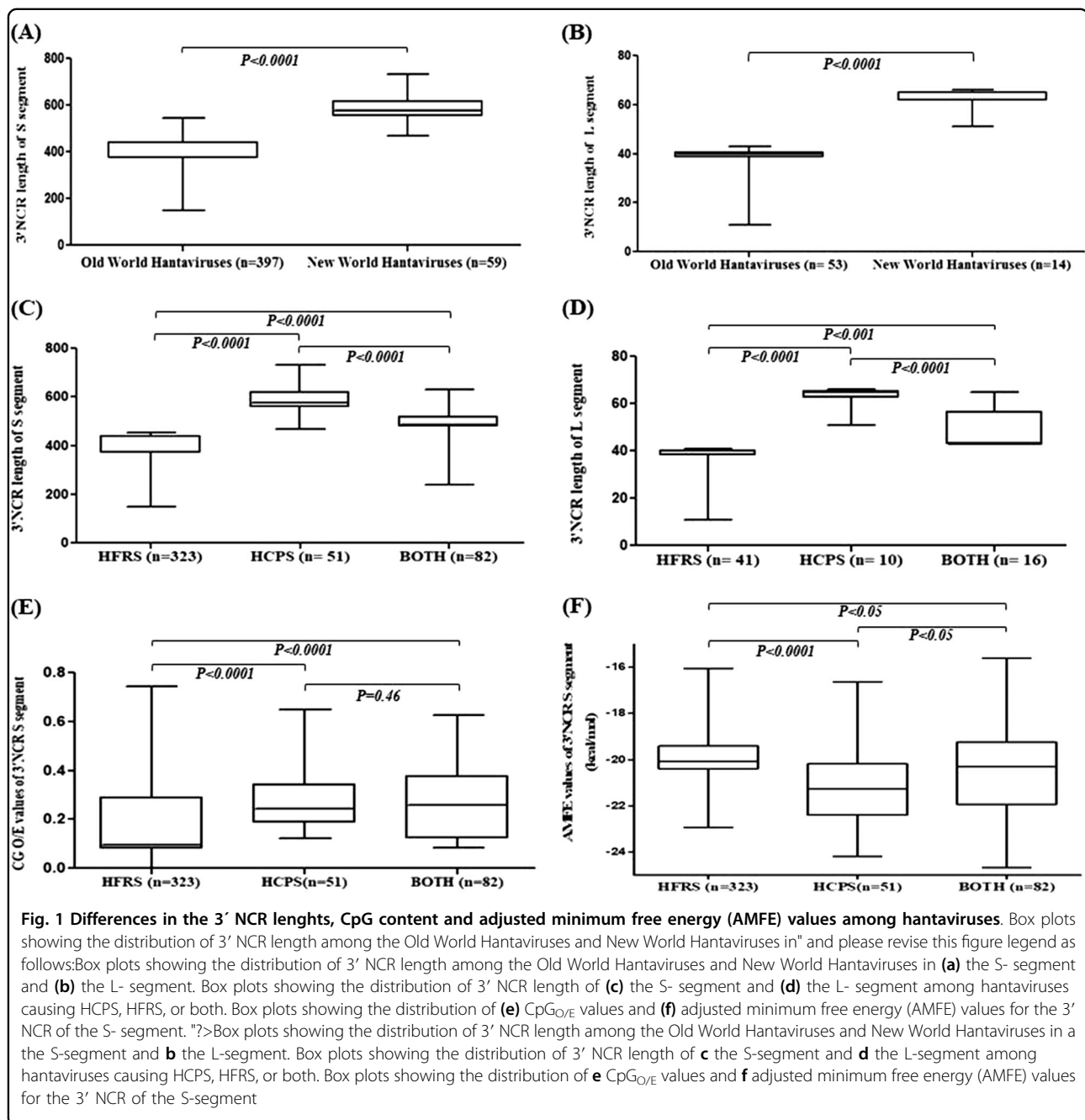
<sup>2</sup>Kusuma School of Biological Sciences, Indian Institute of Technology, New Delhi 110016, India

These authors contributed equally: Sathish Sankar, Jayanta Borkakoti.

© The Author(s) 2018



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.



hantaviruses has not been reported previously. The length of the 3' NCR of the M-segment was comparable between New World hantaviruses and Old World hantaviruses (Supplementary Fig. S1A).

We then investigated, an association, if any, between the length of the 3' NCR of the S-, M-, and L-segments and the ability of hantaviruses to cause HCPS, HFRS, or both. Interestingly, the 3' NCRs of the S- and L-segments for HCPS-causing hantavirus genotypes were significantly longer than that for HFRS-causing hantavirus genotypes (Fig. 1c, d;  $P < 0.0001$  and  $P < 0.0001$ , respectively;

Supplementary Table S3). The length of the 3' NCR of the M-segment was comparable between HFRS-causing and HCPS-causing genotypes (Supplementary Fig. S1B;  $P = 0.77$ ). The number of sequences available for analysis varied greatly across the hantaviruses analyzed. The results from analysis of a maximum of five sequences each for a given hantavirus (Supplementary Fig. S2) are in keeping with that from analyzing all available sequences (Fig. 1c, d). The variability of the 3' NCR sequence lengths in hantaviruses is well-documented for the S-segment<sup>7</sup>; nonetheless, the potential of the 3' NCR length to

distinguish HCPS-causing genotypes from HFRS-genotypes was not known previously.

The average length of the 5' NCR sequences analyzed are tabulated in Supplementary Table S4. The 5' NCRs of the hantavirus M-segment of HCPS-causing genotypes were significantly longer than those of the HFRS-causing genotypes (Supplementary Table S4).

Our findings clearly demonstrate that the 3' NCRs of the S- and the L-segments are longer in HCPS-causing genotypes than in HFRS-causing genotypes. HFRS is caused by Old World hantaviruses, and HCPS is caused by New World hantaviruses<sup>2</sup>. It is possible that the differences in the NCR lengths of hantavirus genotypes causing HFRS and HCPS may reflect the differences between Old World and New World hantaviruses. The hantaviruses that cause both HFRS and HCPS include Old World viruses (Puumala virus and Tula virus) and New World viruses (Andes virus, Black creek Canal virus, and Bayou virus). If the observed differences in the lengths of the NCRs between hantaviruses are linked to the geographic distribution of the reservoir host (Old World vs. New World hantaviruses), one would expect that the NCR lengths of Puumala viruses and Tula viruses (both Old World hantaviruses), which can cause both HFRS and HCPS, would be comparable to those of other Old World viruses that cause HFRS. Interestingly, our results indicate that the 3' NCRs of the S-segment of Puumala viruses and Tula viruses that cause both HFRS and HCPS are significantly longer than those of the rest of the Old World hantaviruses known to cause only HFRS (Supplementary Fig. S3A). In addition, the 3' NCRs of the S-segment of hantaviruses causing HCPS only or both HCPS and HFRS were significantly longer than those causing HFRS only; this was independent of the geographic distribution of the Old World and New World hantaviruses.

Neither the length or the sequence of the 3' NCR segment of a given hantavirus is amenable to major changes, suggesting a functional role for the 3' NCR in hantavirus genomes<sup>1</sup>. Our results indicate major differences in the length of the 3' NCR of the S-segment between hantaviruses causing HCPS and HFRS. We then analyzed (a) the relative abundance of CpG dinucleotides using methods described elsewhere<sup>8</sup> and (b) the minimum free energy (MFE) for RNA folding, calculated using ViennaRNA Package 2.0 (<http://www.tbi.univie.ac.at/RNA>) for the 3' NCR of the S-segment. Subsequently, the adjusted minimum free energy (AMFE) was calculated using the formula  $AMFE = (100 \times MFE) / (\text{length of the RNA})^9$ . The relative abundance of CpG dinucleotides should not be affected by the differences in the length of the 3' NCR of the S-segment. AMFE is adjusted for length, so the differences in the length of the 3' NCR of the S-segment between viruses would affect AMFE values to a lesser extent than MFE values. Both the CpG content<sup>10,11</sup> and

RNA-folding free energy<sup>12</sup> in RNA viruses have been linked to virus pathogenesis. Interestingly, our data suggest the 3' NCR of the S-segment of hantaviruses causing HCPS only or both HCPS and HFRS had higher CpG content (Fig. 1e) compared to those causing HFRS only; this was independent of the geographic distribution of the Old World and New World hantaviruses (Supplementary Fig. S3B). Although the differences in AMFE values between the 3' NCR of the S-segment of hantaviruses causing HCPS only or both HCPS and HFRS were statistically significant (Fig. 1f and Supplementary Fig. S3C), the implications of such small differences on RNA stability is not fully understood.

The roles of the 3' NCRs in virus pathogenesis and adaptation to hosts have been described for other RNA viruses<sup>13,14</sup>. Among hantaviruses, the NCR of the S-segment has the ability to initiate virus replication, that of the M-segment has the strongest promoter activity, and that of the L-segment has the best packaging efficiency<sup>15</sup>. We speculate that the observed differences in the length and RNA-folding free energy of the 3' NCR sequences from the S-segment between HCPS-causing hantaviruses and HFRS-causing hantaviruses may influence the formation and the stability of RNA secondary structures. In addition, the observed differences in CpG content can potentially influence hantavirus adaptation to specific hosts or host cell types.

In sum, our findings have identified major differences in the 3' NCR of the S-segment between hantaviruses causing HFRS and HCPS in (a) the length (b) the relative abundance of CpG dinucleotides, and (c) the RNA-folding free energy. None of these differences have been previously reported, and all of them have implications in virus pathogenesis. Taken together, our findings suggest that the observed differences in the 3' NCR of the S-segment do not merely reflect the geographic distribution of the reservoir host but are linked to the specific syndromes caused by hantavirus genotypes in humans. The association between the tell-tale patterns of the 3' NCR and specific clinical syndromes caused by hantaviruses merits further investigation. We believe that this work provides a novel perspective to our current understanding of the role of NCRs in the pathogenesis of hantaviruses.

#### Acknowledgements

J.B. (Kusuma School of Biological Sciences, IIT Delhi) is a recipient of the Department of Science and Technology National Postdoctoral Fellowship (SERB), [PDF/2016/003852], Government of India.

#### Conflict of interest

The authors declare that they have no conflict of interest.

**Supplementary Information** accompanies this paper at (<https://doi.org/10.1038/s41426-018-0027-z>).

Received: 22 September 2017 Revised: 2 January 2018 Accepted: 2 January 2018

Published online: 21 March 2018

## References

1. Plyusnin, A., Vapalahti, O. & Vaheri, A. Hantaviruses: genome structure, expression and evolution. *J. Gen. Virol.* **77**, 2677–2687 (1996).
2. Jonsson, C. B., Figueiredo, L. T. & Vapalahti, O. A global perspective on hantavirus ecology, epidemiology, and disease. *Clin. Microbiol. Rev.* **23**, 412–441 (2010).
3. Lednicky, J. A. Hantaviruses: a short review. *Arch. Pathol. Lab. Med.* **127**, 30–35 (2003).
4. Leyssen, P., De Clercq, E. & Neyts, J. Molecular strategies to inhibit the replication of RNA viruses. *Antivir. Res.* **78**, 9–25 (2008).
5. Buys, K. K. et al. Maporal virus as a surrogate for pathogenic New World hantaviruses and its inhibition by favipiravir. *Antivir. Chem. Chemother.* **21**, 193–200 (2011).
6. Saiz, M. et al. Deletion or substitution of the aphthovirus 3' NCR abrogates infectivity and virus replication. *J. Gen. Virol.* **82**, 93–101 (2001).
7. Bohlman, M. C. et al. Analysis of hantavirus genetic diversity in Argentina: S segment-derived phylogeny. *J. Virol.* **76**, 3765–3773 (2002).
8. Upadhyay, M. et al. CpG dinucleotide frequencies reveal the role of host methylation capabilities in parvovirus evolution. *J. Virol.* **87**, 13816–13824 (2013).
9. Trotta, E. On the normalization of the minimum free energy of RNAs by sequence length. *PLoS ONE* **9**, e113380 (2014).
10. Wasson, M. K., Borkakoti, J., Kumar, A., Biswas, B. & Vivekanandan, P. The CpG dinucleotide content of the HIV-1 envelope gene may predict disease progression. *Sci. Rep.* **7**, 8162 (2017).
11. Takata, M. A. et al. CG dinucleotide suppression enables antiviral defence targeting non-self RNA. *Nature* **550**, 124–127 (2017).
12. Brower-Sinning, R. et al. The role of RNA folding free energy in the evolution of the polymerase genes of the influenza A virus. *Genome Biol.* **10**, R18 (2009).
13. Roby, J. A. et al. Noncoding subgenomic flavivirusRNA: multiple functions in West Nile virus pathogenesis and modulation of host responses. *Viruses* **6**, 404–427 (2014).
14. Brown, D. M. et al. Cell-dependent role for the poliovirus 3' noncoding region in positive-strand RNA synthesis. *J. Virol.* **78**, 1344–1351 (2004).
15. Flick, K. et al. Functional analysis of the noncoding regions of the Uukuniemi virus (Bunyaviridae) RNA segments. *J. Virol.* **78**, 11726–11738 (2004).