

Short Communication

Correspondence
Alexander N. Lukashev
Alexander_lukashev@
hotmail.com

Received 15 February 2005
Accepted 19 April 2005

Evidence for recombination in *Crimean-Congo hemorrhagic fever virus*

Alexander N. Lukashev

Institute of Poliomyelitis and Viral Encephalitides RAMS, Moscow 142782, Russia

Crimean-Congo hemorrhagic fever (CCHF) virus has attracted considerable attention recently and a number of phylogenetic studies have been published, based mostly on partial sequences of S and M RNA segments. In this study, available full-length S, M and L segment sequences of CCHF virus were checked for recombination. Similarity plots and bootscan analysis of the S segment suggested multiple recombination events between southern European, Asian and African CCHF virus strains, with additional evidence provided by phylogenetic trees, the hidden Markov model and probabilistic divergence measures methods. No unambiguous signs of recombination were observed for M and L segments; however, the results did not exclude the possibility of this. These findings, coupled with a recent report on reassortment in CCHF virus, suggest caution when assessing CCHF virus phylogeny based on short sequence fragments.

Crimean-Congo hemorrhagic fever (CCHF) virus is a member of the genus *Nairovirus* in the family *Bunyaviridae* (Elliott *et al.*, 2000). CCHF is a tick-borne infection, enzootic in livestock in the Middle East, Asia and Africa, and is also recognized in southeast Europe. Transmission to humans via tick bite or by contact with blood or tissues from infected animals or patients results in hemorrhagic fever. The fatality rate of the disease varies from 15 to 80% (Shepherd *et al.*, 1988; Nichol *et al.*, 2000). Maintenance of the CCHF virus in nature involves several vertebrate hosts and tick vectors, although it is clear that *Hyalomma* spp. predominate in those biotopes where transmission is most intensive (Hoogstraal, 1979). CCHF virus is able to replicate in various vertebrates without causing clinical symptoms; only humans (and suckling mice) experience the disease. There is evidence that CCHF virus is present throughout Africa, Bulgaria, the former Yugoslavia, Greece, European Russia, the Arabian peninsula, Iraq, Pakistan, the Central Asian republics of the former Soviet Union and China (Nichol, 2001).

Similar to other members of the family *Bunyaviridae*, CCHF virus possesses a tripartite, single-stranded RNA genome of negative sense that consists of large (L), medium (M), and small (S) segments of approximately 12100, 5050 and 1570 nt (Schmaljohn & Hooper, 2001). The S RNA segment encodes a nucleocapsid protein of 54 kDa. The M segment encodes a polypeptide, which is processed into two surface glycoproteins (Sanchez *et al.*, 2002). Complete sequences of the L segment emerged only last year (Honig *et al.*, 2004; Kinsella *et al.*, 2004). This part of the genome encodes a putative viral polymerase.

Recently, several attempts have been made to conduct phylogenetic analysis of CCHF virus strains from different regions

of China and Europe, and several distinct phylogenetic groups have been identified based on sequence data from different parts of the S segment (Rodriguez *et al.*, 1997; Papa *et al.*, 2002a, b, c; Yashina *et al.*, 2003a, b; Chinikar *et al.*, 2004). Phylogenetic analysis of the M segment has also been performed and this work has divided the Chinese strains into three distinct groups, one of which also included the Nigerian isolate IbAr10200 (Morikawa *et al.*, 2002). Interestingly, the S segment of this isolate made up a separate African phylogenetic group. One recent paper reported convincing evidence of genome reassortment in CCHF virus (Hewson *et al.*, 2004). A comprehensive work on recombination in negative-strand RNA viruses was published in 2003, where 'patterns of sequence variation compatible with the action of recombination' were found in CCHF virus (Chare *et al.*, 2003). In this work, I used a range of phylogenetic methods to provide clear proof on this issue.

All currently available CCHF virus sequences for S, M and L segments (Table 1) were aligned using CLUSTAL_X software (Thompson *et al.*, 1997) and screened for probable recombination using a range of methods. First, I checked the similarity plots produced by SimPlot version 2.5 software (Ray, 1999) to obtain an overview of the situation. A number of the S segment sequences, which were equally similar all over the segment to their closest relative and did not carry a detectable mark of recombination, could be excluded to simplify the analysis. Most of the remaining sequences had similarity plots suggesting recombination (e.g. Fig. 1a). Next, I performed bootscanning analysis (Salminen *et al.*, 1995), also implemented in SimPlot version 2.5. As higher numbers of sequences did not produce reliable bootstrap values in bootscanning, I further limited the analysis to seven sequences that bore the most obvious signs of recombination on similarity plots, namely, strains

Table 1. Sequences used in the work

Strain name*	GenBank accession numbers			Place and source of isolation	Year of isolation	Reference
	S	M	L			
IbAr10200	U88410	U39455	AY389508; NC_005301; AY389361	Nigeria, tick	1969	Honig <i>et al.</i> (2004); Kinsella <i>et al.</i> (2004)
Ap92	U04958			Greece, tick	1976	Marriott <i>et al.</i> (1994)
Ug3010	U88416			Uganda	1956	
Ch66019	AJ010648	AB069669		China, patient	1966	Morikawa <i>et al.</i> (2002)
Ch68031	M86625			China, sheep	1968	Marriott & Nuttall (1992)
Ch7001	AF415236	AB069670		China, patient	1970	Morikawa <i>et al.</i> (2002)
Ch75024	AF362080	AB069671		China, patient	1975	Morikawa <i>et al.</i> (2002)
Ch7803	AF354296	AB069672		China, patient	1978	Morikawa <i>et al.</i> (2002)
Ch79121	AF358784	AB069673		China, jerboa	1979	Morikawa <i>et al.</i> (2002)
Ch8402	AJ010649	AF350449		China, tick	1984	Papa <i>et al.</i> (2002a)
Ch88166	AY029157	AF338470		China, patient	1988	
Dak8194	U88411			Senegal, patient	1987	
Drosdov	U88412			South Russia, patient	1967	
Hodzha	AY223475	AY223476		Uzbekistan, patient	1967	Chumakov <i>et al.</i> (1968); Hewson <i>et al.</i> (2004)
STV/HU29223	AF481802			South Russia, patient	2000	Petrova <i>et al.</i> (2003)
Matin	AF527810	AF467769	AY422208	Pakistan, patient	1976	Sanchez <i>et al.</i> (2002)
Baghdad-12	AJ538196	AJ538197		Iraq, patient	1976	Hewson <i>et al.</i> (2004)
SR3	AJ538198	AJ538199		Pakistan, patient	2000	Hewson <i>et al.</i> (2004)
TADJ/HU896630908		AY179962	AY720893 AY675240	Tajikistan South Russia	1990	Seregin <i>et al.</i> (2004)

*Strain names are truncated on some figures.

Ap92, Dak8194, Ug3010, IbAr10200, Drosdov, Matin and Ch88166. Next, I performed bootscanning among the selected strains. Bootscan analysis slides a window over an alignment and for each window plots the proportion of the bootstrap pseudoreplicates that support phylogenetic grouping of the query strain with the strains used for comparison. Bootstrap values over 70% are generally considered significant. Whenever one reliable grouping changes for another, one may suggest a recombination event. Bootscanning analysis of the strains studied (Fig. 1b–e) provided reasonably convincing evidence of phylogenetic conflict between different parts of the S segment.

To illustrate further the results of bootscanning, I constructed phylogenetic trees (Fig. 2a) for different S segment fragments by using CLUSTAL_X software [using the neighbour-joining (NJ) algorithm and the Kimura distance model] with the ‘exclude positions with gaps’ and ‘correct for multiple substitutions’ options. The Dugbe virus sequence (Genbank accession no. NC_004157; Bridgen *et al.*, 2002) was used as an outgroup to root the phylogenetic trees. Only NJ trees are presented here; however, phylogenetic trees that were created using the maximum-likelihood algorithm implemented in the DNAML module of the PHYLIP software package (Felsenstein, 1989) had very similar topology and reliably supported the same phylogenetic groups as the NJ trees. Conflicting phylogenetic

trees for different S segment fragments (Fig. 2a) further supported the results of bootscanning. First, the outgroup CCHF virus strain was different in these trees. In the alignment region of nt 280–450, and, less reliably, in the nt 420–750 region, the outgroup was Ap92 strain (i.e. all other CCHF virus strains excluding Ap92 grouped together); in the region from nt 750 to the end, the outgroup was strain Dak8194; and in the nt 1–200 alignment region, the outgroup was strain Drosdov. In all these cases, bootstrap values were above or very close to the arbitrary 70% reliability threshold. Minor groups also showed signs of recombination. In the regions nt 1–200 and 280–450, strain Ch88166 grouped with strain IbAr10200, while in regions nt 420–750 and 750–end, it grouped very reliably with strain Drosdov. The Dak8194/Ap92 group, which was observed on the nt 1–200 tree with a highly significant bootstrap value, was not seen in any other alignment region.

Additional methods were run on the four strains that showed the clearest signs of recombination: Dak8194, IbAr10200, Drosdov and Ch88166. The hidden Markov model (HMM) method (Husmeier & McGuire, 2003) implemented in TOPALi version 0.22 (Milne *et al.*, 2004) showed clearly different probabilities of different tree topologies over the genome, with patterns very similar to the bootscan graphs (Fig. 2b). A probabilistic divergence measures (PDM) algorithm (Husmeier & Wright, 2001)

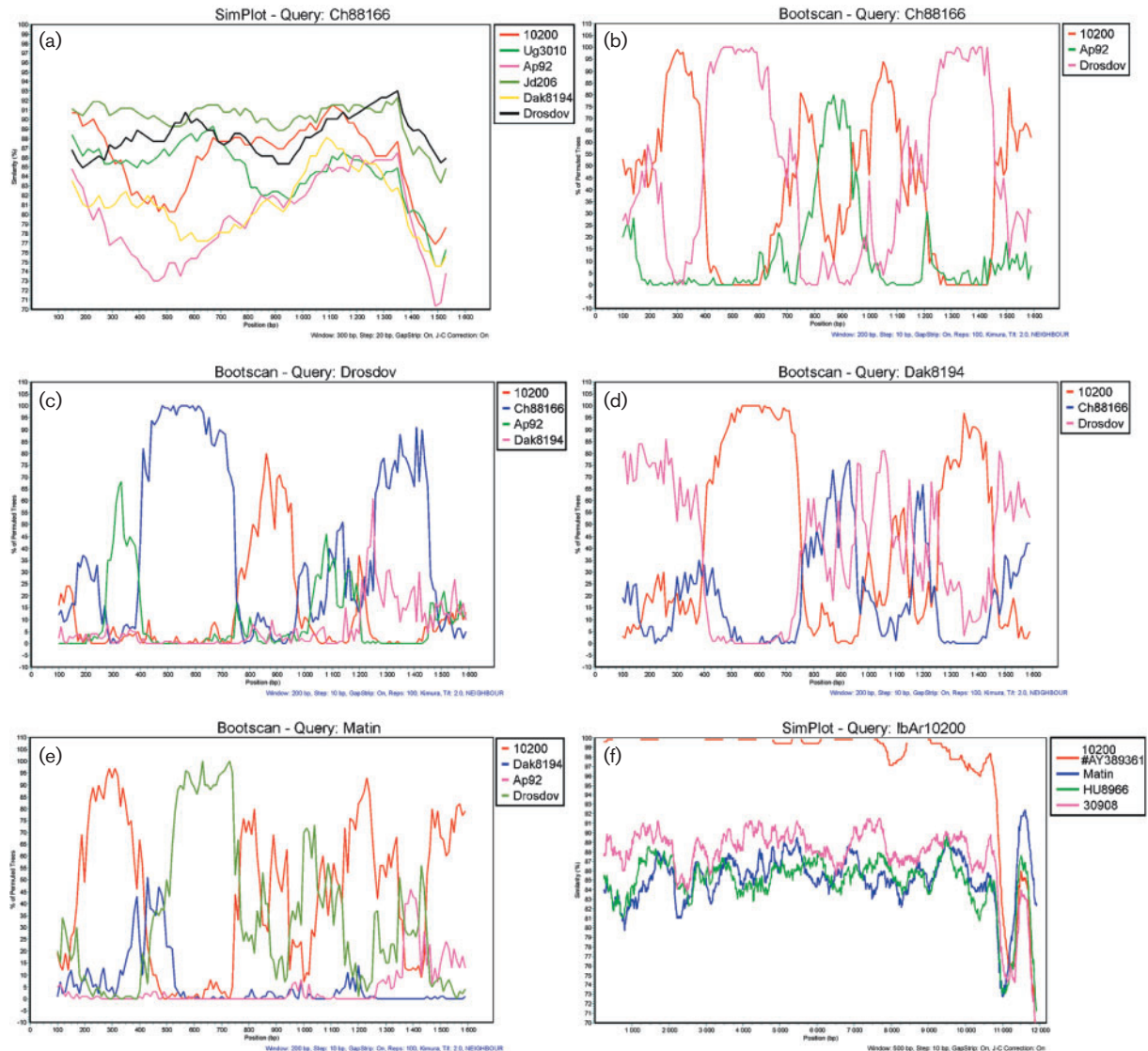


Fig. 1. (a) Similarity plot of the S segment of strain Ch88166 compared with other CCHF virus strains. Window 300 nt, step 20 nt. (b)–(e) Bootscan graphs of strains Ch88166 (b), Drosdov (c), Dak8194 (d) and Matin (e) compared with selected CCHF virus strains. Window 200 nt, step 10 nt. (f) Similarity plot of the L segment of strain IbAr10200 (GenBank accession no. AY389508) compared with other CCHF virus strains. Window 500 nt, step 10 nt.

was also performed using TOPALi version 0.22 software. This method slides a window over the alignment and plots local divergence, i.e. the difference in the phylogeny probability distribution in the current window and in the adjacent windows. Peaks corresponded to the phylogeny changes in neighbouring windows, which most likely resulted from recombination, and the dotted line shows 95 % reliability cut-off. PDM for four S segment sequences (Fig. 2c) indicated several putative recombination points.

Similarity plots of 13 M segment sequences did not reveal any obvious cases of recombination. Next, bootscanning was carried out on all or selected sequences and provided inconclusive results. Only analysis with PDM (window size

200 or 500 nt) clearly supported recombination (Fig. 2d). One can think of many reasons for a virus to give a false recombination signal in one method, for example immune selection pressure or specific virus–host interactions providing subtle convergent changes in different virus strains. No ‘ultimate evidence’, conflicting the phylogenies, could be obtained for the M segment; therefore, recombination in the M segment could neither be excluded nor proven with the currently available sequences.

Currently, there are only five L segment sequences available in GenBank. Interestingly, L segment sequences of strain IbAr10200, submitted independently by two different groups, [GenBank accession nos AY389361 (Kinsella *et al.*,

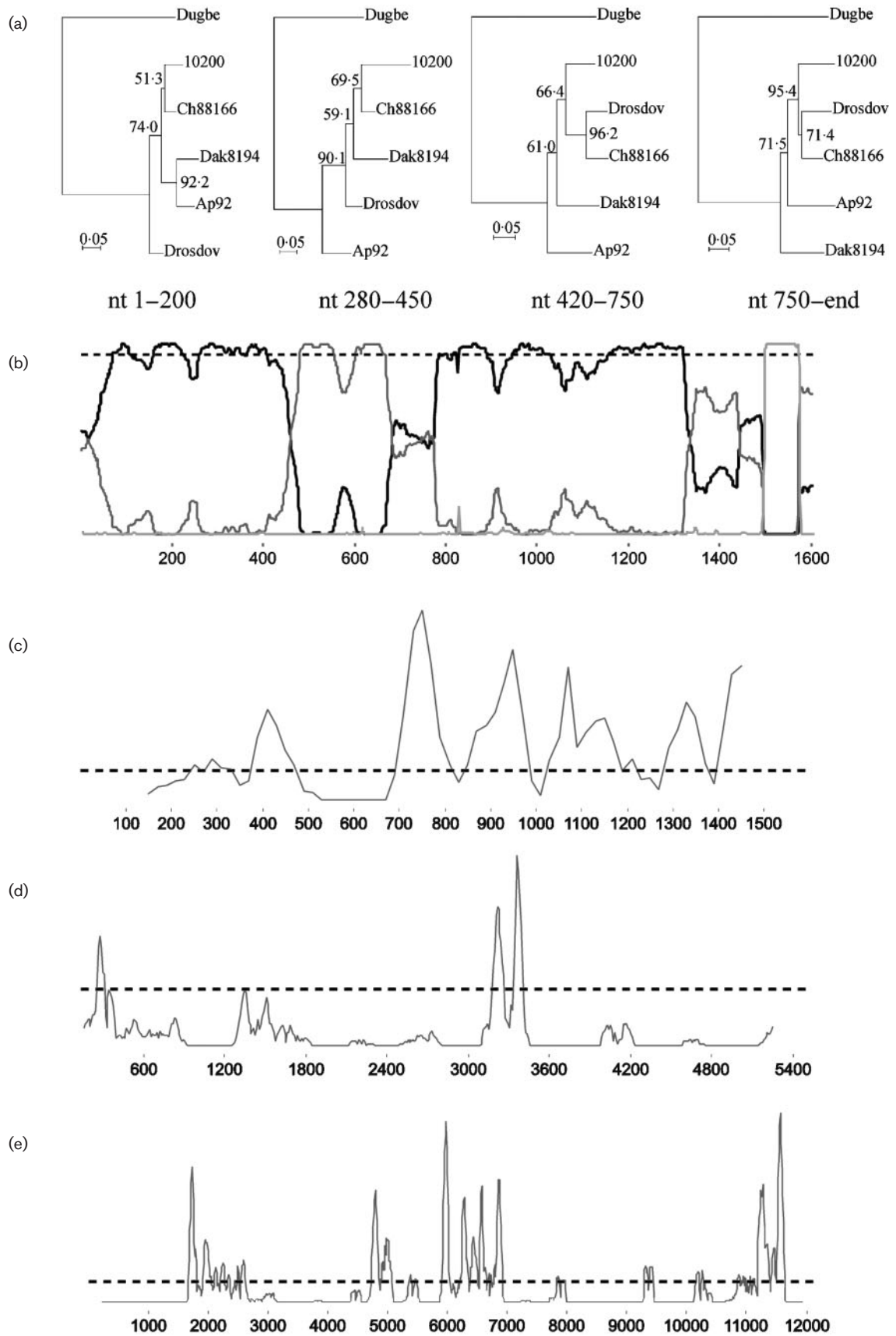


Fig. 2. (a) Phylogenetic NJ trees for different regions of the S segment for selected CCHF virus strains. (b) Results of the HMM method applied to strains Dak8194 (1), Ch88166 (2), IbAr10200 (3) and Drosdov (4). The probabilities of different tree topologies are shown as: black, (1, 4), (2, 3); grey, (1, 3), (2, 4); pale grey, (1, 2), (3, 4). (c) Local divergence measure of alignment of the S segment of strains Dak8194, Ch88166, IbAr10200 and Drosdov as calculated by PDM. Window 200 nt, step 20 nt. The dotted line shows the 95% reliability cut-off point. (d) Local divergence measure of alignment of the M segment of strains Ch7001, Ch7803, TADJ/HU8966 and Matin. Window 200 nt, step 10 nt. (e) Local divergence measure of alignment of the L segment of strains Matin, TADJ/HU8966, 30908 and IbAr10200 (GenBank accession no. AY389508). Window 500 nt, step 10 nt.

2004) and AY389508 (Honig *et al.*, 2004)], did not match. Similarity plots here revealed an unusual situation: sequence AY389508 (the same as NC_005301) had a sharp drop in similarity at approximately nt 11000–11200 compared with other L segment sequences, including another IbAr10200 sequence (AY389361) (Fig. 1f). The same could be seen on the similarity plots for the translated protein sequences (not shown). Strain IbAr10200 differed from other CCHF virus strains by as much as 28–32% of its amino acid sequence, while other CCHF virus strains differ by only 2–5% of amino acid sequence in this region, and strain IbAr10200 differed from other CCHF virus strains by 1–10% of its amino acid sequence in other genome regions. In contrast, according to a BLAST search, this unusual region was still most similar to the CCHF virus. As it stands, a sequencing error cannot be excluded. Therefore, I will abstain from further discussion here. For other L segment sequences, similar to the M segment, PDM, but not other methods, provided evidence of recombination. Obviously, it is not possible at the moment to prove or rule out recombination in the L segment.

Analysis of the CCHF virus sequences has provided solid evidence of recombination in the S segment, although the situation is far less clear in the other genome segments. To my knowledge, this is the first analysis reporting a clear indication of recombination in tick-borne viruses and in *Nairoviruses*, although recombination has been previously described in *Hantaviruses* (Sibold *et al.*, 1999). The possibility of recombination in CCHF virus was discussed recently by Chare *et al.* (2003), but no clear evidence was shown. Several important conclusions can be drawn from the current results. While data for the S segment strongly support recombination, all of the cases detected were phylogenetically ancient. Indeed, the S segment sequences discussed throughout the paper represent the most distant CCHF virus strains isolated from distant locations. With a tick-borne zoonosis such as CCHF, it is not possible to say even approximately when the described recombination occurred. It could well be decades or millennia ago, or could even have happened at the dawn of the virus species. Similarly, it is not possible to estimate the incidence of recombination in CCHF virus. However, only a few complete CCHF virus sequences are currently available, and evidence of recombination often comes as an avalanche with additional new sequences. In any case, the possibility of recombination should always be kept in mind while conducting phylogenetic studies on CCHF virus. In a

number of previous works (Papa *et al.*, 2002b, c; Yashina *et al.*, 2003a, b), a fragment of S segment sequence often as short as 200–250 nt was used, and some authors relied on such results to draw rather far-fetched conclusions. Such an approach does not seem to be very accurate in light of this report. Considering that there is also evidence of reassortment in CCHF virus (Hewson *et al.*, 2004), all three segments and the full sequence for the S segment should be used in a comprehensive molecular epidemiological study on CCHF virus, at least for the key strains.

The exchange of genetic material obviously requires the co-replication of two strains in one organism. The most suitable hosts for such co-infection events are tick vectors, where the virus can persist and undergo vertical transmission, which further increases the probability of recombination. Multiple recombination events involving strains from very different locations would therefore suggest that CCHF virus co-replication is not uncommon in nature. In consequence, this suggests the existence of a global reservoir of CCHF virus with local subreservoirs supporting extensive levels of virus circulation, which permits frequent co-infection. Such a ‘global reservoir’ can be readily explained by the major role of migrant birds in CCHF epidemiology. Even though there has been no clear evidence of CCHF virus infection in migrant birds, they may translocate infected ticks to distant areas (Hoogstraal, 1979). From a practical point of view, the considerable incidence of recombination and reassortment in CCHF virus means that one should always expect the emergence of virus variants that differ genetically and serologically from common isolates in the area. These new variants can be poorly detectable by ‘reliable’ serological tests and may evade protection by future vaccines.

Acknowledgements

I am grateful to Dr Galina Karganova and Dr Ari Hinkkanen for critical reading of the manuscript.

References

- Bridgen, A., Dalrymple, D. A. & Elliott, R. M. (2002). Dugbe nairovirus S segment: correction of published sequence and comparison of five isolates. *Virology* **294**, 364–371.
- Chare, E. R., Gould, E. A. & Holmes, E. C. (2003). Phylogenetic analysis reveals a low rate of homologous recombination in negative-sense RNA viruses. *J Gen Virol* **84**, 2691–2703.

- Chinikar, S., Persson, S. M., Johansson, M. & 7 other authors (2004). Genetic analysis of Crimean-congo hemorrhagic fever virus in Iran. *J Med Virol* **73**, 404–411.
- Chumakov, M. P., Butenko, A. M., Shalunova, N. V. & 10 other authors (1968). New data on the viral agent of Crimean hemorrhagic fever (in Russian). *Vopr Virusol* **13**, 377.
- Elliott, R. M., Bouloy, M., Calisher, C. H., Goldbach, R., Moyer, J. T., Nichol, S. T., Pettersson, R., Plyusnin, A. & Schmaljohn, C. S. (2000). Family *Bunyaviridae*. In *Virus Taxonomy. Seventh Report International Committee for the Taxonomy of Viruses*, pp. 599–621. Edited by M. H. V. van Regenmortel, C. M. Fauquet, D. H. L. Bishop, E. B. Carstens, M. K. Estes, S. Lemon, J. Maniloff, M. A. Mayo, D. McGeogh, C. R. Pringle & R. B. Wickner. San Diego: Academic Press.
- Felsenstein, J. (1989). PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* **5**, 164–166.
- Hewson, R., Gmyl, A., Gmyl, L., Smirnova, S. E., Karganova, G., Jamil, B., Hasan, R., Chamberlain, J. & Clegg, C. (2004). Evidence of segment reassortment in Crimean-Congo haemorrhagic fever virus. *J Gen Virol* **85**, 3059–3070.
- Honig, J. E., Osborne, J. C. & Nichol, S. T. (2004). Crimean-Congo hemorrhagic fever virus genome L RNA segment and encoded protein. *Virology* **321**, 29–35.
- Hoogstraal, H. (1979). The epidemiology of tick-borne Crimean-Congo hemorrhagic fever in Asia, Europe, and Africa. *J Med Entomol* **15**, 307–417.
- Husmeier, D. & Wright, F. (2001). Probabilistic divergence measures for detecting interspecies recombination. *Bioinformatics* **17**, S123–S131.
- Husmeier, D. & McGuire, G. (2003). Detecting recombination in 4-taxa DNA sequence alignments with Bayesian hidden Markov models and Markov chain Monte Carlo. *Mol Biol Evol* **20**, 315–337.
- Kinsella, E., Martin, S. G., Grolla, A., Czub, M., Feldmann, H. & Flick, R. (2004). Sequence determination of the Crimean-Congo hemorrhagic fever virus L segment. *Virology* **321**, 23–28.
- Marriott, A. C. & Nuttall, P. A. (1992). Comparison of the S RNA segments and nucleoprotein sequences of Crimean-Congo hemorrhagic fever, Hazara, and Dugbe viruses. *Virology* **189**, 795–799.
- Marriott, A. C., Polyzoni, T., Antoniadis, A. & Nuttall, P. A. (1994). Detection of human antibodies to Crimean-Congo haemorrhagic fever virus using expressed viral nucleocapsid protein. *J Gen Virol* **75**, 2157–2161.
- Milne, I., Wright, F., Rowe, G., Marshall, D. F., Husmeier, D. & McGuire, G. (2004). TOPALI: software for automatic identification of recombinant sequences within DNA multiple alignments. *Bioinformatics* **20**, 1806–1807.
- Morikawa, S., Qing, T., Xinqin, Z., Saijo, M. & Kurane, I. (2002). Genetic diversity of the M RNA segment among Crimean-Congo hemorrhagic fever virus isolates in China. *Virology* **296**, 159–164.
- Nichol, S. T. (2001). Bunyaviruses. In *Fields Virology*, 4th edn, pp. 1603–1634. Edited by D. M. Knipe & P. M. Howley. Philadelphia: Lippincott Williams & Wilkins.
- Nichol, S. T., Arikawa, J. & Kawaoka, Y. (2000). Emerging viral diseases. *Proc Natl Acad Sci U S A* **97**, 12411–12412.
- Papa, A., Ma, B., Kouidou, S., Tang, Q., Hang, C. & Antoniadis, A. (2002a). Genetic characterization of the M RNA segment of Crimean Congo hemorrhagic fever virus strains, China. *Emerg Infect Dis* **8**, 50–53.
- Papa, A., Bozovi, B., Pavlidou, V., Papadimitriou, E., Pelemis, M. & Antoniadis, A. (2002b). Genetic detection and isolation of *Crimean-Congo hemorrhagic fever virus*, Kosovo, Yugoslavia. *Emerg Infect Dis* **8**, 852–854.
- Papa, A., Bino, S., Llagami, A. & 9 other authors (2002c). Crimean-Congo hemorrhagic fever in Albania, 2001. *Eur J Clin Microbiol Infect Dis* **21**, 603–606.
- Petrova, I. D., Seregin, S. V., Petrov, V. S. & 8 other authors (2003). Genetic characteristics of the S-segment of RNA from two strains of the Crimean-Congo hemorrhagic fever virus isolated in the south of Russia and in Uzbekistan (in Russian). *Vopr Virusol* **48**, 8–11.
- Ray, S. C. (1999). SimPlot for Windows 95/98/NT, 2.5 edn. Distributed by the author: <http://sray.med.som.jhmi.edu/RaySoft/SimPlot/>
- Rodriguez, L. L., Maupin, G. O., Ksiazek, T. G. & 8 other authors (1997). Molecular investigation of a multisource outbreak of Crimean-Congo hemorrhagic fever in the United Arab Emirates. *Am J Trop Med Hyg* **57**, 512–518.
- Salminen, M. O., Carr, J. K., Burke, D. S. & McCutchan, F. E. (1995). Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. *AIDS Res Hum Retroviruses* **11**, 1423–1425.
- Sanchez, A. J., Vincent, M. J. & Nichol, S. T. (2002). Characterization of the glycoproteins of Crimean-Congo hemorrhagic fever virus. *J Virol* **76**, 7263–7275.
- Schmaljohn, C. S. & Hooper, J. W. (2001). *Bunyaviridae: the viruses and their replication*. In *Fields Virology*, 4th edn, pp. 1447–1472. Edited by D. M. Knipe & P. M. Howley. Philadelphia: Lippincott Williams & Wilkins.
- Seregin, S. V., Samokhvalov, E. I., Petrova, I. D. & 8 other authors (2004). Genetic characterization of the M RNA segment of Crimean-Congo hemorrhagic fever virus strains isolated in Russia and Tajikistan. *Virus Genes* **28**, 187–193.
- Shepherd, A. J., Swanepoel, R. & Gill, D. E. (1988). Evaluation of enzyme-linked immunosorbent assay and reversed passive hemagglutination for detection of Crimean-Congo hemorrhagic fever virus antigen. *J Clin Microbiol* **26**, 347–353.
- Sibold, C., Meisel, H., Kruger, D. H., Labuda, M., Lysy, J., Kozuch, O., Pejcoch, M., Vaheri, A. & Plyusnin, A. (1999). Recombination in Tula hantavirus evolution: analysis of genetic lineages from Slovakia. *J Virol* **73**, 667–675.
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**, 4876–4882.
- Yashina, L., Vyshemirskii, O., Seregin, S. & 9 other authors (2003a). Genetic analysis of Crimean-Congo hemorrhagic fever virus in Russia. *J Clin Microbiol* **41**, 860–862.
- Yashina, L., Petrova, I., Seregin, S. & 10 other authors (2003b). Genetic variability of Crimean-Congo haemorrhagic fever virus in Russia and Central Asia. *J Gen Virol* **84**, 1199–1206.