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...
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)
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. 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.
Recombination in CCHF virus

According to a BLAST search, this unusual region was still present in other genome regions. In contrast, other CCHF virus strains differed by 1–10% of their amino acid sequence in this region, and strain IbAr10200 differed from other CCHF virus strains by as much as 28–32% of its amino acid sequence, shown. Strain IbAr10200 differed from other CCHF virus strains isolated from distant locations. With additional new sequences. In any case, the possibility of recombination often comes as an avalanche of evidence 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 was studied. 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


