Molecular epidemiology of African and Asian Crimean-Congo haemorrhagic fever isolates

F. J. BURT* AND R. SWANEPOEL

Special Pathogens Unit, National Institute for Communicable Diseases, Sandringham, South Africa

(Received 20 December 2004)

SUMMARY

Phylogenetic relationships were examined for 70 Crimean-Congo haemorrhagic fever (CCHF) isolates from southern, central and West Africa, the Middle East and Greece using sequence data determined for a region of the S segment of the genome. Analysis revealed up to 18% genetic differences. Tree topology supports previous evidence for the existence of three groups of genetically related isolates, A, B and C. Within group A there are two clades: an African clade and a predominantly Asian clade comprising isolates from Pakistan, China, Iran, Russia and Madagascar. Group B includes isolates from southern and West Africa and Iran, and group C includes a single isolate from Greece. Despite the potential which exists for dispersal of the virus between Africa and Eurasia, it appears that circulation of the virus is largely compartmentalized within the two land masses, and the inference is that the geographic distribution of phylogenetic groups is related to the distribution and dispersal of tick vectors of the virus.

INTRODUCTION

Crimean-Congo haemorrhagic fever (CCHF) virus is a tick-borne virus found in Africa, Asia and eastern Europe. The virus is endemic in southern Africa and the results of antibody surveys on the sera of cattle and wild vertebrates indicate that the distribution of CCHF virus in South Africa broadly corresponds to that of ticks belonging to the genus Hyalomma, the main vectors of the virus [1]. The ticks are xerophilic and are widely distributed in the drier interior of the country, being absent only from the higher rainfall areas of the eastern coastal region [2]. There are three species of Hyalomma in the country, and higher prevalences of antibody to the virus have been demonstrated in sera collected from areas where H. marginatum rufipes and H. marginatum turanicum are present, than from areas where H. truncatum is the sole representative of the genus [3].

A total of 138 outbreaks involving 171 patients, have been recorded from widely scattered locations in southern Africa from the time that the first case was recognized in 1981 up until the end of 2003. The largest group of cases, 75/171, arose from known tick bite, 69/171 arose from known or potential contact with fresh blood or other tissues of livestock and/or ticks, 7/171 were nosocomial infections which arose from contact with blood or fomites of known CCHF patients, while in 20/171 cases there was no direct evidence of contact with livestock or ticks, but the patients lived in or visited a rural environment where such contact was possible. The case-fatality rate fluctuated around 30% for the first few years after CCHF was initially recognized in southern Africa but has since declined to 25% (44/171), probably because greater awareness of the disease leads to earlier recognition and better management of patients in most instances. The severity and outcome of illness vary for

* Author for correspondence: Dr F. J. Burt, Special Pathogens Unit, National Institute for Communicable Diseases, Private Bag X4, Sandringham 2131, Johannesburg, South Africa.
(Email: felicityb@nicd.ac.za)
reasons that are incompletely understood, but age, underlying health, secondary infection, and the timeliness and adequacy of supportive therapy are probably important factors. Initially it was thought that African strains of CCHF virus were less pathogenic for humans than Asian strains [4], although this is no longer believed to be correct [3]. To determine the genetic diversity of the virus, partial nucleotide sequences were determined for 35 isolates from southern Africa plus 11 isolates obtained from serum samples received from human patients in Iran and Pakistan, while sequence data for a further 40 isolates from other locations were retrieved from GenBank for inclusion in the phylogenetic analysis.

METHODS

Virus isolates

Details relating to the origins of the 70 CCHF isolates included in the study are summarized in Tables 1 and 2. Sequence data were determined retrospectively for 19 human isolates collected in southern Africa between 1984 and 1997 and stored at −70 °C in the form of freeze-dried 10% suckling mouse brain suspensions at the level of second or third mouse brain pass (Table 1). The suspensions were inoculated into Vero cell cultures and total RNA was extracted from the infected cells using the acid guanidium thiocyanate–phenol–chloroform method [5]. Prospectively, viral RNA was isolated directly from 27 serum samples from 25 patients, which included 14 patients from southern Africa, three from Iran and eight from Pakistan submitted to the Special Pathogens Unit (SPU) between 1998 and 2002 (Table 1). Viral RNA was extracted from serum samples using the QIAamp viral RNA isolation kit (Qiagen GmbH, Hilden, Germany) according to the manufacturer’s protocol.

The partial nucleotide sequence was determined for isolates obtained from two patients at different times after onset of illness; SPU 234/99 was obtained from blood collected on day 4 and SPU 235/99 on day 5 from the same patient, while SPU 170/00 and SPU 171/00 were obtained from blood collected on day 8 and 22 h later on the same day from a second patient. An isolate of CCHF virus was passaged intracerebrally through mice and the partial nucleotide sequence for the nucleocapsid gene was determined at the second and third pass level.

Sequence data were retrieved from GenBank for 40 CCHF isolates plus the two related nairoviruses, Hazara and Dugbe. Based on the grouping of isolates obtained in a preliminary phylogenetic analysis, sequence data for 24 isolates, which represented each group and each geographically distinct region, were selected for inclusion in the final analysis (Table 2) [6–9].

Reverse-transcriptase polymerase chain reaction and nucleotide sequencing of amplicons

A 536 nucleotide base pair (bp) fragment of the S segment of the viral genome was amplified from viral RNA by reverse-transcriptase polymerase chain reaction (RT–PCR) using primers designated F2 and R3 [10–12]. One step RT–PCR reactions were performed using the Titan One Tube RT–PCR system (Roche Diagnostics GmbH, Mannheim, Germany) according to the manufacturer’s instructions. The nucleotide sequences for the primers (provided by R. Lofts and J. Smith, Virology Division, United States Army Medical Research Institute for Infectious Diseases) were derived from alignment of the S segment RNA of seven geographically distinct CCHF isolates. The nucleotide sequences and primer positions relative to the positive sense strand of the S segment of CCHF virus isolate IbAr 10200 are as follows

F2 nucleotide position 135–535:

5′-TGGACACCTTCACAAACTC-3′

R3 nucleotide position 670–653:

5′-GACAAATTCCCTGCACCA-3′

The RT–PCR reactions were performed on a PerkinElmer GeneAmp 2400 Thermocycler (Applied Biosystems, Warrington, UK). The following cycling conditions were used: 50 °C for 30 min, 94 °C for 2 min, and 30 cycles of 94 °C for 30 s, 47 °C for 30 s and 68 °C for 30 s, with a final incubation of 68 °C for 7 min. The nucleotide sequences of the amplicons were determined using Big Dye™ Terminator Sequencing Ready Reaction kits with AmpliTaq DNA polymerase FS (Applied Biosystems).

Sequence data analysis

Editing and alignment of the nucleotide sequence data were performed using ClustalX, version 1.81. The phylogenetic analysis was performed on a 450-bp region of the amplicons using a weighted maximum parsimony method, with a transition:transversion weighting of 4:1, and phylogenetic analysis using
parsimony (PAUP) software version 4.0b4a for Macintosh [13]. Bootstrap confidence intervals were calculated by 100 heuristic search replicates. Two nairoviruses, Dugbe and Hazara, were included in the analysis as outgroups. Sequence divergence was determined using Molecular Evolutionary Genetics
Analysis (MEGA) version 2.1 to calculate the average $P$ distances within groups and between groups [14].

RESULTS

Histories of patient infection

Tables 1 and 2 summarize the histories of isolates used in this study including information, where available, regarding source of infection, geographic location, year of infection and outcome of illness for human patients. The southern African specimens were collected over a period of 17 years. The majority of the southern African patients (18/33), were infected by tick bite, one patient had a nosocomial infection, 6/33 patients had contact with fresh blood or other tissues of livestock and/or ticks, and 8/33 patients had unknown means of infection but visited or resided in a rural area where exposure to ticks and livestock was possible. A total of 18/33 patients had fatal infection. No correlation could be made between the source of infection and the severity of illness. Detailed histories were not available for the patients of the specimens received from Iran and Pakistan, but most patients were rural people with the potential for exposure to ticks and livestock, while one patient who suffered a fatal infection in Iran was a doctor who putatively acquired nosocomial infection.

Genetic analysis

Preliminary analyses included CCHF isolates representing all the geographic regions available on GenBank for a region of the S segment amplified using the primer pair F3R2 described in previous reports [10, 11]. Isolates from the United Arab Emirates (UAE 9509853) and Kosovo (9553/2001), excluded from the final analysis because of the length of

<table>
<thead>
<tr>
<th>Isolate</th>
<th>Year</th>
<th>Strain origin</th>
<th>Locality</th>
<th>GenBank no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>AP 92</td>
<td>1975</td>
<td><em>Rhipicephalus bursa</em></td>
<td>Greece</td>
<td>U04958</td>
</tr>
<tr>
<td>HD 49199</td>
<td>1988</td>
<td>Human</td>
<td>Mauritania</td>
<td>U15023</td>
</tr>
<tr>
<td>ArD 39554</td>
<td>1984</td>
<td><em>Hyalomma marginatum rufipes</em></td>
<td>Mauritania</td>
<td>U15089</td>
</tr>
<tr>
<td>ArMg 951</td>
<td>1985</td>
<td><em>Boophilus microplus</em></td>
<td>Madagascar</td>
<td>U15024</td>
</tr>
<tr>
<td>ArD 15786</td>
<td>1973</td>
<td><em>H. truncatum</em></td>
<td>Senegal</td>
<td>U15020</td>
</tr>
<tr>
<td>ArD 8194</td>
<td>1969</td>
<td><em>H. marginatum rufipes</em></td>
<td>Senegal</td>
<td>U15021</td>
</tr>
<tr>
<td>ArD 97264</td>
<td>1993</td>
<td><em>H. truncatum</em></td>
<td>Senegal</td>
<td>U15090</td>
</tr>
<tr>
<td>ArD 97268</td>
<td>1993</td>
<td><em>H. nitidum</em></td>
<td>Central African Republic</td>
<td>U15092</td>
</tr>
<tr>
<td>ArB 604</td>
<td>1973</td>
<td><em>H. marginatum rufipes</em></td>
<td>Mauritania</td>
<td>U15024</td>
</tr>
<tr>
<td>HD 38562</td>
<td>1983</td>
<td>Human</td>
<td>Burkin Faso</td>
<td>U15093</td>
</tr>
<tr>
<td>UG3010</td>
<td>1956</td>
<td>Human</td>
<td>Democratic Republic of Congo</td>
<td>U88415</td>
</tr>
<tr>
<td>IbAr10200</td>
<td>1966</td>
<td><em>H. excavatum</em></td>
<td>Nigeria</td>
<td>NC005302</td>
</tr>
<tr>
<td>JD 206</td>
<td>1965</td>
<td><em>H. anatolicum</em></td>
<td>Pakistan</td>
<td>U88414</td>
</tr>
<tr>
<td>729/02</td>
<td>n.a.</td>
<td>n.a.</td>
<td>Iran</td>
<td>AY366375</td>
</tr>
<tr>
<td>714/02</td>
<td>n.a.</td>
<td>n.a.</td>
<td>Iran</td>
<td>AY366376</td>
</tr>
<tr>
<td>786/02</td>
<td>n.a.</td>
<td>n.a.</td>
<td>Iran</td>
<td>AY366378</td>
</tr>
<tr>
<td>ArFeh193-3</td>
<td>1978</td>
<td><em>Alectorobius laiorenis</em></td>
<td>Iran</td>
<td>U15022</td>
</tr>
<tr>
<td>BA8402</td>
<td>1984</td>
<td><em>Hyalomma tick pool</em></td>
<td>China</td>
<td>AJ010649</td>
</tr>
<tr>
<td>HY 13</td>
<td>1968</td>
<td><em>H. asiaticum</em></td>
<td>China</td>
<td>U88413</td>
</tr>
<tr>
<td>BA66019</td>
<td>1965</td>
<td>Human</td>
<td>China</td>
<td>AJ010648</td>
</tr>
<tr>
<td>TI10145</td>
<td>1985</td>
<td><em>H. asiaticum</em></td>
<td>Uzbekistan</td>
<td>AY049082</td>
</tr>
<tr>
<td>STV/HU29223</td>
<td>2000</td>
<td>Human</td>
<td>Russia</td>
<td>AF481802</td>
</tr>
<tr>
<td>POS/T128044</td>
<td>2000</td>
<td><em>H. marginatum</em></td>
<td>Russia</td>
<td>AY277672</td>
</tr>
<tr>
<td>Drosdov</td>
<td>1967</td>
<td>Human</td>
<td>Russia</td>
<td>U88412</td>
</tr>
<tr>
<td>Hazara JC 280</td>
<td>1964</td>
<td><em>Ixodes reikorzevi</em></td>
<td>Pakistan</td>
<td>M86624</td>
</tr>
<tr>
<td>Dugbe ArD 44313</td>
<td>1985</td>
<td><em>Amblyomma variegatum</em></td>
<td>Senegal</td>
<td>M25150</td>
</tr>
</tbody>
</table>

n.a., Not available.
The P distance calculated from nucleotide sequence data between the groups is illustrated below the diagonal, the value calculated from predicted amino-acid sequences is shown above the diagonal. The bold figures illustrated on the diagonal are the P distances within the groups calculated from nucleotide/amo-nucleotide data. n.c., Not calculable.

sequence data available, were found in group A in the Asian clade when tested using a weighted parsimony method (data not shown). Phylogeny was reconstructed using sequence data from 86 CCHF isolates (450 bp, amplified using the primer pair F2 and R3) and finally to simplify the tree, the analysis was performed including only 24/40 isolates from GenBank which represented isolates from distinct geographic regions on major branches. The topology of major branches obtained from analysis of 70 CCHF isolates was identical to that using 86 isolates. A representative tree generated by a weighted maximum parsimony analysis of the partial nucleotide sequence of the S segment of 70 CCHF isolates from 15 countries is shown in the Fig. The node values were generated by 100 bootstrap replications. The tree topology indicates that there are three groups of genetically related isolates, A, B and C. Within group A there are two clades: an African clade and a predominantly Asian clade including isolates from Pakistan, China, Russia, Iran and Madagascar. Group B includes isolates from southern and West Africa, and group C includes a single virus isolated from a tick in Greece. The tree topology shows no obvious correlation between the grouping of isolates and source of infection, year of infection, or pathogenicity for humans, but the geographic distribution of phylogenetic groups appears to relate to the distribution and dispersal of vectors of the virus as discussed below.

The average genetic P distances are shown in Table 3. Base changes were observed throughout the portion of the genome analysed, but the majority of the nucleotide changes were synonymous, despite the 1–18% nucleotide differences observed between isolates in groups A and B, excluding the isolate from

Table 3. Genetic diversity of CCHF isolates determined using average P distances calculated with MEGA (version 2.1) between and within virus subtypes A–C

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.095/0.029</td>
<td>0.052</td>
<td>0.108</td>
</tr>
<tr>
<td>B</td>
<td>0.164</td>
<td>0.033/0.018</td>
<td>0.104</td>
</tr>
<tr>
<td>C</td>
<td>0.2</td>
<td>0.183</td>
<td>n.c.</td>
</tr>
</tbody>
</table>

The average genetic P distances are shown in Table 3. Base changes were observed throughout the portion of the genome analysed, but the majority of
Greece, amino-acid homology was 92.2–100% for the predicted proteins. No nucleotide changes were observed in virus isolates obtained from two patients at different times after onset of illness, or in an isolate of CCHF virus passaged twice in mice (data not shown).

**DISCUSSION**

Factors affecting the pathogenesis of CCHF are incompletely understood, but host factors are probably important determinants of the severity and outcome of illness. To determine whether genetic diversity of the virus correlates with pathogenicity for humans, year and source of infection, or geographic origin, we examined genetic relatedness among southern African and other isolates, using a 450 nucleotide region of the S segment of the virus genome. There were 1–18% differences in nucleotide sequences between groups A and B, which translated into 92.2–100% homology in predicted amino-acid sequences. Our sequence data are consistent with the observations of Marriott & Nuttall [15], who found approximately 20% nucleotide variability between 13 CCHF isolates from Africa, Madagascar, Iran, China and Greece. The high degree of amino-acid homology observed could explain the antigenic similarity of the nucleocapsid protein of CCHF isolates throughout the distribution range of the virus [16–18]. In agreement with previous studies, the tree topology indicates that there are three subtypes of CCHF virus in circulation, which have been designated A, B and C [11, 15]. One lineage of subtype A circulates throughout Africa, while a second lineage of subtype A circulates in Asia and Madagascar. Subtype B circulates in southern and West Africa and Iran, while subtype C is represented by a unique isolate from Greece. There were no obvious correlations between the genetic relationships of isolates and source of infection, year of infection, or pathogenicity for humans, as would be expected with contagious or mosquito-borne viruses prone to rapid evolution and epidemic spread. Instead, the phylogetic groups conform to a broad pattern of geographic distribution which appears to relate to the distribution and dispersal of the tick vectors of the virus. In an investigation of the genetic variability of CCHF in Russia and Central Asia, Yashina et al. [9] found that the greatest genetic variability was seen in isolates from different tick species rather than from geographically distinct areas and the authors suggested that a long-term association with a particular tick species plays a role in genetic variability.

Ixodid ticks have three stages in their life-cycle – larvae, nymphs, and adults – each of which attaches and feeds on a separate vertebrate host before detaching and moulting to the next instar, although some species remain attached during the first or even second moult. Cumulatively, the ticks remain on their vertebrate hosts for only a few days to weeks and, in contrast, spend months away from the hosts during the moult or while the adult females lay eggs which must hatch before the next generation of larvae are ready to feed. Transovarial transmission of infection from female ticks to their progeny occurs with low frequency, but even in the absence of such transmission moulting ticks constitute a reservoir of infection which ensures perpetuation of the virus in the environment from one season or year to the next [4, 19–21]. Thus, tick-borne virus diseases tend to persist within fixed geographic ranges determined by the distribution of the vectors.

CCHF virus has been isolated from 30 species of ixodid ticks, and although it has been demonstrated that members of several genera are capable of transmitting infection, the importance of many species as vectors remains uncertain because virus isolated from engorged ticks may merely have been present in the bloodmeal imbibed from a viraemic host [4, 22, 23]. With the notable exception of Madagascar, however, the distribution of CCHF virus falls exactly within the limits of world distribution of *Hyalomma* ticks, and members of this genus are regarded as the principal vectors.

Mechanisms for the dissemination of the tick vectors and CCHF virus must have operated for millennia, and include bird migration and the movement of livestock and wild animals [3, 4, 24, 25]. Apart from ostriches, birds have been found to be refractory to infection with CCHF virus, but they can support the so-called phenomenon of ‘non-viraemic’ transfer of infection between ticks co-feeding on a host [4, 26–29]. More importantly, birds are known to be parasitized by the immature stage of vectors of CCHF virus, *H. marginatum marginatum* and *H. m. turanicum* of eastern Europe and Asia, plus *H. m. rufipes*, *H. truncatum* and *H. nitidum* of Africa, and can thus serve to disseminate transovarially infected immature ticks on a local and intercontinental basis. It was observed in Egypt that vast numbers of birds migrating from the northern hemisphere in autumn were most heavily infested with *H. m. marginatum* immatures.
whereas spring migrants going north were generally infested with immature *H. m. rufipes* ticks [24, 25], both proven vectors of CCHF virus.

Trade in livestock along the Asian and east African coasts has been associated with outbreaks of human disease in the Near East [11, 30–33]. However intercontinental spread of CCHF virus may be a slow process, taking place over centuries, which accords with the observation that the lineages of CCHF strains appear to circulate largely within continents, Africa or Eurasia, despite mechanisms for movement of ticks between continents. Although large numbers of ticks may be carried between the continents by migrating birds and slaughter animals, the establishment of a tick species in a new environment depends on the availability of a suitable micro-climate, introduction of sufficient numbers of ticks to ensure breeding, and the presence of suitable hosts. In contrast, ticks dispersed within a continent by local movement of birds or domestic and wild animals are being circulated within their endemic distribution range. Nevertheless, it must be acknowledged that on occasion infected immature ticks could detach from migrating birds, moult to the next instar, and transmit infection to a second host in the new environment. However, the low infection rates inherent in ticks, and the high attrition rate in the tick life-cycle, would limit the possibility of this occurring. Moreover, the relatively short periods for which immature ticks remain attached to hosts, would tend to limit the penetration of tick species from one continent into another on migrating birds.

The fact that the two lineages of subtype A of CCHF virus are closer to each other than to subtype B may reflect gradual exchange of virus genetic material between the two land masses. The close relatedness of an isolate from Iran in subtype B may be the result of livestock trade between the continents. Interestingly, the isolate from Madagascar was obtained from a tick species, *Boophilus microplus*, which occurs primarily in Pakistan and India, and is believed to have been introduced into Madagascar on cattle imported from Asia [4]. This could explain the close relationship observed between the CCHF virus isolate from Madagascar and the isolates from Pakistan (Fig.). The uniqueness of the Greek isolate suggests the existence of isolating mechanisms which may relate to bird migratory paths, or to the association of the virus with a particular tick species, *Rhipicephalus bursa*, from which it was isolated. Further investigations focusing on tick species associated with CCHF strains would help to identify the role of tick species and genetic variability.

It can be concluded that the phylogenetic evidence supports the concept of the evolution of CCHF virus strains within specific geographic regions, probably as a consequence of association with particular tick vector species.

**ACKNOWLEDGEMENTS**

The authors acknowledge Dr Athar Saeed Dil and Mr S. Zaidi, National Institute for Health, Islamabad, Pakistan and Dr R. Nategh, Tehran University of Medical Sciences, Iran for submitting specimens from patients with CCHF infections.

**REFERENCES**


