

Next-generation sequencing of southern African Crimean-Congo haemorrhagic fever virus isolates reveals a high frequency of M segment reassortment

D. GOEDHALS¹, P. A. BESTER¹, J. T. PAWESKA^{2,3}, R. SWANEPOEL⁴
AND F. J. BURT^{1*}

¹ *Department of Medical Microbiology and Virology, National Health Laboratory Service/University of the Free State, Bloemfontein, South Africa*

² *Centre for Emerging and Zoonotic Pathogens, National Institute for Communicable Diseases, National Health Laboratory Service, Johannesburg, South Africa*

³ *School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, South Africa*

⁴ *Zoonoses Research Unit, Department of Medical Virology, University of Pretoria, South Africa*

*Received 11 December 2013; Final revision 7 March 2014; Accepted 14 March 2014;
first published online 1 May 2014*

SUMMARY

Crimean Congo haemorrhagic fever virus (CCHFV) is a bunyavirus with a single-stranded RNA genome consisting of three segments (S, M, L), coding for the nucleocapsid protein, envelope glycoproteins and RNA polymerase, respectively. To date only five complete genome sequences are available from southern African isolates. Complete genome sequences were generated for 10 southern African CCHFV isolates using next-generation sequencing techniques. The maximum-likelihood method was used to generate tree topologies for 15 southern African plus 26 geographically distinct complete sequences from GenBank. M segment reassortment was identified in 10/15 southern African isolates by incongruencies in grouping compared to the S and L segments. These reassortant M segments cluster with isolates from Asia/Middle East, while the S and L segments cluster with strains from South/West Africa. The CCHFV M segment shows a high level of genetic diversity, while the S and L segments appear to co-evolve. The reason for the high frequency of M segment reassortment is not known. It has previously been suggested that M segment reassortment results in a virus with high fitness but a clear role in increased pathogenicity has yet to be shown.

Key words: Arboviruses, bunyaviruses, haemorrhagic fever, molecular epidemiology.

INTRODUCTION

Crimean Congo haemorrhagic fever virus (CCHFV) is a member of the family *Bunyaviridae*, genus *Nairovirus* [1]. This tick-borne zoonotic virus is associated with clinical disease ranging from a non-specific febrile illness to severe disease manifesting as haemorrhagic

fever. CCHFV has a negative-sense, single-stranded RNA genome consisting of three segments designated large (L), medium (M) and small (S). The highly conserved complementary terminal nucleotide sequences result in loosely circular RNAs which together with the nucleocapsid protein, make up the three helical nucleocapsids. The L segment encodes the viral RNA polymerase while the S segment encodes the viral nucleocapsid protein. The M segment encodes a polyprotein which undergoes proteolytic processing to yield the viral glycoproteins, Gc and Gn [2].

CCHFV has been documented in more than 30 countries of Africa, Asia, Europe and the Middle

* Author for correspondence: Professor F. J. Burt, Department of Medical Microbiology and Virology, National Health Laboratory Service, Faculty of Health Sciences, University of the Free State, Bloemfontein, 9301, South Africa.
(Email: burtfj@ufs.ac.za)

East, with a distribution following that of *Hyalomma* ticks, the principal vector of the virus [3, 4]. Since the first description of clinical disease due to CCHFV infection in South Africa in 1981, 192 cases have been confirmed [5]. However, complete S, M and L sequences are available for only four isolates collected from humans and one tick isolate from South Africa. These, as well as other published CCHFV sequences, were determined by Sanger sequencing using primer walking [6–18]. In recent years, a number of next-generation sequencing (NGS) methods have been developed which yield large amounts of sequencing data at relatively low cost. Although the specifics differ, these methods all make use of three steps, namely, library preparation, DNA capture and enrichment, and sequencing or detection. In the field of virology, these techniques have been employed for various purposes including the discovery of novel viruses, whole viral genome sequencing, and ‘deep’ sequencing to determine viral quasi-species or genome variability [19]. This study aimed to make use of NGS techniques to obtain whole genome sequences of southern African CCHFV isolates in order to perform genetic analysis including identification of reassortment events.

METHODS

Viral isolates

Ten CCHFV isolates obtained from patients in southern Africa between 1985 and 2008 were sequenced retrospectively. RNA was extracted and supplied by Professor J. T. Paweska, Centre for Emerging and Zoonotic Pathogens, National Institute for Communicable Diseases (NICD), Johannesburg from cell culture preparations or mouse brain suspensions. Total RNA was extracted for isolate SPU44/08 using TRIzol[®] reagent (Invitrogen, USA) according to the manufacturer’s instructions. The remainder of the isolates were stored at –70 °C as freeze-dried 10% suckling mouse brain suspensions at the level of mouse brain passage 2–3. The suspensions were inoculated into Vero cell cultures and total RNA extracted from the infected cells using the acid guanidium thiocyanate-phenol-chloroform method as described previously [20]. The extracted RNA was stored at –70 °C until use.

Reverse transcriptase–polymerase chain reaction (RT–PCR) and sequencing

The complete L segment was amplified using two previously described primer pairs that generated

overlapping amplicons [8]. The complete M segment was amplified using primers designed by Deyde *et al.* [8]. Existing sequences in GenBank were used to design primers SF1 and SR3 (Table 1) for amplification of the S segment. The respective forward primers were used to perform the reverse transcription step for each amplicon using SuperScript[™] III Reverse Transcriptase (Invitrogen). PCR was performed with the Expand Long Template PCR System (Roche Diagnostics GmbH, Germany) using standard cycling conditions according to the manufacturer’s instructions and an annealing temperature of 48 °C. The PCR products were gel extracted and purified using the Wizard[®] SV Gel and PCR Clean-Up System (Promega, USA).

The complete genomes for 3/10 isolates (SPU431/85, SPU383/87, SPU 130/89) were determined using the Ion Torrent PGM[™] sequencer (Life Technologies, USA) by the Central Analytical Facility, Stellenbosch University. For the remaining seven isolates, the S segment data was determined previously in the laboratory using the Big Dye[™] Terminator v. 3.1 Cycle Sequencing kit (Applied Biosystems, USA) according to the manufacturer’s instructions and three overlapping primer sets as described in Table 2. The L and M segments of these isolates were sequenced at the NICD, Johannesburg using the Roche 454 GS Junior[™] sequencing system (Roche Diagnostics GmbH). Additional sequencing of all isolates for incomplete coverage and clarification of ambiguities, including mixed bases or nucleotide substitutions and base insertions or deletions, was performed as required using the Big Dye[™] Terminator v. 3.1 Cycle Sequencing kit (Applied Biosystems). Details regarding the additional primers used are available from the corresponding author upon request.

Data analysis

Raw sequencing data was converted from SFF format to FASTQ files using the `sff_extract` script (available as part of `seq_crumbs` at <http://bioinf.comav.upv.es/>) and trimming and filtering of reads based on length and quality scores was then performed using PRINSEQ [21]. Sequences for the L, M and S segments available in Genbank were used to compile databases and separate the reads into L, M and S segment-related data using `filter_by_blast` (http://bioinf.comav.upv.es/seq_crumbs/available_crumbs.html). *De novo* assembly of the blast-filtered and unfiltered reads was performed using MIRA [22]. The resulting

Table 1. PCR and sequencing primers utilized for generation of amplicons used for sequencing

Primer name	5'-3' position relative to SPU415/85	5'-3' sequence	Reference
L1F	1-22	TCT CAA AGA TAT CAA TCC CCC C	Deyde, 2006 [8]
L1R	7773-7752	TTG GCA CTA TCT TTC ATT TGA C	Deyde, 2006 [8]
L2F	6137-6158	GAA GAG CTA TAT GAC ATA AGG C	Deyde, 2006 [8]
L2R	12157-12134	TCT CAA AGA AAT CGT TCC CCC CAC	Deyde, 2006 [8]
MF	1-18	TCT CAA AGA AAT ACT TGC	Deyde, 2006 [8]
MR	5354-5337	TCT CAA AGA TAT AGT GGC	Deyde, 2006 [8]
SF1	1-21	TCT CAA AGA AAC ACG TGC CGC	In house
SR1*	590-569	GGT TCC TTC TCC TAA TCA TGT C	In house
SF2*	467-487	GGT TTC CGT GTC AAT GCA AAC	In house
SR2*	1114-1095	CAT TGG GGT GCT CAG CAG AG	In house
SF3*	1029-1048	CGA CGG TGT CAC AGT TCC TC	In house
SR3	1673-1651	TCT CAA AGA TAT CGT TGC CGC AC	In house

* Primers designed for sequence determination.

Table 2. Summary of data concerning the CCHFV sequences retrieved from the GenBank database and used in the study

Strain	Accession number			Year and location of isolation	Source of isolate
	L	M	S		
C68031	DQ211616	DQ211629	DQ211642	1968 China	Sheep
Oman	DQ211619	DQ211632	DQ211645	1997 Oman	Human
ArD15786	DQ211614	DQ211627	DQ211640	1972 Senegal	Goat
ArD8194	DQ211613	DQ211626	DQ211639	1969 Senegal	Tick
ArD39554	DQ211615	DQ211628	DQ211641	1984 Mauritania	Tick
SPU415/85	DQ211622	DQ211635	DQ211648	1985 South Africa	Human
SPU97/85	DQ211620	DQ211633	DQ211646	1985 South Africa	Human
SPU103/87	DQ211621	DQ211634	DQ211647	1987 South Africa	Human
Drosdov	DQ211617	DQ211630	DQ211643	1967 Russia	Human
Kashmanov	DQ211618	DQ211631	DQ211644	1967 Russia	Human
Baghdad-12	AY947890	AJ538197	AJ538196	1979 Iraq	Human
Matin	AY422208	AF467769	AF527810	1976 Pakistan	Human
Kosova Hoti	EU044832	EU037902	DQ133507	2001 Kosovo	Human
Turkey 200310849	DQ211623	DQ211636	DQ211649	2003 Turkey	Human
UG3010	DQ211624	DQ211637	DQ211650	1956 DRC	Human
AP92	DQ211612	DQ211625	DQ211638	1975 Greece	Tick
TADJ/HU8966	AY720893	AY179962	AY049083	1990 Tajikistan	Human
ROS/HUVLV-100	AY995166	DQ206448	DQ206447	2003 Russian Federation	Human
Ibar10200	AY947891	AF467768	U88410	1966 Nigeria	Tick
Semunya	DQ076412	DQ094832	DQ076413	1958 Uganda	Human
Turkey-Kelkit06	GQ337055	GQ337054	GQ337053	2006 Turkey	Human
SPU128/81/7	DQ076414	DQ157174	DQ076415	1981 South Africa	Tick
SPU4/81	DQ076417	DQ157175	DQ076416	1981 South Africa	Human
Afg09-2990	HM452307	HM452306	HM452305	2009 Afghanistan	Human
NIVA 118594	JN572092	JN572084	JN572087	2011 India	Tick
NIV 112143	JN572091	JN572085	JN572089	2011 India	Human
NIVA 118595	JN572090	JN572083	JN572088	2011 India	Tick
YL04057	FJ562095	FJ562094	FJ562093	2004 China	Tick
Sudan AB1-2009	HQ378183	HQ378187	HQ378179	2009 Sudan	Human
Sudan Al-Fulah 3-2008	HQ378180	HQ378184	GQ862371	2008 Sudan	Human
79121M18	GU477492	GU477493	GU477494	1979 China	Rodent

Table 3. Summary of southern African CCHFV isolates included in the study

Strain	Accession numbers			Year and location of isolation	Source of human infection	Outcome
	L	M	S			
SPU431/85	KJ682799	KJ682812	KJ682815	1985 Northern Cape	Nosocomial	Fatal
SPU383/87	KJ682801	KJ682806	KJ682816	1987 Free State	Tick	Survived
SPU556/87	KJ682798	KJ682811	KJ682817	1987 Northern Cape	Tick	Fatal
SPU18/88	KJ682803	KJ682810	KJ682818	1988 Northern Cape	Tick	Fatal
SPU45/88	KJ682796	KJ682809	KJ682819	1988 Free State	Tick	Fatal
SPU497/88	KJ682804	KJ682808	KJ682820	1988 Namibia	Livestock/ tick	Fatal
SPU130/89	KJ682802	KJ682807	KJ682821	1989 Northern Cape	Tick	Survived
SPU48/90	KJ682797	KJ682813	KJ682822	1990 North West Province	Unknown	Survived
SPU187/90	KJ682795	KJ682814	KJ682823	1990 North West Province	Abattoir	Fatal
SPU44/08	KJ682800	KJ682805	KJ682824	2008 Free State	Livestock/tick	Survived

contigs from both methods were assembled in Geneious (Geneious v. 4.8.5, Biomatters, <http://www.geneious.com/>) and compared to known complete CCHFV sequences to identify areas of incomplete coverage or ambiguities which required further investigation. Ambiguities and homopolymers were investigated using Gap5 [23] and the majority could be clarified by evaluation of the quality scores at each relevant position. Where necessary, additional Sanger sequencing was performed and sequences were incorporated into the assembly using Geneious and ChromasPro v. 1.42 (Technelysium Pty Ltd, Australia). Alignments were confirmed by visual inspection with reference sequences and comparison with previously published partial sequences.

Complete genome sequence data for 31 isolates were retrieved from GenBank as summarized in Table 2. The sequences were aligned using Clustal X version 2.0 [24] and analysed using Molecular Evolutionary Genetics Analysis (MEGA) version 5 [25] with the bootstrap maximum-likelihood method with 1000 replicates. Sequence divergence was also determined using MEGA by calculating the average *P* distances within and between sequence groups as well as pairwise distances for nucleotide and deduced amino acid sequences.

RESULTS

Sequencing data

Full coverage of complete L, M and S segments was obtained for 10 southern African CCHFV isolates. The details of the 10 isolates are summarized in Table 3. Ion Torrent PGM sequencing yielded raw data of >80000 reads including >13 million bases

for each of the three isolates. Following stringent filtering and trimming of reads based on length as well as quality scores, >35000 reads amounting to ~7 million bases were included in the final analysis of each of these isolates. The 454 GS Junior sequencing data were more variable and duplicate runs were performed for four of the isolates in order to obtain complete coverage. The raw data for these seven isolates yielded between 2404 and 17315 reads including 1–8 million bases per isolate. Following filtering and trimming, the final alignments were performed on between 1105 and 7042 reads with ~500000 to 3.5 million bases per isolate. The alignments generated by MIRA *de novo* assembly of both blast-filtered and unfiltered data corresponded, although improved contiguous quality was achieved with some isolates following the application of filter_by_blast. A comprehensive sequence database and optimization of blast parameters were required to ensure effective filtering of reads without loss of coverage, particularly of the highly variable M segment.

Genetic analysis

The genetic relationship of isolates as determined by the maximum-likelihood method is shown for the S (Fig. 1a), M (Fig. 1b) and L (Fig. 1c) segments. The phylogenetic groups were designated I–VII as defined previously [8, 26]. The complete sequences confirm the phylogenetic grouping of southern African isolates SPU497/88, SPU130/89, SPU383/87, SPU18/88 and SPU45/88 and the occurrence of reassortment of the M segment of these isolates as suggested by partial sequencing [27]. The complete sequences also confirm that the incongruencies in

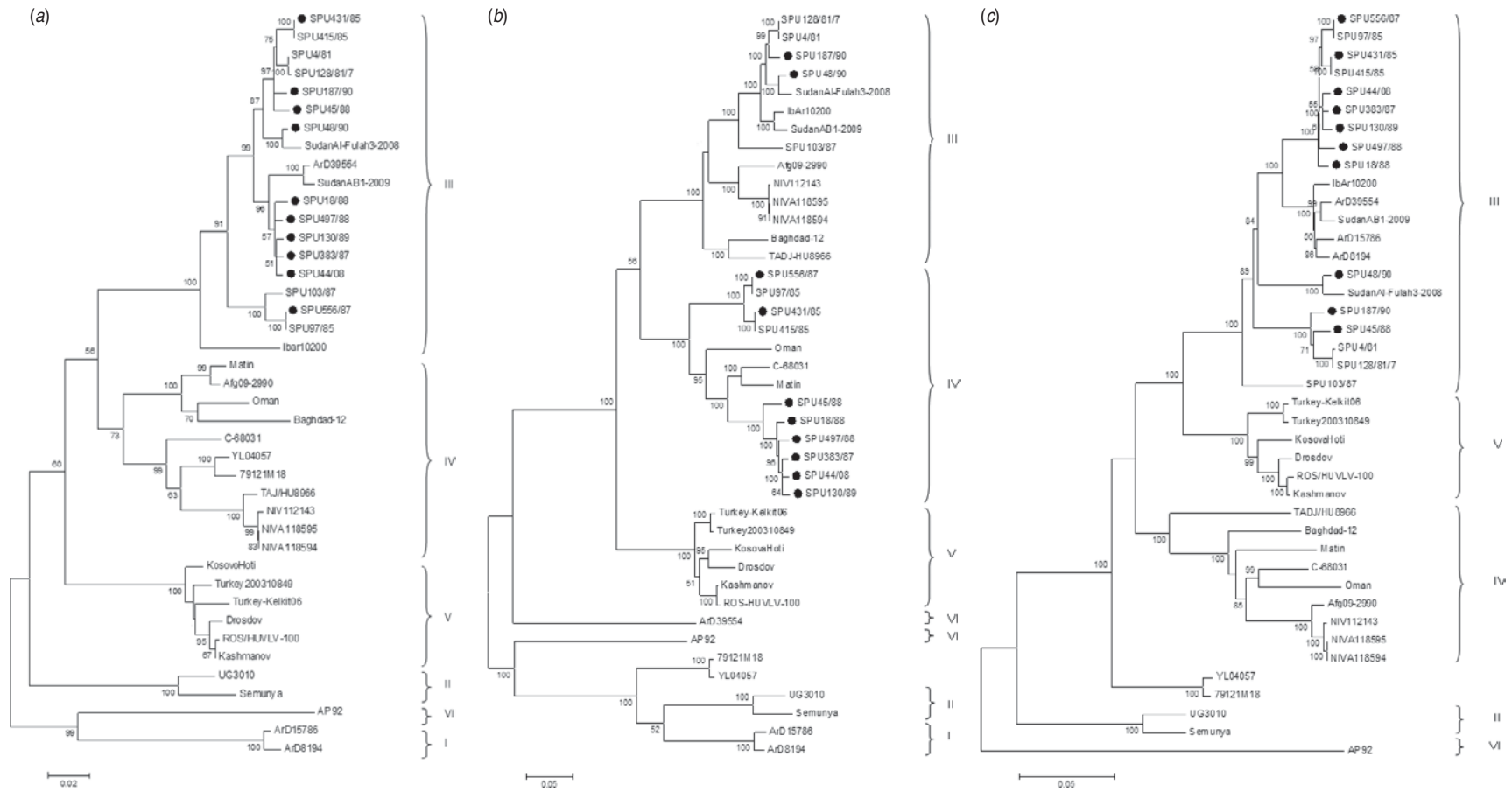


Fig. 1. Phylogenetic analysis of complete coding regions of (a) S segments, (b) M segments, and (c) L segments of CCHFV using a bootstrap maximum-likelihood method with 1000 replicates, with bootstrap values >50% indicated at the relevant nodes. Each sequence is designated by the isolate name and isolates sequenced in the current study are indicated by a solid circle (●).

M segment grouping were due to reassortment rather than recombination events. This pattern of M segment reassortment, showing clustering of the S and L segments with group III strains from South and West Africa while the M segment clusters with group IV isolates from Asia and the Middle East, was also noted for newly sequenced isolates SPU556/87, SPU44/08 and SPU431/85. Isolates SPU48/90 and SPU187/90 showed no evidence of reassortment, with all three segments falling within the group III South/West Africa lineage. Further evidence of M segment reassortment was seen in southern African isolates SPU97/85 and SPU415/85, as described previously [8]. No evidence of S segment reassortment was seen in the available complete sequences, but L segment reassortment was seen in two CCHFV isolates from Senegal, ArD15786 and ArD8194 [8]. The Chinese isolates YL04057 and 79121M18 appear to represent a new group as suggested by Zhou *et al.* [18], with S segments related to group IV Asia and Middle East, while the M and L segments cluster separately from other known groups. Isolate SPU431/85 was obtained from the same patient but from a sample collected 3 days subsequent to previously published SPU415/85. The sequences correlated well with only a single non-synonymous mutation in the coding region of the L segment and four non-synonymous mutations in the coding region of the M segment, as confirmed by Sanger sequencing. RNA from both of these isolates was obtained from cell cultures and these additional passages may have contributed to the accumulated mutations noted.

The geographical distribution of CCHFV groups is illustrated in Figure 2. The S and L segments show a strong correlation in the distinct geographical grouping. The map displaying M segment group distribution clearly illustrates the blending of Asian and African strains, while the remaining groups correlate with S and L segment distributions.

The mean *P* distances within groups and between groups as calculated with MEGA support the phylogenetic groupings. The S segment nucleotide distances within groups were generally low at between 0.1% and 1.4% and similar to the amino acid distances within groups which ranged from 0.1% to 1.8%. The nucleotide and amino acid distances between S segment groups were also similar at 2.5–6.8% and 2.8–8.4%, respectively. The L segment showed a similar degree of diversity within groups at both nucleotide and amino acid levels, but increased diversity between groups. Nucleotide distances within groups ranged

from 0.4% to 2.5% and between groups from 3.8% to 11.3%, while amino acid distances within groups were between 0.5% and 1.8% and amino acid distances between groups were 2.9–10.1%. The M segment showed the greatest genetic diversity, particularly between groups, including numerous non-synonymous mutations. The nucleotide distances within groups ranged from 0.4% to 5.5%, while the nucleotide distances between groups were between 10.9% and 26.8%. At the amino acid level, the distances within groups were between 1.1% and 8.3% and the distances between groups were between 15.1% and 30.1%.

Among the southern African isolates, the S segment was highly conserved with both nucleotide and amino acid mean distances of <1% (range 0–1.3% and 0–1.7%, respectively). The diversity was also low for the L group with mean distances of 1.7% at the nucleotide level (range 0–2.8%) and 1.2% at the amino acid level (range 0–2.1%). The diversity of M segments was higher in group IV isolates than group III, with mean nucleotide distances of 5% and 2.8%, respectively and mean amino acid distances of 7.5% and 4.1%, respectively. Overall, the southern African M segments showed distances of 0–12.4% at the nucleotide level and 0–15.6% at the amino acid level.

Although no association between M segment reassortment and pathogenicity was noted in the southern African isolates included in the current study, the number of both non-fatal infections and non-reassortants included was small. Furthermore, no temporal association was seen with the isolation of reassortants.

DISCUSSION

The S segment of CCHFV was previously investigated as a surrogate for complete sequencing due to its length and the fact that it is relatively conserved in comparison to the M and L segments thereby simplifying the sequencing process. However, the demonstration of both reassortment [8, 20, 27, 28] and, less commonly, recombination [29] among CCHFV genomes confirmed the necessity of at least partial sequencing of all three segments in order to perform accurate genetic analyses. The present study made use of two NGS platforms, namely the Ion Torrent PGM and the Roche 454 GS Junior sequencing systems, to obtain complete CCHFV sequences for ten southern African isolates. The largest collection of complete CCHFV genome sequences to date made use of 25 S, 40 M and 84 L primers in order to obtain

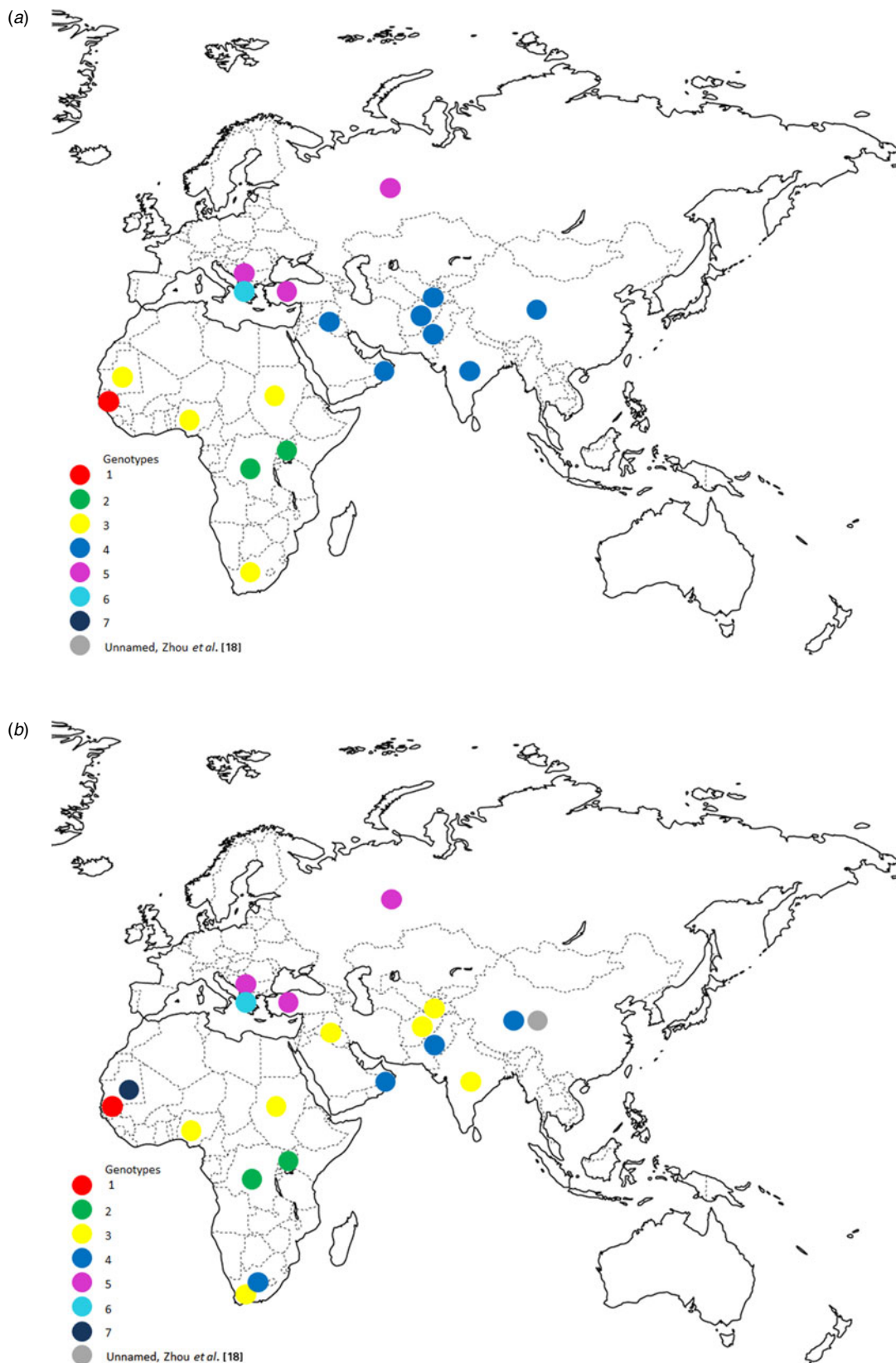


Fig. 2. Geographical distribution of CCHFV groups for the (a) S segment, (b) M segment, and (c) L segment.

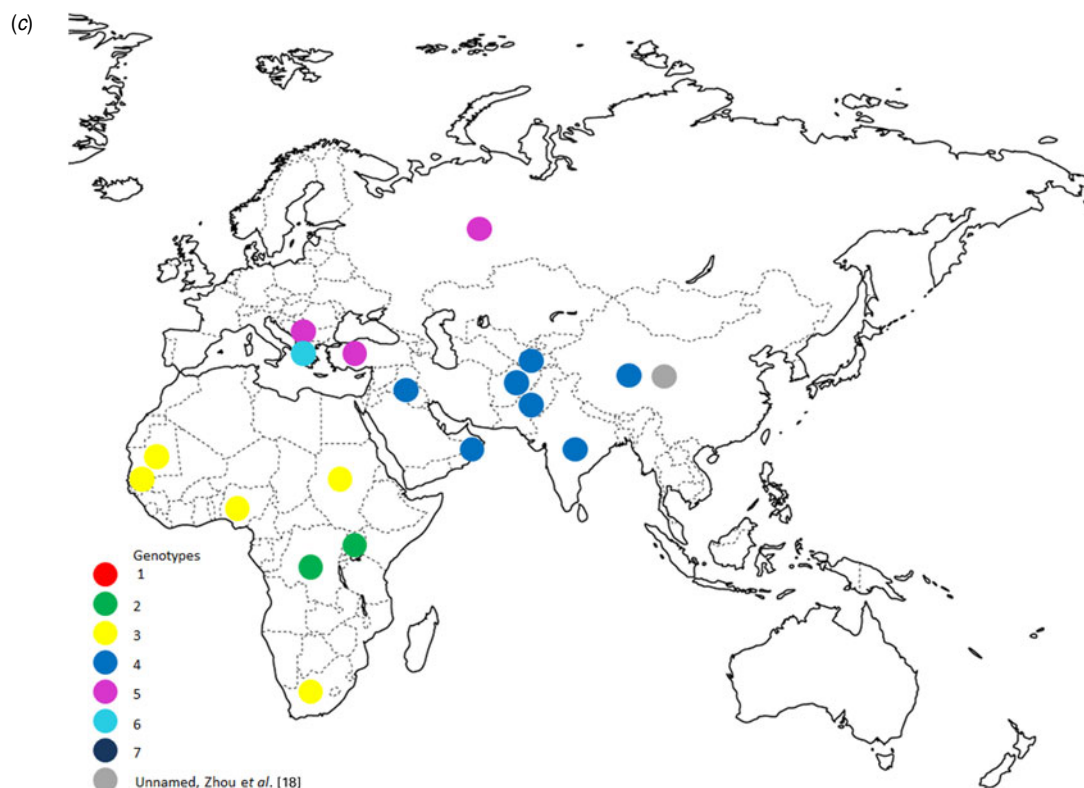


Fig. 2 (c). For legend see previous page.

complete sequence data for 13 geographically distinct isolates [8]. NGS methods present a relatively simple alternative, requiring only appropriate primers for the generation of amplicons by RT-PCR and potentially a limited number of primers to verify isolated regions or bases. This process could be simplified further by performing NGS of CCHFV directly from clinical samples without prior amplification. This would not only negate the need for specific primers and therefore allow sequencing of diverse CCHFV isolates but would also remove bias introduced by PCR errors [30] as well as multiple passages sometimes required to generate adequate viral titres. This method has been successfully used to sequence RNA viruses such as lyssaviruses directly from tissue samples and cell culture lysates [31]. The introduction of errors during PCR can largely be overcome by using high fidelity enzymes. Although the assembly of raw data generated by NGS platforms can be complex, the workflow described in the present study made use almost exclusively of open source software and could be applied to a range of datasets from various sources. Confirmation of the validity of the assemblies obtained was made possible by comparison with a number of partial sequences which were available

from a previous study [27], confirming the accuracy of the methods used.

Genetic evolution of arboviruses, including CCHFV, is a complex process influenced by multiple factors. As with other RNA viruses, the RNA-dependent RNA polymerases lack proofreading activity and show error frequencies of $\sim 10^{-4}$ [32]. This is offset by the effect of alternating infections of arthropods and vertebrates which constrains virus adaptation of arboviruses in comparison to other RNA viruses [33]. Despite this, CCHFV shows a high level of genetic variability. Inclusion of diverse isolates in the current study, particularly from China and central and West Africa, led to nucleotide variation of up to 7%, 27% and 11% for the S, M and L segments, respectively, and amino acid variation of up to 8%, 30% and 10%. This is similar to previous studies with greater variability of the M and L segments compared to the S segment, which is contrary to expectation as the viral RNA polymerase is usually highly conserved [8, 34].

Tree topologies for southern African isolates based on partial S, M and L sequence data correlated with topologies constructed using complete genome data (data available from corresponding author on

request), provide further evidence that incongruencies in grouping are likely reassortant events and not recombination [27]. Genetic relationships indicate movement of CCHFV isolates within and between continents. Isolates from the same group can be found in geographically distinct locations and isolates from different groups can be found to co-circulate in similar regions. Genetic diversity within regions supports movement of the virus by bird migration and livestock trade. Reassortment events have occurred between West African and southern African isolates and between southern African and Asian isolates. Interestingly, the reassortment events between West Africa and southern Africa involved the L segment, whereas reassortment events between southern Africa and Asia involved the M segment. This may be related to the tick species found in these geographical areas. The mechanism of reassortment is unclear; however, it is assumed to occur in the tick host where dual persistent infections are more likely. Few complete sequences are available from West Africa but as three of the four available isolates show either L or M segment reassortment, it appears that reassortant viruses may occur frequently in this area. The two novel Chinese isolates described by Zhou *et al.* cluster independently in an as yet unnamed group, suggesting the occurrence of further genetic groups [18]. Additional complete genome sequence data from geographically distinct isolates are required to corroborate and expand these findings.

Reassortment is widely described for members of the Bunyaviridae family [35]. Both homologous and heterologous reassortment may occur and may result in altered viral phenotypes. Ngari virus is a reassortant of Bunyamwera and Batai viruses and is associated with haemorrhagic fever, in contrast to the mild disease associated with the parent viruses [36]. Both increased neuroinvasiveness and enhanced transmission by insect vectors have been specifically linked to the M segment of La Crosse virus, another member of the Bunyaviridae family [37, 38]. Burt *et al.* suggested a possible association between M segment reassortment of CCHFV and pathogenicity although this finding was not statistically significant [27]. The current study could not confirm this association but given the small number of isolates included, further investigation is warranted. As 10 of the 15 complete sequences of CCHFV from southern Africa show M segment reassortment, it would be of interest to determine whether these M segments provide a competitive advantage such as the increased

transmissibility demonstrated in La Crosse virus reassortants. The postulated co-evolution of S and L segments, possibly related to functional interdependency [7], may account for the relative scarcity of reassortment of these segments.

It is likely that both genetic drift and shift of CCHFV genomes occur chiefly during infection of insect vectors rather than in mammalian hosts due to the longer period of infection and increased likelihood of super-infection with more than one strain [11, 39]. In contrast to the short period of viraemia seen in humans [40], ticks remain infected over much longer periods with transstadial and, more rarely, transovarial transmission playing an important role in virus perpetuation and therefore also virus evolution [39, 41, 42].

In conclusion, NGS methods provide a simple and effective alternative to Sanger sequencing for the determination of complete CCHFV genomes. The application of these methods to clinical samples without prior PCR amplification would provide further advantages and should be validated on CCHFV isolates. Our genetic analysis of complete genome sequences confirms the high frequency of M segment reassortment in southern African CCHFV isolates. Further studies are indicated to elucidate the possible consequences of reassortment particularly relating to viral pathogenicity and transmissibility.

DECLARATION OF INTEREST

None.

ACKNOWLEDGEMENTS

We thank Anelda van der Walt, Central Analytical Facility, Stellenbosch University for advice regarding NGS analysis. This project was funded by the National Health Laboratory Service Research Trust, the Polio Research Foundation, South Africa, and University of the Free State Cluster funding.

REFERENCES

1. Elliott RM, *et al.* Family *Bunyaviridae*. In Regenmortel MHV, Fauquet CM, Bishop DHL, Carstens EB, Estes MK, Lemon SM, Maniloff J, Mayo MA, McGeoch DJ, Pringle CR and Wickner RB, eds. *Seventh Report of the International Committee on Taxonomy of Viruses*. San Diego: Academic Press, 2000, pp. 599–621.
2. Schmaljohn CS, Hooper JW. *Bunyaviridae*: the viruses and their replication. In: Knipe DM and Howley PM,

- eds. *Fields Virology, Volume 1*, 4th edn. Philadelphia: Lippincott, Williams & Wilkins, 2001, pp. 1581–1602.
3. **Whitehouse CA**. Crimean-Congo hemorrhagic fever. *Antiviral Research* 2004; **64**: 145–160.
 4. **Bente DA, et al.** Crimean-Congo hemorrhagic fever: History, epidemiology, pathogenesis, clinical syndrome and genetic diversity. *Antiviral Research* 2013; **100**: 159–189.
 5. **Msimang V, et al.** Update: Crimean-Congo haemorrhagic fever in South Africa. *Communicable Diseases Surveillance Bulletin* 2013; **11**: 62–65.
 6. **Aradaib IE, et al.** Multiple Crimean-Congo hemorrhagic fever virus strains are associated with disease outbreaks in Sudan, 2008–2009. *PLoS Neglected Tropical Diseases* 2011; **5**: e1159.
 7. **Chamberlain J, et al.** Co-evolutionary patterns of variation in small and large RNA segments of Crimean-Congo hemorrhagic fever virus. *Journal of General Virology* 2005; **86**: 3337–3341.
 8. **Deyde VM, et al.** Crimean-Congo hemorrhagic fever virus genomics and global diversity. *Journal of Virology* 2006; **80**: 8834–8842.
 9. **Duh D, et al.** The complete genome sequence of a Crimean-Congo hemorrhagic fever virus isolated from an endemic region in Kosovo. *Virology Journal* 2008; **5**: 7.
 10. **Hewson R, et al.** Evidence of segment reassortment in Crimean-Congo haemorrhagic fever virus. *Journal of General Virology* 2004; **85**: 3059–3070.
 11. **Kinsella E, et al.** Sequence determination of the Crimean-Congo hemorrhagic fever virus L segment. *Virology* 2004; **321**: 23–28.
 12. **Meissner JD, et al.** The complete genome sequence of strain ROS/HUUVL-100, a representative Russian Crimean Congo hemorrhagic fever virus strain. *Virus Genes* 2006; **33**: 87–93.
 13. **Olschläger S, et al.** Complete sequence and phylogenetic characterisation of Crimean-Congo hemorrhagic fever virus from Afghanistan. *Journal of Clinical Virology* 2011; **50**: 90–92.
 14. **Ozdarendeli A, et al.** The complete genome analysis of Crimean-Congo hemorrhagic fever virus isolated in Turkey. *Virus Research* 2010; **147**: 288–293.
 15. **Papa A, et al.** Genetic characterization of the M RNA segment of Crimean Congo hemorrhagic fever virus strains, China. *Emerging Infectious Diseases* 2002; **8**: 50–53.
 16. **Seregin SV, et al.** Genetic characterization of the M RNA segment of Crimean-Congo hemorrhagic fever virus strains isolated in Russia and Tajikistan. *Virus Genes* 2004; **28**: 187–193.
 17. **Yadav PD, et al.** Genetic characterization and molecular clock analyses of the Crimean-Congo hemorrhagic fever virus from human and ticks in India, 2010–2011. *Infection, Genetics and Evolution* 2013; **14**: 223–231.
 18. **Zhou A, et al.** Complete genome sequences of two Crimean-Congo hemorrhagic fever viruses isolated in China. *Genome Announcements* 2013; **1**: e00571-13.
 19. **Capobianchi MR, Giombini E, Rozera G.** Next-generation sequencing technology in clinical virology. *Clinical Microbiology and Infection* 2013; **19**: 15–22.
 20. **Burt FJ, Swanepoel R.** Molecular epidemiology of African and Asian Crimean-Congo haemorrhagic fever isolates. *Epidemiology and Infection* 2005; **133**: 659–666.
 21. **Schmieder R, Edwards R.** Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 2011; **27**: 863–864.
 22. **Chevreur B, Wetter T, Suhai S.** Genome sequence assembly using trace signals and additional sequence information. In *Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB)*, 1999, pp. 45–56.
 23. **Bonfield JK, Smith KF, Staden R.** A new DNA sequence assembly program. *Nucleic Acids Research* 1995; **23**: 4992–4999.
 24. **Larkin MA, et al.** Clustal W and Clustal X version 2.0. *Bioinformatics* 2007; **23**: 2947–2948.
 25. **Tamura K, et al.** MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* 2011; **28**: 2731–2739.
 26. **Carroll SA, et al.** Ancient common ancestry of Crimean-Congo haemorrhagic fever virus. *Molecular Phylogenetics and Evolution* 2010; **55**: 1103–1110.
 27. **Burt FJ, et al.** Genetic relationship in southern African Crimean-Congo haemorrhagic fever isolates: evidence for occurrence of reassortment. *Epidemiology and Infection* 2009; **137**: 1302–1308.
 28. **Grard G, et al.** Re-emergence of Crimean-Congo hemorrhagic fever virus in Central Africa. *PLoS Neglected Tropical Diseases* 2011; **5**: e1350.
 29. **Lukashev AN.** Evidence for recombination in Crimean-Congo hemorrhagic fever virus. *Journal of General Virology* 2005; **86**: 2333–2338.
 30. **Bracho MA, Moya A, Barrio E.** Contribution of *Taq* polymerase-induced errors to the estimation of RNA virus diversity. *Journal of General Virology* 1998; **79**: 2921–2928.
 31. **Marston DA, et al.** Next generation sequencing of viral RNA genomes. *BMC Genomics* 2013; **14**: 444.
 32. **Steinhauer DA, Domingo E, Holland JJ.** Lack of evidence for proofreading mechanisms associated with an RNA virus polymerase. *Gene* 1992; **122**: 281–288.
 33. **Coffey LL, et al.** Arbovirus evolution *in vivo* is constrained by host adaptation. *Proceedings of the National Academy of Sciences USA* 2008; **105**: 6970–6975.
 34. **Anagnostou V, Papa A.** Evolution of Crimean-Congo hemorrhagic fever virus. *Infection, Genetics and Evolution* 2009; **9**: 948–954.
 35. **Briese T, Calisher CH, Higgs S.** Viruses of the family Bunyaviridae: are all available isolates reassortants? *Virology* 2013; **446**: 207–216.
 36. **Bowen MD, et al.** A reassortant bunyavirus isolated from acute hemorrhagic fever cases in Kenya and Somalia. *Virology* 2001; **291**: 185–190.
 37. **Beatty BJ, et al.** Molecular basis of bunyavirus transmission by mosquitoes: role of the middle-sized RNA segment. *Science* 1981; **211**: 1433–1435.

38. **Gonzalez-Scarano F, et al.** Genetics infectivity and virulence of California serogroup viruses. *Virus Research* 1992; **24**: 123–135.
39. **Hoogstraal H.** The epidemiology of tick-borne Crimean-Congo haemorrhagic fever in Asia, Europe, and Africa. *Journal of Medical Entomology* 1979; **15**: 307–417.
40. **Shepherd AJ, et al.** Comparison of methods for isolation and titration of Crimean-Congo hemorrhagic fever virus. *Journal of Clinical Microbiology* 1986; **24**: 654–656.
41. **Gonzalez JP, et al.** Sexual and transovarian transmission of Crimean-Congo haemorrhagic fever virus in *Hyalomma truncatum* ticks. *Research in Virology* 1992; **143**: 23–28.
42. **Gordon SW, Linthicum KJ, Moulton JR.** Transmission of Crimean-Congo hemorrhagic fever virus in two species of *Hyalomma* ticks from infected adults to cofeeding immature forms. *American Journal of Tropical Medicine and Hygiene* 1993; **48**: 576–580.