VIROLOGICA SINICA, August 2011, 26 (4): 229-244
DOI 10.1007/s12250-011-3211-z

Epidemiology and Mutational Analysis of Global Strains of Crimean-Congo Haemorrhagic Fever Virus

Na Han and Simon Rayner**

(Bioinformatics Group, State Key Laboratory for Virology, Wuhan Institute of Virology, Chinese Academy of Sciences, Wuhan, Hubei 430071, China)

Abstract: Crimean-Congo hemorrhagic fever (CCHF) is a severe illness with high fatality. Cases are reported in several countries in Africa, Europe, the Middle East, and Asia. Phylogenetic analyses based on the virus S (nucleocapsid), M (glycoprotein), and L (polymerase) genome segments sequences indicate distinct geographic lineages exist but their specific genetic characteristics require elucidation. In this work we collected all full length S segment sequences and generated a phylogenetic tree based on the alignment of these 62 samples. We then analyzed the alignment using entries from AAIndex, the Amino Acid Index database, to identify amino acid mutations that performed significant changes in charge, pka, hydropathy and side chain volume. Finally, we mapped these changes back to the tree and alignment to identify correlated mutations or sites that characterized a specific lineage. Based on this analysis we are able to propose a number of sites that appear to be important for virus function and which would be good candidates for experimental mutational analysis studies.

Key words: Crimean-Congo hemorrhagic fever virus (CCHFV); Epidemiology; Mutational Analysis

INTRODUCTION

Crimean-Congo haemorrhagic fever (CCHF) is a viral haemorrhagic fever that was first described in Crimea in 1944 and subsequently associated with similar outbreaks that occurred in the Congo. Although CCHF has high mortality in humans, the disease primarily occurs in animals. CCHF is endemic in many countries in Africa, Europe and Asia[5, 6, 15, 21, 33, 37] with outbreaks recently reported in Sudan[3, 4], Kosovo[6, 33], China[12, 28, 36], Russia[19, 21, 37] and India[29]. CCHF is caused by the Crimean-Congo haemorrhagic fever virus (CCHFV), a segmented negative strand RNA virus that is a member of the genus *Nairovirus* of the family *Bunyaviridae*[16, 23]. Bunyaviruses consist of three segments:small (S), medium (M) and large (L) that encode the viral nucleocapsid protein (N), the glycoprotein precursor (GPC) and the polymerase protein (P) respectively[2].

There have been multiple reports that have investigated the epidemiology and phylogeny of the virus, but these have generally concentrated on the phylogenetic relationships amongst sequences from a single segment or have studied all three segments but with a limited number of sequences or concentrated on a specific region. In this work we collected all

Received: 2011-07-04, Accepted: 2011-07-12

** Corresponding author.
Phone: +86-27-87199895,E-mail: s.raynere@wh.iov.cn

publicly available full length S segments for CCHFV and estimated phylogenetic trees based on the nucleotide alignment. Our trees predicted several major clades that were consistent with previous findings and which supported geographical subdivision. We then generated amino acid alignments and selected multiple entries from AAIndex, the Amino Acid Index database[17, 26] and used these to identify mutations that represented major changes in the physical properties of the consensus amino acid at each site. Most of the clades showed few major amino acid replacements with the exception of the Asia 2 clade which showed large numbers of changes associated with charge, volume, salvation energy and hydrophobicity.

MATERIALS AND METHODS

All *Nairovirus* sequences were downloaded from Genbank. 62 Full length S segment sequences were selected and aligned using ClustalX v2.0[32] and gaps were removed to give a final alignment of 1 461 nt. Trees were estimated with the MEGA5 software package[30] using the Neighbourhood Joining method with the Maximum Likelihood Composite method (Tamura-Nei distance matrix) and uniform rates among sites. Sequence accession numbers and background information are listed in Table 1.

To identify sites containing mutations that reflected amino acid changes with significantly different properties, eight specific entries were selected from the Amino Acid Index (AAI) Database[17] that reflected changes in charge, volume, pKa and hydrophobicity. The accession numbers of the selected entries were FAUJ880112-Negative Charge[11], FAUJ880113-Positive Charge[11], FAUJ880114-pK-a value[11], GOLD730102-Residue volume[14], TSAJ990101-Packing Density[34], KRIW790103-Side chain volume[18], EISD840101-Consensus normalized hydrophobicity scale[8], ROSM880105-Hydropathies of amino acid side chains[27]. Each alignment was translated to amino acid and analyzed in turn with each AAI entry. Each sequence was inspected in turn and compared to the consensus for the entire set of CCHFV sequences. For charge entries, any mutation that produced a change in charge from neutral, positive or negative was considered significant. For other entries, the change in a parameter brought about by a mutation at a site was considered significant if

$$\Delta^{I}_{ab} \geqslant \frac{I_{\max} - I_{\min}}{4}$$

where Δ^{I}_{ab} is the change in amino acid index I when amino acid a mutates to amino acid b.

The alignment was analyzed using a custom java program available from the authors on request.

RESULTS

Phylogenetic Analysis

The predicted tree for the S, segment is shown in Fig. 1. Each of the trees exhibit clear geographic subdivision, identifying seven clades that were named Asia 1, Asia 2, Europe 1, Europe 2, Africa 1, Africa 2 and Africa 3 and which are consistent with results from previous studies[2, 4, 5, 13, 15, 16, 22, 28, 29].

Mutation Analysis

We next used entries from the AminoAcidIndex database[17] to analyze the alignment and investigate whether there were specific mutations that were more probable, or regions where mutations were more likely to occur. We investigated mutations that produced changes in charge, hydrophobicity and volume and mapped these mutations to the Asia 1, Asia 2, Europe 1, Europe 2, Europe 3, Africa 1, Africa 2 and Africa 3 clades identified in the previous section. The mutations are listed in Tables 2, 3 and 4. The results for Residue volume and Packing Density were identical so only Residue Volume is shown.

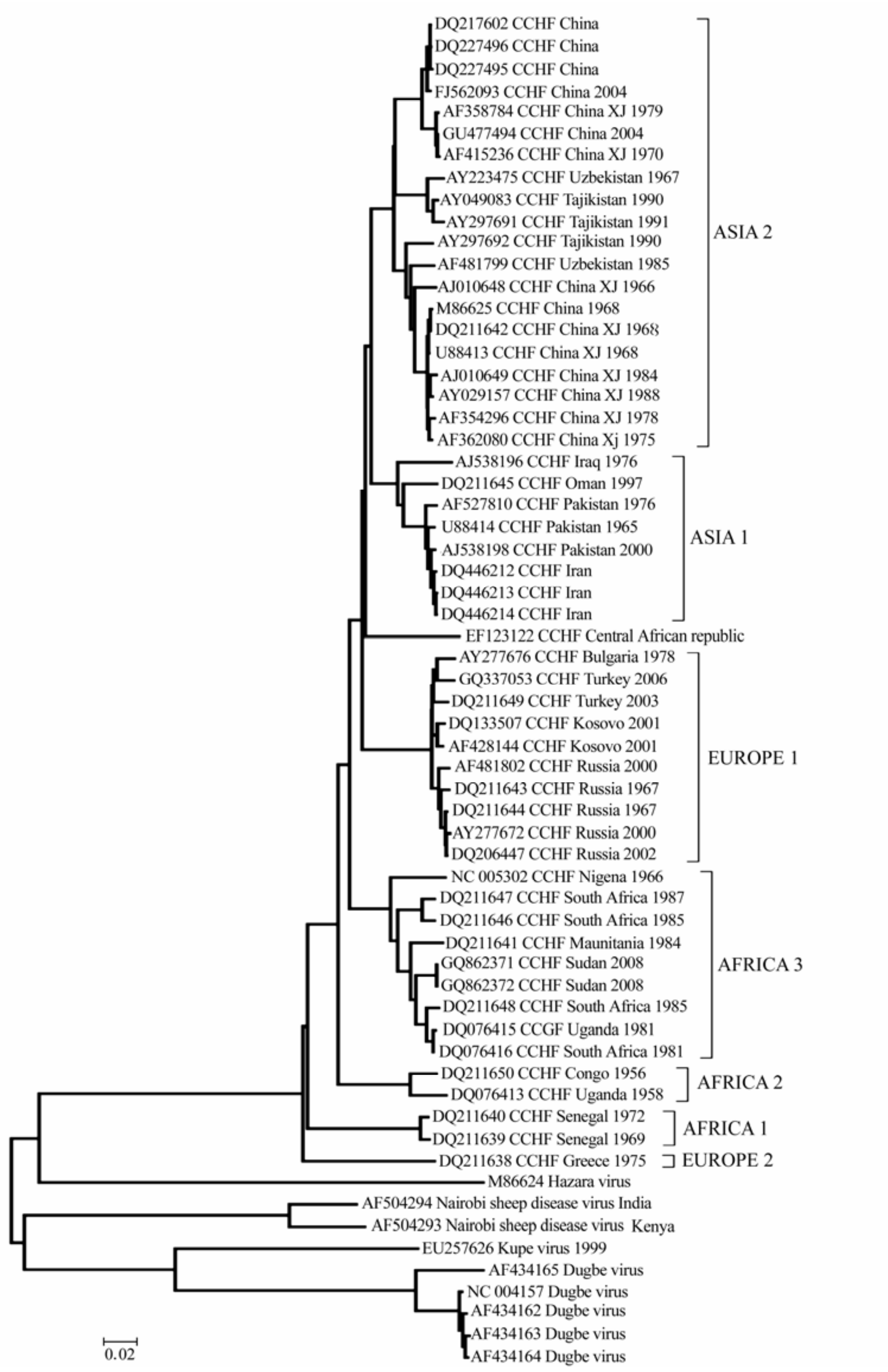
DQ217602 CCHF China
DQ227496 CCHF China
DQ227495 CCHF China
FJ562093 CCHF China 2004
AF358784 CCHF China XJ 1979
GU477494 CCHF China 2004
AF415236 CCHF China XJ 1970
AY223475 CCHF Uzbekistan 1967
AY049083 CCHF Tajikistan 1990
AY297691 CCHF Tajikistan 1991
AY297692 CCHF Tajikistan 1990
AF481799 CCHF Uzbekistan 1985
AJ010648 CCHF China XJ 1966
M86625 CCHF China 1968
DQ211642 CCHF China XJ 1968
U88413 CCHF China XJ 1968
AJ010649 CCHF China XJ 1984
AY029157 CCHF China XJ 1988
AF354296 CCHF China XJ 1978
AF362080 CCHF China Xj 1975
ASIA 2
AJ538196 CCHF Iraq 1976
DQ211645 CCHF Oman 1997
AF527810 CCHF Pakistan 1976
U88414 CCHF Pakistan 1965
AJ538198 CCHF Pakistan 2000
DQ446212 CCHF Iran
DQ446213 CCHF Iran
DQ446214 CCHF Iran
ASIA 1
EF123122 CCHF Central African republic
AY277676 CCHF Bulgaria 1978
GQ337053 CCHF Turkey 2006
DQ211649 CCHF Turkey 2003
DQ133507 CCHF Kosovo 2001
AF428144 CCHF Kosovo 2001
AF481802 CCHF Russia 2000
DQ211643 CCHF Russia 1967
DQ211644 CCHF Russia 1967
AY277672 CCHF Russia 2000
DQ206447 CCHF Russia 2002
EUROPE 1
NC 005302 CCHF Nigena 1966
DQ211647 CCHF South Africa 1987
DQ211646 CCHF South Africa 1985
DQ211641 CCHF Maunitania 1984
GQ862371 CCHF Sudan 2008
GQ862372 CCHF Sudan 2008
DQ211648 CCHF South Africa 1985
DQ076415 CCGF Uganda 1981
DQ076416 CCHF South Africa 1981
AFRICA 3
DQ211650 CCHF Congo 1956
DQ076413 CCHF Uganda 1958
AFRICA 2
DQ211640 CCHF Senegal 1972
DQ211639 CCHF Senegal 1969
AFRICA 1
DQ211638 CCHF Greece 1975
EUROPE 2
M86624 Hazara virus
AF504294 Nairobi sheep disease virus India
AF504293 Nairobi sheep disease virus Kenya
EU257626 Kupe virus 1999
AF434165 Dugbe virus
NC 004157 Dugbe virus
AF434162 Dugbe virus
AF434163 Dugbe virus
AF434164 Dugbe virus
0.02

Fig. 1. NJ tree generated from all full length S segment sequences available in GenBank. Consistent with previous studies, the tree shows seven main clades that exhibit geographic subdivision.

Table 1. Background information of sequences used in phylogenetic analysis and mutation analysis.

Accession	Virus	Country	Date
AF362080	Crimean-Congo_Hemorrhagic_Fever_virus_strain_China	China: Xinjiang	1975
AF354296	Crimean-Congo_Hemorrhagic_Fever_virus_strain_China	China: XinJiang	1978
AY029157	Crimean-Congo_Hemorrhagic_Fever_virus_strain_China	China: Xinjiang	1988
AJ010649	Crimean-Congo_hemorrhagic_fever_virus	China: Xinjiang	1984
U88413	Crimean-Congo_hemorrhagic_fever_virus	China: Xinjiang	1968
DQ211642	Crimean-Congo_hemorrhagic_fever_virus	China: Xinjiang	1968
M86625	Crimean-Congo_hemorrhagic_fever_virus	China	1968
AJ010648	Crimean-Congo_hemorrhagic_fever_virus	China: Xinjiang	1966
AF481799	Crimean-Congo_hemorrhagic_fever_virus	Russia: Uzbekistan	1985
AY297692	Crimean-Congo_hemorrhagic_fever_virus	Tajikistan	1990
GU477494	Crimean-Congo_hemorrhagic_fever_virus	China	2004
AF415236	Crimean-Congo_hemorrhagic_fever_virus	China: Xinjiang	1970
AF358784	Crimean-Congo_Hemorrhagic_Fever_virus_strain_China	China: Xinjiang	1979
FJ562093	Crimean-Congo_hemorrhagic_fever_virus	China	2004
DQ227495	Crimean-Congo_hemorrhagic_fever_virus	China	
DQ227496	Crimean-Congo_hemorrhagic_fever_virus	China	
DQ217602	Crimean-Congo_hemorrhagic_fever_virus	China	
AY223475	Crimean-Congo_hemorrhagic_fever_virus	Russia: Uzbekistan	1967
AY297691	Crimean-Congo_hemorrhagic_fever_virus	Tajikistan	1991
AY049083	Crimean-Congo_hemorrhagic_fever_virus	Tajikistan	1990
AJ538196	Crimean-Congo_hemorrhagic_fever_virus	Iraq: Baghdad	1976
DQ211645	Crimean-Congo_hemorrhagic_fever_virus	Oman	1997
AF527810	Crimean-Congo_hemorrhagic_fever_virus	Pakistan	1976
U88414	Crimean-Congo_hemorrhagic_fever_virus	Pakistan	1965
AJ538198	Crimean-Congo_hemorrhagic_fever_virus	Pakistan: Karachi	2000
DQ446212	Crimean-Congo_hemorrhagic_fever_virus	Iran	
DQ446214	Crimean-Congo_hemorrhagic_fever_virus	Iran	
DQ446213	Crimean-Congo_hemorrhagic_fever_virus	Iran	
EF123122	Crimean-Congo_hemorrhagic_fever_virus	Central African Republic	
GQ337053	Crimean-Congo_hemorrhagic_fever_virus	Turkey	2006
AY277676	Crimean-Congo_hemorrhagic_fever_virus	Bulgaria	1978
DQ211649	Crimean-Congo_hemorrhagic_fever_virus	Turkey	2003
AF428144	Crimean-Congo_hemorrhagic_fever_virus	Kosovo:	2001
DQ133507	Crimean-Congo_hemorrhagic_fever_virus	Kosovo	2001
AF481802	Crimean-Congo_hemorrhagic_fever_virus	Russia: Stavropol	2000
DQ211643	Crimean-Congo_hemorrhagic_fever_virus	Russia: Astrakhan	1967
DQ211644	Crimean-Congo_hemorrhagic_fever_virus	Russia	1967
DQ206447	Crimean-Congo_hemorrhagic_fever_virus	Russia: Rostov region	2002
AY277672	Crimean-Congo_hemorrhagic_fever_virus	Russia: Rostov region	2000
NC_005302	Crimean-Congo_hemorrhagic_fever_virus	Nigeria: Sokoto	1966
DQ211647	Crimean-Congo_hemorrhagic_fever_virus	South Africa	1987
DQ211646	Crimean-Congo_hemorrhagic_fever_virus	South Africa	1985
DQ211641	Crimean-Congo_hemorrhagic_fever_virus	Mauritania	1984
GQ862372	Crimean-Congo_hemorrhagic_fever_virus	Sudan: Al-Fulah City	2008

Table 1. continue

Accession	Virus	Country	Date
GQ862371	Crimean-Congo_hemorrhagic_fever_virus	Sudan: Al-Fulah City	2008
DQ211648	Crimean-Congo_hemorrhagic_fever_virus	South Africa	1985
DQ076416	Crimean-Congo_hemorrhagic_fever_virus	South Africa	1981
DQ076415	Crimean-Congo_hemorrhagic_fever_virus	Uganda	1981
DQ076413	Crimean-Congo_hemorrhagic_fever_virus	Uganda	1958
DQ211650	Crimean-Congo_hemorrhagic_fever_virus	Republic of Congo	1956
DQ211640	Crimean-Congo_hemorrhagic_fever_virus	Senegal	1972
DQ211639	Crimean-Congo_hemorrhagic_fever_virus	Senegal	1969
DQ211638	Crimean-Congo_hemorrhagic_fever_virus	Greece	1975
M86624	Hazara_virus	Pakistan	
AF504294	Nairobi_sheep_disease_virus	India	
AF504293	Nairobi_sheep_disease_virus	Kenya	
EU257626	Kupe_virus	Kenya	1999
AF434165	Dugbe_virus		
AF434164	Dugbe_virus		
AF434163	Dugbe_virus		
AF434162	Dugbe_virus		
NC_004157	Dugbe_virus		

Table 2. Amino acid mutations leading to charge change in the alignment as classified by clades defined in Fig. 1.

Positive Charge				**Negative Charge**			
ASIA 2							
site	**sequence**			**site**	**sequence**		
185	AF354296_China_XJ_1978	S	R	127	DQ217602_China	D	N
212	AF362080_China_XJ_1975	T	K	127	AF358784_China_XJ_1979	D	N
302	AJ010649_China_XJ_1984	G	R	127	DQ227495_China	D	N
323	AY297691_Tajikistan_1991	R	V	127	GU477494_China_2004	D	N
				127	AF481799_Uzbekistan_1985	D	N
				127	FJ562093_China_2004	D	N
				127	AF415236_China_XJ_1970	D	N
				127	DQ227496_China	D	N
				246	AF354296_China_XJ_1978	G	E
				266	DQ217602_China	N	D
				266	AF358784_China_XJ_1979	N	D
				266	DQ227495_China	N	D
				266	GU477494_China_2004	N	D
				266	FJ562093_China_2004	N	D
				266	AF415236_China_XJ_1970	N	D
				266	DQ227496_China	N	D
				403	AY049083_Tajikistan_1990	D	N
				403	AJ010648_China_XJ_1966	D	N
				403	AY223475_Uzbekistan_1967	D	N

Table 2. continue

Positive Charge				Negative Charge			
ASIA 1							
site	**sequence**			**site**	**sequence**		
212	U88414_Pakistan_1965	M	K	449	AJ538196_Iraq_1976	H	D
341	AJ538196_Iraq_1976	H	Q				
449	AJ538196_Iraq_1976	H	D				
EUROPE 1							
site	**sequence**			**site**	**sequence**		
340	GQ337053_Turkey_2006	E	K	23	DQ211643_Russia_1967	E	G
442	GQ337053_Turkey_2006	E	K	23	DQ211644_Russia_1967	E	G
262	**DQ211638_Greece_1975**	**K**	**N**	99	GQ337053_Turkey_2006	D	N
448	**DQ211638_Greece_1975**	**K**	**Q**	340	GQ337053_Turkey_2006	E	K
				442	GQ337053_Turkey_2006	E	K
				263	**DQ211638_Greece_1975**	**D**	**G**
				446	**DQ211638_Greece_1975**	**D**	**N**
AFRICA 1							
Site	**sequence**			**site**	**sequence**		
11	DQ076415_Uganda_1981	H	D	11	DQ076415_Uganda_1981	H	D
262	**DQ211640_Senegal_1972**	**K**	**N**	**118**	**DQ211650_Congo_1956**	**G**	**D**
262	**DQ211639_Senegal_1969**	**K**	**N**	**118**	**DQ076413_Uganda_1958**	**G**	**D**
				260	**DQ211650_Congo_1956**	**D**	**A**
				263	**DQ211640_Senegal_1972**	**E**	**G**
				263	**DQ211639_Senegal_1969**	**E**	**G**
				374	**DQ211650_Congo_1956**	**E**	**Q**
				374	**DQ076413_Uganda_1958**	**E**	**Q**
				403	**DQ211639_Senegal_1969**	**D**	**N**

The most notable result is that in every category the Asia 2 clade appears to contain many more mutations than any of the other clades. Although this clade contains twice as many sequences as the other clades, this still doesn't appear to account for many of the observed differences. For negative charge mutations the numbers of changes were (*Africa 3*: 9 sequences / 3 mutations, *Asia 1*: 8 sequences / 1 mutation, *Asia 2*: 20 sequences / 19 mutations, *Europe 1*: 10 sequences / 1 mutation). Similarly for the *pka* index (*Africa 3*: 9 seqs / 2 muts, *Asia 1*: 8 seqs / 2 muts, *Asia 2*: 20 seqs / 22 muts, *Europe 1*: 10 seqs / 11 muts); *hydrophobicity* index (*Africa 3*: 9 seqs / 3 muts, *Asia 1*: 8 seqs / 1 muts, *Asia 2*: 20 seqs / 32 muts, *Europe 1*: 10 seqs / 4 muts); *volume* index (*Africa 3*: 9 seqs / 4 muts, *Asia 1*: 8 seqs / 2 muts, *Asia 2*: 20 seqs / 13 muts, *Europe 1*: 10 seqs / 4 muts). Even when the Europe 1 and Europe 2 clades and the Africa 1, Africa 2 & Africa 3 clades were consolidated into single European and African clades respectively, they still contained fewer mutations despite their greater genetic diversity (these additional mutations are highlighted in grey in Tables 2, 3 and 4).

Table 3. Amino acid mutations leading to significant changes in pKa and hydropathy values in the alignment as classified by clades defined in Fig. 1. The shaded mutations correspond to mutations that occurred outside the main European (Europe 1) and African (Africa 3) clades

PKA				**Hydrophobicity**			
ASIA 2							
site	**sequence**			**site**	**sequence**		
33	M86625_China_1968	P	S	113	AY297692_Tajikistan_1990	I	T
127	DQ217602_China	D	N	113	M86625_China_1968	I	T
127	AF358784_China_XJ_1979	D	N	113	U88413_China_XJ_1968	I	T
127	DQ227495_China	D	N	113	DQ211642_China_XJ_1968	I	T
127	GU477494_China_2004	D	N	127	DQ217602_China	D	N
127	AF481799_Uzbekistan_1985	D	N	127	AF358784_China_XJ_1979	D	N
127	FJ562093_China_2004	D	N	127	DQ227495_China	D	N
127	AF415236_China_XJ_1970	D	N	127	GU477494_China_2004	D	N
127	DQ227496_China	D	N	127	AF481799_Uzbekistan_1985	D	N
246	AF354296_China_XJ_1978	G	E	127	FJ562093_China_2004	D	N
266	DQ217602_China	N	D	127	AF415236_China_XJ_1970	D	N
266	AF358784_China_XJ_1979	N	D	127	DQ227496_China	D	N
266	DQ227495_China	N	D	156	AY223475_Uzbekistan_1967	S	A
266	GU477494_China_2004	N	D	185	AF354296_China_XJ_1978	S	R
266	FJ562093_China_2004	N	D	212	AF362080_China_XJ_1975	T	K
266	AF415236_China_XJ_1970	N	D	246	AF354296_China_XJ_1978	G	E
266	DQ227496_China	N	D	266	DQ217602_China	N	D
403	AY049083_Tajikistan_1990	D	N	266	AF358784_China_XJ_1979	N	D
403	AJ010648_China_XJ_1966	D	N	266	DQ227495_China	N	D
403	AY223475_Uzbekistan_1967	D	N	266	GU477494_China_2004	N	D
418	DQ227495_China	P	L	266	FJ562093_China_2004	N	D
431	AF415236_China_XJ_1970	P	A	266	AF415236_China_XJ_1970	N	D
				266	DQ227496_China	N	D
				281	AJ010648_China_XJ_1966	T	I
				281	AY223475_Uzbekistan_1967	T	I
				298	AJ010649_China_XJ_1984	L	S
				302	AJ010649_China_XJ_1984	G	R
				311	AY297691_Tajikistan_1991	S	A
				323	AY297691_Tajikistan_1991	R	V
				403	AY049083_Tajikistan_1990	D	N
				403	AJ010648_China_XJ_1966	D	N
				403	AY223475_Uzbekistan_1967	D	N
ASIA 1							
site	**sequence**			**site**	**sequence**		
341	AJ538196_Iraq_1976	H	Q	449	AJ538196_Iraq_1976	H	D
449	AJ538196_Iraq_1976	H	D				
EUROPE 1							
site	**sequence**			**site**	**sequence**		
23	DQ211643_Russia_1967	E	G	23	DQ211643_Russia_1967	E	G
23	DQ211644_Russia_1967	E	G	23	DQ211644_Russia_1967	E	G
99	GQ337053_Turkey_2006	D	N	99	GQ337053_Turkey_2006	D	N
167	DQ211638_Greece_1975	P	S	**126**	**DQ211638_Greece_1975**	**S**	**A**
192	AF481802_Russia_2000	L	P	199	AY277676_Bulgaria_1978	H	R

Table 3. continue

PKA				**Hydrophobicity**			
199	AY277676_Bulgaria_1978	H	R	199	DQ211638_Greece_1975	H	R
199	DQ211638_Greece_1975	H	R	**263**	**DQ211638_Greece_1975**	**D**	**G**
227	AY277672_Russia_2000	P	S	**446**	**DQ211638_Greece_1975**	**D**	**N**
263	DQ211638_Greece_1975	D	G				
285	GQ337053_Turkey_2006	P	L				
446	**DQ211638_Greece_1975**	**D**	**N**				
AFRICA 3							
site	**sequence**			**site**	**sequence**		
11	DQ076415_Uganda_1981	H	D	11	DQ076415_Uganda_1981	H	D
118	**DQ211650_Congo_1956**	**G**	**D**	**118**	**DQ211650_Congo_1956**	**G**	**D**
118	**DQ076413_Uganda_1958**	**G**	**D**	**118**	**DQ076413_Uganda_1958**	**G**	**D**
199	NC_005302_Nigeria_1966	H	R	127	NC_005302_Nigeria_1966	G	N
263	**DQ211640_Senegal_1972**	**E**	**G**	**156**	**DQ211640_Senegal_1972**	**S**	**A**
263	**DQ211639_Senegal_1969**	**E**	**G**	**156**	**DQ211639_Senegal_1969**	**S**	**A**
311	**DQ211650_Congo_1956**	**P**	**A**	**181**	**DQ211650_Congo_1956**	**L**	**Q**
311	**DQ211640_Senegal_1972**	**P**	**A**	**181**	**NC_005302_Nigeria_1966**	**L**	**Q**
311	**DQ211639_Senegal_1969**	**P**	**A**	**181**	**DQ211640_Senegal_1972**	**L**	**Q**
311	**DQ076413_Uganda_1958**	**P**	**A**	**181**	**DQ211639_Senegal_1969**	**L**	**Q**
403	**DQ211639_Senegal_1969**	**D**	**N**	**181**	**DQ076413_Uganda_1958**	**L**	**Q**
				199	NC_005302_Nigeria_1966	H	R
				260	**DQ211650_Congo_1956**	**D**	**A**
				263	**DQ211640_Senegal_1972**	**E**	**G**
				263	**DQ211639_Senegal_1969**	**E**	**G**
				374	**DQ211650_Congo_1956**	**E**	**Q**
				374	**DQ076413_Uganda_1958**	**E**	**Q**
				403	**DQ211639_Senegal_1969**	**D**	**N**

There is no solved structure for the N protein, so it is difficult to determine the significance of these changes, particularly for parameters such as changes in side chain volume. In order to try and identify mutations of possible interest, we next mapped all these changes on to a graphical representation of the alignment. These are shown in Fig. 2 (charge change) and Fig. 3 (pka, hydropathy and volume changes). Again, the greater number of mutations in the Asia 2 clade is clear, but additional features are also apparent. First of all, there is a pair of negative charge mutations that appear to be present in several of the Asia 2 sequences (D127N and N266D). Since the first mutation produces a charge change of +1 and the second produces a change of -1 these two sites may be compensatory. Secondly, the Africa 1 sequences contain two adjacent mutations K262N (charge change -1) and E263G (charge change +1). A similar pair mutation is also present in the Europe 2 sequence, K262N (charge change -1) and D263G (charge change +1) suggesting that these sites also play an important functional or structural role. The schematic for pka, hydropathy and volume changes is more difficult to interpret because these types of changes can occur without producing the same impact as charge changes but it is clear that the Asia 2 clade once again contains more mutations than the other clades. Another interesting feature occurs around AA263 where the Europe and African clades contain a number of mutations that produce significant changes in all three indices.

Table 4. Amino acid mutations leading to charge change in the alignment as classified by clades defined in Fig. 1. The shaded mutations correspond to mutations that occurred outside the main European (Europe 1) and African (Africa 3) clades.

Residual Volume				**Side Chain Volume**			
ASIA 2							
site	**sequence**			**site**	**sequence**		
113	AY297692_Tajikistan_1990	I	T	113	AY297692_Tajikistan_1990	I	T
113	M86625_China_1968	I	T	113	M86625_China_1968	I	T
113	U88413_China_XJ_1968	I	T	113	U88413_China_XJ_1968	I	T
113	DQ211642_China_XJ_1968	I	T	113	DQ211642_China_XJ_1968	I	T
185	AF354296_China_XJ_1978	S	R	185	AF354296_China_XJ_1978	S	R
212	AF362080_China_XJ_1975	T	K	212	AF362080_China_XJ_1975	T	K
246	AF354296_China_XJ_1978	G	E	246	AF354296_China_XJ_1978	G	E
281	AJ010648_China_XJ_1966	T	I	281	AJ010648_China_XJ_1966	T	I
281	AY223475_Uzbekistan_1967	T	I	281	AY223475_Uzbekistan_1967	T	I
298	AJ010649_China_XJ_1984	L	S	298	AJ010649_China_XJ_1984	L	S
302	AJ010649_China_XJ_1984	G	R	302	AJ010649_China_XJ_1984	G	R
323	AF354296_China_XJ_1978	A	V	323	AF354296_China_XJ_1978	A	V
418	DQ227495_China	P	L	418	DQ227495_China	P	L
ASIA 1							
site	**sequence**			**site**	**sequence**		
126	AJ538196_Iraq_1976	F	S	126	AJ538196_Iraq_1976	F	S
449	AJ538196_Iraq_1976	H	D	449	AJ538196_Iraq_1976	H	D
EUROPE 1							
site	**sequence**			**site**	**sequence**		
23	DQ211643_Russia_1967	E	G	23	DQ211643_Russia_1967	E	G
23	DQ211644_Russia_1967	E	G	23	DQ211644_Russia_1967	E	G
113	AY277676_Bulgaria_1978	V	A	113	AY277676_Bulgaria_1978	V	A
192	AF481802_Russia_2000	L	P	192	AF481802_Russia_2000	L	P
262	**DQ211638_Greece_1975**	**K**	**N**	**262**	**DQ211638_Greece_1975**	**K**	**N**
263	**DQ211638_Greece_1975**	**D**	**G**	**263**	**DQ211638_Greece_1975**	**D**	**G**
285	GQ337053_Turkey_2006	P	L	285	GQ337053_Turkey_2006	P	L
				340	GQ337053_Turkey_2006	E	K
				442	GQ337053_Turkey_2006	E	K
AFRICA 3							
site	**sequence**			**site**	**sequence**		
11	DQ076415_Uganda_1981	H	D	11	DQ076415_Uganda_1981	H	D
118	**DQ211650_Congo_1956**	**G**	**D**	**118**	**DQ211650_Congo_1956**	**G**	**D**
118	**DQ076413_Uganda_1958**	**G**	**D**	**118**	**DQ076413_Uganda_1958**	**G**	**D**
127	NC_005302_Nigeria_1966	G	N	127	NC_005302_Nigeria_1966	G	N
260	DQ076415_Uganda_1981	V	A	260	DQ076415_Uganda_1981	V	A
262	**DQ211640_Senegal_1972**	**K**	**N**	**262**	**DQ211640_Senegal_1972**	**K**	**N**
262	**DQ211639_Senegal_1969**	**K**	**N**	**262**	**DQ211639_Senegal_1969**	**K**	**N**
263	**DQ211640_Senegal_1972**	**E**	**G**	**263**	**DQ211640_Senegal_1972**	**E**	**G**
263	**DQ211639_Senegal_1969**	**E**	**G**	**263**	**DQ211639_Senegal_1969**	**E**	**G**
483	DQ211648_South_Africa_1985	V	A	483	DQ211648_South_Africa_1985	V	A

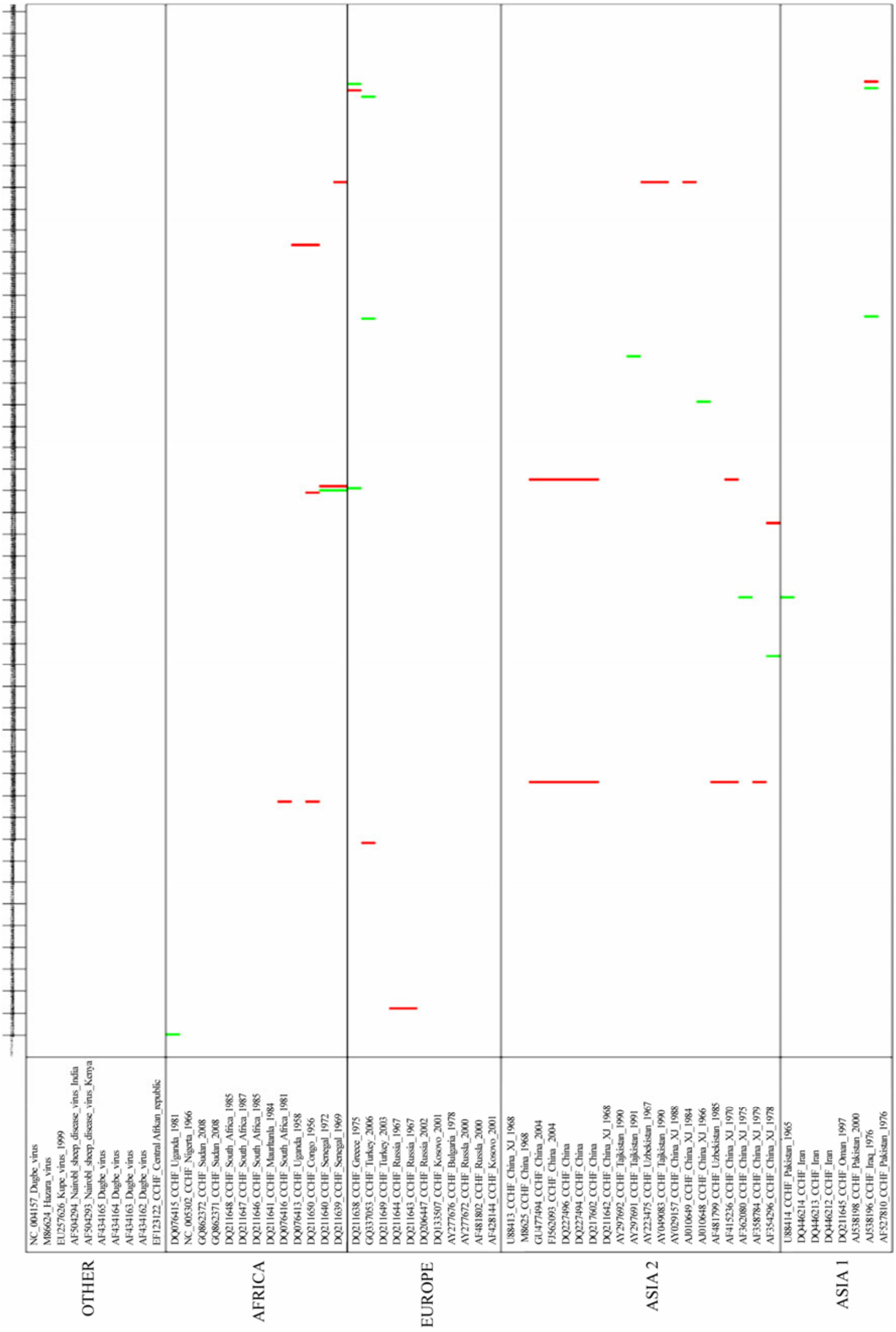

Fig. 2. Graphical representation of positive (green) and negative (red) changes in charge from the consensus sequence. A disproportionate amount of mutations that produce a negative change in charge occur in the Asia 2 clade, with multiple sequence containing the change at two sites (D127N and N266D) suggesting that these may represent compensatory mutations.

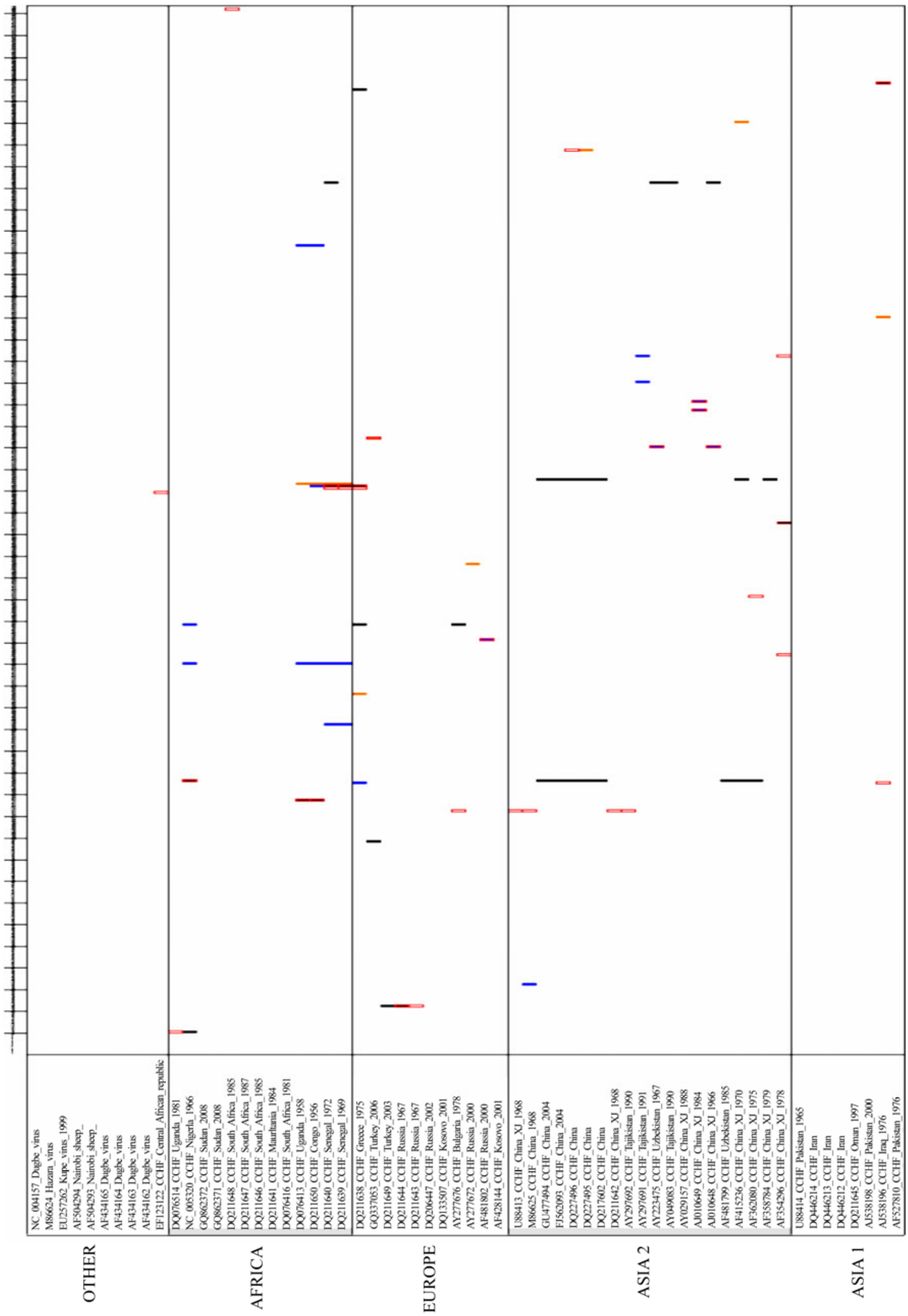

Fig. 3. Graphical representation of mutations that produced significant changes in pka (orange), hydropathy (blue) and side change volume (red boxes) from the consensus sequence. Sites which contain mutations that change both pka and hydropathy are shown as hatched yellow/blue with a red border if there was also a significant change in volume. Consistent with figure 2, a disproportionate amount of mutations occur in the Asia 2 clade, with multiple changes also mapped to the two sites 127 and 266.

Finally, we mapped all the identified changes on to the predicted tree. These are shown in Fig. 4 (charge change) and Fig. 5 (pka, hydropathy and volume changes). The plot identifies site mutations that occur in multiple sequences (vertical lines) and sites that contain mutations that have significant changes in multiple indices (horizontal lines). Sites which occur in multiple sequences and have changes in multiple indices are marked with both horizontal and vertical lines. In Fig. 4, only the Asia 2 clade contains multiple sequences with shared mutations. The compensatory positive/negative mutations in the Senegal sequences (Africa 1-DQ211639 and DQ211640)) are also apparent. In Fig. 4, the Uganda (DQ076413) and Congo (DQ211650) sequences in the Africa 2 clade have two mutations at two different sites that modify all three indices. Other mutations in this Africa 2 clade also modify the same sites in the Africa 1 sequences (Senegal DQ211639 and Senegal DQ211640). The box in the bottom right of the figure that spans the Africa 1, Africa 2 and Europe 2 clades delimit the cluster of mutations that are apparent in Fig. 3 around site AA262 which are shared across the clades and which modify three indices.

DISCUSSION

The S segment of viruses in the *Bunyaviridae* family encodes for the nucleocapsid N protein. This protein plays a role in encapsidating the viral RNA to form ribonucleoprotein complexes (RNP). The N protein is also involved in a range of interactions with other molecules[20] including viral RNA[24, 25], viral polymerase, other viral proteins[31], host proteins[31] as well as forming multimers with themselves[1]. Therefore, trying to identify key sites or domains can provide insight into the specific role of the N protein in these various functions.

In this study we have used a bioinformatics approach to analyze an alignment by (i) estimating the phylogenetic relationship between the sequences, (ii) identifying amino acid changes in each sequence that produce significant modifications to the physical properties of the protein (iii) analyzing these changes with respect to the tree and alignment to investigate whether regions exist where specific mutations are accompanied by compensatory changes elsewhere in sequence. This can be used to gain information regarding sites that share some functional or structural role in the protein.

The most surprising finding in this study is that for all three categories (i.e. charge, hydropathy and volume) the Asia 2 clade had significantly more changes than any other clade in the tree. Although the Asia 2 clade contained twice as many samples as any other clade[20], it seems this alone can not explain the observed differences; even when the much more genetically diverse European and African clades were clustered into single European and African clades they still failed to contain as many mutations.

From this analysis we have identified several sites of interest (AA127, AA181, AA262, AA263 & AA311) that could usefully be investigated experimentally to see their effect on virus fitness. There have been reported mutational analysis studies on the Bunyamwera Orthobunyavirus N gene[7, 35] that identified several key mutations that had detrimental effects on viral replication and fitness. We attempted to relate these findings to our results but the *Bunyaviridae* genera are too diverse to be interpreted here. However, there have been multiple reports that have identified the role

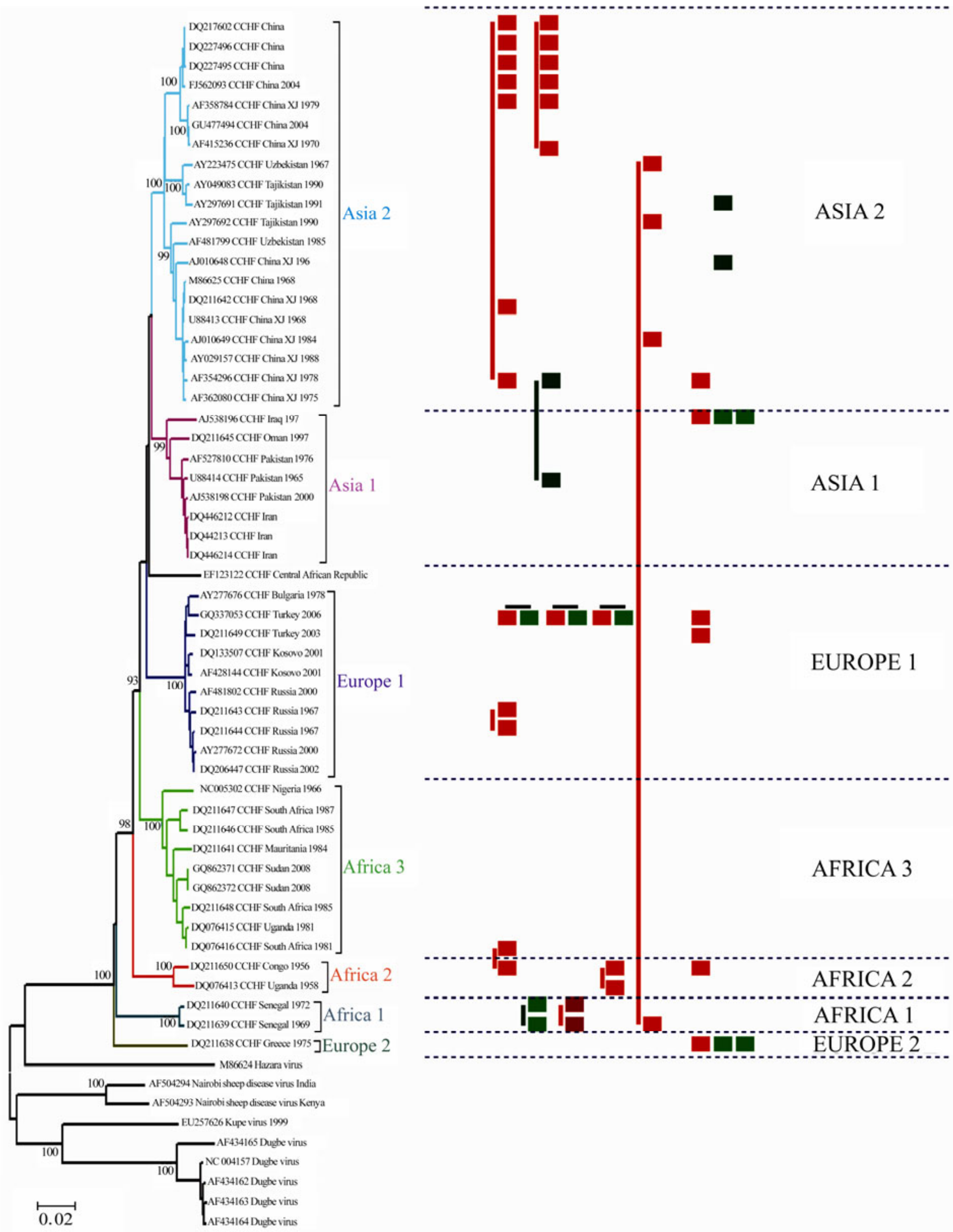
DQ217602 CCHF China
DQ227496 CCHF China
DQ227495 CCHF China
FJ562093 CCHF China 2004
AF358784 CCHF China XJ 1979
GU477494 CCHF China 2004
AF415236 CCHF China XJ 1970
AY223475 CCHF Uzbekistan 1967
AY049083 CCHF Tajikistan 1990
AY297691 CCHF Tajikistan 1991
AY297692 CCHF Tajikistan 1990
AF481799 CCHF Uzbekistan 1985
AJ010648 CCHF China XJ 196
M86625 CCHF China 1968
DQ211642 CCHF China XJ 1968
U88413 CCHF China XJ 1968
AJ010649 CCHF China XJ 1984
AY029157 CCHF China XJ 1988
AF354296 CCHF China XJ 1978
AF362080 CCHF China XJ 1975
Asia 2
AJ538196 CCHF Iraq 197
DQ211645 CCHF Oman 1997
AF527810 CCHF Pakistan 1976
U88414 CCHF Pakistan 1965
AJ538198 CCHF Pakistan 2000
DQ446212 CCHF Iran
DQ44213 CCHF Iran
DQ446214 CCHF Iran
Asia 1
EF123122 CCHF Central African Republic
AY277676 CCHF Bulgaria 1978
GQ337053 CCHF Turkey 2006
DQ211649 CCHF Turkey 2003
DQ133507 CCHF Kosovo 2001
AF428144 CCHF Kosovo 2001
AF481802 CCHF Russia 2000
DQ211643 CCHF Russia 1967
DQ211644 CCHF Russia 1967
AY277672 CCHF Russia 2000
DQ206447 CCHF Russia 2002
Europe 1
NC005302 CCHF Nigeria 1966
DQ211647 CCHF South Africa 1987
DQ211646 CCHF South Africa 1985
DQ211641 CCHF Mauritania 1984
GQ862371 CCHF Sudan 2008
GQ862372 CCHF Sudan 2008
DQ211648 CCHF South Africa 1985
DQ076415 CCHF Uganda 1981
DQ076416 CCHF South Africa 1981
Africa 3
DQ211650 CCHF Congo 1956
DQ076413 CCHF Uganda 1958
Africa 2
DQ211640 CCHF Senegal 1972
DQ211639 CCHF Senegal 1969
Africa 1
DQ211638 CCHF Greece 1975
Europe 2
M86624 Hazara virus
AF504294 Nairobi sheep disease virus India
AF504293 Nairobi sheep disease virus Kenya
EU257626 Kupe virus 1999
AF434165 Dugbe virus
NC 004157 Dugbe virus
AF434162 Dugbe virus
AF434163 Dugbe virus
AF434164 Dugbe virus
100
100
100
100
99
99
93
100
98
100
100
100
100
100
100
100
0.02
ASIA 2
ASIA 1
EUROPE 1
AFRICA 3
AFRICA 2
AFRICA 1
EUROPE 2

Fig. 4. Positive and negative charge mutations in figure 2 mapped on to the predicted tree in figure 1. Mutations that occur at the same site in multiple sequences are connected by a vertical line. Sequences that have a mutation that changes both indices, i.e. a mutation that changes the site from positive to negative (or vice versa), are connected by a horizontal line (e.g. sequence GQ337053 contains three mutations at three different sites that produce positive to negative changes in charge). Only the Asia 2 clade contains multiple sequences with shared mutations. The Senegal sequences (Africa 1 - DQ211639 and DQ211640)) share compensatory positive/negative mutations.

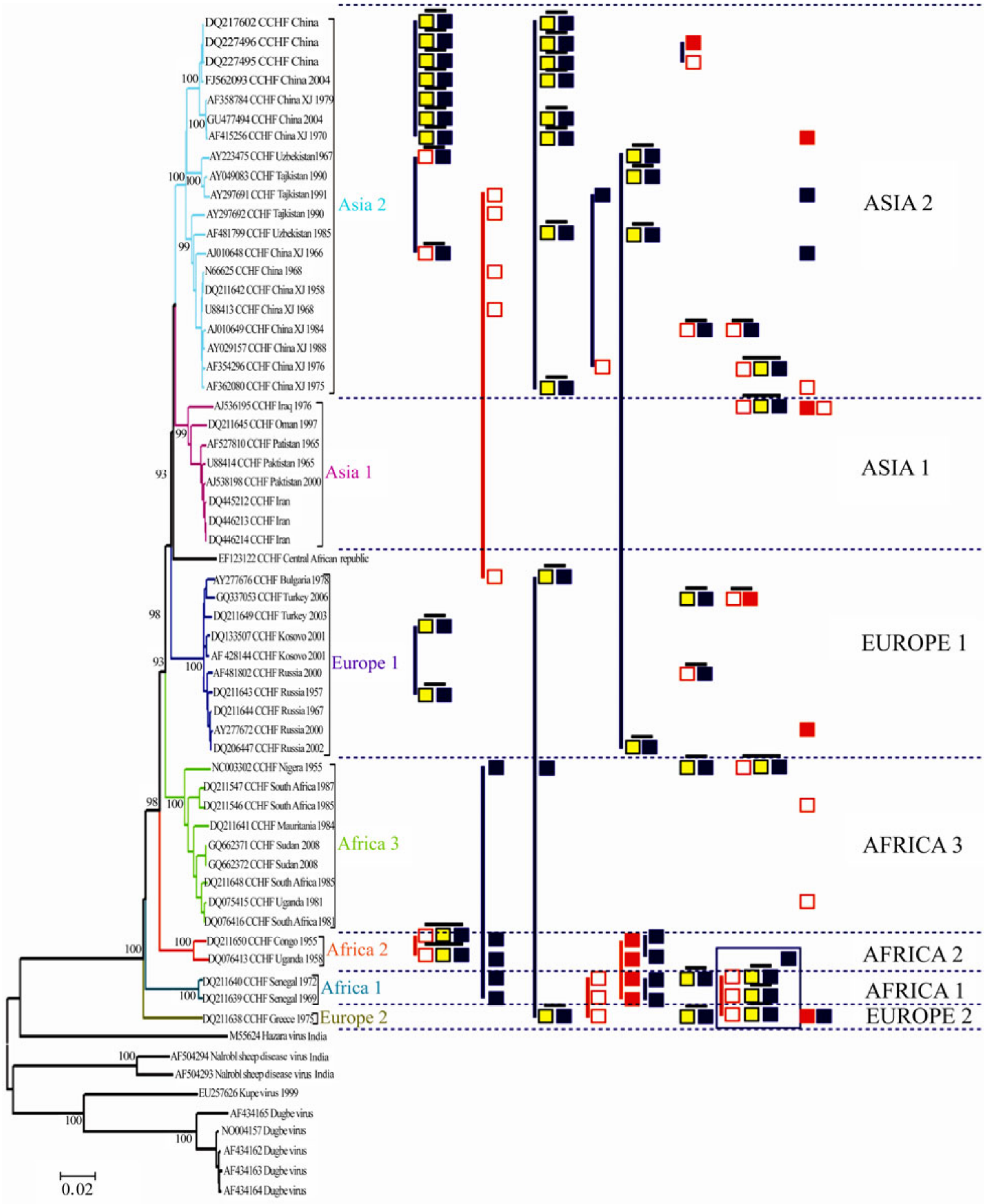
DQ217602 CCHF China
DQ227496 CCHF China
DQ227495 CCHF China
FJ562093 CCHF China 2004
AF358784 CCHF China XJ 1979
GU477494 CCHF China 2004
AF415256 CCHF China XJ 1970
AY223475 CCHF Uzbekistan1967
AY049083 CCHF Tajkistan 1990
AY297691 CCHF Tajkistan 1991
AY297692 CCHF Tajkistan 1990
AF481799 CCHF Uzbekistan 1985
AJ010648 CCHF China XJ 1966
N66625 CCHF China 1968
DQ211642 CCHF China XJ 1958
U88413 CCHF China XJ 1968
AJ010649 CCHF China XJ 1984
AY029157 CCHF China XJ 1988
AF354296 CCHF China XJ 1976
AF362080 CCHF China XJ 1975
AJ536195 CCHF Iraq 1976
DQ211645 CCHF Oman 1997
AF527810 CCHF Patistan 1965
U88414 CCHF Paktistan 1965
AJ538198 CCHF Paktistan 2000
DQ445212 CCHF Iran
DQ446213 CCHF Iran
DQ446214 CCHF Iran
EF123122 CCHF Central African republic
AY277676 CCHF Bulgaria 1978
GQ337053 CCHF Turkey 2006
DQ211649 CCHF Turkey 2003
DQ133507 CCHF Kosovo 2001
AF 428144 CCHF Kosovo 2001
AF481802 CCHF Russia 2000
DQ211643 CCHF Russia 1957
DQ211644 CCHF Russia 1967
AY277672 CCHF Russia 2000
DQ206447 CCHF Russia 2002
NC003302 CCHF Nigera 1955
DQ211547 CCHF South Africa 1987
DQ211546 CCHF South Africa 1985
DQ211641 CCHF Mauritania 1984
GQ662371 CCHF Sudan 2008
GQ662372 CCHF Sudan 2008
DQ211648 CCHF South Africa 1985
DQ075415 CCHF Uganda 1981
DQ076416 CCHF South Africa 1981
DQ211650 CCHF Congo 1955
DQ076413 CCHF Uganda 1958
DQ211640 CCHF Senegal 1972
DQ211639 CCHF Senegal 1969
DQ211638 CCHF Greece 1975
M55624 Hazara virus India
AF504294 Nalrobl sheep disease virus India
AF504293 Nalrobl sheep disease virus India
EU257626 Kupe virus 1999
AF434165 Dugbe virus
NO004157 Dugbe virus
AF434162 Dugbe virus
AF434163 Dugbe virus
AF434164 Dugbe virus
100
100
100
100
99
99
93
98
93
100
98
100
100
100
100
100
100
100
0.02
Asia 2
Asia 1
Europe 1
Africa 3
Africa 2
Africa 1
Europe 2
ASIA 2
ASIA 1
EUROPE 1
AFRICA 3
AFRICA 2
AFRICA 1
EUROPE 2

Fig. 5. Significant pka (yellow), hydropathy (blue) and volume mutations (red) mapped on to the predicted tree in figure 1. Mutations that occur at the same site in multiple sequences are connected by a vertical line. Sequences that have a mutation that changes more than one index at the same site are connected by a horizontal line. For example, the Uganda (DQ076413) and Congo (DQ211650) sequences in the Africa 2 clade have two mutations at two different sites (AA 181 & AA 311) that modify all three indices. Other mutations in this Africa 2 clade also modify the same sites in the Africa 1 sequences (Senegal DQ211639 and Senegal DQ211640). The box in the bottom right of the figure that spans the Africa 1, Africa 2 and Europe 2 clades delimit a set of mutations and sites that are shared across the clades and which modify three indices.

of positively charged amino acids in RNA binding and oligomerization[9, 10, 38] which provides further support to our findings.

Our analysis of CCHFV is our first attempt to perform a mutational analysis of a gene by integrating an alignment, a tree and the AAIndex amino acid database. Given our method achieves reasonable predicts we next plan to analyze a gene for which a solved structure and experimental mutational studies are available.

References

1. **Albertini A A, Wernimont A K, Muziol T, *et al.*** 2006. Crystal structure of the rabies virus nucleoprotein-RNA complex. **Science,** 313:360-363.
2. **Anagnostou V, Papa A.** 2009. Evolution of Crimean-Congo Hemorrhagic Fever virus. **Infect Genet Evol,** 9:948-954.
3. **Aradaib I E, Erickson B R, Karsany M S, *et al.*** 2011. Multiple crimean-congo hemorrhagic Fever virus strains are associated with disease outbreaks in Sudan, 2008-2009. **PLoS Negl Trop Dis,** 5:e1159.
4. **Aradaib I E, Erickson B R, Mustafa M E, *et al.*** 2010. Nosocomial outbreak of Crimean-Congo hemorrhagic fever, Sudan. **Emerg Infect Dis,** 16:837-839.
5. **Chinikar S, Persso n S M, Johansson M, *et al.*** 2004. Genetic analysis of Crimean-congo hemorrhagic fever virus in Iran. **J Med Virol,** 73:404-411.
6. **Drosten C, Minnak D, Emmerich P, *et al.*** 2002. Crimean-Congo hemorrhagic fever in Kosovo. **J Clin Microbiol,** 40:1122-1123.
7. **Eifan S A, Elliott R M.** 2009. Mutational analysis of the Bunyamwera orthobunyavirus nucleocapsid protein gene. **J Virol,** 83:11307-11317.
8. **Eisenberg D.** 1984. Three-dimensional structure of membrane and surface proteins. **Annu Rev Biochem,** 53:595-623.
9. **Elton D, Medcalf E, Bishop K, *et al.*** 1999. Oligomerization of the influenza virus nucleoprotein: identification of positive and negative sequence elements. **Virology,** 260:190-200.
10. **Elton D, Medcalf L, Bishop K, *et al.*** 1999. Identification of amino acid residues of influenza virus nucleoprotein essential for RNA binding. **J Virol**, 73:7357-7367.
11. **Fauchere J L, Charton M, Kier L B, *et al.*** 1988. Amino acid side chain parameters for correlation studies in biology and pharmacology. **Int J Pept Protein Res,** 32:269-278.
12. **Gao X, Nasci R, Liang G.** 2010. The neglected arboviral infections in mainland China. **PLoS Negl Trop Dis,** 4:e624.
13. **Gargili A, Midilli K, Ergonul O, *et al.*** 2011. Crimean-congo hemorrhagic Fever in European part of Turkey: genetic analysis of the virus strains from ticks and a seroepidemiological study in humans. **Vector Borne Zoonotic Dis,** 11:747-752.
14. **Goldsack D E, Chalifoux R C.** 1973. Contribution of the free energy of mixing of hydrophobic side chains to the stability of the tertiary structure of proteins. **J Theor Biol,** 39:645-651.
15. **Hewson R, Chamberlain J, Mioulet V, *et al.*** 2004. Crimean-Congo haemorrhagic fever virus: sequence analysis of the small RNA segments from a collection of viruses world wide. **Virus Res,** 102:185-189.
16. **Hoogstraal H.** 1979. The epidemiology of tick-borne Crimean-Congo hemorrhagic fever in Asia, Europe, and Africa. **J Med Entomol,** 15:307-417.
17. **Kawashima S, Pokarowski P, Pokarowska M, *et al.*** 2008. AAindex: amino acid index database, progress report 2008. **Nucl Acids Res,** 36:D202-205.
18. **Krigbaum W R, Komoriya A.** 1979. Local interactions as a structure determinant for protein molecules: II. **Biochim Biophys Acta,** 576:204-248.
19. **Kuhn J H, Seregin S V, Morzunov S P, *et al.*** 2004. Genetic analysis of the M RNA segment of Crimean-Congo hemorrhagic fever virus strains involved in the recent outbreaks in Russia. **Arch Virol,** 149:2199-2213.
20. **Longhi S.** 2009. Nucleocapsid structure and function. **Curr Top Microbiol Immunol,** 329:103-128.
21. **Meissner J D, Seregin S S, Seregin S V, *et al.*** 2006. A variable region in the Crimean-Congo hemorrhagic fever virus L segment distinguishes between strains isolated from different geographic regions. **J Med Virol,** 78:223-228.
22. **Midilli K, Gargili A, Ergonul O, *et al.*** 2007. Imported Crimean-Congo hemorrhagic fever cases in Istanbul. **BMC Infect Dis,** 7:54.
23. **Mild M, Simon M, Albert J, *et al.*** 2010. Towards an understanding of the migration of Crimean-Congo hemorrhagic fever virus. **J Gen Virol,** 91:199-207.
24. **Mir M A, Panganiban A T.** 2005. The hantavirus

nucleocapsid protein recognizes specific features of the viral RNA panhandle and is altered in conformation upon RNA binding. **J Virol,** 79:1824-1835.

25. **Mohl B P, Barr J N.** 2009. Investigating the specificity and stoichiometry of RNA binding by the nucleocapsid protein of Bunyamwera virus. **RNA,** 15:391-399.
26. **Papa A, Velo E, Papadimitriou E, *et al.*** 2009. Ecology of the Crimean-Congo hemorrhagic fever endemic area in Albania. **Vector Borne Zoonotic Dis,** 9:713-716.
27. **Roseman M A.** 1988. Hydrophilicity of polar amino acid side-chains is markedly reduced by flanking peptide bonds. **J Mol Biol,** 200:513-522.
28. **Sun S, Dai X, Aishan M, *et al.*** 2009. Epidemiology and phylogenetic analysis of crimean-congo hemorrhagic fever viruses in xinjiang, china. **J Clin Microbiol,** 47:2536-2543.
29. **Tahmasebi F, Ghiasi S M, Mostafavi E, *et al.*** 2010. Molecular epidemiology of Crimean- Congo hemorrhagic fever virus genome isolated from ticks of Hamadan province of Iran. **J Vector Borne Dis,** 47:211-216.
30. **Tamura K, Peterson D, Peterson N, *et al.*** 2011. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. **Mol Biol Evol.** 2011 May 4. [Epub ahead of print].
31. **Terribilini M, Lee J H, Yan C, *et al.*** 2006. Prediction of RNA binding sites in proteins from amino acid sequence. **RNA,** 12:1450-1462.
32. **Thompson J D, Higgins D G, Gibson T J.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. **Nucl Acids Res,** 22:4673-4680.
33. **Tonbak S, Aktas M, Altay K, *et al.*** 2006. Crimean-Congo hemorrhagic fever virus: genetic analysis and tick survey in Turkey. **J Clin Microbiol,** 44:4120-4124.
34. **Tsai J, Taylor R, Chothia C, *et al.*** 1999. The packing density in proteins: standard radii and volumes. **J Mol Biol,** 290:253-266.
35. **Walter C T, Bento D F, Alonso A G, *et al.*** 2011. Amino acid changes within the Bunyamwera virus nucleocapsid protein differentially affect the mRNA transcription and RNA replication activities of assembled ribonucleoprotein templates. **J Gen Virol,** 92:80-84.
36. **Xia H, Li P, Yang J, *et al.*** 2011. Epidemiological survey of Crimean-Congo hemorrhagic fever virus in Yunnan, China, 2008. **Int J Infect Dis,** 15:e459-463.
37. **Yashina L, Vyshemirskii O, Seregin S, *et al.*** 2003. Genetic analysis of Crimean-Congo hemorrhagic fever virus in Russia. **J Clin Microbiol,** 41:860-862.
38. **Ye Q, Krug R M, Tao Y J.** 2006. The mechanism by which influenza A virus nucleoprotein forms oligomers and binds RNA. **Nature,** 444:1078-1082.