ORIGINAL ARTICLE

Comparative genome analysis of Alkhumra hemorrhagic fever virus with Kyasanur forest disease and tick-borne encephalitis viruses by the *in silico* approach

Navaneethan Palanisamy^{a,b} , Dario Akaberi^{c,d}, Johan Lennerstrand^{c,d} and Åke Lundkvist^d

^aMolecular and Cellular Engineering Group, University of Freiburg, Institute of Biology II, Freiburg, Germany; ^bThe Hartmut Hoffmann-Berling International Graduate School of Molecular and Cellular Biology (HBIGS), University of Heidelberg, Heidelberg, Germany; ^cSection of Clinical Microbiology, Department of Medical Sciences, Uppsala University, Uppsala, Sweden; ^dDepartment of Medical Biochemistry and Microbiology, Zoonosis Science Center, Uppsala University, Uppsala, Sweden

ABSTRACT

Alkhumra hemorrhagic fever virus (AHFV), a relatively new member of the Flaviviruses, was discovered in Saudi Arabia 23 years ago. AHFV is classified in the tick-borne encephalitis virus serocomplex, along with the Kyasanur forest disease virus (KFDV) and tick-borne encephalitis virus (TBEV). Currently, very little is known about the pathologies of AHFV. In this study, using the available genome information of AHFV, KFDV and TBEV, we have predicted and compared the following aspects of these viruses: evolution, nucleotide and protein compositions, recombination, codon frequency, substitution rate, N- and O-glycosylation sites, signal peptide and cleavage site, transmembrane region, secondary structure of 5' and 3' UTRs and RNA-RNA interactions. Additionally, we have modeled the 3D protease and RNA-dependent RNA polymerase structures for AHFV, KFDV and TBEV. Recombination analysis showed no evidence of recombination in the AHFV genome with that of either KFDV or TBEV, although single break point analysis showed that nucleotide position 7399 (in the NS4B) is a breakpoint location. AHFV, KFDV and TBEV are very similar in terms of codon frequency, the number of transmembrane regions, properties of the polyprotein, RNA-RNA interaction sequences, NS3 protease and NS5 polymerase structures and 5' UTR structure. Using genome sequences, we showed the similarities between these closely-related viruses on several different areas.

KEYWORDS

Alkhumra hemorrhagic fever virus (AHFV); Kyasanur forest disease virus (KFDV); tick-borne encephalitis virus (TBEV); computational tools and servers; prediction

Introduction

Alkhumra hemorrhagic fever virus (AHFV), or Alkhurma, belongs to the *Flavivirus* genus of the *Flaviviridae* family. This *Flavivirus* genus comprises over 70 viruses, nearly half of which are reported to cause infections in humans [1]. Other well-known examples of human pathogenic viruses within this genus include Zika virus (ZIKV), dengue virus (DENV), Japanese encephalitis virus (JEV), West Nile virus (WNV) and yellow fever virus (YFV). AHFV is a relatively new virus discovered only 23 years ago [2]. This virus was initially isolated from a butcher in the Al-Khumra district of the Jeddah metropolitan in Saudi Arabia [2]. The patient was originally misdiagnosed as infected with the Crimean-Congo hemorrhagic fever virus (CCHFV) (*Bunyaviridae* family); however, Dr. Zaki and his colleagues were later able to establish that it was a novel virus [2]. By performing phylogenetic studies using the NS5 sequence, they showed that the virus clustered within the tick-borne encephalitis virus serocomplex. It closely resembled, yet is distinct from the Kyasanur forest

disease virus (KFDV), another tick-borne virus, which was discovered in the late 1950's in Karnataka, India [2].

Like other Flaviviruses, the AHFV genome consists of single-stranded, positive-sense RNA of approximately 10.5 kb, and a single open reading frame. The polyprotein is cleaved into three structural proteins (core, prM and envelope) and eight non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, 2K, NS4B and NS5) by viral and host proteases. The coding sequence is localized between the 5' and 3' untranslated regions (UTRs), which are involved in the pathogenicity of the virus [3]. The NS3 region harbors trypsin-like serine protease and helicase domains, while the NS5 region harbors methyl transferase and RNA-dependent RNA polymerase domains. The NS3 protease activity depends on the presence of NS2B as a cofactor [4,5]. AHFV, like all other Flaviviruses, is an enveloped virus. It can be rendered inactive by exposure to 70% ethanol, dessication or thermal denaturation. An earlier study has shown that AHFV can be completely inactivated at 60 °C when incubated for 3 min, or at 56 °C for 30 min [6].

From univariate and multivariate analyses, the common modes of AHFV transmission included direct contact with the blood of infected animals, like sheep and camel, e.g. while butchering, drinking the raw milk or eating raw meat, or by bites from ticks who have also infected said animals [7]. After the discovery of AHFV, there was no real evidence of the involvement of ticks for the transmission of AHFV, except for the fact that it was phylogenetically clustered within the tick-borne encephalitis virus (TBEV) serocomplex. In 2007, Charrel et al. published the first evidence of the presence of AHFV RNA in *Ornithodoros savignyi*, a soft tick, isolated from a camel resting place near Jeddah, Saudi Arabia [8]. In 2011, Mahdi et al. published detection of AHFV RNA from *O. savignyi* and *Hyalomma dromedarii* (a hard tick) [9]. The study by Charrel et al. also showed that AHFV isolated from ticks form sub-lineages distinct from the ones isolated from humans [8]. Despite detection of AHFV RNA in ticks, transmission from ticks to humans has not yet been confirmed. It is still hypothesized that mosquitoes could potentially transmit AHFV, but this is yet to be proven [10,11]. Furthermore, it is poorly understood how it replicates in sheeps and camels, as inoculation of the virus did not show any obvious symptoms in sheep [12]. Since much is unknown about this virus, AHFV is handled either in BSL-3 or BSL-4 labs, depending on the country.

AHFV-infected patients have been shown to present both hemorrhagic (e.g. thrombocytopenia, leucopenia, kidney and spleen damage, hepatitis, petechia, etc.) and neurological (e.g. seizures, encephalitis, etc.) symptoms [10,11]. The case fatality rate among hospitalized patients was reported to be 25%, making it one of the most deadly pathogens among the Flaviviruses [10]. Not all AHFV-infected persons show severe clinical symptoms, and some individuals show completely sub-clinical infections [7]. Currently, there is no specific drug or vaccine available against AHFV. Treatments are provided only to alleviate the symptoms and maintain the vital signs of the patients. A study about the prevalence of hemorrhagic fever viruses in Saudi Arabia's military personnel has shown that AHFV has a seroprevalence of 1.3%, which is considered high compared to other hemorrhagic fever viruses circulating in Saudi Arabia, like the Rift Valley fever virus (RVFV), DENV and CCHFV [13]. Studies conducted between 2001 and 2003 in Makkah, and between 2003 and 2009 in Najran, have identified 37 and 78 AHFV cases, respectively [7,10]. The case fatality rate in Najran, a highly endemic region in Saudi Arabia, was about 1% [7,11,14]. Until the last decade, AHFV cases were found to be confined solely within Saudi Arabia. Recently, a study showed that an Italian traveler returning from Egypt has been confirmed with AHFV infections and another study has revealed the presence of AHFV RNA in ticks isolated from Djibouti [15]. Due to the increase in travel and in the import and export of food products, there is an urgent need to monitor and curb the transmission of AHFV at the budding stage. We

hypothesized that since AHFV sequences cluster with the TBEV serocomplex in the phylogeny, the genomes (and its architecture) and proteomes (including post-translational modifications) of AHFV, KFDV and classical TBEV should be similar. In this *in silico* study, we have used available genome and proteome information to predict and compare AHFV with its close relatives, KFDV and TBEV, with respect to evolution, nucleotide and protein compositions, recombination, codon frequency, substitution rate, N and O-glycosylation sites, signal peptide and cleavage site, transmembrane region, secondary structure of 5' and 3' UTRs, RNA-RNA interactions, and protease and RNA-dependant RNA polymerase modeled structures. Additionally, we have also studied nucleotide polymorphisms in the 5' and 3' UTRs, and the protein polymorphisms of the CDS in AHFV using the sequences that have been sequenced so far.

Materials and methods

Sequence retrieval, analysis and phylogeny

The complete genome and/or the complete coding sequence (CDS) of 24 AHFV strains, 6 KFDV strains and eight other Flaviviruses were retrieved from the NCBI database on 16 March 2017. For the sequence accession numbers, see Figure 1 or Supplementary Table 1. In case of TBEV, sequence from the Neudoerfl strain was used.

Multiple sequence alignment (MSA) was performed using the ClustalW function in MEGA6 [16] with default settings. Maximum likelihood (ML) trees were constructed from the aligned sequences using the same MEGA6 [16]. While constructing the trees, the bootstrap value was adjusted to 500, the Tamura-Nei model (nucleotide) was chosen as the substitution model, the initial tree for ML was derived from the NJ/BioNJ algorithm and the nearest-neighbor-interchange (NNI) option was chosen in the ML heuristic method. The remaining parameters used default values.

Sequence (nucleotide or protein) identity and similarity were studied using the matrix global alignment tool (MatGAT) with default parameters [17]. While studying the similarity of the protein sequences, the BLOSUM62 matrix was used. Additionally, we also used the sequence identity and similarity (SIAS) server (<http://imed.med.ucm.es/Tools/sias.html>) to verify the MatGAT results.

Recombination analysis

Recombination of sequence between AHFV and other Flaviviruses was studied using the recombination detection program 4.85 (RDP4 Beta 4.85) [18]. The program was downloaded using this link: <http://web.cbio.uct.ac.za/~darren/rdp.html>. ClustalW aligned sequences of AHFV and other Flaviviruses were used as input to the program. RDP is a multiple software suite; it uses algorithms of RDP, GENECONV, Chimaera, MaxChi, Bootscan,

Figure 1. Phylogenetic analysis. Maximum likelihood (ML) tree was built using the complete genome or the complete coding sequence (CDS). AHFV isolates cluster with KFDV and TBEV in the ML tree. Among the AHFV sequences, the virus sequence isolated from ticks (represented by green or red) forms distinct clusters from the human sequences (represented by blue). AHFV sequence from *Ornithodoros savignyi* ticks is represented in green while AHFV sequence from *Hyalomma dromedarii* tick is represented in red. The viruses are represented as NCBI accession number|Name|strain name in the figure. AHFV = *Alkhurma hemorrhagic fever virus*, KFDV = *Kyasanur forest disease virus*, TBEV = *tick-borne encephalitis virus*, ENTV = *Entebbe bat virus*, YFV = *yellow fever virus*, JEV = *Japanese encephalitis virus*, WNV = *West Nile virus*, DENV = *dengue virus*, ZIKV = *Zika virus* and SPOV = *Spondweni virus*.

SiScan, 3Seq and LARD. We used the parameters as described by Lemey and Posda [19].

We also used single break point (SBP) and genetic algorithm recombination detection (GARD) methods in the datamonkey server (<http://www.datamonkey.org/>) to screen for recombinations [20]. A p-value of 0.1 was given as additional input.

Codon usage data and frequency analysis

Codon usage data for *Homo sapiens* (human), *O. savignyi* (tick), *Aedes aegypti* (mosquito), AHFV, KFDV and TBEV were retrieved from the Kazusa codon usage database (<http://www.kazusa.or.jp/codon/>).

Change in the codon frequency between different AHFV sequences was analyzed by using a Javascript previously published [21].

Determination of Ka/Ks ratio

Ka (non-synonymous) and Ks (synonymous) substitutions in the coding sequence of AHFV, KFDV and TBEV sequences were studied using MEGA6 [16]. ClustalW-aligned nucleotide sequences were used as input. While computing the pairwise distances, the Nei-Gojobori (Jukes-Cantor) method was chosen and the rest of the parameters were left with default values. The Nei-Gojobori method calculates the numbers of synonymous and non-synonymous substitutions and the numbers of possible synonymous and possible non-synonymous sites [22].

Positive selection sites within the CDS were further identified using the datamonkey server (<http://www.datamonkey.org/>) [20]. While using this server, single likelihood ancestor counting (SLAC) and fixed-effects likelihood (FEL) methods were chosen and a p-value of 0.1 was given as additional input. The SLAC method was derived from the Suzuki-Gojobori method [23]. SLAC calculates the numbers of synonymous and non-synonymous change and tests how significant non-synonymous numbers from synonymous numbers [23]. The FEL method utilizes the ML algorithm to deduce non-synonymous and synonymous substitution rates at a site [24]. This FEL method assumes that the evolutionary rate is constant [24].

Post-translation analyses

The N- and O-linked glycosylation sites in the protein were predicted using online servers NetNGlyc 1.0 (<http://www.cbs.dtu.dk/services/NetNGlyc/>) and NetOGlyc 4.0 (<http://www.cbs.dtu.dk/services/NetOGlyc/>) [25], respectively.

SignalP 4.1 server (<http://www.cbs.dtu.dk/services/SignalP/>) [26] was used to predict the presence or absence of the signal peptide from the amino acid sequence and also to predict the location of the signal peptide cleavage site. While using this server, all the parameters were left with default values except for the N-terminal truncation, where we instead gave '0' (i.e. no truncation).

The transmembrane region of the protein was predicted using the TMHMM 2.0 server (<http://www.cbs.dtu.dk/services/TMHMM/>).

The general properties, e.g. molecular weight, theoretical pI, residue composition, extinction coefficient, half-life, instability index (II), aliphatic index (AI) and grand average of hydropathicity (GRAVY), of the protein were predicted using the ExPASy ProtParam tool (<http://web.expasy.org/protparam/>).

RNA and protein folding

The secondary structure of RNAs (5' UTR and 3' UTR) was predicted using the RNAfold web server with default values (<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>) [27].

The interactions between 5' UTR and 3' UTR were predicted using the RactIP (RNA-RNA interaction prediction using integer programming) server (<http://rtips.dna.bio.keio.ac.jp/ractip/>) [28] by selecting a CONTRAfold model (default) and leaving the internal and external values to 1.

Homology 3D modeling of proteins was performed by using the protein homology or analogy recognition engine 2.0 (Phyre2) (<http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index>) [29]. While modeling, the *intensive* mode was chosen. For the energy minimization of protein models, the ModRefiner server was used (<http://zhanglab.ccmb.med.umich.edu/ModRefiner/>) [30]. For validating the models, the following servers were used: (1) Ramachandran plot (PROCHECK): an inbuilt of PDBSUM (<http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html>) [31], (2) Verify_3D server (http://services.mbi.ucla.edu/Verify_3D/) [32,33], (3) ERRAT (<http://services.mbi.ucla.edu/ERRAT/>) [34], (4) PROVE (<https://services.mbi.ucla.edu/SAVES/>) [35] and (5) ProSA-web (<https://prosa.services.came.sbg.ac.at/prosa.php>) [36]. The protein models were structurally compared using the iPBA web server (http://www.dsimb.inserm.fr/dsimb_tools/ipba/) [37].

Results

AHFV sequences can be further classified into sub-lineages

Phylogenetic analysis was carried out to study the relation between the different AHFV strains with the KFDV and TBEV sequences. Supplementary Table 1 gives updated information about the AHFV sequences used in this study. To make the analysis more informative, we also included other Flaviviruses, namely JEV, WNV lineage 2, DENV2, YFV, ZIKV, Entebbe bat virus (ENTV) and Spondweni virus (SPOV). The whole genomes or complete CDSs of the viruses were used as input. From the ML tree, it was shown that all AHFV sequences clustered together adjacent to the KFDV sequences with a bootstrap value of 100 (Figure 1). Both the AHFV and KFDV sequences again clustered with TBEV with a bootstrap value of 100. Within the 24 AHFV sequences, it was clear that all tick AHFV sequences except one from *O. savignyi* (Figure 1, represented in green) clustered distinctly from the human AHFV sequences (Figure 1, represented in blue). Interestingly, the remaining AHFV sequence from *O. savignyi* clustered with one AHFV sequence from *Hyalomma dromedarii* (Figure 1, represented in red).

Sequence studies show that AHFV strains are very similar to each other and also similar to KFDV and TBEV

The complete CDS of 24 AHFV strains were given as input to the MatGAT application. Additionally, a value of 16 (default) was given as input for the first gap and 4

(default) was given for the extending gap. The nucleotide identity of the AHFV strains varied between 98.5 and 100%. The lowest identity (i.e. 98.5%) was found between KU884971.1 and JX914663.1 and between KU884971.1 and JX914664.1.

The nucleotide sequences were translated to protein sequences using the EMBOSS Transeq server (http://www.ebi.ac.uk/Tools/st/emboss_transeq/). The first translation frame gave the correct protein sequence without any interrupting stop codons. The translated protein sequences were given as input to the MatGAT application and additional values were given as input: 12 (default) for the first gap and 2 (default) for the extending gap. Protein similarity of AHFV strains varied between 99.5 and 100%. The lowest similarity (i.e. 99.5%) was found between JF416961.1 and JX914663.1 and JX914664.1, as well as between JF416962.1 and JX914663.1 and JX914664.1.

Similarly, we compared nucleotide identity and protein similarity of AHFV with KFDV and TBEV. We used one representative sequence from each virus for comparison. We found that AHFV and KFDV have a nucleotide identity of approximately 92.5% and a protein similarity of approximately 98.6%. Additionally, we found that AHFV and TBEV have a nucleotide identity of approximately 70.5% and a protein similarity of roughly 89.3%. We used the SIAS server to further confirm the results of the MatGAT application (data not shown).

The GC content of CDS in AHFV and TBEV were around 54%, while it was around 55% in KFDV.

No evidence of recombination in the AHFV genome with the studied Flaviviruses

The ClustalW aligned nucleotide sequences (complete genome or CDS) were given as input to the RDP. The program showed that there was no evident recombination in AHFV.

SBP analysis on AHFV CDS using the datamonkey server predicted nucleotide position 7399 (in the NS4B) to be the breakpoint location with 100% confidence. On the contrary, GARD analysis did not predict any breakpoint location in the AHFV CDS.

Minor differences in the codon frequency between AHFV and its vector(s)/host, which may not have any significant negative impact on the viral protein expression level

Viruses hijack the host's protein translation machineries and use them for viral replication. Codon usage bias exists between different organisms and even between different cell types within the same organism [38–40]. Codon usage bias prevents or slows down the replication of foreign genetic material such as viral genomes or transposable elements in the cell. A study has shown that the interferon-inducible schlafen family member 11

Table 1A. Substitution rates in the protein coding sequence of AHFV determined by Jukes-Cantor model.

S. No.	Region	Nucleotide position ^a	Amino acid position ^b	Substitution rate		
				Ka	Ks	Ka/Ks
1	Polyprotein	115–10362	1–3416	0.004	0.022	0.182
2	Core	115–453	1–113	0.005	0.035	0.143
3	prM	454–957	114–281	0.006	0.031	0.194
4	E	958–2445	282–777	0.003	0.021	0.143
5	NS1	2446–3504	778–1130	0.002	0.024	0.083
6	NS2A	3505–4194	1131–1360	0.003	0.029	0.103
7	NS2B	4195–4587	1361–1491	0.001	0.028	0.036
8	NS3	4588–6450	1492–2112	0.001	0.026	0.038
9	NS4A	6451–6828	2113–2238	0.008	0.022	0.364
10	2K	6829–6897	2239–2261	–	0.038	–
11	NS4B	6898–7653	2262–2513	0.002	0.020	0.100
12	NS5	7654–10362	2514–3416	0.007	0.009	0.778

Notes: a = numbering based on accession number: NC_004355.1, b = numbering based on accession number: NP_722551.1. AHFV = *Alkhurma hemorrhagic fever virus*.

Table 1B. Substitution rates in the protein coding sequence of KFDV determined by Jukes-Cantor model.

S. No.	Region	Nucleotide position ^a	Amino acid position ^b	Substitution rate		
				Ka	Ks	Ka/Ks
1	Polyprotein	132–10379	1–3416	0.001	0.003	0.333
2	Core	132–482	1–117	–	–	–
3	prM	483–974	118–281	–	0.003	–
4	E	975–2462	282–777	0.001	0.003	0.333
5	NS1	2463–3524	778–1131	0.001	0.002	0.500
6	NS2A	3525–4211	1132–1360	0.001	0.008	0.125
7	NS2B	4212–4604	1361–1491	–	–	–
8	NS3	4605–6467	1492–2112	0.000	0.004	0.000
9	NS4A	6468–6845	2113–2238	0.001	–	–
10	2K	6846–6914	2239–2261	–	–	–
11	NS4B	6915–7670	2262–2513	0.001	0.003	0.333
12	NS5	7671–10379	2514–3416	0.000	0.005	0.000

Notes: a = numbering based on accession number: HM055369.1, b = numbering based on accession number: ADH95737.1. KFDV = *Kyasanur forest disease virus*.

Table 1C. Substitution rates in the protein coding sequence of TBEV determined by Jukes-Cantor model.

S. No.	Region	Nucleotide position ^a	Amino acid position ^b	Substitution rate		
				Ka	Ks	Ka/Ks
1	Polyprotein	133–10374	1–3414	0.023	0.685	0.034
2	Core	133–468	1–112	0.054	0.414	0.130
3	prM	469–972	113–280	0.025	0.619	0.040
4	E	973–2460	281–776	0.015	0.788	0.019
5	NS1	2461–3516	777–1128	0.022	0.660	0.033
6	NS2A	3517–4206	1129–1358	0.054	0.700	0.077
7	NS2B	4207–4599	1359–1489	0.020	0.678	0.029
8	NS3	4600–6462	1490–2110	0.017	0.728	0.023
9	NS4A	6463–6840	2111–2236	0.023	0.644	0.036
10	2K	6841–6909	2237–2259	0.003	0.760	0.004
11	NS4B	6910–7665	2260–2511	0.023	0.839	0.027
12	NS5	7666–10374	2512–3414	0.021	0.662	0.032

Notes: a = numbering based on accession number: NC_001672.1, b = numbering based on accession number: NP_043135.1. TBEV = *tick-borne encephalitis virus*.

(SLFN11) protein, in a codon usage-dependent manner, selectively inhibits the late stages of HIV-1 production [41]. To determine codon biasness, we studied the codon frequency differences between AHFV and its potential vectors, namely tick (*O. savignyi*) and mosquito (*A. aegypti*), and between AHFV and its host human (*H. sapiens*). Additionally, we also studied the codon frequency differences between different AHFV strains.

When comparing the codon frequencies of AHFV and humans, we found that usage of only two codons, namely GGA (glycine) and UGG (tryptophan) was more than two-fold higher in the AHFV CDS (Figure 2(A)). As

tryptophan is coded by only one codon, the higher number simply means that tryptophan is more abundant in the AHFV CDS. When comparing the codon frequencies of AHFV and *A. aegypti*, we found that usage of six codons, namely AGA (arginine), AGG (arginine), GGG (glycine), ACA (threonine), UGG (tryptophan) and GUG (Valine) was more than two-fold higher in the AHFV CDS (Figure 2(A)). Similarly, by comparing the codon frequencies of AHFV and *O. savignyi*, we found that usage of five codons was more than two-fold higher in the AHFV CDS: AGG (arginine), CGU (arginine), CGG (arginine), CAG (glutamine) and UGG (tryptophan) (Figure 2(A)). These

results clearly indicate that AHFV's replication may not be significantly restricted due to codon usage bias in human or tick. Interestingly, while comparing codon frequencies of AHFV and mosquito, three out of the six above-mentioned codons were more than four-fold higher in AHFV: AGA (arginine), AGG (arginine) and GGG (glycine) (Figure 2(A)).

We also compared the codon frequency of AHFV with KFDV and TBEV separately. We found that there were no significant differences except for codon UUA (leucine), which was used approximately two-fold more in AHFV compared to KFDV (Figure 2(B)).

Finally, we studied codon frequency differences between the different AHFV strains. We found that the mean values were quite similar to the values in the Kazusa database with smaller sample standard deviations (SDs) (Figure 2(C)).

Substitution rate studies show AHFV, KFDV and TBEV undergo purifying or negative selection

We studied non-synonymous (K_a) and synonymous (K_s) substitutions in the CDS of AHFV, KFDV and TBEV. Non-synonymous substitutions in the CDS leads to changes in the amino acid sequence, while synonymous substitution does not lead to any change in the amino acid sequence. This study yielded information about whether a virus acquired beneficial mutation (i.e. when $K_a/K_s > 1$) or neutral mutation (i.e. when $K_a/K_s = 1$) or negative mutation (i.e. when $K_a/K_s < 1$). Using 24 AHFV sequences, we found that all the proteins were undergoing negative selection. Similarly, using 6 KFDV sequences and 150 TBEV sequences separately, we found that all the proteins in these viruses were also undergoing purifying or negative selection. Tables 1A–1C shows K_a , K_s and K_a/K_s values for AHFV, KFDV and TBEV, respectively.

We also studied K_a and K_s substitutions in the CDS of AHFV collected from humans and ticks. Again, we found that all the proteins in these viruses were also undergoing purifying or negative selection (i.e. $K_a/K_s < 1$). Moreover, K_a/K_s for the C, prM and NS2B genes was more than two-fold greater in the AHFV sequences collected from ticks as compared to the AHFV sequences collected from humans, while K_a/K_s for the NS5 gene was greater in the AHFV sequences collected from humans compared to the AHFV sequences collected from ticks (see Figure 3).

We further studied the positive selection sites in the CDS using the datamonkey server. From the FEL analysis, the server predicted codon 102 (located in the core protein) to undergo positive selection in the AHFV polyprotein at a significance level of 0.1. In KFDV, using both SLAC and FEL analyses, no positive selection site was predicted at the same significance level. In the TBEV polyprotein, the server predicted codon 951 (located in the NS1) to undergo positive selection by using both SLAC and FEL analyses at the same significance level.

Number of glycosylation sites was different when comparing AHFV with KFDV and TBEV

The NetNGlyc 1.0 server uses artificial neural networks to predict N-linked glycosylation sites from the protein sequence by scanning for N-X-S/T sequons, where X can be any amino acid. For this prediction, we followed the most stringent condition by choosing only sites where the prediction was '++' or higher in the output. In AHFV, we found two N-linked glycosylation sites (amino acids 145 and 862) (Figure 4(A)). In KFDV, we found only one N-linked glycosylation site (amino acid 145), which is in the same position as AHFV (Figure 4(B)). In contrast to both AHFV and KFDV, NetNGlyc 1.0 server predicted one N-linked glycosylation site in TBEV, but the residue after asparagine was proline, which is known to prevent glycosylation of the preceding residue (Figure 4(C) represented by *).

Similar to the NetNGlyc 1.0 server, the NetOGlyc 4.0 server also uses neural networks to predict O-linked glycosylation sites. NetOGlyc 4.0 server predicted 28, 30 and 28 O-linked glycosylation sites in AHFV, KFDV and TBEV, respectively. For the amino acid positions, see Figure 4.

The SignalP 4.1 server uses a combination of several artificial neural networks to predict signal peptides and cleavage sites. With the server's default input values, it predicted (based on C, S and Y -scores) one signal peptide and one cleavage site in each of the three viruses (Figure 4). When considering only the C-score (≥ 0.5), the SignalP 4.1 server predicted cleavage sites after amino acid positions 262, 281 (end of prM) and 2261 (end of 2 K) in both AHFV and KFDV. Additionally in KFDV, the server also predicted cleavage sites after positions 54 and 122. In TBEV, 9 cleavage sites (i.e. after amino acid positions 54, 73, 120, 280 (end of prM), 283, 399, 776 (end of E), 1402 and 2259 (end of 2 K)) were predicted when considering only the C-score. In both AHFV and KFDV, the cleavage site after the E protein (i.e. amino acid position 777) had a C-score of approximately 0.4.

Number of transmembrane helices is the same in AHFV, KFDV and TBEV

The TMHMM 2.0 server predicted 17 transmembrane helices in each virus (Supplementary Figure 1). The expected number of amino acids in these transmembrane helices, on the other hand, was different: AHFV with approximately 407.78, KFDV with approximately 404.71, and TBEV with 399.37.

Protein properties of AHFV, KFDV and TBEV are similar

The single letter code protein sequence of the polyprotein was given as input to the ExPASy's ProtParam tool. The following protein properties were predicted to be similar (Table 2): molecular weight, theoretical pI, number of positive and negative charged amino acids, extinction

Figure 2. Codon frequency analysis. (A) Comparison of codon frequency of AHFV with human and potential vectors. (B) Comparison of codon frequency of AHFV with KFDV and TBEV. (C) Codon frequency of AHFV sequences compared to Kazusa database. AHFV = *Alkhurma hemorrhagic fever virus*, KFDV = *Kyasanur forest disease virus*, TBEV = *tick-borne encephalitis virus*, HS = *Homo sapiens*, AA = *Aedes aegypti* and OS = *Ornithodoros savignyi*.

Figure 3. Substitution rates of AHFV protein coding genes isolated from humans and ticks. AHFV = Alkhurma hemorrhagic fever virus, K_a = non-synonymous substitution and K_s = synonymous substitution.

Table 2. Polyprotein properties of AHFV, KFDV and TBEV.

Parameter	AHFV	KFDV	TBEV
Accession number	NP_722551.1	AFF18434.1	NP_043135.1
No. of amino acids	3416	3416	3414
Molecular weight (Da)	378121.94	377460.73	378322.11
Theoretical pI	8.64	8.58	8.42
No. of negatively charged residues (Asp + Glu)	376	378	382
No. of positively charged residues (Arg + Lys)	407	406	402
Extinction coefficients ($M^{-1} cm^{-1}$) at 280 nm measured in water, Abs 0.1% (=1 g/l), assuming all pairs of Cys residues form cystines	656490 (Abs 1.736)	653510 (Abs 1.731)	673770 (Abs 1.781)
Extinction coefficients ($M^{-1} cm^{-1}$) at 280 nm measured in water, Abs 0.1% (=1 g/l), assuming all Cys residues are reduced	652240 (Abs 1.725)	649260 (Abs 1.720)	669770 (Abs 1.770)
Estimated half-life in mammalian reticulocytes (<i>in vitro</i>)	30 h	30 h	30 h
Estimated half-life in yeast (<i>in vivo</i>)	>20 h	>20 h	>20 h
Estimated half-life in <i>Escherichia coli</i> (<i>in vivo</i>)	>10 h	>10 h	>10 h
Instability index (II)	36.84 (stable)	35.48 (stable)	35.81 (stable)
Aliphatic index	87.69	86.21	86.54
Grand average of hydropathicity (GRAVY)	-0.148	-0.163	-0.152

AHFV = Alkhurma hemorrhagic fever virus, KFDV = Kyasanur forest disease virus and TBEV = tick-borne encephalitis virus.

coefficients, estimated half-lives in mammalian reticulocytes, yeast (*in vivo*) and *Escherichia coli* (*in vivo*), instability index (II), aliphatic index (AI) and grand average of hydropathicity (GRAVY) of AHFV, KFDV and TBEV.

The 5' UTRs have similar secondary structure while the 3' UTRs have different secondary structure in AHFV, KFDV and TBEV

The RNAfold web server was used to predict the secondary structure of the 5' and 3' UTRs. The server predicted

the 5' UTRs of AHFV, KFDV and TBEV to have typical 'Y' shaped secondary structures (Figure 5, top). In contrast, the server predicted the 3' UTRs to have variable secondary structures (Figure 5, bottom).

Inverted terminal repeat sequences are conserved in AHFV, KFDV and TBEV

The RactIP server was used to predict long-range RNA-RNA interactions. The 5' UTR and 3' UTR sequences were given as input in separate fields. The server predicted the

Figure 4. Prediction of glycosylation sites and signal peptide cleavage site in AHFV (accession number: NP_722551.1) (A), KFDV (accession number: AAQ91607.1) (B) and TBEV (accession number: NP_043135.1) (C). AHFV = *Alkhumra hemorrhagic fever virus*, KFDV = *Kyasanur forest disease virus* and TBEV = *tick-borne encephalitis virus*. Note: illustration not to scale.

following interactions: 5' GGAGAACAAG 3' in the 5' UTR with 5' CUUGUUCUCC 3' in the 3' UTR. These sequences are inverted repeats. We found these sequences conserved in AFHV (114–123 in the 5' UTR and 10695–10704 in the 3' UTR; accession number: JF416957.1), KFDV (114–123 in the 5' UTR and 10694–10703 in the 3' UTR; accession number: JF416958.1 or HM055369.1) and TBEV (115–124 in the 5' UTR and 11062–11071 in the 3' UTR;

accession number: NC_001672.1). This sequence has 50% GC content (total = 10 nucleotides). In all the three viruses, the repeat sequence in the 5' UTR was found in one of the hairpin loops. In the 3' UTR, on the other hand, the repeat sequence was predicted to be in one of the hairpin loops in KFDV (Figure 5, bottom), in one of the broken stems (bulges) in AFDV (Figure 5, bottom), and in one of the internal loops in TBEV (Figure 5, bottom).

Polymorphisms in the 5' and 3' UTR of AHFV isolates

Using the 24 AHFV isolates, we manually noted the polymorphic sites in the 5' and 3' UTRs. In the 5' UTR, 4 polymorphic sites were found. One polymorphism was due to insertion, while the other three polymorphisms were due to substitutions (Figure 6(A)). In the 3' UTR, sixteen polymorphic sites were found and all of them were substitutions (Figure 6(B)).

Polymorphisms in the CDS of AHFV sequences

For the CDS, we studied protein polymorphisms rather than nucleotide polymorphisms. In total, 112 polymorphisms were found. For the complete list of polymorphisms, see Supplementary Table 2. Protein wise, the percentage polymorphism was of the following order: 2K (0) < NS3 (1.93) < E (2.22) < NS2B (3.05) < NS4A (3.17) < NS5 (3.21) < NS4B (3.57) < NS1 (3.68) < NS2A (4.35) < C (5.31) < prM (7.14). The sequence with accession number JF416957.1 has the least number of polymorphisms, while the sequence with the accession number JX914664.1 has the highest number of polymorphisms when compared to the CDS of the reference sequence (i.e. accession number NC_004355.1/AF331718.1) (Figure 6(C)).

Homology modeling of protease and polymerase

Viral proteases and polymerases play a vital role in the replication cycle. By inhibiting these enzymes, the viral replication can be blocked. Until now, no 3D crystal structure of protease and polymerase is available for AHFV, KFDV and TBEV. Few groups have so far modeled TBEV protease and compared different strains [42]. We have modeled the 3D structure of NS3 protease and NS5 RNA-dependent RNA polymerase for AHFV, KFDV and TBEV by homology modeling using Phyre2. The models were energy-minimized using the ModRefiner server. The models were then validated by Ramachandran plot (Supplementary Table 3), PROVE (Supplementary Table 4), ERRAT (Supplementary Table 4), VERIFY_3D (Supplementary Table 4) and ProSA (Supplementary Table 4). When comparing the protease sequences, AHFV had a similarity of 99.4% with KFDV and 89.9% with TBEV, while KFDV had a similarity of 89.3% with TBEV (Figure 7, top). On the 3D structure level, the global root-mean-square deviation (RMSD) of AHFV with KFDV was 0.65 Å and with TBEV was 0.80 Å, while the global RMSD of KFDV with TBEV was 0.89 Å (Figure 7, bottom). Similarly, when comparing the polymerase sequence, AHFV had a similarity of 98.9% with KFDV and 93.1% with TBEV, and KFDV had a similarity of 92.8% with TBEV (Figure 8, top). On the 3D structure level, the global RMSD of AHFV with KFDV was 0.91 Å and with TBEV was 0.81 Å, while the global RMSD of KFDV with TBEV was 0.95 Å (Figure

8, bottom). The models were submitted to the protein model database (PMDb) (<https://bioinformatics.cineca.it/PMDB/>) and can be accessed by the public [43]. For the PMDB ID's, see Table 3.

Can AHFV and KFDV cause microcephaly?

An earlier study hypothesized that retinoic acid response element (RARE) consensus sequence (i.e. 5' AGGTCA 3') repeats in the Zika viral genome might play a role in the development of microcephaly [44]. This hypothesis is contradicted by the fact that unlike retroviruses or viruses of the *Hepadnaviridae* family, there is no intermediate DNA stage in the Flaviviral replication cycle. Therefore, the viral RNA cannot be integrated into the host DNA. In addition, reverse transcriptase and integrase are needed for integration of viral RNA into the host genome. These two enzymes are usually coded for in the viral genome (e.g. in retroviruses). The genes for these enzymes are not present in the Flaviviruses.

Nevertheless, we studied the RARE consensus sequence repeats in AHFV and KFDV. We manually scanned the whole viral genome for the RARE consensus sequence repeats. The number of RARE consensus sequence repeats in AHFV sequences varied between 4 and 7. In the KFDV sequences, on the other hand, we found 6 per isolate. The difference in the number of RARE consensus sequence repeats in AHFV and KFDV sequences may result from polymorphisms. The number of RARE consensus sequence repeats in AHFV and KFDV was almost equal to the number of RARE consensus sequence repeats in ZIKV, as described in the earlier study. These findings suggest that, unless experimentally proven, there is thus far no concrete link between RARE consensus sequence repeats and microcephaly.

Discussion

AHFV, as a fairly new member of the Flavivirus, requires detailed study in order to elucidate putative antiviral drugs. Such studies have high priority for several reasons: (1) the mortality rate from AHFV is the highest among all the Flaviviruses [10]; (2) the infected patients show both hemorrhagic and neurological symptoms, which can cause brain damage; (3) the virus is endemic in Saudi Arabia, especially areas adjoining the Mecca and Jeddah cities (Muslims throughout the world gather at Mecca and Jeddah every year for their Hajj. Therefore, there is a possibility for the virus to efficiently spread to other countries); and (4) the virus has the potential to be weaponized for biological warfare.

In the present study, using available genome and proteome data of AHFV from NCBI, we have compared the virus with KFDV and TBEV. Phylogenetic studies using the complete genome and/or CDS of AHFV strains showed that the AHFV strains can be classified further

Figure 5. (Top) Prediction of RNA secondary structure of 5' UTR in TBEV (accession number: NC_001672.1) (A), KFDV (accession number: JF416958.1) (B) and AHFV (accession number: JF416957.1) (C). (Bottom) Prediction of RNA secondary structure of 3' UTR in TBEV (accession number: NC_001672.1) (A), KFDV (accession number: JF416958.1) (B) and AHFV (accession number: JF416957.1) (C). The color is assigned according to the base pairing probabilities. Below the images: Mountain plot for the predicted secondary structure. Plateaus in the plot represent loops, peaks in the plot represent hairpin loops and slopes in the plot represent helices. *mfe* = minimum free energy structure, *pf* = partition function folding structure and *centroid* = centroid structure. AHFV = Alkhurma hemorrhagic fever virus, KFDV = Kyasanur forest disease virus and TBEV = tick-borne encephalitis virus.

Figure 6. (A) Nucleotide polymorphisms in 5' UTR of AHFV (accession number: JF416957.1). (B) Nucleotide polymorphisms in 3' UTR of AHFV (accession number: JF416957.1) (C) Number of protein polymorphisms in the coding sequence (CDS) of AHFV sequences. AHFV = *Alkhurma hemorrhagic fever virus*.

into sub-lineages [45]. The AHFV strains that have been sequenced so far all have high nucleotide identity among themselves and with KFDV and TBEV. A previous study has predicted that AHFV and KFDV diverged approximately 700 years ago [45]. AHFV and KFDV could be used for further evolutionary studies.

TBEV includes three subtypes. We used TBEV in our study, particularly the Neudoerfl strain (which causes neurological syndrome), instead of Omsk hemorrhagic fever virus (OHFV) which causes hemorrhagic symptoms, for comparison because we aimed to show how AHFV (which claimed to be tick-borne based on phylogenetic grouping with TBEV serocomplex) could be related to the classical tick-borne virus on other features. As

AHFV infections were found to cause neurological syndrome, we determined it to be the right choice for the comparison.

An earlier study on ZIKV has shown that there was a recombination in the NS2B gene with SPOV when pre-epidemic and epidemic strains were compared [46]. Following that study, we checked for recombination in the AHFV polyprotein. Using the ML, RDP, GENECONV, Chimaera, MaxChi, Bootscan, SiScan and 3Seq algorithms independently, we could find no recombination within the studied Flaviviruses. Limitations of computational support restricted us to studying only a few of the Flaviviruses in this study. From the SBP analysis, a breakpoint at nucleotide position 7399 (NS4B) was

Figure 7. (Top) Multiple sequence alignment (MSA) of protease sequence from AHFV, KFDV and TBEV. Amino acids forming the catalytic triad i.e. histidine 54 (H54), aspartic acid 78 (D78) and serine 138 (S138) are represented by upward or downward black triangle. Figure was generated using ESPript 3.0 (<http://esprict.ibcp.fr/ESPript/ESPript/>) [53]. (Bottom) Phyre2 modeled 3D protease structure of AHFV (A), KFDV (B) and TBEV (C). Amino acids forming the catalytic triad i.e. histidine 54 (H54), aspartic acid 78 (D78) and serine 138 (S138) are represented as orange sticks. AHFV = *Alkhurma hemorrhagic fever virus*, KFDV = *Kyasanur forest disease virus* and TBEV = *tick-borne encephalitis virus*.

predicted, but this was not supported by the GARD analysis. Different tools use different algorithms, which vary by sensitivity. However, this has to be verified with large datasets.

Codon usage analysis showed that the frequency of GGA (glycine codon) was more than two-fold higher in the AHFV CDS as compared to humans. It is well known that codon usage bias exists between different cell types even within the same organism [38–40]. Human codon usage data in the Kazusa database is not cell type-specific. It will be more insightful to study codon usage bias between AHFV with respect to different cell types in humans than using the average. Similar studies should also be performed for its vectors. While comparing the CDSs between AHFV and KFDV, only the codon usage of UUA (leucine) surpassed two-fold higher in AHFV compared to KFDV. This could be due to the difference in the GC content between these two viruses. KFDV has 1% higher GC content compared to AHFV in its CDS. Considering the CDS size (roughly 10000 nucleotides), this corresponds to roughly 100 more G/Cs in KFDV. It also suggests that CDS of KFDV has a higher melting temperature as compared to AHFV and TBEV. Codon frequency analysis between different AHFV sequences showed that there were no significant differences with each other. This might suggest that the strains have not been exposed to any artificial selection pressure or have not undergone any accelerated evolution (e.g. by antivirals).

Substitution rate studies showed that the polyprotein and individual protein gene sequences in AHFV, KFDV and TBEV were undergoing purifying or negative selection. Dodd et al. have predicted the absence of positive selection in AHFV, which is in support of our results [45]. On the contrary, Ruiz-Aymá et al. have predicted that the complete protein and the NS3 protein in AHFV have undergone positive selection [47]. Ruiz-Aymá et al. used Kimura 2-parameters instead of the Jukes-Cantor model (used in the present study) to predict the substitution rates [47]. Kimura 2-parameters model has few assumptions; all the four nucleotides have the same frequency and all the sites have the same substitution rates [48]. With these assumptions, Kimura 2-parameters model ranks the synonymous and non-synonymous substitution rates [48]. We also used Kimura 2-parameters to verify the positive selection in the complete protein and the NS3 protein sequences, but we found no positive selection. The occurrence of these contradicting results might be due to the difference in the number of sequences used. A general caveat for such type of studies is the number of sequences used. The substitution rates predicted in this study for KFDV has to be used with caution as only six sequences were analyzed. When comparing the substitution rates of AHFV isolated from humans and ticks, no significant differences were noticed except for the C, prM, NS2B and NS5 genes (>2-fold), which is significant. Again, the sample size is small and using

between these viruses. Certain conserved amino acid positions in all the three viruses were predicted to be glycosylated, however. We used highly stringent parameters in predicting these glycosylation sites using the servers. It should be noted, however, that occupancy of these

Table 3. Protein model database (PMDB) IDs for the Phyre2 modeled protease and RNA-dependent RNA polymerase.

S. No.	Virus	Protein	PMDB ID
1	AHFV	NS3 Protease	PM0081024
2	KFDV	NS3 Protease	PM0081025
3	TBEV	NS3 Protease	PM0081026
4	AHFV	NS5 RNA-dependent RNA Polymerase	PM0081027
5	KFDV	NS5 RNA-dependent RNA Polymerase	PM0081028
6	TBEV	NS5 RNA-dependent RNA Polymerase	PM0081029

AHFV = Alkhumra hemorrhagic fever virus, KFDV = Kyasanur forest disease virus and TBEV = tick-borne encephalitis virus.

glycosylation sites can vary depending upon the type of cell expressing the protein.

The RNAfold server predicted similar 5' UTR secondary structures for AHFV, KFDV and TBEV, while the 3' UTRs were predicted to have variable structures. The secondary structure of RNA can vary depending on the nucleotide composition, length and the environment. The intramolecular RNA-RNA interactions between the 5' UTR and the 3' UTR were predicted and seem to be conserved in all the three viruses with identical inverted repeat sequences. The presence and the role of these inverted repeat sequences (CS-A) in TBEV were experimentally validated in an earlier study [49]. In this study, it was also claimed that based on the location of this CS-A, the tick-borne viruses can be differentiated from the mosquito-borne viruses [49]. If we apply this theory to AHFV, it is predicted to belong to the tick-borne viruses rather than mosquito-borne viruses, albeit it is still possible that mosquitos can spread AHFV.

In a recent report from Ministry of Health in Saudi Arabia, it was claimed that the case fatality rate from AHFV infection was only 1.8%, which was much lower than the earlier published results (a poster from the Ministry of Health, Saudi Arabia). The difference is primarily related to the population studied. For e.g. Madani's study in 2005 [10] reported a high mortality (25%) but this was among hospitalized patients. Madani *et al.* in 2010 subsequently reported a mortality rate in Najran that was notably lower than that in the Makkah outbreak (1.3% vs. 25%) [11]. But this time the study population included both hospitalized and non-hospitalized patients. The notably low mortality observed in Najran outbreak may be secondary to an established endemicity of AHFV in this region with some protective immunity among humans developing as a result of frequent exposure to the virus. Nevertheless, the virus cannot be ignored as it has the potential to become pandemic. There is an urgent need for specific drugs(s) and/or vaccines to tackle this viral disease. Two earlier studies have modeled the AHFV envelope (E) protein and have predicted potential B-cell and T-cell epitopes [50,51]. Beyond that, no further progress has been made in developing a suitable vaccine against AHFV. Drugs (ligands) can be screened and/or developed to

inhibit AHFV enzymes, namely NS3 protease and NS5 polymerase, as these enzymes are vital for viral replication. A study has reported that 6-azauridine (6-azaU), 2'-C-methylcytidine (2'-CMC) and interferon alpha 2a (IFN- α 2a) inhibited the replication of some Flaviviruses, namely AHFV, KFDV, OHFV and Powassan virus [52]. The same study also identified that 6-azaU inhibited pyrimidine biosynthesis, while 2'-CMC inhibited the viral polymerase by binding to the active site [52]. Further studies are needed on these compounds before obtaining FDA approval. Using direct acting antivirals, viral infections like HIV and HCV (a member of *Flaviviridae*) have been successfully controlled, it is possible to cure completely the latter with existing drugs. The NS3 protease and NS5 polymerase protein models developed in this study will accelerate the *in silico* screening of ligands and also aid in the rational designing of ligands. There are several methods or servers available for homology protein modeling. We have used Phyre2, I_TASSER and SWISSMODEL in one of our previous studies on a different protein and we found values for the Phyre2 model to be better rooted in the Ramachandran plot, VERIFY3D, ERRAT etc. (unpublished). We therefore decided to use Phyre2 for this study as well. There is a range of well-known 3D structure predicting servers available each with their own unique algorithms. Because no single algorithm is perfect when it comes to predictions, we performed a few quality control studies (i.e. Ramachandran Plot, ERRAT, VERIFY3D, etc.), after generating the structures via Phyre2 and found the values to be satisfactory. The challenges regarding the usage of specific drugs include the development of resistance and existence of basal polymorphisms that can resist the drugs. These challenges can be overcome by prescribing a combination of drugs that target various processes in the viral replication cycle. The polymorphism data presented in our study could aid in developing better drugs that are not influenced by these basal polymorphisms.

To conclude, we show the molecular and biochemical similarities and differences between the closely related viruses AHFV, KFDV and TBEV using their genome sequences. We further confirm that the genome and proteome of these viruses are highly similar in aspects including e.g. architecture and post-translational modifications. The data obtained in this study will aid in better understanding AHFV, thereby speeding up the process of developing specific drugs, vaccines and management strategies against this virus.

Authors' contributions

N.P. conceived the idea, designed the study, performed the computational analysis and generated the results. N.P., D.A., J.L. and Å.L. wrote the manuscript together. All the authors read the manuscript and approved it.

Acknowledgements

The authors acknowledge the grammatical assistance provided by Sarah Lam and Jelle van der Hilst. N.P. would like to thank the people of the bioinformatics community for developing the wonderful tools and most importantly for making these tools freely accessible. N.P. would also like to thank Nathan Osman (McGill University AIDS Centre) for his help in running the codon frequency script.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Navaneethan Palanisamy <http://orcid.org/0000-0003-0369-2316>

References

- [1] Kuno G, Chang G-JJ, Tsuchiya KR, et al. Phylogeny of the genus *Flavivirus*. *J Virol*. 1998;72:73–83.
- [2] Zaki AM. Isolation of a flavivirus related to the tick-borne encephalitis complex from human cases in Saudi Arabia. *Trans R Soc Trop Med Hyg*. 1997;91:179–181.
- [3] Pijlman GP, Funk A, Kondratieva N, et al. A highly structured, nuclease-resistant, noncoding RNA produced by Flaviviruses is required for pathogenicity. *Cell Host Microb*. 2008;4:579–591.
- [4] Bessaud M, Grard G, Peyrefitte CN, et al. Identification and enzymatic characterization of NS2B-NS3 protease of Alkhurma virus, a class-4 flavivirus. *Virus Res*. 2005;107:57–62.
- [5] Pastorino BAM, Peyrefitte CN, Grandadam M, et al. Mutagenesis analysis of the NS2B determinants of the Alkhurma virus NS2B-NS3 protease activation. *J Gen Virol*. 2006;87:3279–3283.
- [6] Madani TA, Abuelzein E-TME, Azhar EI, et al. Thermal inactivation of Alkhurma hemorrhagic fever virus. *Arch Virol*. 2014;159:2687–2691.
- [7] Alzahrani AG, Al Shaiban HM, Al Mazroa MA, et al. Alkhurma hemorrhagic fever in humans, Najran, Saudi Arabia. *Emerg Infect Dis*. 2010;16:1882–1888.
- [8] Charrel RN, Fagbo S, Moureau G, et al. Alkhurma hemorrhagic fever virus in *Ornithodoros savignyi* ticks. *Emerg Infect Dis*. 2007;13:153–155.
- [9] Mahdi M, Erickson BR, Comer JA, et al. Kyasanur forest disease virus Alkhurma subtype in ticks, Najran Province, Saudi Arabia. *Emerg Infect Dis*. 2011;17:945–947.
- [10] Madani TA. Alkhurma virus infection, a new viral hemorrhagic fever in Saudi Arabia. *J Infect*. 2005;51:91–97.
- [11] Madani TA, Azhar EI, Abuelzein E-TME, et al. Alkhurma (Alkhurma) virus outbreak in Najran, Saudi Arabia: epidemiological, clinical, and laboratory characteristics. *J Infect*. 2011;62:67–76.
- [12] Charrel RN, Lamballerie X de, Zaki AM. Human cases of hemorrhagic fever in Saudi Arabia due to a newly discovered flavivirus, Alkhurma hemorrhagic fever virus. In: Lu, Y, Essex MB, editors. *Emerging infections in Asia*. Springer US; 2008 [cited 2017 Jun 11]. p. 179–192. Available from: http://link.springer.com/chapter/10.1007/978-0-387-75722-3_11.
- [13] Memish ZA, Albarrak A, Almazroa MA, et al. Seroprevalence of Alkhurma and other hemorrhagic fever viruses, Saudi Arabia. *Emerg Infect Dis*. 2011;17:2316–2318.
- [14] Memish ZA, Fagbo SF, Osman Ali A, et al. Is the epidemiology of Alkhurma hemorrhagic fever changing? A three-year overview in Saudi Arabia. *PLoS ONE*. 2014;9:e85564.
- [15] Horton KC, Fahmy NT, Watany N, et al. Crimean Congo hemorrhagic fever virus and Alkhurma (Alkhurma) virus in ticks in Djibouti. *Vector Borne Zoonotic Dis*. Larchmt. N. 2016;16:680–682.
- [16] Tamura K, Stecher G, Peterson D, et al. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*. 2013;30:2725–2729.
- [17] Campanella JJ, Bitincka L, Smalley J. MatGAT: an application that generates similarity/identity matrices using protein or DNA sequences. *BMC Bioinformatics*. 2003;4:29.
- [18] Martin DP, Murrell B, Golden M, et al. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol*. [Internet]. 2015 [cited 2017 Jun 11];1. Available from: <https://academic.oup.com/ve/article/1/1/vev003/2568683/RDP4-Detection-and-analysis-of-recombination>
- [19] Lemey P, Posada D. Recombination. *Phylogenetic Handbook: Practical Approach*. Phylogenetic Anal. Hypothesis Test. [Internet]. 2nd ed. Cambridge: Cambridge University Press; 2009 [cited 2017 Jun 11]. p. 491–549. Available from: <http://www.cambridge.org/catalogue/catalogue.asp?isbn=9780521877107&ss=toc>
- [20] Delport W, Poon AFY, Frost SDW, et al. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinforma Oxf Engl*. 2010;26:2455–2457.
- [21] Palanisamy N, Osman N, Ohnona F, et al. Does antiretroviral treatment change HIV-1 codon usage patterns in its genes: a preliminary bioinformatics study. *AIDS Res Ther*. [Internet]. 2017;14. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5237184/>
- [22] Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol*. 1986;3:418–426.
- [23] Suzuki Y, Gojobori T. A method for detecting positive selection at single amino acid sites. *Mol Biol Evol*. 1999;16:1315–1328.
- [24] Kosakovsky Pond SL, Frost SDW. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol*. 2005;22:1208–1222.
- [25] Hansen JE, Lund O, Tolstrup N, et al. NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. *Glycoconj J*. 1998;15:115–130.
- [26] Petersen TN, Brunak S, von Heijne G, et al. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods*. 2011;8:785–786.
- [27] Gruber AR, Lorenz R, Bernhart SH, et al. The Vienna RNA Websuite. *Nucleic Acids Res*. 2008;36:W70–W74.
- [28] Kato Y, Sato K, Hamada M, et al. RactIP: fast and accurate prediction of RNA-RNA interaction using integer programming. *Bioinforma Oxf Engl*. 2010;26:i460–i466.
- [29] Kelley LA, Mezulis S, Yates CM, et al. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*. 2015;10:845–858.
- [30] Xu D, Zhang Y. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophys J*. 2011;101:2525–2534.
- [31] Laskowski RA. PDBsum: summaries and analyses of PDB structures. *Nucleic Acids Res*. 2001;29:221–222.

- [32] Bowie JU, Lüthy R, Eisenberg D. A method to identify protein sequences that fold into a known three-dimensional structure. *Science*. 1991;253:164–170.
- [33] Lüthy R, Bowie JU, Eisenberg D. Assessment of protein models with three-dimensional profiles. *Nature*. 1992;356:83–85.
- [34] Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci Publ Protein Soc*. 1993;2:1511–1519.
- [35] Pontius J, Richelle J, Wodak SJ. Deviations from standard atomic volumes as a quality measure for protein crystal structures. *J Mol Biol*. 1996;264:121–136.
- [36] Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res*. 2007;35:W407–W410.
- [37] Gelly J-C, Joseph AP, Srinivasan N, et al. iPBA: a tool for protein structure comparison using sequence alignment strategies. *Nucleic Acids Res*. 2011;39:W18–W23.
- [38] Andersson SG, Kurland CG. Codon preferences in free-living microorganisms. *Microbiol Rev*. 1990;54:198–210.
- [39] Grantham R, Gautier C, Gouy M, et al. Codon catalog usage and the genome hypothesis. *Nucleic Acids Res*. 1980;8:r49–r62.
- [40] Hershberg R, Petrov DA. Selection on codon bias. *Annu Rev Genet*. 2008;42:287–299.
- [41] Li M, Kao E, Gao X, et al. Codon-usage-based inhibition of HIV protein synthesis by human schlafen 11. *Nature*. 2012;491:125–128.
- [42] Belikov SI, Kondratov IG, Potapova UV, et al. The relationship between the structure of the tick-borne encephalitis virus strains and their pathogenic properties. *PLoS ONE*. 2014;9:e94946.
- [43] Castrignanò T, Meo D, D'Onorio P, et al. The PMDB protein model database. *Nucleic Acids Res*. 2006;34:D306–D309.
- [44] Kumar A, Singh HN, Pareek V, et al. A possible mechanism of Zika virus associated microcephaly: imperative role of retinoic acid response element (RARE) consensus sequence repeats in the viral genome. *Front Hum Neurosci*. 2016;10:403.
- [45] Dodd KA, Bird BH, Khristova ML, et al. Ancient ancestry of KFDV and AHFV revealed by complete genome analyses of viruses isolated from ticks and mammalian hosts. *PLoS Negl Trop Dis*. 2011;5:e1352.
- [46] Zhu Z, Chan JF-W, Tee K-M, et al. Comparative genomic analysis of pre-epidemic and epidemic Zika virus strains for virological factors potentially associated with the rapidly expanding epidemic. *Emerg Microbes Infect*. 2016;5:e22.
- [47] Ruiz-Aymá G, Tovar-Herrera OE, González-Alvarez R, et al. Mining the Alkhurma hemorrhagic fever virus genomes in 2015. *Southwest Entomol*. 2016;41:87–97.
- [48] Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol*. 1980;16:111–120.
- [49] Kofler RM, Hoenninger VM, Thurner C, et al. Functional analysis of the tick-borne encephalitis virus cyclization elements indicates major differences between mosquito-borne and tick-borne flaviviruses. *J Virol*. 2006;80:4099–4113.
- [50] Mohabatkar H. Computer-based comparison of structural features of envelope protein of Alkhurma hemorrhagic fever virus with the homologous proteins of two closest viruses. *Protein Pept Lett*. 2011;18:559–567.
- [51] Poorinmohammad N, Mohabatkar H. Homology modeling and conformational epitope prediction of envelope protein of Alkhurma haemorrhagic fever virus. *J Arthropod-Borne Dis*. 2014;9:116–124.
- [52] Flint M, McMullan LK, Dodd KA, et al. Inhibitors of the tick-borne, hemorrhagic fever-associated flaviviruses. *Antimicrob Agents Chemother*. 2014;58:3206–3216.
- [53] Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res*. 2014;42:W320–W324.