

Short report

Open Access

The complete genome sequence of a Crimean-Congo Hemorrhagic Fever virus isolated from an endemic region in Kosovo

Darja Duh¹, Stuart T Nichol², Marina L Khristova², Ana Saksida¹, Iva Hafner-Bratkovič³, Miroslav Petrovec¹, Iusuf Dedushaj⁴, Salih Ahmeti⁵ and Tatjana Avšič-Županc*¹

Address: ¹Institute of Microbiology and Immunology, Medical Faculty, Ljubljana, Slovenia, ²Special Pathogens Branch and Biotechnology Core Facility Branch, Centers for Disease Control and Prevention, Atlanta, Georgia, USA, ³National Institute of Chemistry, Ljubljana, Slovenia, ⁴National Institute of Public Health, Pristina, Kosovo and ⁵Clinic of Infectious diseases, Pristina, Kosovo

Email: Darja Duh - darja.duh@mf.uni-lj.si; Stuart T Nichol - stn1@cdc.gov; Marina L Khristova - mik3@cdc.gov; Ana Saksida - ana.saksida@mf.uni-lj.si; Iva Hafner-Bratkovič - iva.hafner@ki.si; Miroslav Petrovec - mirc.petrovec@mf.uni-lj.si; Iusuf Dedushaj - isufdedushaj@hotmail.com; Salih Ahmeti - salih_ahmeti@hotmail.com; Tatjana Avšič-Županc* - tatjana.avsic@mf.uni-lj.si

* Corresponding author

Published: 15 January 2008

Received: 11 December 2007

Virology Journal 2008, **5**:7 doi:10.1186/1743-422X-5-7

Accepted: 15 January 2008

This article is available from: <http://www.virologyj.com/content/5/1/7>

© 2008 Duh et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

The Balkan region and Kosovo in particular, is a well-known Crimean-Congo hemorrhagic fever (CCHF) endemic region, with frequent epidemic outbreaks and sporadic cases occurring with a hospitalized case fatality of approximately 30%. Recent analysis of complete genome sequences of diverse CCHF virus strains showed that the genome plasticity of the virus is surprisingly high for an arthropod-borne virus. High levels of nucleotide and amino acid differences, frequent RNA segment reassortment and even RNA recombination have been recently described. This diversity illustrates the need to determine the complete genome sequence of CCHF virus representatives of all geographically distinct endemic areas, particularly in light of the high pathogenicity of the virus and its listing as a potential bioterrorism threat. Here we describe the first complete CCHF virus genome sequence of a virus (strain Kosova Hoti) isolated from a hemorrhagic fever case in the Balkans. This virus strain was isolated from a fatal CCHF case, and passaged only twice on Vero E6 cells prior to sequence analysis. The virus total genome was found to be 19.2 kb in length, consisting of a 1672 nucleotide (nt) S segment, a 5364 nt M segment and a 12150 nt L segment. Phylogenetic analysis of CCHF virus complete genomes placed the Kosova Hoti strain in the Europe/Turkey group, with highest similarity seen with Russian isolates. The virus M segments are the most diverse with up to 31 and 27% differences seen at the nt and amino acid levels, and even 1.9% amino acid difference found between the Kosova Hoti and another strain from Kosovo (9553-01). This suggests that distinct virus strains can coexist in highly endemic areas.

Findings

Bioinformatics analysis of complete microbial genomes has led to advances in the development of novel diagnostic techniques, in the research of microbial pathogenesis,

and in the control and prevention of infectious diseases. Until the year 2006, only 2 complete genomes of Crimean-Congo hemorrhagic fever virus (CCHFV) had been sequenced [1]. CCHFV, is a tick-borne virus with tri-

partite RNA genome (S, M and L segment), and is the causative agent of a lethal zoonosis named Crimean-Congo hemorrhagic fever (CCHF). The virus is distributed over much of Asia, extending from China to the Middle East and Southern Russia and to the focal endemic areas in Africa and southern Europe, including Kosovo and Turkey [2]. Yearly epidemics, as well as sporadic cases of CCHF are seen in some of these areas, often with high case fatality (approx. 30%) [3]. CCHFV can be transmitted to humans by bites of *Ixodid* ticks and by the contact with blood or tissue from viremic livestock and human patients [2]. Development of diagnostic approaches and potential vaccines is dependent on knowledge of the broad geographic distribution of diverse virus variants and on understanding of the extent of virus genetic reassortment and recombination [3,4]. The analysis of the 16 existing complete CCHFV genomes up to date indicated considerable evolution and high diversity of CCHFV [1,5]. Presumably this reflects the typical high polymerase error rates seen with negative stranded RNA viruses. In addition, previous reports have found evidence of RNA segment reassortment events between CCHFV M segments, and the recombination in CCHFV S segments [1,3,4]. The genetic diversity of CCHFV, its virulence, and its potential as a bioterrorism agent, make it important to obtain the complete genome of CCHFV from all geographically distinct endemic areas.

The Balkan peninsula, and Kosovo in particular, is a well-known endemic region for CCHF, and epidemic outbreaks and sporadic cases have been frequently been recorded [6-8]. Five nucleotide sequences of CCHFV from Kosovo have been published [9-12]. Three of them are partial sequences of S segment, the remaining 2 represent complete sequences of S and M segment of different CCHFV strains, Kosova Hoti and Kosovo 9553-01, respectively. We describe the first complete CCHFV genome sequence of a virus (strain Kosova Hoti) isolated from a hemorrhagic fever case in the Balkans.

The CCHFV Kosova Hoti strain was isolated from a blood of a female fatal case during the epidemic in Kosovo in 2001 [6]. The blood was taken on the 5th day after onset of symptoms. Results of the laboratory analysis showed the presence of IgM antibodies (titer 1:400) and the presence of viral RNA in the concentration of 1.08×10^{10} copies per mL of serum. Virus was grown on Vero E6 cells in BSL-3 laboratory. Viral RNA was extracted with the Trizol reagent from the second passage of the CCHFV in Vero E6 cells, and used for the direct sequencing of the complete genome of the virus. Amplicons of S, M and L full length segments were obtained by following the protocols described previously [1,9,13]. Briefly, a total of 16 S, 40 M and 84 L sequencing primers were used to generate the complete sequence of the S, M and L segments and these

are deposited in the GenBank under the accession numbers [DQ133507](#), [EU037902](#) and [EU044832](#), respectively. Sequence alignment of CCHFV Kosova Hoti strain complete genome with preexisting CCHFV genomes was performed using the CLUSTAL W algorithm of MegAlign module (Lasergene 1999, DNASTAR, USA). Phylogenetic relationships of different CCHFV strains were established with a software package TREECON [14]. The phylogenetic tree was constructed by the neighbor-joining method. The topology of the tree was obtained with the Kimura 80 model and support for the tree nodes was calculated with 500 bootstrap replicates. SignalIP was used to predict the signal sequence cleavage site and TMHMM 2.0 was used to predict transmembrane helices of M segment [15,16]. The amino acid (aa) sequence of L segment was subjected to the PSI-BLAST and PredictProtein server for search of conserved aa motifs [17,18].

The genome size of CCHFV, strain Kosova Hoti, was found to be approximately 19.2 kb in length, consisting of a 1672 nucleotide (nt) S segment, a 5364 nt M segment and a 12150 nt L segment. The open reading frame (ORF) of S segment is 1449 nt in length, encoding a 482 aa (nt position 56 - 1504) nucleocapsid protein. The ORF lengths of M and L segments are 5067 nt/1688 aa (nt position 78-5144) and 11838 nt/3945 aa (nt position 78-11915), respectively.

Three phylogenetic trees were constructed based on the ORF sequences of S, M and L segments of CCHFV (Fig. 1). The general topologies of the trees were consistent with those described previously [1,13]. Seven distinct groups were formed representing the approximate geographic distribution of CCHFV. Based on the analysis of S, M and L segment of Kosova Hoti CCHFV, this strain clustered in group V., which represents the Europe/Turkey geographic lineage [1]. The position of Kosova Hoti strain within group V. was similar in the S and L segment tree (Fig. 1, panels A and C), where it formed a separate lineage ancestral to the three Russian isolates and CCHFV strain 200310849 from Turkey was the most ancestral member of V. group. The group V. topology based on the M segment, was a little different (Fig. 1, panel B). Two Russian strains (VLV-100 and Kashmanov) clustered together with the Turkish CCHFV whereas the Russian Drosdov strain clustered together with both CCHFV strains from Kosovo.

The sequence differences between the CCHFV strains in the group V. are shown in Tables 1, 2, 3. Significant difference was noted between the nt (ORF) and aa sequences of S and L segments, in comparison to the M segment. The majority of nt changes in the S and L segments were synonymous (not amino acid changing) (Tables 1, 3), whereas over 80% of M segment nt changes were non-synonymous (amino acid changing) (Table 2). As seen in ear-

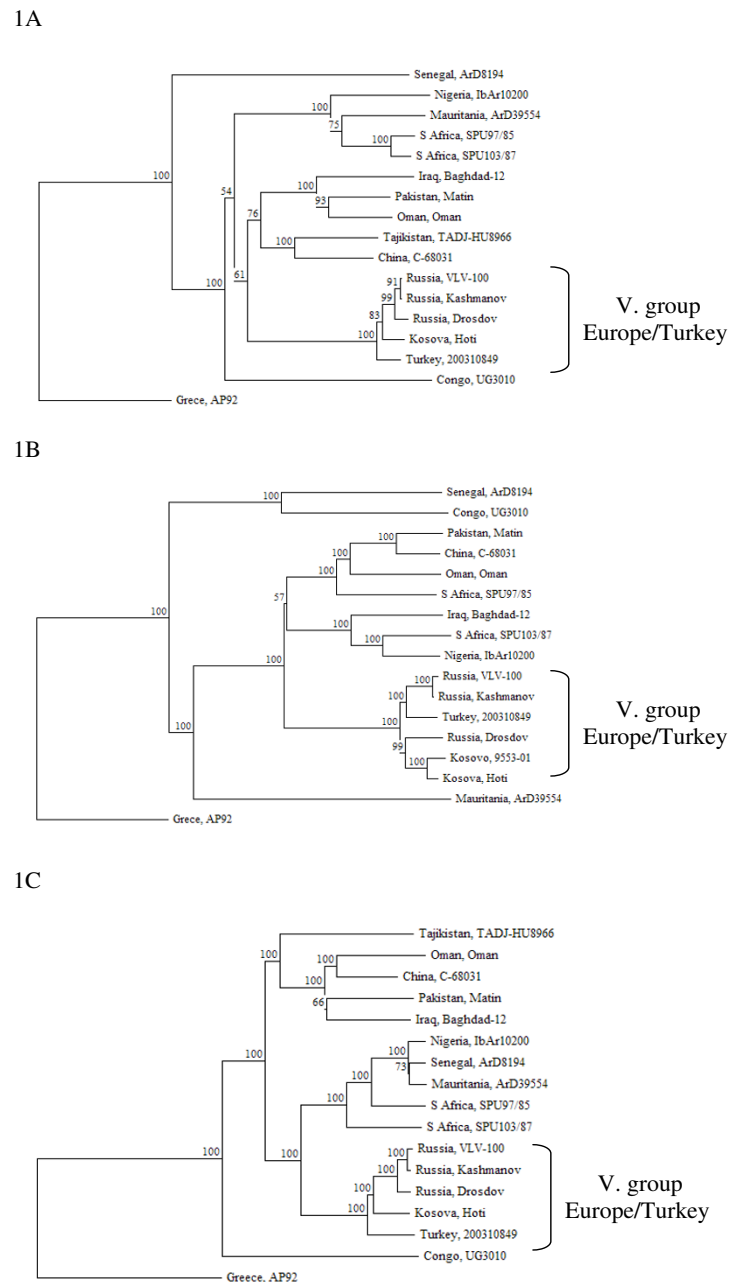


Figure 1

The phylogenetic analysis of the complete genome of CCHFV Kosova Hoti strain. A. Phylogenetic tree based on the alignment of S segment. B. Phylogenetic tree based on the alignment of M segment. C. Phylogenetic tree based on the alignment of L segment. CCHFV strains are presented by the country of origin and the name of the strain. Accession numbers of CCHFV used for the alignment are: Greece AP92 [S segment: GenBank:DO211638, M segment: DO211625, L segment: 211612], Congo UG3010 [DO211650, DO211637, DO211624], Senegal ArD8194 [DO211639, DO211626, DO211613], Nigeria IbaAr10200 [U88410, AF467768, AY389508], Mauritania ArD39554 [DO211641, DO211628, DO211615], S Africa SPU97/85 [DO211646, DO211633, DO211620], S Africa SPU103/85 [DO211647, DO211634, DO211621], Iraq Baghdad-12 [AJ538196, AJ538197, AY947890], Pakistan Matin [AF527810, AF467769, AY422208], Oman Oman [DO211645, DO211632, DO211619], Tajikistan TADJ-HU8966 [AY049083, AY179962, AY720893], China C-68031 [DO211642, DO211629, DO211616], Russia Drosdov [DO211643, DO211630, DO211617], Russia Kashmanov [DO211644, DO211631, DO211618], Russia VLV-100 [DQ206447, DQ206448, AY995166], Turkey 200310849 [DQ211649, DQ211636, DQ211623], Kosovo 9553-01 [M segment: AY675511].

Table 1: The difference between complete S segment of CCHFV strain Kosova Hoti and other strains in the V. group (Europe/Turkey) calculated by the MegAlign module.

S segment, difference (%)	Kosova Hoti				
	CCHFV strain	nt sequence (complete)	nt sequence (ORF)	aa sequence	non-synonymous mutations (%)
Kosovo, 9553-01	NA	NA	NA	NA	NA
Turkey, 2003I0849	2.1	2.1	0.0	0	0
Russia, Drosdov	2.6	2.6	0.6	23	23
Russia, Kashmanov	2.1	2.1	0.4	19	19
Russia, VLV-100	2.0	2.0	0.2	10	10

CCHFV strains are presented by the country of origin and the name of the strain. Accession numbers are shown in the legend of Figure 1A. nt – nucleotide; aa – amino acid, ORF – open reading frame, NA – S sequence of Kosovo 9553-01 strain not available.

Table 2: The difference between complete M segment of CCHFV strain Kosova Hoti and other strains in the V. group (Europe/Turkey) calculated by the MegAlign module. Table includes a separate column for the Mucin-like variable region present in M segment.

M segment, difference (%)	Kosova Hoti				
	CCHFV strain	nt sequence (complete)	nt sequence (ORF)	aa sequence	non-synonymous mutations (%)
Kosovo, 9553-01	2.3	2.2	1.9	86	4.5
Turkey, 2003I0849	5.7	5.6	4.8	85	20.5
Russia, Drosdov	5.2	5.1	4.3	84	15.2
Russia, Kashmanov	5.4	5.2	4.5	86	19.6
Russia, VLV-100	5.7	5.4	4.5	83	20.1

CCHFV strains are presented by the country of origin and the name of the strain. Accession numbers are shown in the legend of Figure 1B. nt – nucleotide; aa – amino acid, ORF – open reading frame.

Table 3: The difference between complete L segment of CCHFV strain Kosova Hoti and other strains in the V. group (Europe/Turkey) calculated by the MegAlign module.

L segment, difference (%)	Kosova Hoti				
	CCHFV strain	nt sequence (complete)	nt sequence (ORF)	aa sequence	non-synonymous mutations (%)
Kosovo, 9553-01	NA	NA	NA	NA	NA
Turkey, 2003I0849	4.4	4.2	1.2	28	28
Russia, Drosdov	3.6	3.4	1.1	32	32
Russia, Kashmanov	3.5	3.4	0.9	26	26
Russia, VLV-100	3.7	3.5	1.0	28	28

CCHFV strains are presented by the country of origin and the name of the strain. Accession numbers are shown in the legend of Figure 1C. nt – nucleotide; aa – amino acid, ORF – open reading frame, NA – L sequence of Kosovo 9553-01 strain not available.

lier studies [12,19,20], considerable glycoprotein amino acid variation was observed, particularly in the mucin-like variable region (Fig. 2), and presumably reflects the biological function of the glycoproteins encoded by the M segment. It is somewhat surprising that the glycoproteins of Kosova Hoti and another strain from Kosovo, 9553-01, differed by 1.9% in complete aa sequence, and up to 4.5% in the mucin-like domain (Table 2). This suggests different genetic strains of CCHFV co-exist in this highly endemic region.

The analysis of the Kosova Hoti strain M segment encoded polyprotein predicted the cleavage of the signal peptide to occur between aa 27 and 28 (AHG-QS). This site is identical to those described for Kosovo 9553-01 and Kashmanov but differs from other strains in group V. (Fig. 2). The mucin-like variable region of Kosova Hoti strain polyprotein stretches from aa 28 to 251 and differs by up to 20.5% from Turkish 200310849 strain (Table 2). Tetrapeptides RSKR251, RKLL523 and RKPL1043 were identified in Kosova Hoti and are identical among all strains in V. group. They represent the cleavage sites for GP38, Gn and Gc proteins, respectively [21,22]. The RKLL523 tetrapeptid of Kosova Hoti is typical for all strains in group V (Europe/Turkey) but it differs from RRLI tetrapeptid in all other CCHFV strains sequenced. However, both tetrapeptides constitute a cleavage recognition site for subtilase SKI-1 [12,22,23]. Five transmembrane helices were predicted for polyprotein of Kosova Hoti as shown on Figure 2.

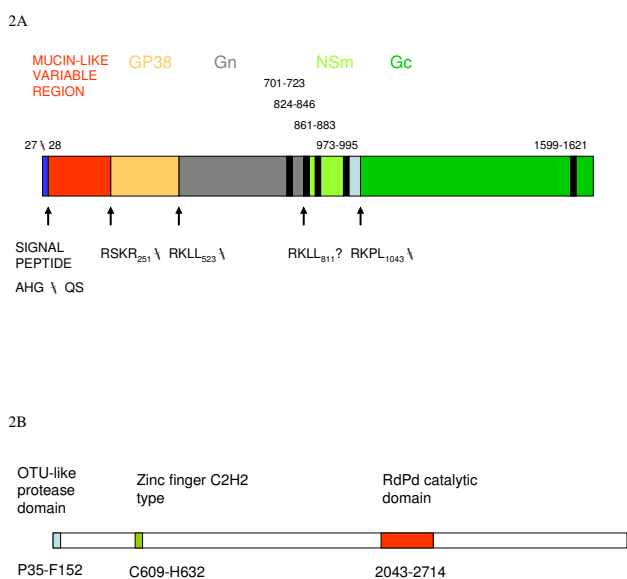


Figure 2
The protein analysis of the complete genome of CCHFV Kosova Hoti strain. A. Scheme of the M polyprotein of Kosova Hoti. B. Scheme of the L protein of Kosova Hoti strain.

Analysis of L protein encoded by the L segment of the Kosova Hoti strain revealed the conserved OTU-like protease domain from aa 35 to 152 (Fig. 2). The identified sequence G₃₇**DGN**₄₀CFYHSIAE.....₁₅₁**HFD** with the catalytic triad (indicated in **bold**) was identical among all CCHFV strains used in the L segment alignment (Fig. 1, panel C). Amino-acids 2043–2714 corresponded to the RNA-dependent RNA polymerase catalytic domain, similarly to the Nigerian IbaR10200 strain [24]. In addition, a zinc finger C2H2-type domain (aa 609–632) was found in the L protein of Kosova Hoti, but a previously identified leucine zipper could not be predicted. A leucine zipper motif (composed of three heptads) previously identified at aa 1386–1407 in the L sequence of a Nigerian strain [24,25], was not identified in the Kosova Hoti L sequence. However, the L sequence of Kosova Hoti (and other strains from group V) in this region differs from the Nigerian strain only in the substitution of the leucine for isoleucine at the position 1386.

Frequently it is observed that arthropod-borne viruses of vertebrates exhibit low genetic diversity which is thought to be due to essentially a double filter in operation, whereby evolution of these viruses is tightly constrained by the need to maintain high fitness in both vertebrate and arthropod host environments [26]. The very high genetic diversity seen in CCHFV is a strikingly exception. Presumably less constraint or greater positive selection is molding the evolutionary pattern of this virus. The complete genome of this representative CCHFV isolate (Kosova Hoti) from a highly endemic region of the Balkans is clearly divergent from strains present in other endemic regions of the world, and considerable sequence difference is even observed among virus strains found within Kosovo. These findings have importance for design of molecular diagnostic tools and vaccine development efforts, as they clearly illustrate the need to consider the high viral diversity and complexity of CCHF viral variant geographic distribution in these efforts.

Competing interests

The author(s) declare that they have no competing interests.

Authors' contributions

DD performed RNA extraction, qualitative and quantitative RT-PCR, analyzed the data and prepared the draft manuscript. MK and STN provided the complete M and L segment sequences and revised the draft manuscript. AS sequenced the complete S segment. IHB performed the protein analysis. MP, ID and SA collected the samples and clinical data. TAZ isolated the virus, supervised the study and revised the final draft. All authors read and approved the final manuscript.

Acknowledgements

We thank Mateja Jelovšek for the serological testing and Miša Korva for the alignments. This work was supported by RiViGene (Contract No. SSPE-CT-2005-022639).

References

- Deyde VM, Khristova ML, Rollin PE, Ksiazek TG, Nichol ST: **Crimean-Congo hemorrhagic fever virus genomics and global diversity.** *J Virol* 2006, **80**:8834-8842.
- Ergonul O: **Crimean-Congo haemorrhagic fever.** *Lancet Infect Dis* 2006, **6**:203-214.
- Hewson R, Gmyl A, Gmyl L, Smirnova SE, Karganova G, Jamil B, Hasan R, Chamberlain J, Clegg C: **Evidence of segment reassortment in Crimean-Congo haemorrhagic fever virus.** *J Gen Virol* 2004, **85**:3059-3070.
- Lukashev AN: **Evidence for recombination in Crimean-Congo hemorrhagic fever virus.** *J Gen Virol* 2005, **86**:2333-2338.
- Meissner JD, Seregin SS, Seregin SV, Vyshemirskii OI, Yakimenko NV, Netesov SV, Petrov VS: **The complete genomic sequence of strain ROS/HUVLV-100, a representative Russian Crimean Congo hemorrhagic fever virus strain.** *Virus Genes* 2006, **33**:87-93.
- Avšič-Zupanc T: **Epidemiology of Crimean-Congo Hemorrhagic Fever in the Balkans.** In *Crimean-Congo Hemorrhagic Fever: A Global Perspective* Edited by: Ergonul OWCA. Dordrecht, Springer; 2007:75-88.
- Ahmeti S, Raka L: **Crimean-Congo haemorrhagic fever in Kosovo: a fatal case report.** *Virology* 2006, **3**:85.
- Vesenjak-Hirjan J, Punda-Polic V, Dobe M: **Geographical distribution of arboviruses in Yugoslavia.** *J Hyg Epidemiol Microbiol Immunol* 1991, **35**:129-140.
- Duh D, Saksida A, Petrovec M, Dedushaj I, Avsic-Zupanc T: **Novel one-step real-time RT-PCR assay for rapid and specific diagnosis of Crimean-Congo hemorrhagic fever encountered in the Balkans.** *J Virol Methods* 2006, **133**:175-179.
- Drosten C, Minnak D, Emmerich P, Schmitz H, Reinicke T: **Crimean-Congo hemorrhagic fever in Kosovo.** *J Clin Microbiol* 2002, **40**:1122-1123.
- Papa A, Bozovi B, Pavlidou V, Papadimitriou E, Pelemis M, Antoniadis A: **Genetic detection and isolation of crimean-congo hemorrhagic fever virus, Kosovo, Yugoslavia.** *Emerg Infect Dis* 2002, **8**:852-854.
- Papa A, Papadimitriou E, Bozovic B, Antoniadis A: **Genetic characterization of the M RNA segment of a Balkan Crimean-Congo hemorrhagic fever virus strain.** *J Med Virol* 2005, **75**:466-469.
- Hewson R, Chamberlain J, Mioulet V, Lloyd G, Jamil B, Hasan R, Gmyl A, Gmyl L, Smirnova SE, Lukashev A, Karganova G, Clegg C: **Crimean-Congo haemorrhagic fever virus: sequence analysis of the small RNA segments from a collection of viruses world wide.** *Virus Res* 2004, **102**:185-189.
- Van de Peer Y, De Wachter R: **TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment.** *Comput Appl Biosci* 1994, **10**:569-570.
- Sonnhammer EL, von Heijne G, Krogh A: **A hidden Markov model for predicting transmembrane helices in protein sequences.** *Proc Int Conf Intell Syst Mol Biol* 1998, **6**:175-182.
- Emanuelsson O, Brunak S, von Heijne G, Nielsen H: **Locating proteins in the cell using TargetP, SignalP and related tools.** *Nat Protoc* 2007, **2**:953-971.
- Rost B, Yachdav G, Liu J: **The PredictProtein server.** *Nucleic Acids Res* 2004, **32**:W321-6.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**:3389-3402.
- Bergeron E, Vincent MJ, Nichol ST: **Crimean Congo hemorrhagic fever virus glycoprotein processing by the endoprotease SKI-I/SIP is critical for virus infectivity.** *J Virol* 2007.
- Erickson BR, Deyde V, Sanchez AJ, Vincent MJ, Nichol ST: **N-linked glycosylation of Gn (but not Gc) is important for Crimean Congo hemorrhagic fever virus glycoprotein localization and transport.** *Virology* 2007, **361**:348-355.
- Sanchez AJ, Vincent MJ, Erickson BR, Nichol ST: **Crimean-congo hemorrhagic fever virus glycoprotein precursor is cleaved by Furin-like and SKI-I proteases to generate a novel 38-kilodalton glycoprotein.** *J Virol* 2006, **80**:514-525.
- Sanchez AJ, Vincent MJ, Nichol ST: **Characterization of the glycoproteins of Crimean-Congo hemorrhagic fever virus.** *J Virol* 2002, **76**:7263-7275.
- Vincent MJ, Sanchez AJ, Erickson BR, Basak A, Chretien M, Seidah NG, Nichol ST: **Crimean-Congo hemorrhagic fever virus glycoprotein proteolytic processing by subtilase SKI-I.** *J Virol* 2003, **77**:8640-8649.
- Honig JE, Osborne JC, Nichol ST: **Crimean-Congo hemorrhagic fever virus genome L RNA segment and encoded protein.** *Virology* 2004, **321**:29-35.
- Kinsella E, Martin SG, Grolla A, Czub M, Feldmann H, Flick R: **Sequence determination of the Crimean-Congo hemorrhagic fever virus L segment.** *Virology* 2004, **321**:23-28.
- Weaver SC: **Evolutionary influences in arboviral disease.** *Curr Top Microbiol Immunol* 2006, **299**:285-314.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

