

Bioinformatics analysis of the recent MERS-CoV with special reference to the virus-encoded spike protein

Mahmoud Kandeel^{1, 2*}

¹ Department of Pharmacology, Faculty of Veterinary Medicine, Kafrelshikh University, Kafrelshikh 33516, Egypt.

² United Graduate School of Drug Discovery and Medical Information Sciences, Gifu University, Yanagido 1-1, Gifu 501-1193, Japan.

Corresponding author:

 mahmoud.kandeel@vet.kfs.edu.eg

Abstract

Coronaviruses (CoVs) are characterized by high recombination frequencies, resulting in sudden outbreak of newly evolved viruses with different pathogenicity, tissue tropism and high genome sequence variability. Recently, an outbreak of CoVs was evolved in the Arabian Peninsula, which is known as Middle East Respiratory Syndrome Coronavirus (MERS-CoV). Full genome sequence analysis of MERS-CoV isolates and its comparison with other CoV full genome sequences revealed a low to medium sequence identity. Furthermore, it showed more sequence identity and phylogenetic relations with bat-derived CoVs and lower values with animal-derived CoVs, indicating low possibility of zoonotic origin and possible incrimination of bats in the spread of MERS-CoV. The higher neighbor homology was evident with BetaCoV and their associated SARS-CoV. The spike protein, which is a highly variable part of CoV genome and responsible for difference in tissue tropism and virus entry to the cell, showed more or less similar profile of the whole genome analysis. Furthermore, the highest identity was with those in bats with Asian origin of CoV and there was lower homology with isolates from other continents. With low human-to-human transmission and low homology with CoV of animal origin, bats are thought to be the source of MERS-CoV, especially those bearing the Asian isolates of CoV.

Introduction

Coronaviruses (CoV) are enveloped single-stranded positive sense RNA viruses infect human and wide variety of animals and birds causing severe illness. Coronaviruses that infect humans include human

coronavirus 1 (alpha), 2 (beta) and 3 (gamma). The severe acute respiratory syndrome (SARS) has been associated with Betacoronavirus, which include four distantly related subgroups. It has been proved that the severe onset of SARS epidemic in 2003 with over 800 fatalities was evolved in a market of life-

caged dogs market in China. Studies confirmed that caged animals and bats were the potential intermediate hosts [1, 2].

The dangerous fate of coronaviruses arises from its ability of high recombination frequencies. The unique viral replication, the low fidelity of coronavirus-encoded polymerases and high recombination allows for unexpected viral evolution to infect other hosts, more clinical deficits and resistance to therapy or vaccination. SARS-CoV is a model virus for studying viral mutation, recombination and evolution. The human virus (SARS-CoV OC43) was thought to be evolved from bovine coronavirus. Furthermore, several reports assured the changes in tissue tropism by evolution of a new porcine respiratory corona virus pathogen from a gastrointestinal ancestor [2, 3]. Such changes in host and tissue tropism elucidate the possibility of emerging new viruses and occurrence of new diseases manifestations associated with the unpredictable fates of coronavirus evolving sequence.

In September 2012, the Middle East Respiratory syndrome Coronavirus (MERS-CoV) was initially identified in Saudi Arabia [4]. Following these first record positive cases were then recorded in United Kingdom, Germany, Tunisia, Jordan, Italy, Qatar, Emirates and France [5-8]. Saudi Arabia is the most heavily impacted with MERS-CoV, recording more than 50% fatality rate by showing 38 deaths out of 64 cases [8].

Among known viruses, coronavirus has a large genome of more than 27 kb. The genome of coronavirus encodes 23 putative proteins including 4 major structural proteins; nucleocapsid [N protein], spike [S protein], membrane [M] and small envelop proteins [E]. The variations in S protein among coronaviruses is associated with different species specificity, serological response and tissue tropism. The S protein is a glycoprotein essential for viral attachment to the cell

surface receptors and translocation into the infected cells. The S protein is cleaved in host cells into S1 and S2 subunits. S1 protein binds the host receptor, while S2 receptor mediates membrane fusion [9]. In this report, the full genome sequence of MERS-CoV was compared with other CoV sequences. Furthermore, the highly variable spike region was also retrieved and analyzed. The phylogenetic relations and its indication are thoroughly discussed.

Materials and methods

Retrieval of full genome sequences

The full-length CoV genomes were retrieved from the nucleotides depository at the National Center for Biotechnology Information (NCBI).

Search for putative domains

The full genome sequence of MERS-CoV was analyzed for all potential domains by the domain search tools at NCBI. The sequence for spike protein was identified and used to BLAST the protein database for knowing the high homology hits.

Multiple sequence alignment

Multiple sequence alignment was constructed by using Clustal omega tool at the European Bioinformatics Institute. The output file was retrieved and manually edited by GeneDoc software. The similarity and homology percent was calculated by Ugene 1.12.2 for mac.

Phylogenetic analysis

The output alignments from Clustal omega were further investigated by Dendroscope software for creation of phylogenetic trees. The phylogenetic tree was obtained by neighbor-joining method in

output formats of radial phylogram or circular cladogram.

Results and discussion

Bioinformatics is a useful tool in analysis of new genes, proteins or whole genome. Annotations and analysis of sequences is a gold standard in the finding of new functions, drug targets and analysis of recently emerged infectious diseases [10-15]. To identify the distinguishing feature of the new MERS-CoV, the full genome sequence of MERS-CoV as well as the sequence of S protein was retrieved, compared to other CoV full genome sequences and analyzed by bioinformatics tools.

Comparison of Coronaviruses whole genome

About 41 full genome sequences representing different isolates of coronaviruses were retrieved from the genome database (**Table 1, Fig. 1**). In general, low to medium identity was evident in comparing MERS-CoV with other CoVs. Taking the four Middle East isolates from Al-Hasa as a reference, the following conclusions can be elucidated. First, it shares about 46% homology with the SARS coronavirus genome sequence. Second, it shares low identity with CoV isolates from animal origins e.g. Sable antelope (18%), rabbit (42%), murine hepatitis (18%), giraffe (18%), bovine (42%), canine (35%), avian infectious bronchitis (12%) and alpaca (35%). Lastly, 34-35%, identity with isolates of CoV from bats.

Table 1. The accession number and identity (%) of the CoV sequences compared with MERS-CoV.

Coronavirus isolate	Accession number	Percent identity
Middle East respiratory syndrome coronavirus isolate Al-Hasa_4_2013	KF186564	100
Middle East respiratory syndrome coronavirus isolate Al-Hasa_3_2013	KF186565	100
Middle East respiratory syndrome coronavirus isolate Al-Hasa_2_2013	KF186566	99
Middle East respiratory syndrome coronavirus isolate Al-Hasa_1_2013	KF186567	100
Human betacoronavirus 2c Jordan-N3/2012	KC776174	99
Alpaca respiratory coronavirus isolate CA08-1/2008	JQ410000	35
Avian infectious bronchitis virus polyprotein	AY692454	12
Avian infectious bronchitis virus	NC_001451	33
Bat coronavirus Cp/Yunnan2011	JX993988	21
Bat coronavirus HKU2	NC_009988	35
Betacoronavirus England 1	NC_019843	99
Bovine coronavirus strain Mebus	BCU00735	42
Bovine coronavirus	NC_003045	42
Calf-giraffe coronavirus US/OH3/2006	EF424624	18
Canine coronavirus strain A76	JN856008	35

Canine respiratory coronavirus strain K37	JX860640	42
Equine coronavirus	NC_010327	27
Feline coronavirus isolate Black	EU186072	35
Giraffe coronavirus US/OH3-TC/2006	EF424622	18
Hipposideros bat coronavirus HKU10 TLC1347A	JQ989273	34
Human betacoronavirus 2c Jordan-N3/2012	KC776174	99
Human coronavirus 229E	NC_002645	34
Human coronavirus NL63 isolate CBJ 037	JX104161	35
Human enteric coronavirus 4408	FJ415324	42
Human group 1 coronavirus associated with pneumonia	AY518894	35
Murine hepatitis virus strain 2	AF201929	18
Murine hepatitis virus	AF208067	18
Murine hepatitis virus strain Penn 97-1	AF208066	18
Porcine epidemic diarrhea virus strain CV777	AF353511	34
Rabbit coronavirus HKU14	NC_017083	42
Rat coronavirus isolate 8190	JF792617	41
Rousettus bat coronavirus HKU10	NC_018871	34
Sable antelope coronavirus US/OH1/2003	EF424621	18
SARS coronavirus CUHK-W1	AY278554	46
SARS coronavirus FRA	AY310120	46
SARS coronavirus HKU-39849 isolate recSARS-CoV HKU-39849	JN854286	46
SARS coronavirus TJF	AY654624	46
SARS coronavirus TOR2	AY274119	46
SARS coronavirus TW11	AY502924	46
SARS coronavirus Urbani	AY278741	46
Scotophilus bat coronavirus 512	NC_009657	35
Transmissible gastroenteritis virus	NC_002306	35

This result highlights the genetic variability of CoV. While keeping low identities, the clinical signs and virus replication is still sharing common mechanisms. Phylogenetic of the whole genome

The phylogenetic relations of the retrieved full genome sequences are shown in Fig. 2. The MERS-CoVs isolated from Al-Hasa and their highly relevant isolated in Jordan and England were shown in blue.

The highly related sequences were highlighted in green. From the figure, MERS-CoV isolates were highly related to SARS CoV isolates and those of Asian origin.

Is MERS-CoV a SARS-CoV?

The clinical signs of MERS-CoV include fever, acute respiratory distress and in some cases renal failure

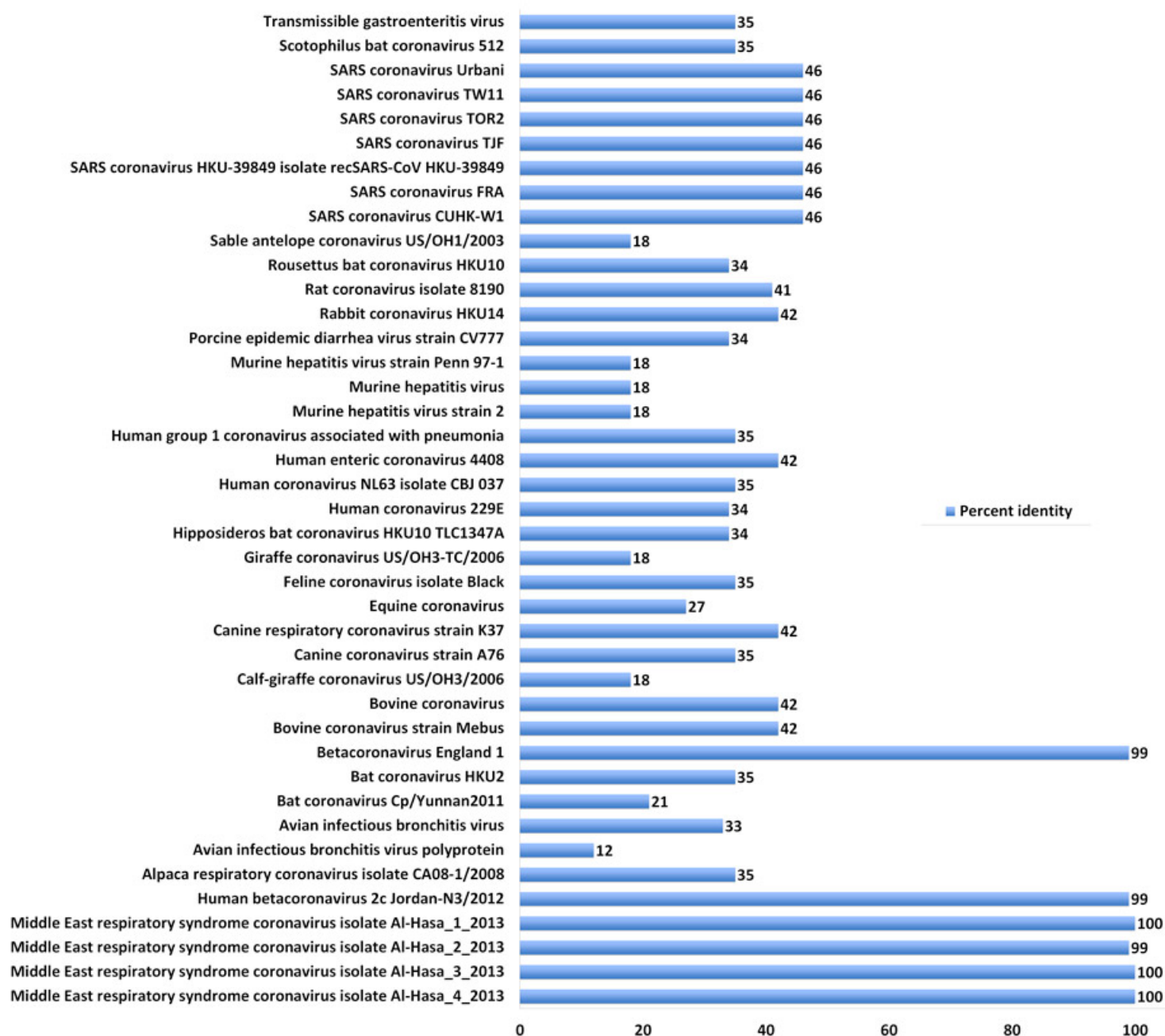
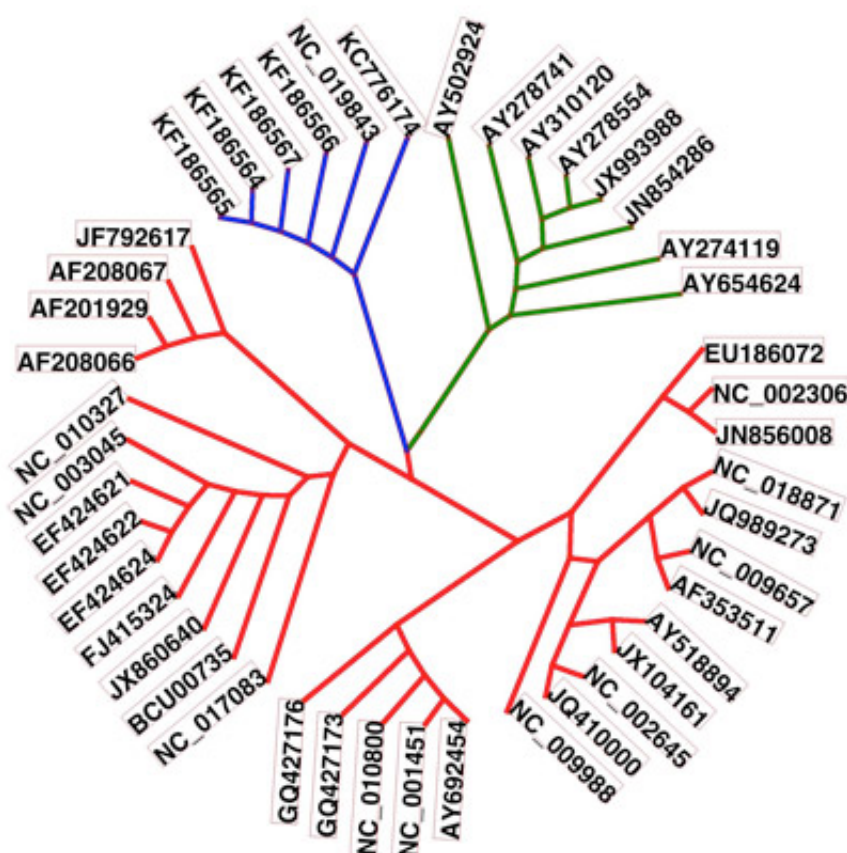


Fig. 1. Percent identity of whole genome sequence of 41 CoVs.

Fig. 2. Phylogram showing the relations of the full genome sequence of CoVs.



[4, 5]. These signs are more or less similar to the signs of SARS-CoV. Based on molecular bases, the new MERS-CoV is not SARS-CoV. MERS-CoV does not use the SARS-CoV receptor during entry into the cells. Furthermore, the mode of binding of MERS-CoV with the cell receptors is different from SARS CoV [16]. Phylogenetic maps showed that both MERS-CoV and SARS-CoV are sharing at least 40% of their sequences.

Phylogenetic relations of S protein

Analysis of the Spike protein of different CoV sequences reveals perfect homology of MERS-CoV sequences from Saudi Arabia with related isolates from England and Jordan. This group of isolates is highly related to bat-derived CoV HKU4 and HKU5 (**Table 2, Fig. 3**). In contrast, it is highly divergent from other bat-derived CoV as Eidolon, Rousettus,

Rhinolophus bats isolates. Interestingly, MERS-CoV was showed poor relations from CoV from animal sources as Sambar deer, bovine, waterbuck and rabbit coronaviruses. This underscores the role animals as carriers for MERS-CoV and highlights the potential incriminations of bats in the initiation of MERS-CoV.

The spike protein sequences show more or less low identity among different coronaviruses, showing as low as 12%. S protein is therefore, a region, which is highly variable among coronaviruses that lead to recognition of different cellular receptors and altered pathogenicity. The low homology of MERS-CoV S protein with the other S proteins reveals a significant change in the molecular interaction of the new virus with host receptors.

Table 2. Sequences producing significant alignments with S protein of MERS-CoV.

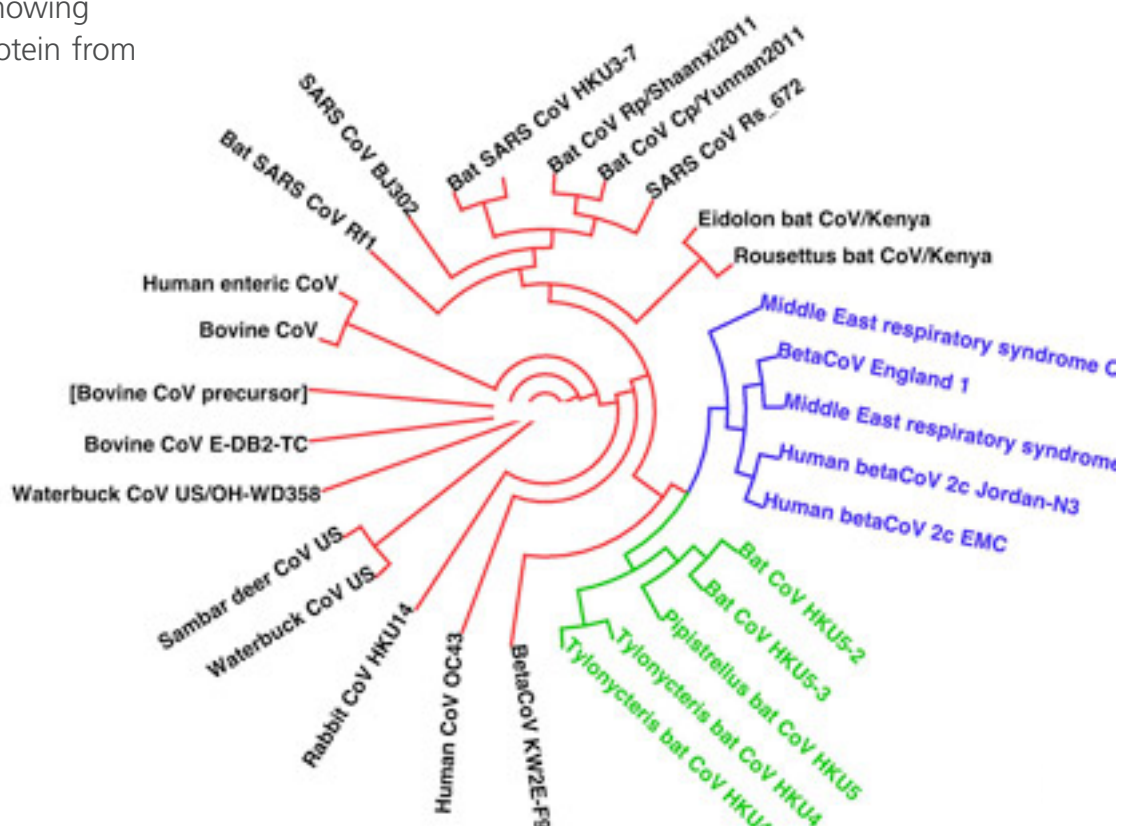
Description	Total score	E value	Ident	Accession
S protein [Middle East respiratory syndrome coronavirus] >gb AGN70951.1	1271	0.0	100%	AGN70929.1
S protein [Human betacoronavirus 2c EMC/2012]	1267	0.0	99%	AFS88936.1
spike glycoprotein [Middle East respiratory syndrome coronavirus]	1266	0.0	99%	AGN52936.1
S protein [Betacoronavirus England 1]	1266	0.0	99%	YP_007188579.1
S protein [Human betacoronavirus 2c Jordan-N3/2012]	1266	0.0	99%	AGH58717.1
spike glycoprotein [Pipistrellus bat coronavirus HKU5]	949	0.0	72%	AGP04936.1
spike glycoprotein [Tylonycteris bat coronavirus HKU4]	912	0.0	75%	AGP04916.1
spike glycoprotein [Betacoronavirus BtCoV/KW2E-F93/Nyc_spec/GHA/2010]	871	0.0	68%	AGC51116.1
spike [Bat coronavirus (BtCoV/A434/2005)]	864	0.0	71%	ABG11962.1
spike glycoprotein [Bovine coronavirus]	485	2e-152	45%	AGO98871.1
spike glycoprotein [Bat SARS coronavirus HKU3-13]	475	1e-149	44%	ADE34823.1
spike glycoprotein [bat SARS coronavirus HKU3-2]	474	2e-149	44%	AAZ41329.1
spike protein [Bat SARS CoV Rm1/2004]	474	3e-149	45%	ABD75332.1
spike protein [Bovine coronavirus]	476	3e-149	45%	ACB30203.1
spike protein [Bat SARS CoV Rf1/2004]	473	4e-149	44%	ABD75323.1
spike protein [Bat coronavirus Cp/Yunnan2011]	472	1e-148	44%	AGC74176.1
spike glycoprotein [Bat SARS coronavirus HKU3-7]	472	2e-148	44%	ADE34755.1
spike protein [Eidolon bat coronavirus/Kenya/KY24/2006]	472	2e-148	43%	ADX59466.1
spike protein [Bat coronavirus Rp/Shaanxi2011]	471	2e-148	44%	AGC74165.1

Zoonotic or human-human transmission of MERS-CoV

Zoonotic exposure implies the transmission of the diseases through contact with animals, in this case as cattle, bats, or cats infected with coronavirus. Until the moment the role of reservoirs or inter-

mediate hosts of MERT-CoV is still not understood. By using phylogenetic analysis, the possible way of transmission of MERS-CoV was deduced from the phylogenetic relation of the newly evolved virus with the previous sequences isolated from bats and other animals and human.

Fig. 3. Phylogram showing the relations of S protein from different CoVs.



A limited human-human transmission was suggested by the analysis of surveillance data in France. The first 2 MERS-CoV-reported cases were in contact together in the hospital for 4 days, however, in a total number of 201 individuals were in contact with 2 cases only 2 were positive and the others were asymptomatic [7].

Camels are thought to be the source for MERS-CoV [17]. Recently, neutralizing antibodies against MERS-CoV were found in all samples of camels in the Middle East [18]. This does not suggest that camels are the source of infection. Confirmatory tests and isolation of the virus from camels is still yet undetermined. Based on the obtained phylogenetic tree, poor relation was found between MERS-CoV and those coronaviruses isolated from bovines and other animals. Bovine CoV was found to be circulating in

cattle causing unnoticed mild symptoms to severe outbreaks. Furthermore, the virus was involved in enteric manifestations and now evolved to include also respiratory symptoms [19-21]. Camels are infected with CoV-like viruses (genetically related with bovine CoV) causing enteric symptoms in calves [20, 22]. However, the presence or absence of respiratory forms of CoV in camels is still not known. The claim that camels might be the source of MERS-CoV still needs further investigations.

Bats are the major reservoir for CoV infections. Recently, 96 bats representing 7 species (*Rhinopoma hardwickii*, *Rhinopoma microphyllum*, *Taphozous perforatus*, *Pipistrellus kuhlii*, *Eptesicus bottae*, *Eidolon helvum*, and *Rousettus aegyptiacus*) were captured and examined for their CoV contents [23]. The prevalence of CoVs was high (~28% of fecal sam-

ples), MERS-CoV was found in only 1 bat. MERS-positive signal was obtained in PCR analysis of the *T. perforatus* bat captured in Bisha near the home and workplace of the MERS index case-patient. This result indicates that bats are the ultimate source

MERS-CoV. To this end, more details about the origin, carriers, intermediate hosts and pathogenesis of MERS-CoV still needs further investigations.

References

1. Perlman, S., Netland, J. Coronaviruses post-SARS: Update on replication and pathogenesis. *Nature reviews. Microbiology* 7 (6): 439-50.
2. Chang, H-W., Egberink, H.F., Halpin, R., Spiro, D.J., Rottier, P.J.M. Spike protein fusion peptide and feline coronavirus virulence. *Emerging Infectious Diseases* 2012; 18 (7): 1089-95.
3. Laude, H., Van Reeth, K., Pensaert, M. Porcine respiratory coronavirus: Molecular features and virus-host interactions. *Veterinary Research* 1993; 24 (2): 125-50.
4. Zaki, A.M., van Boheemen, S., Bestebroer, T.M., Osterhaus, A.D.M.E., Fouchier, R.A.M. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *The New England Journal of Medicine* 2012; 367 (19): 1814-20.
5. Bermingham, A., Chand, M.A., Brown, C.S., Aarons, E., Tong, C., Langrish, C. et al. Severe respiratory illness caused by a novel coronavirus, in a patient transferred to the United Kingdom from the Middle East, September 2012. *Euro surveillance: Bulletin Européen sur les maladies transmissibles = European Communicable Disease Bulletin* 2012; 17 (40): 20290.
6. Buchholz, U., Müller, M.A., Nitsche, A., Sanewski, A., Wevering, N., Bauer-Balci, T. et al. Contact investigation of a case of human novel coronavirus infection treated in a German hospital, October-November 2012. *Euro Surveillance: Bulletin Européen sur les maladies transmissibles = European Communicable Disease Bulletin* 2013; 18 (8).
7. Maillies, A., Blanckaert, K., Chaud, P., van der Werf, S., Lina, B., Caro, V. et al. First cases of Middle East Respiratory Syndrome Coronavirus (MERS-CoV) infections in France, investigations and implications for the prevention of human-to-human transmission, France, May 2013. *Euro Surveillance: Bulletin Européen sur les maladies Transmissibles = European Communicable Disease Bulletin* 2013; 18 (24).
8. Balkhair, A., Al Maamari, K., Alawi, F.B. The Struggle Against MERS-CoV (The Novel Coronavirus). *Oman Medical Journal* 2013; 28 (4): 226-7.
9. Raj, V.S., Mou, H., Smits, S.L., Dekkers, D.H.W., Müller, M.A., Dijkman, R. et al. Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature* 2013; 495 (7440): 251-4.
10. Kandeel, M., Nakanishi, M., Ando, T., El-Shazly, K., Yosef, T., Ueno, Y. et al. Molecular cloning, expression, characterization and mutation of *Plasmodium falciparum* guanylate kinase. *Molecular and Biochemical Parasitology* 2008; 159 (2): 130-3.
11. Kandeel, M., Kitamura, Y., Kitade, Y. The exceptional properties of *Plasmodium* deoxyguanylate pathways as a potential area for metabolic and drug discovery studies. *Nucleic Acids Symposium Series* 2009; (53): 39-40.
12. Kandeel, M., Miyamoto, T., Kitade, Y. Bioinformatics, enzymologic properties, and comprehensive tracking of *Plasmodium falciparum* nucleoside diphosphate kinase. *Biological & Pharmaceutical Bulletin* 2009; 32 (8): 1321-7.
13. Kandeel, M., Ando, T., Kitamura, Y., Abdel-Aziz, M., Kitade, Y. Mutational, inhibitory and microcalorimetric analyses of *Plasmodium falciparum* TMP kinase. Implications for drug discovery. *Parasitology* 2009; 136 (1): 11-25.
14. Kandeel, M., Kitade, Y. Computational analysis of siRNA recognition by the Ago2 PAZ domain and identification of the determinants of RNA-induced gene silencing. *PLoS One* 2013; 8 (2): e57140.
15. Kandeel, M., Kitade, Y. In silico molecular docking analysis of the human Argonaute 2 PAZ domain reveals insights into RNA interference. *Journal of Computer-aided Molecular Design* 2013; 27 (7): 605-14.
16. Chen, Y., Rajashankar, K.R., Yang, Y., Agnihothram, S.S., Liu, C., Lin, Y.L., Baric, R.S., Li, F. Crystal structure of the receptor-binding domain from newly emerged middle East respiratory syndrome coronavirus. *J Virol*. 2013; In press.
17. Hawkes, N. Camels could be the source of MERS coronavirus, research finds. *BMJ (Clinical Research ed.)* 2013; 347: f5052.
18. Reusken, C.B., Haagmans, B.L., Müller, M.A., Gutierrez, C., Godeke, G.-J., Meyer, B. et al. Middle East respiratory syndrome coronavirus neutralising serum antibodies in dromedary camels: A comparative serological study. *The Lancet Infectious Diseases* 2013; In press.
19. Hick, P.M., Read, A.J., Lugton, I., Busfield, F., Dawood, K.E., Gabor, L. et al. Coronavirus infection in intensively managed cattle with respiratory disease. *Australian Veterinary Journal* 2012; 90 (10): 381-6.
20. Boileau, M.J., Kapil, S. Bovine coronavirus associated syndromes. *The Veterinary Clinics of North America. Food animal practice* 2010; 26 (1): 123-46.
21. Saif, L.J. Bovine respiratory coronavirus. *The Veterinary Clinics of North America. Food animal practice* 2010; 26 (2): 349-64.
22. Wünschmann, A., Frank, R., Pomeroy, K., Kapil, S. Enteric coronavirus infection in a juvenile dromedary (*Camelus dromedarius*). *Journal of Veterinary Diagnostic Investigation: Official Publication of the American Association of Veterinary Laboratory Diagnosticians, Inc.* 2002; 14 (5): 441-4.
23. Middle East Respiratory Syndrome Coronavirus in Bats, Saudi Arabia. *Emerging Infectious Disease Journal* 2013; In press.

Comment on this article:



<http://medicalia.org/>

Where Doctors exchange clinical experiences, review their cases and share clinical knowledge. You can also access lots of medical publications for free. **Join Now!**

Publish with iMedPub

<http://www.imedpub.com>

Enzymes received a bulk of research involving medicine, biotechnology, biochemistry, structural biology and many other specialties. Furthermore, it took the interest in many drug companies for development of new drugs. In spite of the great work done in the field of enzymes there no specific journal interested in enzymology and related areas of drug discovery. Therefore, the journal *Molecular Enzymology and Drug Targets* will be unique in its discipline as it can be the first interest to authors working in the field of enzymology and related subjects.

Submit your manuscript here:
<http://www.medt.com.es>