

Genomic sequencing and analysis of the first imported Middle East Respiratory Syndrome Coronavirus (MERS CoV) in China

XIE Qian^{1†}, CAO YuJuan^{1†}, SU Juan², WU XianBo¹, WAN ChengSong¹, KE ChangWen²,
ZHAO Wei^{1*} & ZHANG Bao^{1*}

¹The lab of Biosafety Level 3, School of Public Health and Tropical Medicine, Southern Medical University, Guangzhou 510515, China;

²Research Center for Pathogens Detection Technology of Emerging Infectious Diseases, Guangdong Provincial Center for Disease Control and Prevention, Guangzhou 511430, China

Received June 19, 2015; accepted June 22, 2015

Citation: Xie Q, Cao YJ, Su J, Wu XB, Wan CS, Ke CW, Zhao W, Zhang B. Genomic sequencing and analysis of the first imported Middle East Respiratory Syndrome Coronavirus (MERS CoV) in China. *Sci China Life Sci*, doi: 10.1007/s11427-015-4903-7

Dear Editors,

Middle East Respiratory Syndrome (MERS) has spread to more than 23 countries since the first case of a MERS coronavirus (MERS CoV) infection was confirmed in Sep 2012 [1]. Because MERS CoV has led to mortality as high as 40%, it has become one of the most important emerging pathogens threatening human health. Recently, the outbreak of MERS CoV, which resulted in 172 infections and 27 deaths in South Korea, was the largest outside Saudi Arabia. A 44-year-old man, had been exposed to the first confirmed MERS case who shared a ward with the man's father in this outbreak, travelled from South Korea to China, becoming the first imported MERS case reported in China. Here, we report the sequencing and analysis of the MERS CoV genome from this case. Viral genome sequencing provides information vital to tracing the pathogenic source, predicting viral virulence, and following the pattern of viral evolution, variation, and transmission [2–4].

Twenty-four primer pairs, specific to conserved regions, were used to amplify the complete genome of MERS CoV by PCR. The resulting PCR products had an average length of 1,500 base pairs (bp). The PCR products were recovered and sequenced by the Sanger dideoxy method. The genomic

sequence was assembled with DNASTAR 7.0 software and deposited in GenBank with accession no. KT036372. The deposited sequence was 29,928 bp, which was ~180 bp shorter than the complete genomic sequence, as the technique of 5'- and 3'-RACE (rapid-amplification of cDNA ends), which amplifies genomic ends, was not applied.

The sequence we obtained: KT036372, differed slightly from other MERS CoV sequences submitted. It differed in four nucleotides from the updated sequence of KT006149.2 provided by China Centers for Disease Control (CDC) from the same patient; and in 10 nucleotides from sequence KT029139 submitted by Korea Centers for Disease Control and Prevention. The two sequences used for comparison, KT006149.2 and KT029139, differed from each other by 12 nucleotides. Our analysis shows that the KT036372 sequence is reliable, sharing greater similarity to the KT029139 sequence from South Korea. The differences among the three sequences (Table S1) might have resulted from: viral evolution and mutation; the collection date (KT036372 on May 28; KT006149 on May 27); sample origin (KT036372 and KT006149 from clinical sample; KT029139 from virus culture); and the sequencing method (KT036372 by Sanger dideoxy method; KT006149 by Ion Torrent platform and Sanger dideoxy method; and KT029139 using the Illumina platform).

Complete alignment of the genomic sequence in GenBank indicated that KT036372 shared a high similarity of

[†]Contributed equally to this work

*Corresponding author (email: zhaowei@smu.edu.cn; zhang20051005@126.com)

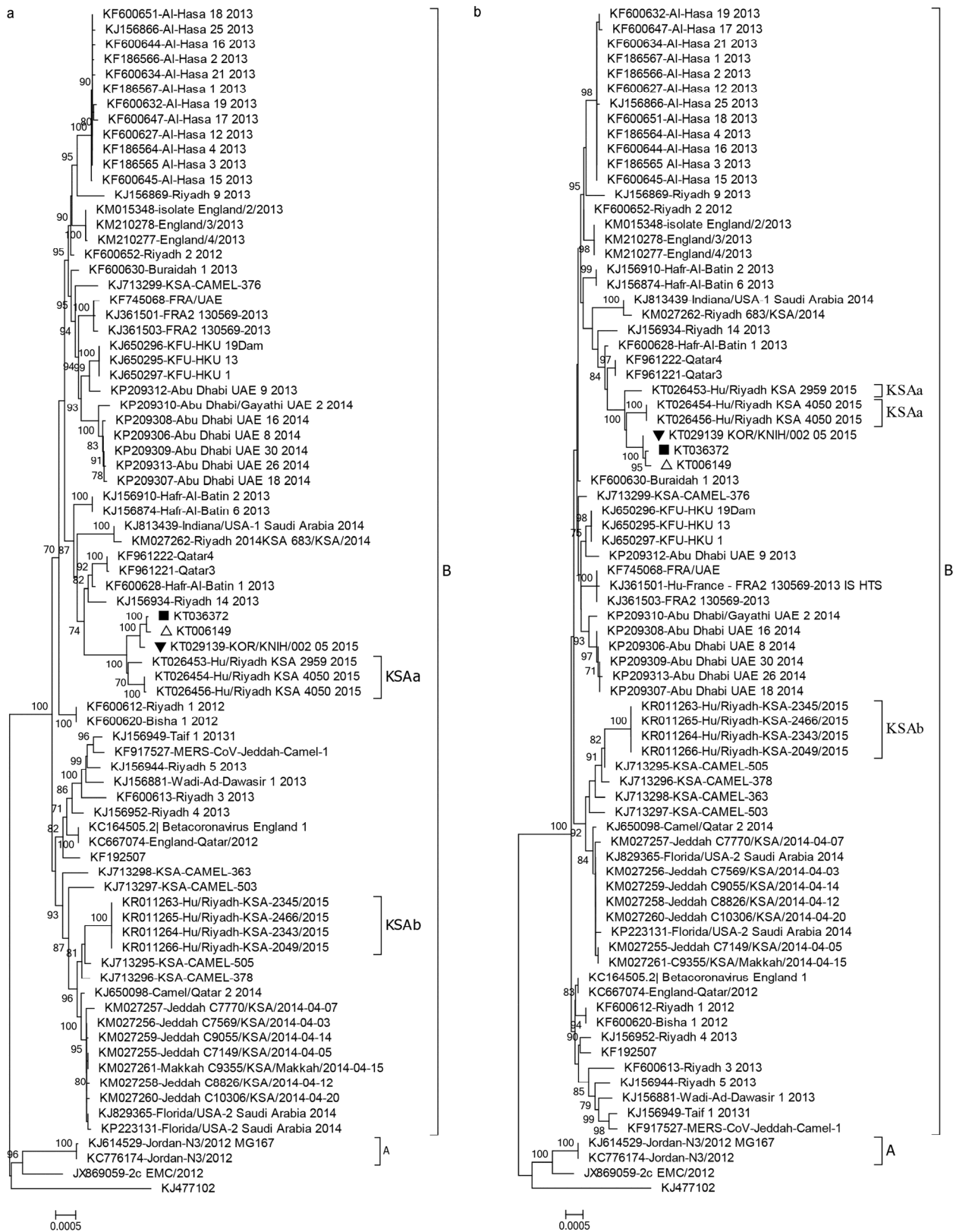


Figure 1 Phylogenetic tree of MERS CoV sequence generated by the Neighbor Joining method using MEGA6.0 and tested with 1,000 bootstrap replicates. a, Phylogenetic tree of MERS CoV based on genomic sequence. b, Phylogenetic tree of MERS CoV based on *ORF1a* gene sequence. ■, The MERS isolate determined in this study; Δ, The MERS strain from the same patient sequenced by China CDC; ▼, The sequence submitted by South Korea. Phylogenetic trees show that MERS-CoV sequences identified in Korea and China cluster closest with viruses obtained from a recent outbreak in Riyadh in 2015 (KSAa).

99.53%–99.92% with the strains isolated in recent years, the highest identity being with the Saudi Arabian isolate sequenced in Mar 2015. The *S* gene of the virus, which is associated with viral factors for host recognition, also shared high similarity of 98.97%–99.93% with the *S* gene of the MERS CoV sequences stored in the database, with highest similarity to KT026453.

We found three novel variants, with single amino acid substitutions: D977G, T1833I, and G6896S, which differ from the strains reported in the database. These mutations are located in the poly-protein 1ab (ORF1ab) that is cleaved into 16 nonstructural proteins (NSP1–16). D977G and T1833I are situated in the nonstructural protein 3 (NSP3), predicted to be a protease responsible for intracellular post-translational processing of viral proteins; while the mutation G6896S is located in the nonstructural protein 16 (NSP16), whose function is unknown. However, no new mutation has emerged in the proteins closely relating to virus assembly and virulence, such as the *S*, *E*, *M*, and *N* proteins.

Phylogenetic analysis of the complete genomic sequence further demonstrated that KT036372 as well as KT029139, belonged to clade B (Figure 1a), like the KSAA strains recently isolated from Saudi Arabia. Sufficient sequence divergence was not seen for them to be assigned to a new clade. Remarkably, there were two different secondary branches KSAA and KSAB, suggesting that two different origins of MERS CoV may prevail in Saudi Arabia. The phylogenetic tree of *ORF1a* gene (Figure 1b) was closer to that of the full-length genome. Nevertheless, *S* gene phylogenetic tree was similar to that of the genome, but distinct in that KSAA and KSAB were clustered together (Figure S1).

The full-length genomic sequence exhibited high similarity above 95% with MERS CoV sequences found in the database. No recombination events were identified in the current epidemic strains by either the Simplot or the RDP software methods.

In conclusion, the sequence of the MERS CoV isolate from the first imported case in China's Guangdong province, was highly similar to that of recent isolates from the Saudi Arabia. We identified three new mutations in the polyprotein ORF1ab; however, no variation in genes associated with viral assembly or the virulence-linked *S*, *E*, *M*

and *N* genes were observed, nor was there any apparent recombination in the genome. These results further suggest that the virulence of the MERS CoV did not increase. Based on current reports, mortality was ~10%, which is far lesser than the 40% previously reported. Simultaneously, the epidemiological characteristics of the virus suggest that it is still limited to human-to-human transmission, with no sign of a large-scale outbreak.

As the case reported here illustrates, globalization and frequent international travel can pose a threat to human health worldwide, and exposure to the infectious agent can occur unpredictably. Hence, excellent technical preparations and mandatory close monitoring are critical for rapid detection of pathogenic viruses like MERS CoV, and to reduce the risk of a pandemic. We were able to achieve this by identification and analysis of the viral genome from the only imported case of MERS CoV, with no contacts being infected in Guangdong up to now [5].

This work was supported by the Guangdong Provincial Science and Technology Department (2010A040302003, 2011B031800163), and the 12th five-year-major-projects of China's Ministry of Public Health (2012zx-10004-213).

- 1 van Boheemen S, de Graaf M, Lauber C, Bestebroer TM, Raj VS, Zaki AM, Osterhaus AD, Haagmans BL, Gorbelenya AE, Snijder EJ, Fouchier RA. Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans. *MBio*, 2012, 3: e00473-12
- 2 Cotten M, Watson SJ, Kellam P, Al-Rabeeh AA, Makhdoom HQ, Assiri A, Al-Tawfiq JA, Alhakeem RF, Madani H, AlRabiah FA, Al Hajjar S, Al-nassir WN, Albarrak A, Flemban H, Balkhy HH, Alsubaie S, Palser AL, Gall A, Bashford-Rogers R, Rambaut A, Zumla AI, Memish ZA. Transmission and evolution of the Middle East respiratory syndrome coronavirus in Saudi Arabia: a descriptive genomic study. *Lancet*, 2013, 382: 1993–2002
- 3 Cotten M, Lam TT, Watson SJ, Palser AL, Petrova V, Grant P, Pybus OG, Rambaut A, Guan Y, Pillay D, Kellam P, Nastouli E. Full-genome deep sequencing and phylogenetic analysis of novel human betacoronavirus. *Emerg Infect Dis*, 2013, 5: 736–742
- 4 Madani TA, Azhar EI, Hashem AM. Evidence for camel-to-human transmission of MERS coronavirus. *N Engl J Med*, 2014, 14: 1360
- 5 Wu J, Yi L, Zou L, Zhong H, Liang L, Song T, Song Y, Su J, Ke C. Imported case of MERS-CoV infection identified in China, May 2015: detection and lesson learned. *Euro Surveill*, 2015, 24: 21158

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Supporting Information

Figure S1 Phylogenetic tree of *S* gene of MERS CoV built with Neighbor Joining method using MEGA6.0 and tested with 1,000 bootstrap replicates. 50 strains with high similarity isolated from Saudi Arabia and United Arab Emirates in 2013–2014 were compressed in the phylogenetic tree of *S* gene.

Table S1 The sequence differences between the three MERS CoVs

The supporting information is available online at life.scichina.com and www.springerlink.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.