SCIENTIFIC REP⦿RTS

OPEN

The recent ancestry of Middle East respiratory syndrome coronavirus in Korea has been shaped by recombination

Jin Il Kim[1],*, You-Jin Kim[2],*, Philippe Lemey[13],*, Ilseob Lee[1], Sehee Park[1], Joon-Yong Bae[1], Donghwan Kim[1], Hyejin Kim[1], Seok-Il Jang[1], Jeong-Sun Yang[2], Hak Kim[2], Dae-Won Kim[4], Jeong-Gu Nam[2], Sung Soon Kim[2], Kisoon Kim[3], Jae Myun Lee[6,12], Man Ki Song[7,12], Daesub Song[8,12], Jun Chang[9,12], Kee-Jong Hong[10,12], Yong-Soo Bae[11,12], Jin-Won Song[1], Joo-Shil Lee[5] & Man-Seong Park[1,12]

Middle East respiratory syndrome coronavirus (MERS-CoV) causes severe cases of human respiratory disease. Since 2012, the victims have mainly come from the Middle East countries or sporadically from some other geographical regions seeded by the travelers who visited the Middle East. Such an introduction through travelling led to the emergence of a MERS-CoV outbreak in Korea in May 2015, which caused more than 140 confirmed human cases in less than a month. Using 70 complete genome sequences of MERS-CoV isolates, including the most recent sequences for the Korean and Chinese isolates, we reconstructed the phylogenetic relationships of the complete genome and the individual protein coding regions. The Korean MERS-CoV strain clustered in the previously established Hafr-Al-Batin-1_2013 clade together with two Saudi Arabian and one Chinese strain sampled in 2015. Although these four strains remained monophyletic in the entire protein-coding region, this clade showed different phylogenetic relationships across the genome, indicating a shared unique recombination pattern that is different from previously reported putative recombination strains. Our findings suggest that the recent ancestor of the Korean and its related MERS-CoV strains is characterized by unique mosaic genome pattern that is different from other putative recombinants.

MERS was first described in 2012 in fatal human cases that were caused by a single-stranded RNA coronavirus[1–3]. Since then, more than 1,000 cases of MERS-CoV infection have been confirmed with an estimated case-fatality rate (CFR) of 39.5%[4], and the elderly and immunocompromised patients appear to be most severely affected[5]. Acute respiratory distress and pneumonia are the major clinical manifestations of MERS-CoV infection[6]. Gastrointestinal symptoms and renal failure are also reported in some cases[5,7]. Dromedary camels are considered to be the intermediate host in MERS zoonotic transmission chains from bats to humans[8,9]. In a previously published study, the spread of MERS-CoV appeared to match transmission routes between animal reservoirs and infected humans[10]. However, it was not readily replicated in a camel transmission study[11]. Despite this contradiction, MERS-CoV spread was

[1]Department of Microbiology, the Institute for Viral Diseases and Korea Bank for Pathogenic Viruses, College of Medicine, Korea University, Seoul 136-705, Republic of Korea. [2]Division of Respiratory Viruses. [3]Division of Influenza Virus, Center for Infectious Diseases. [4]Division of Biosafety Evaluation and Control. [5]Korea National Institute of Health, Osong 361-951, Republic of Korea. [6]Department of Microbiology and Immunology, Yonsei University College of Medicine, Seoul 120-752, Republic of Korea. [7]Laboratory Science Division, International Vaccine Institute, Seoul 151-742, Republic of Korea. [8]College of Pharmacy, Korea University, Sejong 339-700, Republic of Korea. [9]Division of Life & Pharmaceutical Sciences, Ewha Womans University, Seoul 120-750, Republic of Korea. [10]ATGen, Seongnam 463-400, Republic of Korea. [11]Department of Biological Science, Sungkyunkwan University, Suwon 440-746, Republic of Korea. [12]The MERS Task Force of the Korean Society of Virology, Seoul 137-701, Republic of Korea. [13]Department of Microbiology and Immunology, Rega Institute, KU Leuven – University of Leuven, Leuven 3000, Belgium. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to M.-S.P. (email: manseong.park@gmail.com or ms0392@korea.ac.kr)
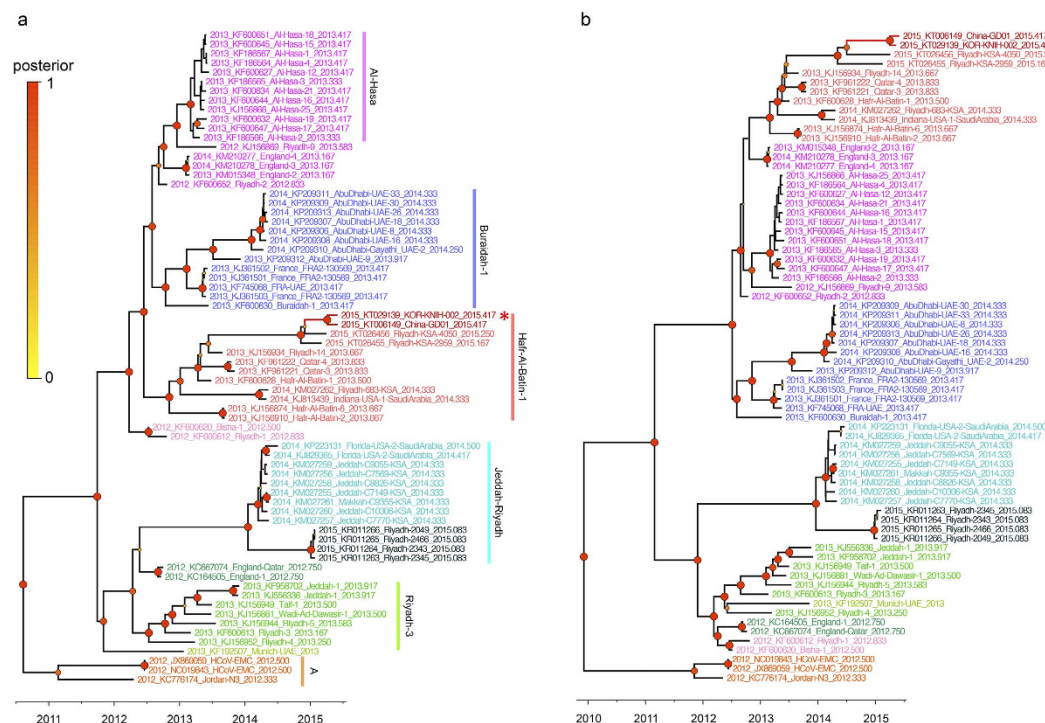
Figure 1. Phylogenetic relationships of MERS-CoV complete genome and ORF1ab sequences. The complete genome (**a**) and ORF1ab (**b**) sequences of 70 MERS-CoV strains were investigated for their phylogenetic relationships. In the complete genome tree, six different evolutionary clades were indicated with different colors (clade A, orange; Riyadh-3, lime green; Jeddah-Riyadh, mint; Hafr-Al-Batin-1, peach; Buraidah-1, lavender; and Al-Hasa, magenta). As the color of circles in the tree nodes, the size of circles in the node represents the posterior probability of their clustering (the bigger size, the higher probability). The tip of the Korean strain (KOR002) was denoted with the red color and asterisk.

demonstrated largely among camels and people in the Middle East[12] and travellers who visited the region could occasionally seed the virus to long-distance destinations in the Europe, Southeast Asia, and North America[9,13,14].

The first case of traveler-associated MERS-CoV outbreak in Korea occurred in May 2015[15]. A 68-year-old index person traveled to four countries in the Middle East and returned to Korea on May 4 without clinical complaints. When clinical symptoms were developed one week later, the index person sought medical attention in two primary clinics and two upper-class hospitals, but a diagnosis was only made on May 20th by confirming MERS-CoV infection. In the meantime, one of the nosocomial contacts of the index patient travelled to China via Hong Kong and was diagnosed with MERS-CoV infection on May 27 in Guangdong. As of June 19, 2015, a total of 166 confirmed cases have been reported in the MERS outbreak in Korea including one Chinese case[16]. Even though the viral spread is mainly limited to hospital-based transmission, as seen in previous cases[17,18] and no more confirmed cases are reported in Korea, this represents the largest outbreak outside the Middle East region. To investigate the evolutionary history of the MERS-CoV strain (GenBank accession No. KT029139; KOR-KNIH-002_2015, KOR002) responsible for the outbreak in Korea, we analyzed 70 complete genome sequences available in the NCBI including the most recent Chinese (KT006149; China-GD01_2015, China01) and Saudi Arabian (KT026455; Riyadh-KSA-2959_2015, KSA2959 and KT026456; Riyadh-KSA-4050_2015, KSA4050) sequences (Table S1).

We first analyzed the phylogenetic relationships between the KOR002 strain and all available MERS-CoV sequences. The complete genome and individual open reading frame (ORF) sequences (ORF1ab, S, ORF3, ORF4a, ORF4b, ORF5, E, N, M, and 8b) were analyzed separately using time-resolved Bayesian phylogenetic inference method implemented in BEAST (v1.8.2)[19]. In agreement with previous analyses[20], six distinct clades were identified in the complete genome tree (Fig. 1a): Clade A, Riyadh-3, Jeddah-Riyadh, Hafr-Al-Batin-1, Buraidah-1, and Al-Hasa. Together with the Chinese China01 strain, the KOR002 strain clusters within the Hafr-Al-Batin-1 clade and shows a relatively close relationship with two Saudi-Arabian strains sampled in 2015 (KSA2959 and KSA4050). Other 2015 strains from Riyadh in Saudi Arabia (KR011264, Riyadh-2343_2015; KR001265, Riyadh02466_2015; KR011266, Riyadh-2049_2015; and KR011263, Riyadh-2345_2015) clustered in a different clade (Fig. 1a). Among the ten individual ORF trees, the tree of ORF1ab exhibited a similar pattern to that of complete genome sequences except for Bisha-1_2012 (GenBank accession number KF600620), Riyadh-1_2012 (KF600612), and Munich-UAE_2013 (KF192507) strains, which clustered together with other strains of the Riyadh-3 clade (Fig. 1b). In the complete genome tree, they were found basal to the Al-Hasa (Bisha-1_2012 and Riyadh-1_2012) and Riyadh-3 (Munich-UAE_2013) clades, respectively (Fig. 1a). Based on the phylogenetic trees of the complete genome and ORF1ab, the Hafr-Al-Batin-1 clade appeared to be more closely related to the Al-Hasa and Buraidah-1 clades whereas the Riyadh-3 and Jeddah-Riyadh clades formed a sister lineages (Fig. 1). In the trees of the S, ORF3,
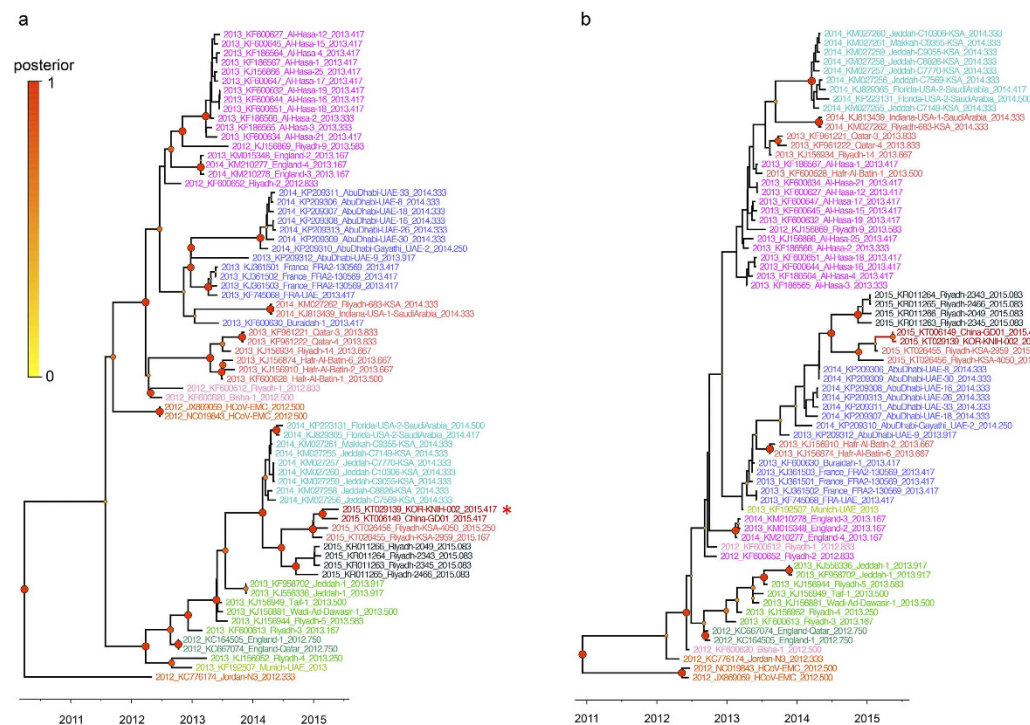
Figure 2. Phylogenetic relationships of MERS-CoV S and N sequences. The S (**a**) and N (**b**) sequences of 70 MERS-CoV strains were investigated for their phylogenetic relationships. See the detailed legend in the Fig. 1.

ORF4a, ORF4b, ORF5, E, and M genes, however, the most recent four sequences (KOR002, China01, KSA2959, and KSA4050) from the complete genome Hafr-Al-Batin-1 clade exhibited a closer relationship to the Jeddah-Riyadh strains compared to the other Hafr-Al-Batin-1, Al-Hasa, and Buraidah-1 clade sequences (Fig. 2a and S1B, S2, S3, and S4A). For the N and ORF8b genes, similar relationships were observed only among the Riyadh 2015 strains from the Jeddah-Riyadh clade sequences (Fig. 2b and S4B). This incongruent clustering pattern in the phylogenetic trees suggests that genetic recombination occurred in the MERS-CoV evolutionary history.

Recombination has been described previously in other coronavirus genomes[21–23], and was also suggested to affect the evolution of MERS-CoV[24]. In our recombination analyses of the MERS-CoV complete genomes, 25 strains (2 strains in 2012, 10 strains in 2013, 9 in 2014, and 4 in 2015) emerged as putative recombinants (Table 1). The 20 strains isolated before 2015 appeared to retain two recombination breakpoints in the linear ORF alignment (Fig. 3a). However, four putative recombinants in 2015 (KOR002, China01, KSA2959, and KSA4050), coinciding with the strains showing the unique relationships noted in the phylogenetic trees above (Figs 1 and 2 and S1 to S4), shared four breakpoints, which resulted in five recombinant fragments (Fig. 3b). Based on these mosaic patterns shared among the four putative recombinants in 2015, we compiled five new datasets representing each non-recombinant fragment and evaluated the phylogenetic relationships of the four putative recombinants in 2015. In each tree (Fig. 3c–e and S5), the four putative recombinants in 2015 always grouped together and showed close relationships with their parental strains as detected in the recombination test (Table 1). A recombination analysis using a larger window size suggested similar strains as putative recombinants, especially for KOR002 and its related 2015 strains (Table S2). Consistent with phylogenetic results above, the trees of each recombination region exhibited similar evolutionary clustering patterns according to the inclusion of corresponding ORF regions: in the trees of recombination region II and IV (Fig. 3d and S1B, S2, S3, and S4A), which represent the ORF1ab and a large part of the S-M protein coding regions, each clade clustered similar to the trees in Fig. 1b and S1A, respectively. In the recombination region III (18,033 to 23,502 region; 5,470 nucleotides in length) (Fig. 3b), which comprises the region of C-terminal ORF1ab and N-terminal S protein genes, the tree pattern appeared to be similar to that of S protein gene (Figs 2a and 3e), which is characterized by a much higher substitution rate than the ORF1ab (Table S3). Even though we used cell culture media of the third passaged Vero cells for the RNA isolation of the KOR002 strain, the possibility for contamination of the original sputum sample with multiple viral clones and subsequent recombination can be excluded because the Chinese and Saudi Arabian strains related to KOR002 all exhibited similar genomic recombination patterns. Taken together, these results suggest that genetic recombination has contributed to the evolutionary dynamics of MERS-CoV genomes and that this has particularly shaped the recent MERS-CoV ancestry of the Korean outbreak.

Based on the phylogenetic clustering patterns and the recombination imprints we detected (Figs 1–3 and Table 1), one of the recombinant strains that evolved from the Hafr-Al-Batin-1 clade was introduced by air travel into Korea. We can only speculate about when the genetic recombination occurred. However, in the Riyadh area, some strains of the Jeddah-Riyadh clade already circulated before May 2015, and considering the close relationships between some of the Hafr-Al-Batin-1 and Jeddah-Riyadh clade sequences shown in the phylogenetic trees,

Accession #	Strain	Recombination region (nucleotide)	Corresponding ORF region	Parental strain	% similarity
KC164505	England-1_2012	22965–24618	S	KM027262_Riyadh-683-KSA_2014	99.5
KC667074	Enland-Qatar_2012	22602–24618	S	KF600620_Bisha-1_2012	99.5
KF600613	Riyadh-3_2013	22965–25594	S	KM027262_Riyadh-683-KSA_2014	99.4
KF600628	Hafr-Al-Batin-1_2013	231–1532	ORF1a	NC019843_HCoV_EMC_2012	99.8
KF958702	Jeddah-1_2013	22965–26708	S, ORF3, ORF4a, ORF4b, ORF5	KM027262_Riyadh-683-KSA_2014	99.5
KF961221	Qatar-3_2013	1–1532; 28497–29529	ORF1a, N, ORF8b	NC019843_HCoV-EMC_2012	99.8
KF961222	Qatar-4_2013	231–1532	ORF1a	NC019843_HCoV-EMC_2012	99.9
KJ156881	Wadi-Ad-Dawasir-1_2013	22965–25594	S	KM027262_Riyadh-683-KSA_2014	99.4
KJ156934	Riyadh-14_2013	231–1532	ORF1a	NC019843_HCoV-EMC_2012	99.9
KJ156944	Riyadh-5_2013	22965–26708	S, ORF3, ORF4a, ORF4b, ORF5	KM027262_Riyadh-683-KSA_2014	99.5
KJ156949	Taif-1_2013	22965–25594	S	KM027262_Riyadh-683-KSA_2014	99.5
KJ556336	Jeddah-1_2013	22963–26706	S, ORF3, ORF4a, ORF4b, ORF5	KM027262_Riyadh-683-KSA_2014	99.5
KJ829365	Florida-USA-2-Saudi Arabia_2014	22964–24397	S	KM027262_Riyadh-683-KSA_2014	99.3
KM027255	Jeddah-C7149-KSA_2014	22965–24397	S	KM027262_Riyadh-683-KSA_2014	99.4
KM027256	Jeddah-C7569-KSA_2014	22965–24397	S	KM027262_Riyadh-683-KSA_2014	99.4
KM027257	Jeddah-C7770-KSA_2014	22965–28690	S, ORF3, ORF4a, ORF4b, ORF5, E, M, N, ORF8b	KM027262_Riyadh-683-KSA_2014	99.7
KM027258	Jeddah-C8826-KSA_2014	22759–24397	S	KM027262_Riyadh-683-KSA_2014	99.4
KM027259	Jeddah-C9055-KSA_2014	22965–24397	S	KM027262_Riyadh-683-KSA_2014	99.4
KM027260	Jeddah-C10306-KSA_2014	22965–24397	S	KM027262_Riyadh-683-KSA_2014	99.4
KM027261	Makkah-C9355-KSA_2014	22965–24397	S	KM027262_Riyadh-683-KSA_2014	99.4
KP223131	Florida-USA-2-Saudi Arabia_2014	22964–24397	S	KM027262_Riyadh-683-KSA_2014	99.3
KT006149	China-GD01_2015	1–1532; 28054–29529	ORF1a, M, N, ORF8b	NC019843_HCoV-EMC_2012	99.8
		18033–23502	ORF1b, S	KF958702_Jeddah-1_2013	99.9
KT026455	Riyadh-KSA-2959_2015	1–1532; 28497–29529	ORF1a, N, ORF8b	NC019843_HCoV-EMC_2012	99.8
		17424–23502	ORF1b, S	KF958702_Jeddah-1_2013	99.9
KT026456	Riyadh-KSA-4050_2015	1–1532; 28497–29529	ORF1a, N, ORF8b	NC019843_HCoV-EMC_2012	99.7
		18033–23502	ORF1b, S	KF958702_Jeddah-1_2013	99.9
KT029139	KOR-KNIH-002_2015	1–1532; 28055–29520	ORF1a, M, N, ORF8b	NC019843_HCoV-EMC_2012	99.8
		18033–23502	ORF1b, S	KF958702_Jeddah-1_2013	99.9

Table 1. Putative recombinant strains (n = 25) detected by the RDP4 method with a default setting (window size = 30).

especially in Fig. 3e, genetic exchange appeared to have occurred among them and affected the phylogenetic evolution of MERS-CoV lineages before the Korean traveler was infected by a productive recombinant strain in the area[24]. However, as discussed previously with regard to the emergence of severe acute respiratory syndrome coronavirus (SARS-CoV) in 2002, other evolutionary aspects, such as mutation rates and selection pressure, should be considered to understand the evolutionary dynamics of MERS-CoV[21,25,26]. Possibly different molecular clock rates of MERS-CoV in animal hosts and humans may also have to be taken into account. As shown by the genomic evolution of influenza A viruses[27], MERS-CoV might experience different evolutionary courses in different hosts. To better understand these dynamics, the chain of MERS-CoV zoonotic transmissions should be further clarified.

Outside the Arabian Peninsula, Korea experienced the biggest outbreak of MERS. Through seeding by only a single patient, MERS-CoV resulted in more than 160 confirmed patients in less than a month and thousands of people were confined under close monitoring. The CFR of MERS outbreak in Korea may appear to be relatively low (approximately 11.7%), compared with the previous outbreaks in the Middle East and no signs of community transmission have been reported. In addition, an announcement regarding the situation assessment of MERS outbreak in Korea issued by the WHO Global Alert and Response program stated that significant virological change was not seen so far and the transmission patterns are unlikely to be different from those previously reported in the Middle East[16]. However, human infections with the MERS-CoV are ongoing in the Middle East countries, and the virus may travel anywhere from the region as seen in the current Korean outbreak and many other previous cases. In support of the struggle against the relatively new MERS-CoV infection, effective medical arsenals should be prepared using the comprehensive measures of epidemiology, pathogenesis, and transmission researches.

In conclusion, we suggest that the MERS-CoV outbreak in Korea appears to be caused by a strain that is closely related to three 2015 strains from the Hafr-Al-Batin-1 clade and that the relatively recent ancestor of these viruses exhibits a unique recombination pattern that is different from other putative recombinants.

Methods
Sequence preparation.   In this study, we investigated a total of 70 complete genome sequences of human MERS-CoV strains that were downloaded from the database of National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/genomes/VirusVariation/Database/nph-select2.cgi). Detailed information
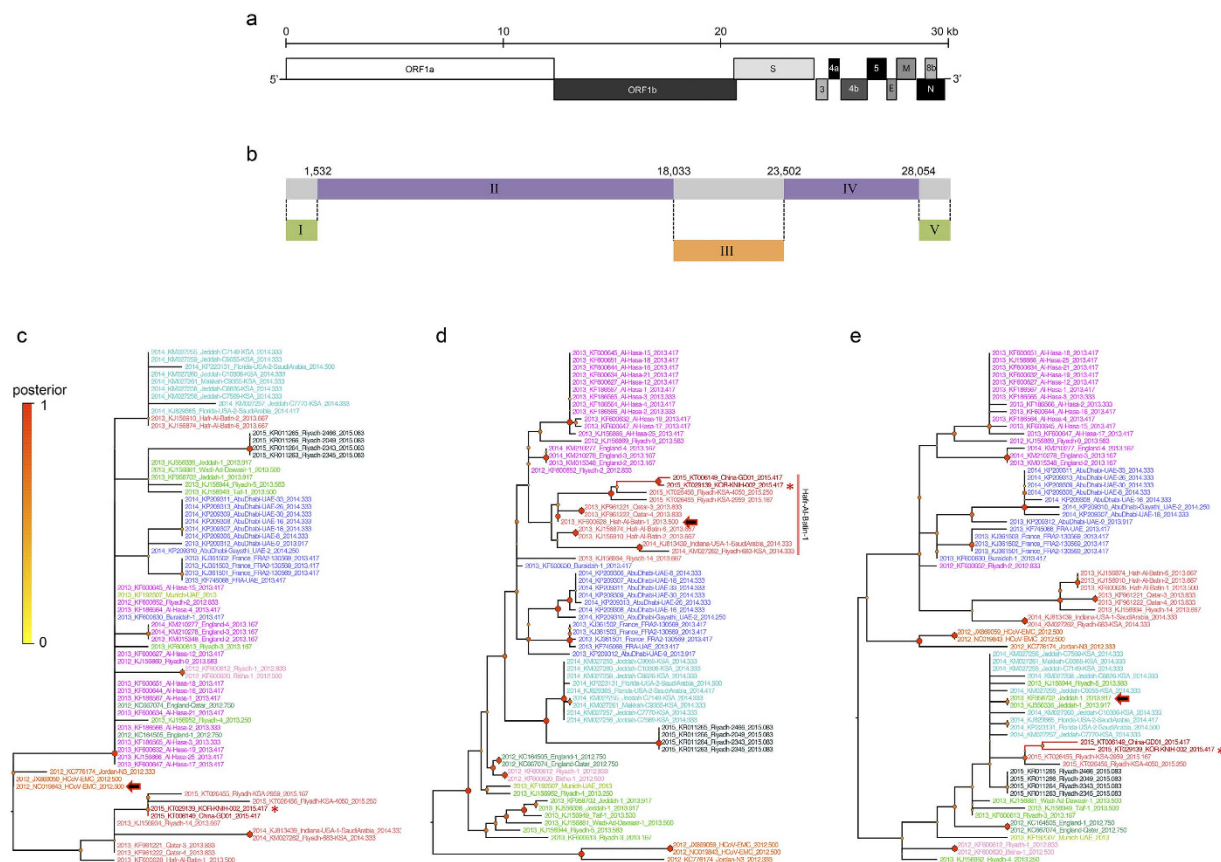
Figure 3. Schematic representation of MERS-CoV coding regions and putative recombinant regions detected in the KOR002 strain. The protein coding regions of MERS-CoV genome (**a**) were represented together with the putative recombinant regions of KOR002 strain (**b**) detected by the RDP method[36] using the complete genome sequences of 70 MERS-CoV strains. Phylogenetic relationships of MERS-CoV complete genomes were reconstructed according to each recombinant region. The trees of recombinant regions I (**c**), II (**d**), and III (**e**) were represented with the same color annotations in the Fig. 1. The tip of the Korean strain (KOR002) was denoted with the red color and asterisk. The tip of a parental strain of KOR002 identified (Table 1) was indicated with an arrow.

of isolation and genomic sequencing of a Korean MERS-CoV strain was published previously[28]. Briefly, a sputum sample was collected in May 20, 2015 from the wife of the index case, who travelled to four Middle East countries including the United Arab Emirates and the Kingdom of Saudi Arabia for 16 days before the onset of his symptom in May 11, 2015. After the third passage in Vero cells, viral RNA was obtained from cell culture media (QIAamp viral RNA mini kit: QIAGEN, Germany) and used for reverse-transcription PCR (Superscript III first-strand synthesis system: Life Technologies, the Netherlands)[29]. The pooled PCR amplicons were then fragmented (~ 300 bp in length) and used for the sequence library construction using an Illumina TruSeq Nano DNA sample prep kit (Illumina, San Diego, CA, USA). After sequencing the library using an Illumina MiSeq 50-bp single-end platform (Illumina), more than 2.6 million sequence reads out of 2,814,805 reads (approximately 93% usage) were mapped to the consensus of human-origin MERS-CoV genome sequences downloaded from GenBank using Bowtie version 2.2.4[30]. The complete genome sequence was obtained based on an average coverage of 3,605.95 and submitted to the NCBI database (accession number KT029139, MERS-CoV/KOR/KNIH/002_05_2015)[28]. The first Chinese strain of MERS-CoV (NCBI accession number KT006149, Middle East respiratory syndrome coronavirus strain ChinaGD01) appeared to be imported from a Korean traveler who visited to Guangdong Province and was sequenced directly from a nasopharyngeal sample (collected in May 27, 2015) using a NGS method[31]. The two most recent Saudi Arabian sequences in 2015 appeared to be obtained after passages in Vero cells of a tracheal aspirate (NCBI accession number KT026455, Hu/Riyadh_KSA_2959_2015)[32] and a respiratory swab (NCBI accession number KT026456, Hu/Riyadh_KSA_4050_2015)[33] samples using the NGS method. After alignment using the MAFFT program (v7.130b)[34], the datasets of complete genome and 10 protein coding regions were established by extracting corresponding sequence regions. The stop codon in the C-terminal region was removed. The resulting MERS-CoV datasets were assigned to complete genome (29,529 nucleotides, nts), ORF1ab (21,234 nts)), S (4,059 nts), ORF3 (309 nts), ORF4a (327 nts), ORF4b (738 nts), ORF5 (672 nts), E (246 nts), M (657 nts), N (1,239 nts), and ORF8b (336 nts). To compare phylogenetic clustering patterns, we also set up another dataset for the regions S through N coding regions (S-N region; 8,352 nts).

Phylogenetic trees and evolutionary dynamics.   Phylogenetic relationships, evolutionary rates (nucleotide substitutions/site/year), and the time (year) of the most recent common ancestor (tMRCA) were estimated using a time-framed Bayesian evolution analysis approach via a Markov Chain Monte Carlo (MCMC) inference method, implemented in the BEAST package (v1.8.2)[19]. We used the GTR+I+Γ substitution model, a lognormal relaxed molecular clock model and a Bayesian skygrid tree prior. For the of ORF4a, ORF4b, and ORF5 datasets, we used the HKY+Γ substitution model and a strict clock model. The evolutionary parameters (only for substitution and molecular clock parameters, not the tree model) were linked for the dataset of E coding region by adjoining those of the complete genome sequences. MCMC analyses were run for 50 million iterations, sampling every 25 thousand iterations after a 10% burn-in. Two or three independent runs for each dataset were combined and assessed to ensure their convergence in Tracer (v1.6)[35]. The MCMC tree samples were used to summarize a maximum clade credibility (MCC) trees for each dataset using TreeAnnotator v1.8.1, which were visualized using FigTree (v1.4.2). The estimates were presented as mean values along with the lower and upper limits of the 95% highest probability density (HPD).

Recombination analysis.   To detect putative recombinant regions in the MERS-CoV genome, we used the RDP4 program (v.4.39)[36] with a default (window size: 30 bp) and a higher window size of 1,000 bp, and the results obtained were confirmed by a manual bootscan method. Using the recombination breakpoints detected in the KOR002 strain by the default setting, we compiled new sequence datasets by dividing the complete genome sequences into five non-recombinant fragments. We subsequently reconstructed the phylogenetic relationships in each region using a maximum likelihood method (GTR+I+Γ, 500 bootstrap replication) implemented in MEGA5[37]. The trees were visualized using FigTree (v1.4.2).

References

1. Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. & Fouchier, R. A. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *The New England journal of medicine* **367,** 1814–1820, doi: 10.1056/NEJMoa1211721 (2012).
2. van Boheemen, S. *et al.* Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans. *mBio* **3,** doi: 10.1128/mBio.00473-12 (2012).
3. Bermingham, A. *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 Europeen sur les maladies transmissibles = European communicable disease bulletin* **17,** 20290 (2012).
4. MERS Corona Map by WHO and Health Ministries. Available at: http://coronamap.com (accessed: 16th June 2015).
5. The WHO MERS-CoV Research Group. State of knowledge and data gaps of Middle East respiratory syndrome coronavirus (MERS-CoV) in humans. doi: 10.1371/currents.outbreaks.0bf719e352e7478f8ad85 fa30127ddb8. *PLoS Currents Outbreaks* **5** (2013).
6. Assiri, A. *et al.* Epidemiological, demographic, and clinical characteristics of 47 cases of Middle East respiratory syndrome coronavirus disease from Saudi Arabia: a descriptive study. *The Lancet. Infectious diseases* **13,** 752–761, doi: 10.1016/S1473-3099(13)70204-4 (2013).
7. Drosten, C. *et al.* Clinical features and virological analysis of a case of Middle East respiratory syndrome coronavirus infection. *The Lancet. Infectious diseases* **13,** 745–751, doi: 10.1016/S1473-3099(13)70154-3 (2013).
8. Reusken, C. B. *et al.* Middle East respiratory syndrome coronavirus neutralising serum antibodies in dromedary camels: a comparative serological study. *The Lancet. Infectious diseases* **13,** 859–866, doi: 10.1016/S1473-3099(13)70164-6 (2013).
9. Al-Tawfiq, J. A. & Memish, Z. A. Middle East respiratory syndrome coronavirus: transmission and phylogenetic evolution. *Trends in microbiology* **22,** 573–579, doi: 10.1016/j.tim.2014.08.001 (2014).
10. Cotten, M. *et al.* Transmission and evolution of the Middle East respiratory syndrome coronavirus in Saudi Arabia: a descriptive genomic study. *Lancet* **382,** 1993–2002, doi: 10.1016/S0140-6736(13)61887-5 (2013).
11. Hemida, M. G. *et al.* Lack of middle East respiratory syndrome coronavirus transmission from infected camels. *Emerging infectious diseases* **21,** 699–701, doi: 10.3201/eid2104.141949 (2015).
12. Muller, M. A. *et al.* Presence of Middle East respiratory syndrome coronavirus antibodies in Saudi Arabia: a nationwide, cross-sectional, serological study. *The Lancet. Infectious diseases* **15,** 559–564, doi: 10.1016/S1473-3099(15)70090-3 (2015).
13. Gautret, P. *et al.* Emerging viral respiratory tract infections--environmental risk factors and transmission. *The Lancet. Infectious diseases* **14,** 1113–1122, doi: 10.1016/S1473-3099(14)70831-X (2014).
14. Bialek, S. R. *et al.* First confirmed cases of Middle East respiratory syndrome coronavirus (MERS-CoV) infection in the United States, updated information on the epidemiology of MERS-CoV infection, and guidance for the public, clinicians, and public health authorities - May 2014. *MMWR. Morbidity and mortality weekly report* **63,** 431–436 (2014).
15. WHO. Middle East respiratory syndrome coronavirus (MERS-CoV) in the Republic of Korea. Situation assessment on June 2, 2015. Available at: http://www.who.int/mediacentre/news/situation-assessments/2-june-2015-republic-of-korea/en/ (accessed: 17th June 2015).
16. WHO. Middle East respiratory syndrome coronavirus (MERS-CoV): Summary and Risk Assessment of Current Situation in the Republic of Korea and China. Available at: http://www.who.int/emergencies/mers-cov/mers-cov-republic-of-korea-and-china-risk-assessment-19-june-2015.pdf?ua=1 (accessed: 19th June 2015).
17. Assiri, A. *et al.* Hospital outbreak of Middle East respiratory syndrome coronavirus. *The New England journal of medicine* **369,** 407–416, doi: 10.1056/NEJMoa1306742 (2013).
18. Oboho, I. K. *et al.* 2014 MERS-CoV outbreak in Jeddah--a link to health care facilities. *The New England journal of medicine* **372,** 846–854, doi: 10.1056/NEJMoa1408636 (2015.).
19. Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular biology and evolution* **29,** 1969–1973, doi: 10.1093/molbev/mss075 (2012).
20. Cotten, M. *et al.* Spread, circulation, and evolution of the Middle East respiratory syndrome coronavirus. *mBio* **5,** doi: 10.1128/mBio.01062-13 (2014).
21. Holmes, E. C. & Rambaut, A. Viral evolution and the emergence of SARS coronavirus. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **359,** 1059–1065, doi: 10.1098/rstb.2004.1478 (2004).
22. Graham, R. L. & Baric, R. S. Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. *Journal of virology* **84,** 3134–3146, doi: 10.1128/JVI.01394-09 (2010).
23. Zhang, Y. *et al.* Genotype shift in human coronavirus OC43 and emergence of a novel genotype by natural recombination. *The Journal of infection* **70,** 641–650, doi: 10.1016/j.jinf.2014.12.005 (2015).
24. Corman, V. M. *et al.* Rooting the phylogenetic tree of middle East respiratory syndrome coronavirus by characterization of a conspecific virus from an African bat. *Journal of virology* **88,** 11297–11303, doi: 10.1128/JVI.01498-14 (2014).

25. Stanhope, M. J., Brown, J. R. & Amrine-Madsen, H. Evidence from the evolutionary analysis of nucleotide sequences for a recombinant history of SARS-CoV. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases* **4,** 15–19, doi: 10.1016/j.meegid.2003.10.001 (2004).

26. Holmes, E. C. & Drummond, A. J. The evolutionary genetics of viral emergence. *Current topics in microbiology and immunology* **315,** 51–66 (2007).

27. Worobey, M., Han, G. Z. & Rambaut, A. A synchronized global sweep of the internal genes of modern avian influenza virus. *Nature* **508,** 254–257, doi: 10.1038/nature13016 (2014).

28. Kim, Y. J. *et al.* Complete Genome Sequence of Middle East Respiratory Syndrome Coronavirus KOR/KNIH/002_05_2015, Isolated in South Korea. *Genome announcements* **3,** doi: 10.1128/genomeA.00787-15 (2015).

29. Cotten, M. *et al.* Full-genome deep sequencing and phylogenetic analysis of novel human betacoronavirus. *Emerging infectious diseases* **19,** 736–742B, doi: 10.3201/eid1905.130057 (2013).

30. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9,** 357–359, doi: 10.1038/nmeth.1923 (2012).

31. Lu, R. *et al.* Complete Genome Sequence of Middle East Respiratory Syndrome Coronavirus (MERS-CoV) from the First Imported MERS-CoV Case in China. *Genome announcements* **3,** doi: 10.1128/genomeA.00818-15 (2015).

32. Queen, K. *et al.* Middle East respiratory syndrome coronavirus isolate Hu/Riyadh_KSA_2959_2015, complete genome. Available at: http://www.ncbi.nlm.nih.gov/nuccore/KT026455 (accessed: 16th June 2015). (2015).

33. Queen, K. *et al.* Middle East respiratory syndrome coronavirus isolate Hu/Riyadh_KSA_4050_2015, complete genome Available at: http://www.ncbi.nlm.nih.gov/nuccore/KT026456 (accessed: 16th June 2015). (2015).

34. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* **30,** 772–780, doi: 10.1093/molbev/mst010 (2013).

35. Rambaut, A., Suchard, M. A. & Drummond, A. J. Tracer v1.6. Available at: http://tree.bio.ed.ac.uk/software/tracer (accessed: 16th June 2015).

36. Martin, D. P., Murrell, B., Golden, M., Khoosal, A. & Muhire, B. RDP4: Detection and analysis of recombination patterns on virus genomes. *Virus Evolution* **1,** doi: doi: 10.1093/ve/vev003 (2015).

37. Tamura, K. *et al.* MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution* **28,** 2731–2739, doi: 10.1093/molbev/msr121 (2011).

Acknowledgements

Author Contributions

M.-S.P. conceived and designed the experiments, J.I.K., Y.-J.K., P.L., J.-Y.B., D.K., I.L., S.P., H.K., S.-I.J., J.-S.Y. and H.K. performed the experiments, J.I.K., Y.-J.K., P.L., D.-W.K., J.-G.N., S.S.K., K.K., J.M.L., M.K.S., D.S., J.C., K.-J.H., Y.-S.B., J.-W.S., J.-S.L. and M.-S.P. analyzed the data, J.I.K., Y.-J.K., P.L. and M.-S.P. wrote the manuscript. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Kim, J. I. *et al.* The recent ancestry of Middle East respiratory syndrome coronavirus in Korea has been shaped by recombination. *Sci. Rep.* **6,** 18825; doi: 10.1038/srep18825 (2016).