Spread, Circulation, and Evolution of the Middle East Respiratory Syndrome Coronavirus

Matthew Cotten, Simon J. Watson, Alimuddin I. Zumla, et al.

2014. Spread, Circulation, and Evolution of the Middle East Respiratory Syndrome Coronavirus . mBio 5(1): . doi:10.1128/mBio.01062-13.

RESEARCH ARTICLE

Spread, Circulation, and Evolution of the Middle East Respiratory Syndrome Coronavirus

Matthew Cotten,[a] Simon J. Watson,[a] Alimuddin I. Zumla,[b,c,d] Hatem Q. Makhdoom,[e] Anne L. Palser,[a] Swee Hoe Ong,[a] Abdullah A. Al Rabeeah,[b] Rafat F. Alhakeem,[b] Abdullah Assiri,[b] Jaffar A. Al-Tawfiq,[f] Ali Albarrak,[g] Mazin Barry,[h] Atef Shibl,[h] Fahad A. Alrabiah,[i] Sami Hajjar,[i] Hanan H. Balkhy,[j] Hesham Flemban,[k] Andrew Rambaut,[l,m] Paul Kellam,[a,c,d] Ziad A. Memish[b,n]

Wellcome Trust Sanger Institute, Hinxton, United Kingdom[a]; Global Centre for Mass Gatherings Medicine (GCMGM), Ministry of Health, Riyadh, Kingdom of Saudi Arabia[b]; Department of Medical Microbiology, University College London Hospitals NHS Foundation Trust, London, United Kingdom[c]; Division of Infection and Immunity, University College London, London, United Kingdom[d]; Jeddah Regional Laboratory, Ministry of Health, Jeddah, Kingdom of Saudi Arabia[e]; Saudi Aramco Medical Services Organization, Saudi Aramco, Dhahran, Kingdom of Saudi Arabia[f]; Prince Sultan Military Medical City, Riyadh, Kingdom of Saudi Arabia[g]; King Saud University, Riyadh, Kingdom of Saudi Arabia[h]; King Faisal Specialist Hospital, Riyadh, Kingdom of Saudi Arabia[i]; King Abdulaziz Medical City, Riyadh, Kingdom of Saudi Arabia[j]; Alhada Military Hospital, Riyadh, Kingdom of Saudi Arabia[k]; Institute of Evolutionary Biology, Ashworth Laboratories, Kings Buildings, West Mains Road, Edinburgh, United Kingdom[l]; Fogarty International Center, NIH, Bethesda, Maryland, USA[m]; College of Medicine, Alfaisal University, Riyadh, Kingdom of Saudi Arabia[n]

M.C., S.J.W., A.I.Z., P.K., and Z.A.M. contributed equally to this work.

ABSTRACT **The Middle East respiratory syndrome coronavirus (MERS-CoV) was first documented in the Kingdom of Saudi Arabia (KSA) in 2012 and, to date, has been identified in 180 cases with 43% mortality. In this study, we have determined the MERS-CoV evolutionary rate, documented genetic variants of the virus and their distribution throughout the Arabian peninsula, and identified the genome positions under positive selection, important features for monitoring adaptation of MERS-CoV to human transmission and for identifying the source of infections. Respiratory samples from confirmed KSA MERS cases from May to September 2013 were subjected to whole-genome deep sequencing, and 32 complete or partial sequences (20 were ≥99% complete, 7 were 50 to 94% complete, and 5 were 27 to 50% complete) were obtained, bringing the total available MERS-CoV genomic sequences to 65. An evolutionary rate of 1.12×10^{-3} substitutions per site per year (95% credible interval [95% CI], 8.76×10^{-4}; 1.37×10^{-3}) was estimated, bringing the time to most recent common ancestor to March 2012 (95% CI, December 2011; June 2012). Only one MERS-CoV codon, spike 1020, located in a domain required for cell entry, is under strong positive selection. Four KSA MERS-CoV phylogenetic clades were found, with 3 clades apparently no longer contributing to current cases. The size of the population infected with MERS-CoV showed a gradual increase to June 2013, followed by a decline, possibly due to increased surveillance and infection control measures combined with a basic reproduction number (R_0) for the virus that is less than 1.**

IMPORTANCE **MERS-CoV adaptation toward higher rates of sustained human-to-human transmission appears not to have occurred yet. While MERS-CoV transmission currently appears weak, careful monitoring of changes in MERS-CoV genomes and of the MERS epidemic should be maintained. The observation of phylogenetically related MERS-CoV in geographically diverse locations must be taken into account in efforts to identify the animal source and transmission of the virus.**

Received 9 December 2013 **Accepted** 16 January 2014 **Published** 18 February 2014

Citation Cotten M, Watson SJ, Zumla AI, Makhdoom HQ, Palser AL, Ong SH, Al Rabeeah AA, Alhakeem RF, Assiri A, Al-Tawfiq JA, Albarrak A, Barry M, Shibl A, Alrabeah FA, Hajjar S, Balkhy HH, Flemban H, Rambaut A, Kellam P, Memish ZA. 2014. Spread, circulation, and evolution of the Middle East respiratory syndrome coronavirus. mBio 5(1):e01062-13. doi:10.1128/mBio.01062-13.

Editor Michael Katze, University of Washington

Address correspondence to Ziad A. Memish, zmemish@yahoo.com.

The Middle East Respiratory Syndrome Coronavirus (MERS-CoV) was first detected in the Kingdom of Saudi Arabia (KSA) in 2012 (1–4), and to date, infection with the virus has been identified in 180 patients with 43% mortality (5). Previously, the SARS coronavirus emerged from an animal reservoir (6), and a zoonotic event may also provide the source of MERS-CoV; however, no consistent pattern of animal exposure has been observed with MERS cases. Serological studies have identified a high prevalence of MERS-CoV reactive antibodies in camels in Oman, the Canary Islands, and Egypt (7, 8), and fragments of MERS-CoV sequence have been reported from bats (9) and camels (10). However, to date, MERS-CoV itself has not been isolated from any nonhuman source. If such an animal reservoir exists, MERS-CoV epidemiology could be explained by intermittent animal-to-human transmission seeding clusters of human-to-human transmission, but with a reproduction number (R_0) of less than 1 (11, 12), these clusters eventually disappear. An alternative hypothesis is that the virus has now infected a sufficient number of humans to account for the observed distribution and diversity of the virus but the infection is asymptomatic in many individuals. A recent serosurvey of 363 individuals in the Saudi Arabia failed, however, to find MERS-CoV-seropositive individuals (13).

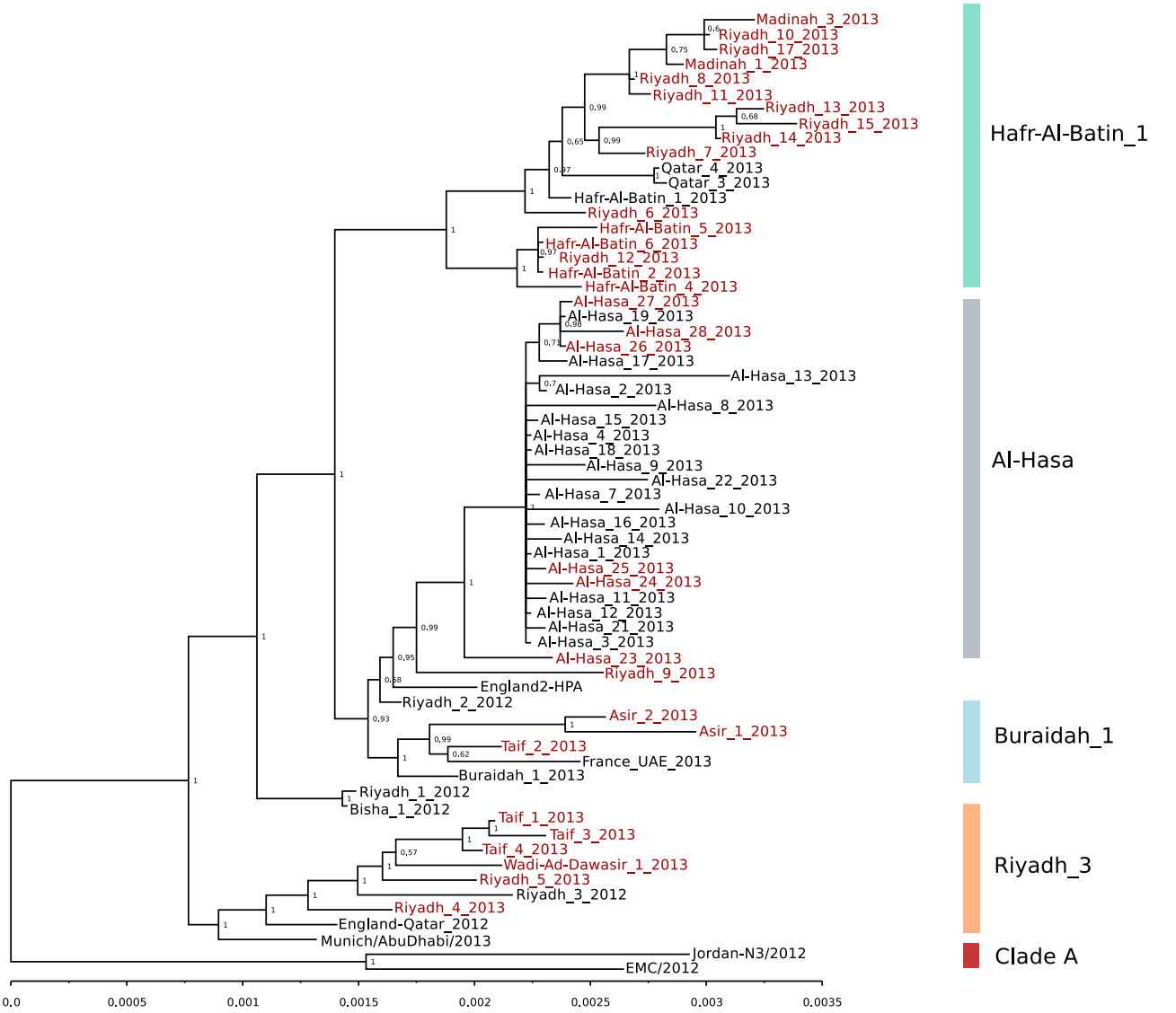

FIG 1 Bayesian-inferred phylogeny of the 32 new MERS-CoV sequences combined with the 33 previously available genomes (EMC/2012 [JX869059], Jordan_N3 [KC776174], Munich_AbuDhabi_2013 [KF192507], England-Qatar_2012 [KC667074], Al-Hasa_1_2013 [KF186567], Al-Hasa_2_2013 [KF186566], Al-Hasa_3_2013 [KF186565], Al-Hasa_4_2013 [KF186564, plus all previously published MERS-CoV sequences 17], England2-HPA [http://www.hpa.org.uk/Topics/InfectiousDiseases/InfectionsAZ/MERSCoV/respPartialgeneticsequenceofnovelcoronavirus/], France_UAE_2013 [KF745068], Qatar_3_2013 [KF961221], and Qatar_4_2013 [KF961222]). All new genome sequences from this study are labeled in red. Clades are marked with vertical bars on the right and (with the exception of clade A and the Al-Hasa clade) named by the initial genome in the clade. The scale bar indicates the genetic distance, in substitutions per site, from the arbitrary midpoint root. Bayesian posterior probabilities for each clade are listed above the relevant node.

A detailed description of MERS-CoV evolution is useful to assess public health risks, to help identify the source of new infections, and to detect viral adaptation to human transmission. In this report, we advance our knowledge of the MERS-CoV outbreak with complete or partial MERS-CoV genome sequences obtained directly from 32 recent MERS patient samples from cases between July and September 2013, bringing the total available MERS-CoV genomic sequences to 65 (37% of the 178 MERS cases reported globally).

RESULTS

Phylogenetic analysis. All PCR-confirmed MERS case samples from Saudi Arabia were processed for whole-genome deep sequencing (14, 15), adding 32 new MERS-CoV genome sequences to the publically available data set. The phylogenetic relationship of all MERS-CoV genomes was inferred from the 33 previously published genomes (2, 10, 14–16) and 32 new sequences (Fig. 1). The previously described Al-Hasa clade (17) has expanded, with 6 new members. The Riyadh_3 clade, which includes virus from a Qatari patient diagnosed in London (15) and a United Arab Emirates patient diagnosed in Munich (16) (Fig. 1), has increased to 9 members since the previous report (17) and includes new viruses from Riyadh, Wadi-Ad-Dawasir, and Ta'if. The Buraidah_1 variant (Fig. 1), first observed with Buraidah_1_2013, has now expanded to include a virus from Ta'if, two viruses from Khamis Mushait in the southern province of Asir, and the UAE_Dubai

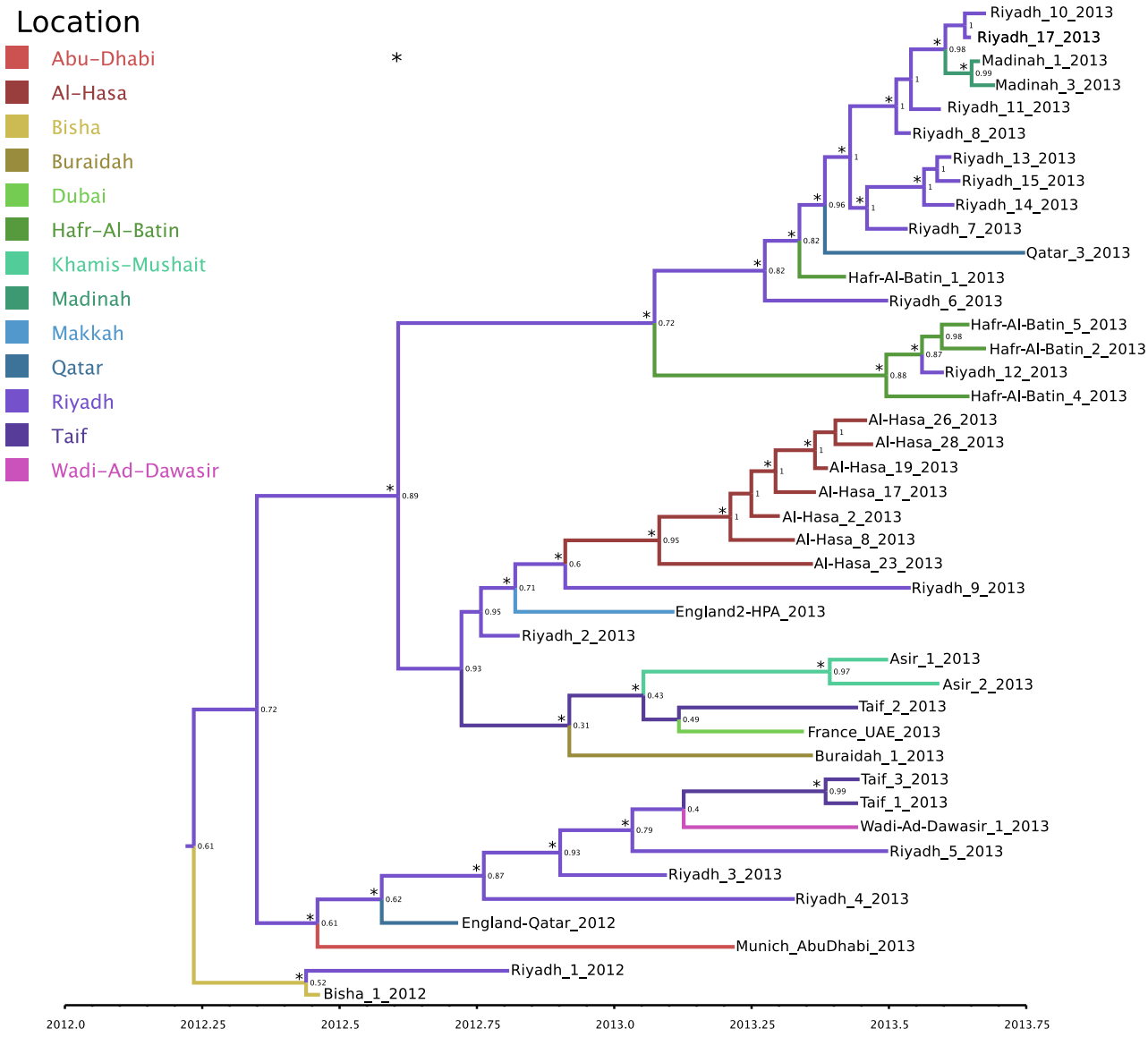

FIG 2 Time-resolved phylogenetic tree of all concatenated coding regions of the 42 phylogenetically distinct MERS-CoV genomes (see Materials and Methods for further details). A discrete traits model implemented in BEAST version 1.7.5 (36) was used to determine the most probable geographical location for each branch; a change in branch color indicates a geographical location change during its evolutionary history. Posterior probabilities for the inferred geographical locations are indicated at the nodes, an asterisk at a node indicates a posterior probability of >0.9 for that clade, and time is indicated on the *x* axis.

_France_patient_1 virus identified in a United Arab Emirates (UAE) patient in Valenciennes, France.

Most of the new genomes cluster with the previous singleton Hafr-Al-Batin_1_2013 genome, which appeared in the northeast of Saudi Arabia on 4 June 2013 (Fig. 1). The later Hafr-Al-Batin cases include a family cluster of MERS cases. Sequences were obtained from three contacts of the index case (Hafr-Al-Batin_4, Hafr-Al-Batin_5, and Hafr-Al-Batin_6) and a contact of Hafr-Al-Batin_6 (Hafr-Al-Batin_2). The four contact sequences cluster together. The close similarity between the Hafr-Al-Batin clade viruses and Riyadh_12_2013 (Fig. 1) indicates a possible link between these cases that has not been revealed epidemiologically. Viruses from recent Madinah cases (Madinah_1_2013 and Madinah_3_2013) and three Riyadh viruses (Riyadh_13-2013, Riyadh_14_2013, and Riyadh_15_2013) cluster closely. In addition, two virus genomes from Qatar MERS patients in October 2013 (10) also cluster in the Hafr-Al-Batin_1 clade. No additional genomes were found in clade A or in the Bisha_1/Riyadh_1 clade.

A time-resolved phylogeny was generated from all epidemiologically unlinked viruses with genome coverage of >30%. The geographical locations of the ancestral viruses were coestimated and marked by color coding in the phylogenetic tree (Fig. 2), leading to a prediction that the ancestors of most of the viral clades originated in Riyadh.

Evolutionary rate. A critical feature of an emerging virus is how quickly it is changing. The evolutionary rate for the updated set of 42 epidemiologically unlinked MERS-CoV genomes was estimated as 1.12×10^{-3} substitutions per site per year (95% credible interval [95% CI], 8.76×10^{-4}; 1.37×10^{-3}), bringing the time to most recent common ancestor (tMRCA) for clade B

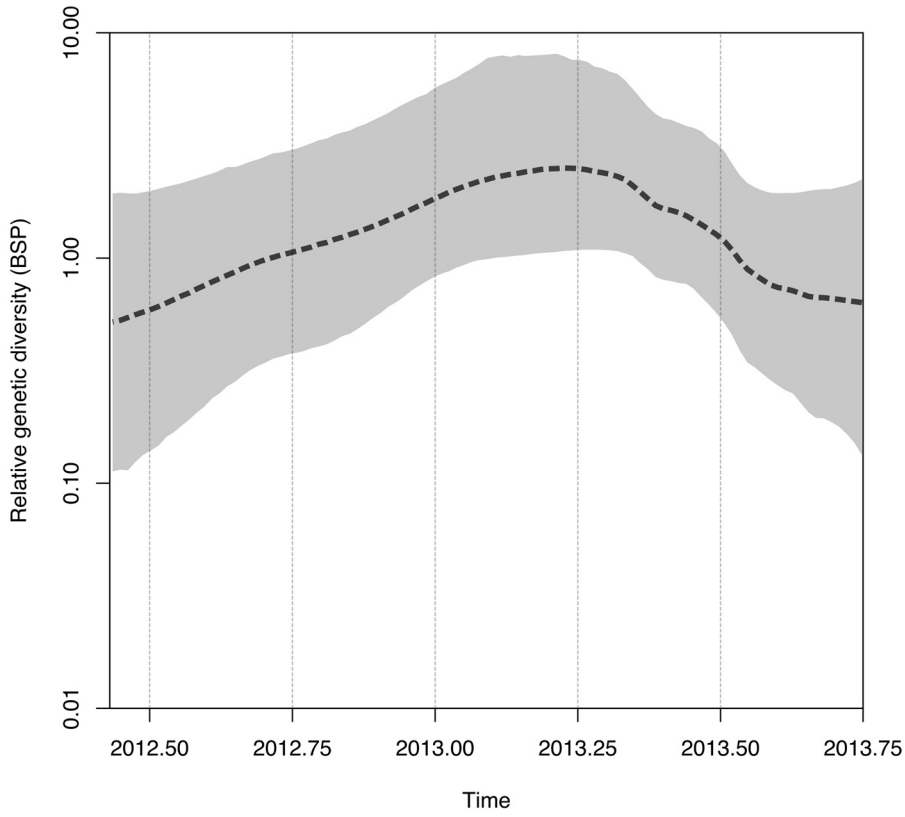

FIG 3 Bayesian skyline plot (BSP) showing the changes in effective population size of MERS-CoV across time. The dashed black line indicates the median population size estimated from the BMCMC used in the inference of the time-resolved phylogeny (see Fig. 2 and Materials and Methods). The gray shading indicates the 95% highest posterior density of the estimated population size.

(all MERS-CoV except clade A) to March 2012 (95% CI, December 2011; June 2012). This is within the credible interval bounds of the previous estimation (17). Two codon positions in the MERS-CoV genome exhibit evidence of episodic selection using mixed effects model of evolution (MEME; see Materials and Methods), spike codon 1020 ($P = 0.014$) and, more weakly, spike codon 158 ($P = 0.059$). Furthermore, under an alternative selection analysis method (fast unconstrained Bayesian approximation [FUBAR]), spike codon 509 is suggested to be under positive selection.

Population size. An estimation of the relative change in the population size of MERS-CoV over time was made from the Gaussian Markov random field (GMRF) Bayesian Skyride coalescent model (18), employed to infer the time-resolved phylogeny. The Bayesian skyline plot (BSP) (Fig. 3) shows that after the first documented MERS case in June 2012, the relative MERS-CoV population size (i.e., the relative number of infections) increased gradually, reaching a plateau at around April 2013. Since then, the effective viral population size has decreased, reflecting the apparent disappearance of multiple lineages (Riyadh_3, Buraidah_1, and Al-Hasa) (Fig. 4A). Plotting genomes by clade and sample time (Fig. 4A) shows that the viral clades appear limited in time, although we note a long time interval between the beginning and end of the Riyadh_3 cluster, suggesting the existence of undetected cases. Under the assumption of limited missing cases, the average time of existence (last observed date to first observed date) (Fig. 4A, see the legend) is 98 days for the four clades, although the last variant, Hafr-Al-Batin_1, was still in circulation at the end of the observation period. All 9 of the recently identified viruses from Riyadh are from the Hafr-Al-Batin_1 clade, and no further Riyadh_3 variants have appeared in Riyadh.

Geography. The variations of MERS-CoV genome sequences combined with sample collection dates and locations can help identify the source of new MERS-CoV infections. Four MERS-CoV monophyletic lineages containing 4 or more cases and persisting for 2 months or more have been detected (Al-Hasa, Riyadh_3, Buraidah_1, and Hafr-Al-Batin_1), and there are 6 sporadic viruses from Bisha, Riyadh, Makkah, and Al Zarqa, Jordan (Fig. 1 and 4B). The geographical locations of all available MERS-CoV genomic sequences, labeled by clade, and the sporadic viruses (clade size, <4) are plotted in Fig. 4B. The Al-Hasa variants (Fig. 4B, gray circles) were not detected in any other part of Saudi Arabia, and the Al-Hasa region has remained free of other virus variants, indicating that the Al-Hasa virus source was constrained to the Al-Hasa region. The more recently emerged Hafr-Al-Batin_1 variant (Fig. 4B, green circles), is now found in three KSA locations (Riyadh, Hafr-Al-Batin, and Madinah), as well as in Qatar. Riyadh_3 viruses (Fig. 4B, orange circles) are geographically dispersed and were found in Riyadh, Wadi Ad-Dawasir, and Ta'if in Saudi Arabia, as well as Qatar/London (15) and Abu Dhabi/Munich (16). The Buraidah_1 clade (Fig. 4B, blue circles) has appeared in Buraidah, Ta'if, Musayt (in the southern province of Asir), and in a patient from Dubai, United Arab Emirates, in Valenciennes, France (19). The geographical dispersion of MERS-CoV lineages suggests a mobile infection source, either as human-

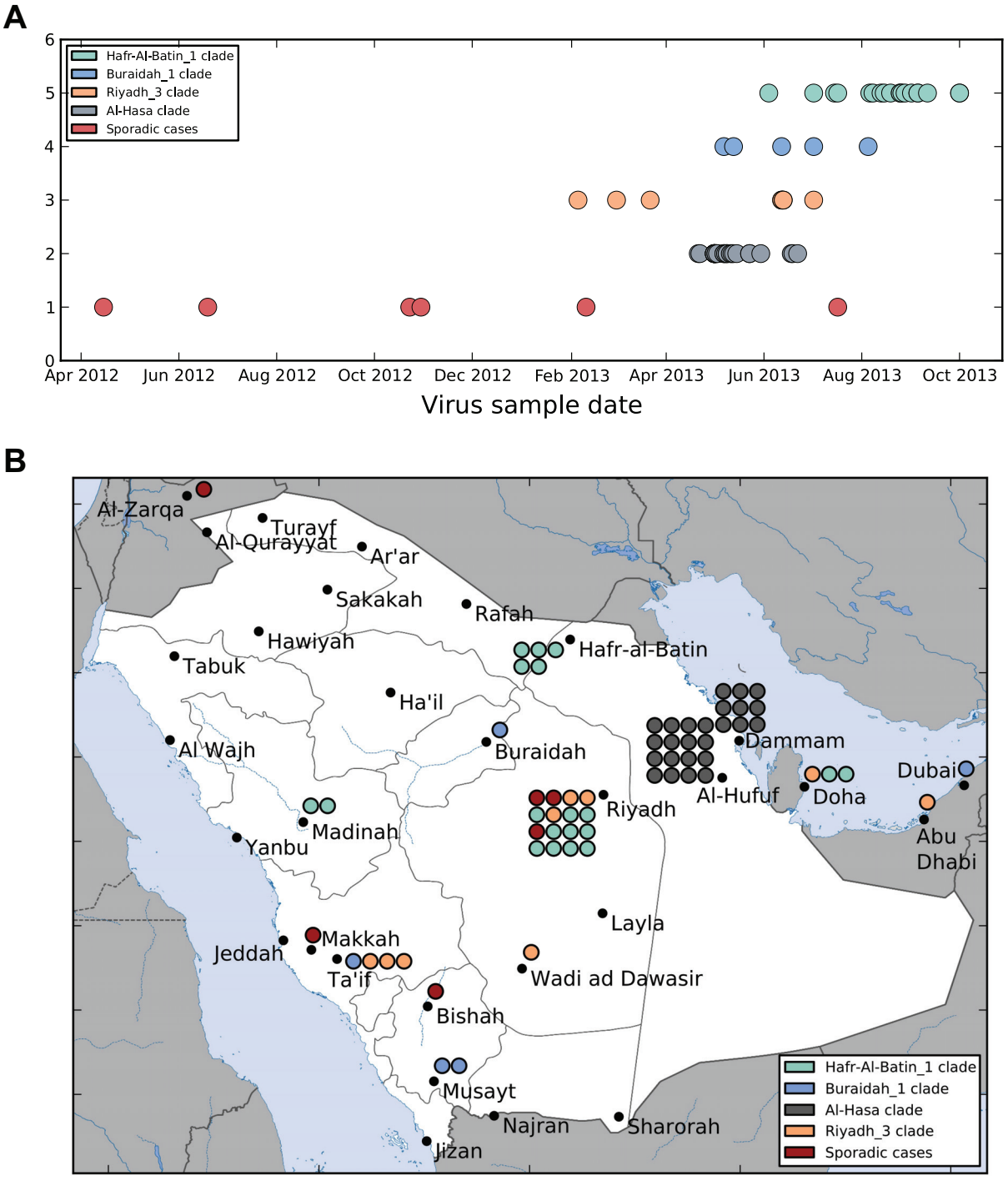

FIG 4 Distribution of MERS-CoV clades in time and space. (A) All available MERS-CoV genomes were stratified by phylogenetic clade (see Fig. 1) and plotted by virus sample date. The length of each clade was determined as the difference in days between the first and last observed sample of that virus and yielded the following values: Al-Hasa (21 April 2013 to 22 June 2013; 62 days), Riyadh_3 (5 February 2013 to 2 July 2013; 147 days), Buraidah_1 viruses (3 May 2013 to 5 August 2013; 84 days), and Hafr-Al-Batin_1 (4 June 2013 to 01 October 2013; 119 days). (B) All available MERS-CoV genomes were stratified by phylogenetic clade (see Fig. 1) and plotted by the case location. Cities are indicated by small black circles, and sequenced viruses by larger circles colored according to phylogenetic clade.

to-human or nonhuman-to-human infections or via transported animal product.

Protein changes in MERS-CoV. The coding regions of the viral genome are evolving at an average rate of 1.12×10^{-3} substitutions per site per year. Substitutions can be nonuniformly distributed, with coding regions constrained by protein function and regions exposed to host innate or adaptive immune responses showing greater levels of substitution. It is important to monitor MERS-CoV amino acid substitutions that could signal adaptation to human transmission, especially in proteins at the virus-host

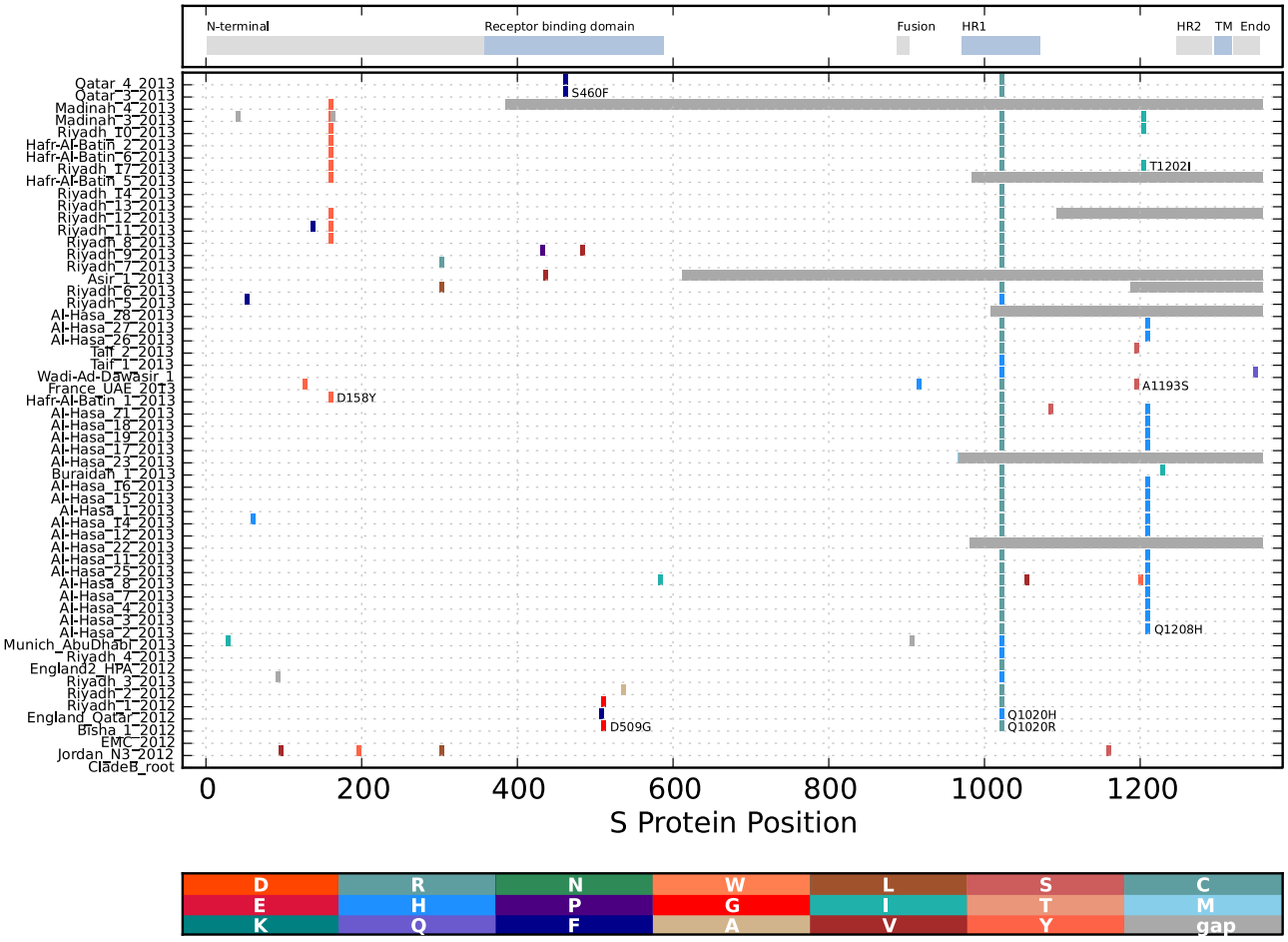

FIG 5 Substitutions in MERS-CoV spike proteins. All available KSA MERS-CoV spike ORFs were translated, the proteins aligned, and amino acid differences from the reconstructed ancestral clade B protein determined; changes observed in more than one genome are marked by vertical colored bars, with the new amino acid residue coded as shown at the bottom. Gray bars indicate a gap in sequence coverage. Functional domains of the spike (S) protein are marked and include the N-terminal domain, the receptor binding domain, the fusion domain (Fusion), heptad repeats 1 and 2 (HR1 and HR2) (20, 42), the transmembrane (TM) domain, and the cytoplasmic (Endo) domain (43).

interface. Changes in all MERS-CoV spike proteins are shown in Fig. 5. Positive selection analysis using the MEME method revealed that spike codon 1020 is under episodic selection, and using the FUBAR method, codon 509 is suggested to be under modest positive selection (see Materials and Methods). The codon 1020 substitution is in heptad repeat 1 (HR1) of the spike protein (Fig. 5; see also Fig. S1 in the supplemental material, right panel), which may influence the membrane fusion activity of the spike protein (20). MERS-CoV genomes in the Al-Hasa and Hafr-Al-Batin_1 clades encode an arginine at this position, while the Riyadh_3 clade genome encodes a histidine. Nine genomes show amino acid substitutions in the receptor-binding domain (RBD) of the spike protein, including a recent genome, Riyadh_9, which has two amino substitutions in the RBD (Fig. 5), and the two recent Qatar genomes. Using a reported crystal structure of the human coronavirus Erasmus Medical Center/2012 (EMC/2012) RBD in complex with the human receptor dipeptidyl peptidase 4 (DDP4) (21) complex (Protein Data Bank [PDB] ID 4L72), nonsynonymous mutations are observed in buried spike protein residues 482, 506, and 534; all are conservative changes in terms of their amino acid properties. A change of aspartic acid to glycine at codon 509 was observed in the Riyadh_1 and Bisha_1 genomes, and this position was found to be under modest positive selection. This residue is not part of, but is immediately adjacent to, the spike-DPP4 binding interface (Fig. S1, left). However, none of the changes in the RBD have been observed in multiple genomes, suggesting limited transmission. Five amino acid substitutions persist in multiple viruses (D158Y, Q1020R or Q1020H, T1202I, Q1208H, and S460F) (Fig. 5), suggesting a neutral or positive consequence of the variant for the virus. These include a Hafr-Al-Batin clade variant with both D158Y and Q1020R. These combined changes first appeared in Riyadh_8 and Hafr-Al-Batin_1 and are also present in the later viruses Hafr-Al-Batin_2, 5, and 6, Riyadh_10, 11, 12, and 17, and Madinah_3. The S460F change in two recent Qatar genomes is close to the spike-DPP4 binding interface. None of these changes reach significance when examined by all positive selection algorithms.

DISCUSSION

The study reported here significantly extends our previous report on 21 MERS-CoV genomes and the observation of three genetically distinct lineages of MERS-CoV circulating in Riyadh. We

concluded previously that it was unlikely that the Riyadh infections were the result of a single continuous human-to-human transmission chain (17) and suggested that transmission within Saudi Arabia was consistent with either movement of an animal reservoir or animal products or movement of infected humans. We now present additional data from 32 new MERS-CoV genomes which show that 4 phylogenetic clades of viruses have been observed and 3 of these clades were no longer detected in cases at the end of the current observation period. This pattern of clade disappearance may be due to the increased MERS surveillance and patient isolation that was implemented during the course of the outbreak (14), combined with an R_0 of less than 1 (11, 12), but it could also reflect undiagnosed asymptomatic spread, and we note the extended pattern of the Riyadh_3 cluster.

Adaptation of a zoonotic virus to a new host often requires sustained replication of the virus in the new host for the selection of amino acid changes that favor transmission. We find only limited evidence of adaptation to human transmission in the form of positively selected amino acids in MERS-CoV lineages. However, none of the MERS-CoV clades have been observed to persist beyond 2 to 3 months, and thus, sustained human transmission may not have occurred yet with MERS-CoV, although with the most recent MERS-CoV Hafr-Al-Batin_1 variant, mortality has been observed in two young healthy patients. It is essential that careful monitoring of virus lineages and genome changes in the epidemic is maintained and that the functional consequences of these substitutions in the spike and other viral proteins should be examined.

The spike amino acid changes to either arginine in the Hafr-Al-Batin clade or histidine in the Riyadh_3 clade codon 1020 are not predicted to change the alpha helical structure of this region (Fig. S1, right); however, the histidine provides an endosomal protonated residue and the arginine provides a potential endosomal protease cleavage site; either of these changes might alter the fusion function of this motif. The combination of HR1 with heptad repeat 2 (HR2) and the fusion domain are essential components of the fusion mechanism of the coronavirus spike protein and allow passage of the virus across the endosomal membrane (20). Changes in HR1 are associated with host range expansion of murine hepatitis virus (22). The external orientation of the spike protein may expose it to immune selection, and such changes are important information when designing reagents for serological testing. Changes in the coronavirus spike have been reported to accompany coronavirus host switches (22) and the SARS coronavirus adaptation to humans (23–25), and such changes should be monitored for their effects on the receptor binding and transmission properties of the virus. In particular, spike changes of D158Y, D509G, Q1020R Q1020H, T1202I, and Q1208H should be tested for altered biological properties.

The MERS-CoV-encoded enzymes are obvious targets for antiviral drugs, and screening efforts should use viral enzymes representative of the currently circulating forms of the virus. The major 3C protease, required for multiple cleavages of the replicase polyproteins, shows a high level of conservation, with only three nonsynonymous changes observed across all known MERS-CoV. The viral papain-like protease (PLP) is required for cleavage of the open reading frame 1A (ORF1a) polyprotein and may antagonize host immune signaling (26, 27). The Al-Hasa lineage shows a sustained A160S substitution in PLP, while the later viruses in the Hafr-Al-Batin lineage also have an R911C substitution in PLP, close to the catalytic CHD triad. In addition, position 90 shows substitutions in Jordan_N3_2012 (K90G) and Wadi-Ad-Dawasir_1, Taif_1_2013, and Taif_4_2013 (K90E), and the changes may be relevant for enzyme activity. The viral ADP-ribose-1″-monophosphatase (ADRP) has two conserved domains required for activity: VNAAN at positions 290 to 294 and GIF at 384 to 386. Wadi-Ad-Dawasir_1 virus shows a change to VNAVN, and a number of sustained amino acid substitutions have occurred in the amino half of ADRP.

Considerable effort has been made to determine an animal source for MERS-CoV. To date, serological evidence for a cross-reactive virus in camels has been reported (7, 8), and a small fragment of MERS-CoV sequence has been identified in a bat from Saudi Arabia (9). Recently, a camel in contact with a case in Saudi Arabia tested positive for MERS-CoV by PCR (28); however, multiple attempts at deep sequencing failed to yield convincing MERS-CoV sequences from the 2 camel nasal samples, despite the availability of a complete genome obtained from the patient (M. Cotten, S. J. Watson, P. Kellam, H. Q. Makhdoom, Z. A. Memish, unpublished results). More recently, 5 fragments of sequence were obtained from a camel cared for by a MERS patient in Qatar (10); these fragments were phylogenetically related to whole-genome sequences of MERS-CoV from two patients in contact with the camel (Qatar_3_2013 and Qatar_4_2013) (Fig. 1), providing support for MERS-CoV infection in camels and suggesting camels as an animal reservoir for the virus. Zoonotic movements from an animal reservoir to humans have occurred with the SARS coronavirus (6, 25, 29). It is unclear to what extent "chatter" occurs between such an animal reservoir and humans before a purely human infection becomes sustained. The strongest argument for a persistent animal reservoir may be that the occurrence of MERS-CoV infections in multiple sites in Saudi Arabia, as well as in Jordan, Qatar, and United Arab Emirates (Dubai and Abu Dhabi), is unlikely to be sustained by the observed limited human-to-human MERS-CoV transmission, and thus, a more widespread population of MERS-CoV in animals could exist. However, the pattern of MERS-CoV lineages we have documented here is not consistent with a uniform gradient of MERS-CoV evolution across the Arabian peninsula. Instead, it is more consistent with the movement of infected livestock or animal products. This conclusion is suggested by the appearance of the Hafr-Al-Batin_1 lineage in Riyadh, Hafr-al-Batin, Madinah, and Qatar or the Riyadh_3 lineage in Riyadh, Wadi Ad-Dawasir, Ta'if, Qatar, and United Arab Emirates. The appearance of phylogenetically related MERS-CoV in geographically distant locations must be taken into account in efforts to identify the animal source and transmission of the virus.

We have estimated the time of the most recent common ancestor (tMRCA) as March 2012, consistent with the initial case detection. It should be noted that the tMRCA only estimates when the currently circulating viruses were last in a single host; it does not tell us what that host was. Although we only have viral sequences isolated from human patients, it is plausible that this virus was in an as-yet-unidentified animal reservoir. The fact that we only have viruses isolated from human cases (and one camel linked to a human case) may simply represent a strong ascertainment bias toward severe human disease.

In conclusion, the rapid identification and isolation of cases, combined with an R_0 of less than 1, may control the human-to-human transmission as long as the virus transmission properties

TABLE 1 Summary of all MERS-CoV genomic sequences used in this study

Genome	Sample collection date	Genome fraction[a]	GenBank accession number or source
Jordan_N3_2012	15 April 2012	1	KC776174
Bisha_1_2012	19 June 2012	1	KF600620
England-Qatar_2012	19 September 2012	1	KC667074
Riyadh_1_2012	23 October 2012	1	KF600612
Riyadh_2_2012	30 October 2012	1	KF600652
Riyadh_3_2013	5 February 2013	1	KF600613
England2-HPA_2013	10 February 2013	1	http://www.hpa.org.uk/Topics/InfectiousDiseases/InfectionsAZ/MERSCoV/respPartialgeneticsequenceofnovelcoronavirus/
Riyadh_4_2013	1 March 2013	1	KJ156952
Munich_AbuDhabi_2013	22 March 2013	1	KF192507
Al_Hasa_2_2013	21 April 2013	1	KF186566
Al_Hasa_3_2013	22 Apr 2013	1	KF186565
Al-Hasa_24_2013	1 May 2013	0.41	KJ156867, KJ156919, KJ156875, KJ156885, KJ156870, KJ156892, KJ156902
Al_Hasa_4_2013	1 May 2013	1	KF186564
Al-Hasa_7_2013	1 May 2013	0.93	KF600623, KF600655
Al-Hasa_8_2013	1 May 2013	0.74	KF600618, KF600626, KF600635, KF600638
Al-Hasa_9_2013	1 May 2013	0.46	KF600622, KF600639, KF600648, KF600649, KF600654
Al-Hasa_25_2013	2 May 2013	1	KJ156866
Al-Hasa_10_2013	2 May 2013	0.32	KF600614, KF600624, KF600629, KF600636, KF600641, KF600642, KF600646, KF600653
Al-Hasa_11_2013	3 May 2013	0.9	KF600629, KF600636, KF600646
Al-Hasa_12_2013	7 May 2013	1	KF600627
Al-Hasa_13_2013	7 May 2013	0.37	KF600616, KF600637, KF600640, KF600650, KF600656
France_UAE_2013	7 May 2013	0.99	KF745068
Al-Hasa_14_2013	8 May 2013	0.75	KF600615, KF600643
Al_Hasa_1_2013	9 May 2013	1	KF186567
Al-Hasa_22_2013	9 May 2013	0.47	KF600617, KF600619, KF600621, KF600625, KF600631, KF600633
Al-Hasa_15_2013	11 May 2013	1	KF600645
Al-Hasa_16_2013	12 May 2013	1	KF600644
Al-Hasa_23_2013	13 May 2013	0.76	KJ156860, KJ156894, KJ156929, KJ156923, KJ156862
Buraidah_1_2013	13 May 2013	1	KF600630
Al-Hasa_17_2013	15 May 2013	1	KF600647
Al-Hasa_19_2013	23 May 2013	1	KF600632
Al-Hasa_18_2013	23 May 2013	1	KF600651
Al-Hasa_21_2013	30 May 2013	1	KF600634
Hafr-Al-Batin_1_2013	4 June 2013	1	KF600628
Taif_1_2013	12 June 2013	1	KJ156949
Wadi-Ad-Dawasir_1_2013	12 June 2013	1	KJ156881
Taif_2_2013	12 June 2013	0.94	KJ156896, KJ156876
Taif_3_2013	13 Jun 2013	0.62	KJ156938, KJ156897, KJ156922, KJ156868, KJ156921, KJ156915, KJ156906
Taif_4_2013	13 June 2013	0.27	KJ156886, KJ156871
Al-Hasa_26_2013	18 June 2013	0.99	KJ156882, KJ156941, KJ156872
Al-Hasa_27_2013	19 June 2013	0.94	KJ156943, KJ156939
Al-Hasa_28_2013	22 June 2013	0.71	KJ156887, KJ156940, KJ156889, KJ156893, KJ156884, KJ156930, KJ156928, KJ156909
Riyadh_5_2013	2 July 2013	1	KJ156944
Riyadh_6_2013	2 July 2013	0.73	KJ156879, KJ156947, KJ156890, KJ156908, KJ156927
Asir_1_2013	2 July 2013	0.44	KJ156948, KJ156925, KJ156903, KJ156883
Riyadh_7_2013	15 July 2013	0.97	KJ156937, KJ156905
Riyadh_9_2013	17 July 2013	1	KJ156869
Riyadh_8_2013	17 July 2013	0.99	KJ156880, KJ156942
Hafr-Al-Batin_2_2013	5 August 2013	1	KJ156910
Riyadh_10_2013	5 August 2013	0.95	KJ156891, KJ156936, KJ156907
Asir_2_2013	5 August 2013	0.65	KJ156863, KJ156899, KJ156912, KJ156900, KJ156898, KJ156945, KJ156932
Riyadh_11_2013	6 August 2013	0.94	KJ156946, KJ156911
Riyadh_12_2013	8 August 2013	0.95	KJ156926, KJ156901
Riyadh_13_2013	13 August 2013	0.97	KJ156888, KJ156873
Riyadh_14_2013	15 August 2013	1	KJ156934
Riyadh_15_2013	19 August 2013	0.49	KJ156914, KJ156877, KJ156878, KJ156859, KJ156933, KJ156953
Hafr-Al-Batin_5_2013	25 August 2013	0.63	KJ156951, KJ156924, KJ156954, KJ156913

(Continued on following page)

TABLE 1 (Continued)

Genome	Sample collection date	Genome fraction[a]	GenBank accession number or source
Hafr-Al-Batin_4_2013	25 August 2013	0.52	KJ156931, KJ156895, KJ156864, KJ156861
Riyadh_17_2013	26 August 2013	1	KJ156918, KJ156920, KJ156865
Hafr-Al-Batin_6_2013	28 August 2013	1	KJ156874
Madinah_1_2013	1 September 2013	0.3	KJ156935, KJ156904, KJ156917
Madinah_3_2013	11 September 2013	1	KJ156950, KJ156916
Qatar_3_2013	1 October 2013	1	KF961221
Qatar_4_2013	1 October 2013	1	KF961222

[a] Fraction of genome obtained compared with a whole-genome value of 30,119 nucleotides.

remain the same. Full control of the MERS epidemic requires identification of the source of infections to prevent the initiation of the observed human-to-human transmission chains.

MATERIALS AND METHODS

Sequence generation. Nucleic acid extracts from PCR-confirmed MERS-CoV-infected patient samples were processed for reverse transcription and PCR amplification as previously described (15). Briefly, nucleic acids were extracted from respiratory tract samples (Table 1) using automated extraction. The MERS-CoV RNA genome was converted to DNA and amplified by PCR in 15 overlapping amplicons. All amplicons for a sample were pooled and processed into Illumina libraries, and sequencing was performed with an Illumina MiSeq instrument to generate 2 million to 5 million 150-nucleotide paired-end reads per sample. The readsets were processed to remove primer and adapter sequences by using QUASR (30) and assembled into whole genomes using *de novo* assembly with SPAdes (31). The assembly fidelity was verified by monitoring intact open reading frames and through comparison with the genome prepared with reference-based assembly using SMALT (version 0.5.0) (32), with differences resolved by examining the raw read data.

Phylogenetic methods. The 32 new genomes were aligned with the 33 published MERS-CoV genomes using MUSCLE (33) implemented in MEGA5 (34). Bayesian inference of the phylogeny was performed with MrBayes version 3.2.1 (35) using a general-time reversible (GTR) substitution model with a 4-category discrete approximation of a gamma distribution (GTR+Γ_4) to represent among-site heterogeneity. For inference of the time-resolved phylogeny, a second, subalignment of 42 genomes was generated by removing epidemiologically linked sequences. Sequences were considered linked if there was epidemiological evidence for contact between the patients that was also supported by the viral genetic data. If the observed number of mutations between the viral genomes fell below the 95% upper confidence interval of the Poisson cumulative distribution function, whose expected value is calculated from the evolutionary rate of the virus, the length of the genome, and the length of time between the samples, then only the index genome was retained. The main coding regions of the genome (encoding ORF1ab, S, E, M, and N) were concatenated, and a codon-partitioning model of evolution applied to the data set. Time-resolved phylogeny was inferred under a codon-partitioned HKY+Γ_4 substitution model (Hasegawa, Kishino, and Yano substitution model with a 4-category discrete approximation of a gamma distribution), with an uncorrelated lognormal molecular clock and a GMRF Bayesian Skyride coalescent model, using a Bayesian Markov-chain Monte Carlo (BMCMC) approach implemented in BEAST version 1.8.0 (36). Ancestral geographical states were coestimated using the Bayesian stochastic search variable selection (37). Models employing reversible or nonreversible transition rate matrices were assessed by comparing the marginal likelihood estimator of the BMCMC chains, produced through the path-sampling approach implemented in BEAST (38). The Bayesian skyline plot, estimating the change in effective population size through time, was generated from the BEAST BMCMC output files using Tracer version 1.5. Hypothetical ancestral sequences were determined using a likelihood-based ancestral reconstruction method implemented in HYPHY version 2.1.2 (39). Nonsynonymous substitutions were determined using custom Python scripts. Codon positions under episodic selection (40) were determined using the mixed effects model of evolution (MEME) (40) or fast unconstrained Bayesian approximation (FUBAR) (41) implemented in HYPHY.

Nucleotide sequence accession numbers. GenBank accession numbers for the new and previously published genomes are listed in Table 1.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.01062-13/-/DCSupplemental.

Figure S1, TIF file, 4.2 MB.

ACKNOWLEDGMENTS

The support of Kingdom of Saudi Arabia Ministry of Health staff at all the hospitals, the regional health directorates for all their efforts in collecting patient data, and the Jeddah, Riyadh, Madinah, and Dammam regional laboratory staff are gratefully acknowledged. We thank the Sanger Illumina C team for the sequencing support.

This work was supported by the Saudi Arabian Ministry of Health, the Wellcome Trust Sanger Institute, and the European Community's Seventh Framework Programme (FP7/2007–2013) under the project EMPERIE, European Community grant agreement number 223498, and under the project PREDEMICS, grant agreement number 278433. A.I.Z. acknowledges support from the National Institute of Health Research Biomedical Research Centre, University College London Hospitals, the EDCTP, and the EC-FW7.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

1. **Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA.** 2012. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. N. Engl. J. Med. **367:**1814–1820. http://dx.doi.org/10.1056/NEJMoa1211721.
2. **van Boheemen S, de Graaf M, Lauber C, Bestebroer TM, Raj VS, Zaki AM, Osterhaus AD, Haagmans BL, Gorbalenya AE, Snijder EJ, Fouchier RA.** 2012. Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans. mBio **3**(6):e00473-12. http://dx.doi.org/10.1128/mBio.00473-12.
3. **Albarrak AM, Stephens GM, Hewson R, Memish ZA.** 2012. Recovery from severe novel coronavirus infection. Saudi Med. J. **33:**1265–1269.
4. **Memish ZA, Zumla AI, Al-Hakeem RF, Al-Rabeeah AA, Stephens GM.** 2013. Family cluster of Middle East respiratory syndrome coronavirus infections. N. Engl. J. Med. **368:**2487–2494. http://dx.doi.org/10.1056/NEJMoa1303729.
5. **WHO.** Accessed 27 January 2014. Middle East respiratory syndrome coronavirus (MERS-CoV). World Health Organization, Geneva, Switzerland. http://www.who.int/csr/don/2014_01_27mers/en/index.html.
6. **Guan Y, Zheng BJ, He YQ, Liu XL, Zhuang ZX, Cheung CL, Luo SW, Li PH, Zhang LJ, Guan YJ, Butt KM, Wong KL, Chan KW, Lim W, Shortridge KF, Yuen KY, Peiris JS, Poon LL.** 2003. Isolation and characterization of viruses related to the SARS coronavirus from animals in

southern China. Science **302:**276–278. http://dx.doi.org/10.1126/science.1087139.

7. **Reusken CB, Haagmans BL, Müller MA, Gutierrez C, Godeke GJ, Meyer B, Muth D, Raj VS, Smits-Vries LS, Corman VM, Drexler JF, Smits SL, El Tahir YE, De Sousa R, van Beek J, Nowotny N, van Maanen K, Hidalgo-Hermoso E, Bosch BJ, Rottier P, Osterhaus A, Gortázar-Schmidt C, Drosten C, Koopmans MP.** 2013. Middle East respiratory syndrome coronavirus neutralising serum antibodies in dromedary camels: a comparative serological study. Lancet Infect. Dis. **13:**859–866. http://dx.doi.org/10.1016/S1473-3099(13)70164-6.
8. **Perera RA, Wang P, Gomaa MR, El-Shesheny R, Kandeil A, Bagato O, Siu LY, Shehata MM, Kayed AS, Moatasim Y, Li M, Poon LL, Guan Y, Webby RJ, Ali MA, Peiris JS, Kayali G.** 2013. Seroepidemiology for MERS coronavirus using microneutralisation and pseudoparticle virus neutralisation assays reveal a high prevalence of antibody in dromedary camels in Egypt, June 2013. Euro Surveill. **18:**20574.
9. **Memish ZA, Mishra N, Olival KJ, Fagbo SF, Kapoor V, Epstein JH, Alhakeem R, Durosinloun A, Al Asmari M, Islam A, Kapoor A, Briese T, Daszak P, Rabeeah AA, Lipkin WI.** 2013. Middle East respiratory syndrome coronavirus in bats, Saudi Arabia. Emerg. Infect. Dis. **19:**1819–1823. http://dx.doi.org/10.3201/eid1911.131172.
10. **Haagmans BL, Al Dhahiry SH, Reusken CB, Raj VS, Galiano M, Myers R, Godeke GJ, Jonges M, Farag E, Diab A, Ghobashy H, Alhajri F, Al-Thani M, Al-Marri SA, Al Romaihi HE, Al Khal A, Bermingham A, Osterhaus AD, Alhajri MM, Koopmans MP.** 2014. Middle East respiratory syndrome coronavirus in dromedary camels: an outbreak investigation. Lancet Infect. Dis. **14:**140–145. http://dx.doi.org/10.1016/S1473-3099(13)70690-X.
11. **Breban R, Riou J, Fontanet A.** 2013. Interhuman transmissibility of Middle East respiratory syndrome coronavirus: estimation of pandemic risk. Lancet **382:**694–699. http://dx.doi.org/10.1016/S0140-6736(13)61492-0.
12. **Cauchemez S, Fraser C, Van Kerkhove MD, Donnelly CA, Riley S, Rambaut A, Enouf V, van der Werf S, Ferguson NM.** 2014. Middle East respiratory syndrome coronavirus: quantification of the extent of the epidemic, surveillance biases, and transmissibility. Lancet Infect. Dis. **14:**50–56. http://dx.doi.org/10.1016/S1473-3099(13)70304-9.
13. **Aburizaiza AS, Mattes FM, Azhar EI, Hassan AM, Memish ZA, Muth D, Meyer B, Lattwein E, Muller M, Drosten C.** 2014. Investigation of anti-MERS-coronavirus antibodies in blood donors and abbatoir workers in Jeddah and Makkah, Kingdom of Saudi Arabia, fall 2012. J. Infect. Dis. **209**(2):243–246. http://dx.doi.org/10.1093/infdis/jit589.
14. **Assiri A, McGeer A, Perl TM, Price CS, Al Rabeeah AA, Cummings DA, Alabdullatif ZN, Assad M, Almulhim A, Makhdoom H, Madani H, Alhakeem R, Al-Tawfiq JA, Cotten M, Watson SJ, Kellam P, Zumla AI, Memish ZA, KSA MERS-CoV Investigation Team.** 2013. Hospital outbreak of Middle East respiratory syndrome coronavirus. N. Engl. J. Med. **369:**407–416. http://dx.doi.org/10.1056/NEJMoa1306742.
15. **Cotten M, Lam TT, Watson SJ, Palser AL, Petrova V, Grant P, Pybus OG, Rambaut A, Guan Y, Pillay D, Kellam P, Nastouli E.** 2013. Full-genome deep sequencing and phylogenetic analysis of novel human betacoronavirus. Emerg. Infect. Dis. **19:**736B–742B. http://dx.doi.org/10.3201/eid1905.130057.
16. **Drosten C, Seilmaier M, Corman VM, Hartmann W, Scheible G, Sack S, Guggemos W, Kallies R, Muth D, Junglen S, Müller MA, Haas W, Guberina H, Röhnisch T, Schmid-Wendtner M, Aldabbagh S, Dittmer U, Gold H, Graf P, Bonin F, Rambaut A, Wendtner CM.** 2013. Clinical features and virological analysis of a case of Middle East respiratory syndrome coronavirus infection. Lancet Infect. Dis. **13:**745–751. http://dx.doi.org/10.1016/S1473-3099(13)70154-3.
17. **Cotten M, Watson SJ, Kellam P, Al-Rabeeah AA, Makhdoom HQ, Assiri A, Al-Tawfiq JA, Alhakeem RF, Madani H, AlRabiah FA, Al Hajjar SA, Al-nassir WN, Albarrak A, Flemban H, Balkhy HH, Alsubaie S, Palser AL, Gall A, Bashford-Rogers R, Rambaut A, Zumla AI, Memish ZA, .** 2013. Transmission and evolution of the Middle East respiratory syndrome coronavirus in Saudi Arabia: a descriptive genomic study. Lancet **382:**1993–2002. http://dx.doi.org/10.1016/S0140-6736(1013)61887-61885.
18. **Minin VN, Bloomquist EW, Suchard MA.** 2008. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. Mol. Biol. Evol. **25:**1459–1471. http://dx.doi.org/10.1093/molbev/msn090.
19. **Guery B, Poissy J, el Mansouf L, Séjourné C, Ettahar N, Lemaire X, Vuotto F, Goffard A, Behillil S, Enouf V, Caro V, Mailles A, Che D, Manuguerra JC, Mathieu D, Fontanet A, van der Werf S, MERS-CoV study group.** 2013. Clinical features and viral diagnosis of two cases of infection with Middle East Respiratory Syndrome coronavirus: a report of nosocomial transmission. Lancet **381:**2265–2272. http://dx.doi.org/10.1016/S0140-6736(13)60982-4.
20. **Gao J, Lu G, Qi J, Li Y, Wu Y, Deng Y, Geng H, Li H, Wang Q, Xiao H, Tan W, Yan J, Gao GF.** 2013. Structure of the fusion core and inhibition of fusion by a heptad-repeat peptide derived from the S protein of MERS-CoV. J. Virol. **87:**13134–13140. http://dx.doi.org/10.1128/JVI.02433-13.
21. **Lu G, Hu Y, Wang Q, Qi J, Gao F, Li Y, Zhang Y, Zhang W, Yuan Y, Bao J, Zhang B, Shi Y, Yan J, Gao GF.** 2013. Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26. Nature **500:**227–231. http://dx.doi.org/10.1038/nature12328.
22. **McRoy WC, Baric RS.** 2008. Amino acid substitutions in the S2 subunit of mouse hepatitis virus variant V51 encode determinants of host range expansion. J. Virol. **82:**1414–1424. http://dx.doi.org/10.1128/JVI.01674-07.
23. **Li W, Zhang C, Sui J, Kuhn JH, Moore MJ, Luo S, Wong SK, Huang IC, Xu K, Vasilieva N, Murakami A, He Y, Marasco WA, Guan Y, Choe H, Farzan M.** 2005. Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. EMBO J. **24:**1634–1643. http://dx.doi.org/10.1038/sj.emboj.7600640.
24. **Sheahan T, Rockx B, Donaldson E, Sims A, Pickles R, Corti D, Baric R.** 2008. Mechanisms of zoonotic severe acute respiratory syndrome coronavirus host range expansion in human airway epithelium. J. Virol. **82:**2274–2285. http://dx.doi.org/10.1128/JVI.02041-07.
25. **Graham RL, Baric RS.** 2010. Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. J. Virol. **84:**3134–3146. http://dx.doi.org/10.1128/JVI.01394-09.
26. **Frieman M, Ratia K, Johnston RE, Mesecar AD, Baric RS.** 2009. Severe acute respiratory syndrome coronavirus papain-like protease ubiquitin-like domain and catalytic domain regulate antagonism of IRF3 and NF-kappaB signaling. J. Virol. **83:**6689–6705. http://dx.doi.org/10.1128/JVI.02220-08.
27. **Devaraj SG, Wang N, Chen Z, Tseng M, Barretto N, Lin R, Peters CJ, Tseng CT, Baker SC, Li K.** 2007. Regulation of IRF-3-dependent innate immunity by the papain-like protease domain of the severe acute respiratory syndrome coronavirus. J. Biol. Chem. **282:**32208–32221. http://dx.doi.org/10.1074/jbc.M704870200.
28. **Memish Z.** 2013. MERS-COV—Eastern Mediterranean (85): animal reservoir, camel, suspected, official. ProMED-Mail, archive no. 20131112.2051424. International Society for Infectious Diseases, Brookline, MA. http://www.promedmail.org/direct.php?id=2051424.
29. **Perlman S, Netland J.** 2009. Coronaviruses post-SARS: update on replication and pathogenesis. Nat. Rev. Microbiol. **7:**439–450. http://dx.doi.org/10.1038/nrmicro2147.
30. **Watson SJ, Welkers MR, Depledge DP, Coulter E, Breuer JM, de Jong MD, Kellam P.** 2013. Viral population analysis and minority-variant detection using short read next-generation sequencing. Philos. Trans. R. Soc. Lond. B Biol. Sci. **368:**20120205. http://dx.doi.org/10.1098/rstb.2012.0205.
31. **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA.** 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. **19:**455–477. http://dx.doi.org/10.1089/cmb.2012.0021.
32. **Ponstingl H.** 2013. SMALT efficiently aligns DNA sequencing reads with a reference genome. Wellcome Trust Sanger Institute, Hinxton, United Kingdom. Current version - SMALT v0.7.5. Released 16th July 2013. http://www.sanger.ac.uk/resources/software/smalt/.
33. **Edgar RC.** 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics **5:**113. http://dx.doi.org/10.1186/1471-2105-5-113.
34. **Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S.** 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. **28:**2731–2739. http://dx.doi.org/10.1093/molbev/msr121.
35. **Huelsenbeck JP, Ronquist F.** 2001. MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics **17:**754–755. http://dx.doi.org/10.1093/bioinformatics/17.8.754.

36. **Drummond AJ, Suchard MA, Xie D, Rambaut A.** 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. **29:**1969–1973. http://dx.doi.org/10.1093/molbev/mss075.
37. **Lemey P, Suchard M, Rambaut A.** 2009. Reconstructing the initial global spread of a human influenza pandemic. PLoS Curr. Influenza **1:**RRN1031. http://dx.doi.org/10.1371/currents.RRN1031.
38. **Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV.** 2012. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. Mol. Biol. Evol. **29:**2157–2167. http://dx.doi.org/10.1093/molbev/mss084.
39. **Pond SL, Frost SD, Muse SV.** 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics **21:**676–679. http://dx.doi.org/10.1093/bioinformatics/bti079.
40. **Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL.** 2012. Detecting individual sites subject to episodic diversifying selection. PLoS Genet. **8:**e1002764. http://dx.doi.org/10.1371/journal.pgen.1002764.
41. **Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, Scheffler K.** 2013. FUBAR: a fast, unconstrained Bayesian approximation for inferring selection. Mol. Biol. Evol. **30:**1196–1205. http://dx.doi.org/10.1093/molbev/mst030.
42. **Bosch BJ, van der Zee R, de Haan CA, Rottier PJ.** 2003. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. J. Virol. **77:**8801–8811. http://dx.doi.org/10.1128/JVI.77.16.8801-8811.2003.
43. **Millet JK, Kien F, Cheung CY, Siu YL, Chan WL, Li H, Leung HL, Jaume M, Bruzzone R, Peiris JS, Altmeyer RM, Nal B.** 2012. Ezrin interacts with the SARS coronavirus spike protein and restrains infection at the entry stage. PLoS One **7:**e49566. http://dx.doi.org/10.1371/journal.pone.0049566.
44. **Wang N, Shi X, Jiang L, Zhang S, Wang D, Tong P, Guo D, Fu L, Cui Y, Liu X, Arledge KC, Chen YH, Zhang L, Wang X.** 2013. Structure of MERS-CoV spike receptor-binding domain complexed with human receptor DPP4. Cell Res. **23:**986–993. http://dx.doi.org/10.1038/cr.2013.92.