AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Virology

Extensive Positive Selection Drives the Evolution of Nonstructural Proteins in Lineage C Betacoronaviruses

Diego Forni,[a] Rachele Cagliani,[a] Alessandra Mozzi,[a] Uberto Pozzoli,[a] Nasser Al-Daghri,[b,c] Mario Clerici,[d,e] Manuela Sironi[a]

Scientific Institute IRCCS E. Medea, Bioinformatics, Bosisio Parini, Italy[a]; Biomarker Research Program, Biochemistry Department, College of Science, King Saud University, Riyadh, Saudi Arabia[b]; Prince Mutaib Chair for Biomarkers of Osteoporosis, Biochemistry Department, College of Science, King Saud University, Riyadh, Saudi Arabia[c]; Department of Physiopathology and Transplantation, University of Milan, Milan, Italy[d]; Don C. Gnocchi Foundation ONLUS, IRCCS, Milan, Italy[e]

ABSTRACT

Middle East respiratory syndrome-related coronavirus (MERS-CoV) spreads to humans via zoonotic transmission from camels. MERS-CoV belongs to lineage C of betacoronaviruses (betaCoVs), which also includes viruses isolated from bats and hedgehogs. A large portion of the betaCoV genome consists of two open reading frames (ORF1a and ORF1b) that are translated into polyproteins. These are cleaved by viral proteases to generate 16 nonstructural proteins (nsp1 to nsp16) which compose the viral replication-transcription complex. We investigated the evolution of ORF1a and ORF1b in lineage C betaCoVs. Results indicated widespread positive selection, acting mostly on ORF1a. The proportion of positively selected sites in ORF1a was much higher than that previously reported for the surface-exposed spike protein. Selected sites were unevenly distributed, with nsp3 representing the preferential target. Several pairs of coevolving sites were also detected, possibly indicating epistatic interactions; most of these were located in nsp3. Adaptive evolution at nsp3 is ongoing in MERS-CoV strains, and two selected sites (G720 and R911) were detected in the protease domain. While position 720 is variable in camel-derived viruses, suggesting that the selective event does not represent a specific adaptation to humans, the R911C substitution was observed only in human-derived MERS-CoV isolates, including the viral strain responsible for the recent South Korean outbreak. It will be extremely important to assess whether these changes affect host range or other viral phenotypes. More generally, data herein indicate that CoV nsp3 represents a major selection target and that nsp3 sequencing should be envisaged in monitoring programs and field surveys.

IMPORTANCE

Both severe acute respiratory syndrome coronavirus (SARS-CoV) and MERS-CoV originated in bats and spread to humans via an intermediate host. This clearly highlights the potential for coronavirus host shifting and the relevance of understanding the molecular events underlying the adaptation to new host species. We investigated the evolution of ORF1a and ORF1b in lineage C betaCoVs and in 87 sequenced MERS-CoV isolates. Results indicated widespread positive selection, stronger in ORF1a than in ORF1b. Several selected sites were found to be located in functionally relevant protein regions, and some of them corresponded to functional mutations in other coronaviruses. The proportion of selected sites we identified in ORF1a is much higher than that for the surface-exposed spike protein. This observation suggests that adaptive evolution in ORF1a might contribute to host shifts or immune evasion. Data herein also indicate that genetic diversity at nonstructural proteins should be taken into account when antiviral compounds are developed.

Middle East respiratory syndrome-related coronavirus (MERS-CoV) (http://talk.ictvonline.org/files/proposals/animal_ssrna_viruses/default.aspx) was identified as the causative agent of a new viral respiratory disease in Saudi Arabia in June 2012 (1). Since then, more than 1,500 cases and 571 deaths have been reported worldwide (as of 12 October 2015 [http://www.who.int/csr/don/12-october-2015-mers-saudi-arabia/en/]), although major outbreaks have been confined to the Middle East and, more recently, to South Korea. The rate of human-to-human transmission of MERS-CoV is relatively low, suggesting that a zoonotic reservoir serves as a major source for transmission (2). Recent studies have indicated that MERS-CoV or a closely related virus originated in bats and possibly spread to humans via transmission from dromedary camels (3).

Like SARS-CoV, which causes severe acute respiratory syndrome (SARS) and evolved in bats as well, MERS-CoV (order *Nidovirales*, family *Coronaviridae*, subfamily *Coronavirinae*) is a positive-sense single-stranded RNA (+ssRNA) virus belonging to the C lineage of the *Betacoronavirus* genus (4). CoVs are exceptional among RNA viruses for having long (~30-kb) genomes, a feature associated with a specific genome architecture and with the acquisition of an RNA 3′-to-5′ exoribonuclease activity (exoN) (5). About two-thirds of the CoV genome consists of two large overlapping open reading frames (ORF1a and ORF1b) that are translated into the polyproteins pp1a and pp1ab (the latter synthesized via a −1 ribosome frameshift at the 3′ end of ORF1a). These polyproteins are subsequently cleaved by viral proteases to generate 16 nonstructural proteins (nsp1 to nsp16), most of which

Received 25 November 2015 **Accepted** 12 January 2016

Accepted manuscript posted online 20 January 2016

Citation Forni D, Cagliani R, Mozzi A, Pozzoli U, Al-Daghri N, Clerici M, Sironi M. 2016. Extensive positive selection drives the evolution of nonstructural proteins in lineage C betacoronaviruses. J Virol 90:3627–3639. doi:10.1128/JVI.02988-15.

Editor: S. Perlman

Address correspondence to Diego Forni, diego.forni@bp.lnf.it.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/JVI.02988-15.

compose the viral replication-transcription complex (RTC). With the exception of nsp1 and nsp2, whose functions are poorly understood, most nsps have been characterized in some detail for MERS-CoV, SARS-CoV, and mouse hepatitis virus (MHV). Thus, nsp3, a large multidomain and multifunctional protein, was shown to play essential roles in the virus replication cycle. In fact, the papain-like protease (PLpro) activity of nsp3 is responsible for the initial processing of pp1a. Also, nsp3, together with nsp4 and nsp6, recruits intracellular membranes to anchor the RTC and to form a reticulovesicular network of double-membrane vesicles (DVMs) and convoluted membranes where viral RNA replication occurs (6). nsp5 is a second viral protease (3C-like protease [3CLpro]) that cleaves both pp1a and pp1ab to the final nsp products. nsp7 to nsp11 provide the primer-making activity and regulate the function of the main RNA-dependent RNA polymerase (RdRp; nsp12 protein). Finally, nsp13 to nsp16 comprise RNA-modifying enzymes, including the exoN activity in nsp14 (6, 7).

Viral RNA represents the major pathogen-associated molecular pattern (PAMP) recognized by the host immune system during CoV infection (8). Both MERS-CoV and SARS-CoV elicit limited interferon (IFN) responses in most cell types, indicating that these viruses have evolved efficient strategies to evade innate immune sensing and/or to block IFN induction (8). Indeed, these viruses express antagonists of the IFN response, including SARS-CoV ORF6, ORF3b, and nucleoprotein, as well as MERS-CoV structural and accessory proteins M, ORF4a, ORF4b, and ORF5 (9–12). Additional immune evasion strategies, though, rely on nsps. In fact, the enzymatic activities of nsp14 and nsp16 endow the viral RNA with a 2′-O-methylated cap structure that mimics cellular mRNAs and avoids activation of the innate immunity receptors RIG-I and MDA5 (8). In analogy to the exoN and endoribonucleases expressed by other viruses, such as Lassa fever virus and pestivirus (13, 14), the RNase activities of nsp14 and nsp15 are also thought to play a role in immune escape by digesting RNA PAMPs (8). Moreover, suppression of IFN responses is mediated by the PLpro in nsp3 through its deubiquitinating and deISGylating activities (15, 16), as well as by nsp1. The latter inhibits IFN-dependent signaling by decreasing the phosphorylation levels of STAT1 and suppresses host protein synthesis (17, 18). Finally, PLpro was shown to physically interact with TRAF3, TBK1, IKKε, STING, and IRF3, which represent key cellular components for IFN response (19).

Therefore, the information encoded by CoV ORF1a and ORF1b is essential for viral replication and for immune evasion. For these reasons, inhibitors that interfere with viral enzymatic activities (e.g., proteases and RdRp) are regarded as promising candidates for therapeutic intervention (20).

From an evolutionary standpoint, different observations suggest that nsps may represent targets of natural selection. First, genes encoding molecules that directly interact with the host immune system are thought to be preferential targets of natural selection as a consequence of host-pathogen arms races (21). Second, adaptation to new hosts in other RNA viruses has been associated with selective changes in polymerase genes (22, 23). Finally, the acquisition of a complex replication machinery is evolutionarily linked to genome expansion in *Nidovirales* (5). Nonetheless, evolutionary studies have mainly focused on the analysis of betacoronavirus (betaCoV) spike proteins, as these are surface exposed and represent major determinants of host range and tissue tropism (24). In this study, we investigated the evolution of ORF1a and ORF1b in MERS-CoV and in lineage C betaCoVs isolated from bats and hedgehogs. Results indicate widespread positive selection, stronger in ORF1a; within this region, nsp3 represents a preferential selection target, and adaptive evolution is ongoing in MERS-CoV strains circulating in the current outbreak.

MATERIALS AND METHODS

Sequences and alignments. ORF1a and ORF1b sequences for 7 lineage C betaCoVs and 87 MERS-CoV strains (available as of July 2015) were retrieved from the NCBI database; a list of accession numbers for the complete genomes is provided in Table S1 in the supplemental material.

Alignment errors are common when divergent sequences are analyzed and can affect evolutionary inference. Thus, we used PRANK (25) to generate multiple sequence alignments and GUIDANCE (26) for filtering unreliably aligned codons (i.e., we masked codons with a score of <0.90), as suggested previously (27).

The alignments were screened for the presence of recombination events using two methods based on distinct data features. (i) GARD (genetic algorithm recombination detection) (28) uses phylogenetic incongruence among segments in the alignment to detect the best-fit number and location of recombination breakpoints; the statistical significance of putative breakpoints is then evaluated through Kishino-Hasegawa (HK) tests. (ii) GENECONV (29) tests for significant clustering of substitutions along sequences; statistical significance is assessed through permutation with multiple-comparison correction. For both methods, recombination breakpoints were considered significant if the *P* value was <0.05. No breakpoint was detected in any analysis.

Detection of positive selection. Gene trees were generated by maximum likelihood using the program phyML with a general time-reversible (GTR) plus gamma-distributed rates model and 4 substitution rate categories (30).

Positive selection can be defined when the nonsynonymous/synonymous rate ratio (ω) is higher than 1; to analyze the presence of episodic positive selection in lineage C betaCoVs, we applied the branch site test (31) from the PAML suite (32). The test is based on the comparison between two nested models: a model (MA) that allows positive selection on one or more lineages (called foreground lineages) and a model (MA1) that does not allow such positive selection. Twice the difference of likelihood for the two models (Δln) is then compared to a χ^2 distribution with one degree of freedom (31). A false-discovery rate (FDR) correction was applied to take into account a multiple-hypothesis issue generated by analyzing different branches on the same phylogeny (33). When the likelihood ratio test suggested the action of positive selection, the Bayes empirical Bayes (BEB) analysis was used to evaluate the posterior probability that each codon belongs to the site class of positive selection on the foreground branch.

BUSTED (branch-site unrestricted statistical test for episodic diversification) (34) is a recently developed software designed to describe episodic positive selection that acts on specific branches in the phylogeny at a proportion of sites within the alignment. An alternative model that allows the action on positive selection on foreground branches is compared with a null model that does not allow an ω of >1. Twice the Δln of the two models is then compared to a χ^2 distribution (degrees of freedom = 2); if the null model is rejected, at least one site is under positive selection on the foreground branch(es). To detect selection at individual sites, twice the difference of the likelihood for the alternative and the null model at each site is compared to a χ^2 distribution (degree of freedom = 1). BUSTED is implemented in the HYPHY package (35).

Conservatively, we considered a site as selected if it showed a *P* value of ≤ 0.05 in BUSTED and a posterior probability of ≥ 0.90 in the BEB analysis.

The site models implemented in PAML were applied for the analysis of nsp3 sequences from MERS-CoV isolates. To detect selection, two different pairs of nested site models (M1a/M2a and M7/M8) were fitted to the data (32); M2a and M8 allow a class of sites to evolve with an ω of >1,

whereas M1a and M7 do not. Positively selected sites were identified using the BEB analysis (from model M8) (36). Sites were validated using MEME (37) (with a cutoff of ≤0.1), which allows the distribution of ω to vary from site to site and from branch to branch at a site.

MEME (37) analyses were performed through the DataMonkey server (38).

Detection of coevolving sites. To detect coevolving sites in the ORF1a and ORF1b alignments, we applied two different methods: Bayesian graphical model (BGM)-Spidermonkey (39) and the Mutual Information Server To Infer Coevolution (MISTIC) (40). Spidermonkey is a tool implemented in the HYPHY package that identifies coevolving sites from an alignment of coding sequences; a BGM is used to evaluate the connection among codons (represented by the nodes of the network). Significant statistical associations between nodes are indicated by the edges of the network, suggesting functional or structural interactions between codons.

MISTIC estimates the relationship between two or more position in an alignment. The coevolutionary association is estimated by Mutual Information (MI), which evaluates how much the information from the amino acid at the first position can help to predict the amino acid identity at the second position.

For BGM-Spidermonkey, sites were filtered based on a minimum count of 4 substitutions across the phylogeny. To be conservative, we considered a pair of residues as coevolving if they showed a posterior probability of >0.75. This threshold corresponds to 0.02% and 1.42% of all analyzed site pairs in ORF1a and ORF1b, respectively. Likewise, for MISTIC, site pairs were required to display an MI rank higher than the 99th percentile calculated using all MI scores from the alignment. Pairs of sites exceeding the thresholds for both methods were declared to be coevolving.

Membrane topology, glycosylation site predictions, and three-dimensional (3D) structure mapping. The membrane protein topology for MERS-CoV nsp3 and nsp4 was predicted by using TMHMM (http://www.cbs.dtu.dk/services/TMHMM/) (41). *N*-Glycosylation sites were predicted with NetNGlyc (http://www.cbs.dtu.dk/services/NetNGlyc/), a program that uses artificial neural networks to examine the sequence context of Asn-X-Ser/Thr motifs.

Sites were mapped onto structures using PyMOL (PyMOL molecular graphics system, version 1.5.0.2; Schrödinger, LLC).

RESULTS

nsp3 in ORF1a is a major selection target in betaCoVs. Lineage C of betaCoVs includes two bat species closely related to MERS-CoV, namely, Ty-BatCoV HKU4 and Pi-BatCoV HKU5, isolated from the lesser bamboo bats (*Tylonycteris pachypus*) and Japanese pipistrelles (*Pipistrellus abramus*), respectively (4). Additional viruses belonging to lineage C of betaCoVs have been described to occur in bats (BtCoV/133 and BtVs-BetaCoV/SC2013) and hedgehogs (hedgehog coronavirus [EriCoV]) (20, 42–44). Recently, a virus belonging to the same species as MERS-CoV was isolated in *Neoromicia* bats (NeoCoV) (45). To investigate the evolutionary history of ORF1a, we obtained sequence information for these viruses and for 6 MERS-CoV strains isolated from either humans or camels and belonging to the major groups described to date (46) (Fig. 1). The sequence alignment was pruned of unreliably aligned codons (see Materials and Methods), a procedure that resulted in the masking of almost the entire acidic domain in nsp3. Indeed, this region was previously shown to be highly divergent among CoVs (47). We next analyzed the alignment for the presence of recombination breakpoints using GARD (genetic algorithm recombination detection) (28) and GENECONV (29). No evidence of recombination was detected.

The phylogenetic tree of ORF1a obtained with phyML was consistent with previously reported ones (45, 46). An estimate of the extent of functional constraint along ORF1a was obtained by identification of negatively selected sites (total number = 903) and calculation of their distribution among nsps. This analysis indicated that the average fraction of negatively selected sites is ~0.24, with the weakest selection in nsp1 and the strongest constraint in nsp6 to nsp9 (Fig. 1).

Evidence of episodic positive selection along the internal branches of the ORF1a phylogeny was searched for using branch site tests. Specifically, we applied two different methods: the branch site unrestricted statistical test for episodic diversification (BUSTED) (34) and the maximum likelihood models (MA/MA1) implemented in the PAML suite (32). These two approaches rely on different assumptions of ω (nonsynonymous/synonymous rate ratio) variation among branches. To be conservative, episodic positive selection at each tested branch was declared only if statistically significant support was obtained with both methods. Using this criterion, we found 4 branches with evidence of episodic selection (Fig. 1; Table 1). Selected sites along these branches were identified using the Bayes empirical Bayes (BEB) procedure from model MA and with BUSTED; again, only sites detected by both methods were considered. A total of 55 selected sites were detected; these were scattered along the entire ORF1a and located in different nsps, with the exclusion of nsp7 and nsp9, which were not targeted by selection (Fig. 1; see also Table S2 in the supplemental material). To determine whether any nsp represented a preferential target of episodic positive selection, we performed random uniform sampling (i.e., assuming a random distribution of selected sites, we calculated the likelihood of identifying a number of sites equal to or higher than the one we observed for each nsp). This analysis indicated that nsp3 was preferentially targeted by episodic selection (Bonferroni corrected *P* value = 0.010) during the evolution of MERS-related CoVs.

Finally, we searched for evidence of coevolution between sites in the ORF1a alignment. To this aim, we applied BGM-Spidermonkey (39) and MISTIC (40) (see Materials and Methods for details). Six pairs of coevolving sites were detected by both methods. Most site pairs were located in nsp3 (Fig. 1; see also Table S3).

Weaker selective pressure for ORF1b. The approach described above was applied to ORF1b sequences. The degree of constraint in nsp12 to nsp16 was comparable to that observed for ORF1a (Fig. 1A). Statistical support of episodic positive selection was detected for 2 branches only (Fig. 1B; Table 1), and fewer selected sites were detected than for ORF1a. Most sites were targeted by selection on the branch leading to bat CoVs (HKU5, BtCoV/133, and HKU4) (Fig. 1; see also Table S2). No nsp resulted a preferential selection target. One pair of coevolving sites was detected by MISTIC and BGM-Spidermonkey (Fig. 1A; see also Table S3).

Ongoing selection at nps3 in MERS-CoV. Given the results described above, we investigated whether positive selection also occurred at nsp3 during the recent evolution of MERS-CoV. To this end, we retrieved 87 sequences from MERS-CoVs isolated from camels or humans (available strains as of July 2015) (see Table S1). No recombination breakpoint was detected by either GARD or GENECONV, and the *codeml* site models were applied (32). Results showed that models of gene evolution that allow a class of codons to evolve with an ω of >1 (NSsite models M2a and M8) better fit the data than the neutral models (NSsite models M1a and M7), strongly supporting the action of positive selection (Table 2). Positively selected sites in nsp3 were identified using

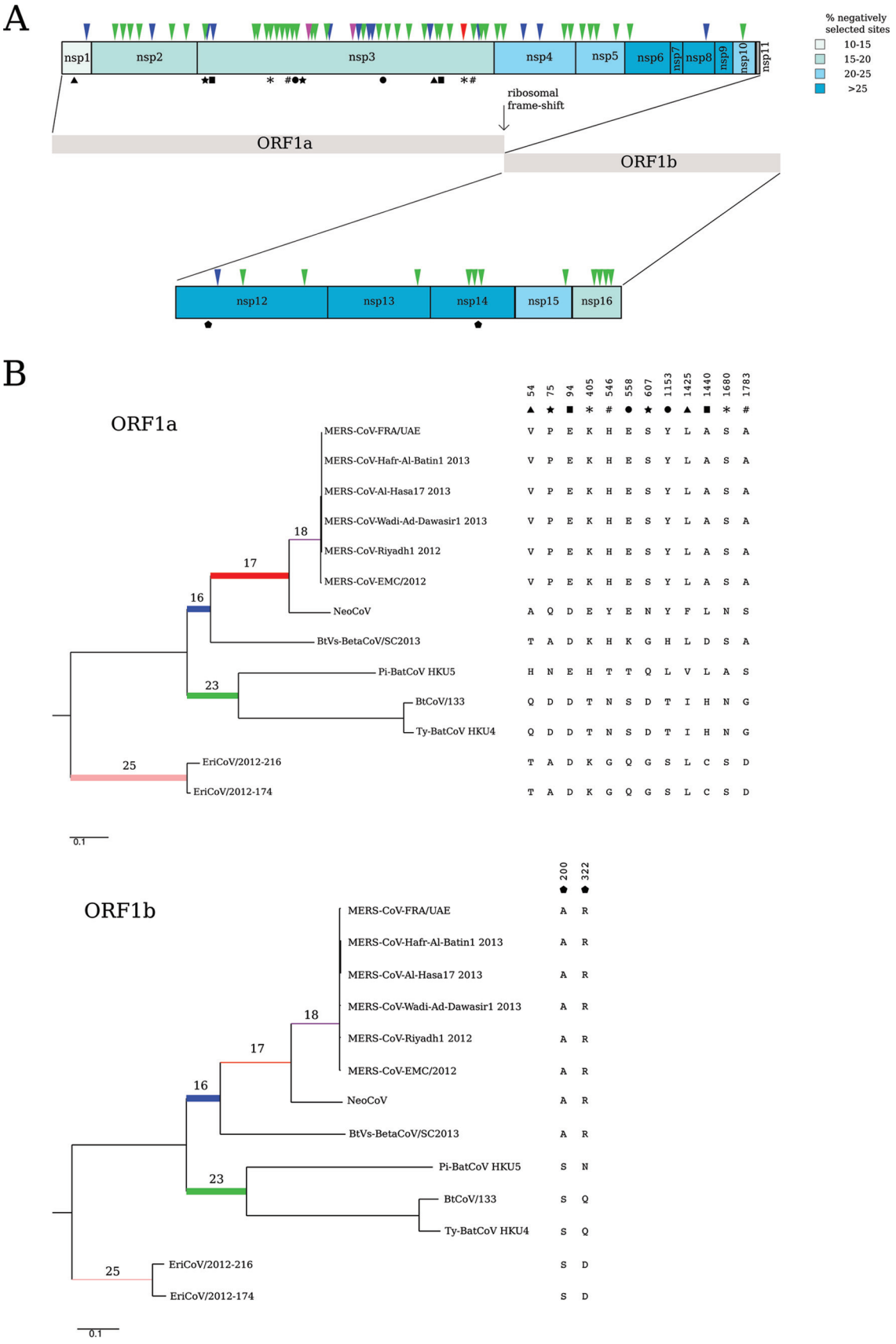

A
% negatively selected sites
10-15
15-20
20-25
>25
nsp1
nsp2
nsp3
nsp4
nsp5
nsp6
nsp7
nsp8
nsp9
nsp10
nsp11
ribosomal frame-shift
ORF1a
ORF1b
nsp12
nsp13
nsp14
nsp15
nsp16
B
ORF1a
54 75 94 405 546 558 607 1153 1425 1440 1680 1783
MERS-CoV-FRA/UAE V P E K H E S Y L A S A
MERS-CoV-Hafr-Al-Batin1 2013 V P E K H E S Y L A S A
MERS-CoV-Al-Hasa17 2013 V P E K H E S Y L A S A
MERS-CoV-Wadi-Ad-Dawasir1 2013 V P E K H E S Y L A S A
MERS-CoV-Riyadh1 2012 V P E K H E S Y L A S A
MERS-CoV-EMC/2012 V P E K H E S Y L A S A
NeoCoV A Q D E Y E N Y F L N S
BtVs-BetaCoV/SC2013 T A D K H K G H L D S A
Pi-BatCoV HKU5 H N E H T T Q L V L A S
BtCoV/133 Q D D T N S D T I H N G
Ty-BatCoV HKU4 Q D D T N S D T I H N G
EriCoV/2012-216 T A D K G Q G S L C S D
EriCoV/2012-174 T A D K G Q G S L C S D
18
17
16
23
25
0.1
ORF1b
200 322
MERS-CoV-FRA/UAE A R
MERS-CoV-Hafr-Al-Batin1 2013 A R
MERS-CoV-Al-Hasa17 2013 A R
MERS-CoV-Wadi-Ad-Dawasir1 2013 A R
MERS-CoV-Riyadh1 2012 A R
MERS-CoV-EMC/2012 A R
NeoCoV A R
BtVs-BetaCoV/SC2013 A R
Pi-BatCoV HKU5 S N
BtCoV/133 S Q
Ty-BatCoV HKU4 S Q
EriCoV/2012-216 S D
EriCoV/2012-174 S D
18
17
16
23
25
0.1

TABLE 1 Likelihood ratio test statistics for branch site tests

Region	Foreground branch[a]	$-2\Delta\ln L^b$ (MA vs MA1)[c]	MA vs MA1 *P* value (FDR corrected *P* value)	$-2\Delta\ln L^b$ (BUSTED)	BUSTED *P* value	No. of sites identified by BEB and BUSTED
ORF1a						
	18	4.35	0.0369 (0.0369)	0.50	0.778	
	17	49.18	2.34×10^{-12} (3.90×10^{-12})	23.18	9.25×10^{-6}	1 (nsp3)
	16	75.22	4.21×10^{-18} (1.05×10^{-17})	85.9	2.58×10^{-19}	13 (1 nsp1, 1 nsp2, 8 nsp3, 2 nsp4, 1 nsp8)
	23	93.39	4.29×10^{-22} (2.14×10^{-21})	71.76	2.61×10^{-16}	41 (6 nsp2, 25 nsp3, 4 nsp4, 4 nsp5, 1 nsp6, 1 nsp10)
	25	26.84	2.20×10^{-7} (2.75×10^{-7})	7.52	0.023	
ORF1b						
	18	0.78	0.38 (0.475)			
	17	8.53	0.0035 (0.0058)	2.8	0.246	
	16	12.70	3.6×10^{-4} (0.0009)	17.36	1.70×10^{-4}	1 (nsp12)
	23	37.83	7.73×10^{-10} (3.86×10^{-9})	31.94	1.16×10^{-7}	11 (2 nsp12, 1 nsp13, 3 nsp14, 1 nsp15, 4 nsp16)
	25	0.23	0.63 (0.63)			

[a] Branches are numbered as in Fig. 1.
[b] 2ΔlnL is twice the difference of the natural logs of the maximum likelihood of the models being compared.
[c] MA and MA1 are branch site models that assume four classes of sites: the MA model allows a proportion of codons to have an ω of ≥1 on the foreground branches, whereas the MA1 model does not.

two methods, the BEB procedure implemented in M8 and MEME (mixed-effects model of evolution). Two sites (G720 and R911) were detected by both methods (Table 2; Fig. 2). Four different residues are present at position 720 in MERS-CoV sequences, and three of them are present in viruses isolated from camels (Fig. 2), suggesting that adaptive evolution at this site started before the spread to humans. Conversely, the R911C substitution occurs only in viruses derived from humans: all of them belong to group 3 and include the virus responsible for the recent South Korean outbreak (MERS-CoV/KOR/KNIH) (Fig. 2).

Positively selected sites in the context of betaCoV biology. To shed light on the functional role of selected sites, we exploited available structural, genetic, and biochemical data for nsps (mainly obtained for SARS-CoV and MHV), as well as *in silico* prediction of transmembrane helices and glycosylation sites.

Most positively selected sites were located in nsp3 (Fig. 3A). In the PLpro domain, four of the positively selected sites, including two that are selected in the viral stains from the current MERS-CoV epidemic, were clustered in a spatially confined region opposed to the catalytic crevice. This region corresponds to the "palm" of the right-handed architecture of the protease, whereas three additional sites are located on the "fingers" (Fig. 3A). One of these three surface-exposed sites, Q830, is part of a noncanonical nsp5 cleavage site, unique for MERS-CoV, that could contribute to the formation of new cleavage products (48) (Fig. 3A).

Additional sites were identified in the other domains of the protein (Fig. 3A). One positively selected site in the Ubl1 domain (C91) precedes a conserved acidic loop that when mutated in SARS-CoV determines a lethal phenotype (49). Among sites in the ADP-ribose (ADPR) domain, none was located within the ADPR binding pocket (Fig. 3A) (50). As for the nucleic acid binding domain (NAB) (Fig. 3A), one of the positively selected sites (T1080) is part of the α2 helix, which, together with the loop preceding the β6 strand, forms a structure possibly interacting with ssRNA via a charged patch (51).

Finally, positively selected sites were found in the so-called transmembrane region of nsp3 (Fig. 3). We performed an *in silico* prediction of transmembrane helices and glycosylation sites. As previously observed for MHV and SARS-CoV (52), the topology model did not fit the general Nendo/Cendo shared structure and was inconsistent with the location of the glycosylation site. Thus, in analogy to other CoVs, some hydrophobic regions predicted to be transmembrane are unlikely to span the lipid bilayer. Intriguingly, we identified two positively selected sites in the luminal loop; one of them (A1386) is located between the first two conserved cysteine residues (Fig. 3B) of the zinc finger motif (ZF). Other positively selected sites were found in the third hydrophobic region (which may or may not span the membrane).

In nsp4, four transmembrane regions were predicted (in SARS-CoV and MHV, all of these are membrane spanning) (53) (Fig. 3C). One of the positively selected sites is located on large luminal loop 1 (Fig. 3C) and does not affect (or introduce) a predicted glycosylation site. Variants in the same loop of MHV nps4 were shown to alter DVM morphology or number, irrespec-

FIG 1 (A) Schematic representation of ORF1a and ORF1b and their nsp products. nsps are colored in hues of blue depending on the percentage of negatively selected sites. nsp11 is shown in gray because it was not analyzed (because it is too short). Positively selected sites are represented by triangles, with colors corresponding to the selected branch in the phylogeny (see panel B). Coevolving sites are shown below the nsp structure, with different symbols indicating each pair of coevolving sites. (B) Maximum likelihood phylogenies for ORF1a (left) and ORF1b (right) in lineage C betaCoVs. Branches set as foreground lineages in independent branch site tests are highlighted with different colors and numbered. Thick branches yielded statistically significant evidence of positive selection. Branch length is proportional to synonymous substitution rate (dS). Coevolving sites are also reported, with different symbols as indicated in panel A. Positions are relative to each nsp; see also Table S3 in the supplemental material.

TABLE 2 Likelihood ratio test statistics for models of variable selective pressure among sites

Region	Likelihood ratio test model	−2ΔlnL[c]	*P* value	% of sites (avg ω)	Positively selected sites (BEB and MEME)
nsp3	M1a vs M2a[a]	15.25	0.00048	0.6 (17.24)	G720, R911
	M7 vs M8[b]	14.15	0.00085	1.6 (11.59)	

[a] M1a is a nearly neutral model that assumes one ω class between 0 and 1 and one class with an ω of 1; M2a (positive selection model) is the same as M1a plus an extra class with an ω of >1.
[b] M7 is a null model that assumes that 0<ω<1 is beta distributed among sites; M8 (positive selection model) is the same as M7 but also includes an extra category of sites with an ω of >1.
[c] 2ΔlnL: twice the difference of the natural logs of the maximum likelihood of the models being compared.

tive of their effect on glycosylation (54). Two additional positively selected sites (G453 and A479) are located in the C-terminal domain of the protein (Fig. 3C). This domain is cytosolic and interacts with viral and host proteins (55).

We also detected 4 positively selected sites in the other viral protease, 3CLpro (nsp5) (Fig. 4). Interestingly, two of them (H8 and V132) were proposed to affect the correct domain orientation required for dimerization of MERS-CoV nsp5 (56). Moreover, site-directed mutagenesis of 3CLpro from MHV indicated that the two corresponding sites affect protease activity in a temperature-dependent fashion (57, 58) (Fig. 4).

Concerning selected sites in nsp10 and ORF1b, interesting findings relate to the nps10-nsp14 and nsp10-nsp16 interactions. Based on the crystal structure of the corresponding SARS-CoV protein complexes, one of the positively selected sites in nsp16 (K249) is located at the direct interaction surface with nsp10 (Fig. 5A). Also, the only positively selected site we detected in nsp10 (S61) is involved in the interaction with both nsp16 (59) and nsp14 (60) (Fig. 5). In fact, mutation of residue 61 in SARS-CoV nsp10 (A61V mutant) strongly reduces interaction with nsp14. In MHV and SARS-CoV, S61 is located on an exposed loop that also includes Q65 (Fig. 5); mutation of Q65 results in a temperature-sensitive (*ts*) phenotype in MHV and, through a poorly understood mechanism, disrupts nsp5 function (61). This observation suggests that this loop may be involved in interaction with other nsps or that amino acid changes in this loop can modify nsp10 conformation (61).

Additional selected sites in nsp14 and nsp16 do not involve the contact interface with nsp10. In nsp14, T228 is located within one of the zinc fingers of the protein (Fig. 5); these domains are essential for the ExoN proofreading activity of nsp14 (62). Another selected residue (S284) flanks a position (F286 in SARS, highly conserved in coronaviruses) that is involved in the hydrophobic interaction between the ExoN and N7-methyltransferase domain (62).

As for nsp16, the T151 selected site is in close proximity to a residue (L153) (Fig. 5) that originates a *ts* phenotype in MHV; specifically, the mutation has an effect on RNA synthesis (63). Finally, the positively selected site 138 (Asn in MERS-CoV) is located on a solvent-exposed flexible loop (not present in the structure) on the RNA binding groove; this loop may be involved in interaction with RNA (64).

DISCUSSION

In recent years, the emergence of SARS-CoV and MERS-CoV as dangerous zoonoses stirred great interest in the ecology and evolution of coronaviruses. Both viruses originated in bats and spread to humans via an intermediate host (3, 65). This clearly highlights

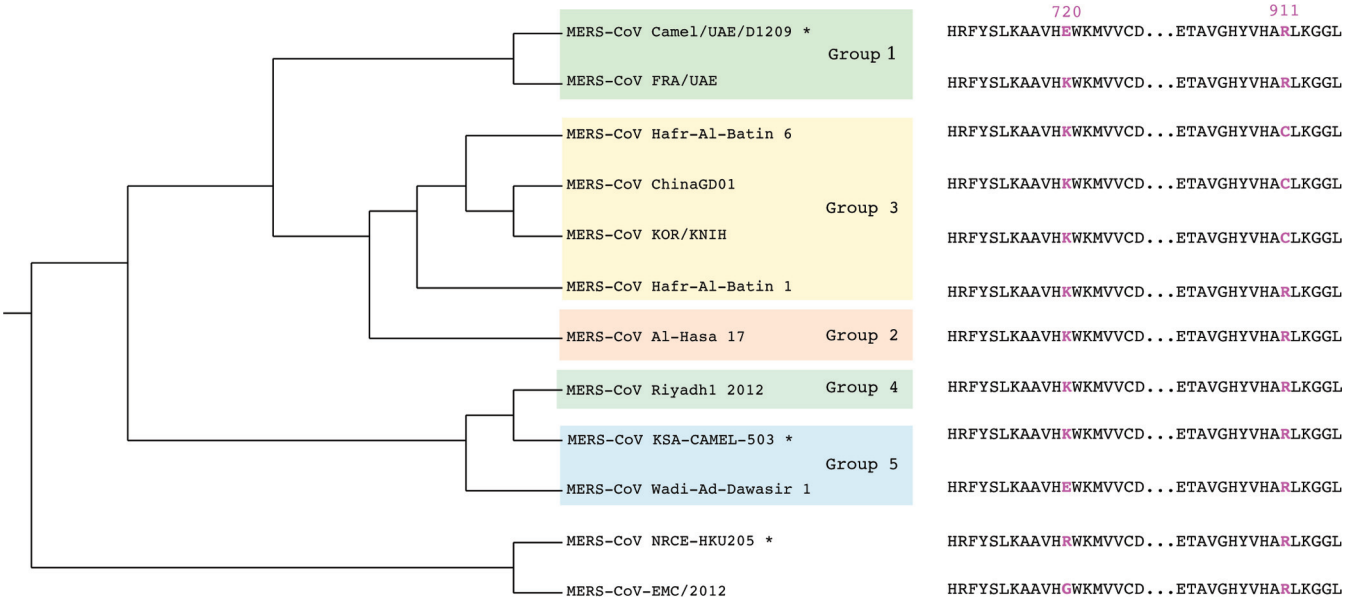


FIG 2 Maximum likelihood phylogeny for nsp3 sequences in a subset of isolates representing MERS-CoV major groups. The amino acid alignment of the region surrounding the two positively selected sites (magenta) in MERS-CoV isolates is also shown. Asterisks indicate viruses isolated from dromedary camels.

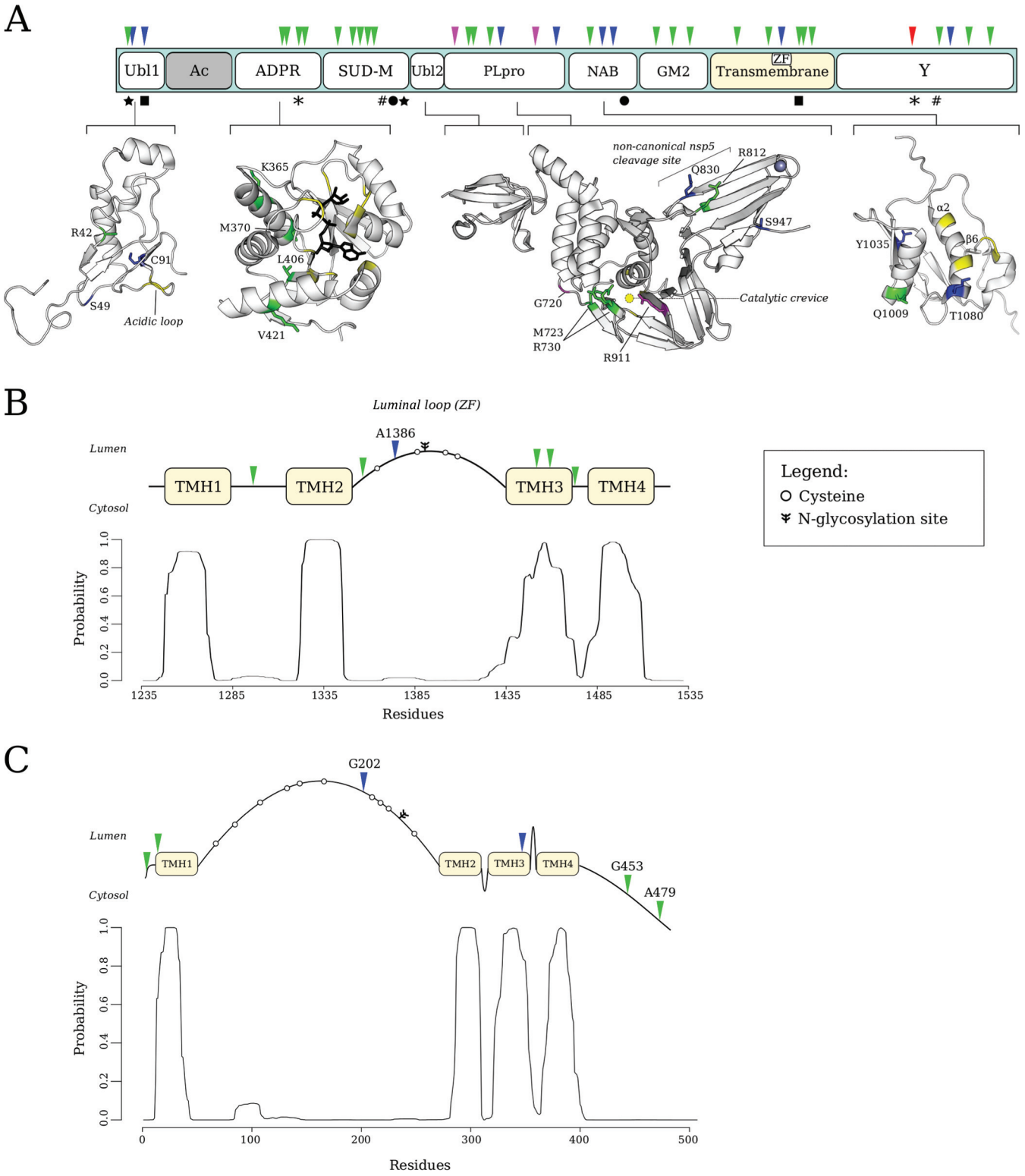
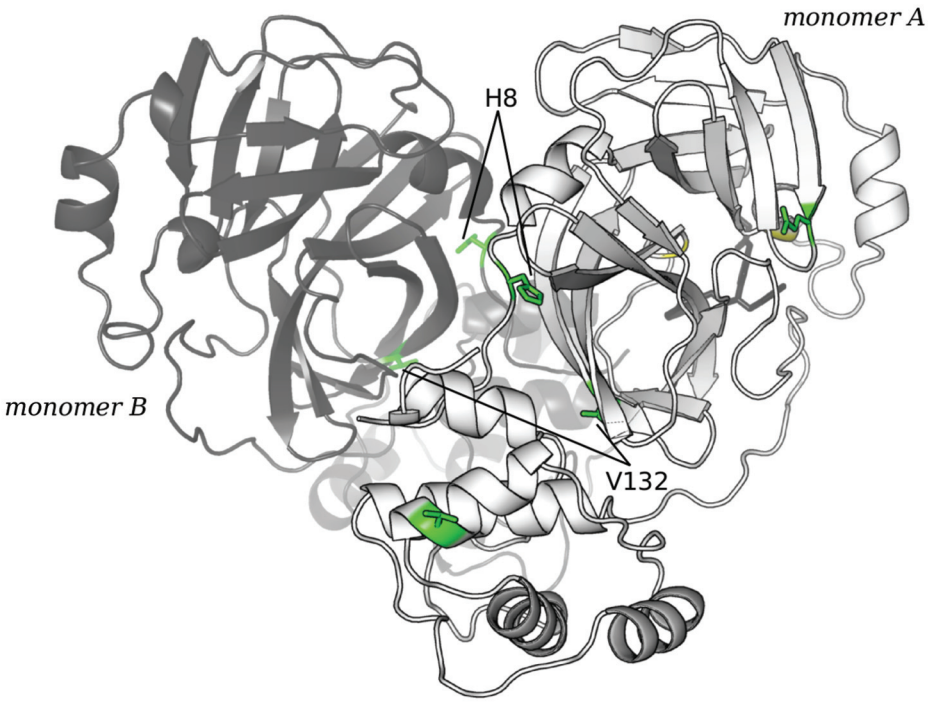


FIG 3 (A) Representation of nsp3 domain architecture. Positively selected sites are indicated by triangles, coevolving sites with symbols (see Fig. 1 legend). In the enlargements, positively selected sites were mapped onto known domain 3D structures of MERS-CoV or SARS-CoV (PDB codes 2GRI, 3EWR, 4RNA, and 2K87). The acidic domain is shown in gray because it was not analyzed (see the text). (B and C) Topology maps and probability diagrams of transmembrane helices for MERS-CoV nsp3 transmembrane domain (B) and MERS-CoV nsp4 (C). The conserved cysteine residues and the predicted N-glycosylation sites are mapped onto the luminal loops. Color codes are as in Fig. 1; yellow indicates protein regions or sites known to be functional and mentioned in the text.

A



B

```
                          1                    32  121                          149
MERS-CoV-EMC/2012         SGLVKMSHPSGDVEACMVQVTCG...YNGRPTGTFTVVMRPNYTIKGSFLCGSCG
MERS-CoV-Al-Hasa17        SGLVKMSHPSGDVEACMVQVTCG...YNGRPTGTFTVVMRPNYTIKGSFLCGSCG
MERS-CoV-Wadi-Ad-Dawasir1 SGLVKMSHPSGDVEACMVQVTCG...YNGRPTGTFTVVMRPNYTIKGSFLCGSCG
MERS-CoV-FRA/UAE          SGLVKMSHPSGDVEACMVQVTCG...YNGRPTGTFTVVMRPNYTIKGSFLCGSCG
MERS-CoV-Riyadh1          SGLVKMSHPSGDVEACMVQVTCG...YNGRPTGTFTVVMRPNYTIKGSFLCGSCG
MERS-CoV-Hafr-Al-Batin1   SGLVKMSHPSGDVEACMVQVTCG...YNGRPTGTFTVVMRPNYTIKGSFLCGSCG
NeoCoV                    SGLVKMSHPSGDVEACMVQVTCG...YNGRPTGTFTVIMRPNYTIKGSFLCGSCG
BtVs-BetaCoV/SC2013       SGLVKMSHPSGAVEACMVQVTCG...YNGRPTGTFTVVMRPNHTIKGSFLCGSCG
Pi-BatCoV HKU5            SGLVKMAAPSGVVENCMVQVTCG...YNGRPTGVFMVNMRQNSTIKGSFLCGSCG
BtCoV/133                 SGLVKMSAPSGAVENCIVQVTCG...YNGKPTGVFTVNLRHNSTIKGSFLCGSCG
Ty-BatCoV HKU4            SGLVKMSAPSGAVENCIVQVTCG...YNGKPTGVFTVNLRHNSTIKGSFLCGSCG
EriCoV/2012-174           SGLVKMAHPSGAVEQCIVQVTCG...YNGRPSGTYTVVMRPNSTIKGSFLCGSCG
EriCoV/2012-216           SGLVKMAHPSGAVEQCIVQVTCG...YNGRPSGTYTVVMRPNSTIKGSFLCGSCG
SARS CoV                  SGFRKMAFPSGKVEGCMVQVTCG...YNGSPSGVYQCAMRPNHTIKGSFLNGSCG
MHV-A59                   SGIVKMVSPTSKVEPCIVSVTYG...YNGRPQGAFHVTLRSSHTIKGSFLCGSCG
```

FIG 4 (A) Structure of the dimeric form of MERS-CoV nsp5 (PDB code 4YLU). Positively selected sites affecting the dimerization process are labeled. Color codes are as in Fig. 1. Catalytic residues are in yellow. (B) Amino acid alignment of the nsp5 regions surrounding selected sites. Residues that confer a temperature-sensitive phenotype when mutated in MHV are underlined (see the text).

the potential for CoV host shifting and the relevance of understanding the molecular events underlying the adaptation to new species. Since the identification of MERS-CoV, a number of related viruses have been isolated from bats (and other mammals) all over the world, suggesting a wide distribution of betaCoVs. Unfortunately, most of these studies reported partial viral sequences, hampering a fully resolved phylogenetic analysis (66–68). Herein we included only viral species or strains with complete information for ORF1a and ORF1b, and the sequences we analyzed are relatively divergent. This may introduce a high false-positivity rate in the inference of positive selection due to two major issues: the unreliability of sequence alignments and the saturation of substitution rates. To circumvent these possible problems, we adopted a stringent alignment filtering criterion and computed substitution rates over phylogenies. The latter procedure allows breaking of long branches, resulting in improved rate estimation. In fact, branch site tests are relatively insensitive to the saturation issue (69). Moreover, we supported all claims of positive selection by the combined use of two methods. Although this may have resulted in a loss of power (that is intrinsically low for branch site tests [31]), we were able to detect several positively selected sites along different branches of the phylogenies. In both ORF1a and ORF1b, episodic selection was particularly strong on the branch leading to bat CoVs (HKU4, HKU5, and BtCoV133), possibly reflecting specific characteristics of these host species (large population sizes, high seroprevalence, and wide geographic distributions).

Recently, a study of the spike (S) protein in MERS-CoV-related

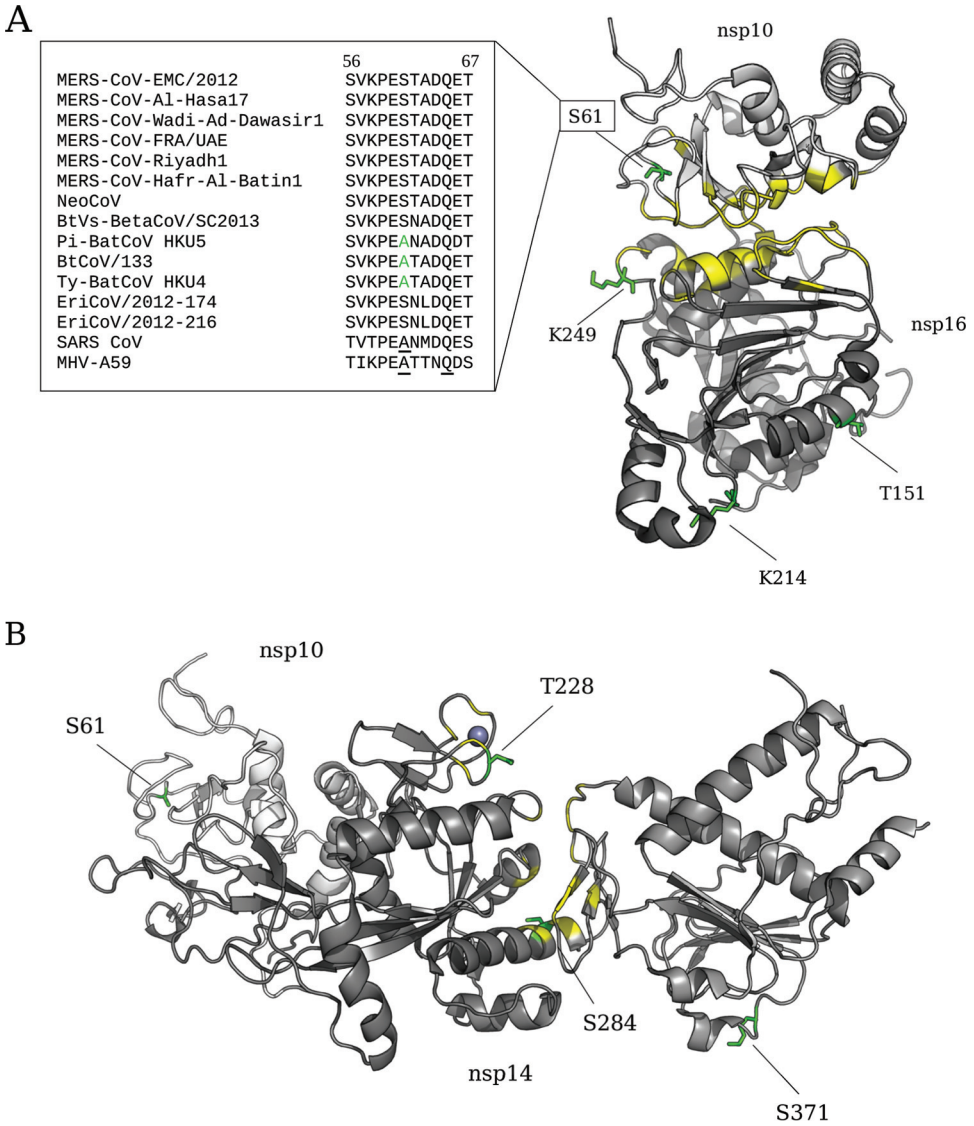


FIG 5 Ribbon representation of SARS-CoV nsp10-nsp16 complex (PDB code 2XYQ) (A) and SARS-CoV nsp10-nsp14 complex (PDB code 5C8U) (B). Positively selected sites in lineage C betaCoVs are shown in green (see Fig. 1), and residues involved in inter- or intraprotein interactions are shown in yellow. An amino acid alignment of the nsp10 exposed loop containing S61 and Q65 is also shown. Functional residues in MHV-CoV and SARS-CoV are underlined (see the text).

viruses analyzed a set of viral sequences very similar to the ones in this study and detected positive selection at 9 sites (24). The proportion of selected sites we identified in the ORF1a region is much higher than that for the *S* gene. Although this is counterintuitive, as the S protein is exposed on the viral surface and functions as a central determinant of host range and of antibody response, this observation indicates that ORF1a represented a major target of positive selection during the evolution of lineage C betaCoVs. Therefore, adaptive evolution in this region might contribute to host shifts or immune evasion. In line with these results, studies on avian influenza A viruses showed that viral surface proteins are not the sole determinants of host range. Conversely, adaptation of avian influenza virus to mammalian hosts was found to be critically dependent on changes in the polymerase genes (22, 23). Recently, an extended analysis indicated that the adaptation of avian influenza virus to swine was accompanied by substitutions in almost all viral genes. Mutation accrual continued for a long time after the host shift, suggesting that multiple mutations progressively optimize viral fitness in the new host (70); many of these mutations were suggested to represent compensatory or epistatic changes (70). Interestingly, we also found evidence of coevolution between site pairs, possibly suggesting epistatic interactions. The vast majority of coevolving sites were located in nsp3, in line with the view that epistatically interacting substitutions are enriched in protein regions undergoing adaptive evolution (71). Epistasis is very common in viruses and is thought to play an important role in the evolution of immune evasion, host shifts, and drug resistance (72, 73); the latter is likely not at play in the case of MERS-CoV and related viruses. Previous studies on RNA viruses have focused mainly on the effect of epistatic interaction on the emergence of drug resistance (e.g., influenza virus hemagglutinin and neuraminidase, HCV NS3 protease, HIV-1 protease, and reverse

transcriptase) or antibody escape (influenza virus hemagglutinin and neuraminidase) (74–77). Little information is available on the possible role of coevolving sites in the context of innate immunity or adaptation to new hosts. Nonetheless, epistatic interaction between two sites in the Chikungunya virus E1 envelope glycoprotein underlies the ability of the pathogen to infect a new mosquito vector (78). These observations suggest that intra- and intergene epistasis contributes to determination of the evolutionary trajectories of viral species whenever the environment (broadly defined, including the host species) changes (73). The coevolving sites we identified are located in distinct domains of nsp3, and the pair in nsp12 and nsp14 does not involve residues at the interaction surface between the two proteins. Thus, we are presently unable to infer the molecular mechanism underlying their interaction. The generation of mutant viruses or recombinant viral proteins carrying different amino acids at coevolving site pairs will be necessary to address the nature of their interaction and the effect on the viral phenotype. We note, however, that long-range interactions are likely to be common in ORF1a and ORF1b, at both the intra- and interprotein levels, as demonstrated by experiments with SARS-CoV and MHV mutants (57, 61). For instance, a temperature-sensitive (*ts*) mutation in MHV nsp10 affects viral replication by blocking the activity of nsp5 (61). Also, *ts* mutations in MHV nsp5 can be rescued by second-site suppressor changes located in physically distant protein regions (57).

The CoV RTC is extremely complex and comprises a number of enzymatic activities that constitute a unique repertoire among RNA viruses. Several molecular details on the assembly and functioning of the RTC are still missing, and the mechanisms underlying the phenotypic effects of specific changes are often poorly understood. For instance, Stokes and coworkers described a conservative mutation (M575I) located in the N-terminal portion of the ADPR domain of MHV nsp3 that determines a *ts* phenotype with impaired RNA synthesis (79). How the mutation exerts its effects, though, remains unknown. We thus interpreted the effect of positively selected sites in light of what is known of CoV biology, and we compared the location of selected sites to the few substitutions (in either SARS-CoV or MHV) with a known effect, as detailed above. Clearly, the selected sites we identified in this study lend themselves to experimental testing to assess their impact on BetaCoV phenotypes. Indeed, evolutionary analyses can provide information on the location and nature of adaptive changes, thus highlighting the presence of functional genetic variants. For instance, the major selection target, nsp3, has multiple enzymatic functions. It would be extremely interesting to use site-directed mutagenesis and biochemical analyses to assess whether the selected sites in the PLpro domain affect the enzyme's specificity not only toward the viral polyprotein but also in terms of deubiquitination and deISGylation activities, which, in turn, may modulate the viral ability to evade host immune responses (16). Likewise, expression of mutant nsp3 and nsp4 carrying different amino acids at the positively selected sites in the large luminal loops will be instrumental to determine whether, as shown for other CoVs (53, 54), these changes affect the formation of membrane rearrangements onto which the RTC assembles. Ultimately, the generation of mutant CoVs followed by *in vitro* infection will clarify the effects of specific changes on viral fitness, at least in cell cultures.

It is noteworthy that we found nsp3 to represent a preferential target of positive selection and that the adaptive process is ongoing in circulating MERS-CoV strains. A previous genome-wide study of positive selection in MERS-CoV detected only one positively selected site in the *S* gene (2); however, fewer strains than those we analyzed in this study were available at the time of that study, and the authors did not include sequences obtained from camels (2). The two selected sites we identified are located in the PLpro domain of nsp3 and, together with two sites selected in bats, map to the "palm" portion of the right-handed structure of the protease. The palm also accommodates the catalytic triad, but the selected sites are located at the opposite side of the crevice, suggesting that they may exert an effect by altering the conformational structure of the protease or via interaction with other viral components or host proteins. In addition to its role in viral replication, PLpro functions as a multitasking inhibitor of IFN responses and physically associates with several host innate immunity molecules (19). Because antagonism of the host immune system is considered a major driver of evolutionary change in viruses (80), the G720K variant is an excellent candidate as a modulator of host responses. However, variability at position 720 is also observed among viruses isolated from camels, suggesting that the selective event ensued in these animals or in a previous host and does not represent a specific adaptation to humans. A similar observation was previously reported for the positively selected site in the S protein (24). Conversely, variation at the second selected site (R911) in PLpro was observed only for viruses isolated from human patients. Although this finding may simply reflect the sparse sampling of camel-derived viruses, adaptation of a zoonotic virus to a new host is expected to result in selective events that optimize viral fitness in terms of replication efficiency, transmissibility, and immune evasion (72).

In general, a deeper understanding of the adaptive events that underlie host shifts in CoVs and other viruses will be pivotal to predict and prevent future zoonoses. Bats alone host a variety of CoVs that represent potential threats to human health (66–68). Data herein suggest that monitoring programs and field surveys of CoV diversity and prevalence should envisage molecular characterization of nsp3.

ACKNOWLEDGMENTS

N.A.-D. was supported by the Deanship of Scientific Research, Prolific Research Group Program (PRG-1436-15), Vice Rectorate for Graduate Studies and Scientific Research, at King Saud University (KSU), Riyadh, Saudi Arabia.

The funder had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

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. **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, Alrabiah 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. http://dx.doi.org/10.1128/mBio.01062-13.
3. **Al-Tawfiq JA, Memish ZA.** 2014. Middle East respiratory syndrome coronavirus: transmission and phylogenetic evolution. Trends Microbiol **22:** 573–579. http://dx.doi.org/10.1016/j.tim.2014.08.001.
4. **de Groot RJ, Baker SC, Baric RS, Brown CS, Drosten C, Enjuanes L, Fouchier RA, Galiano M, Gorbalenya AE, Memish ZA, Perlman S, Poon LL, Snijder EJ, Stephens GM, Woo PC, Zaki AM, Zambon M,**

Ziebuhr J. 2013. Middle East respiratory syndrome coronavirus (MERS-CoV): announcement of the Coronavirus Study Group. J Virol **87:**7790–7792. http://dx.doi.org/10.1128/JVI.01244-13.
5. **Lauber C, Goeman JJ, Parquet Mdel C, Nga PT, Snijder EJ, Morita K, Gorbalenya AE.** 2013. The footprint of genome architecture in the largest genome expansion in RNA viruses. PLoS Pathog **9:**e1003500. http://dx.doi.org/10.1371/journal.ppat.1003500.
6. **Hagemeijer MC, Rottier PJ, de Haan CA.** 2012. Biogenesis and dynamics of the coronavirus replicative structures. Viruses **4:**3245–3269. http://dx.doi.org/10.3390/v4113245.
7. **Neuman BW, Chamberlain P, Bowden F, Joseph J.** 2014. Atlas of coronavirus replicase structure. Virus Res **194:**49–66. http://dx.doi.org/10.1016/j.virusres.2013.12.004.
8. **Kindler E, Thiel V.** 2014. To sense or not to sense viral RNA—essentials of coronavirus innate immune evasion. Curr Opin Microbiol **20:**69–75. http://dx.doi.org/10.1016/j.mib.2014.05.005.
9. **Kopecky-Bromberg SA, Martinez-Sobrido L, Frieman M, Baric RA, Palese P.** 2007. Severe acute respiratory syndrome coronavirus open reading frame (ORF) 3b, ORF 6, and nucleocapsid proteins function as interferon antagonists. J Virol **81:**548–557. http://dx.doi.org/10.1128/JVI.01782-06.
10. **Frieman M, Yount B, Heise M, Kopecky-Bromberg SA, Palese P, Baric RS.** 2007. Severe acute respiratory syndrome coronavirus ORF6 antagonizes STAT1 function by sequestering nuclear import factors on the rough endoplasmic reticulum/Golgi membrane. J Virol **81:**9812–9824. http://dx.doi.org/10.1128/JVI.01012-07.
11. **Yang Y, Zhang L, Geng H, Deng Y, Huang B, Guo Y, Zhao Z, Tan W.** 2013. The structural and accessory proteins M, ORF 4a, ORF 4b, and ORF 5 of Middle East respiratory syndrome coronavirus (MERS-CoV) are potent interferon antagonists. Protein Cell **4:**951–961. http://dx.doi.org/10.1007/s13238-013-3096-8.
12. **Yang Y, Ye F, Zhu N, Wang W, Deng Y, Zhao Z, Tan W.** 2015. Middle East respiratory syndrome coronavirus ORF4b protein inhibits type I interferon production through both cytoplasmic and nuclear targets. Sci Rep **5:**17554. http://dx.doi.org/10.1038/srep17554.
13. **Qi X, Lan S, Wang W, Schelde LM, Dong H, Wallat GD, Ly H, Liang Y, Dong C.** 2010. Cap binding and immune evasion revealed by Lassa nucleoprotein structure. Nature **468:**779–783. http://dx.doi.org/10.1038/nature09605.
14. **Python S, Gerber M, Suter R, Ruggli N, Summerfield A.** 2013. Efficient sensing of infected cells in absence of virus particles by plasmacytoid dendritic cells is blocked by the viral ribonuclease E(rns). PLoS Pathog **9:**e1003412. http://dx.doi.org/10.1371/journal.ppat.1003412.
15. **Báez-Santos YM, St John SE, Mesecar AD.** 2015. The SARS-coronavirus papain-like protease: structure, function and inhibition by designed antiviral compounds. Antiviral Res **115:**21–38. http://dx.doi.org/10.1016/j.antiviral.2014.12.015.
16. **Báez-Santos YM, Mielech AM, Deng X, Baker S, Mesecar AD.** 2014. Catalytic function and substrate specificity of the papain-like protease domain of nsp3 from the Middle East respiratory syndrome coronavirus. J Virol **88:**12511–12527. http://dx.doi.org/10.1128/JVI.01294-14.
17. **Kamitani W, Huang C, Narayanan K, Lokugamage KG, Makino S.** 2009. A two-pronged strategy to suppress host protein synthesis by SARS coronavirus Nsp1 protein. Nat Struct Mol Biol **16:**1134–1140. http://dx.doi.org/10.1038/nsmb.1680.
18. **Wathelet MG, Orr M, Frieman MB, Baric RS.** 2007. Severe acute respiratory syndrome coronavirus evades antiviral signaling: role of nsp1 and rational design of an attenuated strain. J Virol **81:**11620–11633. http://dx.doi.org/10.1128/JVI.00702-07.
19. **Chen X, Yang X, Zheng Y, Yang Y, Xing Y, Chen Z.** 2014. SARS coronavirus papain-like protease inhibits the type I interferon signaling pathway through interaction with the STING-TRAF3-TBK1 complex. Protein Cell **5:**369–381. http://dx.doi.org/10.1007/s13238-014-0026-3.
20. **Chan JF, Lau SK, To KK, Cheng VC, Woo PC, Yuen KY.** 2015. Middle East respiratory syndrome coronavirus: another zoonotic betacoronavirus causing SARS-like disease. Clin Microbiol Rev **28:**465–522. http://dx.doi.org/10.1128/CMR.00102-14.
21. **Sironi M, Cagliani R, Forni D, Clerici M.** 2015. Evolutionary insights into host-pathogen interactions from mammalian sequence data. Nat Rev Genet **16:**224–236. http://dx.doi.org/10.1038/nrg3905.
22. **Hatta M, Gao P, Halfmann P, Kawaoka Y.** 2001. Molecular basis for high virulence of Hong Kong H5N1 influenza A viruses. Science **293:**1840–1842. http://dx.doi.org/10.1126/science.1062882.
23. **Gabriel G, Dauber B, Wolff T, Planz O, Klenk HD, Stech J.** 2005. The viral polymerase mediates adaptation of an avian influenza virus to a mammalian host. Proc Natl Acad Sci U S A **102:**18590–18595. http://dx.doi.org/10.1073/pnas.0507415102.
24. **Forni D, Filippi G, Cagliani R, De Gioia L, Pozzoli U, Al-Daghri N, Clerici M, Sironi M.** 2015. The heptad repeat region is a major selection target in MERS-CoV and related coronaviruses. Sci Rep **5:**14480. http://dx.doi.org/10.1038/srep14480.
25. **Löytynoja A, Goldman N.** 2005. An algorithm for progressive multiple alignment of sequences with insertions. Proc Natl Acad Sci U S A **102:**10557–10562. http://dx.doi.org/10.1073/pnas.0409137102.
26. **Penn O, Privman E, Ashkenazy H, Landan G, Graur D, Pupko T.** 2010. GUIDANCE: a web server for assessing alignment confidence scores. Nucleic Acids Res **38:**W23–W28. http://dx.doi.org/10.1093/nar/gkq443.
27. **Privman E, Penn O, Pupko T.** 2012. Improving the performance of positive selection inference by filtering unreliable alignment regions. Mol Biol Evol **29:**1–5. http://dx.doi.org/10.1093/molbev/msr177.
28. **Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD.** 2006. Automated phylogenetic detection of recombination using a genetic algorithm. Mol Biol Evol **23:**1891–1901. http://dx.doi.org/10.1093/molbev/msl051.
29. **Sawyer S.** 1989. Statistical tests for detecting gene conversion. Mol Biol Evol **6:**526–538.
30. **Guindon S, Delsuc F, Dufayard JF, Gascuel O.** 2009. Estimating maximum likelihood phylogenies with PhyML. Methods Mol Biol **537:**113–137. http://dx.doi.org/10.1007/978-1-59745-251-9_6.
31. **Zhang J, Nielsen R, Yang Z.** 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol Biol Evol **22:**2472–2479. http://dx.doi.org/10.1093/molbev/msi237.
32. **Yang Z.** 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol **24:**1586–1591. http://dx.doi.org/10.1093/molbev/msm088.
33. **Anisimova M, Yang Z.** 2007. Multiple hypothesis testing to detect lineages under positive selection that affects only a few sites. Mol Biol Evol **24:**1219–1228. http://dx.doi.org/10.1093/molbev/msm042.
34. **Murrell B, Weaver S, Smith MD, Wertheim JO, Murrell S, Aylward A, Eren K, Pollner T, Martin DP, Smith DM, Scheffler K, Kosakovsky Pond SL.** 19 February 2015. Gene-wide identification of episodic selection. Mol Biol Evol http://dx.doi.org/10.1093/molbev/msv035.
35. **Pond SL, Frost SD, Muse SV.** 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics **21:**676–679. http://dx.doi.org/10.1093/bioinformatics/bti079.
36. **Anisimova M, Bielawski JP, Yang Z.** 2002. Accuracy and power of Bayes prediction of amino acid sites under positive selection. Mol Biol Evol **19:**950–958. http://dx.doi.org/10.1093/oxfordjournals.molbev.a004152.
37. **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.
38. **Delport W, Poon AF, Frost SD, Kosakovsky Pond SL.** 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics **26:**2455–2457. http://dx.doi.org/10.1093/bioinformatics/btq429.
39. **Poon AF, Lewis FI, Pond SL, Frost SD.** 2007. An evolutionary-network model reveals stratified interactions in the V3 loop of the HIV-1 envelope. PLoS Comput Biol **3:**e231. http://dx.doi.org/10.1371/journal.pcbi.0030231.
40. **Simonetti FL, Teppa E, Chernomoretz A, Nielsen M, Marino Buslje C.** 2013. MISTIC: Mutual Information Server To Infer Coevolution. Nucleic Acids Res **41:**W8–W14. http://dx.doi.org/10.1093/nar/gkt427.
41. **Krogh A, Larsson B, von Heijne G, Sonnhammer EL.** 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol **305:**567–580. http://dx.doi.org/10.1006/jmbi.2000.4315.
42. **Yang L, Wu Z, Ren X, Yang F, Zhang J, He G, Dong J, Sun L, Zhu Y, Zhang S, Jin Q.** 2014. MERS-related betacoronavirus in Vespertilio superans bats, China. Emerg Infect Dis **20:**1260–1262. http://dx.doi.org/10.3201/eid2007.140318.
43. **Tang XC, Zhang JX, Zhang SY, Wang P, Fan XH, Li LF, Li G, Dong BQ, Liu W, Cheung CL, Xu KM, Song WJ, Vijaykrishna D, Poon LL, Peiris JS, Smith GJ, Chen H, Guan Y.** 2006. Prevalence and genetic diversity of coronaviruses in bats from China. J Virol **80:**7481–7490. http://dx.doi.org/10.1128/JVI.00697-06.
44. **Corman VM, Kallies R, Philipps H, Gopner G, Muller MA, Eckerle I,**

Brunink S, Drosten C, Drexler JF. 2014. Characterization of a novel betacoronavirus related to middle East respiratory syndrome coronavirus in European hedgehogs. J Virol **88:**717–724. http://dx.doi.org/10.1128/JVI.01600-13.

45. **Corman VM, Ithete NL, Richards LR, Schoeman MC, Preiser W, Drosten C, Drexler JF.** 2014. Rooting the phylogenetic tree of Middle East respiratory syndrome coronavirus by characterization of a conspecific virus from an African bat. J Virol **88:**11297–11303. http://dx.doi.org/10.1128/JVI.01498-14.
46. **Wang Y, Liu D, Shi W, Lu R, Wang W, Zhao Y, Deng Y, Zhou W, Ren H, Wu J, Wang Y, Wu G, Gao GF, Tan W.** 2015. Origin and possible genetic recombination of the Middle East respiratory syndrome coronavirus from the first imported case in China: phylogenetics and coalescence analysis. mBio **6**(5):e01280-15. http://dx.doi.org/10.1128/mBio.01280-15.
47. **Serrano P, Johnson MA, Almeida MS, Horst R, Herrmann T, Joseph JS, Neuman BW, Subramanian V, Saikatendu KS, Buchmeier MJ, Stevens RC, Kuhn P, Wuthrich K.** 2007. Nuclear magnetic resonance structure of the N-terminal domain of nonstructural protein 3 from the severe acute respiratory syndrome coronavirus. J Virol **81:**12049–12060. http://dx.doi.org/10.1128/JVI.00969-07.
48. **Wu A, Wang Y, Zeng C, Huang X, Xu S, Su C, Wang M, Chen Y, Guo D.** 2015. Prediction and biochemical analysis of putative cleavage sites of the 3C-like protease of Middle East respiratory syndrome coronavirus. Virus Res **208:**56–65. http://dx.doi.org/10.1016/j.virusres.2015.05.018.
49. **Hurst KR, Koetzner CA, Masters PS.** 2013. Characterization of a critical interaction between the coronavirus nucleocapsid protein and nonstructural protein 3 of the viral replicase-transcriptase complex. J Virol **87:**9159–9172. http://dx.doi.org/10.1128/JVI.01275-13.
50. **Xu Y, Cong L, Chen C, Wei L, Zhao Q, Xu X, Ma Y, Bartlam M, Rao Z.** 2009. Crystal structures of two coronavirus ADP-ribose-1″-monophosphatases and their complexes with ADP-ribose: a systematic structural analysis of the viral ADRP domain. J Virol **83:**1083–1092. http://dx.doi.org/10.1128/JVI.01862-08.
51. **Serrano P, Ramajo J, Martinez-Salas E.** 2009. Rescue of internal initiation of translation by RNA complementation provides evidence for a distribution of functions between individual IRES domains. Virology **388:**221–229. http://dx.doi.org/10.1016/j.virol.2009.03.021.
52. **Oostra M, Hagemeijer MC, van Gent M, Bekker CP, te Lintelo EG, Rottier PJ, de Haan CA.** 2008. Topology and membrane anchoring of the coronavirus replication complex: not all hydrophobic domains of nsp3 and nsp6 are membrane spanning. J Virol **82:**12392–12405. http://dx.doi.org/10.1128/JVI.01219-08.
53. **Hagemeijer MC, Monastyrska I, Griffith J, van der Sluijs P, Voortman J, van Bergen en Henegouwen PM, Vonk AM, Rottier PJ, Reggiori F, de Haan CA.** 2014. Membrane rearrangements mediated by coronavirus nonstructural proteins 3 and 4. Virology **458-459:**125–135. http://dx.doi.org/10.1016/j.virol.2014.04.027.
54. **Beachboard DC, Anderson-Daniels JM, Denison MR.** 2015. Mutations across murine hepatitis virus nsp4 alter virus fitness and membrane modifications. J Virol **89:**2080–2089. http://dx.doi.org/10.1128/JVI.02776-14.
55. **Xu X, Lou Z, Ma Y, Chen X, Yang Z, Tong X, Zhao Q, Xu Y, Deng H, Bartlam M, Rao Z.** 2009. Crystal structure of the C-terminal cytoplasmic domain of non-structural protein 4 from mouse hepatitis virus A59. PLoS One **4:**e6217. http://dx.doi.org/10.1371/journal.pone.0006217.
56. **Tomar S, Johnston ML, St John SE, Osswald HL, Nyalapatla PR, Paul LN, Ghosh AK, Denison MR, Mesecar AD.** 2015. Ligand-induced dimerization of Middle East respiratory syndrome (MERS) coronavirus nsp5 protease (3CLpro): implications for nsp5 regulation and the development of antivirals. J Biol Chem **290:**19403–19422. http://dx.doi.org/10.1074/jbc.M115.651463.
57. **Stobart CC, Lee AS, Lu X, Denison MR.** 2012. Temperature-sensitive mutants and revertants in the coronavirus nonstructural protein 5 protease (3CLpro) define residues involved in long-distance communication and regulation of protease activity. J Virol **86:**4801–4810. http://dx.doi.org/10.1128/JVI.06754-11.
58. **Stobart CC, Sexton NR, Munjal H, Lu X, Molland KL, Tomar S, Mesecar AD, Denison MR.** 2013. Chimeric exchange of coronavirus nsp5 proteases (3CLpro) identifies common and divergent regulatory determinants of protease activity. J Virol **87:**12611–12618. http://dx.doi.org/10.1128/JVI.02050-13.
59. **Lugari A, Betzi S, Decroly E, Bonnaud E, Hermant A, Guillemot JC, Debarnot C, Borg JP, Bouvet M, Canard B, Morelli X, Lecine P.** 2010. Molecular mapping of the RNA cap 2′-O-methyltransferase activation interface between severe acute respiratory syndrome coronavirus nsp10 and nsp16. J Biol Chem **285:**33230–33241. http://dx.doi.org/10.1074/jbc.M110.120014.
60. **Bouvet M, Lugari A, Posthuma CC, Zevenhoven JC, Bernard S, Betzi S, Imbert I, Canard B, Guillemot JC, Lecine P, Pfefferle S, Drosten C, Snijder EJ, Decroly E, Morelli X.** 2014. Coronavirus Nsp10, a critical co-factor for activation of multiple replicative enzymes. J Biol Chem **289:**25783–25796. http://dx.doi.org/10.1074/jbc.M114.577353.
61. **Donaldson EF, Graham RL, Sims AC, Denison MR, Baric RS.** 2007. Analysis of murine hepatitis virus strain A59 temperature-sensitive mutant TS-LA6 suggests that nsp10 plays a critical role in polyprotein processing. J Virol **81:**7086–7098. http://dx.doi.org/10.1128/JVI.00049-07.
62. **Ma Y, Wu L, Shaw N, Gao Y, Wang J, Sun Y, Lou Z, Yan L, Zhang R, Rao Z.** 2015. Structural basis and functional analysis of the SARS coronavirus nsp14-nsp10 complex. Proc Natl Acad Sci U S A **112:**9436–9441. http://dx.doi.org/10.1073/pnas.1508686112.
63. **Sawicki SG, Sawicki DL, Younker D, Meyer Y, Thiel V, Stokes H, Siddell SG.** 2005. Functional and genetic analysis of coronavirus replicase-transcriptase proteins. PLoS Pathog **1:**e39. http://dx.doi.org/10.1371/journal.ppat.0010039.
64. **Decroly E, Debarnot C, Ferron F, Bouvet M, Coutard B, Imbert I, Gluais L, Papageorgiou N, Sharff A, Bricogne G, Ortiz-Lombardia M, Lescar J, Canard B.** 2011. Crystal structure and functional analysis of the SARS-coronavirus RNA cap 2′-O-methyltransferase nsp10/nsp16 complex. PLoS Pathog **7:**e1002059. http://dx.doi.org/10.1371/journal.ppat.1002059.
65. **Lau SK, Woo PC, Li KS, Huang Y, Tsoi HW, Wong BH, Wong SS, Leung SY, Chan KH, Yuen KY.** 2005. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. Proc Natl Acad Sci U S A **102:**14040–14045. http://dx.doi.org/10.1073/pnas.0506735102.
66. **Annan A, Baldwin HJ, Corman VM, Klose SM, Owusu M, Nkrumah EE, Badu EK, Anti P, Agbenyega O, Meyer B, Oppong S, Sarkodie YA, Kalko EK, Lina PH, Godlevska EV, Reusken C, Seebens A, Gloza-Rausch F, Vallo P, Tschapka M, Drosten C, Drexler JF.** 2013. Human betacoronavirus 2c EMC/2012-related viruses in bats, Ghana and Europe. Emerg Infect Dis **19:**456–459. http://dx.doi.org/10.3201/eid1903.121503.
67. **Ithete NL, Stoffberg S, Corman VM, Cottontail VM, Richards LR, Schoeman MC, Drosten C, Drexler JF, Preiser W.** 2013. Close relative of human Middle East respiratory syndrome coronavirus in bat, South Africa. Emerg Infect Dis **19:**1697–1699. http://dx.doi.org/10.3201/eid1910.130946.
68. **Anthony SJ, Ojeda-Flores R, Rico-Chavez O, Navarrete-Macias I, Zambrana-Torrelio CM, Rostal MK, Epstein JH, Tipps T, Liang E, Sanchez-Leon M, Sotomayor-Bonilla J, Aguirre AA, Avila-Flores R, Medellin RA, Goldstein T, Suzan G, Daszak P, Lipkin WI.** 2013. Coronaviruses in bats from Mexico. J Gen Virol **94:**1028–1038. http://dx.doi.org/10.1099/vir.0.049759-0.
69. **Gharib WH, Robinson-Rechavi M.** 2013. The branch-site test of positive selection is surprisingly robust but lacks power under synonymous substitution saturation and variation in GC. Mol Biol Evol **30:**1675–1686. http://dx.doi.org/10.1093/molbev/mst062.
70. **Bhatt S, Lam TT, Lycett SJ, Leigh Brown AJ, Bowden TA, Holmes EC, Guan Y, Wood JL, Brown IH, Kellam P, Combating Swine Influenza Consortium Pybus OG.** 2013. The evolutionary dynamics of influenza A virus adaptation to mammalian hosts. Philos Trans R Soc Lond B Biol Sci **368:**20120382. http://dx.doi.org/10.1098/rstb.2012.0382.
71. **Gong LI, Bloom JD.** 2014. Epistatically interacting substitutions are enriched during adaptive protein evolution. PLoS Genet **10:**e1004328. http://dx.doi.org/10.1371/journal.pgen.1004328.
72. **Longdon B, Brockhurst MA, Russell CA, Welch JJ, Jiggins FM.** 2014. The evolution and genetics of virus host shifts. PLoS Pathog **10:**e1004395. http://dx.doi.org/10.1371/journal.ppat.1004395.
73. **Bedhomme S, Hillung J, Elena SF.** 2015. Emerging viruses: why they are not jacks of all trades? Curr Opin Virol **10:**1–6. http://dx.doi.org/10.1016/j.coviro.2014.10.006.
74. **Hinkley T, Martins J, Chappey C, Haddad M, Stawiski E, Whitcomb JM, Petropoulos CJ, Bonhoeffer S.** 2011. A systems analysis of mutational effects in HIV-1 protease and reverse transcriptase. Nat Genet **43:**487–489. http://dx.doi.org/10.1038/ng.795.
75. **Parera M, Martinez MA.** 2014. Strong epistatic interactions within a single protein. Mol Biol Evol **31:**1546–1553. http://dx.doi.org/10.1093/molbev/msu113.
76. **Neverov AD, Kryazhimskiy S, Plotkin JB, Bazykin GA.** 2015. Coordi-

nated evolution of influenza A surface proteins. PLoS Genet **11:**e1005404. http://dx.doi.org/10.1371/journal.pgen.1005404.

77. **Kryazhimskiy S, Dushoff J, Bazykin GA, Plotkin JB.** 2011. Prevalence of epistasis in the evolution of influenza A surface proteins. PLoS Genet **7:**e1001301. http://dx.doi.org/10.1371/journal.pgen.1001301.
78. **Tsetsarkin KA, Chen R, Leal G, Forrester N, Higgs S, Huang J, Weaver SC.** 2011. Chikungunya virus emergence is constrained in Asia by lineage-specific adaptive landscapes. Proc Natl Acad Sci U S A **108:**7872–7877. http://dx.doi.org/10.1073/pnas.1018344108.
79. **Stokes HL, Baliji S, Hui CG, Sawicki SG, Baker SC, Siddell SG.** 2010. A new cistron in the murine hepatitis virus replicase gene. J Virol **84:**10148–10158. http://dx.doi.org/10.1128/JVI.00901-10.
80. **Sawyer SL, Elde NC.** 2012. A cross-species view on viruses. Curr Opin Virol **2:**561–568. http://dx.doi.org/10.1016/j.coviro.2012.07.003.