Journal of Virology

Evidence for ACE2-Utilizing Coronaviruses (CoVs) Related to Severe Acute Respiratory Syndrome CoV in Bats

Ann Demogines, Michael Farzan and Sara L. Sawyer
***J. Virol.* 2012, 86(11):6350. DOI: 10.1128/JVI.00311-12.**
Published Ahead of Print 21 March 2012.

JVI Journals.ASM.org

Evidence for ACE2-Utilizing Coronaviruses (CoVs) Related to Severe Acute Respiratory Syndrome CoV in Bats

Ann Demogines,[a] Michael Farzan,[b] and Sara L. Sawyer[a]

Department of Molecular Genetics and Microbiology, Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, Texas, USA,[a] and Department of Microbiology and Immunobiology, Harvard Medical School, New England Primate Research Center, Southborough, Massachusetts, USA[b]

In 2002, severe acute respiratory syndrome (SARS)-coronavirus (CoV) appeared as a novel human virus with high similarity to bat coronaviruses. However, while SARS-CoV uses the human angiotensin-converting enzyme 2 (ACE2) receptor for cellular entry, no coronavirus isolated from bats appears to use ACE2. Here we show that signatures of recurrent positive selection in the bat *ACE2* gene map almost perfectly to known SARS-CoV interaction surfaces. Our data indicate that ACE2 utilization preceded the emergence of SARS-CoV-like viruses from bats.

Cell-surface receptors often play a key role in defining viral host range. New diseases can emerge when existing viruses evolve the ability to bind the ortholog of their cell-surface receptor in a new species (1, 25, 35). Indeed, the principal genetic component defining host range in coronaviruses is the spike protein on the surface of the virus and, in particular, its receptor-binding domain (RBD) (5, 14). It is believed that the severe acute respiratory syndrome (SARS) epidemic resulted from the zoonotic transmission of a coronavirus from bats to humans (15, 18, 32). The central role of the RBD in the SARS-coronavirus (CoV) zoonosis was crystallized in an experiment in which a bat coronavirus became infectious in primate cells when it was altered to contain the RBD of human SARS-CoV (2).

Bats are thought to have initially infected one or more species of small mammals, such as the palm civet (6, 13, 20, 37). One theory is that this intermediate host provided a selective environment that drove the coronavirus RBD to acquire point mutations that made it compatible with the human ortholog of its cell-surface receptor, angiotensin-converting enzyme 2 (ACE2) (19, 21, 30, 31). However, one key observation has driven the field to favor alternate, more complex theories of emergence. The observation is that while SARS-CoV and closely related viruses from the civet can use ACE2 as a receptor, no bat coronavirus has been shown to use bat, human, or any other orthologs of ACE2 (2, 27). Further,

Received 13 February 2012 Accepted 13 March 2012
Published ahead of print 21 March 2012
Address correspondence to Sara L. Sawyer, saras@mail.utexas.edu.
Supplemental material for this article may be found at http://jvi.asm.org/.

doi:10.1128/JVI.00311-12

TABLE 1 Positive selection of bat *ACE2* codons 1 to 358

ω_o and codon model[a]	Model comparison[b]						dN/dS value (% of codons)[c]	Residues under positive selection[d]
	M1a vs M2a		M7 vs M8		M8a vs M8			
	2ΔlnL	*P* value	2ΔlnL	*P* value	2ΔlnL	*P* value		
0.4, f61	52.7	$P < 0.0001$	56.5	$P < 0.0001$	52.8	$P < 0.0001$	4.3 (11)	Q24**, T27*, K31*, H34*, M82*, L91*, T92, N159*, V212, D213*, D216*, E231*, S280, V298, A301, E329
0.4, f3 × 4	56.3	$P < 0.0001$	56.4	$P < 0.0001$	56.1	$P < 0.0001$	4.3 (11)	Q24**, T27*, K31*, H34*, M82*, L91**, T92, N159*, V212*, D213*, D216*, E231*, S280, V298*, A301, E329
1.6, f61	52.7	$P < 0.0001$	56.3	$P < 0.0001$	52.8	$P < 0.0001$	4.3 (11)	Q24**, T27*, K31*, H34*, M82*, L91*, T92, N159*, V212, D213*, D216*, E231*, S280, V298, A301, E329
1.6, f3 × 4	56.3	$P < 0.0001$	56.4	$P < 0.0001$	56.1	$P < 0.0001$	4.3 (11)	Q24**, T27*, K31*, H34*, M82*, L91**, T92, N159*, V212*, D213*, D216*, E231*, S280, V298*, A301, E329

[a] Initial seed value for ω (dN/dS) and model of codon frequency (f61 or f3 × 4).
[b] Twice the difference in the natural logs of the likelihoods (2ΔlnL) of the two models being compared. This value is used in a likelihood ratio test along with the degrees of freedom. In all cases (M1a versus M2a, M7 versus M8, and M8a versus M8), a model that allows positive selection is compared to a null model. The *P* value indicates the confidence with which the null model can be rejected.
[c] dN/dS value of the class of codons evolving under positive selection in M8 and the percentage of codons falling in that class.
[d] Residues corresponding to codons assigned to the class with a dN/dS ratio of >1 in M8 ($P > 0.90$ by naive empirical Bayes [NEB]). Coordinates correspond to the human protein, although the human sequence was not used in this analysis. Bat numerical coordinates are identical with the exception of three species with single codon insertions or deletions (see alignment in Fig. S1 in the supplemental material). *, $P > 0.95$; **, $P > 0.99$. Three additional codons were identified in the analysis of the full-length gene (see Table S2 in the supplemental material).

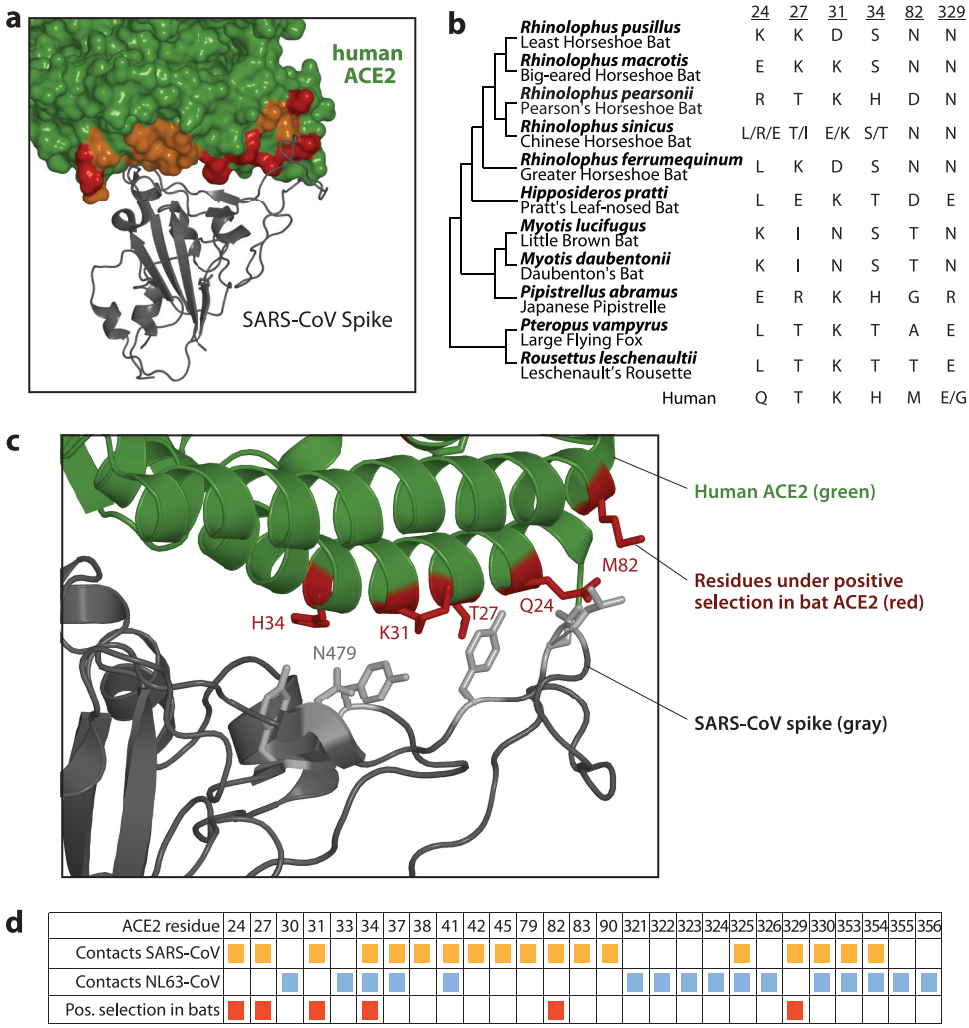

FIG 1 Residues under positive selection in bat ACE2 correspond to human ACE2 residues that interact with the SARS-CoV spike. (a) Six residues under positive selection (red) in bat ACE2 map to the SARS-CoV-binding surface (orange and red) of human ACE2 (green) and are in direct contact with the SARS-CoV spike (gray) in a cocrystal structure (PDB 2AJF) (17). (b) Bat species used in the *ACE2* analysis and the amino acids encoded at the six residue positions that directly contact the SARS-CoV spike and are evolving under positive selection. Bat polymorphisms have been reported at some of these positions (11), and a human polymorphism is found at one of them. (c) Detailed view of the side chains of five of these residues under positive selection (red) in ACE2 (green), along with the side chains of cognate contacts in the SARS-CoV spike (light gray). (d) Cocrystal structures have been solved for human ACE2 in complex with the spike proteins of both SARS-CoV (17) and NL63-CoV (39). ACE2 residues that mediate contact with each virus are indicated. Residues under positive selection in bat ACE2 are indicated in red.

sequence-based studies of the coronaviruses that have been found in bats suggest that their RBDs contain deletions spanning key residues required for mediating contact with ACE2 (5, 15, 18, 20). These observations necessitated alternate models of SARS-CoV emergence, and the currently favored model is one in which a bat coronavirus recombined with the coronavirus of a second, unknown species to create a novel hybrid virus that can use ACE2 (20). Discriminating between these two alternate models of viral emergence (ACE2 usage preexisted in the bat reservoir versus ACE2 usage was acquired outside this reservoir) is important to our understanding of the evolutionary events that generated SARS-CoV. We tested these two models by looking at the evolution of the ACE2 receptor in bats.

Over long periods of time, coevolutionary dynamics can develop between viruses and their hosts (24). For example, host populations will experience natural selection for receptor mutations that reduce virus interaction affinity, and viruses will, in turn, be selected for mutations that increase affinity with new receptor variants. This back-and-forth selection will result in the rapid evolution of both the host receptor and the virus surface protein. The protein evolutionary rate can be analyzed by studying the rates of accumulation of nonsynonymous (dN; changing the encoded amino acid) and synonymous (dS; silent) mutations in the underlying gene (24, 41). Most genes retain far fewer nonsynonymous mutations than synonymous mutations ($dN/dS \ll 1$) because protein-altering mutations tend to be deleterious (24). However, signatures of recurrent positive selection ($dN/dS > 1$) have been shown to accumulate in gene regions corresponding to the physical interaction interface between virus and host proteins, and specifically in codons corresponding to key residues that modulate these interactions (4, 7, 22, 23, 29). Starting with a data set of partial *ACE2* sequences from 11 bat species (codons 1 to 358, containing the SARS-CoV interaction domain of human ACE2) (see Table S1 in the supplemental material) or full-length ACE2

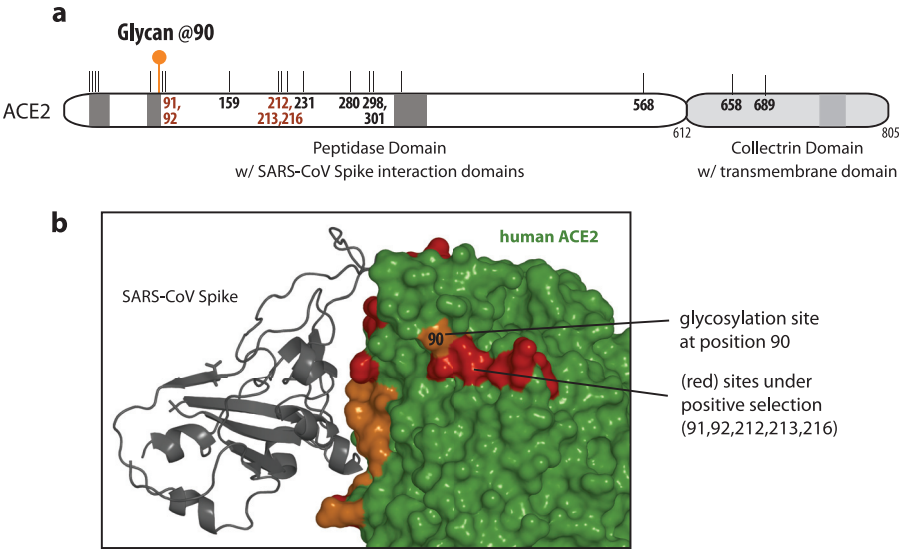

FIG 2 Positive selection of residues at the base of a key ACE2 glycan. (a) A linear schematic of the ACE2 protein is shown. Regions of the protein that interact with the SARS-CoV spike are indicated in dark gray (17). Residue positions found to be under positive selection in bats are shown with black tick marks. Six of these fall in the known surface of interaction with the SARS-CoV spike, and 13 more are indicated with numbers. Of these, five (in red type) are positioned at the base of a key glycan on the receptor that is located at position 90. (b) A rotated view of the structure shown in Fig. 1a, with the main SARS-CoV-binding surface now at the left. The glycosylated asparagine at position 90 is shown in orange, with five residues under positive selection sitting in a ridge adjacent to it (red).

sequences available for 8 of these species, DNA alignments were fit to different models of codon evolution using the codeml program in PAML (40). Some of these models allow certain codons to evolve under positive selection (M2a and M8), while others do not allow positive selection (M1a, M7, and M8a). We found that models of positive selection are highly supported ($P < 0.0001$) in both of these data sets (Table 1; see also Table S2 in the supplemental material). In total, 19 codons were assigned a dN/dS ratio greater than one with high posterior probability, with the partial gene analysis identifying more of these codons because of deeper species representation (Table 1; see also Table S2 in the supplemental material). These 19 codons in bat *ACE2* have experienced recurrent selection for mutations that replace the encoded amino acid. For this reason, these positions are highly variable at the protein level (see Fig. S1 in the supplemental material).

Structures have been solved for human ACE2 (36) and for human ACE2 in complex with the SARS-CoV spike protein (17). Of the 19 *ACE2* codons under positive selection in bats, 17 correlate to residues included in these structures. All 17 of these are surface-exposed residues in human ACE2. Six of these correlate to residues (Q24, T27, K31, H34, M82, and E329) (colored red in Fig. 1a) that make direct contact with the SARS-CoV spike protein (gray structure in Fig. 1a). These six residues are highly variable between and within bat species (Fig. 1b). Five of these residues (colored red in Fig. 1c) comprise a single ridge that intimately contacts the virus spike (gray). Two of the residues in this ridge (K31 and H34) mediate interaction with N479 in the SARS-CoV RBD (17, 20), a key position in the virus that acquired critical mutations during emergence (16, 20, 21, 26, 30). Species-specific differences at four residues in this ridge (residues 27, 31, 34, and 82) are known to contribute to species specificity of receptor usage by SARS-CoV (11, 17). These evolutionary signatures indicate that bats have been coevolving with something that is driving rapid evolution at this ACE2 interface. The footprints left by this interaction track remarkably well with the residues that interact with SARS-CoV.

Additional lines of evidence suggest that the virus driving this evolutionary signature in bat *ACE2* is very similar to SARS-CoV. First, NL63-CoV is another human coronavirus that interacts with the same surface of the ACE2 receptor (8, 9, 38, 39). However, the residues under positive selection in bats track specifically with SARS-CoV-interacting residues rather than with residues shown to mediate interactions with NL63-CoV (Fig. 1d). Second, we noticed that some positions under positive selection in bat ACE2 (numbered tick marks in Fig. 2a) do not correlate to the SARS-CoV-binding surface. However, five of these cluster around a key glycosylation site at position 90 of human ACE2 (Fig. 2b). Although it sits well outside the central SARS-CoV-binding surface (shown at left), this glycan has been shown to alter SARS-CoV binding (21). Position 90 is conserved as an asparagine in many bat species (see Fig. S1 in the supplemental material), and the attached glycan (not shown) faces the virus RBD (gray structure in Fig. 2b) (17). The residues sitting at its base are perhaps experiencing positive selection for amino acid replacements that alter the spatial orientation of this glycan moiety, a process which would constitute a novel genetic mechanism for host adaptation. Because the evolutionary signatures of positive selection recorded in bat ACE2 have accumulated at critical residues in human ACE2 that are known to govern binding by the SARS-CoV spike, we conclude that a virus very similar to SARS-CoV must have left this evolutionary footprint on ACE2 in bats.

These results are consistent with a model in which an ACE2-utilizing bat coronavirus infected civets and/or other intermediate hosts or possibly even transmitted directly to humans. This virus could have preexisted in bats or could have been a newly created virus resulting from recombination between two bat coronaviruses. The data do not support the less parsimonious model that ACE2 utilization was acquired after transmission of a bat coronavirus to another species. Others have also concluded that phylogenetic incongruencies within coronavirus genomes (28, 33, 34) do not necessarily support a model of interhost virus recombination during the emergence of SARS-CoV but may instead simply reflect differences in evolutionary

rates between different coronavirus genes (10). The idea that bats have been coevolving with SARS-CoV-like viruses over long periods of time is supported by the high SARS-CoV antibody prevalence found in bat populations of multiple species isolated from different geographic regions in China (18). This evolutionary analysis of *ACE2* sheds light on the history of emergence of this zoonotic virus from bat reservoirs. Similar insight was recently gained into the emergence of canine parvovirus by analyzing the evolution of its receptor, TfR, in carnivore species from which it arose (12). Likewise, based on evolutionary patterns in the gene encoding the Duffy antigen receptor for chemokines (DARC), we recently proposed that simian primates are an ancient reservoir for malaria-causing *Plasmodium* (3). These are the first examples demonstrating that evolutionary studies of cellular receptors may be broadly useful in understanding disease emergence.

ACKNOWLEDGMENTS

We thank Dianne Lou, Nicholas Meyerson, and Paul Rowley for critical reading of the manuscript.

This work was supported by grants 003658-0250-2009 from the Norman Hackerman Advanced Research Program and R01-GM-093086 from the National Institutes of Health (to S.L.S.) and U54 AI057159 from the New England Regional Center for Excellence/Biodefense and Emerging Infectious Disease and RR000168 from the New England Primate Research Center (to M.F.). A.D. is supported by a postdoctoral fellowship (120612-PF-11-045-01-DMC) from the American Cancer Society. S.L.S. holds a Career Award in the Biomedical Sciences from the Burroughs Wellcome Fund and is an Alfred P. Sloan Research Fellow in Computational and Evolutionary Molecular Biology.

REFERENCES

1. **Baranowski E, Ruiz-Jarabo CM, Domingo E.** 2001. Evolution of cell recognition by viruses. Science **292**:1102–1105.
2. **Becker MM, et al.** 2008. Synthetic recombinant bat SARS-like coronavirus is infectious in cultured cells and in mice. Proc. Natl. Acad. Sci. U. S. A. **105**:19944–19949.
3. **Demogines A, Truong KA, Sawyer SL.** 2012. Species-specific features of DARC, the primate receptor for *Plasmodium vivax* and *Plasmodium knowlesi*. Mol. Biol. Evol. **29**:445–449.
4. **Elde NC, Child SJ, Geballe AP, Malik HS.** 2009. Protein kinase R reveals an evolutionary model for defeating viral mimicry. Nature **457**:485–489.
5. **Graham RL, Baric RS.** 2010. Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. J. Virol. **84**: 3134–3146.
6. **Guan Y, et al.** 2003. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. Science **302**:276–278.
7. **Gupta RK, et al.** 2009. Mutation of a single residue renders human tetherin resistant to HIV-1 Vpu-mediated depletion. PLoS Pathog. **5**:e1000443.
8. **Hofmann H, et al.** 2006. Highly conserved regions within the spike proteins of human coronaviruses 229E and NL63 determine recognition of their respective cellular receptors. J. Virol. **80**:8639–8652.
9. **Hofmann H, et al.** 2005. Human coronavirus NL63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry. Proc. Natl. Acad. Sci. U. S. A. **102**:7988–7993.
10. **Holmes EC, Rambaut A.** 2004. Viral evolution and the emergence of SARS coronavirus. Philos. Trans. R. Soc. Lond. B Biol. Sci. **359**:1059–1065.
11. **Hou Y, et al.** 2010. Angiotensin-converting enzyme 2 (ACE2) proteins of different bat species confer variable susceptibility to SARS-CoV entry. Arch. Virol. **155**:1563–1569.
12. **Kaelber JT, et al.** Evolutionary reconstructions of the transferrin receptor of Caniforms supports canine parvovirus being a re-emerged and not a novel pathogen in dogs. PLoS Pathog., in press.
13. **Kan B, et al.** 2005. Molecular evolution analysis and geographic investigation of severe acute respiratory syndrome coronavirus-like virus in palm civets at an animal market and on farms. J. Virol. **79**:11892–11900.
14. **Kuo L, Godeke GJ, Raamsman MJ, Masters PS, Rottier PJ.** 2000. Retargeting of coronavirus by substitution of the spike glycoprotein ectodomain: crossing the host cell species barrier. J. Virol. **74**:1393–1406.
15. **Lau SKP, et al.** 2005. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. Proc. Natl. Acad. Sci. U. S. A. **102**:14040–14045.
16. **Li F.** 2008. Structural analysis of major species barriers between humans and palm civets for severe acute respiratory syndrome coronavirus infections. J. Virol. **82**:6984–6991.
17. **Li F, Li W, Farzan M, Harrison S.** 2005. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. Science **309**: 1864–1868.
18. **Li W, et al.** 2005. Bats are natural reservoirs of SARS-like coronaviruses. Science **310**:676–679.
19. **Li W, et al.** 2003. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature **426**:450–454.
20. **Li W, et al.** 2006. Animal origins of the severe acute respiratory syndrome coronavirus: insight from ACE2-S-protein interactions. J. Virol. **80**:4211–4219.
21. **Li W, et al.** 2005. Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. EMBO J. **24**:1634–1643.
22. **Lim ES, et al.** 2012. The ability of primate lentiviruses to degrade the monocyte restriction factor SAMHD1 preceded the birth of the viral accessory protein Vpx. Cell Host Microbe **11**:194–204.
23. **Lim ES, Malik HS, Emerman M.** 2010. Ancient adaptive evolution of tetherin shaped the functions of Vpu and Nef in human immunodeficiency virus and primate lentiviruses. J. Virol. **84**:7124–7134.
24. **Meyerson NR, Sawyer SL.** 2011. Two-stepping through time: mammals and viruses. Trends Microbiol. **19**:286–294.
25. **Parrish CR, et al.** 2008. Cross-species virus transmission and the emergence of new epidemic diseases. Microbiol. Mol. Biol. Rev. **72**:457–470.
26. **Qu X-X, et al.** 2005. Identification of two critical amino acid residues of the severe acute respiratory syndrome coronavirus spike protein for its variation in zoonotic tropism transition via a double substitution strategy. J. Biol. Chem. **280**:29588–29595.
27. **Ren W, et al.** 2008. Difference in receptor usage between severe acute respiratory syndrome (SARS) coronavirus and SARS-like coronavirus of bat origin. J. Virol. **82**:1899–1907.
28. **Rest JS, Mindell DP.** 2003. SARS associated coronavirus has a recombinant polymerase and coronaviruses have a history of host-shifting. Infect. Genet. Evol. **3**:219–225.
29. **Sawyer SL, Wu LI, Emerman M, Malik HS.** 2005. Positive selection of primate TRIM5alpha identifies a critical species-specific retroviral restriction domain. Proc. Natl. Acad. Sci. U. S. A. **102**:2832–2837.
30. **Sheahan T, et al.** 2008. Mechanisms of zoonotic severe acute respiratory syndrome coronavirus host range expansion in human airway epithelium. J. Virol. **82**:2274–2285.
31. **Sheahan T, Rockx B, Donaldson E, Corti D, Baric R.** 2008. Pathways of cross-species transmission of synthetically reconstructed zoonotic severe acute respiratory syndrome coronavirus. J. Virol. **82**:8721–8732.
32. **Shi Z, Hu Z.** 2008. A review of studies on animal reservoirs of the SARS coronavirus. Virus Res. **133**:74–87.
33. **Stanhope MJ, Brown JR, Amrine-Madsen H.** 2004. Evidence from the evolutionary analysis of nucleotide sequences for a recombinant history of SARS-CoV. Infect. Genet. Evol. **4**:15–19.
34. **Stavrinides J, Guttman DS.** 2004. Mosaic evolution of the severe acute respiratory syndrome coronavirus. J. Virol. **78**:76–82.
35. **Stehle T, Casasnovas JM.** 2009. Specificity switching in virus-receptor complexes. Curr. Opin. Struct. Biol. **19**:181–188.
36. **Towler P.** 2004. ACE2 X-ray structures reveal a large hinge-bending motion important for inhibitor binding and catalysis. J. Biol. Chem. **279**: 17996–18007.
37. **Wu D, et al.** 2005. Civets are equally susceptible to experimental infection by two different severe acute respiratory syndrome coronavirus isolates. J. Virol. **79**:2620–2625.
38. **Wu K, et al.** 2011. A virus-binding hot spot on human angiotensin-converting enzyme 2 is critical for binding of two different coronaviruses. J. Virol. **85**:5331–5337.
39. **Wu K, Li W, Peng G, Li F.** 2009. Crystal structure of NL63 respiratory coronavirus receptor-binding domain complexed with its human receptor. Proc. Natl. Acad. Sci. U. S. A. **106**:19970–19974.
40. **Yang Z.** 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. **13**:555–556.
41. **Yang Z, Bielawski J.** 2000. Statistical methods for detecting molecular adaptation. Trends Ecol. Evol. **15**:496–503.