JOURNAL OF VIROLOGY, Apr. 2007, p. 4012–4020 Vol. 81, No. 8
0022-538X/07/$08.00+0 doi:10.1128/JVI.02605-06
Copyright © 2007, American Society for Microbiology. All Rights Reserved.

Evolutionary Insights into the Ecology of Coronaviruses[▽]

D. Vijaykrishna,† G. J. D. Smith,† J. X. Zhang, J. S. M. Peiris, H. Chen, and Y. Guan*

State Key Laboratory of Emerging Infectious Diseases, Department of Microbiology, Li Ka Shing Faculty of Medicine, The University of Hong Kong, 21 Sassoon Road, Pokfulam, Hong Kong SAR, China

Received 25 November 2006/Accepted 23 January 2007

Although many novel members of the *Coronaviridae* have recently been recognized in different species, the ecology of coronaviruses has not been established. Our study indicates that bats harbor a much wider diversity of coronaviruses than any other animal species. Dating of different coronavirus lineages suggests that bat coronaviruses are older than those recognized in other animals and that the human severe acute respiratory syndrome (SARS) coronavirus was directly derived from viruses from wild animals in wet markets of southern China. Furthermore, the most closely related bat and SARS coronaviruses diverged in 1986, an estimated divergence time of 17 years prior to the outbreak, suggesting that there may have been transmission via an unknown intermediate host. Analysis of lineage-specific selection pressure also indicated that only SARS coronaviruses in civets and humans were under significant positive selection, also demonstrating a recent interspecies transmission. Analysis of population dynamics revealed that coronavirus populations in bats have constant population growth, while viruses from all other hosts show epidemic-like increases in population. These results indicate that diverse coronaviruses are endemic in different bat species, with repeated introductions to other animals and occasional establishment in other species. Our findings suggest that bats are likely the natural hosts for all presently known coronavirus lineages and that all coronaviruses recognized in other species were derived from viruses residing in bats. Further surveillance of bat and other animal populations is needed to fully describe the ecology and evolution of this virus family.

The majority of human emerging infectious diseases over the past few decades, including AIDS, Ebola fever, avian influenza, and severe acute respiratory syndrome (SARS), have resulted from the interspecies transmission of zoonotic RNA viruses (5, 16, 21, 33). The identification of closely related viruses in animal hosts is a prerequisite for establishing the ecology of viral emergence along with reconstruction of evolutionary pathways; however, complex ecosystems make this difficult (41). This information is critical for the control of these diseases at their sources and for intervention in possible outbreaks.

Coronaviruses have the largest RNA viral genomes, ranging from 26 to 32 kilobases in length (20). They have been known to cause upper and lower respiratory diseases, gastroenteritis, and central nervous system infection in a number of avian and mammalian hosts, including humans (43). Coronaviruses are placed within the family *Coronaviridae*, a member of the order *Nidovirales*, and are classified into three groups based on genetic and antigenic relationships (9, 20).

Uncertainty exists in the classification of SARS coronavirus (SARS-CoV). Initially, SARS-CoV was proposed as a novel group 4 coronavirus (24). Comparative genetic analyses of individual gene segments showed that SARS-CoV was consistently a sister to group 2 CoVs, suggesting that SARS-CoV is an early divergence from other established group 2 viruses. Therefore, it was suggested to be a group 2b CoV (30), and in the current International Committee on Taxonomy of Viruses classification system, SARS-CoV is classified as a group 2 coronavirus. However, phylogenetic and pairwise comparisons of SARS-CoV with other coronaviruses, including those from bats (BtCoVs), showed that they share low homology to both group 2 and group 3 coronaviruses and therefore may belong to a new group (putative group 4) (35). As such, we have classified SARS-CoV in group 4 in this study. The recognition of SARS-like coronaviruses (SARS-like CoVs) in bats suggested that bats might be the most likely natural reservoirs of SARS-CoV (22). Comparative genetic analyses of coronaviruses isolated from bats in China also showed that bat species harbor highly diverse coronaviruses, including a novel lineage (putative group 5) that is exclusive to bats (35). This information suggests that the diversity of coronaviruses is greater than previously recognized and that it is necessary to explore the role of bats in the ecosystem of coronaviruses (35, 45).

Since the onset of the SARS epidemic in early 2003, intense scientific effort has focused on identifying the zoonotic source of SARS-CoV and its transmission route to humans (11, 31, 38, 46). Surveys of domestic and wild animals in southern China revealed that civet cats, raccoon dogs, and ferret badgers from wet markets were vectors for SARS-CoV outbreaks in 2002 and 2003 (11, 31). This was also supported by serological studies that indicated that wildlife traders, animal slaughterers, and restaurant workers with occupational exposure to wild animals, particularly palm civets (*Paguma larvata*), had a higher SARS-CoV antibody prevalence than did market and community controls (11, 46). However, farm populations of palm civets were negative for SARS-CoV, while market populations tested positive, suggesting that market animals were most probably intermediate hosts for SARS-CoV and were possibly infected

* Corresponding author. Mailing address: State Key Laboratory of Emerging Infectious Diseases, Department of Microbiology, The University of Hong Kong, Li Ka Shing Faculty of Medicine Building, 21 Sassoon Road, Pokfulam, Hong Kong SAR, China. Phone: (852) 28199830. Fax: (852) 28171958. E-mail: yguan@hku.hk.

† D. Vijaykrishna and G. J. D. Smith contributed equally to the manuscript.

▽ Published ahead of print on 31 January 2007.

Downloaded from http://jvi.asm.org/ on March 11, 2015 by ST ANDREWS UNIV

from another source in the markets (11, 31, 38). While SARS-like CoVs have been identified from four bat species (genus *Rhinolophus*) in both northern and southern China, phylogenetic relationships and low genetic similarities to human and civet SARS-CoVs indicate that the natural reservoir of these viruses has still not been determined.

In this study, we estimated and compared the divergence times and population dynamics of all coronavirus groups, including new data from wild animals, to better understand the ecology and evolutionary pathways of these viruses. Our results show that BtCoVs are older than coronaviruses from any other animals, while analysis of population dynamics revealed that coronavirus populations in bats have constant population growth and that viruses from all other hosts show epidemic-like increases in population. These results indicate that diverse coronaviruses are endemic in different bat species, with repeated introductions to other animals and occasional establishment in other species. In addition, divergence dates for SARS-CoV and SARS-like BtCoVs indicated that the SARS-CoV precursor may have circulated in an unidentified host before the 2003 epidemic. Lineage-specific positive selection analysis also suggests that the SARS-CoV precursor has not been identified. Our findings suggest that bats are the natural hosts for all coronavirus lineages and that all coronaviruses recognized in other species were derived from viruses residing in bats.

MATERIALS AND METHODS

Sampling and virus detection. During 2003, wild animals were sampled at the live animal markets in Shenzhen and Guangzhou, Guangdong Province, China. Rectal swabs were taken, placed in transport medium, kept in liquid nitrogen for transportation to the laboratory, and then stored at −80°C. Viral RNA extraction, PCR, and sequencing were conducted as previously described (35).

Phylogenetic analysis. For the reconstruction of evolutionary pathways of coronaviruses isolated from bats, coronaviruses representing the three traditional viral groups, SARS-CoVs from humans and civets (group 4), and coronaviruses from bats (group 5) were included in analyses. Two previously unpublished coronavirus sequences from a Chinese ferret badger and raccoon dog (CFBCoV/DM95/2003 and RDCoV/GM43/2003, respectively) were also included in the analysis. The helicase (HEL) domain of the replicase gene, the spike protein (S) gene, and the nucleocapsid protein (N) gene were chosen for analysis based on function and the availability of BtCoV sequences. Nucleotide sequence alignments were carried out using TransAlign (4) with ClustalX (37) to conserve codon positions and were manually optimized wherever necessary using Se-Al, version 2.0 (http://evolve.zoo.ox.ac.uk/). All ambiguously aligned regions were removed prior to phylogenetic analysis. Phylogenetic trees were built using neighbor joining in PAUP*, version 4.0b (34), with the appropriate model of sequence evolution selected by MrModeltest, version 2.2 (J. A. A. Nylander, Evolutionary Biology Center, University of Uppsala, Uppsala, Sweden). Gaps were treated as missing data in all analyses. The same model of evolution was used to calculate Bayesian posterior probability in MrBayes, version 3.1 (15).

Estimation of divergence dates and population dynamics. The divergence dates for coronavirus lineages were estimated based on an alignment of complete HEL domain sequences, using the uncorrelated relaxed clock model in BEAST, version 1.4 (http://evolve.zoo.ox.ac.uk/beast/). Under this model, the rates were allowed to vary at each branch drawn independently from an exponential distribution (8). Sampling dates for each isolate were used as calibration points, and constant population coalescent priors were assumed for all data sets. Depending on the data set, Markov chain Monte Carlo (MCMC) samples chains were run for 50 to 150 million generations, sampling every 1,000 generations under the HKY nucleotide substitution model and allowing γ-rate heterogeneity for all data sets. The convergence of MCMC chains was confirmed for each data set by using the program Tracer, version 1.3 (http://evolve.zoo.ox.ac.uk/tracer/). The times of divergence were estimated utilizing priors obtained from earlier runs, with a discarded burn-in of approximately 10%. The highest posterior density (HPD) was calculated for each of these parameters.

Coronavirus population dynamics were evaluated for eight independent data sets, using the constant, exponential, and logistic population models in BEAST, version 1.4. The posterior likelihoods of the three models were compared using the Akaike information criteria, with one parameter distinguishing the constant and exponential models, while two parameters distinguished the constant and logistic models (28). MCMC chains were run as previously described, and substitution rates and growth rates with HPDs were calculated.

Detection of recombination and lineage-specific selection. Since the presence of recombination among sequences can obscure the detection of positive selection (18), we first tested for recombination in our data set (see below) by identifying breakpoints, using a genetic algorithm for recombination detection (GARD) (17). Identified breakpoints were then assessed by testing the congruence of neighbor-joining trees generated by GARD for each fragment, using the Shimodaira-Hasegawa test in PAUP* (34). Analysis of lineage-specific selection pressure in the S genes of SARS-CoVs was conducted using the Genetic Algorithm (GA) in the program package HYPHY (19). We analyzed an alignment of the complete S genes (3,783 nucleotides and 1,261 codons) of six BtCoVs and three civet and three human SARS-CoVs and specified a neighbor-joining input tree (GTR+I model) generated in PAUP* as described above. The GA in HYPHY assigns four classes of nonsynonymous and synonymous nucleotide substitution rate ratios ($\omega = d_N/d_S$) to each lineage in search of the model of lineage-specific evolution that best fits the data (18). The Akaike information criteria were used to select the best-fitting model of lineage-specific evolution. The probability ($\geq$95%) of ω being >1 along a specific lineage was calculated from the averaged model probability of all models rather than by inference from the single best-fitting model (18). This approach does not require any a priori hypothesis of lineage-specific evolution.

Nucleotide sequence accession numbers. The sequences reported in this paper have been deposited in GenBank under accession numbers EF192155 to EF192160 and EF434376 to EF434381.

RESULTS

Phylogenetic relationships of coronaviruses. Evolutionary relationships of BtCoVs were reconstructed by neighbor-joining and Bayesian analyses of BtCoVs and representatives of all other coronavirus groups. We analyzed the complete HEL domain, generally known by the name nsp 13, of the multidomain replicase gene, the structural spike protein (S) gene, and the nucleocapsid protein (N) gene. Due to the high divergence in the S gene, 1,579 ambiguously aligned characters were removed, resulting in a final alignment size of 2,204 nucleotides. The final sizes of the HEL domain and N gene alignments were 1,797 and 1,812 nucleotides, respectively. The multiple sequence alignments for the HEL domain, S gene, and N gene that were used for analyses are available in Treebase (http://www.treebase.org/treebase/).

Analyses of all three genes showed that all five coronavirus groups are monophyletic, with high statistical support, and that BtCoVs are found only in groups 1, 4, and 5 (Fig. 1). However, support was lesser for some of the deep nodes, particularly in the S gene tree. Gene phylogenies also showed that interrelationships between groups were not consistent between different genes. In the HEL domain and N gene trees, group 4 SARS-CoVs and SARS-like BtCoVs form a sister clade to group 5 BtCoVs, while group 2 coronaviruses are, in turn, the sister clade to both groups 4 and 5 (Fig. 1A and C). However, these relationships change in the S tree, whereby groups 2 and 5 cluster together and group 4 coronaviruses are the sister clade to both groups (Fig. 1B).

The phylogeny of the HEL domain, for which the most sequence data from bats are available, showed that six BtCoVs (BtCoV/Rp3, BtCoV/HKU3, BtCoV/279/05, BtCoV/273/05, BtCoV/Rm1, and BtCoV/Rf1) belong to group 4 and form a sister clade to SARS-CoVs isolated from humans and civet cats

Downloaded from http://jvi.asm.org/ on March 11, 2015 by ST ANDREWS UNIV

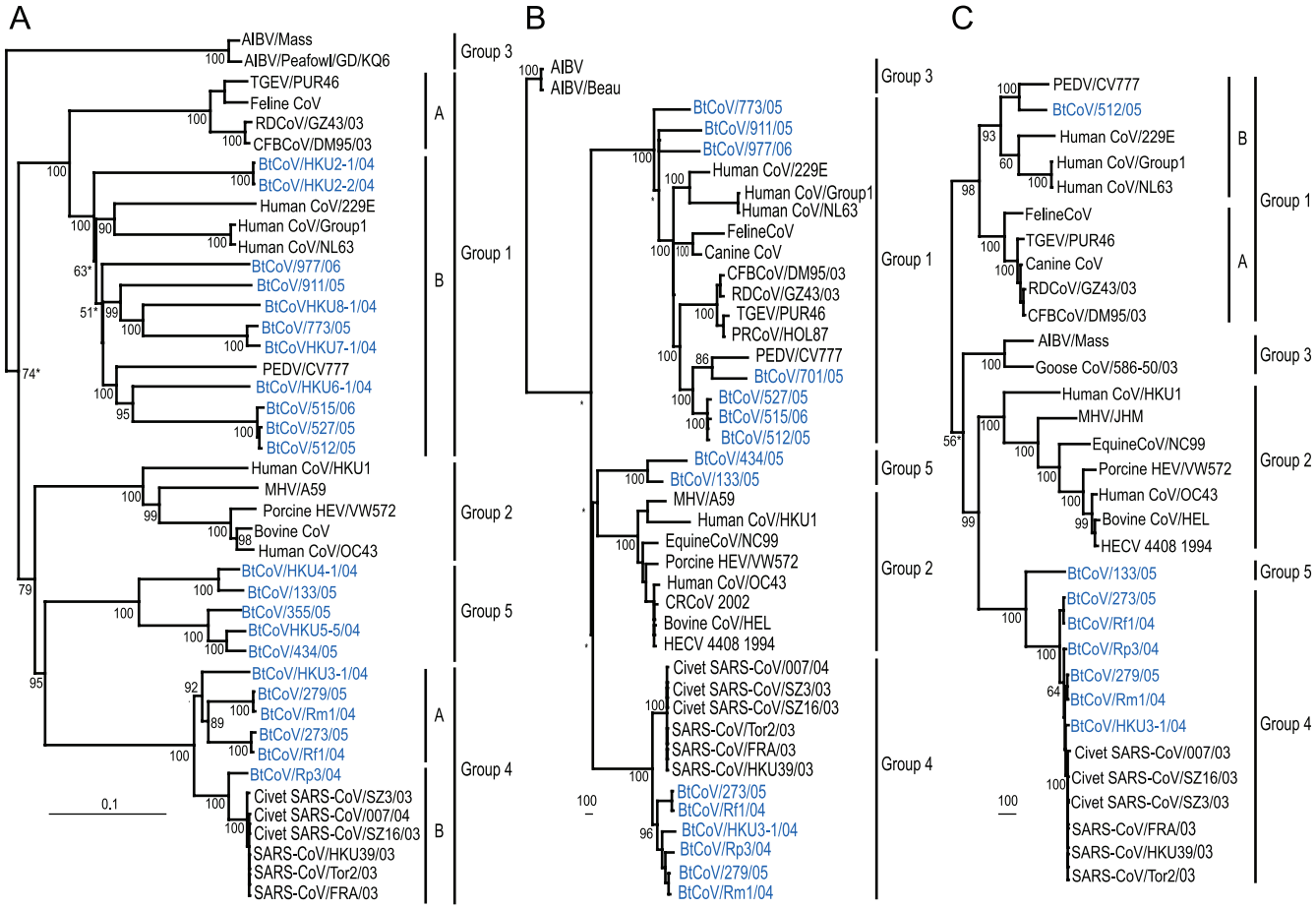

FIG. 1. Phylogenetic relationships of helicase (A), spike (B), and nucleocapsid (C) genes based on alignments of 1,797, 2,204, and 1,812 nucleotides, respectively. Breda virus (AY427798) was used to root the trees. Numbers at branch nodes indicate neighbor-joining bootstrap values of ≥50%. Nodes with Bayesian posterior probabilities of ≤95% are indicated with asterisks. Bar, 0.1 nucleotide substitution per site (A) or 100 nucleotide substitutions per site (B and C).

between 2003 and 2004 (Fig. 1A). Therefore, analysis of available sequence data suggests that bats are the predominant carriers of SARS and SARS-like CoVs. A further five BtCoVs isolated from southern China form group 5 and are only distantly related to other coronavirus groups, as reported earlier (35).

In conjunction with previous studies, phylogenetic analysis of the HEL domain and the N gene showed that group 1 coronaviruses form two separate subgroups (Fig. 1A and C) (10, 11). The first subgroup contains all group 1 BtCoVs, all common cold coronaviruses (HCoVs 229E and NL63) from humans, and porcine epidemic diarrhea virus (PEDV), indicating that bats may be the natural reservoirs for PEDV and common cold coronaviruses. In the HEL domain tree, four BtCoVs, represented by BtCoV/512/05, are most closely related to PEDV (Fig. 1A). This relationship is maintained in the N gene, but it must be noted that N gene sequence data for three of the viruses were not available. The second subgroup contains the two wild animal coronaviruses (raccoon dog CoV/GZ43/03 and Chinese ferret badger CoV/DM95/03) along with canine and feline coronaviruses and porcine transmissible gastroenteritis virus (TGEV). However, these subgroups are not apparent in the S gene tree. Three of the BtCoVs (BtCoV/773/05, BtCoV/911/05, and BtCoV/977/05) now fall in a basal position within group 1, while the remaining BtCoVs, represented by BtCoV/512/05, maintain a close relationship to PEDV (Fig. 1B). Taken together, the above results indicate that BtCoVs are precursors to previously established coronavirus lineages.

The S gene data set displayed the greatest level of divergence between coronavirus groups and may be approaching sequence saturation. Saturation occurs due to multiple nucleotide substitutions at a given nucleotide position over time, masking the true level of divergence and obscuring deeper phylogenetic relationships to the point of making them unrecoverable (1, 2). Since the HEL domain is conserved across all groups, it will most likely reflect the true coronavirus phylogeny, and while the N gene is less conserved than the HEL domain, the regions used in the analysis were also conserved between groups. As such, this may explain the inconsistency between phylogenetic relationships for the S gene and those for the other genes tested, even following removal of highly variable regions and application of an evolutionary model. Therefore, the HEL domain and, to a lesser extent, the N gene are considered the most suitable for calculation of evolutionary rates and divergence times for coronaviruses.

Downloaded from http://jvi.asm.org/ on March 11, 2015 by ST ANDREWS UNIV

TABLE 1. Sampling dates for coronaviruses

Group	Virus	Host	Isolation date (yr)	Reference or GenBank accession no.
1	BtCoV/HKU2/2004	Bats	2004	45
	BtCoV/HKU6/2004	Bats	2004	45
	BtCoV/HKU7/2004	Bats	2004	45
	BtCoV/HKU8/2004	Bats	2004	45
	BtCoV/512/2005	Bats	2005	35
	BtCoV/515/2006	Bats	2006	35
	BtCoV/527/2005	Bats	2005	35
	BtCoV/773/2005	Bats	2005	35
	BtCoV/911/2005	Bats	2005	35
	BtCoV/977/2006	Bats	2006	35
	HCoV-229E	Human	1963	12
	HCoV-NL63	Human	2003	39
	PEDV	Pig	1978	27
	TGEV	Pig	1946	7
	Feline CoV	Cat	1981	25
	RDCoV/GZ43/2003	Raccoon dog	2003	This study
	CFBCoV/DM95/2003	Chinese ferret badger	2003	This study
2	Bovine CoV	Cattle	1998	32
	HCoV-OC43	Human	1967	40
	HCoV-HKU1	Human	2003	44
	Porcine hemagglutinating encephalitis virus	Pig	1972	26
	Murine hepatitis virus	Mouse	1949	23
3	IBV Mass	Chicken	1941	6
	IBV GD/KQ6/2003	Peafowl	2003	AY641576
4	BtCoV/273/2005	Bats	2005	35
	BtCoV/279/2005	Bats	2005	35
	Bat SARS-CoV/Rm1/2004	Bats	2004	22
	Bat SARS-CoV/Rf1/2004	Bats	2004	22
	Bat SARS-CoV/Rp3/2004	Bats	2004	22
	BtCoV/HKU3/2004	Bats	2004	45
	Civet SARS-CoV/SZ3/2003	Civet	2003	11
	Civet SARS-CoV/SZ16/2003	Civet	2003	11
	Civet SARS-CoV/007/2004	Civet	2004	42
	SARS-CoV/Tor2/2003	Human	2003	14
	SARS-CoV/FRA/2003	Human	2003	36
	SARS-CoV/HK-39/2003	Human	2003	47
5	BtCoV/HKU4/2004	Bats	2004	45
	BtCoV/HKU5/2004	Bats	2004	45
	BtCoV/133/2005	Bats	2005	35
	BtCoV/355/2005	Bats	2005	35
	BtCoV/434/2005	Bats	2005	35

Estimation of divergence dates. The most recent common ancestor (MRCA) dates for major coronavirus lineages (Table 1) were calculated for the HEL domain, using the uncorrelated relaxed clock model. However, the large HPDs associated with most of these dates necessitate caution in interpreting these results. The MRCA date for coronaviruses was estimated to be 1586 (HPDs, 1102 to 1878), approximately 500 years before the present, while the major divergence between group 1 and groups 2, 4, and 5 was estimated to occur in 1647 (HPDs, 1247 to 1882) (Fig. 2 and Table 2). Among coronavirus groups, the earliest MRCA date estimates were for group 1 coronaviruses, at 1800 (HPDs, 1570 to 1927), and group 2 coronaviruses, at 1892 (HPDs, 1817 to 1945) (Fig. 2 and Table 2). The group 3 avian IBV coronavirus MRCA date was estimated to be 1925 (HPDs, 1887 to 1941), and those for groups 4 and 5 were 1961 (HPDs, 1918 to 1995) and 1955 (HPDs, 1902 to 1991), respectively (Fig. 2 and Table 2). These results suggest that group 1, which includes a high genetic diversity of BtCoVs, may have evolved much earlier than other coronavirus lineages. Within group 4, the estimated time of divergence for SARS-CoVs from civet cats and humans and the closest SARS-like BtCoV isolate (BtCoV/Rp3) was 1986 (HPDs, 1964 to 2001), a mean divergence time of 17 years before the 2003 outbreak. Furthermore, the MRCA of SARS-CoVs from humans and civet cats was estimated to exist in 1999 (HPDs, 1990 to 2003). These results indicate that the precursor of the virus responsible for the SARS outbreak may not have been determined.

Detection of recombination and lineage-specific selection. We investigated the positive selection pressure along the S gene in an attempt to determine the evolutionary stability of group 4 coronaviruses in different hosts. However, because the presence of recombination among sequences can obscure detection of positive selection, we first tested for recombination in our data set. Five recombination breakpoints were detected in the S gene data set for SARS-CoVs by using GARD (Fig. 3A) (17). However, the Shimodaira-Hasegawa test (29) showed that there was no significant evidence ($P > 0.05$) of incongruence between phylogenies of the six recombinant fragments, and no interhost recombination was observed in the phylogenies (data not shown). Therefore, recombination

Downloaded from http://jvi.asm.org/ on March 11, 2015 by ST ANDREWS UNIV

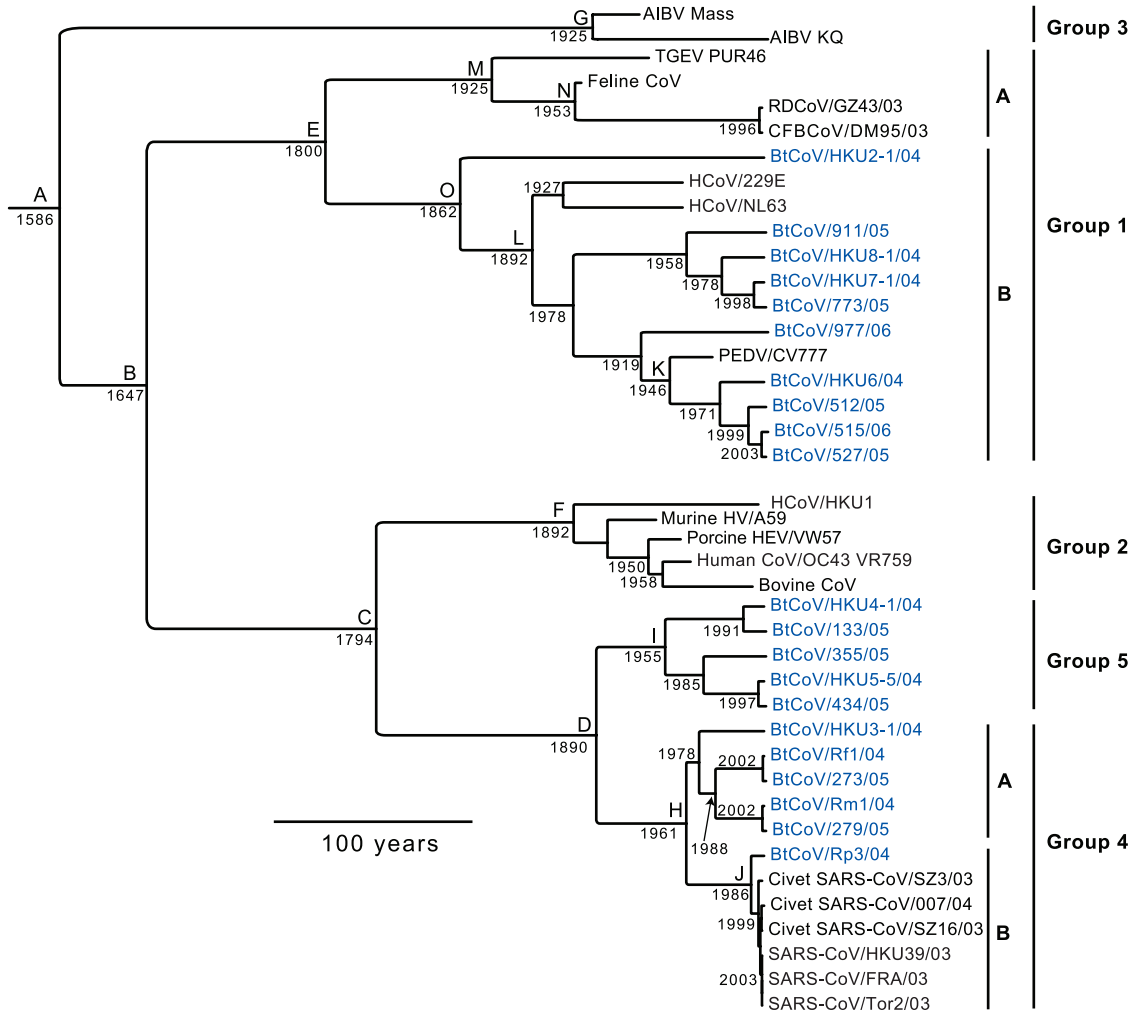

FIG. 2. Divergence dates of coronavirus lineages represented on a phylogenetic tree generated based on 1,797 nucleotides of the complete HEL domain sequence. Breda virus (AY427798) was used to root the tree. Numbers at branch nodes indicate the times of divergence, calculated using the uncorrelated exponential relaxed-clock model in BEAST v1.4. For an explanation of the letters at branch nodes, see Table 2.

TABLE 2. Estimated divergence dates for coronavirus groups

Node[a]	Mean date (95% HPDs)
A	1586 (1102–1878)
B	1647 (1247–1882)
C	1794 (1600–1920)
D	1890 (1785–1969)
E	1800 (1570–1927)
F	1892 (1823–1938)
G	1925 (1887–1941)
H	1961 (1918–1995)
I	1955 (1902–1991)
J	1986 (1964–2002)
K	1946 (1908–1969)
L	1892 (1800–1947)
M	1925 (1882–1946)
N	1953 (1902–1981)
O	1862 (1732–1942)

[a] Nodes indicate the most recent common ancestors of major coronavirus lineages, as shown in Fig. 2.

among these sequences was considered not significant enough to affect the detection of positive selection along lineages.

The averaged model probability of branches showed that only two branches were under significant positive selection (≥95% confidence interval), with one leading to civet SARS-CoVs ($\omega = 2.149$) and one leading to the three human isolates ($\omega = 2.144$) (Fig. 3B). Other branches with ω values of >1 were not under significant positive selection pressure. These findings further support the hypothesis that SARS-CoVs were recently introduced into civet and human populations from an unknown host.

Population dynamics of coronaviruses. Group 1 coronaviruses contain two distinct sublineages, A and B, which were treated independently in these analyses (Fig. 1A and C). Sublineage A (group 1A) consists of TGEV, feline coronavirus, and wild animal coronaviruses, while sublineage B (group 1B) consists of all group 1 BtCoVs, PEDV, and human common cold coronaviruses. The best-fitting population models for groups 1A and 1B were exponential growth and a constant

Downloaded from http://jvi.asm.org/ on March 11, 2015 by ST ANDREWS UNIV

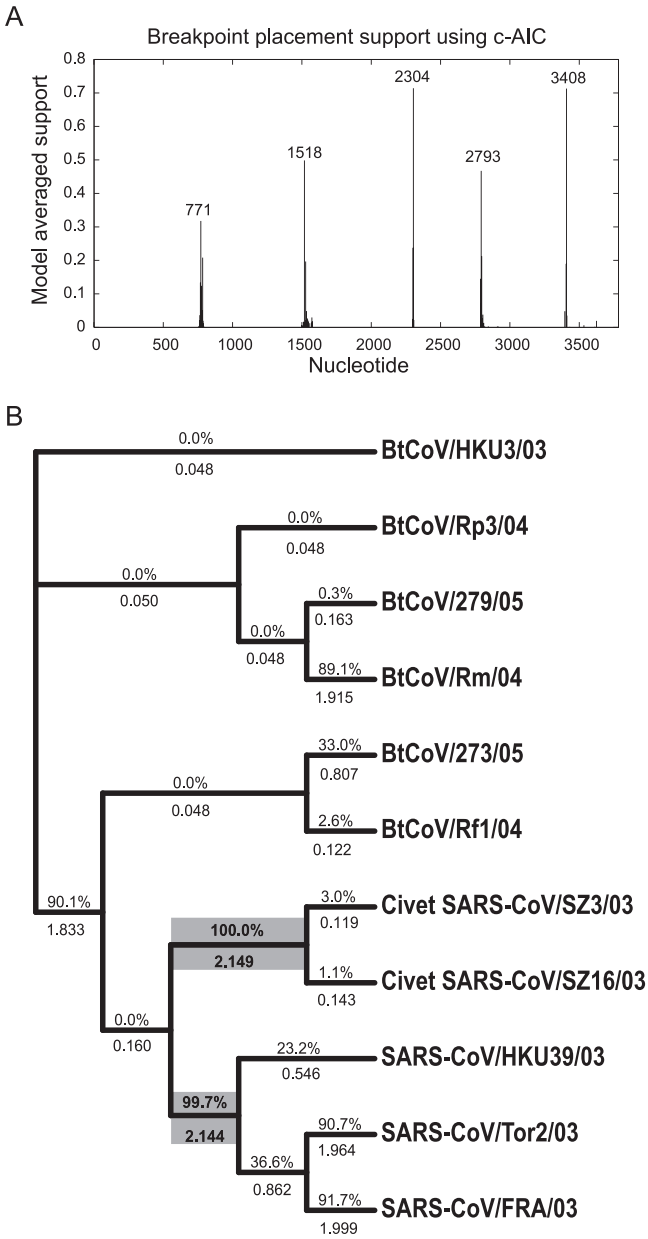

FIG. 3. Detection of recombination (A) and lineage-specific selection pressure (B) on the spike genes of SARS-CoVs. Mean ω values calculated using GA are presented above branches, with ω values of >1 shown in bold. Numbers below branches indicate an averaged model probability of >1 along specific lineages. Branches with *P* values of ≥95% are highlighted in gray. The Akaike information criterion (c-AIC) for the best-fitting model was 23,557.

population size, respectively, suggesting the recent introduction of coronaviruses into domestic and wild animals from a population endemic in bats (Table 3). Even when the group 1B bat viruses were analyzed separately (group 1B bats), the best-fitting population growth model remained a constant population size (Table 3). This likely reflects the early divergence and establishment of PEDV and human cold coronaviruses from BtCoV populations.

Two sublineages, A and B, within group 4 were also independently analyzed. Sublineage A (group 4A) contains only BtCoVs, while sublineage B (group 4B) includes civet and human SARS-CoVs along with BtCoV/Rp3 (Fig. 2). Groups 4A and 5, which consist only of BtCoVs, had a best-fitting model of constant population size, reflecting the endemicity of coronaviruses in bats despite multiple interspecies transmission events between different bat species (Table 3). The best-fitting model for group 4B, which contains BtCoV/Rp3 and SARS-CoVs, was for exponential population growth, with an extraordinarily high mean growth rate compared to those of all other groups (Table 3). Taken together, these results indicate that coronaviruses are endemic in bat populations and that after interspecies transmission to a naïve host, there was a change to epidemic-like population growth, particularly for group 4B, which contains civet and human SARS-CoVs. However, since coalescent estimates improve with increased sample size, especially for complex demographic models, it will be important to reexamine these results as more CoV data become available.

DISCUSSION

The findings of our study show that coronaviruses from bats have the most genetic diversity and are older than all coronaviruses recognized from any other animal species. We also show that coronaviruses in bats have a stable population size, while in other animals there is epidemic-like population growth. All of these attributes consolidate the concept that bats are the natural reservoir of all currently known coronavirus lineages.

Natural reservoirs for RNA viruses typically harbor the most genetic diversity in viruses among all recognized hosts while showing no symptoms of infection (13). A change from a constant virus population size to logistic or exponential population growth has been associated with the interspecies transmission of viruses from their natural reservoir to an alternate host (28). Previously, limited genetic information made it almost impossible to establish the ecology of coronaviruses. However, the accumulated information presented here allows us to better understand the ecology and evolutionary pathways of coronaviruses (Fig. 4).

Although the ancestor of all of the established coronavirus lineages has not been identified, our data indicate that the hypothetical ancestor is likely from a bat. For group 1, there have been two interspecies transmissions from bats to other animals, including humans, and most interestingly between wild animals and domestic pets. However, no precursor from bats has been identified for these group 1A viruses (Fig. 2 and 4). A similar situation exists for group 2 viruses, with interspecies transmissions between domesticated animals, mice, and humans but with no known precursor virus from bats. This is likely because bat species have only recently been studied, and it is possible that after extensive surveillance the evolutionary history of group 2 coronaviruses can be more clearly understood.

The situation with groups 4 and 5 is better resolved, with a clear evolutionary pathway from BtCoVs to SARS-CoV, albeit through an unknown intermediate host, while group 5 viruses appear to circulate among different species of bat (35). Two distantly related coronavirus lineages (groups 1A and 4A) have

Downloaded from http://jvi.asm.org/ on March 11, 2015 by ST ANDREWS UNIV

TABLE 3. Population dynamics of coronaviruses

Group	Best-fitting population model	Mean substitution rate	Substitution rate HPD intervals	Mean growth rate	Growth rate HPD intervals
1A[a]	Exponential	6.08×10^{-3}	1.3×10^{-5}, 1.4×10^{-2}	0.27	0.015, 0.45
1B	Constant	1.4×10^{-2}	2.3×10^{-6}, 4.0×10^{-2}		
1B (bats)	Constant	3.7×10^{-2}	5.5×10^{-5}, 0.1		
2	Exponential	1.8×10^{-2}	3.7×10^{-3}, 3.0×10^{-2}	0.24	0.13, 0.31
3	Exponential	5.35×10^{-3}	2.5×10^{-3}, 1.1×10^{-2}	0.311	0.17, 0.46
4A[b]	Constant	2.0×10^{-2}	8.8×10^{-3}, 3.4×10^{-2}		
4B	Exponential	1.7×10^{-2}	0.1×10^{-4}, 2.5×10^{-2}	10.754	8.02, 13.40
5	Constant	8.4×10^{-2}	2.1×10^{-2}, 0.14		

[a] Group 1A includes the sublineage that contains TGEV, feline CoV, and wild animal coronaviruses from group 1, while group 1B includes all other group 1 coronaviruses; BtCoVs in group 1B were also analyzed separately (see Fig. 2).

[b] Group 4 is subdivided, based on phylogenetic relationships of the HEL gene, into group 4A, which contains only BtCoVs, and group 4B, which includes civet and human SARS-CoVs and BtCoV/Rp3 (see Fig. 2).

been identified from wild animals, with a possible interspecies transmission to domestic animals for group 1A and a jump to humans for group 4B. The exponential population growth in both these groups shows that coronaviruses are not endemic in wild animals and that these animals have acquired the viruses recently. This highlights the importance of further surveillance in a variety of mammals, especially given the emergence of new host populations in China that has resulted from increased consumption of exotic animals in the last 20 years.

Representing more of an enigma are group 3 coronaviruses, which are found exclusively in domestic poultry and are highly divergent from other groups. Given the recent divergence date and exponential population growth, we hypothesize that this group arose from a single introduction from an ancestral BtCoV. A possible transmission pathway from bats to poultry may have occurred via raptors, which are known to prey on bats (3). Obviously, this is speculative and may only be confirmed by virological investigation of bird populations.

The emergence and evolutionary pathway of SARS-CoVs provide a unique opportunity to study the evolutionary dynamics involved with the interspecies transmission of coronaviruses. Comparison of population growth models of group 4A

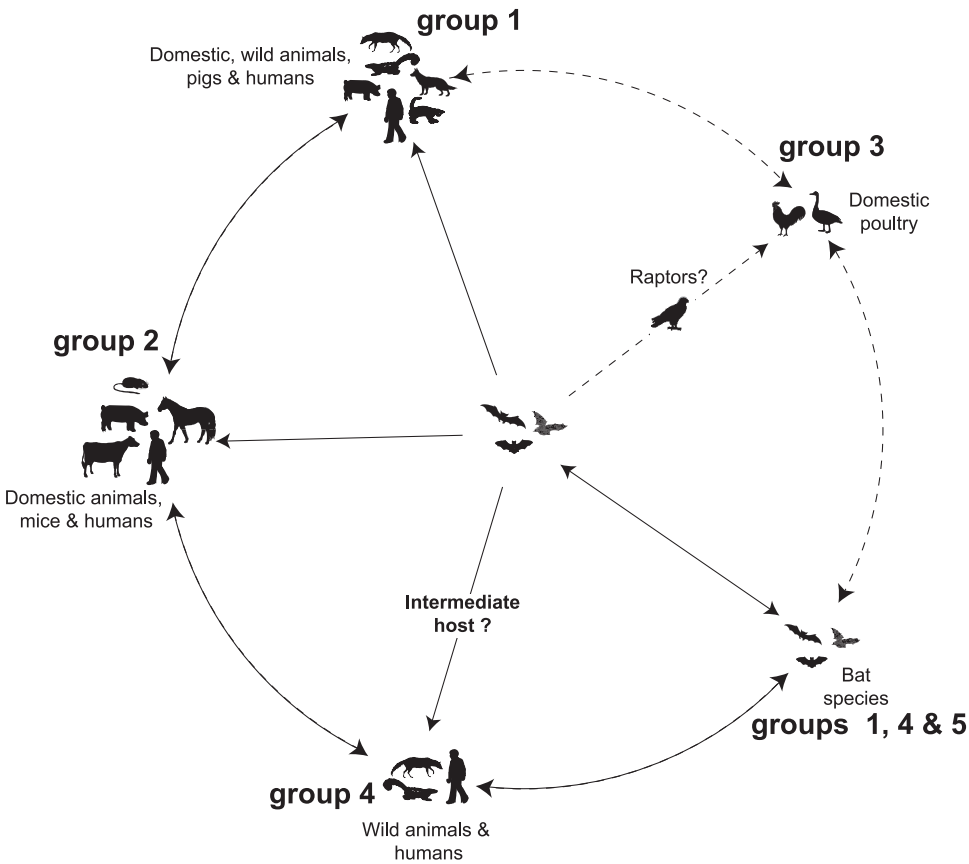

FIG. 4. Postulated ecology of coronaviruses. Solid and dashed lines represent confirmed and hypothetical interspecies transmission pathways, respectively, and arrows represent evolutionary pathways.

Downloaded from http://jvi.asm.org/ on March 11, 2015 by ST ANDREWS UNIV

and 4B viruses indicated a shift from a constant population size to epidemic-like growth following interspecies transmission from the viruses' natural reservoir to alternate hosts. An understanding of the ecology and evolution of coronaviruses will provide crucial information to identify viruses with epidemic potential. It is likely that once more information becomes available, ideas regarding coronaviruses will continue to change radically, as they have since 2003.

ACKNOWLEDGMENTS

This work was supported by the Research Fund for the Control of Infectious Diseases of the Health Welfare and Food Bureau of the Hong Kong SAR Government (no. 03040782), by the National Special Task Force Fund for Identification of the Animal Reservoir of SARS-CoV, and by the Li Ka Shing Foundation.

We thank Andrew Rambaut and Alexei Drummond for their help with the BEAST software package.

REFERENCES

1. **Arbogast, B. S., S. V. Edwards, J. Wakeley, P. Beerli, and J. B. Slowinski.** 2002. Estimating divergence times from molecular data on phylogenetic and population genetic timescales. Annu. Rev. Ecol. Syst. **33:**707–740.
2. **Avise, J. C., J. Arnold, R. M. Ball, E. Bermingham, T. Lamb, J. E. Neigel, C. A. Reeb, and N. C. Saunders.** 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annu. Rev. Ecol. Syst. **18:**489–522.
3. **Baker, J. E.** 1962. The manner and efficiency of raptor depredations on bats. Condor **64:**500–504.
4. **Bininda-Emonds, O. R. P.** 2005. transAlign: using amino acids to facilitate the multiple alignment of protein-coding DNA sequences. BMC Bioinformatics **6:**156.
5. **Chua, K. B., W. J. Bellini, P. A. Rota, B. H. Harcourt, A. Tamin, S. K. Lam, T. G. Ksiazej, P. E. Rollin, S. R. Zaki, W. J. Shieh, C. S. Goldsmith, D. J. Gubler, J. T. Roehrig, B. Eaton, A. R. Gould, J. Olson, H. Field, P. Daniels, A. E. Ling, C. J. Peters, L. J. Anderson, and B. W. J. Mahy.** 2000. Nipah virus: a recently emergent deadly paramyxovirus. Science **288:**1432–1435.
6. **Delaplane, J. P., and H. O. Stuart.** 1941. The modifications of infectious bronchitis virus of chickens as the result of propagation in embryonated chicken eggs. Rhode Island Agr. Exp. Stn. Bull. **284:**4–20.
7. **Doyle, L. P., and L. M. Hutching.** 1946. A transmissible gastroenteritis virus in pigs. J. Am. Vet. Med. Assoc. **108:**257–259.
8. **Drummond, A. J., S. Y. Ho, M. J. Phillips, and A. Rambaut.** 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. **4:**e88.
9. **González, J. M., P. Gomez-Puertas, D. Cavanagh, A. E. Gorbalenya, and L. Enjunes.** 2003. A comparative sequence analysis to revise the current taxonomy of the family coronaviridae. Arch. Virol. **148:**2207–2235.
10. **Gorbalenya, A. E., E. J. Snijder, and W. J. M. Spaan.** 2004. Severe acute respiratory syndrome coronavirus phylogeny: towards consensus. J. Virol. **78:**7863–7866.
11. **Guan, Y., B. J. Zheng, Y. Q. He, X. L. Liu, Z. X. Zhuang, C. L. Cheung, S. W. Luo, P. H. Li, L. J. Zhang, Y. J. Guan, K. M. Butt, K. L. Wong, K. W. Chan, W. Lim, K. F. Shortridge, K. Y. Yuen, J. S. M. Peiris, and L. L. M. Poon.** 2003. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. Science **302:**276–278.
12. **Hamre, D., and J. J. Procknow.** 1966. A new virus isolated from the human respiratory tract. Proc. Soc. Exp. Biol. Med. **121:**190–193.
13. **Haydon, D. T., S. Cleaveland, L. H. Taylor, and M. K. Laurenson.** 2002. Identifying reservoirs of infection: a conceptual and practical challenge. Emerg. Infect. Dis. **8:**1468–1473.
14. **He, R., F. Dobie, M. Ballantine, A. Leeson, Y. Li, N. Bastien, T. Cutts, A. Andonov, J. Cao, T. F. Booth, F. A. Plummer, S. Tyler, L. Baker, and X. Li.** 2004. Analysis of multimerization of the SARS coronavirus nucleocapsid protein. Biochem. Biophys. Res. Commun. **316:**476–483.
15. **Huelsenbeck, P. J., and F. R. Ronquist.** 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17:**754–755.
16. **Keele, B. F., F. V. Heuverswyn, Y. Li, E. Bailes, J. Takehisa, M. L. Saniago, F. Bibollet-Ruche, Y. Chen, L. V. Wain, F. Liegeois, S. Loul, E. M. Ngole, Y. Bienvenue, E. Delaporte, J. F. Y. Brookfield, P. M. Sharp, G. M. Shaw, M. Peeters, and B. Hahn.** 2006. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. Science **131:**523–526.
17. **Kosakovsky Pond, S. L., D. Posada, M. B. Gravenor, C. H. Woelk, and S. D. W. Frost.** 2006. Automated phylogenetic detection of recombination using a genetic algorithm. Mol. Biol. Evol. **23:**1891–1901.
18. **Kosakovsky Pond, S. L., and S. D. W. Frost.** 2005. A genetic algorithm approach to detecting lineage-specific variation in selection pressure. Mol. Biol. Evol. **22:**478–485.
19. **Kosakovsky Pond, S. L., S. D. W. Frost, and S. V. Muse.** 2005. Hyphy: hypothesis testing using phylogenetics. Bioinformatics **21:**676–679.
20. **Lai, M. M. C., and K. V. Holmes.** 2001. Coronaviruses, p. 1163–1185. *In* D. M. Knipe, P. M. Howley, D. E. Griffin, R. A. Lamb, M. A. Martin, B. Roizman, and S. E. Straus (ed.), Fields virology. Lippincott Williams & Wilkins, Philadelphia, PA.
21. **Leroy, E. M., B. Kumulungui, X. Pourrut, P. Rouquet, A. Hassanin, P. Yaba, A. Délicat, J. T. Paweska, J. P. Gonzalez, and R. Swanepoel.** 2005. Fruit bats as reservoirs of Ebola virus. Nature **438:**575–576.
22. **Li, W., Z. Shi, M. Yu, W. Ren, C. Smith, J. H. Epstein, H. Wang, G. Crameri, Z. Hu, H. Zhang, J. Zhang, J. McEachern, H. E. Field, P. Daszak, B. Eaton, S. Zhang, and L. F. Wang.** 2005. Bats are natural reservoirs of SARS-like coronaviruses. Science **310:**676–679.
23. **Manaker, R. A., C. V. Piczak, A. A. Miller, and M. F. Stanton.** 1961. A hepatitis virus complicating studies with mouse leukemia. J. Natl. Cancer Inst. **27:**29–51.
24. **Marra, M. A., S. J. M. Jones, C. R. Astell, R. A. Holt, A. Brooks-Wilson, Y. S. N. Butterfield, J. Khattra, J. K. Asano, S. A. Barber, S. Y. Chan, A. Cloutier, S. M. Coughlin, D. Freeman, N. Girn, O. L. Griffinth, S. R. Leach, M. Mayo, H. McDonald, S. B. Montgomery, P. K. Pandoh, A. S. Petresue, A. G. Tobertson, J. E. Schein, S. Siddiqui, D. E. Smailus, J. M. Stott, and G. S. Yang.** 2003. The genome sequence of the SARS-associated coronavirus. Science **300:**1399–1404.
25. **McKeirnan, A. J., J. F. Evermann, A. Hargis, L. M. Miller, and R. L. Ott.** 1981. Isolation of feline coronaviruses from two cats with diverse disease manifestations. Feline Pract. **11:**16–20.
26. **Pensaert, M. B., and P. E. Callebaut.** 1974. Characteristics of a coronavirus causing vomition and wasting in pigs. Arch. Gesamte Virusforsch. **44:**35–50.
27. **Pensaert, M. B., and P. DeBouck.** 1978. A new coronavirus-like particle associated with diarrhea in swine. Arch. Virol. **58:**243–247.
28. **Shackelton, L. A., C. R. Parrish, U. Truyen, and E. C. Holmes.** 2005. High rate of viral evolution associated with the emergence of carnivore parvovirus. Proc. Natl. Acad. Sci. USA **102:**379–384.
29. **Shimodaira, H., and M. Hasegawa.** 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. **16:**1114–1116.
30. **Snijder, E. J., P. J. Bredenbeek, J. C. Dobbe, V. Thiel, J. Ziebuhr, L. L. M. Poon, Y. Guan, M. Rozanov, W. J. M. Spaan, and A. E. Gorbalenya.** 2003. Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineages. J. Mol. Biol. **331:**991–1004.
31. **Song, H. D., C. C. Tu, G. W. Zhang, S. Y. Wang, K. Zheng, L. C. Lei, Q. X. Chen, Y. W. Gao, H. Q. Zhou, H. Xiang, H. J. Zheng, S. W. Chern, F. Cheng, C. M. Pan, H. Xuan, S. J. Chen, H. M. Luo, D. H. Zhou, Y. F. Liu, J. F. He, P. Z. Qin, L. H. Li, Y. Q. Ren, W. J. Liang, Y. D. Yu, L. Anderson, M. Wang, R. H. Xu, X. W. Wu, H. Y. Zheng, J. D. Chen, G. Liang, Y. Gao, M. Liao, L. Fang, L. Y. Jiang, H. Li, F. Chen, B. Di, L. J. He, J. Y. Lin, S. Tong, X. Kong, L. Du, P. Hao, H. Tang, A. Bernini, X. J. Yu, O. Spiga, Z. M. Guo, H. Y. Pan, W. Z. He, J. C. Manuguerra, A. Fontanet, A. Danchin, N. Niccolai, Y. X. Li, C. I. Wu, and G. P. Zhao.** 2005. Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. Proc. Natl. Acad. Sci. USA **102:**2430–2435.
32. **Storz, J., X. Lin, C. W. Purdy, V. N. Chouljenko, K. G. Kousoulas, F. M. Enright, W. C. Gilmore, R. E. Briggs, and R. W. Loan.** 2000. Coronavirus and *Pasteurella* infections in bovine shipping fever pneumonia and Evans' criteria for causation. J. Clin. Microbiol. **38:**3291–3298.
33. **Subbarao, K., A. Klimov, J. Katz, H. Regnery, W. Lim, H. Hall, M. Perdue, D. Swayne, C. Bender, J. Huang, M. Hemphill, T. Rowe, M. Shaw, X. Xu, K. Fukuda, and N. Cox.** 1998. Characterization of an avian influenza A (H5N1) virus isolated from a child with a fatal respiratory illness. Science **279:**393–396.
34. **Swofford, D. L.** 2001. PAUP*: phylogenetic analysis using parsimony (and other methods) 4.0 beta. Sinauer Associates, Sunderland, MA.
35. **Tang, X. C., J. X. Zhang, S. Y. Zhang, P. Wang, X. H. Fan, L. F. Li, G. Li, B. Q. Dong, W. Liu, C. L. Cheung, K. M. Xu, W. J. Song, D. Vijaykrishna, L. L. M. Poon, J. S. M. Peiris, G. J. D. Smith, H. Chen, and Y. Guan.** 2006. Prevalence and genetic diversity of coronaviruses in bats from China. J. Virol. **80:**7481–7490.
36. **Thiel, V. K., A. Ivanov, A. Putics, T. Hertzig, B. Schelle, S. Bayer, B. Weissbrich, E. J. Snijder, H. Rabenau, H. W. Doerr, A. E. Gorbalenya, and J. Ziebuhr.** 2003. Mechanisms and enzymes involved in SARS coronavirus genome expression. J. Gen. Virol. **84:**2305–2315.
37. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:**4673–4680.
38. **Tu, C., G. Crameri, X. Kong, J. Chen, Y. Sun, M. Yu, H. Xiang, X. Xia, S. Liu, T. Ren, Y. Yu, B. T. Eaton, H. Xuan, and L. F. Wang.** 2004. Antibodies to SARS coronavirus in civets. Emerg. Infect. Dis. **10:**2244–2248.
39. **van der Hoek, L., K. Pyrc, M. F. Jebbink, W. Vermeulen-Oost, R. J. Berkhout, K. C. Wolthers, P. M. Wertheim-van Dillen, J. Kaandorp, J. Spaargaren,**

Downloaded from http://jvi.asm.org/ on March 11, 2015 by ST ANDREWS UNIV

and B. Berkhout. 2004. Identification of a new human coronavirus. Nat. Med. **10:**368–373.

40. **Vijgen, L., E. Keyaerts, E. Moës, I. Thoelen, E. Wollants, P. Lemey, A. M. Vandamme, and M. V. Ranst.** 2005. Complete genomic sequence of human coronavirus OC43: molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event. J. Virol. **79:**1595–1604.
41. **Vijgen, L., E. Keyaerts, P. Lemey, P. Maes, K. Van Reeth, H. Nauwynck, M. Pensaert, and M. Van Ranst.** 2006. Evolutionary history of the closely related group 2 coronaviruses: porcine hemagglutinating encephalomyelitis virus, bovine coronavirus, and human coronavirus OC43. J. Virol. **80:**7270–7274.
42. **Wang, M., M. Yan, H. Xu, W. Liang, B. Kan, B. Zheng, H. Chen, H. Zheng, Y. Xu, E. Zhang, H. Wang, J. Ye, G. Li, M. Li, Z. Cui, Y. F. Liu, R. T. Guo, X. N. Liu, L. H. Zhan, D. H. Zhou, A. Zhao, R. Hai, D. Yu, Y. Guan, and J. Xu.** 2005. SARS-CoV infection in a restaurant from palm civet. Emerg. Infect. Dis. **11:**1860–1865.
43. **Weiss, S. R., and S. Navas-Martin.** 2005. Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus. Microbiol. Mol. Biol. Rev. **69:**635–664.
44. **Woo, P. C. Y., S. K. P. Lau, H. W. Tsoi, R. W. S. Poon, C. M. Chu, R. A. Lee, W. K. Luk, G. K. M. Wong, B. H. L. Wong, V. C. C. Cheng, B. S. F. Tang, A. Wu, R. W. H. Yung, H. Chen, Y. Guan, K. H. Chan, and K. Y. Yuen.** 2005. Clinical and molecular epidemiological features of coronavirus HKU1-associated community-acquired pneumonia. J. Infect. Dis. **192:**1898–1907.
45. **Woo, P. C. Y., S. K. P. Lau, K. S. M. Li, R. W. S. Poon, B. H. L. Wong, H. W. Tsoi, B. C. K. Yip, Y. Huang, K. H. Chan, and K. Y. Yuen.** 2006. Molecular diversity of coronaviruses in bats. Virology **351:**180–187.
46. **Xu, R. H., J. F. He, M. R. Evans, G. W. Peng, H. E. Field, D. W. Yu, C. K. Lee, H. M. Luo, W. S. Lin, P. Lin, L. H. Li, W. J. Liang, J. Y. Lin, and A. Schnur.** 2004. Epidemiological clues to SARS origin in China. Emerg. Infect. Dis. **10:**1030–1037.
47. **Zeng, F. Y., C. W. Chan, M. N. Chan, J. D. Chen, K. Y. Chow, C. C. Hon, K. H. Hui, J. Li, V. Y. Li, C. Y. Wang, P. Y. Wang, Y. Guan, B. Zheng, L. L. Poon, K. H. Chan, K. Y. Yuen, J. S. M. Peiris, and F. C. Leung.** 2003. The complete genome sequence of severe acute respiratory syndrome coronavirus strain HKU-39849 (HK-39). Exp. Biol. Med. **228:**866–873.

Downloaded from http://jvi.asm.org/ on March 11, 2015 by ST ANDREWS UNIV

JOURNAL OF VIROLOGY, Aug. 2007, p. 8371 Vol. 81, No. 15
0022-538X/07/$08.00+0 doi:10.1128/JVI.01135-07

ERRATUM

Evolutionary Insights into the Ecology of Coronaviruses

D. Vijaykrishna, G. J. D. Smith, J. X. Zhang, J. S. M. Peiris, H. Chen, and Y. Guan

State Key Laboratory of Emerging Infectious Diseases, Department of Microbiology, Li Ka Shing Faculty of Medicine, The University of Hong Kong, 21 Sassoon Road, Pokfulam, Hong Kong SAR, China

Volume 81, no. 8, p. 4012–4020, 2007. Page 4013, Materials and Methods, column 1, lines 2–4 from the bottom: "The times of divergence were estimated utilizing priors obtained from earlier runs, with a discarded burn-in of approximately 10%" should read "The times of divergence were estimated with a discarded burn-in of approximately 10%, utilizing optimized operator tuning values to increase the efficiency of sampling for subsequent runs."