JOURNAL OF VIROLOGY, Aug. 2004, p. 7863–7866 Vol. 78, No. 15
0022-538X/04/$08.00+0 DOI: 10.1128/JVI.78.15.7863–7866.2004

GUEST COMMENTARY

Severe Acute Respiratory Syndrome Coronavirus Phylogeny: toward Consensus

Alexander E. Gorbalenya,* Eric J. Snijder, and Willy J. M. Spaan

Department of Medical Microbiology, Leiden University Medical Center, Leiden, The Netherlands

Since the identification of a new coronavirus (severe acute respiratory syndrome coronavirus [SARS-CoV]) as the causative agent of the SARS epidemic in the winter of 2002-2003, the origin of the novel agent has remained a hotly debated topic. Which virus was the immediate ancestor of SARS-CoV, and what are the relationships between SARS-CoV and other previously described coronaviruses? Correct answers to these two questions are vital, as substantiated below, for designing strategies to detect, contain, and combat new outbreaks and for dissecting the fundamentals of the SARS-CoV life cycle.

Major efforts have been invested in a thus far unsuccessful search for a natural SARS-CoV reservoir. In the meantime, and more outside the spotlight, SARS-CoV genome sequences have been used to define the phylogenetic position of SARS-CoV among coronaviruses. These studies have resulted in a lot of controversy whose intricacies may not be very clear to outsiders. Our purpose is to clarify the situation from an insider's point of view.

Originally, coronaviruses were classified on the basis of antigenic cross-reactivity, and in this manner three antigenic groups (1 to 3) were recognized (14). When coronavirus genome sequences began to accumulate, the same groups were evident from phylogenetic analyses of the four structural proteins, N, M, E, and S (19), and of different regions of the giant replicase (3, 22). Group boundaries were also supported by the diversity of small open reading frames (ORFs) encoding accessory proteins, which are dispersed among the structural protein genes in the 3′-proximal region of the genome (Fig. 1). In the middle of the nineties, a first discord between the antigenicity-based and phylogenetic classifications emerged upon the characterization of the coronavirus porcine epidemic diarrhea virus (PEDV) and human coronavirus 229E (HCoV-229E), one of the common cold viruses. These viruses proved not to have antigenic cross-reactivity with members of the established groups (18), yet on the basis of sequence comparisons it was concluded that they segregate into group 1, although they are somewhat separated from porcine transmissible gastroenteritis virus and closely related viruses (subgroup 1b and subgroup 1a, respectively, in Fig. 2) (9). The PEDV and HCoV-229E genomes also share an ORF specific for group 1 in the 3′-proximal region of their genome. The Coronavirus Study Group of the International Committee on Taxonomy of Viruses recognized these viruses as members of group 1 rather than declaring them prototypes of new groups (6). This decision effectively converted the original antigenic groups—which were based essentially on some properties of one or a few viral proteins—into a genetic one based on full-length genome sequences, but this change was never acknowledged explicitly. Consequently, no guidelines were established with respect to handling future disagreements between the classifications based on antigenicity, genome organization, and phylogeny should these arise from the properties of newly identified coronaviruses, and SARS-CoV proved to be quite a classification challenge.

Initial phylogenetic analyses suggested that the novel virus did not cluster with any of the three established coronavirus groups. Accordingly, SARS-CoV also has a unique pattern of small ORFs in the 3′-proximal region of its genome and a unique internal organization of its nonstructural protein 3 (nsp3) replicase subunit, which includes a sizable novel domain (SARS-CoV unique domain SUD) and only one papain-like protease (PL2pro) rather than the two copies commonly found in other coronaviruses (Fig. 1). Although a thorough assessment of the antigenic cross-reactivity of SARS-CoV with other coronaviruses is yet to be published, a proposal to recognize SARS-CoV as a representative of a new, fourth group of coronaviruses seemed most logical (15, 17).

If SARS-CoV indeed represents a new group, then when, relative to other groups, could this lineage have emerged? Several scenarios are theoretically plausible, and one of the most extreme ones, which seems compatible with the unique characteristics of SARS-CoV, places the origin of this lineage next to the ancestor of the other coronaviruses (Fig. 2A). To rigorously infer the origin of SARS-CoV, we conducted a special analysis of the replicase ORF1b region (Fig. 1), the most-conserved part of the coronavirus genome, which accounts for ~20% of its size (20). In this analysis, the equine torovirus—a distant relative of coronaviruses belonging to the genus *Torovirus* of the same *Coronaviridae* family—was used as an outgroup to infer the direction of coronavirus evolution.

Surprisingly, our fully resolved tree demonstrated that the SARS-CoV lineage is an early split-off from the group 2 branch and that the split-off occurred relatively late in coronavirus evolution, after the two bifurcations that gave rise to the three previously established groups (Fig. 2B). This topology is unlikely to be skewed, as it was obtained by using different criteria

* Corresponding author. Mailing address: Department of Medical Microbiology, Center of Infectious Diseases, Leiden University Medical Center, Albinusdreef 2, E4-P, Room L04-036, 2333 ZA Leiden, The Netherlands. Phone: 31-71-526-1652. Fax: 31-71-526-6761. E-mail: a.e.gorbalenya@lumc.nl.

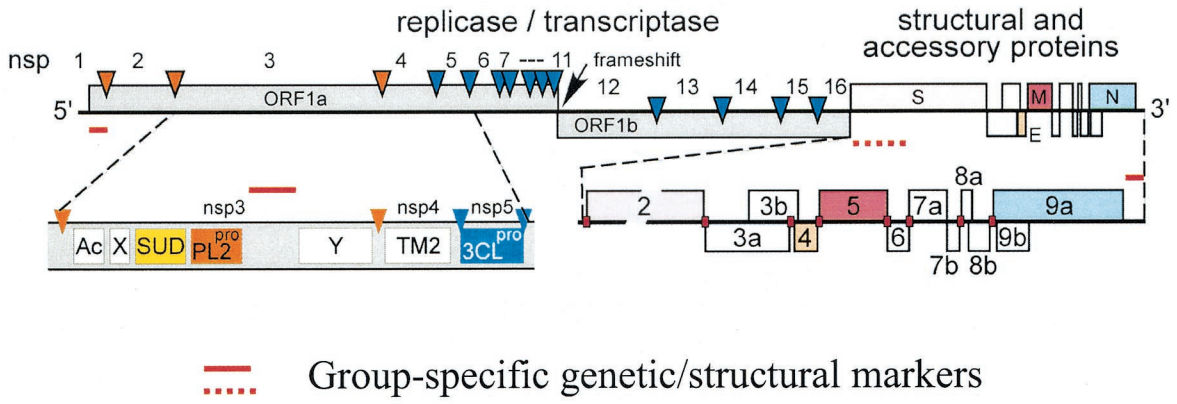

FIG. 1. Overview of the SARS-CoV genome organization and proteins (20). The positions of coronavirus group-specific genetic and structural markers are indicated. SARS-CoV has the markers of group 2 coronaviruses. In the N-terminal part of the S protein, only Cys residues are distributed in a group-specific fashion (5) (dashed line). SUD, SARS-CoV-unique domain; $PL2^{pro}$, group 2 papain-like protease 2; TM2, transmembrane domain 2; $3CL^{pro}$, chymotrypsin-like protease related to $3C^{pro}$ of picornaviruses.

and both protein and nucleotide sequences as well as in an independent study (21). The early split off of the (avian) group 3 coronaviruses from all mammalian coronaviruses also adds to the credibility of the tree. Based on these observations, it was proposed (20) that the group 2 coronaviruses be split into subgroup 2a, which includes the established group 2 coronaviruses like mouse hepatitis virus (MHV) and bovine coronavirus (BCoV), and subgroup 2b, of which SARS-CoV is the prototype.

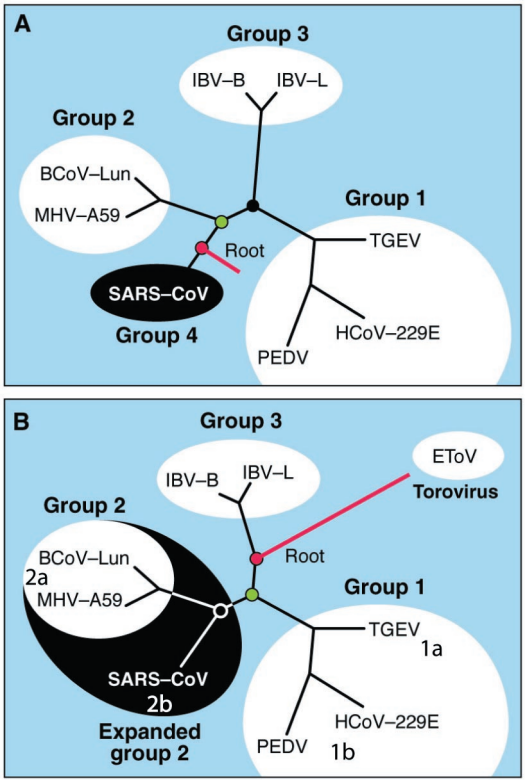

FIG. 2. Two alternative scenarios of SARS-CoV evolution, shown on the same coronavirus tree with alternative roots. The SARS-CoV branch splits, for the sake of comparison, from an ancestor of the other known coronaviruses (A) or, as defined in reference 20, from the group 2 branch (B). IBV, infectious bronchitis virus; TGEV, porcine transmissible gastroenteritis virus; EToV, equine torovirus. BCoV-Lun and other BCoV strains are most closely related to HCoV OC43, which was not included in the analysis.

This classification of SARS-CoV, one of the least expected, appeared to be irreconcilable with other data and was fiercely debated. Some researchers, unaware of the complex relationship between phylogeny- and antigenicity-based classifications of coronaviruses (see above), may have been misguided to believe that this tree (Fig. 2B) implies that SARS-CoV shares antigenic cross-reactivity with established group 2 viruses. Others may have had a hard time reconciling all the unique properties of SARS-CoV with its group 2 membership. Is the latter gap indeed as large as it seems?

Inspection of the published unrooted trees reveals that in the majority, the SARS-CoV branch indeed originates on the group 2 side. Other topologies were originally published only for the M and E structural proteins, but the M-protein tree was later revised in favor of the dominant topology (5). The level of protein conservation behind these topologies is marginally higher between SARS-CoV and group 2 viruses than between group 1 and 3 viruses and, despite its consistent character, may not appear very impressive. However, this margin is significant on the evolutionary scale, since in addition, distinctive group 2-specific genetic and/or structural markers were recently identified for SARS-CoV in the nsp1 and nsp3 replicase subunits (20), the S1 portion of the S protein (5), and, at the RNA level, the 3′ untranslated region (3′UTR) of the genome (8) (Fig. 1).

The conspicuous genetic differences between SARS-CoV and subgroup 2a coronaviruses may not be completely surprising either. Similar, albeit less profound, differences are also evident upon comparison of regions encoding nsp3 and comparable genome proteins encoded downstream of ORF1b in viruses of subgroups 1a and 1b (6, 20). Furthermore, coronaviruses tolerate deletions and reshuffling of ORFs in the 3′-proximal region of their genome in genetic experiments (4). Also, in SARS-CoV the rapid evolution of unique 3′-proximal ORFs has already been reported, both in the field (ORF8a and

ORF8b) (10) and in Vero cell cultures (ORF7b) (20, 23). Thus, there is little in the available data that could not be reconciled with the classification of SARS-CoV as a group 2 coronavirus.

This and any other hierarchical classification is subject to the condition that recombination—a well-known phenomenon in coronaviruses (11, 13)—is mostly restricted by group boundaries. This notion was recently challenged in two studies claiming that the putative RNA-dependent RNA polymerase (RdRp) locus of SARS-CoV (nsp12 in Fig. 1) has resulted from recombination of viruses from groups 1 and 4 (16) and that the M and N protein genes are of group 3 origin (21). The genes identified as being of recombinant origin specify key functions of coronavirus replication and virion biogenesis, respectively, and the above results were obtained with advanced phylogenetic methods which were never before applied to coronaviruses. Taken together, these considerations seem to argue for the mosaic origin of these SARS-CoV major genes and raise the question of whether it will be possible at all to build a meaningful classification for SARS-CoV and other coronaviruses. However, we believe that it is premature to accept these conclusions. For instance, it remains unresolved why the authors of these studies (16, 21) were unable to mutually verify each other's findings and why others, who analyzed the same regions, have missed these recombinations (7). For the RdRp results, it is also worth noting that unless very close relatives are involved recombinants generated by exchange within regions encoding key replicative enzymes of RNA viruses may have poor viability (for examples, see reference 1), emphasizing that claims to the contrary should be validated extensively.

Recombination aside, what is at stake? Does it really matter whether SARS-CoV is a distant member of group 2 or the prototype of a group of its own? It does. The clustering of SARS-CoV within the younger subgroup 2b rather than the older group 4 signifies that relatively small virus diversity can be anticipated in the SARS-CoV lineage, and consequently, the unique genetic properties of SARS-CoV must have evolved relatively fast on the scale of coronavirus evolution. It further predicts that SARS-CoV may be more prone to further gross changes than suggested by its alternative classification in the fourth group. It identifies group 2 viruses, including the extensively characterized MHV, as the most suitable virus models to characterize SARS-CoV-specific genes and/or domains, in particular during dissection of the replicative machinery of the virus and processes involving group-specific determinants. It also indicates that analysis of the close BCoV/HCoV-O43 pair from group 2 could provide useful insights for understanding the adaptation of animal coronaviruses to humans in general and the emergence of SARS-CoV in particular. In summary, decision making in both fundamental and applied research is not going to be the same with SARS-CoV being recognized as a member of subgroup 2b rather than group 4.

What is next? Part of the original and remaining confusion concerning SARS-CoV phylogeny could be attributed to technical aspects, like alignment quality, choice of viruses and genes analyzed, and software used. These issues remain crucial if the ongoing phylogenetic analysis of SARS-CoV is to result in a broad consensus. Particularly, this future accord should accommodate RNA recombination. Although the swapping of conserved genes or domains between distant coronaviruses, including SARS-CoV, is yet to be independently verified (see above), the involvement of recombination in the evolution of the SARS-CoV lineage (subgroup 2b) per se is beyond doubt. This assertion is based on the numerous data identifying RNA recombination as the major mechanism for gross evolution of RNA virus genomes (2, 12). Consequently, the most unique genetic characteristics of SARS-CoV in the nsp3 gene and in the 3′-proximal region of the genome must have evolved by recombination involving either SARS-CoV itself, its subgroup 2b ancestors, or both. In this respect, the identification of the parental sequences of these SARS-CoV-specific genes is clearly very important, and the sequence analysis of new subgroup 2b coronavirus genomes may help in this quest. The functional characterization of group-specific genetic markers, as initiated for the 3′UTR (8), is essential for providing the foundations for a biologically appealing definition of new groups. Finally, the whole controversy arose in the first place because the field had become used to dealing with classification matters on a case-by-case basis and had never before experienced any urgency in this regard. Now it is high time for the coronavirus community to make up its mind with respect to the relative classification value of phylogenetic analyses of replicative, structural, and accessory proteins and to formulate the principles to build a consistent and reliable coronavirus classification system.

REFERENCES

1. **Bell, Y. C., B. L. Semler, and E. Ehrenfeld.** 1999. Requirements for RNA replication of a poliovirus replicon by coxsackievirus B3 RNA polymerase. J. Virol. **73:**9413–9421.
2. **Bujarski, J. J.** 1997. Experimental systems of genetic recombination and defective RNA formation in RNA viruses. Part II. Semin. Virol. **8:**75–76.
3. **Chouljenko, V. N., X. Q. Lin, J. Storz, K. G. Kousoulas, and A. E. Gorbalenya.** 2001. Comparison of genomic and predicted amino acid sequences of respiratory and enteric bovine coronaviruses isolated from the same animal with fatal shipping pneumonia. J. Gen. Virol. **82:**2927–2933.
4. **de Haan, C. A. M., P. S. Masters, X. L. Shen, S. Weiss, and P. J. M. Rottier.** 2002. The group-specific murine coronavirus genes are not essential, but their deletion, by reverse genetics, is attenuating in the natural host. Virology **296:**177–189.
5. **Eickmann, M., S. Becker, H. D. Klenk, H. W. Doerr, K. Stadler, S. Censini, S. Guidotti, V. Masignani, M. Scarselli, M. Mora, C. Donati, J. H. Han, H. C. Song, S. Abrignani, A. Covacci, and R. Rappuoli.** 2003. Phylogeny of the SARS coronavirus. Science **302:**1504–1505.
6. **Enjuanes, L., D. Brian, D. Cavanagh, K. Holmes, M. M. C. Lai, H. Laude, P. S. Masters, P. Rottier, S. Siddell, W. Spaan, F. Taguchi, and P. Talbot.** 2000. Family *Coronaviridae*, p. 835–849. *In* M. H. V. van Regenmortel, C. M. Fauquet, D. H. L. Bishop, E. B. Carstens, M. K. Estes, S. M. Lemon, J. Maniloff, M. A. Mayo, D. J. McGeoch, C. R. Pringle, and R. B. Wickner (ed.), Virus taxonomy: classification and nomenclature of viruses. Seventh report of the International Committee on Taxonomy of Viruses. Academic Press, San Diego, Calif.
7. **Gibbs, A. J., M. J. Gibbs, and J. S. Armstrong.** 2004. The phylogeny of SARS coronavirus. Arch. Virol. **149:**621–624.
8. **Goebel, S. J., B. Hsue, T. F. Dombrowski, and P. S. Masters.** 2004. Characterization of the RNA components of a putative molecular switch in the 3′ untranslated region of the murine coronavirus genome. J. Virol. **78:**669–682.
9. **Gonzaalez, J. M., P. Gomez-Puertas, D. Cavanagh, A. E. Gorbalenya, and L. Enjuanes.** 2003. A comparative sequence analysis to revise the current taxonomy of the family Coronaviridae. Arch. Virol. **148:**2207–2235.
10. **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.
11. **Herrewegh, A. A. P. M., I. Smeenk, M. C. Horzinek, P. J. M. Rottier, and R. J. de Groot.** 1998. Feline coronavirus type II strains 79–1683 and 79–1146 originate from a double recombination between feline coronavirus type I and canine coronavirus. J. Virol. **72:**4508–4514.

12. **Lai, M. M. C.** 1992. RNA recombination in animal and plant viruses. Microbiol. Rev. **56:**61–79.
13. **Lai, M. M. C., R. S. Baric, S. Makino, J. G. Keck, J. Egbert, J. L. Leibowitz, and S. A. Stohlman.** 1985. Recombination between nonsegmented RNA genomes of murine coronaviruses. J. Virol. **56:**449–456.
14. **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.
15. **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. Griffin, S. R. Leach, M. Mayo, H. McDonald, S. B. Montgomery, P. K. Pandoh, A. S. Petrescu, A. G. Robertson, J. E. Schein, A. Siddiqui, D. E. Smailus, J. E. Stott, G. S. Yang, F. Plummer, A. Andonov, H. Artsob, N. Bastien, K. Bernard, T. F. Booth, D. Bowness, M. Czub, M. Drebot, L. Fernando, R. Flick, M. Garbutt, M. Gray, A. Grolla, S. Jones, H. Feldmann, A. Meyers, A. Kabani, Y. Li, S. Normand, U. Stroher, G. A. Tipples, S. Tyler, R. Vogrig, D. Ward, B. Watson, R. C. Brunham, M. Krajden, M. Petric, D. M. Skowronski, C. Upton, and R. L. Roper.** 2003. The genome sequence of the SARS-associated coronavirus. Science **300:**1399–1404.
16. **Rest, J. S., and D. P. Mindell.** 2003. SARS associated coronavirus has a recombinant polymerase and coronaviruses have a history of host-shifting. Infect. Genet. Evol. **3:**219–225.
17. **Rota, P. A., M. S. Oberste, S. S. Monroe, W. A. Nix, R. Campagnoli, J. P. Icenogle, S. Penaranda, B. Bankamp, K. Maher, M. H. Chen, S. X. Tong, A. Tamin, L. Lowe, M. Frace, J. L. Derisi, Q. Chen, D. Wang, D. D. Erdman, T. C. T. Peret, C. Burns, T. G. Ksiazek, P. E. Rollin, A. Sanchez, S. Liffick, B. Holloway, J. Limor, K. McCaustland, M. Olsen-Rasmussen, R. Fouchier, S. Gunther, A. D. M. E. Osterhaus, C. Drosten, M. A. Pallansch, L. J. Anderson, and W. J. Bellini.** 2003. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science **300:**1394–1399.
18. **Sanchez, C. M., G. Jimenez, M. D. Laviada, I. Correa, C. Sune, M. J. Bullido, F. Gebauer, C. Smerdou, P. Callebaut, J. M. Escribano, and L. Enjuanes.** 1990. Antigenic homology among coronaviruses related to transmissible gastroenteritis virus. Virology **174:**410–417.
19. **Siddell, S.** 1995. The *Coronaviridae*. Plenum Press, New York, N.Y.
20. **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 lineage. J. Mol. Biol. **331:**991–1004.
21. **Stavrinides, J., and D. S. Guttman.** 2004. Mosaic evolution of the severe acute respiratory syndrome coronavirus. J. Virol. **78:**76–82.
22. **Stephensen, C. B., D. B. Casebolt, and N. N. Gangopadhyay.** 1999. Phylogenetic analysis of a highly conserved region of the polymerase gene from 11 coronaviruses and development of a consensus polymerase chain reaction assay. Virus Res. **60:**181–189.
23. **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.

The views expressed in this Commentary do not necessarily reflect the views of the journal or of ASM.