Testing the hypothesis of a recombinant origin of the SARS-associated coronavirus

X. W. Zhang¹, Y. L. Yap¹, and A. Danchin²

¹HKU-Pasteur Research Centre, Hong Kong, P.R. China
²Pasteur Institute, Unit Genetics of Bacterial Genomes, Paris, France

Received February 27, 2004; accepted August 16, 2004
Published online October 11, 2004 © Springer-Verlag 2004

Summary. The origin of severe acute respiratory syndrome-associated coronavirus (SARS-CoV) is still a matter of speculation, although more than one year has passed since the onset of the SARS outbreak. In this study, we implemented a 3-step strategy to test the intriguing hypothesis that SARS-CoV might have been derived from a recombinant virus. First, we blasted the whole SARS-CoV genome against a virus database to search viruses of interest. Second, we employed 7 recombination detection techniques well documented in successfully detecting recombination events to explore the presence of recombination in SARS-CoV genome. Finally, we conducted phylogenetic analyses to further explore whether recombination has indeed occurred in the course of coronaviruses history predating the emergence of SARS-CoV. Surprisingly, we found that 7 putative recombination regions, located in Replicase 1ab and Spike protein, exist between SARS-CoV and other 6 coronaviruses: porcine epidemic diarrhea virus (PEDV), transmissible gastroenteritis virus (TGEV), bovine coronavirus (BCoV), human coronavirus 229E (HCoV), murine hepatitis virus (MHV), and avian infectious bronchitis virus (IBV). Thus, our analyses substantiate the presence of recombination events in history that led to the SARS-CoV genome. Like the other coronaviruses used in the analysis, SARS-CoV is also a mosaic structure.

Introduction

SARS, a new disease characterized by high fever, malaise, rigor, headache and non-productive cough, has spread to over 30 countries with around 8% of mortality rate on average. Sequence analysis of SARS coronavirus (SARS-CoV) [17, 25] showed that it is a novel coronavirus [12]. Anand et al. [1] reported a three-dimensional model of SARS-CoV main proteinase and suggested that
modified rhinovirus 3C\textsuperscript{pro} inhibitors could be useful for SARS therapy. Lipsitch et al. [15] developed a mathematical model of SARS transmission to estimate the infectiousness of SARS and the likelihood of an outbreak. Ng et al. [22] suggested that SARS-CoV could have been derived from an innocuous virus or one causing a mild disease, that would become virulent after some mutational event occurring in some carriers. However, the source of SARS-CoV is not yet exactly known, although it has been reported that a virus highly related to SARS-CoV has infected some wild animals, such as masked palm civet, raccoon dog and badger [7].

Recombination, a key evolutionary process, accounts for a considerable amount of genetic diversity in natural populations. The occurrence of high-frequency homologous RNA recombination is one of the most intriguing aspects of coronavirus replication [14, 27, 31, 34]. The first experimental evidence for IBV recombination was found by Kottier et al. [11], although other studies have concluded that recombination is a feature of IBV evolution [4, 5, 10, 36–38]. Recombination in MHV was also experimentally demonstrated [16]. In particular, Snijder et al. [30] indicated that the recombination occurred between a coronavirus/torovirus-like virus and an influenza C-like virus, resulting in a line of coronaviruses that had a haemagglutinin esterase (HE) gene. This prompted us to explore the possible role of recombination in the emergence of SARS-CoV. A recent report indicated that SARS-CoV has been found in a number of wild animals with 99.8% identity [7]. What would be the role of recombination in the event that created this virus, possibly in a predator animal?

Stavrinides and Guttman [32] have suggested that a possible past recombination event between mammalian-like and avian-like parent viruses is responsible for the evolution of SARS-CoV. In order to further test for the recombination hypothesis, we implemented a 3-step strategy. First, we employed BLAST to determine which viruses (coronaviruses or other viruses) should be included in the sample relevant for recombination detection analysis. Second, we used widely used recombination detection techniques to detect the occurrence of recombination between SARS-CoV and other coronaviruses. Finally, we used phylogenetic tree analysis to confirm the presence of recombination events.

**Materials and methods**

**Sequences**

A reference SARS-CoV genome sequence (NC_004718) [17] was downloaded from GenBank. In order to determine which viruses (coronaviruses or other viruses) should be included in the sample relevant for recombination detection analysis, we blasted the whole SARS-CoV sequence against virus database and the result indicated that there are 6 significant hits (at the level of E-value <0.0001. Table 1): Murine hepatitis virus (MHV), Porcine epidemic diarrhea virus (PEDV), Bovine coronavirus (BCoV), Transmissible gastroenteritis virus (TGEV), Avian infectious bronchitis virus (IBV) and Human coronavirus 229E (HCoV). All these sequences were downloaded from GenBank: MHV (AF029248), PEDV (AF353511), BCoV (NC_003045), TGEV (NC_002306), IBV (NC_001451) and HCoV (NC_002645).