

Table 2. List of all single-nucleotide variations (SNVs) with multiple occurrences

Sequence alignments were generated by using CLUSTALW Ver. 1.83 with the default DNA weight matrix for the 96 severe acute respiratory syndrome-coronavirus (SARS-CoV) genomic sequences analyzed in this study (91 human SARS-CoV and 5 palm civet SARS-like-CoV sequences). The names of the genomic sequences (as from INSD/GenBank) are listed on the first column based on clusters determined by our scoring algorithm (1). All of these sequences are further divided by colored lines into subgroups (defined in the text and Fig. 1) based on their host species (human vs. palm civet) and epidemiological phases (year and early, middle, and late phases of the 2002/2003 epidemic).

Characteristics of the SNVs observed are listed in the top rows. From the top, they include the relevant coding DNA sequences (CDSs) and/or predicted ORFs, the variant nucleotides observed at the loci, the affected codon, the resultant amino acid switches, the amino acid coordinate of the CDSs/ORFs, and the nucleotide coordinate of the genome [based on GZ02 (1)], respectively.

All of the 202 SNVs that are observed in more than one of the 99 genomic sequences are listed. Except for the two variations located outside of the coding sequences, 128 out the 200 SNVs within the CDSs were predicted to cause amino acid changes (nonsynonymous variations). Different characters of the amino acid switches were labeled with different-color shadings in the corresponding row: light pink for synonymous variations, yellow for nonsynonymous variations causing amino acid switches in the same physical-chemical group, and greenish yellow for nonsynonymous variations causing drastic amino acid switches. In the rows for individual sequences, the minor forms of the SNVs were shaded with pink, the deletions were shaded with blue, and the undetermined nucleotides (N) were

shaded with yellowish pink.

SNVs that contribute to the grouping of genotypes based on the predominant clustering criteria (1) were further highlighted in the top characteristic rows with different-color shading and bolding of the letters:

1. The positions of the 5-nt motif used to classify the major genotypes of the 2002-2003 epidemic human SARS-CoV (HP03) are shaded as reported (1), including the two loci external to the *S* gene (17564 and 27827, shaded in grey) and the three loci in the *S* gene, namely positions 21721 (yellow), 22222 (blue), and 23823 (green). These variations are all emphasized with bold face.

2. The SNVs distinguishing the SARS-CoVs of the 2003-2004 epidemic (PC04 and HP04) from that of the 2002-2003 epidemic (PC03 and HP03) are shaded in light pink, of which those causing nonsynonymous variations are emphasized with bold face.

3. The SNVs distinguishing the SARS-CoV of the palm civet (PC03 and PC04) and human patient of the 2003-2004 epidemic (HP04) from human patients of the 2002-2003 epidemic (HP03) are shaded in light blue. Because all of them cause nonsynonymous variations, they are all emphasized with bold face.

4. A few special SNVs need illustration:

(i) SNVs in nucleotides 2759 and 2760 cause amino acid switches in the same residue, 832 of Orf1ab. Although they both cause nonsynonymous variations, one change is within the same physical-chemical group (W-F). This group of variation is shaded with dark greenish yellow.

(ii) The SNV in nucleotide 6295 causes residue 2011C of Orf1ab (nsp3) switching to a stop codon. This amino acid switching was presented in red and shaded with dark greenish

yellow.

(iii) SNVs in nucleotides 22927 and 22928 cause nonsynonymous amino acid switches in the same residue, 479 of S. This group of variations is emphasized with bold face and shaded with dark greenish yellow.

(iv) SNVs in nucleotides 23316 and 23317 cause the same nonsynonymous amino acid switches in the same residue, 609 of S. This group of variations is emphasized with bold face and shaded with dark greenish yellow.

(v) SNVs in nucleotides 23718 and 23719 cause nonsynonymous amino acid switches in the same residue, 743 of S. This group of variation is emphasized with bold face and shaded with dark greenish yellow.

1. The Chinese SARS molecular epidemiology consortium (2004) *Science* **303**, 1666-1669.

