Supplemental Data

Comparative study of synonymous codon usage variations
between the nucleocapsid and spike genes of coronavirus, and
C-type lectin domain genes of human and mouse

Insung Ahn¹*, Byeong-Jin Jeong²,³* and Hyeon Seok Son²,³,⁴

Methods
Nucleotide sequences
The nucleocapsid (251 sequences) and spike (284 sequences) genes of Coronavirus genus including SARS-CoVs were collected from the NCBI Taxonomy Browser (www.ncbi.nlm.nih.gov/Taxonomy/) in GenBank format, and then, all the GenBank flat files were parsed into each category such as accession number, species name, gene name and sequence length using JAVA codes to construct a local database to facilitate the further computational works. As for the human and mouse species, we collected the coding sequences of human (homo sapiens) and mouse (mus musculus) from the genome section of NCBI’s FTP site (ftp.ncbi.nih.gov/genomes/), and also parsed and constructed a local database. All the gene names, abbreviations, sequence lengths and their GenBank accession numbers of CTLD genes used in this study are shown in Supplemental Data. Abnormal sequences which include unknown characters except for A, G, C or U were not divided by three - maximum length of a codon unit -were removed. MySQL database management system was used to construct all the local databases on Linux operating system.
Table S1. Gene names, abbreviations, lengths, and their accession numbers of human (*homo sapiens*) and mouse (*mus musculus*) C-type lectin domain genes

| Gene name                      | Abbreviation | Length (bp) | GenBank accession # |
|--------------------------------|--------------|-------------|---------------------|
| Human                          |              |             |                     |
| family 1, member a             | clec1A       | 1791        | NM_016511           |
| family 2, member b             | clec2B       | 2052        | NM_005127           |
| family 4, member a (transcript variant 1) | clec4A_1   | 1284        | NM_016184.3         |
| family 4, member a (transcript variant 2) | clec4A_2   | 1185        | NM_194450.2         |
| family 4, member a (transcript variant 3) | clec4A_3   | 1167        | NM_194447.2         |
| family 4, member a (transcript variant 4) | clec4A_4   | 1068        | NM_194448.2         |
| family 4, member c (transcript variant 1) | clec4C_1   | 1314        | NM_130441           |
| family 4, member c (transcript variant 2) | clec4C_2   | 1221        | NM_203503           |
| family 4, member m (transcript variant 4) | clec4M_4   | 2349        | NM_214677           |
| family 4, member m (transcript variant 5) | clec4M_5   | 2265        | NM_214678           |
| family 4, member m (transcript variant 6) | clec4M_6   | 2202        | NM_214679           |
| family 7, member a (transcript variant 1) | clec7A_1   | 2622        | NM_197947           |
| family 7, member a (transcript variant 2) | clec7A_2   | 2484        | NM_022570           |
| family 7, member a (transcript variant 5) | clec7A_5   | 2385        | NM_197950           |
| family 10, member a (transcript variant 1) | clec10A_1  | 1788        | NM_182906           |
| family 10, member a (transcript variant 2) | clec10A_2  | 1716        | NM_006344           |
| family 14, member a             | clec14A     | 2256        | NM_175060           |
| family 16, member a             | clec16A     | 6786        | NM_015226           |
| Mouse                          |              |             |                     |
| family 3, member b             | clec3B       | 1083        | NM_011606           |
| family 1, member b             | clec1B       | 1134        | NM_019985           |
| family 2, member d             | clec2D       | 1206        | NM_053109           |
| family 4, member n             | clec4N       | 1227        | NM_020001           |
| family 4, member g             | clec4G       | 1230        | NM_029465           |
| family 4, member d             | clec4D       | 1323        | NM_010819           |
| family 2, member h             | clec2H       | 2157        | NM_053165           |
| family 2, member e             | clec2E       | 2268        | NM_153506           |
| family 4, member f             | clec4F       | 2400        | NM_016751           |
| family 3, member a2            | clec4A2      | 2403        | NM_011999           |
| family 14, member a             | clec14A     | 4167        | NM_025809           |