# Wuhan-coronavirus homologue map

Kou Amano

## Background and Objective

**Background:** At the beginning of 2020, global risk of infection of a new coronavirus is spreaded. The pandemic started in 2019 and governments announced a state of emergency. In Japan, the government adopted PCR as a diagnosis method of the infection. But the selection of primers influences the accuracy greatly.
**Objective:** Therefor, I provide a "map" of homological regions of coronavirus genome to other virus and animal genomes to help the primer design.

## Data

### Vertebrates

- Bat (NW_017738920.1 .. NW_017802358.1; 63439frgs)
- Beluga (NW_022097992.1 .. NW_022103895.1; 5904frgs)
- Camel (NC_044511.1 .. NC_044547.1; 37chrs)
- Cat (NC_018723.3 .. NC_018741.3; 19chrs)
- Dog (NC_006583.3 .. NC_006621.3; 39chrs)
- Ferret (NW_004569142.1 .. NW_004576923.1; 7782frgs)
- Human (NC_000001.4 .. NC_000024.3; 24chrs)
- Mouse (NC_000067.6 .. NC_000087.7; 21chrs)
- Rabbit (NC_013669.1 .. NC_013690.1; 22chrs)
- Pig (NC_010443.5 .. NC_010462.3; 20chrs)
- Rock Pigeon (NW_004973171.1 .. NW_004988092.1; 14922frgs)
- Turkey (NC_015011.2 .. NC_015042.2; 32chrs)

### Viruses

- 7554 genomes of viruses or phages
- Wuhan-corona virus genome (MN908947.3)

## Method

### BLAST

DB: `makeblastdb -in <<input file>> -out <<DB name>> -dbtype nucl -parse_seqids`
Query: `megablast -d <<DB name>> -i <<query sequence>> -W 10`

### Self-BLAST

Fragmentation: `fragment bf=<<input file>> S=25 G=25 cs=1`
DB: same as above.
Query: same as above.

### Window-fourier

Fragmentation: 30 fragments; 1000 bases / fragment
Conversion: `"A" -> 1, "T" -> -1, "G" -> I, "C" -> -I`
Fourier transform: $Abs(Ft($`<<each fragment>>`$))$

### Selection of frequent homologues

Vertebrates: `base-hit count >= 5; seq-length >= 5`
Viruses: `base-hit count >= 18; seq-length >= 5`

## Result

### The homologues

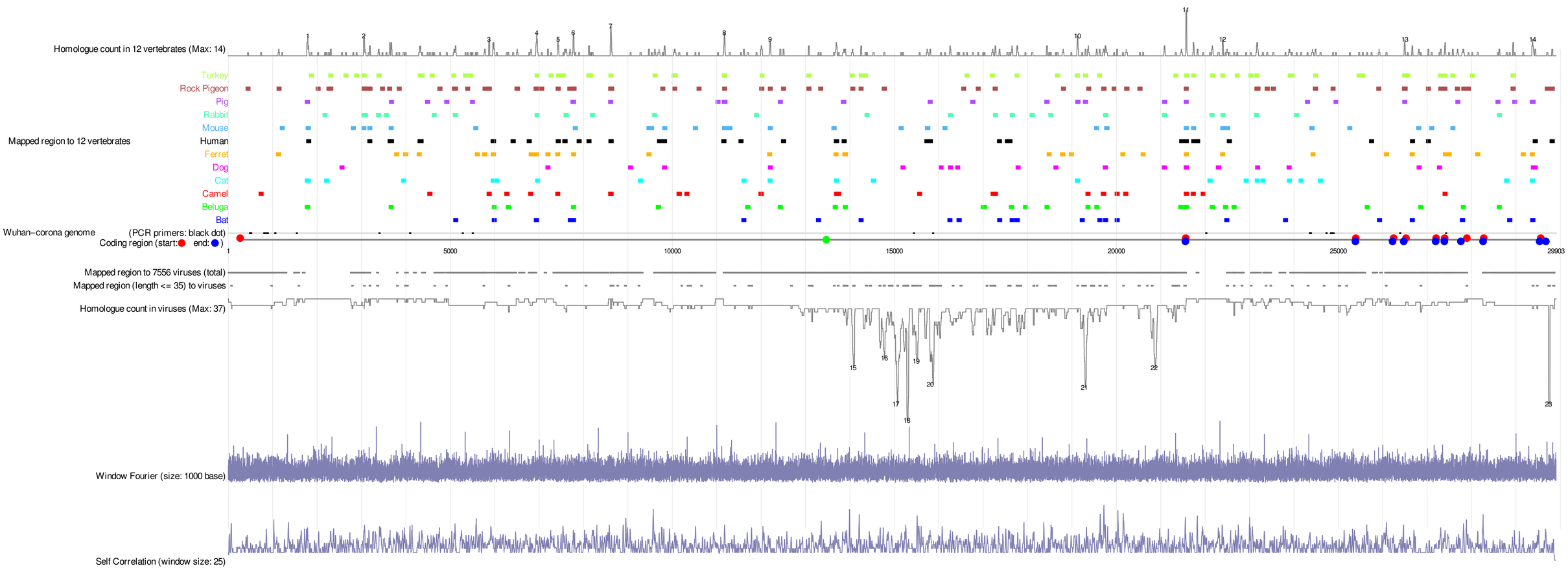
#### Vertebrates

| No. | homologue | start | end | count | in CDS |
|---|---|---|---|---|---|
| 1 | AATTTTA | 1787 | 1793 | 6 | Yes |
| 2 | GATGAGGATGAAGAAGAAGGT | 3047 | 3067 | 6 | Yes |
| 3 | CTACAAAGAAAACAGTTAC | 5866 | 5884 | 5 | Yes |
| 4 | ATAAATATTATAATTTG | 6944 | 6960 | 7 | Yes |
| 5 | TTGCAT | 7425 | 7430 | 5 | Yes |
| 6 | CCATCCATCTTTACTTTGATAAA | 7764 | 7786 | 7 | Yes |
| 7 | TTTTTGTTGCTGCTATTTTCTATTTA | 8607 | 8632 | 9 | Yes |
| 8 | ATTTCTCTGTTTGTTTTTGT | 11 167 | 11 186 | 7 | Yes |
| 9 | AAAAGT | 12 198 | 12 203 | 5 | Yes |
| 10 | ATAAAATAGAAGAA | 19 121 | 19 134 | 6 | Yes |
| 11 | TGTTTGTTTTTCTTGTTTTATTGC | 21 564 | 21 587 | 14 | Yes |
| 12 | TATTAAAATATAATGAA | 22 389 | 22 405 | 5 | Yes |
| 13 | TTTTCT | 26 495 | 26 500 | 5 | No |
| 14 | CCTAA | 29 375 | 29 379 | 5 | Yes |

#### Viruses

| No. | homologue | start | end | count | in CDS |
|---|---|---|---|---|---|
| 15 | CAAGATCTCAATGGTAACTGGTATGATTTCGGTGATT | 14 068 | 14 104 | 21 | Yes |
| 16 | TGGTAATGCTGC | 14 775 | 14 786 | 18 | Yes |
| 17 | ACAAAACGTAATGTCATCCCTACTATAACTCAAATGAATCTTAAGTATGCCATTAGTGCAAAGAATAGAGCTCGCACCGTAGCTGGTGTCTCTAT | 15 031 | 15 125 | 32 | Yes |
| 18 | CTTATGGGTTGGGATTATCCTAAATGTGATAGAGCCATGCCTAA | 15 280 | 15 323 | 37 | Yes |
| 19 | GATGCCACAACTGCTTATGCTAATAGTGTTTTTAACAT | 15 490 | 15 527 | 19 | Yes |
| 20 | CAAAACAATGTTTTTATGTCTGAAGCAAAATGTTGGACTGAGACTGACCTTACTAAAGGACCTCATGAATTTTGCTCTCAACATACA | 15 805 | 15 891 | 26 | Yes |
| 21 | GGTTGTGATGGTGGCAGTTTGTATGTAAATAAACATGCATTCCACACACC | 19 276 | 19 325 | 27 | Yes |
| 22 | TATGAGAGTTATACATTTTGGTGCTGGTTCTGATAAAGGAGTTGCACCAGGTAC | 20 850 | 20 903 | 21 | Yes |
| 23 | CCGAGGCCACGCGGAGTACGATCGAGTGTACAG | 29 732 | 29 764 | 32 | No |

### The map




References are in the proceedings.
JSIMS 2020-07-04 on web meeting.