# 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)

#### Viruses

• 7554 genomes of viruses or phages (FTP bulk download)

- 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)
- Wuhan-corona virus genome (MN908947.3)
- 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)
- Primer (https://www.niid.go.jp/niid/images/lab-manual/2019-nCoV20200217.pdf)

## 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

homologue

#### The homologues

Vertebrates			
start	end	count	in CDS

2	GATGAGGATGAAGAAGAAGGT	3047	3067	6	Yes						
3	CTACAAAGAAAACAGTTAC	5866	5884	5	Yes						
4	ATAAATATTATAATTTG	6944	6960	7	Yes						
5	TTGCAT	7425	7430	5	Yes	No.	homologue	start	end	count	in CDS
6	CCATCCATCTTTACTTTGATAAA	7764	7786	7	Yes	15	CAAGATCTCAATGGTAACTGGTATTCGGTGATT	14 068	14 104	21	Yes
7	TTTTTGTTGCTGCTATTTTCTATTTA	8607	8632	9	Yes	16	TGGTAATGCTGC	14 775	14786	18	Yes
8	ATTTCTCTGTTTGTTTTTGT	11 167	11 186	7	Yes	17	ACAAAACGTAATGTCATCCCTACTATAACTCAAATGAATCTTAAGTATGCCATTAGTGCAAAGAATAGAGCTCGCACCGTAGCTGGTGTCTCTAT	15 031	15 125	32	Yes
9	AAAAGT	12 198	12 203	5	Yes	18	CTTATGGGTTGGGATTATCCTAAATGTGATAGAGCCATGCCTAA	15 280	15 323	37	Yes
10	ATAAAATAGAAGAA	19 121	19 134	6	Yes	19	GATGCCACAACTGCTTATGCTAATAGTGTTTTTAACAT	15 490	15 527	19	Yes
11	TGTTTGTTTTTCTTGTTTTATTGC	21564	21 587	14	Yes	20	CAAAACAATGTTTTTATGTCTGAAGCAAAATGTTGGACTGAGACTGACCTTACTAAAGGACCTCATGAATTTTTGCTCTCAACATACA	15 805	15 891	26	Yes
12	TATTAAAATATAATGAA	22 389	22 405	5	Yes	21	GGTTGTGATGGCAGTTTGTATGTAAATAAACATGCATTCCACACACC	19 276	19325	27	Yes
13	TTTTCT	26 495	26 500	5	No	22	TATGAGAGTTATACATTTTGGTGCTGGTTCTGATAAAGGAGTTGCACCAGGTAC	20 850	20 903	21	Yes
14	CCTAA	29 375	29 379	5	Yes	23	CCGAGGCCACGCGGAGTACGAGTGTACAG	29 732	29 764	32	No

## The map

