Wuhan-coronavirus homologue map

Kou Amano

Background and Objective

Background: At the beginning of 2020, global risk of infection of coronavirus is spreded. 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 prmers infruence the accuracy greatly. Objective: Therefor, I provide a "map" of homological region of coronavirus to other viruses and annimals to help the primer design.

Data

Vertebrates:

- Bat (NW_017738920.1 .. NW_017802358.1; 63439frgs)
 Dog (NC_006583.3 .. NC_006621.3; 39chrs)
- Beluga (NW_022097992.1 .. NW_022103895.1; 5904frgs)

start

1787

3047

end

1793

3067

- Camel (NC_044511.1 .. NC_044547.1; 37chrs)
- Cat (NC_018723.3 .. NC_018741.3; 19chrs) Viruses:
- 7554 genomes of viruses or phages
- 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)

Method

BLAST

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

in CDS

Yes

Yes

Self-BLAST

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

count

Window-fourrier

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

Selection of frequent homologues

Vertebrate: base-hit count >= 5; seq-lenght >= 5 Viruses: base-hit count >= 18; seq-lenght >= 5

Result

homologue

GATGAGGATGAAGAAGAAGGT

AATTTTA

The homologues

		5884	5	Yes						
ATAATTTG	6944	6960	7	Yes						
	7425	7430	5	Yes	No.	homologue	start	end	count	in CDS
TTTACTTTGATAAA	7764	7786	7	Yes	15	CAAGATCTCAATGGTAACTGGTATTCGGTGATT	14 068	14 104	21	Yes
CTGCTATTTTCTATTTA	8607	8632	9	Yes	16	TGGTAATGCTGC	14775	14786	18	Yes
TTTGTTTTTGT	11 167	11 186	7	Yes	17	ACAAAACGTAATGTCATCCCTACTATAACTCAAATGAATCTTAAGTATGCCATTAGTGCAAAGAATAGAGCTCGCACCGTAGCTGGTGTCTCTAT	15 031	15 125	32	Yes
	12 198	12 203	5	Yes	18	CTTATGGGTTGGGATTATCCTAAATGTGATAGAGCCATGCCTAA	15 280	15 323	37	Yes
AAGAA	19 121	19 134	6	Yes	19	GATGCCACAACTGCTTATGCTAATAGTGTTTTTAACAT	15 490	15 527	19	Yes
TTCTTGTTTTATTGC	21564	21587	14	Yes	20	CAAAACAATGTTTTTATGTCTGAAGCAAAATGTTGGACTGAGACTGACCTTACTAAAGGACCTCATGAATTTTGCTCTCAACATACA	15 805	15 891	26	Yes
ATAATGAA	22 389	22 405	5	Yes	21	GGTTGTGATGGCAGTTTGTATGTAAATAAACATGCATTCCACACACC	19 276	19325	27	Yes
	26 495	26 500	5	No	22	TATGAGAGTTATACATTTTGGTGCTGGTTCTGATAAAGGAGTTGCACCAGGTAC	20 850	20 903	21	Yes
	29 375	29 379	5	Yes	23	CCGAGGCCACGCGGAGTACGAGTGTACAG	29732	29 764	32	No
T T	TTACTTTGATAAA TGCTATTTTCTATTTA TTGTTTTTGT AGAA TCTTGTTTTATTGC	7425 TTACTTTGATAAA 7764 TGCTATTTTCTATTTA 8607 TTGTTTTTGT 11 167 12 198 AGAA 19 121 TCTTGTTTTATTGC 21 564 TAATGAA 22 389 26 495	7425 7430 TTACTTTGATAAA 7764 7786 TGCTATTTTCTATTTA 8607 8632 TTGTTTTTGT 11 167 11 186 12 198 12 203 AGAA 19 121 19 134 TCTTGTTTTATTGC 21 564 21 587 TAATGAA 22 389 22 405 26 495 26 500	7425 7430 5 TTACTTTGATAAA 7764 7786 7 TGCTATTTTCTATTTA 8607 8632 9 TTGTTTTTGT 11 167 11 186 7 12 198 12 203 5 AGAA 19 121 19 134 6 TCTTGTTTTATTGC 21 564 21 587 14 TAATGAA 22 389 22 405 5 26 495 26 500 5	7425 7430 5 Yes TTACTTTGATAAA 7764 7786 7 Yes TGCTATTTTCTATTTA 8607 8632 9 Yes TTGTTTTTGT 11 167 11 186 7 Yes 12 198 12 203 5 Yes AGAA 19 121 19 134 6 Yes TCTTGTTTTATTGC 21 564 21 587 14 Yes TAATGAA 22 389 22 405 5 Yes 26 495 26 500 5 No	TTACTTTGATAAA 7764 7786 7 Yes 15 TGCTATTTTCTATTTA 8607 8632 9 Yes 16 TTGTTTTTGT 11167 11186 7 Yes 17 12198 12203 5 Yes 18 AGAA 19121 19134 6 Yes 19 TCTTGTTTTATTGC 21564 21587 14 Yes 20 TAATGAA 22389 22405 5 Yes 21 26495 26500 5 No 22	No. homologue TTACTTTGATAAA 7764 7786 7 Yes 15 CAAGATCTCAATGGTAACTGGTATGATTTCGGTGATT TGCTATTTTCTATTTA 8607 8632 9 Yes 16 TGGTAATGCTGC TTGTTTTTGT 11 167 11 186 7 Yes 17 ACAAAACGTAATGTCATCCCTACTATAACTCAAATGAATCTTAAGTATGCCATTAGTGCAAAGAATAGAGCTCGCACCGTAGCTGGTGTCTCTAT 12 198 12 203 5 Yes 18 CTTATGGGTTGGGATTATCCTAAATGGTAAGGCCATGCCTAA AGAA 19 121 19 134 6 Yes 19 GATGCCACAACTGCTTATGCTAAATGGTTTTTAACAT TCTTGTTTTATTGC 21 564 21 587 14 Yes 20 CAAAACAATGTTTTTATGCTGAAGCAAAAATGTTGGACTGAGCCTTACTAAAGGACCTCATGAATTTTGCTCTCAACATACA TAATGAA 22 389 22 405 5 Yes 21 GGTTGTGATGGGAGTTTGTATGTAAAAACATGCATTCCACACACC 26 495 26 500 5 No 22 TATGAGAGTTATACATTTTGGTGCTGGTTCTGATAAAGGAGTTGCACCAGGTAC	7425 7430 5 Yes No. homologue start TTACTTTGATAAA 7764 7786 7 Yes 15 CAAGATCTCAATGGTAACTGGTATTTCGGTGATT 14 068 TGCTATTTTCATTTA 8607 8632 9 Yes 16 TGGTAATGCTGC 14 775 TTGTTTTTGT 11 167 11 186 7 Yes 17 ACAAAACGTAATGTCATCCCTACTATAACTCAAATGGAATCTTAAGTAGCCATTAGTGCAAAGAATAGAGCTCGCACCGTAGCTGGTGTCTCTAT 15 031 AGAA 19 121 19 134 6 Yes 19 GATGCCACAACTGCTTATGCTAATAGTGTTTTTAACAT 15 490 TCTTGTTTTATTGC 21 564 21 587 14 Yes 20 CAAAACAATGTTTTTATGCTGAAGCAAAATGTTGGACTGAGACTTACTAAAGGACCTCATGAATTTTGCTCAACATACA 15 805 TAATGAA 22 389 22 405 5 Yes 21 GGTTGTGATGGTGGCAGTTTGTATGTAAATAAACATGCATTCCACACAC 20 850 TAAGGAGTTATACATTTTGGTGCTGGTTCTGATAAAGGAGTTGCACCAGGTAC 20 850 20 850 20 850	7425 7430 5 Yes No. homologue start end TTACTTTGATAAA 7764 7786 7 Yes 15 CAAGATCTCAATGGTATGATTTCGGTGATT 14068 14104 TGCTATTTTCTATTTA 8607 8632 9 Yes 16 TGGTAATGCTGC 14775 14786 TTGTTTTTGT 11167 11186 7 Yes 17 ACAAAACGTAATGTCCTACATTAACTCAAATGGATCGCATTAGTGCAAAGAATAGAGCTCGCACCGTAGCTGGTGTCTCTAT 15 031 15 125 AGAA 19 121 19 134 6 Yes 19 GATGCCACAACTGCTTATGCTAATAGTGTTTTTAACAT 15 0323 15 290 15 891 TCTTGTTTTATTGC 21 564 21 587 14 Yes 20 CAAAACAATGTTTTATGTCTGAAGCAAAATGTTGGACTGAGACTTACTAAAGGACCTCATGAATTTTGCTCTCAACATACA 15 805 15 891 TAATGAA 22 389 22 405 5 Yes 21 GGTTGTGATGGTGGCAGTTTGTATGTAAATAAACATGCATTCCACACACC 20 803 20 903	No. No.

The plot

