

Wuhan-coronavirus homologue map

Kou Amano

Background and Objective

Background: At the beginning of 2020, global risk of infection of 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 influence the accuracy greatly.

Objective: Therefor, I provide a "map" of homological region of coronavirus to other viruses and animals to help the primer design.

Data

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

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

Window-fourrier
Fragmentation: 30 fragments – 1000 bases
Conversion: "A" -> 1, "T" -> -1, "G" -> I, "C" -> -I
Fourier transform: $Ft(<<each\ fragment>>)$

Selection of frequent homologues
Vertebrate: `base-hit count >= 5; seq-lenght >= 5`
Viruses: `base-hit count >= 18; seq-lenght >= 5`

The plot

