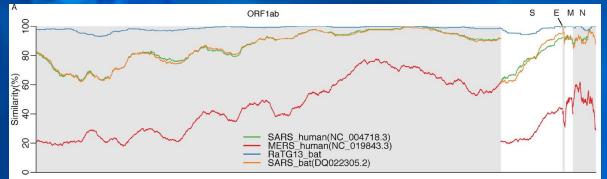
Using RNA Secondary Structure to Analyze Variations in the S (spike) Protein of SARS-CoV-2 and related Coronaviruses



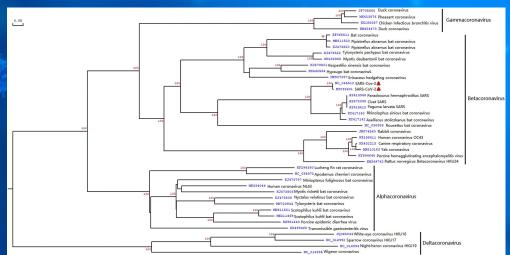
1. Brief Reminder

RNA Secondary Structure, Coronaviruses, S (spike) Protein

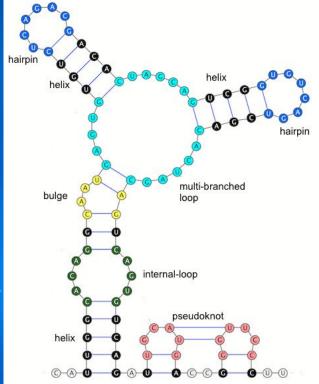




Reference: Wen, Feng, et al. "Identification of the Hyper-Variable Genomic Hotspot for the Novel Coronavirus SARS-CoV-2." *Journal of Infection*, 4 Mar. 2020, doi:10.1016/j.jinf.2020.02.027.



Reference: Li, Chun, et al. "Genetic Evolution Analysis of 2019 Novel Coronavirus and Coronavirus from Other Species." *Infection, Genetics and Evolution*, vol. 82, 10 Mar. 2020, p. 104285., doi:10.1016/j.meegid.2020.104285.



Reference: Mamuye, Adane & Merelli, Emanuela & Tesei, Luca. (2016). A Graph Grammar for Modelling RNA Folding. Electronic Proceedings in Theoretical Computer Science. 231. 31-41. 10.4204/EPTCS. 231.3.

Task / Questions



- Can we link mutation rates / the mutations we see in the spike protein regions across the coronavirus genome with what we see in the corresponding secondary structures?
- What about across the SARS-CoV-2 genome?
- From any links we can establish between variation in the primary and secondary structures, what do they tell us about selection processes?
- Do structural insights help at all with developing a vaccine to exploit potential structural weaknesses?



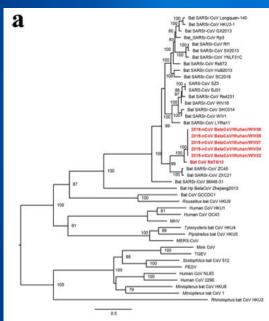


Sequence Selection, Alignment, Software Selection, Output



Sequence Selection across beta CoV genus



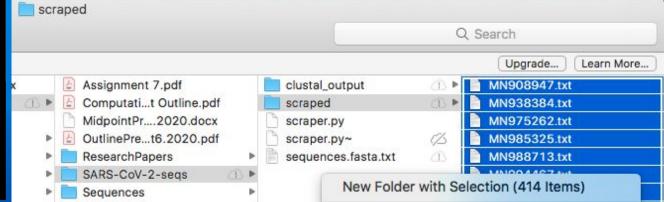


Reference: Zhou, Peng, et al. "A Pneumonia Outbreak Associated with a New Coronavirus of Probable Bat Origin." *Nature News*, Nature Publishing Group, 3 Feb. 2020, www.nature.com/articles/s41586-020-2012-7.

- After reviewing the literature, I have 5 beta coronaviruses I am comparing across the beta coronavirus genus.
 - MN996532.1 (RaTG13)
 - NC_004718.3 (SARS)
 - NC_014470.1 (BM48-31 Bat CoV)
 - NC_019843.3 (MERS)
 - ▶ NC_045512.2 (SARS-CoV-2)
- May add more pending results / literature / availability.
 - e.g. closest pangolin strain, etc.

Sequence Selection within SARS-CoV-2 species

- Wrote a python script that scraped 413 nucleotide sequences that completely encode the S protein, each from a unique case of SARS-CoV-2. With the SARS-CoV-2 reference genome included, this yields 414 S proteins to examine total.
- May add more if I think it's worth it.



Sequence Alignment



percent_identity_mtx.pim.txt

multiple sequence alignment.clustal_num.txt CLUSTAL O(1.2.4) multiple sequence alignment NC 019843.3:21456-25517 ATGATACACTCAGTGTTTCTAC-TGATGTTCTTGTTAACACC-----TACAGAAAG 42 ----ATGT---TTGTTTTT-C-TTGTTTTATTGCCACTAG-----TTTCTAG--TCAG MN996532.1:21545-25354 NC 045512.2:21563-25384 ----ATGT---TTGTTTTT-C-TTGTTTTATTGCCACTAG-----TCTCTAG--TCAG NC 004718.3:21492-25259 ----ATGT-----TTATTTTCTTATTTTCTTACTCTCACTAGTGGTAG 41 NC 014470.1:21391-25170 ATGAAATTT---TTGGCTTTTCTCTGTCTTCTTGGCTTTGCTAACGCTCAAGATGGCAAG 57 NC_019843.3:21456-25517 105 MN996532.1:21545-25354 ----TGTTAATCTAA--CAACTAGAACTCAGTT-----A 72 72 NC 045512.2:21563-25384 -TGTTAATCTTA--CAACCAGAACTCAATT----A NC_004718.3:21492-25259 -CGGTGCACCACTTTTGATGATGTTCAAGC----T 81 NC 014470.1:21391-25170 -TGGTACACTAT--CTAATAAAAGTC---C----A NC 019843.3:21456-25517 CAACAGACTTTCTTTGATAAAACTTGGCCTAGGCCAATTGATGTTTCTAAGGCTGACGGT MN996532.1:21545-25354 CCTC---C--TGCATACACCAACTCATCCA---CCCGTGGTGTCTATTACCCTGACAAA 123 NC_045512.2:21563-25384 CCCC---C--TGCATACACTAATTCTTTCA---CACGTGGTGTTTATTACCCTGACAAA 123 NC 004718.3:21492-25259 CCTAATTA---CACTCAACATACTTCATCTA---TGAGGGGGGTTTACTATCCTGATGAA 135 NC 014470.1:21391-25170 TCTAAGCT---TACTCAGACTCCTTCTTCTA---GGAGGGGTTTTTATTATTTTGATGAC 138 NC 019843.3:21456-25517 ATTAT-ATACCCTCAAGGCCGTACATATTCTAACATAACTATCACTTATCAAGGTCTTTT 224 MN996532.1:21545-25354 166 GTTTTCAGATCTTCA-----GTTTTACA-TTTAACTCAGG-----ATTTGTTTT NC 045512.2:21563-25384 GTTTTCAGATCCTCA-----GTTTTACA-TTCAACTCAGG-----ACTTGTTCT 166 178 NC 004718.3:21492-25259 ATTTTTAGATCAGAC-----ACTCTTTA-TTTAACTCAGG-----ATTTATTTC NC_014470.1:21391-25170 ATTTTTAGGTCTTCA-----ATTCGTGTGCTTACCACTGG-----CCATTTTC 181 NC_019843.3:21456-25517 TCCCTATCAGGGAGACCATGGTGATATGTATGTTTACTCTGCAGGACATGCTACAGGCAC MN996532.1:21545-25354 TACCTTTCT---TCTCCAATGTGACCTGGTTCCA----TGCTATACATGTTTCAGGGAC NC 045512.2:21563-25384 TACCTTTCT---TTTCCAATGTTACTTGGTTCCA----TGCTATACATGTCTCTGGGAC 218 NC_004718.3:21492-25259 TTCCATTTT---ATTCTAATGTTACAGGGTTTCA----TACTATT-----216 NC 014470.1:21391-25170 TTCCTTTTA---ATACTAACCTTACTTGGTATTT----GACTTTAAA-----GTC

Percent Identity Matrix - created by Clustal2.1 1: NC 019843.3 21456-25517 100.00 53.01 52.64 51.69 51.79 2: MN996532.1_21545-25354 53.01 100.00 93.15 73.89 70.24 3: NC 045512.2 21563-25384 52.64 93.15 100.00 73.97 4: NC 004718.3 21492-25259 51.69 73.89 73.97 100.00 71.79 5: NC_014470.1_21391-25170 51.79 70.24 70.40 71.79 100.00



Left: Screenshot of the files output by the Clustal Omega algorithm for multiple sequence alignment.

Bottom: Screenshot of Clustal Omega algorithm's webserver: https://www.ebi.ac.uk/Tools/msa/clustalo/

Clustal Omega

Input form Web services Help & Documentation

Tools > Multiple Sequence Alignment > Clustal Omega

Multiple Sequence Alignment

Clustal Omega is a new multiple sequence alignment program that uses seeded guide trees and HMM profile-profile ted or more sequences. For the alignment of two sequences please instead use our <u>pairwise sequence alignment tools</u>.

Bioinformatics Tools FAQ

Important note: This tool can align up to 4000 sequences or a maximum file size of 4 MB.

Software Selection: Vienna RNA Suite





Font size: A A A

Theoretical Biochemistry Group

Institute for Theoretical Chemistry

■ TBI ■ RNA Software ▼ ■ ViennaRNA Package ■ Documentation ■ Tutorial ■ Changelog

You are here: TBI / Software / ViennaRNA Package

The ViennaRNA Package

The ViennaRNA Package consists of a C code library and several stand-alone programs for the prediction and comparison of RNA secondary structures.

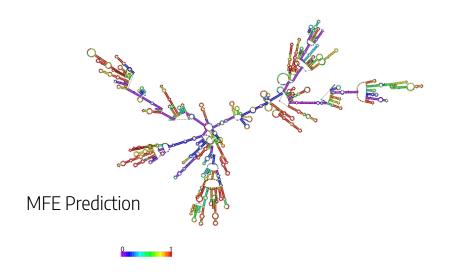
RNA secondary structure prediction through energy minimization is the most used function in the package. We provide three kinds of dynamic programming algorithms for structure prediction: the minimum free energy algorithm of (Zuker & Stiegler 1981) which yields a single optimal structure, the partition function algorithm of (McCaskill 1990) which calculates base pair probabilities in the thermodynamic ensemble, and the suboptimal folding algorithm of (Wuchty et.al 1999) which generates all suboptimal structures within a given energy range of the optimal energy. For secondary structure comparison, the package contains several measures of distance (dissimilarities) using either string alignment or tree-editing (Shapiro & Zhang 1990). Finally, we provide an algorithm to design sequences with a predefined structure (inverse folding).

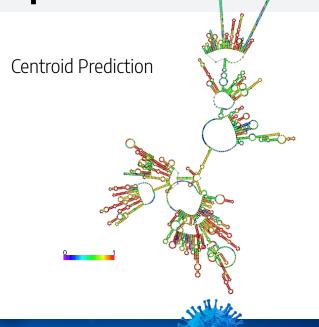
Table of Contents

- About
- News
- Download ▼
- Performance ▼
- Other Packages
- Older Versions
- Comments and Bugs

Screenshot of ViennaRNA package homepage: https://www.tbi.univie.ac.at/R NA/index.html

Vienna RNAfold: Example Output









Old Timeline





Determine and Gather all Kelevant January

By 4/12

Run Software on Seqs.

By 4/18

Synthesize it all + write up findings

By 4/30

By 4/15

Determine Retoyar Software By 4/26

Write + run code to organize + analyse output

4. Plan Going Forward



Gather More Output from Software, Analyse Findings w/code



5. Deliberation



RNAfold gives me a few files that I can manipulate...

Results for minimum free energy prediction

Downloaded output from Vienna RNAfold's webserver: http://rna.tbi.univie.ac.at/cqi-bin/RNAWebSuite/RNAfold.cqi

The optimal secondary structure in <u>dot-bracket notation</u> with a minimum free energy of -1072.20 kcal/mol is given below. [color by base-pairing probability | color by positional entropy | no coloring]

```
160
                        320
480
640
             GAUCUCCCUCAGGGUUUUUUCGCCUUUAGAACCAUUGGUAGAUUUGCCAAUAGGUAUUAACAUCACUAGGUUUCAAACUUUACUUGCUUUACAUAGAAGUUAUUUGACUCCUGGUGAUUCUUCUUCAGGUUGACAGCUGG
800
                                 AGGACUUUUUCUAUUAAAAUAUAAUGAAAUGGAACCAUUACAGAUGUGAGACUGUGCACUUGACCUUUGACCUUCAGAAACAAGUGUACGUUGAAAUCCUUCAAUGUAAAAAAGGAAUCUAUCAAACUUCUAACUUUAAGUUUAAGGAAU
            ARCCARC REARICUATURUUAGAUUUCCUAAUAUUACAAACUUGUGCCCUUUUGGUGAACUUUUAACGCCACAGAUUUGCAUCUGUUUAUGCUUGGAACAGGAAUCAGCAACUGUGUUGCUGAUUA
1120
                                            ugucuccuacuaaauuaaaugaucucugcuuu<mark>ac</mark>uaau<mark>gu</mark>cuaugcagauucauuuguaauuagggggaugaagucagac<del>aaa</del>ucgcuccaggcaaacuggaaaga<mark>uu</mark>gcugauua<u>uaaauuauaaauuac</u>
1280
                    1440
            UAAAAACAAAUGUGUGAAUUUCAACUUCAAUGGUUUAACAGGCACAGGUGUUCUUACUGAGUCUAACAAAAAGUUUUCGACCAACAAUUUGGCAGAGACA<mark>UUGCUGACACUACUGAUGC</mark>UGUCGUGAUCCACAGACAC
1600
                         CUUTUGGUGGUGUALULAA LULAA LULA
          AAUGUUUUUCAAACACGUIGCAGGCUIGUUUAAUAGGGCCUGAACAUGUCAACAACUCAUAUGAGGGUIGACAUAGCCAUUGGUGCAGGUAUAUGCGCUAGUUAUCAGACUCAGACUAAUUCUCCUCGGCGGCACGUAGUUGUAGUUAG
1920
            CUACACUATGIC<mark>ACUAGAGAAAAUUCAGUUGCUUACUCUAAUAACUCUA</mark>UUGCAUACCOCAAAAUUUUACUAUUAGUGUUACCACAGAAAUUCUACCAGAGACUUCAGAGACAUCAGUAGAUUGAGAUGUACAAUGUACAUUUGUGGGGAU
GAAUGCAGCAAUCUUUUGUUGCAAUAUGGCAGUUUUUGUACACAAUUAAACGGUGCUUUAACUGGAAUAGCUGUAACAAGACAAAAAACACCAAGAAAUUUUGCAGAAGG
2080
2240
             UUUUUCACA AUAUUACCAGAUCEAUCAAAACCAAGCAAGAGCUCAUUU AUUGAAGAUCUACUUUUCAACA AGUGACACUUGCAGAUCUGCUGGCUUCAUCAAAACAAUAUGGUGAUUGCCUUGGUGAUAUUGCUGAGACACAUAUGGUGAUAUAUGGUGAUAUUGCUGAGACACAUAUUGGUGACA
2400
              2560
           SCHAUTICGACHUACACAGAAUGUUCUCUAUGAAACCAAAAA UUGAUUCCCAACCAAAAU CAAAAUGUCACAAAAUUCAACAUUUAAA
2720
           2880
3040
3200
          HIGHCCCHIGCACAAGAAAGAACHHCACAACHGCHCCHGCCAHHHIGHCAHGAHGGAAAAGCACAHHHICCHCGHGAAGGHGHCHHHIGH
              GUGUCUGGUAACUGUGAUGUUGUAAUAGGAAUUGUCAACAACACGUUUAUGAUCCUUUGCAACUGAUUAGACUCAUUCAAGGAGGAGUUAGAUUAAAUAUUUU
3360
                 GUUGUAAACAUUGAAAAAAAAAAAUUGACCCCCCXAAUGAGGUUGCCAAGAAUUUUAAAUGAACUCUCCCAUGGUUCGAAGAACUUGGAAAGUAUAUGAGCAGUAUAUAAAAAUGCCCAUGGUACAUUUGGCCUAGGUAUAUAACUCGCCUUGAUU
3520
                       GACAAUUAUGCUUUGCUGUAUGACCAGUUGCUGUAGUUGUCUCAAGGGCUGUUGUUCUUGUGGAGACUGCUGCAAAUU<mark>UGAUGAAG</mark>ACGACUCUGAGCCCAGUGCUCAAAGGAGUCAAAUUACAUUACACAUAA
3680
1
           0 \in \{0, \dots, 0\} \in \{0, \dots, 0\} \cap 
160
          320
480
          640
          800
960
         ((0,0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0),(
1120
1280
         1440
1600
          2080
2240
          2400
         2560
          2880
          3040
```

```
NC 019843.3:21456-25517
                          ATGATACACTCAGTGTTTCTAC-TGATGTTCTTGTTAACACC-----TACAGAAAG
                                                                                      50
MN996532.1:21545-25354
                          ----ATGT---TTGTTTTT-C-TTGTTTTATTGCCACTAG-----TTTCTAG--TCAG
                                                                                      42
                                                                                      42
NC 045512.2:21563-25384
                          ----ATGT---TTGTTTTT-C-TTGTTTTATTGCCACTAG-----TCTCTAG--TCAG
                          ----ATGT-----TTATTTTCTTATTATTTCTTACTCTCACTAGTGGTAG
                                                                                      41
NC 004718.3:21492-25259
NC 014470.1:21391-25170
                          ATGAAATTT---TTGGCTTTTCTCTGTCTTCTTGGCTTTGCTAACGCTCAAGATGGCAAG
                                                                                      57
                                                 * ** **
NC 019843.3:21456-25517
                          TTACGTTGATGTAGGGCCAGATTCTGTTAA----GTCTGCTTGTATTGAGGTTGATATA
                                                                                      105
MN996532.1:21545-25354
                          TG-----A
                                                                                      72
                          TG-----TGTTAATCTTA--CAACCAGAACTCAATT-----A
                                                                                      72
NC 045512.2:21563-25384
NC 004718.3:21492-25259
                          TGACCTTGAC------T
                                                                                      81
                          TG-----TGGTACACTAT--CTAATAAAAGTC---C----A
NC_014470.1:21391-25170
                                                                                      84
NC 019843.3:21456-25517
                          CAACAGACTTTCTTTGATAAAACTTGGCCTAGGCCAATTGATGTTTCTAAGGCTGACGGT
                                                                                      165
MN996532.1:21545-25354
                                                                                      123
                          CCTC---C--TGCATACACCAACTCATCCA---CCCGTGGTGTCTATTACCCTGACAAA
                                                                                      123
NC 045512.2:21563-25384
                          CCCC---C--TGCATACACTAATTCTTTCA---CACGTGGTGTTTATTACCCTGACAAA
NC 004718.3:21492-25259
                                                                                      135
                          CCTAATTA---CACTCAACATACTTCATCTA---TGAGGGGGGTTTACTATCCTGATGAA
NC 014470.1:21391-25170
                          TCTAAGCT---TACTCAGACTCCTTCTTCTA---GGAGGGGTTTTTATTATTTTGATGAC
                                                                                      138
                                               * *
                                                             * * * * ***
NC 019843.3:21456-25517
                          ATTAT-ATACCCTCAAGGCCGTACATATTCTAACATAACTATCACTTATCAAGGTCTTTT
                                                                                      224
MN996532.1:21545-25354
                          GTTTTCAGATCTTCA----ATTTGTTTT
                                                                                      166
NC_045512.2:21563-25384
                          GTTTTCAGATCCTCA-----GTTTTACA-TTCAACTCAGG-----ACTTGTTCT
                                                                                      166
NC 004718.3:21492-25259
                                                                                      178
                          ATTTTTAGATCAGAC-----ACTCTTTA-TTTAACTCAGG-----ATTTATTTC
NC 014470.1:21391-25170
                          ATTTTTAGGTCTTCA-----ATTCGTGTGCTTACCACTGG------CCATTTTC
                                                                                      181
                           ** * * *
NC 019843.3:21456-25517
                          TCCCTATCAGGGAGACCATGGTGATATGTTTACTCTGCAGGACATGCTACAGGCAC
                                                                                      284
                                                                                      218
MN996532.1:21545-25354
                          TACCTTTCT---TCTCCAATGTGACCTGGTTCCA----TGCTATACATGTTTCAGGGAC
NC 045512.2:21563-25384
                          TACCTTTCT---TTTCCAATGTTACTTGGTTCCA----TGCTATACATGTCTCTGGGAC
                                                                                      218
NC 004718.3:21492-25259
                          TTCCATTTT---ATTCTAATGTTACAGGGTTTCA----TACTATT-----
                                                                                      216
NC 014470.1:21391-25170
                          TTCCTTTTA---ATACTAACCTTACTTGGTATTT----GACTTTAAA-----GTC
                                                                                      224
                                       ** ** *
```

>MN996532.1:21545-25354 Bat coronavirus RaTG13, complete genome ATGTTTGTTTTTCTTGTTTTATTGCCACTAGTTTCTAGTCAGTGTGTTAATCTAACAACT AGAACTCAGTTACCTCCTGCATACACCAACTCATCCACCCGTGGTGTCTATTACCCTGAC AAAGTTTTCAGATCTTCAGTTTTACATTTAACTCAGGATTTGTTTTTACCTTTCTTCC AATGTGACCTGGTTCCATGCTATACATGTTTCAGGGACCAATGGTATTAAAAGGTTTGAT AACCCAGTTCTGCCATTCAACGATGGCGTCTATTTTGCTTCCACTGAGAAGTCTAATATA ATAAGAGGATGGATTTTTGGTACTACCTTAGATTCGAAGACCCAGTCTCTACTTATTGTT AATAACGCTACTAATGTTGTTATTAAAGTCTGTGAATTTCAATTTTGTAATGATCCATTT TCTAGTGCGAATAATTGCACTTTTGAGTATGTCTCTCAGCCTTTTCTTATGGACCTTGAA GGAAAACAGGGTAATTTCAAAAATCTTAGGGAATTCGTGTTTAAGAATATTGATGGTTAT TTCAAAATATATTCTAAACATACGCCTATTAATTTAGTGCGTGATCTTCCCCCTGGTTTT TCAGCTTTAGAACCATTGGTAGATCTGCCAATAGGTATTAACATCACTAGGTTTCAAACT TTACTTGCTTTACATAGAAGCTATTTGACTCCTGGTGATTCTTCTTCAGGTTGGACAGCT GGTGCTGCAGCTTATTATGTGGGTTATCTTCAACCAAGGACTTTTCTACTAAAATATAAT GAGAATGGAACCATTACAGATGCTGTAGACTGTGCACTTGACCCTCTTTCAGAAACAAAG TGTACGTTAAAATCCTTCACTGTTGAAAAAGGAATTTATCAAACCTCTAACTTTAGAGTC CAACCAACAGATTCTATTGTTAGATTCCCAAATATTACAAACTTATGTCCTTTTGGTGAA

```
Percent Identity Matrix - created by Clustal2.1
  1: NC 019843.3 21456-25517 100.00
                                      53.01
                                             52.64
                                                     51.69
                                                             51.79
 2: MN996532.1 21545-25354
                              53.01
                                    100.00
                                              93.15
                                                     73.89
                                                             70.24
 3: NC 045512.2 21563-25384
                              52.64
                                      93.15
                                            100.00
                                                     73.97
                                                             70.40
                              51.69
                                      73.89
                                                    100.00
                                                             71.79
 4: NC 004718.3 21492-25259
                                             73.97
  5: NC_014470.1_21391-25170
                              51.79
                                      70.24
                                             70.40
                                                     71.79 100.00
```

```
phylogenetic_tree.ph.txt >
NC 019843.3:21456-25517:0.356531
                                    NC 019843.3 21456-25517:0.33503,
MN996532.1:21545-25354:0.0366142
                                    MN996532.1 21545-25354:0.03404.
                                    NC_045512.2_21563-25384:0.03447)
NC 045512.2:21563-25384:0.0366142
                                    :0.09692.
):0.156459
                                    NC 004718.3 21492-25259:0.12953)
                                    :0.01205.
                                    NC 014470.1 21391-25170:0.14706):
NC 004718.3:21492-25259:0.171842
NC 014470.1:21391-25170:0.171842
1:0.0212317
):0.163458
```

Questions?

