

Science2 Tutorial - 2 (26 April 2023)

- Romica Raisinghani (2021101053)

Ans 1a)

Multiple Sequence Alignment (Genes)

Here,

1. SARS coronavirus
2. SARS CoV-2
3. P-CoV
4. MERS CoV
5. Bat coronavirus RaTG13
6. Bat coronavirus

CLUSTAL O(1.2.4) multiple sequence alignment

NC_019843.3:21456-25517	ATGATAACACTCAGTGTCTACTGATGTTCTGTTAACACCTACAGAAAGTTACGTTGAT	60
NC_048212.1:20814-24623	ATGGTCGT---CTTATTGTTCGTGTGTTAGTTAGTTAGCTGACAGGCCTAATTGT	57
NC_004718.3:21492-25259	---ATGTT---TAT-----TTTCTTATTATTCTTACTCTCACTAGTGGTAGT	42
MT799526.1	ATGTTGTT---TTT-----CTTCTTTTACACTTGCC-----TTAGT	35
NC_045512.2:21563-25384	---ATGTT---TGT-----TTTCTTGTGTTATTGCCA-----CTAGT	32
MN996532.1:21545-25354	---ATGTT---TGT-----TTTCTTGTGTTATTGCCA-----CTAGT	32
	* * ** * * * *	
NC_019843.3:21456-25517	GTAGGGCCAGATTCTGTTAAGTCTGCTTGATTGAGGTTGATATACAACAGACTTCTT	120
NC_048212.1:20814-24623	TATTTCGCCTCA-----GTATTCTTAC-----	78
NC_004718.3:21492-25259	GACCTTGACCG-----GTGCAC---C-----	60
MT799526.1	AAATTC--ACA-----ATGTGT---T-----	51
NC_045512.2:21563-25384	CTCTAG--TCA-----GTGTGT---T-----	48
MN996532.1:21545-25354	TTCTAG--TCA-----GTGTGT---T-----	48
	*	
NC_019843.3:21456-25517	GATAAAAACCTGGCCTAGGCCAATTGATGTTCTAAGGCTGACGGTATTATATACCCCAA	180
NC_048212.1:20814-24623	-----ACTATTCACCAGCAACCATCACTAACACCTCTATT----TTAATGTCATCC	128
NC_004718.3:21492-25259	-----ACTTTTGATGATGTTCAAGCTCTTAATTACACTCAACATACTTCATCTATGAGG	114
MT799526.1	-----AATTTAACAGGTAGAGCTGCTATCCAGCCTTCATTCACCAATTCCCTCAAAGA	105
NC_045512.2:21563-25384	-----AATCTTACAAACAGAACCTCAATTACCCCTGCATACACTAATTCTTCACACGT	102
MN996532.1:21545-25354	-----AATCTAACAACTAGAACCTAGTTACCTCCTGCATACACCAACTCATCCACCGT	102
	* * * *	
NC_019843.3:21456-25517	GGCCGTACATATTCAACATAACTATCACTTATCAAGG----TCTTTTCCCTATCAGGG	236
NC_048212.1:20814-24623	CATTTC CATTGTTGCCCCGAGTTTCAGTTGCGCGCACCAACCTGTTGCGCCAGCG	188
NC_004718.3:21492-25259	GGGGTTTACTATCCTGATGAAATTTTAGATCAGACAC---TCTTTATTAACTCAGGA	170
MT799526.1	GGTGTGTTATTATCCTGACACCATATTAGATCAAACAC---ACTTGTGTTGAGTCAGGG	161
NC_045512.2:21563-25384	GGTGTGTTATTACCCCTGACAAAGTTTCAGATCCTCAGT---TTTACATTCAACTCAGGA	158
MN996532.1:21545-25354	GGTGTCTATTACCCCTGACAAAGTTTCAGATCCTCAGT---TTTACATTAACTCAGGA	158
	* * * * * * ***	
NC_019843.3:21456-25517	AGACCATGGTGATATGATGTTA-----CTCTGCAGGACATGCTACAGGCAC	284
NC_048212.1:20814-24623	TTTACAAGTTGCTGCTT---ATGATTTACTAAGATAACCCATTACACC-----	233
NC_004718.3:21492-25259	TTTATTTCTTCATTTATTCTAATGTTACAGGTTTCA-----	209
MT799526.1	TTACTTTTTACCTTTTATTCTAATGTTAGCTGTTATTATGCATTGAC---AAAAACTAA	218
NC_045512.2:21563-25384	CTTGTCTTACCTTTCTTTCCAATGTTACTGGTTCCATGCTATACATGCTCTGGGAC	218
MN996532.1:21545-25354	TTGTTTTTACCTTTCTTCCTCCAATGTCACCTGGTTCCATGCTATACATGTTCAAGGAC	218
	* *	

NC_019843.3:21456-25517	AACTCCACAAAAGTTGTTGTAGCTA ACTATTCTCAGGACGT AAACAGTTGCTAATGG	344				
NC_048212.1:20814-24623	GCCACACCGTAGTATTT-----	TAATAA				
NC_004718.3:21492-25259	TACTATTAA TACATACGTT -----	TGGCAA				
MT799526.1	CAGTGCTGAAAAGAGGT-----	TGATAA				
NC_045512.2:21563-25384	CAATGGT ACTAACAGAGTT -----	TGATAA				
MN996532.1:21545-25354	CAATGGT ATTAAGAGGT -----	TGATAA				
	*	*				
NC_019843.3:21456-25517	GTTTGTCGTCGT ATAGGAGCAGCTGCCAATTCCACTGGCA CTGTTATTATTAGCCCATC	404				
NC_048212.1:20814-24623	CACTTATTTCGTGTAGGCAGTGGTATTTAGTGAAT ACTCACAT -----	302				
NC_004718.3:21492-25259	CCCTGTC CATACCTTTAAGGATGGT ATTTATTTGCTGCCACAGA-----	278				
MT799526.1	CCCTGTTTGATTTC AAAGACGGT ATTTACTTGCTGCAACTGA-----	287				
NC_045512.2:21563-25384	CCCTGTC CCTACCATTAAATGATGGT TTTATTTGCTCCACTGA-----	287				
MN996532.1:21545-25354	CCCAGTTCTGCCATTCAACGATGGCGT CATTGGCTTCAACTGA -----	287				
	*	*				
NC_019843.3:21456-25517	TACCA CGCGCTACTATACGAAAATTACCC TGTTTATGCTGGTTCTCAGTTGGTAA	464				
NC_048212.1:20814-24623	-GTATC ACAGAACTAACATAGCTACCAACCTAGGTAGCT CTTCTATGG	350				
NC_004718.3:21492-25259	-GAAAT CAAATGTTGTCGGTGGTTGGTTGGTTCTACC ---ATGA	322				
MT799526.1	-AAA ATCTAACATTGTCAGAGGTTGGATCTTGGAACGACT ---CTTG	331				
NC_045512.2:21563-25384	-GAAGT CTAACATAAAAGAGGCTGGATTTTGGTACTACT ---TTAG	331				
MN996532.1:21545-25354	-GAAGT CTAACATAAAAGAGGATGGATTTTGGTACTACC ---TTAG	331				
	*	*	*	**	*	*
NC_019843.3:21456-25517	TTTCTCAGATGGT AAAATGGGCCGCTCTTCATCATACTCTAGTTCTTGC -----	517				
NC_048212.1:20814-24623	TT---GTCAAGAGCCTTTGGTGC-----AGC CTTGGCACACTTTGTTAATGAA	399				
NC_004718.3:21492-25259	AC---AAC AAGTCACAGTCGGTGATTATTAAACAACTCTACTAATGTTATACGA	378				
MT799526.1	AC---AACACATCACAGTC ACTTTGATAGTTAACACGCAACTAATGTTATCATCAA	387				
NC_045512.2:21563-25384	AT---TCGAAGACCAGTC CTACTTATTGTTAAACGCTACTAATGTTATTAAA	387				
MN996532.1:21545-25354	AT---TCGAAGACCAGTC CTACTTATTGTTAAACGCTACTAATGTTATTAAA	387				
	*	*	**			
NC_019843.3:21456-25517	-----CCGATGGATGTGGC ACTTTACTTAGAGCTTTTATTGTTATTCTA	561				
NC_048212.1:20814-24623	CCCATA ACTTAGTTATGTTTATGCAGGCAATGTTGGTAGTTGGCTAGTCTTGTACT	459				
NC_004718.3:21492-25259	GCATGT AACTT---TGAATTGTTGACAACCCCTTCTTGCTGTTCTAAACCCATGGGT	435				
MT799526.1	GTTTG AAATT---CCAGTTGGTATGACCCCTAACCTTAGTGGTTATTATCATAACAT	444				
NC_045512.2:21563-25384	GTCTGT GAATT---TCAATTGTTAATGATCCATTGGTGGTTATTACACACAAAAAC	444				
MN996532.1:21545-25354	GTCTGT GAATT---TCAATTGTTAATGATCCATTGGTGGTTATTACACACAAAAAC	444				
	*	*				

NC_019843.3:21456-25517	GAGCCTCGCTCTGGAAATCATTGTCCTGCTGGCAAT ---TCCTATACTCTTTGCCACT	618
NC_048212.1:20814-24623	CAGCAGCAAGCTAACATCACCATAGTGTCTTGATAACGCCACTTGTGTGCTAACCCC	519
NC_004718.3:21492-25259	A---CACAGACACATAC -----TATGATATTGATAATGCATTAAATTGCACT	480
MT799526.1	-----AAAACGTGGAGCAGCAGAGAGTTGCTGTTATTCCCTTATGCCAATTGCACT	498
NC_045512.2:21563-25384	A---ACAAAAGTTGGATGGAAAGTGAAGTTAGCTAGTGCAGATAATTGCACT	501
MN996532.1:21545-25354	A---ACAAAAGTTGGATGGAAAGTGAAGTTACTAGTGCAGATAATTGCACT	501
	*	* * *
NC_019843.3:21456-25517	TATCACACTCCTGCAACAGATTGTTCTGATGGCAATTACAATCGTA-ATGCCAGTCTGAA	677
NC_048212.1:20814-24623	TTCTT-----TCTAAG--ATGGGGTCCAGGTGTCAACGCAGTTTAC	560
NC_004718.3:21492-25259	TTCGAGTACATA-TCTGATGCCCTTCGCTTGTGTTAGAAAAAGTCAGGTAAATTAA	539
MT799526.1	TTTGAGTATGTG-TCTAAGTCCTTATGCTAGATATACTGGCAAAGTGGCTTATTGA	557
NC_045512.2:21563-25384	TTTGAAATATGTC-TCTCAGCCTTTCTTATGGACCTTGAGGAAAACAGGGTAATTCAA	560
MN996532.1:21545-25354	TTTGAGTATGTG-TCTCAGCCTTTCTTATGGACCTTGAGGAAAACAGGGTAATTCAA	560
	*	*
NC_019843.3:21456-25517	CTCTTTAAGGAGTATTTAATTACGTAACTGCACCTTATGTACACTATAACATTAC	737
NC_048212.1:20814-24623	TATAAGTAAATGCATTCAGTGTCACTGTTAATTACATTAACTGACACT -----	609
NC_004718.3:21492-25259	ACACTTACGAGAGTTGTGTTAAAAATAAAG-----	571
MT799526.1	CACATTAAGAGAGTTGTTCCGAAATGTCG-----	589
NC_045512.2:21563-25384	AAATCTTAGGAAATTGTGTTAAGAATATTG-----	592
MN996532.1:21545-25354	AAATCTTAGGAAATTGTGTTAAGAATATTG-----	592
	* *	*
NC_019843.3:21456-25517	CGAAGATGAGATT-TTAGAGTGGTTGGCATTACACAAACTGCTCAAG-GTGTTCACCTC	795
NC_048212.1:20814-24623	-AAACTTGTAAATTTCAGTGTCACTGCTGCTATAATTAGCTTTACTTTGCAGA	668
NC_004718.3:21492-25259	-----ATGGGTTCTCTATGTTATAAGGGCTATCAACCT-----AT---AGATGTAGT	617
MT799526.1	-----ACGGATATTCAGATTACTAAACACACCT-----GT---TAATGTAAA	635
NC_045512.2:21563-25384	-----ATGGTTATTTAAAATATATTCTAACGACACACGCT-----AT---TAATTAGT	638
MN996532.1:21545-25354	-----ATGGTTATTTAAAATATATTCTAACACACGCT-----AT---TAATTAGT	638
	* * *	*
NC_019843.3:21456-25517	TTCTCATCTCGGTATGTTGATTGTACGGCGGCAATATGTTCAATTGCCACCTTGCCT	855
NC_048212.1:20814-24623	TGGTGACATTTTATGTAATGCTGCACTGGGTGATTCTTGAAATTACCTACTGCTTA	728
NC_004718.3:21492-25259	TCGTGATCTACCTCTGGTTAACACTTGAAACCTATTAAAGTTGCCCTTGGTAT	677
MT799526.1	TAGTAATTACCTATAGGTTTCAGCACTTGAACCTTGTGAAATTCCAGCTGGCAT	695
NC_045512.2:21563-25384	GCGTGATCTCCCTCAGGGTTTCGGCTTAGAACCAATTGGTAGATTGCCAATAGGTAT	698
MN996532.1:21545-25354	GCGTGATCTCCCCCTGGTTTCAGCTTAAAGAACCAATTGGTAGATTGCCAATAGGTAT	698
	* * *	* * * *

NC_019843.3:21456-25517	GT TTATGATACTATTAAAGTATTATTCTATCATTCC TCA-----	893
NC_048212.1:20814-24623	CCC GCTGCAGCC CAT TTACGT AT ACCTGCTGG TTAAACCAGATGTTGACAGTTTT	788
NC_004718.3:21492-25259	TAACATT ACA A TTT AGGCCATT CTTACAGC TTT-----	715
MT799526.1	AAATATT ACTA A TTT AGAAC ACTC CT ACTACATA -----	733
NC_045512.2:21563-25384	TAACAT CA T AGG TTCAAAC TTACTG CTTACATA-----	736
MN996532.1:21545-25354	TAACAT CA T AGG TTCAAAC TTACTG CTTACATA-----	736
	* * * *	
NC_019843.3:21456-25517	-CA-----GTATT---CGTTCTATCCAAAGTGA TAGAAAAGCTTGG GC	932
NC_048212.1:20814-24623	CCAGGCCATAACACGCCCA ATGGT GAAG TGCTG AGCTTCTAGCACCGT	848
NC_004718.3:21492-25259	-CA-----CCTGTC AA GAC ATTTGGG CA GTC AGCTGC	749
MT799526.1	-GAGGAGACCCA-----TGC CTA -----ATAATGGCTGGACAGCTGGTGC	776
NC_045512.2:21563-25384	-GAAGTTATTG A -----CTCCTGGT ATT CTTCTTCAGGTTGGACAGCTGGTGC	788
MN996532.1:21545-25354	-GAAGCTATTG A -----CTCCTGGT ATT CTTCTTCAGGTTGGACAGCTGGTGC	788
	*	
NC_019843.3:21456-25517	TG CTTCTACGT AT ATAAA CTCAACGGTAA CTTCTG TTGGATTTCTGTTGATGG	992
NC_048212.1:20814-24623	TTCTGGT ATG TGCGGGTGT ATT ATAAAG GCT ACTGGTTGGT ATG TTCTTACAG	908
NC_004718.3:21492-25259	AGCCT ATTTG TTGGCT ATTTAA AGCCA ACTAC ATTTATGCT AA G TATG AAAATGG	809
MT799526.1	TGCTT ATTACG TGGGCT ATTAG C TCACG T ACATT T GTTAA TT AA TT ATA ATGAAAATGG	836
NC_045512.2:21563-25384	AGCTT ATTATG GGGTT ATCTC AA CCTAGG ACTTTCT ATTAAA AT AA ATGAAAATGG	848
MN996532.1:21545-25354	AGCTT ATTATG GGGTT ATCTC AA CCAAGG ACTTTCT ACTAAA AT AA TGAGAATGG	848
	* * * ** * * * * * * * * *	
NC_019843.3:21456-25517	TTATATACG CAGAG C TATAG ACTG TGG TTTAA TGATTG T CAAACTCC ACTG CTCATA	1052
NC_048212.1:20814-24623	TA GTATCG T TA AT GTTAG TTT GTTCTC AG TGCTG AA AGT G AGTTG C AGTGCTT G TT	968
NC_004718.3:21492-25259	TACAAT ACAGA T GCTG T TTGATTG T CTC AAA ATCC ACT TGCTG AA CT AA TGCT GT	869
MT799526.1	TACAATA ACAGA T GCTG T TTGATTG CC CTAG T CCCTATCTG AG GGCT AA TG AC ATT	896
NC_045512.2:21563-25384	AACCATT ACAGA T GCTG T AGACTG T GCACTT G CCCTCTC AG AAACAAAGT G TACGTT	908
MN996532.1:21545-25354	AACCATT ACAGA T GCTG T AGACTG T GCACTT G CCCTCTT AG AAACAAAGT G TACGTT	908
	** * * *** * * * * * * * * *	
NC_019843.3:21456-25517	TGAAT CTTCG AT GTTG A CTGG AG TTTATT C AGTTG C CTTCG AA GCAAAAC CTTC	1112
NC_048212.1:20814-24623	GGGTAC TTGCGCC AG CTA GG TTTATT C ACTCAG T AA TT TCG AG TAGTCCC AC	1028
NC_004718.3:21492-25259	TAAGAG CTTGAG TTG ACAAAGGAA TT ACCGAC CT CTA TT CAGGGT T TTCC TC	929
MT799526.1	AAA ATCCTTAA CT GTTG AAA AGGAA T CTATCAG A CTTCTA AC TTAGAGTC AA CCAAAC	956
NC_045512.2:21563-25384	GAA ATCCTTAC G TGAGAAAAAGGAA T CTATCAA A CTTCTA AC TTAGAGTC AA CCAAAC	968
MN996532.1:21545-25354	AAA ATCCTTAC G TGTTG AAA AGGAA TT TATCAA AC CTCTA AC TTAGAGTC AA CCAAAC	968
	** * * * * * * * * * * *	

NC_019843.3:21456-25517	TGGCTCAGTTGTGAAACAGGCTGAAGGTGTTAATGTGAT---TTTCACCTCTCTGTC	1169
NC_048212.1:20814-24623	TGACAATGTACGTATCACTAACAGTGCCTCTCTGTAGTGTACCTTATTCTGTTTGAG	1088
NC_004718.3:21492-25259	AGGAGATGTTGAGATTCCCTAATTACAAAATTGTGTCCTTTGGAGAGGTTTTAA	989
MT799526.1	TGAATCTATAGTTAGGTTCCAATTACAAAATTGTGTCCTTTGGTAAGTTTCAA	1016
NC_045512.2:21563-25384	AGAACTATTGTTAGATTCCAATTACAAAATTGTGTCCTTTGGTAAGTTTAA	1028
MN996532.1:21545-25354	AGATTCTATTGTTAGATTCCAATTACAAAATTGTGTCCTTTGGTAAGTTTAA	1028
	* * * * * * * ***	
NC_019843.3:21456-25517	TGG---CACACCTCTCAGGTTATAATTCAAGCGTTGGTTTACCAATTGCAATT	1226
NC_048212.1:20814-24623	TCGGCCACCTTAC-----TTTGATGGAGCGTTATGCCATTCCATTGTAAAGTT	1142
NC_004718.3:21492-25259	TGCTACTAAATTCCCTCTGCTATGCATGGAGAGAAAAAAATTCTAATTGTGTTGC	1049
MT799526.1	TGCAACCACCTTGCATCTGTTATGCTGGAAAGAAAGAGAATCAGTAACGTGTTGC	1076
NC_045512.2:21563-25384	CGCCACCAGATTGCATCTGTTATGCTGGAAAGGAAGAAATCAGCAACTGTGTTGC	1088
MN996532.1:21545-25354	CGCCACCACATCGCATCAGTTATGCTGGAAAGAAAGAGAATTAGCAACTGTGTTGC	1088
	* * * * * * * ***	
NC_019843.3:21456-25517	TAATCTTACCAAATTGCTTCACCTTTCTGTGAATGATTTACTGTAGTCAAATATC	1286
NC_048212.1:20814-24623	TGATTTTCAAGGCATTGCTTAGCATTGCTCACATTCAACTTCGCTTTGGCATATC	1202
NC_004718.3:21492-25259	TGATTACTCTGTCCTACAACCTAACATTCTCTAACATTCAATGTTATGGAGTTTC	1109
MT799526.1	TGATTACTCTGTTCTTACAACCTAACCTCTCTAACATTCAATGTTATGGAGTTTC	1136
NC_045512.2:21563-25384	TGATTATTCTGTCCTATATAATTCCGCATCATTCTAACCTTAAAGTGTATGGAGTGC	1148
MN996532.1:21545-25354	TGATTACTCTGTCCTATATAATTCCACTTCACCTTAAAGTGTATGGAGTGC	1148
	* ** * * * * * ***	
NC_019843.3:21456-25517	TCCAGCAGCAATTGCTAGCAACTGTTATTCTCAGTGAATTGGATTACTTTCATACCC	1346
NC_048212.1:20814-24623	GCCTACTAAGCTTGCCTACTATGTGTTTGGCACTGTCACTTACAGATTATGTTGGTTAA	1262
NC_004718.3:21492-25259	TGCCACTAAGTTGATGATCTTGCTCTCCAAATGTCTATGCAGATTCTTGTAGTCAA	1169
MT799526.1	ACCAACCAAACTAAATGATCTGCTTAACTAACGTTATGCAGACTCATTGTAGTTAG	1196
NC_045512.2:21563-25384	TCCTACTAAATTAAATGATCTGCTTAACTAATGTCTATGCAGATTCAATTGTAAATTAG	1208
MN996532.1:21545-25354	TCCTACTAAATTAAATGATCTGCTTAACTAATGTTATGCAGACTCATTGTGATTAC	1208
	* * * * * * * ***	
NC_019843.3:21456-25517	ACTTAGTATGAAATCCGATCTCAGTGTAGTTCTGCTGGTCAAATATCCCAGTTAA	1406
NC_048212.1:20814-24623	TGTTACTCATTACAATAACCTCTTAAATGATGTTCTGTAGTTAGCCTTATAATTAA	1322
NC_004718.3:21492-25259	GGGAGATGATGTAAGACAAATAGCGCCAGGACAAACTGGTGTATTGCTGATTATAATTAA	1229
MT799526.1	AGGTGATGAAAGTCAGACAAATTGCTCCAGGACAAACAGGAAGAATTGCTACTATAATTAA	1256
NC_045512.2:21563-25384	AGGTGATGAAAGTCAGACAAATCGCTCCAGGGCAAACTGGAAAGATTGCTATTATAATTAA	1268
MN996532.1:21545-25354	AGGTGATGAAAGTCAGACAAATTGCGCCAGGACAAACAGGAAGATTGCTACTACAATTAA	1268
	* * * * * * * *****	

NC_019843.3:21456-25517	TAAACAGTCCTTTCTAATCCCACATGTTGATTTAGCGACTGTTCCCTCATAACCTTAC	1466	
NC_048212.1:20814-24623	TCAATTACCGCGTAAATTCTACGGTTGTTGCAT---AGTTATTATTTACCTAATG---A	1376	
NC_004718.3:21492-25259	TAATTGCCAGATGATTTCATGGTTGTGTCCTT---GCTTGAATACTAGGAACATTGA	1286	
MT799526.1	TAAACTCCCTGATGATTTCACAGGTTGTAAATA---GCTTGAATTCTAACAAACCTTGA	1313	
NC_045512.2:21563-25384	TAATTACCAAGATGATTTACAGGCTGCGTTATA---GCTTGAATTCTAACAAACCTTGA	1325	
MN996532.1:21545-25354	TAACACTACCAGATGATTTACTGGTTGTGTTATA---GCTTGAATTCTAACAGATATTGA	1325	
	* * * * *	*	
NC_019843.3:21456-25517	TACTATTACTAAGCCTTAAAGTACAGCTATTTAACAAAGTGCTCTCGTCTTCTTCTGA	1526	
NC_048212.1:20814-24623	CACTGCA-----TTTAGTTAATCTGTA-----GC	1400	
NC_004718.3:21492-25259	TGCTACT---TCAACTGGTAAATTAAATTATAAAATAGGTATCTTA-----GA	1332	
MT799526.1	TTCTAAG---GTTGGTGGTAAATTAAATTACCTGTATAGATTGTTA-----GA	1359	
NC_045512.2:21563-25384	TTCTAAG---GTTGGTGGTAAATTAAATTACCTGTATAGATTGTTA-----GG	1371	
MN996532.1:21545-25354	TGCAAAA---GAGGGCGGTAAATTAACTATCTTACCGCTCTTTA-----GA	1371	
	*	*	*
NC_019843.3:21456-25517	TGATCGTACTGAAGTACCTCAGTTAGTGAACGCTAATCAAACTCACCCGTGTATCCAT	1586	
NC_048212.1:20814-24623	TTCGCGCATACGTTACCCCT---TCTTGG-----GTGCATTC	1433	
NC_004718.3:21492-25259	CATGGCAAGCTTAGGCCCT---TTGAGAGAGACATATCTAATGTGCCTTCTCCCTGA	1388	
MT799526.1	AAGTCCAACCTCAAACCTT---TTGAAACGAGACATTCTACAGAAATATACCAAAGCTGG	1415	
NC_045512.2:21563-25384	AAGTCTAACTCAAACCTT---TTGAGAGAGATATTCAACTGAAATCTATCAGGCCGG	1427	
MN996532.1:21545-25354	AAAGCTAACTTAAACCTC---TTGAGAGGGATATCTCAACTGAAATTACCAAGCAGG	1427	
	*	*	*
NC_019843.3:21456-25517	TGTCCCATCCACTGTGTGGGAGACGGTGATTATTAGGAAACAATCTCCACTTGA	1646	
NC_048212.1:20814-24623	TATCACACCTGGTGGTAGACAACCCG---TTGGCCCTTCTTA-----GA	1475	
NC_004718.3:21492-25259	TGGCAACCTTGCACCCC---ACCTGCTTAAATTGTTATTGG-----CCATTAAA	1436	
MT799526.1	TAGTACACCCCTGCAATGGGGTTGAAGGTTTAACTGTTACTTT-----CCTCTACA	1466	
NC_045512.2:21563-25384	TAGCACACCTTGTAAATGGTGTGAAGGTTTAAATTGTTACTTT-----CCTTTACA	1478	
MN996532.1:21545-25354	CAGCAAACCTTGTAAATGGTCAAACCTGGTCAATTGCTACTAC-----CCACTTTA	1478	
	*	*	*
NC_019843.3:21456-25517	AGGTGGTGGCTGGCTTGTGCTAGTGGCTCAACTGTTGCCATGACTGAGCAATTACAG--	1704	
NC_048212.1:20814-24623	T----TCC-TTACAAAGTTCC-----AGCAAGCCTTGTACCGGATCATGTTT	1517	
NC_004718.3:21492-25259	TGATTATGGTTTTACACCACT-----ACTGGCATTGGCTACCAACCTTA---	1481	
MT799526.1	ATCTTATGGTTTCCACCCACT-----AATGGTGTGGTTACCAACCTTA---	1511	
NC_045512.2:21563-25384	ATCATATGGTTTCCACCCACT-----AATGGTGTGGTTACCAACCTA---	1523	
MN996532.1:21545-25354	TAGATATGGATTTACCCACT-----GATGGTGTGGTCAACCAACCTTA---	1523	
	*	*	*

NC_019843.3:21456-25517	ATGGGCTTGGTATTACAGTTCAATATGGTACAGACACC-AATAGTGTTCGCCCCAAGCT	1763	
NC_048212.1:20814-24623	AGGTTAGCCGTTATTCTTAAAGTAGCCTCGGAAACAAAATTAGTTGTCCAG----	1573	
NC_004718.3:21492-25259	CAGAGTTGTAGTACTTCTTTGAACCTTAAATGCACCGGCCACGGTTGTGGAC----	1537	
MT799526.1	TAGAGTAGTACTTGTCAATTGAACCTTAAATGCACCTGCTACTGTGTGGAC----	1567	
NC_045512.2:21563-25384	CAGAGTAGTACTTCTTTGAACCTTCAATGCCACCGCAACTGTGTGGAC----	1579	
MN996532.1:21545-25354	TAGGGTAGTACTTCTTTGAACCTTAAATGCACCGCAACTGTGTGGAC----	1579	
	* ** * * * ** *****		
NC_019843.3:21456-25517	TGAATTGCTAATGACACAAAAATTGCCCTCAATTAGGCAATTGCGTGGAAATTCCCT	1823	
NC_048212.1:20814-24623	-----TAGGCAATGACACAGATAATTGCGCTGACACTTGTCACACTACAATAT	1622	
NC_004718.3:21492-25259	-----CAAAATTATCCACTGACCTTAAAGAACAGTGTGTCAATTAAATT	1586	
MT799526.1	-----CTAACACAGTCCACTAACCTAGTTAAAAACAAATGTGTCAACTTCAATT	1616	
NC_045512.2:21563-25384	-----CTAAAAAGTCTACTAATTGGTTAAAAACAAATGTGTCAATTCAACTT	1628	
MN996532.1:21545-25354	-----CTAAGAAGTCTACTAACCTGGTTAAAAATAATGTGTCAATTCAACTT	1628	
	* ** * * * *		
NC_019843.3:21456-25517	CTATGGTGTTCGGGCGTGGTGTTCAGAATTGACAGCTGTAGGTGTGACAGCA	1883	
NC_048212.1:20814-24623	TTATGGTTATCAAGGCACTGGTGTATTAGTAGGTCAAATTACACTTGCCTCTTCAA	1682	
NC_004718.3:21492-25259	TAATGGACTCACTGGTACTGGTGTAACTCCTCTTCAAGAGATTTCAACCATTCA	1646	
MT799526.1	TAATGGTCAACAGGCACAGGTGTTCTACAGTCTAGCAAAAGTTTGCCTTCCA	1676	
NC_045512.2:21563-25384	CAATGGTTAACAGGCACAGGTGTTCTACTGAGTCAACAAAGTTTGCCTTCCA	1688	
MN996532.1:21545-25354	TAATGGTTAACGGCACAGGTGTCCTCACAGTCTAAATAAAAGTTTACCTTCCA	1688	
	**** ** ***** * * *		
NC_019843.3:21456-25517	GCGCTTTGTTATGA-----TGCCTGTTGCTGCTGTTAGTGTGTTCTGTCATCTATGATAA	1934	
NC_048212.1:20814-24623	GGTTTCTCGCTTATGTTCTGGTGAGTTGACTGTCTTGAGTT-----GTTCTC	1736	
NC_004718.3:21492-25259	ACAATTGGCCGTGATTTCTGATTTACTGATTCCGTTGAGATCCTAAACATCTGA	1706	
MT799526.1	ACAATTGGCAGAGATAATTGGCGACACTACTGATGCTGTCCTGATCCACAGACACTTGA	1736	
NC_045512.2:21563-25384	ACAATTGGCAGAGACATTGCTGACACTACTGATGCTGTCCTGATCCACAGACACTTGA	1748	
MN996532.1:21545-25354	ACAATTGGTAGAGACATTGCAAGACACTACTGATGCCGTCCTGATCCACAGACACTTGA	1748	
	** * * * * *		
NC_019843.3:21456-25517	TGGCAACTACTACTGTTGCGTGTGTTAGTGTGTTCTGTCATCTATGATAA	1994	
NC_048212.1:20814-24623	TTTTTACAGCTATCTCCGTGTCATTGCTCTATATCTGCA-----CTTTTA	1787	
NC_004718.3:21492-25259	AATATTAGACATTTCACCCGTGTTCTTGGGGGTGAAAGTGTAAATTACACCTGGAACAAA	1766	
MT799526.1	AATTCTTGATATCACACCGTGTGTTCTTGGGGGTGTCAGTGTATAACACCAAGGAAACAAA	1796	
NC_045512.2:21563-25384	GATTCTTGACATTACACCATGTTCTTGGGGGTGTCAGTGTATAACACCAAGGAAACAAA	1808	
MN996532.1:21545-25354	GATTCTTGACATTACACCATGTTCTTGGGGGTGTCAGTGTATAACACCTGGAACAAA	1808	
	*	* * * ** * * *	*

NC_019843.3:21456-25517	AGAAACTAAAACCCACGCTACTCTATTGGTAGTGTGCATGTAAACACATTCTTCTAC	2054
NC_048212.1:20814-24623	TAAGGGTTACACTACTTCTTGTGTTAATAACCTCCTTGTCCACCACATCGTGC	1847
NC_004718.3:21492-25259	TGCTTCATCTGAAGTTGCTGTTCTATATCAAGATGTTAACGTGACTGATGTTCTACAGC	1826
MT799526.1	CACTTCTAACCAAGTGGCTGTTCTTATCAGGATGTTAACGTGACTGAAGTCCCTGTTGC	1856
NC_045512.2:21563-25384	TACTTCTAACCAGGTTGCTGTTCTTATCAGGATGTTAACGTGACAGAAGTCCCTGTTGC	1868
MN996532.1:21545-25354	TGCCCTAACCAGGTTGCTGTTCTTATCAGGATGTTAACGTGACAGAAGTCCCTGTTGC	1868
	** * * * * * * * * *	
NC_019843.3:21456-25517	CATGTCTCAATACTCCCCTTACCGCATCAATGCTTAAACGGCGAGATTCTACATATGG	2114
NC_048212.1:20814-24623	AGT---CGTGGAACCTGTTAGTGCTTATTGGCGTCG-----TAGTGTGCGACAAATAA	1898
NC_004718.3:21492-25259	AATTATGCAGATCAACTCACACCAGCTTGGCGCAT-----ATATTCTACTGAAACAA	1880
MT799526.1	TATTATGCAGATCAATTAAACCCAACCTGGCGTGT-----TTACTCTACAGGTTCAA	1910
NC_045512.2:21563-25384	TATTATGCAGATCAACTTACTCTACTTGGCGTGT-----TTATTCTACAGGTTCAA	1922
MN996532.1:21545-25354	TATCCATGCAGACCAACTTACTCCCACCTTGGCGTGT-----TTACTCCACAGGTTCAA	1922
	* * * * *	
NC_019843.3:21456-25517	CCCCCTTCAGACACCTGTTGGTTGTGCCTAGGACTTAAATTCCCTT---TGTTCGT	2171
NC_048212.1:20814-24623	CACGTTTGACACAACCTGCTGGGTGCAATTAAATGCGTATAATTAACTAGCATTGTA	1958
NC_004718.3:21492-25259	TGTATTCCAGACTCAAGCAGGCTGTTATAGGAGCTGAGCATGTCG-----CACTTC	1934
MT799526.1	TGTTTTTCAACCGCGTGCAGGCTGTTAATAGGGCTGAAACATGTTAA-----CAACAC	1964
NC_045512.2:21563-25384	TGTTTTTCAACACAGTGCAGGCTGTTAATAGGGCTGAAACATGCAA-----CAACTC	1976
MN996532.1:21545-25354	TGTTTTTCAACACAGTGCAGGTTGTTAATAGGGCTGAAACATGCAA-----CAACTC	1976
	* * *** * * * * * *	**
NC_019843.3:21456-25517	AGAGGACTGCAAGTTGCCTCTGGTCATCTCTGTGCTCTTCTGACACACCTAGTAC	2231
NC_048212.1:20814-24623	TAATCAGTGTGATTACCTATTGGTGTAGTTATTGTTA-----CAACCATC-C---C	2008
NC_004718.3:21492-25259	TTATGAGTGCACATTCCTATTGGAGCTGGCATTGTCAGTTACCATACAGTTT---C	1991
MT799526.1	TTACGAGTGTGACATACCAATTGGTGCAGGAATATGTCAGTTACAGACTCAA---C	2021
NC_045512.2:21563-25384	ATATGAGTGTGACATACCCATTGGTGCAGGTATATGCGCTAGTTACAGACTCAGA---C	2033
MN996532.1:21545-25354	GTATGAGTGTGACATACCTATTGGTGCAGGAATATGCGCCAGTTACAGACTCAA---C	2033
	* * *** * * * * * *	
NC_019843.3:21456-25517	TCTCACACCTCGCAGTGTGCGCTCTGTTCCAGGTGAAATGCGCTTGGCATTCCATTGCTT	2291
NC_048212.1:20814-24623	TTATTAAAGG-----GTTTTGAGCCACATTATCATTGGTGTACGTACAACCCATT	2057
NC_004718.3:21492-25259	TTTATTACG-----TAGTACTAGCCAAAATCTATTGTCAGCTTATGTC	2039
MT799526.1	TAATTCACT-----TAGTGTGTCAGTCAAGCTTATTGCGCTACACTATGTC	2069
NC_045512.2:21563-25384	TAATTCTCCTCGCGGGCACGTAGTGTAGCTAGTCATCCATCATTGCGCTACACTATGTC	2093
MN996532.1:21545-25354	TAATTCACT-----GTAGTGTGGCCAGTCAATCTATTGCGCTACACTATGTC	2081
	*	*

NC_019843.3:21456-25517	TAATCATCCTATTCAAGGTTGATCAACCTTAATAGTAGTTATTTAAATTAAAGTATAACCCAC	2351
NC_048212.1:20814-24623	GCCT----GAT----TCTTTAACACCTATCACACCGGTTTCAAGTTTCAGTGCTAC	2108
NC_004718.3:21492-25259	TTA-----GGTGCT-GATAGTCATTGCTTACTCTAATAACACCATTGCTATACCTAC	2093
MT799526.1	ACTT-----GGTGCA-GAAAATTCAAGTTGCTTATGCTAATAACTCTATTGCCATACCTAC	2123
NC_045512.2:21563-25384	ACTT-----GGTGCA-GAAAATTCAAGTTGCTTACTCTAATAACTCTATTGCCATACCCAC	2147
MN996532.1:21545-25354	ACTT-----GGTGCA-GAAAATTCAAGTTGCTTATTCTAATAACTCTATTGCCATACCTAC	2135
	* * * * *	*** **
NC_019843.3:21456-25517	TAATTTTCCTTGGTGTGACTCAGGAGTACATTCAAGACAACCATTCAAGAAAGTTACTGT	2411
NC_048212.1:20814-24623	CAATTTTACTTTGGTAGCTTCACTGAAATATATTCAAACCTTATGCATCAAAATTCTAT	2168
NC_004718.3:21492-25259	TAACTTTCATTAGCATTACTACAGAAGTAATGCCTTCTATGGCTAAACCTCCGT	2153
MT799526.1	AAATTTTACTATTAGTGTGACCACTGAATTCTACCAAGTGTCTATGACAAAGACATCAGT	2183
NC_045512.2:21563-25384	AAATTTTACTATTAGTGTGACCACTGAATTCTACCAAGTGTCTATGACCAAGACATCAGT	2207
MN996532.1:21545-25354	AAATTTTACTATTAGTGTGACCACTGAATTCTACCTGTGTCTATGACAAAGACATCGGT	2195
	*** *** * * * * ** * * * * *	** * *
NC_019843.3:21456-25517	TGATTGTAAACAGTACGTTGCAATGGTTCCAGAAGTGTGAGCAATTACTGCGCGAGTA	2471
NC_048212.1:20814-24623	TGATTGTGCTAAGTATTATGTGGTGATTCTTACAGTGCAAGAACGGTGTGTTGAGTA	2228
NC_004718.3:21492-25259	AGATTGTAAATGTACATCTGCGGAGATTCTACTGAATGTGCTAATTGCTTCTCCAATA	2213
MT799526.1	AGATTGTACAATGTACATTGTTGGTGACTCAATAGAGTGCAGCACCTTTGCTCCAATA	2243
NC_045512.2:21563-25384	AGATTGTACAATGTACATTGTTGGTGATTCAACTGAATGCAAGCAATTGTTGCAATA	2267
MN996532.1:21545-25354	AGACTGTACAATGTATATTGTTGGTGATTCAACTGAAGTGCAGCACACCTTTGTTGCAATA	2255
	*** *** * * * * * * * * * *	** * **
NC_019843.3:21456-25517	TGGCCAGTTTGTCCAAAATAACCAGGCTCTCATGGGCCATTACGCCAGGATGA	2531
NC_048212.1:20814-24623	CGGCACATTGGTAAATGATGTTAATGTTGCTTAAAGAGGGTTTACGTTGCTTACAA	2288
NC_004718.3:21492-25259	TGGTAGCTTGCACACAACTAATCGTCACTCTCAGGTATTGCTGCTAACAGGATCG	2273
MT799526.1	TGGTAGTTTGCACACAACCTAATCGTCTTAACTGGAATTGCTGTTAACAGAACAA	2303
NC_045512.2:21563-25384	TGGCAGTTTGTACACAATTAAACCGTCTTAACTGGAATAGCTGTTAACAGAACAA	2327
MN996532.1:21545-25354	TGGTAGTTTGCACACAATTAAATCGTCTTAACTGGAATAGCTGTTAACAGAACAA	2315
	** ***** * * * * * * * * * *	** * **
NC_019843.3:21456-25517	TTCTGTACGTAATTGTTGCGAGCGTGAAAAGCTCTCAATCATCTCTATCA-TACCAAG	2590
NC_048212.1:20814-24623	TTCACTCGTTGATACGTTTCACTCCTGAA---AGTCCACAGCACCAAGTTAGCTT	2344
NC_004718.3:21492-25259	CAACACACGTGAAAGTGTGCTCAAGTCACAAACAAATGTACAAAACCCAACTT-TGAAAT	2332
MT799526.1	AAACACACAGGAAAGTTTGCACAAAGTTAAACAAATTACAGACACCACCAA-TAAAGG	2362
NC_045512.2:21563-25384	AAACACCCAAAGAAAGTTTGCACAAAGTCACAAACAAATTACAGACACCACCAA-TAAAG	2386
MN996532.1:21545-25354	AAATACTCAAGAAAGTTTGCACAGTTAAACAAATTATAAGACACCACCAA-TAAAG	2374
	* * * * * * * * * * * *	*

NC_019843.3:21456-25517	GTTTTGGAGGTGACTTAATTGACACTTC---TAGAACCTGTTCTATATCTACTGGCA	2647
NC_048212.1:20814-24623	ATACG---GGTGATTTAATTCACTCGTTAGTGGCTGTATTGGTACTGATTGTGATT	2401
NC_004718.3:21492-25259	ATTTT---GGTGGTTTAATTTCACAAATATTA-----CCTGACCCCTCAA	2377
MT799526.1	ATTTT---GGTGGTTCAACTTTCTCAAAATATTA-----CCAGATCCATCAA	2407
NC_045512.2:21563-25384	ATTTT---GGTGGTTTAATTTCACAAATATTA-----CCAGATCCATCAA	2431
MN996532.1:21545-25354	ATTTT---GGTGGTTCAATTTCACAAATATTA-----CCAGATCCATCAA	2419
	* ***** * * * * *	
NC_019843.3:21456-25517	GTCGTAGTCACGTAGTGCTATTGAGGATTGCTATTTGACAAAGTCACTATAGCTGATC	2707
NC_048212.1:20814-24623	CTAAGTCACATAGGCCGTTGTCAGATCTATTGTTAGAAGGTTAGTGTGCTGACC	2461
NC_004718.3:21492-25259	AGCCAACTAAGAGGCTTTATTGAGGACTGCTCTTTATAAGGTGACACTCGCTGATG	2437
MT799526.1	AACCAAGCAAGAGGTCAATTATTGAAGATTTACTCTTCACAAAGTGACACTGCTGATG	2467
NC_045512.2:21563-25384	AACCAAGCAAGAGGTCAATTATTGAAGATCTACTTTCAACAAAGTGACACTGCGATG	2491
MN996532.1:21545-25354	AACCAAGCAAGAGGTCAATTATTGAAGATTTACTTTCAATAAGTGACACTGCTGATG	2479
	* * * * * * * * * * * * * * *	
NC_019843.3:21456-25517	CTGGTTATATGCAAGGTTACGATGATTGCA---GCAGCAAGGTCCAGCATCAGCTGTG	2764
NC_048212.1:20814-24623	CAGGTTTCAATGCAATCTTACCAACAATGTTGGATGCACAGTGGGTGGTAATATACGTG	2521
NC_004718.3:21492-25259	CTGGCTTCATGAAGCAATA-----TGGCGAATGCCTAGGTGATATTATGCTAGAG	2488
MT799526.1	CTGGCTTCATCAAAACAATA-----TGGTGAATTGCTTGGTGATATTGCCGCTAGAG	2518
NC_045512.2:21563-25384	CTGGCTTCATCAAAACAATA-----TGGTGAATTGCTTGGTGATATTGCTGCTAGAG	2542
MN996532.1:21545-25354	CTGGCTTCATCAAAACAATA-----TGGTGAATTGCTTGGTGATATTGCTGCTAGGG	2530
	* *** * * * * * * * * * *	
NC_019843.3:21456-25517	ATCTTATTTGTGCTCAATATGTTGGCTGGTTACAAAGTATTACCTCTTATGGATGTTA	2824
NC_048212.1:20814-24623	ATTGCTTTGACTCAGACGTTCAATGGCATCTCAGTATTGCCACCCATTGATTCCTT	2581
NC_004718.3:21492-25259	ATCTCATTTGTGCGCAGAAGTCAATGGACTTACAGTGTGCTCACCTGCTACTGATG	2548
MT799526.1	ATCTTATTTGTGCAACAAAGTTAATGGCTTACTGTTCTGCCACCTTGCTCACAGATG	2578
NC_045512.2:21563-25384	ACCTCATTTGTGCAACAAAGTTAACGGCCTTACTGTTTGCCACCTTGCTCACAGATG	2602
MN996532.1:21545-25354	ATCTTATTTGTGCTCAAAAGTTCAATGGCCTTACTGTTCTGCCACCTTGCTCACAGATG	2590
	* * ***** * * * * * * * * * *	
NC_019843.3:21456-25517	ATATGGAAGCCCGGTATACTTCATCTTGCTTGGCAGCATAGCAGGTGTTGGACTG	2884
NC_048212.1:20814-24623	CCATGCAAGGCTCTATAACGACCGCCTAGTCGGTGGCATTGCAGCCTCTGGTTTACTT	2641
NC_004718.3:21492-25259	ATATGATTGCTGCCTACACTGCTGCTCTAGTTAGTGGTACTGCCACTGCTGGATGGACAT	2608
MT799526.1	AAATGATTGCTCAATACACCTCTGCACTACTTGCAAGGGACAATCACATCAGGTTGGACCT	2638
NC_045512.2:21563-25384	AAATGATTGCTCAATACACTTCTGCACTGTTAGCAGGGTACAATCACATTCTGGTTGGACCT	2662
MN996532.1:21545-25354	AAATGATCGCTCAATACACTTCTGCACTATTAGCAGGTACAATCACATTCTGGTTGGACCT	2650
	*** * * * * * * * * * * * * * *	

NC_019843.3:21456-25517	CTGGCTTATCCTCCTTGCTGCTATTCCATTGACAGAGTATCTTTATAGGTTAACG	2944
NC_048212.1:20814-24623	TTGGTGTGTCATCTGCTGCTGTAATTCCCTTGCACACAGTTGCAGTCAGACTTAATG	2701
NC_004718.3:21492-25259	TTGGTGTGCTGGCGCTGCTCTCAAATACCTTTGCTATGCAAATGGCATATAGGTTCAATG	2668
MT799526.1	TTGGTGTGCTGGTGCAGCATTACAGATACCATTGCTATGCAAATGGCTTATAGGTTAATG	2698
NC_045512.2:21563-25384	TTGGTGCAGGTGCTGCATTACAATACCAATTGCTATGCAAATGGCTTATAGGTTAATG	2722
MN996532.1:21545-25354	TTGGTGCAGGTGCTGCTTACAATACCAATTGCCATGCAAATGGCTTATAGGTTAATG	2710
	*** * *** * **** * * * * * * *	
NC_019843.3:21456-25517	GTTGTTGGCATTACTCAACAGGTTCTTCAGAGAACCAAAAGCTTATTGCCAATAAGTTA	3004
NC_048212.1:20814-24623	GGTTAGGTGTGACCACTAATGTTTATGGAGAACAACTATTGATAGCTAATGCTTTA	2761
NC_004718.3:21492-25259	GCATTGGAGTACCCAAAATGTTCTATGAGAACCAAAACAAATGCCAACCAATTAA	2728
MT799526.1	GTATTGGAGTACACAAAATGTTCTACAGAACCAAAACTATTGCAAACCAATTCA	2758
NC_045512.2:21563-25384	GTATTGGAGTACACAGAACATGTTCTATGAGAACCAAAATTGCAACCAATTAA	2782
MN996532.1:21545-25354	GTATTGGAGTACACAGAACATGTTCTATGAGAACCAAAATTGCAACCAAGTTA	2770
	* * *** * * * * * * * * * * * * * * * *	
NC_019843.3:21456-25517	ATCAGGCTCTGGGAGCTATGCAAACAGGCTTCACTACAACATAATGAAGCTTTAGAAGG	3064
NC_048212.1:20814-24623	ATAAGGCTCTGTTTGATACAGGAAGGTTTACAGCCACTAATCAGGCTCTAAACAAA	2821
NC_004718.3:21492-25259	ACAAGGCGATTAGTCACAAATTCAAGAACACTTACAACAAACATCACTGCACTGGCAAGC	2788
MT799526.1	ACAGTGCAATTGGCAAAATTCAAGATTCACTTCATCTACTGCAAGTGCACTTGGAAAC	2818
NC_045512.2:21563-25384	ATAGTGCTATTGGCAAAATTCAAGAACACTCACCTTCTTCCACAGCAAGTGCACTTGGAAAC	2842
MN996532.1:21545-25354	ATAGTGCTATTGGCAAAATTCAAGAACACTCACCTTCTTCACTGCAAGTGCACTTGGAAAC	2830
	* * *** * * * * * * * * * * * * * * * *	
NC_019843.3:21456-25517	TTCAGGATGCTGTGAAACACAAATGCACAGGCTCTATCCAAATTAGCTAGCGAGCTATCTA	3124
NC_048212.1:20814-24623	TCCAAACTGTTGTCATAACAAATGCCTTGCAATTGCAAGTGTAGTGCAGCAGTTGGTA	2881
NC_004718.3:21492-25259	TGCAAGACGTTGTAACCAAAATGCAAGGTTAAACACACTTGTAAACAACTTAGCT	2848
MT799526.1	TTCAAGATGTTGTCACCACAAATGCAAGGTTAAACACACTTGTAAACAACTTAGCT	2878
NC_045512.2:21563-25384	TTCAAGATGTTGTCACCACAAATGCAAGGTTAAACACGCTTGTAAACAACTTAGCT	2902
MN996532.1:21545-25354	TTCAAGATGTTGTCACCACAAATGCAAGGTTAAACACGCTTGTAAACAACTTAGCT	2890
	* * *** * * * * * * * * * * * * * * * *	
NC_019843.3:21456-25517	ATACTTTGGTGTCTATTCCGCCTCTATTGGAGACATCATACACGCTTGATGTTCTG	3184
NC_048212.1:20814-24623	ATACATTTGGTGTCTATCAGCATCTGTTAATGAAATTGATAGCTCTCGACTTGTG	2941
NC_004718.3:21492-25259	CTAATTTGGTGCACATTCAAGTGTGCTAAATGATATCCTTCGCGACTTGTAAAGTCG	2908
MT799526.1	CTAATTTGGAGCCATTTCAGTGTGTTAAATGACATTCTTCACGTCTTGACAAAGTTG	2938
NC_045512.2:21563-25384	CCAATTTGGTGCACATTCAAGTGTGTTAAATGATATCCTTCACGTCTTGACAAAGTTG	2962
MN996532.1:21545-25354	CCAATTTGGAGCTATTCTAGCGTGTAAATGATATCCTTCACGTCTTGACAAAGTTG	2950
	* ***** * * * * * * * * * * * * * * * *	

NC_019843.3:21456-25517	AACAGGACGCCAATAGACAGACTTATTAAATGCCGTTGACAACACTAAATGCTTTG	3244
NC_048212.1:20814-24623	AGGCTAATGCCGAAAGTTGATAGCTGATTCTGGTCGTATGGTTGACTTAACACATATG	3001
NC_004718.3:21492-25259	AGGCGGAGGTACAAATTGACAGGTTAATTACAGGCAGACTTCAAAGCCTCAAACCTATG	2968
MT799526.1	AGGCTGAAGTCAAATTGACAGGTTGATCACTGGCAGATTACAAAGTTGCAGACATACG	2998
NC_045512.2:21563-25384	AGGCTGAAGTCAAATTGATAGGTTGATCACAGGCAGACTCAAAGTTGCAGACATATG	3022
MN996532.1:21545-25354	AGGCTGAAGTGCAGATTGACAGGTTGATCACAGGCAGACTCAAAGCTTGCAAGACATATG	3010
	* * * * * * * * * * * * * * * *	
NC_019843.3:21456-25517	TTGCACAGCAGCTTGTGTTCCGAATCAGCTGCTCTTCCGCTCAATTGGCTAAAGATA	3304
NC_048212.1:20814-24623	TGACTCAGTTGTTATTCAAGCTTCTGAACCTCGGTCTAGGCAGATTGGCTAAGCAA	3061
NC_004718.3:21492-25259	TAACACAACAACAAATCAGGGCTGCTGAAATCAGGGCTCTGCTAAATCTTGCTGCTACTA	3028
MT799526.1	TGACTCAACAACAAATTAGAGCCGAGAAATTAGAGCTTCTGCTAAATCTTGCCGCACTA	3058
NC_045512.2:21563-25384	TGACTCAACAATTAAATTAGAGCTGCAAGAAATCAGAGCTTCTGCTAAATCTTGCTGCTACTA	3082
MN996532.1:21545-25354	TGACTCAACAATTAAATTAGAGCTGCAAGAAATCAGAGCTTCTGCCAATCTTGCTGCTACTA	3070
	* * * * * * * * * * * * * * * *	
NC_019843.3:21456-25517	AAGTCATGAGTGTGTCAGGGACAACTCAAAGCGTTCTGGATTTGCGGTCAAGGCACAC	3364
NC_048212.1:20814-24623	AGATGTCCGAATGTGTTAAGTCTCAATCTTGCTGATGACTTCTGTGGAACGGGACTC	3121
NC_004718.3:21492-25259	AAATGTCTGAGTGTGTTCTGGACAACTCAAAAAGAGTTGACTTTGTGGAAAGGGCTACC	3088
MT799526.1	AGATGTCTGAATGTGTTCTGGACAACTCAAAAAGAGTTGACTTTGTGGAAAGGGCTACC	3118
NC_045512.2:21563-25384	AAATGTCAAGAGTGTGACTTGGACAACTCAAAAAGAGTTGATTTTGTTGGAAAGGGCTATC	3142
MN996532.1:21545-25354	AAATGTCAAGAGTGTGACTTGGACAACTCAAAAAGAGTTGATTTTGTTGGAAAGGGCTATC	3130
	* * * * * * * * * * * * * * * *	
NC_019843.3:21456-25517	ATATAGTGTCTTTGTTGTAATGCCCTAATGCCCTTACTTCATGCTGATGGTTATT	3424
NC_048212.1:20814-24623	ATGTTCTTAGCATACCTCAGTTAGCTCTAATGGCATGCTGTTCATACATTATAGTTATC	3181
NC_004718.3:21492-25259	ACCTTATGTCTTCCCACAAAGCAGCCCCGATGGTTGCTTCTCATATGTCACGTATG	3148
MT799526.1	ACCTTATGTCTTCCCAGTCAGCACCTCATGGTGTAGTCTTCTGCTGATGTGACTTATG	3178
NC_045512.2:21563-25384	ATCTTATGTCTTCCCAGTCAGCACCTCATGGTGTAGTCTTCTGCTGATGTGACTTATG	3202
MN996532.1:21545-25354	ATCTTATGTCTTCCCAGTCAGCACCTCATGGTGTAGTCTTCTGCTGATGTGACATATG	3190
	* * * * * * * * * * * * * * * *	
NC_019843.3:21456-25517	ACCCCTAGCAACCACATTGAGGTTGTTCTGCTTATGGTCTTGCAGTCAGCTAACCTTA	3484
NC_048212.1:20814-24623	AACCTACTAAGTATGCACAGGTGATACCAACGGCGGGCTATGCTCAATGGTACAGGCT	3241
NC_004718.3:21492-25259	TGCCATCCCAGGAGAGGAACCTCACACAGCGCCAGCAATTGTCATGAAGGCAAGCAT	3208
MT799526.1	TTCCCATCTCAAGAAAAGAATTTACTACTACCCCTGCCATTGTCATGAAGGAAAAGCAC	3238
NC_045512.2:21563-25384	TCCCTGCACAAGAAAAGAACCTCACAACTGCTCTGCCATTGTCATGATGGAAAAGCAC	3262
MN996532.1:21545-25354	TCCCTGCACAAGAAAAGAACCTCACAACTGCTCTGCCATTGTCATGATGGAAAAGCAC	3250
	** * * * * * * * * * * * * * * *	

NC_019843.3:21456-25517	CTAATTGTATGCCCTGTTATGGCTACTTATTAAACATAAACAGTAGGATTGTTG	3544
NC_048212.1:20814-24623	TTGTTCCCTAGGGATGGTTGTTGTCAGGAAAACA-----ATGAATCC-----C	3286
NC_004718.3:21492-25259	ACTTCCCCTCGTAAGGTGTTTGTTTAAATGGCA-----CTTCT-----	3249
MT799526.1	ACTTCCCCTCGTAAGGTGTTTGTTTAAACGGCA-----CGCAC-----	3279
NC_045512.2:21563-25384	ACTTCCCCTCGTAAGGTGTTTGTTTAAATGGCA-----CACAC-----	3303
MN996532.1:21545-25354	ACTTCCCACGTGAAGGTGTTTGTTTAAATGGCA-----CACAC-----	3291
	* * * *	*
NC_019843.3:21456-25517	ATGAGTGGTCATAACTGGCTCGTCCTCTATGCACCTGAGCCCCATTACCTCCCTTAATA	3604
NC_048212.1:20814-24623	AAGTTTGGTATTTACTAAGGCAAGTTTACAATCCTGTAAGTTGCTTATGAGAAC	3346
NC_004718.3:21492-25259	-----TGGTTTATTACACAGAGGAACCTCTTCTCCAACAAATAATTACTACAGACAATA	3304
MT799526.1	-----TGGTTTGTAAACACAAAGGAATTCTATGAACCACAAATTATTACACGGACAATA	3334
NC_045512.2:21563-25384	-----TGGTTTGTAAACACAAAGGAATTCTATGAACCACAAATTACTACAGACAACA	3358
MN996532.1:21545-25354	-----TGGTTTGTAAACACAAAGGAATTCTATGAACCACAAATTATTACACAGACAACA	3346
	**** ** ** * ** * * ***	
NC_019843.3:21456-25517	CTAAGTATG-TTGCAACCACAGGTGACATACCAAAACATTCTACTAACCTCCCTCCT	3663
NC_048212.1:20814-24623	CTCATTTATTAGACACTTGTGGTGTTAATTACACTACAGTC---ATAATAGTGTGTTG	3402
NC_004718.3:21492-25259	CATTTGTCTCAGGAATTGTGATGTCGTTATTGGCATCATT---AACACACAGTTAT	3360
MT799526.1	CTTTGTCTCTGGTAGCTGTGATGTTGATTGGATTGTC---AACACACAGTTAT	3390
NC_045512.2:21563-25384	CATTTGTCTCTGGTAACCTGTGATGTTGTAATAGGATTGTC---AACACACAGTTAT	3414
MN996532.1:21545-25354	CATTTGTCTCTGGTAGCTGTGATGTTGTAATAGGATTGTC---AACACACAGTTAT	3402
	* * * ** * * **	
NC_019843.3:21456-25517	CTTCCTGGCAATTCCACCGGGATTGACTTCAAGATGAGTGGATGAGTTTTCAAAAT	3723
NC_048212.1:20814-24623	AACCCAATTGAA CCTCTTAATTATAATTTCAGGAAGAGTCGATAAGTATTCAAGAAC	3462
NC_004718.3:21492-25259	GATCCTCTGCAACCTGAGCTTGACTCATTCAGAAGAGCTGGACAAGTACTTCAAAAT	3420
MT799526.1	GATCCTTGTCAACCAGAACCTGATTGATTCAAGGAGGAGTTGGACAATATTAAAAAT	3450
NC_045512.2:21563-25384	GATCCTTGTCAACCAGAACCTGATTAGACTCATTCAGGAGGAGTTAGATAAAATTTAAGAAT	3474
MN996532.1:21545-25354	GATCCTTGTCAACCAGAACCTGATTCAAGGAGGAGTTGGATAAAATCTTTAAAAT	3462
	** * ** *	
NC_019843.3:21456-25517	GTTAGCACCGTATACCTAACTTGGTCCCTAACACAGATTAAACTACATTACTCGAT	3783
NC_048212.1:20814-24623	CAGTCCTCTCAGTTAAATATCCTTGTGATTTCTCAGTTAATGTTTCTATTGTTAAT	3522
NC_004718.3:21492-25259	CATACATCACCAAGATGTTGATCTGGCGACATTCAGGCATTACCGCTTGTGTCAAC	3480
MT799526.1	CATACATCACCAAGATGTTGATTTAGGTGACATTCAGGCGATTACCGCTTGTGTCAAC	3510
NC_045512.2:21563-25384	CATACATCACCAAGATGTTGATTTAGGTGACATCTCTGGCATTATGCTTCAAGTTAAC	3534
MN996532.1:21545-25354	CATACATCACCTGATGTAGATTAGGTGACATTCTGGCATTATGCTTCAAGTTGTCAAT	3522
	* *	

NC_019843.3:21456-25517	CTTACCTACGAGATGTTGTCCTCAACAAGTTAAAGGCCCTAATGAGTCTACATA	3843
NC_048212.1:20814-24623	TTGAATGAAACAGATGGCTGCTTAGATTCAAGTCGTTAAAGAGTCAATGAATCTTCATT	3582
NC_004718.3:21492-25259	ATTCAAAAAGAAATTGACCCTCAATGAGGTCGCTAAAAATTAAATGAATCACTCATT	3540
MT799526.1	ATTCAAGAAAGAAATTGACCCTCAACGAGGTTGCCAAAAATCTAAATGAATCTCATC	3570
NC_045512.2:21563-25384	ATTCAAAAAGAAATTGACCCTCAATGAGGTTGCCAAGAATTAAATGAATCTCATC	3594
MN996532.1:21545-25354	ATTCAAAAAGAAATTGACCCTCAATGAGGTTGCCAAAAATCTAAATGAATCTCATC	3582
	* * * * *	*
NC_019843.3:21456-25517	GACCTTAAAGAGCTTGCATTATACTTATTACAACAAATGGCCGTGGTACATTTGGCTT	3903
NC_048212.1:20814-24623	GACCTCAAGAACGCTAGGTGTCACCGCAGCACCAAACACACCCTGGTATGCGTGGTTA	3642
NC_004718.3:21492-25259	GACCTTCAGAAATTGGGAAAATATGAGCAATATATTAAATGGCCTTGGTATGTTTGGCTC	3600
MT799526.1	GACCTCCAGAAACTTGGAAAGTATGAGCAGTATATAAAATGGCCATGGTATTTGGCTA	3630
NC_045512.2:21563-25384	GATCTCCAGAAACTTGGAAAGTATGAGCAGTATATAAAATGGCCATGGTACATTTGGCTA	3654
MN996532.1:21545-25354	GATCTCCAGAAACTTGGAAAGTATGAGCAGTATATAAAATGGCCATGGTACATTTGGCTA	3642
	* * * * * * * * *	** * * * * * * * *
NC_019843.3:21456-25517	GGTTTCATTGCTGGCTTGTGCTTAGCTCTAGCTCTTCATACTGTGCTGCACT	3963
NC_048212.1:20814-24623	GGTATGATTGAGGTTAGTTGGACTCGCCTTAGCTGTTTATGCTTGTGCAATGACC	3702
NC_004718.3:21492-25259	GGCTTCATTGCTGGACTAATTGCCATCGTCATGGTTACAATCTGCTTGTGCACT	3660
MT799526.1	GGATTATTGCAAGGCTTGATAGCTATAATCATGGTTACAATCATGTTATGCTGATGACC	3690
NC_045512.2:21563-25384	GGTTTTATAGCTGGCTGATTGCCATAGTAATGGTGAACATTATGCTTGTGATGACC	3714
MN996532.1:21545-25354	GGTTTTATAGCTGGCTGATTGCCATAATAATGGTCACGATTATGCTTGTGATGACC	3702
	* * * * * * * * *	** * * * * * * * *
NC_019843.3:21456-25517	GGTTGTGGCACAAACTGTATGGGAAACCTAAAGTGTAACTGTTGTTGATAGATACGAG	4023
NC_048212.1:20814-24623	AAC TGCTGCAGTGGATTCAAGGGGCATATGCTCTGTAAAGCAATGCCAGTA--TGATGAT	3759
NC_004718.3:21492-25259	AGTTGTTGCAGTTGCCTCAAGGGTGCACTGCTCTGTGGTTCTGCTGCAA--GTTGAT	3717
MT799526.1	AGTTGCTGCAGTTGCTCAAGGGCTGTTCTGTGGCTCTGCTGTA--ATTGAT	3747
NC_045512.2:21563-25384	AGTTGCTGAGTTGCTCAAGGGCTGTTCTGTGGATCCTGCTGCAA--ATTGAT	3771
MN996532.1:21545-25354	AGTTGCTGCAGTTGCTCAAGGGCTGTTCTGTGGATCTGCTGCAA--ATTGAT	3759
	* * * * * * * * *	** * * * * * * * *
NC_019843.3:21456-25517	GAATAACGACCTCGAGCCGCATAAGGTTCATGTTCACTAA-----	4062
NC_048212.1:20814-24623	TATGCTGACGTTTACCCAGCTGTTCGAGTTAGTGGTAAACGAACAGTATGA	3810
NC_004718.3:21492-25259	GAGGATGACTCTGAGCCAGTCTCAAGGGTGCAAATTACATTACACATAA	3768
MT799526.1	GAAGACGACTCTGAGCCAGTACTCAAAGGAGTCAAATTACATTACACATAA	3798
NC_045512.2:21563-25384	GAAGACGACTCTGAGCCAGTCTCAAGGGAGTCAAATTACATTACACATAA	3822
MN996532.1:21545-25354	GAAGACGACTCTGAGCCAGTCTCAAGGGAGTCAAATTACATTACACATAA	3810
	* *** * **	*

Percent Identity Matrix (Genes)

```

#
#
# Percent Identity Matrix - created by Clustal2.1
#
#
1: NC_019843.3_21456-25517 100.00 44.16 45.49 45.55 46.19 45.96
2: NC_048212.1_20814-24623 44.16 100.00 47.18 46.94 47.71 47.12
3: NC_004718.3_21492-25259 45.49 47.18 100.00 73.12 73.35 73.26
4: MT799526.1 45.55 46.94 73.12 100.00 84.14 83.39
5: NC_045512.2_21563-25384 46.19 47.71 73.35 84.14 100.00 93.12
6: MN996532.1_21545-25354 45.96 47.12 73.26 83.39 93.12 100.00

```

Here,

1. SARS coronavirus
2. SARS CoV-2
3. P-CoV
4. MERS CoV
5. Bat coronavirus RaTG13
6. Bat coronavirus

Guide Tree

```
(  
NC_019843.3:21456-25517:0.358004  
,  
(  
NC_048212.1:20814-24623:0.346637  
,  
(  
NC_004718.3:21492-25259:0.175723  
,  
(  
MT799526.1:0.0961032  
,  
(  
NC_045512.2:21563-25384:0.0366142  
,  
MN996532.1:21545-25354:0.0366142  
):0.059489  
):0.07962  
):0.170914  
):0.011367  
)  
;
```

Phylogenetic Tree

```
(  
(  
(  
(  
NC_019843.3_21456-25517:0.28638,  
NC_048212.1_20814-24623:0.27198)  
:0.12247,  
NC_004718.3_21492-25259:0.13498)  
:0.05199,  
MT799526.1:0.08290)  
:0.04506,  
NC_045512.2_21563-25384:0.03191,  
MN996532.1_21545-25354:0.03686);
```

Multiple Sequence Alignment Proteins

Here,

1. SARS coronavirus
2. SARS CoV-2
3. P-CoV
4. MERS CoV
5. Bat coronavirus RaTG13
6. Bat coronavirus

CLUSTAL O(1.2.4) multiple sequence alignment

YP_009047204.1	MIHSVFLLMFLLTPTESYVDVGPDSVKSACIEVDIQQTFFDKTW-PRPIDVSKADGIIYP	59
YP_009824990.1	-MVVLLFVMFLSLVVA-----DRPNCYLQPQSYTISPATITNTSYFNVNPIIVP	49
NP_828851.1	--MFIFLFLTLTSGS-----DLDRCTTFDDVQAPNY-----TQHTSSMRGVYYP	43
QLR06869.1	-MLFFFFLHFALVNS-----QCVNLTGRAAIQP-----SFTNSSQRGVYYP	40
YP_009724390.1	--MFVFLVLLPLVSS-----QCVNLTTQLP-----AYTNSFTRGVYYP	39
QHR63300.2	--MFVFLVLLPLVSS-----QCVNLTTQLP-----AYTNSSTRGVYYP	39
	.:: : * . *	
YP_009047204.1	QGRRTYSNITITYQGLFPYQGDHGDMYVVSAGHATGTPQKLFVANYSQDVVKQFANGFVVR	119
YP_009824990.1	EFSVARTNLVVRQRLQVAAYD-----FTKIPIT-PPHRSIFN----NTYFRVGDGILVN	98
NP_828851.1	DEIFRSDTLYLTQDLFLPFYS--NVTGFHTN-----HTFG---NPVIPFKDGIVFA	90
QLR06869.1	DTIFRSNTLVLQSQGYFLPFYS--NVSWYYALTAKTN-SAEKRVD---NPVLDFKDGIYFA	93
YP_009724390.1	DKVFRSSVLHSTQDLFLPFFS--NVTWFHAIHVSGTNGTKRFD---NPVLVPNDGVYFA	93
QHR63300.2	DKVFRSSVLHLTQDLFLPFFS--NVTWFHAIHVSGTNGIKRFD---NPVLVPNDGVYFA	93
	: * . :	
YP_009047204.1	IGAAAANSTGTVIISPST-SATIRKIYPAFMLGSSVGN-----FSDGKMGRRFF----N	166
YP_009824990.1	THMYHR---TNIATNLGSSFYGCQEPPGAAGNTFVNNEPITLVMFYAGNVGSWSSLVTQQ	155
NP_828851.1	ATE-----KS-----NVVRGWVFGSTMNNKSQ-----SVIIINN	119
QLR06869.1	ATE-----KS-----NIVRGWIFGTTLDNTSQ-----SLLIVNN	122
YP_009724390.1	STE-----KS-----NIIRGWIFGTTLDSTKQ-----SLLIVNN	122
QHR63300.2	STE-----KS-----NIIRGWIFGTTLDSTKQ-----SLLIVNN	122
	: : : : * : :	
YP_009047204.1	HTLVLLPDGCGTLL--RAFYCILEPRSGNHCPAGNSYTSFATYHTPATDCSDGNYNRNAS	224
YP_009824990.1	QANITIVSCDNATLCANPFLRWGP-----GVIRSFT	187
NP_828851.1	STNVVIR-ACNFELCDNPFFAVSKP-----MGT---	146
QLR06869.1	ATNVIIK-VCNFQFCYDPYLSGYHH-----N-NKTWS	152
YP_009724390.1	ATNVIIK-VCEFQFCNDPFLGVYYH-----KNNKSWM	153
QHR63300.2	ATNVIIK-VCEFQFCNDPFLGVYYH-----KNNKSWM	153
	: : : : : : :	
YP_009047204.1	LNSFKEYFNLRNCTFMYTYNITEDEILEWFGITQTAQG---VHLFSSRYVDLYGGNMFI-	279
YP_009824990.1	I-----SNAFQCHGNYTFYDTKL--VNFTVSTAR---YNLAFTFADGDFIMY--YAAL	233
NP_828851.1	QTHTMIFDNAFNCTFEYISDAF-S--LDVSEKSGNFKHLREFVKFKNDGFLYVYKGYQPI	203
QLR06869.1	TREFAVYSSYANCTFEYVSKSFM--LDIAGKSGLFDTLREFVFRNVDGYFKIYSKYTPV	209
YP_009724390.1	ESEFRVYSSANNCTFEYVSQPF-L--MDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPI	210
QHR63300.2	ESEFRVYSSANNCTFEYVSQPF-L--MDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPI	210
	. : * * :: : *	

YP_009047204.1	-QFATLP-VYDTIKYSSIIIPHSI-----RSIQSDRKAW---AAFYVYKL	318
YP_009824990.1	GDSLNLPHTAY-PLQPYLRIAGFKPDVVQFFQAITRPN--GESAEVACASSTVSWYVSRL	290
NP_828851.1	DVVRDLPSGFNTLKPIFKLPLGINITNFRAILTAFS----PAQDIWGTSAAYFVGYL	257
QLR06869.1	NVNSNLPIGFSALEPLVLEIPAGINIKFRTLLTIHRGDPMP--NNGWTVFSAAYYVGYL	266
YP_009724390.1	NLVRDLPQGFSALEPLVLDLPIGINITRFQTLLALHRSYLTGPGDSSSGWTAGAAAYYVGYL	270
QHR63300.2	NLVRDLPQGFSALEPLVLDLPIGINITRFQTLLALHRSYLTGPGDSSSGWTAGAAAYYVGYL	270
	** : :: :* . : .::* * .::: * *	
YP_009047204.1	QPLTFLLDFSVVDGYIRRайдCGFNDLSQLHCSYESFDVEGVYSVSSFEAKPSGSVVEQA	378
YP_009824990.1	YYKELLVGYDSYSSIVNVSCSSDAESELQCLLGTAFANGVYLSNFRSSPTDNV-RIT	349
NP_828851.1	KPTTFLMLKYDENGTITDAVDCSQNPLAELKCSVKSFEIDKGIFYQTSNFRVVPNGDVVRFP	317
QLR06869.1	APRTFMLNYNENGTTITDAVDCALDPLSEAKCTLKSLTVKGIYQTSNFRVQPTESIVRFP	326
YP_009724390.1	QPRTFLLKYNENGTTITDAVDCALDPLSETKCTLKSLTVKGIYQTSNFRVQPTESIVRFP	330
QHR63300.2	QPRTFLLKYNENGTTITDAVDCALDPLSETKCTLKSLTVKGIYQTSNFRVQPTDSIVRFP	330
	:: : . . * . * . : :: :* : :: . : : . * : * . : .	
YP_009047204.1	E-GVECDFSPS-LSGTPPVYNFKRLVFTNCNYNLTKLLSLFSVNDFTCSQISPAAIASN	436
YP_009824990.1	NSASSCSVPYS-VLSRPPLPFVWKRYAISNCKDFQALLSHLPTFQLRCFGISPTKLATM	408
NP_828851.1	NITNLCPFGEVFNATKFPSVYAWERKKISNCVADYSVLYNSTFSTFKCYGVSATKLNDL	377
QLR06869.1	NITNLCPFGEVFNATTFAFASVYAWNRKRISNCVADYSVLYNSTSFSTFKCYGVSPTKLNDL	386
YP_009724390.1	NITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDL	390
QHR63300.2	NITNLCPFGEVFNATTFAFASVYAWNRKRISNCVADYSVLYNSTSFSTFKCYGVSPTKLNDL	390
	: * . : : :* : :** : * . : * : * : :	
YP_009047204.1	CYSSLILDYFSYPLSMKSDLSVSSAGPISQFNYKQSFSNPTCLILATVPHNLTTITKPLK	496
YP_009824990.1	CFGTVTLIMLVNVTHYNNLLNDVPPDDFSLYNYQLPRNFTYGLHSYYLPNTA-----	461
NP_828851.1	CFSNVYADSFVVKGDDVRQIAPGQTGVIADYNYKLPDDFMGCVLAWSNTRNIDATSTGNYN	437
QLR06869.1	CFTNVYADSFVVRGDEVRQIAPGQTGRIADYNYKLPDDFTGCVIAWNSNNLDSKVGGYN	446
YP_009724390.1	CFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGYN	450
QHR63300.2	CFTNVYADSFVITGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNKHIDAKEGNFN	450
	*: : . * : : : . . : :** : . * : : :	
YP_009047204.1	YSYINKCS-RLLSDDRTEV-----PQLV--NANQYSPCIVPS-TVWEDGDYYRKQL	545
YP_009824990.1	FSYTVASRIRYPSWVHSITPGGRQPVGPFLDSLQSSSKPCTGS-----	504
NP_828851.1	YKYRLRGKLRPFERDIS-----NVPFSPDGKPCTP-PALNCYW-----	476
QLR06869.1	YLYRLFRKSNLKPFERDIS-----TEIYQAGSTPCNGVEGFNCYF-----	486
YP_009724390.1	YLYRLFRKSNLKPFERDIS-----TEIYQAGSTPCNGVEGFNCYF-----	490
QHR63300.2	YLYRLFRKANLKPFERDIS-----TEIYQAGSKPCNGQTGLNCYY-----	490
	: * . : : **	

YP_009047204.1	SPLEGGGWLVASGSTVAMTEQLQMGFGITVQYGTDNSVCPLKFANDTKIASQLGNCVE	605
YP_009824990.1	-----CLGLAVI----SLSIASANKLVCVPVN-----DTDIVPDTCVN	538
NP_828851.1	-PLNDYGFYTTTGIGYQPYRVVVL---SFELLNAPATVCGPKL-----STDLIKNCVN	526
QLR06869.1	-PLQSYGFHPTNGVGQPYRVVVL---SFELLNAPATVCGPKQ-----STNLVKNKCVN	536
YP_009724390.1	-PLQSYGFQPTNGVGQPYRVVVL---SFELLNAPATVCGPKK-----STNLVKNKCVN	540
QHR63300.2	-PLYRYGFYPTDGVGHQPYRVVVL---SFELLNAPATVCGPKK-----STNLVKNKCVN	540
	: : . . ** . . . **:	
YP_009047204.1	YSLYGVSGRGVFQNCTAVGVRQQRFVYDAYQNL-VGYYSDDGNNYCLRACSVPVSVIYD	664
YP_009824990.1	YNIYGYQGTGVISRNSYTLPPSSKVFSLSSGE--LTVFAVGSSFYQLSPCAFAPISAAFY	596
NP_828851.1	FNFNGLTGTGVLTTPSSKRFQPFQQFGRDVSDFTSDVRDPKTSEILDISPACAFGGVSITP	586
QLR06869.1	FNFNGLTGTGVLTTESSKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSITP	596
YP_009724390.1	FNFNGLTGTGVLTESNKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSITP	600
QHR63300.2	FNFNGLTGTGVLTESNKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSITP	600
	: . * * **: .. : * . . : * . : * : * . :	
YP_009047204.1	KE-TKTHATLFGSVAEHISSMSQYSRSTRMLKRRDSTYGPLQTPVGCVLGLVNS-S	721
YP_009824990.1	---KGYTTSLLFNNLPCSNRHRAV-VEP--VSAYWRRSVADNNTFDTTAGCIFNAYNLTS	650
NP_828851.1	GTNASSEVAVLYQDVNCTDVSTAIIHADQ--LTPAWRIYSTGNNVFQTQAGCLIGAEHVDT	644
QLR06869.1	GTNTSNQVAVLYQDVNCTEVPAIHADQ--LTPTRWVYSTGSNVFQTRAGCLIGAEHVNN	654
YP_009724390.1	GTNTSNQVAVLYQDVNCTEVPAIHADQ--LTPTRWVYSTGSNVFQTRAGCLIGAEHVNN	658
QHR63300.2	GTNASNQVAVLYQDVNCTEVPAIHADQ--LTPTRWVYSTGSNVFQTRAGCLIGAEHVNN	658
	: *: : * . :: : : : . : * . : * . : * . : :	
YP_009047204.1	LFVEDCKLPLGQSLCALPDTPSTLTPRSVRSVPGEMRLASIAFNHPIQV-DQLNSSYFKL	780
YP_009824990.1	IVVNQCDLPIGDSYCLQPSLIKGE-AT-----LSLVTYNP-LADSLTPITPVYQV	699
NP_828851.1	--SYECDIPIGAGICASYHTVSL---LRSTSQ---KSIVAYTMSLGADSSIAYSNNTI	694
QLR06869.1	--TYECDIPIGAGICASYQTQTN-----RSVSS---QAIIAYTMSLGAENSVAYANNSI	704
YP_009724390.1	--SYECDIPIGAGICASYQTQTNSP-RRARSVAS---QSIIAYTMSLGAENSVAYSNNSI	712
QHR63300.2	--SYECDIPIGAGICASYQTQTN-----RSVAS---QSIIAYTMSLGAENSVAYSNNSI	708
	: *: : * . * . : : : . : : . : :	
YP_009047204.1	SIPTNFSFGVTQEYIQTTEIQKVTVDCKQYVCNGFQKCEQLLREYGQFCSKINQALHGANL	840
YP_009824990.1	SVPTNFTLVASTEYIQTYYASKISIDCAKYLCGDSSQCRTVLLQYGTFCNDVNVALTRVFT	759
NP_828851.1	AIPTNFSISITTEVMPVSMAKTSVDCNMYICGDSTECANLLLQYGSFCTQLNRALSGIAA	754
QLR06869.1	AIPTNFTISVTTEILPVSMKTSVDCTMYICGDSTECNSNLLQYGSFCTQLNRALTGIAV	764
YP_009724390.1	AIPTNFTISVTTEILPVSMKTSVDCTMYICGDSTECNSNLLQYGSFCTQLNRALTGIAV	772
QHR63300.2	AIPTNFTISVTTEILPVSMKTSVDCTMYICGDSTECNSNLLQYGSFCTQLNRALTGIAV	768
	: : *: : * . * : * . : * : * . : * : * : * . : * : * . : * : * . : *	

YP_009047204.1	RQDDSVRNLFASVKSSQSSPIIPFGGGDFNLTLEPVSISTGSRARSARSAIEDLLFDKVTL	900
YP_009824990.1	LLDNSLVDTFSSLKSTA-PVQLAYTGFDFNFTSLVGCGIGTDCDSKSHRSALS DLLFSKVSV	818
NP_828851.1	EQDRNTREVFAQVKQMYKTPTLKYGFFNFSQLI-----PDPLKPTKRSFIEDLLFNKVTL	810
QLR06869.1	EQDKNTQEVF AQVKQIYKT PPIKDFGGFNFSQLI-----PDPSKPSKRSFIEDLLFNKVTL	820
YP_009724390.1	EQDKNTQEVF AQVKQIYKT PPIKDFGGFNFSQLI-----PDPSKPSKRSFIEDLLFNKVTL	828
QHR63300.2	EQDKNTQEVF AQVKQIYKT PPIKDFGGFNFSQLI-----PDPSKPSKRSFIEDLLFNKVTL	824
	* . : *.*. : * . :: . . : ** .****.*::	
YP_009047204.1	ADPGYMQGYDDCMQQG-PASARDLICAQYVAGYKVLPLMDVNMEAAYTSSLLGSIAGVG	959
YP_009824990.1	ADPGFMQSYYQQCLDAQWGGNIRDLLCTQT FNGISVLPIVSPSMQALYTTALVGGIAASG	878
NP_828851.1	ADAGFMKQYGECLGDI---NARDLICAQKFNGLTVLPPLL TDEMIAAYTAALVSGTATAG	867
QLR06869.1	ADAGFIKQYGDCLGDI---AARDLICAQKFNGLTVLPPLL TDEMIAQYTSALLAGTITSG	877
YP_009724390.1	ADAGFIKQYGDCLGDI---AARDLICAQKFNGLTVLPPLL TDEMIAQYTSALLAGTITSG	885
QHR63300.2	ADAGFIKQYGDCLGDI---AARDLICAQKFNGLTVLPPLL TDEMIAQYTSALLAGTITSG	881
	** *::: * *: ***: * . * .****: . * * *::*: * .	
YP_009047204.1	WTAGLSSFAAIPFAQSIFYRLNGVGITQQVLS ENQKLIANKF NQALGAMQTGTTTNEAF	1019
YP_009824990.1	FTFGVSSAAVIPFA TQLQFRLNGLGVTTNVLMENQQLIANAFN KALVSIQEGFTATNQAL	938
NP_828851.1	WTFGAGAALQIPFAMQMAYRFNGIGVTQNVL YENQKQIANQFNKAISQIQESLTTSTAL	927
QLR06869.1	WTFGAGAALQIPFAMQMAYRFNGIGVTQNVL YENQKLIANQFNSAIGKIQDSSLSTASAL	937
YP_009724390.1	WTFGAGAALQIPFAMQMAYRFNGIGVTQNVL YENQKLIANQFNSAIGKIQDSSLSTASAL	945
QHR63300.2	WTFGAGAALQIPFAMQMAYRFNGIGVTQNVL YENQKLIANQFNSAIGKIQDSSLSTASAL	941
	:* * .: **** .: :*::*: *: ***: *** ***: *: * .::: * *:	
YP_009047204.1	QKVQDAVNNAQALSKLASLSNTFGAISASI GDI IQRLDVLEQDAQIDRLINGRLTTLN	1079
YP_009824990.1	NKIQTVVNNNALQ LQVLVQQLGNTFGAISASVNEIF SRLDLLEANAEVDRLISGRMVVLN	998
NP_828851.1	GKLQDVVNQNAQALNTLVKQLSSNFGAISSV LNDILSRLDKVEAEVQIDRLITGRLQSLQ	987
QLR06869.1	GKLQDVVNQNAQALNTLVKQLSSNFGAISSV LNDILSRLDKVEAEVQIDRLITGRLQSLQ	997
YP_009724390.1	GKLQDVVNQNAQALNTLVKQLSSNFGAISSV LNDILSRLDKVEAEVQIDRLITGRLQSLQ	1005
QHR63300.2	GKLQDVVNQNAQALNTLVKQLSSNFGAISSV LNDILSRLDKVEAEVQIDRLITGRLQSLQ	1001
	*: * .***: * . * .:::*****: .::: * .***: *: * .:::*****: *: *:	
YP_009047204.1	AFVAQQLVRSESAALSAQLAKDKVNECVKAQSKRGFCGQGTHIVSFVVNAPNGLYFMHV	1139
YP_009824990.1	TYVTQQLIQAELRSQ AELAKQKMSEC VKSQSLRNDFCGNGTHVLSIPQLAPNGMLFIHY	1058
NP_828851.1	TYVTQQLIRAAEIRASANLAATKMSEC VLGQS KRVD FCGKGYHLM SFPQAAPHGVFLHV	1047
QLR06869.1	TYVTQQLIRAAEIRASANLAATKMSEC VLGQS KRVD FCGKGYHLM SFPQSAPHGVFLHV	1057
YP_009724390.1	TYVTQQLIRAAEIRASANLAATKMSEC VLGQS KRVD FCGKGYHLM SFPQSAPHGVFLHV	1065
QHR63300.2	TYVTQQLIRAAEIRASANLAATKMSEC VLGQS KRVD FCGKGYHLM SFPQSAPHGVFLHV	1061
	:*:* *::: . . *:** *:*** .** * .***: * *::: * *::: * *:	

YP_009047204.1	GYYPSNHIEVV SAYGLCDAANPTNCIAPVNGYFIKTNTRIVDEWSYTGSSFYAPEPITS	1199
YP_009824990.1	SYQPTKYAQVYTTAGLCFNGT---GFVPRDGLFVRENNE--SQWYFTKASFYNPVNLSY	1113
NP_828851.1	TYVPSQERNFTTAPAICHEGK---AYFPREGVFVFN----GTSWFITQRNFFSPQIITT	1099
QLR06869.1	TYVPSQEKNFTTPAICHEGK---AHFPREGVFVSN----GTHWFVTQRNFYEPQIITT	1109
YP_009724390.1	TYVPAQEKNFTTAPAICHDGK---AHFPREGVFVSN----GTHWFVTQRNFYEPQIITT	1117
QHR63300.2	TYVPAQEKNFTTAPAICHDGK---AHFPREGVFVSN----GTHWFVTQRNFYEPQIITT	1113
	* *:: : . :: .*: .. * ;* *: * * . * : * ::	
YP_009047204.1	LNTKYVAPQ-VTYQNISTNLPPPLLGNSTGIDFQDELDEFFKNVSTSIPNFGSLTQINTT	1258
YP_009824990.1	ENTHLLDTCGVNYTTVNNSVLNPIE--PPNYNFQEEFDKYFKNQSSQFNITFDSSQFNVS	1171
NP_828851.1	DNTFVSGNCDDVVIGIINNTVYDPLQ--PELDSFKEELDKYFKNHTSPDVLDGDISGINAS	1157
QLR06869.1	DNTFVSGSCDVVIGIVNNTVYDPLQ--PELDSFKEELDKYFKNHTSPDVLDGDISGINAS	1167
YP_009724390.1	DNTFVSGNCDDVVIGIVNNTVYDPLQ--PELDSFKEELDKYFKNHTSPDVLDGDISGINAS	1175
QHR63300.2	DNTFVSGSCDVVIGIVNNTVYDPLQ--PELDSFKEELDKYFKNHTSPDVLDGDISGINAS	1171
	** * . * : * . * : * * : * * * : * * : . : . * . :	
YP_009047204.1	LLDLTYEMLSLQQVVKALNESYIDLKELGNYTYYNKWPWYIWLGFIAGLVALALCVFFIL	1318
YP_009824990.1	IVNLNEQMAALDSVVKSLNESFIDLKKGKVYEQPNTPWYAWLGMIAGLVGLALAVFMLC	1231
NP_828851.1	VVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYVWLGFIAGLIAIMVTILLC	1217
QLR06869.1	VVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIMVTIMLC	1227
YP_009724390.1	VVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIMVTIMLC	1235
QHR63300.2	VVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIMVTIMLC	1231
	: . . . : * : . * . * * * * : * * : * : * : * : * : * : . . . :	
YP_009047204.1	CCTGCGTNCMGKLKCNRCCDRYEYDLEPHKVHVH----	1353
YP_009824990.1	CMTNCCSGFRGICCSCKQCQYD-DYADVYPAVRVSGKRTV	1269
NP_828851.1	CMTSCCSCLKGACSCGSCCKF-DEDDSEPVLKGVKLHYT	1255
QLR06869.1	CMTSCCSCLKGCCSCGSCCKF-DEDDSEPVLKGVKLHYT	1265
YP_009724390.1	CMTSCCSCLKGCCSCGSCCKF-DEDDSEPVLKGVKLHYT	1273
QHR63300.2	CMTSCCSCLKGCCSCGSCCKF-DEDDSEPVLKGVKLHYT	1269
	* *.* : * . * * : * * : * : * :	

Percent Identity Matrix (Proteins)

```

#
#
# Percent Identity Matrix - created by Clustal2.1
#
#


 1: YP_009047204.1  100.00   30.63   32.53   32.32   32.10   32.29
 2: YP_009824990.1   30.63  100.00   34.68   34.00   33.72   33.89
 3: NP_828851.1     32.53   34.68  100.00   76.82   77.22   77.68
 4: QLR06869.1      32.32   34.00   76.82  100.00   90.59   89.64
 5: YP_009724390.1   32.10   33.72   77.22   90.59  100.00   97.71
 6: QHR63300.2      32.29   33.89   77.68   89.64   97.71  100.00

```

Here,

1. SARS coronavirus

2. SARS CoV-2
3. P-CoV
4. MERS CoV
5. Bat coronavirus RaTG13
6. Bat coronavirus

Guide Tree

```

(
YP_009047204.1:0.375387
(
YP_009824990.1:0.367809
(
NP_828851.1:0.129681
(
QLR06869.1:0.0535573
(
YP_009724390.1:0.0126084
,
QHR63300.2:0.0126084
):0.040949
):0.076124
):0.238127
):0.00757837
)
;

```

Phylogenetic Tree

```

(
(
(
YP_009047204.1:0.35565,
YP_009824990.1:0.33806)
:0.20606,
NP_828851.1:0.11104)
:0.06749,
QLR06869.1:0.05062,
(
YP_009724390.1:0.00985,
QHR63300.2:0.01300)
:0.03685);

```

Ans 1b)

From the results above (Percentage Identity Matrix) we can infer that:

The source of **SARS-COV-2** is likely **SARS-COV**

The source of **MERS-COV-2** is likely a **Bat Coronavirus RaTG13**

QUES 2)

Programs used are:

Parisimony: dnaps

Distance-based: dnadist → neighbor

Maximum Likelihood Method: dnaml

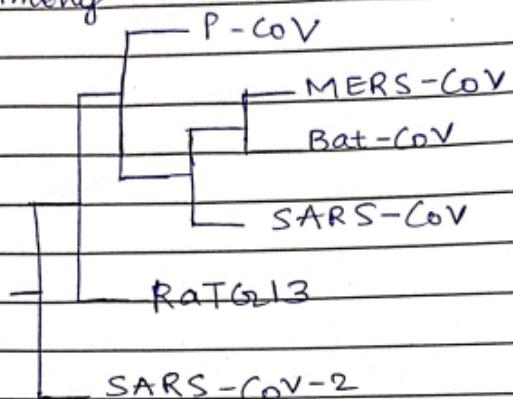
For bootstrapping: seqboot

Phylogenetic Trees:

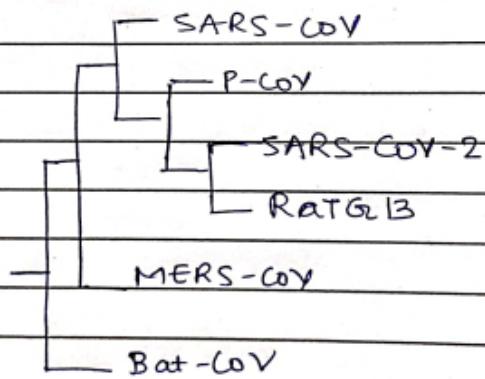
DNA

With Bootstrapping

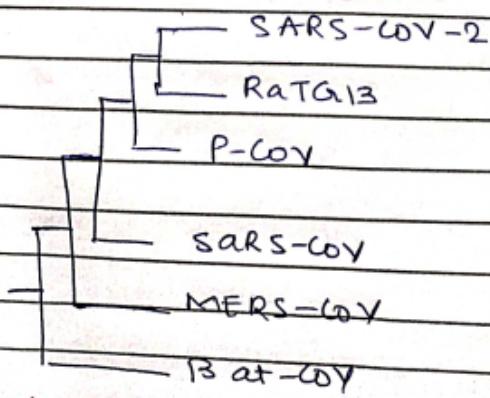
→ Parsimony



→ Distance



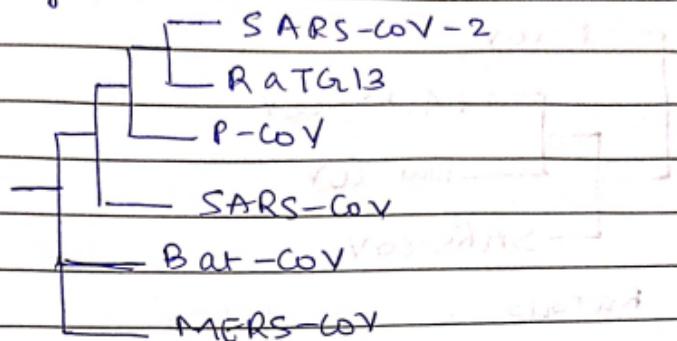
→ Max.-likelihood



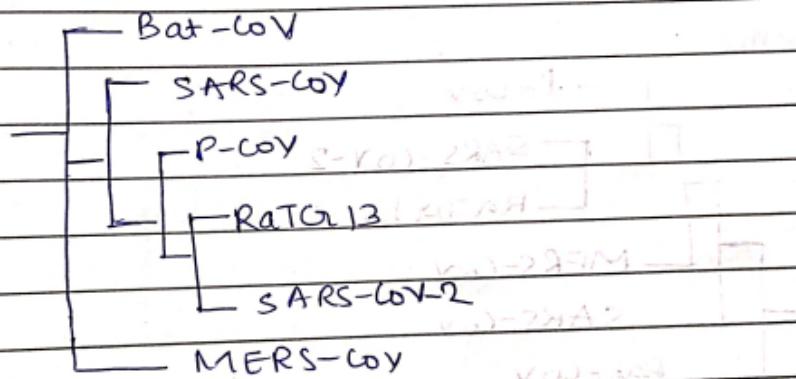
DNA

without Bootstrapping

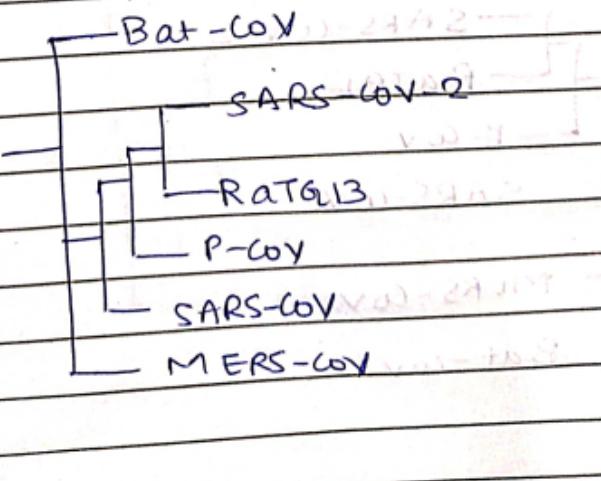
→ Parsimony



→ Distance

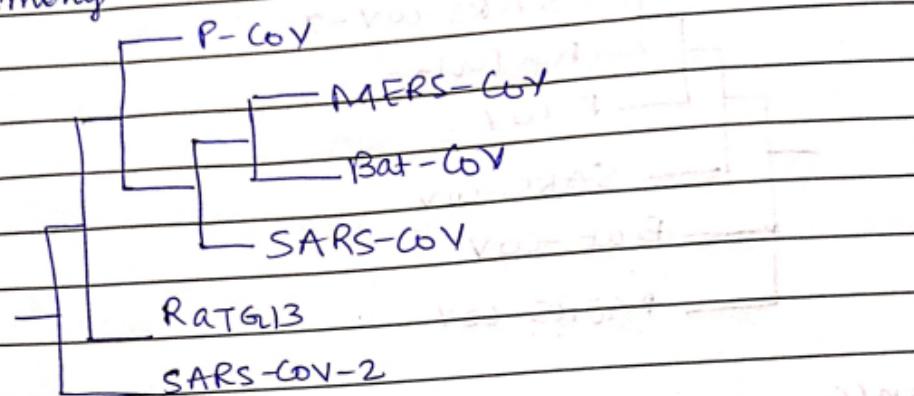


→ Max-likelihood

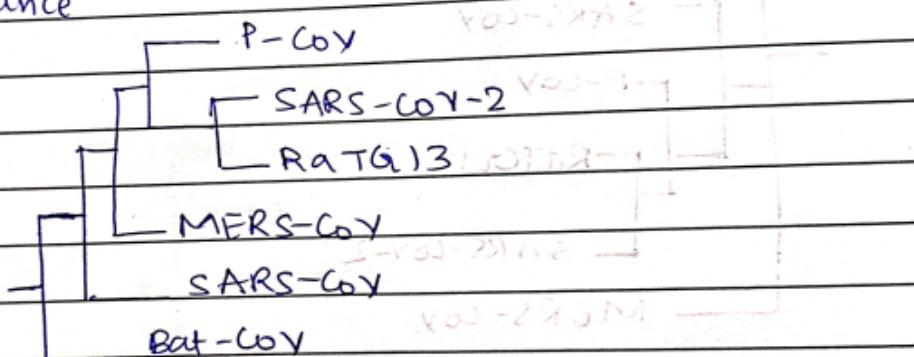


Protein
with Bootstrapping

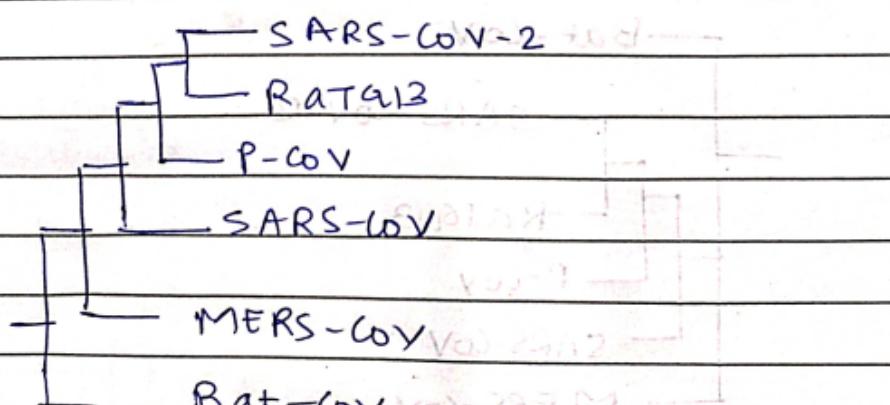
→ Parsimony



→ Distance



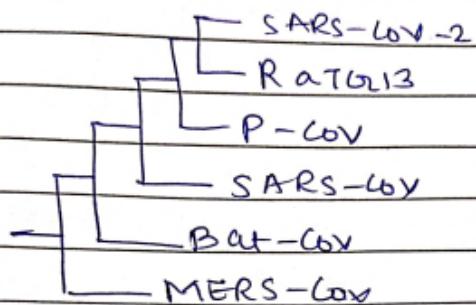
→ max-likelihood



protein

without Bootstrapping

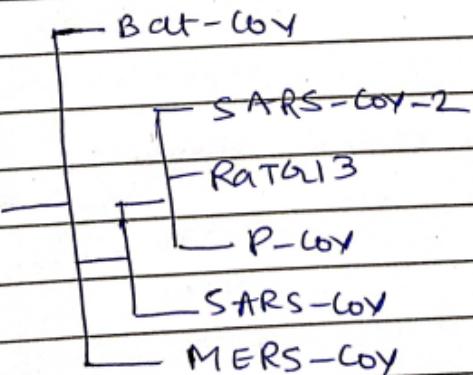
→ Parsimony



→ Distance



→ Max-Likelihood



Ans 2a)

By observation from the above phylogenetic trees, we can say that the trees agree in topological aspect.

Ans 2b)

When comparing the trees with and without bootstrapping, there are differences in the inferred relationships and the support for the branching patterns. Without bootstrapping, the inferred relationships may be based on a single alignment and may not account for uncertainty in the data. With bootstrapping, the inferred relationships are based on multiple alignments and provide a more accurate representation of the data.

Bootstrapping helps to assess the statistical significance of the inferred relationships and provides a measure of confidence in the branching patterns of the phylogenetic tree. It helps to prevent overinterpreting the results based on a single alignment and provides a more accurate representation of the evolutionary relationships among the sequences.

Ans 2c)

On observation, from the trees in both DNA and protein sequences, our inference from Q1b is correct. That is:

SARS-CoV2 is more closely linked to BAR RatG13

MERS-CoV is more closely linked to Bat-CoV

QUES 3)

Sequence1:

CCACACAAAAAAAGTCGAGAAAATATGAGGAACAAAAAGCGAAGACGACAAAAAAAAAAACTCTGTTTGTATCTCTATCTGAGAGGCACACATTGCTCTTCTCAATTATTTA TTGGTTCTCCTACCTATCTCGATCTCAATTCTCCCATTCTCTCTCAAGTCAAATTCTTGAG AATTAGCTTACCGAATTCTGCTCCGATAACTAGTGGATGATGATTACCCCAAATCTTCTGCT CAAGGTAATTCTGAGAAATTCTCAAATTCAAACCGCATGGTCCGGAGTCGGGGTAGTGGC GGTGGCGTGGCGTGGCGGGAGAAGAAGAACCGTCGTCAAGTCACACTCTAATAACCGAAGAG GAGGAGAACAGCTAACGTCGGGAACGAAATTCTCAGACCAAGAACACTGAATGAAGCAA AGCAATTCAACAGTACACCGTCGACGCAAGACTCACGCCGTTTCAACAATCGGGAAATCAGGAA TCATTGACTACTACAATCACTCAAAAGACGACGTAACGGTCTCTGTACTTGAGAACAGATCACAG CTTATCTCTCGAATCCAGGAGGTGGTACATTCAAGCTTCTGGATGATGATGCCGTGATGAATC CAGTTCCGGATCATGGTTACAGTGGAAACGCCAGAGAAATTGTTAGGGATTATGCCCAATCTGTTCT ACTCTGAGAACACTGAGATTCTAGCTATGGGAACGACTGATGAGATCTTGTACTTCAAGCTGA TTCTACTCGAGCGTCTTCGTTGCTGAGAGATTACCTGTTAAATCCGTTGGATCCATTCAAGAA TACTGGTAAACCCTTACGCCATTCTCATAGGATTGATTTGGTGTGTTATTGATTAGAGCCAGCT AGAACTGAAGATCCTGCGCTTCTATTGCTGGTGTGTTCAATCGAGAAACTCGCGTGTGCGATT CTCAGTTACAGGCTCTCTGGAGATATTAGCTTGTGAGACTGTCGTGAAAGTGTGAGGGAA CTTGACTGTTATGATCGTGTATGGTTATAAGTTCATGAAGATGAGCATGGAGAACGTTAGCTGAG AGTAAACGAGATGATTTAGAGCCTTATTTGGACTGCAATTCTGCTACTGATATTCTCAAGCGTCAA GGTTCTGTTAAGCAGAACCGTGTCCGAATGATAGTAGATTGCAATGCCAACCTGTTGTTCA GA CGATAGGCTAACATCAGTCTATGTCGTGGTTCTACTTCTAGGGCTCTCATGGTTGACTCT CAGTATATGGCTAACATGGGATCTATTGCTCTTAGCAATGGGGTTATAATCAATGGAAATGAAGATG ATGGGAGCAATGAGCTAGTGGAGAACGCTGATGAGGCTTGGGGTTGGTGTGCCATCACACTC TTCTCGCTGCATACCGTTCCGTAAGGTATGCTGTGAGTTTGATGCAAGCTTGGTTACAGTT AACATGGAATTGCAAGTGGTTGCAAATGTCAGAGAACCGCTTGGAGAACGCAAGACACTGTTATGT ATATGCTTCTGCTGACTCGCTGCTGGATTGTTACACAGACTCCAGTATCATGGACTTAGTGAATG TGACGGTGCAGCTTCTTACACGGGAGTTACCGCTGGGTGCTCTAGTGAAGTTAGATA AACAGATGTTGGAGGTGGTCTGCTGGAATCATGGGATTCAACCGGATTAAGCACTGATAGTTAGCG ATGCGGGGTACCGGGTGCAGCTGCTGGGGATGCTGTGCGGTATGGCAGTTGCAATATACAAA AAAGAGACTTTCTTCTGGTGTGATCTACACTGCGAAAGAAATCAATGGGGAGGCGCTAACATCAT CCGGAGGATAAGATGATGGCACGAAATGCAATCTCGTGTGCTCTTCAGGCTTCTGAAGTTGTTA AGAGCCGGAGTCAGCATGGGAAACTCGGAAATGGATGGGATTCACTCGCTCAGCTTATTCTGAGAGA CTCTTTAAAGAATGAGGCGCTATGACTCTAAAGTTGAGGTTGAGCTGGGATGTTGAGGCTATTGAGAGA ATGGCGGGGGAACAGGGGATTGATGAGTTAGGTGAGCTGGGATGTTGAGGCTATTGAGACTG CAACTGTTCTATATTGCTGTGGATGCCGGAGGCTGCATCAATGGATGGAACGCTAACATTGAGAGTT GACAGGTCTCAGTTGAAGAAGCTATGGGAAGTCTCTGGTTCTGATTTAATATAACAAAGAGAATGAA GCAACTGTCAATAAGCTTCTCTCGTCTTGAGAGGTATATTCACTTCAAGCTATGTTGATCTGC GGTGTATATCCAATTGCGGGTATTGATTATTGCTGATTGCAATTGCAATGCAAGGGGACGAGGAAAGA ATGTTGAGGTTAACGCTAACACTTCAAGCCCAGACTACAAGGGAAAGCAGTTTGTGGTTGAAATGC TTGTTCCAGCAAGGACTACTGAACAAATTGCGCGTTGTTGGACAAGACGTTACTAGTCAG AAAATCGTAATGGATAAGTTCATCAACATACAAGGAGATTACAAGGCTATTGATCATAGCCAAACCTC TAATCCGCCAATTGCTGCTGACGAGAACACGTGCTGCTGGAAATGGAACATGGCAGTGGAAAGCT TACGGGTTGCTGCTGAGTGAAGTGGAAAATGATTGCTGGGGAAAGTGTGTTGGAGCTGTTGATG CTAAAGGGTCTGATGCTTAAACAGTTCATGATTGATTTGCTAATGCGATGGGGTGGCCAAGATAACGG ATAAGTTCCCTTCCATTGACCGCAATGGGAAGTTGTTTCAAGGCTCTTCTGTTGCAAAACAAGCG GGTTAGCCTCGAGGGAAAGTTATTGGGCTTCTGTTCTGCAAAATCCGAGGCCCTGAGCTGCGCAA GCTTCTAGCAGTCCAAAGGGAGGAGCACACAGGTGTTCAAGGCAAAGAGTTGGCTTATTTGTC AGGTGATAAAAGAATCTTGAGGGTATGCTGGGGAGGCTCAAGGCTAACAGACTTGAACGA GGACCGAGAACGAGTTACTTGAAACAAGTGGTTCTTGCGAGAACAGATCTCAAGGATGTCGGGACATG GATCTGAAAGCATTGAAGACGGTGAGTATGTTAGAATTCTAGAAGCTAGTTGCTTACTTCAC AAAATGTGACCAAATCCAAATTGTTTCTTGTGATGTTAGCTAAAGGGACAGAGGCTTCAG CTTCTGGGAAGTGTCTAAACCGCAGGAGATCAAATCAATAGAGGTTTGGAGGACAGATAAGGATTCAACAGC CTGATCCGTGACATTCCGAAGAGATCAAATCAATAGAGGTTTGGAGGACAGATAAGGATTCAACAGC TCCCTGGCTGAGTTCTGCTGAGTATAATCCGTATGCAACCATCTCAAGAGTGGGGAGATCCATTAAAG CCAACTTTCAAGCAAATGGCTGATGGATTGCGGCCATCCGACAGAATTCAAGGATCTTCAATTGTC CCGCTGTTGCTCCACATATCCATAACCAAATTATGCAATCCGGTTTGGTCTTATTTGACA TAAAGAAAATGAATTGGTTGGTTAATTACGAATTGATTTAGGCGTTAAAGAATTGAGGTTAAC CAATTCACTATTGTTGGTTATTGTTAGTGGAAACCTAGATTAGTTGATTTGTTATTGGTTAG TCGACTTGGGAACCTTGTAGACACATCCATAGGCCCTAGAATTAGCAGTCAGGAATGTAATGTTTCAAT TGATGAAAACAGCTCAAAGTGTAAACACTGGGTTCTGTTGCTGAGCTAGTCCGAGACATGTTCCATAGCAG TTGTTGCGAGATGGCGTCCAGGTGAGGTCTGCCCTCAGAGCTAGTCCGAGACATGTTCCATAGCAG CAGGTGGACAAGCCCTGAAGGTTAGGCTAAGCGTATGCGAAAGATTAAAGCTAATGAAACGGTGA GTTCAATACATCCGAGAATCAGAACCGTCTATTCTCATCTTCTGCAACTCCCTGACCTCGAAAGC GACATTGTCACGCTAGTGGAGGCTGAGCATGTCGATGTCGACATGGCTACATTAGTCACACTTCAGTT GTAGGAGGTTGATCTTGTGAGTTGATGAGTTGTTGCTGCAACGACGTCGGAGGAGGATAGAAAGTTT TTTTGTTCGGTAGGATTAGTAGAGAACAGGGAGATTATTGCGTTAGCTCAGCTCAGCTCGCCGGAAAAAA AACGTAACAGTAGTTGAGAGAACAGACTTTGTTGCTGTTGTAATTGACAACACTCCGAGAGAA ACAAAACATGAGATAAGAACAGAGCATATTATGATGACCAATTCTTAAATTGTTACTCGTTACAT CAATTCAAATTATTA

Sequence 2:

```
ACTTTGATGTCACTTAGCTTAAATTCTGTTAATTGCATTAAATCAAAAACAAACAAAA  
AACAAAAGAAAAACAAAAAAATCAAACAAACAAAAACAAAATAAGGCTGAAAAGC  
CCGAAGACTCTTTGTTTTGTATTGTATTGTATTGTATTATATACAT  
AATCTGACTAATGTAAGATTGAATCAAATTAGCCTGTGACAAGCAGAC
```

On performing Blast Search for **Sequence 1** we get:

Job Title Seq1
RID 4JT789MT016 Search expires on 04-28 01:52 am Download All ▾
Program BLASTN ⓘ Citation ▾
Database nt See details ▾
Query ID lclQuery_1535
Description None
Molecule type dna
Query Length 4706
Other reports Distance tree of results MSA viewer ⓘ

Filter Results
Organism only top 20 will appear exclude
Type common name, binomial, taxid or group name
+ Add organism

Percent Identity E value Query Coverage
[] to [] [] to [] [] to []
Filter Reset

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments Download Select columns Show 100 ⓘ

Description		Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	Arabidopsis thaliana genome assembly_chromosome_2	Arabidopsis thaliana	8691	8691	100%	0.0	100.0%	20021703	LR792543.1
<input checked="" type="checkbox"/>	Arabidopsis thaliana chromosome_2	Arabidopsis thaliana	8691	8691	100%	0.0	100.0%	22468181	CP116281.1
<input checked="" type="checkbox"/>	Arabidopsis thaliana genome assembly_chromosome_2	Arabidopsis thaliana	8691	8691	100%	0.0	100.0%	23012915	QX298803.1
<input checked="" type="checkbox"/>	Arabidopsis thaliana chromosome_2	Arabidopsis thaliana	8691	8691	100%	0.0	100.0%	22868883	CP087127.2
<input checked="" type="checkbox"/>	Arabidopsis thaliana isolate i21_salk_col chromosome_2	Arabidopsis thaliana	8691	8691	100%	0.0	100.0%	22217084	CP096025.1
<input checked="" type="checkbox"/>	Arabidopsis thaliana chromosome_2	Arabidopsis thaliana	8691	8691	100%	0.0	100.0%	19698289	CP002685.1
<input checked="" type="checkbox"/>	Arabidopsis thaliana chromosome 2 clone MSF3 map CIC06E08_complete sequence	Arabidopsis thaliana	8691	8691	100%	0.0	100.0%	88413	AC005724.3
<input checked="" type="checkbox"/>	Arabidopsis thaliana genome assembly_chromosome_2	Arabidopsis thaliana	8685	8685	100%	0.0	99.98%	19416174	LR699771.1
<input checked="" type="checkbox"/>	Arabidopsis thaliana genome assembly_chromosome_2	Arabidopsis thaliana	8685	8685	100%	0.0	99.98%	19252707	LR215053.1
<input checked="" type="checkbox"/>	Arabidopsis thaliana genome assembly_chromosome_2	Arabidopsis thaliana	8685	8685	100%	0.0	99.98%	21328600	QX291572.1
<input checked="" type="checkbox"/>	Arabidopsis thaliana ecotype 9412 chromosome 2 sequence	Arabidopsis thaliana	8685	8685	100%	0.0	99.98%	20497022	CP086735.1
<input checked="" type="checkbox"/>	Arabidopsis thaliana genome assembly_chromosome_2	Arabidopsis thaliana	8685	8685	100%	0.0	99.98%	19454276	LR797793.1
<input checked="" type="checkbox"/>	Arabidopsis thaliana genome assembly_chromosome_2	Arabidopsis thaliana	8674	8674	100%	0.0	99.94%	21027966	QX461156.1
<input checked="" type="checkbox"/>	Arabidopsis thaliana genome assembly_chromosome_2	Arabidopsis thaliana	8608	8608	100%	0.0	99.72%	23046220	QX461161.1

Description: Arabidopsis thaliana genome assembly, chromosome: 2

Scientific Name(Organism): Arabidopsis thaliana

On carefully observing the results of the blast search, the sequence1 has **molecule type** dna with a **query length** of 4706

On performing Blast Search for **Sequence 2** we get:

Job Title Seq2

RID 4JTF4PC6013 Search expires on 04-28 01:56 am Download All

Program BLASTN Citation

Database nt See details

Query ID IclQuery_64749

Description None

Molecule type dna

Query Length 261

Other reports Distance tree of results MSA viewer ?

Filter Results

Organism only top 20 will appear exclude
Type common name, binomial, taxid or group name
+ Add organism

Percent Identity E value Query Coverage
[] to [] [] to [] [] to []

Filter Reset

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments Download Select columns Show 100 ?

✓ select all 100 sequences selected GenBank Graphics Distance tree of results MSA Viewer

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
✓	Tomato spotted wilt virus intergenic region, isolate TSWV-IT-CE, genomic RNA	Tomato spotted ...	483	483	100%	9e-132	100.00%	261	AJ309924.1
✓	Tomato spotted wilt orthopspovirus isolate T1012 segment M, complete sequence	Tomato spotted ...	470	470	100%	7e-128	99.23%	4762	QN840011.1
✓	Tomato spotted wilt orthopspovirus isolate 104DOT22S segment M, complete sequence	Tomato spotted ...	464	464	100%	3e-126	98.85%	4762	Q0534350.1
✓	Tomato spotted wilt orthopspovirus isolate 102SEC22S segment M, complete sequence	Tomato spotted ...	464	464	100%	3e-126	98.85%	4762	Q0534347.1
✓	Tomato spotted wilt orthopspovirus isolate 71SET22S segment M, complete sequence	Tomato spotted ...	464	464	100%	3e-126	98.85%	4762	Q0534344.1
✓	Tomato spotted wilt orthopspovirus isolate 106DOT22S segment M, complete sequence	Tomato spotted ...	464	464	100%	3e-126	98.85%	4762	Q0534341.1
✓	Tomato spotted wilt orthopspovirus isolate 108DOT22S segment M, complete sequence	Tomato spotted ...	464	464	100%	3e-126	98.85%	4762	Q0507128.1
✓	Tomato spotted wilt orthopspovirus isolate 107DOT22S segment M, complete sequence	Tomato spotted ...	464	464	100%	3e-126	98.85%	4762	Q0507125.1
✓	Tomato spotted wilt orthopspovirus isolate 105DOT22S segment M, complete sequence	Tomato spotted ...	464	464	100%	3e-126	98.85%	4762	Q0507122.1
✓	Tomato spotted wilt orthopspovirus isolate 86DOW22S segment M, complete sequence	Tomato spotted ...	464	464	100%	3e-126	98.85%	4762	Q0507119.1
✓	Tomato spotted wilt orthopspovirus isolate 85DOPt22S segment M, complete sequence	Tomato spotted ...	464	464	100%	3e-126	98.85%	4762	Q0507116.1
✓	Tomato spotted wilt virus isolate intergenic region, isolate TSWV-IT-TQ	Tomato spotted ...	464	464	100%	3e-126	98.85%	260	AJ428574.1
✓	Tomato spotted wilt virus RNA segment M, complete sequence	Tomato spotted ...	448	448	100%	3e-121	97.36%	4768	AB190818.1
✓	Tomato spotted wilt orthopspovirus isolate P349 nonstructural protein (NSm) gene, complete cds	Tomato spotted ...	446	892	100%	1e-120	97.70%	1564	QMB867574.1

Description: Tomato spotted wilt virus intergenic region, isolate TSWV-IT-CE, genomic RNA

Scientific Name(Organism): Tomato spotted wilt orthopspovirus

On carefully observing the results of the blast search, the sequence2 has **molecule type** dna with a **query length** of 261

Clearly, Tomato spotted wilt virus is an intergenic sequence and not a gene itself.

Hence, sequence 1 is a protein coding genic sequence and sequence 2 is a non-genic sequence.