# Exploring the Mutations that Comprise the Genetically Linked Haplotype of the SARS-CoV-2 Virus





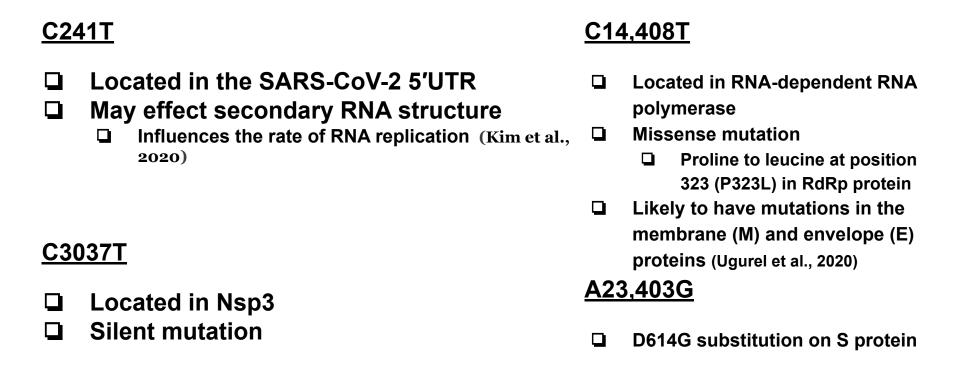
### **Outline**

- 1. The D614G Mutation is Almost Always Observed with Three Other Mutations
- 2. The Effects and Locations of these Four Mutations
- The Sequences were Collected and a Sequence Alignment and Phylogenetic Tree were Generated
- 4. The Sequence Alignment Only Yielded One Sequence with Irregular Mutations
- The phylogenetic tree did not show a relationship between the location collected and date collected
- 6. Future Directions

# The D614G Mutation is Almost Always Observed with Three Other Mutations

- In March of 2020, the D614G mutation was observed globally and quickly became the dominant form of the SARS-CoV-2 virus
- The D614G mutation increases the flexibility of the SARS-CoV-2 spike protein and has the ability to influence the dynamics of the fusion peptide
  - D614G is just a part of the story
- Four mutations comprise a "genetically linked haplotype" (Korber et al., 2020)
  - C241T (UTR)
  - C3,037T (silent mutation)
  - C14,408T (RdRp)
  - A23,403G (D614G)
- Global sequences: CCCA -> TTTG
- In some cases TTCG and CCCG sequences have been observed
- Did each mutation of the genetically linked haplotype arise independently, and if so, in what order?

#### The Effects and Locations of these Four Mutations



## Regions Across the World Experienced Switch to G614 Dominance at Different Times

	Onset	Before-	Delay	After	Last	Delta Fisher
Location	Date	G/(G+D) G+D	Date	G/(G+D) G+D	Sample	G/(G+D) p-val
Africa	Mar 13	0.870 23	Mar 27	0.974 2780	Nov 18	0.10 0.021409
Asia	Jan 28	0.009 344	Feb 11	0.658 9292	Nov 15	0.65 0.000000
Europe	Jan 29	0.235 17	Feb 12	0.942 127747	Nov 24	0.71 0.000000
North-America	Feb 28	0.051 99	Feb 20	0.911 37866	Nov 16	0.86 0.000000
Oceania	Mar 4	0.059 51	Mar 18	0.929 10408	Dec 2	0.87 0.000000
South-America	Mar 4	0.611 18	Mar 18	0.971 1364	Nov 26	0.36 0.000001

- Frequency switchover date
  - o Asia- Feb 11
  - North America- Feb 20
- This informed collection date periods to be considered when compiling sequences

# The Sequences were Collected and a Sequence Alignment and Phylogenetic Tree were Generated.

- 48 sequences were collected from the NCBI sequence and NCBI virus Sequence database.
  - 12 Sequences were collected from China; 12 before 2/11 and 12 after 2/11
  - 12 sequences were collected from the US; 12 before 2/20 and 12 after 2/20
- The sequences were aligned using EMBL clustal omega sequence aligner tool and a phylogenetic tree was generated using the same tool.
- The mutation locations were observed in all the sequences observing the:
  - C to T mutation at the 3037 nucleotide location
  - C to T mutation at the 14,408 nucleotide location
  - A to G at the 23, 403 nucleotide location
  - C to T at the 241 nucleotide location

#### **Sequence Alignment Showing the C241T and C3037T Mutation**

	_				_		
MT079843.1		TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	300	MT079843.1	TGATGAGTCTGGTGAGTTTAAATTGGCTTCACATATGTATTGTTCTT C PA		3060
MT412307.1		TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	282	MT412307.1	TGATGAGTCTGGTGAGTTTAAATTGGCTTCACATATGTATTGTTCTT T tA	CCCTCCAGA	3042
MT510727.1	ACATCTAGGTT1C	TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	288	MT510727.1	TGATGAGTCTGGTGAGTTTAAATTGGCTTCACATATGTATTGTTCTT CTA	CCCTCCAGA	3048
MT281577.1	ACATCTAGGTT1 C	TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	288	MT281577.1	TGATGAGTCTGGTGAGTTTAAATTGGCTTCACATATGTATTGTTCTT C TA	CCCTCCAGA	3048
MW301121.1	ACATCTAGGTT: I	TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	282	MW301121.1	TGATGAGTCTGGTGAGTTTAAATTGGCTTCACATATGTATTGTTCTT T tA	CCCTCCAGA	3042
MT510728.1	ACATCTAGGTT: C	TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	288	MT510728.1	TGATGAGTCTGGTGAGTTTAAATTGGCTTCACATATGTATTGTTCTT CTA	CCCTCCAGA	3048
MT412134.1	ACATCTAGGTT1 C	TCCGGGTGTGACCGAAAGGTAAGATGGGGAGCCTTGTCCCTGGTTT	264	MT412134.1	TGATGAGTCTGGTGAGTTTAAATTGGCTTCACATATGTATTGTTCTT C TA	CCCTCCAGA	3024
MW011766.1	ACATCTAGGTT1 C	TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	288	MW011766.1	TGATGAGTCTGGTGAGTTTAAATTGGCTTCACATATGTATTGTTCTT C TA	CCCTCCAGA	3048
MW011765.1	ACATCTAGGTT1 C	TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	288	MW011765.1	TGATGAGTCTGGTGAGTTTAAATTGGCTTCACATATGTATTGTTCTT CTA	CCCTCCAGA	3048
MW011764.1	ACATCTAGGTT1	TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	288	MW011764.1	TGATGAGTCTGGTGAGTTTAAATTGGCTTCACATATGTATTGTTCTT CTA	CCCTCCAGA	3048
MT252727.1	ACATCTAGGTT1 C	TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	281	MT252727.1	TGATGAGTCTGGTGAGTTTAAATTGGCTTCACATATGTATTGTTCTT C TA	CCCTCCAGA	3041
MT622319.1	ACATCTAGGTT1	TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	286	MT622319.1	TGATGAGTCTGGTGAGTTTAAATTGGCTTCACATATGTATTGTTCTT CTA	CCCTCCAGA	3046
MT412338.1	ACATCTAGGTT:	TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	264	MT412338.1	TGATGAGTCTGGTGAGTTTAAATTGGCTTCACATATGTATTGTTCTT CTA	CCCTCCAGA	3024
MT412339.1	ACATCTAGGTT: C	TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	264	MT412339.1	TGATGAGTCTGGTGAGTTTAAATTGGCTTCACATATGTATTGTTCTT CTA	CCCTCCAGA	3024
MT407653.1	ACATCTAGGTT1 C	TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	240	MT407653.1	TGATGAGTCTGGTGAGTTTAAATTGGCTTCACATATGTATTGTTCTT C ta	CCCTCCAGA	3000
MT412340.1	ACATCTAGGTT: 1	TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	289	MT412340.1	TGATGAGTCTGGTGAGTTTAAATTGGCTTCACATATGTATTGTTCTT T TA	CCCTCCAGA	3049
MT407649.1	ACATCTAGGTT: C	TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	240	MT407649.1	TGATGAGTCTGGTGAGTTTAAATTGGCTTCACATATGTATTGTTCTT C TA	CCCTCCAGA	3000
MT407652.1	ACATCTAGGTT:	TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	240	MT407652.1	TGATGAGTCTGGTGAGTTTAAATTGGCTTCACATATGTATTGTTCTT C ta	CCCTCCAGA	3000
MT568636.1	ACATCTAGGTT: C	TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	276	MT568636.1	TGATGAGTCTGGTGAGTTTAAATTGGCTTCACATATGTATTGTTCTT C TAG	CCCTCCAGA	3036
MT846460.1(USA:	ACATCTAGGTT: 1	TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	287	MT846460.1(USA:		CCCTCCAGA	3047
MT568634.1	ACATCTAGGTT1	TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	276	MT568634.1		CCCTCCAGA	3036
MT598633.1	ACATCTAGGTT:	TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	275	MT598633.1	TGATGAGTCTGGTGAGTTTAAATTGGCTTCACATATGTATTGTTCTT		3035
MT407651.1	ACATCTAGGTT1	TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	240	MT407651.1	TGATGAGTCTGGTGAGTTTAAATTGGCTTCACATATGTATTGTTCTT		3000
MT407655.1	ACATCTAGGTT1	TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	240	MT407655.1	TGATGAGTCTGGTGAGTTTAAATTGGCTTCACATATGTATTGTTCTT CTA		3000
MT407650.1	ACATCTAGGTT:	TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	240	MT407650.1	TGATGAGTCTGGTGAGTTTAAATTGGCTTCACATATGTATTGTTCTT		3000
MT252714.1	ACATCTAGGTT:	TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	239	MT252714.1	TGATGAGTCTGGTGAGTTTAAATTGGCTTCACATATGTATTGTTCTT		2999
MT252724.1	ACATCTAGGTT:	TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	248	MT252724.1	TGATGAGTCTGGTGAGTTTAAATTGGCTTCACATATGTATTGTTCTT CTA		3008
MT791896.1	ACATCTAGGTT: I	TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	286	MT791896.1	TGATGAGTCTGGTGAGTTTAAATTGGCTTCACATATGTATTGTTCTT		3046
MT534630.1	ACATCTAGGTT:	TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	288	MT534630.1	TGATGAGTCTGGTGAGTTTAAATTGGCTTCACATATGTATTGTTCTTTCT		3048
MT252720.1	ACATCTAGGTT:	TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	214	MT252720.1	TGATGAGTCTGGTGAGTTTAAATTGGCTTCACATATGTATTGTTCTTTCT		2974
MT834735.1(USA:	ACATCTAGGTT' I	TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	279	MT834735.1(USA:		CCCTCCAGA	3039
MT890279.1	ACATCTAGGTT1	TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	239	MT890279.1		CCCTCCAGA	2999
MT506885.1	ACATCTAGGTT' I	TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	234	MT506885.1		CCCTCCAGA	2994
MW064579.1	ACATCTAGGTT' I	TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	234	MW064579.1		CCCTCCAGA	2994
MT252722.1	ACATCTAGGTTT T	TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	204	MT252722.1		CCCTCCAGA	2964
MW161259.1	ACATCTAGGTT1	TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	238	MW161259.1		CCCTCCAGA	2998
MT952134.1	ACATCTAGGTTTC	TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	238	MT952134.1	TGATGAGTCTGGTGAGTTTAAATTGGCTTCACATATGTATTGTTCTT CTA		2998
MT646079.1		TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	234	MT646079.1	TGATGAGTCTGGTGAGTTTAAATTGGCTTCACATATGTATTGTTCTT TTA		2994
MT706175.1		TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	234	MT706175.1	TGATGAGTCTGGTGAGTTTAAATTGGCTTCACATATGTATTGTTCTT TA		2994
MW190442.1	ACATCTAGGTT' T	TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	234	MW190442.1	TGATGAGTCTGGTGAGTTTAAATTGGCTTCACATATGTATTGTTCTT' TTA		2994
MW190547.1	ACATCTAGGTTT	TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	234	MW190547.1	TGATGAGTCTGGTGAGTTTAAATTGGCTTCACATATGTATTGTTCTT TTA		2994
MW064856.1		TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	164	MW064856.1		CCCTCCAGA	2924
MT370913.1		TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	233	MT370913.1	TGATGAGTCTGGTGAGTTTAAATTGGCTTCACATATGTATTGTTCTT' TTA		2993
MT371033.1		TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	233	MT371033.1		CCCTCCAGA	2993
MT834704.1		TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	239	MT834704.1		NNNNNNNN	2999
MT252718.1		ATGGAGAGCCTTGTCCCTGGTTT	23	MT252718.1		CCCTCCAGA	2783
MT811658.1		ATGGAGAGCCTTGTCCCTGGTTT	23	MT811658.1		CCCTCCAGA	2783
MT856370.1			0	MT856370.1	TGATGAGTCTGGTGAGTTTAAATTGGCTTCACATATGTATTGTTCTT CTA		2268
MT396241.1	ACATCTAGGTT	TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTT	288	MT396241.1	TGATGAGTCTGGTGAGTTTAAATTGGCTTCACATATGTATTGTTCTT		3045
			200				

### Sequence Alignment Showing the C14,408T and C23,403T Mutation

		_					
MT079843.	.1 TTCTCTACAGTGTTCCCAC	Cracaagttttggaccactagtgagaaaaatatttgttgat	14460	MT079843.1	TGTTCTTTATCAGOA	GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23460
MT412307.	.1 TTCTCTACAGTGTTCCCAC	Tracaagttttggaccactagtgagaaaaatatttgttgat	14442	MT412307.1		GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23442
MT510727.		CTACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14448	MT510727.1		GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23448
MT281577.		CTACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14448	MT281577.1		GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23448
MW301121.		T'ACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14442	MW301121.1		GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23442
MT510728.		C ACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14448	MT510728.1		GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23448
MT412134.		C ACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14424	MT412134.1		GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23424
MW011766.		C ACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14448	MW011766.1		GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23448
MW011765.		C ACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14448	MW011765.1		GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23448
MW011764.		C ACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14448	MW011764.1		GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23448
MT252727.		C ACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14441	MT252727.1		GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23441
MT622319.		CACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14446	MT622319.1		GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23446
MT412338.		C ACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14424	MT412338.1		GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23424
MT412339.		CACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14424	MT412339.1		GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23424
MT407653.		CACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14400	MT407653.1		GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23400
MT412340.		T ACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14449	MT412340.1		GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23449
MT407649.		CACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14400	MT407649.1		TGTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23400
MT407652.		CACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14400	MT407652.1		GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23400
MT568636.		C ACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14436	MT568636.1		GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23436
MT846460.		T ACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14447	MT846460.1(USA:		GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23447
MT568634.		C ACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14436	MT568634.1		TGTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23436
MT598633.		C ACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14435	MT598633.1		GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23435
MT407651.		C ACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14400	MT407651.1		GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23400
MT407651.			14400	MT407655.1		GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23400
MT407650.		C PACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT C PACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14400	MT407650.1		GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23400
MT252714.			14399	MT252714.1		GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23399
MT252714.		CACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14408	MT252724.1		GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23408
MT791896.		C ACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14446	MT791896.1		GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23446
MT534630.		T ACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT C ACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14448	MT534630.1	TGTTCTTTATCAGO A	GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23448
MT252720.		CTACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14374	MT252720.1		GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23374
MT834735.		T ACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14439	MT834735.1(USA:	TGTTCTTTATCAGGG	GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23439
MT890279.		C ACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14399	MT890279.1		GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23399
MT506885.		T ACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14394	MT506885.1		GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23394
MW064579.		T ACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14394	MW064579.1		GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23394
MT252722.		T ACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14364	MT252722.1	TGTTCTTTATCAG( G	GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23364
MW161259.		C ACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14398	MW161259.1	TGTTCTTTATCAGO A	GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23398
MT952134.		C ACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14398	MT952134.1	TGTTCTTTATCAG(A	GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23398
MT646079.			14394	MT646079.1		GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23394
MT706175.		T ACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT T ACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14394	MT706175.1	TGTTCTTTATCAG(G	GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23394
				MW190442.1	TGTTCTTTATCAGGG	GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23394
MW190442. MW190547.		T ACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14394	MW190547.1	TGTTCTTTATCAGC G	FGTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23394
MW064856.		T ACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14394 14324	MW064856.1	TGTTCTTTATCAG(G	TGTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23324
		T ACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT		MT370913.1	TGTTCTTTATCAGC G	FGTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23393
MT370913.		T ACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14393	MT371033.1	TGTTCTTTATCAGG G	GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23393
MT371033.		T ACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14393	MT834704.1	TGTTCTTTATCAG( G	GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23399
MT834704.		T ACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14399	MT252718.1	TGTTCTTTATCAG( G	GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23183
MT252718.		T ACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14183	MT811658.1		GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23183
MT811658.		T ACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14183	MT856370.1	TGTTCTTTATCAG(A	GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	22668
MT856370.		CACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	13668	MT396241.1	TGTTCTTTATCAG(A	GTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACT	23445
MT396241.		CTACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGAT	14445		********	**********	
	*********	**********			_		

# The Sequence Alignment Only Yielded One Sequence With Irregular Mutations

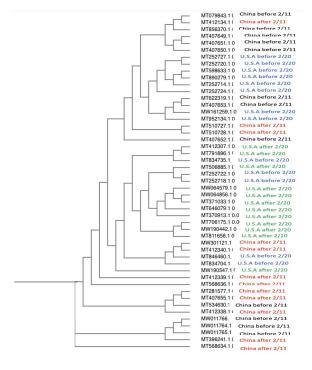
			EN .		
MT079843.1	CCCA	MT412307.1	TTTG	MT856370.1	NCTA
MT510727.1	CCCA	MT281577.1	TTTG		THE STATE IS NOT
MW301121.1	CCCA	MT412340.1	TTTG	6	
MT510728.1	CCCA	MT846460.1	TTTG	6	
MT412134.1	CCCA	MT252714.1	TTTG	8	
MW011766.1	CCCA	MT534630.1	TTTG		
MW011765.1	CCCA	MT834735.1	TTTG	E.	
MW011764.1	CCCA	MT890279.1	TTTG	E.	
MT252727.1	CCCA	MT506885.1	TTTG	8	
MT412338.1	CCCA	MW161259.1	TTTG	5	
MT412339.1	CCCA	MT646079.1	TTTG		
MT407653.1	CCCA	MT706175.1	TTTG		
MT407649.1	CCCA	MW190442.1	TTTG		
MT407652.1	CCCA	MW190547.1	TTTG		
MT568636.1	CCCA	MW064856.1	TTTG		
MT568634.1	CCCA	MT370913.1	TTTG	8	
MT598633.1	CCCA	MT371033.1	TTTG	6 e	
MT407651.1	CCCA	MT834704.1	TNTG	e e	
MT407655.1	CCCA	MT252718.1	NTTG	© 2	
MT252724.1	CCCA	MT811658.1	NTTG		
MT252724.1	CCCA			70	
MT791896.1	CCCA				
MT252720.1	CCCA				
MW064570 1	CCCA	l			

- ~55% of the sequences were CCCA
- ~44% of the sequences were TTTG, TNTG, or NTTG
- ~1% of the sequences were NCTA
- The sequence alignment results provide support for the "genetically linked haplotype" (Korber et al., 2020)
  - All four mutations arose at once
- May need more sequences to view more irregular mutations

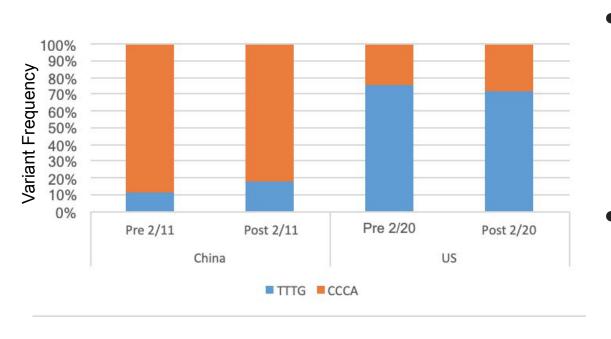
# Phylogenetic Tree does not Show that Sequences collected Around the Same Time in the Same location have the Nearest Common Ancestor.

- Some sequences collected from the same area and time are grouped together.

No clear trends are observable.



### CCCA Variant is Dominant Strain in China While TTTG is Dominant in USA On Both Ends of Conversion Date



- This data is confounding when considering the frequency of G614 increases to >90% after the conversion dates
- Spreading the range of collection dates could produce more accuracy of frequencies

# 55 percent Sequences Collected demonstrated the D614 variant.

- Of the 48 Sequences collect, 26 of the sequences showed the D614 Mutation.
- The sequences were randomly collected and the variants were not searched for.
- The Phylogenetic Tree did not show an overall trend with sequences collected from the same area and time not sharing the most common ancestor.
- After the conversion date the G614 variant was the dominant strand.

### **Future Directions**

- Gather more sequences and find a software that allows for the sequence alignment of a large number of nucleotides
  - Sequences spanning a large geography and time range
  - Limited by database and alignment software
- Utilize more automated databases and pipelines such as Korber et al.'s database



### **Acknowledgments**

 We wanted to thank Dr. Dahlquist for helping us find a resourceful article and aiding us with a couple figures.

We wanted to thank our T/A Annika Dinulos.

 We also wanted to thank the LMU library for giving us access to articles and databases and the LMU biology department.

### References

Korber, B., Fischer, W. M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., ... & Hastie, K. M. (2020). Tracking changes in SARS-CoV-2 Spike: evidence that D614G increases infectivity of the COVID-19 virus. Cell, 182(4), 812-827, retrieved from <a href="https://doi.org/10.1016/j.cell.2020.06.043">https://doi.org/10.1016/j.cell.2020.06.043</a>.

Ugurel, O. M., Ata, O., & Turgut-Balik, D. (2020). An updated analysis of variations in SARS-CoV-2 genome. Turkish journal of biology = Turk biyoloji dergisi, 44(3), 157–167. <a href="https://doi.org/10.3906/biy-2005-111">https://doi.org/10.3906/biy-2005-111</a>

Kim, D., Lee, J. Y., Yang, J. S., Kim, J. W., Kim, V. N., & Chang, H. (2020). The architecture of SARS-CoV-2 transcriptome. *Cell*.