Evolutionary Shift from Purifying Selection towards Divergent Selection of SARS-CoV2 Favors its Invasion into Multiple Human Organs

Running title: SARS-CoV2 adopts divergent selection

Amit K Maiti

Department of Genetics and Genomics, Mydnavar, 2645 Somerset Boulevard, Troy MI 48084, USA, Email: akmit123@yahoo.com, amit.maiti@mydnavar.com, Phone: +1 248 379 3129

Abstract

SARS-CoV2 virus is believed to be originated from a closely related bat Coronavirus RaTG13 lineage after gaining insertions of RBD of spike (S) protein by exchanged recombination with pangolin virus Pan_SL_COV_GD. SARS-CoV2 uses its entry-point key residues in S1 protein to attach with human ACE2 receptor. SARS-CoV2 evolution comprises any of these possibilities: it entered human from bat with its poorly developed entry-point residues much before its known appearance with slower mutation rate; or recently with efficiently developed entry-point residues having more infective power with higher mutation rate; or through an intermediate host. RaTG13 has 96.3% identity with SARS-CoV2 genome implying that it substituted ~1106 nucleotides to evolute as present-day virus. Temporal analysis of SARS-CoV2 genome from December 2019 shows that its nucleotide substitution rate is as low as 27nt/year with an evolutionary rate of $9x10^{-4}$/site/year, which is a little less than other retrovirus (10^{-4} to 10^{-6}/site/year). Estimation of TMRCA of SARS-CoV2 from bat RaTG13 lineage appears to be in between 9-14 years. Furthermore, evolution of a critical entry-point residue Y493Q needs two substitutions with an intermediate virus carrying Y493H (Y>H>Q), although such an intermediate virus has not been identified in known twenty-nine bat CoV virus. Genetic codon analysis indicates that SARS-CoV2 evolution from RaTG13 lineage strictly follows neutral evolution with strong purifying selection whereas its propagation in human disobeys neutral evolution as nonsynonymous mutations surpasses synonymous mutations with the increase of ω (d_n/d_s) signifying its proceedings towards divergent selection predictably for its infection power to evade multiple organs.

Keywords: SARS-CoV2; Covid-19; mutation; evolution; divergent selection

Introduction

Novel coronavirus SARS-CoV2 is believed to be originated in Wuhan, China in 2019 and has genomic identities to earlier SARS-COV virus with 79.8% and with MERS-COV virus with 59.1% (Ren, Wang et al. 2020, Zhou, Yang et al. 2020). Comparing highest identities of SARS-CoV2 genome with related CoV virus genome of Bat (ZC-45 (87.7%), RaTG13 (96.3%)), Pangolin (Pan-SL_CoV_GD (Guangdong, China) (91.2%), Pan_SL_CoV_GX (Guanxhi) (85.4%)), Li et al (Li, Giorgi et al. 2020) proposed that SARS-CoV2 arose from bat RaTG13 after gaining three insertions in the vicinity of RBM (Receptor Binding Motif) at RBD (Receptor Binding Domain) in S1 region by exchanged recombination with Pan_SL_COV_GD of pangolin. Due to higher dissimilarities with Pan-SL_CoV_GD in other genes except RBD of S1 protein, they suggested that pangolin could not be an intermediate host of SARS-CoV2 but RatG13 of bat (*Rhinolophus affinis*) is the most probable ancestors of SARS-CoV2 of human and, this views are supported by several phylogenetic studies (Forster, Forster et al. 2020, Gonzalez-Reiche, Hernandez et al. 2020, Latinne, Hu et al. 2020). Mainly based on the presence of furin cleavage site (RPRR motif) in SARS-CoV2 that is not observed in any other CoV virus, although present in some beta coronavirus (Hoffmann, Kleine-Weber et al. 2020), Andersen et al (2020) suggested that SARS-CoV2 was originated naturally from the homologue of SARS-CoV2 that was circulating long time in the bat along with RaTG13 and, had a common ancestor (Andersen, Rambaut et al. 2020).

S (spike) protein of SARS-CoV2 virus resides on their protein coat membrane and is cleaved into two small proteins S1 and S2 by the human host enzymes. The cleavage occurs at the two sites: one in between S1/S2 site by furin and other in S2 site by a serine protease, TMPRSS2 (Hoffmann, Kleine-Weber et al. 2020, Hoffmann, Kleine-Weber et al. 2020). S1 forms a claw like structure with seven key entry-point residues 449Y, 455L, 486F, 489Y, 493Q, 500T and, 501N at the RBM and, attaches with K31, E35, D38, M82 and K353 of the host ACE2 (Angiotensin Converting Enzyme 2) receptor (Wan, Shang et al. 2020, Wang, Zhang et al. 2020) whereas S2 mediates membrane fusion with the host cell. Among these residues K31-493Q and K353-501N interactions are most important for SARS-CoV2 infection to human host and, provide more chemically favorable interactions than SARS-COV K31-479L/N (homologue of SARS-CoV2

493Q) and K353-487S/T (homologue of 501N) binding, which gave SARS-CoV2 more infection power over SARS-COV (Wan, Shang et al. 2020).

Apart from uncertainty about the origin of SARS-CoV2, evidences are accumulating that RaTG13 of bat is zoonotically evoluted to SARS-CoV2 of human without the recombination of RBD of S1 protein region (Boni, Lemey et al. 2020) and, remains as a significant ancestral lineages of SARS-CoV2. Thus, an attempt to estimate evolutionary time frame, TMRCA (Time from Most Common Recent Ancestor) from its nearest ancestor RaTG13 could give an insight into SARS-CoV2 origin which could even be longer if both share a common unknown ancestor.

Estimating the TMRCA to evolute SARS-CoV2 from RaTG13 is intricate and, depends on mutation rate with other factors. Especially, the retrovirus evolution is complicated as it depends on the forces that drive the mutation rate per site nucleotide in the genome for its extra step of reverse transcription (Temin 1989). The optional mutation rate is context dependent at which rate the errors are made during replication of the viral genome. Apart from depending on the size of the genome, it also depends on the fidelity of RDRP (RNA Directed RNA Polymerase), proofreading activity and, selection pressure (Peck and Lauring 2018). Although, all retrovirus RDRP do not possess proofreading activities, but Coronavirus have strong proofreading activities with very different sequences. As for example, SARS-CoV2 and Ebola RDRP are completely different (no significant similarities, data not shown) although SARS-CoV2 RDRP bears considerable identities with SARS-COV (Ren, Wang et al. 2020, Zhou, Yang et al. 2020). Thus, a general consensus about a mutation rate in SARS-CoV2 cannot be reached although the mutation rate for positive strand retrovirus have been estimated as 10^{-4} to 10^{-6}/s(substitution)/n(nucleotide)/c (cell infection). Cell infection estimates the viral generation (Holmes 2009, Sanjuán, Nebot et al. 2010, Peck and Lauring 2018).

The estimation of evolutionary time, TMRCA, from genetic data mostly are fitted to molecular clock theory (Zuckerkandl and Pauling 1965) and later refined by neutral theory of evolution (Kimura 1979). For retrovirus evolution, the neutral theory based on strict molecular clock appears to be correct for influenza A virus (Gojobori, Moriyama et al. 1990). Although, HIV1 evolution initially rejected the neutral theory (Posada and Crandall 2001) but removal of datasets

carrying recombination event from the nucleotide data showed HIV evolution also followed strict molecular clock (Liu, Nickle et al. 2004). In oppose to molecular clock theory, rate of molecular evolution can vary over time and strict molecular clock can fail due to vast data (Lee, Rodrigue et al. 2015) and in such cases, relaxed molecular clock could be fitted. However, such relaxed molecular clock (Ho and Duchêne 2014) like rate smoothing method (Sanderson 2002), Bayesian methods (Thorne, Kishino et al. 1998, Huelsenbeck, Larget et al. 2000, Drummond and Suchard 2010), likelihood distribution (Felsenstein 1981, Paradis, Tedesco et al. 2013) or partitioned molecular clock (Thorne, Kishino et al. 1998) have been widely used especially for phylogenetic separation of species. Local clocks (Yoder and Yang 2000) or molecular dating (Lepage, Bryant et al. 2007) techniques are also choice of methods that are used for species separations.

Here, we estimated the evolutionary rate of SARS-CoV2 virus in human by analyzing the nucleotide base substitutions of one sixty-six SARS-CoV2 genome from December 2019 to September 2020 in temporal manner. Comparing the evolutionary rate by synonymous amino acid changes following neutral evolution, we estimated that it would take approximately 9-14 years to appear as SARS-CoV2 in human from RaTG13 lineage of bat. Detail analysis of nucleotide evolution of RBD of S protein does not support recombination of S1 RBD region from Pan_SL_CoV_GD. Evolution of a key entry-point residue Y493Q indicates that Y must be mutated twice to give rise to Q as Y>H>Q but such an intermediate CoV virus is never identified in known twenty-nine bat CoV virus. Comparing the changes in genetic codons of synonymous and nonsynonymous amino acids we also show that SARS-CoV2 evolution from RaTG13 lineage followed strict molecular clock with neutral evolution but after its appearance in human, SARS-CoV2 does not follow strict neutral evolution. As the infection time progresses, the increase of ω (d_n/d_s, the rate of the proportion of nonsynonymous mutations to the synonymous mutations) value indicates that SARS-CoV2 is proceeding towards divergent selection from purifying selection predictably with the invasion of multiple organs of the human body before reaching saturation or equilibrium.

Materials and methods

Genomic Sequences

SARS-CoV2 genomic sequences are obtained from covid-19 data portal (www.covid19dataportal.org; ENA browser, European nucleotide archive) of European institute and from NIH repository (www. https://www.ncbi.nlm.nih.gov/nuccore/) [**Supplementary Table 1].** Collection date and place of collection are recorded for each sequence and, these viral genomes are grouped by their collection date within 1st and 10th of each month to use for analysis so that sequences should represent gaps of at least approximately of one month. Also, in each month group, SARS-CoV2 genomes were collected from different places in the world to maintain diversification.

Blast and Alignments

Virus genome sequences are compared for identity differences using 2-nucleotide blasts (Needleman-Wunsch Global Align Nucleotide Sequences) at the NCBI website using the SARS-CoV2 reference genome (NC_045512, Wuhan-Hu-1). This genome has 100% identity with the genome that was collected on 12/01/2019 (MN908947). From the blast results the numbers of nucleotide identity differences are noted or counted removing the gaps and, other artifacts in alignments [**Supplementary Table 1**]. Average nucleotide differences are calculated for each month by using mean differences in nucleotides of all the genome collected in that month. Average nucleotide difference of a month group over the average nucleotide difference of previous month is considered as the base substitution rate in that month.

Synonymous and nonsynonymous assignment of the codon

Synonymous and nonsynonymous amino acid changes are identified using blast (https://blast.ncbi.nlm.nih.gov/) and protein translation tools (www.expasy.ch). For identifying amino acid changes in SARS-CoV2, multiple SARs_CoV2 genome sequences (nucleotide) from September (1st-10th) isolates were aligned with CLUSTALW (https://npsa-prabi.ibcp.fr/) with a reference genome of SARS-CoV2 (MN908947, 29903nt). Mismatch nucleotide (nt) sequences are traced back to the amino acid translation of the same reference SARS-CoV2 genome and codon changes in synonymous and nonsynonymous amino acids are noted and listed [**Supplementary table 2**]. According to the type

of nucleotide changes in a codon, α (transitions, UC, CU, GA, AG), ß (transversions, AU, UA, CG, GC) and Γ (transversions, UG, GU, CA, AC) are assigned. Each nucleotide mismatch at the various positions of a codon (1^{st}, 2^{nd}, 3^{rd}) are deduced for each synonymous (α, ß, Γ) and nonsynonymous (α1, ß1 and Γ1) amino acid changes respectively.

Similarly, for RATG13-SARS-CoV2 evolution, we used poly 1ab protein region of RaTG13 (acc no: MN996532.2; 266nt-21555nt) and SARS-CoV2 (NC_045512.2, 266nt-21555nt) for identification of synonymous and nonsynonymous amino acid changes. 689 homologous amino acids (aa) replacements are comparable in both genomes and substituted nucleotides in each codon of synonymous (583aa) or nonsynonymous (106aa) amino acid changes are assigned as earlier as α, ß, Γ and α1, ß1 and Γ1 respectively [**Supplementary Table 3**].

Estimation of evolutionary rate in SARS-CoV2 in human using neutral evolution

The evolutionary rate per site of SARS-CoV2 genome is estimated following three substitution model (3ST) (Kimura 1981). SARS-CoV2 genome of 29903nt mostly has coding sequences of 29244nt that codes for 9748aa. The probability of transitions as P (base transitions) is calculated as the total number of transitions in each position of the genetic code to the total number of homologous codes compared (9748aa). For example, in 1^{st} codon position, P = (α+α1)/9748. Simultaneously, Q (a transversion type of changes) at the 1^{st} codon position should be (ß+ß1)/9748 and R (other type of transversion) at the 1^{st} site of codon position equals to (Γ+Γ1)/9748. In this way, subsequently P, Q and R are calculated in other positions (2^{nd}, 3^{rd}) of each codon [**Supplementary table 2**].

When we put the P, Q and R value in the equation of 3ST base substitution model, we obtain the total rate of base substitution (K) per site.

$$K= -(1/4)\ln[(1- 2P – 2Q)(1-2P – 2R)(1 – 2Q – 2R)],$$

The K is calculated for each position of the codon as for 1^{st} codon position, K=0.0028; For 2^{nd} codon position, K=0.0027 and for 3^{rd} codon position, K=0.0070.

Using the formula of neutral theory of evolution,

K=2Tknuc (T, time period in years, knuc=evolutionary rate per site)

knuc=K/2T (T=9 months as 0.75 years for SARS-CoV2 from January, 2020 to September, 2020), thus the evolutionary rate at 1st codon position is $Knuc_1$=0.0027/(2*0.0027)=0.001787=1.78E-03/site/year.

Similarly, the $Knuc_2$ and knuc for 2nd and 3rd codon position should be 1.8E-03 (0.00185)/site/year and 4.7E-03 (0.00467)/site/year

Estimation of evolutionary time TMRCA from RaTG13 using neutral evolution

Evolutionary time of SARS-CoV2 from RaTG13 is estimated using neutral evolution (Kimura 1979) and (Nei and Gojobori 1986). For synonymous codons, α, ß, Γ, and for nonsynonymous codons α1, ß1 and Γ1 are calculated as earlier during estimation of evolutionary rate calculation for SARS-CoV2 in human. The total base substitution for each position at the codon are calculated using the same formula for neutral evolution K= -(1/4)ln[(1- 2P - 2Q)(1-2P - 2R)(1 - 2Q - 2R)] and the K corresponds to 0.0018, 0.0069 and 0.08 for 1st , 2nd and 3rd codon respectively [**Supplementary Table 3].**

The evolutionary time frames, TMRCA are calculated using the evolutionary rate in each codon position that had been earlier estimated for SARS-CoV2 in human (knuc for 1st codon position, 0.00178/site/year, for 2nd position, 0.00185/site/year and for 3rd position, 0.00467/site/year) by using same formula for neutral evolution, K=2Tknuc.

Hence, by using same formula (T=K/2knuc), we determined T, the time taken to evolute 1st codon corresponds to 3.3 years (=0.0018/2*0.0018), 1.9 years (0.0069/2*0.00178) and for 3rd codon position 9.57 years (0.08/2* 0.00467).

Estimation of w (d_n/d_s)

Synonymous and nonsynonymous amino acid changes are estimated according Nei and Gozobori (1986) (Nei and Gojobori 1986) with Zukes and cantor (1969)(Zukes and Cantor 1969), Yang and Nelson (2000)(Yang and Nielsen 2000)

and Li et al (1986)(Li, Wu et al. 1985). These estimations were essentially done using software PAML (phylogenetic Analysis by Maximum Likelihood) (https://abacus.gene.ucl.ac.uk/software/paml.html). Although all methods gave similar values for ω, we focused on Nei and Gojobori (1986) estimation method [**Supplementary table 5**]. Briefly, for RaTG13-SARS-CoV2, this estimation based on the respective sum of all 1-nt and 2nt (16 codons) changes of all synonymous (s_{dj}) and nonsynonymous (n_{dj}) codons as Sd ($\Sigma^{r}j=_{1}sdj$) and Nd ($\Sigma^{r}j=_{1}ndj$) are determined for r number of j-th codon. The proportion of synonymous (P_s) and nonsynonymous (P_n) codons are determined by S_d/S and N_d/N when S or N are the total expected alteration site in whole sequence length. The rate of synonymous (d_s) and nonsynonymous (d_n) changes are estimated by the formula d= -3/4 Log e (1-(4/3)p), when d is d_s or d_n and p is p_s or p_n. ω is calculated as d_n/d_s and designated as $\omega_{\text{-rtg13-sars}}$ for RaTG13 -SARS-CoV2 analysis.

For estimation of ω between SARS-CoV2, December and September isolate we created an artificial SARS-CoV2 sequence of September by incorporating 126 mutations that are identified in September isolates and compared with reference sequence of December 2019 in PAML

For ω analysis of RBD domain of S1 protein region of human SARS-CoV2, bat RaTG13 and pangolin Pan_SL_COV_GD, the amino acid sequences of (438-506aa) are compared in PAML.

Protein alignment

Alignments of ACE2 protein sequences from all animals and S1 protein region of RBD are done using CLUSTALW at https://npsa-prabi.ibcp.fr/.

Results

Estimation of base substitution rate in SARS-CoV2 in human

RaTG13 shows highest 96.3% nucleotide homology with SARS-CoV2 (Zhou, Yang et al. 2020) implying that approximately 1106nt (29903 x 3.7 (100-96.3)) has to be replaced over the years to evolute to SARS-CoV2 from RaTG13 lineage or even more if they shared a common ancestors. We estimated the evolutionary time frame, TMRCA from RaTG13, using genomic

sequences of SARS-CoV2 after its emergence from collection date of December 2019 to September 2020 and, temporally analyzed to get an estimation that how rapidly the virus was changing [**Supplementary table 1**]. Pairwise sequence analysis of one hundred and sixty-six SARS-CoV2 genomic sequences with reference genome, we calculated the average base substitution rate [**Fig. 1A**] of the virus. The average nucleotide changes occurred ~2.22 bp/month [**Fig1B**] and the typical average nucleotide substituted from December 2019 to September (1st-10th) for 9 months is 20.16 nt (~20 nucleotides). This observed base substitution with a selection constraint at this rate in human host, a simple extension of this calculation gives us 27 nucleotide (2.22 x 12) substitutions per year, which ultimately gives ~9.00^{-4} nt/site/yr (27nt/29903nt/0.75 years).

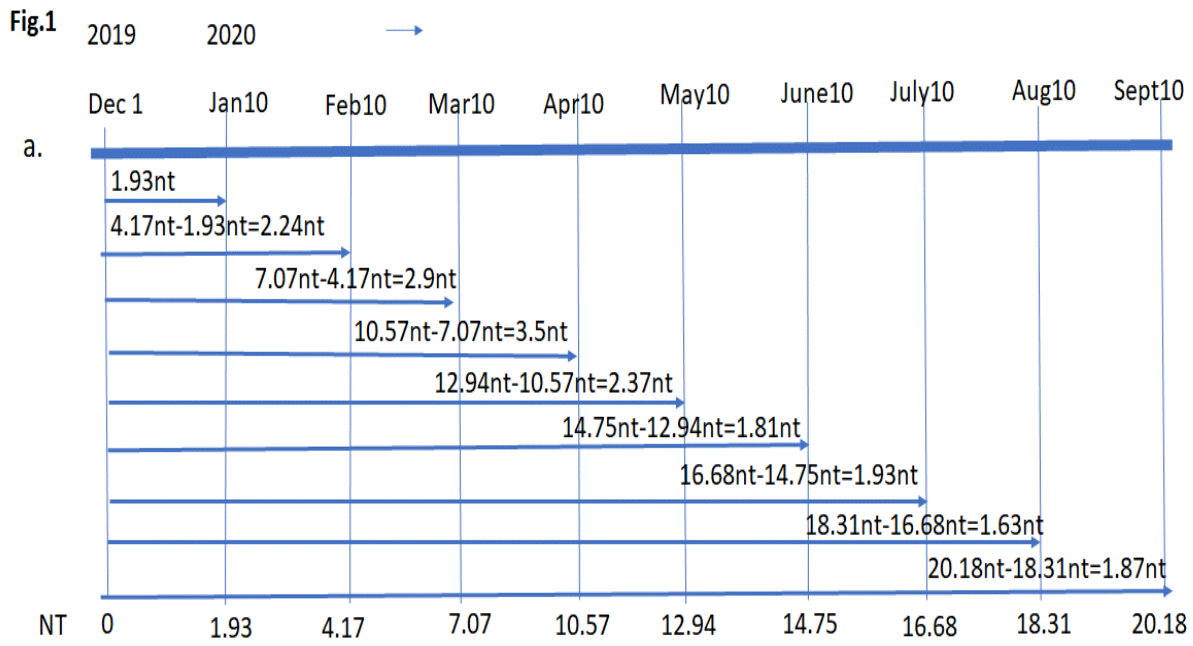
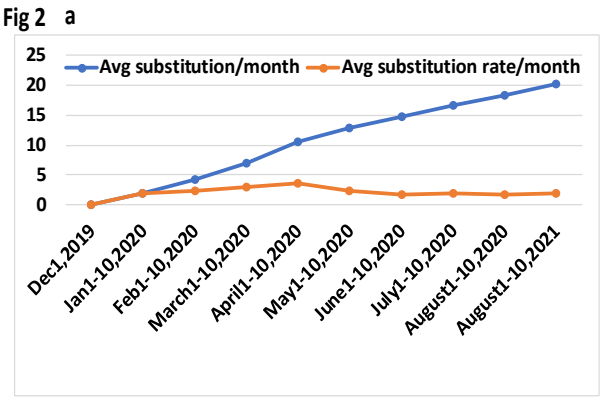
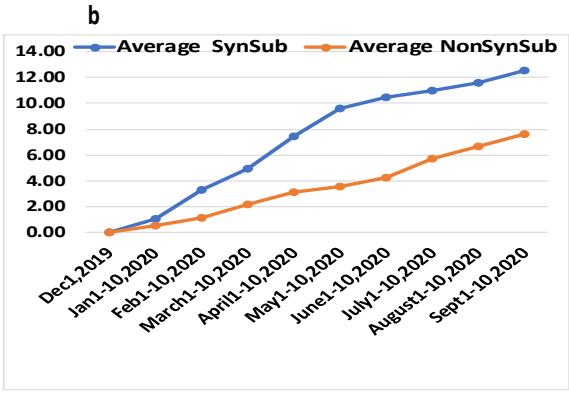
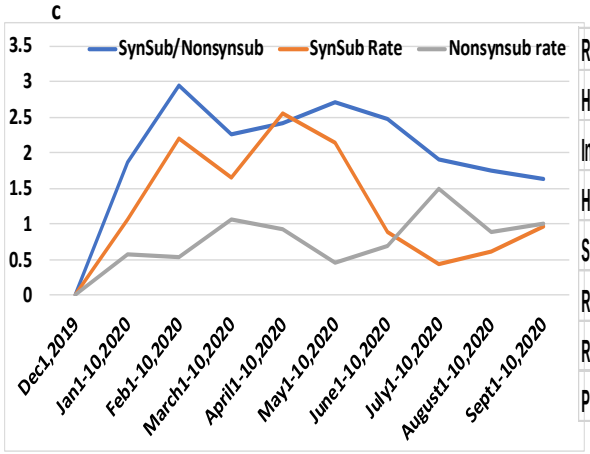


Figure 1. Average Nucleotide changes in SARS-CoV2 from January, 2020 to September, 2020. Each genomic sequences of SARS-CoV2 that is collected within 1st 10 days of each month were compared with 2-sequences bLAST with reference genomeand nucleotide differences were counted except artfacts and gaps. Rate of substitution for each months was considered by substracting the NT differences of previous months.

Although the total average nucleotide substitution increases temporally from January 2020, the average substitution rate is almost similar in all months [**Fig 2a]** suggesting that evolutionary rate of SARS-CoV2 in human are more or less constant in these months. Estimation of synonymous and nonsynonymous substitution rates indicate that synonymous amino acid changes are much greater (synonymous aa (0.000591689) / nonsynonymous aa (0.000340282) =1.735) [**Fig 2b, 2c, Supplementary table 1]** than nonsynonymous amino acid changes emphasizing SARS-CoV2 evolution is following the neutral evolution with strict molecular clock, which assumes that substitution rate will be equal to the mutation rate (Kimura 1979).







d

Retrovirus	Gene	Synonymous rat	nonsynonymous	Reference
HIV1	gag	13.08×10^{-3}	3.92×10^{-3}	Li et al 1986
Influenza A	Hemaglutinin	13.10×10^{-3}	3.59×10^{-3}	Nei and Gojobori 19
HBV	P	4.57×10^{-5}	1.45×10^{-5}	Orito et al
SARS-CoV2	whole	2.11×10^{-3}	2.0×10^{-3}	This study
RATG13-SARS-CoV2	Poly 1ab (266-21555nt)	2.8×10^{-2}	5×10^{-3}	This study
RATG13-SARS-CoV2	S1 RBD (449aa-518aa)	1.2×10^{-1}	6×10^{-2}	This study
Pan_SL_CoV_GD-SAR	S1 RBD (449aa-518aa)	4.4×10^{-1}	4.8×10^{-3}	This study

Figure 2. Average rate of nucleotide substitution in various months. a)Comparison of total and rate of substitution of NT. b)Average synonymous and nonsynonymous amino acid changes in NT level. Average synonymous substitution is always higher than nonsynonymous substitution in all months. c) Rate of synonymous substitution is also higher in all months (except April) than nonsynonymous substition. d)Rate of synonymous and nonsynonymous substition are compared with other RNA virus.(synSub, Synonymous substitution; NonSynSub; Nonsynonymous substitution

Based on strict molecular clock, the estimation of the TMRCA to evolute SARS-CoV2 from RaTG13 indicates (1106/27) 40.96 years to replace 1106nt of RaTG13. Duchene et. al (2014) (Duchêne, Holmes et al. 2014) suggested that the evolutionary time-scales in retroviruses require to consider substitution rates as a dynamic, rather than as a static,

especially when molecular data sampled over a short timeframe that are often appear to evolve at higher rates than those sampled over a longer time period. The inadequate 'correction' of multiple substitutions at single nucleotide sites is necessary to explain the evolutionary time-scale of a retrovirus, hence three orders of magnitude lower than that estimated using virus samples should be used (Worobey, Telfer et al. 2010, Ho and Duchêne 2014). Thus, after all corrections, it would take approximately 13.6 (40.96/3) years to evolve SARS-CoV2 from RaTG13 lineage or more from a common ancestor at the rate of evolution that are occurring in present day SARS-CoV2 virus in human.

Estimation of evolutionary rate of SARS-CoV2 in human using strict molecular clock with neutral evolution

Molecular clock based on neutral evolution suggests that synonymous mutations that are not in a selection constraint always must be greater than nonsynonymous mutations (Kimura 1979, Gojobori, Moriyama et al. 1990). As the synonymous amino acid changes are nonconstrained, the spontaneous nucleotide changes at the 3rd position of a codon that leads to mostly synonymous amino acid changes and, that would be greater than the evolutionary rate of other two positions of the codon (72%) (1st position gives 5% and 2nd position is zero as all 2nd position nt changes gives rise to nonsynonymous mutation) (Nei and Gojobori 1986). We estimated the evolutionary rate of SARS-CoV2 in human by measuring the synonymous and nonsynonymous amino acid changes from the same SARS-CoV2 genome in September isolates (1st-10th), as these changes are persisted in the population after 9 months of propagation from its appearance in December 2019. We used same sequences for identification of synonymous and nonsynonymous mutations to preserve their proportions in contrast to random mutation collections because deleterious nonsynonymous mutation also abolishes synonymous mutations in the same sequence. We identified 126 amino acid changes (63 synonymous and 63 nonsynonymous amino acid changes) in SARS-CoV2 genome (29903nt, 9748aa) and estimated the evolutionary rate, knuc for the 1st (0.00178/site/year), 2nd (0.0185/site/year) and 3rd (0.00467/site/year) codon positions by taking the value of T as the time taken to change these amino acid equals to 9 months (0.75 years) [**Supplementary table 2**]. Obviously the knuc for 3rd position of the codon is much higher than other two positions of the codon as it mostly leads to synonymous amino acid changes (Nei and Gojobori 1986).

Estimation of evolutionary time TMRCA of SARS-CoV2 from bat RaTG13

We calculated the TMRCA to evolute SARS-CoV2 from RaTG13 lineage that would provide insight into the origin of SARS-CoV2. For comparing mismatch nucleotide sequences between two virus and estimation of TMRCA using strict molecular clock following neutral evolution, we excluded S protein region from the whole genomic sequences of SARS-CoV2 and RaTG13 virus genome as this region is ambiguously believed to be inserted by recombination from Pan_SL_CoV_GD(Li, Giorgi et al. 2020). In addition, Liu et al (2004) (Liu, Nickle et al. 2004) showed that when recombination regions were removed from dataset, HIV1 evolution based on only mismatch nucleotide analysis exactly followed strict molecular clock. We considered SARS-CoV2 poly 1ab gene (266nt-21555nt, 7097aa) including RDRP gene for estimating synonymous and nonsynonymous changes of amino acids among 7096 aa (one is omitted due to a codon insertion). Total 689 aa are changed and comparable with 106aa and 583 aa are nonsynonymous and synonymous aa respectively. The synonymous aa changes are also much higher (583/106=5.5) than nonsynonymous aa changes and, could be fitted to strict molecular cloak (Kimura 1979). After estimation of the probability of transitions (P) and transversions (Q and R) of these amino acids and using the evolutionary rate at each position of the codon in SARS-CoV2, we determined the evolutionary time frame T (in years) following neutral evolution. The time for the nucleotide in 1^{st} position of the codon, 2^{nd} position of the codon and 3^{rd} position of the codon are estimated using the knuc value [**Fig.3a]** that are obtained for SARS-CoV2 evolution in human for each position and are 3.31 years, 1.9 years and 9.6 years respectively. As the 3^{rd} codon replacements mostly (72%) stands for synonymous amino acid changes (Nei and Gojobori 1986) and evolution under zero constraints, 13.3 years ((9.6/72) X 100) could be an reasonable TMRCA to arise SARS-CoV2 from RaTG13 or more than 9.6-13.3 years from an unknown ancestor of both of them. However, this time frame is in close approximation of 13.6 years that has been estimated by using strict molecular clock using base substitution model.

SARS-CoV2 is proceeding towards divergent selection from purifying selection

We estimated the proportion of synonymous and nonsynonymous differences, ω (d_n/d_s) in SARS-CoV2 in human propagation and from RaTG13 to SARS-CoV2 evolution. Although the ω is below (1< ω) in both the cases [**Fig 3b**] for RaTG13-SARS-CoV2 ($\omega_{\text{-rtg13-sars}}$ =0.04) and SARS-CoV2 in human ($\omega_{\text{-sars-hu}}$=0.30) signifying purifying or negative selection, but they have large differences in values. In fact, SARS-CoV2 in human has more than 6 fold ($\omega_{\text{-sars-hu/}}$ $\omega_{\text{-rtg13-sars}}$=0.30/0.04=7.5) higher ω value than RaTG13-SARS-CoV2 implying that selection constraints are tremendously driving the SARS-CoV2 evolution in

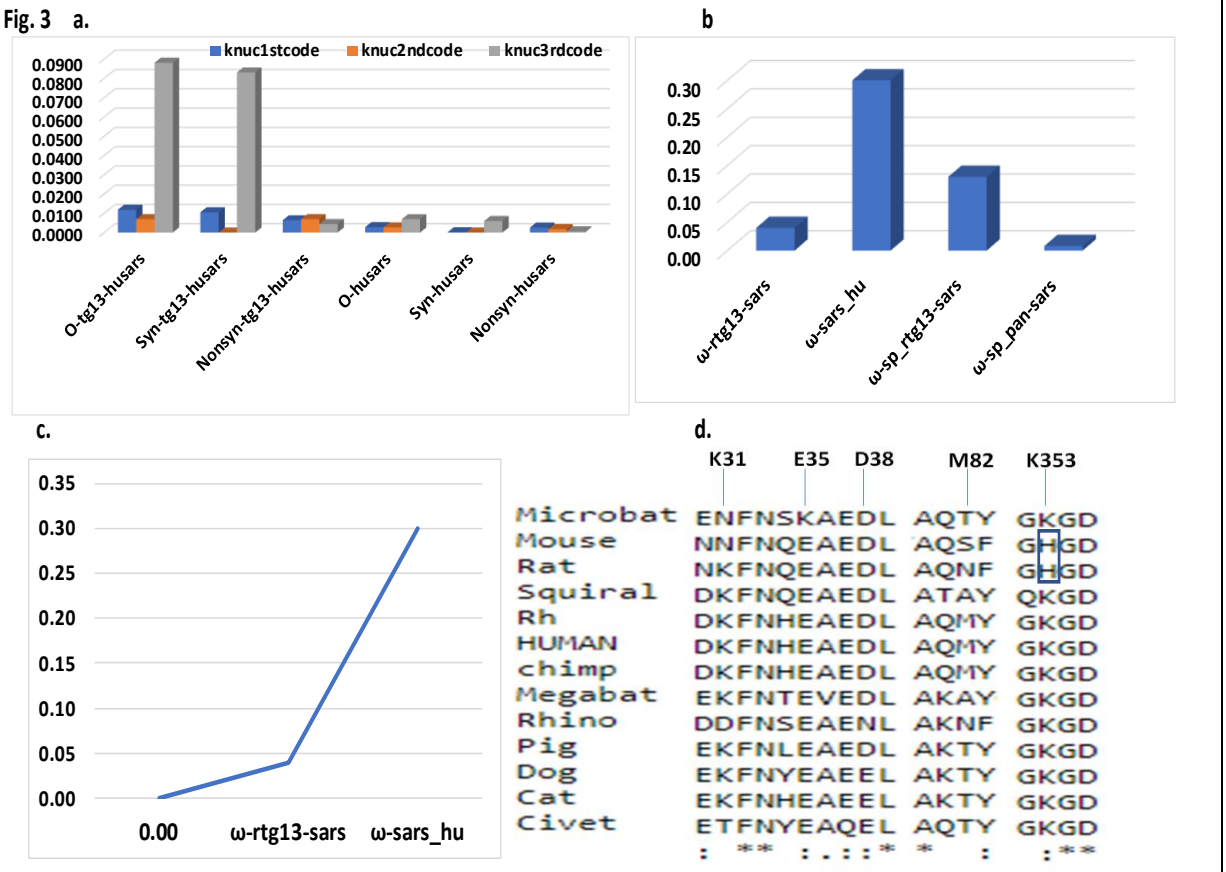


Figure 3. Evolutionary rate in codon position and ω (d_n/d_s) value in various conversion. a) Evolutionary rate (k_{nuc}) at each codon position (e.g 1st, 2nd and 3rd) in Overall, synonymous and nonsynonymous amino acid changes in SARS-CoV2 in human propagation from December, 2019-September, 2020 and from RATG13 of bat to the appearance in human. b**)** ω of various conversion are compared. It is notable that if RBD domain of S1 protein (sp) had come from pangolin virus into SARS-CoV2, Pan_SL_CoV_GD by recombination and then accumulated mutation, the ω value is very low (ω <0.0078) in compare to other conversion implying that RBD region might originated zoonotically from RATG13 of bat. c) Sudden increase of ω implying increase of nonsynonymous substitution over synonymous substitution in SARS-CoV2 during human propagation. d) Alignment of five key entry point residues in ACE2 protein of various animals. SARS-CoV2 shows poor infectivity due to absence of K353 residue in mouse and rat (shown in blue box).

human from purifying selection towards diversifying selection (approaching ω >=1) (Yang and Bielawski 2000). The purifying selection often could be a force for viral evolution as they proceed towards fixation by deleting the nonsynonymous and deleterious mutations (Kryazhimskiy and Plotkin 2008) (Lin, Bhattacharjee et al. 2019) and, it appears that SARS-CoV2 is proceeding such evolutionary .trajectories towards diversifying selection.

Nucleotide evolution does not support recombination of S1 protein RBD region from Pangolin CoV virus

RBD (438aa-506aa) of S1 protein contains key entry-point residues to attach human ACE2 receptor in SARS-CoV2 and, is believed to be originated from recombination from the Pan_SL_COV_GD (Li, Giorgi et al. 2020). The main reason of such a conclusion came from amino acid alignment [**Fig 4a, 4b]** of this region of SARS-CoV2 with Pan_SL_CoV_GD and RaTG13 where only 1 nonsynonymous amino acid differs with pangolin virus whereas 17 aa differs with RaTG13. Almost all other evidence of phylogenetic reports are based on protein sequences made this view stronger **[Fig. 4]** (Forster, Forster et al. 2020, Latinne, Hu et al. 2020, Li, Giorgi et al. 2020). However, the nucleotide alignment in this region shows that both RaTG13 and

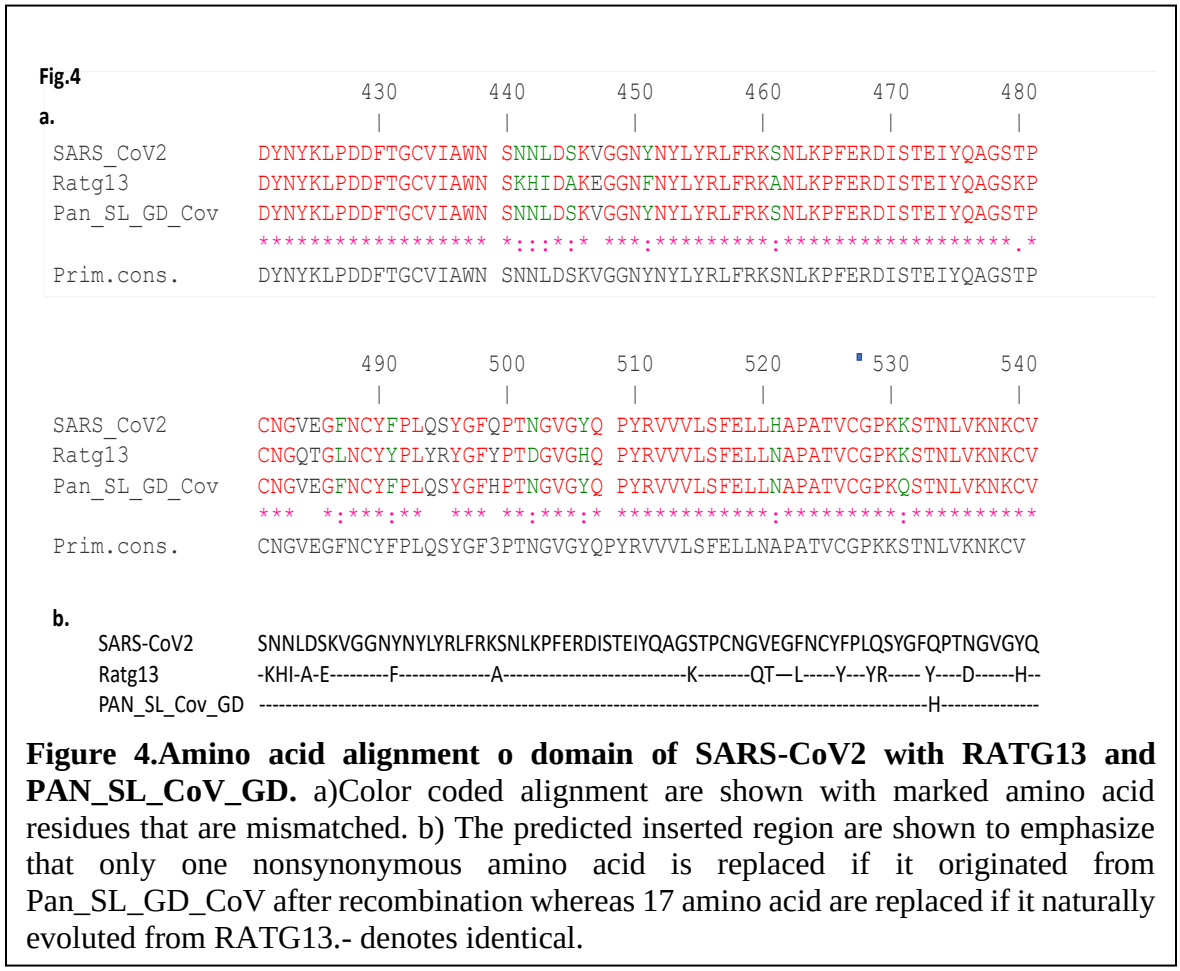


Figure 4.Amino acid alignment o domain of SARS-CoV2 with RATG13 and PAN_SL_CoV_GD. a)Color coded alignment are shown with marked amino acid residues that are mismatched. b) The predicted inserted region are shown to emphasize that only one nonsynonymous amino acid is replaced if it originated from Pan_SL_GD_CoV after recombination whereas 17 amino acid are replaced if it naturally evoluted from RATG13.- denotes identical.

Pan_SL_CoV_GD have 58 and 41 nucleotide changes respectively from SARS-CoV2 in 69 codons in this region **[Fig.5, supplementary table 4].** Among 58 codons of RaTG13-SARS-CoV2, 41 codons are synonymous and 17 codons (41aa+11aa=58) are nonsynonymous whereas among 41 codons of Pan_SL_CoV_GD-SARS-CoV2, 40 codons are synonymous and only 1 codon are nonsynonymous (40aa +1aa). At the nucleotide level, if 41nt replacement can occur after recombination of Pan_SL_CoV_GD RBD region into SARS-CoV2 leading to mostly synonymous aa changes, 58nt replacement with synonymous and nonsynonymous changes could also occur naturally in RaTG13 to SARS-CoV2.

However, the large differences observed in the number of nonsynonymous aa (1aa vs 17aa) in these two cases but not with synonymous amino acid changes (40aa

Fig.5

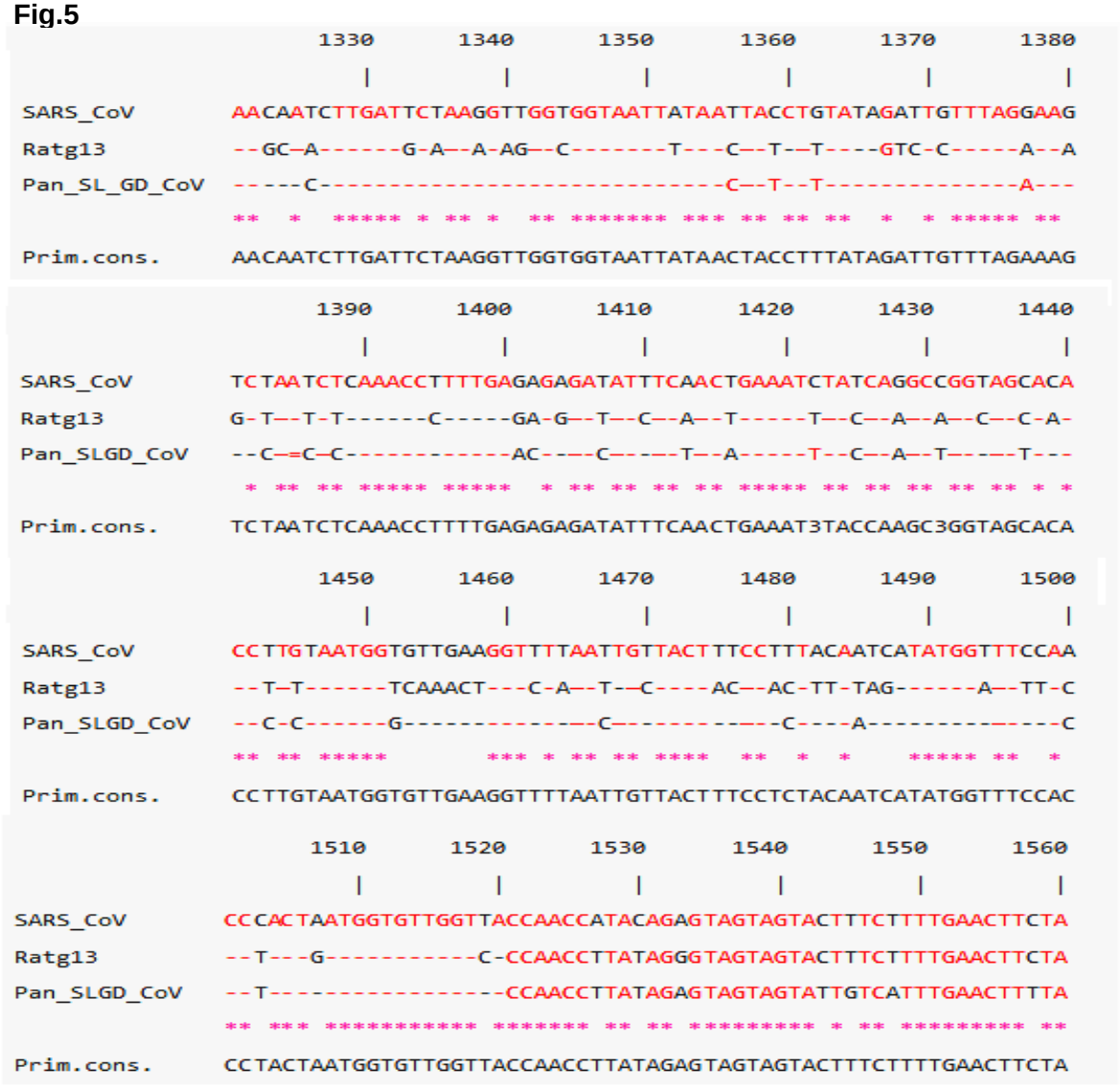


Figure 5. Nucleotide alignment of the same RBD region of SARS-CoV2, RATG13 and Pan_SL_CoV_GD. There are at least 28 NT changes although only one nonsynonymous amino acid replacement implying that amino acid identities between SARS-CoV2 and Pan_SL_CoV-GD are apparent.

vs 41aa). Although arguably it is possible to assume that S1 region could arise by recombination from Pan_SL_CoV_GD and then synonymous (neutral) mutations were accumulated in course of SARS-CoV2 evolution when neutral theory states that the rate of synonymous or silent substitutions is usually larger than that of nonsynonymous (i.e., amino acid-altering) substitutions (Gojobori, Moriyama et al. 1990). But if we observe the synonymous and nonsynonymous mutation rate in other retrovirus [**Table 2d],** only 2-to-3-fold differences generally occurred and, not with such a large difference as 1 vs. 17 nonsynonymous amino acids. It is again questionable after recombination why 40 synonymous aa changes could occur in this region with only one nonsynonymous aa change unless this region is refractory to nucleotide changes. But the present data of SARS-CoV2 mutation profile does not support that this region is refractory to nucleotide changes to generate nonsynonymous mutation as Long et al (2020) and Hoque et al.(2020) identified numerous nonsynonymous mutations (C480F, Y495S, L517F and G476S, V483A,Y508H respectively) in this region in SARS-CoV2 (Islam, Hoque et al. 2020, Long, Olsen et al. 2020). Nevertheless, when ω is calculated for this region of S1 protein (438aa-506aa), the Pan_SL_CoV_GD (after recombination followed by natural synonymous mutation) shows very low value (ω =0.008) than other conversions [**Fig3b**, **Supplementary Table 5, Supplementary table 6**] (Lin, Bhattacharjee et al. 2019). Such a low ω in compare to other evolution cast doubt on whether this region of S1 protein at all came by recombination and, indicate that it might come from RaTG13 itself zoonotically and accumulated both 41 neutral and 17 nonsynonymous mutation.

In addition, the concept of recombination in this region of S1 protein from Pan_SL_CoV_GD ambiguously determined with Simplot analysis (Li, Giorgi et al. 2020) meant for similar sequence identification but not recombination although Wu et al (2020)(Wu, Zhao et al. 2020), who first noted recombination of this region compared only several bat CoV virus with SARS-CoV2 but not with any pangolin CoV virus. Moreover, using extensive recombination breakpoint analysis at the S1 protein of SARS-CoV2, Pan_SL_CoV_GD and other bat coronavirus, Boni et al (2020) (Boni, Lemey et al. 2020) recently showed that S1 RBD region is unlikely came from Pan_SL_CoV_GD by recombination, instead, SARS-CoV2 naturally evoluted from RaTG13 or other related bat virus.

Unavailability of Intermediate Host between bat and Human

SARS-CoV2 virus uses key entry-point residues of RBD in S1 protein to bind with the ACE2 receptor of human through K31, E35, D38, M82 and, K353 (Wan, Shang et al. 2020). Among them, K31 and, K353 of ACE2 are the most important residues for effective SARS-CoV2 binding. Analysis of these residues in ACE2 receptors in various animals suggests that mouse and, rat possess poor ACE2 receptors (H353 in both animals instead of K353; also, mouse has N31 instead of K31) for SARS-CoV2 attachment [**Fig. 3c]** (Wan, Shang et al. 2020). By cloning and infectivity experiments, they also showed that in Civet cats, presence of T31 instead of K31 but with intact K353 in ACE2 receptor allows a moderate SARS-CoV2 infection but not in mouse or rat (absence of K353) and, indicated that K353 of ACE2 may be the most crucial residue in terms of SARS-CoV2 attachment. Other animals like Chimp, Rhesus monkey, monkey, cat, dog and pig have high identity with human ACE2 receptor protein sequence and possess both K31 and, K353 residues in their ACE2 receptor that could serve as an excellent attachment point for SARS-CoV2 RBM, thus, could efficiently serve as an intermediate host before infecting human. Although these animals are artificially infectible with SARS-CoV2 virus, none of these animals are found to be naturally harbored any SARS-CoV2 or its nearby genetically related COV virus. Thus, the conjecture remains to be elucidated whether such an intermediate host between human and bat would be existed or be explored in future in nature.

Evolution of SARS-CoV2 Entry-point Residues Interacting with ACE2 Receptor

K353-501N attachment site of human ACE2-SARS-CoV2 is the most efficient and crucial entry-point. In RaTG13 of bat from where SARS-CoV2 is believed to be originated, the homologue at 501N position is aa D (code GAU). An amino acid changes from D (code GAU) to N (code AAU) at this position in SARS-CoV2 enables them to infect human host. Thus, a single substitution in 1st codon from G>A nucleotide could give rise aaN from aaD at the 501 position in the RBD of SARS-CoV2 for K353-501N salt bridge formation for important attachment site and, almost gave RaTG13 a passport to infect human efficiently.

Similarly, 493Q residue in SARS-CoV2 for K31-493Q interaction, which is the second most important entry-point attachment is evoluted from amino acid Y, which is present in RaTG13 of bat. However, Y is coded by UAU and to become Q (code CAA) of SARS-CoV2, the codon needs to mutate at least twice i.e., mutation in two nucleotides in 1st and 3rd codon. The 1st codon must be U>C mutation and the second mutation at the 3rd codon could be U>A. If the 3rd codon mutation occurred earlier than 1st codon mutation in the bat virus, it would lead to nonsense (stop) code (UAA) and immaturely terminate S protein formation. Thus, 1st codon mutation (U>C) had to be created earlier than 3rd codon mutation for survival of this present-day virus. Eventually, 1st codon mutation (U>C) would create intermediate code CAU in ancestors of SARS-CoV2 virus that would code for H (Histidine) at this position. Thus, the conversion of Y > Q had to be in the course of pathway Y >H>Q. In that case, 493H carrying intermediate ancestor virus must be existed in any of the related bat virus strain. Until now whole genomic sequences from twenty-six types of bat CoV virus are known and analyzed (Forster, Forster et al. 2020, Latinne, Hu et al. 2020, Li, Giorgi et al. 2020) but no such ancestral viral strain was identified with a 493H in RBD. Although 493H carrying region might come from Pan_SL_CoV_GD by recombination as believed by Li et al(Li, Giorgi et al. 2020) but the recombination event is ambiguously projected as it is ruled out (Boni, Lemey et al. 2020).Thus, beside bat, there should be a SARS-CoV2 ancestor virus carrying 493H that remains to be identified unless the existence of such an ancestor virus still could be explored in bat.

For other remaining entry-point residues of 449Y, 455L, 486F, 489Y and 500T, three residues 455L, 489Y and 500T of SARS-CoV2 did not need any nucleotides substitutions. But 449Y of SARS-CoV2 needs a single nucleotide substitution from bat RATG13 (aa F>aa Y, UUU > UAU, 2nd codon, U>A). Similarly, for 486F (aaL > aaF, CUA >CUU, 3rd codon A>U) . Thus, in 449Y and 486F both cases, a single nucleotide substitution from bat also could give rise to these SARS-CoV2 entry-point residues.

Discussion

Effective vaccines or medicines development and, testing them in animals, it is necessary to know the origin of this virus and, how it is behaving in the human population by changing their genetic profile with evolutionary dynamics.

The creation of new mutations with evolutionary trajectories driving these pandemic could be replicated *in vitro* that could predict the reversion of vaccine effect and, its virulency as shown in other RNA virus (Stern, Yeh et al. 2017).

Most of the retrovirus followed neutral evolution with strict molecular clock, such as Influenza A virus (Gojobori, Moriyama et al. 1990), HIV1 virus (Liu, Nickle et al. 2004), Rous sarcoma virus and Herpes simplex virus (Sanjuán, Nebot et al. 2010). As we observed synonymous amino acid changes are much greater in both SARS-CoV2 in human than from RaTG13 lineage to SARS-CoV2, the evolution of these viruses is predictably following neutral evolution with strict molecular clock. Here we also did not consider any relaxed molecular clock as we did not compare here divergence of species from related bat coronavirus nor we focused on phylogenetic relationship with other related virus. Our analysis suggests that the mutation rate of SARS-CoV2 (~9.0 X 10^{-4}/site/year) is slightly lower than other retrovirus as it has been estimated as in the order of 10^{-3} to 10^{-4}/site/year (E.C. 2009, Duchêne, Holmes et al. 2014) or 10^{-4} to 10^{-6} /site/nucleotide/cell infection (Peck and Lauring 2018). In HIV1 Virus, the mutation rate for synonymous and nonsynonymous amino acid changes are (13.2 X10^{-3}/site/year) and, (6.9 X 10^{-3}/site/year) respectively (Li, Tanimura et al. 1988) **[Figure 2d].**

It is surprising that SARS-CoV2 appears to be undergoing higher number of generations as it created pandemic in world human population, yet the mutation rate is lower than other retroviruses. Thus SARS-CoV2 mutation rate could be better explained by the rate upon context dependent (Peck and Lauring 2018). Mutation frequency can be confounded by selection and genetic drift. In that case mutation rate could be lowered as the deleterious mutation drives the mutation rate lower as a principle criteria for purifying or negative selection (Peck and Lauring 2018). Between two models as speed vs adaptability of viral mutation rate, here it appears that SAR-COV2 evolution fits with adaptability model, which states that after a long adaptation to evade immune system, the selection pressure is relatively low and the supply of beneficial mutation frequency is reduced, thus population favors a low mutation rate. When the mutation reaches to an optimum level simply because selection is acting on it long time within the context of immune escape to reach the maximum mutation fitness (Orr 2000, Sanjuán 2010, Sanjuán, Nebot et al. 2010, Peck and Lauring 2018).

When purifying selection occurs in the population, the ω value is low (ω <1) but in case of higher value (ω =>1) significant divergent and aggressive evolution is warrant. The ω value from RaTG13 lineage-SARS-CoV2 evolution (ω =0.04) reached to much higher value (ω=0.30) after its appearance and, propagation in human suggesting that the virus is proceeding towards divergent selection (ω =>1) from purifying selection (ω =<1). A virus undergoes divergent selection when novel conditions are created by i) change of environment over short time, ii)heterogenous environment with multiple niches (Elena and Sanjuán 2005). SARS-CoV2 attaches with key entry-point residues in ACE2 receptor that is highly expressed in pneumocytes in lung, endocytes in gut and nasal goblet cells but also expresses in almost all tissues in the human body with a moderate level (Ziegler, Allon et al. 2020). Thus, SARS-CoV2 has ability to invade multiple organs having ACE2 receptor expressions and faces constant challenges from tissue specific immune surveillance with multiple niches of environmental conditions. Invasion of SARS-CoV2 into various human organs is predictably increasing the nonsynonymous mutation over synonymous mutations that are persisted over the deleterious mutation and, increases its selection fitness (Elena and Sanjuán 2005, Peck and Lauring 2018). *In vitro* studies confirmed the attainment of high ω value for HIV1 env gene when propagated in multiple tissue specific propagation (Sanjuán, Codoñer et al. 2004). Indeed, it has been shown that the turnover rate of the type of infected cells (tissue specificity) is positively associated with the rate of HIV1 viral evolution (Hicks and Duffy 2014).

Our estimation of evolutionary time taken by the SARS-CoV2 from bat RaTG13 lineage using strict molecular clock based on both base substitution model and neutral evolution are in close proximation of time frame of 9.6 years to 13.6 years. SARS-CoV2 might directly came from RaTG13 of bat or through intermediate host with efficient entry-point residues but highly adapted to replicate with slower mutation rate and, survive in a specialized immune system of the human body. After adaptation in human host, it gained more virulence by further substitution followed by selection pressure. Virulency of SARs-CoV2 is further evidenced by a strain containing D614G mutation believed to be the reason of widespread infection in USA and Europe. This mutation creates an extra serine protease cleavage site at the S1/S2 junction of the spike protein and facilitate further infectivity in Caucasians with a Del C (rs35074065)

genotypic background in the intergenic region between TMPRSS2 and MX1 gene (Korber, Fischer et al. 2020). Zhang et al (2020) showed that 614G mutated protein reduces S1 shedding and increase infectivity (Zhang, Jackson et al. 2020). However, van Dorp et al (2020) (van Dorp, Richard et al. 2020) did not find any evidence of a particular strain of SARS-CoV2 that has over-infecting ability in the population although Long et al (2020) convincingly showed the widespread infectivity of this strain in southern USA population (Long, Olsen et al. 2020),.However, very recently a super-infective b.1.1.7 strain carries a set of spike protein mutations including D614G, N501Y and P681H (Rambaut, Loman et al. 2020). This N501Y (AAT->TAT) conversion is believed to provide further stronger K353-501Y attachment to enable more infective power. P681H resides in the furin cleavage site in this strain predictably with improved function. However, these three mutations were observed independently but not together in one strain like b.1.1.7. We also detected P681H in a strain (isolated in New Mexico, USA, in September isolate,) in several people (e.g., Acc no. MW075768) and D614G independently. It appears that other spike mutations are selected on D614G carrying background to give better adaptability in different geographical places to evade challenging human immune environment.

Y493Q needs mutation in two nucleotides but could occur through a genetic drift. However, such an intermediate CoV virus are yet to be identified in bat and, also the silent presence of SARS-CoV2 related virus is neither documented in human for long time nor with any primate population that are suffered due to this viral attack (recent mink infection is by a mutated SARS-CoV2). Although, with the current genomic and amino acid sequences of SARS-COV2 having 493Q and 501N in the RBM suggests that SARS-CoV2 could infect any of the primate or higher order mammals as intermediate host having K31 and K353 residues in their ACE2 receptor gene. Li et al (2004) (Li, Giorgi et al. 2020) also suggested that such an intermediate host can never be identified. Although, it is impossible to conclude that such an intermediate host or virus can never be found, systematic investigations are expected.

Limitations in our study may involve a biased sampling of a particular variant strain that could represent repetitively over other low mutating strain or inclusion of a single genome consists of repeated mismatch. Also, we

wanted to assess here the average mutation rate in SARS-CoV2 virus undergoing substitutions to evolute to become a better strain. Another important consideration is that we did not observe any recombination or big insertions in a particular strain as frequent occurrence of those event could increase the mutation rate further. In September isolates, during substitution rate estimation, synonymous rate is higher than nonsynonymous mutations, but they are almost equal during evolutionary rate estimation as later is accounted for proportionate calculation in the same sequences. Lastly, we estimated the evolutionary rate of SARS-CoV2 in human host and, those values are used to calculate the TMRCA from bat RaTG13 lineage. Nevertheless, to take less time than our estimation time to evolute in bat than human (<9-14 years) could presume that bat system must have higher mutation rate than human which further assumes that it had to face much more challenging environment in bat than human but that is not expected as RaTG13 is native virus (long time adaptation) or circulating in bat for long time (Andersen, Rambaut et al. 2020).

Conclusion

SARS-CoV2 evolution comprises any of these three possibilities: it entered human host from bat early with its poorly developed entry-point residues much before its known appearance and remained silent for long time with slower mutation rate to evade human immune system or recently with efficiently developed entry-point residues having more infective power but adapted with higher mutation rate or recently through an intermediate host having human like conditions. Taken together, our analysis predicts the presence of SARs-COV2 virus in human for a long time (9-14 years) even it could be silent but do not satisfy any of these conditions such as very high mutation rate of SARS-CoV2 or a must needed intermediate host carrying intermediate virus with 493H. Furthermore, after its known appearance in human, it is proceeding towards aggressive and divergent selection predictably due to invasion to multiple organs in diverse population of the world. Due to excessive constrained immune and environmental challenges, SARS-CoV2 is evolving rapidly with proportionately more nonsynonymous amino acid changes over synonymous amino acids and is acquiring excessive selective fitness to be stable to create pandemic. Moreover, nucleotide evolution suggests that

SARS-CoV2 is likely to be originated from RaTG13 of bat without the recombination of S1 RBD region from pangolin CoV virus.

Funding information: This work is not funded

Declaration: There is no conflict of interest

Data availability: All data in this manuscript will be available to public upon acceptance of the paper.

Contribution: Amit K Maiti conceived the idea, performed analysis and wrote the manuscript.

Acknowledgement

I am indebted to Dr. Sudhiranjan Gupta for critically reading the manuscript and Mr. S. Hasan and U. Biswas for assistance in preparing the manuscript.

References

Andersen, K. G., A. Rambaut, W. I. Lipkin, E. C. Holmes and R. F. Garry (2020). "The proximal origin of SARS-CoV-2." Nat Med **26**(4): 450-452.

Boni, M. F., P. Lemey, X. Jiang, T. T. Lam, B. W. Perry, T. A. Castoe, A. Rambaut and D. L. Robertson (2020). "Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic." Nat Microbiol **5**(11): 1408-1417.

Drummond, A. J. and M. A. Suchard (2010). "Bayesian random local clocks, or one rate to rule them all." BMC Biol **8**: 114.

Duchêne, S., E. C. Holmes and S. Y. Ho (2014). "Analyses of evolutionary dynamics in viruses are hindered by a time-dependent bias in rate estimates." Proc Biol Sci **281**(1786).

E.C., H. (2009). The evolution and emergence of RNA viruses. . New York, NY: Oxford University Press.

Elena, S. F. and R. Sanjuán (2005). "Adaptive value of high mutation rates of RNA viruses: separating causes from consequences." J Virol **79**(18): 11555-11558.

Felsenstein, J. (1981). "Evolutionary trees from DNA sequences: a maximum likelihood approach." J Mol Evol **17**(6): 368-376.

Forster, P., L. Forster, C. Renfrew and M. Forster (2020). "Phylogenetic network analysis of SARS-CoV-2 genomes." Proc Natl Acad Sci U S A **117**(17): 9241-9243.

Gojobori, T., E. N. Moriyama and M. Kimura (1990). "Molecular clock of viral evolution, and the neutral theory." Proc Natl Acad Sci U S A **87**(24): 10015-10018.

Gonzalez-Reiche, A. S., M. M. Hernandez, M. J. Sullivan, B. Ciferri, H. Alshammary, A. Obla, S. Fabre, G. Kleiner, J. Polanco, Z. Khan, B. Alburquerque, A. van de Guchte, J. Dutta, N. Francoeur, B. S. Melo, I. Oussenko, G. Deikus, J. Soto, S. H. Sridhar, Y. C. Wang, K. Twyman, A. Kasarskis, D. R. Altman, M. Smith, R. Sebra, J. Aberg, F. Krammer, A. García-Sastre, M. Luksza, G. Patel, A. Paniz-Mondolfi, M. Gitman, E. M. Sordillo, V. Simon and H. van Bakel (2020). "Introductions and early spread of SARS-CoV-2 in the New York City area." Science.

Hicks, A. L. and S. Duffy (2014). "Cell tropism predicts long-term nucleotide substitution rates of mammalian RNA viruses." PLoS Pathog **10**(1): e1003838.

Ho, S. Y. and S. Duchêne (2014). "Molecular-clock methods for estimating evolutionary rates and timescales." Mol Ecol **23**(24): 5947-5965.

Hoffmann, M., H. Kleine-Weber and S. Pöhlmann (2020). "A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells." Mol Cell **78**(4): 779-784.e775.

Hoffmann, M., H. Kleine-Weber, S. Schroeder, N. Krüger, T. Herrler, S. Erichsen, T. S. Schiergens, G. Herrler, N. H. Wu, A. Nitsche, M. A. Müller, C. Drosten and S. Pöhlmann (2020). "SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor." Cell.

Holmes, E. C. (2009). "RNA virus genomics: a world of possibilities." J Clin Invest **119**(9): 2488-2495.

Huelsenbeck, J. P., B. Larget and D. Swofford (2000). "A compound poisson process for relaxing the molecular clock." Genetics **154**(4): 1879-1892.

Islam, M. R., M. N. Hoque, M. S. Rahman, A. S. M. R. Alam, M. Akther, J. A. Puspo, S. Akter, M. Sultana, K. A. Crandall and M. A. Hossain (2020). "Genome-wide analysis of SARS-CoV-2 virus strains circulating worldwide implicates heterogeneity." Sci Rep **10**(1): 14004.

Kimura, M. (1979). "The neutral theory of molecular evolution." Sci Am **241**(5): 98-100, 102, 108 passim.

Kimura, M. (1981). "Estimation of evolutionary distances between homologous nucleotide sequences." Proc Natl Acad Sci U S A **78**(1): 454-458.

Korber, B., W. M. Fischer, S. Gnanakaran, H. Yoon, J. Theiler, W. Abfalterer, N. Hengartner, E. E. Giorgi, T. Bhattacharya, B. Foley, K. M. Hastie, M. D. Parker, D. G. Partridge, C. M. Evans, T. M. Freeman, T. I. de Silva, C. McDanal, L. G. Perez, H. Tang, A. Moon-Walker, S. P. Whelan, C. C. LaBranche, E. O. Saphire, D. C. Montefiori and S. C.-G. Group (2020). "Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus." Cell **182**(4): 812-827.e819.

Kryazhimskiy, S. and J. B. Plotkin (2008). "The population genetics of dN/dS." PLoS Genet **4**(12): e1000304.

Latinne, A., B. Hu, K. J. Olival, G. Zhu, L. Zhang, H. Li, A. A. Chmura, H. E. Field, C. Zambrana-Torrelio, J. H. Epstein, B. Li, W. Zhang, L. F. Wang, Z. L. Shi and P. Daszak (2020). "Origin and cross-species transmission of bat coronaviruses in China." Nat Commun **11**(1): 4235.

Lee, H. J., N. Rodrigue and J. L. Thorne (2015). "Relaxing the Molecular Clock to Different Degrees for Different Substitution Types." Mol Biol Evol **32**(8): 1948-1961.

Lepage, T., D. Bryant, H. Philippe and N. Lartillot (2007). "A general comparison of relaxed molecular clock models." Mol Biol Evol **24**(12): 2669-2680.

Li, W. H., M. Tanimura and P. M. Sharp (1988). "Rates and dates of divergence between AIDS virus nucleotide sequences." Mol Biol Evol **5**(4): 313-330.

Li, W. H., C. I. Wu and C. C. Luo (1985). "A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes." Mol Biol Evol **2**(2): 150-174.

Li, X., E. E. Giorgi, M. H. Marichannegowda, B. Foley, C. Xiao, X. P. Kong, Y. Chen, S. Gnanakaran, B. Korber and F. Gao (2020). "Emergence of SARS-CoV-2 through recombination and strong purifying selection." Sci Adv **6**(27).

Li, X., E. H. Giorgi, M. H. Marichannegowda, B. Foley, C. Xiao, X.-P. Kong, Y. Chen, S. Gnanakaran, B. Korber and F. Gao (2020). Emergence of SARS-CoV-2 through recombination and strong purifying selection. Sci. Adv.

Lin, J. J., M. J. Bhattacharjee, C. P. Yu, Y. Y. Tseng and W. H. Li (2019). "Many human RNA viruses show extraordinarily stringent selective constraints on protein evolution." Proc Natl Acad Sci U S A **116**(38): 19009-19018.

Liu, Y., D. C. Nickle, D. Shriner, M. A. Jensen, G. H. Learn, J. E. Mittler and J. I. Mullins (2004). "Molecular clock-like evolution of human immunodeficiency virus type 1." Virology **329**(1): 101-108.

Long, S. W., R. J. Olsen, P. A. Christensen, D. W. Bernard, J. J. Davis, M. Shukla, M. Nguyen, M. O. Saavedra, P. Yerramilli, L. Pruitt, S. Subedi, H. C. Kuo, H. Hendrickson, G. Eskandari, H. A. T. Nguyen, J. H. Long, M. Kumaraswami, J. Goike, D. Boutz, J. Gollihar, J. S. McLellan, C. W. Chou, K. Javanmardi, I. J. Finkelstein and J. M. Musser (2020). "Molecular Architecture of Early Dissemination and Massive Second Wave of the SARS-CoV-2 Virus in a Major Metropolitan Area." mBio **11**(6).

Nei, M. and T. Gojobori (1986). "Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions." Mol Biol Evol **3**(5): 418-426.

Orr, H. A. (2000). "The rate of adaptation in asexuals." Genetics **155**(2): 961-968.

Paradis, E., P. A. Tedesco and B. Hugueny (2013). "Quantifying variation in speciation and extinction rates with clade data." Evolution **67**(12): 3617-3627.

Peck, K. M. and A. S. Lauring (2018). "Complexities of Viral Mutation Rates." J Virol **92**(14).

Posada, D. and K. A. Crandall (2001). "Selecting models of nucleotide substitution: an application to human immunodeficiency virus 1 (HIV-1)." Mol Biol Evol **18**(6): 897-906.

Rambaut, A., N. Loman, O. Pybus , W. Barclay, T. J. Barrett, A. Carabelli, N. Connor, T. Peacock, D. Robertson, E. Volz and o. b. o. C.-G. C. U. (CoG-UK) (2020). Preliminary genomic characterization of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. virological.org.

Ren, L. L., Y. M. Wang, Z. Q. Wu, Z. C. Xiang, L. Guo, T. Xu, Y. Z. Jiang, Y. Xiong, Y. J. Li, X. W. Li, H. Li, G. H. Fan, X. Y. Gu, Y. Xiao, H. Gao, J. Y. Xu, F. Yang, X. M. Wang, C. Wu, L. Chen, Y. W. Liu, B. Liu, J. Yang, X. R. Wang, J. Dong, L. Li, C. L. Huang, J. P. Zhao, Y. Hu, Z. S. Cheng, L. L. Liu, Z. H. Qian, C. Qin, Q. Jin, B. Cao and J. W. Wang (2020). "Identification of a novel coronavirus causing severe pneumonia in human: a descriptive study." Chin Med J (Engl).

Sanderson, M. J. (2002). "Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach." Mol Biol Evol **19**(1): 101-109.

Sanjuán, R. (2010). "Mutational fitness effects in RNA and single-stranded DNA viruses: common patterns revealed by site-directed mutagenesis studies." Philos Trans R Soc Lond B Biol Sci **365**(1548): 1975-1982.

Sanjuán, R., F. M. Codoñer, A. Moya and S. F. Elena (2004). "Natural selection and the organ-specific differentiation of HIV-1 V3 hypervariable region." Evolution **58**(6): 1185-1194.

Sanjuán, R., M. R. Nebot, N. Chirico, L. M. Mansky and R. Belshaw (2010). "Viral mutation rates." J Virol **84**(19): 9733-9748.

Stern, A., M. T. Yeh, T. Zinger, M. Smith, C. Wright, G. Ling, R. Nielsen, A. Macadam and R. Andino (2017). "The Evolutionary Pathway to Virulence of an RNA Virus." Cell **169**(1): 35-46.e19.

Temin, H. M. (1989). "Retrovirus variation and evolution." Genome **31**(1): 17-22.

Thorne, J. L., H. Kishino and I. S. Painter (1998). "Estimating the rate of evolution of the rate of molecular evolution." Mol Biol Evol **15**(12): 1647-1657.

van Dorp, L., D. Richard, C. C. S. Tan, L. P. Shaw, M. Acman and F. Balloux (2020). "No evidence for increased transmissibility from recurrent mutations in SARS-CoV-2." Nat Commun **11**(1): 5986.

Wan, Y., J. Shang, R. Graham, R. S. Baric and F. Li (2020). "Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus." J Virol **94**(7).

Wang, Q., Y. Zhang, L. Wu, S. Niu, C. Song, Z. Zhang, G. Lu, C. Qiao, Y. Hu, K. Y. Yuen, H. Zhou, J. Yan and J. Qi (2020). "Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2." Cell **181**(4): 894-904.e899.

Worobey, M., P. Telfer, S. Souquière, M. Hunter, C. A. Coleman, M. J. Metzger, P. Reed, M. Makuwa, G. Hearn, S. Honarvar, P. Roques, C. Apetrei, M. Kazanji and P. A. Marx (2010). "Island biogeography reveals the deep history of SIV." Science **329**(5998): 1487.

Wu, F., S. Zhao, B. Yu, Y. M. Chen, W. Wang, Z. G. Song, Y. Hu, Z. W. Tao, J. H. Tian, Y. Y. Pei, M. L. Yuan, Y. L. Zhang, F. H. Dai, Y. Liu, Q. M. Wang, J. J. Zheng, L. Xu, E. C. Holmes and Y. Z. Zhang (2020). "A new coronavirus associated with human respiratory disease in China." Nature **579**(7798): 265-269.

Yang, Z. and J. P. Bielawski (2000). "Statistical methods for detecting molecular adaptation." Trends Ecol Evol **15**(12): 496-503.

Yang, Z. and R. Nielsen (2000). "Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models." Mol Biol Evol **17**(1): 32-43.

Yoder, A. D. and Z. Yang (2000). "Estimation of primate speciation dates using local molecular clocks." Mol Biol Evol **17**(7): 1081-1090.

Zhang, L., C. B. Jackson, H. Mou, A. Ojha, H. Peng, B. D. Quinlan, E. S. Rangarajan, A. Pan, A. Vanderheiden, M. S. Suthar, W. Li, T. Izard, C. Rader, M. Farzan and H. Choe (2020). "SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity." Nat Commun **11**(1): 6013.

Zhou, P., X. L. Yang, X. G. Wang, B. Hu, L. Zhang, W. Zhang, H. R. Si, Y. Zhu, B. Li, C. L. Huang, H. D. Chen, J. Chen, Y. Luo, H. Guo, R. D. Jiang, M. Q. Liu, Y. Chen, X. R. Shen, X. Wang, X. S. Zheng, K. Zhao, Q. J. Chen, F. Deng, L. L. Liu, B. Yan, F. X. Zhan, Y. Y. Wang, G. F. Xiao and Z. L. Shi (2020). "A pneumonia outbreak associated with a new coronavirus of probable bat origin." Nature **579**(7798): 270-273.

Ziegler, C. G. K., S. J. Allon, S. K. Nyquist, I. M. Mbano, V. N. Miao, C. N. Tzouanas, Y. Cao, A. S. Yousif, J. Bals, B. M. Hauser, J. Feldman, C. Muus, M. H. Wadsworth, S. W. Kazer, T. K. Hughes, B. Doran, G. J. Gatter, M. Vukovic, F. Taliaferro, B. E. Mead, Z. Guo, J. P. Wang, D. Gras, M. Plaisant, M. Ansari, I. Angelidis, H. Adler, J. M. S. Sucre, C. J. Taylor, B. Lin, A. Waghray, V. Mitsialis, D. F. Dwyer, K. M. Buchheit, J. A. Boyce, N. A. Barrett, T. M. Laidlaw, S. L. Carroll, L. Colonna, V. Tkachev, C. W. Peterson, A. Yu, H. B. Zheng, H. P. Gideon, C. G. Winchell, P. L. Lin, C. D. Bingle, S. B. Snapper, J. A. Kropski, F. J. Theis, H. B. Schiller, L. E. Zaragosi, P. Barbry, A. Leslie, H. P. Kiem, J. L. Flynn, S. M. Fortune, B. Berger, R. W. Finberg, L. S. Kean, M. Garber, A. G. Schmidt, D. Lingwood, A. K. Shalek, J. Ordovas-Montanes, H. L. B. N. E. a. lung-network@humancellatlas.org and H. L. B. Network (2020). "SARS-CoV-2 Receptor ACE2 Is an Interferon-Stimulated Gene in Human Airway Epithelial Cells and Is Detected in Specific Cell Subsets across Tissues." Cell **181**(5): 1016-1035.e1019.

Zuckerkandl, E. and L. Pauling (1965). "Molecules as documents of evolutionary history." J Theor Biol **8**(2): 357-366.

Zukes, T. H. and C. R. Cantor (1969). Evolution of protein molecules, H. N. MUNRO, ed. Mammalian protein metabolism III. Academic Press, New York. **2:** 1-132.

Supplementary Table 1: Mutation rate of SARS-CoV2 from January 2020 to Sept, 2020								
EUROPEAN NUCLEOTIDE ARCHIEVE (ENA) (www.covid19dataportal.org);NCBI (https://www.ncbi.nlm.nih.gov/nuccore/)								
Collection Date	Reference	Identity with ref. genome	NT differences	Special feature	Accession No	Place of collection	Synonymous mutation	Nonsynonymous mutation
12/1/1919	NC_045512.2	100%	0	None	MN908947	Wuhan, China		
Dec 30, 2019-Jan10, 2020								
2/1/2020	NC_045512.2	100%	0.00	None	MN988668	Wuhan	0.00	0.00
2/1/2020	NC_045512.2	100%	0.00	None	MN988669	Wuhan	0.00	0.00
1/1/2020	NC_045512.2	99%	5.00	None	LC522973	Japan	4.00	1.00
1/2/2020	NC_045512.2	99%	4.00	None	LC522972	Japan	3.00	1.00
1/1/2020	NC_045512.2	99%	2.00	None	MN996531	Wuhan	0.00	2.00
1/1/2020	NC_045512.2	99%	2.00	None	MN996527	Wuhan	1.00	1.00
12/30/2019	NC_045512.2	99%	2.00	None	MN996529	Wuhan	0.00	2.00
12/30/2019	NC_045512.2	100%	0.00	None	MN996528	Wuhan	0.00	0.00
1/1/2020	NC_045512.2	99%	3.00	None	LC529905	Japan	0.00	0.00
8/1/2020	NC_045512.3	100%	0.00	None	MT093631	China	0.00	0.00
8/1/2020	NC_045512.4	99%	4.00	None	LC522974	Japan	3.00	1.00
12/30/2019	NC_045512.5	100%	0.00	None	MN996530	Wuhan	0.00	0.00
12/30/2019	NC_045512.6	100%	3.00	None	LR757995	Wuhan	2.00	0.00
12/26/2019	NC_045512.7	100%	2.00	None	LR757998	Wuhan	2.00	0.00
Total NT difference			27				15.00	8.00
Average NT difference			1.93	14			1.07	0.57
std			1.73				1.88	
Substitution Rate							1.07	0.57
Feb, 1st-Feb, 2020								
2/6/2020	NC_045512.2	99%	7	None	MT093571	Sweden	5	2.00
2/10/2020	NC_045512.2	99%	1	None	MT106053	CA, USA	1	0.00
2/1/2020	NC_045512.2	99%	2	None	MT365031	Hongkong	1	1.00
2/1/2020	NC_045512.2	99%	2	None	MT276597	Israel	1	1.00
2/10/2020	NC_045512.2	99%	6	None	LC528232	japan	5	1.00
2/6/2020	NC_045512.2	99%	4	None	MT106052	CA, USA	4	0.00
2/2/2020	NC_045512.2	99%	2	None	MT106053		2	0.00
2/2/2020	NC_045512.2	99%	9	None	MT121215	shanghai, China	8	3.00
2/3/2020	NC_045512.2	99%	4	None	LC542976	Japan	4	0.00
2/10/2020	NC_045512.2	99%	7	None	MT106054	TX, USA	6	1.00

2/5/2020	NC_045512.3	99%	2	None	MT066176	Taiwan	1	1.00
2/5/2020	NC_045512.4	99%	4	None	MT374101	Taiwan	3	1.00
1/31/2020	NC_045512.5	99%	1	None	MT039887	WI, USA	0	1.00
1/31/2020	NC_045512.6	99%	4	None	MT365032	Hongkong	3	1.00
1/31/2020	NC_045512.7	99%	3	None	MT365030	Hongkong	2	1.00
1/31/2020	NC_045512.8	99%	6	None	MT050493	Kerala, India	3	3.00
2/5/2020	NC_045512.9	99%	3	None	MT123290	GuangDong, China	3	0.00
2/6/2020	NC_045512.10	99%	8	None	MT198652	Valencia, Spain	7	3.00
Total NT difference			75				59	20.00
Average NT difference			4.17	18			3.277777778	1.11
std			2.46				2.95	
Substitution rate			2.24				2.21	0.54
March,1st-March10, 2020								
3/1/2020	NC_045512.2	99%	6.00	None	MT322401	Virginia, USA	5.00	1.00
3/6/2020	NC_045512.2	99%	6.00	None	MT325580	Kansas, USA	5.00	1.00
3/9/2020	NC_045512.2	99%	10.00	None	MT325581	Louisiana, USA	5.00	5.00
3/9/2020	NC_045512.2	99%	6.00	None	MT325579	Indiana, USA	4.00	2.00
3/8/2020	NC_045512.2	99%	12.00	None	MT325608	S Carolina, USA	7.00	5.00
3/5/2020	NC_045512.2	99%	5.00	None	MT325609	Utah, USA	3.00	2.00
3/6/2020	NC_045512.2	99%	4.00	None	MT325616	Massachusetts, USA	3.00	1.00
3/6/2020	NC_045512.2	99%	6.00	None	MT325627	NY, USA	4.00	2.00
3/8/2020	NC_045512.2	99%	4.00	None	MT325620	Iowa, USA	3.00	1.00
3/5/2020	NC_045512.2	99%	4.00	None	MT344948	Hawaii, USA	3.00	1.00
3/7/2020	NC_045512.2	99%	5.00	None	MT344962	MN, USA	3.00	2.00
3/7/2020	NC_045512.2	99%	7.00	None	MT344960	RI, USA	5.00	2.00
3/8/2020	NC_045512.2	99%	13.00	None	MT252805	WA, USA	11.00	2.00
3/1/2020	NC_045512.2	99%	6.00	None	MT320538	France	4.00	2.00
3/10/2020	NC_045512.2	99%	7.00	None	MT412135	MI, USA	4.00	3.00
3/1/2020	NC_045512.3	99%	7.00	None	MT322394	Virginia, USA	5.00	2.00
3/1/2020	NC_045512.4	99%	10.00	None	MT322398	Virginia, USA	6.00	4.00
3/1/2020	NC_045512.5	99%	7.00	None	MT322402	Virginia, USA	6.00	1.00
3/1/2020	NC_045512.6	99%	6.00	None	MT322404	Virginia, USA	4.00	2.00

3/1/2020	NC_045512.7	99%	8.00	None	MT322405	Virginia, USA	7.00	1.00
3/1/2020	NC_045512.8	99%	9.00	None	MT322406	Virginia, USA	6.00	3.00
3/1/2020	NC_045512.9	99%	6.00	None	MT322412	Virginia, USA	4.00	2.00
3/1/2020	NC_045512.10	99%	8.00	None	MT322415	Virginia, USA	6.00	2.00
3/5/2020	NC_045512.11	99%	5.00	None	MT325566	FL, USA	3.00	2.00
3/6/2020	NC_045512.12	99%	9.00	None	MT325570	GA, USA	6.00	3.00
3/5/2020	NC_045512.13	99%	7.00	None	MT374102	Taiwan	4.00	3.00
3/5/2020	NC_045512.10	99%	8.00	None	MT198652	Valencia, Spain	7.00	1.00
Total NT difference			191				133.00	58.00
Average NT difference			7.07				4.93	2.18
std			2.29				2.26	
Substitution Rate			2.91				1.65	1.07
April, 1st-April 10, 2020								
4/1/2020	NC_045512.2	99%	7.00	None	MT419821	Puerto Rico, USA	5.00	2.00
4/1/2020	NC_045512.2	99%	8.00	None	MT350246	VA, USA	5.00	3.00
4/1/2020	NC_045512.2	99%	11.00	None	MT358401	LA,USA	6.00	5.00
4/1/2020	NC_045512.2	99%	12.00	None	MT412313	WA, USA	8.00	4.00
4/1/2020	NC_045512.2	99%	10.00	None	MT412319	CT,USA	8.00	2.00
4/1/2020	NC_045512.2	99%	9.00	None	MT418883	VA,USA	6.00	3.00
4/1/2020	NC_045512.2	99%	8.00	None	MT358637	Rajkot, India	7.00	1.00
4/1/2020	NC_045512.2	99%	11.00	None	MT358644	WA, USA	8.00	3.00
4/4/2020	NC_045512.2	99%	8.00	None	MT385433	CA, USA	5.00	3.00
4/4/2020	NC_045512.2	99%	14.00	None	MT412255	WA, USA	10.00	4.00
4/1/2020	NC_045512.3	99%	14.00	None	MT358656	WA, USA	11.00	3.00
4/1/2020	NC_045512.4	99%	10.00	None	MT358666	WA, USA	7.00	3.00
4/1/2020	NC_045512.5	99%	12.00	None	MT358744	WA, USA	7.00	5.00
4/1/2020	NC_045512.6	99%	6.00	None	MT358743	WA, USA	4.00	2.00
4/7/2020	NC_045512.7	99%	11.00	None	MT375463	WA, USA	9.00	2.00
4/7/2020	NC_045512.8	99%	8.00	None	MT375470	WA, USA	6.00	2.00
4/1/2020	NC_045512.9	99%	11.00	None	MT350257	VA, USA	8.00	3.00
4/1/2020	NC_045512.10	99%	12.00	None	MT358650	WA, USA	10.00	2.00
4/6/2020	NC_045512.4	99%	15.00	None	MT535509.1	UT,USA	13.00	2.00
4/6/2020	NC_045512.5	99%	13.00	None	MT535508.1	UT,USA	10.00	3.00
4/5/2020	NC_045512.6	99%	12.00	None	MT535507.1	UT,USA	4.00	8.00
Total NT difference			222				157.00	63.00
Average NT difference			10.57	21			7.48	3.10

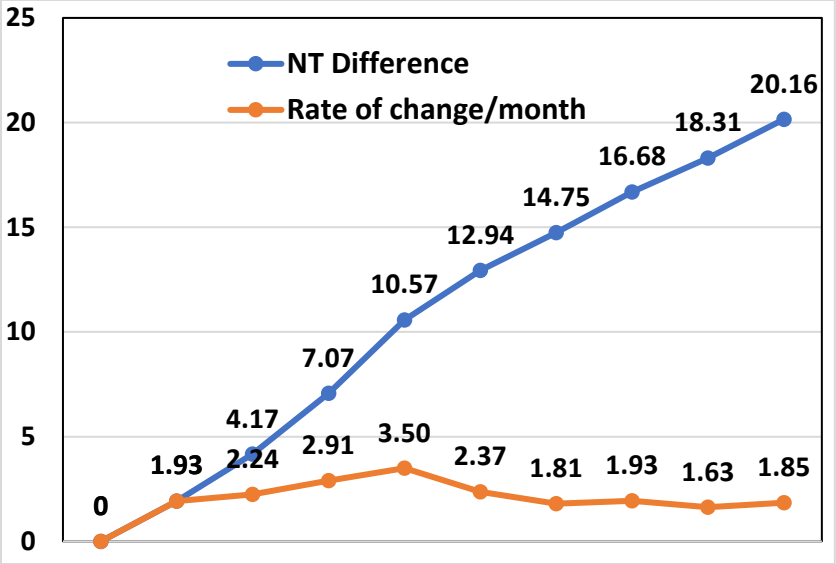
std			2.46				2.42	
Substitution Rate			3.50				2.55	0.92
April, 30th-May,10, 2020								
5/6/2020	NC_045512.2	99%	9.00	None	MT481933.1	CA, USA	5.00	4.00
5/6/2020	NC_045512.2	99%	9.00	None	MT481932.1	CA, USA	5.00	4.00
4/30/2020	NC_045512.2	99%	10.00	None	MT482146.1	VA, USA	7.00	3.00
4/30/2020	NC_045512.2	99%	16.00	None	MT482145.1	VA, USA	13.00	3.00
4/6/2020	NC_045512.3	99%	11.00	None	MT481928.1	CA, USA	11.00	4.00
4/6/2020	NC_045512.4	99%	15.00	None	MT535509.1	UT,USA	13.00	2.00
5/12/2020	NC_045512.2	99.94%	17	33nt (bad seq deleted	MW040613	FL,USA	12.00	5.00
5/6/2020	NC_045512.3	99.93%	18.00	None	MW023417	FL,USA	14.00	4.00
5/9/2020	NC_045512.2	99.97%	10	None	MW039108	TX, USA	6.00	4.00
5/9/2020	NC_045512.2	99.98%	10	None	MW039111	TX, USA	6.00	4.00
5/9/2020	NC_045512.2	99.97%	14	None	MW039113	TX, USA	9.00	5.00
5/9/2020	NC_045512.2	99.98%	10	None	MW039114	TX, USA	6.00	4.00
5/9/2020	NC_045512.2	99.98%	10	None	MW039116	TX, USA	7.00	3.00
5/7/2020	NC_045512.2	99.81%	17	None	MT873050	MA, USA	15.00	2.00
5/9/2020	NC_045512.2	99.81%	12	None	MT873051	MA, USA	10.00	2.00
5/4/2020	NC_045512.2	99.80%	20	None	MT873079	MA, USA	13.00	7.00
5/1/2020	NC_045512.2	99.81%	14	None	MT873081	MA, USA	12.00	2.00
5/2/2020	NC_045512.2	99.95%	11	None	MT873082	MA, USA	9.00	2.00
Total NT difference			233				173.00	64.00
Average NT difference			12.94				9.61	3.56
std			3.39				2.70	
Substitution rate			2.37				2.13	0.46
June, 1st-Jun10, 2020								
5/31/2020	NC_045512.2	99.95%	13		MW035935	CA, USA	7.00	6.00
5/31/2020	NC_045512.2	99.96%	14		MW035933	CA, USA	8.00	6.00
6/2/2020	NC_045512.2	99.81%	16		MW036056	CA, USA	8.00	8.00
6/4/2020	NC_045512.2	99.95%	12		MW023429	FL,USA	6.00	6.00
6/4/2020	NC_045512.2	99.94%	17		MW023431	FL,USA	14.00	3.00
6/8/2020	NC_045512.2	99.98%	12		MW023433	FL,USA	10.00	2.00
6/4/2020	NC_045512.2	99.99%	11		MW023425	FL,USA	0.00	11.00
6/4/2020	NC_045512.2	99.99%	10		MW023426	FL,USA	8.00	2.00
6/4/2020	NC_045512.2	99.97%	14		MW023428	FL,USA	12.00	2.00

6/2/2020	NC_045512.2	99.97%	13		MW023424	FL,USA	11.00	2.00
6/4/2020	NC_045512.2	99.97%	16		MW075709.	TX,USA	10.00	6.00
6/8/2020	NC_045512.2	99.97%	13		MW075711	TX,USA	11.00	2.00
6/9/2020	NC_045512.2	99.97%	15		MW070038	NM, USA	13.00	2.00
6/9/2020	NC_045512.2	99.96%	17		MW070032	NM, USA	13.00	4.00
6/9/2020	NC_045512.2	99.93%	22		MW070033	NM, USA	18.00	4.00
6/9/2020	NC_045512.2	99.93%	21		MW070034	NM, USA	19.00	2.00
							168.00	68.00
Total NT difference			236				10.50	4.25
Average NT difference			14.75				2.47	
std			3.34				0.89	0.69
Substitution rate			1.81					
July, 1st-July10								
7/10/2020	NC_045512.2	99.95%	15		MW042923	SC, USA	9.00	6.00
7/10/2020	NC_045512.2	99.94%	18		MW018448	Puerto Rico	13.00	5.00
7/5/2020	NC_045512.2	99.91%	20		MW030237	Peru	13.00	7.00
7/1/2020	NC_045512.2	99.93%	16		MW030238	Peru	9.00	7.00
7/4/2020	NC_045512.2	99.93%	16		MW030240	Peru	9.00	7.00
7/4/2020	NC_045512.2	99.93%	15		MW030241	Peru	11.00	4.00
7/4/2020	NC_045512.2	99.93%	16		MW030242	Peru	11.00	5.00
7/1/2020	NC_045512.2	99.93%	16		MW070043	NM, USA	12.00	4.00
7/1/2020	NC_045512.2	99.92%	19		MW006587	NM, USA	12.00	7.00
7/9/2020	NC_045512.2	99.94%	15		MT969870	Victoria, Australia	13.00	2.00
7/9/2020	NC_045512.2	99.94%	15		MT969874	Victoria, Australia	13.00	2.00
7/9/2020	NC_045512.2	99.94%	15		MT969886	Victoria, Australia	3.00	12.00
7/9/2020	NC_045512.2	99.94%	16		MT972745	Victoria, Australia	10.00	6.00
7/7/2020	NC_045512.2	99.94%	15		MT876555	Bangladesh	10.00	5.00
7/9/2020	NC_045512.2	99.93%	17		MT876556	MO, USA	12.00	5.00
7/10/2020	NC_045512.2	99.93%	23		MT795897.1	WI, USA	13.00	10.00
7/4/2020	NC_045512.2	99.93%	15		MT795908	WI, USA	9.00	6.00
7/7/2020	NC_045512.2	99.93%	17		MT762396	Bangladesh	12.00	5.00
7/7/2020	NC_045512.2	99.93%	18		MT762398	Bangladesh	14.00	4.00
Total NT difference			317				208.00	109.00
Average NT difference			16.68				10.95	5.74
std			2.14				1.91	
Substitution rate			1.93				0.45	1.49

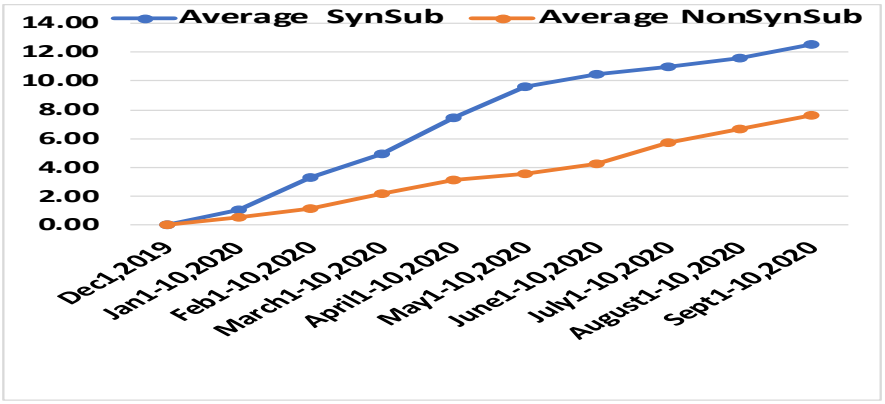
August ,1st-August,10th								
8/6/2020		99.93%	18		MW040625/	FL, USA	11.00	7.00
8/5/2020		99.93%	17		MW040621/	FL,USA	10.00	7.00
8/6/2020		99.93%	17		MW040623/	FL,USA	10.00	7.00
8/5/2020		99.92%	18		MW035371/	WI, USA	11.00	7.00
8/4/2020		99.92%	19		MW035375/	WI,USA	7.00	12.00
8/5/2020		99.92%	19		MW035377/	WI, USA	12.00	7.00
8/6/2020		99.92%	18		MW035379/	WI, USA	14.00	4.00
8/3/2020		99.92%	18		MT928991/	WI, USA	14.00	4.00
8/3/2020		99.92%	18		MT929005/	WI, USA	13.00	5.00
8/4/2020		99.92%	19		MW035381	WI, USA	13.00	6.00
8/4/2020		99.92%	19		MT958261/	MN, USA	11.00	8.00
8/5/2020		99.92%	17		MT956718	FL, USA	10.00	7.00
8/7/2020		99.92%	19		MW023467	FL, USA	11.00	7.00
8/3/2020		99.94%	18		MW023463	FL, USA	15.00	3.00
8/3/2020		99.93%	19		MW023466	FL, USA	11.00	8.00
8/3/2020		99.92%	20		MW023399	FL, USA	12.00	7.00
Total NT difference			293				185.00	106.00
Average NT difference			18.31				11.56	6.63
std			0.87				1.75	
Substitution rate			1.63				0.62	0.89
Sept, 1st-Sept,10								
9/3/2020	NC_045512.2	99.95%	17		MW040645/	FL,USA	11.00	6.00
9/8/2020	NC_045512.2	99.96%	15		MW040655/	FL,USA	9.00	6.00
9/3/2020	NC_045512.2	99.92%	22		MW035558	WI,USA	13.00	9.00
9/4/2020	NC_045512.2	99.92%	19		MW035560	WI,USA	11.00	8.00
9/5/2020	NC_045512.2	99.92%	23		MW035570	WI,USA	15.00	8.00
9/6/2020	NC_045512.2	99.92%	20		MW035542	WI,USA	14.00	6.00
9/4/2020	NC_045512.2	99.94%	18		MW035571	WI,USA	11.00	7.00
9/3/2020	NC_045512.2	99.92%	21		MW035572	WI,USA	13.00	8.00
9/5/2020	NC_045512.2	99.92%	21		MW035573	WI,USA	13.00	8.00
9/5/2020	NC_045512.2	99.92%	21		MW035574	WI,USA	13.00	8.00
9/4/2020	NC_045512.2	99.92%	19		MW035575	WI,USA	11.00	8.00
9/4/2020	NC_045512.2	99.93%	17		MW035576	WI,USA	11.00	6.00
9/4/2020	NC_045512.2	99.92%	21		MW035578	WI,USA	10.00	11.00
9/7/2020	NC_045512.2	99.92%	20		MW035556	MO,USA	12.00	8.00

9/2/2020	NC_045512.2	99.92%	23		MW035476	CA, USA	16.00	7.00
9/2/2020	NC_045512.2	99.92%	24		MW035466	CA, USA	17.00	7.00
9/1/2020	NC_045512.1	99.92%	22		MW035459	NY, USA	16.00	6.00
9/9/2020	NC_045512.1	99.92%	21		MW079419	Malta	12.00	9.00
9/3/2020	NC_045512.1	99.90%	19		MW070107.	NM, USA	10.00	9.00
Total NT difference			383				238.00	145.00
Average NT difference			20.16	17.00			12.53	7.63
std			2.32				1.64	
Substitution rate			1.85				0.96	1.01
Total sample				166.00				
Avg NT sub/month		2.21775		Avg Syn sub/ month	1.47	Avg NSyn sub/ month	0.85	
Avg NT sub/yr		26.61305		Avg Syn sub/yr	17.6932888	Avg Nsyn subs/ yr	10.18	
Avg NT sub/site/yr		0.00089		Avg Syn/ site /yr	0.00059169	Avg Nsyn subs/ site/yr	0.00034	
Evol. rate/site/yr		0.00089		Evol. rate /site/ yr	0.00059	Evol. rate/site/ yr	0.00034	
		~ 9.0E-04						

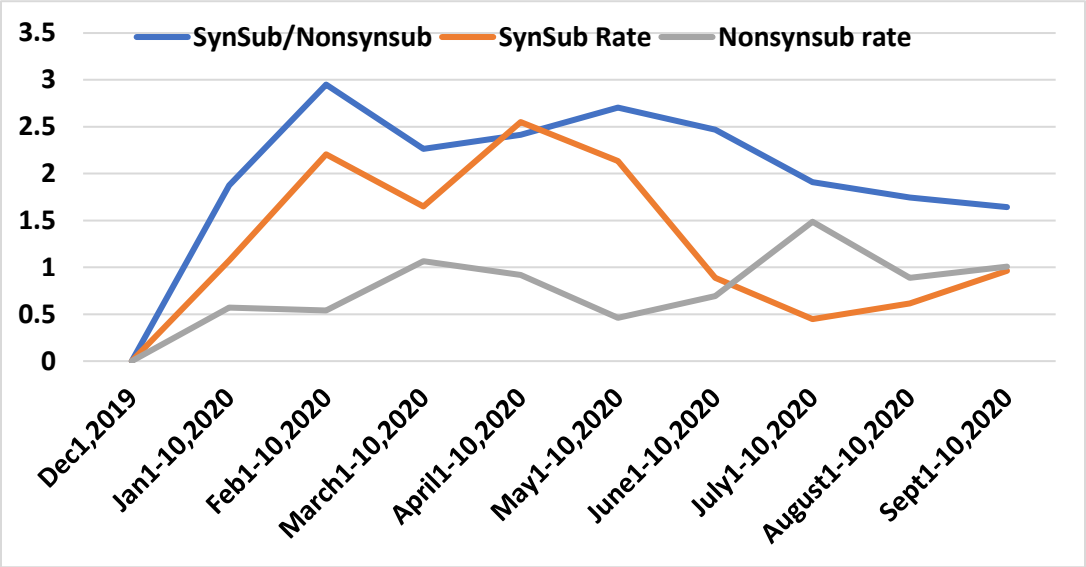
Avg. substitutions per month	Difference	Rate						
Dec1,2019	0	0						
Jan1-10,2020	1.93	1.93						
Feb1-10,2020	4.17	2.24						
March1-10,2020	7.07	2.91						
April1-10,2020	10.57	3.50						
May1-10,2020	12.94	2.37						
June1-10,2020	14.75	1.81						
July1-10,2020	16.68	1.93						
August1-10,2020	18.31	1.63						
August1-10,2021	20.16	1.85						



Months	Average SynSub	Average NonSynSub	Average SynSub	Average NonSynSub		Rate			
Dec1,2019	0.00	0.00	0	0.00		0.00			
Jan1-10,2020	1.07	0.57	15.00	8.00		1.88			
Feb1-10,2020	3.28	1.11	59.00	20.00		2.95			
March1-10,2020	4.93	2.18	133.00	58.00		2.26			
April1-10,2020	7.48	3.10	157.00	63.00		2.42			
May1-10,2020	9.61	3.56	173.00	64.00		2.70			
June1-10,2020	10.50	4.25	168.00	68.00		2.47			
July1-10,2020	10.95	5.74	208.00	109.00		1.91			
August1-10,2020	11.56	6.63	185.00	106.00		1.75			
Sept1-10,2020	12.53	7.63	238.00	145.00		1.64			



Months	SynSub/Nonsynsub rate	SynSub Rate	NonSynSub Rate
Dec1,2019	0	0	0
Jan1-10,2020	1.88	1.07	0.57
Feb1-10,2020	2.95	2.21	0.54
March1-10,2020	2.26	1.65	1.07
April1-10,2020	2.42	2.55	0.92
May1-10,2020	2.70	2.13	0.46
June1-10,2020	2.47	0.89	0.69
July1-10,2020	1.91	0.45	1.49
August1-10,2020	1.75	0.62	0.89
Sept1-10,2020	1.64	0.96	1.01



Supplementary Table 2: Estimation of evolutionary rate and time of SARS-CoV2 in human from January, 2020-Sept 2020										
Nucleotide	1-29903									
Strain	NT.ACC	Isolation date								
SARS-CoV2	NC_045512	12/30/2019)								
SARS-CoV2 -CA	MW035476	09/02/2020								
							\T/U<->C=α	T/U<->A=ß	T/U<->G=Γ	
SARS-CoV2 -CA	MW035476	9/2/2020					A<->G=α	C<->G=ß	C<->A=Γ	
SARS-CoV2 -FL	MW040645	9/3/2020								

SARS-CoV2 -WI	MW035558	9/3/2020									
SARS-CoV2 -MO	MW035556	9/7/2020									
SARS-CoV2 -WI	MW035570.1	9/5/2020									
SARS-CoV2 -WA	MW077468	9/13/2020									
SARS-CoV2 -WA	MW077477	9/13/2020									
SARS-CoV2 -NM	MW070096.1	9/13/2020									
SARS-CoV2 -WI	MW035542	9/6/2020									
SARS-CoV2 -CA	MW035466.1	9/2/2020									
SARS-CoV2 -NM	MW070107.1	9/7/2020									
SARS-CoV2 -NM	MW070096	9/1/2020									
SARS-CoV2 -NM	MW075732	9/3/2020									
SARS-CoV2 -WI	MW031044	9/7/2020									
SARS-CoV2 - Australia	MW155352	9/10/2020									
SARS-CoV2 - Australia	MW155192	9/7/2020									
SARS-CoV2 -WI	MW030990	9/8/2020									
SARS-CoV2 -IL	MW031006	9/12/2020									
SARS-CoV2 -NJ	MW035532	9/13/2020									
SARS-CoV2 - Australia	MW157210	9/1/2020									
SARS-CoV2 - Australia	MW185627	9/2/2020									
SARS-CoV2 - Australia	MW155442	9/13/2020									
SARS-CoV2 - Australia	MW155194	9/8/2020									
SARS-CoV2 -WI	MW030992	9/8/2020									
SARS-CoV2 -NM	MW075751	9/7/2020									
SARS-CoV2 - Australia	MW155198	9/8/2020									
SARS-CoV2 - Australia	MW155185	9/8/2020									
SARS-CoV2 - Australia	MW157210	9/9/2020									
SARS-CoV2 - Australia	MW157208	9/10/2020									
SARS-CoV2 - Australia	MW157205	9/11/2020									
SARS-CoV2 -NM	MW075768	9/6/2020									
SARS-CoV2 -NM	MW075757.1	9/7/2020									
SARS-CoV2 -NM	MW075760	9/8/2020									
SARS-CoV2 -NM	MW075763	9/4/2020									

SARS-CoV2 -WI	MW031002	9/9/2020									
SARS-CoV2 - Australia	MW155334	9/10/2020									
SARS-CoV2 - Australia	MW155301	9/11/2020									
SARS-CoV2 gene	Codon	NT position	1st codon		2nd codon		3rd codon			Syn or nonsyn	
F924F,orf1ab	TTT>TTC	3036					T>C	α	1	S	
T2274I,orf1ab	ACT>ATT	7086			C>T	α1			1	NS	
P4715L,orf1ab	CCT>CTT	14408			C>T	α1			1	NS	
V6241F,orf1ab	GTT>TTT	18988	G>T	Γ1					1	NS	
S521A, S protein	GCA>TCA	23126	G>T	Γ1					1	NS	
F541F,S protein	TTT>TTC	23184					T>C	α	1	S	
G614D,S protein	GGT>GAT	23403			G>A	ß1			1	NS	
Y196S, M protein	TAT>TCT	27111			A>C	Γ1			1	NS	
D5109E,orf1ab	GAT>GAG	15590					T>G	Γ1	1	NS	
Y5276Y,orf1ab	TAC>TAT	16096					C>T	α	1	S	
C5272C,orf1ab	TGT>TGC	16079					T>C	α	1	S	
V5819V.orf1ab	GTT>GTG	17721					T>G	Γ	1	S	
F3628V,orf1ab	TTT>GTT	11147	T>G	Γ1					1	NS	
Y6317Y,orf1ab	TAT>TAC	19215					T>C	α	1	S	
T4616R,orf1ab	ACA>AGA	14110			C>G		ß1		1	NS	
I1018I, S protein	ATT>ATC	24616					T>C	α	1	S	
L106L, orf3a	CTC>CTT	25710					T>C	α	1	S	
G172V, orf3a	GGT>GTT	25907			G>T	Γ1			1	NS	
P67S, N protein	CCT>TCT	28472	C>T	α1					1	NS	
A90A, N protein	GCT>GCC	28543					T>C	α	1	S	
S152A, N protein	GCT>TCT	28727	G>T	Γ1					1	NS	
P199L, N protein	CCA>CTA	28869			C>T	α1			1	NS	
I265T, orf1ab	ATC>ACC	1059			T>C	α1			1	NS	
V665I, orf1ab	GTC>ATC	2258	G>A	α1					1	NS	
A1049V, orf1ab	GCT>GTT	3411			C>T				1	NS	
R1170C, orf1ab	CGC>TGC	3773	C>T	α2					1	NS	
R1464W, orf1ab	CGG>TGG	4655	C>T	α1					1	NS	
S2625S, orf1ab	TCC>TCT	8140					C>T	α	1	S	
Q2702H, orf1ab	CAG>CAT	8371					G>T	Γ1	1	NS	
Y3204H, orf1ab	TAT>CAT	9875	T>C	α1					1	NS	
T3255I, orf1ab	ACC>ATC	10029			C>T	α1			1	NS	
L3350F, orf1ab	CTT>TTT	10313	C>T	α1					1	NS	
L4643L, orf1ab	TTA>CTA	14191	T>C	α					1	S	
G5071G, orf1ab	GGA>GGT	15477					A>T	ß	1	S	
D5129D, orf1ab	GAC>GAT	15654					C>T	α	1	S	

L5737L, orf1ab	CTT>CTA	17476					A>T	ß	1	S	
A6044V, orf1ab	GCT>GTT	18395			C>T		ß1		1	NS	
N6054D, orf1ab	AAT>GAT	18427	A>G	α1					1	NS	
T6056T, orf1ab	ACA>ACG	18431					A>G	α	1	S	
L6652L, orf1ab	TTA>TTG	20220					A>G	α	1	S	
K6786N, orf1ab	AAA>AAT	20622					A>T	ß1	1	NS	
D6787F, orf1ab	GAT>TTT	20623	G>A	α1	A>T	ß1			2	NS	
V6876V, orf1ab	GTT>GTA	20892					T>A	ß	1	S	
N7096N, orf1ab	AAC>AAT	21552					C>T	α	1	S	
V3V, S protein	GTT>GTG	21571					T>G	Γ	1	S	
T723I, S protein	ACC>ATC	23730			C>T	α1			1	NS	
P897P, S protein	CCA>CCT	24253					A>T	ß	1	S	
V1061V, S protein	GTC>GTA	24744					C>A	Γ	1	S	
S183Y, N gene, orf9	TCT>TAT	28822			C>A	Γ1			1	NS	
K257K, N gene, orf9	AAG>AAA	29043					G>A	α	1	S	
T332I, N protein	ACA>ATA	29272			C>T	α1			1	NS	
G2667C, orf1ab	GGC>TGC	8264	G>T	α1					1	NS	
V2955V, orf1ab	GTG>GTT	9130					G>T	α	1	S	
L3337L, orf1ab	CTC>TTC	10277	C>T						1	S	
E5136D, orf1ab	GAG>GAT	15671					G>T	α1	1	NS	
L5762L, orf1ab	CTC>CTT	17550							1	S	
A522S, S protein	GCA>TCA	23126	G>T	α1					1	NS	
F543F, S protein	TTC>TTT	23191					C>T	α	1	S	
E780Q, S protein	GAA>CAA	23900	G>C	ß1					1	NS	
T1238T, S protein	ACC>ACT	25276					C>T	α	1	S	
A23V, orf3a	GCT>GTT	25460			C>T	α1			1	NS	
A110S, orf3a	GCT>TCT	25720	G>T	Γ1					1	NS	
T223I, orf3a	ACT>ATT	26060			C>T	α1			1	NS	
S98F, orf7a	TCT>TTT	27686			C>T	α1			1	NS	
D2043D, orf1ab	GAT>GAC	6394					T>C	α	1	S	
F2338F, orf1ab	TTC>TTT	7278					C>T	α	1	S	
Y3803Y, orf1ab	TAC>TAT	11674					C>T	α	1	S	
L4185L, orf1ab	TTA>TTG	12820					A>G	α	1	S	
I4205I, orf1ab	ATC>ATT	12880					C>T	α	1	S	
A4918L, orf1ab	GCA>TTA	15016	G>T		C>T	α1			2	NS	
Y5223Y, orf1ab	TAC>TAT	15933					C>T	α	1	S	
F5348F, orf1ab	TTC>TTT	16308					C>T	α	1	S	
T5477I, orf1ab	ACT>ATT	16694			C>T	α1			1	NS	
L5500F, orf1ab	CTT>TTT	16762	C>T	α1					1	NS	
E5568D, orf1ab	GAG>GAT	16968					G>T	Γ1	1	NS	
V6215F, orf1ab	GTT>TTT	18907	G>T	Γ1					1	NS	
N6525N, orf1ab	AAT>AAC	19842					T>C	α	1	S	

G6602V, orf1ab	GGT>GTT	20069			G>T	Γ1			1	NS	
A6802A, orf1ab	GCG>GCA	20670					G>A	α	1	S	
A6966V, orf1ab	GCT>GTT	21161			C>T	α1			1	NS	
L17F, orf5	CTT>TTT	26571	C>T	α1					1	NS	
V100V, orf8	GTG>GTT	28193					G>T	Γ	1	S	
D128D, orf9	GAC>GAT	28657					C>T	α	1	S	
R203K, orf9	AGG>AAA	28881			G>A	α1	G>A	α1	2	NS	
G204R, orf9	GGA>CGA	28883	G>C	ß1					1	NS	
T205I, orf9	ACT>ATT	28887			C>T	α1			1	NS	
H81Y, orf1ab	CAT>TAT	507	C>T	α1					1	NS	
I300F, orf1ab	ATT>TTT	1163	A>T	ß1					1	NS	
K1202N, orf1ab	AAG>AAT	3871					G>T	Γ1	1	NS	
V1222V, orf1ab	GTT>GTC	3931					T>C	α	1	S	
P1803S, orf1ab	CCT>TCT	5672	C>T	α1					1	NS	
K2029K, orf1ab	AAG>AAA	6352					G>A	α	1	S	
T2425T, orf1ab	ACT>ACC	7540					T>C	α	1	S	
C2445C, orf1ab	TGC>TGT	7600					C>T	α	1	S	
K2497N, orf1ab	AAG>AAT	7756					G>T	Γ1	1	NS	
D4166D, orf1ab	GAC>GAT	12763					C>T	α	1	S	
Q4635L, orf1ab	CTA>CAA	14169			T>A	ß1			1	NS	
C5332C, orf1ab	TGC>TGT	16260					C>T	α	1	S	
T5461T, orf1ab	ACG>ACT	16647					G>T	Γ	1	S	
A6035A, orf1ab	GCC>GCT	18365					C>T	α	1	S	
L6981L, orf1ab	CTC>CTT	21207					C>T	α	1	S	
C662C, S protein	TGT>TGC	23548					T>C	α	1	S	
L219L, orf1ab	CTG>CTT	922					G>T	Γ	1	S	
P804S, orf1ab	CCT>TCT	2686	C>T	α1					1	NS	
S1856L, orf1ab	TCA>TTA	5835			C>T	α1			1	NS	
L2620F, orf1ab	CTT>TTT	8123	C>T	α1					1	NS	
A3070V, orf1ab	GCT>GTT	9474			C>T	α1			1	NS	
K5110K, orf1ab	AAG>AAA	15594					G>A	α	1	S	
A5376V, orf1ab	GCT>GTT	16391			C>T	α1			1	NS	
T5690T, orf1ab	ACG>ACT	17334					G>T	Γ1	1	S	
D137N, S protein	GAT>AAT	21974	G>A	α1					1	NS	
G181R, S protein	GGA>AGA	22103	G>A	α1					1	NS	
S3S, orf4	TCA>TCC	26253					A>C	Γ	1	S	
L2527L, orf1ab	CTG>TTG	7846	C>T	α					1	S	
D6664NL, orf1ab	GAC>AAC	20254	G>A	Γ1					1	NS	
V83V, S protein	GTC>GTT	21811					C>T	α	1	S	
N121N, S protein	AAT>AAC	21925					T>C	α	1	S	
P681H, S protein	CCT>CAT	23604			T>A	ß1			1	NS	
G838G, S protein	GGT>GGC	24076					T>C	α	1	S	

V491V, orf1ab	GTG>GTT	1738					G>T	Γ	1	S	
L3238L, orf1ab	CTT>CTC	9977					T>C	α	1	S	
D1440D, orf1ab	GAT>GAC	4585					C>T	α	1	S	
D6097D, orf1ab	GAC>GAT	18555					C>T	α	1	S	
Q613Q, S protein	CAG>CAA	23401					G>A	α	1	S	
N824N, S protein	AAC>AAT	24034					C>T	α	1	S	
L331L, N protein	TTG>TTA	29266					G>A	α	1	S	
			α+ß+Γ	2		0		58		126	
			α1+ß1+Γ1	27		26		9			
			α+ß+Γ+α1+ß1+Γ1	29		26		67	122	366	
		Total	α	2		0		46			
			ß	0		0		4			
			Γ	0		0		8			
			α1	17		18		2			
			ß1	3		4		1			
			Γ1	5		4		6			
Total homologous codon=		9748									
Total NonSyn codon=		63									
Total Syn codon=		63									
Overall											
At codon position 1						**At codon position 2**		**At codon position 3**			
	P (probablity of transition)		0.0019			0.00185		0.0049241			
	Q (probablity of tranversion1)		0.0003			0.00041		0.0005129			
	R (probablity of transversion2)		0.0005			0.00041		0.0014362			
K=	= -(1/4)ln[(1- 2P - 2Q)(1-2P - 2R)(1 - 2Q - 2R)]					= -(1/4)ln[(1- 2P - 2Q)(1-2P - 2R)(1 - 2Q - 2R)]		= -(1/4)ln[(1- 2P - 2Q)(1-2P - 2R)(1 - 2Q - 2R)]			
	0.0028					0.00267		0.0069104			
knuc=	K/2T	/1st codon site/year				K/2T	/2nd codon site/ year	K/2T	/3rd codon site/per year		

	0.0018505	/1st codon site/year				0.00178	/2nd codon site/year	0.0046069	/3rd codon site/ year		
	1.78E-03					1.80E-03		4.60E-03			
FOR SYNONYMOUS											
	P (probablity of transition)		0.0002			0		0.0047189			
	Q(probablity of tranversion1)		0			0		0.0004103			
	R(probablity of transversion2)		0			0		0.0008207			
K=	= -(1/4)ln[(1- 2P - 2Q)(1-2P - 2R)(1 - 2Q - 2R)]										
K=	0.0002				K=	0.0000	K	0.0059794			
K=	2Tkn1										
knuc1=K/2T					knuc2		Knuc3				
	0.0001368					0		0.0039863			
FOR NONSYNONYMOUS											
	P (probablity of transition)		0.0017			0.00185		0.0002052			
	Q(probablity of tranversion1)		0.0003			0.00041		0.0001026			
	R(probablity of transversion2)		0.0005			0.00041		0.0006155			
K=	= -(1/4)ln[(1- 2P - 2Q)(1-2P - 2R)(1 - 2Q - 2R)]										
K=	0.0026				K=	0.0027	K	0.0009239			
K=	2Tkn1										
knuc1=	K/2T				knuc1		knuc3				
	0.0017131					0.00178		0.0006159			

Supplementary Table 3:Evolutionary rate and time estimation in SARS-CoV2 from RaTG13										
RaTG13	MN996532.2				Bat (host)	protein sequence		ORF1ab polyprotein including RDRP		
SARS-CoV2	NC_045512		YP_009724389		Human (host)	protein sequence		ORF1ab polyprotein including RDRP		
Nucleotide	266-21555									
Protein	1-7097aa		MESLVP--VLVNN							
			T/U<->C=α		**T/U<->A=ß**		**T/U<->G=Γ**			
			A<->G=α		**C<->G=ß**		**C<->A=Γ**			
Codon Substitution; S=Synonymous substitution; NS=Nonsynonymous substitution										
Amino Acid	Codon	NT position (bp)	1st codon		2nd codon		3rd codon		No. of NT changes	Syn or nonsyn
V20V	GTC>GTT	325					C>T	α	1	S
A38V	GCT>GTC	378			C>T	α1	T>C	α1	2	NS
G59G	GGC>GGT	441					C>T	α	1	S
S74S	TCT>TCG	487					T>G	Γ	1	S
G82G	GGC>GGT	511					C>T	α	1	S
L92L	CTT>CTC	539					T>C	α	1	S
N93E	AAT>GAA	540	A>G	α1			T>A	ß1	2	NS
Y96Y	TAC>TAT	555					C>T	α	1	S
L104L	CTC>CTT	577					C>T	α	1	S
Y110H	TAT>CAT	593	T>C	α1					1	NS
T114I	ACA>ATA	605			C>T	α1			1	NS
V117A	GTT>GCT	614			T>C	α1			1	NS
A138A	GCT>GCC	678					T>C	α	1	S
S142S	TCG>TCA	691					G>A	α	1	S
V169V	GTC>GTT	772					C>T	α	1	S
D172E	GAT>GAA	801					T>A	ß1	1	NS
N178N	AAT>AAC	798					T>C	α	1	S
A208A	GCT>GCA	882					T>A	ß	1	S
K235K	AAA>AAG	940					A>G	α	1	S
R236R	AGA>AGG	943					A>G	α	1	S
C230C	TGT>TGC	957					T>C	α	1	S
E237E	GAG>GAA	975					G>A	α	1	S
T265T	ACA>ACC	1060					C>A	Γ	1	S
F266F	TTT>TTC	1063					T>C	α	1	S
F275F	TTC>TTT	1089					C>T	α	1	S
L277L	CTA>TTA	1093	C>T	α					1	S
T280I	ACA>ATA	1104			C>T	α1			1	NS
A307A	GCT>GCG	1186					T>G	Γ	1	s
T333T	ACA>ACG	1264					A>G	α	1	S
C341C	TGT>TGC	1288					T>C	α	1	S
E347E	GAA>GAG	1306					A>G	α	1	NS

E352E	GAG>GAA	1321					G>A	α	1	S
L360L	CTA>TTA	1343	C>T	α					1	S
V366V	GTC>GTT	1363					C>T	α	1	S
H374H	CAT>CAC	1386					T>C	α	1	S
P376S	CCA>TCA	1390	C>T	α1					1	NS
A385A	GCT>GCC	1420					T>C	α	1	S
Y387Y	TAT>TAC	1426					T>C	α	1	S
P395T	CCC>ACC	1447	C>A	Γ1					1	NS
A405A	GCT>GCC	1480					T>C	α	1	S
S412S	TCC>TCT	1500					C>T	α	1	S
Y417H	TAC>CAT	1514	T>C	α1			C>T	α1	2	NS
N418N	AAT>AAC	1518					T>C	α	1	S
I424V	ATT>GTT	1535	A>G	α1					1	NS
A429A	GCC>GCT	1552					T>C	α	1	S
C433C	TGC>TGT	1563					C>T	α	1	S
N434N	AAT>AAC	1566					T>C	α	1	S
S443S	TCC>TCT	1595					C>T	α	1	S
D448D	GAT>GAC	1609					T>C	α	1	S
L454L	CTT>CTC	1626					T>C	α	1	S
S483S	TCT>TCC	1714					T>C	α	1	S
T498A	ACA>GCA	1756	A>G	α1					1	NS
C506C	TGC>TGT	1782					C>T	α	1	S
N508N	AAC>AAT	1788					C>T	α	1	S
K513K	AAG>AAA	1803					G>A	α	1	S
G514G	GGG>GGA	1806					G>A	α	1	S
K515K	AAG>AAA	1809					G>A	α	1	S
A516A	GCA>GCT	1812					A>T	ß	1	S
K517K	AAG>AAA	1815					G>A	α	1	S
A520A	GCT>GCC	1825					T>A	ß	1	S
N522N	AAC>AAT	1831					C>T	α	1	S
Q526Q	CAA>CAG	1843					A>G	α	1	S
L530L	TTG>CTG	1853	T>C	α					1	S
A537A	GCG>GCA	1875					G>A	α	1	S
V544V	GTT>GTA	1897					T>A	ß	1	S
S558S	TCC>TCT	1938					C>T	α	1	S
A561V	GCC>GTT	1947			C>T	α1	C>T	α1	2	NS
Y576Y	TAC>TAT	1992					C>T	α	1	S
V591A	GTT>GCT	2036			T>C	α1			1	NS
L595L	CTG>CTA	2049					G>A	α	1	S
A599A	GCT>GCC	2063					T>C	α	1	S
S610S	TCA>TCG	2095					A>G	α	1	S
T614T	ACA>ACT	2107					A>T	ß	1	S

N615N	AAT>AAC	2110					T>C	α	1	S
E622E	GAG>GAA	2131					G>A	α	1	S
P626P	CCT>CCC	2143					T>C	α	1	S
V627V	GTT>GTC	2146					T>C	α	1	S
L631L	CTC>CTT	2158					C>T	α	1	S
K634K	AAA>AAG	2167					A>G	α	1	S
F635F	TTT>TTC	2170					T>C	α	1	S
D644D	GAT>GAC	2197					T>C	α	1	S
T654T	ACT>ACC	2226					T>C	α	1	S
K672K	AAA>AAG	2280					A>G	α	1	S
N683N	AAC>AAT	2314					C>T	α	1	S
L688L	TTA>TTG	2329					A>G	α	1	S
S692S	TCC>TCT	2341					C>T	α	1	S
I694I	ATC>ATT	2346					C>T	α	1	S
T710T	ACA>ACG	3397					A>G	α	1	S
Y717Y	TAT>TAC	2415					T>C	α	1	S
P723S	CCC>TCC	2432	C>T	α1					1	NS
K724R	AAA>AGA	2436			A>G	α1			1	NS
L729L	TTA>CTA	2450	T>C	α					1	S
L732L	CTG>CTA	2464					G>A	α	1	S
A735A	GCT>GCC	2470					T>C	α	1	S
I740I	ATT>ATC	2485					T>C	α	1	S
P748P	CCT>CCC	2506					T>C	α	1	S
A776A	GCC>GCT	2592					C>T	α	1	S
C784C	TGT>TGC	2617					T>C	α	1	S
L788L	CTC>CTT	2629					C>T	α	1	S
I793I	ATT>ATC	2643					T>C	α	1	S
L815L	CTT>CTC	2709					T>C	α	1	S
K822K	AAA>AAG	2732					A>G	α	1	S
I831I	ATT>ATA	2452					T>A	ß	1	S
Y835Y	TAT>TAC	2772					T>C	α	1	S
D846D	GAT>GAC	2803					T>C	α	1	S
K851K	AAG>AAA	2820					G>A	α	1	S
N854N	AAC>AAT	2829					C>T	α	1	S
T859A	ACC>GCC	2839	A>G	α1					1	NS
F871F	TTT>TTC	2878					T>C	α	1	S
A872A	GCT>GCC	2881					T>C	α	1	S
V874V	GTA>GTG	2890					A>G	α	1	S
L883L	TTA>TTG	2915					A>G	α	1	S
I891T	ATA>ACA	2939			T>C1	α1			1	NS
G901S	GGT>AGT	2966	G>A	α1					1	NS
E913E	GAA>GAG	3003					A>G	α	1	S

Y921Y	TAC>TAT	3028					C>T	α	1	S
F924F	TTT>TTC	3037					T>C	α	1	S
Y925Y	TAT>TAC	3040					T>C	α	1	S
E931E	GAG>GAA	3058					G>A	α	1	S
D940E	GAT>GAG	3085					T>G	Γ1	1	NS
E942E	GAA>GAG	3091					A>G	α	1	S
P944S	CCA>TCA	3095	C>T	α1					1	NS
S959P	TCT>CCT	3139	T>C	α1					1	NS
A967V	GCT>GTC	3165			C>T	α1	T>C	α1	2	NS
T968A	ACT>GCT	3166	A>G	α1					1	NS
P969L	CCT>CTT	3171			C>T	α1			1	NS
L976Q	CTA>CAA	3188			T>A	ß1			1	NS
V988V	GTC>GTT	3229					C>T	α	1	S
V989G	GTT>GGT	3231			T>G	Γ1			1	NS
E991Q	GAA>CAA	3236	G>C	ß1					1	NS
D992D	GAT>GAC	3241					T>C	α	1	S
D993G	GAC>GGC	3243			A>G	α1			1	NS
E995E	GAA>GAG	3250					A>G	α	1	S
V996D	GTT>GAC	3253			T>A	ß1	T>C	α1	2	NS
I1001T	ATC>ACT	3267			T>C	α1	C>T	α1	2	NS
T1002I	ACT>ATT	3270			C>T	α1			1	NS
S1004T	TCA>ACA	3275	T>A	ß1					1	NS
A1006V	GCT>GTT	3282			C>T	α1			1	NS
Q1011Q	CAG>CAA	3298					G>A	α	1	S
P1016L	CCT>CTT	3306			C>T	α1			1	NS
1023I*	0> ATT (INS)	3331	0>A*		0>T*		0>T*		0	
L1027F	CTT>TTT	3346	C>T	α1					1	NS
I1056V	ATA>GTG	3430	A>G	α1			A>G	α1	2	NS
V1058V	GTC>GTT	3438					C>T	α	1	S
N1059N	AAC>AAT	3441					C>T	α	1	S
A1072A	GCG>GCA	3485					G>A	α	1	S
A1074A	GCT>GCC	3491					T>C	α	1	S
N1080N	AAT>AAC	3505					T>C	α	1	S
S1087S	TCA>TCT	3526					A>T	ß	1	S
H1089D	CAT>GAT	3530	C>G	ß1					1	NS
A1091A	GCC>GCT	3541					C>T	α	1	S
T1093T	ACC>ACT	3544					C>T	α	1	S
G1102S	GGC>AGT	3569	G>A	α1			C>T	α1	2	NS
L1105L	TTG>TTA	3580					G>A	α	1	S
S1106S	AGT>AGC	3583					T>C	α	1	S
N1113H	AAC>CAC	3601	A>C	Γ1					1	NS
V1118V	GTT>GTC	3618					T>C	α	1	S

G1119G	GGT>GGC	3621					T>C	α	1	S
R1124K	AGA>AAA	3636			G>A	α1			1	NS
D1127D	GAT>GAC	3646					T>C	α	1	S
D1142E	GAT>GAA	3652					T>A	ß1	1	NS
L1145L	CTC>CTT	3700					C>T	α	1	S
L1148L	CTA>TTA	3707	C>T	α					1	S
A1150A	GCC>GCT	3718					C>T	α	1	S
V1159I	GTA>ATA	3741	G>A	α1					1	NS
D1167D	GAC>GAT	3766					C>T	α	1	S
L1175L	CTA>TTA	3787	C>T	α					1	S
L1191L	TTA>TTG	3838					A>G	α	1	S
T1203I	ACT>ATC	3873			C>T	α1	T>C	α1	2	NS
E1205E	GAA>GAG	3880					A>G	α	1	S
S1213P	TCA>CCA	3901	T>C	α1					1	NS
S1214F	TCT>TTT	3905			C>T	α1			1	NS
I1215I	ATT>ATA	3909					T>A	ß	1	S
L1220P	CTT>CCT	3924			T>C	α1			1	NS
E1223E	GAG>GAA	3934					G>A	α	1	S
Q1226K	CAA>AAA	3941	C>A	ß1					1	NS
V1228D	GTT>GAT	3948			T>A	ß1			1	NS
F1248F	TTT>TTC	4009					T>C	α	1	S
N1252N	AAT>AAC	4021					T>C	α	1	S
L1254L	CTC>CTT	4029					C>T	α	1	S
N1272S	AAT>AGT	4078			A>G	α1			1	NS
P1284P	CCC>CCA	4127					C>A	Γ	1	S
V1299V	GTA>GTG	4162					A>G	α	1	S
G1308G	GGT>GGC	4189					T>C	α	1	S
T1314A	ACG>GCG	4204	A>G	α1					1	NS
A1316A	GCC>GCT	4212					C>T	α	1	S
D1303D	GAT>GAC	4234					T>C	α	1	S
N1304N	AAT>AAC	4237					T>C	α	1	S
P1310P	CCA>CCG	4255					A>G	α	1	S
G1331G	GGC>GGT	4259					C>T	α	1	S
Y1336Y	TAT>TAC	4275					T>C	α	1	S
R1343K	AGG>AAG	4293			G>A	α1			1	NS
E1383E	GAG>GAA	4414					G>A	α	1	S
T1385T	ACT>ACA	4420					T>A	ß	1	S
R1386R	CGT>CGC	4423					T>C	α	1	S
M1393V	ATG>GTG	4440	A>G	α1					1	NS
A1397A	GCT>GCC	4455					T>C	α	1	S
Y1406Y	TAC>TAT	4483					C>T	α	1	S
A1420A	GCC>GCT	4525					C>T	α1	1	NS

Y1425Y	TAT>TAC	4540					T>C	α	1	S
A1432A	GCA>GCG	4560					A>G	α	1	S
N1435N	AAT>AAC	4572					T>C	α	1	S
L1438L	CTC>CTT	4578					C>T	α	1	S
E1443E	GAG>GAA	4594					G>A	α	1	S
G1451G	GGT>GGC	4632					T>C	α	1	S
Y1486Y	TAC>TAT	4732					C>T	α	1	S
G1508G	GGC>GGT	4792					C>T	α	1	S
S1509S	TCT>TCC	4795					T>C	α1	1	S
Y1510Y	TAC>TAT	4798					C>T	α1	1	S
Q1519Q	CAG>CAA	4821					G>A	α	1	S
V1535V	GTG>GTA	4869					G>A	α	1	S
N 1540N	AAC>AAT	4885					C>T	α	1	S
I1542T	ATT>ACC	4890			T>C	α1	T>C	α1	2	NS
F1544F	TTT>TTC	4897					T>C	α	1	S
I1558T	ATA>ACA	4937			T>C	α1			1	NS
K1563R	AAA>AGA	4954			A>G	α1			1	NS
V1565V	GTT>GTG	4960					T>G	Γ	1	S
L1579L	CTT>CTC	5011					T>C	α	1	S
G1595G	GGC>GGT	5044					C>T	α	1	S
T1617T	ACG>ACA	5135					G>A	α	1	S
L1621L	TTG>TTA	5128					G>A	α	1	S
L1627L	TTA>CTA	5144	T>C	α					1	S
A1629V	GCT>GTT	5151			C>T	α1			1	NS
V1661V	GTC>GTT	5248					C>T	α	1	S
L1680l	TTA>TTG	5305					A>G	α	1	S
L1683L	CTA>CTC	5314					A>C	Γ	1	S
Q1684Q	CAG>CAA	5317					G>A	α	1	S
K1689K	AAA>AAG	5314					A>G	α	1	S
N1691N	AAC>AAT	5331					C>T	α	1	S
L1695L	TTA>CTA	5347	T>C	α					1	S
A1698A	GCC>GCT	5358					C>T	α	1	S
Y1699Y	TAC>TAT	5361					C>T	α	1	S
R1701R	AGG>AGA	5368					G>A	α	1	S
R1703R	AGA>AGG	5374					A>G	α	1	S
N1709N	AAT>AAC	5291					T>C	α	1	S
N1719N	AAT>AAC	5419					T>C	α	1	S
N1733S	AAT>AGT	5448			A>G	α1			1	NS
N1749N	AAT>AAC	6152					T>C	α	1	S
T1760T	ACT>ACA	5512					T>A	ß	1	S
T1761T	ACT>ACC	5548					T>C	α	1	S
L1779F	CTT>TTT	5599	C>T	α1					1	NS

K1781K	AAG>AAA	5608					G>A	α	1	S
T1788T	ACT>ACG	5629					T>G	α	1	S
Q1795K	CAA>AAA	5647			C>A	Γ1			1	NS
Q1799Q	CAA>CAG	5665					A>G	α	1	S
S1808S	TCG>TCA	5665					G>A	α	1	S
I1822T	ATT>ACT	5731			T>C	α1			1	NS
S1857S	TCT>TCA	5835					T>A	ß	1	S
E1858E	GAG>GAA	5838					G>A	α	1	S
I1866V	ATT>GTT	5861	A>G	α1					1	NS
N1871N	AAT>AAC	5878					T>C	α	1	S
S1872S	AGC>AGT	5881					C>T	α	1	S
Y1873Y	TAT>TAC	5884					T>C	α	1	S
I1874T	ATA>ACA	5886			T>C	α1			1	NS
I1881T	ATT>ACT	5907			T>C	α1			1	NS
D1893D	GAT>GAC	5943					T>C	α	1	S
V1915V	GTG>GTA	6009					G>A	α	1	S
P1919P	CCT>CCA	6021					T>A	α	1	S
P1921P	CCG>CCA	6028					G>A	α	1	S
N1922N	AAT>AAC	6031					T>C	α	1	S
F1928F	TTC>TTT	6049					C>T	α	1	S
N1932N	AAC>AAT	6066					C>T	α	1	S
L1937F	CTT>TTT	6073	C>T	α1					1	NS
Y1947Y	TAC>TAT	6106					C>T	α	1	S
F1954L	TTT>CTT	6127	T>C	α1					1	NS
P1959P	TTC>TTT	6141					C>T	α	1	S
P1977P	CCT>CCC	6195					T>C	α	1	S
V1990V	GTC>GTT	6237					C>T	α	1	S
N1999N	AAC>AAT	6259					C>T	α	1	S
A2000A	GCT>GCC	6268					T>C	α	1	S
N2006N	AAC>AAT	6283					C>T	α	1	S
T2007T	ACT>ACC	6286					T>C	α	1	S
R2011R	CGC>CGT	6298					C>T	α	1	S
N2015S	AAC>AGC	6342			A>G	α1			1	NS
N2023N	AAC>AAT	6334					C>T	α	1	S
S2024S	TCT>TCG	6337					T>G	Γ	1	S
K2029K	AAA>AAG	6352					A>G	α	1	S
T6023A	ACG>GCG	6362	A>G	α1					1	NS
I2061V	ATT>GTT	6346	A>G	α1					1	NS
D2082N	GAT>AAT	6509	G>A	α1					1	NS
G2083S	GGT>AAT	6512	G>A	α1	G>A	α1			2	NS
L2084L	TTG>TTA	6517					G>A	α1	1	NS
E2088E	GAG>GAA	6529					G>A	α	1	S

G2091G	GGT>GGC	6535					T>C	α	1	S
L2095L	CTA>TTA	6613	C>T	α					1	S
V2123A	GTT>GCT	6633			T>C	α1			1	NS
V2133V	GTT>GTC	6663					C>T	α	1	S
P2134P	CCG>CCT	6666					G>T	Γ	1	S
N2147N	AAT>AAC	6706					T>C	α	1	S
S2151S	AGC>AGT	6718					C>T	α	1	S
N2155N	AAT>AAC	6718					T>C	α	1	S
V2157V	GTC>GTT	6735					T>C	α	1	S
V2206V	GTT>GTC	6883					T>C	α	1	S
C2210C	TGC>TGT	6895					C>T	α	1	S
Y2217Y	TAT>TAC	6915					T>C	α	1	S
L2218L	CTG>TTG	6916	C>T				T>C	α	1	S
F2224S	TTT>TCT	6936			T>C	α1			1	NS
L2226L	TTG>CTG	6941	T>C	α					1	S
I2227I	ATT>ATA	6946					T>A	ß1	1	S
C2239C	TGC>TGT	6981					C>T	α1	1	S
L2240L	TTA>CTA	6982	T>C	α					1	S
L2250L	CTA>TTA	7013					C>T	α	1	S
G2265G	GGC>GGT	7060					C>T	α	1	S
T2278I	ACA>ATT	7097			C>T	α1	A>T	ß1	2	NS
I2280T	ATC>ACC	7104			T>C	α1			1	NS
D2298D	GAT>GAC	7161					T>C	α	1	S
F2334F	TTT>TTC	7266					T>C	α	1	S
F2337F	TTC>TTT	7275					C>T	α	1	S
I2346I	ATT>ATC	6303					T>C	α1	1	NS
P2375P	CCA>CCG	7392					A>G	α	1	S
R2382R	AGG>AGA	7411					G>A	α	1	S
I2385I	ATT>ATC	7420					T>C	α	1	S
D2402D	GAT>GAC	7471					T>C	α	1	S
T2405N	ACT>AAT	7479			C>A	Γ1			1	NS
V2421V	GTT>GTC	7528					T>C	α	1	S
H2448H	CAT>CAC	7609					T>C	α	1	S
D2456D	GAC>GAT	7632					C>T	α	1	S
S2488S	TCC>TCT	7729					C>T	α	1	S
I2490I	ATT>ATC	7735					T>C	α	1	S
D2491V	GAT>GTC	7740					T>C	α1	1	NS
N2526N	AAC>AAT	7843					C>T	α	1	S
L2527L	TTG>CTG	7844	T>C	α					1	S
S2530N	AGT>AAT	7854			G>A	α1			1	NS
Y2559Y	TAT>TAC	7945					T>C	α	1	S
T2587A	ACA>GCG	8015	A>G	α1			A>G	α1	2	NS

V2604V	GTT>GTA	8015					T>A	ß	1	S
L2609L	CTA>CTC	8272					A>C	α	1	S
S2631S	TCC>TCT	7158					C>T	α	1	S
T2632T	ACC>ACT	7161					C>T	α	1	S
G2667G	GGT>GGC	8266					T>C	α	1	S
P2685P	CCT>CCC	8320					T>C	α	1	S
N2708N	AAT>AAC	8389					T>C	α	1	S
A2710A	GCG>GCT	8395					G>T	Γ	1	S
F2718F	TTT>TTC	8418					T>C	α	1	S
L2721L	CTG>TTG	8426	C>T	α					1	S
A2733A	GCC>GCT	8463					C>T	α	1	S
N2736N	AAC>AAT	8472					C>T	α	1	S
L2737L	TTG>TTA	8478					G>A	α	1	S
K2757K	AAA>AAG	8535					A>G	α	1	S
L2778L	CTC>CTT	8598					C>T	α	1	S
I2786I	ATC>ATT	8623					C>T	α	1	S
I2806I	ATT>ATC	8683					T>C	α	1	S
G2808G	GGG>GGA	8689					G>A	α	1	S
I2812I	ATT>ATC	8701					T>C	α	1	S
S2839S	AGT>AGC	8781					T>C	α	1	S
C2851C	TGT>TGC	8818					T>C	α	1	S
L2869L	CTT>TTG	8870	C>T	α			T>G	α	2	S
C2896C	TGC>TGT	8952					C>T	α	1	S
T2910T	ACT>ACA	8955					T>A	ß	1	S
T2936T	ACT>ACC	8995					T>C	α	1	S
A2944A	GCA>GCT	9072					A>T	ß	1	S
L2948L	CTA>TTA	9097	C>T	α					1	S
R2949R	CGT>CGC	9102					T>C	α	1	S
D2951D	GAT>GAC	9107					T>C	α	1	S
V2954V	GTA>GTG	9119					A>G	α	1	S
Y2983Y	TAT>TAC	9214					T>C	α	1	S
R2985R	AGA>AGG	9220					A>G	α	1	S
G2987G	GGT>GGC	9226					T>C	α	1	S
N3007N	AAT>AAC	9286					T>C	α	1	S
F3017F	TTT>TTC	9318					T>C	α	1	S
F3031F	TTC>TTT	9357					C>T	α	1	S
I3037I	ATC>ATT	9378					T>C	α	1	S
i3047i	ATT>ATA	9406					T>A	ß	1	S
A3049A	GCC>GCT	9392					C>T	α	1	S
G3050G	GGC>GGT	9395					C>T	α	1	S
V3053V	GTT>GTA	9404					T>A	ß	1	S
Y3063Y	TAC>TAT	9455					C>T	α	1	S

S3075S	AGC>AGT	9489					C>T	α	1	S
F3100F	TTT>TTC	9565					T>C	α	1	S
G3103G	GGC>GGT	9574					C>T	α	1	S
F3122F	TTC>TTT	9632					C>T	α	1	S
I3126i	ATC>ATT	9643					C>T	α	1	S
V3143A	GTT>GCT	9693			T>C	α1			1	NS
V3145I	GTC>ATC	9698	G>A	α1					1	NS
F3153F	TTT>TTC	9723					T>C	α	1	S
Y3154Y	TAC>TAT	9741					T>C	α	1	S
N3159N	AAC>AAT	9741					C>T	α	1	S
K3162K	AAA>AAG	9751					A>G	α	1	S
R3163R	AGG>AGA	9754					G>A	α	1	S
L3180L	TTA>CTG	9805	T>C	α			A>G	α	2	S
T3182T	ACT>ACC	9811					T>C	α	1	S
F3183F	TTC>TTT	9814					C>T	α	1	S
L3184L	TTA>TTG	9817					A>G	α	1	S
K3187K	AAG>AAA	9825					A>G	α	1	S
K3192K	AAA>AAG	9840					A>G	α	1	S
V3197V	GTA>GTG	9855					A>G	α	1	S
L3198L	CTT>CTA	9858					T>A	ß	1	S
L3199L	CTA>TTA	9859	C>T	α					1	S
Y3207Y	TAT>TAC	9886					T>C	α	1	S
A3209A	GCC>GCT	9892					C>T	α	1	S
A3220A	GCC>GCA	9925					C>A	Γ	1	S
L3234L	CTT>CTC	9966					T>C	α	1	S
K3236K	AAA>AAG	9972					A>G	α	1	S
D3240D	GAT>GAC	9985					T>C	α	1	S
N3243N	AAT>AAC	9995					T>C	α	1	S
Y3250Y	TAC>TAT	10015					T>C	α	1	S
S3259S	TCG>TCA	10041					T>C	α	1	S
C3279C	TGC>TGT	10101					T>C	α	1	S
V3283V	GTT>GTA	10114					T>A	ß	1	S
T3288T	ACC>ACT	10129					C>T	α	1	S
N3281N	AAT>AAC	10138					C>T	α	1	S
D3287D	GAT>GAC	10155					T>C	α	1	S
V3289V	GTC>GTT	10161					T>C	α	1	S
N3314N	AAT>AAC	10206					T>C	α	1	S
D3319D	GAC>GAT	10221					C>T	α	1	S
L3321L	CTT>CTC	10228					T>C	α	1	S
I3322I	ATC>ATT	10231					C>T	α	1	S
R3322R	CGC>CGT	10234					T>C	α	1	S
L3338L	CTT>CTC	10272					T>C	α	1	S

R3339R	AGA>AGG	10281					A>G	α	1	S
V3349V	GTG>GTA	10312					G>A	α	1	S
V3353V	GTA>GTT	10322					A>T	ß	1	S
T3356T	ACT>ACA	10332					T>A	α	1	S
T3359P	ACT>CCT	10339	A>C	Г1					1	NS
T3361T	ACG>ACA	10348					G>A	α	1	S
P3371P	CCT>CCA	10373					T>A	α	1	S
Q3390Q	CAG>CAA	10445					G>A	α	1	S
R3394R	AGA>AGG	10417					A>G	α	1	S
P3395P	CCT>CCC	10420					T>C	α	1	S
N3396N	AAC>AAT	10423					C>T	α	1	S
F3397F	TTT>TTC	10426					T>C	α	1	S
T3398T	ACA>ACT	10429					A>T	ß	1	S
S3407S	TCT>TCA	10486					T>A	ß	1	S
T3443N	ACC>AAC	10593			C>A	Г1			1	NS
F3444F	TTC>TTT	10597					C>T	α	1	S
R3451R	AGA>AGG	10618					A>G	α	1	S
I3463I	ATC>ATT	10654					C>T	α	1	S
V3465V	GTC>GTT	10660					C>T	α	1	S
D3479D	GAT>GAC	10740					T>C	α	1	S
L3483L	CTC>CTT	10709					C>T	α	1	S
Q3507Q	CAG>CAA	10786					G>A	α	1	S
T3520T	ACT>ACC	10825					T>C	α	1	S
L3530L	CTA>TTA	10856	C>T	α					1	S
V3572V	GTA>GTG	10980					A>G	α	1	S
G3578G	GGC>GGT	10998					C>T	α	1	S
H3580H	CAT>CAC	11008					T>C	α	1	S
L3584L	CTT>CTC	11020					T>C	α	1	S
L3612L	TTG>TTA	11100					G>A	α	1	S
V3629V	GTT>GTC	11152					T>C	α	1	S
F3635F	TTC>TTT	11169					C>T	α	1	S
F3639F	TTC>TTT	11181					C>T	α	1	S
A3645A	GCT>GCC	11199					T>C	α	1	S
Y3649Y	TAT>TAC	11209					T>C	α	1	S
R3662R	CGC>CGT	11251					C>T	α	1	S
F3677F	TTC>TTT	11315					C>T	α	1	S
A3686A	GCC>GCT	11329					C>T	α	1	S
R3707R	AGG>AGA	11386					G>A	α	1	S
V3714V	GTC>GTT	11418					C>T	α	1	S
A3729A	GCT>GCC	11455					T>C	α	1	S
S3745S	TCG>TCA	11500					G>A	α	1	S
F3753F	TTC>TTT	11523					C>T	α	1	S

F3789F	TTC>TTT	11632					C>T	α	1	S
G3795G	GGT>GGC	11649					T>C	α	1	S
R3802R	CGT>CGC	11671					T>C	α	1	S
Y3813Y	TAT>TAC	11703					T>C	α	1	S
L3814L	TTG>TTA	19706					G>A	α	1	S
S3816S	TCC>TCT	11712					C>T	α	1	S
Q3818Q	CAA>CAG	11718					A>G	α	1	S
R3821R	AGG>AGA	11728					G>A	α	1	S
L3828L	TTG>CTA	11747	T>C	α			G>A	α	2	S
P3831P	CCA>CCC	11758					A>C	α	1	S
F3838F	TTT>TTC	11778					T>C	α	1	S
I3842I	ATC>ATT	11832					C>T	α	1	S
V3848V	GGA>GGT	11808					A>T	ß	1	S
G3848G	GGT>GGC	11811					T>C	α	1	S
I3853I	ATT>ATC	11823					T>C	α	1	S
T3868T	ACG>ACA	11869					G>A	α	1	S
L3879L	CTT>CTC	11901					T>C	α	1	S
V3892V	GTT>GTC	11940					T>C	α	1	S
I3897I	ATA>ATT	11958					A>T	ß	1	S
V3917V	GTC>GTT	12015					C>T	α	1	S
L3935L	CTT>CTG	12069					T>G	Γ	1	S
S3949S	TCT>TCC	12115					T>C	α	1	S
T3959T	ACT>ACC	12141					T>C	α	1	S
D3972D	GAC>GAT	12180					C>T	α	1	S
V3976V	GTC>GTT	12192					C>T	α	1	S
R3999R	CGC>CGT	12259					C>T	α	1	S
E4019E	GAA>GAG	12321					A>G	α	1	S
P4058P	CCA>CCG	12438					A>G	α	1	S
L4070L	CTA>TTA	12473	C>T	α					1	S
S4093S	TCG>TCA	12543					G>A	α	1	S
V4102V	GTC>GTA	12571					C>A	ß	1	S
S4115S	AGC>AGT	12609					C>T	α	1	S
L4184L	TTG>CTG	12815	T>C	α					1	S
N4235N	AAT>AAC	12971					T>C	α	1	S
A4247A	GCT>GCC	13009					T>C	α	1	S
L4251L	TTA>CTA	13018	T>C	α					1	S
L4267L	CTA>TTA	13064	C>T	α					1	S
A4276A	GCC>GCT	13092					C>T	α	1	S
A4279A	GCG>GCT	13101					G>T	Γ	1	S
D4282D	GAC>GAT	13111					C>T	α	1	S
P4311P	CCA>CCG	13200					A>G	α	1	S
C4332C	TGT>TGC	13255					T>C	α	1	S

T4354T	ACG>ACA	13326					G>A	α	1	S
A4422V	GCA>GTA	13520			C>T	ß1			1	NS
V4422A	GTG>GTA	13533					G>A	α1	1	NS
L4441L	TTA>CTA	13585	T>C	α					1	S
K4450K	AAA>AAG	13617					A>G	α	1	S
E4448E	GAG>GAA	13623					G>A	α	1	S
H4490H	CAC>CAT	13732					C>T	α	1	S
D4526D	GAC>GAT	13845					C>T	α	1	S
G4528G	GGC>GGT	13851					C>T	α	1	S
D4532D	GAT>GAC	13855					T>C	α	1	S
R4535K	AGA>AAA	13863			G>A	α1			1	NS
D4545D	GAC>GAT	13899					C>T	α	1	S
F4551F	AAA>AAG	13919					A>G	α	1	S
D4590N	GAT>AAT	14037	G>A	α1					1	NS
V4594V	GTA>GTT	14051					A>T	α	1	S
L4597L	CTA>CTG	14054					A>G	α	1	S
D4613D	GAC>GAT	14102					C>T	α	1	S
T4616T	ACT>ACG	14117					T>G	Γ	1	S
I4625V	ATT>GTT	14138	A>G	α1					1	NS
L4637L	CTA>TTA	14167	C>T	α					1	S
T4638T	ACT>ACC	14177					T>C	α	1	S
L4639L	TTA>TTG	14180					A>G	α	1	S
T4640T	ACT>ACC	14183					T>C	α	1	S
A4642A	GCA>GCT	14191					A>T	ß	1	S
P4656P	CCC>CCT	14231					C>T	α	1	S
F4676F	TTT>TTC	14265					T>C	α	1	S
N4706N	AAC>AAT	14382					C>T	α	1	S
P4720P	CCC>CCT	13424					C>T	α	1	S
R4741R	AGG>AGA	14487					G>A	α	1	S
L4779L	TTA>CTA	14598	T>C	α					1	S
F4787F	TTC>TTT	14628					C>T	α	1	S
A4841A	GCC>GCT	14787					C>T	α	1	S
L4861L	CTC>CTA	14847					C>A	Γ	1	S
L4906L	CTC>CTT	14982					C>T	α	1	S
Y4908Y	TAC>TAT	14988					C>T	α	1	S
Y4913Y	TAC>TAT	15003					C>T	α	1	S
I4981I	ATC>ATT	15207					C>T	α	1	S
F4985F	TTT>TTC	15222					C>T	α	1	S
H4991H	CAT>CAC	15237					T>C	α	1	S
Y4998Y	TAC>TAT	15258					C>T	α	1	S
H5005H	CAT>CAC	15279					C>T	α	1	S
Y5011Y	TAC>TAT	15297					C>T	α	1	S

C5037C	TGC>TGT	15374					C>T	α	1	S
V5085V	GTC>GTT	15519					C>T	α	1	S
V5092V	GTT>GTC	15540					T>C	α	1	S
H5111Y	CAC>TAT	15595	C>T	α1			C>T	α1	2	NS
L5150L	CTT>CTC	15714					T>C	α	1	S
D5152D	GAT>GAC	15720					T>C	α	1	S
F5174F	AAG>AAA	15788					G>A	α	1	S
Y5179Y	TAC>TAT	15800					C>T	α	1	S
N5182N	AAC>AAT	15814					C>T	α	1	S
L5221L	CTC>CTT	15927					C>T	α	1	S
P5223P	CCG>CCA	15936					G>A	α	1	S
R5227R	CGA>AGA	15946	C>A	Γ					1	S
L5230L	TTA>CTA	15952	T>C	α					1	S
A5232A	GCT>GCC	15960					T>C	α	1	S
L5246L	CTG>CTT	16002					G>T	Γ	1	S
F5251F	TTT>TTC	16021					T>C	α	1	S
H5274H	CAC>CAT	16085					T>C	α	1	S
L5283L	TTA>CTA	16111	T>C	α					1	S
H5289H	CAT>CAC	15134					T>C	α	1	S
C5350C	TGC>TGT	16314					C>T	α	1	S
H5357H	CAC>CAT	16334					C>T	α	1	S
P5406P	CCG>CCA	16482					G>A	α	1	S
K5418K	AAG>AAA	16517					G>A	α	1	S
N5431N	AAC>AAT	16557					C>T	α	1	S
D5437D	GAT>GAC	16596					T>C	α	1	S
F5457F	TTC>TTT	16634					C>T	α	1	S
A5476A	GCC>GCT	16689					C>T	α	1	S
L5487L	TTG>TTA	16725					G>A	α	1	S
V5533V	GTC>GTT	16865					C>T	α	1	S
L5543L	CTA>TTA	16891	C>T	α					1	S
D5547D	GAC>GAT	16885					C>T	α	1	S
I5582I	ATC>ATT	17010					C>T	α	1	S
Q5605Q	CAA>CAG	17079					A>G	α	1	S
L5618L	TTA>CTA	17118	T>C	α					1	S
Y5630Y	TAC>TAT	17154					C>T	α	1	S
Y5679Y	TAC>TAT	17300					C>T	α	1	S
T5683T	ACT>ACA	17337					T>A	ß	1	S
A5713A	GCT>GCC	17372					T>C	α	1	S
P5730P	CCA>CCT	17404					A>T	ß	1	S
L5751L	CTC>CTT	17519					C>T	α	1	S
L5762L	CTT>CTC	17550					T>C	α	1	S
L5785L	CTG>CTT	17619					G>T	Γ	1	S

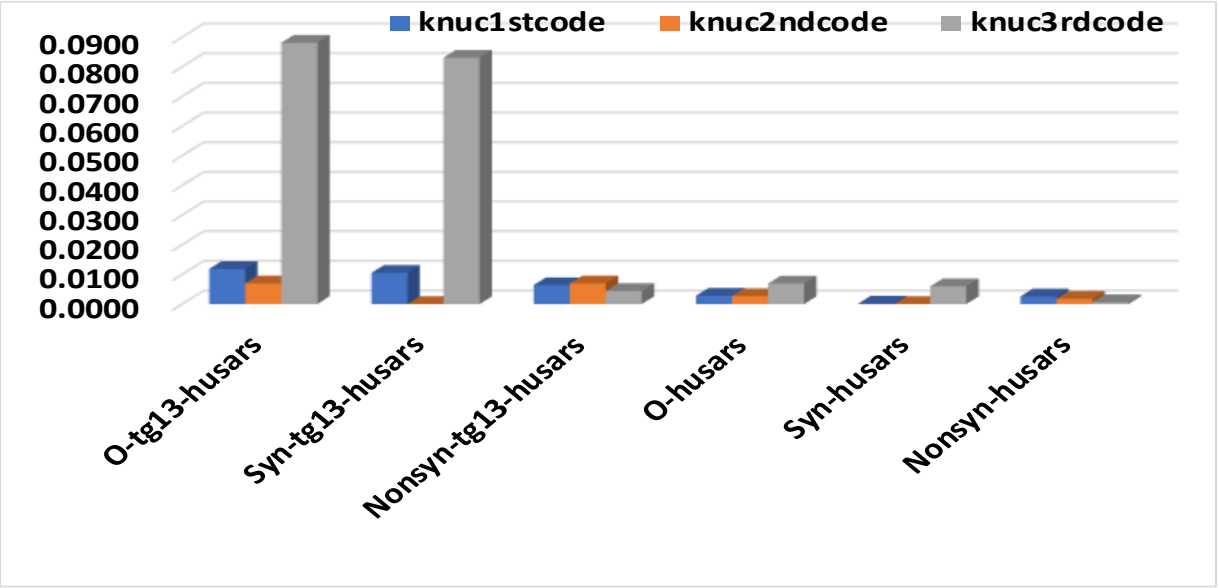
I5805I	ATT>ATC	17716					T>C	α	1	S
T5806T	ACT>ACG	17719					T>G	Γ	1	S
F5829F	TTC>TTT	17733					C>T	α	1	S
N5839N	AAT>AAC	17745					T>C	α	1	S
T5828A	ACT>GCT	17749	A>G	α					1	NS
F5835F	TTC>TTT	17768					C>T	α	1	S
G5862G	GGT>GGC	17851					T>C	α	1	S
E5875E	GAG>GAA	17888					G>A	α	1	S
T5890T	ACT>ACC	17934					T>C	α	1	S
E5915E	GAG>GAA	18013					G>A	α	1	S
L5932L	CTT>CTC	18059					T>C	α	1	S
K5934K	AAG>AAA	18065					G>A	α	1	S
H5944H	CAC>CAT	18096					C>T	α	1	S
G5969G	GGT>GGC	18171					T>C	α	1	S
F5985F	TTC>TTT	18214					C>T	α	1	S
Y5989Y	TAC>TAT	18226					C>T	α	1	S
N5996N	AAT>AAC	18251					T>C	α	1	S
R6001R	CGT>CGC	18267					T>C	α1	1	NS
E6002E	GAG>GAA	18270					G>A	α	1	S
A6004A	GCC>GCT	18276					C>T	α	1	S
F6014F	TTT>TTC	18305					T>C	α	1	S
I6026V	ATT>GTT	18340	A>G	α1					1	NS
T6028T	ACT>ACC	18348					T>C	α	1	S
G6035G	GGC>GGT	18368					C>T	α	1	S
G6048G	GGC>GGT	18408					C>T	α	1	S
P6066P	CCA>CCG	18462					A>G	α	1	S
Q6070Q	CAG>CAA	18470					G>A	α	1	S
H6073H	CAT>CAC	18482					T>C	α	1	S
L6074L	CTT>CTC	18485					T>C	α	1	S
Y6079Y	TAT>TAC	18500					T>C	α	1	S
D6097D	GAT>GAC	18484					T>C	α	1	NS
V6107V	GTG>GTA	18585					G>A	α	1	S
E6116E	GAA>GAG	18611					A>G	α	1	S
L6117L	TTA>TTG	18614					A>G	α	1	S
T6131T	ACT>ACC	18653					T>C	α	1	S
K6137R	AAA>AGA	18673			A>G	α1			1	NS
Y6161Y	TAT>TAC	18744					T>C	α	1	S
Q6187Q	CAG>CAA	18825					G>A	α	1	S
R6203R	AGA>AGG	18828					A>G	α	1	S
Y6221Y	TAC>TAT	18873					C>T	α	1	S
F6251F	TTT>TTC	19016					T>C	α	1	S
H6255H	CAT>CAC	19029					T>C	α	1	S

D6277D	GAC>GAT	19095					C>T	α	1	S
D6389D	GAC>GAT	19432					C>T	α	1	S
C6398C	TGC>TGT	19457					C>T	α	1	S
L6420L	CTT>CTC	19524					T>C	α	1	S
Y6423Y	TAC>TAT	19533					C>T	α	1	S
Y6442Y	TAC>TAT	19590					C>T	α	1	S
N6446N	AAT>AAC	19602					T>C	α	1	S
K6464K	AAA>AAG	11656					A>G	α	1	S
S6476S	TCC>TCT	19694					C>T	α	1	S
L6494L	CTA>TTG	19743	C>T	α			A>G	α	2	S
L6509L	CTC>CTT	19790					C>T	α	1	S
N6514N	AAT>AAC	19874					T>C	α	1	S
I6537I	ATT>ATC	19874					T>C	α	1	S
N6566T	AAT>ACG	19961			A>C	Γ1	T>G	Γ1	2	NS
F6575F	TTC>TTT	19989					C>T	α1	1	S
V6579V	GTC>GTT	20000					C>T	α	1	S
N6578D	AAT>GAT	20001	A>G	α1					1	NS
L6585L	TTG>TTA	20019					G>A	α1	1	S
P6609P	CCT>CCC	20180					T>C	α	1	S
A6612A	GCC>GCT	20100					C>T	α	1	S
L6619L	TTG>TTA	20120					G>A	α	1	S
L6624V	TTA>GTA	20134	T>G	Γ1					1	NS
Y6631Y	TAT>TAC	20153					T>C	α	1	S
N6635D	AAT>GAT	20163	A>G	α1					1	NS
V6638V	GTT>GTC	20177					T>C	α	1	S
L6641L	CTA>TTA	20185	C>T	α					1	S
T6647T	ACC>ACT	20205					C>T	α	1	S
Q6648Q	CAA>CAG	20208					A>G	α	1	S
K6653Q	AAA>CAA	19820	A>C	Γ1					1	NS
F6655F	TTC>TTT	20228					C>T	α	1	S
L6679L	CTA>TTA	20298	C>T	α					1	S
G6681G	GGT>GCC	20307					T>C	α	1	S
V6688V	GTC>GTA	20328					C>A	Γ	1	S
R6695S	AGG>AGT	20348					G>T	Γ1	1	NS
S6710F	TCT>TTT	20393			C>T	α1			1	NS
L6715F	CTT>TTT	20406	C>T	α1					1	NS
N6729N	AAT>AAC	20451					T>C	α	1	S
T6730T	TAC>TAT	20453					C>T	α	1	S
S6760S	TCG>TCC	20544					G>C	ß	1	S
Q6761Q	CAG>CAA	20547					G>A	α	1	S
V6771V	GTG>GTT	20571					G>T	Γ	1	S
T6777T	ACC>ACT	20883					C>T	α	1	S

G6789G	GGA>GGC	20628					A>C	Г	1	S
V6790V	GTT>GTA	20634					T>A	ß	1	S
L6812L	CTC>CTT	20699					C>T	α	1	S
L6819L	TTA>CTA	20718	T>C	α					1	S
L6820L	CTA>TTA	20721	C>T	α					1	S
Q6850Q	CAG>CAA	20814					G>A	α	1	S
N6853N	AAT>AAC	20823					T>C	α	1	S
L6855L	CTT>TTA	20827	C>T	α			T>A	ß	2	S
T6856T	ACT>ACA	20830					T>A	ß	1	S
R6864R	AGG>AGA	20856					G>A	α	1	S
P6878P	CCT>CCA	20933					T>A	ß	1	S
T6890T	ACA>ACG	20937					A>G	α	1	S
L6891L	CTA>CTG	20940					A>G	α	1	S
D6931D	GAT>GAC	21057					T>C	α	1	S
D6942D	GAT>GAC	21090					T>C	α	1	S
G6946G	GGA>GGT	20402					A>T	ß	1	S
G6953G	GGA>GGG	21122					A>G	α	1	S
A6960A	GCC>GCT	21143					C>T	α	1	S
S6964S	TCT>TCC	21156					C>T	α	1	S
V6965V	GTA>GTG	21159					A>G	α	1	S
E6971E	GAG>GAA	21176					G>A	α	1	S
A6976A	GCC>GCT	21191					C>T	α	1	S
L6981L	CTC>CTT	21207					C>T	α	1	S
H6984H	CAT>CAC	21216					T>C	α	1	S
T6989T	ACT>ACA	21231					T>A	ß	1	S
A6990A	GCT>GCC	21234					T>C	α	1	S
V6995V	GTA>GTG	21248					A>G	α	1	S
A6997A	GCA>GCG	21254					T>A	ß	1	S
S6998S	TCT>TCA	21257					T>A	ß	1	S
									710	583
										106
Total Synonymous			α+ß+Г	38	α+ß+Г	0	α+ß+Г	547	585	2136
Total NonSynonymous			α1+ß1+Г1	45	α1+ß1+Г1	49	α1+ß1+Г1	31	125	106
Total Syn+NonSyn			α+ß+Г+α1+ß1 +Г1	83	α+ß+Г+α1+ ß1+Г1	49	α+ß+Г+α1+ ß1+Г1	578	710	2136
Total			α	37		0		491	710	
			ß	0		0		36		
			Г	1		0		20		

			α1	36		40		23		
			ß1	4		4		5		
			Γ1	5		5		3		
Total homologous codon		7096.6666	7097							
1 codons inserted		7096								
2 nt changes in 5 codon										
Total nucleotide		21290								
Overall Evolutionary Rate										
At codon position 1						**At codon position 2**		**At codon position 3**		
P	(probablity of transition)		0.010288936			0.00564		0.0724454		
Q	(probablity of tranversion1)		0.000563777			0.00056		0.0057787		
R	(probablity of transversion2)		0.000845666			0.0007		0.0032417		
The total no of base sub/site (K)										
1st codon (K)					**2nd codon (K)**			**3rd codon (K)**		
K	= -(1/4)ln[(1- 2P - 2Q)(1-2P - 2R)(1 - 2Q - 2R)]									
K_{1stcod}	0.0118				K_{2ndcod}	0.00695	K_{3rdcod}	0.0881191		
When T is time taken to evolute										
T	K/2knuc	/1st codon site/year				K/2knuc	/2nd codon site/year	K/2knuc	/3rd codon site/per year	
$T_{o1stcode}$=	3.3208025				$T_{o2ndcode}$=	1.92965	$T_{o3rdcode}$=	9.57817		
FOR SYNONYMOUS										
	P (probablity of transition)		0.00521494			0		0.0692037		
	Q(probablity of tranversion1)		0			0		0.005074		
	R(probablity of transversion2)		0.00521494			0		0.0028189		
K=	= -(1/4)ln[(1- 2P - 2Q)(1-2P - 2R)(1 - 2Q - 2R)]									
$K_{syn1stcode}$=	0.0105				$K_{syn12ndcode}$=	0.0000	$K_{syn3rdcode}$=	0.0830675		
K=	2Tknuc									

T	K/2knuc	/1st codon site/year				K/2knuc	/2nd codon site/year	K/2knuc	/3rd codon site/year	
$T_{s1stcode=}$	38.4202884				$T_{s2ndcode=}$	#DIV/0!	$T_{s3rdcode=}$	10.41921		
FOR NONSYNONYMOUS										
	P (probablity of transition)		0.005073996			0.00564		0.0032417		
	Q(probablity of tranversion1)		0.000563777			0.00056		0.0007047		
	R(probablity of transversion2)		0.000704722			0.0007		0.0004228		
K=	= -(1/4)ln[(1- 2P - 2Q)(1-2P - 2R)(1 - 2Q - 2R)]									
$K_{nsyn1stcode=}$	0.0064				$K_{nsyn2ndcode=}$	0.0069	$K_{nsyn3rdcode=}$	0.0043845		
K=	2Tkn1									
T=	K/2knuc	/1st codon site/year				K/2knuc	/2nd codon site/year	K/2knuc	/3rd codon site/ year	
$T_{ns1stcode=}$	1.86100863				$T_{ns2ndcode=}$	1.94939	$T_{ns3rdcode=}$	3.5591875		
The evolutionary rate per site per year (as per previous excel sheet base substitution rate=9.0E-04										
Evolutionary rate	O-tg13-husars	Syn-tg13-husars	Nonsyn-tg13-husars	O-husars	syn-husars	nonsyn-husars				
$k_{nuc1stcode}$	0.0118	0.0105	0.0064	0.0027	0.000136	0.0025				
$k_{nuc2ndcode}$	0.00694675	0.0000	0.0069	0.0026	0	0.0017				
$k_{nuc3rdcode}$	0.08811913	0.0830675	0.004384487	0.0069	0.005979	0.0006				

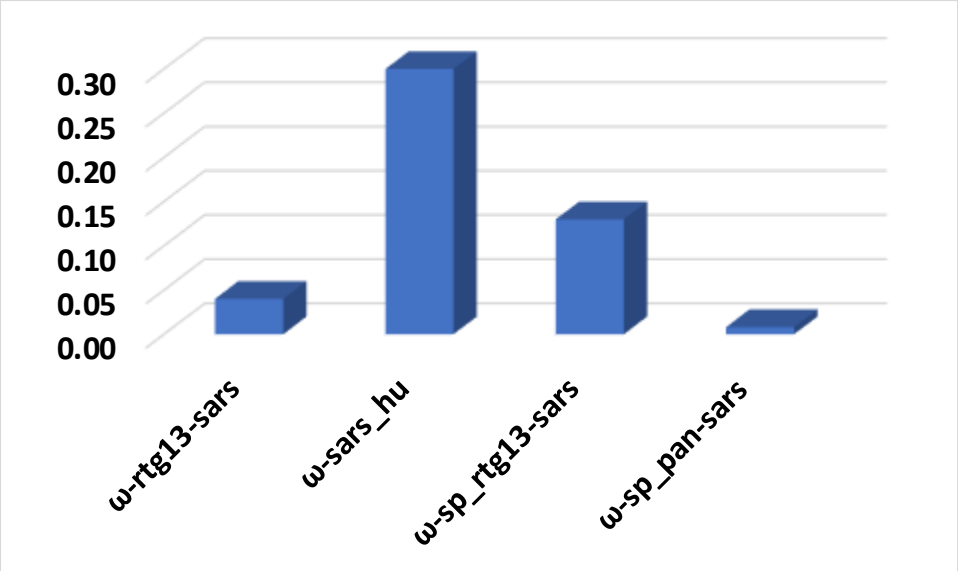


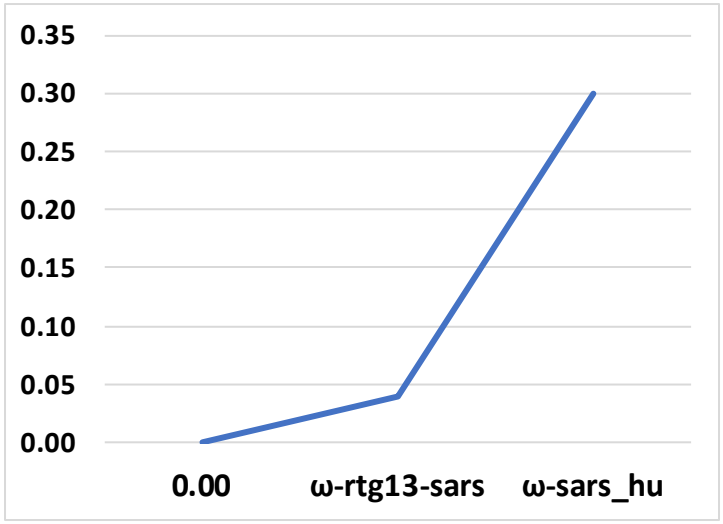
Supplementary Table 4: Nucleotide and amino acid differences in RaTG13, Pan_SL_Cov_GD and SARS_CoV2												
S Protein	RBD	438aa-506aa										
Bat-RaTG13	Pan_SL_CoV_GD	SARS-CoV2	RaTG13	Pan_SL_CoV_GD	SARS_CoV2	Key entry residues	PAN to SARS-CoV2	Type	RaTG13 to SARS-CoV2	Type	S score	NS score
TCT	TCT	TCT			S							
AAG	AAC	AAC	K		N				G>C	NS	0	1
CAT	AAC	AAT	H		N		C>T	S	C>A	NS	0	1
ATT	CTT	CTT	I		L							
GAT	GAT	GAT			D							
GCA	TCT	TCT	A		S							
AAA	AAG	AAG			K							
GAG	GTT	GTT	E		V							
GGC	GGT	GGT			G							
GGT	GGT	GGT			G							
AAT	AAT	AAT			N							
TTT	TAT	TAT	F		Y	449			T>A	NS	0	1
AAC	AAC	AAT			N		C>T	S	C>T	S	1	0
TAT	TAC	TAC			Y				T>C	S	1	0
CTT	CTT	CTG			L		T>G	S	T>G	S	1	0
TAC	TAT	TAT			Y				C>T	S	1	0
CGT	AGA	AGA			R				C>A, T>A	S	1	1
CTC	TTG	TTG			L	455			C>T, C>G	S	2	0
TTT	TTT	TTT			F							
AGA	AGA	AGG			R		A>G	S	A>G	S	1	0
AAA	AAG	AAG			K							
GCT	TCC	TCT	A		S		C>T	S	G>T	NS	0	1
AAT	AAC	AAT			N		C>T	S				
CTT	CTC	CTC			L				T>C	S	1	0
AAA	AAA	AAA			K							
CCC	CCT	CCT			P				C>T	S	1	0
TTT	TTT	TTT			F							
GAG	GAA	GAG			E		A>G	S				
AGG	CGA	AGA			R		C>A	S	G>A	S	1	0
GAT	GAC	GAT			D				C>T	S	1	0
ATC	ATT	ATT			I				C>T	S	1	0
TCA	TCT	TCA			S		T>A	S				
ACT	ACA	ACT			T		A>T	S				

GAA	GAA	GAA			E							
ATT	ATA	ATC			I		A>C	S	T>C	S	1	0
TAT	TAT	TAC			Y		T>C	S				
CAA	CAA	CAG			Q		A>G	S				
GCA	GCT	GCC			A		T>C	S	A>C	S	1	0
GGC	GGT	GGT			G				C>T	S	1	0
AGC	AGT	AGC			S		T>C	S				
AAA	ACA	ACA	K		T				A>C	NS	0	1
CCT	CCC	CCT			P		C>T	S				
TGT	TGC	TGT			C		C>T	S				
AAT	AAT	AAT			N							
GGT	GGG	GGT			G		G>T	S				
CAA	GTT	GTT	Q		V				C>G, A>T,A>T	NS	0.5	2.5
ACT	GAA	GAA	T		E				A>G, C>T,T>A	NS	0.5	2.5
GGT	GGT	GGT			G							
CTA	TTT	TTT	L		F	486			C>T, A>T	NS	1	1
AAT	AAC	AAT			N		C>T	S				
TGC	TGT	TGT			C							
TAC	TAT	TAC			Y	489	T>C	S				
TAC	TTT	TTT	Y		F				A>T, C>T	NS	1	1
CCA	CCT	CCT			P				A>T	S	1	0
CTT	CTA	CTA			L				T>A	S	1	0
TAT	CAA	CAA	Y		Q	493			T>C, T>A	NS	0	2
AGA	TCT	TCA	R		S		T>A	S	A>T,G>C	NS	0	2
TAT	TAT	TAT			Y							
GGA	GGT	GGT			G				A>T	S	1	0
TTT	TTC	TTC			F				T>C	S	1	0
TAC	CAC	CAA	Y	H	Q		T>C	NS	T>C, C>A	NS	0	2
CCT	CCT	CCC			P		T>C	S				
ACT	ACT	ACT			T	500						
GAT	AAT	AAT	D		N	501			G>A	NS	0	1
GGT	GGT	GGT			G							
GTT	GTT	GTT			V							
GGT	GGT	GGT			G							
CAC	TAC	TAC	H		Y				C>T	S		
CAA	CAA	CAA			Q							
							53.13	153.9				
		aa Match	aa match	Aa Mismatch								

Total codon compared		69										
All match	Yellow	17										
SARS+Pan	Blue	28	45	24	=S16+NS1							
PAN+RaTG13	Green	6	23									
SARS+RaTG13	Maroon	11	28	41	=S24+NS17							
Total mismatch	White	7	62									

Supplementary Table 5: ω value in various conversion					
				Luo et al (1985)	
	Nei and Gozobori (1986)	Yang and Neilson (2000)	LWL85	LWL85m	LWL93
	0.00				
$\omega_{\text{-rtg13-sars}}$	0.04	0.04	0.04	0.07	0.07
$\omega_{\text{-sars\_hu}}$	0.30	0.28	0.27	0.30	0.35
$\omega_{\text{-sp\_rtg13-sars}}$	0.13	0.17	0.14	0.15	0.16
$\omega_{\text{-sp\_pan-sars}}$	0.008	0.002	0.03	0.03	0.03
ω-sp_tg13-sars/ω-sp_pan-sars	16.25				
ω-rtg13-sars/ω-sp_pan-sars	5.00				




0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0.00
ω-rtg13-sars
ω-sars_hu

Supplemintary Table 6: ω (dn/ds) in RNA virus			
Type	Virus	ω	
Retrovirus	HIV virus	0.314	Lin et al (2019)
	HTLV1	0.113	Lin et al (2019)
SS +RNA	HCV	0.088	Lin et al (2019)
	West Nile virus	0.059	Lin et al (2019)
	Rubella	0.046	Lin et al (2019)
	Hepatitis E (HEV-4)	0.04	Lin et al (2019)
	$\omega_{RATG13\text{-}SARS}$	0.04	This study
	$\omega_{Hu\text{-}sars}$	0.3	This study
	$\omega_{S1\text{-}ratg13\text{-}sars}$	0.13	This study
	$\omega_{s1\text{-}Pan\text{-}sars}$	0.0078	This study
	Dengue virus (DENV-1)	0.03	Lin et al (2019)
SS -RNA	Influenza A (H3N2)	0.08	Lin et al (2019)
	Influenza A (H1N1)	0.077	Lin et al (2019)
	Ebola	0.077	Lin et al (2019)