microbiology research

MDPI

Article

In-Silico Pangenomics of SARS-CoV-2 Isolates Reveal Evidence for Subtle Adaptive Expression Strategies, Continued Clonal Evolution, and Sub-Clonal Emergences, Despite Genome Stability

Kamaleldin B. Said [1,2,3,*], Ahmed Alsolami [4], Anas Fathuldeen [4], Fawwaz Alshammari [4], Walid Alhiraabi [1], Salem Alaamer [1], Hamad Alrmaly [1], Fahad Aldamadi [1], Dakheel F. Aldakheel [5], Safia Moussa [5], Ahmed Al Jadani [4] and Abdulhafiz Bashir [6]

[1] Department of Pathology, College of Medicine, University of Ha'il, Ha'il 55476, Saudi Arabia; s201706514@uoh.edu.sa (W.A.); s201706582@liveuohedu.onmicrosoft.com (S.A.); s201712249@liveuohedu.onmicrosoft.com (H.A.); s2017171@uoh.edu.sa (F.A.)
[2] Genomics, Bioinformatics and Systems Biology, Carleton University, 1125 Colonel-By Drive, Ottawa, ON K1S 5B6, Canada
[3] ASC Molecular Bacteriology, McGill University, 21111 Lakeshore Rd, Montreal, QC H9X 3L9, Canada
[4] Department of Internal Medicine, College of Medicine, University of Ha'il, Ha'il 55476, Saudi Arabia; vicedean.medicine@uoh.edu.sa (A.A.); a.fathuldeen@uoh.edu.sa (A.F.); fawwazf@uoh.edu.sa (F.A.); a.aljadani@uoh.edu.sa (A.A.J.)
[5] Ministry of Health, Ha'il Region, Ha'il 55476, Saudi Arabia; daldakheel@moh.gov.sa (D.F.A.); safiamoussa89@yahook.com (S.M.)
[6] Department of Physiology, College of Medicine, Ha'il University, Ha'il 55476, Saudi Arabia; ah.bashir@uoh.edu.sa
* Correspondence: kbs.mohamed@uoh.edu.sa; Tel.: +966-500771459

Citation: Said, K.B.; Alsolami, A.; Fathuldeen, A.; Alshammari, F.; Alhiraabi, W.; Alaamer, S.; Alrmaly, H.; Aldamadi, F.; Aldakheel, F.; Mosa, S.; et al. In-Silico Pangenomics of SARS-CoV-2 Isolates Reveal Evidences for Subtle Adaptive Expression Strategies, Continued Clonal Evolution, and Sub-Clonal Emergences, Despite Genome Stability. *Microbiol. Res.* **2021**, *12*, 204–233. https://doi.org/10.3390/microbiolres12010016

Received: 13 February 2021
Accepted: 9 March 2021
Published: 17 March 2021

Abstract: The devastating SARS-CoV2 pandemic is worsening with relapsing surges, emerging mutants, and increasing mortalities. Despite enormous efforts, it is not clear how SARS-CoV2 adapts and evolves in a clonal background. Laboratory research is hindered by high biosafety demands. However, the rapid sequence availability opened doors for bioinformatics. Using different bioinformatics programs, we investigated 6305 sequences for clonality, expressions strategies, and evolutionary dynamics. Results showed high nucleotide identity of 99.9% among SARS-CoV2 indicating clonal evolution and genome. High sequence identity and phylogenetic tree concordance were obtained with isolates from different regions. In any given tree topology, ~50% of isolates in a country formed country-specific sub-clusters. However, abundances of subtle overexpression strategies were found including transversions, signature-sequences and slippery-structures. Five different short tracks dominated with identical location patterns in all genomes where Slippery-4 *AAGAA* was the most abundant. Interestingly, transversion and transition substitutions mostly affected the same amino acid residues implying compensatory changes. To ensure these strategies were independent of sequence clonality, we simultaneously examined sequence homology indicators; tandem-repeats, restriction-site, and 3'UTR, 5' UTR-caps and stem-loop locations in addition to stringent alignment parameters for 100% identity which all confirmed stability. Nevertheless, two rare events; a rearrangement in two SARS-CoV2 isolates against betacoronavirus ancestor and a polymorphism in S gene, were detected. Thus, we report on abundance of transversions, slippery sequences, and ON/OFF molecular structures, implying adaptive expressions had occurred, despite clonal evolution and genome stability. Furthermore, functional validation of the point mutations would provide insights into mechanisms of SARS-CoV2 virulence and adaptation.

Keywords: SARS-CoV2-phylogenetics; SARS-CoV2-genomics; SARS-CoV2-bioinformatics; Coronavirus pandemic; Covid-19 pandemic

1. Introduction

The global community is under siege due to the on-going devastating pandemic by one of the most serious zoonotic coronavirus lineages of all times. In a matter of months, Covid-19, the disease caused by the Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV2), has reached every corner on earth dramatically changing the way humans live. As of January 12, 2021, 88,387,352 cases and about 2 million deaths were documented in the WHO situation report [https://www.who.int/publications/m/item/weekly-epidemiological-update---12-january-2021] accessed on 12 January 2021. The unprecedented speed in the spread of SARS-CoV2 and the subsequent evolution of mutants have resulted in increased hospitalizations, mortality are rates, and paralyzed global healthcare systems. Significant gaps have been created in our knowledge of evolutionary dynamics, epidemicity, and adaptive transcriptomic of coronavirus pandemics. It is not fully clear how the recurring zoonotic coronavirus lineages of including SARS-CoV2, successfully adapted, evolved, host-jumped, and maintained transmissions into humans.

Coronaviruses are enveloped, positive-sense, single-stranded RNA viruses that belong to the family Coronaviridae. Based on phylogenetic relationships and genomic structures, there are four genera of CoVs, namely, *Alphacoronavirus* (αCoV), *Betacoronavirus* (βCoV), *Deltacoronavirus* (δCoV), and *Gammacoronavirus* (γCoV) [1]. It has been widely established that most α- and βCoVs infect mammals, bats and rodents, where they usually cause respiratory illness in humans and gastroenteritis in animals, while avian species carry δCoVs and γCoVs. There are four known human coronaviruses (HCoV-NL63, HCoV-229E, HCoV-OC43 and HKU1). These were mostly known to cause mild infections in immunocompetent people until the outbreak of severe acute respiratory syndrome (SARS) in 2002 and 2003 in Guangdong province, China [2–5]. A decade later, another highly pathogenic coronavirus, the Middle East respiratory syndrome coronavirus (MERSCoV) emerged in Middle Eastern countries [6]. The current emergence of the SARS-CoV2 implies occurrence of significant changes in the evolutionary dynamics of the virus-host interactions in coronaviruses.

The molecular mechanisms underlying evolution of new coronavirus lineages with enhanced ability in cross-host transmissions is not well understood. Domain variations and receptor binding specificities were reported. All of the three viruses—SARS-CoV, MERS-CoV, and SARS-CoV2—showed frequent crosses to species barriers leading to rapid adaptation and human to human transmissions [7–9]. The massive number of coronaviruses carried by different bat species and the increasing human exploitation of the wild implied continued transmissions from bats to animals to humans. SARS-CoV and MERS-CoV were transmitted directly to humans from market civets and dromedary camels, respectively [10,11]; however both viruses were thought to have originated in bats [12,13]. In human, transmission was found facilitated by the SARS-CoV high affinity for angiotensin-converting enzyme 2 (ACE2) [14,15]. However, variations in binding affinities of SARS CoV for human ACE2 affected infectivity of human cells [16]. In the 2002–2003 outbreak strain hTor02 had a high affinity for ACE2 and efficiently transmitted between humans. However, palm civets strains cSz02 and cHb05 showed low affinity and low infectivity in human cells [10,17]. On the other hand, MERS-CoV uses dipeptidyl peptidase 4 (DPP4; also known as CD26) as a receptor and infects unciliated bronchial epithelial cells and type II pneumocytes [18,19]. While the S1-CTD is common in RBD of both MERS-CoV and SARS-CoV, the latter uses ACE2 and the former uses DPP4. These observations indicate potential changes in subtle translation and functional mechanisms, rather than sequence variations of accessory genes.

Domain variations within the binding structures of previous coronaviruses have been reported. Although SARS-CoVs and bat SARSr-CoVs share high sequence identity in the S2 domain, they mainly vary in three regions: (1) the spike protein (S) (both the S1 amino-terminal domain (S1-NTD) and the S1 receptor binding domain [20] (2) ORF8 (8a and b) and (3)

ORF3 (3a and 3b). Compared with human and civet SARS-CoV, bat SARSr-CoV S1 can be separated out into two clades. Clade 1, Yunnan province specific and has the same size S protein as human and civet isolates and clade 2 which is universal in many locations has a shorter size S protein due to deletions of 5, 12 or 13 amino acid residues [21]. The second variation occurs in the 366 or 369 nucleotides *orf8* locus retained intact in all bat SARSr-CoVs and shared among themselves and with civet and human SARS CoVs. Analysis of the 2002–2003 outbreak transmission pattern explained how the *orf8* gene underwent phases of adaptations during transmission from animals to humans. The outbreak occurred in three phases; in the early-phase (localized limited cases) patients had two genotypes of *orf8* on their viral genomes; (one with a complete *orf8* of 369 nucleotides, and the other with 82-nucleotide deletion}. Whereas, in late-phase (pandemic) and most middle-phase (super-reader) patients had a split *orf8* (*orf8a* and *orf8b*) on their viral genomes because of a 29-nucleotide deletion [17]. However, two exceptions were found in the middle-phase genomes, one that had an 82-nucleotide deletion in *orf8* and the other with the whole *orf8* deleted. The human isolates from 2004 and all civet SARS-CoV genomes had a complete *orf8* except one civet strain with an 82-nucleotide deletion. Thus, *orf8* genes are accessory adaptation markers that aid in transmission from animals to humans and thus, are potential targets as gene candidates. Interestingly, the European bat SARSr-CoV has completely lost *orf8* [22] implying that the *orf8* accessory genes in bat SARSrCoVs are potential targets for continuous evolution in their natural hosts like bats, civets and humans and warranting further investigations for diagnosis and therapeutics. The third variable locus is in ORF3. On SARS CoV genome, this is a 154-amino acid that codes for an interferon antagonist. While the ORF3a locus is highly similar in both SARSr-CoVs and SARS-CoV (96.4–98.9% amino acid identity), the SARSr-CoVs have different sizes of ORF3b [21]. These observations indicate functional changes in addition to sequence variations. Thus, there seems to be a significant knowledge gap in how coronaviruses exploit functional mechanisms, RNA modifications, and coding potentials to adapt to different hosts, despite RNA stability.

RNA modifications and signature sequences play important roles in prokaryotic adaptive expressions. Differential over expression of chromosomal genes, mutation in accessory genes, slippage on signature sequences, and repeat based hot spots that serve as ON/OFF switches are some of the mechanisms involved. In bacteria, these are mostly found associated to successfully re-emerging and outbreak pathogens such as *Hemophilus influenza*, streptococcus, and methicillin-resistant *Staphylococcus aureus*. In E. coli prfB mRNA for instance, the U CUU UGA slippery site contains the in frame UGA termination codon which is recognized by RF2 [23]. Although adaptive replication and transcription mechanisms are common in prokaryotes, it is still unclear whether the general rule applies to SARS-CoV-2. The translational reading frame establishment and maintenance with a universal start AUG codon (encoding methionine), has become a more complicated scenario in polycistronic prokaryotes mRNA translation processes. Relatively recent data on ribosome profiling suggest large amounts of translation initiation events occur at only a subset of non-AUG codons [24]. Enabling multiple coding mechanism to manipulate RNA structure-based strategies for translation is a selective advantage for rapid adaptive expressions [25]. An example of these are production of overlapping Open Reading Frames (ORFs) through Programmed Ribosomal Frameshifting (PRF) recoding, ON/OFF molecular switches, and differential expressions, to enhance the coding capacity of a single RNA template. These mechanisms are all 'programmed' at specific cis-acting elements on mRNAs, and at higher orders of magnitude more frequent than non-programmed events. Some of the cis-acting signals located in mRNA sequences are known regulatory loci that improve gene expression. It involves classes of loci that direct elongating ribosomes shifting the reading frame by one (or more) base in either directions; the 5′ (−1) or 3′ (+1). The process is largely programed for over expression to compensate for the compact

genomes size and hence it is called Programmed Ribosomal Frameshifting (PRF). Its common functional property is to stimulate ribosomes to pause at specific 'slippery' sequences. These PRF events were found to operate at efficiencies as high as 80% increasing the coding capacity of genomes at the same time regulating mRNA stability [26]. However, ribosome can also shift by –2, –4, +2, +5, or +6 nucleotides in rare events [27,28]. In coronaviruses, alterations in triplet decoding of the messenger RNA by the elongating ribosome is influenced by many factors including RNA dimerization [29]. As a result, protein synthesis does not proceed through the usual translation steps. Instead, while the first major polyprotein produced by the ORF1a that encodes non-structural proteins is translated normally, PRF acts on the elongation process. Slippery signals just before the termination codon of ORF1a redirect some species of translating ribosomes making them to bypass the stop codon. This allows translation to continue in the −1 reading frame thereby creating the larger ORF1ab polyprotein [30–32]. As a result, the ORF1a produces polypeptide 440–500 kDa protein that is cut into ~11 non structural proteins (nsps). The -1 PRF occurs right upstream of the ORF1a stop codon allowing translation to continue giving the large polypeptide ORF1b measuring 740–810 kDa that is cut into 16 nsps. Nevertheless, it is not yet clear what are the types, numbers, and locations of specific protective modification events that stabilize RNA and the types of changes that occur in the protein. In eukaryotic viruses, the slippery site has the heptameric motif N NNW WWH, where the incoming reading frame is indicated, and N = any three identical nucleotides, W = AAA or UUU, and N ≠ G [33]. The slippery site can also be presented as: X XXY YYZ, where the underlined codons denote the 0 reading frame. The ribosome slips on these sites to change register by 1 nucleotide in the 5′ direction, followed by a pseudoknot. DNA nanoball sequencing has revealed that the transcriptome of SARS-CoV2 is highly complex due to numerous discontinuous transcription events [34]. However, it is known that RNA is protected through series of mechanisms such as 5′ cap, a highly methylated modification, and a Poly A tail to each end of a pre-mRNA molecule to protects the new mRNA. While genomic RNA is used to assemble progeny virions, shorter sgRNAs encode conserved structural proteins. These are namely: the spike surface glycoprotein (S), small envelope protein (E), matrix protein (M), and nucleocapsid protein (N). The S protein is involved in binding to receptors on the host cell and determines host tropism [35,36]. Although SARS-CoV and MERS-CoV are related, their spike proteins bind different receptors indicating different mechanisms under common genomic backgrounds. In addition, SARS-CoV-2 is known to have at least eight accessory proteins (3a, 3b, p6, 7a, 7b, 8b, 9b, and orf14) [37] (GenBank: NC_045512.2). However, not all ORFs are experimentally verified for expression.

Regular update on the evolutionary process and phylogenetics of SARS-CoV2 is imperative due to the changing epidemiology of the virus. Early genomic analysis suggested evolutionary association of 2019-nCoV and SARS like bat coronaviruses and MERS [38]. In depth comparative analysis of the first determined genomes of the novel coronavirus 2019-nCoV with comparison to related coronaviruses strains determined that the three strains were almost identical. Comparisons included Wuhan/IVDC-HB-01/2019 (GISAID accession ID:EPI_ISL_402119) (HB01), Wuhan/IVDCHB-04/2019(EPI_ISL_402120) (HB04), and Wuhan/IVDC-HB-05/2019 (EPI_ISL_402121) (HB05). Analysis included 1008 human SARS CoV, 338 bat SARS-like CoV, and 3131 human MERS-CoV, whose genomes were published before 12 January 2020 [37]. The notable difference are the 8a protein is present in SARS-CoV and absent in 2019-nCoV; the 8b protein is 84 amino acids in SARS-CoV, but longer in 2019-nCoV, with 121 amino acids; the 3b protein is 154 amino acids in SARS-CoV, but shorter in 2019-nCoV, with only 22 amino acids [37]. It is still not clear how these differences affect the functionality and pathogenesis of 2019-nCoV. Similarly, recent genomic characterization and evolutionary origins of SARS-CoV-2 viruses isolated from a number of patients indicated sequence identity of about 99.9%, potentially implying a very recent host jump into humans

[39–41]. Based on phylogenetic and evolutionary dynamics, the novel coronavirus Covid 19 was found to belong to genus *Betacoronavirus*, subgenus *Sarbecovirus* (previously lineage 2b of Group 2 coronavirus) with 86% similarity to SARSr-CoV [42]. According to International Committee on Taxonomy of Viruses (ICTV) criteria, only the strains found in *Rhinolophus* bats in European countries, Southeast Asian countries and China are SARSr-CoV variants. These data indicate that SARSr-CoVs have wide geographical spread and might have been prevalent in bats for a very long time. In the absence of a direct ancestor of SARS-CoV in bat populations despite nearly two decades of research and since RNA coronaviruses recombination are frequent, there is a high potentiality that SARS-CoV2 have newly emerged from bat SARSr-CoVs by recombination. Recombination of two existing bat strains, WIV16 and Rf4092 giving rise to the civet SARS-CoV strain SZ3 have been reported [21]. Furthermore, because α- and β-CoVs infect livestock animals such as porcine transmissible gastroenteritis virus, porcine enteric diarrhoea virus (PEDV) and the recently emerged swine acute diarrhea syndrome coronavirus (SADS-CoV), they pose significant risk to human. Selection pressures and high adaptability in humans might be driving further adaptive mutations in this strain. Population genetic analyses of 103 SARS-CoV-2 genomes nucleotides have shown SARS-CoV2 evolution into two major types. These include a more prevalent and aggressive L type that evolved during the outbreak from the ancestral less aggressive clone S [43].

Thus, the main objectives of this work were to understand how SARS-CoV2 quickly adapts and evolves in humans, despite the genome stability. How the knowledge gap in pathogenicity, epidemicity, and evolutionary dynamics created by the unprecedented behavior of this virus, were different from earlier outbreaks. What were the successful mechanisms of and strategies underlying the success of the lineage in causing the pandemic. How successful bioinformatics programs in providing in silico evidences in the light of demands for higher biosafety level protocols. Thus, sequence based comparative genomics and functional analysis were successful in identifying abundance of signature-sequences as subtle expression strategies and provided evidences for sub-clonal differentiations at different regions despite genome stability. We have simultaneously examined sequence stability indicators including tandem-repeats, restriction site, and 3′UTR, 5′ UTR caps and stem-loop locations in addition to stringent alignment parameters for 100% identity, which all confirmed stability. Nevertheless, two extremely rare events were detected; one involved a segment inversion against betacoronavirus ancestor and the second was a significant polymorphism in the S gene. Furthermore, abundance of transversions, slippery sequences, and ON/OFF molecular structures, implied adaptive expressions had occurred in a clonal background. Future analysis of new isolates and experimental validation of these findings will provide more insights into the adaptive evolution of SARS-CoV2.

2. Materials and Methods

The dataset comprised of all currently available (n = 6314) full genome sequences from the current (2019–2020) SARS-CoV2 pandemic, as well as closely related MERS (n = 35), SARS-CoV (n = 220 of which 32 were selected for alignments), and SARS CoV related bat strains (SARS-like CoV) (n = 29). These were retrieved from NCBI (http://www.ncbi.nlm.nih.gov/genbank/) and also confirmed in the GISAID (https://www.gisaid.org/) databases. Before genomic analysis, all sequences were subjected to proofing and validations for integrity of sequences using available software and programs.

Pan-genome Sequence Analysis and phylogenetic analysis was performed using Clustal Omega 1.2.2 in Geneious Bioinformatics Package using Clustal size set at 180 in mBed Algorithm and Guide Tree and validated by MUSCLE 3.8.425, and Geneious Alignment in Geneious primer 2020.2.3 Java version 11.0.4+11 (64 bit). In alignment and guide phylogenetic tree view, strains showing 100% identity within the same clusters were dimmed in color and only

one or a few representative strains were highlighted to show genomic structural details to allow maximum visibility and to accommodate all strains within different clusters tested. Analysis of the alignments was done in Geneious Bioinformatics Package (Geneious Primer 2020.2.3 Biomatters Ltd., Biomatters, New Zealand). Consensus sequence threshold was set at 100% and highlighted agreement to consensus. The following analysis were done within Geneious and in standalone programs:

In the whole genome multiple alignment view the following analysis were shown out: annotations, tracks of repeats, and slippery sequences were shown as color-coded. The following tracks of slippery sequences were identified by the search criteria: 100% match, zero mismatches, and both directions (double sided arrow indicate both directions), at approximate intervals and locations shown on all genomes on number alignment window: As an example, tracks of Strain NC_014470 will be described:

Geneious annotations and tracks: (details and locations of all of these are found in Additional Files attached) gene annotations and names (green color bars) were set to place on the sequence lines. 3′UTR, 5′ UTR, and Stem loops were shown alignments.

Slippery sequences (Geneious primer software) (brown arrows, one and two sided): slippery sequence 1: 7 nt (UUUAAC, TTTAAAC); slippery sequence 2: 7 nt (UUUAAAA, TTTAAAA); slippery sequence 3: 7 nt (UUUAAAU, TTTAAAT); slippery sequence 4: 5 nt (AAGAA); slippery sequence 5: 7 nt (UUUUUUA, TTTTTTA)

Tandem repeat tracks (using Phobos Tandem repeat finder 3.3.11 2006–2010) (each of the red color vertical arrows along genome sequence indicated a tandemly repeat unit track. The number of vertical arrows in a location indicated the number of times repeat unit copies are repeated). The program was set to "Perfect Repeats" with unit length (1–500 bp), "100% identity", "No gaps" "high score", "high index values"

Restriction sites: Geneious program was used to identify restriction enzyme sites on the genomes. Commonly available enzymes were selected in addition to a long list sites. Restriction site cleavage abilities were validated against methylation or variability, ambiguity, or deletion of targets.

Phylogenetic guide-trees were built by standalone Clustal Omega 1.2.2 and Geneious Tree Builder using Tamura-Nei genetic distance model and Neighbor-Joining method of tree building with and without outgroups (consensus method used for clustering was Majority Greedy clustering with (%) consensus support at nodes. Tree branches were cladogram based in a decreasing order with a line weight of 2 and tip labeled with standard strain names displaying country of isolation. The output trees were repeated several times using different programs to confirm the topology. Circular tree layout showing the "origins" of evolutionary lines consistent with the original bat strains from Europe, Bulgaria (NC_014470; GU190215; as well as a Chinese strain MT084071) from reference strains indicated by bold blue nodes were selected as the best output. All identical clusters within lineages were color coded. Similar nucleotide substitutions within clusters were also shown.

Whole genome alignment for pairwise alignment of specific genomes of interest was performed using progressive Mauve algorithm [44] updated versions. Mauve Genome Alignment had a special viewer which enabled detection of genome rearrangements and locally aligned blocks at a glance. Each sequence was represented by one horizontal panel of blocks. Each colored block represented a region of sequence that aligned to part of another genome, and was presumably homologous and free from internal rearrangements. These were called locally collinear blocks (LCBs). The alignment display was organized into one horizontal "panel" per input genome sequence. Each genome's panel contained the name of the genome sequence, a scale showing the sequence coordinates for that genome, and a single black horizontal center line. Colored block outlines appeared above and possibly below the center line. Each of these block outlines surrounded a region of the genome sequence that aligned to part

of another genome, and was presumably homologous and internally free from genomic rearrangement. When a block lay above the center line the aligned region was in the forward orientation relative to the first genome sequence. Blocks below the center line indicated regions that aligned in the reverse complement (inverse) orientation. Regions outside blocks lacked detectable homology among the input genomes. Inside each block Mauve drew a similarity profile of the genome sequence. The height of the similarity profile corresponded to the average level of conservation in that region of the genome sequence. Areas that were completely white were not aligned and probably contained sequence elements specific to a particular genome. The height of the similarity profile was calculated to be inversely proportional to the average alignment column entropy over a region of the alignment.

The backbone color scheme in Mauve: colors regions conserved among all genomes differently than regions conserved among subsets of the genomes. Conserved regions among all genomes are termed "backbone," which are drawn in mauve color. The colored blocks in each genome are connected by vertical lines.

The Tandem Repeat Finder program available at [https://tandem.bu.edu/trf/trf.html] accessed on 25 September 2020, was used to confirm repeat unit sizes, copies, consensus sequences, scores, and flanking sequences.

Map to references is used for the 6312 genomes was used with the default setting except for the following criteria that were selected: Consensus set to "100% identity" Seq Coverage = 29.89, Genenious Mapper with "High Sensitivity", five times "Fine Tuning" after alignment, saved "assembly report", and "Do not trim" before mapping. Due to the alignment size produced, only visible regions were saved and presented in this study. Extensive analysis were carried out to understand the extent of transitions and transversions. A total of 95 seqs that showed increased transversions were extracted and used for independent analysis to identify common features, if any. The reference assembly algorithm used is a seed and expand style mapper followed by an optional fine tuning step to better align reads around indels to each other rather than the reference sequence. Various optimizations and heuristics are applied at each of analysis. The final optional fine tuning step at the end, shuffles the gaps around so that they reads better align to each other rather than the reference sequence

Dot Plot analysis integrated in Geneious Package was used to detect and identify significantly long repeats. In addition, authenticity of the repeats was confirmed by this program. In addition, chromosomal rearrangements, inversions and translocations were identified.

Full genome sequences of the pandemic isolate SARS-CoV2 (7000 genomes by September–October 2020) as well as selected new isolates and closely related MERS (n = 35), SARS-CoV (n = 220 of which 32 were selected for alignments), and SARS CoV related bat isolates (SARS-like CoV) (n = 29) were downloaded from public databases. These were retrieved from NCBI (http://www.ncbi.nlm.nih.gov/genbank/, accessed on 26 May 2020) and GISAID (https://www.gisaid.org/, accessed on 1 January 2020) databases. Before genomic analysis, all sequences were subjected to proofing and validations for sequence integrity using available software and programs.

Pan-genome sequence analysis and phylogenetics was performed using Clustal Omega 1.2.2 in Geneious Bioinformatics Package using Clustal size set at 180 in mBed Algorithm and Guide Tree and validated by MUSCLE 3.8.425, and Geneious Alignment in Geneious primer 2020.2.3 Java version 11.0.4+11 (64 bit). In alignment and guide phylogenetic tree view, isolates within a cluster were colord. Analysis of the alignments was done in Geneious Bioinformatics Package (Geneious Primer 2020.2.3 Biomatters Ltd.). Consensus sequence threshold was set at 100%. Two operations executed using two settings: in one setting "agreement to consensus" employed to highlight nucleotide agreements amongst all genomes; in another, the disagreements to consensus was used to highlight locations of genomic disagreements between the aligned genomes.

Phylogenetic guide-trees were built by standalone Clustal Omega 1.2.2 and Geneious Tree Builder using Tamura–Nei genetic distance model and Neighbor-Joining method of tree building with and without outgroups (consensus method used for clustering was Majority Greedy clustering with (%) consensus support at nodes. Tree branches were cladogram based in a decreasing order with a line weight of 1 and tip labeled with standard isolate names displaying country of isolation. The output trees were repeated several times using different programs to confirm the topology. Circular tree layout showing the "origins" of evolutionary lines consistent with the original bat isolates from Europe, Bulgaria (NC_014470; GU190215; as well as a Chinese isolate MT084071) from reference isolates indicated by green nodes was selected as the best output. All identical clusters within lineages were color coded. Similar nucleotide substitutions within clusters were also shown. SARS-CoV2 genomic identity and alignments were further confirmed by using Geneious Prime Dotplot; visual genome-wide sequence comparisons using Dotplot High Sensitivity settings; probabilistic score matrix and adjusting window size and threshold and a smaller tile size at 1000.

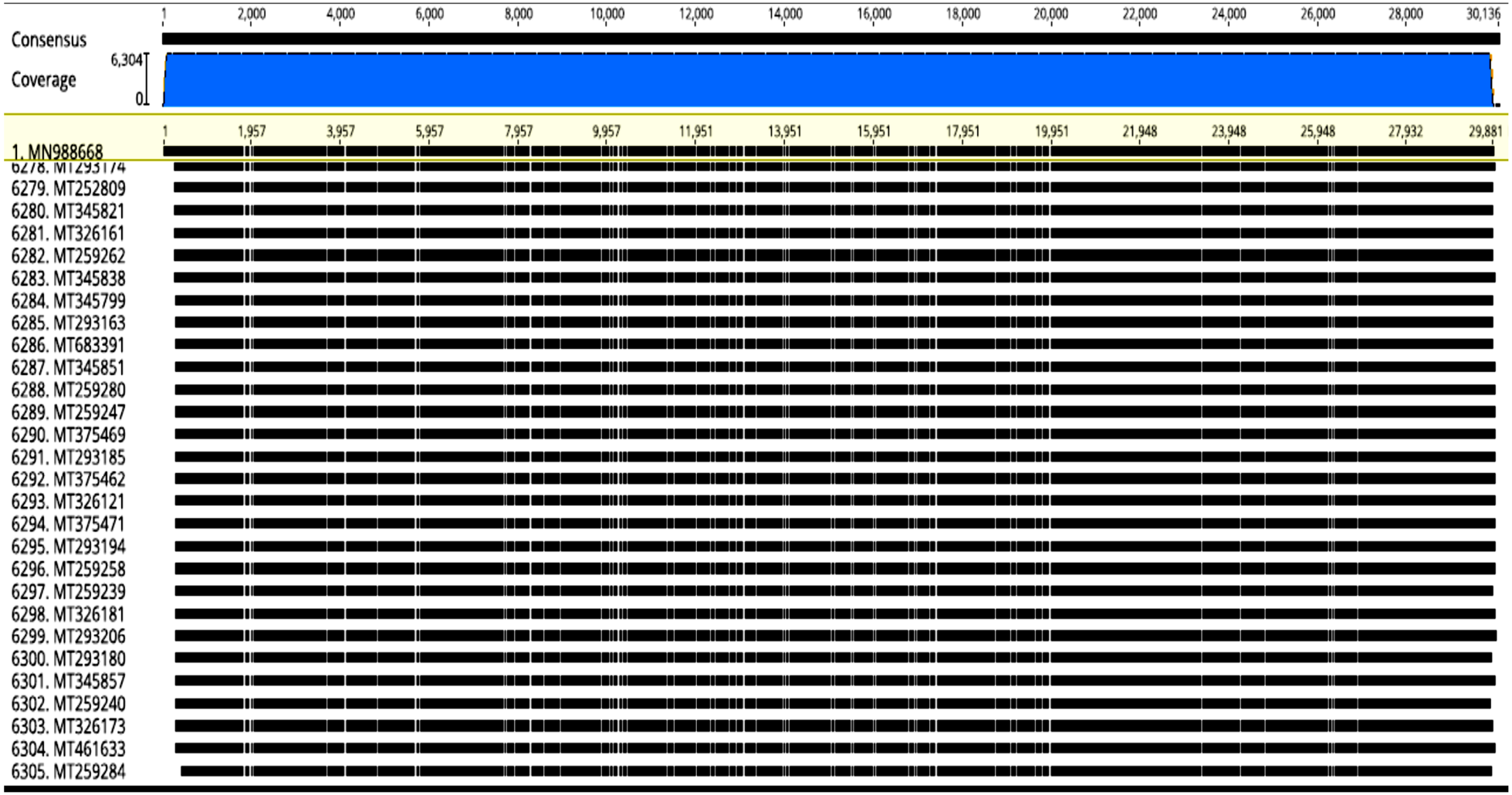
3. Results

3.1. Clonal Genome and Evolution

Complete genome sequences of global SARS-CoV2 pandemic that were made available until towards the end of 2020 were used in this study. As indicated in the figures, distinct genomic clusters were identified that aligned together. Of these, 6305 genomes representing the breadth of the virus used in whole genome analysis were mapped to the Wuhan-1 reference (MN988668) with 99.9% nucleotide identity (Figure 1a). As indicated in this figure, sequence coverage was 1–6305, however, only the last visible regions in the analysis window are shown from 6280 to 6305 for simplicity. This level of identity was maintained throughout different test parameters against the references and consensus sequence including gaps, transitions, transversions, agreement and disagreement to consensus sequences and to references. However, significant changes in alignment was observed when transversions to reference was selected (Figure 1b). To further examine, 174 genome sequences, from different geographic regions that showed abundant transversions, were re-aligned and examined against the reference NC_045512. As shown in Figure 2a, a high degree of genome sequence homology was seen with 99.95% nucleotide identity. Many transversion were highlighted (black vertical dashes); however, isolates in country-specific sub-clusters showed identical transversion patterns.


1
2,000
4,000
6,000
8,000
10,000
12,000
14,000
16,000
18,000
20,000
22,000
24,000
26,000
28,000
30,136
Consensus
Coverage
6,304
0
Identity
1
1,957
3,957
5,957
7,957
9,957
11,951
13,951
15,951
17,951
19,951
21,948
23,948
25,948
27,932
29,881
1. MN988668
6280. MT345821
6281. MT326161
6282. MT259262
6283. MT345838
6284. MT345799
6285. MT293163
6286. MT683391
6287. MT345851
6288. MT259280
6289. MT259247
6290. MT375469
6291. MT293185
6292. MT375462
6293. MT326121
6294. MT375471
6295. MT293194
6296. MT259258
6297. MT259239
6298. MT326181
6299. MT293206
6300. MT293180
6301. MT345857
6302. MT259240
6303. MT326173
6304. MT461633
6305. MT259284

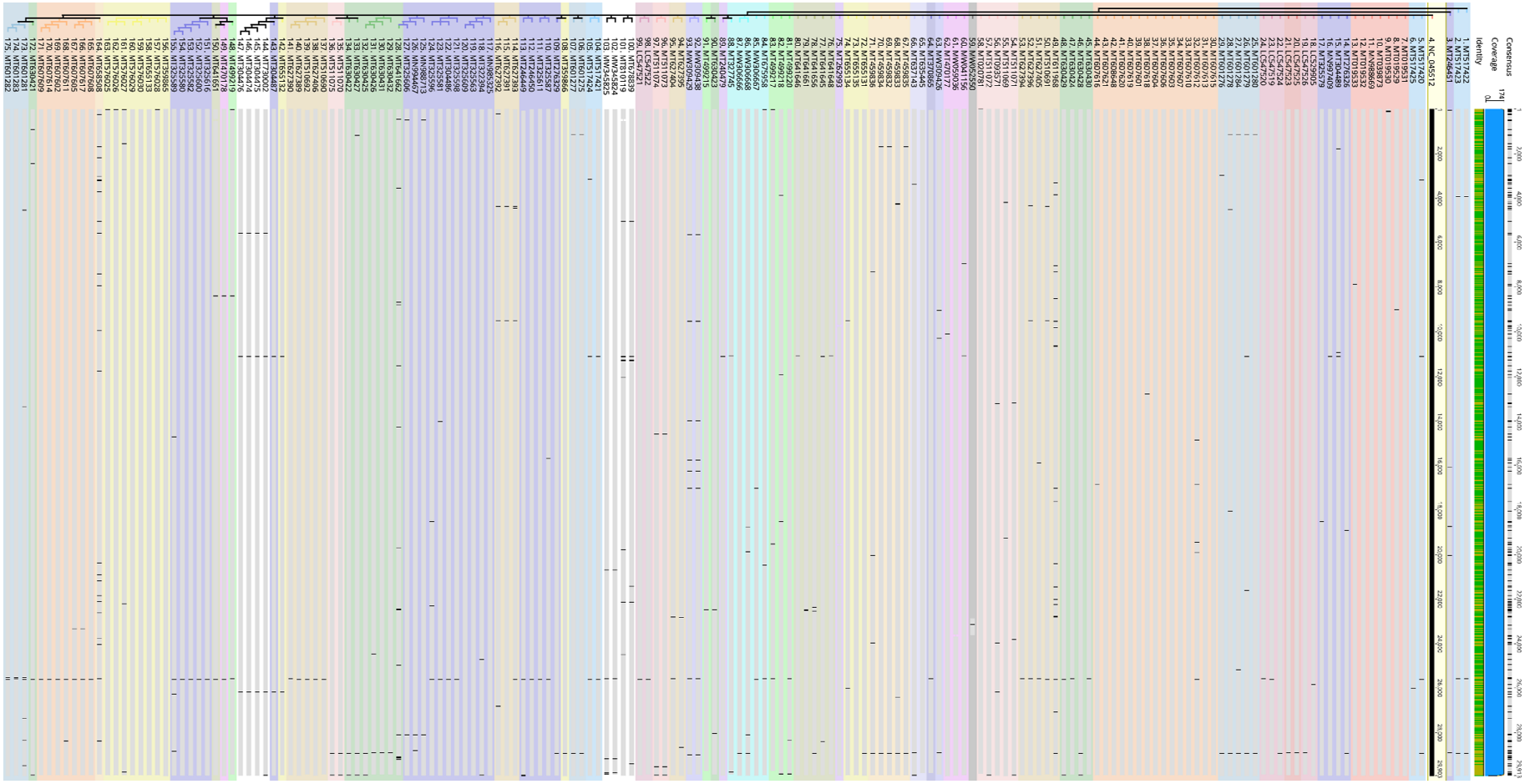
(**a**)



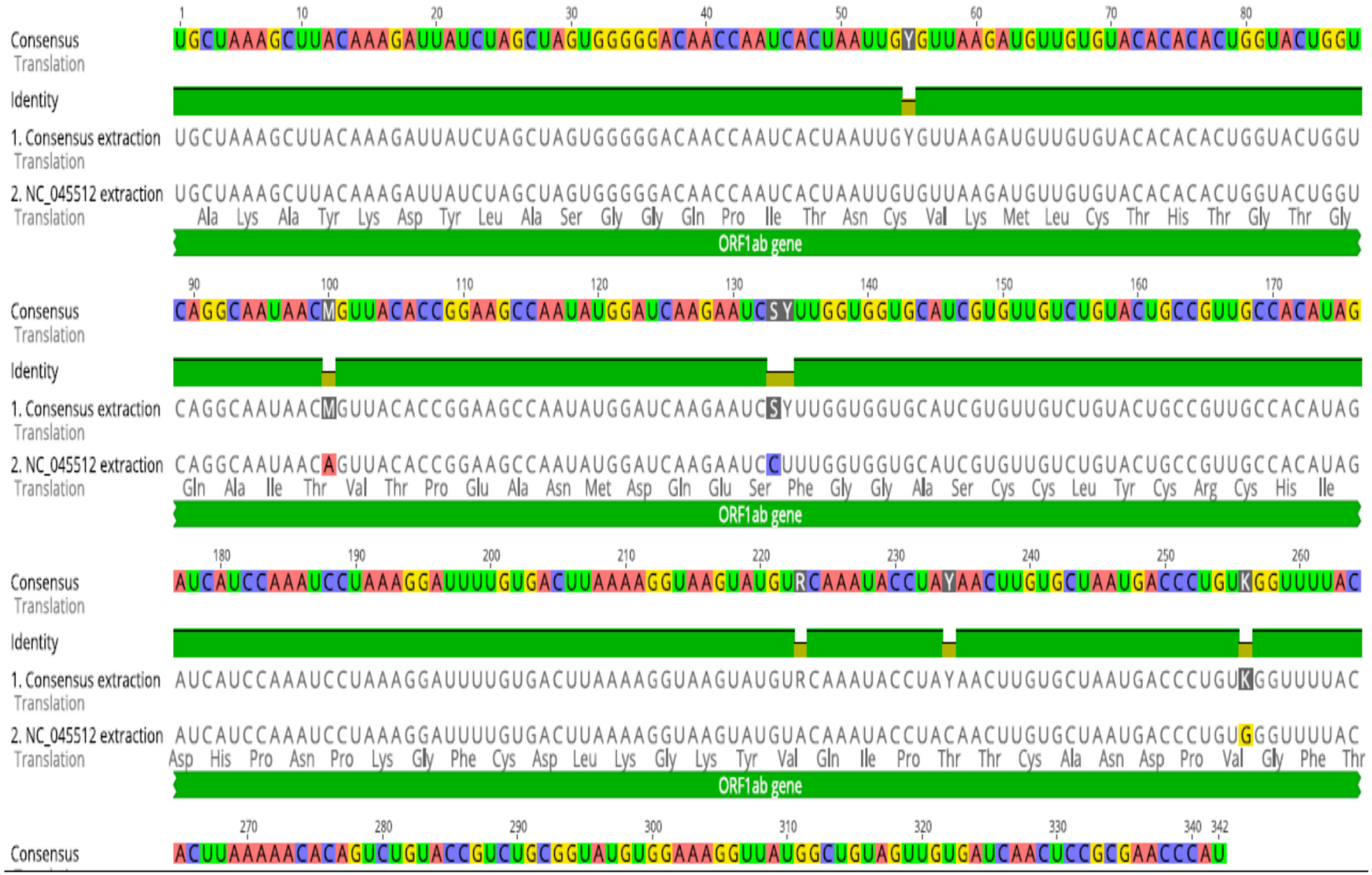
(**b**)

Figure 1. Map to reference whole genome alignment of SARS-CoV2 sequences representing different geographic regions. Strain MN988668 6305 was used in the Map to reference program. The default setting was applied except for the following criteria: consensus set to "100% identity" Full Seq Coverage = 29.89, Genenious Mapper with

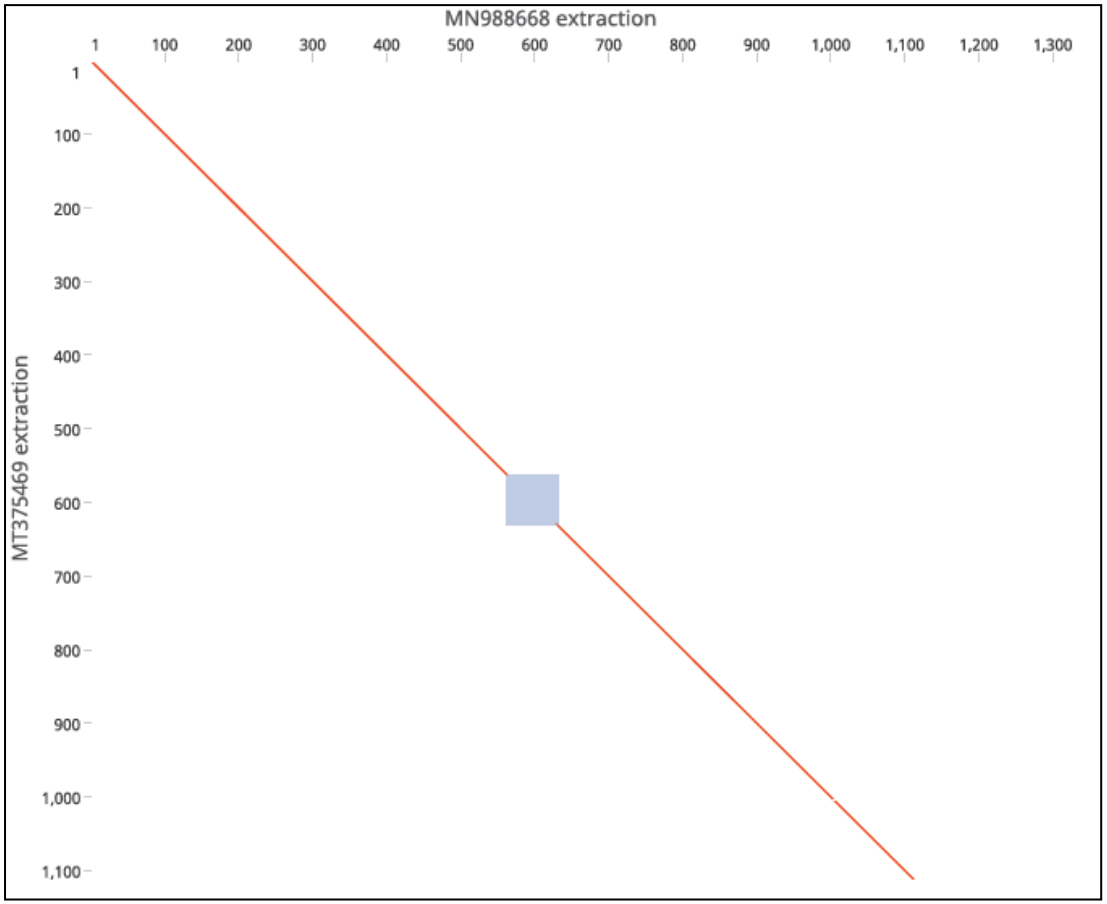
"High Sensitivity", five times "Fine Tuning" after alignment, saved "assembly report", and "Do not trim" before mapping. To minimize the image size, only visible regions were saved. (**a**) Highlight agreement to consensus sequence and to reference. (**b**) Highlight transversions to consensus sequence and to reference.



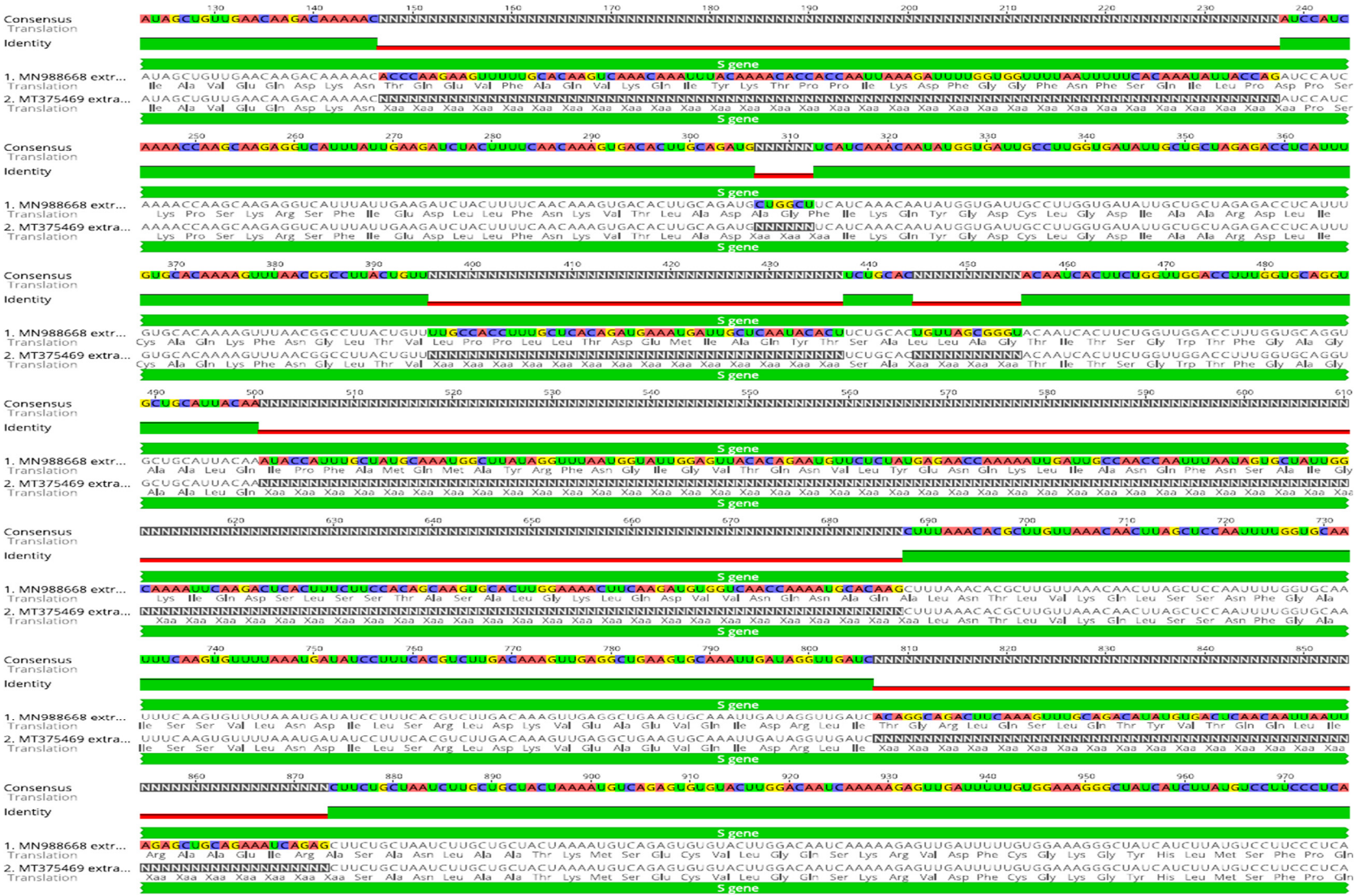
(**a**)


Consensus
Translation
Identity
1. Consensus extraction
Translation
2. NC_045512 extraction
Translation
Ala Lys Ala Tyr Lys Asp Tyr Leu Ala Ser Gly Gly Gln Pro Ile Thr Asn Cys Val Lys Met Leu Cys Thr His Thr Gly Thr Gly
ORF1ab gene
Gln Ala Ile Thr Val Thr Pro Glu Ala Asn Met Asp Gln Glu Ser Phe Gly Gly Ala Ser Cys Cys Leu Tyr Cys Arg Cys His Ile
ORF1ab gene
Asp His Pro Asn Pro Lys Gly Phe Cys Asp Leu Lys Gly Lys Tyr Val Gln Ile Pro Thr Thr Cys Ala Asn Asp Pro Val Gly Phe Thr
ORF1ab gene

(**b**)


MN988668 extraction
1
100
200
300
400
500
600
700
800
900
1,000
1,100
1,200
1,300
MT375469 extraction
1
100
200
300
400
500
600
700
800
900
1,000
1,100

(**c**)



(**d**)

Figure 2. (**a**) Whole genome alignment with guide tree of 174 SARS-CoV2 sequences showing transversions to references and consensus. Clustal Omega 1.2.2 in Geneious Bioinformatics Package using Clustal size set at 180 in mBed Algorithm and Guide Tree, validated by MUSCLE 3.8.425,and Geneious Alignment in Geneious primer 2020.2.3

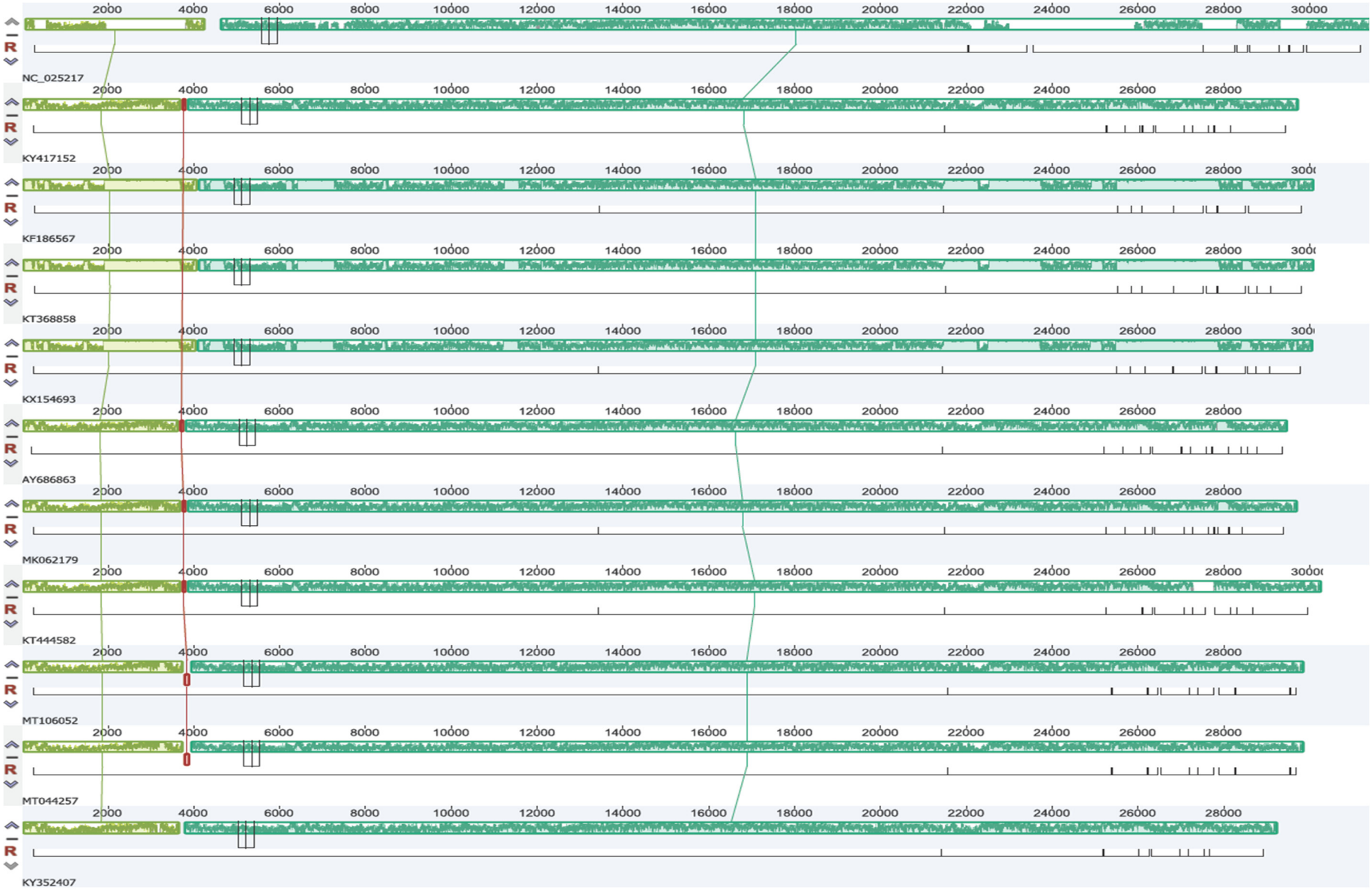
Java version 11.0.4+11 (64bit). Consensus sequence threshold was set at 100% and highlighted transversion to consensus. (**b**) Zoomed analysis of consensus sequence extraction containing regions of transversion substitutions in 342 bp region (13093 > 13434) on ORF1ab (extracted from the 92 most abundant genomes in the 6305 genomes analyzed. (**c**) Dot plot comparisons of the strain MT375469 against the reference MT988668 from the sequence alignment extractions (~1.1 and 1.3 kb) indicating differences at the S gene region towards the end of the genomes. (**d**) Strains MT375469 isolate (Connecticut CT-UW) a 431 bp region in the S gene containing segments of length polymorphisms against reference MS988668.

3.2. Analysis of Point Mutations Affecting Substitutions

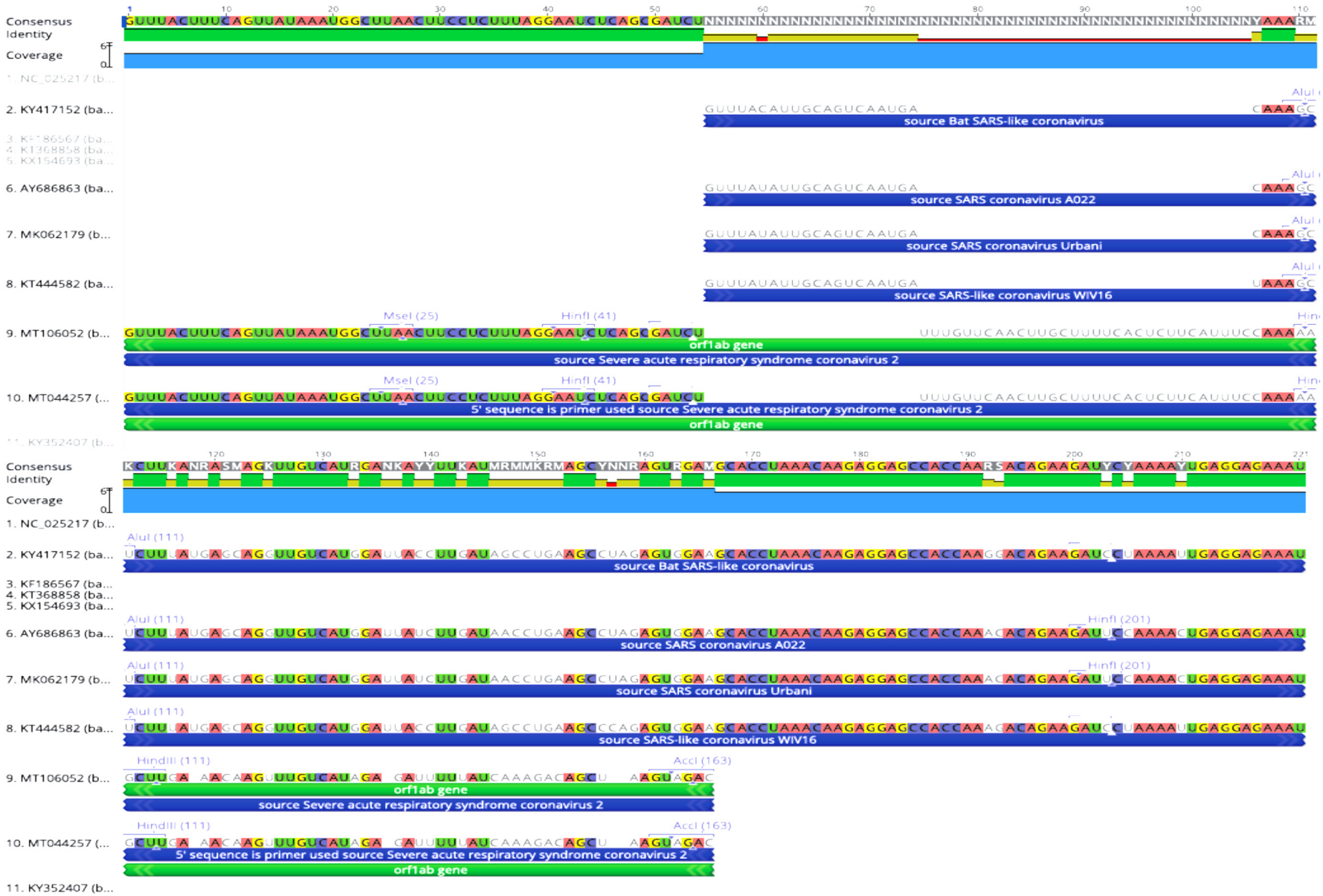
Vertical analysis of the locations and the extent of potential functional mutations affecting transversions were verified by the extraction and analysis of 92 events (black tracks on Figure 2a) showing most prominent transversion events in the 6305 genome alignment. An example of the analysis is shown in Figure 2b where aligned extractions at position (13,093 > 13,434,342 bp) on ORF1ab from 92 sequences mapped to reference genome (NC_045512), revealed seven events. These were four transitions (3 Y pyrimidine transitions affecting amino acids Ser, Cys, and Thr, and 1 R purine transition affecting amino acid Val), and three transversions M, S, K affecting amino acid Thr, Ser, and Val, respectively. In addition, the MT375469 isolate from Connecticut state (CT-UW) was the most distant amongst all genomes mapped to the reference isolate MN988668. Dot-plot view and cupped alignment analysis using the latter reference revealed that this isolate had deletions of 431 bp in S gene starting at the location 21,500 bp (Figure 2c,d). These differences and the fact that the MT375469 shared similarity with the Australian cluster might have potentially implied its introduction in the Connecticut state.

3.3. Whole Genome Comparative Genomics for Rearrangements in the SARS-CoV2 Genomes

All types of regulatory structures governing the genomic balance through rearrangements involving inversions, translocations, genome mapping, and deletions of larger blocks were examined using different programs including Mauve for whole genome analysis (Figure 3a). No extensive rearrangements were detected further indicating SARS-CoV2 as a highly clonal lineage that evolved recently, except for only two SARS-CoV2 isolates MT106052 (USA-CA7) and MT044257 (USA-IL2) against reference NC_025517 betacoronavirus isolate. A genomic inversion was detected at position 4000 affecting ORF1ab in these two isolates (Figure 3a). These regions are indicated by the off-alignment red boxes below the genome line at 4000bp. Examination of the Locally Collinear Blocks (LCBs) in MAUVE topology identified a segment inversion of 221 bp at position 4000 despite high genome conservation indicated by native color codes (Figure 3b). Dot-plot and nucleotide analysis of the region revealed abundance of point mutations at a unique hot-spot that is prone to genomic rearrangement (Figure 3c,d).


2000
4000
6000
8000
10000
12000
14000
16000
18000
20000
22000
24000
26000
28000
30000
NC_025217
KY417152
KF186567
KT368858
KX154693
AY686863
MK062179
KT444582
MT106052
MT044257
KY352407

a


Consensus
Identity
Coverage
1. NC_025217 (b...
2. KY417152 (ba...
3. KF186567 (ba...
4. KT368858 (ba...
5. KX154693 (ba...
6. AY686863 (ba...
7. MK062179 (b...
8. KT444582 (ba...
9. MT106052 (b...
10. MT044257 (...
11. KY352407 (b...
source Bat SARS-like coronavirus
source SARS coronavirus A022
source SARS coronavirus Urbani
source SARS-like coronavirus WIV16
MseI (25)
HinfI (41)
orf1ab gene
source Severe acute respiratory syndrome coronavirus 2
5' sequence is primer used source Severe acute respiratory syndrome coronavirus 2
AluI (111)
HinfI (201)
HindIII (111)
AccI (163)

b

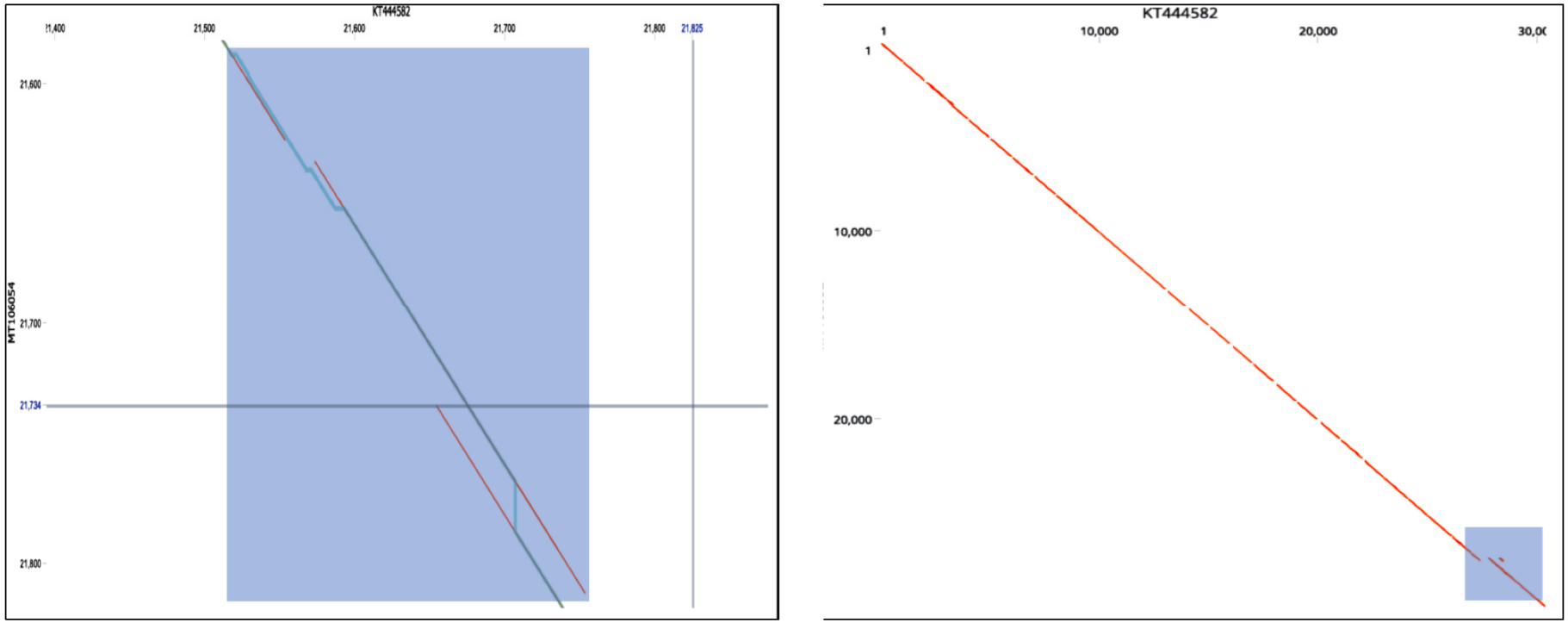
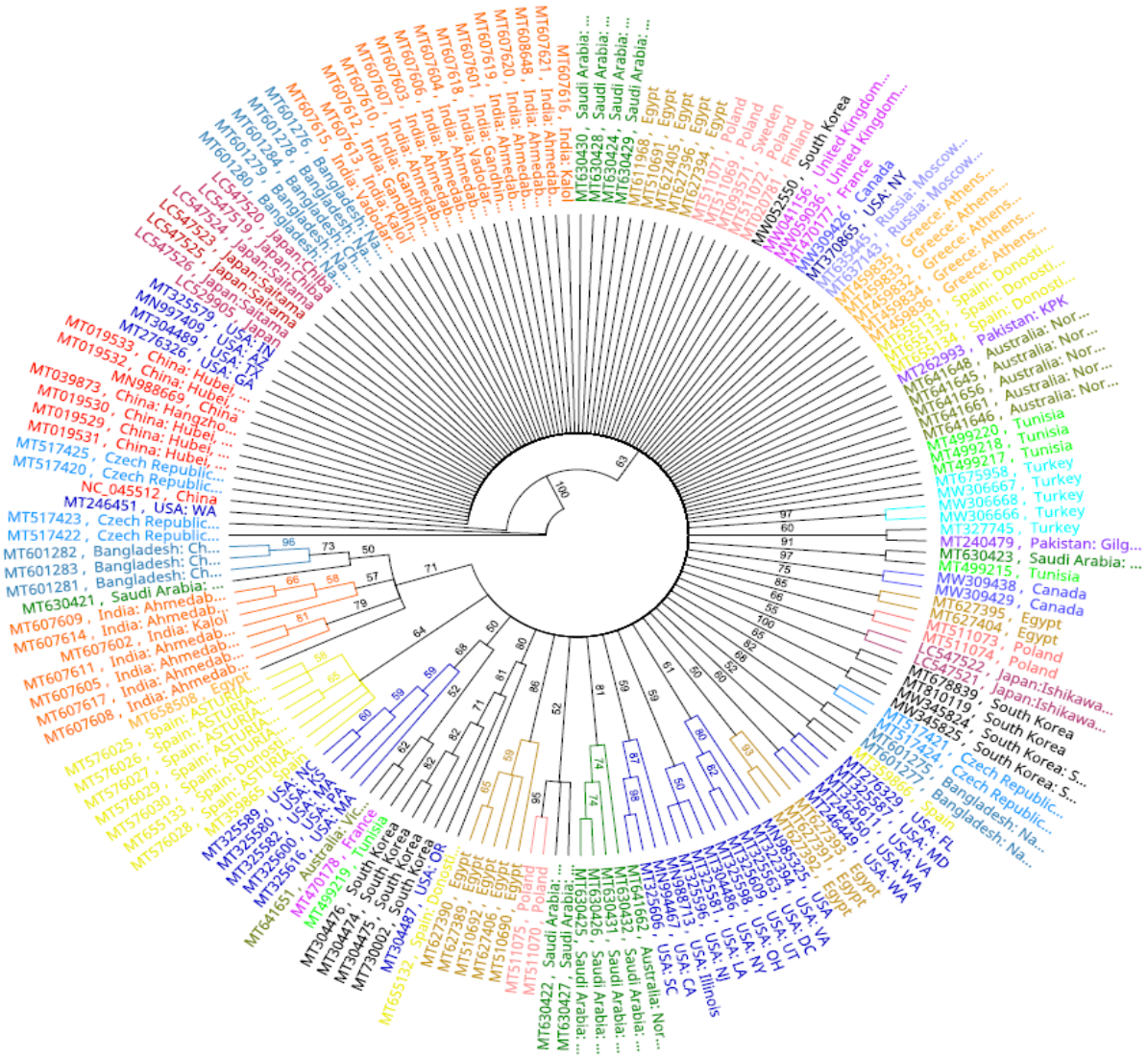


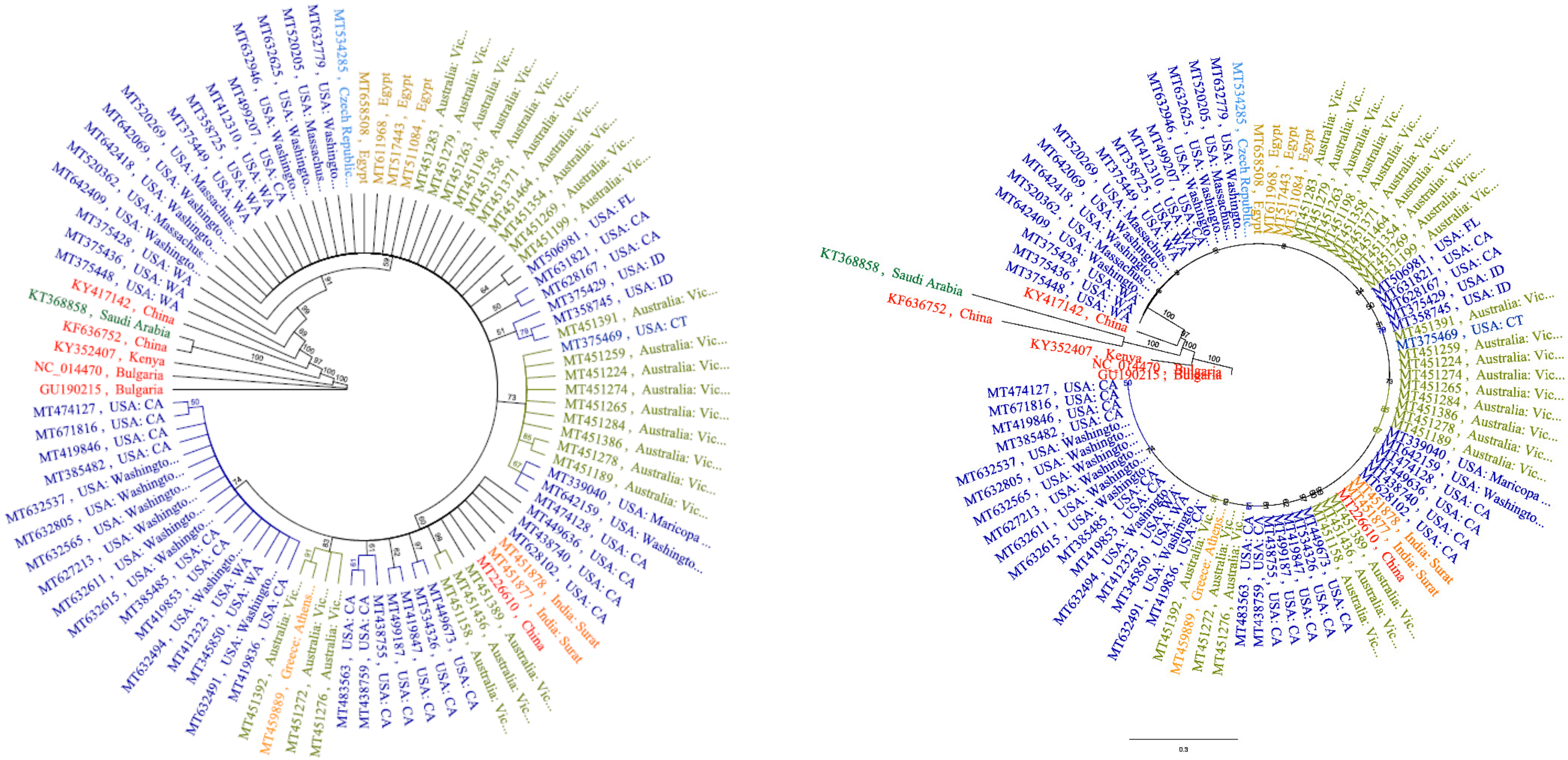
Figure 3. Whole genome comparative genomic analysis of SARS-CoV2 sequences. (**a**) ProgressiveMauve output showing inversion event in the SARS-CoV2 isolates MT106052 (USA-CA7) and MT044257 (USA-IL2) against the reference NC_025517 betacoronavirus isolate at a position [4000 bp] in ORF1ab of the two isolates (red colored box against mauve native color modes that is out of register below the genome line). (**b**) View of the Locally Collinear Blocks (LCBs) revealing a 221 bp inverted segment zoomed to reveal nucleotides and restriction sites. (**c**) Dotplot pairwise comparisons revealing a hot-spot for potential recombination-translocation event at the inverted locus.

3.4. Distant Similarities and Phylogenomics of SARS-CoV2

The genetic distance similarity with an overall scale of 0.2, as well as transformed cladogram depicting phylogenetic lines revealed a tree with two major types of clusters. As shown in Figure 4a, all of the SARS-CoV2 isolates were separated out as a single large cluster supported by 100% bootstrap value at the branch node. In this mega cluster, two major types of isolates were identified in different countries forming two distinct country-specific topologies of the tree. Nearly 50% of the 174 (85 genomes) formed independent lineages while the other half showed identical sub-clusters within the same country. The former contained only independent lineages with those from the same country more closely related, and the latter mainly contained sub-clusters of identical isolates also mostly from the same country. Attempt to construction a more concise phylogenetic tree with lesser number of genomes (92 genomes) from remote geographic regions that confirm country-specific sub-clusters, also yielded the same topology, except for the Connecticut isolate MT375469 that clustered in Australian cluster. Similarly, all SARS-CoV2 clustered in a single large branch out of the reference group with a higher percent of bootstrap support at the nodes (100%). Within this large SARS-CoV2 genomic cluster, five sub-cluster groups representing five different countries were formed. They included Saudi Arabia (green), India (orange), Egypt (red), Tunisia (blue), and single isolate from Turkey (Figure 4b,c).



(**a**)



(**b**) (**c**)

Figure 4. Distance similarity and phylogenetic analysis of SARS-CoV2 isolates. (**a**) Genome alignments with integrated guide-tree of 174 genomes sequences selected from 6305 genomes of SARS-CoV2 isolated at different countries. (**b**) The topology showing independent lineages and sub clonal populations at different geographic regions is a circular tree layout showing the "origins" of evolutionary lineages from Europe, Bulgaria (NC_014470; GU190215; and Chinese strain MT084071). Tree view of Clustal Omega 1.2.2 and Geneious Primer bioinformatics package Tree Builder using Tamura-Nei genetic distance model and Neighbor-Joining method of tree building transformed to cladogram (consensus method used for clustering was Majority Greedy clustering with (%) consensus support at nodes. Tree branches were cladogram based in a decreasing order with a line weight of 2 and tip labeled with standard strain names displaying country of isolation. (**b**,**c**) 92 genome sequence alignments.

3.5 Abundance of Signature Sequences for Adaptive Expression Strategies on the Clonal Genomes of SARS-CoV2

Intragenic coding tandem repeats are known to regulate adaptive pathogenic mechanisms in prokaryotes by serving as ON/OFF switches of transcription-translation machinery. Because they are in-frame for translation, additions or deletions of repeat units does not alter the protein, but cause length polymorphism that affects functionality while keeping genome stable. To test whether repeats may been used in the rapid emergence and adaptation of SARS-CoV2 under a clonal background, we examined repeat units and restriction sites on stably aligned genomes. Profiles of only one representative isolate was shown in a given alignment cluster, the rest of isolates were dimmed. Full repeat profiles and details are found in Additional Files (attached). As shown in Figure 5, the most common restriction sites were indicated while the majority were represented by downward blue color arrow on the genome sequences. Except for the references, which had their own pattern of restriction sites, all isolates had identical sites common to all genomes and cluster-specific sites shared only by isolates within given phylogenetic clusters (Additional Files, attached). Similarly, with the exception of a few, the patterns of tandem repeats on all genomes implied a high degree of genome clonality in the SARS-CoV2 lineage. However, while the reference genomes of betacoronaciruses shared most of the repeat patterns, they significantly differed in repeat contents, patterns, and locations. Reference strains shared repeat patterns identical to the SARS-CoV2 lineage isolates at positions: 3125, 4630, 8743, 9741, 11445, 13262, 132688, 4806, 17166, 20566, 20657, 21099, 21405 (trans), 24,333, 25728,26637,27403, and 28,709. Numerous repeat patterns on the references were either missing or identical to those found at different locations on the lineage isolates. The programs were set to detect only “Perfect Repeats”, “100% identity”, “No gaps” “high score”, “high index values”.

To understand the mechanism behind the success of SARS-CoV2 lineage in the rapid jump into humans and adaptation, despite high core genome conservation and RNA stability, we investigated regulatory signatures on primary sequences. The abundance of transversion, point- and frameshift-mutations, RNA modifications, and signature sequences play important roles in the adaptive expression of prokaryotic genomes. Abundance of five different types of regulatory sequences were identified on the genome of SARS-CoV2. These were in the order of their abundance on the genomes: Slippery 4: 5bp length track: AAGAA; Slippery 2: 7bp length track:TTTAAAA (motif UUUAAAA); Slippery 1: 7bp Length track: TTTAAAC (motif UUUAAAC); Slippery 3: 7bp Length track: TTTAAAT (motif UUUAAAU); and Slippery 5: 7 bp length track: TTTTTTA (motif UUUUUUA). Detailed profiles and structural profiles of these are found in additional Files (attached). Slippery 4 pattern was the most abundant and had a unique identical universal pattern in all genomes including the reference, unlike restriction sites and tandem repeats. However, the rest of the four slippery loci differed significantly in their distribution, number, and loci, except for eight that were common between the reference and all isolates. These included: Slippery 2 and 5 at 13540; Slippery 2 at 13539; Slippery 1 at 16747; Slippery 2 at 17727; Slippery 1 at 18552; Slippery 3 at 20894; and Slippery 2 at 25134. In this study, all of the SARS-CoV2 isolates were revealed as belonging to a single identical clone, except for the references and the pangolin isolate MT084071. Profiles of only one representative isolate was shown in any given cluster, the rest of the isolates were dimmed.

Figure 5. Annotations and signature sequence patterns on whole genome alignments with guide tree of 34 SARS-CoV2 genomes. Gene annotations and names (horizontal green color bars). 3'UTR, 5' UTR, and Stem loops. Slippery sequences (Brown arrows, direct and reverse): Slippery sequence 1: 7 nt (UUUAAC, TTTAAAC); slippery sequence 2: 7 nt (UUU-AAAA, TTTAAAA); Slippery sequence 3: 7 nt (UUUAAAU, TTTAAAT); Slippery sequence 4: 5 nt (AAGAA); Slippery sequence 5: 7 nt (UUUUUUA, TTTTTTA). Profiles of only one representative isolate was shown in a given cluster, the rest of isolates were dimmed. Tandem repeat tracks (Using Phobos Tandem repeat finder 3.3.11 2006-2010). The program was set to "Perfect Repeats" with unit length (1-500bp), "100% identity", "No gaps" "high score", "high index values" Each of the stacked red color arrows along genome sequence indicated a tandemly repeating unit track. Restriction sites: Geneious Prime. Commonly used enzymes were highlighted. Restriction site cleavage abilities were validated against methylation or variability, ambiguity, or deletion of targets. The most common restriction sites were indicated while the majority were represented by downward blue color arrow on the genome sequences. Clustal Omega 1.2.2 in Geneious Bioinformatics Package using Clustal size set at 180 in mBed Algorithm and Guide Tree (colored nodes), validated by MUSCLE 3.8.425, and Geneious Alignment in Geneious primer 2020.2.3 Java version 11.0.4+11 (64bit). Consensus sequence threshold was set at 100% and highlighted.

4. Discussion

It has been quite puzzling how SARS-CoV2 undergoes rapid paradigm shifts between clonal emergence and adaptational changes, despite the stable genome. To understand, we examined genomes sequences for genetic distance similarities, evolutionary phylogenomic, and *in-silico* functional predictions. In this study, selected whole genome sequence alignment of 6305 SARS-CoV2 from different countries yielded a high degree of nucleotide identity (99.99%) (Figure 1a). However, in spite of the high sequence homology, abundant transversions were detected throughout the genome sequences that slightly changed the topology (Figure 1b). These results were in agreement with earlier finding [39–41] implying that SARS-CoV2 lineage continues to have a clonal evolution and genome as it spreads throughout the world, despite identifications of recent point mutations (Mercatelli and Giorgi 2020; https://www.bmj.com/content/bmj/371/bmj.m4857.full.pdf), accessed on 12 February 2021]. Coronavirus core genome integrity is not affected by point mutations due to the protective proof-reading activity of the non-structural protein nsp14 [45,46]. Differential expression of genes as a means for adaptation while maintaining genome stability was the mechanisms used during decades of pandemic outbreaks by the community acquired *Staphylococcus aureus* [47,48].

In this study, vertical analysis of the transversion loci and potential functional mutations was verified by extracting and analyzing the 92 genomes with most transversion loci amongst the 6305 genomes as well as in 174 genomes from different geographic regions against the reference NC_045512 (Figure 2a). Interestingly, isolates within country-specific sub-clusters showed identical transversion patterns. In addition, all genomes maintained high level of sequence identity further supporting clonal evolution and genome. An example on abundance of point mutations is shown within a 342 bp region (13,093 > 13,434) in ORF1ab implying potential functional changes (Figure 2b). Interestingly, these were of two different types of substitutions (transversion and transitions) affecting the same amino acids in all instances tested. In this representative example (Figure 2b), four transitions (Ser, Cys, Thr, and Val) and three transversions (Thr, Ser, and Val) affected the same amino acids. Throughout the sequences, including S protein, M, S, K transversions and Y transitions were common. (Anahina for new paper on point mutations) Compensations have been experimentally confirmed to occur either for recovery [49,50] or to reduce deleterious effects [51–53]. Our findings remain to be verified experimentally as to whether they were compensatory or were induced by selective pressure. Deleterious or genomic rearrangements affecting sequence homology or clustering patterns were rare in this study, except for two novel events. Strains MT375469 isolate (Connecticut CT-UW) with a sum of deletions measuring up to 431 bp in the S gene (Figure 2c,d), was the only isolate with a significant distance mapped to the reference isolates (MN988668). It was also the only isolate clustered in Australian group, in contrast to country-specific grouping. The S protein was thought to be highly conserved among SARS-CoV2 isolates and thereby of a therapeutic importance and a vaccine candidate. The S protein length polymorphism is often helpful in dictating ancestral origins. For instance, clade 2 bat SARSr-CoV S protein which is universally present in many locations has a shorter size S protein due to deletions of 5, 12 or 13 amino acid residues [21]. While the overwhelming majority of genomes we analyzed had conserved regions, these findings might be of considerations in therapeutics applications [54]. Another novel inversion was detected in the SARS-CoV2 isolates MT106052 (USA-CA7) and MT044257 (USA-IL2) against the reference NC_025517 betacoronavirus isolate. ProgressiveMauve, other programs, and manual analysis revealed an inversion at position (~4000 bp) in ORF1ab of the two isolates (Figure 3a–c). Examination of Locally Collinear Blocks (LCBs) identified a 221 bp segment inversion despite high genome conservation indicated by native color codes (Figure 3a). Dot-plot and nucleotide analysis of the region revealed a unique hot-spot with abundance of point mutations prone to genomic rearrangement (Figure 3c). Finally, it has been found that coronaviruses maintain greater RNA structuredness across their genomes compared to MERS-

CoV and high nucleotide variability had no impact on RNA secondary structure stability which suggested selective pressures against its disruption [55].

Analysis of phylogenomic trees with cladogram and genetic distance similarity revealed identical tree topologies. All SARS-CoV2 isolates were separated as a single clone in a large cluster supported by 100% bootstrap value at the branch node. Within this mega cluster, SARS-CoV2 isolates from countries were of two categories; independent lineages and country-specific sub-clusters, potentially implying subclonal evolution (Figure 4a). Construction of a more concise tree with lesser number of genomes representing different geographic regions and different substitution patterns also yielded the same identical country-specific topology, except for isolate MT375469 that clustered with Australian cluster (Figure 4b,c). The recent finding of re-infection a patient by two SARS-CoV2 strains with significant genetic differences and varying disease severities, supported the notion of sub-clonal evolutions under clonal a background [56]. The latter study confirmed re-infection and ruled out in vivo evolution in the patient, owing to the greater genetic discordances between the two strains. This indicates that the variant, even though not evolved in the same patient, originated remotely supporting the country- or region-specific sub-clonal evolutions obtained in this study. It is possible that regional sub-clusters could have been due to multiple introductions. However, for this to be true, isolates from unique super-spreader region(s) or epicenters ought to be identified that shared in all local clusters. For instance, clues to multiple introductions events, before strict lockdowns were enforced, were verified by sharing epicenter isolates with Europeans [57]. However, in this study all groups mostly contained local isolates only. In addition, the possibility that all sub-clones simultaneous originated in the epicenter in Wuhan, China, is remote.

In this study, we report on in-silico evidences for adaptive overexpression strategies under a clonal SARS-CoV2 lineage background. We tested regulatory events that are prone to induce over-expression of viral proteins in SARS-CoV2 lineage. These included indels, recombination and translocations, restriction sites, AU-rich transcription regulating sequences, tandem repeats, signature sequences, and stem-loops (Figure 5). Except for the reference and two events described below, almost all genomes showed identical patterns in these criteria in addition to sub-cluster-specific patterns of each criteria shared within phylogenetic groups consistent with clonal evolution. This supports the notion that the remarkable coronavirus core-genome conservation is independent of replication and adaptive expressions in infectious variants. It has been experimentally verified that translocations of S, M, E, and N, viral proteins in murine coronaviruses mutants had no effect in genome stability and that mutants variants remained viable in cell cultures [58]. The inversion over the 5′ terminus region in ORF1ab gene, identified in this study, would indicated adaptive strategy rather than lethal mutation. We did not find large recombination events similar to that occurred in the two existing bat strains, WIV16 and Rf4092, giving rise to the civet SARS-CoV strain SZ3 [21]. In the aforementioned inversion detected in this study, sites for HindIII (A^{+}AGCUU) and AluI (AGCU) and Hinf1 (GAUUC) lied only a few stretches apart. However, a strong effect of *recA* independent homologous recombination event was reported to occur even at distantly separated loci [59]. Interesting, in contrast to stabilizing GC-rich regions, abundance of AT-rich tandem repeats and intervening slippery sequences were seen at the inversion locus (Figure 5). These included multiple copies of repeating units (ATC, AAG, AATGTC, AAAAGT, AAG, AAACAG, AAAG, ATCC). The full list of repeats is found in the Additional Supplementary Material Files (attached). Recent studies indicated that *recA*-independent recombination can be as efficient as *recA*-dependent in the absence of exonucleases activity [60]. Future analysis of more newly sequenced isolates and experimental validations would reveal more insights into the evolutionary adaptive mechanisms of SARS-CoV2. Regular updating and sequence analysis of SARS-CoV2 has become imperative as frequent evolution of mutants with particular signatures some of which is associated to patient mortality, is on the rise [61,62]. As the time of writing the manuscript, evolution of the following variants is being monitored; B.1.1.7 lineage originated in the UK, B.1.351 lineage in South Africa, and P.1

lineage in Brazil are being monitored, in addition to some new variants being reported in the USA at the time of writing the manuscript, (available at: https://www.cdc.gov/coronavirus/2019-ncov/more/science-and-research/scientific-brief-emerging-variants.html accessed on 2 February 2021).

Thus, in this study we present evidences for subtle adaptive strategies that occurred during the rapid evolution of SARS-CoV2 pandemic. Comparative genomics of whole genome alignments, tandem repeat screening, restriction site mapping, and pairwise genomic comparisons indicated recent evolution and clonality of SARS-CoV2 lineage. However, despite the high degree of core genome conservation and strong evidence for clonal emergence, limited diversifications were believed to occur during outbreak resulting in sub-clonal populations at different regions. Validation of an extremely rare event of adaptive rearrangement in SARS-CoV2 against remote betacoronavirus ancestor would provide details of evolutionary mechanisms. In addition, detection of an extremely rare but a substantial difference in the S protein due to potential deletions would be of therapeutic importance. More importantly, abundance of transversions, slippery sequences, and ON/OFF molecular structures implied subtle adaptive expressions had occurred in a clonal background. Future analysis of additional newly sequenced isolates and experimental validation of selected in-silico findings will provide more insights into the adaptive evolution of SARS-CoV2.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Table S1 Restriction sties and their sequence patterns on SARS-CoV2 genomes. Table S2 Slippery 1, (UUUAAC, TTTAAAC); sequence locations and patterns identified on SARS-CoV2 genomes, Table S3 Slippery 2, (UUUAAAA, TTTAAAA) sequence locations and patterns identified on SARS-CoV2 genomes. Table S4; Slippery 3, (UUUAAAU, TTTAAAT) sequence locations and patterns identified on SARS-CoV2 genomes. Table S5, Slippery 4, (AAGAA) sequence locations and patterns identified on SARS-CoV2 genomes. Table S6, Slippery 5, (UUUUUUA, TTTTTTA) sequence locations and patterns identified on SARS-CoV2 genomes

Author Contributions: K.B.S., A.A., A.F., F.A. (Fawwaz Alshammari), W.A., S.A., H.A., F.A. (Fahad Aldamadi), D.F.A., S.M., A.A.J, and A.B., are all made substantial contributions to the conception, design of the work, data acquisition, analysis, interpretation of data; optimization of software, deposition of genomes and bioinformatics analysis. They all have edited, drafted, revised, and approved the submitted version of the manuscript are fully aware and accountable for the work carried out as a team. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Deanship of Graduate Studies and Scientific Research at the University of Ha'il-Saudi Arabia through project number Covid RG-1917 to KB.

Institutional Review Board Statement: This study has been reviewed and approved by the Research Ethical Commiftee (REC at the Universlty of Hail dated: 18/82O2O and approved by university president letter number Nr. 55456/5/41 dated 29/12/ t44I H. on the project number H-2020-20. Extensive review is not applicable for this study does not involve animal or human cells.

Data Availability Statement: This section is excluded because only additional data submitted is available and no other data is available to report anywhere

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chan, J.F.; To, K.K.; Tse, H.; Jin, D.Y.; Yuen, K.Y. Interspecies transmission and emergence of novel viruses: Lessons from bats and birds. *Trends Microbiol.* **2013**, *21*, 544–555, doi:10.1016/j.tim.2013.05.005.
2. Zhong, N.S.; Zheng, B.J.; Li, Y.M.; Poon, Z.H.;Chan, K.H.; Li, P.H.; Tan, S.Y.; Chang, Q.; Xie, J.P.; Liu, X.Q.; et al. Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003. *Lancet* **2003**, *362*, 1353–1358.
3. Drosten, C.; Günther, S.; Preiser, W.; Van Der Werf, S.; Brodt, H.R.; Becker, S.; Rabenau, H.; Panning, M.; Kolesnikova, L.; Fouchier, R.A.M.; et al. Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome. *N. Engl. J. Med.* **2003**, *348*, 1967–1976, doi:10.1056/nejmoa030747.

4. Fouchier, R.A.M.; Kuiken, T.; Schutten, M.; van Amerongen, G.; van Doornum, G.J.J.; van den Hoogen, B.G.; Peiris, M.; Lim, W.; Stöhr, K.; Osterhaus, A.D.M.E. Aetiology Koch's postulates fulfilled for SARS virus. *Nature* **2003**, *423*, 240.
5. Ksiazek, T.G.; Erdman, D.; Goldsmith, C.S.; Zaki, S.R.; Peret, T.; Emery, S.; Tong, S.; Urbani, C.; Comer, J.A.; Lim, W.; et al. SARS Working Group. A novel coronavirus associated with severe acute respiratory syndrome. *N. Engl. J. Med.* **2003**, *348*, 1953–1966.
6. Zaki, A.; Van Boheemen, S.; Bestebroer, T.; Osterhaus, A.; Fouchier, R. Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia. *New Engl. J. Med.* **2012**, *367*, 1814–1820, doi:10.1056/nejmoa1211721.
7. Berger, A.; Drosten, H.W.; Stürmer, D.M.; Preiser, W. Severe Acute Respiratory Syndrome (SARS)—Paradigm of an Emerging Viral Infection. *J. Clin. Virol.* **2004**, *29*, 13–22.
8. Chan, J.F.W.; Lau, S.K.P.; To, K.K.W.; Cheng, V.C.C.; Woo, P.C.Y.; Yuen, K.-Y. Middle East Respiratory Syndrome Coronavirus: Another Zoonotic Betacoronavirus Causing SARS-Like Disease. *Clin. Microbiol. Rev.* **2015**, *28*, 465–522, doi:10.1128/cmr.00102-14.
9. Cheng, V.C.C.; Lau, S.K.P.; Woo, P.C.Y.; Yuen, K.Y. Severe Acute Respiratory Syndrome Coronavirus as an Agent of Emerging and Reemerging Infection. *Clin. Microbiol. Rev.* **2007**, *20*, 660–694, doi:10.1128/cmr.00023-07.
10. Guan, Y.J.; Zheng, B.J.; He, Y.Q.; Liu, X.L.; Zhuang, Z.X.; Cheung, C.L.; Luo, S.W.; Li, P.H.; Zhang, L.J.; Butt, K.M.; et al. Isolation and Characterization of Viruses Related to the SARS Coronavirus from Animals in Southern China. *Science* **2003**, *302*, 276–278, doi:10.1126/science.1087139.
11. Hemida, M.G.; Chu, D.K.; Poon, L.L.; Perera, R.A.; Alhammadi, M.A.; Ng, H.-Y.; Siu, L.Y.; Guan, Y.; Alnaeem, A.; Peiris, M. MERS Coronavirus in Dromedary Camel Herd, Saudi Arabia. *Emerg. Infect. Dis.* **2014**, *20*, 1231–1234, doi:10.3201/eid2007.140571.
12. Lau, S.K.P.; Woo, P.C.Y.; Li, K.S.M.; Huang, Y.; Tsoi, H.-W.; Wong, B.H.L.; Wong, S.S.Y.; Leung, S.-Y.; Chan, K.-H.; Yuen, K.-Y. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 14040–14045, doi:10.1073/pnas.0506735102.
13. Ge, X.-Y.; Li, J.-L.; Yang, X.-L.; Chmura, A.A.; Zhu, G.; Epstein, J.H.; Mazet, J.K.; Hu, B.; Zhang, W.; Peng, C.; et al. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. Nature **2013**, *503*, 535–538, doi:10.1038/nature12711.
14. Li, W.; Moore, M.J.; Vasilieva, N.; Sui, J.; Wong, S.K.; Berne, M.A.; Somasundaran, M.; Sullivan, J.L.; Luzuriaga, K.; Greenough, T.C.; et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **2003**, *426*, 450–454, doi:10.1038/nature02145.
15. Qian, Z.; Travanty, E.A.; Oko, L.; Edeen, K.; Berglund, A.; Wang, J.; Ito, Y.; Holmes, K.V.; Mason, R.J. Innate Immune Response of Human Alveolar Type II Cells Infected with Severe Acute Respiratory Syndrome–Coronavirus. *Am. J. Respir. Cell Mol. Biol.* **2013**, *48*, 742–748, doi:10.1165/rcmb.2012-0339oc.
16. Li, F.; Li, W.; Farzan, M.; Stephen, S.C. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* **2005**, *309*, 1864–1868.
17. Chinese, S.M.E.C. Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China. *Science* **2004**, *303*, 1666–1669.
18. Raj, V.S.; Mou, H.; Smits, S.L.; Dekkers, D.H.W.; Müller, M.A.; Dijkman, R.; Muth, D.; Demmers, J.A.A.; Zaki, A.; Fouchier, R.A.M.; et al. Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature* **2013**, *495*, 251–254, doi:10.1038/nature12005.
19. Lau, S.K.P.; Zhang, L.; Luk, H.K.H.; Xiong, L.; Peng, X.; Li, K.S.M.; He, X.; Zhao, P.S.-H.; Fan, R.Y.Y.; Wong, A.C.P.; et al. Receptor Usage of a Novel Bat Lineage C Betacoronavirus Reveals Evolution of Middle East Respiratory Syndrome-Related Coronavirus Spike Proteins for Human Dipeptidyl Peptidase 4 Binding. *J. Infect. Dis.* **2018**, *218*, 197–207, doi:10.1093/infdis/jiy018.
20. Li, W.; Zhang, C.; Sui, J.; Kuhn, J.H.; Moore, M.J.; Luo, S.; Wong, S.-K.; Huang, I.-C.; Xu, K.; Vasilieva, N.; et al. Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J.* **2005**, *24*, 1634–1643, doi:10.1038/sj.emboj.7600640.
21. Xiao-Shuang, Z.; Zeng, L.-P.; Yang, X.-L.; Ge, X.-Y.; Zhang, W.; Lin-Fa, W.; Xie, J.-Z.; Dong-Sheng, L.; Zhang, Y.-Z.; Wang, N.; et al. Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *PLOS Pathog.* **2017**, *13*, e1006698, doi:10.1371/journal.ppat.1006698.
22. Drexler, J.F.; Gloza-Rausch, F.; Glende, J.; Corman, V.M.; Muth, D.; Goettsche, M.; Seebens, A.; Niedrig, M.; Pfefferle, S.; Yordanov, S.; et al. Genomic Characterization of Severe Acute Respiratory Syndrome-Related Coronavirus in European Bats and Classification of Coronaviruses Based on Partial RNA-Dependent RNA Polymerase Gene Sequences. *J. Virol.* **2010**, *84*, 11336–11349, doi:10.1128/jvi.00650-10.
23. Craigen, W.J.; Cook, R.G.; Tate, W.P.; Caskey, C.T. Bacterial peptide chain release factors: Conserved primary structure and possible frameshift regulation of release factor 2. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 3616–3620, doi:10.1073/pnas.82.11.3616.
24. Kearse, M.G.; Wilusz, J.E. Non-AUG translation: A new start for protein synthesis in eukaryotes. *Genes Dev.* **2017**, *31*, 1717–1731, doi:10.1101/gad.305250.117.
25. Jaafar, Z.A.; Kieft, J.S. Viral RNA structure-based strategies to manipulate translation. *Nat. Rev. Genet.* **2018**, *17*, 110–123, doi:10.1038/s41579-018-0117-x.
26. Caliskan, N.; Wohlgemuth, I.; Korniy, N.; Pearson, M.; Peske, F.; Rodnina, M.V. Conditional Switch between Frameshifting Regimes upon Translation of dnaX mRNA. *Mol. Cell* **2017**, *66*, 558–567.e4, doi:10.1016/j.molcel.2017.04.023.
27. Fang, Y.; Treffers, E.E.; Li, Y.; Tas, A.; Sun, Z.; Van Der Meer, Y.; De Ru, A.H.; Van Veelen, P.A.; Atkins, J.F.; Snijder, E.J.; et al. Efficient -2 frameshifting by mammalian ribosomes to synthesize an additional arterivirus protein. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, E2920–E2928, doi:10.1073/pnas.1211145109.

28. Yan, S.; Wen, J.-D.; Bustamante, C.; Tinoco, I. Ribosome Excursions during mRNA Translocation Mediate Broad Branching of Frameshift Pathways. *Cell* **2015**, *160*, 870–881, doi:10.1016/j.cell.2015.02.003.
29. Ishimaru, D.; Plant, E.P.; Sims, A.C.; Yount, B.L.; Roth, B.M.; Eldho, N.V.; Pérez-Alvarado, G.C.; Armbruster, D.W.; Baric, R.S.; Dinman, J.D.; et al. RNA dimerization plays a role in ribosomal frameshifting of the SARS coronavirus. *Nucleic Acids Res.* **2012**, *41*, 2594–2608, doi:10.1093/nar/gks1361.
30. Plant, E.P.; Pérez-Alvarado, G.C.; Jacobs, J.L.; Mukhopadhyay, B.; Hennig, M.; Dinman, J.D. A Three-Stemmed mRNA Pseudoknot in the SARS Coronavirus Frameshift Signal. *PLoS Biol.* **2005**, *3*, e172, doi:10.1371/journal.pbio.0030172.
31. Su, M.C.; Chang, C.T.; Chu, C.H.; Tsai, C.H.; Chang, K.Y. An atypical RNA pseudoknot stimulator and an upstream attenuation signal for -1 ribosomal frameshifting of SARS coronavirus. *Nucleic Acids Res.* **2005**, *33*, 4265–4275.
32. Thiel, V.; Ivanov, K.A.; Putics, Á.; Hertzig, T.; Schelle, B.; Bayer, S.; Weißbrich, B.; Snijder, E.J.; Rabenau, H.; Doerr, H.W.; et al. Mechanisms and enzymes involved in SARS coronavirus genome expression. *J. Gen. Virol.* **2003**, *84*, 2305–2315, doi:10.1099/vir.0.19424-0.
33. Harger, J.W.; Meskauskas, A.; Dinman, J.D. An integrated model of programmed ribosomal frameshifting. *Trends Biochem. Sci.* **2002**, *27*, 448–454.
34. Kim, D.; Lee, J.-Y.; Yang, J.-S.; Kim, J.W.; Kim, V.N.; Chang, H. The Architecture of SARS-CoV-2 Transcriptome. *Cell* **2020**, *181*, 914–921.e10.
35. Lin, C.-M.; Saif, L.J.; Marthaler, D.; Wang, Q. Evolution, antigenicity and pathogenicity of global porcine epidemic diarrhea virus strains. *Virus Res.* **2016**, *226*, 20–39, doi:10.1016/j.virusres.2016.05.023.
36. Zhou, P.; Fan, H.; Lan, T.; Yáng, X.-L.; Shi, W.-F.; Zhang, W.; Zhu, Y.; Zhang, Y.-W.; Xie, Q.-M.; Mani, S.; et al. Fatal swine acute diarrhoea syndrome caused by an HKU2-related coronavirus of bat origin. *Nature* **2018**, *556*, 255–258, doi:10.1038/s41586-018-0010-9.
37. Wu, A.; Peng, Y.; Huang, B.; Ding, X.; Wang, X.; Niu, P.; Meng, J.; Zhu, Z.; Zhang, Z.; Wang, J.; et al. Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. *Cell Host Microbe* **2020**, *27*, 325–328, doi:10.1016/j.chom.2020.02.001.
38. Zhu, N.; Zhang, D.; Wang, W.; Li, X.; Yang, B.; Song, J.; Zhao, X.; Huang, B.; Shi, W.; Lu, R.; et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.* **2020**, *382*, 727–733, doi:10.1056/NEJMoa2001017.
39. Lu, R.; Zhao, X.; Li, J.; Niu, P.; Yang, B.; Wu, H.; Wang, W.; Song, H.; Huang, B.; Zhu, N.; et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: Implications for virus origins and receptor binding. *Lancet* **2020**, *395*, 565–574, doi:10.1016/s0140-6736(20)30251-8.
40. Zhou, P.; Yang, X.L.; Wang, X.G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H.R.; Zhu, Y.; Li, B.; Huang, C.L.; et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **2020**, *579*, 270–273.
41. Ren, L.-L.; Wang, Y.-M.; Wu, Z.-Q.; Xiang, Z.-C.; Guo, L.; Xu, T.; Jiang, Y.-Z.; Xiong, Y.; Li, Y.-J.; Li, X.-W.; et al. Identification of a novel coronavirus causing severe pneumonia in human: A descriptive study. *Chin. Med. J.* **2020**, *133*, 1015–1024, doi:10.1097/cm9.0000000000000722.
42. Chan, J.F.-W.; Yuan, S.; Kok, K.-H.; To, K.K.-W.; Chu, H.; Yang, J.; Xing, F.; Liu, J.; Yip, C.C.-Y.; Poon, R.W.-S.; et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: A study of a family cluster. *Lancet* **2020**, *395*, 514–523, doi:10.1016/s0140-6736(20)30154-9.
43. Tang, X.; Wu, C.; Li, X.; Song, Y.; Yao, X.; Wu, X.; Duan, Y.; Zhang, H.; Wang, Y.; Qian, Z.; et al. On the origin and continuing evolution of SARS-CoV-2. *Natl. Sci. Rev.* **2020**, *7*, 1012–1023, doi:10.1093/nsr/nwaa036.
44. Darling, A.C.E.; Mau, B.; Blattner, F.R.; Perna, N.T. Mauve: Multiple Alignment of Conserved Genomic Sequence with Rearrangements. *Genome Res.* **2004**, *14*, 1394–1403, doi:10.1101/gr.2289704.
45. Minskaia, E.; Hertzig, T.; Gorbalenya, A.E.; Campanacci, V.; Cambillau, C.; Canard, B.; Ziebuhr, J. Discovery of an RNA virus 3'->5' exoribonuclease that is critically involved in coronavirus RNA synthesis. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 5108–5113, doi:10.1073/pnas.0508200103.
46. Ferron, F.; Subissi, L.; De Morais, A.T.S.; Le, N.T.T.; Sevajol, M.; Gluais, L.; Decroly, E.; Vonrhein, C.; Bricogne, G.; Canard, B.; et al. Structural and molecular basis of mismatch correction and ribavirin excision from coronavirus RNA. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E162–E171, doi:10.1073/pnas.1718806115.
47. Kennedy, A.D.; Otto, M.; Braughton, K.R.; Whitney, A.R.; Chen, L.; Mathema, B.; Mediavilla, J.R.; Byrne, K.A.; Parkins, L.D.; Tenover, F.C.; et al. Epidemic community-associated methicillin-resistant Staphylococcus aureus: Recent clonal expansion and diversification. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 1327–1332, doi:10.1073/pnas.0710217105.
48. Li, M.; Diep, B.A.; Villaruz, A.E.; Braughton, K.R.; Jiang, X.; DeLeo, F.R.; Chambers, H.F.; Lu, Y.; Otto, M. Evolution of virulence in epidemic community-associated methicillin-resistant Staphylococcus aureus. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 5883–5888, doi:10.1073/pnas.0900743106.
49. Hoffman, N.G.; Schiffer, C.A.; Swanstrom, R. Covariation of amino acid positions in HIV-1 protease. *Virology* **2005**, *331*, 206–207.
50. Knies, J.L.; Dang, K.K.; Vision, T.J.; Hoffman, N.G.; Swanstrom, R.; Burch, C.L. Compensatory Evolution in RNA Secondary Structures Increases Substitution Rate Variation among Sites. *Mol. Biol. Evol.* **2008**, *25*, 1778–1787, doi:10.1093/molbev/msn130.
51. Poon, A.; Chao, L. The rate of compensatory mutation in the DNA bacteriophage φX174. *Genetics* **2005**, *170*, 989–999.
52. Poon, A.; Otto, S.P. Compensating for our load of mutations: Freezing the meltdown of small populations. *Evolution* **2000**, *54*, 1467–1479.

53. Burch, C.; Chao, L. Evolution by small steps and rugged landscapes in the RNA virus ϕ6. *Genetics* **1999**, *151*, 921–927.
54. Huang, Y.; Yang, C.; Xu, X.-F.; Xu, W.; Liu, S-W. Structural and functional properties of SARS-CoV-2 spike protein: Potential antivirus drug development for COVID-19. *Acta Pharmacol. Sin.* **2020**, *41*, 1141–1149.
55. Sanders, W.; Fritch, E.J.; Madden, E.A.; Graham, R.L.; Vincent, H.A.; Heise, M.T.; Ralph, S.; Baric, R.S.; Moorman, N.J. Comparative analysis of coronavirus genomic RNA structure reveals conservation in SARS-like coronaviruses. *BioRxiv* **2020**, doi:10.1101/2020.06.15.153197.
56. Tillett, R.L.; Sevinsky, J.R.; Hartley, P.D.; Kerwin, H.; Crawford, N.; Gorzalski, A.; Laverdure, C.; Verma, S.C.; Rossetto, C.C.; Jackson, D.; et al. Genomic evidence for reinfection with SARS-CoV-2: A case study. *Lancet Infect. Dis.* **2020**, *21*, 52–58.
57. Stefanelli, P.; Faggioni, G.; Presti, A.L.; Fiore, S.; Marchi, A.; Benedetti, E.; Fabiani, C.; Anselmo, A.; Ciammaruconi, A.; Fortunato, A.; et al. Whole genome and phylogenetic analysis of two SARS-CoV-2 strains isolated in Italy in January and February 2020: Additional clues on multiple introductions and further circulation in Europe. *Eurosurveillance* **2020**, *25*, 2000305, doi:10.2807/1560-7917.es.2020.25.13.2000305.
58. De Haan, C.A.M.; Volders, H.; Koetzner, C.A.; Masters, P.S.; Rottier, P.J.M. Coronaviruses Maintain Viability despite Dramatic Rearrangements of the Strictly Conserved Genome Organization. *J. Virol.* **2002**, *76*, 12491–12502, doi:10.1128/jvi.76.24.12491-12502.2002.
59. Lovett, S.T.; Gluckman, T.J.; Simon, P.J.; Sutera, V.A., Jr.; Drapkin, P.T. Recombination between repeats in Escherichia coli by a recA-independent, proximity-sensitive mechanism. *Mol. Gen. Genet. MGG* **1994**, *245*, 294–300.
60. Dutra, B.E.; Sutera, V.A., Jr.; Lovett, S.T. RecA-independent recombination is efficient but limited by exonucleases *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 216–221.
61. Troyano-Hernáez, P.; Reinosa, R.; Holguín, Á. Evolution of SARS-CoV-2 Envelope, Membrane, Nucleocapsid, and Spike Structural Proteins from the Beginning of the Pandemic to September 2020: A Global and Regional Approach by Epidemiological Week. *Viruses* **2021**, *13*, 243, doi:10.3390/v13020243.
62. Dumonteil, E.; Fusco, D.; Drouin, A.; Herrera, C. Genomic Signatures of SARS-CoV-2 Associated with Patient Mortality. *Viruses* **2021**, *13*, 227, doi:10.3390/v13020227.