JIDC | THE JOURNAL OF INFECTION IN DEVELOPING COUNTRIES

Coronavirus Pandemic

Phylogeography and genomic analysis of SARS-CoV-2 delta variant in Morocco

Safae El Mazouri[1 #], Houda Bendani[1 #], Nasma Boumajdi[1 #], Tarik Aanniz[1], Ilham Kandoussi[1], Saaïd Amzazi[2], Lahcen Belyamani[3], Azeddine Ibrahimi[1], Mouna Ouadghiri[1]

[1] Laboratory of Biotechnology, Medical and Pharmacy School, Mohammed V University, Rabat, Morocco
[2] Laboratory of Human Pathologies Biology, Faculty of Sciences, Mohammed V University, Rabat, Morocco
[3] Emergency Department, Military Hospital Mohammed V, Rabat Medical and Pharmacy School, Mohammed V University, Rabat, Morocco

# Authors contributed equally to this work.

Abstract

Introduction: Since the COVID-19 pandemic began in December 2019, the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has continuously evolved with many variants of concern emerging across the world.

Methodology: In order to monitor the evolution of these variants in Morocco, we analyzed a total of 2130 genomes of the delta variant circulating around the world. We also included 164 Moroccan delta variant sequences in our analysis.

Results: Our findings suggest at least four introductions from multiple international sources and a rise of a dominant delta sub-lineage AY.33 in Morocco. Moreover, we report three mutations in the N-terminal domain of the S protein specific to the Moroccan AY.33 isolates, T29A, T250I and T299I. The effect of these mutations on the secondary structure and the dynamic behavior of the S protein N-terminal domain was further determined.

Conclusions: We conclude that these mutations might have functional consequences on the S protein of SARS-CoV-2.

Key words: SARS-CoV-2; delta variant; AY.33; mutation analysis; spike; phylogeography.

J Infect Dev Ctries 2022; 16(8):1258-1268. doi:10.3855/jidc.16513

(Received 24 February 2022 – Accepted 09 June 2022)

Introduction

COVID-19, caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus, first appeared in the Wuhan province in China in 2019. Since then, the ongoing pandemic has prompted the international community to take measures to stop the spreading of the virus. The symptoms of COVID-19 are similar to other viral upper respiratory illnesses and include fever, cough, fatigue and dyspnea [1]. SARS-CoV-2 genomic sequence analysis can be used to determine the origin and transmission patterns of the virus after it enters a new population, and is proving to be a key tool in developing pandemic management decisions.

Morocco reported its first COVID-19 case on the 2nd of March 2020. New positive cases and deaths from COVID-19 were reported a few days later. The number of infected cases had increased to approximately 951,482 and 14,796 deaths (http://www.covidmaroc.ma) by mid-December 2021. A Moroccan genomic monitoring network composed of several public and private laboratories was constructed in response to the COVID-19 pandemic to track the changes in the SARS-CoV-2 mutation profile in order to identify keystone events in the evolution of the virus and assess the outcomes of COVID-19 prevention measures taken by the Moroccan government.

Since the identification of the first complete genome of the virus, whole genomes of multiple strains have been deposited in public databases such as GISAID [2] and GENBANK [3]. Many tools have been developed to monitor the evolution of SARS-CoV-2. The most commonly used tool is the Phylogenetic Assignment of Named Global Outbreak Lineages (PANGOLIN) software [4] that assigns lineage to a given SARS-CoV-2 sequence consistently with the Pango dynamic lineage nomenclature scheme. Nextclade is an interesting project developed under the Nextstrain SARS-CoV-2 resources [5]. It is an open-source project for viral genome alignment, mutation calling, clade assignment, quality checks, and phylogenetic placement.

Since the beginning of the COVID-19 pandemic, genomic analysis of Moroccan samples has revealed the existence of variants of concern (VOCs). The emergence of these variants has been accompanied by waves of infection; the first wave was associated with the dominance of the alpha variant (B.1.1.7) in early February 2021 [6], and the second wave was associated with the appearance of the delta variant (B.1.617.2) in June 2021.

These variants have gained evolutionary advantages which include high virulence and increased infectivity due to multiple spike glycoprotein mutations, in particular, the D614G mutation that has shown the most efficient interaction with the Angiotensin-converting enzyme 2 (ACE2) receptor. This mutation changes the conformation of the receptor-binding domain, cleavage patterns of S-glycoprotein and replication fitness of SARS-CoV-2 variants [7]. Following the emergence of D614G, another mutation was identified in the receptor-binding motif (RBM), N439K, that enhances the ACE2-binding affinity and reduces the neutralizing activity of some monoclonal and polyclonal antibodies present in sera of individuals who have recovered from infection [8]. Another RBM mutation, Y453F, was associated with increased ACE2-binding affinity received considerable attention following its identification in sequences associated with infections in humans and mink in Denmark [9]. This pattern suggests that the virus accumulates mutations in order to reduce or avoid recognition by antibodies while maintaining or increasing binding to ACE2 [10].

The mutation profile of the delta variant provided in the outbreak.info project describes the mutations in at least 75% of the delta sequences. Overall, delta variant contains multiple mutations including 7 in the spike region, 10 in ORF1ab, 3 in the N, and 2 in ORF8 (www.outbreak.info) [11]. The spike gene mutations in B.1.617.2 variant are T19R, L452R, T478K, D614G, P681R, and D960N, with deletions at positions 157 and 158 [12]. Those mutations are suspected to increase the transmissibility which is in accordance with the high reproductive value (R0 = 5.08) of the delta variant [13].

In this study, we analyzed all Moroccan delta genomes available in GISAID and compared them to delta genomes from all over the world in order to identify the origin of the delta variant in Morocco and track its evolution.

Methodology

Genomic analysis

Full genome nucleotide sequences of SARS-CoV-2 delta variant deposited between July 2021 and November 2021, were retrieved from GISAID repository and grouped into two sets. The first set contained all Moroccan sequences (164 sequences in total out of which 48 were sequenced in the Laboratory of Biotechnology). The other set contained 1966 complete delta sequences deposited by multiple countries: Brazil, USA, Chile, Denmark, Australia, Italy, Canada, France, South Africa, Ghana, Germany, India, Indonesia, Japan, South Korea, Thailand, Wales, England and Scotland. Our final dataset contained 2130 delta genomes.

Variant calling

Sequences of SARS-CoV-2 delta variant were retrieved from GISAID repository and mapped to the reference sequence MN908947.3 using Minimap [14]. The BAM files were sorted by SAMtools [15], and then used to call the genetic variants in variant call format (VCF) by SAMtools mpileup and bcftools. The files were then annotated and their impact was predicted using SnpEff [16].

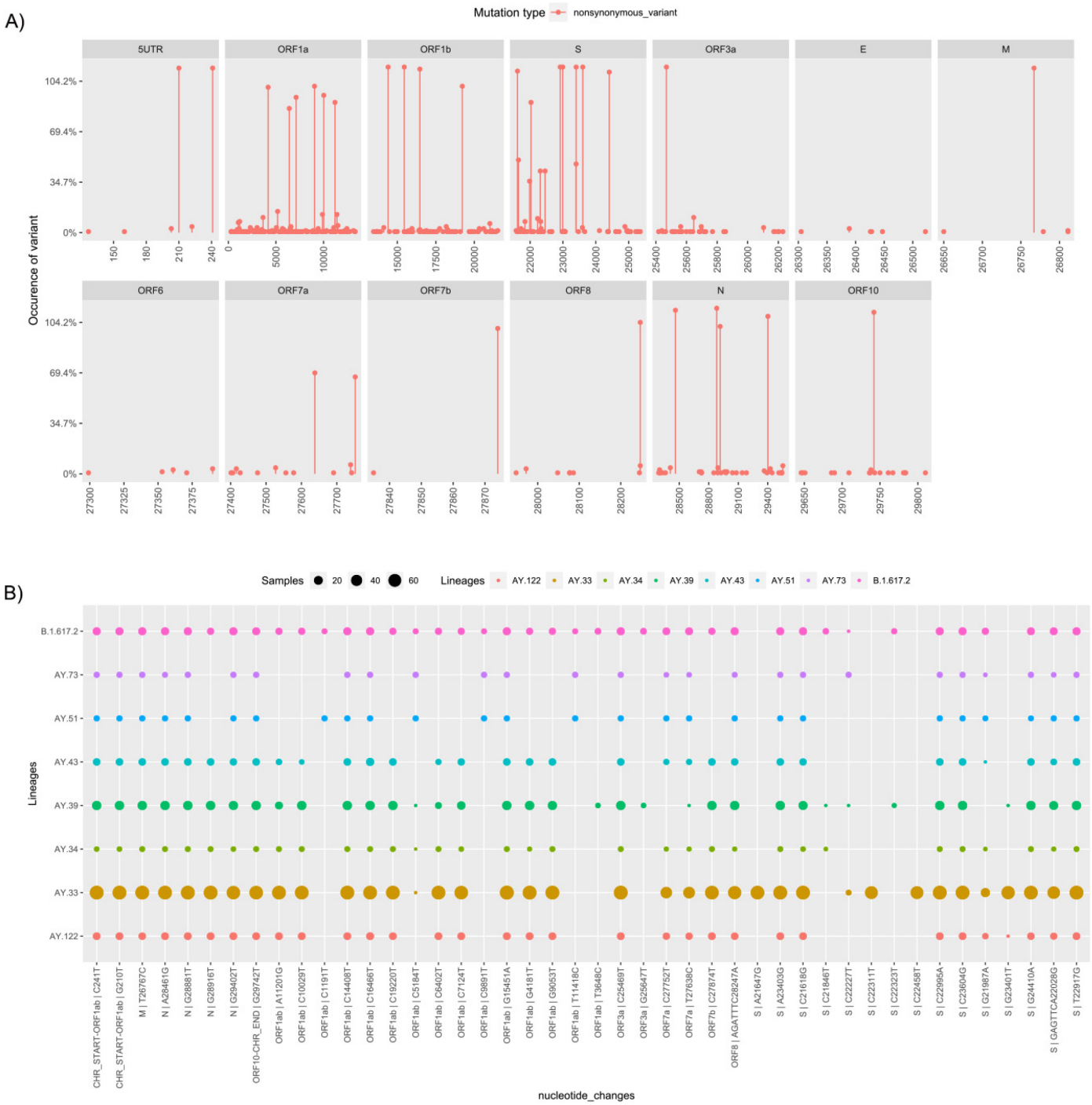
Global alignment and phylogenetic analysis

In order to build a global phylogeny of delta genomes, a total of 2130 sequences were included in the final alignment. MAFFT (v7.487) [17] was used to align the sequences with MN908947.3 as the reference sequence. The maximum likelihood tree was constructed using IQTREE [18]; based on the performance of IQTREE model finder, GTR+F+I was selected as the best substitution model for our dataset. The phylogenetic results were graphically visualized using Figtree (v1.4.4) (http://tree.bio.ed.ac.uk/software/figtree/).

Ancestral reconstruction

Ancestral reconstruction was performed using PastML [19]. The program takes as input a rooted tree and/or dated tree (in Newick format) and an annotation table containing the state of each tree tip. At first, we annotated our dataset with the date of collection and the country from where samples were collected. Then, we used the Least Square Dating (LSD) for ancestral events dating and for rooting the resulting phylogenetic tree based on dates [20]. Maximum likelihood marginal Posterior Probabilities Approximation (MPPA) was used as the prediction method with Felsenstein 1981 (F81) as a model. Default values were used for the reaming parameters. The PastML generated full tree was visualized and edited using Itol [21].

Figure 1. Nucleotide analysis of the 164 Moroccan samples.



A. The distribution of mutations found in the 164 Moroccan samples according to their nucleotide position. **B.** Occurrence of nucleotide mutations per lineage with presence in more than 6% of sequenced samples (n = 164). The color of the dots represents the lineage and the size of the dots represents the number of the samples showing the presence of the nucleotide mutation.

Mutation effect prediction

We used Chou & Fasman Secondary Structure Prediction Server (CFSSP) to predict the secondary structure of SARS-CoV-2 spike N-terminal domain (NTD) [22]. To investigate the effect of the mutations on this domain's structural conformation, its molecular stability and flexibility, we used DynaMut software (University of Melbourne, Australia) [23]. We first downloaded the crystallographic structure of neutralizing antibody 2-51 in complex with SARS-CoV-2 spike N-terminal domain (NTD) from RCSB (PDB ID: 7L2C). Next, we removed the antibody from the structure and prepared a mutant version of the NTD structure using PyMOL. Finally, the SARS-CoV-2 spike N-terminal domain structure was uploaded onto DynaMut software and the effect of mutations in various protein structure stability parameters was determined.

Results

Mutation profile of delta sequences in Morocco

Analysis of the 164 Moroccan sequences of SARS-CoV-2 delta variant revealed a total of 6403 mutations of which 872 (13%) were synonymous mutations. We were only interested in the 5531 (87%) non-synonymous mutations; we extracted all distinct non-synonymous mutations found in at least one sequence and obtained 418 mutations. Overall, 72 (17%) substitution mutations were found in the spike gene (61 missense, 8 frameshifts, and 3 disruptive in-frame deletions), 32 (7%) missense mutations were observed in the nucleocapsid (N) gene, 6 in the envelope (E) gene and 20 in the NSP12.

For a more specific analysis, we extracted the frequency of each mutation to get an idea on the common mutations between Moroccan sequences (Supplementary Table 1). Only 19 mutations were classified as common in 90% of the sequences, and 347 were rare mutations with an occurrence rate of 1-2% as shown in Figure 1A. From the 19 common mutations, six were observed in the S gene (C21618G/T19R, T22917G/L452R, C22995A/T478K, A23403G/D614G, C23604G/P681R, G24410A/D950N), 3 in the N protein (A28461G/D63G, G28881T/R203M, G29402T/D377Y), 2 in the NSP12 (C14408T/P4715L, G15451A/G5063S) and 2 in the 5'UTR (G210T, C241T), 1 in the NSP13 (C16466TT/P5401L), in M (T26767C/I82T), in ORF10 (G29742T) and in ORF3a (C25469T/S26L). In addition, we found 2 deletions, one in the S gene 156/158 and the other in the ORF8 gene 119/120 in 79% and 91% of Moroccan sequences respectively.

Dominance of AY.33 delta sublineage in Morocco

The variants from the Moroccan sequences were clustered using the Pangolin web service in 14 SARS-CoV-2 delta sublineages. AY.33 was the dominant sublineage (74 sequences) followed by AY.39 (24 sequences) and B.1.617.2 (15 sequences) (Supplementary Table 2, Supplementary Figure 1).

To understand the abundance of the delta variant sublineages in Morocco, we plotted their cumulative count against sample collection date (Supplementary Figure 2). We observed that the variant B.1.617.2 first appeared in Morocco in June 2021, delta sublineages quickly appeared in the following months. After July 2021, the majority of samples were delta sublineages, specifically AY.33, which was the dominant sublineage by the end of July 2021.

We observed a total of 418 unique mutation events from the 164 SARS-CoV-2 isolates in Morocco and we plotted the occurrence of nucleotide mutations (frequency > 6%) per lineage (Figure 1B). The AY.33 sublineage had the highest number of mutations in the S gene (13 mutations) whereas the average number of mutations in the other sublineages was 10. We also observed 6 distinct mutations in this sublineage out of which three were missense mutations; T29A, T250I and T299I were found in the spike gene and were present in nearly 45% of the sequences. Two synonymous mutations were found in ORF1ab, P748P belonging to AY.33 and AY.73 and L1356L belonging to AY.33 and AY.39.

The B.1.617.2 and AY.39, had nearly similar mutation profiles (82%), with an occurrence of 9% and 15% respectively. There were approximately 7 distinct mutations between B.1.617.2 and AY.39 occurring mostly in the S and ORF1ab genes.

Moroccan specific AY.33 mutations, T29A, T250I and T299I, cause alteration in secondary structure of SARS-CoV-2 spike N-terminal domain

The analysis of the AY.33 Moroccan sequences revealed three unique mutations in SARS-CoV-2 spike N-terminal domain, T29A, T250I and T299I. We studied their effect on the secondary structure of the protein to investigate the impact of these mutations. Our data revealed that the mutation T29A caused a change in the secondary structure as shown in Figure 2. The threonine amino acid at position 29 was substituted by alanine. Our analysis showed that there was a loss of sheet at positions 28, 29 and 30, and a replacement of

sheet by coils at position 28 and 29 due to the mutation T29A (Figure 2A, compare i and ii). Similarly, in T250I threonine was substituted by isoleucine. Due to this substitution, there was a loss of turn at position 251 and its replacement by a coil (Figure 2B, compare i and ii). Further, the substitution of threonine by isoleucine at position 299 resulted in considerable changes in the secondary structure (Figure 2C, compare i and ii). The detailed analysis revealed that there was a replacement of coils by helix at positions 294, 295, 296, 297 and 299, while there was a replacement of turn by helix at position 298. Altogether, the substitution of threonine to alanine at position 29 and to isoleucine at positions 250 and 299 in the mutant spike N-terminal domain resulted in change in the secondary structure of the protein which may have had functional consequences.

T29A alters the stability dynamics of tertiary structure of SARS-CoV-2 spike N-terminal domain

To understand the impact of mutations on the tertiary structure of SARS-CoV-2 spike N-terminal domain, we used DynaMut to get information about the alteration in protein stability and flexibility due to each mutation in the native protein structure. Our data revealed that there was a change in vibrational entropy energy (ΔΔSVibENCoM) and free energy differences (ΔΔG) between the wild type and the mutant isolate (Table 1). Vibrational entropy represents an average of the configurational entropies of the protein within a single minimum of the energy landscape [24]. The negative ΔΔSVibENCoM of mutant spike N-terminal domain represents the rigidification of the protein structure and positive ΔΔSVibENCoM represents gain in flexibility. Here, our data showed that the mutations T29A, T250I and T299I lead to an increase of molecule flexibility. Further, we also calculated the free energy differences, ΔΔG, between wild-type and mutant. The ΔΔG caused by mutation had been correlated with the structural changes, such as changes in packing density, cavity volume and accessible surface area and therefore, it measures the effect of the mutation on protein stability [25]. In general, a ΔΔG below zero means that the mutation causes destabilization and above zero represents protein stabilization. Here, our analysis showed positive ΔΔG for T250I and T299I suggesting that these mutations were stabilizing the

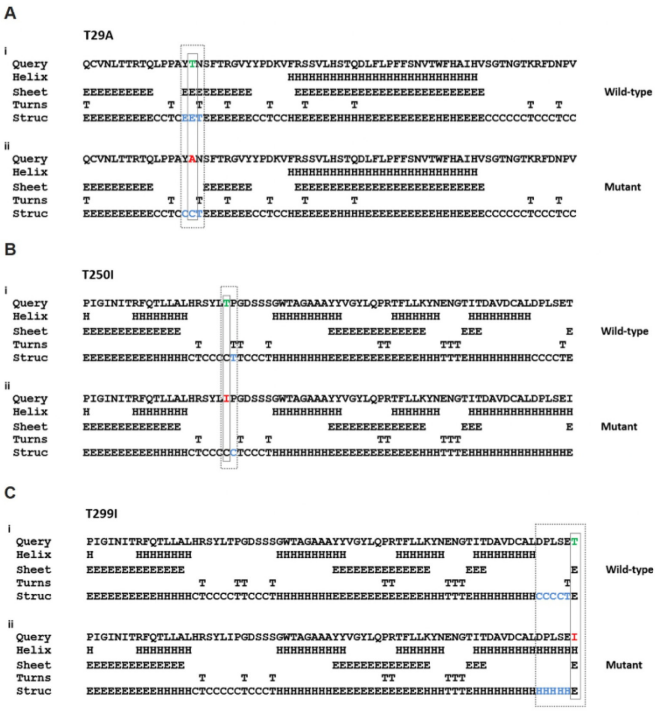


Figure 2. Prediction of secondary structure of SARS-CoV-2 spike N-terminal domain. Effect of mutations on secondary structure of the protein.

(**A–C**) demonstrate three mutations observed in Moroccan isolates; (i) represents the sequence of the Wuhan isolate and (ii) represents a sequence of Moroccan isolates. The small rectangular box shows the mutant residue. The difference of secondary structure between Wuhan and Moroccan isolates are highlighted with the position of the dashed box in respective panels.

mutant structure as compared to the wild-type; however, we observed negative ΔΔG for T29A mutation indicating its destabilizing behavior (Supplementary Figure 3). The ΔΔG for T299I mutant was 0.150 kcal/mol which was significantly higher than others. Accordingly, we closely analyzed the changes in the intramolecular interactions due to these three mutations in SARS-CoV-2 spike N-terminal domain. The substitution of threonine with mutant residues alters the side chain leading to change of intramolecular bonds in the pocket of SARS-CoV-2 spike N-terminal domain (Figure 3). Therefore, it can be consecutively stated that T29A is affecting the stability and intramolecular interactions in the protein which may have functional consequences.

Table 1. The values of change in ΔΔSVibENCoM (kcal mol-1 K-1) and ΔΔG (kcal.mol-1) due to the mutations in the spike N-terminal domain of SARS-CoV-2.

Mutation	ΔΔS$_{Vib}$ENCoM	Effect	ΔΔG DynaMut	Effect
T29A	0.293	Increase of molecule flexibility	-1.555	Destabilizing
T250I	0.023	Increase of molecule flexibility	0.057	Stabilizing
T299I	0.123	Increase of molecule flexibility	0.150	Stabilizing

Phylogenetic analysis

In order to identify the possible sources of the delta variant's introduction to Morocco, we searched for Moroccan sequences that were nested within sequences from various countries in the global phylogenetic tree containing 2131 leaves, including the reference genome (Supplementary Figure 4).

We discovered several clades where a Moroccan sequence was near other delta genomes circulating in the world (Figure 4). We propose three distinct possible groups for the phylogenetic position of Moroccan samples in the tree:

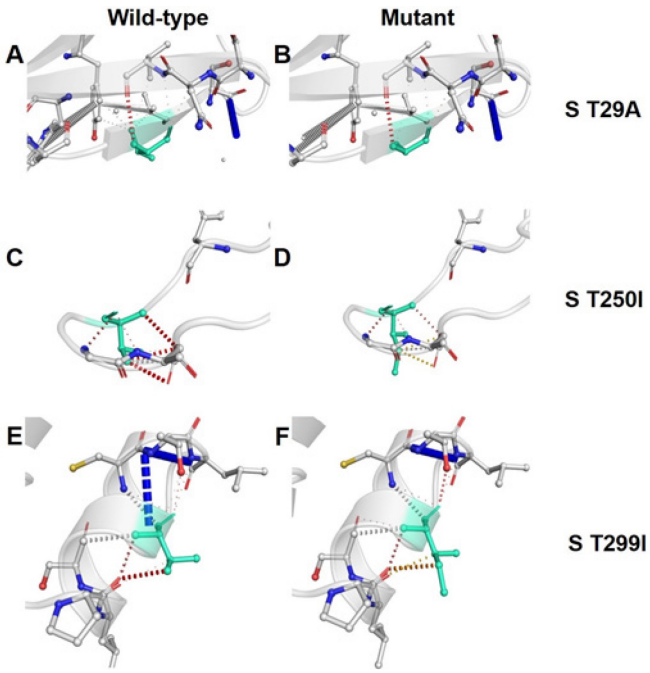
Group 1 (Moroccan clades): groups made up mostly of Moroccan sequences.

Group 2 (Heterogeneous clades): clades containing a set of Moroccan sequences and a set of sequences from other countries.

Group 3 (Unique sequence): single Moroccan sequence discovered inside a clade of sequences from other nations, but is not related to any other Moroccan sequences.

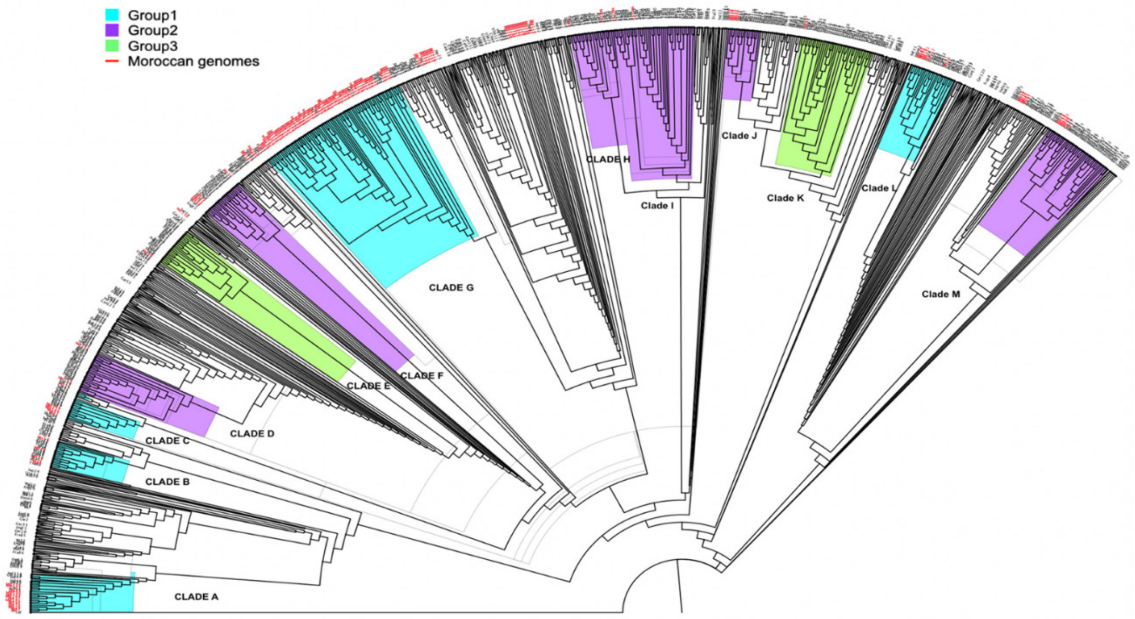
After the clades were constructed, 128 out of 164 Moroccan genomes were grouped into clades. Five clades (A, B, C, G and L) including 105 sequences belonging to Group 1 (Supplementary Figure 5), six clades (D, F, H, I, J and M) including 21 sequences belonging to Group 2 (Supplementary Figure 6) and two clades (E and K) including 2 unique sequences belonging to Group 3 (Supplementary Figure 7).

Figure 3. Effect of amino acid substitution on interatomic interactions. Interatomic interactions mediated by T29A, T250I and T299I of SARS-CoV-2 spike N-terminal domain.



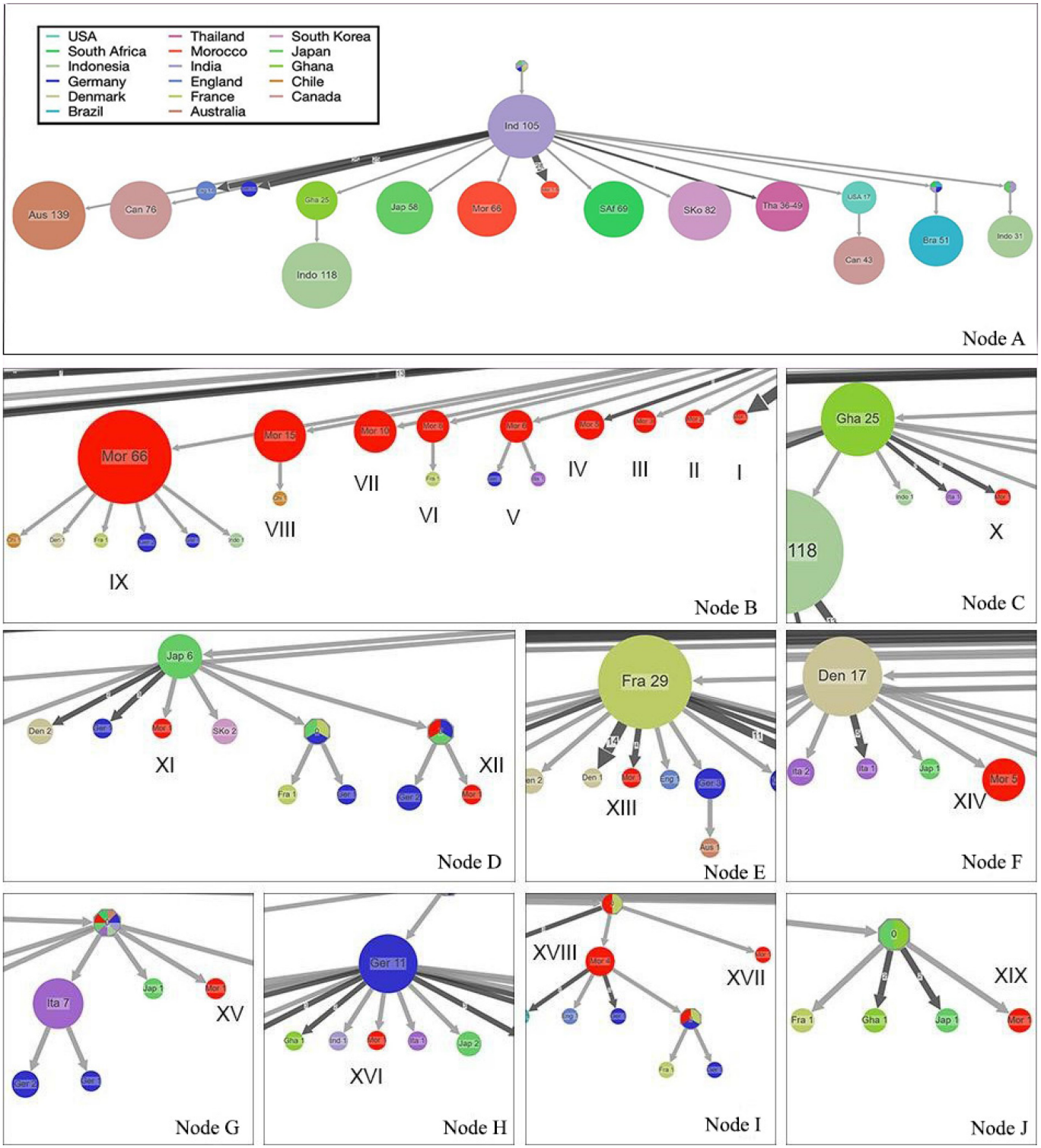
(**A, B**) represent threonine to alanine substitution at 29th position, (**C, D**) represent threonine to isoleucine substitution at 250th position, (**E, F**) represent threonine to isoleucine substitutions at 299th position. Wild-type and mutant residues are colored in light-green and are also represented as sticks alongside with the surrounding residues which are involved in any type of interactions.

Figure 4. Cladogram constructed with 2131 sequences, including the reference genome.



Moroccan samples are indicated in red. The phylogenetic clades are distributed in three groups, indicated by three different colors.

Figure 5. Ancestral reconstruction of the 2130 delta sequences by country origin.



A. shows the compressed visualization of the tree. (**B-J**) show the Moroccan nodes extracted from the tree produced by PastML using MPPA with an F81 model. Different colors correspond to different countries. Nodes denote genetic clusters of samples with the same state. The state and sample size of clusters are indicated for each circle. An arrow between two circles denotes events of transmission from the top node to the bottom node. The size and the number on top of the arrows indicate that the arrows represent multiple transmission events leading to clusters of similar sizes.

Thereafter, we referred to the genetic profile of each clade as summarized in Supplementary Table 3, in order to verify the existence of unique mutations.

Ancestral reconstruction of the delta variant sequences

Summarized ancestral scenarios for location reconstruction on the delta tree is shown in Figure 5A. In order to get more information about all the events, we used the complete visualization with 2976 nodes and 71 unresolved nodes. All the probabilities of the clades are provided in Supplementary Table 4. The results indicated that approximately all nodes showed a parent node originated from India (probability = 0.86). From there, the delta variant was introduced to Morocco in multiple ways as it is shown by the 19 Moroccan nodes ($0.68 < \text{probability} < 1$) (Figure 5B to 5J). Nine separate nodes (Figure 5B: node I to IX) diverge directly from the Indian parent, 5 nodes (Figure 5: X, XI, XIII, XIV and XVI) diverge from an ancestor that itself derives from India. The remaining 5 nodes are called "unresolved nodes"; these nodes have multiple colors indicating several corresponding states having similar marginal probabilities.

Possible delta introduction sources

In order to extract the possible sources of delta variant's introduction, we considered the groups 2 and 3, where the majority of the sequences composing the clades belonged to a country other than Morocco.

India: As represented in the ancestral, India was the main ancestor for all the nodes where B.1.617.2 variant was first detected. As for Moroccan nodes (Figure 5B: I to IX), nine nodes diverged directly from India with approximately 119 sequences. The mutation profile analysis also confirmed this as we observed the same major mutations in those nodes.

France: The node in Figure 5E indicated an introduction from France to Morocco. This introduction was also confirmed in the clade H (Figure 4) with a shared mutation (Y6160Y/ORF1ab).

Japan: Two introductions were observed in the ancestral reconstruction from India to Japan and from Japan to Morocco (Figure 5D). It concerned the following mutations: V4887V/ORF1ab, I850L/S and K16T/ORF3a that were shared between 2 sequences from Morocco and others from Japan as described in clade D (Figure 4).

Germany: An introduction is shown in Figure 5H, and confirmed in clade K (Figure 4) via a single mutation (T6891T/ORF1ab).

Possible delta transmission from Morocco to other countries

We suggest that there is a possibility of transmission of the delta variant from Morocco, when a foreign sequence is located in a clade composed of mostly Moroccan sequences.

Chile: Transmission confirmed in clade G (Figure 4); a single mutation in the spike gene leading to T29A was shared between a sequence from Chile and 72 sequences from Morocco. The transmission was also confirmed in the ancestral reconstruction (Figure 5B: node IX).

France: Multiple transmissions from Morocco to France were found. This concerns 5 mutations found in 3 different nodes from the clades in Figure 4: Q677H/S and E239E/ORF3a shared exclusively between a sequence from France and 5 Moroccan sequences from clade B which was confirmed in the ancestral node (Figure 5I), P207L/N and P46S/N are found in one sequence from France and 6 from Morocco belonging to clade L which is confirmed by the ancestral node VI (Figure 5B) and the last mutation generating the T29A is found in 72 Moroccan sequences and in 1 sequence from France as represented in the clade G (Node IX in Figure 5B).

Germany: Transmission of mutations from Morocco to Germany was observed in 3 nodes. The first transmission concerns 2 mutations (Q677H/S and E239E/ORF3a) found in 2 German and 5 Moroccan sequences as illustrated in clade B (Figure 4) and in the ancestral reconstruction (Figure 5I). The second one was located in clade M (Figure 4) where 2 mutations (V3986V/ORF1ab and E102V/ORF3a) were found only in 5 sequences from Morocco and one from Germany (node V in Figure 5B). The last mutation concerns the T26A/S that was dominant in Morocco and spread to Germany as confirmed in clade G (Figure 4) and in the ancestral (node IX in Figure 5B).

England: The mutation analysis of clade B sequences revealed specific mutations shared between 5 Moroccan sequences and one sequence from England, Q677H/S and E239E/ORF3a. Those two mutations were confirmed by the ancestral tree as transferred from Morocco to England (Figure 5I).

Italy: The ancestral construction revealed a transmission from Morocco to a single sequence from Italy (node V in Figure 5B), where the sequence has accumulated 2 mutations, V3986V/ORF1ab and E102V/ORF3a as illustrated in clade M (Figure 4).

Denmark: We noticed one single transmission from Morocco (node IX in Figure 5B), which was the case of

the mutation T29A in the spike gene found in clade G (Figure 4) and in the majority of Moroccan clades.

Discussion

The epidemiologic curve of COVID-19 in Morocco showed two distinct waves; the first wave caused by the alpha variant extended from August 2020 to January 2021, and the delta variant wave extended from July 2021 to the end of October 2021. In the current study, we analyzed delta Moroccan genome sequences and compared them to delta genomes from all over the world in order to monitor the evolution of this variant in Morocco.

Considering only the mutations present in 75% of the Moroccan sequences (25 mutations), the major delta variant mutations found in Morocco were 84% similar to the reference profile. Six mutations were not detected in our analysis (4 in ORF1, E156G in the S and S84L in ORF8). This indicated a slight diversity of the delta variant in Morocco. This result has been proven in another SARS-CoV-2 genetic diversity study from Morocco [26]. A recent study reported that Morocco is among the North African countries with a higher rate of positivity and significant genetic diversity (46 Pango lineages identified) [27].

Each country is characterized by a specific delta mutation profile [12]. In the case of Morocco, we found that in addition to the major mutations found in other countries, some specific mutations, such as the I1128T mutation in the ORF1ab gene, was found exclusively in a set of Moroccan sequences. Another mutation, M920I, at position 3025 of the ORF1ab gene was found only in five Moroccan sequences.

This study revealed the dominance of the AY.33 sublineage in Morocco; this sublineage was first sequenced in Morocco in June 2021 (EPI_ISL_3253426) [2]. Global phylogenetic analysis of delta variant grouped Moroccan sequences belonging to this sublineage in one clade (Clade G in Figure 4). Thereafter, the results of mutation analysis deduced the existence of specific spike N-terminal domain (NTD) mutations in this clade; T29A, T250I, and T299I, and their effect on the protein structure and dynamics was subsequently analyzed.

The NTD has been reported to interact with the tyrosine-protein kinase UFO (AXL) host receptor facilitating the virus's entrance into human cells [28]. Since the antibody-mediated protection depends on the target antigen structure, any mutation causing the target antigen's productive conformational shift may reduce its binding effect and degrade its protective function [29,30]. Our data demonstrate that T29A, T250I and T299I mutations lead to significant changes in the protein secondary structure.

Several mutations of the SARS-CoV-2 spike protein have been found and their effects are being explored in immune system evasion and increased transmission [31]. The mutations in delta variants distribute in a similar way to those of other VOCs, clustering in the RBD and the NTD of the spike protein targeted by most neutralizing antibodies [32-34]. However, some of these monoclonal antibodies have shown an impairment in binding to the B.1.617.2 spike due to diverse mutations in the two aforementioned domains, suggesting that the delta variant partially but significantly escape neutralizing antibodies [35].

In this study, we proposed the different possibilities of introduction and transmission of the delta variant and its sublineages between Morocco and other countries based on the phylogenetic groups, genetic profiles and ancestral reconstruction results. Since the beginning of the SARS-CoV-2 pandemic, the Moroccan government has taken quick decisions to prevent the spread of the virus. These measures included suspension of flights with several countries, especially when a new variant is detected somewhere. This decision was not only aimed at reducing the spread of the virus, but also at preventing the introduction of new variants in the country.

Conclusions

This study revealed the dominance of AY.33 sublineage, and the existence of three specific Moroccan mutations in the spike N-terminal domain protein in this lineage. We have also demonstrated by phylogenetic and network analysis the possibility that the delta variant was introduced into Morocco from different countries, including India, France, Japan, and Germany. Moreover, we suggest that there is a possible transmission of the delta variant from Morocco to Chile, France, Germany, England and Italy.

The limitation of our study is that the number of sequenced and deposited Moroccan SARS-CoV-2 delta sequences is not proportional to the number of cases. In addition, it was not feasible to include more countries due to the unavailability of data at the time of the preparing this manuscript.

Acknowledgements

This work was carried out with the assistance of national funding from the Moroccan Ministry of Higher Education and Scientific Research (COVID-19 program) to A.I. and scholarship of excellence from the National Center for Scientific and Technical Research in Morocco.

References

1. Center for Disease Control and Prevention (2019) Novel coronavirus (2019-nCoV) frequently asked questions and answers. Available: https://www.cdc.gov/coronavirus/2019-ncov/faq.html. Accessed: 19 May 2022.
2. Khare S, Gurry C, Freitas L, Schultz MB, Bach G, Diallo A, Akite N, Ho J, Lee RT, Yeo W, Team GCC, Maurer-Stroh S (2021) GISAID's Role in Pandemic Response. China CDC Wkly 3: 1049–1051.
3. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2013) GenBank. Nucleic Acids Res. 41: D36-42.
4. O'Toole Á, Scher E, Underwood A, Jackson B, Hill V, McCrone JT, Colquhoun R, Ruis C, Abu-Dahab K, Taylor B, Yeats C, du Plessis L, Maloney D, Medd N, Attwood SW, Aanensen DM, Holmes EC, Pybus OG, Rambaut A (2021) Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. Virus Evol. 7: veab064.
5. Aksamentov I, Roemer C, Emma BH and Neher RA (2021) Nextclade: clade assignment, mutation calling and quality control for viral genomes. Journal of Open Source Software, 6: 3773.
6. Ouadghiri M, Aanniz T, Essabbar A, Seffar M, Kabbaj H, El Amin G, Zouaki A, Amzazi S, Belyamani L, Ibrahimi A (2021) Report of SARS-CoV-2 B1.1.7 lineage in Morocco. Microbiol Resour Announc 10: e00240-21.
7. Tomaszewski T, DeVries RS, Dong M, Bhatia G, Norsworthy MD, Zheng X, Caetano-Anollés G (2020) New pathways of mutational change in SARS-CoV-2 proteomes involve regions of intrinsic disorder important for virus replication and release. Evol Bioinform Online 16: 1176934320965149.
8. Thomson EC, Rosen LE, Shepherd JG, Spreafico R, Filipe A da S, Wojcechowskyj JA, Davis C, Piccoli L, Pascall DJ, Dillen J, Lytras S, Czudnochowski N, Shah R, Meury M, Jesudason N, Marco AD, Li K, Bassi J, O'Toole A, Pinto D, Colquhoun RM, Culap K, Jackson B, Zatta F, Rambaut A, Jaconi S, Sreenu VB, Nix J, Zhang I, Jarrett RF, Glass WG, Beltramello M, Nomikou K, Pizzuto M, Tong L, Cameroni E, Croll TI, Johnson N, Iulio JD, Wickenhagen A, Ceschi A, Harbison AM, Mair D, Ferrari P, Smollett K, Sallusto F, Carmichael S, Garzoni C, Nichols J, Galli M, Hughes J, Riva A, Ho A, Schiuma M, Semple MG, Openshaw PJM, Fadda E, Baillie JK, Chodera JD, Rihn SJ, Lycett SJ, Virgin HW, Telenti A, Corti D, Robertson DL, Snell G (2021) Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity. Cell. 184: 1171-1187.e20.
9. Starr TN, Greaney AJ, Hilton SK, Ellis D, Crawford KHD, Dingens AS, Navarro MJ, Bowen JE, Tortorici MA, Walls AC, King NP, Veesler D, Bloom JD (2020) Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. Cell 182: 1295-1310.e20.
10. Baj A, Novazzi F, Drago Ferrante F, Genoni A, Tettamanzi E, Catanoso G, Dalla Gasperina D, Dentali F, Focosi D, Maggi F (2021) Spike protein evolution in the SARS-CoV-2 Delta variant of concern: a case series from Northern Lombardy. Emerg Microbes Infect. 10: 2010-2015.
11. Gangavarapu K, Latif AA, Mullen JL, Alkuzweny M, Hufbauer E, Tsueng G, Haag E, Zeller M, Aceves CM, Zaiets K, Cano M, Zhou J, Qian Z, Sattler R, Matteson NL, Levy JI, Lee RT, Freitas L, Maurer-Stroh S, Team G core and curation, Suchard MA, Wu C, Su AI, Andersen KG, Hughes LD (2022) Outbreak.info genomic reports: scalable and dynamic surveillance of SARS-CoV-2 variants and mutations. medRxiv: 2022.01.27.22269965.
12. Shiehzadegan S, Alaghemand N, Fox M, Venketaraman V (2021) Analysis of the delta variant B. 1.617. 2 COVID-19. Clinics and Practice 11: 778-784.
13. Liu Y, Rocklöv J (2021) The reproductive number of the Delta variant of SARS-CoV-2 is far higher compared to the ancestral SARS-CoV-2 virus. Journal of Travel Medicine 28: taab124.
14. Li H (2021) Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34: 3094–3100.
15. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, (2009) 1000 genome project data processing subgroup. The sequence alignment/map format and SAM tools. Bioinformatics 25: 2078–2079.
16. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly 6: 80–92.
17. Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast fourier transform. Nucleic Acids Res 14: 3059-3066.
18. Nguyen LT, Schmidt HA, von Haeseler AV, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular Biology and Evolution 32: 268–274.
19. Ishikawa SA, Zhukova A, Iwasaki W, Gascuel O (2019) A fast likelihood method to reconstruct and visualize ancestral scenarios. Molecular Biology and Evolution 36: 2069–2085.
20. To TH, Jung M, Lycett S, Gascuel O (2016) Fast dating using least-squares criteria and algorithms. Syst Biol 65: 82–97.
21. Letunic I, Bork P (2021) Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Research 49: W293–W296.
22. Ashok Kumar T (2013) CFSSP: Chou and Fasman secondary structure prediction server. Wide Spectrum 1: 15–19.
23. Rodrigues CH, Pires DE, Ascher DB (2018) DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. Nucleic Acids Res 46:W350–W355.
24. Goethe M, Fita I, Rubi JM (2017) Testing the mutual information expansion of entropy with multivariate Gaussian distributions. J. Chem. Phys 147: 224102.
25. Eriksson AE, Baase WA, Zhang XJ, Heinz DW, Blaber M, Baldwin EP, Matthews BW (1992) Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. Science 255: 178–183.
26. Laamarti M, Essabbar A, Alouane T, Kartti S, Boumajdi N, Bendani H, Laamarti R, Allam L, Ouadghiri M, Chemao-Elfihri MW, Ghrifi F, Smyej I, Rahoui J, Benrahma H, Diawara I, Aanniz T, Hafidi NE, Jaoudi RE, Nejjari C, Amzazi S, Mentag R, Belyamani L, Ibrahimi A (2020) Do the Moroccan SARS-CoV-2 genetic diversity hamper the use of the developed universal vaccines in Morocco? bioRxiv: doi: 10.1101/2020.06.30.181123.
27. Menasria T, Aguilera M (2022) Genomic diversity of SARS-CoV-2 in Algeria and North African countries: what we know so far and what we expect? Microorganisms. 10: 467.
28. Wang S, Qiu Z, Hou Y, Deng X, Xu W, Zheng T, Wu P, Xie S, Bian W, Zhang C, Sun Z, Liu K, Shan C, Lin A, Jiang S, Xie Y, Zhou Q, Lu L, Huang J, Li X (2021) AXL is a candidate receptor for SARS-CoV-2 that promotes infection of pulmonary and bronchial epithelial cells. Cell Res. 31: 126-140y.

29. Kwarteng A, Asiedu E, Sylverken AA, Larbi A, Sakyi SA, Asiedu SO (2021) Molecular characterization of interactions between the D614G variant of SARS-CoV-2 S-protein and neutralizing antibodies: a computational approach. Infect Genet Evol 91: 104815.
30. Li Q, Wu J, Nie J, Zhang L, Hao H, Liu S, Zhao C, Zhang Q, Liu H, Nie L, Qin H, Wang M, Lu Q, Li X, Sun Q, Liu J, Zhang L, Li X, Huang W, Wang Y (2020) The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. Cell. 182: 1284-1294.e9.
31. Mengist HM, Kombe AJ, Mekonnen D, Abebaw A, Getachew M, Jin T (2021) Mutations of SARS-CoV-2 spike protein: implications on immune evasion and vaccine-induced immunity. Semin Immunol 55: 101533.
32. Cerutti G, Guo Y, Zhou T, Gorman J, Lee M, Rapp M, Reddem ER, Yu J, Bahna F, Bimela J, Huang Y, Katsamba PS, Liu L, Nair MS, Rawi R, Olia AS, Wang P, Zhang B, Chuang G-Y, Ho DD, Sheng Z, Kwong PD, Shapiro L (2021) Potent SARS-CoV-2 neutralizing antibodies directed against spike N-terminal domain target a single supersite. Cell Host Microbe 29: 819-833.e7.
33. McCallum M, De Marco A, Lempp FA, Tortorici MA, Pinto D, Walls AC, Beltramello M, Chen A, Liu Z, Zatta F, Zepeda S, di Iulio J, Bowen JE, Montiel-Ruiz M, Zhou J, Rosen LE, Bianchi S, Guarino B, Fregni CS, Abdelnabi R, Foo S-YC, Rothlauf PW, Bloyet L-M, Benigni F, Cameroni E, Neyts J, Riva A, Snell G, Telenti A, Whelan SPJ, Virgin HW, Corti D, Pizzuto MS, Veesler D (2021) N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. Cell 184:2332-2347.e16.
34. Suryadevara N, Shrihari S, Gilchuk P, VanBlargan LA, Binshtein E, Zost SJ, Nargi RS, Sutton RE, Winkler ES, Chen EC, Fouch ME, Davidson E, Doranz BJ, Chen RE, Shi P-Y, Carnahan RH, Thackray LB, Diamond MS, Crowe JE (2021) Neutralizing and protective human monoclonal antibodies recognizing the N-terminal domain of the SARS-CoV-2 spike protein. Cell 184: 2316-2331.e15.
35. Planas D, Veyer D, Baidaliuk A, Staropoli I, Guivel-Benhassine F, Rajah M, Planchais C, Porrot F, Robillard N, Puech J, Prot M, Gallais F, Gantner P, Velay A, Guen JL, Kassis-Chikhani N, Edriss D, Belec L, Seve A, Pere H, Courtellemenont L, Hocqueloux L, Fafi-Kremer S, Prazuck T, Mouquet H, Bruel T, Simon-Loriere E, Rey F, Schwartz O (2021) Reduced sensitivity of infectious SARS-CoV-2 variant B.1.617.2 to monoclonal antibodies and sera from convalescent and vaccinated individuals. Nature 596, 276–280.

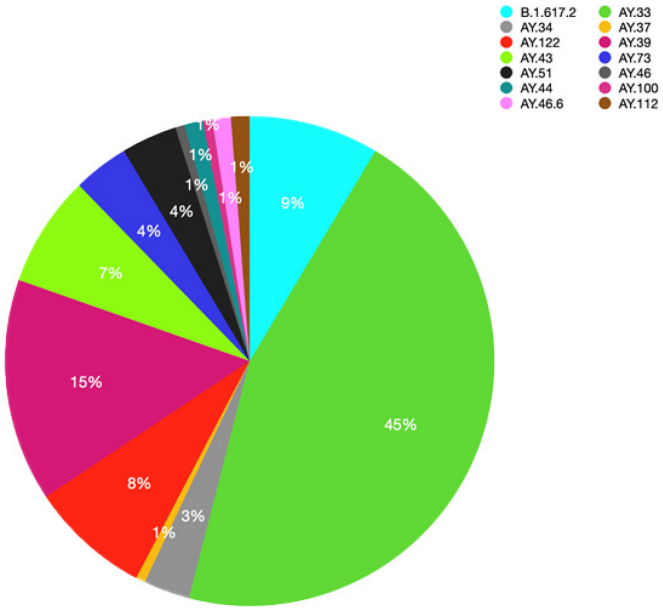
Corresponding author

Pr. Mouna Ouadghiri
Laboratory of Biotechnology, Medical and Pharmacy School, Mohammed V University, Rabat, 10100, Morocco
Tel: +212666138766
Email: m.ouadghiri@um5r.ac.ma

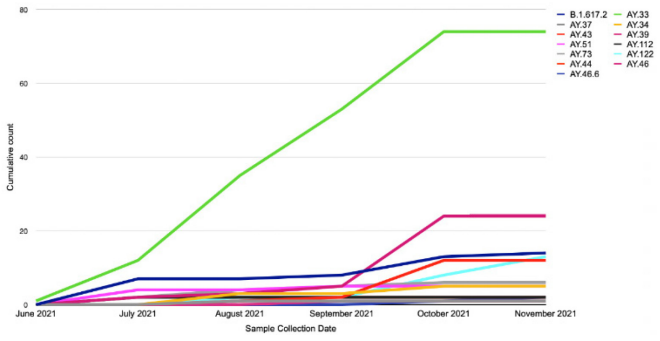
Conflict of interests: No conflict of interests is declared.

Annex – Supplementary Items

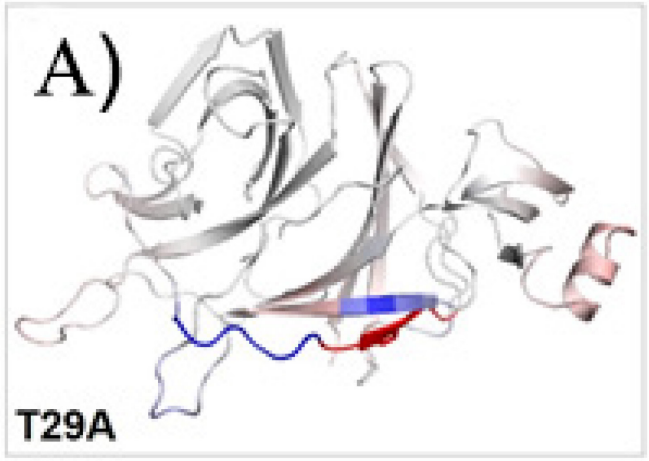
Supplementary Figure 1. Percentage distribution of circulating delta lineages in Morocco.

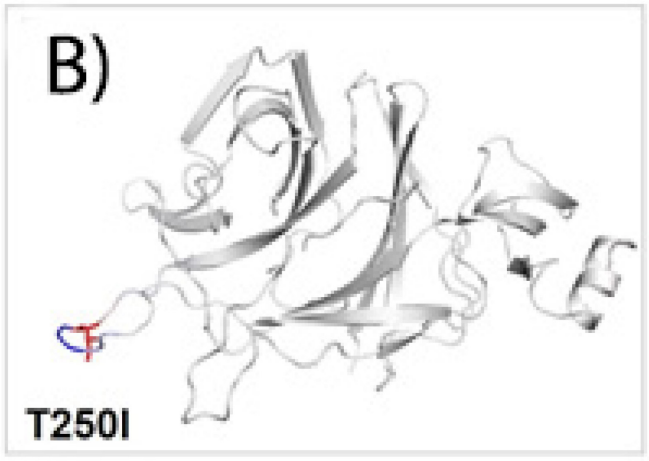


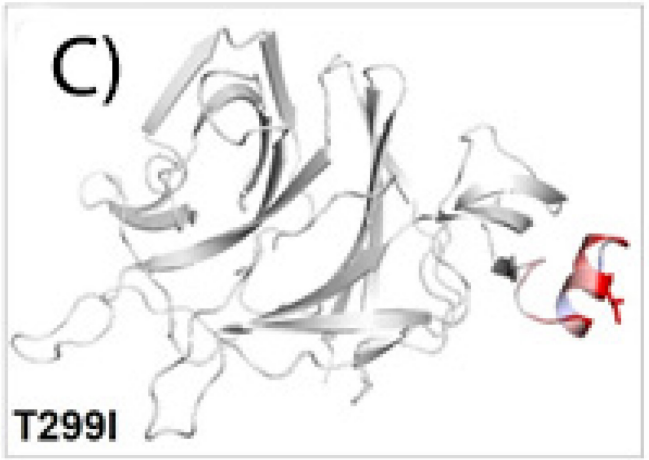
Supplementary Figure 2. Cumulative count of sub-lineages plotted against sample collection date.



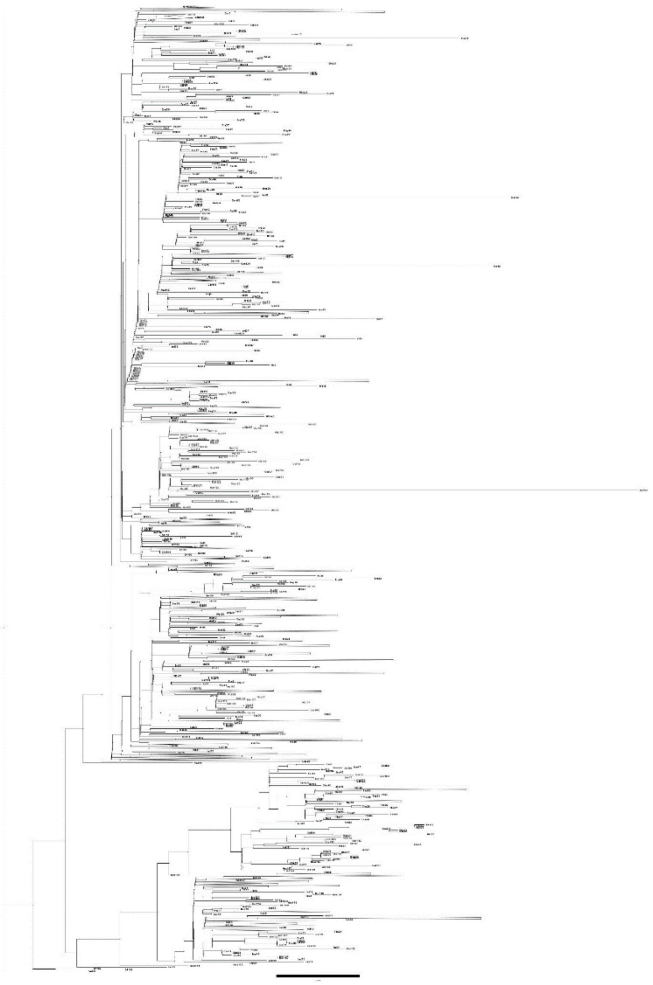
Supplementary Figure 3. Effect of mutations on structural dynamics of SARS-CoV-2 spike N-terminal domain. Analysis of the protein dynamicity and flexibility. (Panels **A, B** and **C**) Δ Vibrational Entropy Energy between Wild-Type and Mutant S protein, amino acids are colored according to the vibrational entropy change as a consequence of mutation of S protein. Blue represents a rigidification of the structure and red a gain in flexibility.



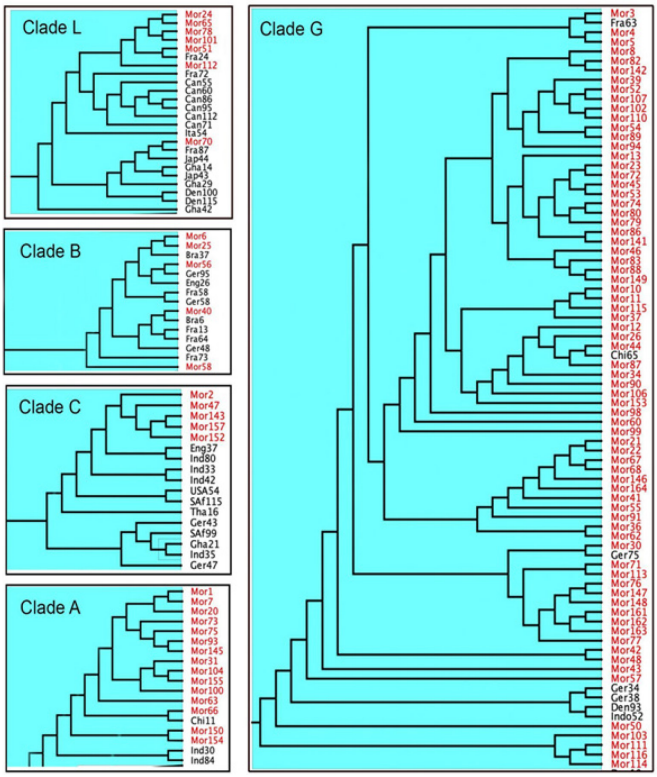




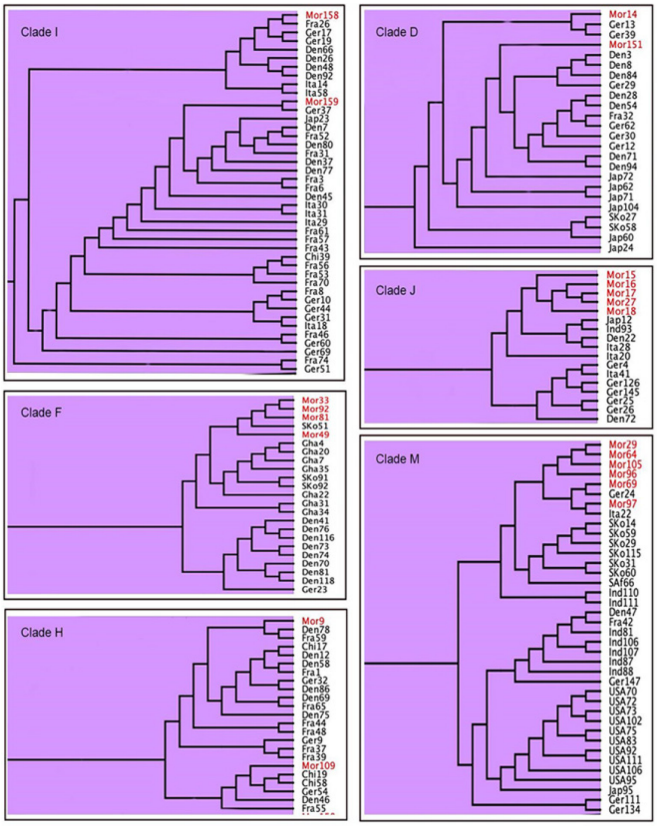
Supplementary Figure 4. Global ML phylogenetic tree containing 2131 branches, including the reference genome. Moroccan samples are indicated in red.



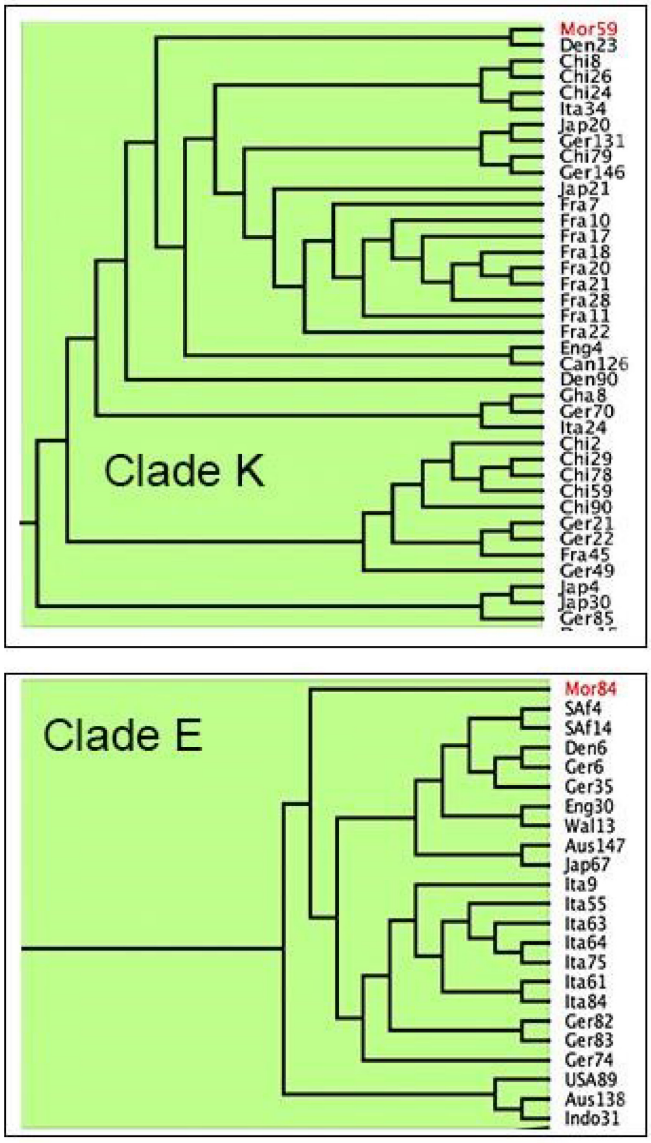
Supplementary Figure 5. Clades containing the majority of Moroccan genomes extracted from the phylogenetic tree (Group 1).



Supplementary Figure 6. Heterogeneous clades containing Moroccan genomes and genomes from different countries, extracted from the phylogenetic tree (Group 2).



Supplementary Figure 7. Clades containing one Moroccan sequence and a set of sequences from different countries extracted from the phylogenetic tree (Group 3).



Supplementary Table 1. Occurrence of the Moroccan mutations.

Nucleotide position	Gene	Residue change	Nucleotide change	Effect	peptide	Count
14408	ORF1b	P4715L	14144C > T	missense_variant	NSP12	164
15451	ORF1b	G5063S	15187G > A	missense_variant	NSP12	164
22917	S	L452R	1355T > G	missense_variant	--	164
22995	S	T478K	1433C > A	missense_variant	--	164
23403	S	D614G	1841A > G	missense_variant	--	164
23604	S	P681R	2042C > G	missense_variant	--	164
25469	ORF3a	S26L	77C > T	missense_variant	--	164
28881	N	R203M	608G > T	missense_variant	--	164
210	5UTR	Not found	210G > T	intergenic_region	--	163
241	5UTR	Not found	241C > T	intergenic_region	--	163
26767	M	I82T	245T > C	missense_variant	--	163
28461	N	D63G	188A > G	missense_variant	--	162
16466	ORF1b	P5401L	16202C > T	missense_variant	NSP13	162
21618	S	T19R	56C > G	missense_variant	--	160
29742	ORF10	Not found	29742G > T	intergenic_region	--	160
24410	S	D950N	2848G > A	missense_variant	--	159
29402	N	D377Y	1129G > T	missense_variant	--	156
28247	ORF8	Asp119_Phe120del	355_360del	conservative_inframe_deletion	--	150
28916	N	G215C	643G > T	missense_variant	--	146
9053	ORF1a	V2930L	8788G > T	missense_variant	NSP4	145
19220	ORF1b	A6319V	18956C > T	missense_variant	NSP14	145
4181	ORF1a	A1306S	3916G > T	missense_variant	NSP3	144
27874	ORF7b	T40I	119C > T	missense_variant	--	144
10029	ORF1a	T3255I	9764C > T	missense_variant	NSP4	136
7124	ORF1a	P2287S	6859C > T	missense_variant	NSP3	134
11201	ORF1a	T3646A	10936A > G	missense_variant	NSP6	129
22028	S	Glu156_Arg158delinsGly	467_472del	disruptive_inframe_deletion	--	129
6402	ORF1a	P2046L	6137C > T	missense_variant	NSP3	123
27638	ORF7a	V82A	245T > C	missense_variant	--	100
27752	ORF7a	T120I	359C > T	missense_variant	--	96
21647	S	T29A	85A > G	missense_variant	--	72
23401	S	Q613H	1839G > T	missense_variant	--	68
22311	S	T250I	749C > T	missense_variant	--	61
22458	S	T299I	896C > T	missense_variant	--	61
21987	S	G142D	425G > A	missense_variant	--	51
5184	ORF1a	P1640L	4919C > T	missense_variant	NSP3	21
9891	ORF1a	A3209V	9626C > T	missense_variant	NSP4	18
11418	ORF1a	V3718A	11153T > C	missense_variant	NSP6	18
3648	ORF1a	I1128T	3383T > C	missense_variant	NSP3	15
25647	ORF3a	L85F	255G > T	missense_variant	--	15
22227	S	A222V	665C > T	missense_variant	--	14
1191	ORF1a	P309L	926C > T	missense_variant	NSP2	11
22323	S	S254F	761C > T	missense_variant	--	11
21846	S	T95I	284C > T	missense_variant	--	11
1048	ORF1a	K261N	783G > T	missense_variant	NSP2	10
21010	ORF1b	V6916L	20746G > T	missense_variant	NSP16	9
27739	ORF7a	L116F	346C > T	missense_variant	--	9
29554	N	Not found	29554G > T	intergenic_region	--	8
28247	ORF8	D119None	355_358del	frameshift_variant	--	8
11514	ORF1a	T3750I	11249C > T	missense_variant	NSP6	7
27527	ORF7a	P45L	134C > T	missense_variant	--	6
18176	ORF1b	P5971L	17912C > T	missense_variant	NSP14	6
25562	ORF3a	Q57R	170A > G	missense_variant	--	6
25697	ORF3a	E102V	305A > T	missense_variant	--	6
222	5UTR	Not found	222C > T	intergenic_region	--	6
6573	ORF1a	S2103F	6308C > T	missense_variant	NSP3	6
28409	N	P46S	136C > T	missense_variant	--	6
28893	N	P207L	620C > T	missense_variant	--	6
24887	S	F1109L	3325T > C	missense_variant	--	6
5051	ORF1a	P1596S	4786C > T	missense_variant	NSP3	5

6539	ORF1a	H2092Y	6274C > T	missense_variant	NSP3	5
14122	ORF1b	G4620S	13858G > A	missense_variant	NSP12	5
27972	ORF8	Gln27*	79C > T	stop_gained	--	5
29427	N	R385K	1154G > A	missense_variant	--	5
11083	ORF1a	L3606F	10818G > T	missense_variant	NSP6	5
2306	ORF1a	L681F	2041C > T	missense_variant	NSP2	5
3025	ORF1a	M920I	2760G > T	missense_variant	NSP3	5
26104	ORF3a	D238Y	712G > T	missense_variant	--	5
27416	ORF7a	A8V	23C > T	missense_variant	--	5
23593	S	Q677H	2031G > C	missense_variant	--	5
27390	ORF6	Not found	27390G > T	intergenic_region	--	5
203	5UTR	Not found	203C > T	intergenic_region	--	4
5690	ORF1a	A1809T	5425G > A	missense_variant	NSP3	4
12455	ORF1a	L4064I	12190C > A	missense_variant	NSP8	4
1137	ORF1a	K291R	872A > G	missense_variant	NSP2	4
21724	S	L54F	162G > T	missense_variant	--	4
22335	S	W258L	773G > T	missense_variant	--	4
26389	E	V49L	145G > T	missense_variant	--	4
15952	ORF1b	L5230I	15688C > A	missense_variant	NSP12	4
28299	N	Q9L	26A > T	missense_variant	--	4
20320	ORF1b	H6686Y	20056C > T	missense_variant	NSP15	4
6497	ORF1a	K2078E	6232A > G	missense_variant	NSP3	4
20611	ORF1b	L6783F	20347C > T	missense_variant	NSP15	4
27361	ORF6	A54S	160G > T	stop_gained	--	4
29738	ORF10	Not found	29738C > T	intergenic_region	--	4
17746	ORF1b	P5828S	17482C > T	missense_variant	NSP13	4
9441	ORF1a	C3059F	9176G > T	missense_variant	NSP4	4
16466	ORF1b	Not found	*230C >T	downstream_gene_variant	NSP13	3
20375	ORF1b	I6704T	20111T > C	missense_variant	NSP15	3
3510	ORF1a	A1082V	3245C > T	missense_variant	NSP3	3
936	ORF1a	T224I	671C > T	missense_variant	NSP2	3
18255	ORF1b	M5997I	17991G > T	missense_variant	NSP14	3
5822	ORF1a	L1853I	5557C > A	missense_variant	NSP3	3
29364	N	P364L	1091C > T	missense_variant	--	3
29409	N	T379I	1136C > T	missense_variant	--	2
28703	N	D144H	430G > C	missense_variant	--	2
12242	ORF1a	R3993C	11977C > T	missense_variant	NSP8	2
21575	S	L5F	13C > T	missense_variant	--	2
26811	M	I97None	290_315del	frameshift_variant	--	2
21137	ORF1b	K6958R	20873A > G	missense_variant	NSP16	2
24110	S	I850L	2548A > C	missense_variant	--	2
25439	ORF3a	K16T	47A > C	missense_variant	--	2
2991	ORF1a	D909G	2726A > G	missense_variant	NSP3	2
13140	ORF1a	T4292N	12875C > A	missense_variant	NSP10	2
2631	ORF1a	M789None	2367_2394del	frameshift_variant	NSP2	2
16393	ORF1b	P5377S	16129C > T	missense_variant	NSP13	2
8131	ORF1a	K2622N	7866G > T	missense_variant	NSP3	2
16877	ORF1b	T5538I	16613C > T	missense_variant	NSP13	2
21800	S	D80Y	238G > T	missense_variant	--	2
27353	ORF6	Q51L	152A > T	missense_variant	--	2
28724	N	P151S	451C > T	missense_variant	--	2
6601	ORF1a	E2112D	6336A > C	missense_variant	NSP3	2
21520	ORF1b	V7086F	21256G > T	missense_variant	NSP16	2
4916	ORF1a	I1551V	4651A > G	missense_variant	NSP3	2
28985	N	G238C	712G > T	missense_variant	--	2
2574	ORF1a	T770I	2309C > T	missense_variant	NSP2	2
10323	ORF1a	K3353R	10058A > G	missense_variant	NSP5	2
24969	S	T1136K	3407C > A	missense_variant	--	2
15746	ORF1b	T5161I	15482C > T	missense_variant	NSP12	2
25699	ORF3a	Pro104_Phe105insProPro	311_312insCCCCCC	disruptive_inframe_insertion	--	2
8915	ORF1a	F2884L	8650T > C	missense_variant	NSP4	2
28961	N	L230F	688C > T	missense_variant	--	2
4601	ORF1a	V1446I	4336G > A	missense_variant	NSP3	2

15451	ORF1b	Not found	*2010G > A	downstream_gene_variant	NSP12	1
15451	ORF1b	Not found	*1971G > A	downstream_gene_variant	NSP12	1
28290	N	P6L	17C > T	missense_variant	--	1
1821	ORF1a	G519D	1556G > A	missense_variant	NSP2	1
6336	ORF1a	S2024L	6071C > T	missense_variant	NSP3	1
11693	ORF1a	V3810I	11428G > A	missense_variant	NSP6	1
21911	S	L117I	349C > A	missense_variant	--	1
26428	E	V62F	184G > T	missense_variant	--	1
29767	ORF10	Not found	29767A > T	intergenic_region	--	1
6213	ORF1a	A1983V	5948C > T	missense_variant	NSP3	1
7059	ORF1a	G2265V	6794G > T	missense_variant	NSP3	1
9419	ORF1a	I3052V	9154A > G	missense_variant	NSP4	1
19735	ORF1b	D6491Y	19471G > T	missense_variant	NSP15	1
29783	ORF10	Not found	29783G > C	intergenic_region	--	1
8759	ORF1a	A2832T	8494G > A	missense_variant	NSP4	1
19855	ORF1b	I6531V	19591A > G	missense_variant	NSP15	1
12410	ORF1a	I4049V	12145A > G	missense_variant	NSP8	1
23007	S	G482V	1445G > T	missense_variant	--	1
28300	N	Q9H	27G > T	missense_variant	--	1
1284	ORF1a	T340I	1019C > T	missense_variant	NSP2	1
17464	ORF1b	L5735None	17201dupC	frameshift_variant	NSP13	1
19859	ORF1b	A6532V	19595C > T	missense_variant	NSP15	1
26650	M	Asn43_Trp55delinsLys	129_164del	disruptive_inframe_deletion	--	1
20010	ORF1b	Q6582H	19746A > C	missense_variant	NSP15	1
2619	ORF1a	N786None	2356_2398del	frameshift_variant	NSP2	1
15594	ORF1b	K5110N	15330G > T	missense_variant	NSP12	1
24710	S	M1050L	3148A > T	missense_variant	--	1
26779	M	C86F	257G > T	missense_variant	--	1
28245	ORF8	L118V	352T > G	missense_variant	--	1
29183	N	I304V	910A > G	missense_variant	--	1
29686	ORF10	Not found	29686C > T	intergenic_region	--	1
29709	ORF10	Not found	29709T > C	intergenic_region	--	1
14488	ORF1b	E4742None	14224_14225insC	frameshift_variant	NSP12	1
14829	ORF1b	M4855I	14565G > T	missense_variant	NSP12	1
21203	ORF1b	K6980None	20940_20971del	frameshift_variant	NSP16	1
21707	S	H49Y	145C > T	missense_variant	--	1
21812	S	P85None	253_259del	frameshift_variant	--	1
23054	S	Gln498_Tyr508delinsHis	1494_1523del	disruptive_inframe_deletion	--	1
25012	S	E1150D	3450G > T	missense_variant	--	1
25599	ORF3a	W69C	207G > T	missense_variant	--	1
27299	ORF6	I33T	98T > C	missense_variant	--	1
29646	ORF10	V30A	89T > C	missense_variant	--	1
10815	ORF1a	S3517F	10550C > T	missense_variant	NSP5	1
6177	ORF1a	D1971G	5912A > G	missense_variant	NSP3	1
2700	ORF1a	T812I	2435C > T	missense_variant	NSP2	1
3344	ORF1a	S1027G	3079A > G	missense_variant	NSP3	1
13912	ORF1b	N4550Y	13648A > T	missense_variant	NSP12	1
27371	ORF6	P57L	170C > T	missense_variant	--	1
7282	ORF1a	V2340None	7018_7063del	frameshift_variant	NSP3	1
10265	ORF1a	G3334S	10000G > A	missense_variant	NSP5	1
10833	ORF1a	A3523V	10568C > T	missense_variant	NSP5	1
18261	ORF1b	Thr6000_Ala6004del	17999_18013del	disruptive_inframe_deletion	NSP14	1
18397	ORF1b	V6045L	18133G > T	missense_variant	NSP14	1
21148	ORF1b	Gly6962_Leu6978delinsVal	20885_20932del	disruptive_inframe_deletion	NSP16	1
24381	S	S940F	2819C >T	missense_variant	--	1
29445	N	T391I	1172C > T	missense_variant	--	1
8189	ORF1a	V2642I	7924G > A	missense_variant	NSP3	1
17119	ORF1b	L5619I	16855C > A	missense_variant	NSP13	1
26305	E	L21F	61C > T	missense_variant	--	1
1878	ORF1a	S538L	1613C > T	missense_variant	NSP2	1
29041	N	K256N	768G > T	missense_variant	--	1
22658	S	S366A	1096T > G	missense_variant	--	1
25555	ORF3a	V55F	163G > T	missense_variant	--	1

27578	ORF7a	Gln62_Phe63delinsHis	186_188del	disruptive_inframe_deletion	--	1
1473	ORF1a	T403I	1208C > T	missense_variant	NSP2	1
1545	ORF1a	A427V	1280C > T	missense_variant	NSP2	1
16687	ORF1b	I5475V	16423A > G	missense_variant	NSP13	1
28048	ORF8	R52I	155G > T	missense_variant	--	1
28308	N	A12V	35C > T	missense_variant	--	1
21077	ORF1b	T6938I	20813C > T	missense_variant	NSP16	1
25770	ORF3a	R126S	378G > T	missense_variant	--	1
5437	ORF1a	E1724D	5172G >T	missense_variant	NSP3	1
13712	ORF1b	K4483R	13448A > G	missense_variant	NSP12	1
14890	ORF1b	D4876N	14626G > A	missense_variant	NSP12	1
17027	ORF1b	S5588N	16763G > A	missense_variant	NSP13	1
17236	ORF1b	I5658V	16972A > G	missense_variant	NSP13	1
18395	ORF1b	A6044V	18131C > T	missense_variant	NSP14	1
25062	S	G1167V	3500G > T	missense_variant	--	1
25445	ORF3a	G18V	53G > T	missense_variant	--	1
25844	ORF3a	T151I	452C > T	missense_variant	--	1
28725	N	P151L	452C > T	missense_variant	--	1
29736	ORF10	Not found	29736G > T	intergenic_region	--	1
16466	ORF1b	Not found	*3025C > T	downstream_gene_variant	NSP13	1
16466	ORF1b	Not found	*2986C > T	downstream_gene_variant	NSP13	1
24378	S	S939F	2816C > T	missense_variant	--	1
24959	S	V1133F	3397G > T	missense_variant	--	1
27948	ORF8	Glu19*	55G > T	stop_gained	--	1
10998	ORF1a	G3578D	10733G > A	missense_variant	NSP6	1
22031	S	F157I	469T > A	missense_variant	--	1
22032	S	F157C	470T > G	missense_variant	--	1
22034	S	R158G	472A > G	missense_variant	--	1
25641	ORF3a	L83F	249G > T	missense_variant	--	1
27556	ORF7a	A55S	163G > T	missense_variant	--	1
5621	ORF1a	P1786S	5356C > T	missense_variant	NSP3	1
8858	ORF1a	V2865I	8593G > A	missense_variant	NSP4	1
11230	ORF1a	M3655I	10965G > T	missense_variant	NSP6	1
26811	M	A98None	291_315del	frameshift_variant	--	1
28359	N	N29T	86A > C	missense_variant	--	1
368	ORF1a	V35L	103G > T	missense_variant	NSP1	1
3560	ORF1a	V1099L	3295G > T	missense_variant	NSP3	1
4232	ORF1a	D1323Y	3967G > T	missense_variant	NSP3	1
11108	ORF1a	M3616None	10845_10887del	frameshift_variant	NSP6	1
11837	ORF1a	V3858L	11572G > T	missense_variant	NSP6	1
16005	ORF1b	M5247I	15741G > T	missense_variant	NSP12	1
22104	S	G181A	542G > C	missense_variant	--	1
23061	S	N501None	1500dupT	frameshift_variant&stop_gained	--	1
25505	ORF3a	Q38R	113A > G	missense_variant	--	1
27742	ORF7a	Lys117*	349A > T	stop_gained	--	1
4764	ORF1a	H1500R	4499A > G	missense_variant	NSP3	1
22408	S	N282K	846T > A	missense_variant	--	1
25286	S	S1242G	3724A > G	missense_variant	--	1
26426	E	R61H	182G > A	missense_variant	--	1
18394	ORF1b	A6044S	18130G > T	missense_variant	NSP14	1
21372	ORF1b	Q7036H	21108G > T	missense_variant	NSP16	1
25714	ORF3a	L108F	322C > T	missense_variant	--	1
27427	ORF7a	L12F	34C > T	missense_variant	--	1
29513	N	A414S	1240G > T	missense_variant	--	1
3720	ORF1a	Ile1153_Ala1156del	3457_3468del	conservative_inframe_deletion	NSP3	1
3838	ORF1a	L1191F	3573G > T	missense_variant	NSP3	1
7770	ORF1a	H2502R	7505A > G	missense_variant	NSP3	1
9866	ORF1a	L3201F	9601C > T	missense_variant	NSP4	1
21804	S	P82None	245_282del	frameshift_variant	--	1
23055	S	P499None	1496_1524del	frameshift_variant	--	1
27691	ORF7a	Phe101_Val104del	301_312del	conservative_inframe_deletion	--	1
29761	ORF10	Not found	29761A > G	intergenic_region	--	1
24620	S	A1020S	3058G > T	missense_variant	--	1

7027	ORF1a	M2254I	6762G > T	missense_variant	NSP3	1
8991	ORF1a	A2909V	8726C > T	missense_variant	NSP4	1
11280	ORF1a	T3672N	11015C > A	missense_variant	NSP6	1
18261	ORF1b	R6001None	18000_18025del	frameshift_variant	NSP14	1
18394	ORF1b	A6044T	18130G > A	missense_variant	NSP14	1
25549	ORF3a	L53F	157C > T	missense_variant	--	1
17964	ORF1b	M5900I	17700G > T	missense_variant	NSP13	1
6077	ORF1a	A1938S	5812G > T	missense_variant	NSP3	1
12081	ORF1a	A3939V	11816C > T	missense_variant	NSP7	1
25273	S	M1237I	3711G > T	missense_variant	--	1
26171	ORF3a	M260K	779T > A	missense_variant	--	1
6428	ORF1a	P2055S	6163C > T	missense_variant	NSP3	1
25511	ORF3a	S40L	119C > T	missense_variant	--	1
25684	ORF3a	A98T	292G > A	missense_variant	--	1
28975	N	M234I	702G > T	missense_variant	--	1
21364	ORF1b	P7034S	21100C > T	missense_variant	NSP16	1
22094	S	D178N	532G > A	missense_variant	--	1
15925	ORF1b	L5221F	15661C > T	missense_variant	NSP12	1
4410	ORF1a	A1382V	4145C > T	missense_variant	NSP3	1
23472	S	S637Y	1910C > A	missense_variant	--	1
24362	S	I934F	2800A > T	missense_variant	--	1
25463	ORF3a	T24I	71C > T	missense_variant	--	1
26522	E	Not found	26522C > T	intergenic_region	--	1
29743	ORF10	Not found	29743C > A	intergenic_region	--	1
18740	ORF1b	D6159G	18476A > G	missense_variant	NSP14	1
19666	ORF1b	D6468Y	19402G > T	missense_variant	NSP15	1
376	ORF1a	E37D	111G > T	missense_variant	NSP1	1
13610	ORF1b	Q4449L	13346A > T	missense_variant	NSP12	1
3393	ORF1a	A1043V	3128C > T	missense_variant	NSP3	1
12033	ORF1a	G3923D	11768G > A	missense_variant	NSP7	1
13329	ORF1a	T4355I	13064C > T	missense_variant	NSP10	1
15760	ORF1b	G5166C	15496G > T	missense_variant	NSP12	1
16957	ORF1b	V5565L	16693G > C	missense_variant	NSP13	1
22289	S	A243S	727G > T	missense_variant	--	1
410	ORF1a	G49C	145G > T	missense_variant	NSP1	1
2086	ORF1a	Q607H	1821G > T	missense_variant	NSP2	1
2185	ORF1a	E640D	1920G > T	missense_variant	NSP2	1
13274	ORF1a	P4337S	13009C > T	missense_variant	NSP10	1
15908	ORF1b	G5215D	15644G > A	missense_variant	NSP12	1
22661	S	V367L	1099G > C	missense_variant	--	1
23058	S	P499R	1496C > G	missense_variant	--	1
29081	N	V270L	808G > T	missense_variant	--	1
29541	N	Not found	29541C > T	intergenic_region	--	1
29648	ORF10	D31Y	91G > T	missense_variant	--	1
4856	ORF1a	G1531C	4591G > T	missense_variant	NSP3	1
29750	ORF10	Not found	29750C > T	intergenic_region	--	1
29784	ORF10	Not found	29784C > T	intergenic_region	--	1
10737	ORF1a	N3491S	10472A > G	missense_variant	NSP5	1
26447	E	S68F	203C > T	missense_variant	--	1
29140	N	Q289H	867G > T	missense_variant	--	1
25726	ORF3a	V112F	334G > T	missense_variant	--	1
29535	N	Not found	29535C > A	intergenic_region	--	1
3047	ORF1a	D928H	2782G > C	missense_variant	NSP3	1
5628	ORF1a	T1788M	5363C > T	missense_variant	NSP3	1
16700	ORF1b	R5479L	16436G > T	missense_variant	NSP13	1
25528	ORF3a	I47None	139_141del	conservative_inframe_deletion	--	1
160	5UTR	Not found	160G > A	intergenic_region	--	1
685	ORF1a	Lys141_Phe143del	421_429del	conservative_inframe_deletion	NSP1	1
4655	ORF1a	R1464W	4390C > T	missense_variant	NSP3	1
21600	S	S13I	38G > T	missense_variant	--	1
29690	ORF10	Not found	29690G > T	intergenic_region	--	1
6683	ORF1a	Y2141None	6421_6423del	conservative_inframe_deletion	NSP3	1
5310	ORF1a	T1682I	5045C > T	missense_variant	NSP3	1

12106	ORF1a	E3947D	11841G > T	missense_variant	NSP8	1
21608	S	V16F	46G > T	missense_variant	--	1
23064	S	N501I	1502A > T	missense_variant	--	1
8367	ORF1a	A2701V	8102C > T	missense_variant	NSP3	1
17561	ORF1b	R5766Q	17297G > A	missense_variant	NSP13	1
26198	ORF3a	T269M	806C > T	missense_variant	--	1
888	ORF1a	A208D	623C > A	missense_variant	NSP2	1
2912	ORF1a	L883M	2647T > A	missense_variant	NSP3	1
3180	ORF1a	E972G	2915A > G	missense_variant	NSP3	1
6449	ORF1a	L2062F	6184C > T	missense_variant	NSP3	1
7764	ORF1a	S2500F	7499C > T	missense_variant	NSP3	1
28332	N	P20L	59C > T	missense_variant	--	1
10754	ORF1a	A3497S	10489G > T	missense_variant	NSP5	1
6372	ORF1a	M2036T	6107T > C	missense_variant	NSP3	1
22241	S	V227L	679G > T	missense_variant	--	1
28925	N	A218S	652G > T	missense_variant	--	1
21810	S	L84None	250_254del	frameshift_variant	--	1
27835	ORF7b	I27T	80T > C	missense_variant	--	1
9092	ORF1a	V2943I	8827G > A	missense_variant	NSP4	1
18826	ORF1b	V6188F	18562G > T	missense_variant	NSP14	1
22592	S	A344S	1030G > T	missense_variant	--	1
29810	ORF10	Not found	29810G > T	intergenic_region	--	1
1002	ORF1a	E246G	737A > G	missense_variant	NSP2	1
9783	ORF1a	S3173N	9518G > A	missense_variant	NSP4	1
28086	ORF8	A65S	193G > T	missense_variant	--	1
3304	ORF1a	E1013D	3039G > T	missense_variant	NSP3	1
5492	ORF1a	S1743A	5227T > G	missense_variant	NSP3	1
23009	S	V483F	1447G > T	missense_variant	--	1
26230	ORF3a	Not found	26230G > T	intergenic_region	--	1
28892	N	P207T	619C > A	missense_variant	--	1
13625	ORF1b	D4454A	13361A > C	missense_variant	NSP12	1
18530	ORF1b	I6089T	18266T > C	missense_variant	NSP14	1
5943	ORF1a	D1893G	5678A > G	missense_variant	NSP3	1
20759	ORF1b	A6832V	20495C > T	missense_variant	NSP16	1
17330	ORF1b	E5689G	17066A > G	missense_variant	NSP13	1
26204	ORF3a	T271I	812C > T	missense_variant	--	1
21410	ORF1b	P7049L	21146C > T	missense_variant	NSP16	1
22328	S	S256None	767_770del	frameshift_variant	--	1
25571	ORF3a	S60F	179C > T	missense_variant	--	1
3096	ORF1a	S944L	2831C > T	missense_variant	NSP3	1
17122	ORF1b	A5620S	16858G > T	missense_variant	NSP13	1
1655	ORF1a	V464I	1390G > A	missense_variant	NSP2	1
21005	ORF1b	A6914V	20741C > T	missense_variant	NSP16	1
29394	N	K374T	1121A > C	missense_variant	--	1
9996	ORF1a	S3244L	9731C > T	missense_variant	NSP4	1
20224	ORF1b	E6654K	19960G > A	missense_variant	NSP15	1
12738	ORF1a	T4158I	12473C > T	missense_variant	NSP9	1
15906	ORF1b	Q5214H	15642G > T	missense_variant	NSP12	1
25522	ORF3a	G44R	130G > A	missense_variant	--	1
10889	ORF1a	R3542C	10624C > T	missense_variant	NSP5	1
21817	S	F86None	258delT	frameshift_variant	--	1
23045	S	G496None	1485_1518del	frameshift_variant	--	1
25623	ORF3a	H78None	234_267del	frameshift_variant	--	1
3768	ORF1a	T1168I	3503C > T	missense_variant	NSP3	1
9182	ORF1a	V2973I	8917G > A	missense_variant	NSP4	1
13721	ORF1b	P4486L	13457C > T	missense_variant	NSP12	1
29651	ORF10	V32L	94G > T	missense_variant	--	1
25418	ORF3a	T9K	26C > A	missense_variant	--	1
28851	N	S193I	578G > T	missense_variant	--	1
282	ORF1a	P6L	17C > T	missense_variant	NSP1	1
335	ORF1a	R24C	70C > T	missense_variant	NSP1	1
4296	ORF1a	T1344I	4031C > T	missense_variant	NSP3	1
25218	S	G1219V	3656G > T	missense_variant	--	1

27401	ORF7a	I3S	8T > G	missense_variant	--	1
26176	ORF3a	P262S	784C > T	missense_variant	--	1
28077	ORF8	V62L	184G > T	missense_variant	--	1
5526	ORF1a	T1754I	5261C > T	missense_variant	NSP3	1
6633	ORF1a	A2123V	6368C > T	missense_variant	NSP3	1
20251	ORF1b	I6663V	19987A > G	missense_variant	NSP15	1
22975	S	E471D	1413A > C	missense_variant	--	1
23664	S	A701V	2102C > T	missense_variant	--	1
27406	ORF7a	L5F	13C > T	missense_variant	--	1
7405	ORF1a	M2380I	7140G > T	missense_variant	NSP3	1
19086	ORF1b	K6274N	18822G > T	missense_variant	NSP14	1
28077	ORF8	V62M	184G > A	missense_variant	--	1
28727	N	A152S	454G > T	missense_variant	--	1
29545	N	Not found	29545C > T	intergenic_region	--	1
10977	ORF1a	A3571V	10712C > T	missense_variant	NSP6	1
14999	ORF1b	S4912N	14735G > A	missense_variant	NSP12	1
18225	ORF1b	M5987I	17961G > T	missense_variant	NSP14	1
20762	ORF1b	T6833I	20498C > T	missense_variant	NSP16	1
29542	N	Not found	29542A > T	intergenic_region	--	1
29737	ORF10	Not found	29737G > T	intergenic_region	--	1
16852	ORF1b	G5530C	16588G > T	missense_variant	NSP13	1
24374	S	L938F	2812C > T	missense_variant	--	1
24608	S	A1016T	3046G > A	missense_variant	--	1
17012	ORF1b	S5583L	16748C > T	missense_variant	NSP13	1
25685	ORF3a	A98V	293C > T	missense_variant	--	1
7038	ORF1a	G2258A	6773G > C	missense_variant	NSP3	1
27476	ORF7a	T28I	83C > T	missense_variant	--	1
935	ORF1a	T224P	670A > C	missense_variant	NSP2	1
12513	ORF1a	T4083M	12248C > T	missense_variant	NSP8	1
7548	ORF1a	N2428I	7283A > T	missense_variant	NSP3	1
18646	ORF1b	P6128S	18382C > T	missense_variant	NSP14	1
21614	S	L18I	52C > A	missense_variant	--	1
21615	S	L18H	53T > A	missense_variant	--	1
127	5UTR	Not found	127G > T	intergenic_region	--	1
22329	S	S256L	767C > T	missense_variant	--	1
25355	S	L1265F	3793C > T	missense_variant	--	1
25690	ORF3a	G100C	298G > T	missense_variant	--	1
23052	S	Phe497_Gln498delinsLeu	1491_1493delCCA	disruptive_inframe_deletion	--	1
13458	ORF1b	S4398L	13193C > T	missense_variant	NSP12	1
29781	ORF10	Not found	29781G > T	intergenic_region	--	1
6672	ORF1a	D2136G	6407A > G	missense_variant	NSP3	1
16750	ORF1b	P5496S	16486C > T	missense_variant	NSP13	1
25855	ORF3a	D155H	463G > C	missense_variant	--	1

Supplementary Table 2. Distribution of the Moroccan mutations per lineages.

Gene | Nucleotide changes	Lineages	Number of samples
S | GAGTTCA22028G	AY.37	1
ORF1ab | A11201G	AY.122	13
ORF1ab | C16466T	AY.122	14
S | C22458T	AY.33	61
S | C22323T	AY.112	1
ORF7a | C27752T	AY.33	45
ORF1ab | G15451A	AY.34	5
ORF1ab | C14408T	AY.122	13
CHR_START-ORF1ab | G210T	AY.46	1
S | GAGTTCA22028G	AY.34	3
ORF10-CHR_END | G29742T	AY.51	6
M | T26767C	AY.112	2
ORF7b | C27874T	AY.46.6	2
N | G29402T	AY.33	69
ORF1ab | T3648C	B.1.617.2	8

S | T22917G	AY.51	6
ORF1ab | C6402T	AY.34	5
S | A23403G	AY.112	2
S | C23604G	AY.122	13
ORF1ab | C5184T	AY.33	1
CHR_START-ORF1ab | C241T	AY.33	74
S | GAGTTCA22028G	AY.46.6	2
ORF1ab | C14408T	AY.73	6
N | G29402T	AY.122	13
S | G21987A	AY.34	2
S | C22995A	B.1.617.2	15
N | A28461G	AY.51	6
ORF1ab | G15451A	AY.33	76
ORF1ab | C7124T	B.1.617.2	10
ORF1ab | G9053T	AY.43	12
S | G24410A	AY.43	12
S | C23604G	AY.46.6	2
ORF1ab | C5184T	AY.34	1
ORF1ab | C19220T	B.1.617.2	10
S | C23604G	B.1.617.2	15
ORF1ab | C5184T	B.1.617.2	5
S | G24410A	AY.112	2
ORF1ab | C10029T	AY.34	5
ORF1ab | C10029T	AY.44	2
ORF1ab | G15451A	AY.43	12
N | A28461G	AY.46.6	2
S | G23401T	AY.39	1
S | C23604G	AY.100	1
ORF7a | T27638C	AY.34	3
S | C22995A	AY.43	12
M | T26767C	AY.37	1
CHR_START-ORF1ab | G210T	AY.122	13
ORF10-CHR_END | G29742T	AY.33	71
S | C21846T	AY.34	2
N | G28881T	AY.39	24
ORF7a | C27752T	AY.100	1
S | C22227T	AY.37	1
M | T26767C	AY.44	2
N | G29402T	AY.46	1
S | GAGTTCA22028G	AY.112	1
S | T22917G	AY.34	5
S | C21618G	AY.122	13
ORF1ab | A11201G	B.1.617.2	10
ORF1ab | C16466T	AY.51	6
CHR_START-ORF1ab | G210T	AY.73	6
ORF1ab | C19220T	AY.44	2
ORF7a | C27752T	B.1.617.2	13
ORF10-CHR_END | G29742T	AY.37	1
CHR_START-ORF1ab | G210T	AY.44	2
ORF1ab | C16466T	AY.37	1
N | A28461G	AY.44	2
ORF1ab | G9053T	AY.46.6	2
ORF3a | C25469T	AY.39	24
ORF3a | C25469T	AY.122	13
N | A28461G	AY.33	73
S | G24410A	AY.122	12
ORF7a | C27752T	AY.46.6	2
ORF1ab | G4181T	B.1.617.2	10
ORF1ab | C6402T	AY.46	1
ORF7a | T27638C	AY.44	2
ORF8 | AGATTTC28247A	AY.122	12
ORF1ab | C5184T	AY.73	6
ORF7b | C27874T	AY.39	24

ORF8 | AGATTTC28247A	AY.33	65
ORF1ab | C9891T	AY.51	6
ORF1ab | C10029T	AY.112	2
S | C22323T	AY.39	3
ORF1ab | C10029T	AY.39	24
ORF7a | C27752T	AY.112	2
ORF1ab | C7124T	AY.46	1
M | T26767C	AY.46	1
N | G28881T	AY.33	74
N | G28881T	AY.44	2
S | C21618G	AY.112	2
ORF1ab | A11201G	AY.100	1
N | G28916T	AY.100	1
ORF1ab | C14408T	AY.112	2
ORF1ab | C16466T	AY.39	24
ORF1ab | C5184T	AY.51	6
ORF1ab | C6402T	AY.39	9
ORF1ab | G15451A	AY.44	2
ORF3a | C25469T	AY.51	6
S | G21987A	B.1.617.2	10
CHR_START-ORF1ab | G210T	AY.39	24
ORF1ab | C6402T	AY.100	1
S | G21987A	AY.51	5
S | T22917G	AY.73	6
ORF7a | T27638C	AY.73	4
S | C21618G	AY.51	6
S | A23403G	AY.46.6	2
S | C22311T	AY.33	61
ORF8 | AGATTTC28247A	AY.44	2
ORF1ab | C6402T	AY.44	2
M | T26767C	AY.39	24
ORF3a | C25469T	AY.33	74
N | G28881T	AY.51	6
N | G29402T	AY.51	6
CHR_START-ORF1ab | G210T	B.1.617.2	15
N | G28881T	AY.112	2
ORF1ab | C16466T	AY.73	6
S | A23403G	AY.33	74
CHR_START-ORF1ab | C241T	AY.122	13
S | C21618G	AY.100	1
ORF1ab | C10029T	B.1.617.2	10
N | G28916T	B.1.617.2	10
N | G28881T	AY.73	6
ORF1ab | A11201G	AY.43	7
S | C21618G	AY.39	22
ORF1ab | C14408T	AY.43	12
ORF1ab | C14408T	B.1.617.2	15
S | GAGTTCA22028G	AY.51	6
S | T22917G	B.1.617.2	15
ORF3a | C25469T	AY.73	6
ORF3a | G25647T	AY.112	2
ORF7a | C27752T	AY.73	4
N | G29402T	AY.37	1
S | T22917G	AY.37	1
ORF7a | T27638C	AY.33	47
ORF7b | C27874T	AY.34	5
N | A28461G	AY.112	2
ORF7a | T27638C	AY.51	6
N | G28916T	AY.112	2
ORF1ab | C5184T	AY.39	1
ORF1ab | C9891T	AY.37	1
N | G29402T	AY.34	4
S | A21647G	AY.33	72

ORF1ab | G9053T	AY.100	1
N | G28881T	B.1.617.2	15
M | T26767C	AY.51	6
N | A28461G	AY.46	1
S | C23604G	AY.73	6
ORF3a | C25469T	AY.100	1
S | C21618G	AY.73	6
S | C22995A	AY.33	74
ORF7b | C27874T	AY.44	2
ORF1ab | C16466T	AY.34	5
ORF1ab | C10029T	AY.46.6	2
ORF1ab | C6402T	AY.122	13
N | A28461G	AY.43	12
ORF1ab | C6402T	AY.33	73
S | A23403G	AY.37	1
S | C22323T	AY.44	1
N | G29402T	AY.73	6
M | T26767C	AY.46.6	2
M | T26767C	B.1.617.2	15
N | G28881T	AY.100	1
ORF7b | C27874T	AY.122	13
ORF1ab | C16466T	B.1.617.2	15
S | C22227T	AY.39	1
N | G28881T	AY.46	1
S | C21618G	AY.44	2
ORF1ab | A11201G	AY.112	2
ORF1ab | C10029T	AY.43	4
ORF8 | AGATTTC28247A	AY.34	3
N | G28916T	AY.46	1
ORF1ab | A11201G	AY.46.6	1
ORF1ab | G15451A	AY.73	6
M | T26767C	AY.73	6
ORF1ab | G4181T	AY.43	12
N | A28461G	AY.34	5
CHR_START-ORF1ab | C241T	AY.43	12
ORF3a | C25469T	B.1.617.2	15
ORF1ab | G15451A	AY.46	1
ORF3a | G25647T	AY.39	4
ORF8 | AGATTTC28247A	AY.46	1
ORF3a | G25647T	AY.44	1
ORF1ab | C14408T	AY.46	1
ORF1ab | C7124T	AY.39	18
S | C22995A	AY.100	1
S | C22323T	B.1.617.2	6
CHR_START-ORF1ab | C241T	AY.112	2
ORF1ab | C19220T	AY.43	12
S | G24410A	AY.37	1
ORF1ab | G15451A	AY.100	1
ORF1ab | C7124T	AY.43	11
ORF1ab | A11201G	AY.44	2
ORF7a | C27752T	AY.51	6
ORF1ab | C16466T	AY.44	2
S | A23403G	AY.46	1
ORF7b | C27874T	AY.46	1
ORF1ab | G4181T	AY.34	3
ORF1ab | C7124T	AY.44	2
CHR_START-ORF1ab | G210T	AY.33	74
S | G24410A	B.1.617.2	14
ORF1ab | C7124T	AY.46.6	2
S | G21987A	AY.43	1
S | G24410A	AY.73	6
S | C21618G	AY.46.6	2
ORF1ab | C7124T	AY.34	4

ORF7a | T27638C	AY.39	1
ORF1ab | C9891T	AY.73	6
S | G24410A	AY.100	1
ORF1ab | A11201G	AY.46	1
ORF1ab | C14408T	AY.44	2
ORF1ab | C14408T	AY.39	24
S | C23604G	AY.37	1
S | C23604G	AY.112	2
ORF10-CHR_END | G29742T	AY.122	13
ORF1ab | C7124T	AY.122	13
S | C22227T	AY.73	6
ORF1ab | C1191T	AY.51	6
S | T22917G	AY.43	12
ORF1ab | C19220T	AY.112	2
N | G29402T	AY.43	11
S | G24410A	AY.44	2
CHR_START-ORF1ab | G210T	AY.37	1
ORF7a | T27638C	B.1.617.2	14
ORF1ab | C10029T	AY.33	72
S | A23403G	AY.51	6
ORF1ab | G9053T	AY.33	74
S | C21618G	AY.43	11
ORF1ab | T11418C	AY.37	1
S | C23604G	AY.46	1
CHR_START-ORF1ab | C241T	AY.46	1
S | A23403G	B.1.617.2	15
S | T22917G	AY.33	74
ORF10-CHR_END | G29742T	AY.39	24
ORF8 | AGATTTC28247A	B.1.617.2	15
ORF1ab | C19220T	AY.34	5
CHR_START-ORF1ab | G210T	AY.34	4
CHR_START-ORF1ab | C241T	AY.39	24
ORF7a | C27752T	AY.44	2
ORF1ab | C14408T	AY.37	1
S | A23403G	AY.43	12
ORF1ab | C6402T	B.1.617.2	10
ORF1ab | C10029T	AY.46	1
ORF3a | C25469T	AY.37	1
N | G29402T	B.1.617.2	15
ORF1ab | C9891T	B.1.617.2	5
S | C23604G	AY.34	5
S | G21987A	AY.122	7
S | G24410A	AY.34	5
S | G23401T	AY.122	1
ORF1ab | C6402T	AY.46.6	1
ORF1ab | C19220T	AY.39	24
S | G24410A	AY.46.6	2
S | T22917G	AY.44	2
ORF7a | C27752T	AY.122	11
S | A23403G	AY.34	5
S | G24410A	AY.33	72
ORF7a | T27638C	AY.46.6	2
N | G29402T	AY.44	2
ORF1ab | C19220T	AY.33	73
S | C23604G	AY.43	12
ORF7a | T27638C	AY.100	1
ORF3a | C25469T	AY.112	2
N | A28461G	AY.37	1
M | T26767C	AY.43	12
ORF8 | AGATTTC28247A	AY.43	11
ORF1ab | G4181T	AY.46	1
ORF1ab | C6402T	AY.112	2
S | A23403G	AY.44	2

S | C21618G	AY.33	74
S | G24410A	AY.39	24
N | G28916T	AY.43	12
S | GAGTTCA22028G	AY.73	4
ORF1ab | G9053T	AY.34	5
M | T26767C	AY.34	5
ORF8 | AGATTTC28247A	AY.100	1
ORF1ab | C6402T	AY.43	6
N | G28881T	AY.46.6	2
ORF7b | C27874T	AY.33	72
ORF1ab | A11201G	AY.33	73
ORF1ab | C14408T	AY.34	5
N | G28881T	AY.122	13
N | A28461G	AY.39	24
S | C21618G	AY.34	4
ORF10-CHR_END | G29742T	AY.34	5
ORF1ab | C19220T	AY.46	1
CHR_START-ORF1ab | G210T	AY.43	12
ORF1ab | C19220T	AY.46.6	2
CHR_START-ORF1ab | C241T	AY.34	4
S | T22917G	AY.46.6	2
ORF10-CHR_END | G29742T	AY.100	1
S | T22917G	AY.39	24
ORF1ab | C14408T	AY.100	1
ORF10-CHR_END | G29742T	B.1.617.2	14
ORF7a | T27638C	AY.37	1
ORF7b | C27874T	AY.112	2
ORF10-CHR_END | G29742T	AY.46.6	2
S | G24410A	AY.51	5
S | GAGTTCA22028G	B.1.617.2	14
ORF1ab | G15451A	AY.122	13
ORF7a | T27638C	AY.122	11
S | GAGTTCA22028G	AY.46	1
ORF1ab | G4181T	AY.39	24
N | A28461G	AY.73	5
M | T26767C	AY.33	74
S | C21618G	AY.37	1
ORF1ab | G9053T	AY.46	1
ORF1ab | T11418C	B.1.617.2	5
S | C21846T	AY.39	1
ORF1ab | A11201G	AY.39	14
ORF1ab | C14408T	AY.46.6	2
ORF1ab | C14408T	AY.51	6
ORF1ab | C16466T	AY.46.6	2
ORF7a | C27752T	AY.43	6
N | G29402T	AY.46.6	2
S | GAGTTCA22028G	AY.33	57
N | G28881T	AY.34	5
ORF1ab | G15451A	AY.37	1
N | G28916T	AY.33	74
S | A23403G	AY.39	24
ORF1ab | G9053T	B.1.617.2	10
CHR_START-ORF1ab | C241T	B.1.617.2	15
ORF10-CHR_END | G29742T	AY.73	6
ORF1ab | C7124T	AY.33	70
ORF10-CHR_END | G29742T	AY.43	12
ORF1ab | C10029T	AY.122	13
N | G28881T	AY.37	1
S | C22995A	AY.39	24
CHR_START-ORF1ab | C241T	AY.51	6
N | G29402T	AY.100	1
ORF8 | AGATTTC28247A	AY.39	24
S | C23604G	AY.51	6

S | C23604G	AY.44	2
ORF1ab | C14408T	AY.33	74
S | C22995A	AY.46	1
ORF1ab | T11418C	AY.73	6
ORF1ab | G9053T	AY.112	2
ORF1ab | T3648C	AY.39	4
ORF1ab | C16466T	AY.33	72
S | A23403G	AY.100	1
S | C22995A	AY.37	1
ORF1ab | C16466T	AY.46	1
ORF7b | C27874T	B.1.617.2	10
S | C21618G	B.1.617.2	15
S | GAGTTCA22028G	AY.122	10
S | G21987A	AY.73	2
ORF1ab | C16466T	AY.43	16
ORF3a | C25469T	AY.43	12
S | C23604G	AY.39	24
ORF3a | G25647T	B.1.617.2	8
ORF1ab | G9053T	AY.122	13
S | G24410A	AY.46	1
ORF1ab | T3648C	AY.112	2
S | A23403G	AY.122	13
ORF1ab | C10029T	AY.100	1
S | C22227T	AY.33	5
ORF8 | AGATTTC28247A	AY.112	2
ORF1ab | C19220T	AY.122	13
N | G28916T	AY.34	5
CHR_START-ORF1ab | C241T	AY.73	6
S | C21846T	B.1.617.2	7
ORF8 | AGATTTC28247A	AY.51	6
ORF3a | C25469T	AY.46	1
ORF1ab | C1191T	B.1.617.2	5
ORF1ab | G15451A	B.1.617.2	15
ORF1ab | G15451A	AY.46.6	2
S | C22995A	AY.44	2
ORF7a | T27638C	AY.43	6
M | T26767C	AY.122	12
ORF1ab | G15451A	AY.112	2
ORF3a | C25469T	AY.34	5
N | G28881T	AY.43	12
ORF10-CHR_END | G29742T	AY.112	2
ORF7b | C27874T	AY.100	1
ORF7b | C27874T	AY.43	12
ORF1ab | C7124T	AY.112	2
CHR_START-ORF1ab | G210T	AY.112	2
S | G21987A	AY.46.6	1
S | T22917G	AY.112	2
ORF1ab | G4181T	AY.44	2
S | C22995A	AY.46.6	2
N | G28916T	AY.39	24
CHR_START-ORF1ab | G210T	AY.51	6
S | G21987A	AY.33	23
S | C22995A	AY.51	6
S | C21618G	AY.46	1
ORF1ab | G4181T	AY.46.6	2
M | T26767C	AY.100	1
ORF8 | AGATTTC28247A	AY.46.6	2
S | C23604G	AY.33	74
N | A28461G	B.1.617.2	15
CHR_START-ORF1ab | G210T	AY.46.6	2
ORF1ab | G9053T	AY.39	23
ORF1ab | A11201G	AY.34	5
ORF7a | C27752T	AY.37	1

S | T22917G	AY.100	1
ORF1ab | G15451A	AY.39	24
S | GAGTTCA22028G	AY.39	21
S | C22995A	AY.73	6
ORF1ab | G9053T	AY.44	2
ORF1ab | G4181T	AY.100	1
CHR_START-ORF1ab | G210T	AY.100	1
N | G29402T	AY.39	23
S | C21846T	AY.112	1
ORF1ab | T3648C	AY.44	1
S | C22995A	AY.34	5
ORF1ab | C19220T	AY.100	1
CHR_START-ORF1ab | C241T	AY.37	1
CHR_START-ORF1ab | C241T	AY.44	2
ORF1ab | C16466T	AY.100	1
ORF1ab | G4181T	AY.33	74
S | T22917G	AY.122	13
N | G28916T	AY.44	2
N | G28916T	AY.46.6	2
ORF8 | AGATTTC28247A	AY.73	5
ORF1ab | T11418C	AY.51	6
N | A28461G	AY.100	1
ORF1ab | C16466T	AY.112	2
S | GAGTTCA22028G	AY.100	1
ORF1ab | G4181T	AY.112	2
CHR_START-ORF1ab | C241T	AY.100	1
ORF3a | C25469T	AY.46.6	2
ORF10-CHR_END | G29742T	AY.44	2
N | G29402T	AY.112	2
S | C22227T	B.1.617.2	1
S | GAGTTCA22028G	AY.43	8
ORF1ab | C7124T	AY.100	1
S | C22995A	AY.112	2
ORF1ab | G4181T	AY.122	13
N | G28916T	AY.122	13
ORF7a | T27638C	AY.112	2
S | G23401T	AY.33	66
S | C22995A	AY.122	13
ORF3a | C25469T	AY.44	2
ORF1ab | C5184T	AY.37	1
ORF8 | AGATTTC28247A	AY.37	1
ORF10-CHR_END | G29742T	AY.46	1
ORF7a | C27752T	AY.34	3
N | A28461G	AY.122	13
CHR_START-ORF1ab | C241T	AY.46.6	2
S | T22917G	AY.46	1
S | A23403G	AY.73	6
ORF1ab | G15451A	AY.51	6

Supplementary Table 3. Clades mutation analysis.

Residue change	Nucleotide position	Effect	Gene	Nucleotide change	Clade	Number of Moroccan samples with the mutation	Number of samples with the mutation	Description of the samples
Y6160Y	18744	synonymous_variant	ORF1ab	18480C > T	cladeI	2	74	13Den_3Bra _13Chi_21Fra _7Ita_2Mor _10Ger_2USA _3Jap_1Eng_1Indo
V4887V	14925	synonymous_variant	ORF1ab	14661C > T	cladeD	2	24	1Ita_7Jap_ 5Ger_2SKo _2Mor_8Den_1Fra
I850L	24110	missense_variant	S	2548A > C	cladeD	2	22	6Jap_6Ger _2SKo_2Mor _7Den_1Fra
K16T	25439	missense_variant	ORF3a	47A > C	cladeD	2	21	6Jap_5Ger _2SKo_2Mor _7Den_1Fra
T6891T	20937	synonymous_variant	ORF1ab	20673G > T	cladeK	1	43	5Jap_12Fra _8Ger_1Eng _9Chi_1Bra _1Gha_2Den _2Ita_1SKo _1Mor_1Can
T29A	21647	missense_variant	S	85A > G	cladeG	72	7	72Mor_3Ger _1Den_1Chi _1Indo_1Fra
E239E	26109	synonymous_variant	ORF3a	717G > A	cladeB	5	10	4Fra_2Bra _5Mor_1Eng _3Ger
Q677H	23593	missense_variant	S	2031G > C	cladeB	5	16	4Fra_2Bra _1Eng_1Aus _2Ind_4Ger _5Mor_2USA
P207L	28893	missense_variant	N	620C > T	cladeL	6	1	6Mor_1Fra
P46S	28409	missense_variant	N	136C>T	cladeL	6	1	6Mor_1Fra
V3986V	12223	synonymous_variant	ORF1ab	11958G > T	cladeM	6	2	6Mor_1Ita_1Ger
E102V	25697	missense_variant	ORF3a	305A > T	cladeM	6	2	6Mor_1Ita_1Ger

Supplementary Table 4. Probabilities of main nodes in the ancestral construction.

Nodes code	predicted probability	Country	Nodes
n357357	0.68	Ind	Node 19
n106	0.76	Mor	Node 18
n0	0.86	Ind	Node 0
n49	0.89	Den	Node 14
n5858	0.93	Fra	Node 13
n88	0.94	Mor	Node 9
n7272	0.94	Ger	Node 16
n212	0.96	Mor	Node 2
n126126	0.98	Jap	Node 11
n4	0.99	Mor	Node 8
n135	0.99	Mor	Node 3
n23	1.00	Ger	Node 9
n4545	1.00	Mor	Node 4
n415	1.00	Mor	Node 5
n330	1.00	Mor	Node 6
n399	1.00	Mor	Node 4
n4949	1.00	Mor	Node 14
n256	1.00	Gha	Node 10

n143143	1.00	Mor	Node 7
Mor117	1.00	Mor	Node 1
Mor119	1.00	Mor	Node 1
Mor120	1.00	Mor	Node 1
Mor121	1.00	Mor	Node 1
Mor122	1.00	Mor	Node 1
Mor124	1.00	Mor	Node 1
Mor125	1.00	Mor	Node 1
Mor126	1.00	Mor	Node 1
Mor127	1.00	Mor	Node 1
Mor128	1.00	Mor	Node 1
Mor129	1.00	Mor	Node 1
Mor130	1.00	Mor	Node 1
Mor131	1.00	Mor	Node 1
Mor132	1.00	Mor	Node 1
Mor133	1.00	Mor	Node 1
Mor134	1.00	Mor	Node 1
Mor135	1.00	Mor	Node 1
Mor136	1.00	Mor	Node 1
Mor137	1.00	Mor	Node 1
Mor138	1.00	Mor	Node 1
Mor140	1.00	Mor	Node 1
Mor156	1.00	Mor	Node 1
Mor38	1.00	Mor	Node 1
Mor49	1.00	Mor	Node 1
Mor118	1.00	Mor	Node 10
Mor123	1.00	Mor	Node 10
Mor19	1.00	Mor	Node 10
Mor151	1.00	Mor	Node 11
Mor14	1.00	Mor	Node 12
Mor109	1.00	Mor	Node 13
Mor158	1.00	Mor	Node 13
Mor159	1.00	Mor	Node 13
Mor9	1.00	Mor	Node 13
Mor84	1.00	Mor	Node 15
Mor59	1.00	Mor	Node 16
Fra13	1.00	Fra	Node 17
Fra64	1.00	Fra	Node 17
Fra73	1.00	Fra	Node 17
Mor58	1.00	Mor	Node 17
Bra37	1.00	Bra	Node 18
Bra6	1.00	Bra	Node 18
Eng26	1.00	Eng	Node 18
Fra58	1.00	Fra	Node 18
Ger48	1.00	Ger	Node 18
Ger58	1.00	Ger	Node 18
Ger95	1.00	Ger	Node 18
Fra87	1.00	Fra	Node 19
Gha14	1.00	Gha	Node 19
Gha29	1.00	Gha	Node 19
Jap43	1.00	Jap	Node 19
Jap44	1.00	Jap	Node 19
Mor70	1.00	Mor	Node 19
Ger24	1.00	Ger	Node 5
Ita22	1.00	Ita	Node 5
Fra24	1.00	Fra	Node 6
Chi11	1.00	Chi	Node 8
Chi65	1.00	Chi	Node 9
Den93	1.00	Den	Node 9
Fra63	1.00	Fra	Node 9
Ger75	1.00	Ger	Node 9
Indo52	1.00	Indo	Node 9