Genetic analysis of SARS-CoV-2 strains collected from North Africa: viral origins and mutational spectrum

Meriem LAAMARTI[1], Souad KARTTI[1], Tarek ALOUANE[1], Rokia LAAMARTI[2], Loubna ALLAM[1],Mouna OUADGHIRI[1], M.W. CHEMAO-ELFIHRI[1], Imane SMYEJ[3], Jalila RAHOUI[3], Houda BENRAHMA[3], Idrissa DIAWARA[3], Abdelomunim ESSABBAR[1], Nasma BOUMAJDI[1], Houda BENDANI[1], El Mehdi BOURICHA[1], Tarik AANNIZ[1], Jalil ELATTAR[4], Naima EL HAFIDI[1], Rachid EL JAOUDI[1], Laila SBABOU[5], Chakib NEJJARI[6], Saaid AMZAZI[7], Rachid MENTAG[8], Lahcen BELYAMANI[9] and Azeddine IBRAHIMI1[1]*

[1] Medical Biotechnology Laboratory (MedBiotech), Bioinova Research Center, Rabat Medical & Pharmacy School, Mohammed Vth University in Rabat, Morocco.
[2] Medical Biotechnology Center, Moroccan Foundation for Science, Innovation & Research (MAScIR), Rabat, Morocco.
[3] National Reference Laboratory, Mohammed VI University of Health Sciences (UM6SS), Casablanca, Morocco.
[4] Laboratoire Riad, hay Riad, Rabat, Morocco.
[5] Research Center of Plants and Microbial Biotechnologies, Biodiversity and Environment. of Microbiology and Molecular Biology Team, Faculty of Sciences, Mohammed V University in Rabat, Morocco.
[6] International School of Public Health, Mohammed VI University of Health Sciences (UM6SS), Casablanca, Morocco.
[7] Laboratory of Human Pathologies Biology, Faculty of Sciences, Mohammed V University in Rabat, Morocco.
[8] Biotechnology Unit, Regional Center of Agricultural Research of Rabat, National Institute of Agricultural Research, Rabat, Morocco.
[9] Emergency Department, Military Hospital Mohammed V, Rabat Medical & Pharmacy School, Mohammed Vth University in Rabat, Morocco.

Short title: Genomic analysis of North African isolates

*** Corresponding author: a.ibrahimi@um5s.net.ma**

Abstract

In Morocco two waves of SARS-CoV-2 infections have been recorded. The first one occurred from March 02, 2020 with infections mostly imported from Europe and the second one dominated by local infections.

At the time of writing, the genetic diversity of Moroccan isolates of SARS-CoV-2 has not yet been reported. The present study aimed to analyze first the genomic variation of the twenty-eight Moroccan strains of SARS-CoV-2 isolated from March 03, 2020 to May 15, 2020, to compare their distributions with twelve other viral genomes from North Africa as well as to identify their possible sources.

Our finding revealed 61 mutations in the Moroccan genomes of SARS-CoV-2 compared to the reference sequence Wuhan-Hu-1/2019, of them 23 (37.7%) were present in two or more genomes. Focusing on non-synonymous mutations, 29 (47.54%) were distributed in five genes (ORF1ab, spike, membrane, nucleocapsid and ORF3a) with variable frequencies. The non-structural protein coding regions nsp3-Multi domain and nsp12-RdRp of the ORF1ab gene harbored more mutations, with six for each. The comparison of genetic variants of fourty North African strains revealed that two non-synonymous mutations D614G (in spike) and Q57H (in ORF3a) were common in four countries (Morocco, Tunisia, Algeria and Egypt), with a prevalence of 92.5% (n = 37) and 42.5% (n = 17), respectively, of the total genomes.

Phylogenetic analysis showed that the Moroccan and Tunisian SARS-CoV-2 strains were closely related to those from different origins (Asia, Europe, North and South America) and distributed in different distinct subclades. This could indicate different sources of infection with no specific strain dominating yet in in these countries. These results have the potential to lead to new comprehensive investigations combining genomic data, epidemiological information and the clinical characteristics of patients with SARS-CoV-2.

Keywords: SARS-CoV-2, Morocco, North African strains, mutations, spike protein, RdRp, Phylogeny.

Introduction

The new coronavirus 2019, also known as Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) (1) is the causative agent of COVID-19, a new type of pneumonia that caused in late December, 2020, an epidemic in Wuhan, China, and then spread to 215 countries around the world. In February, 2020, COVID-19 was emerged in North African countries, notably in Egypt, Tunisia, Algeria and Morocco (2, 3). The first case was reported in Egypt on February 14, followed by Algeria on February 25, then Morocco and Tunisia on the same day, March 2, 2020 (2, 3). Due to the rapid transmission of viruses in the 5 continents and the large number of confirmed cases, the World Health Organization (WHO) has declared (March 11, 2020) COVID-19 as a global pandemic (4). As of June 26th 9,473,214 and 484,249 (5.11%) of confirmed and deceased cases, respectively, have been reported worldwide (5). It should be noted that mortality from SARS-CoV-2 differs considerably according to the geographic region. USA has the largest population of confirmed cases (2,367,064) and deaths (121,645) (5). Meanwhile, South America and Europe were also hit hard with 1,188,631 and 620,794 confirmed cases in Brazil and Russia, on their respective continents, while the African region had the least number of cases, with 258,752 (5).

SARS-CoV-2 is a single-stranded positive-sense RNA virus, coding for four structural proteins (spike (S), envelope (E), membrane (M) and nucleocapsid (N)), 16 non-structural proteins (nsp1 to nsp16) and several accessory proteins (ORF3a, ORF6, ORF7a, ORF7b, and ORF8) (6,7). Protein S which s responsible for binding to membrane receptors in host cells (ACE2) *via* its receptor-binding domain (RBD), therefore is considered as the most important target for candidate vaccines (8,9,10).

It is known that the mutation rate of the RNA virus contributes to viral adaptation, creating a balance between the integrity of genetic information and the variability of the genome, thus allowing viruses to escape host immunity and develop drug resistance (11,12). Our recent study (13) based on the analysis of 30,983 genomes of SARS-CoV-2 variants belonging to 80 countries, revealed 5.67% of total mutations with a frequency greater than 1% of all the sequences analyzed suggesting that this virus is not yet adapted to its host.

The genetic variants of the Moroccan strains of SARS-CoV-2 and their distribution

along the viral genome are not yet documented. In the present study, we investigated the genomic diversity of twenty-eight SARS-CoV-2 strains that emerged in Morocco between March 3th and May 15th, 2020 with s six new genomes presented for the first time. Next, we compared the distribution of these SARS-CoV-2 variants with twelve other genomes from North Africa (Tunisia, Algeria and Egypt). In addition, the identification of the possible source of the Moroccan strains was carried out by comparing them with genomes from Africa, Asia, Europe, North and South America and Oceania.

Materials and Methods

Genomes sequencing

From the viral RNA extracted from six clinical samples, the cDNA was synthesized using reverse transcriptase with random hexamers, then amplified for genomes enrichment using Q5 Hot Start High- Fidelity DNA Polymerase (NEB) using a set of primers targeting regions of the SARS-CoV-2 genome designed by ARTIC network (https://artic.network/ncov-2019). The PCR products weres purified by adding equal volume of AMPure XP beads (Beckman Coulter). The sequencing was performed according to the eight-hour routine workflow and amplicons were repaired with NEBNext FFPE Repair Mix (NEB), followed by the DNA ends preparation using NEBNext End repair/ dA-tailing Module (NEB) before adding native barcodes and sequencing adapters supplied in the EXP-NBD104/114 kit (Nanopore) to the DNA ends. After priming the flow cell, 60 ng DNA per sample were pooled with a final volume of 65 µL. Following the ligation sequencing kit (SQK-LSK109) protocol, MinION Mk1B was used to perform the genome sequencing on an R9.4.1 flow cell.

Variant calling analysis

A set of 40 SARS-CoV-2 genomes: 28 from Morocco, including six sequenced in the present study, 7 from Tunisia, 3 from Algeria, and 2 from Egypt, were downloaded from GISAID database (http://www.gisaid.org/) (14) (**Table 1**).

The reads generated by MinION Nanopore-Oxford of the six isolates were mapped to the reference sequence genome Wuhan-Hu-1/2019 using BWA-MEM v0.7.17-r1188 (15) with default parameters, while the data downloaded from GISAID database was mapped using Minimap v2.12-r847 (16).

The BAM files were sorted using SAMtools (17) and were subsequently used to call the genetic variants in variant call format (VCF) by BCFtools (17). The final call set of the 40 genomes, was annotated and their impact was predicted using SnpEff v 4.3t (18). First, the SnpEff databases were built locally using annotations of the reference genome NC_045512.2 obtained in GFF format from NCBI database. Then, the SnpEff database was used to annotate SNPs and with putative functional effects according to the categories defined in the SnpEff manual

(http://snpeff.sourceforge.net/SnpEff_manual.html).

Phylogenetic analysis and spatio-dynamic analysis

We performed multiple sequence alignment using Muscle v 3.8 (19) for the 28 Moroccan strains with 229 genomes of SARS-CoV-2 circulating in the world from different geographical areas (Africa, Asia, Europe, North and South America and Oceania) (**Table S2**). Maximum-likelihood trees were inferred with IQ-TREE v1.5.5 under the GTR model (20). Generated trees were visualised using FigTree 1.4.3 (http://tree.bio.ed.ac.uk/software/figtree).

Results

Genetic variants in twenty-eight SARS-CoV-2 genomes from Morocco

In order to identify the genetic variants of the SARS-COV-2 moroccan genomes, 28 genomes were studied, including six sequenced in the present study and twenty-two others available in GISAID database (**Table 1**). 94.9 % to 99.93 % of the reads produced for the six genomes were mapped on the reference sequence Wuhan-Hu-1/2019 (**Table S1**). In all Moroccan SARS-CoV-2 genomes, the analysis of genetic variants revealed 61 mutations compared to the reference sequence (**Fig 1**), including 29 non-synonymous mutations, of them 27 (93.10%) having missense effects and 2 (6.90%) producing a lost stop, 27 synonymous mutations and 5 mutations localized in the intergenic regions. The distribution of these mutations along the viral genome revealed that five genes (ORF1ab, S, M, N and ORF3a) harbored mutations with varying frequencies. It is interesting to note that 37.7% (n = 23) of mutations were present in two or more genomes, while the remainder were singleton mutations. Focusing on non-synonymous mutations, 75.86% (n = 22) were located in the ORF1ab gene and distributed in eight non-structural proteins, including 6 (D1036E, L1249H, V2047F, P2110L, A2637V and T2648I) in nsp3-Multi- domain, 6 (C4588F, S4611L, C4772F, T5020I, A5039S and R5314M) in nsp12-RNA-dependent RNA polymerase (RdRp), three (P6000S, T6249P and M6345V) in nsp14-Exonuclease, 2 (T7083I and Ter6967Y) in nsp16 Methyltransferase and one for each of the five other nsp; nsp2 (T265I), nsp5-main proteinase (V3388I), nsp10-CysHis (R4387S), nsp13-Helicase (T5448I) and nsp15-EndoRNAse (Ter6668W). The remaining non-synonymous mutations (24.14%), were distributed in S (V6F, D614G, M1237I), N (R203K and D348H), M (L13F) and ORF3a (Q57H).

Distribution of genetic variants in four North African countries

In order to characterize the genetic variants of the circulating strains in North Africa, a set of 40 genomes, including 28 from Morocco, 7 from Tunisia, 3 from Algeria and 2 from Egypt, were compared to the reference sequence Wuhan-Hu-1/2019. A total of 118 mutations were detected, of them 58 non-synonymous mutations (91.38% have missense effects, 6.90% produce a lost stop and 1.72% produce a stop gained), 48 syn-

onymous mutations and 12 intergene mutations (**Fig 2A**). These mutations have been distributed in seven genes, (ORF1ab, S, E, M, N, ORF3a and ORF8) with variable frequencies. As regard to non-synonymous mutations (**Fig 2B**), we observed that four genes carried at least one non-synonymous mutation. ORF1ab harbored two-thirds of mutations (67.24%; n = 39), distributed in thirteen non-structural proteins; nsp3-Multi-domain: 17.24%, nsp12-RdRp: 12.07%, nsp4-transmembrane domain-2: 8.62%, nsp14-Exonuclease: 8.62%, nsp2: 3.45%, nsp13-Helicase: 3.45%, nsp16-Methyltransferase: 3.45 %, nsp5-main proteinase: 1.72%, nsp6-transmembrane domain: 1.72%, nsp7: 1.72%, nsp8: 1.72%, nsp10-CysHis: 1.72%, nsp15-EndoRNAse: 1.72%. Followed by S, N and ORF3a proteins, with 12.07%, 10.34%, 5.17%, respectively. Whereas E, M and ORF8 proteins, having 1.72% of non-synonymous mutations each.

It is interesting to note that among the 58 non-synonymous mutations, 13 (22.41%) were recurrent in two or more genomes (**Fig 2A**). The most frequent one was the D614G mutation (in S protein) with a prevalence of 92.5% (n = 37) among the 40 genomes included in this study, the second one was Q57H (in ORF3a) with a prevalence of 42.5% (n = 17). These two mutations have been observed within the four north African countries (**Fig 2B**). However, the eleven other mutations were variable between these four countries, for example, T265I (in nsp2) was found in 25% of the genomes, including those of Moroccan, Algerian and Tunisian origins. Likewise, T5020I mutation (in nsp12-RdRp) was found with a prevalence of 17.5% within genomes belonging to Morocco and Tunisia. In addition, K2798R mutation (in nsp4-transmembrane domain-2) was present in 10% of the genomes from Tunisia and Egypt. In addition, six mutations, R203K (in N protein), D1036E, V2047F, A2637V, T2648I (in nsp3-Multi-domain) and C4588F (in nsp12-RdRp) were recurrent in Moroccan genomes. Whereas, the two remaining mutations S202N (in N protein) as well as L84S (in ORF8) were recurrent in genomes of SARS-CoV-2 from Tunisia.

Phylogenetic analysis of the SARS-CoV-2 Moroccan genomes with other genomes from different geographical areass

The phylogenetic analysis wa carried out using a set of 256 genomes from different countries representing the 6 continents (**Table S2**) in order to study the possible source of SARS-CoV-2 strains circulating in North Africa, with a focus on Moroccan strains.

The phylogenetic tree revealed five main clades: two clades (represented by mauve and green colors) particularly contained strains from Asia, while the other clades contained strains belonging to different continents. We observed that approximately 70% of the strains belonged to the clade colored in light orange and showing the mutation D614G (in S protein). With the exception of three strains from Tunisia, all of the North African strains harboring the D614G mutation belonged to this clade, which is also subdivided into several subclades. Among the twenty-eight Moroccan strains, five (Morocco/6893, Morocco/6906, Morocco/15N, Morocco/6902 and Morocco/6888) were close to the strains from South America (Chile or Brazil). Likewise, four strains (Morocco/RMPS-01, Morocco/RMPS-04, Morocco/RMPS-06 and Morocco/9601) seem to share a close sequence similarity with the strains from Israel (Asia), USA (North America) and the isolates from Algeria (Africa), while Morocco/RMPS-02 and Morocco/6891 were close to those from the USA (North America) and Tunisia (Africa). In addition, three strains (Morocco/6899 Morocco/6900 and Morocco/6905) were grouped with strains from Gambia (West Africa), and Italy (Europe) for Morocco/6906.

Remarkably, Moroccan and Tunisian strains were closely related to those from different continents, which could indicate different sources of infection with no single dominant strain circulating yet in Morocco.

Discussion

The appearance and monitoring of genetic variants plays a major role in orienting the therapeutic approach for the development of candidate vaccines in order to limit this SARS-CoV-2 pandemic (21). To date, the genetic diversity of SARS-CoV-2 strains from North Africa is poorly documented. In this study, we performed a genetic analysis of forty SARS-Cov-2 genomes from North Africa, including twenty-eight from Morocco (6 newly sequenced), seven from Tunisia, three from Algeria and two from Egypt, to provide new information on genetic diversity and transmission of SARS-CoV-2.

Genetic diversity could potentially increase the physical shape of the viral population and make it difficult to fight, or reverse, make the virus weaker, which could be correlated with the loss of their virulence and a decrease in the number of critical cases (22). Compared to the reference sequence of Wuhan-Hu-1/2019, strains from North Africa harbored 4 to 15 genetic variants, of which 1 to 11 are involved in the change of amino acids. These results are consistent with the mutation rate previously reported in SASR-CoV-2 from different geographic areas (13, 23-25).

In Morocco, Tunisia, Algeria and Egypt, five non-synonymous mutations were common within at least two countries. Among them, D614G (in S protein) and Q57H (in OR3a) were observed in strains from the four countries. The D614G mutation is proximal to the S1 cleavage domain of advanced glycoprotein (26) and was of great interest due to their predominance in the six continents (27, 28). Alouane et al. (13) showed that this mutation appeared for the first time on January 24, 2020 in the Asian region (China), after a week it was also observed in Europe (Germany). The Q57H mutation was taken away end of February in Africa (Senegal), Europe (France and Belgium) and North America (USA and Canada). Likewise, our previous study (13) showed that D614G had no impact on the two-dimensional or three-dimensional structure of advanced glycoprotein. Of the other three non-synonymous mutations that are variable between the strains from the four countries, T265I mutation (in nsp2) was identified Moroccan, Tunisian and Algerian strains, while K2798R (in nsp4-transmembrane domain-2) and T5020I mutations (in nsp12-RdRp) were observed within the strains from Tunisia-Egypt and Tunisia-Morocco, respectively. Among these

five mutations, four (D614G, Q57H, T265I and T5020I) have been considered as hotspot mutations in a large population (13, 23).

Focusing on SARS-CoV-2 Moroccan strains, four mutations were common within all strains, which is in agrrement with recent studies showing a low frequency of recurrent mutations in thousands of SARS-CoV-2 genomes (13, 23).

The ORF1ab polyprotein is known to be cleaved into 16 non-structural proteins (nsp1-nsp16) (6). We observed two domains rich in non-synonymous mutations, the first, nsp3-Multi domain due to its large size compared to other non-structural proteins and previously described as playing a different role in SARS-CoV-2 infection (29). Likewise, nsp12-RdRp displays the same number of non-synonymous mutations although it has a smaller size and considered as a key element of the replication/ transcription mechanism (30).

Phylogenetic analysis using a set of 256 strains representing the six continents revealed five main clades. The most important clade contained approximately three-quarters of all the strains. All SARS-CoV-2 strains from North Africa harboring the D614G mutation belonged to this clade, with the exception of three Tunisian strains. It is interesting to note that the Moroccan and Tunisian strains were closely related to those from Asia, Europe, South and North America, which could indicate different sources of SARS-CoV-2 infection in these two countries. Whereas, the sources of strains from Algeria and Egypt could probably be from two main countries namely Israel (Asia) and USA (North America).

These results provide valuable information on the genetic diversity of North African strains and their possible origins with a focus on the new Moroccan strains of SARS-CoV-2. This finding could lead to further comprehensive investigations combining genomic data, and clinical epidemiology of SARS-CoV -2 patients in North Africa.

Conflict of interest

The authors declare that they have no competing interests.

Acknowledgments

We sincerely thank the authors and laboratories around the world who have sequenced and shared the full genome data for SARS-CoV-2 in the GISAID database. All data authors can be contacted directly via www.gisaid.org.

This work was carried out under National Funding from the Moroccan Ministry of Higher Education and Scientific Research (Covid-19 Program) to AI. This work was also supported by a grant to AI from Institute of Cancer Research and the PPR-1 program to AI.

References

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients 218 with Pneumonia in China, 2019. N Engl J Med. 382(8):727-33.
2. Kobia, F., & Gitaka, J. (2020). COVID-19: Are Africa's diagnostic challenges blunting response effectiveness?. AAS Open Research, 3:4.
3. Gilbert, M., Pullano, G., Pinotti, F., Valdano, E., Poletto, C., Boëlle, P. Y., ... & Gutierrez, B. (2020). Preparedness and vulnerability of African countries against importations of COVID-19: a modelling study. The Lancet, 395(10227), 871-877.
4. Cucinotta, D., & Vanelli, M. (2020). WHO declares COVID-19 a pandemic. Acta bio-medica: Atenei Parmensis, 91(1), 157-160.
5. WHO, "Coronavirus disease 2019 (COVID-19) Situtation report," https://covid19.who.int/ (accessed June 26, 2020 at 3:23pm).
6. Wu, A., Peng, Y., Huang, B., Ding, X., Wang, X., Niu, P., ... & Sheng, J. (2020). Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. Cell host & microbe.
7. Malik, Y. A. (2020). Properties of Coronavirus and SARS-CoV-2. The Malaysian Journal of Pathology, 42(1), 3-11.

8. Wang, N., Shang, J., Jiang, S., & Du, L. (2020). Subunit vaccines against emerging pathogenic human coronaviruses. Frontiers in microbiology, 11, 298.

9. Tai, W., He, L., Zhang, X., Pu, J., Voronin, D., Jiang, S., ... & Du, L. (2020). Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. Cellular & molecular immunology, 17(6), 613-620.

10. Chen, W. H., Strych, U., Hotez, P. J., & Bottazzi, M. E. (2020). The SARS-CoV-2 vaccine pipeline: an overview. Current tropical medicine reports, 1-4.

11. Domingo, E. (2000). Viruses at the edge of adaptation. Virology, 270(2), 251-253.

12. Domingo, E. J. J. H., & Holland, J. J. (1997). RNA virus mutations and fitness for survival. Annual review of microbiology, 51(1), 151-178.

13. Alouane, T., Laamarti, M., Essabbar, A., Hakmi, M., Bouricha, E. M., Chemao-Elfihri, M. W., ... & Ghrifi, F. (2020). Genomic diversity and hotspot mutations in 30,983 SARS-CoV-2 genomes: moving toward a universal vaccine for the" confined virus"?. bioRxiv. DOI: 10.1101/2020.06.20.163188

14. Shu, Y., & McCauley, J. (2017). GISAID: Global initiative on sharing all influenza data–from vision to reality. Eurosurveillance, 22(13), 30494.

15. Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. bioinformatics, 25(14), 1754-1760.

16. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics, 34(18), 3094-3100.

17. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... & Durbin, R. (2009). The sequence alignment/map format and SAMtools. Bioinformatics, 25(16), 2078-2079.

18. Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., ... & Ruden, D. M. (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(2), 80-92.

19. Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research, 32(5), 1792-1797.

20. Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular biology and evolution, 32(1), 268-274.

21. Tu, Y. F., Chien, C. S., Yarmishyn, A. A., Lin, Y. Y., Luo, Y. H., Lin, Y. T., ... & Wang, M. L. (2020). A review of SARS-CoV-2 and the ongoing clinical trials. International journal of molecular sciences, 21(7), 2657.

22. Parrish, C. R., Holmes, E. C., Morens, D. M., Park, E. C., Burke, D. S., Calisher, C. H., ... & Daszak, P. (2008). Cross-species virus transmission and the emergence of new epidemic diseases. Microbiology and Molecular Biology Reviews, 72(3), 457-470.

23. Laamarti, M., Alouane, T., Kartti, S., Chemao-Elfihri, M. W., Hakmi, M., Essabbar, A., ... & El Jaoudi, R. (2020). Large scale genomic analysis of 3067 SARS-CoV-2 genomes reveals a clonal geodistribution and a rich genetic variations of hotspots mutations. bioRxiv. DOI: 10.1101/2020.05.03.074567

24. Hassan, S.S.; Pal Choudhury, P.; Roy, B. SARS-CoV2 Envelope Protein: Non-Synonymous Mutations and Its Consequences. Preprints 2020, 2020060072 (doi: 10.20944/preprints202006.0072.v1).

25. Yu, W. B., Tang, G. D., Zhang, L., & Corlett, R. T. (2020). Decoding the evolution and transmissions of the novel pneumonia coronavirus (SARS-CoV-2/HCoV-19) using whole genomic data. Zoological Research, 41(3), 247.

26. Veljkovic, V., Perovic, V., & Paessler, S. (2020). Prediction of the effectiveness of COVID-19 vaccine candidates. F1000Research, 9(365), 365.

27. Eaaswarkhanth, M., Al Madhoun, A., & Al-Mulla, F. (2020). Could the D614 G substitution in the SARS-CoV-2 spike (S) protein be associated with higher COVID-19 mortality?. International Journal of Infectious Diseases.

28. Kiyotani, K., Toyoshima, Y., Nemoto, K., & Nakamura, Y. (2020). Bioinformatic prediction of potential T cell epitopes for SARS-Cov-2. Journal of Human Genetics, 65(7), 569-575.

366 29. Angeletti, S., Benvenuto, D., Bianchi, M., Giovanetti, M., Pascarella, S., & Cic-
367 cozzi, M. (2020). COVID 2019: the role of the nsp2 and nsp3 in its pathogenesis.
368 Journal of medical virology, 92(6), 584-588.

369 30. Pachetti M, Marini B, Benedetti F, Giudici F, Mauro E, Storici P, et al. Emerging
370 SARS-CoV-2 Mutation Hot Spots Include a Novel RNA-dependent-RNA Poly-
371 merase Variant. J Transl Med. 2020;18: 179.

372

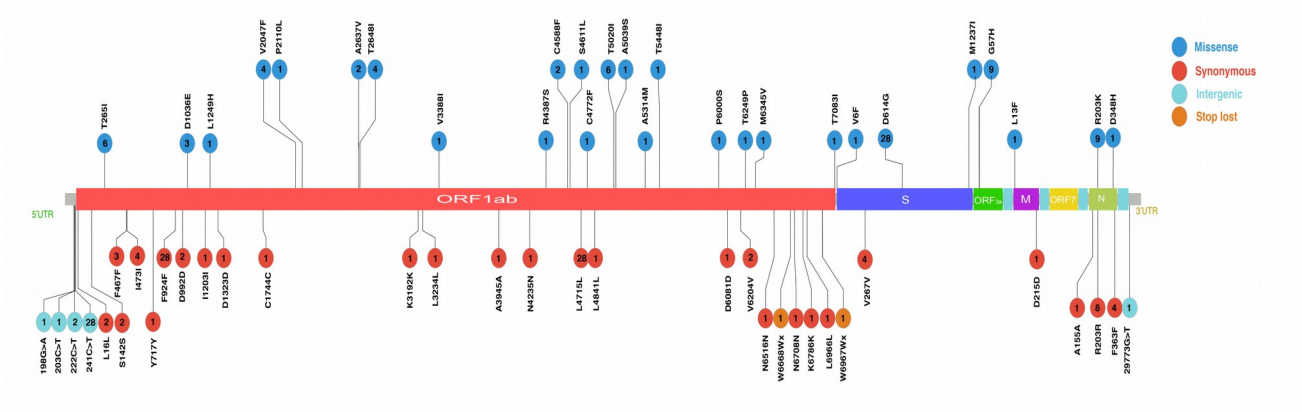
Table and Figures

Table I: Full genomes of SARS-CoV-2 from North Africa used in this study.

Virus name	Accession ID	Collection date	Originatinglab	Platform
Morocco/RMPS-01	EPI_ISL_469017	2020-04-13	MedBiotech (Morocco)	MinIONNanopore-Oxford
Morocco/RMPS-02	EPI_ISL_469049	2020-03-30		
Morocco/RMPS-03	EPI_ISL_469051	2020-04-03		
Morocco/RMPS-04	EPI_ISL_469052	2020-03-30		
Morocco/RMPS-05	EPI_ISL_469053	2020-03-30		
Morocco/RMPS-06	EPI_ISL_469054	2020-04-01		
Morocco/6887	EPI_ISL_459965	2020-03-03	Pasteur Institute (Morocco)	Illumina-NextSEQ500
Morocco/6888	EPI_ISL_459966	2020-03-15		
Morocco/6889	EPI_ISL_459967	2020-03-15		
Morocco/6890	EPI_ISL_459968	2020-03-17		
Morocco/6891	EPI_ISL_459969	2020-03-20		
Morocco/6892	EPI_ISL_459970	2020-03-17		
Morocco/6893	EPI_ISL_459971	2020-03-18		
Morocco/6894	EPI_ISL_459972	2020-03-20		
Morocco/6895	EPI_ISL_459973	2020-03-20		
Morocco/6896	EPI_ISL_459974	2020-03-20		
Morocco/6897	EPI_ISL_459975	2020-03-21		
Morocco/6898	EPI_ISL_459976	2020-03-16		
Morocco/6899	EPI_ISL_459977	2020-04-21		
Morocco/6900	EPI_ISL_459978	2020-04-20		
Morocco/6901	EPI_ISL_459979	2020-04-19		
Morocco/6902	EPI_ISL_459980	2020-04-19		
Morocco/6903	EPI_ISL_459981	2020-04-19		
Morocco/6904	EPI_ISL_459982	2020-04-18		
Morocco/6905	EPI_ISL_459983	2020-04-21		
Morocco/6906	EPI_ISL_459984	2020-04-06		
Morocco/OUA677-19	EPI_ISL_451400	2020-04-23	LRAM (Morocco)	Illumina-MiSeq
Morocco/15N	EPI_ISL_458150	2020-05-15	ANOUAL laboratory (Morocco)	AppliedBiosystems PGM
Tunisia/MHT_2	EPI_ISL_458286	2020-03-24	IMB, (Germany/ Submitting lab)	NanoporeGridION
Tunisia/COV0010-12	EPI_ISL_463001	2020-03-18	Institut Pasteur (Tunisia)	Illumina
Tunisia/COV0880	EPI_ISL_463002	2020-03-28		
Tunisia/COV1339	EPI_ISL_463003	2020-03-30		

Tunisia/COV1663	EPI_ISL_463004	2020-04-01		
Tunisia/COV0425	EPI_ISL_463005	2020-03-27		
Tunisia/COV1482	EPI_ISL_463006	2020-03-31		
Algeria/G0638_2264	EPI_ISL_418241	2020-03-02	Pasteur Institute (Algeria)	Illumina-NextSEQ500
Algeria/G0640_2265	EPI_ISL_418242	2020-03-08		
Algeria/G0860_2262	EPI_ISL_420037	2020-03-02		
Egypt/NRC-03	EPI_ISL_430819	2020-03-18	CSEIV-NRC (Egypt)	Illumina-MiSeq
Egypt/NRC-01	EPI_ISL_430820	2020-03-18		

375



376 **Figure 1: SARS-CoV-2 genomes landscape illustration representing mutations**
377 **identified in 28 Moroccan genomes**. Colored circles represent gene distribution across
378 the genomes. Lollipop stick represents individual mutation. Colored circles represent
379 the type of mutation and the blue box represents the number of genomes harboring the
380 mutations.

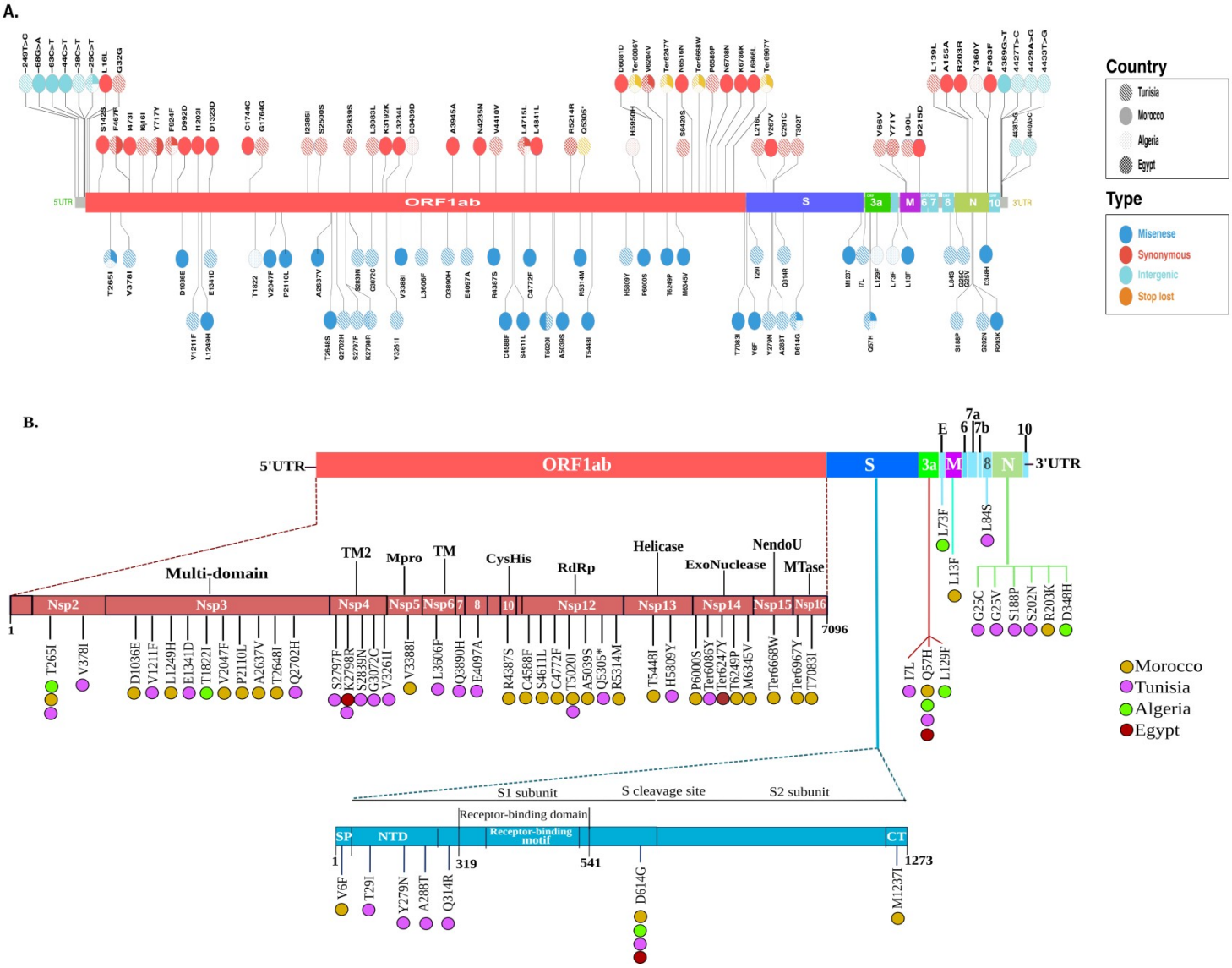
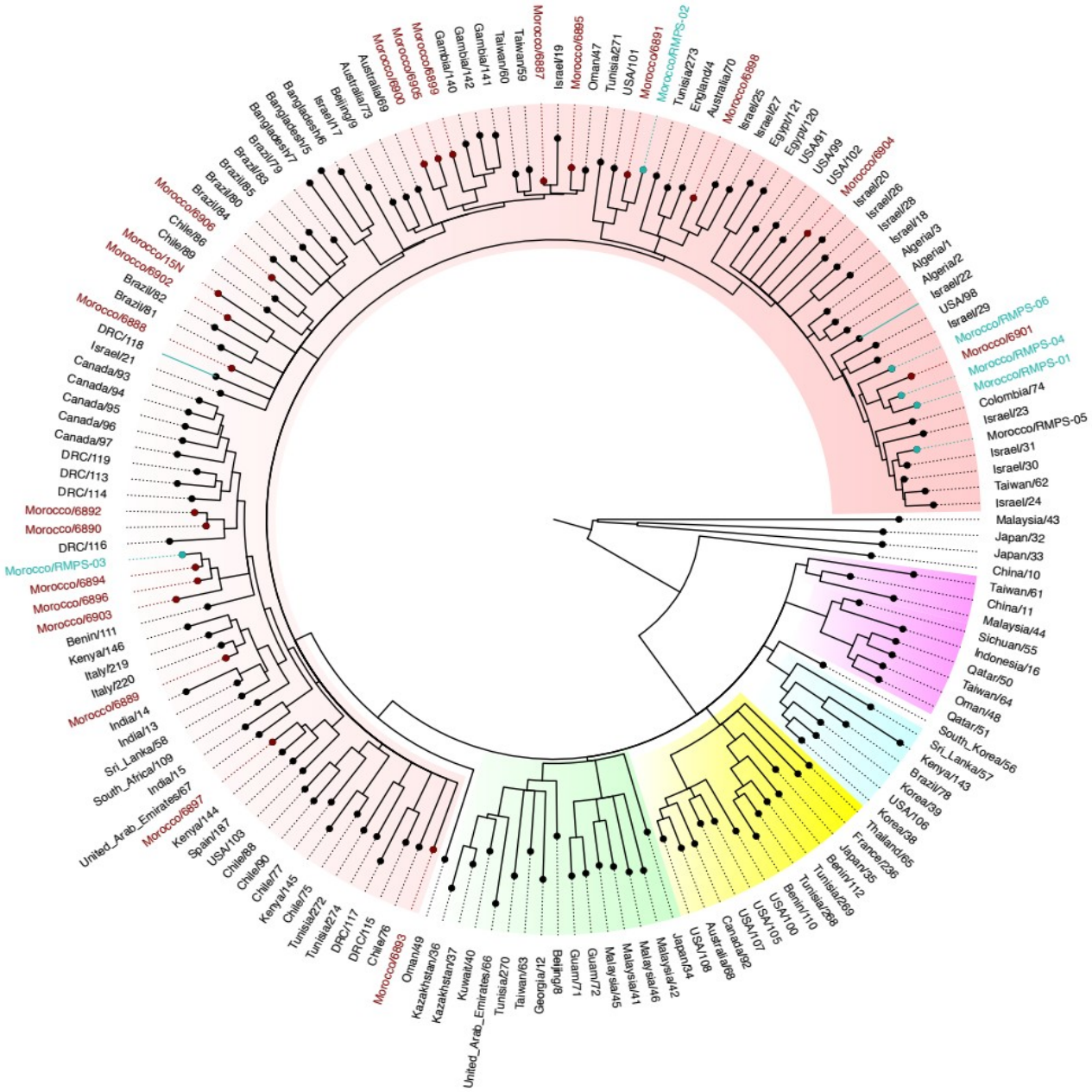


Figure 2: SARS-CoV-2 genomes landscape illustration representing mutations identified in North African genomes. (A) Distribution of all types of mutations (Nons-synonymous, synonymous and intergenic) identified in the genomes of North Africa (Algeria, Egypt, Tunisia and Morocco). The lollipop stick represents individual mutations. The colored circles represent the type of mutation across the genomes. The fill pattern corresponds to each country, as indicated in the key box. **(B)** Distribution of non-synonymous mutations along the viral genome. Vertical lines: 58 non-synonymous mutations. Colored circles: the different countries assessed.

389



390 **Figure 3: Phylogenetic tree based on 256 complete SARS-COV2 genomes from dif-**
391 **ferent geographic areas.** The scale bar shows the length of the branch which repre-
392 sents the change of nucleotides in the genome. The six Moroccan isolates newly se-
393 quenced in this study, are represented by turquoise and the other genomes from the
394 same country (retrieved from the GISAID database), represented by red.