Preprint DOI: https://doi.org/10.1099/acmi.0.000853.v2
Posted on July 17, 2024

Genetic diversity and genomic epidemiology of SARS-CoV-2 during the first three years of the pandemic in Morocco: Comprehensive sequence analysis, including the unique lineage B.1.528 in Morocco

Soulandi Djorwé[1, 2*], Abderrahim Malki[1], Néhémie Nzoyikorera[3, 4, 5], Joseph Nyandwi[6, 7], Samuel Privat Zebsoubo[8], Kawthar Bellamine [2], and Amale Bousfiha [1]

[1]Laboratory of Physiopathology and Molecular Genetics, Faculty of Sciences Ben M'Sik, Hassan II University of Casablanca (Morocco). Avenue Cdt Driss El Harti, PB 7955 Sidi Othman Casablanca (Morocco)

[2]Bourgogne Laboratory of Medical and Scientific Analysis, 136, residence belhcen, Bd Bourgogne, Casablanca (Morocco)

[3]National Reference Laboratory, National Institute of Public Health, Burundi

[4]Higher Institute of Biosciences and Biotechnology, Mohammed VI University of Health Sciences (UM6SS), Casablanca, Morocco

[5]Laboratory of Microbial Biotechnology and Infectiology Research, Mohammed VI Center for Research & Innovation, Mohammed VI University of Health Sciences (UM6SS), Casablanca, Morocco

[6]Département de Médecine, Faculté de Médecine, Université du Burundi, Burundi

[7]Ministère de la Santé Publique et de la Lutte contre le Sida, Institut National de Santé Publique de Bujumbura, Burundi

[8]School of Advanced Studies in Biotechnology and Private Health (EHEB), 183 Bd de la Resistance, Casablanca 20250 (Morocco)

Corresponding author: Email: **soulandidjorwe@gmail.com**

ABSTRACT

During the three years following the emergence of the COVID-19 pandemic, the African continent, like other regions of the world, was substantially impacted by COVID-19. In Morocco, the COVID-19 pandemic has been marked by the emergence and spread of several SARS-CoV-2 variants, leading to a substantial increase in the incidence of infections and deaths. Nevertheless, there has been a persistent lack of comprehensive understanding regarding the genetic diversity, evolution, and epidemiology of several viral lineages. This study sought to deepen the understanding of the genomic epidemiology of SARS-CoV-2 through a retrospective analysis. The main objective of this study was to analyze the genetic diversity of SARS-CoV-2 and identify distinct lineages, as well as assess their evolution during the pandemic in Morocco, using genomic epidemiology approaches. Furthermore, several key mutations in the functional proteins across different viral lineages were highlighted along with an analysis of the genetic relationships amongst these strains to better understand their evolutionary pathways. A total of 2274 genomic sequences of SARS-CoV-2, acquired from Morocco during the period of 2020 to 2023, were extracted from the GISAID EpiCoV database and subjected to analysis. Lineages and clades were classified according to the nomenclature of GISAID, Nextstrain, and Pangolin. The study was conducted and reported in accordance with STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines. An exhaustive analysis of 2274 genomic sequences led to the identification of 157 PANGO lineages, including notable lineages such as B.1, B.1.1, B.1.528 and B.1.177, as well as variants such as B.1.1.7, B.1.621, B.1.525, B.1.351, B.1.617.1, B.1.617.2 and its notable sub-lineages AY.33, AY.72, AY.112, AY.121 that evolved over time before being supplanted by Omicron in December 2021. Among the 2274 sequences analyzed, Omicron and its sub-variants had a prevalence of 59.5%. The most predominant clades were 21K, 21L, and 22B, which are respectively related phylogenetically to BA.1, BA.2, and BA.5. In June 2022, Morocco rapidly observed a recrudescence of cases of infection, with the emergence and concurrent coexistence of subvariants from clade 22B such as BA.5.2.20, BA.5, BA.5.1, BA.5.2.1, and BF.5, supplanting the subvariants BA.1 (Clade display 21K) and BA.2 (Clade display 21L), which became marginal. However, XBB (clade 22F) and its progeny such XBB.1.5(23A), XBB.1.16(23B), CH.1.1(23C), XBB.1.9(23D), XBB.2.3(23E), EG.5.1(23F), and XBB.1.5.70 (23G) have evolved sporadically. Furthermore, several notable mutations, such as H69del/V70del, G142D, K417N, T478K, E484K, E484A, L452R, F486P, N501Y, Q613H, D614G, and P681H/R, known for their implications in transmissibility by increase infectivity and their ability to escape antibodies, were identified. This study has identified several distinct lineages, various mutations involved in the genetic diversity of Moroccan isolates, and analyzed their evolutionary trends. Furthermore, the lineages identified in Morocco were genetically closely linked to the original lineages, highlighting the importance of maintaining continuous surveillance of COVID-19 and the expansion of vaccine coverage would help protect patients against more severe clinical disease. These findings also provide a robust basis for better understanding the distinct mutations and their role in transmission, immune evasion, and reinfection.

Keywords: SARS-CoV-2, Genetic Diversity, Variant, Lineage, Omicron, Genomic sequence

Data summary

The sequence data used in this investigation were extracted from the GISAID EpiCov database. All the sequences analyzed in this study are available under the accession number (GISAID identifier): EPI_SET_240225ry (supplementary material 1). Supplementary materials are provided alongside the article in PDF and Excel formats. Access to the supplementary data files is available on Figshare via the following DOI link: https://doi.org/10.6084/m9.figshare.25748970.v1.

Abbreviations

ACE2: Angiotensin converting enzyme 2; COVID-19: Coronavirus disease 2019; SARS-CoV-2: severe acute respiratory syndrome coronavirus 2; CT: Cytoplasmic domain; FP: Fusion peptide; HR1/HR2: Heptad repeats 1 and 2; SNP: Single Nucleotide Polymorphism; NSP: Non-Structural Protein; NTD: N-terminal domain; RBD: Receptor-Binding Domain; RBM: Receptor-Binding Motif; TM: Transmembrane domain; VOC: Variants of Concern; VOI: Variants of Interest; VUM: Variants Under Monitoring.

INTRODUCTION

Appeared in 2019 in Wuhan, in the Hubei province, China, COVID-19 is one of the most devastating pandemics of this century, involving millions of cases and deaths worldwide. In February 2020, a case of COVID-19 was reported for the first time in Africa in Egypt, and the first case in Morocco was reported on 02 March 2020 [1]. In Morocco, a national SARS-CoV-2 genomic surveillance consortium has been set up to sequence Moroccan isolates in order to inform the health authorities about circulating variants. As of 6 August 2023, more than 769 million cases of infection and over 6.9 million deaths have been reported worldwide, according to the World Health Organization (WHO) [2]. In Morocco, during the last quarter of 2023, over 1,278,055 confirmed cases and 16,298 deaths were recorded [3].

The SARS-CoV-2, an RNA virus of the coronavirus family responsible for the global COVID-19 pandemic, has a genome length of approximately 29.9 kilobases (kb) [4]. The genome consists of 15 open reading frames (ORFs), with two important polyproteins involved in replication: pp1a and pp1ab (ORF1ab). These polyproteins are cleaved to produce 16 non-structural proteins. Additionally, the genome codes for eleven accessory proteins and four structural proteins, including the spike glycoprotein, envelope, membrane, and nucleocapsid proteins [4–6]. Generally, within the SARS-CoV-2 genome, mutations can progressively accumulate over time, in particular key mutations in structural proteins, of which the spike glycoprotein is the most closely monitored. Moreover, some mutations in the SARS-CoV-2 genome can significantly impact its fundamental properties. These mutations can increase contagiousness, enhance the severity and virulence of the infection, reduce the effectiveness of drugs and vaccines, allow the virus to evade the immune system, cause reinfections, and escape molecular detection during laboratory diagnostics. Therefore, variants that exhibited the aforementioned characteristics are classified either as Variants of Concern (VOC), either as Variants Under Monitoring (VUM) or Variants of Interest (VOI). However, these variants required close monitoring [7,8].

This study aimed to retrospectively analyze data from sequences derived from Moroccan isolates during the COVID-19 pandemic, in order to better understand the genomic epidemiology of SARS-CoV-2 in Morocco. The main aim of this study was to analyze the genetic diversity of SARS-CoV-2 and identify distinct lineages, as well as to assess their evolution during the pandemic in Morocco, using genomic epidemiology approaches. Furthermore, several key mutations in the functional proteins across different viral lineages were highlighted along with an analysis of the genetic relationships amongst these strains to better understand their evolutionary pathways.

The study conducted will provide access to in-depth data analysis with detailed large-scale analysis. Furthermore, this study will also shed light on the genetic trajectory of SARS-CoV-2, specifically the appearance and distribution of variants during the three years after the pandemic's onset in Morocco. This study will extend our understanding of the elements affecting the genetic variation of SARS-CoV-2 in Morocco, such as mutation rates and transmission dynamics. Finally, it will provide a strong basis for understanding the virus's evolution, facilitating the development of more precise preventive measures in future research. However, expanding the analysis by incorporating global GISAID sequences would contribute to a better development of preventative measures.

MATERIALS AND METHODS

Study design

This study was conducted and reported in accordance with STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines [9].

Sequence data acquisition

The complete genomic sequences of SARS-CoV-2 isolates, collected in Morocco from 2020 to 2023, were extracted in FASTA format from the GISAID EpiCoV database (https://gisaid.org/, accessed on 23 January 2024) [10]. The genomic sequences of the Moroccan isolates were compared with the Wuhan-Hu-1 reference genome, identified by accession number NC_045512.2 in the GenBank database.

Supplementary Material 1 contains the digital object identifier (DOI) and EPI_SET identifier of the 2,274 SARS-CoV-2 genomic sequences used in this study. The collection dates range from 02 February 2020 to 03 November 2023.

Sequence alignment and phylogenetic analysis of Moroccan genomes

In this study, we used standard dynamic classification systems to assign genetic lineages and viral clades. The classification of genomic sequences into lineages was achieved by Pangolin COVID-19 lineage assigner version 4.3, which is a Phylogenetic Assignment tool for Named Global Outbreak LINeages developed by the Centre for Genomic Pathogen Surveillance (https://cov-lineages.org/resources/pangolin.html, accessed on 23 January 2024) and/or Nextstrain web tool version 3.3.1 (https://clades.nextstrain.org/, accessed on 23 January 2024) [11–13]. Rigorous quality checks and the assignment of viral clades were performed using the Nextclade Web and GISAID. The phylogenetic tree was generated using the UCSC UShER Web interface (https://genome.ucsc.edu/util.html, accessed on 25 January 2024), Microreact (https://microreact.org/upload, accessed on 03 February 2024), and Nextclade. Viral clades were defined based on shared mutation profiles among the analyzed genomic sequences [13,14].

Analysis of mutation profiles and assignment of lineages and clades

The GISAID database, the Nextclade and Coronapp web tool (http://giorgilab.unibo.it/coronannotator/, accessed on 20 February 2024) [15] were used to detect and annotate all mutations, thus establishing the single nucleotide polymorphism (SNP) profile of the 2274 genomic sequences. This was achieved by identifying substitutions (amino acid), deletions, or insertions (Indels) in structural protein regions, as well as in some regions of non-structural protein (ORF1ab) [NSP1 to NSP16]. Furthermore, we used widely applied systems of classification and dynamic nomenclature such as Nextclade Web Tool, Pangolin COVID-19 Lineage Assigner, GISAID, and WHO to assign lineages and genomic clades [10,12,13].

RESULTS

Genomic diversity and demographic distribution of SARS-CoV-2 sequences

A set of 2274 genomic sequences of SARS-CoV-2 collected in Morocco over the three years following the pandemic has been analyzed, revealing several variants and lineages. The table 1 shows the temporal distribution of variants and lineages among the 2274 sequences analyzed. Of the 2274 sequences, 3.9% (89/2274) of isolates were sequenced in 2020, 22.4% (511/2274) in 2021, 53.5% (1217/2274) in 2022, and 20.1% (457/2274) in 2023. Among the 2274 sequences analyzed, 20.2% (460/2274) were assigned to lineages other than the Alpha, Beta, Delta, Eta, Kappa, Mu, and Omicron variants. Of the 460 sequences analyzed, 19.3% (89/460) were identified in 2020, 40% (184/460) in 2021, 4.1% (19/460) in 2022, and 36.5% (168/460) in 2023 (Table 1). The Alpha variant had a prevalence of 7.7% (176/2274) among all sequences analyzed, of which 81.2% (143/176) of sequences were detected in 2021, 5.1% (9/176) in 2022, and 13.6% (24/176) in 2023. The Delta variant and its sub-variants accounted for 11.3% (257/2274) of the analyzed sequences, among which 62.6% (161/257) were identified in 2021, 28.4% (73/257) in 2022, and 8.9% (23/257) in 2023. The Omicron variant and its sub-variants were identified in 59.5% (1353/2274) of all sequences analyzed, of which 1.1% (15/1353) were identified in 2021, 82.3% (1114/1353) in 2022, and 16.5% (224/1353) in 2023. Furthermore, the other variants such as Beta, Eta, Kappa, and Mu were less predominant (Table 1). These findings illustrate the dynamic evolution of SARS-CoV-2 variants and lineages during the pandemic, highlighting periods of predominance of variants and lineages identified over time.

Of the 2274 sequences analyzed, 54.7% (1244/2274) were recorded in the Casablanca region, followed by 15.1% (343/2274) in Rabat, 1.3% (30/2274) in Ouarzazate, 1.2% (27/2274) in Fès, 1% (21/2274) in Mohammedia, 0.7% (17/2274) in Marrakech, while the remaining 26% (592/2274) were from other cities in Morocco.

Table 1: Temporal distribution of SARS-CoV-2 variants and lineages among the 2274 sequences analyzed: Annual and cumulative prevalence of variants and lineages identified from 2020 to 2023.

Variants Year	Alpha	Beta	Delta	Eta	Kappa	Mu	Omicron	Other lineages	Total
2020	0	0	0	0	0	0	0	19.3% (89/460)	3.9% (89/2274)
2021	81.2% (143/176)	16.6% (1/6)	62.6% (161/257)	57.1% (4/7)	100% (1/1)	14.3% (2/14)	1.1% (15/1353)	40% (184/460)	22.4% (511/2274)
2022	5.1% (9/176)	16.6% (1/6)	28.4% (73/257)	14.2% (1/7)	0	0	82.3% (1114/1353)	4.1% (19/460)	53.5% (1217/2274)
2023	13.6% (24/176)	66.6% (4/6)	8.9% (23/257)	28.5% (2/7)	0	85.7% (12/14)	16.5% (224/1353)	36.5% (168/460)	20.1% (457/2274)
Total	176	6	257	7	1	14	1353	460	2274

Frequency distribution of the 2274 SARS-CoV-2 sequences into lineages

Table 2 shows the distribution of the proportions of variants and lineages according to their dominance among the 2,274 genomic sequences analyzed, and their classification in the clades defined by GISAID and Nextstrain. Among these sequences, ten clades were identified according to GISAID nomenclature and thirty based on Nextstrain classification. The GRA (Omicron) clade had a prevalence of 59.5% (1353/2274) of all sequences analyzed, making it the predominant clade in this study (Table 2). Other lineages, such as B, B.1, B.1.1, B.1.528, B.1.258, B.1.177, B.1.1.487, B.1.1.378, and their sub-lineages, represented 20.23% (460/2274) of the sequences, distributed in GISAID clades G-GH-GR-GV-S-V. Furthermore, the GK clade (Delta) was mainly composed of the B.1.617.2 and its sublineages (AY. *) with a frequency of 11.30%, followed by the GRY clade (Alpha), dominated by the B.1.1.7 variant, with a prevalence of 7.74%. On the other hand, other variants such as Beta, Eta, Mu, and Kappa had frequencies below 1% (Table 2).

Table 2: Frequencies of variants and lineages among the 2,274 genomic sequences analyzed, including the assignment of GISAID and Nextstrain clades for each variant or lineage.

WHO nomenclature (lineage)	Clade_GISAID	Clade_Nextstrain	Number of sequences	Frequency (%)
Omicron (B.1.1.529/BA. */BQ.1/BQ.1. */BQ1.1)	GRA	21K-21L-22A-22B-22C-22D-22E	1246/2274	54.79%
Unassigned name (B, B.1, B.1.1, B.1.528, B.1.258, B.1.177, B.1.1.487, B.1.1.378 and other lineages)	G-GH-GR-GV-S-V	19A-19B-20A-20B-20C-20D-20E	460/2274	20.23%
Delta (B.1.617.2/AY. *)	GK	21A-21I-21J	257/2274	11.30%
Alpha (B.1.1.7)	GRY	20I	176/2274	7.74%
Omicron (XBB.1.9.1/XBB.1.9.1. *)	GRA	23D	37/2274	1.63%
Omicron (XBB.1.5/XBB.1.5. *)	GRA	23A-23G	25/2274	1.10%
Omicron (XBB/XBB. * Excluding XBB.1.5, XBB.1.16, XBB.1.9.1, XBB.1.9.2 and XBB.2.3)	GRA	22F	20/2274	0.88%
Mu (B.1.621)	GH	21H	14/2274	0.62%
Omicron (XBB.2.3/XBB.2.3. *)	GRA	23E	8/2274	0.35%
Eta (B.1.525)	G/484K.V3	21D	7/2274	0.31%
Omicron (XBB.1.16/XBB.1.16. *)	GRA	23B	7/2274	0.31%
Beta (B.1.351)	GH/501Y.V2	20H	6/2274	0.26%
Omicron (BA.2.75/BA.2.75. *)	GRA	22D	4/2274	0.18%
Omicron (XBB.1.9.2/XBB.1.9.2. *)	GRA	23D	3/2274	0.13%
Omicron (EG.5/EG.5. *)	GRA	23F	2/2274	0.09%
Kappa (B.1.617.1)	G/452R.V3	21B	1/2274	0.04%
Omicron (GP.2/CH.1.1/CH.1.1. *)	GRA	23C	1/2274	0.04%

* Indicates sublineage (s) or subvariant.

The analysis of genomic sequences in this study has highlighted a convergent evolution of several variants particularly the Omicron variant and its related subvariants (Table 1). In this regard, increasing diversification has been observed within the Omicron and delta variant, and their sublineages. Over time, Omicron and its subvariants has shown increasing diversification gradually supplanted previous lineages and variants such as B.1, B.1.528, B.1.177, as well as the B.1.1.7, Delta and its sub-variants (AY. *), B.1.621, B.1.525, B.1.617.1, and B.1.351 variants (Figure 1a). The most notable peaks of Omicron and its derivatives were observed in June 2022, a period during which several co-circulating lineages entered into competition to establish their dominance (Figure 1a and 1b).

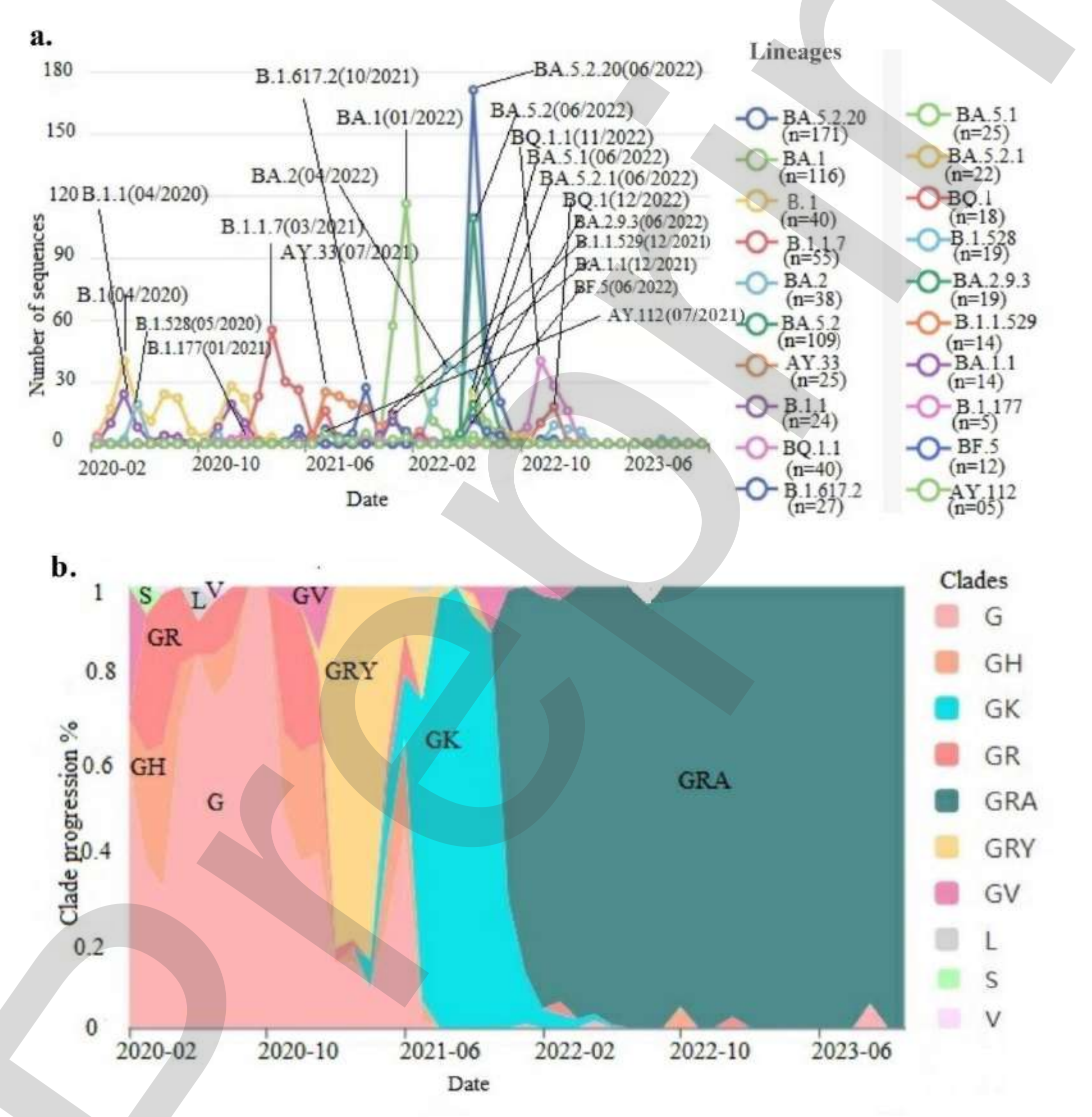

Figure 1: Temporal evolutionary dynamics of SARS-CoV-2 lineages and sub-lineages: **a.** GISAID graph illustrating the gradual emergence of notable peaks of SARS-CoV-2 variants and lineages observed during the pandemic. **b**. Progression of GISAID clades over the years (2020-2023).

In this study, an unequal distribution of various variants and subvariants was observed, with varying proportions over the three years of the pandemic. Some subvariants, such as BA.5.2.20 (11%), BA.1 (10.1%), and B.1 (9.8%), stood out due to their high frequency. Variants B.1.1.7 (7.7%), BA.2 (7.1%), BA.5.2 (6.6%), and AY.33 (5%) also exhibited notable frequencies. Other subvariants such as B.1.1, BQ.1.1, and B.1.617.2, BA.5.1, BA.5.2.1, BQ.1, B.1.528 were also noted with non-negligible frequencies (Table 3).

Table 3: Relative frequency of the highest genomic sequences

Pango lineage	Number of sequences	Frequency (%)
BA.5.2.20 (Omicron)	248/2274	11%
BA.1 (Omicron)	230/2274	10.1%
B.1 (lineage B.1)	223/2274	9.8%
B.1.1.7 (Alpha)	176/2274	7.7%
BA.2 (Omicron)	161/2274	7.1%
BA.5.2 (Omicron)	151/2274	6.6%
AY.33 (Delta)	112/2274	5%
B.1.1 (lineage B.1)	99/2274	4.3%
BQ.1.1 (Omicron)	96/2274	4.2%
B.1.617.2 (Delta)	75/2274	3.3%
BA.5.1 (Omicron)	44/2274	2%
BA.5.2.1 (Omicron)	39/2274	1.7%
BQ.1 (Omicron)	30/2274	1.3%
B.1.528 (lineage B.1.528)	25/2274	1.1%

Genomic variation and mutational patterns

The comparison of 2274 genomic sequences with the reference genome Wuhan-Hu-1 revealed a notable genetic diversity, characterized by a substantial number of mutations, both synonymous and nonsynonymous (Figure 3).

In total, 123,530 mutations were observed relative to the reference genome. Among them, an important frequency of SNP was observed, representing 64.30% (79,433/123,530) of the mutations, while the proportion of silent nucleotide substitutions or variations (silent SNPs) in transcribed regions of the genome was 21.70% (26,814/123,530). Insertions or deletions (indels) were also common, with 5.4% (6,719/123,530) of deletions, 2.1% (2,583/123,530) of frameshift deletions, followed by 0.2% (229/123,530) of insertion events, and 0.6% (799/123,530) of frameshift insertions. Mutations in extragenic regions (UTR) had a prevalence of 10.7% (13,192/123,530) (Supplementary material 2).

Transition substitution was the most common among the 2,274 sequences, with the C>T transition being the most frequent substitution event, occurring with a prevalence of 35.5% (43,862/123,530). This genomic variation was followed by G>A and A>G with frequencies of

10% (12,303/123,530) and 9.8% (12,125/123,530), respectively. The transition substitutions were followed by T>G transversion with a proportion of 5.8% (7,201/123,530), G>T with a proportion of 5.6% (6,997/123,530), and T>C, A>C, A>T, T>A substitutions with prevalences ranging from 1.7% to 4.5% (Figure 2d).

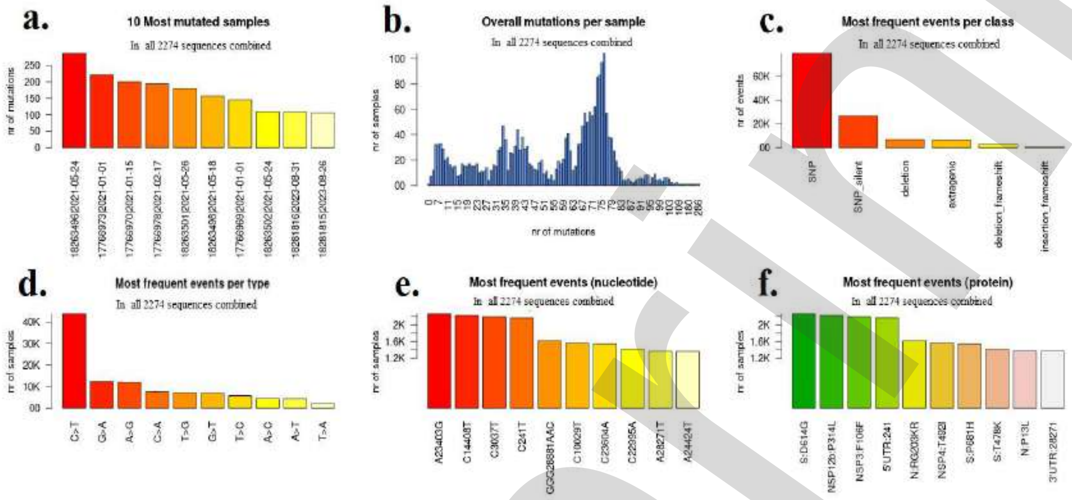

Figure 2: Distribution of mutational events in the 2274 analyzed genomic sequences: (a) Selection of the most mutated sequences among the 2274, (b) Overall number of mutations per sequence, (c) Distribution of genomic variations (SNPs and Indels), (d) Frequency of the different types of substitutions observed, (e) Identification of the most frequently encountered point, (f) Analysis of the most common amino acid mutations.

The genetic variations observed in the genomic sequences of SARS-CoV-2 revealed several notable events, notably the 23403A>G (D614G) genetic variation in the Spike protein, present in 99.5% (2263/2274) of the analyzed genomic sequences. This mutation was followed by the 14408C>T variation (P314L) affecting 98.3% (2237/2274) of sequences and resulting in a change in the NSP12 protein (P323L). Silent mutations were also observed, such as 3037C>T (F106F) [97%; 2,206/2,274] targeting NSP3 (Figure 2e-2f). Furthermore, several other mutations affecting various protein sequences in this study were observed at notable frequencies. Among these mutations, the RG203KR mutation identified in the nucleocapsid protein (N), with a prevalence of 71% (1615/2274) and the P13L mutation with a prevalence of 60.4% (1370/2274). These mutations are followed by NSP4: T492I, with a frequency of 68.7% (1564/2274), Figure 2f. Additionally, genetic mutations in the spike glycoprotein gene, notably P681H (68%; 1545/2274) and T478K (62.4%; 1420/2274) were frequently observed. Mutations in non-coding regions were also identified, notably the 241C>T mutation in 5'UTR (95.6%; 2175/2274) and the 28271A>T mutation in 3'UTR, detected in 60% (1363/2274) of the sequences analyzed (Figure 2).

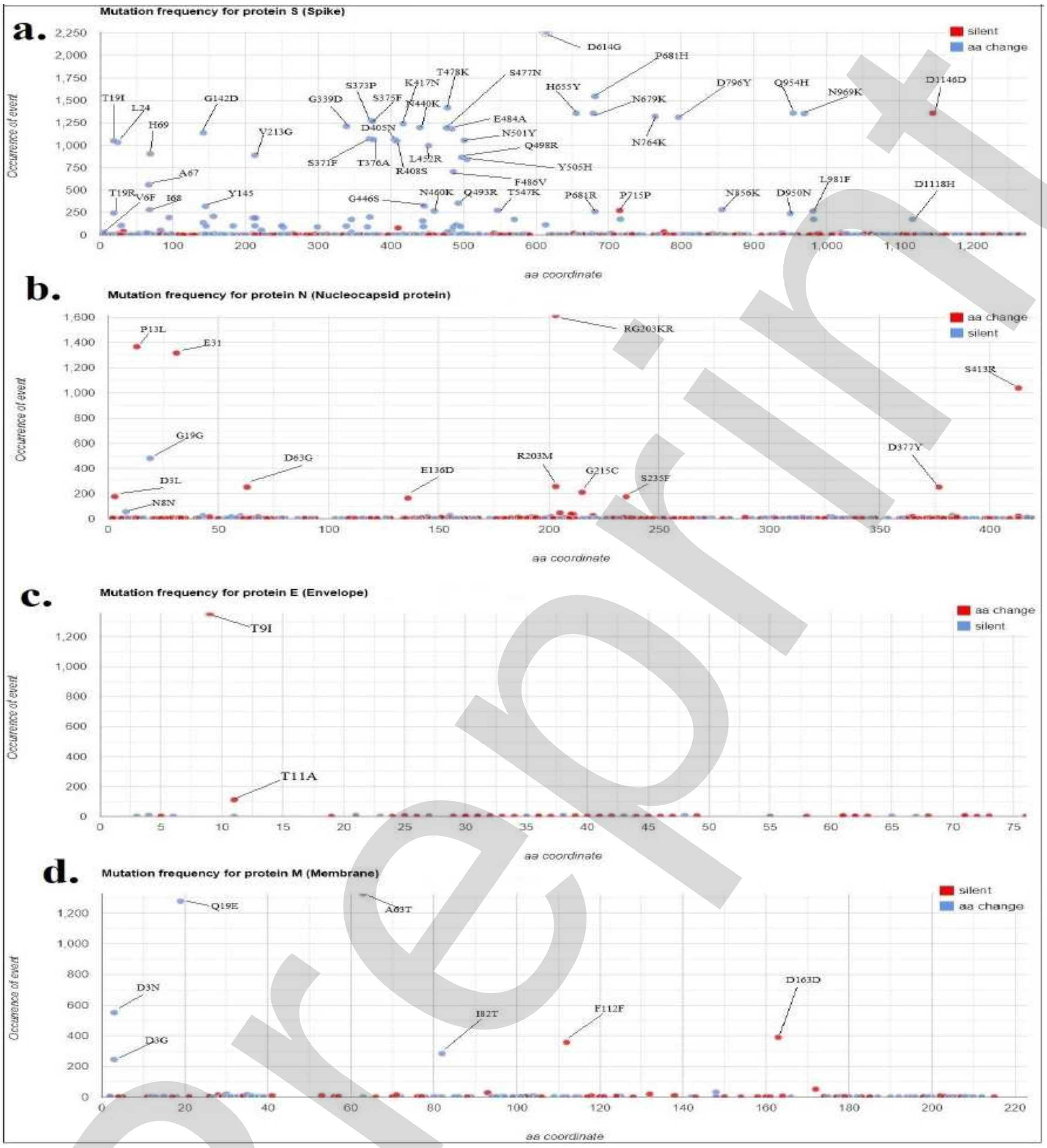

Figure 3: Distribution of synonymous and non-synonymous mutation frequencies of amino acids in structural proteins of the 2274 analyzed genomic sequences. The coordinates on the Y-axis represent the different scales of mutation frequencies observed in the structural proteins of the 2274 analyzed genomic sequences. The X-axis shows the position of each amino acid mutation along the sequences constituting the structural proteins: In the Spike protein (**a**) with ~1,273 amino acids and the membrane protein (**d**) with ~222 amino acids, the blue dots represent nonsynonymous mutations while the red dots indicate synonymous mutations. In the nucleocapsid protein (**b**) with ~419 amino acids and the envelope protein (**c**) with ~75 amino acids, the red dots show nonsynonymous mutations while the blue dots represent synonymous mutations.

Besides the substantial mutations observed in the spike glycoprotein, notable genetic mutations have also been highlighted in the viral envelope and membrane protein (Figure 3). Although less frequent than those observed in the spike protein and nucleocapsid protein, these mutations have been identified at notable rates. For example, in the E protein, mutations T9I (59.6%; 1356/2274) and T11A (4.9%; 111/2274) were the most notable, while in the M protein, mutations such as Q19E (56.3%; 1280/2274), A63T (58.1%; 1323/2274), D3N (24.2%; 550/2274), and D3G (10.7%; 243/2274) were also observed with notable frequencies (Figure 3c and 3d).

Characteristics of the most important mutations in the predominant SARS-CoV-2 lineages isolated in Morocco

Among the various variants identified in this study, Omicron and its subvariants harbored the highest rate of mutations. These mutations are localized in different regions of the genome, among which those requiring particular attention are found in the spike glycoprotein. In the N-terminal domain (NTD) of the S1 region are observed, mutations such as T19I, L24del, P25del, P26del, A27S, Q52H, A67V, H69del/V70del, T95I, G142D, Y144del, H146Q, E180V, Q183E, N211del, L212I, V213E, ins214EPE, G252V, and D253G. In the RBD/RBM domain (S1 region), some sequences exhibited mutations such as G339D/H, R346T, S371F, T376A, D405N, R408S, K417N, N440K, K444T, V445P, G446S, L452R, F456L, N460K, S477/S477N, T478K, E484A, F486P/S/V, F490S, Q493R, Q498R, N501Y, and Y505H were identified. Mutations are also present in the SD1 and SD2 domains (S1 region) outside RBD/RBM domain such as P521S, D614G, A626V, H655Y, N679K, P681H, and S704L while the S2 region harbors mutations such as N764K, D796Y, N969K, and Q954H (Figure 3a, Figure 4a, and Supplementary material 3).

As for the Delta variant and its subvariants (AY. *), they harbored notable mutations such as G142D, T478K, and L452R (Figure 4a). Other notable mutations were also identified, including T19R, T29A, F157del, R158del, E156G, T250I, T299I, Q613H, P681R, and D950N. The Alpha and Beta variants carried mutations such as H69del/V70del, Y144del, N501Y, A570D, D614G, P681H, and D1118H for the Alpha variant, while the Beta variant displayed mutations like D614G, N501Y, L18F, D80A, G142D, L242del/L244del, K417N, and E484K (Figure 3a, Figure 4a, and Supplementary material 2).

However, the B.1.525, B.1.617.1, and B.1.621 variants stood out due to mutations such as A67V, H69del/V70del, Y144del, E484K, and D614G, hosted mutations in the spike protein of the Eta variant, and E154K, E484Q, G142D, L452R, D614G, P681R, Q1071H, and T95I for Kappa. The Mu variant was characterized by mutations such as T95I, A262S, D614G, D950N, E484K, N501Y, P681H, R346K, and Y144del (Figure 3a and Figure 4a).

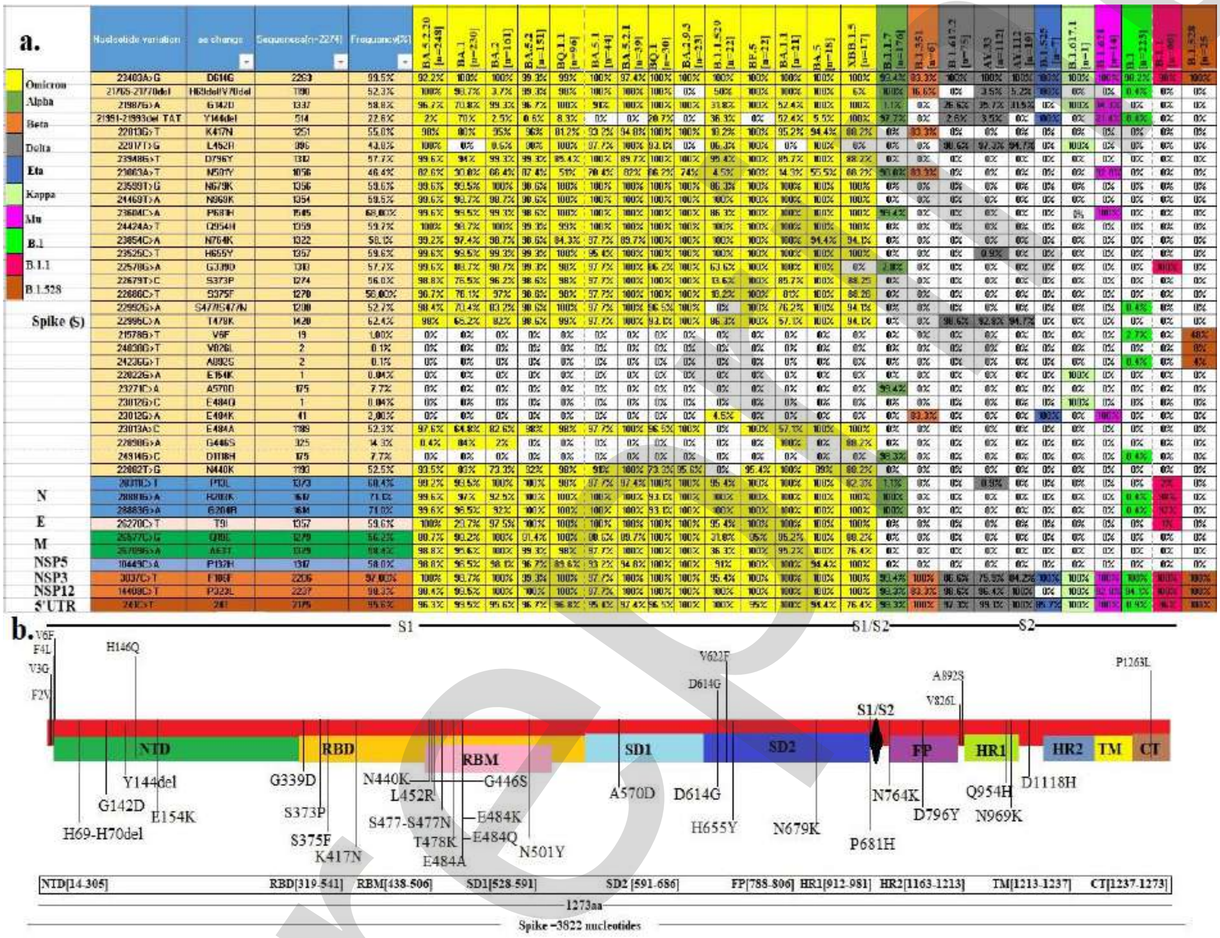

Figure 4: Frequencies of key amino acid mutations observed in the structural proteins (S, M, N, and E) of the variants and the predominant lineages of SARS-CoV-2, as well as the localization of the mutations in the spike glycoprotein, displayed on a schematic genomic map of SARS-CoV-2. **a**) The amino acid mutations observed in the variants and certain predominant lineages among the 2274 genomic sequences analyzed are visually represented. Each colored box indicates the proportion of specific amino acid substitutions in these variants and lineages, while the non-colored (white) box indicates the absence of the corresponding amino acid mutation in the respective variant or lineage. The prevalence of each mutation is assessed by determining the number of specific mutations hosted within the totality of sequences assigned to each variant or lineage. **b**) Diagram illustrating the location of some mutations distributed in the different regions of the spike glycoprotein. These regions include: N-terminal domain (NTD), receptor-binding domain (RBD), receptor-binding motif (RBM), fusion peptide (FP), HR1 and HR2 (heptad repeats 1 and 2), transmembrane domain (TM), and CT (cytoplasmic tail). Mutations identified in the upper part of figure b correspond to mutations in the unique B.1.528 lineage.

Phylogenetic analysis of the 2274 genomic sequences

To analyze the phylogenetic characteristics of the 2274 SARS-CoV-2 sequences, a phylogenetic tree was generated using the maximum likelihood method, aligned with the reference genomic sequence. This analytical method has been used to classify genetic sequences according to their genetic relationship with others, in order to understand the evolutionary genetic lineage of SARS-CoV-2. In addition, phylogenetic analysis revealed the presence of 157 PANGO lineages distributed within thirty distinct clades (Figure 5a and 5b, Table 4). Among the 157 lineages, one was not assigned to a specific clade and was classified as recombinant (Figure 5a and Table 4).

Table 4: Distribution of the different clades and phylogenetically related lineages.

Clade_display	**Pango lineage**
19A	B, B.39
19B	A.5
20A	B.1, B.1.22, B.1.36, B.1.78, B.1.153, B.1.160, B.1.221, B.1.258, B.1.416, B.1.528, B.1.619
20B	B.1.1, B.1.1.112, B.1.1.118, B.1.1.318, B.1.1.447
20C	B.1.356, B.1.575, B.1.597
20D	C.17
20E	B.1.177, B.1.177.51, B.1.177.73, B.1.177.77
20H (Beta)	B.1.351
20I (Alpha)	B.1.1.7
21A (Delta)	B.1.617.2 (13 sequences), AY.51
21B (Kappa)	B.1.617.1
21D (Eta)	B.1.525
21H (Mu)	B.1.621
21I (Delta)	B.1.617.2 (2 sequences), AY.37, AY.72, AY.73, AY.133
21J (Delta)	B.1.617.2 (60 sequences), AY.4, AY.5, AY.33, AY.34, AY.34.1, AY.34.1.1, AY.43, AY.46, AY.98, AY.112, AY.121, AY.122, AY.126
21K (BA.1)	BA.1, BA.1.1, BA.1.1.1, BA.1.8, BA.1.13, BA.1.15, BA.1.17.1, BA.1.18
21L (BA.2)	BA.2, BA.2.3, B.1.1.529, BA.2.3.9, BA.2.5, BA.2.9, BA.2.9.3, BA.2.10, BA.2.10.1, BA.2.11, BA.2.14, BA.2.36, BA.2.40.1, BA.2.49, BA.2.56, BA.2.72, BA.2.79, CM.7.1
22A (BA.4)	BA.4, BA.4.1, BA.4.3
22B (BA.5)	BA.5, BA.5.1, BA.5.1.2, BA.5.1.5, BA.5.1.22, BA.5.1.23, BA.5.2, BA.5.2.1, BA.5.2.3, BA.5.2.20, BA.5.3.1, BA.5.6, BE.1, BE.1.1, BE.1.3, BE.2, BE.3, BF.5, BF.10, BF.15, BF.19, BV.2, CR.1, DB.1, DZ.1
22C (BA.2.12.1)	BA.2.12.1
22D (BA.2.75)	BA.2.75.5, BN.1.2, CJ.1.1
22E (BQ.1)	BQ.1, BQ.1.1, BQ.1.1.4, BQ.1.1.6, BQ.1.1.15, BQ.1.1.18, BQ.1.1.47, BQ.1.1.56, BQ.1.1.58, BQ.1.1.69, BQ.1.1.72, BQ.1.5, BQ.1.13.1, BQ.1.23, BQ.1.31, EF.1
22F (XBB)	XBB, XBB.1, XBB.1.22, XBB.1.38, XBB.2.4, XBB.6
23A (XBB.1.5)	XBB.1.5, XBB.1.5.6, XBB.1.5.23, XBB.1.5.24, XBB.1.5.89, XBB.1.5.75
23B (XBB.1.16)	XBB.1.16, XBB.1.16.10, XBB.1.16.11, XBB.1.16.15, XBB.1.16.21
23C (CH.1.1)	GP.2
23D (XBB.1.9)	XBB.1.9, EG.3, FL.1.5, FL.1.5.1, FL.5, FL.10, FL.26, XBB.1.9.1, XBB.1.9.2
23E (XBB.2.3)	XBB.2.3.11
23F (EG.5.1)	EG.5.1, EG.5.1.3
23G (XBB.1.5.70)	GK.2
Unassigned clade	Recombinant XW

The phylogenetic analysis revealed a gradual diversification of the virus over time, characterized by the emergence of one of the earliest mutations, D614G, in the spike glycoprotein. This mutation greatly enhanced the virus's transmission capacity, marking the onset of high-frequency spread of its first evolutionary clone. Initially (In 2020) seven distinct clades were identified with lineage B (Clade 19A), considered as the parental lineage of the reference strain Wuhan-Hu-1, followed by subsequent lineages like B.1 and B.1.528 from clade 20A; B.1.1/B.1.1. * from clade 20B; B.1, B.1.356, B.1.575 from clade 20C; C.17 (Clade 20D), and B.1.177, B.1.177.73, B.1.177.77 from clade 20E (Figure 5a and 5b). Lineage C.17 (Clade 20D), partially aliased to B.1.1.1.17, is similar to the lineage that emerged in Egypt, suggesting an introduction event from Egypt. Substantial spread was observed from clades 20A and 20B through community transmission. Notably, lineage B.1.528 formed monophyletic clusters nationwide. The lineages and sub-lineages from the aforementioned clades became less frequent in the last quarter of 2020. Over time, other variants have been identified and classified as VOC. Among them, one of the first VOC introduced in Morocco was the Alpha variant (Clade 20I), isolated during the first quarter of 2021, representing 7.7% of the analyzed genomic sequences (Table 2). Over time, Morocco has witnessed the emergence of other substantial variants, as shown in Figure 5a and 5b.

The Beta variant from clade 20H had a relatively low proportion with a prevalence of 0.26%, while the Eta variant from clade 21D had a proportion of 0.31% of all genomic sequences. Additionally, the Kappa variant from clade 21B resulted from an introduction event in Morocco with a single identified sequence (0.04%). This sequence was closely related to the B.1.617.1 sequence identified in India. On the other hand, the Mu variant from clade 21H, had a proportion of 0.62% of the analyzed sequences and was closely related to B.1.621 sequences from Colombia. As for the Delta variant and its subvariants (AY. *), their prevalence was 11.30%. Furthermore, the phylogenetic analysis of the Delta variant and its subvariants revealed varied distribution in three distinct clades, namely 21A, 21I, and 21J in comparison with the original strain (Figure 5a), highlighting several phylogenetically distinct clusters. Towards the end of the year 2021, Omicron and its subvariants (Figure 5a) emerged spectacularly, representing 59.5% of all genomic sequences analyzed. This emergence has been marked by the presence of fifteen distinct clades (21K, 21L, 22A, 22B, 22C, 22D, 22E, 22F, 23A, 23B, 23C, 23D, 23E, 23F, 23G) constituting several phylogenetically distinct groups, suggesting multiple noteworthy genetic divergences and different introduction events

(Figure 6a). In the following lines, we will examine in more detail the emergence of Omicron and its sub-variants.

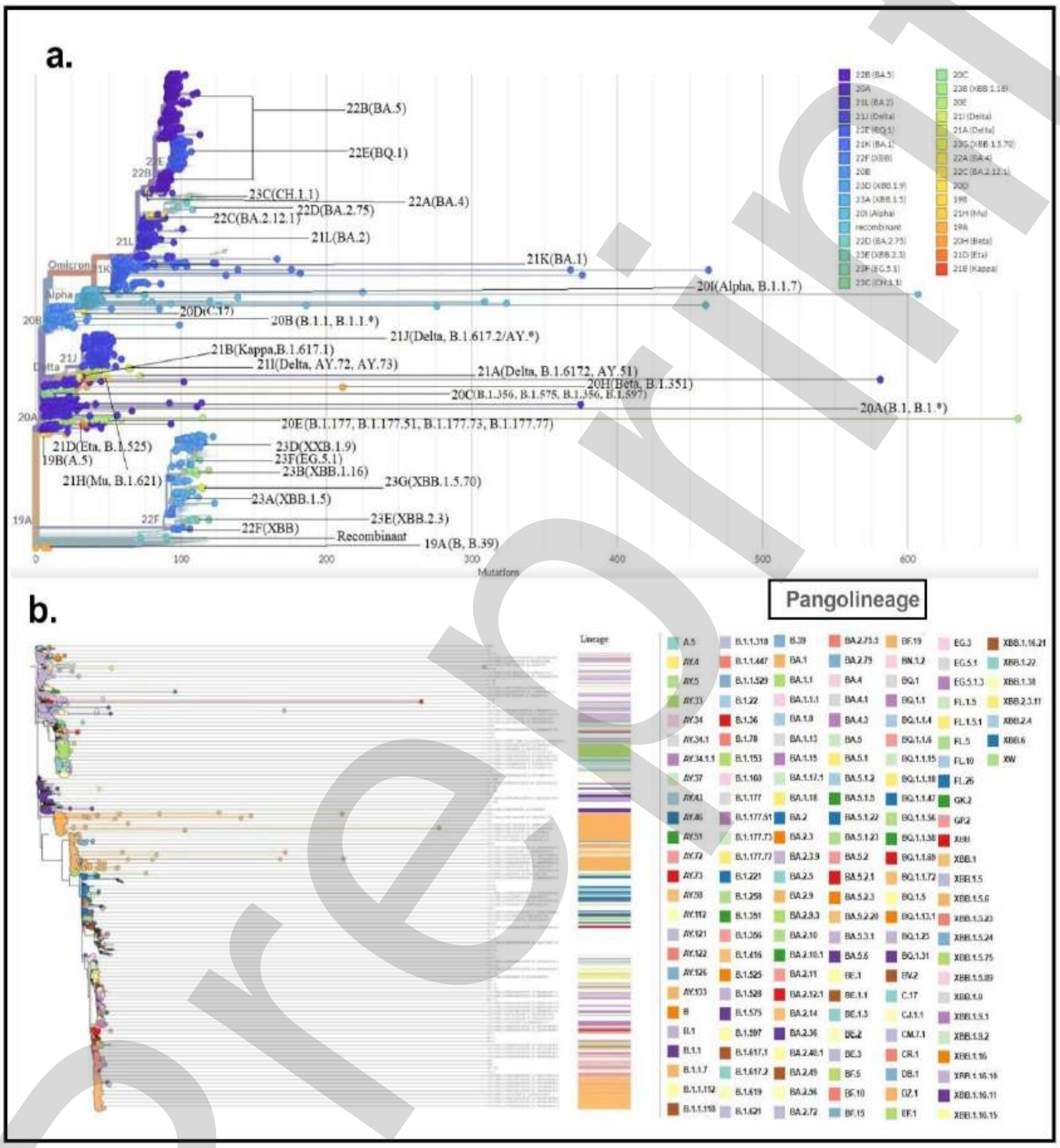

Figure 5: Phylogenetic tree generated from 2274 genomic sequences of Moroccan isolates. **a**. Maximum likelihood cladistic phylogenetic tree generated by Nextclade. **b**. Phylogenetic tree of the 2274 genomic sequences and attribution of different Pango lineages performed using Microreact.

Phylogenetic analysis of genomic diversity of the Omicron variant

Among the 2274 genomic sequences analyzed, 1353 sequences were attributed to the Omicron variant and its subvariants (Table 1). A total of 96450 mutation events were identified, with the majority being SNPs representing 66.2% (63899/96450) of all events, followed by silent SNPs with a prevalence of 21.5% (20808/96450). SNPs involving a stop codon represented only 0.1% (10/96450) of all events. Transition-type substitutions were the most frequent. The C>T transition was the most frequent event with a prevalence of 34.9% (33,673/96,450). Subsequently, G>A and A>G transitions were observed with frequencies of 11.5% (11114/96450) and 9.7% (9386/96450), respectively. Among transversion substitutions, the most notable event was T>G, accounting for 6.8% (6621/96450) of all mutations.

The phylogenetic analysis of Omicron evolution and its subvariants, as illustrated in Figure 6a and 6b, revealed a total of 106 Pango lineages. The analysis revealed fifteen Nextstrain clades phylogenetically related to Omicron and its subvariants, along with one case of recombinant lineage with unassigned clade. Clades 21K (BA.1), 21L (BA.2), 22A (BA.4), 22B (BA.5), 22C (BA.2.12.1), 22D (BA.2.75), 22E (BQ.1), and 23C (CH.1.1) are the progeny of clade 20B, as illustrated in Figure 3a. In particular, clades 21K, 21L, and 22B, phylogenetically related to BA.1, BA.2, and BA.5 lineages respectively, showed a remarkable evolution in Morocco throughout the year, with a notable number of infection cases recorded. In addition, Omicron variant BQ.1 and one of its sublineages, BQ.1.1, both descendants of BA.5, were among the dominant subvariants emerging towards November 2022. Clade 22F, phylogenetically related to recombinant lineage XBB, commonly known as «gryphon », is one of Omicron's subvariants with several progenies such as XBB.1 (clade 22F) «hippogryph», 23A (XBB.1.5) « kraken», 23B (XBB.1.16) «Arcturus» etc. These sub-lineages are respectively considered grandparent, parent, and sub-lineage, respectively. Furthermore, clades such as 23D (XBB.1.9), hosting various subvariants including XBB.1.9.2, as well as clade 23F, hosting variant EG.5.1 named «Eris» and its sub-lineages (Figure 6a). Conversely, clades 22F (phylogenetically related to XBB), 23A (XBB.1.5), 23B (XBB.1.16), 23D (XBB.1.9), 23E (XBB.2.3), 23F (EG.5.1), 23G (XBB.1.5.70) as well as the recombinant variant (XW) are subvariants sporadically identified in Morocco between the last quarter of 2022 and 2023.

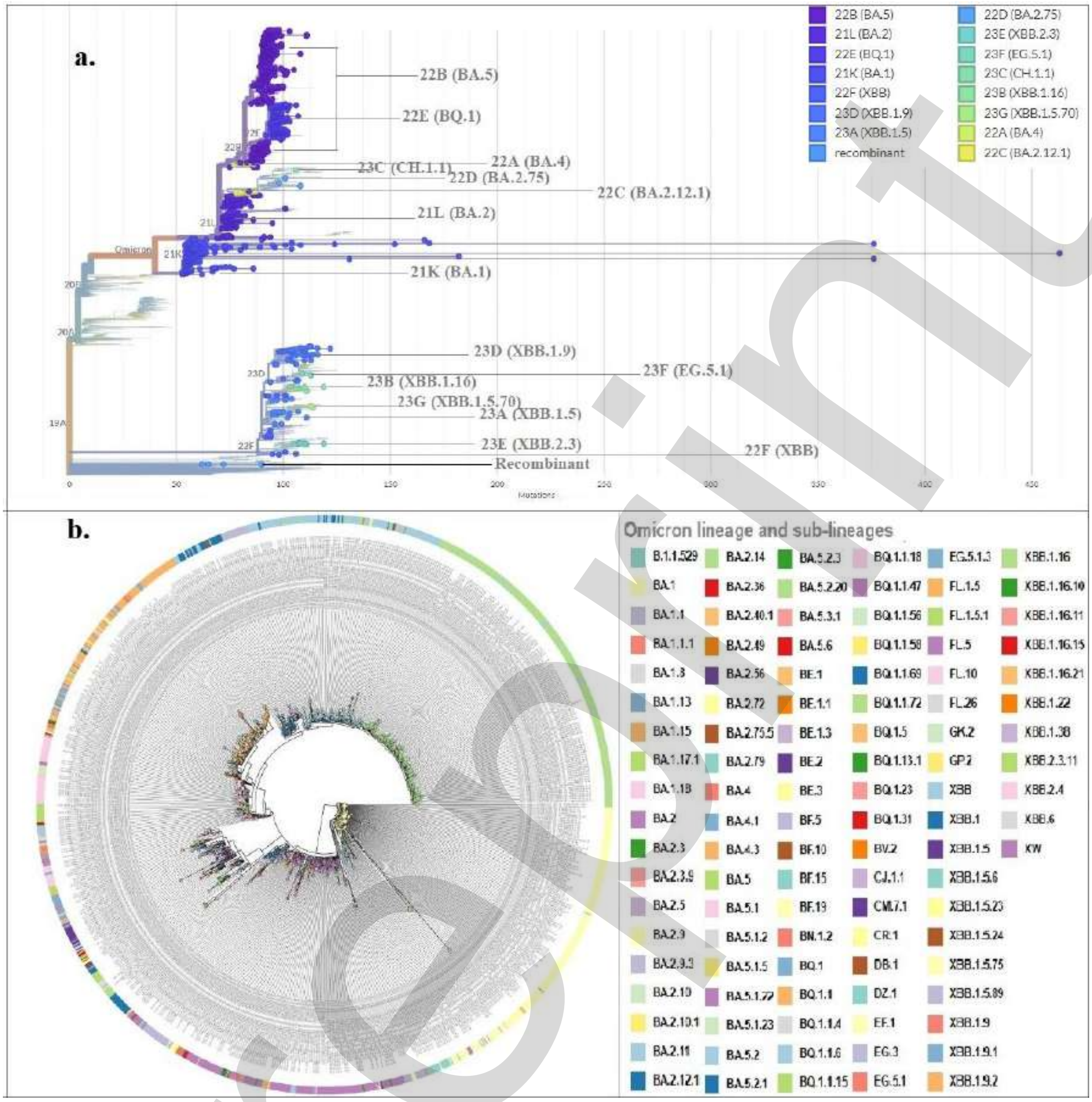

Figure 6: Estimated maximum likelihood phylogeny of convergent evolution of Omicron strain and its progeny isolated in Morocco. a) Phylogenetic tree of Omicron isolates generated by Nextclade, illustrating their classification into different clades based on their evolutionary relationships. b) Phylogenetic tree of the Omicron lineage and its derivatives generated by Microreact. The phylogenetic tree generated from the 1353 sequences attributed to Omicron and its sub-variants shows how these variants are linked by common ancestors, illustrating their genetic diversity and evolutionary relationships.

Genomic analysis of the lineage B.1.528

The analysis of the 25 genomic sequences of the B.1.528 lineage revealed an important proportion of genetic polymorphism in comparison with the reference sequence. A total of 231 mutation events were identified, with SNPs representing 44.6% (103/231) of all events.

Silent SNPs were observed with a frequency of 27.2% (63/231), while deletions and insertions had a prevalence of 3.9% (9/231) respectively. In addition, mutations in extragenic regions (UTR region) had a prevalence of 20.3% (47/231). Transition-type substitutions were the most common event. The C>T substitution was the most frequent event with 49.3% (114/231). This genomic variation was followed by T>A transversion events with a frequency of 12.1% (28/231). The notable mutations detected in the B.1.528 lineage are listed in Table 5.

Table 5: Notable mutations observed in the lineage B.1.528.

Genomic region	Nucleotide variation	Amino acid change	number of sequences	Frequency (%)	Genetic variations	Mutation type
Spike protein	A23403G	S: D614G	25	100%	SNP	Nonsynonymous
Spike protein	G24038T	S: V826L	2	8%	SNP	Nonsynonymous
Spike protein	G21578T	S: V6F	5	20%	SNP	Nonsynonymous
Spike protein	C25350T	S: P1263L	2	8%	SNP	Nonsynonymous
Spike protein	G23426T	S: V622F	1	4%	SNP	Nonsynonymous
Spike protein	T21572C	S: F4L	1	4%	SNP	Nonsynonymous
Spike protein	C22000A	S: H146Q	1	4%	SNP	Nonsynonymous
Spike protein	T21566G	S: F2V	1	4%	SNP	Nonsynonymous
Spike protein	T21570G	S: V3G	1	4%	SNP	Nonsynonymous
Spike protein	G24236T	S: A892S	1	4%	SNP	Nonsynonymous
Nucleocapsid protein	C29421T	N: P383L	2	8%	SNP	Nonsynonymous
NSP2	G2458T	ORF1ab/NSP2: M551I	2	8%	SNP	Nonsynonymous
NSP3	C3037T	ORF1ab /NSP3: F106F	25	100%	SNP	Synonymous
NSP6	C11139T	ORF1ab /NSP6: A56V	1	4%	SNP	Nonsynonymous
NSP6	G11306A	ORF1ab /NSP6: D112N	1	4%	SNP	Nonsynonymous
NSP6	T11309A	ORF1ab /NSP6: C113S	1	4%	SNP	Nonsynonymous
NSP12	C14408T	ORF1ab /NSP12: P323L	25	100%	SNP	Nonsynonymous
NSP14	C18508T	ORF1ab /NSP14: L157F	25	100%	SNP	Nonsynonymous
NSP14	T19420A	ORF1ab /NSP14: S461T	1	4%	SNP	Nonsynonymous
NSP14	T19420C	ORF1ab /NSP14: S461P	1	4%	SNP	Nonsynonymous
NSP14	C18568T	ORF1ab /NSP14: L177F	1	4%	SNP	Nonsynonymous
NSP15	G19656T	ORF1ab /NSP15: K12N	1	4%	SNP	Nonsynonymous
5'UTR	C241T	5'UTR: 241(NA)	25	100%	Extragenic	Non-coding

SNP= Single Nucleotide Polymorphism, NSP= Non-Structural Protein

The lineage « B.1.528 » constituted a monophyletic cluster within the Moroccan population (Figure 7a), with a proportion of 1.1% (25/2274). Twenty-four sequences of the B.1.528 lineage were detected in Ouarzazate and one sequence in Rabat. The analysis of the phylogenetic tree based on 89 sequences isolated in 2020 reveals that lineage B.1.528 evolved from the ancestral lineage, displaying several additional distinct mutations in addition to the initial D614G mutation, from which many other lineages have emerged (Figure 7a and 7b). Moreover, the emergence of the B.1.528 lineage has sparked increasing interest due to the identification of specific mutations in its genomic sequence. These include the D614G mutation located in the S1 domain, followed by mutations such as F2V, V3G, F4L, and V6F in the NTD domain (S1). Additionally, mutations V826L and A892S have been observed in the S2 domain, along with the amino acid substitution of P1263L in the C-terminal region (Figure 7b). The figure 7b provides a detailed insight into the mutations present in the 25 sequences of B.1.528, illustrating the genetic diversity and evolutionary dynamics of this lineage. These mutations can be traced through the tree, providing substantial information on the evolution and differentiation of key mutations within the lineage B.1.528 over time.

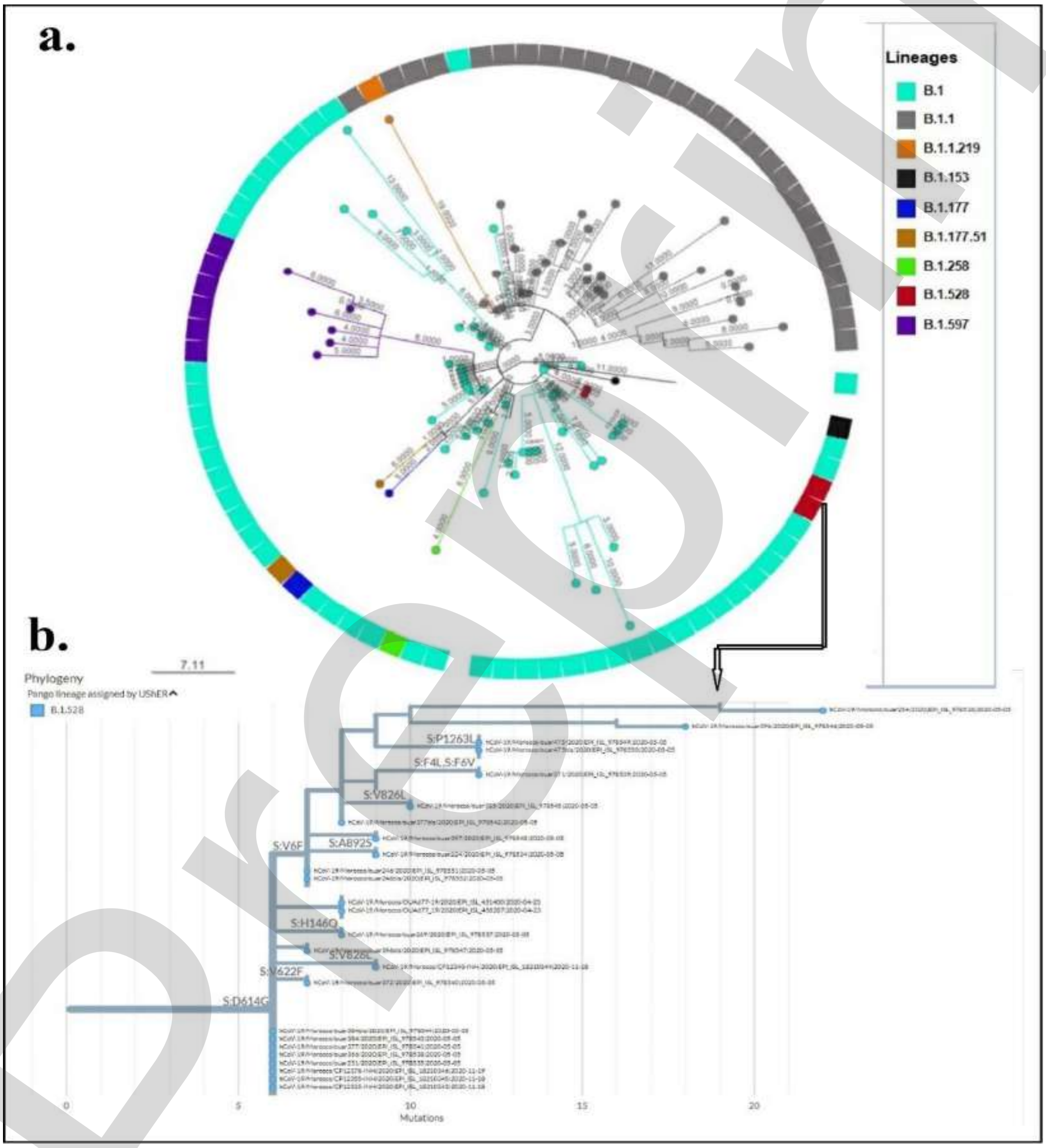

Figure 7: Maximum likelihood phylogenetic tree analysis of 89 genomic sequences of SARS-CoV-2 isolated and sequenced in Morocco in 2020. **a**) Phylogenetic tree of lineages isolated in 2020, **b**) The lineage « B.1.528 » with notable amino acid substitution observed in the 25 isolates.

DISCUSSION

This study stands out as one of the first to investigate the complete genomic divergence of 2,274 sequences of severe acute respiratory syndrome coronavirus 2 in Morocco over the three years following the COVID-19 pandemic. Furthermore, this study highlights the impact of SARS-CoV-2 evolution on public health by identifying several distinct mutational signatures that characterize notable variants, thus providing a solid basis for future research.

During this pandemic, the year 2020 witnessed a turning point in the evolution of SARS-CoV-2 with the emergence of the D614G mutation [16]. Some mutations in the genome of SARS-CoV-2 have been recognized as determining mutations in variations in transmissibility (more or less contagious), antigenicity variation (immune evasion capacity and/or risk of reinfections), as well as variations in pathogenicity (virulence variation with more or less severe events) [16–18]. Mutations in the RBD/RBM region could impact how the virus interacts with the ACE2 receptor, potentially influencing the virus's transmissibility and could assist the virus to escape immune recognition. Moreover, mutations near the furin cleavage site (S1/S2) could accelerate the Spike protein cleavage, enhancing virus infectivity (Figure 4b) [16–18].

In Morocco in 2020, dominant lineages such as, B.1, B.1.528, B.1.177, and B.1.1 with the D614G mutation were prevalent, mainly in clade G. All subsequent SARS-CoV-2 lineages/variants in Morocco are derived from lineage B.1 (clade 20A). These lineages emerged approximately at the same time, indicating multiple introduction events in Morocco, except for the lineage « B.1.528 », which is closely related to clade 20A. This lineage formed a distinct cluster nationally and carried notable mutations in spike glycoprotein, including D614G, V826L, and A892S (Figure 4b and table 5) [19]. These amino acid substitutions, located outside the predicted immunological epitope region, could affect transmissibility and immune evasion [19]. Furthermore, this study highlighted the P1263L mutation, located in the C-terminal region, within B.1.528 sequences. This mutation has also been reported in certain lineages in Europe and Oceania [20]. Interestingly, previous studies have suggested that the P1263L mutation, could potentially alter the structure and function of the SARS-CoV-2 spike protein, leading to decreased infectivity of the virus [20,21]. It should also be noted that non-synonymous mutations such as F2V, V3G, F4L, V6F, and H146Q in the B.1.528 sequences are located within the region of the predicted epitopes [22]. Previous studies have reported that these mutations could contribute to increased transmissibility and pathogenicity of SARS-CoV-2 [22]. Moreover, their location within epitopes could potentially confer to the virus the ability to escape the immune system [21,22]. Consequently, all these key mutations in the spike glycoprotein, particularly those associated with the D614G mutation present in the sequences of lineage B.1.528 (Figure 7b and Table 5), played a crucial role in the emergence of this lineage within clade 20A. These mutations distinguish it from other lineages within the same clade, as well as from lineages of ancestral clades such as 19A, 19B, 20B, and 20C, where the dominant mutation in the Spike protein was sometimes only the D614G mutation (Table 4).

In January 2021, the Alpha variant (B.1.1.7), previously designated as VOC 202012/01, and commonly known as GRY clade (formerly GR/501Y.V1), was identified in Sidi Lahcen, Morocco. Known for its increased transmissibility, alpha variant led to strengthened health measures in Morocco. The N501Y mutation in the Spike protein, found in the Alpha, Beta, Mu, and Omicron variants (Figure 4a), is known to enhance virus transmissibility [16,23]. In addition, it has been reported that the deletions such as H69del/V70del and Y144del in the NTD region would be involved to immune evasion and increased virulence of SARS-CoV-2 [17,23]. The P681H substitution in the S1/S2 cleavage loop, observed in several variants, would favor the processivity of spike protein cleavage [16,18,24]. Moreover, the coexistence of the N501Y and P681H mutations would increase transmission and disease severity [16,18,24]. However, it is interesting to note that the introduction of the Alpha variant (B.1.1.7) marked a major change in the evolutionary dynamics of SARS-CoV-2 in Morocco. Furthermore, our observations corroborate the findings of a previous study, which reported that at the beginning of 2021, the Alpha variant was dominant in Tunisia, Algeria, and Morocco, while the Eta variant predominated in Egypt, Libya, and Mauritania [25]. In contrast, while Hamzaoui and colleagues observed a frequency of less than 1% for the Alpha variant and an absence of sequences attributed to the Beta variant, our study showed a prevalence of 7.74% for the Alpha variant and 0.31% for the Beta variant [25].

In May 2021, the Beta variant (B.1.351) of SARS-CoV-2 has been detected in Casablanca, Morocco. This variant is characterized by key mutations in the RBD/RBM (N501Y, E484K, K417N) and NTD (L242del/L244del) regions. Previous studies have suggested that these mutations may contribute to SARS-CoV-2's immune evasion and increased risk of reinfections [18,23,26]. However, despite concerns about B.1.351 in some countries, the impact of B.1.351 in Morocco has been relatively minor [27,28]. Morocco later detected the Eta, Kappa, and Mu variants. Eta, formerly known as G/484K.V3, with the first case identified in Morocco in March 2021, carries the E484K mutation known for its role to circumvent host immune responses [16,18,23]. The Mu variant (B.1.621), previous VOI, was identified in Morocco on May 22, 2021. This variant (B.1.621) contains several notable mutations, including T95I, A262S, D614G, D950N, E484K, N501Y, P681H, R346K, and Y144del, which are likely to increase immune evasion and virus transmissibility [29,30]. In general, the Beta, Eta, Kappa, and Mu variants did not have a major impact in Morocco, possibly due to effective health measures or competition with more widespread variants such as Alpha, Delta, and Omicron.

In the last quarter of 2020, the epidemiological dynamics of COVID-19 have evolved substantially. India, in particular, has faced an outbreak of several SARS-CoV-2 variants that have spread to affect many countries. Among these variants, B.1.617.1 and B.1.617.2 emerged, both sharing key mutations such as G142D, L452R, and P681R in the spike glycoprotein, along with other distinct mutations likely to increase virus transmission and promote immune evasion [16,18,31–34]. It is important to note that the Kappa and Delta variants have been labeled under the generic name of «double mutants», due to the presence of key mutations L452R and E484Q (Figure 4a), which had never been observed together before the emergence of these variants [16,18,31–34]. However, the epidemiological situation in Morocco in 2021 was marked by the emergence of B.1.617.2 variant and its sub-variants (B.1.617.2/AY. *) which are phylogenetically related to clades 21A, 21I, and 21J. The first strain of the delta variant (B.1.617.2) was isolated on 22 April 2021, in Casablanca. Following their introduction, they progressively became the predominant variant, supplanting other variants. Between June and September 2021, B.1.617.2 underwent rapid evolution, giving rise to a major evolutionary node with the emergence of subvariants such as AY.33, AY.112, and AY.122, which became dominant nationwide. Compared to the double mutant B.1.617.1, the B.1.617.2 variant and its subvariants harbored mutations in the NTD domain such as T19R, T29A, G142D, F157del, E156G, R158del, T250I, T299I, and in the RBD domain with mutations L452R and T478K. The SD2 region also revealed mutations such as Q613H, D614G, D950N, as well as an amino acid substitution around the furin cleavage site, P681R (a mutation shared by Delta and Kappa). All these mutations differentiated them from other variants by their ability to enhance their transmissibility and their capacity to escape the immune system. Although the delta variant is devoid of the E484Q mutation, analogous to the immune escape mutation « E484K » found in other variants such as B.1.351 and B.1.525, respectively. The T478K mutation hosted in the spike glycoprotein could accelerate the replication and transmissibility of the pathogen [16,35]. Analogously to the D614G mutation, it is hypothesized that the Q613H mutation may enhance the virulence of SARS-CoV-2 by facilitating the cleavage of the spike protein, thereby enhancing fusion between the viral envelope and the cell membrane [36]. Interestingly, among the mutations of the aforementioned delta sub-variants, mutations such as T29A, T250I, T299I, and Q613H were specific to subvariants AY.33, AY.112, and AY.122. These mutations may confer to theses subvariant a selective advantage by enhancing their transmission capacity and their ability to escape the immune system [37]. Moreover, previous studies have reported that the F157del and R158del deletions, as well as the L452R and T478K mutations (Figure 4a), would confer to Delta and its sub-variants the ability to escape neutralizing antibodies [18,35]. Consequently, all these characteristics make the Delta variant and its subvariants (AY.*) one of the most virulent variants of SARS-CoV-2 [38]. Furthermore, it should be noted that during the emergence of the Delta variant and its sub-variants, the AY.33 and B.1.617.2 lineages predominated in Morocco (Figure 1a and Table 3), while in Egypt and Algeria, the B.1.617.2 variant was dominant. In contrast, the AY.4, AY.122, and AY.34.1 subvariants were the most dominant in Libya, Tunisia, and Mauritania, respectively [25]. These observations highlight the distinct genetic diversity, reflecting the broader global trends in the evolution of the Delta variant and its sub-variants [25].

The epidemiological evolution observed in Morocco during the last quarter of 2021 was marked by a transitional period with the emergence of the Omicron variant and its derivatives designated under the GRA nomenclature in GISAID (formerly GR/484A). As of 2022, 82.3% of genomic sequences were assigned to omicron and its sub-variants (table 1). The introduction of the Omicron variant and its subvariants in Morocco resulted in a major epidemiological shift, relegating the previously dominant Delta variant to a marginal position. Among the Omicron strains that have led to substantial infection in Morocco, we note BA.1 and BA.1.1 (displays clade 21K), as well as BA.2 and B.1.1.529 (clade 21L), identified in Morocco in December 2021. However, in June 2022, Morocco rapidly observed a surge in infection cases, with the emergence and concurrent coexistence of subvariants from clade 22B such as BA.5.2.20, BA.5, BA.5.1, BA.5.2.1, and BF.5 (subvariants emerged from Botswana, Hong Kong, and South Africa), succeeding the lineages BA.1 and BA.2, respectively (Figure 1a and Table 4). Several derivatives of BA.1 sub-variant (BA.1 phylogenetically related to clade 21K) were also identified, notably the sub-lineage BA.1.1 (BA.1+ R346K), distinguished from BA.1 by the additional presence of mutation R346K. In Morocco, alongside lineage BA.1, lineage BA.2 and its sub-lineages emerged, identified in February

2022, with a peak reached on April 2022. Lineage BA.1 harbors certain mutations, such as A67V, H69del/V70del, T95I, Y144del, L212I, N211del, ins214EPE, S371L, G446S, G496S, T547K, N856K, L981F, distinguishing it distinctly from lineage BA.2. Conversely, lineage BA.2 is distinguished from BA.1 by specific mutations such as T19I, L24del, P25del/P26del, A27S, V213G, S371F, T376A, D405N, R408S. Our analysis results corroborate the conclusions presented by Balupuri and colleagues in 2023 [39].

Interestingly, variants BA.1, BA.2, BA.4, and BA.5 share several mutation similarities. The mutations L452R and F486V are specific to BA.4 and BA.5, while H69del/V70del are specific to BA.1, BA.4, and BA.5. Additionally, mutation Q493R is specific to BA.1 and BA.2 [40,41]. Moreover, sub-variants BA.4 and BA.5 also display mutation similarities in their spike protein with previous VOCs. These mutations include the N501Y mutation hosted by the Alpha, Beta, and Mu variants; H69del/V70del hosted by the Alpha variant; K417N observed in sequences of Beta variant; E484K present in sequences of the Beta and Eta variants, while L452R and T478K are observed in sequences of the Delta variant (Figure 4a). Unlike the Beta variant, which features the E484K mutation, sub-variants BA.4 and BA.5 harbor the E484A mutation (Figure 4a). The mutation profiles described in this study align with findings reported in previous studies [26,40–44]. Over time, the BA.2 variant has gradually emerged as the predominant sub-variant, relegating the BA.1, BA.1.1, and B.1.1.529 variants to a marginal position (Figure 1a). Furthermore, a notable descendant of the BA.2 lineage, the sub-variant BA.2.12.1 (BA.2+L452Q+S704L), is characterized by an additional mutation, L452Q and S704L, absent in BA.2. Nevertheless, most lineages such as BA.2 and its sub-lineages, XBB and its derivatives, the Beta, Kappa, Mu, B.1, Delta variants and their derivatives were mostly devoid of the H69del/V70del mutation, unlike some omicron sub-variants, the Alpha and Eta variant (Figure 3a). These observations correlate with previous research findings [40,45,46]. In addition to the previously mentioned subvariants of Omicron, BA.5 and BA.4 are among those being closely monitored. They are characterized by notable mutations such as L452R and F486V. In the same perspective, the BA.5.2.20 subvariant, a progeny of BA.5, has been observed with the highest proportion of infections in the Moroccan population (Figure 1 and Table 3). Furthermore, it is important to note that subvariants BA.2.12.1, BA.4, and BA.5 have also shown a higher rate of infection than lineage BA.2 in the Moroccan population, in part due to mutations such as S371F, D405N, and R408S, which could allow these subvariants to evade antibody control [42,47]. Remarkably, in Morocco, the evolution of SARS-CoV-2 has been characterized by substantial genetic diversity, with a particularly marked predominance of the Omicron variant and its sub-variants over time. This evolutionary trend has also been observed in several countries during the pandemic [25,48]. In Morocco, Omicron and its sub-variants accounted for 59.5% of sequences, while Delta was observed at frequency of 11.30%. Approximate prevalences of Omicron and Delta variants have been reported elsewhere [25]. Hamzaoui and colleagues reported that in Egypt, 55.3% of sequences were assigned to Omicron and 12.1% to Delta. In addition, Omicron sub-variants such as BA.5.2.20 and BA.1 were dominant in Morocco (Figure 1a), accounting for 11% and 10.1% of sequences respectively (Table 3). Our observations are consistent with those of Sarkar and colleagues, who reported a predominance of BA.1 in Europe in late 2021 and early 2022 [48]. This trend was also observed in the USA, Brazil and South Africa, in which BA.1 prevailed over BA.2. Furthermore, in June 2022, a trend of BA.4 and BA.5 sub-variants was observed. This observation is consistent with previous studies [25,48]. However, unlike some North African countries, such as Egypt, the BA.5.2 (13.75%) and B.1.617.2 (8.74%) variants were the most prevalent, while in Algeria, BA.5.2 (17.27%) and B.1.617.2 (12%) were predominant [25]. Nevertheless, despite the notable spread of Omicron worldwide, some pre-Omicron lineages persisted and were dominant in other countries [25,48]. Notably, in Libya, B.1.525 (44.94%) and A (9.55%) predominated; in Tunisia, AY.122 (24.23%) and B.1.1.7 (21.67%) were dominant; and in Mauritania, B.1.525 (31.03%) and AY.34.1 (27.58%) were the most dominant [25]. In general, the peaks of predominance and the heterogeneity of SARS-CoV-2 lineages observed in this study and those reported in previous studies highlight the genetic diversity of SARS-CoV-2 during its worldwide spread [25,48].

Furthermore, in Morocco, the BQ.1 lineage (BA.5+R346T+N460K) and one of its sub-lineages BQ.1.1 (BQ.1+R346T), both descendants of BA.5, which emerged from Botswana/Hong Kong/South Africa, have attracted particular interest and were among the monitored sub-variants. These lineages (BQ.1 and BQ.1.1), partially aliased to BA.5.3.1.1.1.1.1.1 were identified in September and November 2022, respectively. These subvariants stood out by several key mutations, notably R346T, K417N, N440K, K444T, L452R, N460K, F486V, E484A, and N501Y, known for their ability to escape antibodies. In particular, the mutations R346T, K444T, and F486V appear to promote evasion of humoral immune response induced by vaccination [47,49–51]. Furthermore, the mutations L452R, N460K, and R346T observed in the spike glycoprotein would be correlated with the virus's ability to bind more efficiently to ACE2 receptors. Consequently, this could increase the virus's transmissibility or its ability to infect cells. These observations have been reported in previous studies [49–51]. It is worth noting that in Morocco, BQ.1.1 and BQ.1, rapidly reached their epidemic peak in December 2022 (Figure 1a). Similar epidemiological trends have been described elsewhere, such as in the United States and Europe [52]. While BQ.1 and BQ.1.1 were declining in Morocco towards the end of 2022, new sub-variants emerged simultaneously, among which was the XBB (22F) lineage and its progeny (Figure 6a, Table 4). Unlike what was observed in Europe and the United States, cases of XBB emerged with rapid increases. In contrast, our study revealed sporadic circulation of XBB and its progeny [52]. In this regard, the ancestor (XBB) of these new subvariants was characterized by mutations such as D614G, D405N, E484A, F486S, F490S, G446S, K417N, N440K, N460K, P681H, R408S, S371F, S477N, T376A, T478K, and V445P, all known for their ability to increase transmissibility and evade the immune system [43,53,54]. Among the descendants of XBB, the GP.2 subvariant of clade 23C (CH.1.1) stood out due to the presence of the key mutation K444T (XBB+ K444T), enhancing the virus's evasive capacity. Furthermore, the mutation G446S observed in XBB and BA.2.75.5 lineage related to BA.2.75 conferred them a resistance to antibodies [49]. Additionally, subvariants such as XBB.1.5 (XBB+ F486P), XBB.1.16 (XBB+E180V+F486P), and EG.5.1 (XBB.1+ Q52H + F456L + F486P / EG.5 + Q52H), harbored mutations that conferred them an evasive capacity against neutralizing antibodies, notably R408S, G446, G339D, D405N, F490S, K417N, and Q498R (Table 4) [54]. It should be noted that EG.5.1 (Eris) from clade 23F, descendant of XBB.1.9.2 (Clade 23D), has a mutational profile in the spike glycoprotein similar to that of XBB.1.5 [54]. However, according to updated data from the WHO on 09 February 2024, these previously mentioned lineages are classified as VOIs currently circulating worldwide [55]. Moreover, subvariants such as XBB.1.9.1 (XBB.1 + F486P) and XBB.2.3.11 (XBB + T478R + P521S), derived from the XBB.2.3 lineage (23E) (XBB.1 + D253G + F486P + P521S), analyzed in this study were also less frequent (Table 4). Nevertheless, these lineages are classified by the WHO as VUMs currently in circulation [55]. Furthermore, this study supports previous observations suggesting that Omicron subvariants such as BA.1.1, BQ.1, BQ.1.1, XBB, and its subvariants present distinct characteristics and are associated with variable epidemiological trends worldwide [48]. According to our results, these subvariants which appeared between the end of 2022 and 2023 (XBB and its derivatives) did not lead to large-scale epidemics, as observed elsewhere [48,56]. This observation could be attributed to the implementation of health safety measures and protocols, and the mass vaccination campaign in Morocco aimed at achieving herd immunity in order to curb the spread of SARS-CoV-2.

In addition to the key mutations observed in the spike glycoprotein of different lineages of SARS-CoV-2, some notable mutations were also observed in other structural proteins, particularly in the nucleocapsid, such as P13L and R203K/G204R (Figure 4a). These mutations are associated with increased infectivity, structural stability, and virulence of the virus, which could confer on it a better ability to escape the immune response [57,58]. In general, our analysis revealed a prevalence of 71% of R203K/G204R mutations in our isolates, significantly higher than the 55% prevalence reported in the study of Ahmad and his colleagues [59]. These mutations are observed at high frequency in the omicron and its sub-variants, as well as in the Alpha variant and the B.1.1 lineage (Figure 4a).

However, this study had several limitations. This study included both high and low coverage genomic sequences. Although low-coverage sequences can provide additional information on the genomic diversity of the virus, this can also pose methodological challenges and limitations in the interpretation of results, with possible under-representation of certain viral strains or genome regions. In addition, in-depth documentation of mutations in structural and non-structural proteins could provide a better understanding of the genomic diversity of Moroccan isolates of SARS-CoV-2 and their impact on the dynamics of the virus.

CONCLUSION

This study provided a detailed analysis of the genomic epidemiology and genetic diversity of SARS-CoV-2 lineages identified in Morocco during the three years of the pandemic, enabling a better understanding of the evolution and phylogenetic relationships among different lineages. Several lineages identified in Morocco were closely related to those observed worldwide, except for lineage B.1.528, before their local spread, highlighting the impact of human mobility on the introduction and spread of these lineages during the pandemic. Viral dynamics in Morocco, characterized by a predominance of Alpha, Delta, Omicron variants, and their subvariants, reflected global trends in their evolution. However, the epidemiological trends of some Delta and Omicron subvariants showed variable patterns compared to those observed in other countries. Additionally, several key mutations identified within the lineages analyzed were correlated with variations in transmissibility, pathogenicity and antigenicity, which could have contributed to affecting vaccine efficacy and pandemic management. However, the set-up of the SARS-CoV-2 genomic surveillance consortium in Morocco and vaccination campaigns have contributed to control and reduce infection rates and severe forms of COVID-19, thus mitigating the impact of infections at national level.

Ethics Approval

This study did not require ethical clearance and approval, since the genome sequence data used were extracted from the GISAID EpiCoV database.

Conflicts of Interest

The authors declare no competing interests.

Funding information

None.

Availability of Data

The results presented in this study are derived from the analysis of 2274 genomic sequences of metadata available on:

GISAID Identifier: EPI_SET_240225ry

DOI: 10.55876/gis8.240225ry

Author contributions

Study Design: AB, AM, and KB. Data Collection and analysis: SD, NN and SPZ. Data

analyses and interpretation: SD, NN, JN and SPZ. Manuscript Writing: SD and NN. All authors

contributed to revisions and have read and approved the final manuscript.

Acknowledgment

We gratefully acknowledge all data contributors, including the authors and their respective laboratories for sample collection, as well as the submitting laboratories for producing the genetic sequences and metadata, and for sharing them through the GISAID Initiative (accessible via GISAID's EpiCoV database at http://www.gisaid.org), upon which this study is based.

REFERENCES

1. Zhang Y, Belayachi J, Yang Y, Fu Q, Rodewald L, Li H, et al. Real-world study of the effectiveness of BBIBP-CorV (Sinopharm) COVID-19 vaccine in the Kingdom of Morocco. BMC Public Health. 20 août 2022;22(1):1584.

2. Weekly epidemiological update on COVID-19 - 10 August 2023 [Internet]. [cité 18 avr 2024]. Disponible sur: https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---10-august-2023

3. Najimi N, Tajount L, Regragui Z, Remz C, Ait-Lhaj-Mhand R, Kadi C, et al. Pre-pandemic antibodies screening against SARS-CoV-2 and virus detection among children diagnosed with eruptive fevers. Int J Immunopathol Pharmacol. 1 janv 2024;38:03946320241260633.

4. Voskarides K. SARS-CoV-2: tracing the origin, tracking the evolution. BMC Med Genomics. 18 mars 2022;15(1):62.

5. Yu M, Li J, Gao W, Li Z, Zhang W. Multiple E3 ligases act as antiviral factors against SARS-CoV-2 via inducing the ubiquitination and degradation of ORF9b. J Virol. 6 mai 2024;98(6):e01624-23.

6. Al-Qaaneh AM, Alshammari T, Aldahhan R, Aldossary H, Alkhalifah ZA, Borgio JF. Genome composition and genetic characterization of SARS-CoV-2. Saudi J Biol Sci. 1 mars 2021;28(3):1978-89.

7. Saxena SK, Kumar S, Ansari S, Paweska JT, Maurya VK, Tripathi AK, et al. Characterization of the novel SARS-CoV-2 Omicron (B.1.1.529) variant of concern and its global perspective. J Med Virol. avr 2022;94(4):1738-44.

8. Chouikha A, Fares W, Laamari A, Haddad-Boubaker S, Belaiba Z, Ghedira K, et al. Molecular Epidemiology of SARS-CoV-2 in Tunisia (North Africa) through Several Successive Waves of COVID-19. Viruses. 17 mars 2022;14(3):624.

9. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet Lond Engl. 20 oct 2007;370(9596):1453-7.

10. Khare S, Gurry C, Freitas L, Schultz MB, Bach G, Diallo A, et al. GISAID's Role in Pandemic Response. China CDC Wkly. 3 déc 2021;3(49):1049-51.

11. Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol. nov 2020;5(11):1403-7.

12. O'Toole Á, Scher E, Underwood A, Jackson B, Hill V, McCrone JT, et al. Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. Virus Evol. 1 déc 2021;7(2):veab064.

13. Aksamentov I, Roemer C, Hodcroft EB, Neher RA. Nextclade: clade assignment, mutation calling and quality control for viral genomes. J Open Source Softw. 30 nov 2021;6(67):3773.

14. Fernandes JD, Hinrichs AS, Clawson H, Gonzalez JN, Lee BT, Nassar LR, et al. The UCSC SARS-CoV-2 Genome Browser. Nat Genet. oct 2020;52(10):991-8.

15. Mercatelli D, Giorgi FM. Geographic and Genomic Distribution of SARS-CoV-2 Mutations. Front Microbiol. 22 juill 2020;11:1800.

16. Lina B. Les différentes phases de l'évolution moléculaire et antigénique des virus SARS-CoV-2 au cours des 20 mois suivant son émergence. Bull Acad Natl Med. janv 2022;206(1):87-99.

17. Wang Q, Guo Y, Iketani S, Nair MS, Li Z, Mohri H, et al. Antibody evasion by SARS-CoV-2 Omicron subvariants BA.2.12.1, BA.4 and BA.5. Nature. août 2022;608(7923):603-8.

18. Ghosh N, Nandi S, Saha I. A review on evolution of emerging SARS-CoV-2 variants based on spike glycoprotein. Int Immunopharmacol. 1 avr 2022;105:108565.

19. Nazario-Toole AE, Xia H, Gibbons TF. Whole-genome Sequencing of SARS-CoV-2: Using Phylogeny and Structural Modeling to Contextualize Local Viral Evolution. Mil Med. 1 janv 2022;187(1-2):e130-7.

20. Gonçalves RL, Leite TCR, Dias B de P, Caetano CC da S, de Souza ACG, Batista U da S, et al. SARS-CoV-2 mutations and where to find them: an in silico perspective of structural changes and antigenicity of the spike protein. J Biomol Struct Dyn. 3 mai 2022;40(7):3336-46.

21. Li Q, Wu J, Nie J, Zhang L, Hao H, Liu S, et al. The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity. Cell. 3 sept 2020;182(5):1284-1294.e9.

22. Odongo S, Okella H, Ndekezi C, Okee M, Namayanja M, Mujuni B, et al. Retrospective in silico mutation profiling of SARS-CoV-2 structural proteins circulating in Uganda by July 2021: Towards refinement of COVID-19 disease vaccines, diagnostics, and therapeutics. PLOS ONE. 22 déc 2022;17(12):e0279428.

23. Yi H, Wang J, Wang J, Lu Y, Zhang Y, Peng R, et al. The Emergence and Spread of Novel SARS-CoV-2 Variants. Front Public Health [Internet]. 2 août 2021 [cité 19 mars 2024];9. Disponible sur: https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2021.696664/full

24. Tao K, Tzou PL, Nouhin J, Gupta RK, de Oliveira T, Kosakovsky Pond SL, et al. The biological and clinical significance of emerging SARS-CoV-2 variants. Nat Rev Genet. déc 2021;22(12):757-73.

25. Hamzaoui Z, Ferjani S, Medini I, Charaa L, Landolsi I, Ben Ali R, et al. Genomic surveillance of SARS-CoV-2 in North Africa: 4 years of GISAID data sharing. IJID Reg. 1 juin 2024;11:100356.

26. Zhou D, Dejnirattisai W, Supasa P, Liu C, Mentzer AJ, Ginn HM, et al. Evidence of escape of SARS-CoV-2 variant B.1.351 from natural and vaccine-induced sera. Cell. 29 avr 2021;184(9):2348-2361.e6.

27. Ong SWX, Chiew CJ, Ang LW, Mak TM, Cui L, Toh MPHS, et al. Clinical and Virological Features of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Variants of Concern: A Retrospective Cohort Study Comparing B.1.1.7 (Alpha), B.1.351 (Beta), and B.1.617.2 (Delta). Clin Infect Dis. 1 juill 2022;75(1):e1128-36.

28. Basheer A, Zahoor I. Genomic Epidemiology of SARS-CoV-2 Divulge B.1, B.1.36, and B.1.1.7 as the Most Dominant Lineages in First, Second, and Third Wave of SARS-CoV-2 Infections in Pakistan. Microorganisms. 17 déc 2021;9(12):2609.

29. Laiton-Donato K, Franco-Muñoz C, Álvarez-Díaz DA, Ruiz-Moreno HA, Usme-Ciro JA, Prada DA, et al. Characterization of the emerging B.1.621 variant of interest of SARS-CoV-2. Infect Genet Evol. nov 2021;95:105038.

30. Chatterjee D, Tauzin A, Laumaea A, Gong SY, Bo Y, Guilbault A, et al. Antigenicity of the Mu (B.1.621) and A.2.5 SARS-CoV-2 Spikes. Viruses. janv 2022;14(1):144.

31. McCallum M, Marco AD, Lempp FA, Tortorici MA, Pinto D, Walls AC, et al. N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. Cell. 29 avr 2021;184(9):2332-2347.e16.

32. Wang P, Nair MS, Liu L, Iketani S, Luo Y, Guo Y, et al. Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. Nature. mai 2021;593(7857):130-5.

33. Cerutti G, Guo Y, Zhou T, Gorman J, Lee M, Rapp M, et al. Potent SARS-CoV-2 neutralizing antibodies directed against spike N-terminal domain target a single supersite. Cell Host Microbe. 12 mai 2021;29(5):819-833.e7.

34. Morang'a CM, Ngoi JM, Gyamfi J, Amuzu DSY, Nuertey BD, Soglo PM, et al. Genetic diversity of SARS-CoV-2 infections in Ghana from 2020-2021. Nat Commun. 6 mai 2022;13(1):2494.

35. Planas D, Veyer D, Baidaliuk A, Staropoli I, Guivel-Benhassine F, Rajah MM, et al. Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. Nature. août 2021;596(7871):276-80.

36. Alkhatib M, Svicher V, Salpini R, Ambrosio FA, Bellocchi MC, Carioti L, et al. SARS-CoV-2 Variants and Their Relevant Mutational Profiles: Update Summer 2021. Microbiol Spectr. 17 nov 2021;9(3):e01096-21.

37. Hemlali M, Chouati T, Ghammaz H, Melloul M, Alaoui Amine S, Rhoulam S, et al. Coding-Complete Genome Sequences of a Delta Subvariant (AY.33) of SARS-CoV-2 Obtained from Moroccan COVID-19 Patients. Microbiol Resour Announc. 3 févr 2022;11(2):e01099-21.

38. Sahni C, Chowdhury PBR, Devadas D, Ashish A, Singh NK, Yadav A, et al. SARS-CoV-2 Mutations Responsible for Immune Evasion Leading to Breakthrough Infection. Cureus [Internet]. 24 sept 2022 [cité 15 févr 2024];14(9). Disponible sur:

https://www.cureus.com/articles/114827-sars-cov-2-mutations-responsible-for-immune-evasion-leading-to-breakthrough-infection

39. Balupuri A, Kim JM, Choi KE, No JS, Kim IH, Rhee JE, et al. Comparative Computational Analysis of Spike Protein Structural Stability in SARS-CoV-2 Omicron Subvariants. Int J Mol Sci. janv 2023;24(22):16069.

40. Shrestha LB, Foster C, Rawlinson W, Tedla N, Bull RA. Evolution of the SARS-CoV-2 omicron variants BA.1 to BA.5: Implications for immune escape and transmission. Rev Med Virol. sept 2022;32(5):e2381.

41. Tuekprakhon A, Nutalai R, Dijokaite-Guraliuc A, Zhou D, Ginn HM, Selvaraj M, et al. Antibody escape of SARS-CoV-2 Omicron BA.4 and BA.5 from vaccine and BA.1 serum. Cell. 7 juill 2022;185(14):2422-2433.e13.

42. Yu J, Collier ARY, Rowe M, Mardas F, Ventura JD, Wan H, et al. Neutralization of the SARS-CoV-2 Omicron BA.1 and BA.2 Variants. N Engl J Med. 21 avr 2022;386(16):1579-80.

43. Huang SH, Chen YT, Lin XY, Ly YY, Lien ST, Chen PH, et al. In silico prediction of immune-escaping hot spots for future COVID-19 vaccine design. Sci Rep. 18 août 2023;13(1):13468.

44. Habib MT, Rahman S, Afrad MH, Howlader AM, Khan MH, Khanam F, et al. Natural selection shapes the evolution of SARS-CoV-2 Omicron in Bangladesh. Front Genet [Internet]. 9 août 2023 [cité 31 mars 2024];14. Disponible sur: https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2023.1220906/full

45. Thakur S, Sasi S, Pillai SG, Nag A, Shukla D, Singhal R, et al. SARS-CoV-2 Mutations and Their Impact on Diagnostics, Therapeutics and Vaccines. Front Med. 22 févr 2022;9:815389.

46. Carabelli AM, Peacock TP, Thorne LG, Harvey WT, Hughes J, de Silva TI, et al. SARS-CoV-2 variant biology: immune escape, transmission and fitness. Nat Rev Microbiol. mars 2023;21(3):162-77.

47. Cao Y, Yisimayi A, Jian F, Song W, Xiao T, Wang L, et al. BA.2.12.1, BA.4 and BA.5 escape antibodies elicited by Omicron infection. Nature. août 2022;608(7923):593-602.

48. Sarkar A, Omar S, Alshareef A, Fanous K, Sarker S, Alroobi H, et al. The relative prevalence of the Omicron variant within SARS-CoV-2 infected cohorts in different countries: A systematic review. Hum Vaccines Immunother. 2 janv 2023;19(1):2212568.

49. Tamura T, Ito J, Uriu K, Zahradnik J, Kida I, Anraku Y, et al. Virological characteristics of the SARS-CoV-2 XBB variant derived from recombination of two Omicron subvariants. Nat Commun. 16 mai 2023;14(1):2800.

50. Ito J, Suzuki R, Uriu K, Itakura Y, Zahradnik J, Kimura KT, et al. Convergent evolution of SARS-CoV-2 Omicron subvariants leading to the emergence of BQ.1.1 variant. Nat Commun. 11 mai 2023;14(1):2671.

51. Makowski EK, Schardt JS, Smith MD, Tessier PM. Mutational analysis of SARS-CoV-2 variants of concern reveals key tradeoffs between receptor affinity and antibody escape. PLoS Comput Biol. 31 mai 2022;18(5):e1010160.

52. Aguilar Ticona JP, Xiao M, Li D, Nery N, Hitchings M, Belitardo EMMA, et al. Extensive transmission of SARS-CoV-2 BQ.1* variant in a population with high levels of hybrid immunity: A prevalence survey. Int J Infect Dis. 1 févr 2024;139:159-67.

53. Liu Z, Li J, Pei S, Lu Y, Li C, Zhu J, et al. An updated review of epidemiological characteristics, immune escape, and therapeutic advances of SARS-CoV-2 Omicron XBB.1.5 and other mutants. Front Cell Infect Microbiol [Internet]. 14 déc 2023 [cité 21 mai 2024];13. Disponible sur: https://www.frontiersin.org/articles/10.3389/fcimb.2023.1297078

54. Giancotti R, Lomoio U, Puccio B, Tradigo G, Vizza P, Torti C, et al. The Omicron XBB.1 Variant and Its Descendants: Genomic Mutations, Rapid Dissemination and Notable Characteristics. Biology. févr 2024;13(2):90.

55. Tracking SARS-CoV-2 variants [Internet]. [cité 8 mars 2024]. Disponible sur: https://www.who.int/activities/tracking-SARS-CoV-2-variants

56. Cobar O, Cobar S. XBB.2.3, Prevalence, Structural, Genomic, and Pathogenic Properties. Med Res Arch [Internet]. 29 juill 2023 [cité 1 avr 2024];11(7.2). Disponible sur: https://esmed.org/MRA/mra/article/view/4137

57. Wu H, Xing N, Meng K, Fu B, Xue W, Dong P, et al. Nucleocapsid mutations R203K/G204R increase the infectivity, fitness, and virulence of SARS-CoV-2. Cell Host Microbe. 8 déc 2021;29(12):1788-1801.e6.

58. Muradyan N, Arakelov V, Sargsyan A, Paronyan A, Arakelov G, Nazaryan K. Impact of mutations on the stability of SARS-CoV-2 nucleocapsid protein structure. Sci Rep. 11 mars 2024;14(1):5870.

59. Ahmad W, Ahmad S, Basha R. Analysis of the mutation dynamics of SARS-CoV-2 genome in the samples from Georgia State of the United States. Gene. 20 oct 2022;841:146774.

Response to editor

Journal: Access Microbiology

Title: Genetic diversity and genomic epidemiology of SARS-CoV-2 during the first three years of pandemic in Morocco: Comprehensive sequence analysis, including the unique B.1.528 lineage in Morocco

Article Type: Research paper

Authors' statements

First of all, we would like to thank you for the time and attention you devoted to reading and evaluating our manuscript entitled « Genetic diversity and genomic epidemiology of SARS-CoV-2 during the first three years of pandemic in Morocco: Comprehensive sequence analysis, including the unique B.1.528 lineage in Morocco ». Your detailed and constructive comments were extremely valuable to us and helped us to improve the quality of our work.

We have taken each of your comments very seriously, and have made the necessary changes to address your concerns and suggestions. You will find below our detailed answers to each of your questions and comments, as well as a description of the changes made to the manuscript.

We hope that these revisions will meet your expectations and that our revised article will bring you renewed satisfaction.

The reviewers' comments and suggestions have been incorporated directly into the manuscript. We have addressed all questions to the best of our abilities and understanding of the issues raised.

Reviewer 1:

We would like to thank you for your pertinent and constructive comments. We understand your concerns about a possible overestimation of the spread of the virus in Morocco. Please allow us to clarify and justify our conclusions with the following points:

In Morocco a Consortium of genomic surveillance laboratories for SARS-CoV-2 has been set up by the Ministry of Health to monitor the epidemiology of COVID-19 effectively. The majority of sequences submitted to GISAID come from the genomic surveillance consortium set up by the Ministry of Health, enabling the temporal evolution of COVID-19 epidemiology to be monitored.

The sequenced isolates come from different cities in Morocco, as mentioned in the results section entitled « Genomic diversity and demographic distribution of SARS-CoV-2 sequences », ensuring a diverse representation of regions and epidemiological periods in Morocco.

By comparing our results with those of countries with similar surveillance systems, we observe consistent epidemiological trends. Infection and mortality rates in Morocco, with 1,278,992 positive cases and over 16,000 deaths, are comparable to those of other nations with a similar COVID-19 surveillance strategy.

We have also taken into account the remark in the manuscript that this study is retrospective.

1- Lines 67-70. Authors make the claim that extending vaccine coverage may prevent new variants emerging. The vaccines have been shown to not prevent infection but protect against more severe clinical disease. This must be altered accordingly.

The statements in lines 67-70 have been modified according to the suggestion of Reviewer 1. The expansion of vaccine coverage helps protect patients against more severe clinical forms of the disease.

2- Line 108. Remove 'of contamination' that is unnecessary

Line 108: « contamination » has been deleted.

3- Line 123. 10 ORFs encoding 26 genes? This must be changed.

Moreover, the SARS-CoV-2 genome contains 14 open reading frames (ORFs), including two replicative polyproteins (pp1a and pp1ab) totaling approximately 13,218 bases (ORF1ab), which are cleaved into 16 non-structural proteins. In addition, the genome encodes 11 accessory proteins and 4 structural proteins (lines 115-121).

4- Lines 148-150. This is a strongly worded statement. Yes, the claim is accurate but specifically for Morocco. A global analysis of GISAID sequences would allow for better development of preventative measures. Suggestion correction to tone down this language

The recommendations relating to lines 148-150 have been incorporated (line 146-147). It should also be noted that all sequences included in this study are those of the Moroccan sequence submitted during the three years of the pandemic, as specified in supplementary material 1.

5- Lines 201-203. This is too brief for the methodology and does not cite appropriate references for the various tools of classification. EG - Pangolin.

In line 201-203 references have been added for greater clarity.

6- Line 251. Whilst I am sure the study was meticulous; this is an unnecessary word for a results section. Remove

The word meticulous has been removed.

7- Lines 317-320. Notable has been used repeatedly, swap one of these.

The amendments to lines 317-320, concerning the change in the term « notable », have been made.

8- Line 491. Elaboration is not the right word here. Development may suit better

Line 491: The word « elaboration » has been removed to improve the clarity of the sentence.

9- Lines 546-551. The claim that alpha variant seems to have increased replication capacity due to a mutation is true, but I am not sure how your analysis provides any further evidence given you have done no In vitro analysis of this particular mutation.

We have revised the manuscript to clarify this point by emphasizing that these observations are consistent with evidence from previous in vitro and observational studies. We have updated lines 546-551 to include the following clarification (lines 556-559):

The Alpha variant harbors the P681H mutation, known for its ability to increase SARS-CoV-2 replication. Previous observational studies have reported that this mutation could enhance the virus's potential for replication by increasing the processivity of spike protein cleavage, thereby increasing the challenges for pandemic control. Although our study does not include specific in vitro analyses of this mutation, our conclusions are based on robust observational and epidemiological data that corroborate these findings.

10- Lines 578 & 580. Nomenclature was changed to Greek letters to stop certain variants being associated with geographic locations. You should remove the term 'Indian variants' and should stick with using Delta and Kappa.

Following the reviewers' recommendations, we have revised the manuscript in its entirety by removing the term « Indian variants » and opting instead to use Delta and Kappa (Greek-letter variant nomenclatures) to avoid associating specific variants with geographical locations.

11- Line 636. Full stop missing at end of sentence

The missing full stop on line 636 has been added at the end of the sentence.

12- Lines 646-648. You have repeated the same thing in two sentences. Correct this. Given that this is a retrospective study based on publicly available data, it seems obvious to this reviewer that mutation profiles will have been seen elsewhere first.

The corrections have been made to lines 646-648, and the reworded text highlights that the mutation profiles described in our study correlate with those reported in previous studies.

- Figure issues in general:

I have looked at every additional figure at the end of this article and I have come to the same conclusion. They are too busy, particularly Figure 3A which is also hindered by the small size. The text on the figures is also small and hard to read thus hindering full clarity on the data presented. Suggest the authors address the figure issue.

Figure 3 is a representation generated automatically when the 2274 sequences were analyzed using the Coronapp application. Following the reviewers' suggestions, we opted to use real values at regular intervals. This figure illustrates the frequency of mutations observed in the structural proteins of SARS-CoV-2.

Figure 3 shows the distribution of synonymous and non-synonymous mutation frequencies of amino acids in structural proteins of the 2274 analyzed genomic sequences. The coordinates on the Y-axis represent the different scales of mutation frequencies observed in the structural proteins of the 2274 analyzed genomic sequences. The X-axis shows the position of each amino acid mutation along the sequences constituting the structural proteins: In the Spike protein (a) with ~1,273 amino acids and the membrane protein (d) with ~222 amino acids, the blue dots represent nonsynonymous mutations while the red dots indicate synonymous mutations. In the nucleocapsid protein (b) with ~419 amino acids and the envelope protein (c) with ~75 amino acids, the red dots show nonsynonymous mutations while the blue dots represent synonymous mutations.

Reviewer 2:

-We are grateful for your constructive comments and suggestions to improve the quality of our manuscript. We have taken your recommendations into account by strengthening the manuscript by also highlighting specific aspects of the progression of the pandemic in Morocco, compared to those observed worldwide. In the manuscript, we have clarified the key points, highlighting how the epidemiological situation in Morocco differs from the global progression. We have contextualized these differences and highlighted the unique features observed in our study. We hope that these adjustments will strengthen the relevance and robustness of our manuscript.

- In response to your comments regarding the discussion on the B.1.528 lineage, we have expanded this section to include a detailed analysis of what distinguishes the B.1.528 lineage from other related or co-occurring lineages, both phylogenetically and genetically. We have also emphasized the potential impacts of the key mutations that define this lineage, to better contextualize its importance and unique characteristics.

- We have taken into account the suggestion to modify the use of geographically named lineages. We understand and share the WHO's concerns regarding the potential stigmatization of populations associated with these regions when using such designations. Consequently, we have revised the manuscript to adopt the technical nomenclatures recommended by the WHO, using the variant nomenclature assigned in Greek letters, such as Delta and Omicron, to prevent certain variants from being associated with geographic locations.

- To avoid any ambiguity in the use of the term « significant », we ensured that appropriate terms were used throughout the manuscript. Regarding the Omicron variant and its sub-variants, we also supported our results by emphasizing their convergent and divergent evolution during the pandemic, comparing our observations with those of previous studies.

1- Presentation of results

* Figure legends for tables specifically require more detail, the role of a figure legend is to walk the reader through the figure and guide their understanding.

We understand the importance of providing detailed figure legends for tables to facilitate the understanding of the presented data. We have revised the table legends to include more details, as suggested. You will find these improvements in the revised manuscript.

* Table 1 needs a much more detailed figure legend, what does the colour mean? What are the denominators of the percentages? Etc.

In the manuscript (in the section: genomic diversity and demographic distribution of SARS-CoV-2 sequences), we have provided more details for Table 1 to guide its understanding. The table 1 illustrates the temporal distribution of SARS-CoV-2 variants and lineages in the 2274 sequences analyzed: Annual (represented by rows, showing the annual prevalence of each variant and lineage) and cumulative (represented by columns, showing the total variants and lineages identified each year and their cumulative) prevalence of identified variants and lineages from 2020 to 2023. In the tables, colors were generated automatically during the formatting of our tables to represent our results. To avoid any ambiguity regarding the colors, they have been removed to simplify the table.

- The denominators for the percentages of each variant are the totals for each variant (indicated in the last row of the respective column), and for the annual total percentages, the denominator is 2274.

-Determination of percentages in the table:

Using the total number of a lineage as the denominator for percentages provides a complete overview of the distribution of each variant across the 2274 sequences sequenced during the pandemic. This method establishes a consistent reference point to compare the relative abundance of different variants over time. In the context of longitudinal comparison, our aim was to illustrate the evolution and changes in variant prevalence throughout the pandemic. By using the total number of sequences as the denominator, we can better highlight trends and fluctuations in the dominance or decline of specific variants over several years (2020-2023). Furthermore, the cumulative approach helps understand the cumulative impact and spread of each variant over the entire pandemic period, offering a broader perspective that complements year-by-year analysis. It demonstrates the overall trend of each variant in the context of the pandemic progression, thereby contributing to understanding global epidemiological trends.

*Similar issues with table 2, it is not immediately clear what the denominator within the second column is in isolation

We would like to clarify that the denominator used for the percentages in Table 2 refers to all 2274 sequences analyzed during the pandemic. This table is designed to illustrate the distribution of variants and lineages, showing data from the most dominant to the least dominant variants and lineages over the three years of the pandemic. In addition, Table 3 shows the variants and lineages that reached a notable proportion among the 2274 sequences analyzed during the pandemic, providing a detailed view of their relative prevalence.

*The log10 scale of figure 3 is very strange, please consider using a log transformed scale with the actual numbers on it e.g. 1, 10, 100 equally spaced.

Figure 3 is a representation generated automatically when the 2274 sequences were analyzed using the Coronapp application. Following the reviewers' suggestions, we opted to use real values at regular intervals. This figure illustrates the frequency of mutations observed in the structural proteins of SARS-CoV-2. The coordinates on the Y-axis, spaced at regular intervals of real values, represent the different scales of mutation frequencies observed in the structural proteins of the 2274 analyzed genomic sequences. This facilitated the identification of the highest or lowest amino acid mutations, as well as the position of each amino acid mutation along the sequences that constituted the structural proteins (x-axis). The x-axis (X) shows the positions of amino acid mutations along the structural proteins: the Spike protein (a) with ~1273 amino acids (Blue dots: nonsynonymous mutations; red dots: synonymous mutations), nucleocapsid protein (b) with ~419 amino acids (Red dots : nonsynonymous mutations ; blue dots : synonymous mutations), envelope protein (c) with ~75 amino acids (Red dots : nonsynonymous mutations ; blue dots : synonymous mutations), and membrane protein (d) with ~222 amino acids (Blue dots: nonsynonymous mutations; red dots: synonymous mutations).

*Similar issues as above are present with figure 4, it is also worth noting that it is not a heatmap as is stated in the figure legend

In Figure 4, we have taken into account the reviewers' comments and modified the term 'heat map' to eliminate any ambiguity. Additionally, we have included detailed explanations in the manuscript to better clarify Figure 4 and provide more information about it. This figure presents a diagram detailing amino acid mutations in the structural proteins of SARS-CoV-2. Figure 4a specifically illustrates the frequency of each mutation observed among all sequences attributed to each variant or lineage, thereby determining the specific proportion of mutations present in each variant and lineage. In other words, this diagram shows the variation of mutations among different variants or lineages, indicating which mutations are present or absent between these variants and lineages.

* The b phylogeny in figure 6 is too busy to be interpretable, what are you seeking to convey with the inclusion of this figure generally?

We included phylogeny b from Figure 6 to highlight the diversity and evolutionary relationships within the Omicron sub-variants based on the 1353 sequences assigned to Omicron and its sub-variants. The objective of this phylogenetic tree is to provide a visual representation of the evolutionary links among the different sub-variants, demonstrating how they are connected through common ancestors. Additionally, it aims to graphically depict the genetic diversity within the Omicron lineage and their evolutionary relationships. We acknowledge that the density of information may make interpretation challenging, and it is important to note that the figure was generated automatically.

* Figure 7 reads far better than the above, it is clear how the a and b panels relate to another but it would be desirable for the figure to be better utilised in the text, there is only one reference to it and the text doesn't seem to rely on the figure at all.

We have ensured better integration of Figure 7 into the manuscript. It is now cited more appropriately in the Results and Discussion sections to reinforce the analyses carried out and clarify the relationship between Figures 7a and 7b, to ensure its better understanding.

2- How the style and organization of the paper communicates and represents key findings

The manuscript does not often clearly highlight what is a key finding, for example, the long sections discussing defining mutations for various well described lineages don't make it clear what is a new finding vs new evidence of previously observed and described trends.

The manuscript authors might benefit from attempting to distil the manuscript into a small number of key findings and refocusing the manuscript on just these while giving adequate regional and global context for those findings.

- We have improved the manuscript taking into account the reviewers' suggestions. We have clarified key conclusions and distinguished important mutations that have a significant impact on the diversity of SARS-CoV-2 lineages, as well as on the new conclusions based on evidence of previously observed trends. In addition, we have distilled the manuscript into a small number of key conclusions and refocused the text on these, while providing an appropriate regional and global context for these conclusions.

3-Literature analysis or discussion

Several sections of the results section are not mentioned at all within the discussion, if a result does not merit any further discussion, I question on what grounds it is included?

When discussing the impacts of specific mutations it would be desirable to see a more fleshed out discussion of the evidence for and against their potential impact on the virus. Especially in sections such as the end of the second paragraph of the discussion (lines 510-513) where there are no citations supporting the speculation about the potential function of the P1263L mutation.

- We would like to emphasize that, in this study, certain sections of the results support the ideas to explain the key concepts more clearly. These key ideas were selected for their role in developing and clarifying an appropriate conclusion. To avoid cluttering the Discussion section with excessive details, only the essential results were discussed in depth, thereby contributing to a global and coherent understanding of the main ideas presented in the results sections. In the revised version of the manuscript, we made sure to make these connections more explicit to demonstrate how each result fits into our analysis and strengthens the main ideas.

- We have taken into account the reviewers' suggestions by deepening the Discussion section to include the impact of specific mutations, presenting evidence both for and against their potential influence on the virus. This particularly includes the P1263L mutation. We have enriched this section by citing previous studies that have examined the impact of the P1263L mutation on the attenuation of infectivity in the lineages that carry this mutation (Line 531-537).