


## Article

# Genomic Characterization and Phylogenetic Analysis of SARS-CoV-2 in Libya

Silvia Fillo <sup>1,\*</sup>, Francesco Giordani <sup>1</sup>, Anella Monte <sup>1</sup>, Giovanni Faggioni <sup>1</sup>, Riccardo De Santis <sup>1</sup>, Nino D'Amore <sup>1</sup>, Stefano Palomba <sup>2</sup>, Taher Hamdani <sup>3</sup>, Kamel Taloa <sup>3</sup>, Atef Belkhir Jumaa <sup>3</sup>, Siraj Bitrou <sup>3</sup>, Ahmed Alaruusi <sup>3</sup>, Wadie Mad <sup>3</sup>, Abdulaziz Zorgani <sup>4</sup>, Omar Elahmer <sup>4</sup>, Badereddin Annajr <sup>4</sup>, Abdalla Bashein <sup>4</sup> and Florigio Lista <sup>1</sup>

- <sup>1</sup> Scientific Department, Army Medical Center, 00184 Rome, Italy; franc.giordani@gmail.com (F.G.); nellymonte88@gmail.com (A.M.); giovanni.faggioni@gmail.com (G.F.); riccardo.desantis@gmail.com (R.D.S.); nino.damore@esercito.difesa.it (N.D.); romano.lista@gmail.com (F.L.)
- <sup>2</sup> Medical Situation Awareness Branch, General Directorate of Military Medical Service, 00184 Rome, Italy; stefano.palomba@gmail.com
- <sup>3</sup> Public Health Emergency Office, National Center for Disease Control, Tripoli 71171, Libya; Taher.elhamdani10@gmail.com (T.H.); Ktaloa2014@gmail.com (K.T.); Atefbelkhir@gmail.com (A.B.J.); Sirajbitrou@yahoo.com (S.B.); ahmed.arusi88@gmail.com (A.A.); wadiemadi@yahoo.com (W.M.)
- <sup>4</sup> Public Health Reference Laboratory, National Center for Disease Control, Tripoli 71171, Libya; zorgania@yahoo.com (A.Z.); oelahmer@yahoo.co.uk (O.E.); bbannajar@gmail.com (B.A.); abashein\_95@yahoo.com (A.B.)
- \* Correspondence: silviafillo@gmail.com; Tel.: +39-06-777-039-135

**Abstract:** The COVID-19 epidemic started in Libya in March 2020 and rapidly spread. To shed some light on the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) strains circulating in Libya, viruses isolated from 10 patients in this country were sequenced, characterized at the genomic level, and compared to genomes isolated in other parts of the world. As nine genomes out of 10 belonged to the SS1 cluster and one to SS4, three datasets were built. One included only African strains and the other two contained internationally representative SS1 and SS4 genomes. Genomic analysis showed that the Libyan strains have some peculiar features in addition to those reported in other world regions. Considering the countries in which the strains are genetically more similar to the Libyan strains, SARS-CoV-2 could have entered Libya from a North African country (possibly Egypt), sub-Saharan Africa (e.g., Ghana, Mali, Nigeria), the Middle East (e.g., Saudi Arabia), or Asia (India, Bangladesh).

**Keywords:** SARS-CoV2; SNP (single-nucleotide polymorphism) analysis; molecular epidemiology; Africa; Libya



**Citation:** Fillo, S.; Giordani, F.; Monte, A.; Faggioni, G.; De Santis, R.; D'Amore, N.; Palomba, S.; Hamdani, T.; Taloa, K.; Jumaa, A.B.; et al. Genomic Characterization and Phylogenetic Analysis of SARS-CoV-2 in Libya. *Microbiol. Res.* **2021**, *12*, 138–149. <https://doi.org/10.3390/microbiolres12010010>

Received: 28 January 2021

Accepted: 26 February 2021

Published: 3 March 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

On 31 December 2019, an outbreak of atypical pneumonia cases was reported in Wuhan, in Hubei province of China, but the causative agent was not identified or confirmed until 9 February 2020 [1]. Once identified, the agent was named SARS-CoV-2 (severe acute respiratory syndrome coronavirus-2), and the human disease it causes was named COVID-19 [2,3]. By 24 March, when the first case of COVID-19 was diagnosed in Libya, the disease had already spread to 194 countries, with a total of 372,755 reported cases (<https://www.who.int> (accessed on 4 January 2021)).

The spread of COVID-19 in Libya came late. This could be attributed to several factors. First, most people live in single-family houses. Second, according to the Bureau of Statistics and Census in Libya, 68% of the population is aged 15–54 years and only 5% of Libyans are older than 65 years. The Libyan government took very early precautionary measures. In early March, the Ministry of Education adjourned all schools and universities. On 16 March, all borders, airports, and seaports were closed to avoid entry of infections from abroad. Mosques were ordered to shut down, and all social congregations were banned.

On 20 March, a curfew was imposed from 6:00 p.m. to 6:00 a.m., and only grocery stores and pharmacies were kept open during the day.

In spite of those early measures, the first case of coronavirus was reported on 24 March, a pilgrim returning from Saudi Arabia via Tunisia. The infection curve rose slowly, reaching only 75 cases by 24 May [4]. However, after that, the number of daily reported cases increased significantly, and, by 28 July, there were 3017 cases, with 579 recoveries and 67 reported deaths. Moreover, in the beginning of the infection spread, the infections were located mainly in the coastal areas and limited to the cities of Tripoli, Misurata, and Benghazi. However, in the last 2 months, it has involved more than 40 cities and towns, and, during those months, the city of Sabha in the south became the main focus of the infection nationwide.

Several published studies on the evolution of the virus, even during the short time after its appearance, have examined the genomic mutations in order to determine the path of COVID-19 diffusion in the various world regions [5–16]. In late February, the lineages characterized by specific mutations were related to particular geographical distributions and were designated as A, B, and C. Lineage A corresponds to the ancestral genome, lineage B remained predominant in China, and lineages C and A were transmitted outside China [8]. Then, on the basis of a larger number of publicly shared sequenced genomes, four distinct viral clusters exhibiting high potential to undergo global transmission were identified [16]. The four clusters were defined as super-spreader clusters 1 (SS1), 2 (SS2), 3 (SS3), and 4 (SS4). These are recognizable by the following specific signature mutations: SS1 (C8782T and T28144C), SS2 (G26144T), SS3 (G11083T), and SS4 (C241T, C3037T and A23403G). The SS clusters identified by Yang and colleagues [16] partially fit the lineages of Forster and colleagues [8]. Lineages B and C correspond to SS1 and SS2, respectively. The SS1 genomes were transmitted mainly in Asian countries (China, Vietnam, Japan, South Korea, Taiwan, and Singapore) but were also detected in North America, especially in California and Washington. SS2 was disseminated in various Asian countries, North America (United States of America (USA)), Europe, South America (Brazil), and Australia. SS3 was transmitted to several Asian countries, including Singapore and Japan, as well as Europe, USA, and Australia. SS4 was transmitted mainly to Europe, where it was responsible for the explosive increase in COVID-19 incidence in March. The first genome of this cluster was reported in Germany in late January, where the majority of SS4 strains had acquired a fourth mutation, C14408T. To explain the rapid and massive expansion of cluster SS4, mutation A23403G, which causes the D614G substitution in the S protein, was hypothesized to enhance virus fitness [17].

## 2. Materials and Methods

### 2.1. Patients, Samples, and Whole-Genome Sequencing

Fifty-seven nasopharyngeal/oropharyngeal swabs found to be PCR-positive for SARS-CoV-2 at the Libyan National Center of Disease Control were sent to the Scientific Department of the Army Medical Center in Rome, Italy (AMC) for whole-genome sequencing. Viral RNA was extracted from 125 µL of each swab sample by using the RNeasy Mini Kit (Qiagen, Hilden, Germany) and eluted in 18 µL of nuclease-free water. To confirm the positivity of the samples, RT-PCR was repeated by AMC using the Novel Coronavirus (2019-nCoV) Nucleic Acid Diagnostic Kit (Sansure Biotech Inc., Changsha, China) on an LC480 instrument (Roche Diagnostics, Mannheim, Germany). To optimize the sequencing procedures, 10 samples were selected according to the viral titer estimated from the RT-PCR cycle threshold (Ct range 16–25) and their geographical origin in Libya (Table 1).

**Table 1.** Samples analyzed in this study; <sup>a</sup> cycle threshold for N gene; <sup>b</sup> cycle threshold for ORF gene. M, male; F, female; ND, not determined.

Internal Numbering	Sample	Gender	Age	Origin	Date	Ct N Gene <sup>a</sup>	Ct ORF Gene <sup>b</sup>	GenBank Accession Number	Coverage
2068	10,230	M	54	Tripoli	13 June 2020	24	25	MW018433	10×
2084	7700	F	50	Tripoli	04 June 2020	18	20	MW018435	2738×
2093	12,156	F	35	Tripoli	19 June 2020	19	20	MW018429	4017×
2095	11,040	F	40	Tripoli	15 June 2020	19	21	MW018431	964×
2101	75	ND	ND	Sabha	ND	22	23	MW018437	143×
2103	449	ND	ND	Sabha	ND	19	20	MW018436	12×
2115	12,371	M	47	Tripoli	ND	25	25	MW018428	43×
2117	958	M	4	Tripoli	ND	21	21	MW018434	557×
2119	11,106	M	32	Kabaw	15 June 2020	26	28	MW018430	11×
2123	10,298	M	47	Tripoli	13 June 2020	20	23	MW018432	1779×

Genomic RNA was reverse-transcribed using the SuperScript III Reverse Transcriptase kit (Invitrogen, Carlsbad, CA, USA). Double-stranded DNA was subsequently synthesized by using the Klenow enzyme (Roche, Basel, Switzerland) according to the manufacturer's instructions. The Nextera XT kit was used for library preparations, and whole-genome sequencing was performed using the Illumina Miseq V3 flow cell (2 × 150 cycles) on a MiSeq sequencer following the manufacturer's instructions (Illumina, San Diego, CA, US). The reads were trimmed for quality (qscore = 20) and minimum length (100 bps) using the BBDuk trimmer integrated in Geneious Prime [18] (<http://www.geneious.com> (accessed on 4 January 2021)). High-quality reads were assembled by mapping to the reference genome Wuhan-hu-1 (GenBank accession number: NC\_045512.2) with the bowtie2 mapping algorithm also integrated in Geneious Prime. All 10 viral genomes were deposited in Genbank (Table 1).

All the data in this study were anonymized by deleting all sensitive information. According to the Italian Data Protection Code (Legislative Decree of 30 June 2003, n. 196), this made it unnecessary to obtain ethics approval (Legislative Decree of 24 June 2003, n. 211—article 6).

## 2.2. SNV Analysis

SNVs (single-nucleotide variants) of each sequenced genome were obtained by applying the Find Variation/SNPs tool, a Geneious Prime feature, to the read mappings, using a minimum variant frequency of 0.85. The resulting information was integrated with amino-acid translation (Figure 1).

A set of ad hoc scripts (available on demand), written in python language and working in pipeline, was created to identify the genomes in the GISAID (global initiative on sharing all influenza data—<https://www.gisaid.org> (accessed on 4 January 2021)) database carrying the same mutations of the new sequenced genomes. The examined part of the GISAID database consisted of 48,782 sequences comprising all those submitted until 29 July 2020 with a collection date until 30 June 2020 that passed the filters “complete”, “high coverage”, and “low coverage exclusion”. They are available for download from the GISAID site. For analysis, the sequences were divided in groups of no more than 3000 members, and each group was aligned with the same reference (Wuhan-hu-1, NC\_045512.2) using MAFFT (multiple alignment using fast Fourier transform) v7.453 with strategy FFT-NS-1 [19]. From the resulting alignments, a script determined the mutation profile of each genome, using a uniform code related to the positions of the reference sequence. A second script compared a query SNV profile with the created SNV profile database, identifying the genomes with mutations common to the query.

protein w mutation	position	ancestral allele	mutate allele	locus/gene	CDS position	Protein effect	Codon Number	Nsp3_2720/8554											Nsp12_13442/16236					Nsp14_18040/19620			Nsp15_19621/20658			
								241	361	939	1943	3037	5359	6337	8097	8327	8420	8782	11083	11575	12223	13629	13960	14218	14408	15407	18877	19086	19955	20339
		C	A	A	C	C	T	G	C	C	A	C	G	C	G	G	G	C	C	C	C	G	C	T	T	A				
		T	G	G	T	T	C	T	T	T	T	G	T	T	T	A	T	T	T	T	T	T	del	del	del					
		S' UTR																												
		96	674	1678	2772	5094	6072	7832	8062	8155	8517	10818	11310	11958	13365	13696	13964	14144	15143	18613	18822	19691	20075	20076	20077					
		None	K->R	R->C	None	None	None	T->I	L->F	M->V	None	L->F	None	None	None	V->I	D->Y	P->L	A->V	None	K->N	T->I	del	del	del					
		32	225	560	924	1698	2024	2612	2688	2719	2839	3606	3770	3986	4455	4566	4652	4715	5047	6204	6274	6564								
Sample	isolation loc.																													
10230	TRIPOLI	T	G		T			T						T				T	T	T										
958	TRIPOLI		G		G				T		T								T											
7700	TRIPOLI		G		G				T		T								T											
75	SABHA		G		G				T		T																			
449	SABHA		G		G				T		T													del	del	del				
11106	KABAW		N		T				T		T																			
12156	TRIPOLI		G		T				T		G																			
11040	TRIPOLI		G		T				T		T																			
10298	TRIPOLI		G	G	T				T		T																			
12371	TRIPOLI		G		T				T		T																			

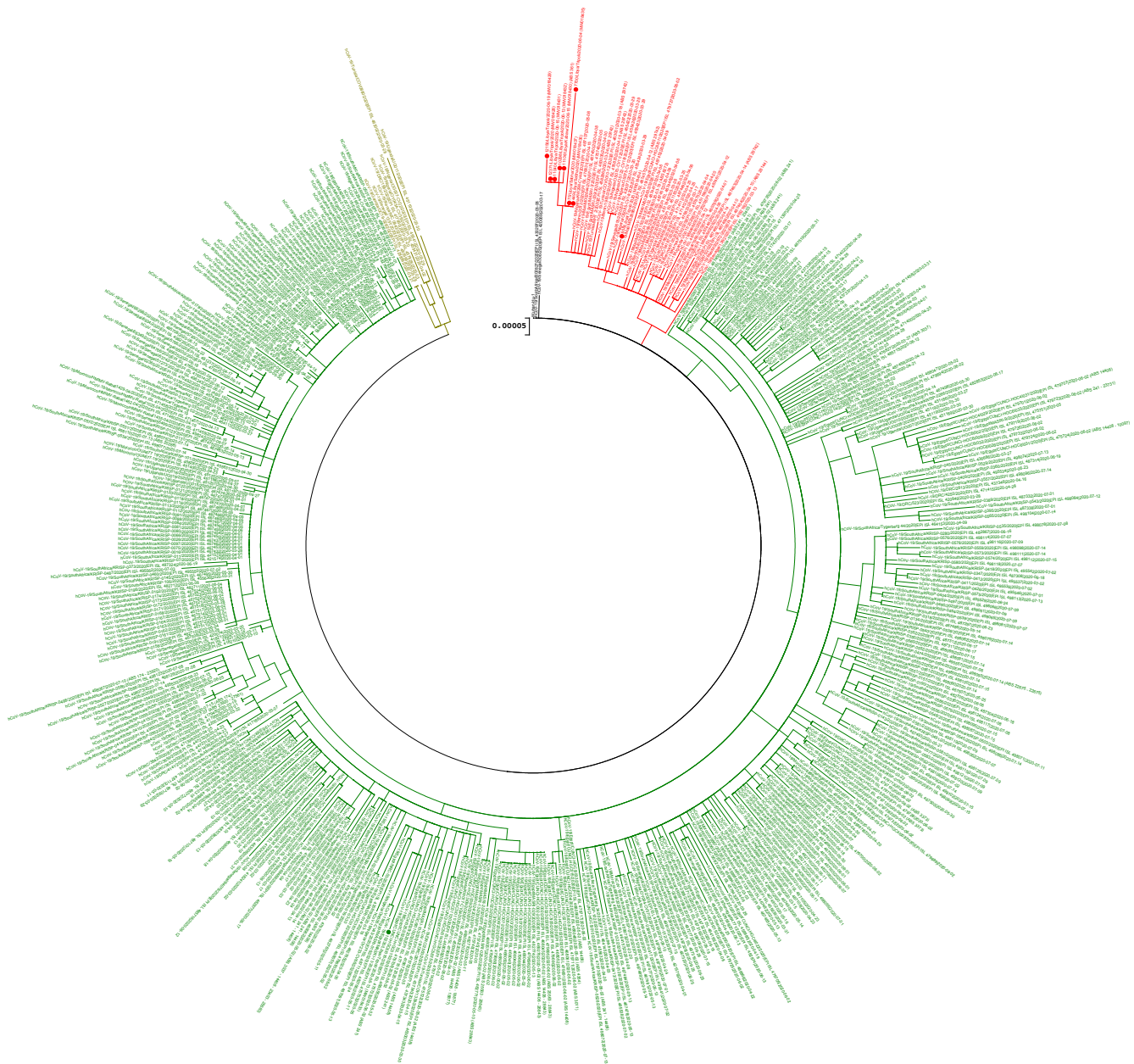
protein w mutation	position	ancestral allele	mutate allele	locus/gene	CDS position	Protein effect	Codon Number	S_21563/25384											ORF3_25393/26220				E_26245/26472		M_26523/27191		ORF6_27202/27387			ORF7_27394/27887		ORF8_27894/28259		N_28274/29533			ORF10_29559/29874	
								22444	22459	22468	23403	23539	25249	25563	25593	25641	26461	26735	27208	27382	27384	27863	28144	28854	28878	29449	29642	29742	29757									
		C	A	G	A	A	G	G	G	C	C	G	T	T	T	C	C	C	T	A	T	T	C	G	G	G	G											
		T	T	T	G	T	T	T	T	T	T	T	T	T	T	C	C	C	T	A	T	T	T	A	T	A	T											
		882	897	906	1841	1977	3687	171	201	249	217	213	7	181	183	108	251	581	605	1176	85																	
		None	None	None	D->G	None	M->I	Q->H	K->N	L->F	L->F	None	H->Y	D->Y	None	None	L->S	S->L	S->N	None	TRUNCATION																	
		294	299	302	614	659	1229	57	67	83	72	71	3	61	61	36	84	193	201	392																		
Sample	isolation loc.																																					
10230	TRIPOLI	T	T		G						N	T	T			C	C	T																				
958	TRIPOLI		T		T																																	
7700	TRIPOLI		T		T																																	
75	SABHA		T		T																																	
449	SABHA		T		T																																	
11106	KABAW		T		T																																	
12156	TRIPOLI		T		T																																	
11040	TRIPOLI		T		T																																	
10298	TRIPOLI		T		T																																	
12371	TRIPOLI		T		T																																	

**Figure 1.** Sample single-nucleotide variants (SNVs) and AA changes. Colors: blue = super-spreader 1 (SS1) signature mutations; green = recurrent mutations in SS1 genomes; yellow = SS4 signature mutations; orange = SS3 signature mutations.

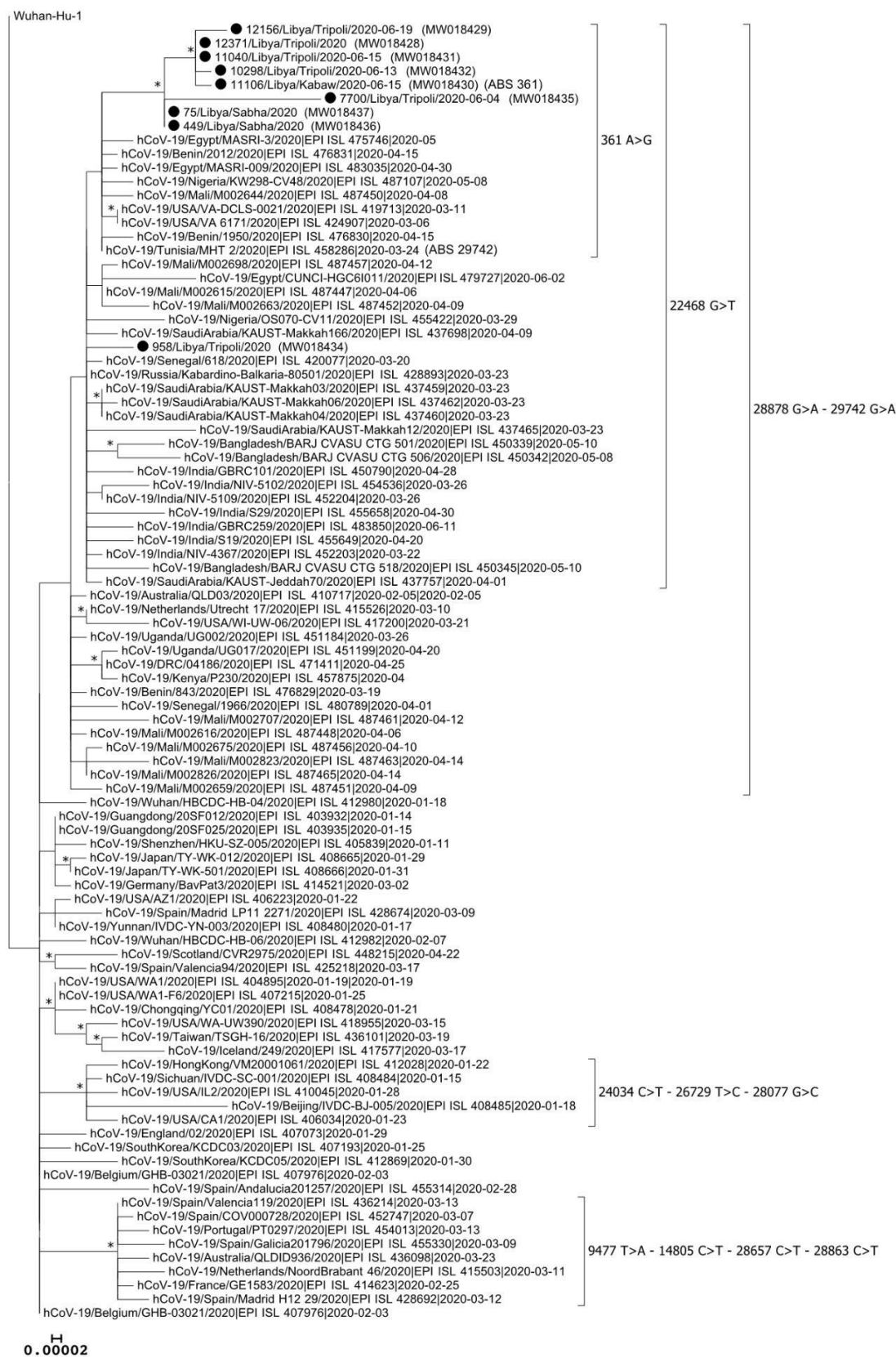
### 2.3. Phylogenetic Analysis

In silico phylogenetic analysis was performed on the newly sequenced Libyan viral genomes against three different genome datasets retrieved from GISAID until 30 June 2020. The first was related to the African high-coverage Sars-CoV-2 complete genomes, while the second and third consisted of international genomes coming from all five continents and representative of SS1 and SS4 clusters.

Raw datasets were aligned to the reference genome NC\_045512.2 by using MAFFT v7.453 with default settings [19]. Then, to obtain a higher-quality dataset, the flanking regions of the aligned sequences were trimmed to the consensus range of 54 bps to 29,783 bps according to the reference sequence. Moreover, sequences containing >0.1% of N were detected with an ad hoc script (available on demand) and removed from the dataset. The final high-quality datasets consisted of 501 sequences (first dataset), 84 sequences (second dataset), and 98 sequences (third dataset). All the phylogenetic trees were calculated with PhyML v3.0 [20] (<http://www.atgc-montpellier.fr/phyml/> (accessed on 4 January 2021)) on the basis of the maximum likelihood principle, with AIC (Akaike information criterion) as the substitution model [21,22], and with 200 bootstraps as the branch support test. The accession numbers of the Libyan genomes are reported in Table 1; those of all analyzed genomes are shown in dendrograms of Figures 2–4.



**Figure 2.** Dendrogram comparing the Libyan new sequenced genomes (marked by a dot) with a selection of representative African strains. The SNVs characterizing the main groups are shown as group names. The notation “ABS position” in brackets, adjacent to the name of a sample, means that the sample lacks the mutation in that position, albeit inserted in a group depicted as containing that mutation. Note that SS4 refers to the evolved SS4 comprising also the mutation C14408T, unless indicated otherwise. The asterisks indicate the supported clusters (bootstrap value  $\geq 60\%$ ).



**Figure 3.** Dendrogram comparing the Libyan new sequenced genomes (marked by a dot) with international SS1 strains. The SNVs characterizing the main groups are shown as group names. The notation “ABS position” in brackets, adjacent to the name of a sample, means that the sample lacks the mutation in that position, albeit inserted in a group depicted as containing that mutation. The asterisks indicate the supported clusters (bootstrap value  $\geq 60\%$ ).



**Figure 4.** Dendrogram comparing the Libyan new sequenced genomes (marked by a dot) with international SS1 strains. The SNVs characterizing the main groups are shown as group names. The notation “ABS position” in brackets, adjacent to the name of a sample, means that the sample lacks the mutation in that position, albeit inserted in a group depicted as containing that mutation. The asterisks indicate the supported clusters (bootstrap value  $\geq 60\%$ ).

### 3. Results and Discussion

#### 3.1. SNV Analysis

To investigate the genetic diversity of SARS-CoV-2 in Libya, the 10 genomes isolated from Libyan patients living in different regions were sequenced and analyzed. Seven of these were from Tripoli, one was from Kabaw in the northwest, and two were from Sabha in the south. Genome analysis was performed relative to the Wuhan-Hu-1 reference sequence (NC\_045512.2). The mutational profiles were classified following the genetic cluster scheme, suggested by Yang and colleagues, consisting of four clusters (SS1–SS4) [16]. The resulting genomic SNVs and the corresponding amino-acid positions and variations inside the proteins are shown in Figure 1. Due to their unique features, two samples (10,230 and 11,106) were included in this analysis although they had N percentage >0.1%: genome 10,230 was the only one belonging to the SS4 cluster and 11,106 was the only one isolated in Kabaw. The mutational analysis revealed 47 SNVs, of which 20 were silent mutations (synonymous), 23 were missense, three were deletions (20339delTTA20441), and one is a stop codon in ORF (open reading frame) 10. The ratio between synonymous and non-synonymous mutations was 1:1.77 across the entire set of Libyan genomes. Furthermore, 30 SNVs were found only once (63.8%), 10 were recurrent (21.3%), and seven were SS signature mutations.

During the mutational analysis, we found that nine genomes belonged to the SS1 group (signature mutations C8782T and T28144C) and one (10230, from Tripoli) belonged to SS4. This genome had the SS4 signature mutations (C241T, C3037T, A23403G) plus C14408T, as well as 11 other SNVs, of which six were missense, five produced an AA (amino acid) substitution in ORF1ab protein (K225R, T2612I, K6274N), one was in ORF3a (Q57H), one was in ORF6 (D61Y), and one was in N (S193L).

All nine SS1 Libyan strains were characterized by a haplotype consisting of the mutations G22468T, G28878A, and G29742A. Moreover, all SS1 strains except 958 also shared five other mutations; among them, A361G was the most relevant in the comparison with genomes originating in other regions, as explained below (Figure 3).

One of the SS1 genomes (7700) was different from all the others in this group. Although sharing with them 10 SNVs, it also had the SS3 mutation G11083T, a stop codon on ORF10 (K85StopCodon), and an AA mutation on E (L72F).

Moreover, the SS1 genomes also showed sporadic SNVs, some of them resulting in AA changes.

Two genomes coming from Sabha had the same SNV profile; however, in one of them (genome 75), there was a deletion of three nucleotides (20339delTTA20441). This deletion was reported in two other genomes coming from the Netherlands (EPI\_ISL\_455224) and the United States (EPI\_ISL\_485840). As these genomes did not share any other mutations with the Libyan genome 75, it seems likely that these deletions occurred independently. Mutation 20339delTTA20441 resulted in an amino-acid loss (phenylalanine) and the replacement of a serine (polar) with a cysteine (nonpolar).

We searched for mutational profiles resembling those in the Libyan genomes within a large section of the GISAID database. The search covered 48,782 sequences corresponding to those submitted until 29 July 2020, with a collection date until 30 June 2020, and passing the quality filters available in the GISAID site.

The SS1 Libyan genome haplotype (G22468T, G28878A, G29742A plus the SS1 signature mutations) was found in 80 genomes. The subgroup most similar to the SS4 Libyans (eight out of nine) was constituted by 16 genomes that also contained the mutation A361G, prevalently originating from African countries (Egypt, Benin, Mali, Ghana, Nigeria), four from Belgium (but isolated from a patient who had recently traveled to Niger: EPI\_ISL\_487433-36), and three from USA. Of all those, the genome with the earliest isolation date was EPI\_ISL\_424907 from USA (06 March 2020), followed by EPI\_ISL\_419713 from USA (11 March 2020) and EPI\_ISL\_422402 from Ghana (30 March 2020); no genome showed the basic mutational state (no mutations outside of the six characterizing the whole group), but the genomes with the lowest number of mutations (seven) were the aforementioned



three genomes with the earliest isolation date and the two genomes originating from Benin and Egypt. The remaining 66 genomes, those lacking the mutation A361G, came from India (29), Saudi Arabia (18), Bangladesh (five), Australia (one), Nigeria (four), Egypt (two), Mali (two), and Russia (two). The earliest isolated genomes were EPI\_ISL\_469243 from Saudi Arabia (15 March 2020) and EPI\_ISL\_420077 from Senegal (20 March 2020), while there were five genomes with the basic SNV: three from Saudi Arabia and two from Russia.

The most similar genome to 10,230, the only SS4 Libyan sample, originated from India (EPI\_ISL\_461484, isolated on 27 May 2020) and shared 10 mutations with 10,230: the four SS4 signature mutations plus C18877T, G19086T, C22444T, G25563T, C26735T and C28854T. A total of 200 sequences showed a similarity of nine mutations (those mentioned, except for G19086T), which originated mainly from Asian countries (India (144), Bangladesh (15), Oman (10), and Saudi Arabia (nine)) and more rarely from other countries (Ireland (nine), Australia (three), and Egypt (three)). Of these, the genomes with the earliest isolation dates were EPI\_ISL\_490003 (16 February 2020) and EPI\_ISL\_490004 (17 February 2020), both from Saudi Arabia, while there were eight genomes with the basic nine SNVs: five from India, two from Saudi Arabia, and one from the Netherlands.

### 3.2. Phylogenetic Analysis

To obtain a more comprehensive representation of the genetic diversity of the SARS-CoV-2 isolates included in this study, four phylogenetic trees were constructed. The first (Figure 2) compares the Libyan genomes to other African samples (501). The second and the third make comparisons with the SS1 (84; Figure 3) and SS4 (98; Figure 4) international strains, respectively. SS4 was the SS cluster gathering most African strains (445 out of 511; 87%), followed by SS1 (45; 9%) and SS2 (8; 1.6%) (Figure 2). Cluster SS1 was found in countries in the northern, southern, and central parts of the continent. All the African SS1 strains but one (EPI\_ISL\_418216 from Senegal) also carried mutations G28878A and G29742A. A subgroup of SS1 (26 samples) also contained G22468T in addition to the aforementioned mutations. Within this subgroup, 15 sequences also had mutation A361G, including seven Libyan sequences (however, in sample 11,106, no base could be confirmed in this position) and the other eight from Egypt, Nigeria, Mali, Benin, and Tunisia. One Libyan strain had G22468T but not A361G, while the 10 sequences with the same mutational profile originated in Nigeria (four), Senegal, Mali, and Egypt (Figure 2).

The SS4 genomes are ubiquitous in Africa. Almost all SS4 African strains analyzed here contained the C14408T SNV, which characterizes also the majority of SS4 strains in Europe [14]. Subclusters within SS4 were distinguished by the main additional mutations: the three subsequent and associated mutations G28881A, G28882A, and G28883C (156 samples), frequently present also in strains originating in other continents (America, Europe) [10], mutation G25563T (87 samples), also frequent outside Africa, especially in Europe [10], and mutations C15324T (44 samples), C16376T (38 samples), and A20268G (32 samples) (Figure 2). Libyan sample 10,230 was contained in an SS4 subcluster characterized by the previously mentioned G25563T and C18877T mutations. The other 27 members of this subcluster were prevalently from North Africa (Egypt, Morocco, Tunisia), as well as from sub-Saharan countries (Nigeria, Democratic Republic of Congo (DRC)) (Figure 2).

The SS1 international reference dendrogram (Figure 3) shows that the additional mutations found in African genomes were also disseminated in other continents. The mutations G28878A and G29742A were both present in strains isolated in Australia, the Netherlands, and USA. Strains with G22468T associated with the previous two mutations were found in Asian (India, Saudi Arabia, and Bangladesh) and European (Russia) countries.

Likewise, as the SS4 international reference dendrogram shows (Figure 4), genomes with mutations in common with Libyan sample 10,230 were disseminated outside Africa. Mutations G25563T and C18877T were present, coupled, in genomes from Europe and America (USA, Austria, Belgium, Greece, England, and Finland). Around the world, the genomes most similar to 10,230 (with shared mutations C26735T, C22444T, and C28854T, plus the previously mentioned G25563T and C18877T) originated primarily from Asian countries (India, Saudi Arabia, Oman, and Turkey), as well as from the Netherlands.

#### 4. Conclusions

On the basis of the genomic features found in this study, the following can be affirmed about SARS-CoV-2 in Libya:

1. SS1 (type B) seems to be predominant in Libya (as in East Asian countries) and likely entered before the first case was reported; genome 958 had only a few SNVs in addition to the SS1 signature mutations.
2. In Libya, the SS1 lineage evolved through human-to-human transmission and acquired mutations, but the parental strain remained within the population. The SNVs increased in number from five to 10, forming a SNV pattern typical of Libyan strains.
3. Genome 7700 had the SS3 signature mutation G11083T, which was found in only one of our patients, as well as the signature mutations of SS1 and other mutations common with Libyan SS1 genomes. Thus, it can be considered an SS1 with a subsequent G11083T mutation rather than a true SS3. It probably represents the evolution of an SS1 branch already present in Libya rather than a sign of an independent entry of SS3 into Libya.
4. A second separate entrance is suggested by the finding of an SS4 genome, 10,230.

The countries from which SARS-CoV-2 entered Libya can be inferred from the regions of the world where the strains are genetically most similar to the strains isolated in Libya. Our results indicate that the SS1 Libyan genomes carrying A361G descended from a strain that was introduced from sub-Saharan or North African countries (Egypt) (Figure 3). It could also represent the evolution of a strain lacking A361G coming from a Middle Eastern country (likely Saudi Arabia) or Asia (India) (Figure 3). Libyan strain 958, which lacks A361G, could have come from other African countries or directly from Middle Eastern regions. Two strains isolated in March 2020, originating in the USA and having four mutations in common with the SS1 Libyan strain (including A361G), suggest an alternative hypothesis. Strain 10,230 of cluster SS4 was likely introduced from Asia or the Middle East. India is the most probable origin, followed by Bangladesh, Saudi Arabia, and Oman. However, given that a few strains with a high similarity to this Libyan strain have been isolated in Egypt, it is not possible to exclude an origin from adjacent African countries (Figure 4).

In conclusion, our study demonstrated the simultaneous circulation of distinct variants of SARS-CoV-2 in Libya and local SS1 lineage evolution.

**Author Contributions:** Conceptualization, S.F., F.L., T.H., K.T., A.B.J., S.B., and W.M.; methodology, S.F.; formal analysis, A.A.; investigation, S.F., F.G., A.M., G.F., R.D.S., and N.D.; resources, O.E. and A.B.; data curation, F.G. and A.A.; writing—original draft preparation, S.F. and F.G.; writing—review and editing, A.Z.; visualization, T.H. and O.E.; supervision, K.T.; project administration, A.A., B.A., and A.B.; funding acquisition, F.L. and S.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Italian Ministry of Defense.

**Institutional Review Board Statement:** All the data in this study were anonymized by deleting all sensitive information. According to the Italian Data Protection Code (DL 196/2003), this made it unnecessary to obtain ethics approval (art. 6 DL 211/2003).

**Informed Consent Statement:** Before participating in the present study, all patients signed the informed consent statement.

**Data Availability Statement:** All the genomes sequenced in the present study were deposited in the NIH genetic sequence database (Genbank: <https://www.ncbi.nlm.nih.gov/genbank/> (accessed on 4 January 2021)) and will be publicly available once the study is published. All the other genomes analyzed in the present study are available at the genome database of GISAID (<https://www.gisaid.org> (accessed on 4 January 2021)), the global science initiative for genomic data of influenza and COVID-19 viruses.

**Acknowledgments:** The authors would like to thank the Public Health Laboratory reference staff, Tripoli, Libya, for their contributions and the COVID-19 Italian Army Study Group for their valuable contribution. The COVID-19 Italian Army Study Group includes Gaetano Alfano, Alessandra Amoroso, Anna Anselmo, Manfredo Bortone, Giandomenico Cerreto, Andrea Ciammaruconi, Angelo De Domenico, Nino D'Amore, De Santis Riccardo, Giovanni Faggioni, Silvia Fillo, Antonella Fortunato, Valeria Franchini, Francesco Giordani, Florigio Lista, Filippo Molinari, Anella Monte, Diego Munzi, Lucia Nicosia, Giancarlo Petralito, Maria Anna Spinelli, Stefano Palomba, Elisa Regarbuto, and Vanessa Vera Fain. We gratefully acknowledge the authors in the originating and submitting laboratories for their sequence and metadata shared through GISAID and NCBI.

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

## References

- Zhu, N.; Zhang, D.; Wang, W.; Li, X.; Yang, B.; Song, J.; Zhao, X.; Huang, B.; Shi, W.; Lu, R.; et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.* **2020**, *382*, 727–733. [[CrossRef](#)] [[PubMed](#)]
- Helmy, Y.A.; Fawzy, M.; Elasad, A.; Sobieh, A.; Kenney, S.P.; Shehata, A.A. The COVID-19 Pandemic: A Comprehensive Review of Taxonomy, Genetics, Epidemiology, Diagnosis, Treatment, and Control. *J. Clin. Med.* **2020**, *9*, 1225. [[CrossRef](#)]
- Sun, J.; He, W.-T.; Wang, L.; Lai, A.; Ji, X.; Zhai, X.; Li, G.; Suchard, M.A.; Tian, J.; Zhou, J.; et al. COVID-19: Epidemiology, Evolution, and Cross-Disciplinary Perspectives. *Trends Mol. Med.* **2020**, *26*, 483–495. [[CrossRef](#)]
- Alhudiri, I.M.; Ebrahim, F.O.; Sassi, A.; El Meshri, S.E.; Fellah, A.A.; Asloug, A.K.; Elmaghrabi, A.M.; Sheriff, D.; Elzagheid, A.I. Overview of Epidemiology, Pathogenesis and Clinical Aspects of COVID-19: A Libyan Perspective. *IOSR J. Dent. Med. Sci.* **2020**, *19*, 28–41.
- Bartolini, B.; Rueca, M.; Gruber, C.E.M.; Messina, F.; Carletti, F.; Giombini, E.; Lalle, E.; Bordi, L.; Matusali, G.; Colavita, F.; et al. SARS-CoV-2 Phylogenetic Analysis, Lazio Region, Italy, February–March 2020. *Emerg. Infect. Dis.* **2020**, *26*, 1842–1845. [[CrossRef](#)]
- Castillo, A.E.; Parra, B.; Tapia, P.; Acevedo, A.; Lagos, J.; Andrade, W.; Arata, L.; Leal, G.; Barra, G.; Tambley, C.; et al. Phylogenetic analysis of the first four SARS-CoV-2 cases in Chile. *J. Med. Virol.* **2020**, *92*, 1562–1566. [[CrossRef](#)]
- Deng, X.; Gu, W.; Federman, S.; Du Plessis, L.; Pybus, O.G.; Faria, N.R.; Wang, C.; Yu, G.; Bushnell, B.; Pan, C.-Y.; et al. Genomic surveillance reveals multiple introductions of SARS-CoV-2 into Northern California. *Science* **2020**, *369*, 582–587. [[CrossRef](#)]
- Forster, P.; Forster, L.; Renfrew, C.; Forster, M. Phylogenetic network analysis of SARS-CoV-2 genomes. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 9241–9243. [[CrossRef](#)] [[PubMed](#)]
- Kim, J.S.; Jang, J.H.; Kim, J.M.; Chung, Y.S.; Yoo, C.K.; Han, M.G. Genome-Wide Identification and Characterization of Point Mutations in the SARS-CoV-2 Genome. *Osong Public Health Res. Perspect.* **2020**, *11*, 101–111. [[CrossRef](#)]
- Liu, S.; Shen, J.; Fang, S.; Li, K.; Liu, J.; Yang, L.; Hu, C.-D.; Wan, J. Genetic Spectrum and Distinct Evolution Patterns of SARS-CoV-2. *Front. Microbiol.* **2020**, *11*, 593548. [[CrossRef](#)]
- Saha, I.; Ghosh, N.; Maity, D.; Sharma, N.; Sarkar, J.P.; Mitra, K. Genome-wide analysis of Indian SARS-CoV-2 genomes for the identification of genetic mutation and SNP. *Infect. Genet. Evol.* **2020**, *85*, 104457. [[CrossRef](#)]
- Stefanelli, P.; Faggioni, G.; Presti, A.L.; Fiore, S.; Marchi, A.; Benedetti, E.; Fabiani, C.; Anselmo, A.; Ciammaruconi, A.; Fortunato, A.; et al. Whole genome and phylogenetic analysis of two SARS-CoV-2 strains isolated in Italy in January and February 2020: Additional clues on multiple introductions and further circulation in Europe. *Eurosurveillance* **2020**, *25*, 2000305. [[CrossRef](#)]
- Ugurel, O.M.; Ata, O.; Turgut-Balik, D. An updated analysis of variations in SARS-CoV-2 genome. *Turk. J. Biol.* **2020**, *44*, 157–167. [[CrossRef](#)] [[PubMed](#)]
- Yin, C. Genotyping coronavirus SARS-CoV-2: Methods and implications. *Genomics* **2020**, *112*, 3588–3596. [[CrossRef](#)]
- Walker, A.; Houwaart, T.; Wienemann, T.; Vasconcelos, M.K.; Strelow, D.; Senff, T.; Hülse, L.; Adams, O.; Andree, M.; Hauka, S.; et al. Genetic structure of SARS-CoV-2 reflects clonal superspreading and multiple independent introduction events, North-Rhine Westphalia, Germany, February and March 2020. *Eurosurveillance* **2020**, *25*, 2000746. [[CrossRef](#)]
- Yang, X.; Dong, N.; Chan, E.W.-C.; Chen, S. Genetic cluster analysis of SARS-CoV-2 and the identification of those responsible for the major outbreaks in various countries. *Emerg. Microbes Infect.* **2020**, *9*, 1287–1299. [[CrossRef](#)] [[PubMed](#)]
- Korber, B.; Fischer, W.M.; Gnanakaran, S.; Yoon, H.; Theiler, J.; Abfalterer, W.; Hengartner, N.; Giorgi, E.E.; Bhattacharya, T.; Foley, B.; et al. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* **2020**, *182*, 812–827.e19. [[CrossRef](#)]
- Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649. [[CrossRef](#)]

19. Katoh, K.; Rozewicki, J.; Yamada, K.D. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Briefings Bioinform.* **2019**, *20*, 1160–1166. [[CrossRef](#)]
20. Guindon, S.; Dufayard, J.-F.; Lefort, V.; Anisimova, M.; Hordijk, W.; Gascuel, O. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst. Biol.* **2010**, *59*, 307–321. [[CrossRef](#)]
21. Lefort, V.; Longueville, J.-E.; Gascuel, O. SMS: Smart Model Selection in PhyML. *Mol. Biol. Evol.* **2017**, *34*, 2422–2424. [[CrossRef](#)] [[PubMed](#)]
22. Posada, D.; Buckley, T.R. Model Selection and Model Averaging in Phylogenetics: Advantages of Akaike Information Criterion and Bayesian Approaches Over Likelihood Ratio Tests. *Syst. Biol.* **2004**, *53*, 793–808. [[CrossRef](#)]