



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



## Tracking of SARS-CoV-2 Alpha variant (B.1.1.7) in Palestine

Abdelmajeed Nasereddin<sup>a</sup>, Amer Al-Jawabreh<sup>b,c,\*</sup>, Kamal Dumaidi<sup>c</sup>, Ahmed Al-Jawabreh<sup>a,b</sup>, Hanan Al-Jawabreh<sup>b</sup>, Suheir Erekat<sup>a</sup>

<sup>a</sup> Biochemistry and Molecular Biology Department, Faculty of Medicine, Al-Quds University, East Jerusalem, Palestine

<sup>b</sup> Leishmaniasis Research Unit, Jericho, Palestine

<sup>c</sup> Arab American University, Jenin, Palestine

### ARTICLE INFO

#### Keywords:

SARS-CoV-2  
 COVID-19  
 Whole genome sequence  
 Phylogenetic tree  
 Genetic variation  
 Palestine

### ABSTRACT

As surges of the COVID-19 pandemic continue globally, including in Palestine, several new SARS-CoV-2 variants have been introduced. This expansion has impacted transmission, disease severity, virulence, diagnosis, therapy, and natural and vaccine-induced immunity. Here, 183 whole genome sequences (WGS) were analyzed, of which 129 were from Palestinian cases, 62 of which were collected in 11 Palestinian districts between October 2020 and April 2021 and sequenced completely. A dramatic shift from the wild type to the Alpha variant (B.1.1.7) was observed within a short period of time. Cluster mapping revealed statistically significant clades in two main Palestinian cities, Al-Khalil (Monte Carlo hypothesis test-Poisson model,  $P = 0.00000000012$ ) and Nablus (Monte Carlo hypothesis test-Poisson model,  $P = 0.014$  and  $0.015$ ). The phylogenetic tree showed three main clusters of SARS-CoV-2 with high bootstrap values ( $>90$ ). However, population genetics analysis showed a genetically homogenous population supported by low Wright's F-statistic values ( $F_{st} < 0.25$ ), high gene flow ( $N_m > 3$ ), and statistically insignificant Tajima's D values (Tajima's test, neutrality model prediction,  $P = 0.02$ ). The Alpha variant, rapidly replaced the wild type, causing a major surge that peaked in April 2021, with an increased COVID-19 mortality rate, especially, in the Al-Khalil and Nablus districts. The source of introduction remains uncertain, despite the minimal genetic variation. The study substantiates the use of WGS for SARS-CoV-2 surveillance as an early warning system to track down new variants requiring effective control.

### 1. Introduction

The recently introduced serious respiratory illness termed COVID-19 was caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). It is an enveloped, positive sense, single-stranded RNA virus, belonging to the family *Coronaviridae*, genus *Betacoronavirus*. In December 2019, SARS-CoV-2 was firstly identified and reported in Wuhan, the capital city of Hubei province, China (Wang et al., 2020a). It has rapidly spread across the world (Vassallo et al., 2021) with a sharp increase in the number of cases (Velavan and Meyer, 2020). On January 30, 2020, the World Health Organization (WHO) declared SARS-CoV-2 an outbreak and public health emergency of international concern. On March 11, 2020, COVID-19 was declared a pandemic. More than 236 million cases had been reported with more than 4.8 million deaths up to October 6, 2021 (Worldmeter, 2021). In Palestine, including the West Bank, Gaza strip and East Jerusalem, the first cases were reported on

March 4, 2020. The number has grown since then, reaching more than 439 thousands laboratory-confirmed SARS-COV-2 cases and almost 4400 deaths up to October 6, 2021 (<http://site.moh.ps/index/covid19/LanguageVersion/1/Language/ar>).

At the beginning of the pandemic, SARS-CoV-2 was classified into two lineages; A and B. Both of these lineages originated in China. Later on, lineage A spread to Asia and then to the rest of the world whereas lineage B spread to Europe (Rambaut et al., 2020). However, although SARS-CoV-2 appears to have a relatively stable genome, its spilling into humans and the failure to contain its spread in many countries, several lineages and genetic diversities have been reported (Eden et al., 2020). In April 2020, Korber et al. (2020) reported a virus with a high frequency containing a spike D614G mutation, which became the dominant type of the pandemic. In addition, they showed that the G614 variant is associated with greater infectivity with clinical evidence associating it with higher viral loads (Korber et al., 2020). Later, a

Abbreviations: SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; COVID-19, coronavirus disease of 2019; WGS, whole genome sequences.

\* Corresponding author at: Jericho, Al-Quds st., The West Bank, Palestine.

E-mail address: [amer.aljawabreh@aaup.edu](mailto:amer.aljawabreh@aaup.edu) (A. Al-Jawabreh).

<https://doi.org/10.1016/j.meegid.2022.105279>

Received 17 December 2021; Received in revised form 15 March 2022; Accepted 29 March 2022

Available online 4 April 2022

1567-1348/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

N439K mutation in the spike receptor-binding domain (RBD) that was reported in some European countries and the United States. The N439K mutation was reported to confer resistance against several neutralizing monoclonal antibodies and reduced the activity of some polyclonal sera from recovered COVID-19 persons (Thomson et al., 2021). Furthermore, the new variant SARS-CoV-2 lineage B.1.1.7, recently called Alpha variant (Konings et al., 2021), which was first reported in England, had both the N501Y and the 69–70 deletion mutations in the spike region of the viral genome. This variant showed an increased viral transmissibility rate above that of the previous circulating variants but with no evidence of increased clinical severity or vaccine escape capability (Tang et al., 2021).

Later, Tegally et al. (2021) described a new SARS-CoV-2 variant (B.1.351-501Y.V2), which was firstly seen in South Africa (Beta variant) (Konings et al., 2021), and had eight mutations in the spike protein, three of which were located at the RBD (K417N, E484K and N501Y), which may have functional significance. The site of two of the RBD mutations, K417N and E484K, are considered the key for binding neutralizing antibodies. The primary model of this variant (501Y.V2) showed 50% extra transmissibility than previous variants (Tegally et al., 2021). Other recent variants of concern are the gamma variants, B.1.1.28–484 K.V2 and B.1.1.248, which have in addition to the E484K mutation several other unique mutations in the spike glycoprotein. These two variants may enhance transmissibility similarly to both Alpha and beta variants as they share a similar pattern of mutations (Toovey et al., 2021). The Delta variant, first reported in India, carried two unique mutations, L452R and T478K, on the spike region. All the aforementioned variants became a public health concern and, therefore, defined as variants of concern (VOC). These variants showed significant effects on disease distribution (surges of the pandemic), severity and transmissibility as evidenced by increased numbers of hospitalizations, case fatality rate, reduction of neutralizing antibody levels post-vaccination and reduced susceptibility to drugs (CDC.gov, 2021). In addition, the VOC list was extended by the variants of interest (VOI) which emerged and caused significant increase in community transmission in several countries and found sporadically in others, i. e., as Eta (B. 1.525), Iota (B. 1.526), Kappa (1. 617.1) and Lambda (C.37) (Canton et al., 2021). Qutob et al. studied the genomic epidemiology of the first surge of the SARS-CoV-2 pandemic in Palestine, showing the genomes with unassigned variants (wild type) (Qutob et al., 2021). Such findings highlight the importance of active genomic surveillance of the SARS-CoV-2. The aim of this study was to describe the genetic structure of the whole genome sequence of the SARS-CoV-2 circulating in the Palestinian West Bank, and expose viral mutants and characterize their genomic variation, and their genetic relationship to the reference and selected global strains.

## 2. Methods

### 2.1. Sample collection

In a descriptive cross-sectional study, nasopharyngeal swabs (Poctman, Guandong Poctman Life Technology Co., Ltd., China) were conveniently collected from a pool of 62 Palestinian patients after being officially confirmed as COVID-19 cases with varying degrees of classical symptoms. All samples were transported at 4 °C in a sealed container for RNA extraction and processing. Samples were collected between October 2020 and April 2021, during the second surge of the COVID-19 pandemic.

### 2.2. Ethical consideration

The study received ethical approval from the Research Ethics Committee of Al-Quds University under the reference number 154/REC/2020. Verbal informed consent was taken from the patients or their guardians if less than 18 years old.

### 2.3. RNA extraction and DNA flex library preparation

Viral RNA was extracted using a QIAamp® Viral RNA Extraction Kit (Qiagen Diagnostics GmbH, Germany) according to manufacturer's instructions. DNA library preparation and the sequencing method were based on ARTIC v3 amplicon and on IDT ARTIC nCoV-2019 V3 Panel (ARTIC Network, 2019). LunaScript RT SuperMix (New England Biolabs, Ipswich, MA, USA) was used for cDNA synthesis. ARTIC v3 Amplicons specific amplification was done, using two separate ARTIC v3 primer pools. PCR products (Pools 1 and 2) were mixed together, purified with 1.3× AMPureXP beads (Beckman Coulter, Brea, CA, USA), and quantified by Qubit Fluorometer DNA assay (Thermo Fisher Scientific, Waltham, MA, USA) and TapeStation capillary electrophoresis (Agilent, Santa Clara, CA, USA). DNA libraries were prepared, using Illumina Nextera DNA Flex library kit according to the manufacturer's recommended protocol. Pooled DNA libraries were sequenced with a NextSeq 500/550 Mid Output Kit v2.5 (150 Cycles) on NextSeq 500 machine (Illumina, San Diego, CA, USA).

### 2.4. Viral whole genome assembly

Binary Base Call (BCL) files as output from the Nextseq 500 sequencing machine were converted to FASTQ format using BCL to Fastq command line (bcl2fastq v2.20.0.422 Copyright (c) 2007–2017 Illumina, Inc.). FASTQ files were analyzed using the galaxy program (Galaxy Version 0.7.17.1) (<https://usegalaxy.eu/>) (Afgan et al., 2018). SARS-CoV-2 consensus sequences were obtained by mapping reads with BWA-MEM - map medium and long reads (> 100 bp) against the Wuhan-Hu-1 SARS-CoV-2 reference isolate (hCoV-19/Wuhan/Hu-1/2019 and GenBank accession number NC\_045512.0, GISAID accession ID: EPI\_ISL\_402125) (Gohl et al., 2020; Quick et al., 2017). The mapping was displayed with a local Integrative Genomics Viewer (IGV) and the consensus sequence from the IGV was copied and accepted as the sequence of each sample.

### 2.5. Emerging, PANGO lineages and GISAID clades identification

Three free online program platforms were used for the analysis of emerging, PANGO lineages, and GISAID clades; The first, being the GISAID CoVsurver (<https://www.gisaid.org/epiflu-applications/covsurver-mutations-app/>) shows phenotypic or epidemiological amino acid changes compared to hCoV-19/Wuhan/WIV04/2019. The second, the Genome Detective Coronavirus typing tool version 1.132- (<https://www.genomedetective.com/app/typingtool/cov/>), to identifies coronavirus types, genotypes and lineages of nucleotide sequence, using the basic local alignment search tool (BLAST) and phylogenetic methods and used for this purpose on the Palestinians isolates.

Lastly; Nextstrain, a website for real-time tracking of pathogen evolution, was used to identify the Palestinian SARS-CoV-2 isolates; ([https://nextstrain.org/ncov/asia?f\\_country=Palestine](https://nextstrain.org/ncov/asia?f_country=Palestine)) (Nextstrain, 2021).

### 2.6. Population genetics analyses

Population genetic analyses included phylogenetic tree construction and genetic diversity. The GISAID-deposited and GenBank-deposited study genomes along with randomly-retrieved genomes from the GenBank and GISAID were aligned, using the Multiple Alignment Program for amino acid or nucleotide sequences (MAFFT - Multiple Alignment using Fast Fourier Transform) (Galaxy Version 7.221.3, <https://usegalaxy.org/>). The fasta output from the galaxy program was uploaded to the MEGA version X program (Kumar et al., 2018). The phylogenetic tree was inferred, using the UPGMA method (Felsenstein, 1985; Kumar et al., 2018; Saitou and Nei, 1987; Tamura et al., 2004).

The randomly-retrieved complete SARS-CoV-2 genomes are evenly distributed all over the continents. Maximum likelihood (ML) and neighbor-joining (NJ) phylogenetic trees of the complete genomes with

1000 iterations for bootstrapping were constructed, using MEGA version X. The analysis included the Palestinian study genomes, retrieved genomes, Wuhan reference strain, and an out-group genome of the bat Coronavirus from China (GenBank: MN996532.2).

Population genetic parameters included mean genetic diversity (Hd) number of haplotypes (h), nucleotide diversity per site ( $\pi$ ) (Saitou and Nei, 1987), total number of mutations (Eta), average number of nucleotide differences (k) (Tajima, 1983), and number of variable/segregating sites. Eta and neutrality estimators of mutation rate included Tajima's D, Fu Li's D, and Fu Li's F tests. Genetic differentiation parameters were calculated, using DnaSP ver. 6.12.03 (Rozas et al., 2017). These estimators included Wright's F-statistics (Fst) as pairwise genetic distance (Wright, 1951), number of migrant (Nm), gene flow and population migration among populations,  $Nm = (1-Fst)/2Fst$  haploid,  $Nm = (1-Fst)/4Fst$  diploid (Hudson et al., 1992; Wright, 1951). The average

number of nucleotide differences between populations 1 and 2 (Kxy), the average number of nucleotide substitutions per site between populations 1 and 2 (Dxy) (Saitou and Nei, 1987), the number of net nucleotide substitutions per site between populations 1 and 2 (Da) (Saitou and Nei, 1987). The genetic differentiation index is based on the frequency of haplotypes (Gst) (Nei, 1973). Also, Hudson-Kreitman-Aguadé,  $HKA(X^2)$ , a neutrality test to assess fitness of data to neutral evolution that assumes the same polymorphism and divergence, was estimated.

2.7. Cluster mapping

Cluster mapping of SARS-CoV-2 cases from Palestine with statistical inference was conducted, using two software packages. The first was SaTScan™ v9.7 for the spatial scan statistics of case clustering, while the

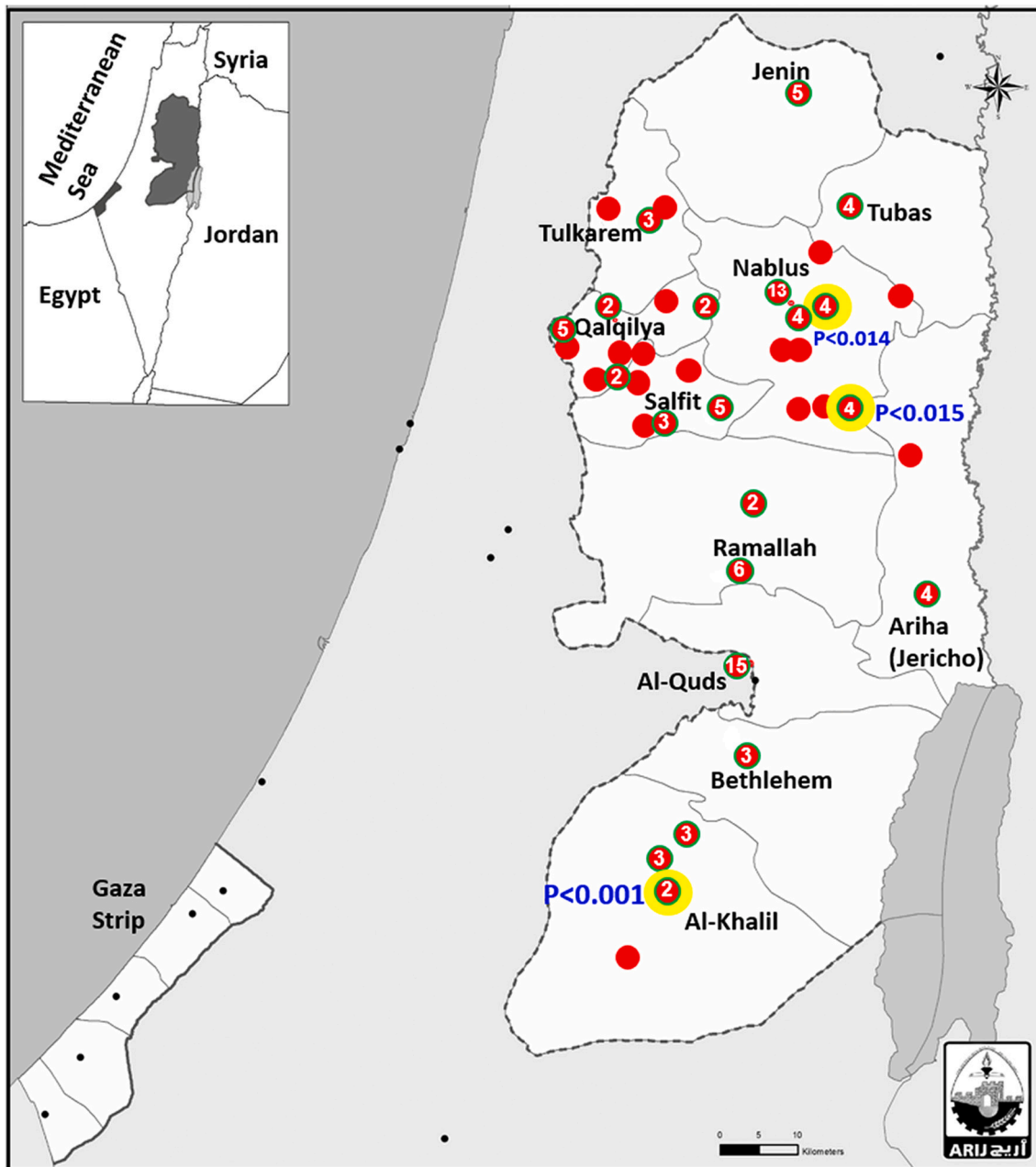


Fig. 1. Cluster mapping of COVID-19 in the study area, using SaTScan software: red circles indicate the clusters. Number inside the red circles indicates the number of COVID-19 cases in that locality. Red circles inside yellow ones are the statistically significant clusters. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

second was Epi Info™ 7 statistical package (CDC free-software) for accurate mapping. SaTScan™ v9.7 Freeware was used to detect statistical evidence for purely spatial clustering of COVID-19 cases in Palestine. The input files included the number of cases per locality, year of infection, population size of location in the year of infection and the exact latitude-longitude coordinates of each location. Data were analyzed based on the discrete Poisson model, using Monte Carlo hypothesis testing with the level of statistical significance considered at  $P$ -value  $\leq 0.05$  (Kulldorff et al., 2005).

### 3. Results

#### 3.1. Study population and cluster mapping

The study included 62 samples from Palestinian symptomatic and asymptomatic COVID-19 patients from the 11 Palestinian districts in the West Bank (Fig. 1), excluding East Jerusalem. In addition, whole genome sequences (WGS) of SARS-CoV-2 from 67 Palestinian cases were retrieved from the GISAID websites and were included in the study for comparison. The study samples were collected during three different periods of time that correlated with pandemic surges (October–November 2020, January–February 2021, and April 2021). The median age of the subjects studied ( $n = 62$ ) was 34 years (1–78) and a male: female ratio of 1:1. Altogether there were 129 patients in the study group and GISAID-retrieved group, which came from the 11 districts: Jericho (5), Al-Quds (Jerusalem) (17), Bethlehem (11), Al-Khalil (9) Salfit (11), Tubas (4), Nablus (31), Jenin (6), Tulkarem (5), Ramallah (16) and Qalqilia (14) (Fig. 1). The patients showed mild to severe COVID-19 symptoms. The 62-genome sequences, sequenced in this study-, were deposited in the GenBank and GISAID. Of the 129 patients, 113 were from 37 geographical localities spread over the 11 districts. With a known number of COVID-19 patients and population of each locality in 2020, purely spatial cluster analysis, using SaTScan, showed the villages

Majdal Bani Fadil and Deir-al-Hatab in the northern district of Nablus and Al-Khalil city in the southern district, the most populous districts, showed significant clusters (Monte Carlo hypothesis test-Poisson model  $P = 0.014$ ,  $P = 0.15$ ,  $P = 0.0000000012$ , respectively).

#### 3.2. Identification of the emerging lineages, GISAID clades and PANGO lineages

The most frequent Pango lineage in the study area during the second surge of the pandemic was B.1.1.50 (55%), followed by the B.1.1.7 Alpha variant (27%). In parallel, the GISAID clade GR was the dominant one (125/145, 86%).

This study showed that since the start of WGS SARS-CoV-2 sequencing in Palestine as of October 2020, the main predominantly known lineages were B.1.1.50 and 20B. For six months, these two lineages continued to predominate until mid-February 2021 when new lineages started to appear at the expense of others. The newly-emerging lineages were B.1.1.7, B.1.1.7 (Alpha), and 20I (Alpha, V1) (Fig. 2).

#### 3.3. Phylogenetic analysis

A total of 183 SARS-CoV-2 WGS were used to construct a consensus phylogenetic tree (Fig. 3) of which, 129 were Palestinian sequences, 62 from this study and 67 from the Palestinian Ministry of Health (Qutob et al., 2021), and 52 international GISAID-retrieved strains from various countries in all continents, including neighborhood sequences from Jordan and Egypt. All positions containing gaps and missing data were eliminated, using the complete deletion option. There were a total of 23,926 positions in the final dataset (Kumar et al., 2018).

The maximum likelihood phylogenetic tree (Fig. 3) showed that the 183 complete genome sequences grouped into three distinct genetic clusters. The lateral oval-shaped radiation form of the tree when placed as layer beneath the circular form produced a similar matching pattern.

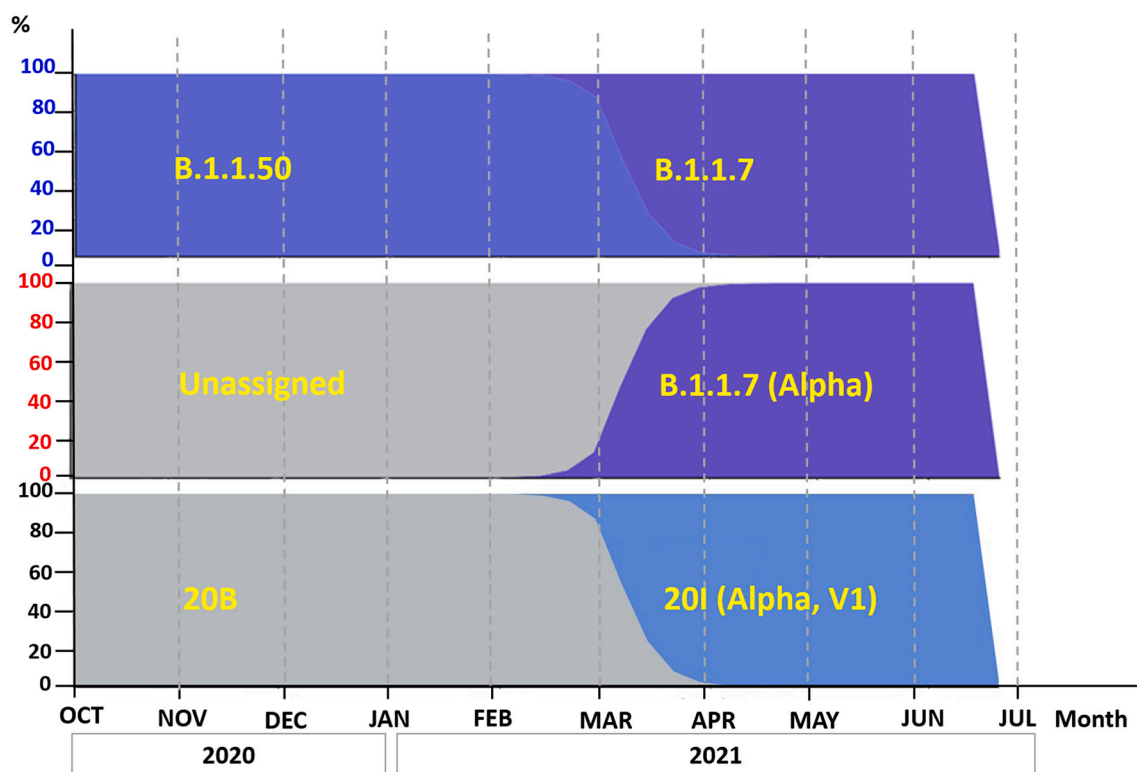
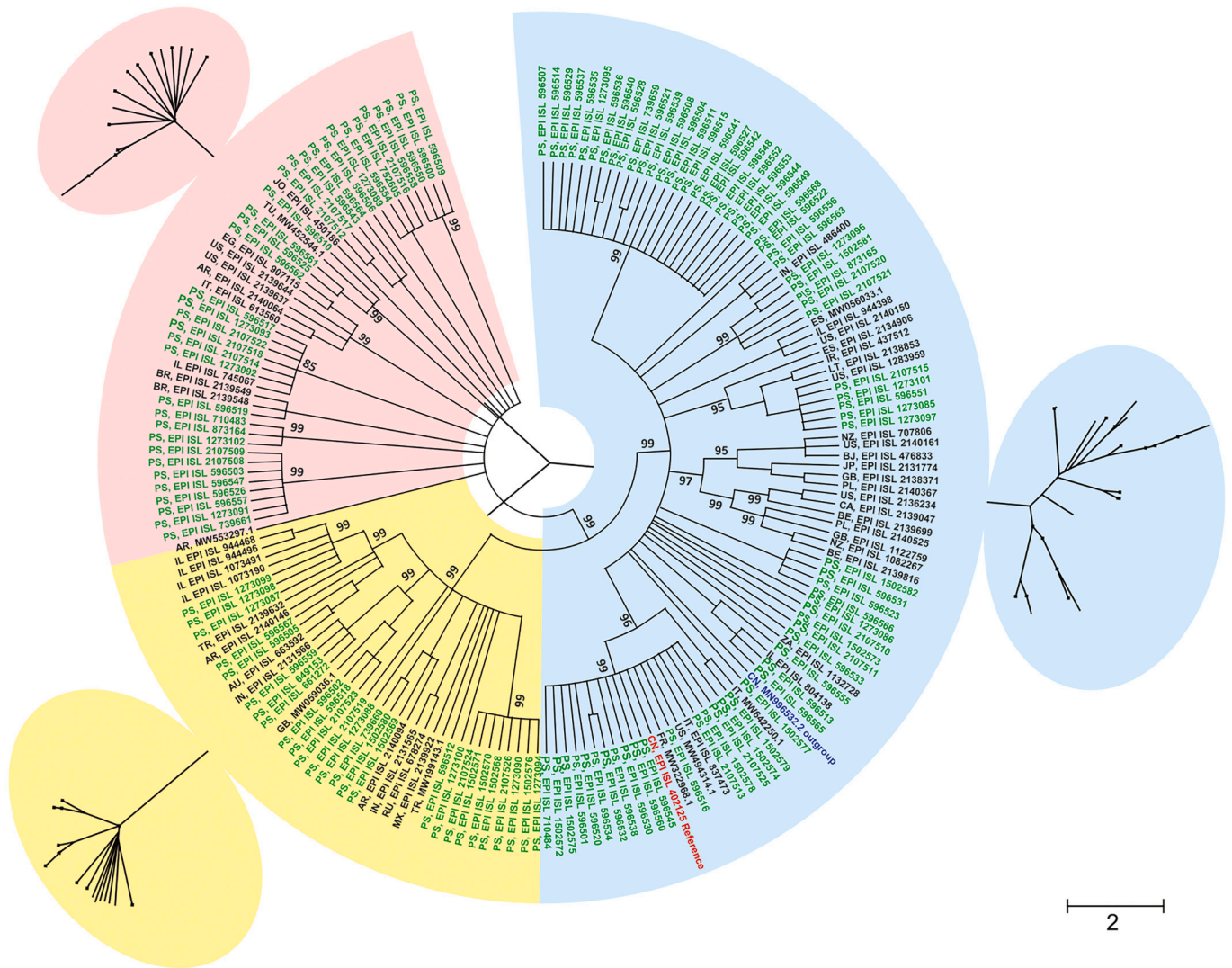


Fig. 2. Time-line of the frequencies of the emerging lineages of SARS-CoV-2 in Palestine starting from October 2020 until June 2021. Colors represent the emerging lineages with normalization to 100% at each time point for nine out of a total of 3654 tips. The graph is compiled based on data and graphs from Nextstrain online platform ([https://nextstrain.org/ncov/gisaid/asia?f\\_country=Palestine](https://nextstrain.org/ncov/gisaid/asia?f_country=Palestine)).





**Fig. 3.** Combined circular and radiation consensus Maximum-likelihood (ML) phylogenetic trees with a bootstrap value of 100 replicates based on SARS-CoV-2 WGS deposited in the GenBank and GSIAD, using Tamura-Nei model (Saitou and Nei, 1987; Tamura et al., 2004). The bootstrap values with a percentage above 80 are shown next to the branches (Felsenstein, 1985). Branches less than that are collapsed. The codes starting with ‘EPI’ indicate Accession ID from GSIAD while the others are Accession numbers from the GenBank. Green Accession IDs are Palestinian sequences, black accession IDs are international sequences, red accession ID is the Wuhan reference, and the blue accession number is the bat SARS-CoV-2 outgroup (*Rhinolophus affinis*). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The congruent trees supported the existence of the three clusters. Cluster-I (Blue) formed the largest group with 96 sequences. More than half (53%) of the Palestinian sequences grouped in cluster I along with the Wuhan reference strain and the bat SARS-CoV-2 sequence was used as an outgroup. Cluster II (Yellow) and III (red) contained 26 and 35 Palestinian genome sequences, respectively. The genome sequences did not follow any geographical clustering pattern.

**3.4. Genetic variation of SARS-CoV-2**

In the second SARS-CoV-2 surge of the pandemic in Palestine, population diversity indices and neutrality tests were calculated for the 182 complete SARS-CoV-2 genomes based on the clustering shown in the maximum likelihood phylogenetic tree (Tables 1 and 2). Genome sequences were distributed throughout the three clusters. Cluster-I ( $n = 96$  genomes, 53%) had the highest number of haplotypes ( $h = 90$ )

**Table 1**  
Genetic variation parameters and neutrality indices of the three clusters of the studied SARS-CoV-2 genomes.

Cluster	Haplotype- nucleotide diversity							Neutrality tests		
	n	H	Eta	Hd ± SD	$\pi \pm SD$	K	S	Tajima’s D	Fu-Li’s F	Fu-Li’s D
Cluster-I	95	90	77,089	0.99 ± 0.002	0.621 ± 0.01	16,649.04	26,773	0.37	1.62	2.13*
Cluster-II	41	41	78,710	1.00 ± 0.005	0.652 ± 0.01	17,858.83	27,358	-0.11	1.07	1.40
Cluster-III	46	45	77,167	0.99 ± 0.005	0.661 ± 0.01	18,138.93	27,472	0.12	1.37*	1.77*
<b>Total</b>	<b>182</b>	<b>176</b>	<b>71,779</b>	<b>1.00 ± 0.001</b>	<b>0.699 ± 0.005</b>	<b>16,734.50</b>	<b>23,927</b>	<b>1.13</b>	<b>2.59*</b>	<b>na</b>

n: Number of sequences, h: Number of Haplotypes, Hd: Haplotype (gene) diversity,  $\pi$ : Nucleotide diversity (per site), K: Average number of nucleotide differences between two randomly chosen sequences from within in the population, S: Number of variable/segregating sites. Eta: Total number of mutations. \*:  $P < 0.05$ .

**Table 2**  
Population genetic differentiation and gene flow indices between the three SARS-CoV-2 probable clusters.

Pop 1	Pop 2	Fst	Nm	Kxy	Dxy	Gst	Da	HKA(X <sup>2</sup> )
Cluster-I	Cluster-II	0.14	3.07	17,723.5	0.74	0.0016	0.10	0.000
Cluster-I	Cluster-III	0.13	3.35	17,779.7	0.74	0.0014	0.10	0.000
Cluster-II	Cluster-III	0.12	3.67	17,774.5	0.74	0.0003	0.09	0.000

compared to clusters II and III. However the ratio of the number of genomes to the number of haplotypes (h:n) remained constant across the population ( $\approx 1$ ). The number of mutations (Eta) was almost identical in the three populations. The nucleotide diversity ( $\pi$ ) in the three clusters was low (Average of 0.664 and total of  $0.669 \pm 0.005$ ), while haplotype diversity (Hd) was high (Total of  $1.00 \pm 0.001$ ) (Table 1). The DnaSP ver. 6.12.03 estimated the average number of nucleotide differences between any two sequences (k) and the number of segregating (polymorphic) sites (S) were the lowest in cluster I, 16649.04 and 26,773, respectively. Cluster I contained the Wuhan reference genome (hCoV-19/Wuhan/Hu-1/2019). The SARS-CoV-2 study genome sequences ( $n = 129$ ) were distributed in the three clusters. Tajima's D did not show any significant departure from neutrality (Tajima's test, neutrality model prediction,  $P = 0.02$ ) in any of the three clusters. Fu-Li's F and Fu-Li's D tests showed statistically significant differences from neutral expectations of cluster-III, while cluster-I was only significant by the Fu-Li's D test.

The Wright's F-statistics pairwise genetic distance (Fst) between any two of the three SARS-CoV-2 clusters was small ( $< 0.25$ ). However, Fst values depend not only on the amount of differentiation among populations but also on diversity, which is largely affected by effective population size and mutation rate with larger populations decreasing Fst values. At the same time the migration number (Nm) estimates, gene flow between population, were high and exceeded 3 (Table 2). However, other inter-population genetics indices, including the average number of nucleotide differences between clusters (Kxy), the average number of nucleotide substitutions per site between clusters (Dxy), the number of net nucleotide substitutions per site between clusters (Da), HKA-X2, and genetic differentiation index based on the frequency of haplotypes (Gst) were low as shown in Table 2.

#### 4. Discussion

This study almost duplicated the number of available SARS-CoV-2 genomes for Palestine in the GISAID and GenBank databases, providing insights into the dynamics of the pandemic and the real time conversion from unassigned emerging lineages to the variant of concern's emerging lineage (Alpha, V1).

The Alpha variant was first identified in Palestine late in January 2021 with a frequency of 1% of the positively-tested samples. In mid-February 2021, the Alpha variant frequency reached 97% of the tested positive samples, replacing the wild type with a single lineage (Alpha V1) in about three weeks' time. These results are in agreement with results from neighboring areas reporting dominance of the Alpha variant in February 2021 (Mor et al., 2021). This explains the rapid spread of the COVID-19 in Palestine during that outbreak. No Alpha variants were reported in the previous Palestinian study (Qutob et al., 2021), indicating the need for continuous surveillance by WGS sequencing and phylogenetic analysis. These key strategies appear to be the most effective means of studying and tracking circulating viral lineages, identifying their transmission pathways, and screening for the introduction and spread of new variants of concern, especially of the highly virulent ones like the Alpha variant. A study conducted by Vassallo et al. (2021) showed that patients infected with the Alpha variant had 3.8-fold higher risk of death or transfer to the intensive care unit (ICU) compared to those who became infected with the original wild strain (Vassallo et al., 2021). Further studies should be done to see if this variant could still be a threat with high risk of severe COVID-19 occurring in Palestine

and globally.

On March 2021, the country went into a nationwide lockdown owing to the Alpha variant. The Palestinian health authorities confirmed 27 deaths and 2884 new infections from COVID-19 in one day and with a peak of 34 deaths on 22nd of April 2021. On the 15th April 2021, there were 33,275 active cases (PalestineMinistryofHealth(PMOH), 2021; Worldmeter, 2021). These high numbers were probably owing to the Alpha variant (Fig. 2). The Palestinian Ministry of Health received its first shipment of COVID-19 vaccines on March 2021, and a shortage of doses and vaccination hesitancy left the Palestine public vulnerable to the Alpha variant. The time distribution of the Alpha variant in Palestine was different from that in the Israeli community. In the later, 90% of the Alpha variants were during December–February, while in Palestine Alpha variant started to appear in March and dominated the scene in April to July (Fig. 2) (Munitz et al., 2021). However, the genomes of the Alpha variants in both close-by communities did not show any pattern in the phylogenetic tree (Fig. 3), but rather randomly distributed across the three cluster.

Of the 37 geographical localities in the Palestinian West Bank-, only three were significant purely spatial COVID-19 clusters, one of which was in Al-Khalil, the highest populated city in Palestine and the most affected city with the highest number of COVID-19 deaths. The other two foci were small rural marginalized villages in the Nablus District, the second highest populated district in Palestine, which bore the second highest number of COVID-19 casualties after Al-Khalil. The 182 sequences, of which 129 were from Palestine, were distributed within three major clusters (Fig. 3) without showing any geographical pattern of distribution. Cluster I was the largest of the three ( $n = 96$ ) and contained the Wuhan reference strain, indicating that the wild type was still dominant in the study area during the second surge of the pandemic, reflecting a very recent population expansion. The genetic variation seen in SARS-CoV-2 is low as evidenced by the constant ratio of the number of genomes-to-number of haplotypes (h:n) across clusters and equal number of mutations (Eta). Moreover, the low nucleotide diversity ( $\pi$ ) with concomitantly high values of haplotype diversity (Hd) is another indication of low genetic variation among the studied SARS-CoV-2 genome sequences. The Tajima's D values did not significantly depart from neutrality (Tajima's test, neutrality model prediction,  $P = 0.02$ ) in any of the three clusters, which supports the low genetic variation across the genomes studied (Table 1). Another core proof of the low genetic diversity among the three clusters is the low pairwise genetic distance (Fst) between the three clusters ranging from 0.12 to 0.14 ( $< 0.25$ ) and the high gene flow and population migration between the clusters ( $Nm > 3$ ). The high gene flow is expected to increase homogeneity between clusters and reduce genetic differentiation (Slatkin, 1987). The most plausible explanation is the recent expansion of SARS-CoV-2, which did not allow any population differentiation within this relatively short period of time. Similarly, the high mobility of human hosts between clusters in a short period of time owing to extensive international travel allowed rapid exchange of genetic material leading to homogeneity, albeit less likely (Bolnick and Nosil, 2007; Slatkin, 1987). These results of high homogeneity at the nucleotide level are in complete congruence with other studies in the USA and China, reaching a homogeneity level of 99.99 to 100% (Kaushal et al., 2020; Wang et al., 2020b). The other genetic differentiation parameters were almost equally low, supporting homogeneity and low genetic diversity (Table 2).

## 5. Conclusion

The study revealed the prompt introduction of the Alpha variant of the SARS-CoV-2 virus during the second surge of the COVID-19 pandemic, which caused serious sickness and death. Two districts in the Palestinian West Bank had the highest burden of disease, the Nablus District and the Al-Khalil District. At the time of this study, the pandemic was relatively recent and the studied sequences were still homogenous with minimal genetic variation. The study emphasized the importance of WGS surveillance in monitoring SARS-CoV-2 in the community in terms of the spread of the disease and in pinpointing early cases caused by highly transmissible emerging variants.

## Declaration of Competing Interest

The authors declare that there is no conflict of interests.

## Acknowledgments

This study received partial funding from the Deanship of Scientific Research of Al-Quds University and an in-kind donation of Illumina sequencing kits from Medipharm Co. Ltd. in Palestine.

## References

- Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Cech, M., Chilton, J., Clements, D., Coraor, N., Gruning, B.A., Guerler, A., Hillman-Jackson, J., Hiltemann, S., Jalili, V., Rasche, H., Soranzo, N., Goecks, J., Taylor, J., Nekrutenko, A., Blankenberg, D., 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* 46, W537–W544.
- ARTIC Network, 2019. ARTIC Network 2019. ARTIC nanopore protocol for nCoV2019 novel coronavirus.
- Bolnick, D.I., Nosil, P., 2007. Natural selection in populations subject to a migration load. *Evolution* 61, 2229–2243.
- Canton, R., De Lucas Ramos, P., Garcia-Botella, A., Garcia-Lledo, A., Gomez-Pavon, J., Gonzalez Del Castillo, J., Hernandez-Sampelayo, T., Martin-Delgado, M.C., Martin Sanchez, F.J., Martinez-Selles, M., Molero Garcia, J.M., Moreno Guillen, S., Rodriguez-Artalejo, F.J., Ruiz-Galiana, J., Bouza, E., 2021. New variants of SARS-CoV-2. *Rev. Esp. Quimioter* 34 (5), 419–428. <https://doi.org/10.37201/req/071.2021>.
- CDC.gov, 2021. Centers for Disease Control and Prevention 2021: SARS-CoV-2 Variant Classifications and Definitions.
- Eden, J.S., Rockett, R., Carter, L., Rahman, H., de Ligt, J., Hadfield, J., Storey, M., Ren, X., Tulloch, R., Basile, K., Wells, J., Byun, R., Gilroy, N., O'Sullivan, M.V., Sintchenko, V., Chen, S.C., Maddocks, S., Sorrell, T.C., Holmes, E.C., Dwyer, D.E., Kok, J., nCoV, V.S.G., 2020. An emergent clade of SARS-CoV-2 linked to returned travellers from Iran. *Virus Evol.* 6, veaa027.
- Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791.
- Gohl, D.M., Garbe, J., Grady, P., Daniel, J., Watson, R.H.B., Auch, B., Nelson, A., Yohe, S., Beckman, K.B., 2020. A rapid, cost-effective tailed amplicon method for sequencing SARS-CoV-2. *BMC Genomics* 21, 863.
- Hudson, R.R., Slatkin, M., Maddison, W.P., 1992. Estimation of levels of gene flow from DNA sequence data. *Genetics* 132, 583–589.
- Kaushal, N., Gupta, Y., Goyal, M., Khaiboullina, S.F., Baranwal, M., Verma, S.C., 2020. Mutational frequencies of SARS-CoV-2 genome during the beginning months of the outbreak in USA. *Pathogens* 9.
- Konings, F., Perkins, M.D., Kuhn, J.H., Pallen, M.J., Alm, E.J., Archer, B.N., Barakat, A., Bedford, T., Bhiman, J.N., Caly, L., Carter, L.L., Cullinane, A., de Oliveira, T., Druce, J., El Masry, I., Evans, R., Gao, G.F., Gorbalenya, A.E., Hamblion, E., Herring, B.L., Hodcroft, E., Holmes, E.C., Kakkar, M., Khare, S., Koopmans, M.P.G., Korber, B., Leite, J., MacCannell, D., Marklewitz, M., Maurer-Stroh, S., Rico, J.A.M., Munster, V.J., Neher, R., Munnink, B.O., Pavlin, B.I., Peiris, M., Poon, L., Pybus, O., Rambaut, A., Resende, P., Subissi, L., Thiel, V., Tong, S., van der Werf, S., von Gottberg, A., Ziebuhr, J., Van Kerkhove, M.D., 2021. SARS-CoV-2 variants of interest and concern naming scheme conducive for global discourse. *Nat. Microbiol.* 6, 821–823.
- Korber, B., Fischer, W.M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Hengartner, N., Giorgi, E.E., Bhattacharya, T., Foley, B., Hastie, K.M., Parker, M.D., Partridge, D.G., Evans, C.M., Freeman, T.M., de Silva, T.I., Sheffield, C.-G.G., McDanal, C., Perez, L.G., Tang, H., Moon-Walker, A., Whelan, S.P., LaBranche, C.C., Saphire, E.O., Montefiori, D.C., 2020. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* 182 (812–827), e819.
- Kulldorff, M., Heffernan, R., Hartman, J., Assuncao, R., Mostashari, F., 2005. A space-time permutation scan statistic for disease outbreak detection. *PLoS Med.* 2, e59.
- Kumar, S., Stecher, G., Li, M., Nnyaz, C., Tamura, K., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549.
- Mor, O., Mandelboim, M., Fleishon, S., Bucris, E., Bar-Ilan, D., Linial, M., Nemet, I., Kliker, L., Lustig, Y., Israel National Consortium For Sars-Co, V.S., Mendelson, E.S., Zuckerman, N.S., 2021. The rise and fall of a local SARS-CoV-2 variant with the spike protein mutation L452R. *Vaccines (Basel)* 9.
- Munitz, A., Yechezkel, M., Dickstein, Y., Yamin, D., Gerlic, M., 2021. BNT162b2 vaccination effectively prevents the rapid rise of SARS-CoV-2 variant B.1.1.7 in high-risk populations in Israel. *Cell Rep. Med.* 2, 100264.
- Nei, M., 1973. Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. U. S. A.* 70, 3321–3323.
- Nextstrain, 2021. Nextstrain 2021. Genomic epidemiology of novel coronavirus - Global subsampling.
- Palestine Ministry of Health (PMOH), 2021. COVID-19 in Palestine. PMOH, Ramallah.
- Quick, J., Grubaugh, N.D., Pullan, S.T., Claro, I.M., Smith, A.D., Gangavarapu, K., Oliveira, G., Robles-Sikisaka, R., Rogers, T.F., Beutler, N.A., Burton, D.R., Lewis-Ximenez, L.L., de Jesus, J.G., Giovanetti, M., Hill, S.C., Black, A., Bedford, T., Carroll, M.W., Nunes, M., Alcantara Jr., L.C., Sabino, E.C., Baylis, S.A., Faria, N.R., Loose, M., Simpson, J.T., Pybus, O.G., Andersen, K.G., Loman, N.J., 2017. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat. Protoc.* 12, 1261–1276.
- Qutob, N., Salah, Z., Richard, D., Darwish, H., Sallam, H., Shtayah, I., Najjar, O., Ruzayqat, M., Najjar, D., Balloux, F., van Dorp, L., 2021. Genomic epidemiology of the first epidemic wave of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in Palestine. *Microb. Genom.* 7.
- Rambaut, A., Holmes, E.C., O'Toole, A., Hill, V., McCrone, J.T., Ruis, C., du Plessis, L., Pybus, O.G., 2020. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* 5, 1403–1407.
- Rozas, J., Ferrer-Mata, A., Sanchez-DelBarrio, J.C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S.E., Sanchez-Gracia, A., 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* 34, 3299–3302.
- Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425.
- Slatkin, M., 1987. Gene flow and the geographic structure of natural populations. *Science* 236, 787–792.
- Tajima, F., 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105, 437–460.
- Tamura, K., Nei, M., Kumar, S., 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc. Natl. Acad. Sci. U. S. A.* 101, 11030–11035.
- Tang, J.W., Tambyah, P.A., Hui, D.S., 2021. Emergence of a new SARS-CoV-2 variant in the UK. *J. Inf. Secur.* 82, e27–e28.
- Tegally, H., Wilkinson, E., Giovanetti, M., Iranzadeh, A., Fonseca, V., Giandhari, J., Doolabh, D., Pillay, S., San, E.J., Msomi, N., Mlisana, K., von Gottberg, A., Walaza, S., Allam, M., Ismail, A., Mohale, T., Glass, A.J., Engelbrecht, S., Van Zyl, G., Preiser, W., Petruccione, F., Sigal, A., Hardie, D., Marais, G., Hsiao, N.Y., Korsman, S., Davies, M.A., Tyers, L., Mudau, I., York, D., Maslo, C., Goedhals, D., Abrahams, S., Laguda-Akingba, O., Alisoltani-Dehkordi, A., Gondzik, A., Wibmer, C. K., Sewell, B.T., Lourenco, J., Alcantara, L.C.J., Kosakovsky Pond, S.L., Weaver, S., Martin, D., Lessells, R.J., Bhiman, J.N., Williamson, C., de Oliveira, T., 2021. Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature* 592, 438–443.
- Thomson, E.C., Rosen, L.E., Shepherd, J.G., Spreafico, R., da Silva Filipe, A., Wojcechowskyj, J.A., Davis, C., Piccoli, L., Pascall, D.J., Dillen, J., Lytras, S., Czudnochowski, N., Shah, R., Meury, M., Jesudasan, N., De Marco, A., Li, K., Bassi, J., O'Toole, A., Pinto, D., Colquhoun, R.M., Culpak, K., Jackson, B., Zatta, F., Rambaut, A., Jaconi, S., Sreenu, V.B., Nix, J., Zhang, I., Jarrett, R.F., Glass, W.G., Beltramo, M., Nomikou, K., Pizzuto, M., Tong, L., Cameron, E., Croll, T.I., Johnson, N., Di Iulio, J., Wickenhagen, A., Ceschi, A., Harbison, A.M., Mair, D., Ferrari, P., Smollett, K., Sallusto, F., Carmichael, S., Garzoni, C., Nichols, J., Galli, M., Hughes, J., Riva, A., Ho, A., Schiama, M., Sempile, M.G., Openshaw, P.J.M., Fadda, E., Baillie, J.K., Chodera, J.D., Investigators, I.C., Consortium, C.-G.U., Rihn, S.J., Lycett, S.J., Virgin, H.W., Telenti, A., Corti, D., Robertson, D.L., Snell, G., 2021. Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity. *Cell* 184 (1171–1187), e1120.
- Toovey, O.T.R., Harvey, K.N., Bird, P.W., Tang, J.W.W., 2021. Introduction of Brazilian SARS-CoV-2 484K.V2 related variants into the UK. *J. Inf. Secur.* 82, e23–e24.
- Vassallo, M., Manni, S., Klotz, C., Fabre, R., Pini, P., Blanchouin, E., Sindt, A., Lotte, L., Dubertrand, J.M., Liguori, S., Berkane, N., Duval, Y., Rolland, F., Pradier, C., 2021. Patients admitted for variant alpha COVID-19 have poorer outcomes than those infected with the old strain. *J. Clin. Med.* 10.
- Velavan, T.P., Meyer, C.G., 2020. The COVID-19 epidemic. *Tropical Med. Int. Health* 25, 278–280.
- Wang, C., Horby, P.W., Hayden, F.G., Gao, G.F., 2020a. A novel coronavirus outbreak of global health concern. *Lancet* 395, 470–473.
- Wang, C., Liu, Z., Chen, Z., Huang, X., Xu, M., He, T., Zhang, Z., 2020b. The establishment of reference sequence for SARS-CoV-2 and variation analysis. *J. Med. Virol.* 92, 667–674.
- Worldmeter, 2021. COVID-19 Corona Virus Pandemic.
- Wright, S., 1951. The genetical structure of populations. *Ann. Eugenics* 15, 323–354.