




Molecular Epidemiology of SARS-CoV-2 in Greece Reveals Low Rates of Onward Virus Transmission after Lifting of Travel Restrictions Based on Risk Assessment during Summer 2020

Evangelia Georgia Kostaki,^a Georgios A. Pavlopoulos,^{b,c} Kleio-Maria Verrou,^b Giannis Ampatziadis-Michailidis,^b Vaggelis Harokopos,^c Pantelis Hatzis,^{b,c} Panagiotis Moulos,^{b,c} Nikolaos Sifakas,^d Spyridon Pournaras,^d Christos Hadjichristodoulou,^e Fani Chatzopoulou,^f Dimitrios Chatzidimitriou,^{f,g} Periklis Panagopoulos,^h Panagiota Lourida,ⁱ Aikaterini Argyraki,ⁱ Theodoros Lytras,^j Spyros Sapounas,^k Gerasimos Gerolymatos,^k Georgios Panagiotakopoulos,^k Panagiotis Prezerakos,^l Sotirios Tsiodras,^m Vana Sypsa,^a Angelos Hatzakis,^a Cleo Anastassopoulou,ⁿ Nikolaos Spanakis,ⁿ Athanasios Tsakris,ⁿ Meletios Athanasios Dimopoulos,^{b,o} Anastasia Kotanidou,^p Petros Sfikakis,^{b,q} Georgios Kollias,^{b,r,s} Gkikas Magiorkinis,^a  Dimitrios Paraskevis^a

^aDepartment of Hygiene, Epidemiology and Medical Statistics, Medical School, National and Kapodistrian University of Athens, Athens, Greece

^bCenter of New Biotechnologies & Precision Medicine, Medical School, National and Kapodistrian University of Athens, Athens, Greece

^cInstitute for Fundamental Biomedical Research, Biomedical Sciences Research Center "Alexander Fleming," Vari, Greece

^dLaboratory of Clinical Microbiology, ATTIKON University Hospital, Medical School, National and Kapodistrian University of Athens, Athens, Greece

^eLaboratory of Hygiene and Epidemiology, Faculty of Medicine, Larisa, Greece

^fLabnet, Laboratories, Thessaloniki, Greece

^gDepartment of Microbiology, Medical School, Aristotle University of Thessaloniki, Thessaloniki, Greece

^h2nd Department of Internal Medicine, General Hospital of Alexandroupoli, Democritus University of Thrace, Alexandroupoli, Greece

ⁱInfectious Diseases Clinic A, Thoracic Diseases General Hospital Sotiria, Athens, Greece

^jSchool of Medicine, European University Cyprus, Nicosia, Cyprus

^kNational Public Health Organization, Athens, Greece

^lDepartment of Nursing, University of Peloponnese, Tripoli, Greece

^m4th Department of Internal Medicine, Attikon University Hospital, Medical School, National and Kapodistrian University of Athens, Athens, Greece

ⁿDepartment of Microbiology, Medical School, National and Kapodistrian University of Athens, Athens, Greece

^oDepartment of Therapeutics, Medical School, National and Kapodistrian University of Athens, Athens, Greece

^p1st Intensive Care Unit, General Hospital Evangelismos, National and Kapodistrian University of Athens, Athens, Greece

^q1st Department of Propaedeutic Internal Medicine, Medical School, National and Kapodistrian University of Athens, Athens, Greece

^rInstitute for Bioinnovation, Biomedical Sciences Research Center "Alexander Fleming," Vari, Greece

^sDepartment of Physiology, Medical School, National and Kapodistrian University of Athens, Athens, Greece

Gkikas Magiorkinis and Dimitrios Paraskevis made equal contributions to this work.

ABSTRACT The novel coronavirus severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spread rapidly during the first months of 2020 and continues to expand in multiple areas across the globe. Molecular epidemiology has provided an added value to traditional public health tools by identifying SARS-CoV-2 clusters or providing evidence that clusters based on virus sequences and contact tracing are highly concordant. Our aim was to infer the levels of virus importation and to estimate the impact of public health measures related to travel restrictions to local transmission in Greece. Our phylogenetic and phylogeographic analyses included 389 full-genome SARS-CoV-2 sequences collected during the first 7 months of the pandemic in Greece and a random collection in five replicates of 3,000 sequences sampled globally, as well as the best hits to our data set identified by BLAST. Phylogenetic trees were reconstructed by the maximum likelihood method, and the putative source of SARS-CoV-2 infections was inferred by phylogeographic analysis. Phylogenetic analyses revealed the presence of 89 genetically distinct viruses

Citation Kostaki EG, Pavlopoulos GA, Verrou K-M, Ampatziadis-Michailidis G, Harokopos V, Hatzis P, Moulos P, Sifakas N, Pournaras S, Hadjichristodoulou C, Chatzopoulou F, Chatzidimitriou D, Panagopoulos P, Lourida P, Argyraki A, Lytras T, Sapounas S, Gerolymatos G, Panagiotakopoulos G, Prezerakos P, Tsiodras S, Sypsa V, Hatzakis A, Anastassopoulou C, Spanakis N, Tsakris A, Dimopoulos MA, Kotanidou A, Sfikakis P, Kollias G, Magiorkinis G, Paraskevis D. 2021. Molecular epidemiology of SARS-CoV-2 in Greece reveals low rates of onward virus transmission after lifting of travel restrictions based on risk assessment during summer 2020. *mSphere* 6:e00180-21. <https://doi.org/10.1128/mSphere.00180-21>.

Editor Christina F. Spiropoulou, U.S. Centers for Disease Control and Prevention

Copyright © 2021 Kostaki et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Dimitrios Paraskevis, dparask@med.uoa.gr.

Received 25 February 2021

Accepted 6 June 2021

Published 30 June 2021

identified as independent introductions into Greece. The proportion of imported strains was 41%, 11.5%, and 8.8% during the three periods of sampling, namely, March (no travel restrictions), April to June (strict travel restrictions), and July to September (lifting of travel restrictions based on thorough risk assessment), respectively. The results of phylogeographic analysis were confirmed by a Bayesian approach. Our findings reveal low levels of onward transmission from imported cases during summer and underscore the importance of targeted public health measures that can increase the safety of international travel during a pandemic.

IMPORTANCE Our study based on current state-of-the-art molecular epidemiology methods suggests that virus screening and public health measures after the lifting of travel restrictions prevented SARS-CoV-2 onward transmission from imported cases during summer 2020 in Greece. These findings provide important data on the efficacy of targeted public health measures and have important implications regarding the safety of international travel during a pandemic. Our results can provide a roadmap about prevention policy in the future regarding the reopening of borders in the presence of differences in vaccination coverage, the circulation of the virus, and the presence of newly emergent variants across the globe.

KEYWORDS SARS-CoV-2, molecular epidemiology, public health, phylogeography, travel restrictions

In December 2019, a new respiratory disease was reported in Wuhan, China, which was found to be caused by a novel coronavirus named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (1). The new virus spread globally and caused a pandemic associated with increased morbidity and mortality rates (2). In the absence of an effective vaccine in the first year of the pandemic, nonpharmaceutical interventions (NPIs), such as social distancing, use of masks in the community, travel restrictions, and school and nonessential shop closures were implemented to control community transmission (3). The health, social, and economic consequences of the pandemic are continuing; thus, a better understanding of the characteristics of SARS-CoV-2 transmission is needed to minimize its consequences.

Molecular epidemiology analyses of SARS-CoV-2 full-genome sequences have been extensively performed to classify viral diversity into groups or lineages/sublineages (4), to provide continuous monitoring of virus dispersal patterns, and to obtain insights into critical epidemiological or public health issues related to the geographic origin and dating of viral transmission (5–7). For example, the results of phylogenetic studies revealed that the origin of transmission during the first pandemic wave in Italy and in Seattle, Washington, were from different sources in Asia (8). Phylogenetic analysis of virus samples revealed SARS-CoV-2 clusters and tourism-associated virus dispersal of the first wave in Austria (9). In the United Kingdom, where virus genetic diversity has been systematically surveyed (5), a detailed description of the characteristics of transmission by means of the number and size of local clusters has been performed, as well as quantification of the spatiotemporal characteristics of viral diversity (10). The origin and dynamics of virus importation patterns during the first wave in the United Kingdom were also mapped (10). Importantly, a study from Iceland reported high concordance between the contacts identified by contact tracing and molecular data, suggesting that the latter can be used to control viral transmission in the community (11). Recently, genomic surveillance has been of interest due to reports that certain new lineages found to rapidly spread across the United Kingdom, South Africa, Brazil, and other areas in recent months (i.e., B.1.1.7, B.1.351, and P.1) may confer different biological characteristics to the virus (12–14).

In Greece, the first pandemic wave was mild due to the early implementation of public health measures and the high compliance of the population with the imposed lockdowns (15, 16). Public health measures included the suspension of all educational activities and all services of religious worship, and the closure of bars, cafes, restaurants, retail shops,

museums, and sports facilities in the country (<https://eody.gov.gr>). Moreover, a travel ban for travellers coming from abroad and a 14-day posttravel quarantine period were implemented in the middle of March. It has been shown that, during this period, an outbreak in Athens, Greece, was seeded from different countries (17). By the end of October 2020, the country was experiencing rapid increases in the number of SARS-CoV-2 cases in the metropolitan area of Thessaloniki and other areas of northern Greece. In the meantime, between the lifting of the first measures in May and this second wave, the number of cases remained relatively low, even after travel restrictions were lifted at the beginning of July 2020. To date, several issues remain unanswered, such as how SARS-CoV-2 was introduced in the country at different time periods, what the patterns of virus dispersal were, and importantly, what the impact of the lifting of travel restrictions was on SARS-CoV-2 transmission.

By applying molecular epidemiology methods, we aimed to quantify the levels of virus importation in comparison with surveillance data during these different time periods, to investigate the patterns of SARS-CoV-2 dispersal, and to estimate the impact of public health measures related to travel restrictions to local transmission in Greece.

RESULTS

Our study data comprised of 389 unique full-genome SARS-CoV-2 sequences, of which 280 were newly generated and 109 were available on the GISAID database, collected in Attica, Greece, until 1 December 2020 (17). The vast majority of our samples had been collected in Attica ($n = 353$, 90.7%), which is the largest metropolitan area and comprises approximately 40% of the population in Greece and the largest airport and major transit hub in the country. The total number of coronavirus disease 2019 (COVID-19)-related deaths during our sampling period in Attica equalled 40% of the total deaths across the country, suggesting that SARS-CoV-2 cases were distributed proportionally across Attica and the rest of Greece. To investigate the patterns of SARS-CoV-2 infection in the areas of northeastern Greece and Thessaly, where virus surges were reported in March and May, respectively, we analyzed 17 samples drawn from Alexandroupoli, Kavala, Komotini, and Xanthi in northeastern Greece and 13 samples from the Nea Smirni area in Larissa, Thessaly (see Fig. S1 in the supplemental material). A few samples ($n = 4$) analyzed as part of routine diagnostic testing in Attica were also available from three Aegean islands (Fig. S1).

As mentioned in Materials and Methods, the sampling process covered three time periods. The study samples included 156 of 1,565 diagnosed cases (10%) for the first period, 101 of 1,873 cases (5.4%) for the second period, and 132 of 15,869 cases (0.8%) for the third period. The lower proportion for the third period was due to the number of tests performed increasing gradually with time (i.e., the average number of tests per month was approximately $10\times$ higher in the third period versus the first period), suggesting that the last period was not underrepresented in our sample. To estimate the proportion of samples per the actual number of infections, we used as a proxy the total number of COVID-19-associated deaths estimated for each time period plus 11 days, which is the median time of death since the diagnosis date (<https://eody.gov.gr>). Specifically, the ratio of samples per deaths were 156 of 90 deaths (1.73) for the first period, 101 of 91 deaths (1.11) for the second period, and 132 of 244 deaths (0.54) for the third period, suggesting that sampling size compared to the total number of deaths was 3 and 2 times higher in the first and second period than the third period, respectively. These differences were smaller than the differences in proportions estimated using the number of diagnosed cases.

The results of the classification of viral sequences into lineages, as estimated using the pangolin program, are presented in Table 1. The most frequent lineages were B.1.1 (European lineage, 40.6%), B.1.1.152 (Russian lineage, 19.5%), B.1.1.38 (the United Kingdom lineage, 11.8%), B.1 (a European lineage that corresponds to the spring outbreak in Italy, 5.7%), B (basal lineage from China with many global exports, 4.4%), and B.40 (lineage dominant in the United Kingdom and Australia, 5.1%), while the A

TABLE 1 Lineages of the study sequences per time period

Lineage	No. of sequences (%) in the following time period:			
	First	Second	Third	All (total)
A.2	2 (1.28)			2 (0.51)
A.5	3 (1.92)			3 (0.77)
B	14 (8.97)	3 (2.97)		17 (4.37)
B.1	13 (8.33)	2 (1.98)	7 (5.3)	22 (5.66)
B.1.1	81 (51.92)	50 (49.5)	27 (20.45)	158 (40.62)
B.1.1.1	1 (0.64)		1 (0.76)	2 (0.51)
B.1.1.38	5 (3.21)	13 (12.87)	28 (21.21)	46 (11.83)
B.1.1.70			1 (0.76)	1 (0.26)
B.1.1.100			1 (0.76)	1 (0.26)
B.1.1.102			1 (0.76)	1 (0.26)
B.1.1.145	1 (0.64)			1 (0.26)
B.1.1.152	8 (5.13)	23 (22.77)	45 (34.09)	76 (19.54)
B.1.1.237		1 (0.99)	2 (1.52)	3 (0.77)
B.1.1.291			9 (6.82)	9 (2.31)
B.1.1.315			1 (0.76)	1 (0.26)
B.1.5	4 (2.56)	1 (0.99)	2 (1.52)	7 (1.80)
B.1.22		1 (0.99)		1 (0.26)
B.1.36			3 (2.27)	3 (0.77)
B.1.36.6			1 (0.76)	1 (0.26)
B.1.98	2 (1.28)			2 (0.51)
B.1.160			1 (0.76)	1 (0.26)
B.1.177			1 (0.76)	1 (0.26)
B.1.255	4 (2.56)			4 (1.03)
B.1.319			1 (0.76)	1 (0.26)
B.3	2 (1.28)			2 (0.51)
B.4	1 (0.64)			1 (0.26)
B.28	1 (0.64)			1 (0.26)
B.39	1 (0.64)			1 (0.26)
B.40	13 (8.33)	7 (6.93)		20 (5.14)
Total	156 (100)	101 (100)	132 (100)	389 (100)

lineages originally detected at the early stages of the pandemic in Asia were present at low frequencies in Greece (A2, 0.5%; A5, 0.8%).

To investigate the patterns of SARS-CoV-2 pandemic dispersal in Greece, we performed phylogenetic analyses on five different data sets, including as reference a random collection of globally sampled sequences and the best hits identified by BLAST. The selection of sequences was performed over all SARS-CoV-2 lineages to avoid any biases stemming from lineage classification. Phylogenetic analyses on the different alignments revealed similar results, with at least 89 genetically distinct viruses identified as independent introductions in Greece. This number corresponds to the number of sequences ($n=63$) not falling within phylogenetic clusters with other sequences from Greece, here named singletons, as previously reported (10), plus the number of different clusters ($n=26$) comprising sequences from Greece. Given that our sample pool corresponds to 10% of the diagnosed cases and, also, that the actual number of SARS-CoV-2 infections was likely severely underdiagnosed, the number of different lineages introduced to Greece should be higher than our estimation. The characteristics of SARS-CoV-2 phylogenetic clusters are depicted in Fig. 1, where, in addition to the 63 sequences that were not associated with onward transmission in Greece, we found several small clusters consisting of 2 to 6 sequences and two larger ones of 31 and 221 sequences (Fig. 2A to C). The second largest cluster included 31 identical sequences sampled at the early stage of the pandemic from 10 to 31 March 2020 in Greece. The largest cluster consisted of 37 (16.7%; first period), 72 (32.6%; second period), and 112 (50.7%; third period) sequences collected during the respective sampling periods (Fig. 2C). Notably, samples from the second and third periods dominated in the largest

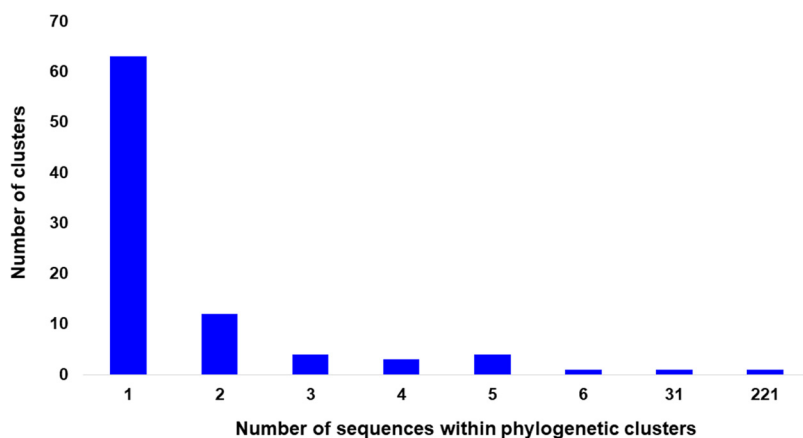


FIG 1 Distribution of the number of sequences per phylogenetic cluster in Greece. The horizontal axis indicates the number of sequences within clusters, and the vertical axis indicates the number of the corresponding clusters.

local cluster. Furthermore, this cluster was underpinned by high levels of phylogenetic support (Shimodaira-Hasegawa [SH] support > 0.9) and was similarly detected in the phylogenetic tree performed using the random sampling of 15,000 GISAID sequences (SH support > 0.85). Specifically, in the latter tree, the composition of clusters was almost identical, with 31 and 217 sequences falling within the two largest clusters.

The putative numbers and sources of virus importation during the three time periods were inferred by means of phylogeographic analyses. The patterns of SARS-CoV-2 importation differed greatly between the three time periods: the proportion of imported infections peaked during the first period (mean value over the five data sets, 41%), while it remained low in the second (mean value over the five data sets, 11.5%) and third (mean value over the five data sets, 8.8%) periods (Fig. 3A). The numbers of imported infections were similar across the different data sets and matched the proportion of imported cases reported from SARS-CoV-2 surveillance (Fig. 3A). The corresponding figures were 31.2%, 15.5%, and 13.8% for the three periods, respectively (Fig. 3A). Implementation of travel restrictions and quarantine measures were applied in the middle of March, causing a decline in international arrivals, and were maintained until June (Fig. 3B). Notably, the proportion of imported infections remained low after the lifting of restrictions on international travel implemented on 1 July in Greece (Fig. 3B), and although a virus surge was detected in August, it was not associated with an increased proportion of imported infections (Fig. 3C).

To investigate the significance of the pattern of virus importation, we compared the previous estimates with the expected proportions of imported infections under a scenario of panmixis. The results of our analysis suggested that the estimated proportions of imported infections were much lower than those expected by chance (Fig. S2), even during the first period ($P < 0.001$); however, the differences were more pronounced in the second and third periods ($P < 0.001$) (Fig. S2). These findings suggest that local transmission eventually dominated during the SARS-CoV-2 pandemic but was less pronounced at the early stages when travel restrictions had not yet been implemented.

The putative geographic origin of the imported infections showed that the majority originated from Europe and specifically from the United Kingdom (23 out of 43 imported cases; 53.5%) in the first period (Table 2). However, a considerable number of transmissions originated from non-European countries (17 out of 43 imported cases; 39.5%) (Table 2). Subsequent analysis revealed that these cases were imported from America and Asia. Specifically, 71.4% of the imported infections outside Europe were from the United States, 14.3% were from Japan, and the remaining 14.3% were from Malaysia. During the second period, most of the imported infections were inferred to

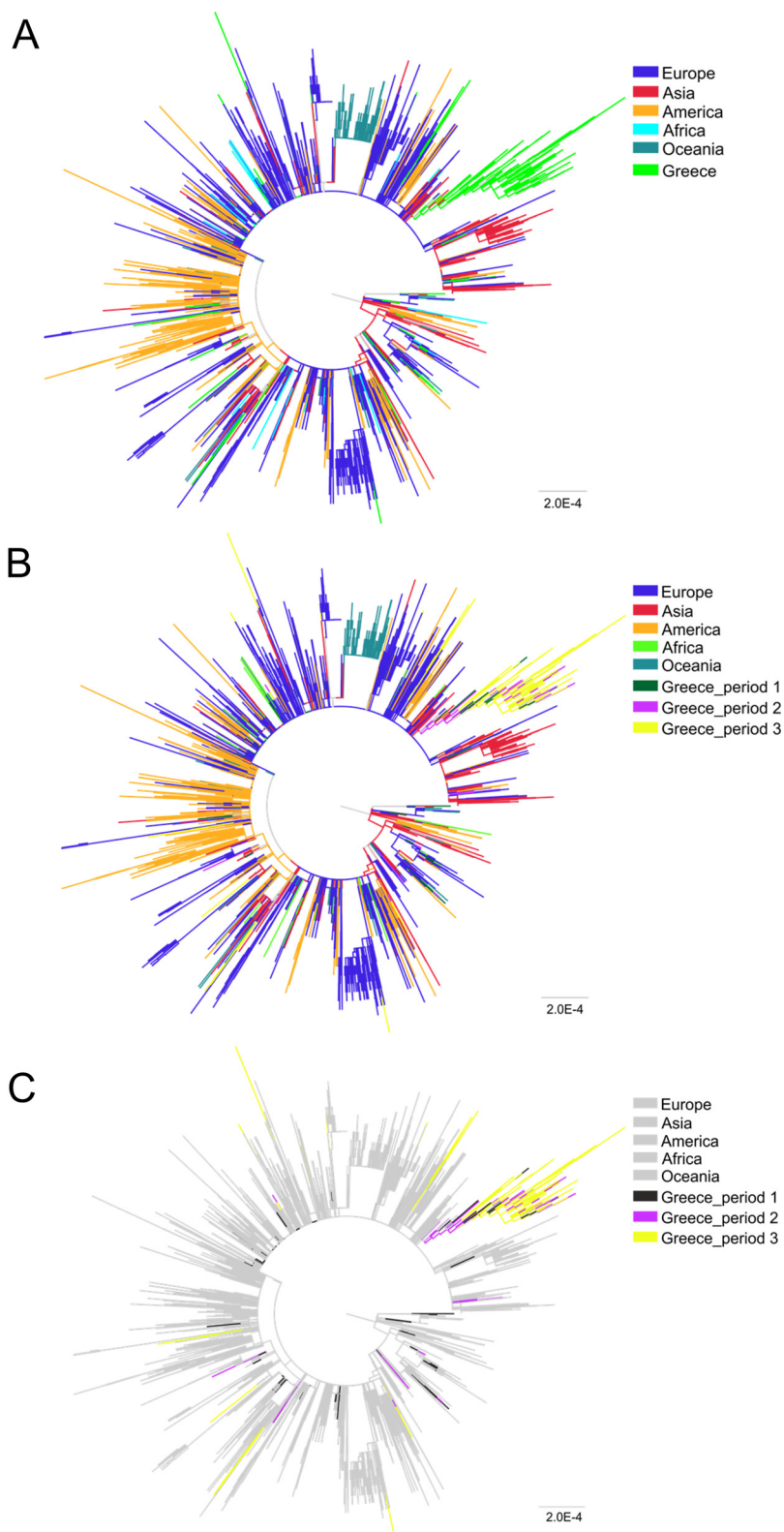


FIG 2 Unrooted phylogenetic tree estimated by FastTree (version 2) of SARS-CoV-2 sequences from Greece ($n=389$) and a global reference data set $n=4,647$). (A) All sequences from Greece are colored in light green. (B) Sequences from Greece are marked in dark green (sampling period 1, 29/2/2020 to 31/03/2020 [no travel restrictions]), purple (sampling period 2, 1/4/2020 to 30/6/2020 [travel restrictions]), and yellow (sampling period 3, 1/7/2020 to 29/9/2020 [lifting of travel restrictions]). (C) Sequences from Greece sampled from different time periods are shown in different colors, and all reference sequences are shown in gray.

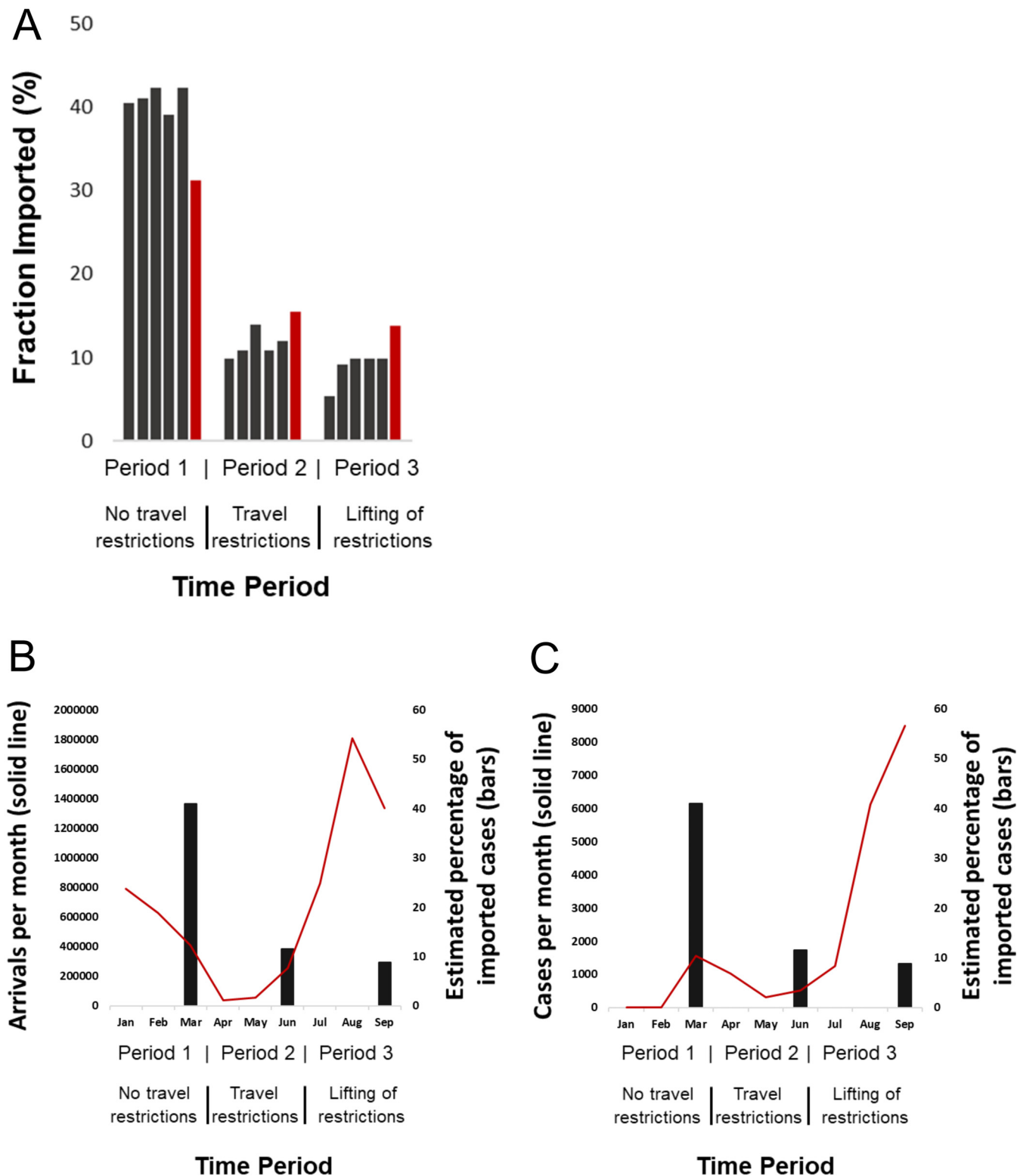


FIG 3 Proportion of virus importation estimated by phylogeographic analysis over the three sampling periods (sampling period 1, 29/2/2020 to 31/3/2020 [no travel restrictions]; sampling period 2, 1/4/2020 to 30/06/2020 [travel restrictions]; sampling period 3, 1/7/2020 to 29/9/2020 [lifting of restrictions]). (A) Proportions of virus importation inferred by phylogeographic analysis using five different data sets (black) and surveillance data (red). Black bars indicate the proportion of virus importation inferred by phylogeographic analysis (mean value estimated from the five different data sets) in combination with the number of international arrivals per month (red line) (B) and the number of SARS-CoV-2 cases per month in Greece (red line) (C).

TABLE 2 Estimated number of imported cases (migration events)

Country	No. of imported cases in the following time period:		
	Period 1 (Feb. 29 to Mar. 31) (<i>n</i> = 156 sequences)	Period 2 (Apr. 1 to June 30) (<i>n</i> = 101 sequences)	Period 3 (July 1 to Sept. 29) (<i>n</i> = 132 sequences)
Non-European countries	17	8	3
United Kingdom	23	1	2
Denmark	0	0	1
Germany	3	0	0

have originated from non-European countries (8 out of 9; 88.9%), specifically from Japan, and the rest from the United Kingdom (1 out of 9; 11.1%) (Table 2). During the third period, half of the imported cases were traced to countries outside Europe (66.7% from the United States and 33.3% from Saudi Arabia), and the remaining 33.3% and 16.7% were from the United Kingdom and Denmark, respectively (Table 2). According to the surveillance data, the highest number of imported SARS-CoV-2 cases were from the United Kingdom, at 36.1% and 16.8% for the first and second periods, respectively, proportions that were similar to those estimated by phylogeographic analysis. Data about individuals' travel history were collected from interviews during sample collection. No information about the origin of potential imported cases was available for the third period.

The patterns of imported infections in the three time periods were verified using a Bayesian phylogeographic approach on three different data sets, including 50 sequences from Greece and 150 globally sampled reference sequences. The median number of imported infections was 13.9 (95 highest posterior density [HPD], 11.0 to 17.3) corresponding to 27.8%, 6.2 (95 HPD, 5.0 to 7.1) corresponding to 12.4%, and 5.2 (95 HPD, 5.0 to 6.0) corresponding to 10.4% for the three time periods, respectively. These figures were similar to our previous phylogeography inferences using the criterion of parsimony and confirmed that viral dispersal pattern differed between the three time periods. Based on the dated tree of the first sampling period corresponding until the end of March 2020, the time of origin of the most recent common ancestor (tMRCA) of the largest monophyletic cluster of SARS-CoV-2 sequences from Greece was estimated to be 15 February 2020 (95% HPD, 2 February 2020 to 26 February 2020) (Fig. 4). This date corresponds to the putative origin of SARS-CoV-2 infections in Greece.

DISCUSSION

In the current study, using SARS-CoV-2 genomes from three distinct time periods, we showed that imported lineages were responsible for 41% of transmissions during the first pandemic wave in Greece. Moreover, we found that levels of virus importation significantly decreased during the period of travel restrictions and quarantine measures and, notably, remained low even after the opening of borders during the 3 months of the peak tourist season. Our results were robust across different reference data sets and correlated strongly with the surveillance data regarding both the proportion of imported infections and the putative origin of SARS-CoV-2 lineages. The transmission patterns were further confirmed using a Bayesian phylogeographic approach and although the analysis was performed after downsampling, the pattern of imported infections was similar to the statistical phylogeography estimations, with the number of imported cases being lower in the period of travel ban and quarantine and after the lifting of travel restrictions. Our findings suggest that imported infections dominated at the early stage of the pandemic before the implementation of travel bans.

More importantly, we found that virus importation remained low and did not substantially contribute to SARS-CoV-2 onward transmission even after the lifting of travel restrictions. Since 1 July 2020, all incoming travelers, including Greek citizens, need to

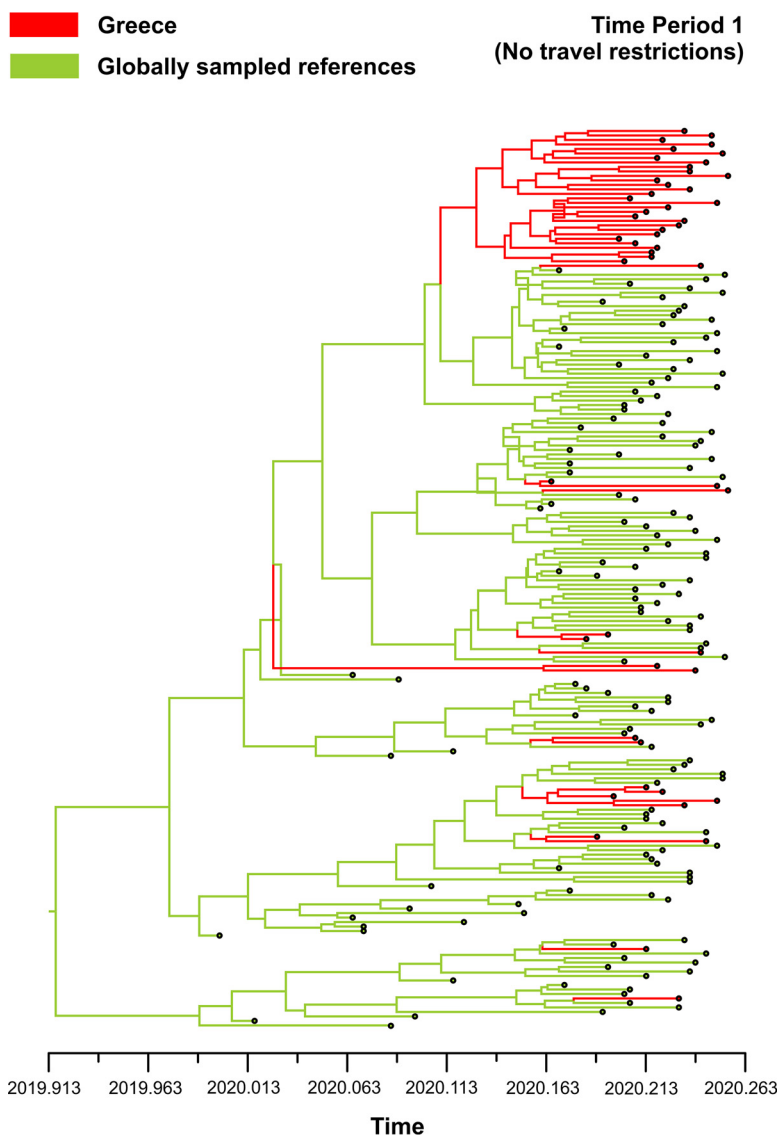


FIG 4 Dated phylogenetic tree of the first sampling phase until the end of March 2020. The time of the most recent common ancestor (tMRCA) of the largest monophyletic cluster including SARS-CoV-2 sequences from Greece is indicated at the node of the corresponding cluster. The monophyletic cluster is shown in red, and reference sequences are shown in green.

have completed a passenger locator form (PLF) 48 h before entering Greece. Health screening procedures have been put in place at airports and other ports of entry, where targeted testing has been performed guided by an artificial intelligence algorithm termed EVA. The algorithm combines information from previous tests performed at entry points in the country, as well as data obtained from the PLFs, creating an importation risk profile for each visitor according to the country of origin. Health authorities can utilize this profile to determine border molecular testing prioritization, thus enhancing public health protection. Risk assessment for all countries was continuously performed daily, and measures were modified accordingly, such as when entry restrictions were tightened for some countries when an increase in the number of SARS-CoV-2-positive cases was observed (a negative SARS-CoV-2 PCR test being required for entry). Additional public health measures were implemented (social distancing, local lockdowns, compulsory use of face masks in public spaces, etc.) locally or nationwide as necessary by local and national authorities.

Our study suggests that the impact of travelers to SARS-CoV-2 local transmission in Greece was low during the summer. To our knowledge, this is one of few molecular epidemiology studies showing that the lifting of travel restrictions after the first pandemic wave was not associated with onward transmission driven by imported SARS-CoV-2 cases. This was most likely due to virus screening at entry points and public health measures implemented during the summertime and afterwards, which helped to control virus spread in the community. Notably, except for a few islands (i.e., Paros, Mykonos), no virus surges were detected during the summer period in Greece, and the effective reproductive number R remained around 1.1 to 1.2 during this period (National Public Health Organization, unpublished data).

Early importation events observed during the first period resulted in large clusters only in one case; local clusters, potentially seeded after the lifting of travel restrictions, may have remained undetected in our study. However, if this hypothesis were true, we would expect to have observed a high number of singleton (imported) infections, which was not the case. This suggests that the former hypothesis does not provide the most plausible explanation for the SARS-CoV-2 dispersal pattern during the third described period in Greece. Moreover, although our samples were not collected at tourist destinations, they were drawn from Attica, where almost 40% of the total Greek population resides, and which fuels tourism in these destinations during the summertime. Therefore, if new strains were associated with high levels of local transmission, we should have been able to detect them through our sampling.

Our findings are similar to those of previous studies in Europe and the Americas showing that the levels of imported infections declined after the implementation of travel restrictions during the first pandemic wave (10). The scale of virus importation in Greece was in accordance with that in Boston, Massachusetts, prior to 28 March 2020 (approximately 35%) and thereafter, when travel restrictions were implemented (median, 9.3%) (7). We also showed that the majority of strains during the first pandemic wave were imported from Europe, and specifically from the United Kingdom, but a significant proportion of virus importation originated in non-European countries. This pattern matched the origin of cases associated with travel during the first phase, suggesting that, although phylogeographic accuracy can be compromised due to potentially nonrepresentative sequencing, in our analyses, the putative origin of imported cases estimated by phylogeography matched that estimated by surveillance data.

Bayesian phylogeographic analysis revealed that the tMRCA of the largest SARS-CoV-2 was 15 February 2020, considered the putative origin of SARS-CoV-2 infections in Greece. This date precedes the first case diagnosed in 26 February 2020 in Greece, suggesting that virus was circulating for at least 10 days before it became diagnosed. Although the hypothesis for earlier introduction cannot be rejected, if it happened, it was most likely at low levels and was probably not associated with onward transmission in the population.

Our study has several limitations. Our sampling was not representative and was not performed across Greece. On the other hand, as discussed above, our analysis was based on 389 full-length genomes collected at different time points from Attica, Greece, suggesting that our results reflect a large proportion of the population. Moreover, our study data included a dense sample of diagnosed cases during the first and second phases; sampling proportion was lower in the third period, but this was due to the enhanced testing performed over time. In addition, if viral transmission pattern had been different during the second and third phases (i.e., the number of imported cases had remained as high as in the first phase), we would have been able to identify these changes even with lower sampling proportions. Notably, the proportion of imported cases remained robust after a downsampling and using a Bayesian approach, thus suggesting that viral transmission patterns can be inferred even with lower sampling. Regarding the putative limitation of nonsampling from tourist destinations during the third phase, if SARS-CoV-2 was continuously transmitted from viral lineages imported during the summertime in Greece, we would be able to detect them

in Attica residents, a large proportion of whom visit different places in Greece during the summertime. Our findings suggest that imported cases did not contribute substantially to SARS-CoV-2 local spread between July and September 2020. Importantly, our results on the effects of virus importation correlated with those estimated from surveillance data, thus enhancing the robustness of our findings. We should note that our findings are relevant to the summer period in the Mediterranean region and may not be generalizable for areas with different climatic conditions.

In conclusion, our molecular epidemiology study showed that the estimated proportion of imported cases during the first pandemic wave in Greece was 41% and that virus screening and public health measures after the lifting of travel restrictions prevented SARS-CoV-2 onward transmission from imported cases during summer 2020. These findings provide important insights on the efficacy of targeted public health measures and have important implications regarding the safety of international travel during a pandemic.

MATERIALS AND METHODS

Analyzed SARS-CoV-2 samples. The SARS-CoV-2 samples analyzed in the context of the current study were collected from 29 February to 19 September 2020 in the Attica, Larisa, and Thrace regions of Greece, from two SARS-CoV-2 reference centers. Specifically, all SARS-CoV-2-positive samples available with reverse transcription-PCR (RT-PCR) threshold detection cycles (C_T) ≤ 30 until the end of August 2020 from Attikon University Hospital (first reference center) were included in our analysis. Similarly, samples fulfilling the previous criterion available in September 2020 at the Department of Hygiene, Epidemiology and Medical Statistics of the School of Medicine at National and Kapodistrian University of Athens (second reference center) were included in our analysis. SARS-CoV-2 samples obtained at border control areas from travelers arriving in Greece were excluded from the analysis. Specimen types included rhino-oropharyngeal swabs. Sample inactivation/RNA extraction was performed according to different protocols available across the different laboratories from which the RNA samples were made available.

The study was approved by the Ethics and Bioethics Committee of the Medical School of the National and Kapodistrian University of Athens (protocol 300/25-05-2020).

The SARS-CoV-2 samples selected were part of the routine diagnostic procedures performed in Greece at the two reference centers in Attica or elsewhere in Greece. SARS-CoV-2 testing in Greece is prioritized for symptomatic individuals, high-risk contacts of SARS-CoV-2-positive cases, or populations at high risk for SARS-CoV-2 infection (i.e., health care workers, elderly and/or disabled nursing home residents). Samples collected in September included those performed by the National Public Health Organization (<https://eody.gov.gr/en/>) as part of volunteer testing of the population in Attica, Greece.

Our sampling comprised three time periods: between 29 February and 31 March 2020, between 1 April and 30 June 2020, and between 1 July and 29 September 2020. The three time periods were defined according to the status of travel restrictions implemented in Greece during the SARS-CoV-2 pandemic. Specifically, a general travel ban and quarantine measures (i.e., 14-day quarantine) for all travelers upon arrival from abroad were implemented on March 14 and 16, respectively, and, given the length of the SARS-CoV-2 incubation period, the first phase was extended until the end of March 2020. Similarly, restrictions on all nonessential movement throughout the country were implemented on 23 March 2020. The second period corresponded to when the international travel restrictions were in place, including a quarantine period for arriving travellers, and the third corresponded to the period after the lifting of travel restrictions, based on a thorough risk assessment for all international travellers.

Next-generation sequencing (NGS). RNA samples were processed using the CleanPlex SARS-CoV-2 panel (Paragon Genomics) according to the manufacturer's instructions. Samples were quantitated (Qubit RNA HS [high-sensitivity] assay kit; ThermoFisher), and 50 to 100 ng of total RNA was used for library preparation, with a final PCR amplification of 24 to 26 cycles. The resulting libraries were analyzed on a Bioanalyzer system (high-sensitivity DNA kit; Agilent), quantified (Qubit double-stranded DNA [dsDNA] HS assay kit; ThermoFisher), and multiplexed; they were sequenced on a NextSeq 550 System (Illumina), using the Mid Output kit v2.5 (300 cycles), in paired-end mode.

The quality of FASTQ files was assessed using FastQC (version 0.11.9) (18). Prior to more exhaustive quality controls, the workflow recommended by Paragon Genomics (for full details, see <https://github.com/moulos-lab/greek-covid19-assembly>) was applied for adapter and primer sequence trimming. Further potential adapter leftovers and poor-quality bases were trimmed with TrimGalore (version 0.6.6) (19), which also deploys Cutadapt (version 2.8) (20), keeping reads with length of at least 50 bp. Subsequently, the remaining paired short reads were normalized to 100 \times uniform coverage using BBnorm from the BBmap suite (21) and then subjected to guided *de novo* genome assembly using SPAdes (version 3.14.1) (22) with the *-careful* option. The guided SARS-CoV-2 genome assembly is achieved by using a reference genome with the *-trusted-contigs* option of SPAdes. The reference SARS-CoV-2 genome for the guided assembly was retrieved from University of California at Santa Cruz (23). The quality of the assemblies was assessed using QUAST (version 5.0.2) (24). By using the guided approach, the vast majority of the assemblies was complete. For the few assemblies that were not complete, the MEDUSA scaffolder was deployed in order to complete the assemblies (25). The short reads were also mapped to the SARS-CoV-2 genome retrieved from UCSC using BWA (version 0.7.17) (26) in

order to further assess the quality of virus sequencing and visually inspect coverage and potential virus mutations.

The complete *de novo* genome assembly and assessment procedure can be found online (<https://github.com/moulos-lab/greek-covid19-assembly>). The bioinformatics analysis was performed using the computational infrastructure of the Center of New Biotechnologies & Precision Medicine (pMedGR), School of Medicine, National and Kapodistrian University of Athens, Greece (<https://www.precisionmedicine.gr/>).

Sequence alignment and phylogenetic analysis. The dispersal patterns of SARS-CoV-2 in Greece were investigated by means of phylogenetic analysis. Classification of SARS-CoV-2 sequences in different lineages was implemented in the pangolin webtool (<https://cov-lineages.org/pangolin.html>). Our data consisted of the following: (i) five data sets of 3,000 randomly selected sequences until 30 September 2020 sampled from the GISAID database plus the best hits identified by a BLAST search using as queries all sequences of our study population against the GISAID database sampled at the same time period, and (ii) a data set of 15,000 randomly selected sequences sampled until 30 September 2020 plus the best hits identified by a BLAST search. We collected 173,991 sequences of high quality and of length >29,000 nucleotides (nt) from GISAID until 30 September 2020 and created a BLAST Database. We queried the 389 sequences and set a threshold to report the first 50 best hits (ranked by E value and bit-score). BLAST reported pairwise 20,517 hits matching at different regions. These hits correspond to 2,059 unique sequences. The BLAST search was performed using only the coding region of SARS-CoV-2 (29,410 nt). The total sizes of unique sequence data sets after the inclusion of the best hits, the random sets of 3,000 sequences, and our study population, were 5,039 (data set 1), 5,036 (data set 2), 5,038 (data set 3), 5,039 (data set 4), and 5,036 (data set 5). The size of the data set including the 15,000 randomly selected sequences was 16,919 unique sequences. Analysis was performed without taking into account the classification of SARS-CoV-2 into lineages.

Multiple sequence alignments were performed using the multithreaded version of the MAFFT program (27). This was run using XSEDE, available from the cyberinfrastructure for phylogenetic research (The CIPRES Science Gateway, version 3.3; <https://www.phylo.org/>), and the infrastructure at pMedGR.

Phylogenetic analyses were carried out by the maximum likelihood (ML) method using the IQ-Tree (version 2.1.1) (28) and the FastTree (version 2) (29) programs. Due to constraints in computation time, phylogenetic trees for sequence alignments larger than 3,000 sequences were imputed using FastTree. Phylogenetic analysis was performed using the best-fit nucleotide substitution with ModelFinder and the Akaike information criterion (AIC) as implemented in IQ-Tree (28). The nucleotide substitution model selected more often was GTR+I+G4. For FastTree, GTR+G4 was selected as the nucleotide substitution model. All runs were performed at the CIPRES Science Gateway (version 3.3) and the infrastructure at pMedGR.

The resulting phylogenies were visualized and annotated by the FigTree (version 1.4) and the Dendroscope (version 3.7.2) programs.

Estimation of the number of imported SARS-CoV-2 infections. To infer the dispersal patterns of SARS-CoV-2 (i.e., the number of imported infections versus the within country infection events during different time periods), we performed phylogeographic analysis on all five data sets of 5,036 to 5,039 sequences each. Phylogeographic analysis was performed on the ML tree reconstructed by phylogenetic analysis conducted on the FastTree program in the previous analysis step. Specifically, we estimated the number of SARS-CoV-2 infections (viral migration events) between different geographic areas/countries around the world and Greece (imported infections) during three time periods: between (i) 29 February and 31 March 2020, (ii) 1 April and 30 June 2020, and (iii) 1 July 1 and 29 September 2020. Additionally, we estimated infections occurring locally (local infections) between individuals for whom viral samples were taken at these three time periods. The viral migration events were quantified between the different geographic areas/countries by character reconstruction using the criterion of parsimony as implemented in PAUP*4.0 (30).

We assessed whether the inferred migration events (imported or local infections) were different from those expected by chance (panmixis). Hypothesis testing was performed by character reconstruction using the criterion of parsimony on the Mesquite (version 3.61) program (31). Further details on the methodology of viral migration event estimation have been described in detail elsewhere (32–35).

Bayesian phylogeographic analysis. We further conducted phylogeographic analysis by using the Bayesian approach to verify our findings. Analysis was performed separately on three data sets representing the three periods of sampling (up to March, April to June, and July to September). Each data set consisted of 200 randomly selected sequences (50 sequences from Greece and 150 globally sampled reference sequences). Analysis was performed using the general time reversible (GTR) nucleotide substitution model with the G heterogeneity model, an uncorrelated log normal relaxed clock of molecular clock model with TipDates, a Bayesian skyline nonparametric plot demographic model with 10 groups, and a discrete phylogeographic approach with a nonreversible continuous-time Markov chain (CTMC) model, a Bayesian stochastic search variable selection (BSSVS), and a Markov jump count parameter (36, 37), as implemented in BEAST v1.8.0 (38). Noninformative priors were used for the Markov chain Monte Carlo (MCMC) runs. MCMC analysis was run for each data set for 100×10^6 to 200×10^6 generations, sampled every 10,000 to 20,000 steps (burn-in, 10%). The MCMC convergence and the effective sample sizes (ESS) were checked using the program Tracer v1.7.1. The maximum clade credibility (MCC) tree was selected from the posterior tree distribution by the TreeAnnotator v1.8 program (38) and visualized by the FigTree v1.3.1 program.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, TIF file, 5.6 MB.

FIG S2, TIF file, 1.1 MB.

ACKNOWLEDGMENTS

We acknowledge the support for next-generation sequencing and bioinformatic analysis by the project “The Greek Research Infrastructure for Personalised Medicine (pMED-GR)” (MIS 5002802) which is implemented under the Action “Reinforcement of the Research and Innovation Infrastructure,” funded by the Operational Programme “Competitiveness, Entrepreneurship and Innovation” (NSRF 2014–2020) and cofinanced by Greece and the European Union (European Regional Development Fund) and UniPharma S.A. We also acknowledge support of National Public Health Association, the Greek Shipowners’ Social Welfare Company SYN-ENOSIS, and the Public Benefit Foundation “John S. Latsis.”

REFERENCES

- Lake MA. 2020. What we know so far: COVID-19 current clinical knowledge and research. *Clin Med (Lond)* 20:124–127. <https://doi.org/10.7861/clinmed.2019-coron>.
- Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P, Zhan F, China Novel Coronavirus Investigating and Research Team. 2020. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* 382:727–733. <https://doi.org/10.1056/NEJMoa2001017>.
- Flaxman S, Mishra S, Gandy A, Unwin HJT, Mellan TA, Coupland H, Whittaker C, Zhu H, Berah T, Eaton JW, Monod M, Imperial College COVID-19 Response Team, Ghani AC, Donnelly CA, Riley S, Vollmer MAC, Ferguson NM, Okell LC, Bhatt S. 2020. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* 584:257–261. <https://doi.org/10.1038/s41586-020-2405-7>.
- Rambaut A, Holmes EC, O’Toole A, Hill V, McCrone JT, Ruis C, Du Plessis L, Pybus OG. 2020. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol* 5:1403–1407. <https://doi.org/10.1038/s41564-020-0770-5>.
- COVID-19 Genomics UK (COG-UK) consortium. 2020. An integrated national scale SARS-CoV-2 genomic surveillance network. *Lancet Microbe* 1:e99–e100. [https://doi.org/10.1016/S2666-5247\(20\)30054-9](https://doi.org/10.1016/S2666-5247(20)30054-9).
- Lu J, Du Plessis L, Liu Z, Hill V, Kang M, Lin H, Sun J, Francois S, Kraemer MUG, Faria NR, McCrone JT, Peng J, Xiong Q, Yuan R, Zeng L, Zhou P, Liang C, Yi L, Liu J, Xiao J, Hu J, Liu T, Ma W, Li W, Su J, Zheng H, Peng B, Fang S, Su W, Li K, Sun R, Bai R, Tang X, Liang M, Quick J, Song T, Rambaut A, Loman N, Raghwanji J, Pybus OG, Ke C. 2020. Genomic epidemiology of SARS-CoV-2 in Guangdong Province, China. *Cell* 181:997–1003.e9. <https://doi.org/10.1016/j.cell.2020.04.023>.
- Lemieux JE, Siddle KJ, Shaw BM, Loreth C, Schaffner SF, Gladden-Young A, Adams G, Fink T, Tomkins-Tinch CH, Krasnikova LA, DeRuff KC, Rudy M, Bauer MR, Lagerborg KA, Normandin E, Chapman SB, Reilly SK, Anahar MN, Lin AE, Carter A, Myhrvold C, Kembell ME, Chaluvadi S, Cusick C, Flowers K, Neumann A, Cerrato F, Farhat M, Slater D, Harris JB, Branda JA, Hooper D, Gaeta JM, Baggett TP, O’Connell J, Gnirke A, Lieberman TD, Philippakis A, Burns M, Brown CM, Luban J, Ryan ET, Turbett SE, LaRocque RC, Hanage WP, Gallagher GR, Madoff LC, Smole S, Pierce VM, Rosenberg E, et al. 2021. Phylogenetic analysis of SARS-CoV-2 in Boston highlights the impact of superspreading events. *Science* 371:eabe3261. <https://doi.org/10.1126/science.abe3261>.
- Worobey M, Pekar J, Larsen BB, Nelson MI, Hill V, Joy JB, Rambaut A, Suchard MA, Wertheim JO, Lemey P. 2020. The emergence of SARS-CoV-2 in Europe and North America. *Science* 370:564–570. <https://doi.org/10.1126/science.abc8169>.
- Popa A, Genger JW, Nicholson MD, Penz T, Schmid D, Aberle SW, Agerer B, Lercher A, Endler L, Colaco H, Smyth M, Schuster M, Grau ML, Martinez-Jimenez F, Pich O, Borena W, Pawelka E, Keszei Z, Senekowitsch M, Laine J, Aberle JH, Redlberger-Fritz M, Karolyi M, Zoufaly A, Maritschnick S, Borkovec M, Hufnagl P, Nairz M, Weiss G, Wolfinger MT, von Laer D, Superti-Furga G, Lopez-Bigas N, Puchhammer-Stockl E, Allerberger F, Michor F, Bock C, Bergthaler A. 2020. Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2. *Sci Transl Med* 12:eabe2555. <https://doi.org/10.1126/scitranslmed.abe2555>.
- du Plessis L, McCrone JT, Zarebski AE, Hill V, Ruis C, Gutierrez B, Raghwanji J, Ashworth J, Colquhoun R, Connor TR, Faria NR, Jackson B, Loman NJ, O’Toole A, Nicholls SM, Parag KV, Scher E, Vasylyeva TI, Volz EM, Watts A, Bogoch II, Khan K, COVID-19 Genomics UK (COG-UK) Consortium, Aanensen DM, Kraemer MUG, Rambaut A, Pybus OG. 2021. Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. *Science* 371:708–712. <https://doi.org/10.1126/science.abc2946>.
- Gudbjartsson DF, Helgason A, Jonsson H, Magnusson OT, Melsted P, Norddahl GL, Saemundsdottir J, Sigurdsson A, Sulem P, Agustsdottir AB, Eiriksdottir B, Fridriksdottir R, Gardarsdottir EE, Georgsson G, Gretarsdottir OS, Gudmundsson KR, Gunnarsdottir TR, Gylfason A, Holm H, Jensson BO, Jonasdottir A, Jonsson F, Josefsdottir KS, Kristjansson T, Magnusdottir DN, Le Roux L, Sigmundsdottir G, Sveinbjornsson G, Sveinsdottir KE, Sveinsdottir M, Thorarensen EA, Thorbjornsson B, Love A, Masson G, Jonsdottir I, Moller AD, Gudnason T, Kristinsson KG, Thorsteinsdottir U, Stefansson K. 2020. Spread of SARS-CoV-2 in the Icelandic population. *N Engl J Med* 382:2302–2315. <https://doi.org/10.1056/NEJMoa2006100>.
- Rambaut A, Loman N, Pybus O, Barclay W, Barrelet J, Carabelli A, Connor T, Peacock T, Robertson DL, Volz E, COVID-19 Genomics Consortium UK (CoG-UK). 2020. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutation. ARTIC Network.
- Faria NR, Moraes Claro I, Candido D, Moyses Franco LA, Andrade PS, Coletti TM, Silva CAM, Sales FC, Manuli ER, Aguiar RS, Gaburo N, Camilo CDC, Fraiji NA, Esashika Crispim MA, Carvalho MDPSS, Rambaut A, Loman N, Pybus OG, Sabino EC, CADDE Network. Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings.
- Kosakovsky Pond SL, Wilkison E, Weaver S, James SE, Tegally H, de Oliveira T, Martin D. 2020. A preliminary selection analysis of the South African V501.V2 SARS-CoV-2 clade.
- Sypsa V, Roussos S, Paraskevis D, Lytras T, Tsiodras S, Hatzakis A. 2021. Effects of social distancing measures during the first epidemic wave of severe acute respiratory syndrome infection, Greece. *Emerg Infect Dis* 27:452–462. <https://doi.org/10.3201/eid2702.203412>.
- Tsitsilonis OE, Paraskevis D, Lianidou E, Pterros V, Akalestos A, Kastiritis E, Moutsatsou P, Scorilas A, Sphicopoulos T, Terpos E, Thomaidis N, Tsakris A, Voulgaris N, Daskalaki CC, Evangelakou Z, Fouki C, Gianniou DD, Gumeni S, Kostaki EG, Kostopoulos IV, Manola MS, Orogas-Stavrou N, Panteli C, Papanagnou ED, Rousakis P, Sklirou AD, Smilkou S, Stergiopoulou D, Trougakos IP, Tsiodras S, Sfikakis PP, Dimopoulos MA. 2020. Seroprevalence of antibodies against SARS-CoV-2 among the personnel and students of the National and Kapodistrian University of Athens, Greece: a preliminary report. *Life (Basel)* 10:214. <https://doi.org/10.3390/life10090214>.
- Spanakis N, Kassaia K, Dovrolis N, Bampali M, Gatzidou E, Kafasi A, Froukala E, Stavropoulou A, Lilakos K, Velezta S, Tsiodras S, Tsakris A, Karakasiliotis I. 2021. A main event and multiple introductions of SARS-Cov2 initiated the COVID-19 epidemic in Greece. *J Med Virol* 93:2899–2907. <https://doi.org/10.1002/jmv.26778>.

18. Anonymous. fastqc. Babraham Bioinformatics. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
19. Anonymous. trim_galore. Babraham Bioinformatics. https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.
20. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal* 17:10–12. <https://doi.org/10.14806/ej.17.1.200>.
21. Anonymous. BBMap. <https://sourceforge.net/projects/bbmap/>.
22. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
23. University of California at Santa Cruz. COVID-19 Research at UCSC. University of California at Santa Cruz. <https://genome.ucsc.edu/covid19.html>.
24. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUILT: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>.
25. Bosi E, Donati B, Galardini M, Brunetti S, Sagot MF, Lio P, Crescenzi P, Fani R, Fondi M. 2015. MeDuSa: a multi-draft based scaffold. *Bioinformatics* 31:2443–2451. <https://doi.org/10.1093/bioinformatics/btv171>.
26. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
27. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780. <https://doi.org/10.1093/molbev/mst010>.
28. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 37:1530–1534. <https://doi.org/10.1093/molbev/msaa015>.
29. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. <https://doi.org/10.1371/journal.pone.0009490>.
30. Wilgenbusch JC, Swofford D. 2003. Inferring evolutionary trees with PAUP*. *Curr Protoc Bioinformatics* Chapter 6:Unit 6.4. <https://doi.org/10.1002/0471250953.bi0604s00>.
31. Maddison DR, Maddison WP. 2019. Mesquite: a modular system for evolutionary analysis. Version 3.61. <http://www.mesquiteproject.org>.
32. Angelis K, Albert J, Mamais I, Magiorkinis G, Hatzakis A, Hamouda O, Struck D, Vercauteren J, Wensing AM, Alexiev I, Asjo B, Balotta C, Camacho RJ, Coughlan S, Griskevicius A, Grossman Z, Horban A, Kostrikis LG, Lepej S, Liitsola K, Linka M, Nielsen C, Otelea D, Paredes R, Poljak M, Puchhammer-Stockl E, Schmit JC, Sonnerborg A, Stanekova D, Stanojevic M, Boucher CA, Kaplan L, Vandamme AM, Paraskevis D. 2015. Global dispersal pattern of HIV type 1 subtype CRF01_AE: a genetic trace of human mobility related to heterosexual sexual activities centralized in Southeast Asia. *J Infect Dis* 211:1735–1744. <https://doi.org/10.1093/infdis/jiu666>.
33. Paraskevis D, Kostaki E, Nikolopoulos GK, Sypsa V, Psychogiou M, Del Amo J, Hodges-Mameletzis I, Paraskeva D, Skoutelis A, Malliori M, Williams L, Friedman SR, Daikos GL, Hatzakis A. 2017. Molecular tracing of the geographical origin of human immunodeficiency virus type 1 infection and patterns of epidemic spread among migrants who inject drugs in Athens. *Clin Infect Dis* 65:2078–2084. <https://doi.org/10.1093/cid/cix717>.
34. Magiorkinis G, Angelis K, Mamais I, Katzourakis A, Hatzakis A, Albert J, Lawyer G, Hamouda O, Struck D, Vercauteren J, Wensing A, Alexiev I, Asjo B, Balotta C, Gomes P, Camacho RJ, Coughlan S, Griskevicius A, Grossman Z, Horban A, Kostrikis LG, Lepej SJ, Liitsola K, Linka M, Nielsen C, Otelea D, Paredes R, Poljak M, Puchhammer-Stockl E, Schmit JC, Sonnerborg A, Stanekova D, Stanojevic M, Stylianou DC, Boucher CAB, SPREAD program, Nikolopoulos G, Vasylyeva T, Friedman SR, van de Vijver D, Angarano G, Chaix ML, de Luca A, Korn K, Loveday C, Soriano V, Yerly S, Zazzi M, Vandamme AM, Paraskevis D. 2016. The global spread of HIV-1 subtype B epidemic. *Infect Genet Evol* 46:169–179. <https://doi.org/10.1016/j.meegid.2016.05.041>.
35. Kostaki EG, Nikolopoulos GK, Pavlitina E, Williams L, Magiorkinis G, Schneider J, Skaathun B, Morgan E, Psychogiou M, Daikos GL, Sypsa V, Smyrnov P, Korobchuk A, Malliori M, Hatzakis A, Friedman SR, Paraskevis D. 2018. Molecular analysis of human immunodeficiency virus type 1 (HIV-1)-infected individuals in a network-based intervention (Transmission Reduction Intervention Project): phylogenetics identify HIV-1-infected individuals with social links. *J Infect Dis* 218:707–715. <https://doi.org/10.1093/infdis/jiy239>.
36. Minin VN, Suchard MA. 2008. Counting labeled transitions in continuous-time Markov models of evolution. *J Math Biol* 56:391–412. <https://doi.org/10.1007/s00285-007-0120-8>.
37. Minin VN, Suchard MA. 2008. Fast, accurate and simulation-free stochastic mapping. *Philos Trans R Soc Lond B Biol Sci* 363:3985–3995. <https://doi.org/10.1098/rstb.2008.0176>.
38. Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 29:1969–1973. <https://doi.org/10.1093/molbev/mss075>.