Journal of Infection and Public Health 17 (2024) 102496

Contents lists available at ScienceDirect

Journal of Infection and Public Health

journal homepage: www.elsevier.com/locate/jiph

The impact of insularity on SARS-CoV-2 diffusion: Recapitulating three years of COVID-19 pandemic in the island of Sardinia

Nicole Grandi [a], Roberto Cusano [b], Giovanna Piras [c], Maura Fiamma [c], Maria Itria Monne [c], Tatiana Fancello [c], Jessica Milia [b], Sandro Orrù [d], Sante Scognamiglio [a], Caterina Serra [e,f], Giuseppe Mameli [c,g], Sergio Uzzau [e,f], Germano Orrù [d], Angelo Domenico Palmas [c], Salvatore Rubino [e,f], Enzo Tramontano [a,*]

[a] Lab. of Molecular Virology, Dept. of Life and Environmental Sciences, University of Cagliari, Italy
[b] CRS4 - Center for Advanced Studies, Research and Development in Sardinia, Science and Technology Park Polaris, Cagliari, Italy
[c] ASL Nuoro, Ospedale San Francesco, Italy
[d] Dept. of Surgical Sciences, University of Cagliari, Italy
[e] Dept. of Biomedical Sciences, University of Sassari, Italy
[f] S.C. Microbiologia e Virologia, AOU Sassari, Italy
[g] ASL Sassari Ospedale Civile Alghero, Italy

ARTICLE INFO

Article history:
Received 22 February 2024
Received in revised form 2 July 2024
Accepted 4 July 2024

Keywords:
SARS-CoV-2
COVID-19
Sardinia
Insularity
Genomic diversity

ABSTRACT

Background: Italy has been the first European Country dealing with SARS-CoV-2, whose diffusion on the territory has not been homogeneous. Among Italian regions, Sardinia represented one of the lowest incidence areas, likely due to its insular nature. Despite this, the impact of insularity on SARS-CoV-2 genetic diversity has not been comprehensively described.

Methods: In the present study, we performed the high throughput sequencing of 888 SARS-CoV-2 genomes collected in Sardinia during the first 23 months of pandemics. In addition, 1439 high-coverage SARS-CoV-2 genomes circulating in Sardinia along three years (December 2019 – January 2023) were downloaded from GISAID, for a total of 2327 viral sequences that were characterized in terms of phylogeny and genomic diversity.

Results: Overall, COVID-19 pandemic in Sardinia showed substantial differences with respect to the national panorama, with additional peaks of infections and uncommon lineages that reflects the national and regional policies of re-opening and the subsequent touristic arrivals. Sardinia has been interested by the circulation of at least 87 SARS-CoV-2 lineages, including some that were poorly represented at national and European level, likely linked to multiple importation events. The relative frequency of Sardinian SARS-CoV-2 lineages has been compared to other Mediterranean Islands, revealing a unique composition.

Conclusions: The genomic diversity of SARS-CoV-2 in Sardinia has been shaped by a complex interplay of insular geography, low population density, and touristic arrivals, leading on the one side to the importation of lineages remaining rare at the national level, and resulting on the other side in the delayed entry of otherwise common variants.

1. Introduction

SARS-CoV-2 is the second severe acute respiratory syndrome coronavirus, responsible for COVID-19 pandemic. The first European Country reached by SARS-CoV-2 has been Italy, with the imported cases of a tourists couple from Wuhan (January 2020) followed by the earliest autochthonous diagnosis (February 2020) [1]. The rapid increase of cases led to restrictive measures culminated in a national lockdown (March 9–May 3, 2020). To date, Italy registered > 26,4 million cases on ~59 million inhabitants, with a 0,7 % lethality rate (Italian National Institute of Health, https://www.iss.it/coronavirus, accessed on 3rd November 2023). COVID-19 diffusion in the Italian territory showed a clear gradient from highly-affected Northern regions to lower-incidence Southern areas [2,3] that included Sicily and Sardinia major Mediterranean islands. Sardinia has a higher geographic isolation as compared to Sicily, presenting also limited

* Corresponding author.
E-mail address: tramon@unica.it (E. Tramontano).

https://doi.org/10.1016/j.jiph.2024.102496

internal connectivity and low population density [4,5]. Moreover, Sardinian population is among the elderest in Italy, with 26 % of over-65 inhabitants of which 677 over-100, (Italian National Institute of Statistics, 2023) as a peculiar risk-factor for severe COVID-19 forms. The first COVID-19 case in Sardinia dates back to March 2nd 2020 in Cagliari main city [6]. Following the national lockdown, from March 14th, the region ordered the closure of air and marine connections [6], also to counteract the arrival of Sardinian citizens and northern Italian inhabitants from highly affected regions [3]. The subsequent day, after few isolated cases (n = 77), the first COVID-19 outbreak was officially declared in North Sardinia (n = 35) and the infection progressively spread in the whole Island [6]. In this scenario, few studies analyzed the impact of insularity on SARS-CoV-2 genetic variability in Italy [5,7,8] and other Countries [9–18].

The present study aimed to provide a comprehensive description of the dynamics of diffusion and genetic diversity of SARS-CoV-2 in Sardinia island across the pandemic, to evaluate the impact of insularity with respect to mainland Italy and compare it with other Mediterranean islands. To this purpose, we took into account a total of 2327 SARS-CoV-2 genome sequences circulating in Sardinia along 38 out of 42 months of COVID-19 pandemic.

2. Materials and methods

2.1. Sampling and viral RNA extraction

Between February 2020 and January 2022, 888 SARS-CoV-2 positive nasopharyngeal swabs were selected by three clinical centers (Supplementary Fig. 1) based on CT values < 32 and temporal-geographic distribution. Viral RNA was extracted (Seegene Nimbus system, STARMag Universal Cartridge kit), validated (RealStar SARS-CoV-2 RT-PCR Kit 1.0), and sent to NGS Core for sequencing.

2.2. Next generation sequencing

Viral RNA was mixed with Random Hexamer Primer Mix, dNTPs and nuclease-free water (Thermo Fisher Scientific, total volume = 10,5 μL) and incubated at 65 °C for 5′ followed by 1′ in ice. We then added 5 ×SSIV buffer, DTT, RNaseOUT, SSIV enzyme and nuclease-free water (Thermo Fisher Scientific, total volume=20 μL) and heated to 23 °C-10′, 52 °C-10′, and 80 °C-10′ to obtain cDNA, processed using ARTIC nCoV-2019 V3 amplicon panel. Libraries were prepared from 90 ng amplicons (Illumina DNA Prep), analyzed on Agilent Bioanalyzer 2100, quantified using Qubit dsDNA BR Assay kit, and pooled at equimolar concentrations. Pools were size-selected to obtain ~450 bp fragments that were sequenced on HiSeq3000 (Illumina) to generate ~3–4 millions 150 bp paired-end reads, processed with DRAGEN pipeline (BaseSpace Sequence Hub, Illumina).

2.3. External resources

1439 SARS-CoV-2 sequences circulating in Sardinia (December 2019 to January 2023, low coverage excluded) were downloaded from GISAID (https://gisaid.org, Supplementary Tab. S1) and classified with Nextclade (https://clades.nextstrain.org). The distribution and characteristic mutations of SARS-CoV-2 lineages were inferred with CoVe Tracker [19] and GISAID.

2.4. Phylogenetic analyses

Maximum likelihood tree was inferred with FastTree applying a generalized time-reversible model [20], local support values were computed with the Shimodaira– Hasegawa test and analyzed with Dendroscope, v.3.7.2 [21].

3. Results

3.1. COVID-19 trend in Sardinia island

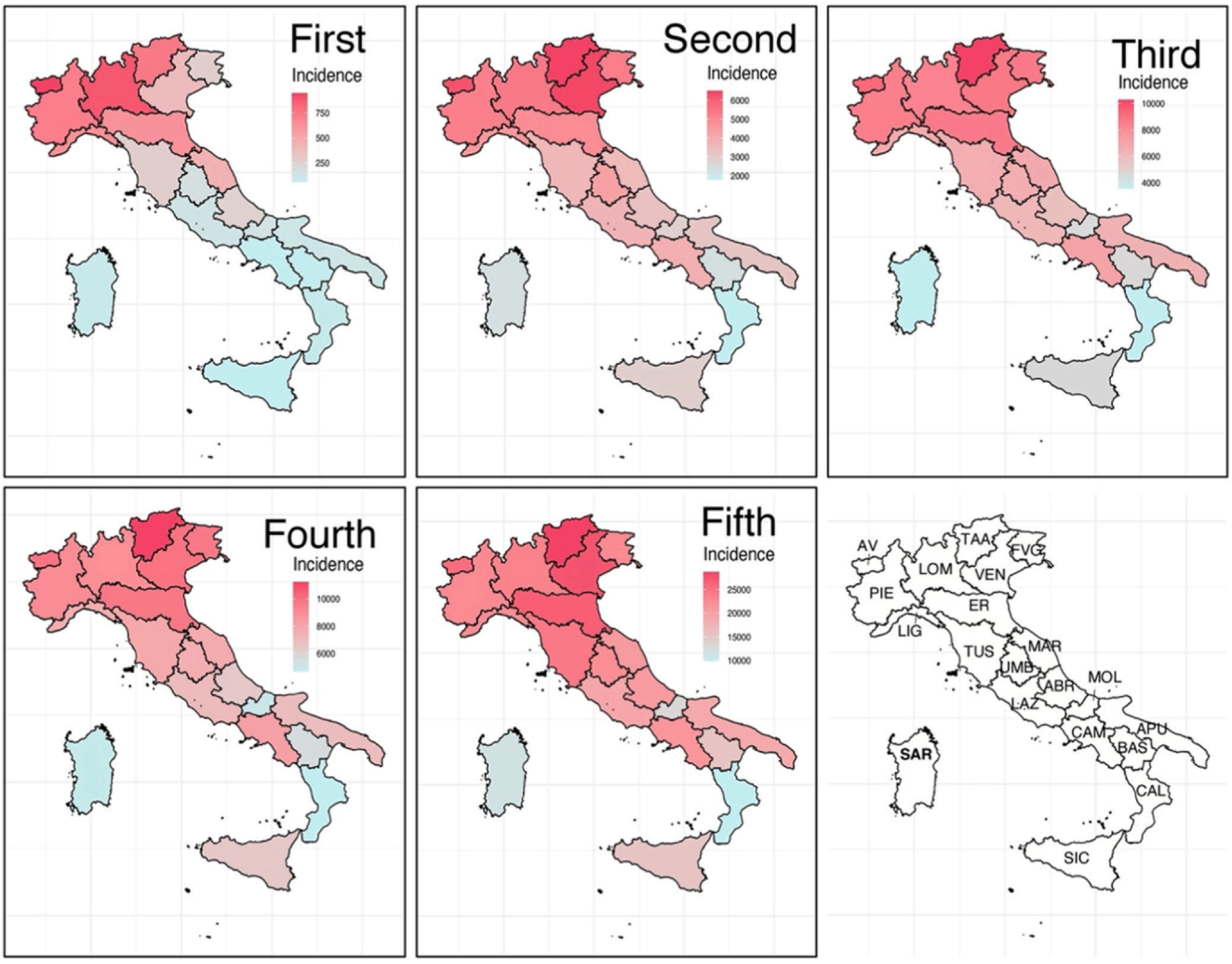
Between January 2020 and March 2023, Sardinia registered 527.670 cases for 1.587.413 inhabitants (prevalence=33 %) while a severely affected area such as Lombardy counted 4.125.708 cases for 9.943.004 inhabitants (prevalence=41 %) (inhabitants source: latest report of National institute for Statistics). Also, the temporal distribution of COVID-19 cases in Sardinia showed substantial differences as compared to the individual Italian regions and the overall national trend (Figs. 1 and 2). The first pandemic phase (February – May 2020) had a slightly delayed start in Sardinia, with the first imported case at the beginning of March and local cases increasing from mid-month. In this period, Sardinia and Sicily islands were among the 5 Italian administrative areas with the lowest COVID-19 cumulative incidence, according to an evident decreasing North-to-South gradient (Fig. 1). The second phase (October 2020 – January 2021) had instead an earlier start in Sardinia, with an additional peak in the second half of summer 2020 and an expanded fall-winter peak (Fig. 2). Accordingly, SARS-CoV-2 cumulative incidence had a slight increase, still remaining the second lowest one (Fig. 1). The third phase (February-May 2021) looks more similar to the national panorama, likely due to the positive impact of the vaccination campaign: during this period Sardinia had the lowest case incidence in Italy (Figs. 1 and 2). As for 2020, the summer period June-October 2021 (fourth phase) in Sardinia was associated to an additional peak of cases, in contrast to an almost flat epidemiological curve in Italy and Lombardy, with a slight increase in cumulative incidence that returned the second lowest one (Figs. 1 and 2). Finally, the fifth phase (November 2021 – February 2022) was associated to Omicron VoC, which had a lower virulence but a very high transmissibility and led to a pronounced increase in COVID-19 cases, becoming rapidly dominant [22]. Accordingly, in Italy and Lombardy the peak was reached in ~1 month, while Sardinian trend showed a correspondent initial peak followed by a temporary slowdown, with a plateau of cases that started to increase again in February 2022 (Fig. 2). After the end of the fifth phase, two subsequent peaks were observed until the end of summer 2022, reaching only in Sardinia a comparable height to the previous.

3.2. SARS-CoV-2 genomes in the study

To investigate the diversity of SARS-CoV-2 in Sardinia, within CarGen4CoV collaborative project we sequenced 888 SARS-CoV-2 genomes (“CarGen dataset”) from nasopharyngeal swabs collected between February 2020 and December 2021 (23 months) by three clinical centers (Supplementary Fig. 1) that were strategically located to assure a good coverage of SARS-CoV-2 genomes since the beginning of COVID-19 diffusion, i.e before than the regional and national surveillance made epidemiological data publically available. In addition, we downloaded from GISAID all 1439 high-coverage SARS-CoV-2 genomes collected in Sardinia along 38 months of pandemic (December 2019-January 2023, “GISAID dataset”) (Supplementary Table 1).

Hence, considering both datasets, a total of 2327 complete SARS-CoV-2 genomes were included (Table 1). Noteworthy, for some clades, the sequences from CarGen dataset represent the majority of the reported SARS-CoV-2 genomes, with a major contribution to their characterization in Sardinia (Table 1). For example, the major European clade 20E included 581 genomes, of which 457 (79 %) were sequenced during the present study. Indeed, the clades not found among CarGen genomes were either rarely found in Sardinia (e.g 20 H/β VoC, 21D/η, 21 F/ι, and 21 H/μ) or occurred after the end of CarGen4CoV sampling period (as for ο VoC clades subsequent to 21 K) (Supplementary Table 2).

A



B

first		second		third		fourth		fifth	
Campania	59,42	Calabria	1781,52	Sardinia	3545,06	Calabria	4658,56	Calabria	9807,38
Sicily	61,42	Sardinia	2345,59	Calabria	3556,64	Sardinia	4790,5	Sardinia	11731,26
Basilicata	64,67	Basilicata	2362,21	Molise	4587,13	Molise	4950,99	Molise	13153,73
Calabria	78,82	Molise	2833,79	Sicily	4692,95	Basilicata	5547,51	Sicily	15319,13
Sardinia	82,28	Sicily	2843,36	Basilicata	4732,83	Sicily	6467,01	Basilicata	15500,78
Apulia	111,66	Apulia	3167,1	Abruzzo	5763,37	Abruzzo	6470,07	Apulia	18891,62
Lazio	132,32	Abruzzo	3338,02	Lazio	6129,36	Apulia	6960,6	Lazio	19316,86
Molise	142,66	The Marches	3570,57	Apulia	6383,1	Lazio	7084,19	Abruzzo	20660,4
Umbria	162,24	Tuscany	3649,61	Tuscany	6571,51	Umbria	7619,77	The Marches	20826,48
Abruzzo	248,17	Lazio	3692,39	Umbria	6576,9	Liguria	7629,79	Campania	21292,42
Tuscany	265,63	Campania	3847,12	The Marches	6766,32	The Marches	7736,55	Umbria	22148,11
Friuli Venezia Giulia	269,33	Umbria	4307,57	Liguria	6816,84	Tuscany	7905,61	Liguria	22404,35
Veneto	390,72	Liguria	4595,28	Campania	7090,46	Campania	8259,66	Friuli Venezia Giulia	22637,02
The Marches	440,58	Emilia Romagna	4940,95	Lombardy	8372,52	Lombardy	8975,44	Piedmont	22674,66
Emilia Romagna	624,02	Piedmont	5222,75	Piedmont	8474,43	Piedmont	9067,28	Lombardy	23495,3
Liguria	628,51	Friuli Venezia Giulia	5272,52	Emilia Romagna	8635	Friuli Venezia Giulia	9689,9	Tuscany	24318,81
Piedmont	714,81	Lombardy	5384,04	Friuli Venezia Giulia	8636,89	Emilia Romagna	9740,21	Aosta Valley	25471,57
Lombardy	886,68	Aosta Valley	6247,9	Veneto	8742,65	Aosta Valley	9894,67	Emilia Romagna	26715,73
Aosta Valley	949,34	Veneto	6423,07	Aosta Valley	9307,07	Veneto	9912,95	Veneto	27473,69
Trentino	746,02	Trentino	6504,78	Trentino	10300,6	Trentino	11245,22	Trentino	28714,86

Fig. 1. Cumulative incidence of COVID-19 cases during the considered pandemic phases. The cumulative incidence (for 100.000 inhabitants) of reported COVID-19 cases in the different Italian regions is plotted in the map with a color gradient (pink-high to ciano-low) according to the 5 pandemic phases considered in the study (as defined in Fig. 1) (A). The correspondent data are also reported and were used to sort the Italian regions according to the increasing cumulative incidence (B).

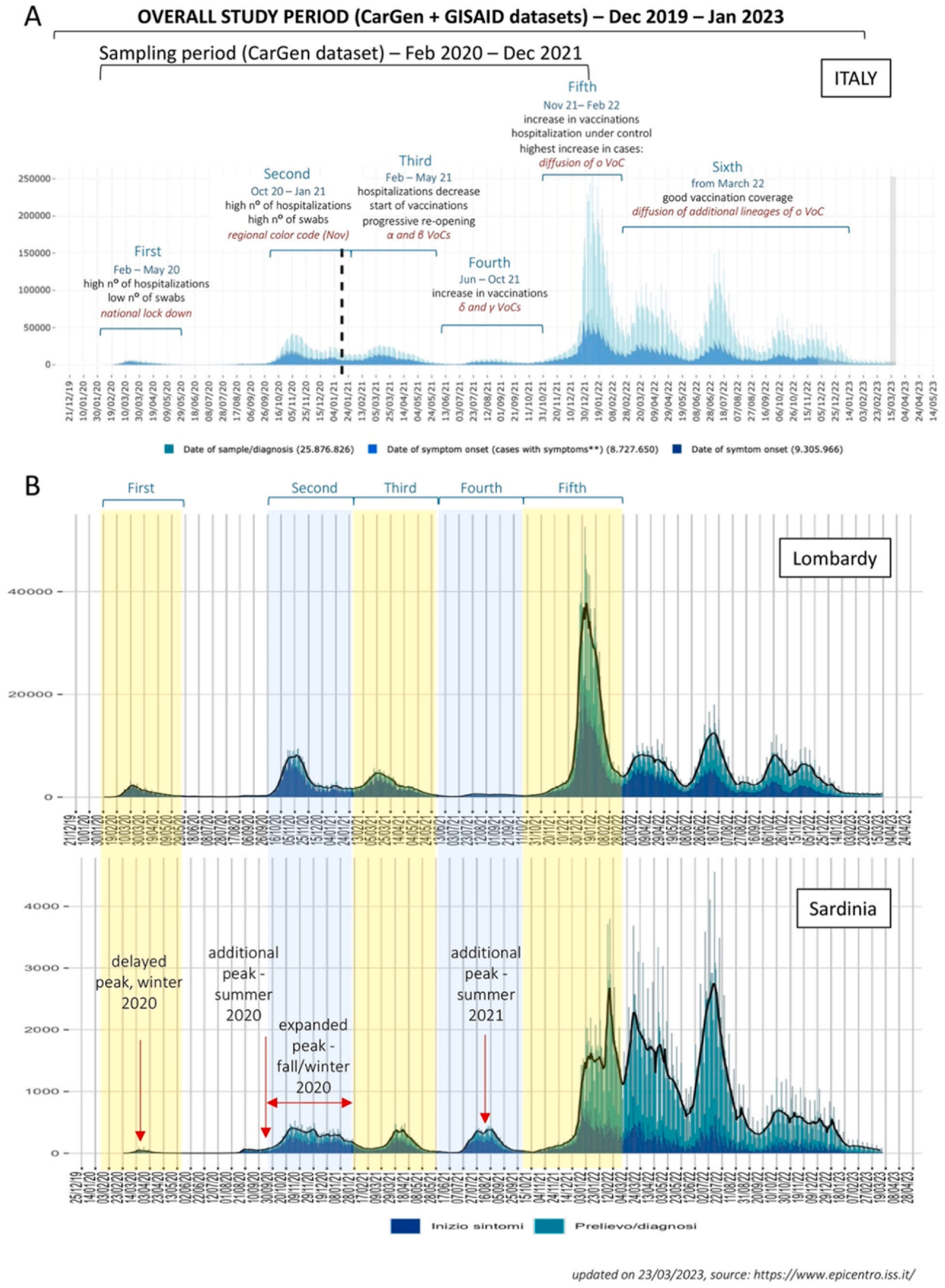


Fig. 2. Epidemiological trend of COVID-19 pandemics The bar plots depict the dynamics of COVID-19 cases as reported in Italy (A) and in two Italian regions (B) including a highly affected one (Lombardy, 4.125.708 cases) and Sardinia (527.670 cases). The overall time period considered in the present study is indicated in the top of the figure, while the different pandemic phases are highlighted with blue brackets and colored boxes. The main differences of Sardinian epidemiological curve as compared to the national and Lombardy ones are indicated with red arrows. All plots are adapted from the ones provided by the Italian Institute for Public Health website https://www.epicentro.iss.it/, accessed on 23/03/2023.

Table 1
Number and clades of the SARS-CoV-2 genomes circulating in Sardinia, as generated in this study and retrieved from Gisaid database.

clade*	pandemic wave**	This study (Feb 2020 - Dec 2021, 23 months)	GISAID – Sardinia (Dec 19 – Jan 23, 38 months)	Total n° of sequences
19 A	First	1 (50 %)	1	**2**
19B	First	6 (40 %)	9	**15**
20 A	First	22 (46 %)	26	**48**
20B	First	29 (57 %)	22	**51**
20 C	Second	7 (87 %)	1	**8**
20D	Second	13 (81 %)	3	**16**
20E	Second	457 (79 %)	124	**581**
20I (Alpha)	Third	113 (14 %)	694	**807**
20 H (Beta)	Third	-	16	**16**
20 J (Gamma)	Fourth	1 (8 %)	12	**13**
21 A, 21I and 21 J (Delta)	Fourth	199 (37 %)	345	**541**
21D (Eta)	Fourth	-	1	**1**
21 F (Iota)	Fourth	-	2	**2**
21 H (Mu)	Fourth	-	5	**5**
21 K, 21 L, 22 A, 22B, 22E (Omicron)	Fifth	40° (33 %)	80	**120**
CRF	-	-	101	**101**
TOTAL		888 (38 %)	1439	**2327**

°includes only 21 K (sampling ended in December 2021)
For the SARS-CoV-2 genomes generated in this study, the percentage as referred to the total number of sequences in Sardinia (this study + Gisaid) is reported between brackets
* as assigned with Nextclade
** as defined in the present study. First: February - May 2020; second: October 2020 - January 2021; third: February - May 2021; fourth: June - October 2021; fifth: November 2021 - February 2022; sixth: from March 2022

3.3. Phylogeny of SARS-CoV-2 genomes in Sardinia

To define the phylogeny of SARS-CoV-2 genomes, we built a Maximum Likelihood tree including CarGen and GISAID datasets, identifying 64 supported clusters, (Fig. 3, Supplementary Fig. 2).

We found only 17 sequences from 19 A and 19B haplotypes (1 %, cluster 37), responsible for the beginning of SARS-CoV-2 diffusion in December 2019, in line with the low incidence of the infection in Sardinia during the first pandemic wave. An analogous situation was observed for clade 20 A (2 %, subcluster 41a), whose origin broadly corresponds to the northern Italy outbreak in early 2020, and clade 20B (2 %, subcluster 38a), characterized by the onset of three single nucleotide substitutions (G28881A, G28882A, G28883GC) (Table 1, Fig. 3). Similarly, in the second wave, a minority of sequences belonged to clades 20 C (0,3 %, subcluster 41c), which led to the differentiation of the beta, iota and epsi variants; and 20D (0,6 %, subcluster 38b), spread mostly in England (Fig. 3). Clusters 38 and 41 also includes some minor lineages derived from the above clades and mainly found in Northern Europe, such as B.1.160 and B.1.389 (41a with 20 A) and B.1.1.232 (38a with 20B) (Fig. 3). The fact that different lineages form shared clusters likely depends on the fact that most spike mutations appeared later on, increasing inter-lineage diversity.

The major European clade 20E formed the second supercluster in terms of members, with 581 SARS-CoV-2 genomes divided in 21 clusters (16 to 36) (Table 1, Fig. 3). The size of this cluster reflects the impact of summer 2020, when the reopening and relaxation of social isolation measures drastically changed the pandemic trend of the island, which until that moment was only marginally affected (Fig. 2). In particular, sequences belonging to the main lineage B.1.177 (n = 363) were distributed in 9 clusters, while the descendent Italian lineages B.1.177.75 (n = 122) and B.1.177.83 (n = 93) - circulating from August - September 2020 - were grouped in 8 (24–31) and 3 (21–23) clusters, respectively (Fig. 3).

The alpha VoC (20I/B.1.1.7) was firstly reported in England in September 2020 and resulted the most spread in Sardinia, in line with the increase of infections in fall 2020 third wave (Fig. 2), including 807 genomes (Table 1). These sequences are divided in 24 clusters, sharing similar times of diffusion (January-June 2021) (Fig. 3).

Contrarily, in the fourth wave, beta and gamma VoCs showed a minor presence in Sardinia, in line with their low diffusion in the whole Country (Table 1) [23]. The delta VoC is indeed the second most represented VoC in Sardinia, forming a supercluster that includes 11 supported clusters with 544 genomes, classified into 3 clades (21 A, 21I, 21 J) and 46 lineages (Fig. 3). The delta VoC was identified in India from spring 2021 and rapidly spred due to multiple spike mutations. Its prevalence on the territory is in line with the increase in COVID-19 cases in Sardinia in spring-summer 2021 (Fig. 2).

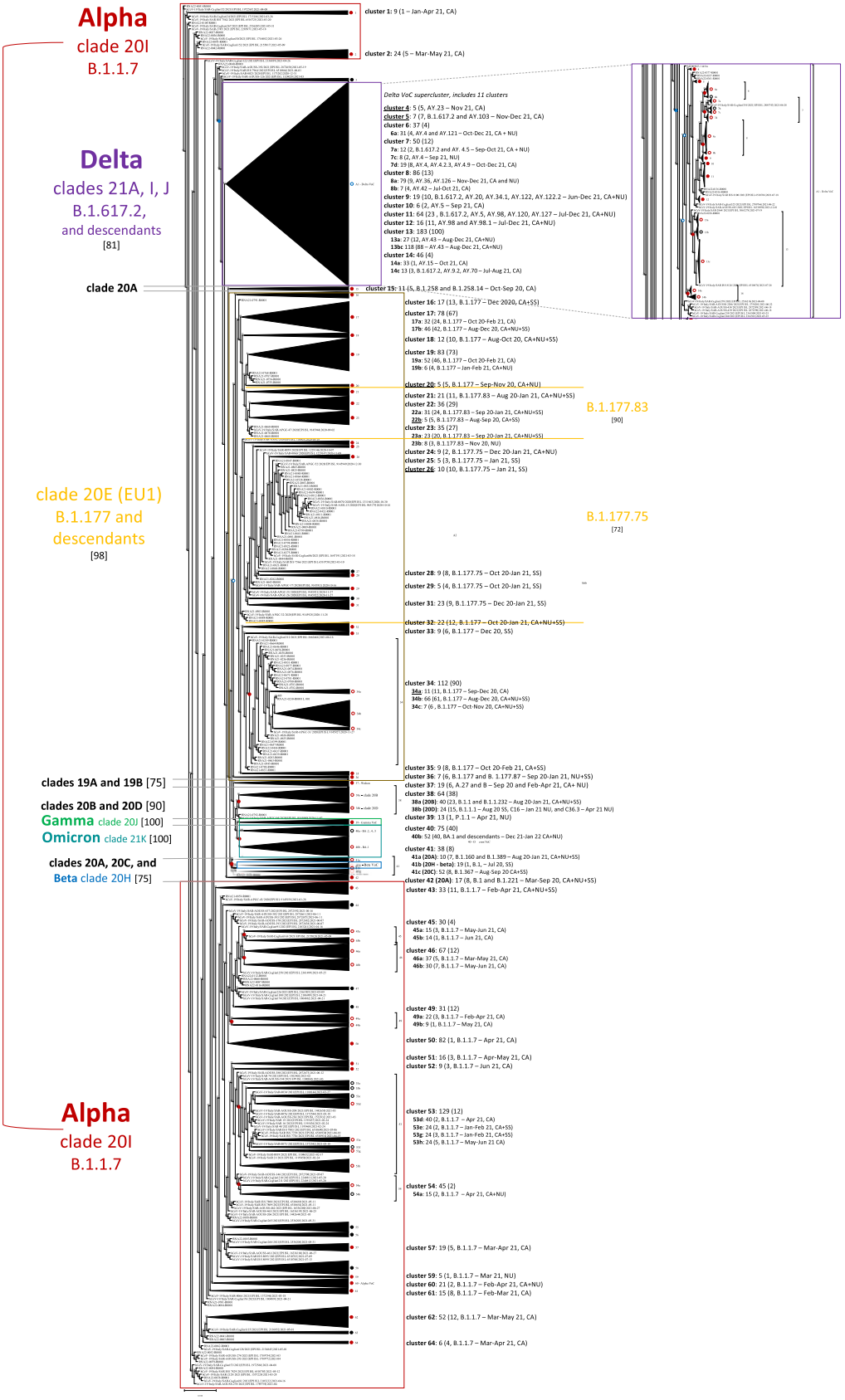
Finally, in the fifth wave, omicron VoC has been associated with the highest contagiousness, showing in turn a lower virulence that contributed to its massive spread in Europe and Italy from autumn-winter 2021. In this case, its minority presence in CarGen population is because its spread began in South-Africa at the end of CarGen sampling period (November 24th 2021) that included hence only the first sub-lineage (BA.1.14, clade 21 K) (Table 1). Nevertheless, it represents the 5 % of the Sardinian sequences (120 sequences, clades 21 K, 21 L, 22 A, 22B, and 22E), confirming a wide diffusion on the territory (Table 1).

3.4. Relative frequency

To gain more insights about SARS-CoV-2 diversity, we generated frequency plots showing the relative percentage of each lineage and clade from February 2020 to December 2021 (Fig. 4). Moreover, given the hypotesis that such distribution could be influenced by insularity, we performed the same analysis for the other major Italian island (Sicily) and a highly affected non-insular region (Lombardy) (Fig. 4). Finally, we generated frequency plots for other 4 major Mediterranenan Islands: Corse (France), Crete (Greece), Mallorca (Spain), and Malta (Supplementary Fig. 3). While Sardinian plot included both CarGen and GISAID datasets, the others were based on GISAID data only, and present some missing period during which no sequences were uploaded.

Interestingly, Sardinia plots shows a peculiar distribution of circulating lineages and clades. First, certain lineages were found here but not in the other areas: B.1.22, B.1.36, and B.1.558 (clade 20 A); B.1.1.25, B.1.1.198, and B.1.1.294 (20B); C.16 (20D); B.1.177.83 and B.1.177.87 (20E); AY.34.1.1, AY.78, and AY.90 (21 J, delta VoC); and BA.1.14 (21 J, omicron VoC) (Fig. 4, Supplementary Fig. 3). The global presence of these lineages was assessed in GISAID (December 2019 – January 2023) and some were even not reported in Italy (B.1.22 and B.1.1.198, mostly in Netherlands and UK, respectively) or confirmed to be present in Sardinia region only (B.1.558 and C.16, diffused in North America/Africa and France, respectively) (Table 2). Importantly, various lineages were not present in Sardinia according to GISAID data, and were hence firstly reported thanks to CarGen dataset (B.1.22, B.1.36, B.1.1.198, B.1.1.294, AY.34.1.1, and BA.1.14) (Table 2). On the counterpart, entire clades were not present in Sardinia, including 20 F (in Sicily, lineage D.2) as well as 20 G and 21 G/lambda (in Lombardy, lineages B.1.2 and C.37/C37.1, respectively) (Fig. 4). The search of such clades in GISAID confirmed their low circulation in Italy - with only 1 (20 F-D.2), 3 (20 G-B.1.2), and 19 (21 G-C.37/C.37.1) enties (data not shown) - and in the analyzed Mediterranean Islands, except for a single 20 G-B.1.2 entry in Mallorca (Supplementary Fig. 3).

Second, Sardinia showed lineages with higher prevalence that in some months made up almost the totality of sequences, with a less heterogeneous composition as compared to Sicily and Lombardy:



(caption on next page)

Fig. 3. Phylogenetic tree of SARS-CoV-2 genomes circulating in Sardinia The full-length nucleotide sequences of the 888 SARS-CoV-2_CarGen genomes plus all the sequences reported in Gisaid for Sardinia since the beginning of the pandemics (December 2019 – January 2023, 38 months) were used to generate a neighbor joining phylogenetic tree. The latter is presented here in a condensed version and is provided as full-length figure in Supplementary Material (Fig. S2). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) are shown next to the branches. The evolutionary distances were computed using the p-distance method and are in the units of the number of nucleotide differences per site. Reference SARS-CoV-2 genomes for the original haplotype and the variants of concerns (VoCs) are also included. Phylogenetic groups corresponding to the different VoC clades are highlighted with colored boxes, and their bootstrap support is reported between squared brackets. Cluster and subclusters including sequences from this study are marked with filled and unfilled red dots, respectively, and are annotated with progressive numbers and a brief description reporting the number of total members and, between round brackets, the number of sequences from our study along with the corresponding lineages, time and site of sampling. Clusters and subclusters formed exclusively by sequences from the present study are underlined. Criteria used to define a cluster or subcluster were a threshold of bootstrap support of at least 70 and a minimum number of members of 5 sequences.

B.1.177 (clade 20E) during Summer-Fall 2020 and the alpha VoC (B.1.1.7, 20I) in Winter-Spring 2020–2021 (Fig. 4). Then, the delta VoC arrival fastly led to the prevalence of 21 A, 21I, and 21 J clades from Summer 2021, albeit showing a lower heterogenicity of circulating lineages in Sardinia (n = 45) as compared to Sicily (n = 63) and Lombardy (n = 93) (Fig. 5, Supplementary Table 3). Most of such lineages had a relative frequency < 1 %: in Sardinia, 11 lineages accounted for the 89 % of all delta sequences, while in Sicily and Lombardy 12 and 14 lineages made up the 88 % and 83 % of the total delta members, respectively (Supplementary Table 3). Six of these were shared by the three regions: while the founder B.1.617.2 lineage showed a comparable presence (10 % in Sardinia and Sicily, 15 % in

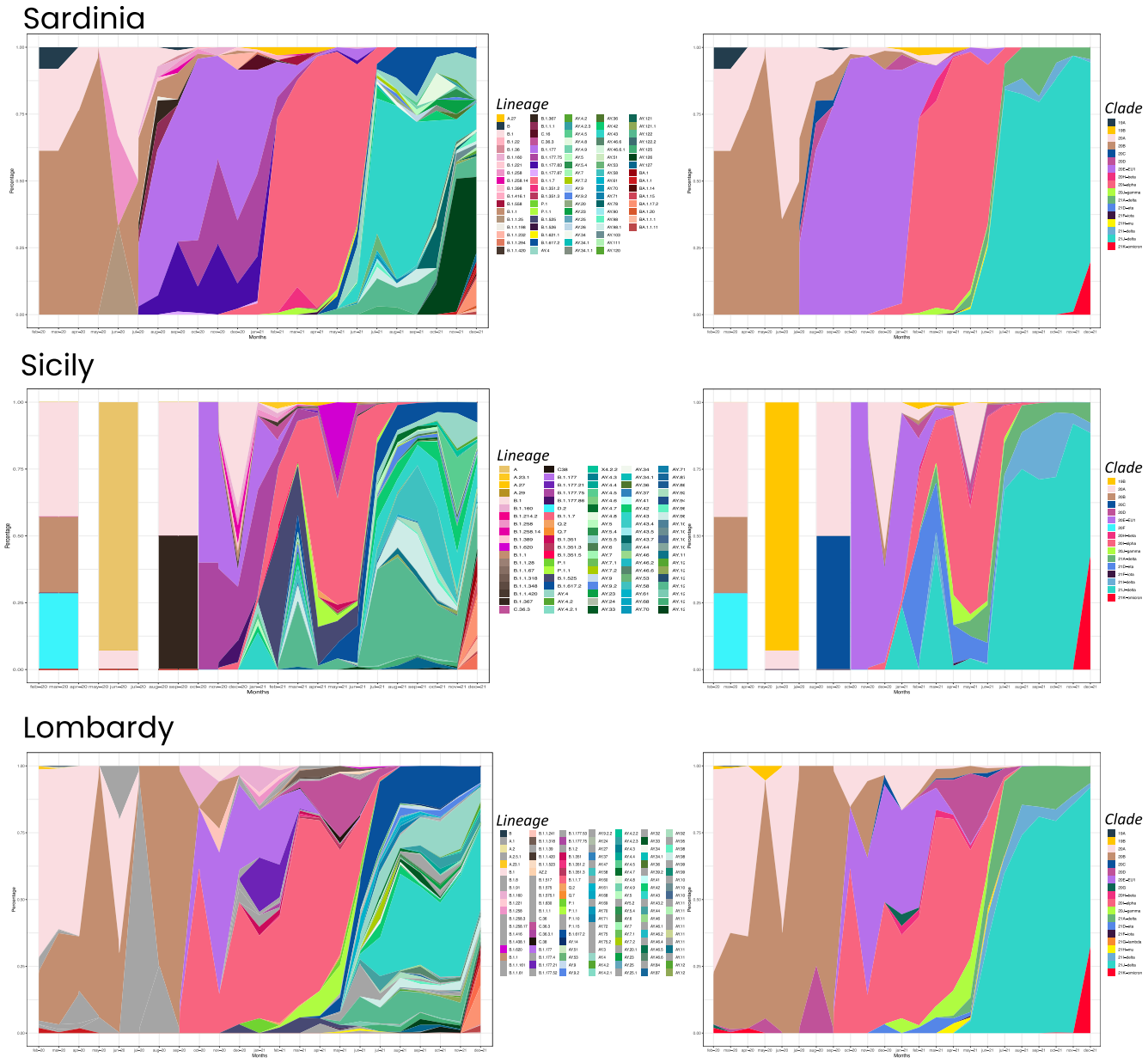


Fig. 4. Frequency plots of SARS-CoV-2 lineages The relative frequency of the SARS-CoV-2 lineages circulating in Sardinia, Sicily, and Lombardy regions have been plotted along 23 months of pandemics, corresponding to the sampling period of the CarGen4CoVstudy (February 2020 – December 2022). Frequencies are expressed as percentage among the total of cases, as retrieved from GISAID database and from our dataset in the case of Sardinia. For Lombardy, lineages being not present in the two main islands have been represented in grey, to facilitate the comparison.

Table 2
SARS-CoV-2 lineages found only in Sardinia.

clade*	wave**	Lineage*	Sardinia		Italy and rest of the World			
			This study[a]	*GISAID*[b]	*First report*	*Main diffusion*	*n° of sequences (Italy)*	*n° of sequences (global)*
20 A	1st	B.1.22	1	-	03/2020	Netherlands (698)	0	851
		B.1.36	1	-	02/2020	India (4438), North America (2435)	5 (Lazio, Apulia)	9258
		B.1.558	-	7	04/2020	North America (336)	7 (Sardinia)	343
20B	1st	B.1.1.198	1	-	03/2020	UK (461)	0	626
		B.1.1.25	-	1	04/2020	Bangladesh (1230), North America (755)	26 (Lazio, Emilia-Romagna, Apulia, Campania)	2350
		B.1.1.294	1	-	03/2020	Russia (235), Denmark (148), Germany (148)	5 (Emilia-Romagna, Veneto)	906
20D	2nd	C.16	5	2	11/2020	Africa (278), France (167), North America (128), Portugal (121)	2 (Sardinia)	1067
20E	2nd	B.1.177.83	68	25	07/2020	Italy (542), Germany (129), Croatia (73)	542 (Campania, Abruzzo, Lazio, Sardinia, Emilia-Romagna, Apulia)	792
		B.1.177.87	2	1	08/2020	UK (1281)	5 (Campania, Sardinia, Apulia, Emilia-Romagna)	1373
21 J (Δ)	4th	AY.34.1.1	1	-	06/2021	South America (2555)	8 (Apulia)	2667
		AY.78	-	9	06/2021	Turkey (238), Germany (151)	1 (Emilia-Romagna)	487
		AY.90	-	1	05/2021	UK (1838)	2 (Sardinia, Lazio)	1875
21 K (O)	5th	BA.1.14	7	-	10/2021	Belgium (2446), UK (1258), Denmark (920), Germany (753)	32 (Lazio, Veneto, Campania, Lombardy, Piedemont, Trentino)	11621

* as assigned with Nextclade
** as defined in the present study. 1st: February - May 2020; 2nd: October 2020 - January 2021; 3rd: February - May 2021; 4th: June - October 2021; 5th: November 2021 - February 2022; 6th: from March 2022
[a] sampling period: February 2020 - December 2021, 23 months
[b] search period: December 2019 – January 2023, 38 months

Lombardy), the prevalent delta lineage was AY.43 in Sardinia and Lombardy (34 % and 32 %, respectively, vs < 12 % in Sicily) and AY.122 in Sicily (30 % vs 10 % and 8 % in Sardinia and Lombardy, respectively) (Fig. 5). The remaining lineages were instead specifically found in one or two regions. Whereas in Lombardy they all showed individual percentages < 2,5 %, in Sardinia AY.126 was the second most abundant (15 % vs 1 % and 0,9 % in Lombardy and Sicily, respectively), while in Sicily lineage AY.58 was the fourth in prevalence (8 % vs 0,1 % in Lombardy, not reported in Sardinia) (Fig. 5).

4. Discussion

Viral spread is influenced by several variables, including fitness and socio-geographical factors. Islands are interesting sites to study viral evolution, as their geography reduces the complexity of population connections, especially in the presence of restrictive measures. Here we explored SARS-CoV-2 diversity in sardinia island along 38 out of 42 months of pandemics (December 2029- January 2023). Some previous studies provided a snapshot of SARS-CoV-2 genetic diversity in Sardinia, including 13 (March-May 2020, central inner Sardinia [3]); 92 (March 2020-May 2021, central inner Sardinia [5]); and 55 (March 2020-January 2021, North Sardinia [4]) SARS-CoV-2 genomes. Our study considered instead a total of 2327 SARS-CoV-2 genome sequences circulating in the different Sardinia areas, as obtained through the sequencing of clinical samples from three main clinical centres of the island (located in Cagliari, Sassari, and Nuoro) as well as from GISAID repository. The study however has some limitations: despite the high number of SARS-CoV-2 genomes analysed (of which the ones sequenced in this study represent the 38 %), the obtained dataset does not cover all the viral sequences that circulated in Sardinia, and especially the lineages with a low prevalence, that remain underrepresented or even undetected. Similarly, the sampling period ended at the beginning of the omicron variant spread, and is hence poorly representative of this variant's diversity, being accounted only by the GISAID sequences. About the latter, it is difficult to obtain specific information on the composition of the different clusters of infection within the territory, and hence to track the moment of entry and the dynamics of diffusion of the various viral clades. Nevertheless, our study represents the most updated and comprehensive description of SARS-CoV-2 diversity in Sardinia, and provides useful information about its diffusion in an insular territory.

In general, SARS-CoV-2 incidence in Sardinia has been lower as compared to the regional and national panorama, likely due to insular geography and other socio-geographic features, including large rural areas, limited connectivity between major cities and low population density [4,5]. This trend is particularly evident in the first months, coincident with the national lockdown: the contenitive measures in Sardinia were strengthened by the natural isolation, with a lower and delayed peak of infections sustained by the original haplotype 19 A. Then, the progressive re-opening policy caused two additional peaks in Summer seasons 2020 and 2021, plus an expanded peak in Fall-Winter 2020. These summer peaks were linked to the high amount of turistic arrivals from Italy and Europe, as also suggested by the epidemiology of COVID-19 cases.

Particularly, the main European clade 20E/EU1 (Summer 2020) had a key role in increasing Sardinian cumulative incidence, and forms the major cluster in SARS-CoV-2 phylogeny, as previously reported for the central-inner and northern part of the Island [4,5]. This outbreak was claimed to be responsible for the start of the second wave in Italy through tourists returning from the Island, but previous works found such hypotesis unlikely [4]. Accordingly i) two of the three main lineages found in Sardinia (B.1.177 and B.1.177.75) circulated in mainland prior to reach the Island, making more likely their importation through national touristic arrival, and ii) some lineages present in Sardinia were absent in Italy, including B.1.22 and B.1.558 (clade 20 A, Netherlands and North America, respectively) and B.1.1.198 (clade 20B, UK and Luxembourg). Also other lineages occasionally found in Sardinia had a very low circulation in the rest

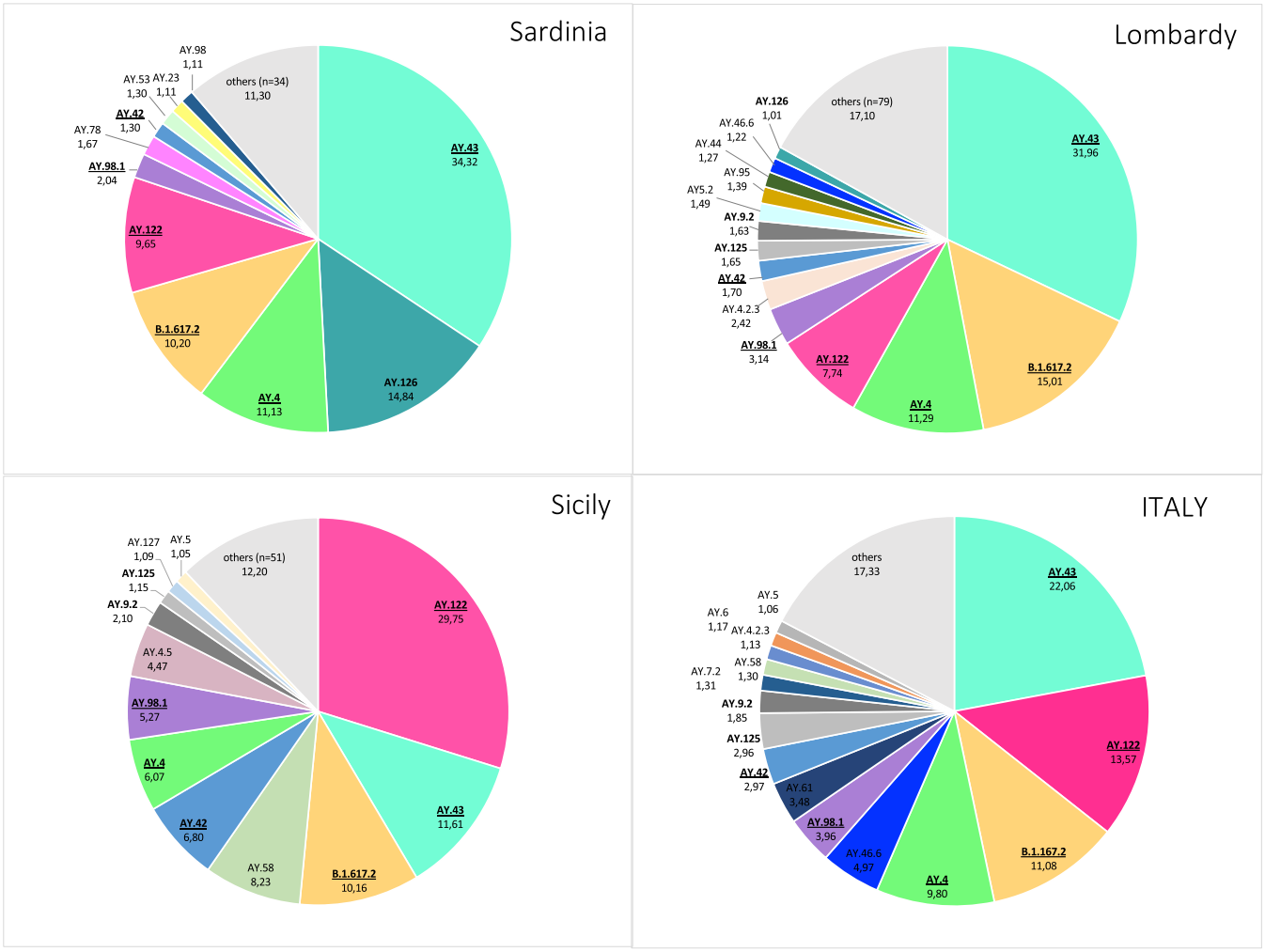


Fig. 5. Delta lineage composition The plots show the composition of the Delta VoC in Sardinia, Sicily, Lombardy and at the national level, indicating the percentage of the different lineages along 23 months of pandemics, corresponding to the sampling period of the study (February 2020 – December 2022). The names of lineages that were found in all the three Italian regions are underlined and in bold, the ones shared between two regions are in bold. Lineages representing less than the 1 % are grouped under the category "others".

of Italy, further suggesting their independent introduction by the resumption of European travels from June 2020 [24,25].

The increase in infections in winter 2020 was associated to the alpha VoC, rapidly became prevalent thanks to its increased transmissibility [26], costituting ~35 % of Sardinian sequences. The corresponding phylogenetic clusters are splitted in the upper (clusters 1 and 2) and lower (clusters 43–64) parts of the tree based on the presence of 3 nucleotide substitutions in all members of the two upper clusters. While one is a synonymous mutation in nsp15/endoRNAse (C20148T), two non-synonymous substitutions (C16468T/P78S and T17009G/I258S) were found in nsp13/helicase. P78S was reported in USA in March 2020, but had a major diffusion in Italy in March-May 2021 [19], in line with the date of CarGen sampling; while I258S was firstly described in Saudi Arabia in April 2020, and reported in Italy in March- May 2021 [19]. Interestingly, when we searched the above mutations in GISAID (accessed on August 24th 2023), the 60 % of entries harboring P78S mutation were from Sardinia (93/154), with the remaining distributed in other 12 Italian regions; while I258S was found in 107 Italian sequences, of which 96 % from Sardinia, with only 4 sequences in 3 other regions (data not shown). This mutation was confirmed to be specifically found in Sardinia also at the global level, with only 85 sequences in the rest of the Countries, being likely arose from the intensive circulation of apha VoC during summer-fall 2020. Another lineage strictly associated to the Island was A.27 (19B): arose in Spain, it was firstly observed in Italy in a Sardinian family in January 2021 [27], and its subsequent diffusion was limited to Sardinia and Sicily (data not shown).

Summer 2021 peak was linked too to the numerous tourist entrances, favoring the spread of the highly-infectious delta VoC, i.e. the third cluster in terms of numbers and the most complex for phylogenetic composition.

Beside the major deviations from the national curve, Sardinia also showed a peculiar frequency of SARS-CoV-2 lineages. The previous works reported 10 to 15 lineages in Sardinia [4,5]: according to our study, a total of 21 clades and 87 lineages were circulating in the Island between December 2019 and January 2023, including some clades that would have been otherwise underestimated or even not reported based on the sole GISAID dataset. The completeness of the study dataset allowed to generate frequency plots depicting the relative percentages of lineages and clades, and compare it to a highly affected region (Lombardy) and the other major island (Sicily). Sardinia showed again a peculiar trend, with the complete replacement of one clade by the subsequent, indicating a low co-circulation of different clades. The first major replacement occurred between co-circulating clades ^{20}A–^{20}B and clade 20E (July 2020), then replaced by the alpha VoC (January 2021) that was itself cleared by the Delta clades (June 2021). A comparable sweep between lineages was reported for some oceanic islands [9] and observed in the other Mediterranean islands, despite the presence of whole months with no sequences available, further suggesting that SARS-CoV-2 lineage

composition in isular territories presents common signatures, especially during the months of turistic arrivals, but also maintains a certain degree of specificity, being likely influenced by the respective mainlands.

The delta variant showed a complex phylogeny due to the progressive accumulation of substitutions [28]. The lower eterogenicity was observed in Sardinia, with 45 lineages against 93 in Lombardy; while Sicily showed an intermediate situation, likely reflecting its partial isolation. In fact, Sicily has hystorically more exchanges with the nearby peninsula and, importantly, is a main point of arrival for African migratory routes, representing a costant movement of people despite travel restrictions [8,29]. Finally, our data suggest that insular geography allowed the expansion of imported lineages that remained rare at the national level, as reported for other islands [9].

5. Conclusions

Our results demonstrates that the genomic diversity of SARS-CoV-2 in Sardinia has been shaped by a complex interplay of insular geography, low population density, and touristic arrivals, as seen for other islands [12,17,18]. Insular areas, above all the ones with touristic vocation, are on the one side an hotspot for the import of lineages that remain rare or absent in mainland, with a potential key role in their evolution [9]. On the other side, especially with restrictive measures, islands can constitute bottlenecks resulting in the delayed entry of otherwise common variants [12,30].

Ethical Approval statement

All samples were collected as part of clinical diagnostic routine following official procedure (ISS Working group Diagnostics and microbiological surveillance of COVID-19). All samples and data collected are directly related to Italian pandemic control and were previously anonymised as required by the Italian Data Protection Code (Legislative Decree 196/2003) and the general authorisations issued by the Data Protection Authority.

Funding source

The present work has been funded by Sardegna Ricerche agency within the collaborative research project “CarGen4CoV: Genetic characterization of SARS-CoV-2 circulating in Sardinia”, grant n. F24I20000190002.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The present work was supported by Sardegna Ricerche Biomedical Research Support Unit. We would like to thank Franco Cappai, Caterina Carboni, and Giuseppe Serra for their project management.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jiph.2024.102496.

References

[1] Prezioso C, Pietropaolo V. COVID-19: update of the Italian Situation. J Neurovirol 2020;26:834–7. https://doi.org/10.1007/s13365-020-00900-w
[2] Prezioso, C.; Marcocci, M.E.; Palamara, A.T.; De Chiara, G.; Pietropaolo, V. The “Three Italy” of the COVID-19 Epidemic and the Possible Involvement of SARS-CoV-2 in Triggering Complications Other than Pneumonia., doi:10.1007/s13365–020-00862-z/Published.
[3] Piras G, Grandi N, Monne M, Asproni R, Fancello T, Fiamma M, et al. Early diffusion of SARS-CoV-2 infection in the inner area of the Italian Sardinia Island. Front Microbiol 2021;11. https://doi.org/10.3389/fmicb.2020.628194
[4] Rocchigiani AM, Ferretti L, Ledda A, Di Nardo A, Floris M, Bonelli P, et al. Origin, genetic variation and molecular epidemiology of SARS-CoV-2 strains circulating in Sardinia (Italy) during the first and second COVID-19 epidemic waves. Viruses 2023;15. https://doi.org/10.3390/v15020277
[5] Malune P, Piras G, Monne M, Fiamma M, Asproni R, Fancello T, et al. Molecular characterization of severe acute respiratory syndrome coronavirus 2 isolates from central Inner Sardinia. Front Microbiol 2022;12. https://doi.org/10.3389/fmicb.2021.827799
[6] Puci MV, Loi F, Ferraro OE, Cappai S, Rolesu S, Montomoli C. COVID-19 trend estimation in the elderly Italian Region of Sardinia. Front Public Health 2020;8. https://doi.org/10.3389/fpubh.2020.00153
[7] Rocchigiani AM, Ferretti L, Ledda A, Di Nardo A, Floris M, Bonelli P, et al. Origin, genetic variation and molecular epidemiology of SARS-CoV-2 strains circulating in Sardinia (Italy) during the first and second COVID-19 epidemic waves. Viruses 2023;15. https://doi.org/10.3390/v15020277
[8] Padilla-Blanco M, Gucciardi F, Guercio A, Rubio V, Princiotta A, Veses V, et al. Pilot investigation of SARS-CoV-2 variants in the Island of Sicily Prior to and in the second wave of the COVID-19 pandemic. Front Microbiol 2022;13. https://doi.org/10.3389/fmicb.2022.869559
[9] Wilkinson DA, Mercier A, Turpin M, Simbi MA, Turpin J, Lebarbenchon C, et al. Genomic evolution of SARS-CoV-2 in Reunion Island. Infect, Genet Evol 2022;106. https://doi.org/10.1016/j.meegid.2022.105381
[10] Ha YR, Kim HJ, Park JS, Chung YS. Genomic surveillance of genes encoding the SARS-CoV-2 spike protein to monitor for emerging variants on Jeju Island, Republic of Korea. Front Microbiol 2023;14. https://doi.org/10.3389/fmicb.2023.1170766
[11] López-Causapé C, Fraile-Ribot PA, Jiménez-Serrano S, Cabot G, del Barrio-Tofiño E, Prado MC, et al. A genomic snapshot of the SARS-CoV-2 pandemic in the Balearic Islands. Front Microbiol 2022;12. https://doi.org/10.3389/fmicb.2021.803827
[12] Viala VL, Slavov SN, de Lima LPO, Lima ARJ, Ribeiro G, Martins AJ, et al. The divergent pattern of SARS-CoV-2 variant predominance and transmission dynamics in the Brazilian Island of Ilhabela. Viruses 2022;14. https://doi.org/10.3390/v14071481
[13] Subiros M, De Latour CR, Parenton F, Soulaimana I, Hassani Y, Blondé R, et al. Epidemiological profile of COVID-19 in the French Overseas Department Mayotte, 2020 to 2021. Eurosurveillance 2022;27. https://doi.org/10.2807/1560-7917.ES.2022.27.34.2100953
[14] Grout L, Gottfreðsson M, Kvalsvig A, Baker MG, Wilson N, Summers J. Comparing COVID-19 pandemic health responses in two high-income island nations: Iceland and New Zealand. Scand J Public Health 2023;51(5):797–813. https://doi.org/10.1177/14034948221149143
[15] Kamiya T, Alvarez-Iglesias A, Ferguson J, Murphy S, Sofonea MT, Fitz-Simon N. Estimating time-dependent contact: a multi-strain epidemiological model of SARS-CoV-2 on the island of Ireland. Glob Epidemiol 2023;5. https://doi.org/10.1016/j.gloepi.2023.100111
[16] Cazelles B, Nguyen-Van-Yen B, Champagne C, Comiskey C. Dynamics of the COVID-19 epidemic in ireland under mitigation. BMC Infect Dis 2021;21. https://doi.org/10.1186/s12879-021-06433-9
[17] Chrysostomou AC, Vrancken B, Koumbaris G, Themistokleous G, Aristokleous A, Masia C, et al. A Comprehensive Molecular Epidemiological Analysis of Sars-Cov-2 Infection in Cyprus from April 2020 to January 2021: Evidence of a Highly Polyphyletic and Evolving Epidemic. Viruses 2021;13. https://doi.org/10.3390/v13061098
[18] Richter J, Fanis P, Tryfonos C, Koptides D, Krashias G, Bashiardes S, et al. Molecular epidemiology of SARS-CoV-2 in Cyprus. PLoS One 2021;16. https://doi.org/10.1371/journal.pone.0248792
[19] Sathyaseelan C, Magateshvaren Saras MA, Prasad Patro LP, Uttamrao PP, Rathinavelan T. CoVe-tracker: an interactive SARS-CoV-2 pan proteome evolution tracker. J Proteome Res 2023;22:1984–96. https://doi.org/10.1021/acs.jproteome.3c00068
[20] Price MN, Dehal PS, Arkin AP. FastTree 2 - approximately maximum-likelihood trees for large alignments. PLoS One 2010;5. https://doi.org/10.1371/journal.pone.0009490
[21] Huson DH, Richter DC, Rausch C, Dezulian T, Franz M, Rupp R. Dendroscope: an interactive viewer for large phylogenetic trees. BMC Bioinforma 2007;8. https://doi.org/10.1186/1471-2105-8-460
[22] Stefanelli P, Trentini F, Petrone D, Mammone A, Ambrosio L, Manica M, et al. Tracking the progressive spread of the SARS-CoV-2 omicron variant in Italy, December 2021 to January 2022. Eurosurveillance 2022;27. https://doi.org/10.2807/1560-7917.ES.2022.27.45.2200125
[23] Stefanelli P, Trentini F, Guzzetta G, Marziano V, Mammone A, Schepisi MS, et al. Co-Circulation of SARS-CoV-2 Alpha and Gamma variants in Italy, February and March 2021. Eurosurveillance 2022;27. https://doi.org/10.2807/1560-7917.ES.2022.27.5.2100429

[24] Hodcroft EB, Zuber M, Nadeau S, Vaughan TG, Crawford KHD, Althaus CL, et al. Spread of a SARS-CoV-2 variant through Europe in the Summer of 2020. Nature 2021;595:707–12. https://doi.org/10.1038/s41586-021-03677-y
[25] Alessandretti L. What human mobility data tell us about COVID-19 spread. Nat Rev Phys 2022;4:12–3.
[26] Cocherie T, Zafilaza K, Leducq V, Marot S, Calvez V, Marcelin AG, et al. Epidemiology and characteristics of SARS-CoV-2 variants of concern: the impacts of the spike mutations. Microorganisms 2023;11.
[27] Presti AL, Coghe F, Di Martino A, Fais S, Cappai R, Marra M, et al. First detection of SARS-CoV-2 lineage A.27 in Sardinia, Italy. Ann Ist Super Sanita 2022;58:1–5. https://doi.org/10.4415/ANN_22_01_01
[28] Klink GV, Safina KR, Nabieva E, Shvyrev N, Garushyants S, Alekseeva E, et al. The rise and spread of the SARS-CoV-2 AY.122 lineage in Russia. Virus Evol 2022;8. https://doi.org/10.1093/ve/veac017
[29] Grandi N, Paglietti B, Cusano R, Ibba G, Lai V, Piu C, et al. Genomic snapshot of SARS-CoV-2 in migrants entering through Mediterranean Sea routes. Front Public Health 2022;10. https://doi.org/10.3389/fpubh.2022.846115
[30] Ciuffreda L, González-Montelongo R, Alcoba-Florez J, García-Martínez de Artola D, Gil-Campesino H, Rodríguez-Pérez H, et al. Tracing the trajectories of SARS-CoV-2 variants of concern between December 2020 and September 2021 in the Canary Islands (Spain). Front Cell Infect Microbiol 2022;12. https://doi.org/10.3389/fcimb.2022.919346