GENOMIC EPIDEMIOLOGY OF THE MAIN SARS-CoV-2 VARIANTS CIRCULATING IN ITALY DURING THE OMICRON ERA

Annalisa Bergna[1], A. Lai[1], Fabio Sagradi[2], Stefano Menzo[3], Nicasio Mancini[4], Bianca Bruzzone[5], Stefano Rusconi[1], Greta Marchegiani[6], Nicola Clementi[7], Daniela Francisci[8], Ilaria Vicenti[9], Silvia Ronchiadin[10], Harcel Djaya Mbissam[1], Carla Della Ventura[1], Leonardo Lanfranchi[2], Sophie Testa[2], Sara Caucci[3], Carla Acciarri[3], Luca Carioti[6], Alessandro Occhionero[11], Federica Novazzi[4], Angelo Genoni[4], Francesca Drago Ferrante[4], Vanessa De Pace[5], Monica Ferraris[5], Matilde Ogliastro[12], Arianna Gabrieli[1], Massimo De Paschale[13], Giada Canavesi[13], Maria Concetta Bellocchi[6], Marco Iannetta[14], Loredana Sarmati[14], F. Ceccherini-Silberstein[6], Agostino Riva[1], Spinello Antinori[1], and Gianguglielmo Zehender[1]

[1]Universita degli Studi di Milano Dipartimento di Scienze Biomediche e Cliniche Luigi Sacco
[2]Azienda Socio Sanitaria Territoriale di Cremona
[3]Universita Politecnica delle Marche Dipartimento di Scienze Biomediche e Sanita Pubblica
[4]Universita degli Studi dell'Insubria
[5]IRCCS Ospedale Policlinico San Martino
[6]Universita degli Studi di Roma Tor Vergata Dipartimento di Medicina Sperimentale
[7]Universita Vita Salute San Raffaele
[8]Azienda Ospedaliera di Perugia
[9]Universita degli Studi di Siena Dipartimento di Biotecnologie Mediche
[10]Intesa Sanpaolo Innovation Center
[11]Azienda Ospedaliero Universitaria delle Marche
[12]Universita degli Studi di Genova Dipartimento di Scienze della Salute
[13]Ospedale Nuovo di Legnano
[14]Universita degli Studi di Roma Tor Vergata Dipartimento di Medicina dei Sistemi

August 26, 2024

Abstract

Since early 2022 the Omicron variant has rapidly spread worldwide, becoming the dominant variant to date. The study aimed to investigate the clinical and epidemiological characteristics of COVID-19 patients and to reconstruct the genomic epidemiology of main SARS-CoV-2 Omicron sub-lineages in Italy in 2022. 8,970 SARS-CoV-2 samples were studied, and phylogenetic analyses were focused on BA.1, BA.2 and BA.5 sub-variants. More than half of subjects received three doses of vaccine and experienced a reinfection. A significant larger proportion of unvaccinated subjects presented reinfection compared to vaccinated. Clusters presented a tMRCA between September-November 2021 (BA.1), November 2021-January 2022 (BA.2) and October 2021-May 2022 (BA.5). R_e values showed the highest level between September-October, January-February 2022, and May 2022 for BA.1, BA.2 and BA.5, respectively. Limited number of studied variant sequences are included in clusters. The analyses dissect the

epidemiological dynamics of Omicron sub-lineages in Italy over a period of great epidemiological changes in the COVID-19 epidemic. The spread rate of the studied variant exceeded its evolutionary rate. No single sub-lineage had sufficient time to differentiate into large clusters, but only into small and fragmented groups sharing the same recent ancestor. These analyses dissect the epidemiological dynamics of Omicron sub-lineages in Italy over a period of great epidemiological changes in the COVID-19 epidemic.

1 INTRODUCTION

Since its first identification in late December 2020, in the city of Whuan (China), COVID-19 pandemic caused more than 700 million cases all over the world with more than 6 million of deaths.

The first genomes of SARS-CoV-2 were characterized and publicly shared as early as in January 2020 reaching in January 2024 the number of more than 16,5 million of genomes available for the scientific community. This large publicly available data, and the development of efficient methods for phylogenetic and phylodynamic analyses, allowed to track the evolution of the virus genome and identify emerging variants providing important public health tools for the surveillance of SARS-CoV-2 during the pandemic.[1]

After the spillover event (or events) and the appearance of the first variant at highest transmissibility (D614G) in 2020,[2]the virus circulated as a heterogeneous population of genomic sub-lineages all derived by the original lineages (called A and B, following the Pango classification).

The Variants of Concern (VOCs), carrying an unusual number of mutations especially in the Spike protein, conferring to the mutant an increased transmissibility, were described for the first time in December 2020 (VOC Alpha, Beta, Gamma followed by Delta and then Omicron) and spread all over the world, with a mechanism of variant replacement.[3]

At present, Omicron remains the dominant variant circulating globally, somewhat distantly related to previous variants of concern (VOCs)[4] since it carries the highest number of mutations ever found in other VOCs. Moreover, it resulted associated with increased infectivity and enhanced immunoevasive properties.[5] Omicron was first identified in mid-November 2021 in South Africa and was designated as a VOC on 26 November 2021 (https://www.who.int/news/item/28-11-2021-update-on-omicron).[6]However, retrospective analyses revealed that Omicron was present in Europe 10 days before its discovery in South Africa with no obvious transmission link between the two locations.[7,8]

Omicron comprises 5 distinct sub-lineages (BA.1-5) that were discovered almost simultaneously, in November 2021, despite each sub-lineage being as different from each other as Alpha, Beta, Gamm a and Delta are from each other.[5] It has been suggested that BA.4 and BA.5 may have diverged via a recombination event, with a breakpoint suggested between the E and M genes.[9]

These characteristics gave the Omicron variant an adaptive advantage and a high spreading potential, as many as 67% (about 400 million) of global cases occurred after January 2022, since then it became the most prevalent variant worldwide. Epidemiological studies showed a global increase in the infection-induced seroprevalence after the emergence of Omicron variant in Europe.[10] Moreover, the seroprevalence among children increased in Europe from 37% during the fifth wave up to more than 56% after the sixth wave characterized by the circulation of Delta plus Omicron. [11]

Thus, at least in industrialized countries, it is possible to identify a pre-Omicron era, characterized by a low level of population immunization and the rise of variants with increased transmissibility, and an Omicron era, characterized by the emergence of lineages with a greater immunoevasive capacity selected by extensive natural and/or artificial immunization.[12-14]

It has been debated whether Omicron causes less severe disease; since, the lower incidence of hospitalisation and deaths observed with Omicron, is biased by the high vaccination rate and by previously infection.[15] However, after adjusting for the confounding effects of age, sex, ethnicity, prior infection, vaccination status, comorbidities, effect of province and effect of public/private sector, a recent WHO report showed evidence of reduced severity and lower mortality of the Omicron variant compared with the Delta variant

(https://www.who.int/publications/i/item/9789240051829). Both Pfizer and Moderna introduced booster vaccine dose including the S protein of Omicron variant claiming a better protection (Wuhan-like variant with an Omicron BA.1, BA.4/BA.5).[16]

SARS-CoV-2 recombinants emerging during the different waves of COVID-19 pandemic raised significant concerns, primarily due to their potential to accelerate immune evasion by means of antigenic shift. Among these recombinants, the first observed was named "Deltacron",[17] which originated in early 2022 from the recombination of Delta and Omicron BA.1 lineages, however it exhibited limited spread. More recently, the newly identified XBB lineage, also called the Kraken variant,[18]has gained considerable attention originating through the recombination of two highly diversified lineages, BJ.1 and BA.2.75.2, both derivatives of the Omicron BA.2 lineage. Remarkably, the XBB variants swiftly spread across populations worldwide, including those who had been vaccinated and those with hybrid immunity. In September 2023, the new updated vaccine targeting Omicron XBB became available.[19] Aims of this work were to study the clinical characteristics of COVID-19 patients and to reconstruct the genomic epidemiology and phylodynamic of main SARS-CoV-2 Omicron lineages circulating in Italy in 2022 period.

2 MATERIAL AND METHODS

2.1 Sample Collection

Between 1 January and 31 December 2022, the Italian Centres participating in the SCIRE (SARS-CoV-2 Italian Research Enterprise) collaborative group characterized a total of 8,970 SARS-CoV-2 positive samples obtained from either hospitalised or asymptomatic subjects tested in screening programs. The demographic characteristics of patients, as well as information about COVID-19 vaccination status, hospitalization and SARS-CoV-2 genotype, were collected at each centre for surveillance or for research purposes. This study was conducted in accordance with the principles of the 1964 Declaration of Helsinki and approved by the Sacco Hospital ethics committee (protocol n. 47866, 9 September 2020).

2.2 Virus characterization

Variant assessment was performed by different methods: RT-PCR variant specific screening assays (n=4,640, 51.7%), spike sequencing (n=1,164, 13%), and whole-genome sequencing (WGS, n=3,166, 35.3%). Viral RNA extraction, RT-PCR genotyping, amplification, and sequencing were obtained using different commercial kits or home-made procedures as previously described.[20] The SARS-CoV-2 lineage and clade were assigned to all Spike or Whole Genome (WG) sequences using the Pangolin COVID-19 Lineage Assigner v. 4.3 (https://pangolin.cog-uk.io/) and Nextclade v. 2.14.1 (https://clades.nextstrain.org/). Mutations were identified using Nextclade.

2.3 Statistical Analysis

Statistical analyses were performed with the IBM SPSS Statistics version 29. Descriptive analyses of data are presented as a median and an inter-quartile range (IQR) when quantitative and as a proportion when qualitative. To compare normally distributed, non-normally distributed continuous, and categorical variables, parametric tests (t-test and ANOVA), nonparametric tests (Mann–Whitney and Kruskal–Wallis), and the Pearson 2 test (or Fisher exact test, when necessary) were used, respectively. A $p\text{-}value < 0.05$ was considered statistically significant.

2.4 SARS-CoV-2 data sets

To study the major lineages of Omicron variant circulating during 2022, isolates of Omicron BA.1, BA.2 and BA.5 (BA.1, n= 268; BA.2, n= 677; BA.5, n=713) were selected and aligned with other Italian sequences of the same lineage, available in GISAID (https://gisaid.org). Genomes were selected based on the following criteria: 10 whole genomes for each Italian region and sampling month, according with the circulation period of each sub-variant, with a maximum of two sequences for region/week, excluding identical genomes and those with more than 5% of gaps. Three Italian datasets were set up: BA.1 (including a total of 880 isolates), BA.2 (n=1,627) and BA.5 (n=1,761) Omicron sub-variants. The dataset composition and regional

distribution of Italian sequences are summarized in **Tables S1** and **S2** . To place the Italian sequences in the international contest, an additional dataset was set up for each variant, selecting five genomes for each European and non-European countries and sampling month. Identical strains or those with more than 5% of gaps were excluded (**Table S1**).

Alignment of multiple sequences was obtained using MAFFT (https://mafft.cbrc.jp/alignment/server/) and the alignment was manually cropped using BioEdit v. 7.2.6.1 (https://bioedit.software.informer.com/). at the same length (29,774 bp).

The isolates included in the Omicron BA.1 dataset dated between November 2021 and April 2022, the BA.2 sub-variant isolates dated between January 2021 and December 2022, while the genomes included in the BA.5 sub-variant dataset had a date between April 2021 and December 2022.

2.5 Phylogenetic analysis

The statistically significant clusters (including more than three sequences) were identified in the International ML trees by Cluster Picker v.1.2.3 using 70% bootstrap support and a mean genetic distance of 0.1% as thresholds. Epidemiological characteristics of the identified clusters were further investigated using Cluster Matcher v. 1.2 23 which allows the identification of clusters meeting given criteria. Clusters were classified as mixed (M), containing both Italian and non-Italian isolates in different proportions, pure Italian (IT), including only Italian genomes, or European (EU), containing only European genomes.

The maximum likelihood trees of the three Italian datasets were estimated using IQ-TREE v. 1.6.12 (http://www.iqtree.org/).[21] The GTR+F+R3 (General time reversible + empirical base frequencies + three number of categories) model was used for BA.1 and BA.2 variants, while GTR+F+R6 models (General time reversible + empirical base frequencies + six number of categories) was used for BA.5. 1,000 parametric bootstrap replicates were performed to support the nodes ([?]60% bootstrap support).

For Italian datasets, the statistically significant clusters (including more than 3 sequences) were identified in the ML tree by Cluster Picker v.1.2.3 using 60% bootstrap support and a mean genetic distance of 0.1% as thresholds. Preliminary maximum likelihood tree was constructed including all the variants' significant clusters.

2.6 Phylodynamic analysis

To characterize the epidemiological and evolutionary history of the different SARS-CoV-2 Omicron variants in Italy, only clusters including at least 10 sequences were considered for each Italian dataset, by using the coalescent and the birth-death models.

Bayesian analysis was performed by BEAST v. 1.10.4 (https://beast.community/)[22] with the same substitution model and molecular clock employed for the previously described analyses.[23,24] Evolutionary rates were estimated using a Log Normal prior (mean, M=8E-4; variance, S=1.25) in real space using a strict clock and Bayesian Skygrid model, a nonparametric coalescent model that estimates the effective population size over time.

MCMC (Markov chain Monte Carlo) analyses were run for 60 million generations and sampled every 3,000. Convergence was assessed by estimating Effective Sampling Size (ESS) after applying a 10% burn-in through Tracer v.1.7 software (http://tree.bio.ed.ac.uk/software/tracer/),[25]accepting ESS of at least 200. The uncertainty of estimates was indicated with 95% highest prior density (HPD) intervals.

The final tree was selected based on the maximum posterior probability (pp) value after performing a 10% burn-in using Tree Annotator v.10.4 software (included in the BEAST package). Posterior probabilities greater than 0.7 were considered significant. Finally, all trees were visualized and edited in FigTree v. 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

The birth-death skyline model implemented in Beast v. 2.7 was used to infer changes in the effective reproductive number (R_e), and other epidemiological parameters such as the death/recovery rate (δ), the

transmission rate (λ), the origin of the epidemic, and the sampling proportion (ρ).[26] Given that the samples were collected during a short period of time, a "birth-death skyline serial" model was used.

For the birth-death analysis, one and two intervals and a lognormal prior to R_e, with a mean (M) of 0.0 and a variance (S) of 1.8 were chosen, which allows the R_e values to change between less than 1 and more than 7.

A normal prior with M = 48.8 and S = 15 (IC95%: 24.0-73.4) was used for the rate of becoming uninfectious. These values are expressed as units per year and reflect the inverse of the time of infectiousness (5.3-19 days; mean, 7.5) according to the serial interval estimated by Li et al.[27]

Sampling probability (ρ) was estimated assuming a prior β (α = 1.0 and β = 1,500), estimated based on available genomes in the analyses (normalizing to 1) and numbers of COVID-19 active cases at pick of the studied period.

For all sub-variants, origin of the epidemic was estimated using a lognormal prior with M = 0.1 and S = 0.3. The mean growth rate was calculated based on the birth and recovery rates (r = λ - d), and the doubling time was estimated by the equation: doubling time = ln(2)/r.[28]

3 RESULTS

3.1 Population characteristics

Analysed samples were collected from Italian centres located in Liguria (n=802), Lombardy (n=5,368), Umbria (n=66), Marche (n=2,506) and Lazio (n=228). Females accounted for 54% (n=4,788/8,867) and the median age was 58 years (IQR: 37-76) without any significant differences between sexes. Significant differences were observed in the median age over different months ($p<0.001$), with an increase in median age overtime (from 51 years in January to 73 years in December). Despite information of previous exposure to SARS-CoV-2 infection was available for a limited number of subjects (n=478), 54.8% (n=262) experienced a reinfection. Around one third of subjects had a known clinical status (n=3,102); 61.5% presented mild infections (n=1,907), followed by 31.7% of moderate/severe infections requiring hospitalisation (n=984). Among hospitalised patients, 6.3% (n=62) required intensive care and 8 died. Hospitalized patients showed a higher median age compared to asymptomatic or mildly symptomatic subjects ($p<0.001$, 73 vs. 52 and 49, respectively). Only a minority of the subjects were asymptomatics (211, 6.8%). These data are summarized in **Table 1** .

Table1. Characteristics of studied subjects.

Characteristics	**Overall**
Study population n(%)	8970
Methodology	
RT-PCR	4640 (51.7)
WGS NGS	3166 (35.3)
WGS/Spike Sanger	1164 (13)
Sex n(%)	
Males	4079 (46)
Females	4788 (54)
Median age (IQR)	58 (37-76)
Regions n(%)	
Liguria	802 (8.9)
Lombardy	5368 (59.8)
Umbria	66 (0.73)
Marche	2506 (27.9)
Lazio	228 (2.5)
Clinical status	

Characteristics	Overall
Non hospitalized	2118 (68.3)
Hospitalized	976 (31.5)
Deaths	8 (0.26)
Vaccination status	
Vaccinated	2814 (67.3)
Unvaccinated	1365 (32.7)
Doses administered	
1	449 (23.3)
2	390 (20.2)
3	1030 (53.4)
4	61 (3.2)
Type of vaccine	
Vaxzevria	22 (2.3)
Spikevax	220 (22.8)
Comirnaty	719 (74.7)
Jcovden	2 (0.9)

3.2 COVID-19 vaccinated *vs.* non-vaccinated

Among subjects with known COVID-19 vaccination status (n=4,179), 67.3% (n=2,814) received at least one dose of vaccine. More than half of the studied subjects received 3 doses (53.4%, n=1,030/1,930) of vaccine, only 3.1% received four doses (n=61) and 78.8% received BNT162b2 vaccine (n=719). No differences were observed in the proportion of vaccinated or non-vaccinated subjects in the study period, however, the median age of vaccinated individuals was lower compared to that of non-vaccinated ones (57.2 vs. 60, $p<0.001$). A significant larger proportion of unvaccinated subjects presented reinfection compared to vaccinated (53% *vs.* 25.1% $p<0.001$). Significant differences in the distribution of clinical status were present with the highest proportion of non-hospitalized subjects in vaccinated compared to unvaccinated (75.3% *vs.* 60.9%,$p<0.0001$). The proportion of deaths was significantly higher in unvaccinated than in vaccinated (0.6% *vs.* 0.4%,$p<0.0001$). No significant differences were observed in the gender distribution between vaccinated and unvaccinated subjects.

3.3 Lineages and clades

Globally the main observed variant was Omicron (8740/8970, n=97.4%) and its sub-lineages showing a prevalence of 44.6%, 26.8%, 1.8% and 26.8% for BA.1 (n=3898), BA.2 (n=2338), BA.4 (n=160) and BA.5 (n=2344), respectively. The Delta variant was observed until August, when the last case was observed, and globally accounted for 1.8% (n=160). Recombinants represented 0.7% of total sequences and included XC, XAZ, XBB, XBG, XBF, XQ and XT. First cases of XBB recombinants were observed in November (n=10, 2.6%). BA.1 remained the dominant until March (90.3%, 82.5% and 57.1% in January, February, and March, respectively) and completely disappeared since November. From April (95.2%) to June (48.6%), BA.2 became prevalent and was then replaced by BA.5. BA.5 reached the highest prevalence in September (83.9%) and then decreased to 24% in December when BQ.1 and descendants prevailed (52.6%). First case of BQ.1 was observed in September. Accordingly, the main clades included 21K (39.7%, n=3,469), 21L (25.4, n=2,279) and 22B (22.6%, n=2030) (**Figure 1**). By considering clinical status between subjects infected with Delta *versus* Omicron variant, a significantly higher proportion of non-hospitalized subjects was observed in subjects infected by Omicron (68.4% *vs.* 58.7%) and a significant higher proportion of deaths was found in Delta patients (2.2% *vs.* 0.2%; $p=0.017$). Of note, considering vaccination status, significant higher proportions of hospitalization and deaths were present in vaccinated patients carrying Delta variant compared to vaccinated or unvaccinated subjects with Omicron (72%*vs.* 23.2% and 38.8% for hospitalization and 4% *vs.* 0.3% and 0.6% for deaths, $p=0.02$). Considering Delta versus Omicron sub-lineages, the highest proportions of hospitalization were observed in Delta and BA.5 compared to BA.1 and BA.2 (53.1% and

51.5%*vs.* 11.8% and 28.8%; *p<0.0001*), while the proportion of deaths was significantly higher in subjects affected by Delta compared to those by Omicron sub-lineages (3.1% vs. 1.4%, 0 and 0.4%).

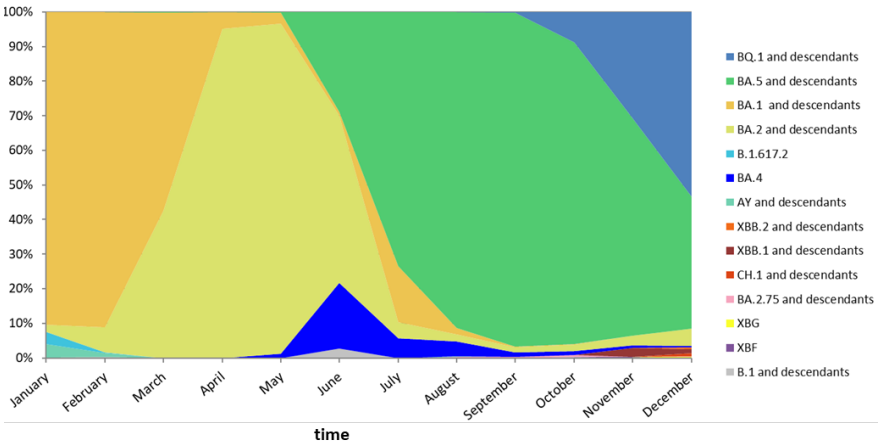


Figure 1. Dynamics of the SARS-CoV-2 epidemic in Italy, in term of lineages, during 2022.

3.4 Mutation analyses of the Italian sequences

Table S3 shows the sub-lineage composition of Italian Omicron BA.1 dataset.

The comparison between genomes from Italy and the reference sequence showed 49 aminoacidic substitutions and 7 deletions present in at least 10% of isolates. More than 30 mutations were present in the spike protein. Over 90% of the sequences had characteristics mutations of this variant and its descendants (Table S4). The V1187I mutation in ORF1a, characteristic of sub-lineages BA.1.17 and BA.1.17.2, was globally found in 32.4% (n=285) of genomes but was present in 90.5% (67/74) and 97.6% (204/209) of BA.1.17 and BA.1.17.2, respectively. The R346K mutation in the spike protein, found in 39.2% (n=345) of sequences, was present in 94.8% (n=275) of BA.1.1 genomes (n=290) and in almost all (88.4%, n=61) BA.1.1.1 (n=69) and descendants isolates. The G446S mutation in the Spike protein, typical of the BA.1.1 and descending sub-lineages, was found in a total of 665 isolates (75.6%), of which 73.3% (n=545) of BA.1.1 and descendants (n=746). The A701V mutation in the Spike protein, present in 20.5% (n=180) of genomes, was found in 85.6% (179/209) of the sequences belonging to this sub-lineage BA.1.17.2. In addition, the totality of the sequences of sub-lineages BA.1.15 and descendants (17/17), had the additional mutation D343G in the protein N, distinctive of this sub-lineage (Table S4).

The sub-lineage composition of the sequences included in the Omicron BA.2 dataset was shown in Table S5.

In the BA.2 and descendants dataset, only mutations/deletions typical of this lineage were found, as shown in Supporting Information (Table S6). In the ORF3a region, the H78Y mutation, present in a total of 16.5% (n=268) of the isolates, was prevalent in the BA.2.9 sub-lineage isolates (88.5%, 234/265). In the ORF1a region C655R, A2909V and Q3966H mutations (observed in less than 5% in dataset global) were identified in almost 100% of the isolates of sub-lineages BA.2.52.2 (33/33), BA.2.3 (44/44) and BA.2.22 (16/18), where these substitutions are characteristic. Similarly, S959P mutation in ORF1b, was present in 95.2% (20/21) of the BA.2.10 isolates. The L140F mutation in ORF3a, was found in almost all BA.2.3 and descendant sequences (97.7%, 42/43). In protein S, two mutations characteristics of the

sub-lineage BA.2.12.1, L452Q and S704L, were identified in 94.5% (63/66) and 98.5% (65/66) of its isolates, respectively.

Table S7 shows the sub-lineage composition of Omicron BA.5 dataset.

A total of 50 mutations and 5 deletions were found in at least 10% of the sequences analysed (Table S8). These substitutions have been identified in more than 80% of the isolates, except for the T1050N mutation, in the ORF1b region, with a global prevalence of 20.8% (n=367) but which was present in almost all isolates BA.5.2, BA.5.2.2 and descendant sub-lineages, and mutation D16G in the ORF9b region; this substitution, present in almost 50% (n=844) of the isolates, was found in all BA.5.2 isolates (n=350) and in 97.6% (322/330) of BA.5.2.1 sequences. In protein S, in addition to mutations typical of this variant, 98.1% (n=1721) of the sequences bore the mutation G142Y. Substitutions with a global frequency of less than 10% but characteristics of different sub-lineages have been identified in the ORF1a, ORF1b, S and N regions. In the ORF1a region, mutations S302F, Q556K, K3839R and T4161I were observed in all BA.5.1.23 (n=27), BA.5.3.1 (n=16), BA.5.1.10 (n=101) and BA.5.1.8 (n=37) sequences, respectively.

3.5 Phylogenetic analysis and dating of the Italian clusters

3.5.2 Maximum likelihood analyses of international datasets

Maximum Likelihood analysis of the international datasets showed that the majority of whole genomes of BA.1, BA.2, and BA.5 (ranging from 72.5% to 87.6%) were scattered throughout the trees, while 12.4% (228/1,837) of BA.1, 19.9% (649/3246) of BA.2, and 27.5% (885/3219) of BA.5 genomes formed significant clusters, from 3 to 17, 30 and 60 sequences (for BA.1, BA.2 and BA.5, respectively) and mainly localized at the external nodes of the trees. In detail, 46.2% (24/52) of BA.1 clusters included sequences exclusively from Italy, as well as 37.9% (53/140) of BA.2 clusters and 36.5% (69/189) of BA.5 clusters, while mixed clusters were 9.6% (5/52), 27.9% (39/140), and 49.7% (94/189) respectively.

3.5.1 Maximum likelihood analyses on Italian datasets

The phylogenetic analysis conducted on the Italian sequences of the sub-variant Omicron BA.1 showed the presence of 30 clusters, characterized by more than three sequences (min 4-max 21), which included 24.3% (n=214) of total analysed sequences (n=880); four (8.7%) clusters included more than 10 genomes. There was no change in the pattern of clustering based on the sampling area (northern, southern, central Italy and islands).

Analysis of the Omicron BA.2 sub-variant showed the presence of 60 clusters (min-max: 4-30 sequences), which included 22.6% of the isolates analysed (368/1,627); 7 (6.3%) clusters included more than 10 genomes. No different clustering pattern was found based on the sampling area.

The 26.5% (n=467) of isolates included in the BA.5 dataset (n=1,761) grouped into 68 clusters, of which 10 (7.4%) clusters included a number greater/equal to 10 genomes. Sequences from the islands clustered more frequently than those from northern, central, and southern Italy (52.3% vs. 37.6%, 35.6%, 37.2%;$p<.05$).

Maximum Likelihood analysis conducted on isolates included in all BA.1, BA.2 and BA.5 clusters showed that these lineages formed three highly significant monophyletic groups (Table S9 and Figure S1).

Preliminary analysis by root-to-tip regression revealed a linear relationship between genetic diversity and time (correlation coefficient=0.81 and R squared= 0.66) (**Figure S2**).

3.5.2 Omicron BA.1

Given the limited number of clusters containing more than 10 isolates, the Bayesian phylogenetic analysis was conducted on a dataset that included all sequences forming clusters with at least 4 sequences (n=214).

Bayesian analysis estimated a mean substitution rate of 4.84x10^{-4} s/s/y (95%HPD: 3.76-5.98x10-4 s/s/y) and showed that all sequences grouped within twelve statistically supported clusters (pp>0.9) in the tree (**Figure S3**).

The tMRCA of each cluster was dated between September and November 2021 (95% HPD: June-December 2021) (**Table 2**). These clusters contained an average of 17.8 genomes (minimum of 4 and maximum of 71) with a persistence between 2 and 7 months (**Table 2**). Earlier clusters (dated September 2021) showed the larger size (20.5 *vs.* 6.5 isolates) and the longer persistence (7 months*vs.* 4.5 months) than later clusters (dated October/November).

Table 2 . **tMRCA estimation of the main clusters of BA.1 dataset with the relative 95% HPD and duration in the time.**

	n. of sequences	*pp*[a]	**data**	**95%HPD**[b] ***Lower***	**95%HPD *Upper***	**More recent samples**	**mon**
#1	71	0.98	05/09/2021	21/06/2021	23/10/2021	05/04/2022	7
#11	20	0.99	06/09/2021	02/07/2021	29/10/2021	04/04/2022	7
#10	16	0.97	14/09/2021	13/07/2021	02/11/2021	31/01/2022	4
#12	16	0.99	16/09/2021	21/07/2021	04/11/2021	08/02/2022	5
#8	21	0.99	20/09/2021	16/07/2021	05/11/2021	02/04/2022	7
#7	21	0.92	25/09/2021	21/07/2021	09/11/2021	03/04/2022	7
#9	9	0.99	02/10/2021	04/08/2021	14/11/2021	07/04/2022	6
#2	19	1	11/10/2021	30/08/2021	13/11/2021	28/03/2022	5
#6	7	1	23/10/2021	14/09/2021	26/11/2021	12/03/2022	5
#3	6	1	31/10/2021	28/09/2021	29/11/2021	28/12/2021	2
#5	4	1	03/11/2021	01/10/2021	03/12/2021	31/01/2022	2
#4	4	1	06/11/2021	01/10/2021	09/12/2021	07/03/2022	4

[a]*pp: posterior probability.*

[b]*HPD: highest posterior density.*

3.5.3 BA.2

Bayesian phylogenetic analysis was conducted on a dataset that included all sequences included in the 7 largest Italian clusters containing more than 10 sequences for a total of 111 genomes.

The Bayesian analysis estimated a mean evolutionary rate of 3.99x10^{-4}s/s/y (95%HPD: 2.70x10-4-5.33 x10^{-4}).

Clusters contained an average of 15.9 genomes (min-max: 10-30), dated between November 2021 and January 2022 (95%HPD: August 2021 and February 2022) and a persistence between 4 and 8 months (**Figure S4; Table 3**) without any relationship between the clusters size and persistence.

Table 3 . tMRCA estimation of the main clusters of BA.2 dataset with the relative confidence intervals and duration in the time.

	n. of sequences	*pp*[a]	**data**	**95%HPD**[b] ***Lower***	**95%HPD *Upper***	**More recent samples**	**mo**
#249	11	0.99	26/11/2021	28/08/2021	30/01/2022	07/06/2022	7
#114	12	0.99	11/12/2021	16/10/2021	22/01/2022	07/06/2022	5
#40	16	0.99	27/12/2021	13/11/2021	03/02/2022	08/09/2022	8
#52	30	1	01/01/2022	17/11/2021	08/02/2022	07/06/2022	5
#133	10	1	03/01/2022	25/11/2021	05/02/2022	04/05/2022	4

	n. of sequences	*pp*[a]	data	95%HPD[b] *Lower*	95%HPD *Upper*	More recent samples	mo
#117	17	1	03/01/2022	28/11/2021	05/02/2022	07/06/2022	5
#57	15	1	13/01/2022	29/11/2021	23/02/2022	04/07/2022	6

[a]*pp: posterior probability.*

[b]*HPD: highest posterior density.*

3.5.4 BA.5

Bayesian analysis of the Omicron BA.5 variant was conducted on the 164 sequences included in the 10 clusters containing more than 10 sequences.

The estimated evolutionary rate showed an average of 4.56×10^{-4} s/s/y (95%HPD: 3.72×10^{-4}-5.44×10^{-4}). Clusters' tMRCAs dated from October 2021 to May 2022 (95%HPD: July 2021 and July 2022) (**Figure S5**), but most of them dated in March and April 2022. Clusters contained an average of 16.4 genomes (min-max: 10-39) and showed a persistence of a mean 8.3 months (range: 5 11 months) (**Table 4**). No relationship was observed between the clusters size and persistence.

Table 4 . tMRCA estimation of the main clusters of BA.5 dataset with the relative confidence intervals and duration in the time.

	n. of sequences	*pp*[a]	data	95%HPD[b] *Lower*	95%HPD *Upper*	More recent samples	mo
#208	11	0.99	24/10/2022	25/07/2021	16/01/2022	25/07/2022	9
#141	18	0.99	08/01/2022	02/11/2021	08/01/2022	08/11/2022	11
#42	20	1	13/02/2022	21/12/2021	04/04/2022	10/12/2022	10
#113	21	1	21/03/2022	01/02/2022	04/05/2022	07/11/2022	8
#92	39	1	26/03/2022	02/02/2022	11/05/2022	09/12/2022	9
#105	12	1	31/03/2022	05/02/2022	17/05/2022	17/12/2022	9
#133	10	0.99	05/04/2022	16/02/2022	18/05/2022	21/11/2022	7
#192	10	1	17/04/2022	13/03/2022	20/05/2022	27/09/2022	5
#62	10	1	20/04/2022	11/03/2022	28/05/2022	13/12/2022	8
#213	13	1	24/05/2022	15/04/2022	14/07/2022	15/12/2022	7

[a]*pp: posterior probability.*

[b]*HPD: highest posterior density.*

3.6 Bayesian phylodynamic analysis

3.6.1 BA.1 phylodynamic

The Bayesian phylodynamic analysis of the BA.1 Italian clusters, showed that the number of infections progressively grew since the origin of the epidemic (September 2021); a spike growth started in November 2021 reaching the plateau in January 2022 lasting until March 2022, when the effective number of infections started to decrease (**Figure 2A**).

In agreement with this dynamic, the estimate of R_e was close to the threshold 1 until December when the effective reproduction number reached 1.45, followed by a decline in January 2022 (when the number of infections reached the plateau) to the initial values (**Figure 2B**).

Hosted file

`image2.emf` available at https://authorea.com/users/639130/articles/1218908-genomic-epidemiology-of-the-main-sars-cov-2-variants-circulating-in-italy-during-the-omicron-era

Figure 2. (A) Bayesian Skygrid plot of BA.1 variant. The y-axis indicates the effective population (N_e), the x-axis shows the time expressed in dates. The thick line in the graph indicates the median of the value of the estimate, while the blue area indicates 95% HPD. (B) Birth-death skyline plot of BA.1 variant, in relation to time (x-axis) and the effective reproduction rate (R_e) (y-axis).

3.6.2 BA.2 phylodynamics

In the case of BA.2, an exponential increase of the effective number of infections was observed only in January/February 2022 and the plateau was reached between March and April 2022, when an initial decline in the number of infections was observed, followed by a rebound during summer (**Figure 3A**). Similarly, the estimate of the Re has shown values around 1 since the beginning of the epidemic, but the peak (1.42) was reached between January and February 2022, returning to values around the unity in February-March showing a more pronounced reduction between May and July 2022 (**Figure 3B**).

Hosted file

`image3.emf` available at https://authorea.com/users/639130/articles/1218908-genomic-epidemiology-of-the-main-sars-cov-2-variants-circulating-in-italy-during-the-omicron-era

Figure 3. (A) Bayesian Skygrid plot of BA.2 variant. The y-axis indicates the effective population (N_e), the x-axis shows the time expressed in dates. The thick line in the graph indicates the median of the value of the estimate, while the blue area indicates 95% HPD. (B) Birth-death skyline plot of BA.2 variant, in relation to time (x-axis) and the effective reproduction rate (R_e) (y-axis).

3.6.3 BA.5 phylodynamics

The curve showing the effective number of BA.5 infections exhibits two growth phases, with the initial phase in January 2022, being flatter, followed by a subsequent steeper increase starting from May 2022 and reaching a peak of cases around July. The decrease began in the second half of the same month of July or August, with a more pronounced decline starting from October 2022 (**Figure 4A**).

Similarly, the estimate of the Re showed values above 1 from October 2021, although the highest values were observed from May 2022 (1.28) to July, when the estimates of the effective reproduction number dropped around 1, where they remained until the end of the study (**Figure 4B**).

Hosted file

`image4.emf` available at https://authorea.com/users/639130/articles/1218908-genomic-epidemiology-of-the-main-sars-cov-2-variants-circulating-in-italy-during-the-omicron-era

Figure 4. (A) Bayesian Skygrid plot of BA.5 variant. The y-axis indicates the effective population (N_e), the x-axis shows the time expressed in dates. The thick line in the graph indicates the median of the value of the estimate, while the blue area indicates 95% HPD. (B) Birth-death skyline plot of BA.5 variant, in relation to time (x-axis) and the effective reproduction rate (R_e) (y-axis).

4 DISCUSSION

After the first report in Botswana in late 2021, since early 2022 the Omicron variant has rapidly spread worldwide, becoming the dominant variant to date with its derived sub-lineages.[29,30] The higher transmissibility, the lower neutralizing efficacy of antibodies stimulated by previous infections or vaccination, the less severe clinical spectrum (possibly because infecting subjects protected by cell-mediated immunity), the shorter incubation period (4.5-5 days for Alpha and Delta variants *vs* . 3.5 days for Omicron variant) and the higher replicative efficiency, made this variant very different from those that circulated previously, requiring an update of the vaccines in use. In addition, the presence in the meantime of a population almost entirely vaccinated with at least one dose (more than 85% in Italy; https://www.lombardianotizie.online/vaccinati-over-80/), and the progressively easing of containment measures until their total elimination, radically changed the epidemiology of the infection leading to a distinction between a past "pre-Omicron" and a current Omicron era.

Compared to previous variants, the Omicron variant showed a high number of mutations, with an average of 50 substitution throughout the genome more than half of which are localized in the spike protein. These characteristics are responsible for a higher binding affinity to the ACE2 receptor and greater immune escape from neutralizing and monoclonal antibodies.[12,14,31]

This variant has rapidly further evolved giving rise to an array of multiple lineages and sub-lineages[32]with specific mutational profiles. The intrinsic characteristic of the Omicron variant and the current absence of containment measures have allowed for a rapid spread[33] and evolution of Omicron.

This study included more than 8,000 samples collected during the year 2022 in five Italian regions accounting for more than 33% of the Italian population (19 million out of a total population of 58 million individuals in 2024); 2 regions comprising at least 72.5% of the population living in North-Western (around 72 million out of 115 million inhabitants) and 3 and in the central part of the Country, accounting for 68% (8 million out of 69 million, data from ISTAT: http://dati.istat.it/Index.aspx?DataSetCode=DCIS_POPRES1).

Due to the predominant inclusion of patients from clinical centres and the relative microbiology laboratories, only a small number of asymptomatic subjects were included, while the proportion of subjects who had received at least one dose of vaccine resulted 67%, lower that that reported in Italy in the same period.

However, as expected, comparing the characteristics of vaccinated and unvaccinated subjects, the unvaccinated subjects showed a significant higher proportion of subjects requiring hospitalization (24.7% among vaccinated *vs* . 39.1% among unvaccinated), a double frequency of patients experiencing reinfection (53% *vs* . 25%.1%) and a higher frequency of deaths (0.6% *vs.* 0.4%), confirming previous data on the efficacy of COVID-19 vaccines. Moreover, these data confirmed that Omicron is less virulent than the Delta variant but also indicated, differently from previous reports[34,35] a lower lethality both in vaccinated and unvaccinated subjects infected by Omicron. Despite the small number of individuals infected by Delta variant, due to the time-period analysed in this study, a relevant difference in the proportion of deaths was also observed considering different Omicron sub-lineages such as BA.1, BA.2 and BA.5. An important limit to these observations could be the changing criteria that during the study period were adopted by the clinical centres for the admission of COVID-19 patients. In fact, as vaccination became more prevalent in the population, the COVID-19 pressure on hospitals was alleviated. This allowed the structures to admit also less severe infections and, more frequently, to admit patients with mild or asymptomatic COVID-19 requiring hospitalization for unrelated pathologies. Alleviations of prophylactic measures also favoured the nosocomial circulation of the virus in patients and health workers. All these phenomena are clearly testified by the decreasing age of the study subjects throughout the course of the study. We suspect that this kind of bias applies also to all other studies describing clinical differences between SARS-CoV-2 variants, given the impossibility to set-up rigorous prospective studies.

The present study also provides a comprehensive description of the spread of the SARS-CoV-2 variants in Italy in 2022, starting from the replacement of the previously dominant variant of concern Delta with the Omicron variant since January 2022.

The Omicron variant prevailed in this study, representing more than 97% while Delta variant was observed only in less than 2% of subjects until August, when it disappeared similarly to what observed in the rest of the world. A similar replacement mechanism was also observed in the succession of Omicron sub-lineages overtime. Lineage BA.1 was prevalent until March 2022, when it was replaced by BA.2 which remained the dominant lineage until June, when BA.5 prevailed until November. Only at the end of the study (December 2022) the lineage BQ.1, derived by BA.5, became the most frequent.

While the BA.1 variant has become extinct a few months after its spread, BA.2 and BA.5 variants continue (February 2024) to circulate with different derived sub-lineages and recombinant forms due to the increased transmissibility.[31] The circulation of recombinants started in 2022 due to the co-circulation of Delta and Omicron VOCs, however their large spread matches with the identification of XBB recombinants resulted from recombination between two lineages of BA.2, which were first identified only at the end of study period (November 2022).

By analysing the mutational profile of these sub-lineages, we found a high number of mutations, mainly located in the S gene, all identical to those previously described.

The analysis of the international dataset showed the presence of only small clusters at the external nodes of the tree, including only few isolates probably closely epidemiologically related, thus preventing the observation of larger significant transmission clusters.[36] This is due to the scanty, relatively rare, and dispersed sampling, along with the strategies adopted to reduce the magnitude of genomes introduced in the analysis.

It is also likely that, as the spread rate of the variant exceeded its evolutionary rate (Omicron became prevalent worldwide within a couple of months) grace to its insensitivity to herd immunity and the lowering of restrictions, single sub-lineage did not have sufficient evolutionary pressure for being selected to form large clusters, but only small and fragmented groups sharing the same recent ancestor.[37,38]

Analysing the sub-variants within the international context, there was a tendency for Italian isolates to group in the tree in the same regions, even without forming distinct and significant clades. For this reason, the phylogenetic analysis was conducted considering only the national context in which we found a partial formation of clusters on a local basis, mainly in the case of the BA.1 and BA.2 lineages.

Indeed, thanks to the neutral evolution phenomena, such as the "founding effect", it was possible to identify significant clusters only at local (national) level, while in the international context, due to the high degree of evolutionary correlation between strains of a single variant (BA.1, BA.2 and BA.5), it is difficult to identify more closely related groups of sequences that share a recent common ancestor.[37,38]

The estimated time to most recent common ancestor for BA.5 suggests this lineage would have been circulating throughout the period dominated by BA.1 and then BA.2 without any transmission advantage. According to literature data, maximum likelihood estimations suggest that BA.5 could have descended from BA.2.[9] From clusters analysis, most of BA.1 clusters dated in autumn 2021, those of the BA.2 mainly between late 2021 and early 2022, while the estimated origin of the BA.5 clusters covered a wide time span, ranging from October 2021 to May 2022.

For these variants, higher R_e values were observed than those estimated for previous variants,[23] confirming Omicron's enhanced transmissibility. Although the literature data indicate a higher transmissibility of BA.2 than BA.1,[39,40] the values we estimated were comparable. These estimates are in line with the official ones (https://covid19.infn.it/), except for the peak in January 2022, corresponding to the simultaneous circulation of the Delta and Omicron variants. Although the mean R_e value estimated at peak for BA.5 was lower than those estimated for BA.1 and BA.2 (1.28*vs.* 1.45 and 1.42, respectively), there was a persistence of values above 1 over time since the beginning of the epidemic, which would account for the peak of infected individuals observed in the 2022 wave (https://covid19.infn.it/).

Analyses show a good temporal correspondence between the trend in the number of infections estimated from the Skyline graph and the estimated R_e by birth-death Skyline.

Although this work is the result of the collaboration of many clinical/diagnostic centres located throughout the country, some regions are less represented leading to a potential sampling bias.[41,42] In conclusion, these data allowed an accurate description of the epidemiological dynamics of Omicron sub-lineages in Italy over a period of great epidemiological changes in the COVID-19 epidemic.

AUTHOR CONTRIBUTIONS

Annalisa Bergna, Alessia Lai, Gianguglielmo Zehender conceived the project; Fabio Sagradi, Stefano Menzo, Nicasio Mancini, Bianca Bruzzone, Stefano Rusconi, Greta Marchegiani, Nicola Clementi, Daniela Francisci, Ilaria Vicenti, Silvia Ronchiadin, Carla della Ventura, Leonardo Lanfranchi, Sophie Testa, Sara Caucci, Carla Acciarri, Luca Carioti, Alessandro Occhionero, Federica Novazzi, Angelo Paolo Genoni, Francesca Drago Ferrante, Vanessa De Pace, Monica Ferraris, Matilde Ogliastro, Arianna Gabrieli, Massimo De Paschale, Giada Canavesi, Maria Concetta Bellocchi, Marco Iannetta, Loredana Sarmati, Francesca Ceccherini-

Silberstein, Agostino Riva, Spinello Antinori collected the samples and information; Alessia Lai, Harsel Djaya Mbissam performed statistical data analyses; Annalisa Bergna, Alessia Lai, Gianguglielmo Zehender performed phylogenetic analyses; Annalisa Bergna, Alessia Lai, Gianguglielmo Zehender interpreted the results; Fabio Sagradi, Bianca Bruzzone, Greta Marchegiani, Nicola Clementi, Ilaria Vicenti, Carla della Ventura, Leonardo Lanfranchi, Sara Caucci, Carla Acciarri, Luca Carioti, Federica Novazzi, Angelo Paolo Genoni, Francesca Drago Ferrante, Vanessa De Pace, Monica Ferraris, Matilde Ogliastro, Massimo De Paschale, Giada Canavesi, Maria Concetta Bellocchi contributed to the sequencing; Annalisa Bergna, Alessia Lai, Stefano Menzo, Gianguglielmo Zehender wrote the first draft of the manuscript. All authors read, revised and approved the final version of the manuscript.

COLLABORATIVE GROUP

SCIRE collaborative Group: Claudia Balotta[1], Mario Corbellino[1], Massimo Galli[1], Valentina Ricucci[5], Federica Stefanelli[5], Nadia Randazzo[5], Giada Garzillo[13], Massimo Clementi[8], Maurizio Zazzi[10], Lia Fiaschi[10], Massimo Andreoni[15], Arianna Miola[11], Valeria Ricci[11], Laura Li Puma[11], Luigi Ruggerone[11].

FUNDING STATEMENT

This work was supported by the European Union's Horizon Europe Research and Innovation Actions under grant no. 101046041, and from the Ministero dell'Università e della Ricerca (PRIN 202022GZEHE_01) and partly supported by EU funding within the NextGeneration EU-MUR PNRR Extended Partnership initiative on Emerging Infectious Diseases (Project no. PE00000007, INF-ACT). The funders had no role in study design, data collection, and analysis, the decision to publish or preparation of the manuscript.

ACKNOWLEDGEMENTS

The authors thank the UNIMI GSA-IDEA project.

DATA AVAILABILITY STATEMENT

All analytical data are available within the article, Figures, and Supplementary Data. Whole-genome sequences were submitted to GISAID.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict to interest.

SUPPORTING INFORMATION

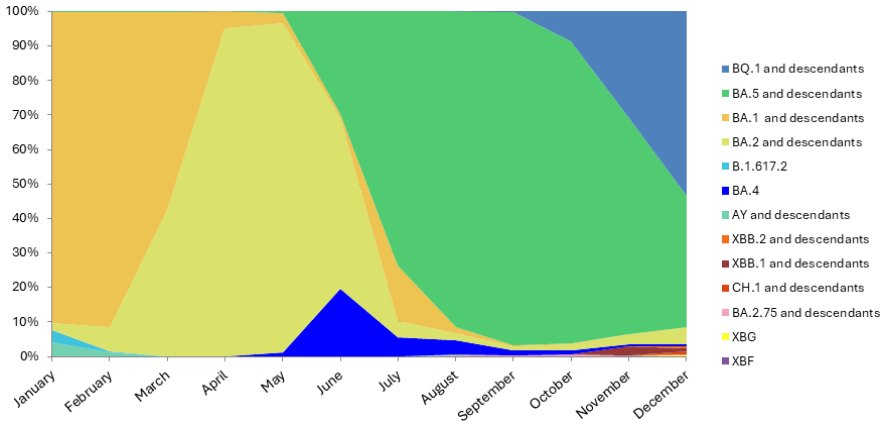
Supporting information S1 PDF.

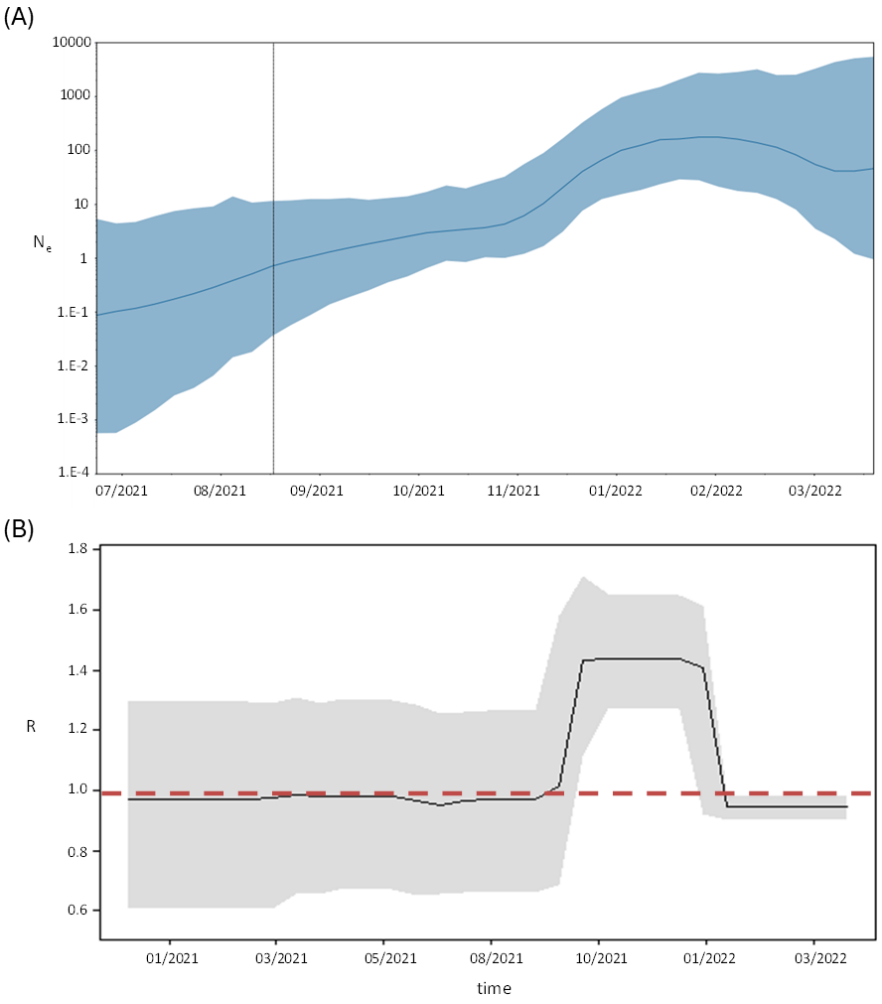
REFERENCES

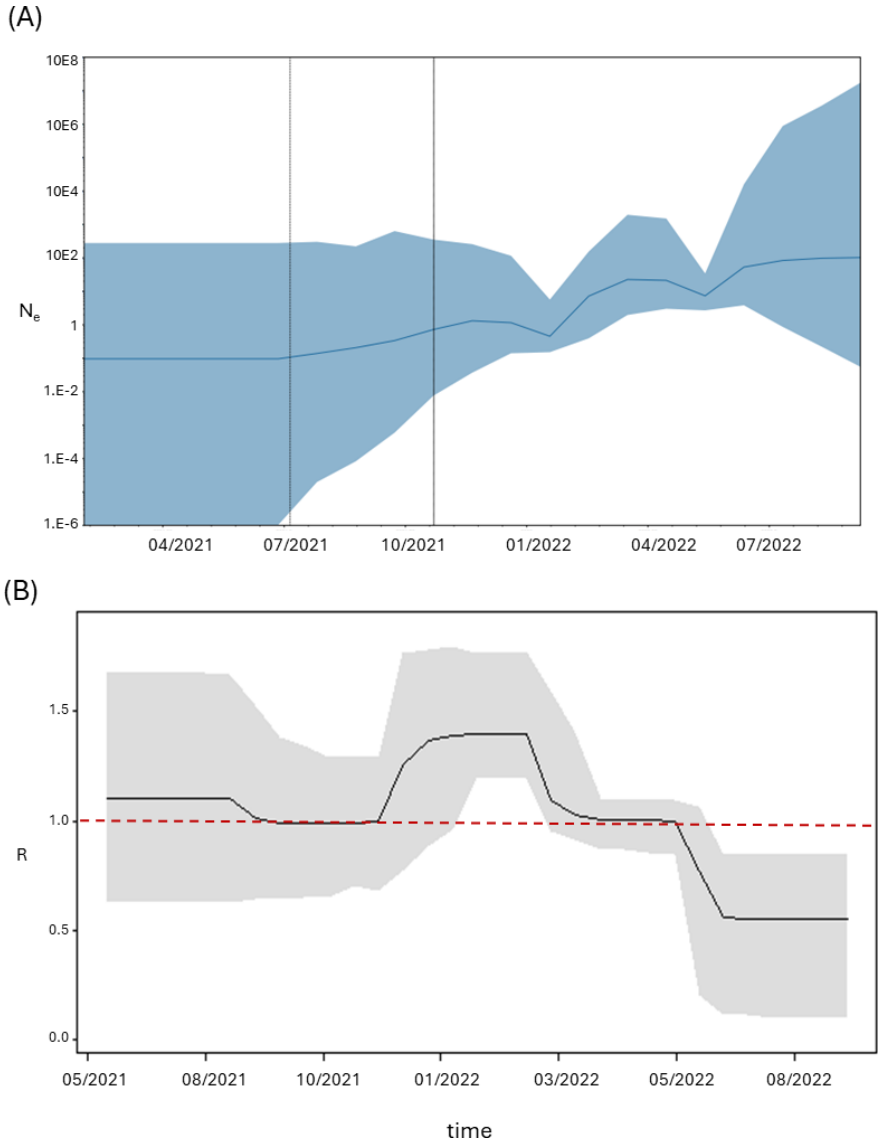
1. Attwood SW, Hill SC, Aanensen DM, Connor TR, Pybus OG. Phylogenetic and phylodynamic approaches to understanding and combating the early SARS-CoV-2 pandemic. *Nature reviews Genetics.* 2022;23(9):547-562.2. Korber B, Fischer WM, Gnanakaran S, et al. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell.* 2020;182(4):812-827 e819.3. Telenti A, Hodcroft EB, Robertson DL. The Evolution and Biology of SARS-CoV-2 Variants. *Cold Spring Harbor perspectives in medicine.* 2022;12(5).4. Hadfield J, Megill C, Bell SM, et al. Nextstrain: real-time tracking of pathogen evolution.*Bioinformatics (Oxford, England).* 2018;34(23):4121-4123.5. Berkhout B, Herrera-Carrillo E. SARS-CoV-2 Evolution: On the Sudden Appearance of the Omicron Variant. *Journal of virology.* 2022;96(7):e0009022.6. Viana R, Moyo S, Amoako DG, et al. Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa.*Nature.* 2022;603(7902):679-686.7. Poudel S, Ishak A, Perez-Fernandez J, et al. Highly mutated SARS-CoV-2 Omicron variant sparks significant concern among global experts - What is known so far?*Travel medicine and infectious disease.* 2022;45:102234.8. Tsui JL, McCrone JT, Lambert B, et al. Genomic assessment of invasion dynamics of SARS-CoV-2 Omicron BA.1. *Science (New York, NY).* 2023;381(6655):336-343.9. Tegally H, Moir M, Everatt J, et al. Emergence of SARS-CoV-2 Omicron lineages

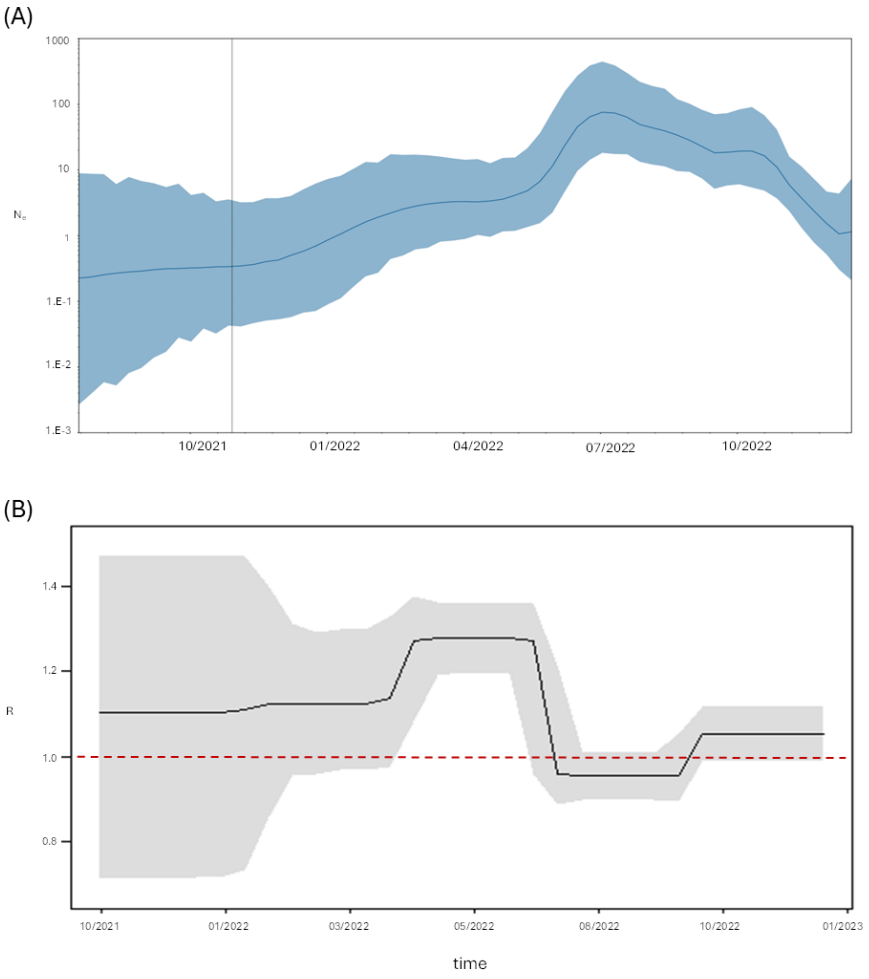
BA.4 and BA.5 in South Africa.*Nature medicine.* 2022;28(9):1785-1790.10. Bergeri I, Whelan MG, Ware H, et al. Global SARS-CoV-2 seroprevalence from January 2020 to April 2022: A systematic review and meta-analysis of standardized population-based studies. *PLoS medicine.* 2022;19(11):e1004107.11. Naeimi R, Sepidarkish M, Mollalo A, et al. SARS-CoV-2 seroprevalence in children worldwide: A systematic review and meta-analysis.*EClinicalMedicine.* 2023;56:101786.12. Chen L, He Y, Liu H, Shang Y, Guo G. Potential immune evasion of the severe acute respiratory syndrome coronavirus 2 Omicron variants. *Frontiers in immunology.* 2024;15:1339660.13. Gili R, Burioni R. SARS-CoV-2 before and after Omicron: two different viruses and two different diseases? *Journal of translational medicine.* 2023;21(1):251.14. Guo K, Barrett BS, Mickens KL, et al. Interferon Resistance of Emerging SARS-CoV-2 Variants. *bioRxiv : the preprint server for biology.* 2021.15. Le Page M. Understanding omicron. *New scientist (1971).* 2022;253(3369):8-9.16. Liu J, He Q, Gao F, et al. Heterologous Omicron-adapted vaccine as a secondary booster promotes neutralizing antibodies against Omicron and its sub-lineages in mice. *Emerging microbes & infections.* 2023;12(1):e2143283.17. Farheen S, Araf Y, Tang YD, Zheng C. The Deltacron conundrum: Its origin and potential health risks. *Journal of medical virology.* 2022;94(11):5096-5102.18. Parums DV. Editorial: The XBB.1.5 ('Kraken') Subvariant of Omicron SARS-CoV-2 and its Rapid Global Spread. *Medical science monitor : international medical journal of experimental and clinical research.* 2023;29:e939580.19. Link-Gelles R, Ciesla AA, Roper LE, et al. Early Estimates of Bivalent mRNA Booster Dose Vaccine Effectiveness in Preventing Symptomatic SARS-CoV-2 Infection Attributable to Omicron BA.5- and XBB/XBB.1.5-Related Sublineages Among Immunocompetent Adults - Increasing Community Access to Testing Program, United States, December 2022-January 2023. *MMWR Morbidity and mortality weekly report.* 2023;72(5):119-124.20. Lai A, Bergna A, Della Ventura C, et al. Epidemiological and Clinical Features of SARS-CoV-2 Variants Circulating between April-December 2021 in Italy. *Viruses.* 2022;14(11).21. Minh BQ, Schmidt HA, Chernomor O, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era.*Molecular biology and evolution.* 2020;37(5):1530-1534.22. Bouckaert R, Vaughan TG, Barido-Sottani J, et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS computational biology.* 2019;15(4):e1006650.23. Bergna A, Lai A, Ventura CD, et al. Genomic epidemiology of the main SARS-CoV-2 variants in Italy between summer 2020 and winter 2021. *Journal of medical virology.* 2023;95(11):e29193.24. Lai A, Bergna A, Toppo S, et al. Phylogeography and genomic epidemiology of SARS-CoV-2 in Italy and Europe with newly characterized Italian genomes between February-June 2020.*Scientific reports.* 2022;12(1):5736.25. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic biology.* 2018;67(5):901-904.26. Stadler T, Kühnert D, Bonhoeffer S, Drummond AJ. Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proceedings of the National Academy of Sciences of the United States of America.* 2013;110(1):228-233.27. Li Q, Guan X, Wu P, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia.*The New England journal of medicine.* 2020;382(13):1199-1207.28. Walker PR, Pybus OG, Rambaut A, Holmes EC. Comparative population dynamics of HIV-1 subtypes B and C: subtype-specific differences in patterns of epidemic growth. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases.* 2005;5(3):199-208.29. Carabelli AM, Peacock TP, Thorne LG, et al. SARS-CoV-2 variant biology: immune escape, transmission and fitness. *Nature reviews Microbiology.* 2023;21(3):162-177.30. Roemer C, Sheward DJ, Hisner R, et al. SARS-CoV-2 evolution in the Omicron era. *Nature microbiology.* 2023;8(11):1952-1959.31. Chavda VP, Bezbaruah R, Deka K, Nongrang L, Kalita T. The Delta and Omicron Variants of SARS-CoV-2: What We Know So Far. *Vaccines.* 2022;10(11).32. Wang Y, Long Y, Wang F, Li C, Liu W. Characterization of SARS-CoV-2 recombinants and emerging Omicron sublineages. *International journal of medical sciences.* 2023;20(1):151-162.33. Tegally H, Wilkinson E, Tsui JL, et al. Dispersal patterns and influence of air travel during the global expansion of SARS-CoV-2 variants of concern. *Cell.* 2023;186(15):3277-3290 e3216.34. Radhakrishnan N, Liu M, Idowu B, et al. Comparison of the clinical characteristics of SARS-CoV-2 Delta (B.1.617.2) and Omicron (B.1.1.529) infected patients from a single hospitalist service.*BMC infectious diseases.* 2023;23(1):747.35. Van Goethem N, Chung PYJ, Meurisse M, et al. Clinical Severity of SARS-CoV-2 Omicron Variant Compared with Delta among Hospitalized COVID-19 Patients in Belgium during Autumn and Winter Season 2021-2022. *Viruses.* 2022;14(6).36. Francesconi M, Giovanetti M, De Florio L, et al. Genomic Epidemiology Unveil the Omicron Transmission

Dynamics in Rome, Italy.*Pathogens (Basel, Switzerland).* 2022;11(9).37. Ragonnet-Cronin M, Hodcroft E, Hué S, et al. Automated analysis of phylogenetic clusters.*BMC bioinformatics.* 2013;14:317.38. Xie R, Edwards KM, Adam DC, et al. Resurgence of Omicron BA.2 in SARS-CoV-2 infection-naive Hong Kong. *Nature communications.* 2023;14(1):2422.39. Lyngse FP, Kirkeby CT, Denwood M, et al. Household transmission of SARS-CoV-2 Omicron variant of concern subvariants BA.1 and BA.2 in Denmark.*Nature communications.* 2022;13(1):5760.40. Zhou Y, Zhi H, Teng Y. The outbreak of SARS-CoV-2 Omicron lineages, immune escape, and vaccine effectivity. *Journal of medical virology.* 2023;95(1):e28138.41. Han AX, Kozanli E, Koopsen J, et al. Regional importation and asymmetric within-country spread of SARS-CoV-2 variants of concern in the Netherlands. *eLife.* 2022;11.42. Kucharski AJ, Jit M, Logan JG, et al. Travel measures in the SARS-CoV-2 variant era need clear objectives.*Lancet (London, England).* 2022;399(10333):1367-1369.









Hosted file

Table 1.docx available at https://authorea.com/users/639130/articles/1218908-genomic-epidemiology-of-the-main-sars-cov-2-variants-circulating-in-italy-during-the-omicron-era

Hosted file

Table 2.docx available at https://authorea.com/users/639130/articles/1218908-genomic-epidemiology-of-the-main-sars-cov-2-variants-circulating-in-italy-during-the-omicron-era

Hosted file

Table 3.docx available at https://authorea.com/users/639130/articles/1218908-genomic-epidemiology-of-the-main-sars-cov-2-variants-circulating-in-italy-during-the-omicron-era

Hosted file

Table 4.docx available at https://authorea.com/users/639130/articles/1218908-genomic-epidemiology-of-the-main-sars-cov-2-variants-circulating-in-italy-during-the-omicron-era