

UNIVERSITÀ DEGLI STUDI DI MILANO

SCUOLA DI DOTTORATO IN SCIENZE BIOMEDICHE, CLINICHE E SPERIMENTALI

Dipartimento di Scienze Cliniche e Sperimentali

CORSO DI DOTTORATO IN “SCIENZE PER LA SANITÀ PUBBLICA”

XXXVI CICLO

Dipartimento di Scienze Biomediche e Cliniche



TESI DI DOTTORATO DI RICERCA

**EPIDEMIOLOGIA GENOMICA DELLE PRINCIPALI VARIANTI DI SARS-CoV-2  
CIRCOLANTI IN ITALIA PRIMA E DURANTE L'ERA OMICRON**

Settore scientifico disciplinare MED/42

**DOTTORANDA: ANNALISA BERGNA**

**TUTOR: PROF. GIANGUGLIELMO ZEHENDER**



Gianguglielmo  
Zehender  
Universita' degli  
Studi di Milano  
28.12.2023  
16:33:08  
GMT+01:00

**COORDINATORE DEL DOTTORATO: PROF. CARLO LA VECCHIA**

**A.A. 2022/2023**

## Sommario

<b>1. INTRODUZIONE</b>	<b>1</b>
<b>1.1 ORIGINE DI SARS-CoV-2</b>	<b>1</b>
<b>1.2 TASSONOMIA E GENOMA DI SARS-CoV-2</b>	<b>2</b>
<b>1.3 CLASSIFICAZIONE MOLECOLARE DI SARS-CoV-2</b>	<b>7</b>
<b>1.4 CLASSIFICAZIONE DELLE VARIANTI DI SARS-CoV-2</b>	<b>9</b>
1.4.1 <i>VARIANTS OF INTEREST (VOI)</i>	10
1.4.2 <i>VARIANTS OF CONCERN (VOC)</i>	10
1.4.3 <i>VARIANT UNDER MONITORING (VUM)</i>	11
1.4.4 <i>FORMERLY MONITORING VARIANTS</i>	12
<b>1.5 PRINCIPALI LINEAGES/VARIANTI CIRCOLANTI DA AGOSTO 2020 AD OGGI</b>	<b>13</b>
1.5.1 <i>CLADE 20E.EU1</i>	13
1.5.2 <i>VARIANTE ALPHA (B.1.1.7)</i>	13
1.5.3 <i>VARIANTE DELTA (B.1.617.2)</i>	14
1.5.4 <i>VARIANTE OMICRON (B.1.1.529)</i>	14
1.5.5 <i>RICOMBINANTI DI SARS-CoV-2</i>	17
<b>1.6 EPIDEMIOLOGIA</b>	<b>18</b>
<b>1.7 LA PANDEMIA DA COVID-19 NEL MONDO</b>	<b>19</b>
<b>1.8 SITUAZIONE IN ITALIA</b>	<b>21</b>
<b>1.9 VACCINI E COPERTURA VACCINALE IN ITALIA</b>	<b>25</b>
<b>2. PRIMI DATI IN ITALIA</b>	<b>30</b>
<b>3. SCOPO</b>	<b>33</b>
<b>4. MATERIALI E METODI</b>	<b>34</b>
4.1 <b>POPOLAZIONE IN STUDIO</b>	<b>34</b>
4.2 <b>SEQUENZIAMENTO DEI GENOMI VIRALI</b>	<b>34</b>
4.3 <b>CRITERI DI ALLESTIMENTO DEI DATASET</b>	<b>35</b>
4.4 <b>ANALISI FILOGENETICA</b>	<b>37</b>
4.5 <b>ANALISI FILODINAMICA E STIMA DEL TASSO RIPRODUTTIVO DI BASE</b>	<b>38</b>
<b>5. RISULTATI (1)</b>	<b>42</b>
5.1 <b>ANALISI DELLE MUTAZIONI NEI GENOMI DI SARS-CoV-2 ITALIANI</b>	<b>42</b>
5.1.1 <i>CLADE 20E.EU1</i>	42
5.1.2 <i>VARIANTE ALPHA</i>	43
5.1.3 <i>VARIANTE DELTA</i>	44
5.2 <b>ANALISI FILOGENETICA E DATAZIONE DEI CLUSTER ITALIANI</b>	<b>45</b>

5.2.1	CLADE 20E.EU1 .....	45
5.2.2	VARIANTE ALPHA .....	47
5.2.3	VARIANTE DELTA.....	49
<b>5.3</b>	<b>ANALISI FILODINAMICA DEI CLUSTER ITALIANI .....</b>	<b>52</b>
5.3.1	CLADE 20E.EU1 .....	52
5.3.2	VARIANTE ALPHA .....	53
5.3.3	VARIANTE DELTA.....	54
<b>6. RISULTATI (2) .....</b>		<b>55</b>
<b>6.1</b>	<b>ANALISI DELLE MUTAZIONI NEI GENOMI DI SARS-COV-2 .....</b>	<b>55</b>
6.1.1	OMICRON BA.1.....	55
6.1.2	OMICRON BA.2.....	57
6.1.3	OMICRON BA.5.....	59
<b>6.2</b>	<b>ANALISI E DATAZIONE FILOGENETICA DELLE VARIANTIOMICRON .....</b>	<b>61</b>
<b>6.3</b>	<b>ANALISI BAYESIANA DEI CLUSTER ITALIANI .....</b>	<b>61</b>
6.3.1	OMICRON BA.1.....	61
6.3.2	OMICRON BA.2.....	65
6.3.3	OMICRON BA.5.....	68
<b>6.4</b>	<b>ANALISI FILODINAMICA DEI CLUSTER ITALIANI .....</b>	<b>71</b>
6.4.1	OMICRON BA.1.....	71
6.4.2	OMICRON BA.2.....	72
6.4.3	OMICRON BA.5.....	73
<b>7. DISCUSSIONE .....</b>		<b>74</b>
<b>8. BIBLIOGRAFIA .....</b>		<b>81</b>

# 1. INTRODUZIONE

## 1.1 ORIGINE DI SARS-COV-2

Alla fine di dicembre 2019, nella città di Wuhan, nella provincia di Hubei, Cina Centrale, a seguito dell'identificazione di un focolaio di 29 casi di polmoniti acute ad eziologia non nota, è stato descritto un nuovo *coronavirus*, successivamente rinominato *Severe Acute Respiratory Syndrome Coronavirus 2* (SARS-CoV-2), responsabile della malattia definita COVID-19 (<https://www.who.int/emergencies/diseases/novel-coronavirus-2019>).

I soggetti malati presentavano principalmente sintomi simil-influenzali come febbre e mialgia, sebbene una parte dei pazienti presentasse un quadro clinico più grave caratterizzato da problemi respiratori e lesioni polmonari osservabili tramite radiografia toracica.

A seguito dell'indagine epidemiologica iniziata il 31 dicembre 2019 dal dipartimento cinese dell'Organizzazione Mondiale della Sanità (OMS), i casi di polmonite severa acuta furono sin da subito associati al mercato di Wuhan (*Wuhan's South China Seafood City market*) nel quale venivano venduti diversi animali selvatici, tra cui serpenti, uccelli, pipistrelli, conigli, rane e numerosi piccoli mammiferi.<sup>1</sup>

Dalla Cina il virus si diffuse rapidamente nelle nazioni confinanti (in particolare Sud Corea, Hong Kong e Singapore) raggiungendo regioni più distanti come l'Europa, gli Stati Uniti e l'Iran, dove i primi casi furono registrati a fine gennaio. A differenza di quanto accaduto per i *coronavirus* umani SARS-CoV (*Severe Acute Respiratory Syndrome Coronavirus*) e MERS-CoV (*Middle East Respiratory Syndrome Coronavirus*), SARS-CoV-2 risultava aver acquisito un'efficiente capacità di trasmissione interumana, che ne permise una rapida diffusione.

Sebbene infatti i primi casi severi di polmonite acuta furono da subito associati al mercato di Wuhan, successive indagini rilevarono la presenza di soggetti infetti senza alcun legame con tale mercato, ipotizzando una trasmissione non solo di origine zoonotica, ma anche interumana.

Il 30 gennaio 2020, a seguito dell'identificazione di casi di importazione in Paesi diversi dalla Cina, l'Organizzazione Mondiale della Sanità (WHO) dichiarò il nuovo *coronavirus* un'emergenza di rilevanza mondiale in termini di Salute Pubblica. In questa data, infatti, SARS-CoV-2 si era già diffuso in 20 Paesi con 10,000 casi e 200 morti confermate. I primi casi in Italia risalgono al 21 febbraio 2020 e quasi in contemporanea avvenne la diffusione in Europa, America del Nord e successivamente nel resto del mondo. L'11 marzo 2020 l'Organizzazione mondiale della Sanità dichiarò lo stato di pandemia.<sup>2</sup> A distanza di poco più di 3 anni, dopo oltre 6.5 milioni di morti e

gravi sconvolgimenti economici e sociali a livello mondiale, il 5 maggio 2023, la stessa organizzazione dichiara ufficialmente terminata l’Emergenza Internazionale di salute pubblica (Pheic) per la COVID-19 sottolineando allo stesso tempo l’importanza nel continuare a monitorare questo virus che può ancora rappresentare una minaccia in termini di salute globale.

## 1.2 TASSONOMIA E GENOMA DI SARS-CoV-2

I *coronavirus* sono virus appartenenti al dominio *Riboviria*, ordine *Nidovirales*, famiglia *Coronaviridae* e sottofamiglia *Ortho-coronavirinae*. L’ordine dei *Nidovirales* (**Figura 1**) è rappresentato da virus che possiedono una molecola di RNA lineare a polarità positiva. Il prefisso “Nido-” indica che il loro RNA si avvolge a formare una struttura simile ad un nido. La sottofamiglia *Ortho-coronavirinae* è costituita da 4 diversi generi: *Alphacoronavirus* (dimensione media del genoma 28kb), *Beta-coronavirus* (dimensione media del genoma 30kb), *Gammacoronavirus* (dimensione media del genoma 28kb) e *Deltacoronavirus* (dimensione media del genoma 26kb).<sup>3</sup> Mentre i mammiferi rappresentano l’*habitat* naturale dei generi *Alphacoronavirus* e *Betacoronavirus*, i *Gammacoronavirus* e *Deltacoronavirus* sono stati identificati nelle specie aviarie.

A seguito dell’immediato sequenziamento di questo nuovo virus, l’analisi della sequenza genomica ne ha indicato l’appartenenza al genere *β-coronavirus*, sottogenere *Sarbecovirus*, indicando un’identità di sequenza nucleotidica del 79.6% con il genoma di SARS-CoV<sup>4</sup> e per questo rinominato SARS-CoV-2.

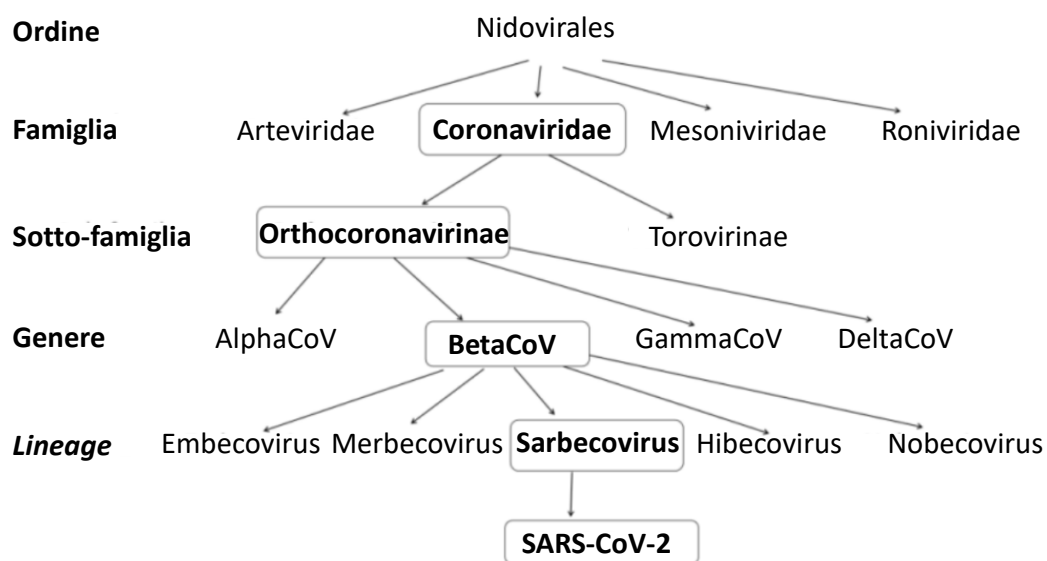


Figura 1: Tassonomia dei Coronavirus. Fonte: modificato da Anup Som, 2022.

I *β-coronavirus* sono anche chiamati *bat-coronavirus* in quanto il loro serbatoio naturale è rappresentato dai pipistrelli.

Diversi studi hanno sin da subito supportato l'ipotesi che, analogamente agli altri *β-coronavirus* umani, definiti HCoV (quali HCoV-OC43, HCoV-HKU1, SARS-CoV, MERS-CoV), anche SARS-CoV-2 avesse come *reservoir* i pipistrelli; in particolare, è stata riscontrata un'identità genomica del 96% tra il genoma di SARS-CoV-2 e il genoma del *coronavirus* BatCov RaTG13 isolato dal pipistrello *Rhinolophus affinis* nella provincia dello Yunnan.<sup>5</sup>

Ulteriori studi hanno inoltre riportato come, analogamente a SARS-CoV e MERS-CoV che si servono di ospiti intermedi per effettuare il salto di specie all'uomo, anche per SARS-CoV-2 esiste un ospite intermedio nel quale il virus si replica prima del passaggio alla specie umana, probabilmente rappresentato dal pangolino.<sup>6</sup> Similitudini sono state riscontrate tra la regione ad anello variabile della proteina spike di SARS-CoV-2 e quella osservata nel ceppo *coronavirus* derivato dal Pangolino malese (Pangolin-CoV-2020).<sup>7</sup> Tuttavia, l'origine diretta dal pangolino è stata messa in discussione dalla scoperta di altri virus simili SARS-CoV-2, trovati in pipistrelli provenienti da diverse località del sud-est asiatico.<sup>4</sup> All'interno di questi ospiti intermedi non definiti, SARS-CoV-2 ha acquisito gran parte delle mutazioni necessarie per un'efficiente trasmissione e replicazione nell'uomo.<sup>8</sup>

SARS-CoV-2 è un virus sferico di circa 60-140 nm rivestito da un *envelope* da cui sporgono delle strutture superficiali osservabili attraverso microscopia elettronica, che conferiscono alla superficie del virus un aspetto simile ad una corona dovuta alla presenza della glicoproteina *spike* (S) che forma dei raggi di 9-12 nm di lunghezza sulla superficie virale.

Il genoma a singolo filamento a RNA a polarità positiva, di circa 30 kb, è costituito da un *cap* metilato in 5' e una coda di poliA in 3' (regioni non tradotte, UTRs) che contengono strutture secondarie di RNA ad azione *cis*, essenziali per la sintesi dell'RNA, ed è caratterizzato dalla presenza di 4 geni molto conservati tra i *coronavirus*, codificanti per le proteine strutturali: E (*envelope*), N (*nucleocapside*), S (*spike*), M (*membrane*) (**Figura 2**).

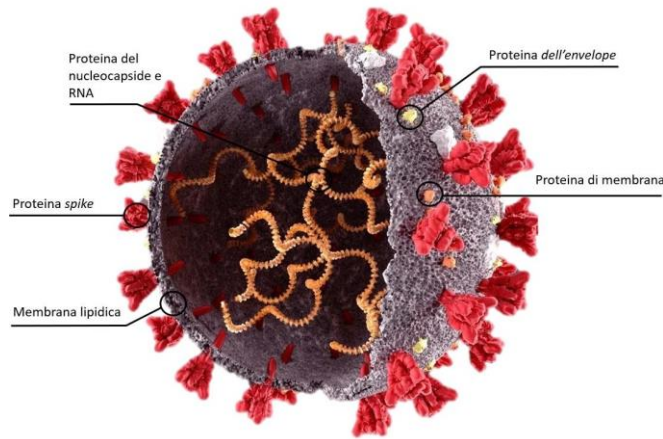


Figura 2. Rappresentazione schematica di SARS-CoV-2. Fonte: modificato da <https://www.shutterstock.com//it>.

Il genoma di SARS-CoV-2 è caratterizzato da 14 *open reading frames* (ORFs): ORF1a e 1b, che occupano i due terzi dell'intero genoma, ORF2 (*spike*), ORF3a, ORF3b, ORF4 (*envelope*), ORF5 (*membrane*), ORF6, ORF7a, ORF7b, ORF8, ORF9a, ORF9b (*nucleocapside*) e ORF10 che codificano per proteine accessorie e proteine strutturali (Figura 3).

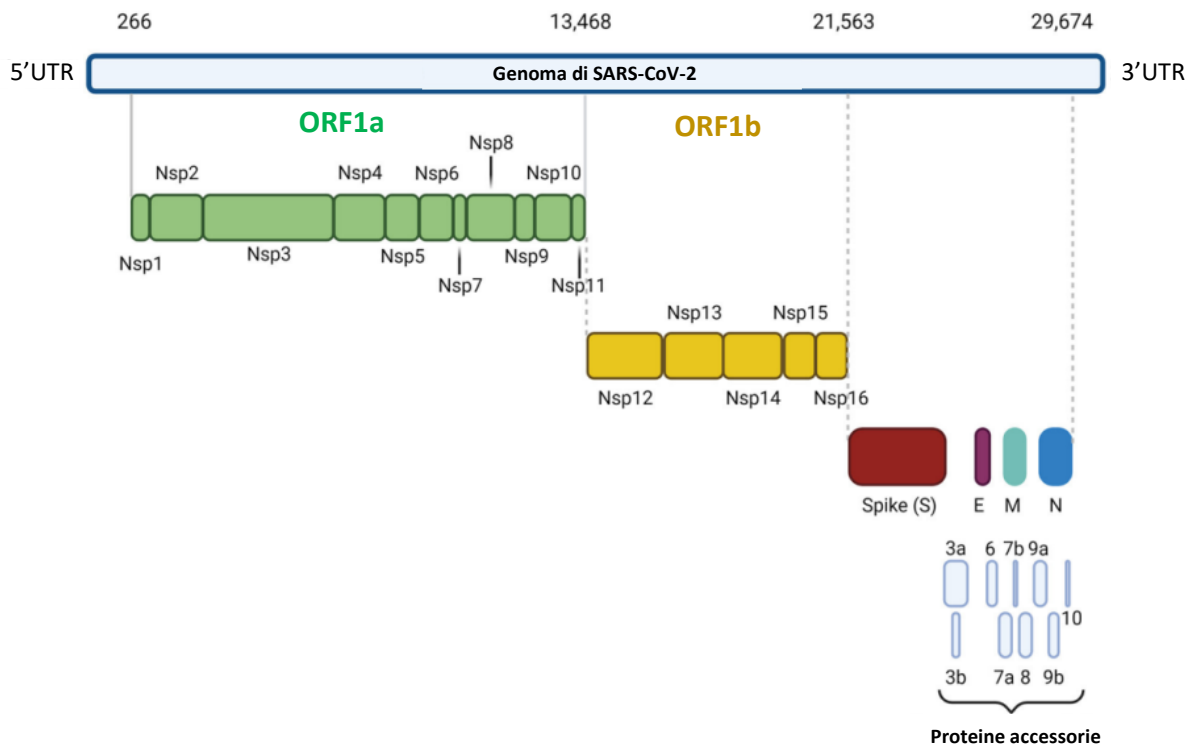


Figura 3. Organizzazione genomica di SARS-CoV-2. Fonte: modificata da Pizzato M, 2022.<sup>9</sup>

Le regioni **ORFs 1a** e **1b** codificano per due polipetidi, pp1a e pp1ab, che a seguito di processi post-traduzionali verranno divisi ad opera di proteasi in 16 proteine non strutturali (Nsp) 1-16, che andranno poi a formare il complesso di replicazione-trascrizione. Tra queste, la proteina non strutturale Nsp12 è la RNA polimerasi RNA-dipendente (RdRp) e rappresenta la proteina più conservata in tutti i *coronavirus*. Essa ha un ruolo centrale nel complesso di replicazione e trascrizione del genoma virale e la sua funzione è quella di copiare l'RNA virale.

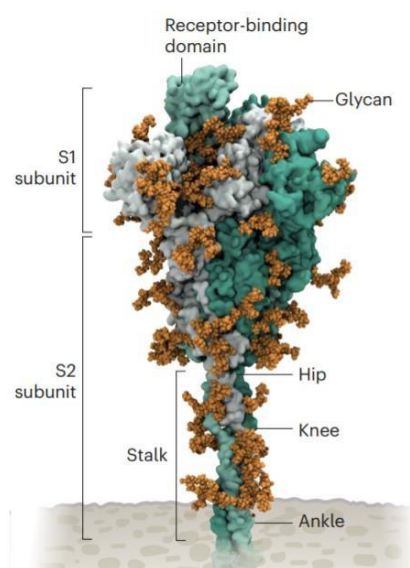
La proteina non strutturale Nsp14 svolge un ruolo cruciale nella sintesi dell'RNA virale. Attraverso l'attività esoribonucleasica 3'-5' di *proofreading* è implicata nel sistema di riparo dei *mismatch* incrementando la fedeltà della sintesi dell'RNA, e mediante l'attività metil-transferasica, necessaria per la formazione del *cap* del virus, protegge l'RNA dalla degradazione e dall'identificazione da parte del sistema immunitario dell'ospite.

Nove proteine accessorie, denominate ORF3a, 3b, 6, 7a, 7b, 8, 9a, 9b e 10, codificate dalle omonime ORF, sebbene ritenute non essenziali per la replicazione del virus, sulla base di studi *in vitro*, sembrano esercitare importanti funzioni nella modulazione del metabolismo della cellula ospite e dell'immunità antivirale.<sup>10</sup>

Tra le più rilevanti proteine strutturali, la **proteina S**, codificata dalla ORF2, è una glicoproteina trimerica presente sulla superficie del virus e consiste in un ectodominio, una porzione transmembrana e un breve dominio intracellulare (**Figura 4**). L'ectodominio è formato da due subunità (S1 e S2), dove la S1, contenente il dominio di legame al recettore (*Receptor Binding Domain*, RBD), è la porzione responsabile sia dell'attacco con il recettore della cellula ospite sia della fusione con la cellula stessa.

Tra le subunità S1 e S2 è presente un sito di clivaggio della furina (RRAR), che consente un taglio da parte della furina e di altre proteasi, permettendo la fusione del virus con la cellula ospite.

La proteina S (Spike) riconosce e lega la proteina ACE2 (*angiotensin-converting enzyme 2*) presente sulla superficie della cellula ospite, determinando il tropismo del virus e rappresentando l'antigene principale di SARS-CoV-2.



**Figura 4.** Rappresentazione della proteina spike di SARS-CoV-2. Fonte: *Nature* 2021.<sup>11</sup>



La **proteina strutturale E** (*envelope*) è la più piccola proteina integrale di membrana presente in SARS-CoV-2, codificata dalla ORF4 e costituita da un breve dominio N-terminale, un lungo dominio transmembrana idrofobo ed un lungo dominio C-terminale idrofilo. Svolge un ruolo centrale nella morfogenesi e nell'assemblaggio del virus. Nel dominio C-terminale di questa proteina sono infatti presenti tre cisteine che subiscono processi di ubiquitinazione e palmitoilazione, reazione necessaria per l'assemblaggio del virus; questa proteina è inoltre in grado di oligomerizzare, formando delle viroporine capaci di trasportare selettivamente gli ioni calcio e partecipare all'assemblaggio e al rilascio di particelle virali nella cellula ospite. Le viroporine sono anche essenziali per la formazione dell'ERGIC (*Endoplasmatic-Reticulum-Golgi intermediate Compartment*) all'interno del quale avviene l'assemblaggio dei virioni.

La **glicoproteina M** (*membrane*), implicata nella formazione dell'*envelope*, è codificata dalla ORF5 e rappresenta una delle proteine più abbondanti in SARS-CoV-2. È costituita da tre regioni transmembrana caratterizzate da un breve dominio NH<sub>2</sub> terminale che protrude all'esterno del virus e da un lungo dominio citoplasmatico COOH all'interno del virione. Questa proteina è presente in due forme, una allungata e l'altra compatta; la proteina spike si inserisce prevalentemente a livello della proteina in forma lunga, facendogli assumere una forma sferica, circondando così la ribonucleoproteica.

La proteina M insieme alla **glicoproteina strutturale N** (*nucleocapside*), codificata dalla ORF9, è implicata nell'assemblaggio dei nuovi virioni e nella loro gemmazione. Quest'ultima, presenta tre domini altamente conservati tra i *coronavirus*: un dominio N-terminale di legame all'RNA (NTD, N1b), un dominio di dimerizzazione C-terminale (CTD, N2b) e un dominio *linker* ricco in arginina e serina (SR, N3). Essa gioca un ruolo fondamentale nella formazione della struttura del virione in quanto è localizzata sia nella regione di replicazione-trascrizione sia nel sito di assemblaggio del virione ERGIC. La proteina N può legarsi alla proteina Nsp3 favorendo il legame con il complesso di replicazione-trascrizione RTC (*Replication-Transcription Complex*) e, legandosi direttamente all'RNA virale, contribuisce alla sua stabilità. La proteina N di SARS-CoV-2 è espressa principalmente durante le prime fasi dell'infezione e rappresenta un *marker* diagnostico in quanto induce una forte risposta immunitaria.

### 1.3 CLASSIFICAZIONE MOLECOLARE DI SARS-CoV-2

L'elevata variabilità genomica di SARS-CoV-2 è evidenziata dalle sequenze virali depositate che, sulla base della classificazione di Pangolin COVID-19 Lineage Assigner (<https://pangolin.cog-uk.io/>),<sup>12</sup> si raggruppano in due *lineages* principali, A e B, separati all'inizio dell'epidemia di Wuhan, ognuno contenente un ampio numero di *sub-lineages* (A.1, B.1, B.1.1, B.1.1.1 ecc.) ([https://cov-lineages.org/lineage\\_list.html](https://cov-lineages.org/lineage_list.html)).

Il *lineage* B.1 ha rappresentato nelle prime fasi dell'epidemia, il più rilevante *lineage* europeo, la cui origine sembra corrispondere all'epidemia che ha coinvolto l'Italia settentrionale all'inizio del 2020.

Sulla base di un diverso sistema di classificazione (<https://nextstrain.org/>) (**Figura 5**), solo parzialmente sovrapponibile a quello descritto prima, che assegna a ogni *clade* una lettera sulla base della diversificazione del ceppo rispetto al *clade* progenitore e un numero corrisponde all'anno di comparsa del nuovo *clade* (19, 20, 21, 22 e 23), sono stati identificati nel corso del 2020 cinque principali *clades*: 19A, 19B, 20A, 20B e 20C. I criteri per la classificazione di un nuovo *clade* sono due: la frequenza globale deve essere almeno del 20% con almeno 2 nuove mutazioni rispetto al *clade* originario.

I *clade* 19A e 19B (corrispondenti ai *lineages* B e A) sono emersi a Wuhan e hanno caratterizzato l'epidemia iniziale in Cina. Il *clade* 19A rappresenta il progenitore e include la sequenza di riferimento Wuhan Hu-1 (*Accession Code* MN908947, campionata il 26 dicembre 2019) da cui è emerso il *clade* 20A che ha dominato l'epidemia europea a marzo diffondendosi poi a livello globale. Dal *clade* 20A sono emersi i *clades* 20B e 20C, geneticamente distinti.

A questi si sono aggiunti altri 6 nuovi *clades* emersi nell'estate del 2020, identificati come 20D-20I, che includono le varianti classificate come "*Variants of concern*" (VoC), quali 20H/501Y.V2 (*lineage* B.1.351, Beta) identificata a ottobre del 2020 in Sud Africa, 20I/501Y.V1 (*lineage* B.1.1.7, Alpha) identificata per la prima volta nel Regno Unito a fine settembre 2020 e 20J/501.V3 (*lineage* P.1, Gamma) identificata in Brasile a dicembre 2020.

I *clades* emersi nel 2021, 21A-21F, includono la VoC Delta (*lineage* B.1.617.2, 21A) identificata nell'ottobre 2020 in India e varianti classificate inizialmente come "*Variants of Interest*" (VoI) quali 21D (B.1.525, Eta), 21F (B.1.526, Iota), 21B (B.1.617.1, Kappa), 21G (C.37, Lambda).

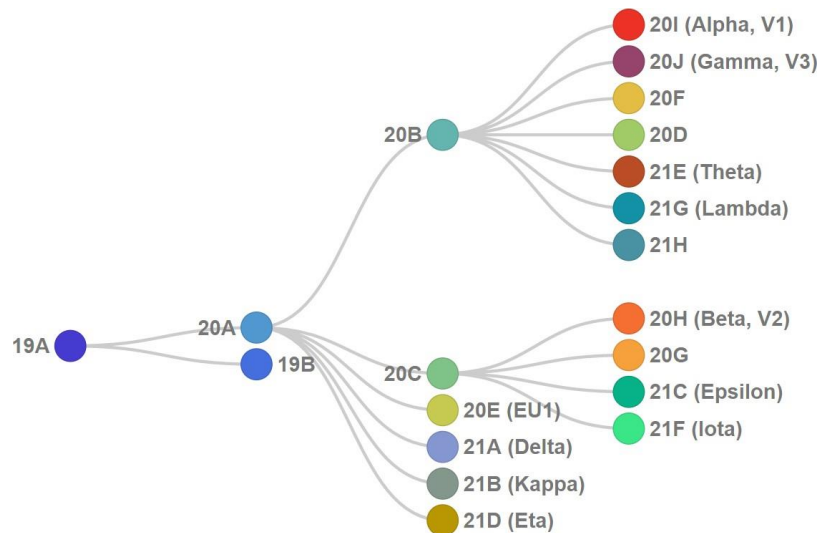


Figura 5. Illustrazione della relazione filogenetica tra i diversi clade di SARS-CoV-2 emersi negli anni 2020-2021, in base alla classificazione di Nextstrain. Fonte: <https://clades.nextstrain.org/>.

A questi si aggiunge la VoC Omicron (*clade* 21M, *lineage* B.1.1.529), inizialmente identificata nel novembre del 2021 in Botswana e Sud Africa, diventata nel corso del 2022 l'unica variante circolante a livello mondiale con i *clades* 21K (BA.1), 21L (BA.2), 22A (BA.4), 22B (BA.5), i *sub-clades* 22C (BA.2.12.1) e 22D (BA.2.75) e le forme ricombinanti XBB (22F) e BQ.1 (22E). I *clades* emersi nel 2023 includono le sottovarianti di Omicron 23A-F (XBB.1.5, XBB.1.16, CH.1.1, XBB.1.9, XBB.2.3 e EG.5.1 (Figura 6).

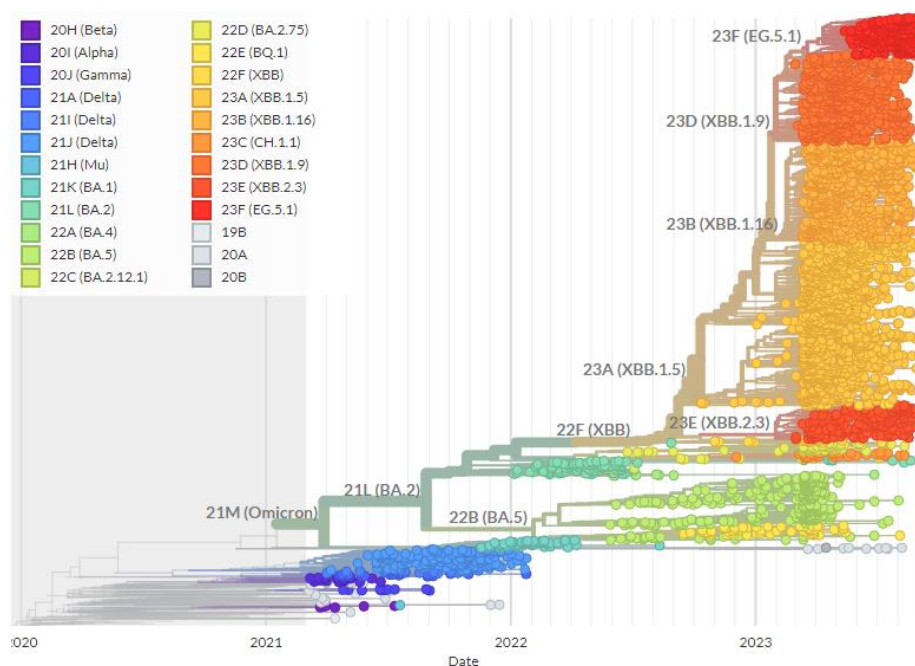


Figura 6. Illustrazione della relazione filogenetica tra i diversi *clade* di SARS-CoV-2 emersi negli anni 2021-2023, in base alla classificazione di Nextstrain. Fonte: <https://clades.nextstrain.org/>.

## 1.4 CLASSIFICAZIONE DELLE VARIANTI DI SARS-CoV-2

L'Organizzazione Mondiale della Sanità, da gennaio 2020, insieme ad altre organizzazioni internazionali, monitora costantemente l'evoluzione di SARS-CoV-2. Nonostante l'intrinseca limitata variabilità genetica, tipica dei *coronavirus* e dovuta all'attività di *proofreading* della proteina non strutturale Nsp14 che assesta il numero di mutazioni intorno a 1-2 al mese, l'ampia e rapida circolazione del virus favorita dalla trasmissione aerea dell'infezione, ha portato ad una rapida e massiccia evoluzione di SARS-CoV-2 in una serie, in continua espansione, di nuove varianti. Sebbene la maggior parte dei cambiamenti abbia un impatto minimo o nullo, alcune mutazioni possono influenzarne le proprietà, come la facilità di diffusione, la severità della malattia, l'efficacia dei vaccini e delle strategie terapeutiche e il rilevamento diagnostico, ed avere così un impatto sulla Sanità Pubblica globale.

A partire dalla fine del 2020, il continuo emergere di varianti che destavano preoccupazione per la Salute Pubblica, ha spinto la loro classificazione in tre grandi gruppi in base alle loro peculiarità e al livello di pericolosità. Vengono pertanto divise in "*Variants of Interest*" (VoI), "*Variants of Concern*" (VoC) e "*Variants Under Monitoring*" (VuM), allo scopo di tenere sotto controllo e orientare la ricerca verso le varianti che sembrano essere più pericolose (<https://who.int/activities/tracking-SARS-CoV-2-variants>). La proteina spike è la regione maggiormente colpita da mutazioni, e le varianti circolanti di SARS-CoV-2 sono classificate sulla base delle variazioni di tale proteina, rispetto al ceppo ancestrale.<sup>13</sup>

Nel maggio 2021, l'OMS assegnava lettere greche alle varianti allora classificate come VoC e alle VoI, mentre da marzo 2023, le lettere greche sono assegnate solo alle *Variants of Concern*, mentre per le *Variants of Interest* vengono utilizzati i sistemi di nomenclatura Nextstrain e Pango (ad esempio XBB.1.5/23A) (<https://www.who.int/publications/m/item/updated-working-definitions-and-primary-actions-for--sars-cov-2-variants>).

Recentemente, gli organismi internazionali OMS ed ECDC hanno aggiornato i loro sistemi di classificazione delle varianti di SARS-CoV-2, per valutare in modo indipendente i molteplici *lineages* discendenti di Omicron in circolazione e classificare le nuove varianti in modo più chiaro. La classificazione ora in vigore non prevede *Variants of Concern* (VoC), ma solo *Variants of Interest* (VoI) e *Variants under Monitoring* (VuM), come conseguenza della situazione epidemiologica stabile (<https://www.epicentro.iss.it/coronavirus/sars-cov-2-monitoraggio-varianti-rapporti-periodici>).

### 1.4.1 VARIANTS OF INTEREST (VOI)

Le varianti di interesse sono contraddistinte da cambiamenti genetici prevedibili o noti che influenzano le caratteristiche del virus quali trasmissibilità, gravità della malattia, *immune escape*, efficacia diagnostica o terapeutica. Hanno un impatto sulla situazione epidemiologica rappresentando un rischio emergente per la Salute Pubblica Globale.

Tra le Voi, nel corso del 2021 vennero incluse le varianti Epsilon (B.1.427), Zeta (P.2), Eta (B.1.525), Theta (P.3), Iota (B.1.56), Kappa (B.1.617.1), Lambda (C.37) e Mu (B.1.621).

A fine novembre 2022, nessuna variante viene inclusa dall'OMS nelle Voi, mentre L'ECDC include in questa categoria la variante Omicron BQ.1, definita *Cerberus*, con mutazioni S:K444T e S:N460K, e la variante Omicron BA.2.75, definita *Centaurus*, osservata per la prima volta in India a maggio 2022. Per entrambe non vi sono però evidenze di una maggiore trasmissibilità e di una maggiore severità della malattia ma hanno ambedue un impatto sull'immunità.

A settembre 2023, l'OMS ha incluso tra le Voi due forme ricombinanti BA.2.10.1/BA.2.75 quali XBB.1.5 (23A, nominata *Kraken*), con mutazione addizionale S:F486P, e XBB.1.16 (23B, denominata *Arcturus*) con le mutazioni aggiuntive S:E180V, S:K478R e S:F486P; a queste si aggiunge la variante EG.5, chiamata anche *Eris*, con mutazione addizionale S:F456L. A livello globale, EG.5 risulta la variante di interesse più diffusa, rappresentando il 26.1% delle sequenze depositate nella settimana tra il 7 e il 13 agosto 2023, superando il ricombinante XBB.1.16, la cui prevalenza era del 22.7% nella stessa settimana. La variante XBB.1.5, continua a mostrare un *trend* in calo.

### 1.4.2 VARIANTS OF CONCERN (VOC)

Una variante viene definita “variante di preoccupazione” quando, oltre a rispettare i requisiti che la classificherebbero tra le “varianti di interesse”, mostra un aumento della trasmissibilità o un cambiamento nell'epidemiologia, oppure un aumento della virulenza o un cambiamento nel quadro clinico della malattia, o ancora una diminuzione dell'efficacia delle terapie, dei vaccini e della diagnosi.

Attualmente, l'unica variante classificata come VoC è la variante Omicron (B.1.1.529) insieme ai *sub-lineage* BA.1, BA.2, BA.3, BA.4, BA.5 e discendenti e le forme ricombinanti BA.1/BA.2 come il ricombinante XE. Una “variante di interesse” o una “variante di preoccupazione” che ha dimostrato in modo chiaro di non rappresentare più un rischio per la salute pubblica può essere

riclassificata come “*previously circulating VoC/Vol*”, riflettendo in questo modo il forte declino della circolazione della variante.

Fino a marzo 2022 rientravano in questa categoria le varianti Alpha (B.1.1.7), Beta (B.1.351) e Gamma (P.1). La variante Delta (B.1.617.2) è stata riclassificata nel mese di giugno 2022.

### 1.4.3 *VARIANT UNDER MONITORING (VUM)*

Le “varianti sotto monitoraggio” sono varianti che presentano cambiamenti genetici che si sospetta possano influenzare le caratteristiche del virus e che potrebbero pertanto rappresentare un rischio per il futuro. Per queste varianti non ci sono ancora prove chiare sull’impatto fenotipico o epidemiologico, e quindi richiedono un monitoraggio potenziato e una valutazione continua. Le varianti precedentemente classificate come Vol/VoC possono essere monitorate nel corso del tempo nell’ambito di questa categoria.

A settembre 2023, sono classificate come tali le varianti BA.2.75 (22D) discendente di BA.2, CH.1.1 (22D, definita Orthrus) derivata da BA.2.75 e identificata per la prima volta nel Regno Unito nel novembre 2022, la variante XBB\* (22F) e le tre forme ricombinanti BA.2.10.1/BA.2.75 quali XBB.1.9.1 (23D), XBB.1.9.2 (23D) e XBB.2.3 (23E).

Dal 17 agosto 2023, nonostante la limitata disponibilità di sequenze, rientra in questa categoria anche la variante BA.2.86, nota come Pirola, caratterizzata da un elevato numero di mutazioni addizionali nella proteina spike rispetto al *sub-lineage* parentale BA.2 ed ai *lineages* XBB attualmente circolanti, il cui impatto è attualmente ancora in fase di valutazione. Al 30 agosto 2023, 21 sequenze di questa variante sono state segnalate da sette Paesi (cinque nella regione europea, uno nella regione africana e uno nella regione delle Americhe) e caricate sulla banca dati GISAID (*Global Initiative on Sharing All Influenza Data*). A fine settembre 2023, questa variante è stata isolata e sequenziata per la prima volta in Italia, a Brescia.

Un solo caso ha avuto storia di viaggio da un paese nella regione del Pacifico occidentale, dove BA.2.86 non è ancora stata segnalata. Ad oggi, non sono stati segnalati decessi all'OMS tra i casi rilevati con questa variante. L'impatto potenziale dell'elevato numero di mutazioni in BA.2.86 è attualmente sconosciuto ed è in fase di valutazione.

La **Tabella 1** mostra il numero di Paesi che segnalano la presenza di Vol e VuM e la loro prevalenza.

**Tabella 1. Prevalenza settimanale delle VoI e VuM, settimana 32-36 2023.**  
<https://www.who.int/publications/m/item/covid-19-epidemiological-update---29-september-2023>.

Lineage	Countries <sup>§</sup>	Sequences <sup>§</sup>	2023-32	2023-33	2023-34	2023-35	2023-36
<b>VOIs</b>							
XBB.1.5*	125	287 491	12.0	10.7	9.7	9.8	8.6
XBB.1.16*	120	69 600	23.5	22.4	21.3	21.3	18.9
EG.5*	73	31 905	25.9	28.1	30.2	30.5	33.6
<b>VUMs</b>							
BA.2.75*	128	128 180	1.7	1.6	1.7	2.0	2.7
BA.2.86 <sup>†</sup>	21	198					
CH.1.1*	99	41 367	0.1	0.1	0.1	0.0	0.0
XBB*	136	76 999	6.2	5.9	5.2	4.7	4.1
XBB.1.9.1*	114	65 528	10.3	9.8	9.5	9.9	8.7
XBB.1.9.2*	92	34 137	6.3	7.6	8.4	8.7	11.2
XBB.2.3*	89	19 158	7.7	7.1	7.1	7.4	6.9
Unassigned	92	149 746	1.3	1.5	1.6	0.8	0.1
Other <sup>‡</sup>	211	6 778 535	4.6	4.7	4.7	4.4	4.7

<sup>§</sup> Number of countries and sequences are since the emergence of the variants.

\* Includes descendant lineages, except those individually specified elsewhere in the table. For example, XBB\* does not include XBB.1.5, XBB.1.16, EG.5, XBB.1.9.1, XBB.1.9.2, and XBB.2.3.

<sup>‡</sup> "Other" represents other circulating lineages excluding the VOI, VUMs, BA.1\*, BA.2\*, BA.3\*, BA.4\*, BA.5\*. Due to delays in or retrospective assignment of variants, caution should be taken when interpreting the prevalence of the "Other" category.

<sup>†</sup> Prevalence for BA.2.86 cannot be calculated due to the very small number of sequences; this is also represented by the shaded cells in the table. The number of sequences and countries were derived from GISAID as of 8:00 AM CEST on 25 September 2023.

#### *Omicron subvariant under monitoring*

L’OMS ha aggiunto nel corso del 2022 questa categoria, attualmente non più in uso, al sistema di tracciamento delle varianti con lo scopo di avvisare le Autorità Sanitarie pubbliche a livello globale, su quali delle sotto-varianti Omicron possano richiedere attenzione e monitoraggio prioritari. La diffusa trasmissione della variante Omicron a livello globale ha portato ad una significativa evoluzione intra-VoC con la comparsa di *lineages* derivati, caratterizzati da *pattern* mutazionali differenti rappresentando infatti un potenziale rischio per la Salute Pubblica globale.

Rientrava in questa categoria una variante che, secondo l'analisi filogenetica, apparteneva ad una VoC in circolazione, mostrava un vantaggio di trasmissione rispetto alla VoC originaria e presentava ulteriori cambiamenti in aminoacidi noti o sospettati di avere un impatto sull'epidemiologia e un vantaggio in termini di *fitness*.

#### **1.4.4 FORMERLY MONITORING VARIANTS**

Le “varianti precedentemente monitorate” (Ex VuM) includono le varianti che sono state riclassificate in quanto non circolano più ad elevati livelli per essere definite di importanza globale per la salute pubblica oppure circolano da molto tempo senza alcun impatto sulla situazione epidemiologica globale.

## 1.5 PRINCIPALI *LINEAGES*/VARIANTI CIRCOLANTI DA AGOSTO 2020 AD OGGI

### 1.5.1 *CLADE* 20E.EU1

Il *clade* 20E.EU1 (EU1), conosciuto anche come *lineage* B.1.177, è stato identificato in Spagna all'inizio dell'estate 2020 per poi diffondersi in tutta l'Europa occidentale, diventando il *lineage* più diffuso. Il *cluster* è stato inizialmente identificato come un gruppo monofiletico di sequenze derivanti dal *clade* 20A con sostituzioni amminoacidiche nelle posizioni S: A222V, ORF10: V30L e N: A220V, corrispondenti alle mutazioni nucleotidiche C22227T, C28932T e G29645T. La mutazione non sinonima dell'amminoacido 222 nella proteina spike da alanina a valina (A222V) è una mutazione localizzata nel dominio N-terminale (NTD), riscontrata successivamente anche nella variante Delta, suggerendone un effetto funzionale. Inoltre, il *clade* 20E.EU1 (EU1) differisce dai suoi antenati per le mutazioni sinonime T445C, C6286T e C26801G.<sup>14</sup> Sebbene non ci siano evidenze di un'aumentata trasmissibilità, si ritiene che condizioni epidemiologiche favorevoli quali l'allentamento delle misure di contenimento, la ripresa dei viaggi e la diminuzione di *screening* nella popolazione, siano state responsabili del successo di questo *lineage*, causa della seconda ondata epidemica nel nostro Paese nell'autunno del 2020.

### 1.5.2 VARIANTE ALPHA (B.1.1.7)

Identificata per la prima volta a dicembre 2020 in alcune regioni Sud-orientali del Regno Unito, anche se studi retrospettivi ne hanno dimostrato la circolazione già due mesi prima con un rapido incremento nelle settimane successive, la variante Alpha (*lineage* B.1.1.7, *clade* 20I/501Y) presenta 14 cambiamenti genetici localizzati prevalentemente nella proteina spike. Queste mutazioni includono la mutazione N501Y, situata a livello di uno dei residui chiave nel sito RDB, correlata ad un aumento dell'affinità della proteina per il recettore. La seconda mutazione, P681H, si localizza in uno dei quattro residui responsabili della formazione del sito di clivaggio della furina, tra le subunità S1 e S2, necessario per promuovere l'ingresso del virus nelle cellule epiteliali respiratorie. Queste due mutazioni erano già state osservate in altre varianti, ma non erano mai state identificate contemporaneamente. La variante Alpha presenta, inoltre, la delezione di due amminoacidi consecutivi nei siti 69-70 del gene S, in aggiunta alla  $\Delta Y144$ , nel dominio N-terminale della spike.<sup>15</sup> Diversi studi ne hanno suggerito una trasmissibilità superiore al 50% rispetto agli altri ceppi. La sua comparsa è stata da subito associata ad un forte incremento delle infezioni diventando la variante predominante a livello globale. È stato inoltre evidenziato un potenziale



aumento della severità della malattia in base ai tassi di ospedalizzazione ed ai quadri clinici riscontrati.<sup>16, 17</sup> L'efficacia del trattamento con anticorpi monoclonali è risultato efficace, mentre è stata riscontrata una minima diminuzione della neutralizzazione con siero convalescente o post-vaccinazione (<https://www.who.int/> (2022)).

### 1.5.3 VARIANTE DELTA (B.1.617.2)

La variante Delta (*lineage* B.1.617.2, *clade* 20A/S:478K), è stata identificata per la prima volta nell'ottobre 2020 in India. Il primo isolamento è avvenuto da un tampone nasofaringeo di un paziente sintomatico ritornato in Francia dopo un viaggio in India. Dalla sua identificazione, questo *lineage* è diventato dominante prima in diverse regioni dell'India, per poi diffondersi in molti Stati, fino a sovrastare da giugno 2021, la variante Alpha, allora predominante. Il *lineage* B.1.617, derivato dal B.1, comprende tre sottotipi principali: B.1.617.1 (variante Kappa), B.1.617.2 (variante Delta) e B.1.617.3. Presenta dieci mutazioni/delezioni nella spike: T19R, G142D, E156G, Δ157-158, L452R, T478K, E484Q, D614G, P681R, D950N. Tra queste, le mutazioni L452R e E484Q sembrano essere associate all'*escape* immunitario ed avere un ruolo nella stabilizzazione dell'interazione proteina spike-recettore ACE-2 aumentando l'infettività del virus, mentre la mutazione P618R, trovandosi vicino al sito di scissione della furina, sembra essere coinvolta nell'aumento della trasmissibilità. L'elevata trasmissibilità è da attribuirsi anche alla maggiore quantità di RNA virale riscontrabile nelle vie aeree dei pazienti, dovuta ad una maggiore permanenza rispetto al ceppo *wild type*.<sup>18, 19</sup> È risultata inoltre associata ad un aumentato rischio di ospedalizzazione in particolare tra individui parzialmente vaccinati o non vaccinati.

Questa variante presenta un numero elevato di *sub-lineages* discendenti, denominati AY.<sup>20</sup>

Nel rapporto dell'Istituto Superiore di Sanità del mese di ottobre 2021, relativo alla prevalenza delle varianti in Italia, il 98.4% dei tamponi sequenziati è risultato positivo alla variante Delta.

Quasi tutti i ceppi designati come Delta sono suscettibili ai trattamenti con anticorpi monoclonali, con l'eccezione dei ceppi derivanti AY.1 e AY.2.

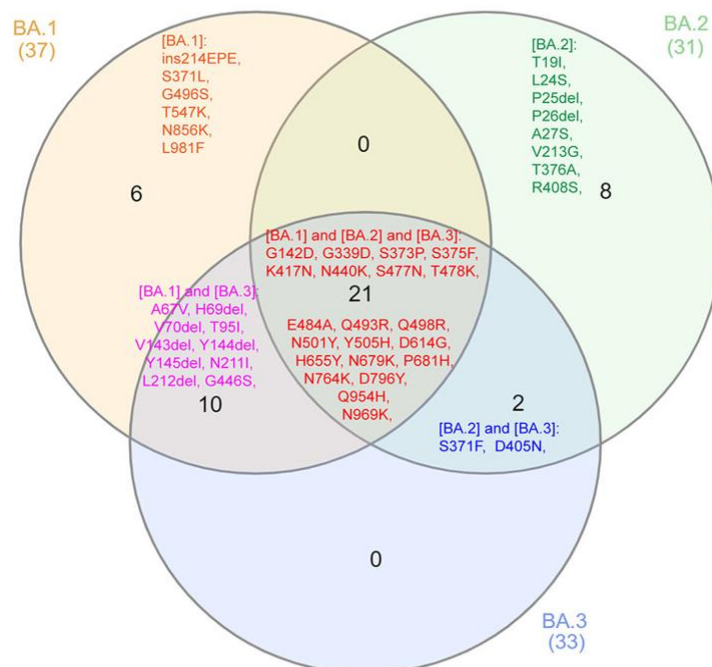
### 1.5.4 VARIANTE OMICRON (B.1.1.529)

Identificata per la prima volta a novembre 2021 in Sud Africa e Botswana, la variante Omicron (*lineage* B.1.1.529/*clades* 21K, 21L e 21M) è caratterizzata da un elevato numero di mutazioni, oltre 30, e delezioni a livello del gene S (A67V, Δ69-70, T95I, G142D, Δ143-145, N211I, Δ212, ins215EPE, G339D, S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R,

G496S, Q498R, N501Y, Y505H, T547K, D614G, H655Y, N679K, P681H, N764K, D796Y, N856K, Q954H, N969K, L981F), un numero mai riscontrato nelle precedenti varianti circolanti.

Studi sui potenziali effetti della variante confermano un forte impatto sull'aumento della trasmissibilità del virus rispetto al ceppo *wild type*, alla base dell'incremento del numero di casi di positività, riscontrato da dicembre 2021. Caratteristica peculiare che differenzia la variante Omicron dalle altre è la sintomatologia simil-influenzale, stanchezza, prurito che interessa naso e gola, dolore alla schiena e tachicardia, assenti nelle altre varianti.

La variante è stata inizialmente suddivisa in 3 *lineages* (BA.1, BA.2 e BA.3) con sostanziali differenze genomiche, in particolare tra Omicron BA.1 (21K) e BA.2 (21L) (<https://www.who.int/>) (**Figura 7**).



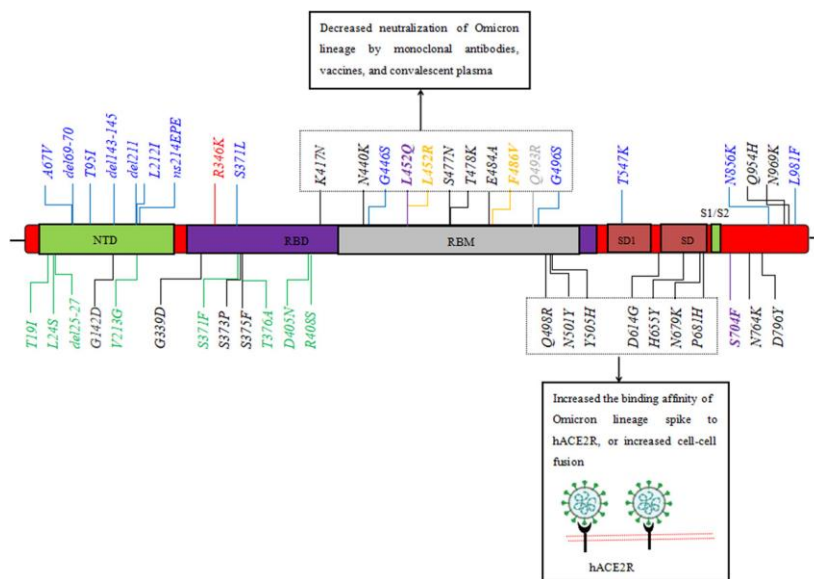
**Figura 7. Mutazioni nella proteina spike delle varianti Omicron BA.1, BA.2 e BA.3.**

La prima sequenza BA.2 è stata resa disponibile dal Sud Africa il 27 novembre 2021. La proteina spike di questo *lineage* contiene 29 sostituzioni aminoacidiche e due delezioni (T19I, L24S, P25del, P26del, A27S, G142D, V213G, G339D, S371F, S373P, S375F, T376A, D405N, R408S, K417N, N440K, S477N, T478K, E484A, Q493R, Q498R, N501Y, Y505H, D614G, H655Y, N679K, P681H, N764K, D796Y, Q954H, N969K). Nonostante condividano 21 mutazioni aminoacidiche nella proteina spike (**Figura 7**), le varianti Omicron BA.1 e BA.2 differiscono complessivamente per circa 40 mutazioni e delezioni, tra cui le mutazioni chiave nell'N-terminale e nel dominio di legame del recettore del gene *spike*, regioni che influenzano la risposta immunitaria. Rispetto a BA.1, il *sub-lineage* BA.2

presenta un maggior tasso di replicazione nelle cellule epiteliali nasali umane e studi *in-vivo* ne hanno mostrato inoltre una maggiore carica di RNA virale nello stesso *timepoint*.<sup>21</sup> Data la maggiore trasmissibilità, la variante BA.2 ha raggiunto a livello globale una maggiore prevalenza rispetto al BA.1, soppiantandolo. Per la variante Omicron BA.3, identificata per la prima volta in Sudafrica nel novembre del 2021, non si hanno invece evidenze su trasmissibilità, immunità e severità della malattia.

Da gennaio 2022 sono state segnalate in Sud Africa i *lineages* BA.4 (22A) e BA.5 (22B) con lo stesso profilo di mutazioni della variante BA.2 nel gene S, ma diverse sostituzioni nel restante genoma. Alcuni studi mostrano l'associazione di questa variante ad un aumentato rischio di reinfezione, di circa 5.4 volte più alto rispetto alla variante Delta (<https://www.imperial.ac.uk/media/imperial-college/medicine/mrc-gida/2021-12-16-COVID19-Report-49.pdf>) e *breakthrough*.<sup>22</sup>

Tutti i 5 *lineages* di Omicron condividono 20 mutazioni nella proteina spike (G142D, G339D, S373P, S375F, K417N, N440K, S477N, T478K, E484A, Q498R, N501Y, Y505H, D614G, H655Y, N679K, P681H, N764K, D796Y, Q954H, N969K) (**Figura 8**). Studi *in vitro* hanno indicato come le mutazioni, D614G, H655Y, N679K, P681H, N501Y, Q498R e la mutazione Y505H potrebbero aumentare l'affinità di RBD all'enzima ACE2, incrementando così l'infettività delle varianti Omicron.<sup>23</sup> Le mutazioni K417N, N440K, S477N, T478K, E484A, e N501Y sono associate ad *immune-escape* e studi *in vitro* hanno confermato il dato rispetto alla capacità neutralizzante di sieri convalescenti, alcuni anticorpi monoclonali (mAbs) e sieri immuni.<sup>23,24</sup>



**Figura 8.** Mutazioni rilevanti nella proteina spike proprie dei differenti *lineages* di Omicron BA.1 (in blu), BA.1.1 (in rosso), BA.2 (in verde), BA.2.12.1 (in viola), BA.4 e BA.5 (giallo). In nero sono indicate le 20 mutazioni aminoacidiche comuni ai 5 *lineages*. Fonte: *J Med Virol*, 2023.<sup>23</sup>

Le varianti Omicron non solo hanno accumulato un gran numero di mutazioni nella proteina spike, ma anche nelle regioni ORF1ab, N, E e M (inclusi NSP1: S135R; PLpro:T24I e G489S; NSP4: L264F, T327I, L438F, T492I; 3CL: P132H; NSP6: del105-108 e F108L; NSP12: P323L; NSP13: R392C; NSP14: I42V; NSP15: T112I; OR3a: T223I; E: T91I; M: Q19E, N: P13L, del31-33, R203K, G204R, S413R).

Complessivamente, i diversi *lineages* di Omicron presentano il maggior numero di mutazioni in tutto il genoma, responsabili dell'aumento della trasmissibilità e dell'*escape* immunitario di questa variante rispetto al WT e alle precedenti varianti Alpha, Beta, Gamma, Delta.

Omicron è stata la variante dominante a livello globale nel corso del 2022, rappresentando oltre il 98% delle sequenze virali condivise su GISAID dopo febbraio 2022 (<https://www.who.int/activities/tracking-SARS-CoV-2-variants>). Dati di sorveglianza indicano come dall'inizio di luglio 2022, le varianti BA.4/5 e nuovi *lineages* discendenti insieme alle forme ricombinanti, stiano guidando la nuova ondata epidemica. Tra questi, BA.5.2, BA.5.2.6, BA.5.3.1.1.1.1.1 (BQ.1), BA.5.3.1.1.1.1.1.1 (BQ.1.1), BA.4.6, BA.5.2.1.7 (BF.7) e BA.5.1.7, insieme ai *lineages* di BA.2 quali BA.2.75 e BA.2.75.2 e ai ricombinanti intra-BA.2 XBB\*. (WHO.<https://www.who.int/news/item/27-10-2022-tag-ve-statement-on-omicron-sublineages-bq.1-and-xbb>).

### 1.5.5 RICOMBINANTI DI SARS-CoV-2

Con la progressione della pandemia e l'evoluzione di *lineages/sub-lineages* di SARS-CoV-2 geneticamente divergenti, la ricombinazione genetica è il meccanismo evolutivo chiave che sta continuamente "rimodellando" la diversità genetica SARS-CoV-2. Alcune tra le varianti che destano preoccupazione sono infatti il risultato della ricombinazione tra *lineages* differenti che co-circolano e replicano simultaneamente in individui suscettibili all'infezione.<sup>25</sup>

Tra i ricombinanti inter-VoC sono stati identificati, nel corso del 2022, il ricombinante XC, derivato dalla ricombinazione tra la variante Delta AY.29 e la variante Alpha B.1.1.7, e 5 ricombinanti generati dalle varianti Delta e Omicron, denominati "Deltacron" (XD, XF, XS, XAW e XAY).<sup>26</sup>

Molteplici sono stati gli eventi di ricombinazione *intra-Omicron* tra le varianti BA.1 e BA.2 (XE, XG, XH, XJ, XK, XL, XN, XT, XU, XV, XW, XY, XAA, XAB, XAC, XAF, XAG, XAH, XAK, XAP, XAQ, XAR), BA.1.1 e BA.2 (XM, XP, XQ, XR, XZ, XAD, XAE), BA.2 e BA.4 (XAJ), BA.2 e BA.5 (XAV) e BA.4 e BA.5 (XAS). In aggiunta, le varianti XAM e XAU derivano dalla combinazione di BA.1.1 e BA.2.9, mentre XAL, XAN e XAT rispettivamente dalla ricombinazione tra BA.2.12.1/BA.4, BA.2/BA.5.1 e BA.2.3.13/BA.1. La ricombinazione tra BA.2.5, BA.5 e BA.2.5 ha dato origine alla variante XAZ.

Nell'ottobre del 2022, l'OMS registrava un aumento del nuovo *lineage* ricombinante XBB e dei suoi *sublineages* (indicati come XBB\*), ricombinanti di *sub-lineages* BA.2.10.1 e BA.2.75.

I processi di ricombinazione, che contribuiscono alla diversità genetica dei virus a RNA, rappresentano un fenomeno diffuso e frequente anche nei coronavirus umani e animali e quindi, sebbene alcuni ricombinanti di SARS-CoV-2 abbiano mostrato una circolazione limitata a livello di popolazione, rimane possibile che in futuro emergano nuovi *lineages* ricombinanti più trasmissibili e/o patogeni.<sup>25</sup>

## 1.6 EPIDEMIOLOGIA

Come per tutte le pandemie, anche per quella causata da SARS-CoV-2 è stata essenziale la definizione di alcuni parametri epidemiologici fondamentali, che hanno permesso la valutazione dell'andamento della pandemia.

Uno dei primi parametri che necessitava di essere definito era il tasso riproduttivo di base ( $R_0$ ), inteso come numero di casi secondari attesi da un singolo soggetto infetto, in una popolazione completamente suscettibile all'infezione. Il tasso riproduttivo di base dipende dalla durata dell'infettività, dalla probabilità di trasmissione tra un soggetto infetto e uno suscettibile e dal tasso di attacco del virus. Quando il valore di  $R_0$  è inferiore a 1 l'epidemia tende ad esaurirsi, se il valore di  $R_0$  è uguale a 1 la situazione diventa stazionaria, invece, se il valore di  $R_0$  è maggiore di 1 l'epidemia crescerà. Durante il corso della pandemia, il valore di  $R_0$  ha subito diverse variazioni nella stima, in quanto è strettamente correlato alla capacità di trasmissione del singolo individuo.

Un secondo parametro associato a  $R_0$  è il tasso di riproduzione effettivo ( $R_t$ ), che rappresenta il numero medio di persone che è in grado di contagiare un soggetto infetto nel tempo ( $t$ ). Questo parametro serve a definire la trasmissibilità del virus, indicando come cambia la sua contagiosità in una determinata area al variare del tempo. Questo parametro è utile per capire se le misure messe in atto per contrastare l'epidemia sono efficaci o meno e, analogamente al parametro  $R_0$ , varia nel corso della pandemia in relazione al "*replacement*" di una variante con la successiva, con picchi corrispondenti alle differenti ondate epidemiche.<sup>27, 28</sup>

Un ulteriore parametro da valutare è il tasso di letalità (numero di decessi sul totale di ammalati), che definisce la gravità dell'epidemia. Nel caso di SARS-CoV-2, a causa dell'elevato numero di infezioni asintomatiche che non permettono di stabilire in modo certo la popolazione infetta, questo valore non risulta facile da calcolare, ma si attesta intorno a 1.5%.<sup>29</sup>

## 1.7 LA PANDEMIA DA COVID-19 NEL MONDO

Durante i primi tre mesi della pandemia, quasi due terzi di tutti i casi di infezione segnalati nei singoli Paesi, riguardavano soggetti che riportavano viaggi recenti in soli tre stati quali Cina, Iran e Italia, a dimostrazione di come gli spostamenti internazionali in luoghi con un'elevata circolazione del virus possano avere avuto un impatto enorme sulla diffusione del patogeno.

L'eterogenea situazione epidemiologica tra i continenti è risultata correlata all'adozione di diverse strategie di contenimento da parte dei vari Stati (Statista. <https://www.statista.com/statistics/1102209/coronavirus-cases-development-europe>).

A partire dalla seconda metà di febbraio 2020, in Europa si è verificata una prima ondata di contagi, che hanno interessato in particolare Italia, Francia, Germania, Regno Unito e Spagna, riconducibili a soggetti di ritorno dalla Cina. L'Italia è stato il primo Paese europeo a documentare un caso di infezione da SARS-CoV-2 non riconducibile a un viaggio, nonché il primo a registrare la prima ondata della pandemia.

Da marzo 2020 ad oggi, si sono alternate fasi caratterizzate da un aumento del numero di nuovi casi di infezione, fasi con picchi di contagio e fasi che registravano una diminuzione del numero di nuovi casi. In particolare, con il periodo estivo in molti degli Stati europei si è verificata una "fase di transizione" caratterizzata da una diminuzione della circolazione di SARS-CoV-2. La conseguente diminuzione delle misure di contenimento ha determinato, all'inizio dell'autunno 2020, la seconda ondata della pandemia, caratterizzata dalla diffusione della variante Alpha di SARS-CoV-2 (*lineage* B.1.1.7), dotata di una più elevata trasmissibilità rispetto al ceppo WT, con una conseguente crescita delle curve dei contagi di tutti i Paesi europei.

Con l'avvio della campagna vaccinale, iniziata a gennaio 2021, dapprima in Israele e in USA e subito dopo in gran parte dei Paesi europei, si è osservato nuovamente un decremento del numero di nuovi casi e il conseguente allentamento delle misure di prevenzione. Una terza ondata si è verificata a partire dalla seconda metà di febbraio 2021, con il picco di contagi raggiunto tra marzo e aprile 2021, a causa della diffusione della variante Delta (*lineage* B.1.617.2). Alla fine di luglio 2021, in quasi tutti i Paesi europei si è verificata una situazione simile quella dell'anno precedente a conclusione della prima ondata, con minimi livelli di casi e di decessi. Dall'estate 2021 alla fine del 2022, in Europa si è assistito ad una quarta ondata, caratterizzata dapprima dalla circolazione dei *lineage* discendenti della variante Delta, e da novembre 2021 dalla comparsa della variante Omicron, responsabile dell'esponenziale aumento dei contagi a partire da dicembre 2021.

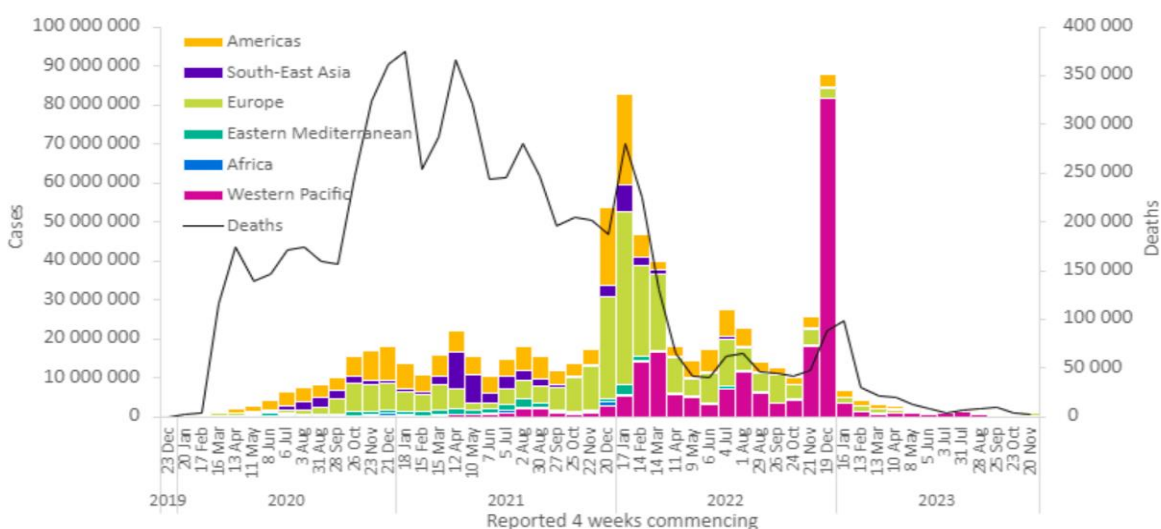
Questa fase ha avuto un andamento differente nei Paesi interessati, sulla base del livello e della durata di copertura vaccinale della popolazione e delle misure di contenimento adottate.

Il WHO (<https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19>), al 23 gennaio 2022, riportava un aumento a livello globale del numero di infezioni con oltre 21 milioni di nuovi casi confermati. Il maggior numero di nuove diagnosi di infezione è stato rilevato nelle regioni del Mediterraneo orientale, nel Sud-Est asiatico, in Europa e in America.

A settembre 2022 Omicron, con i *sub-lineage* BA.4 e BA.5, rappresentava la variante predominante circolante in Europa e nel mondo ([www.gisaid.com](http://www.gisaid.com)).

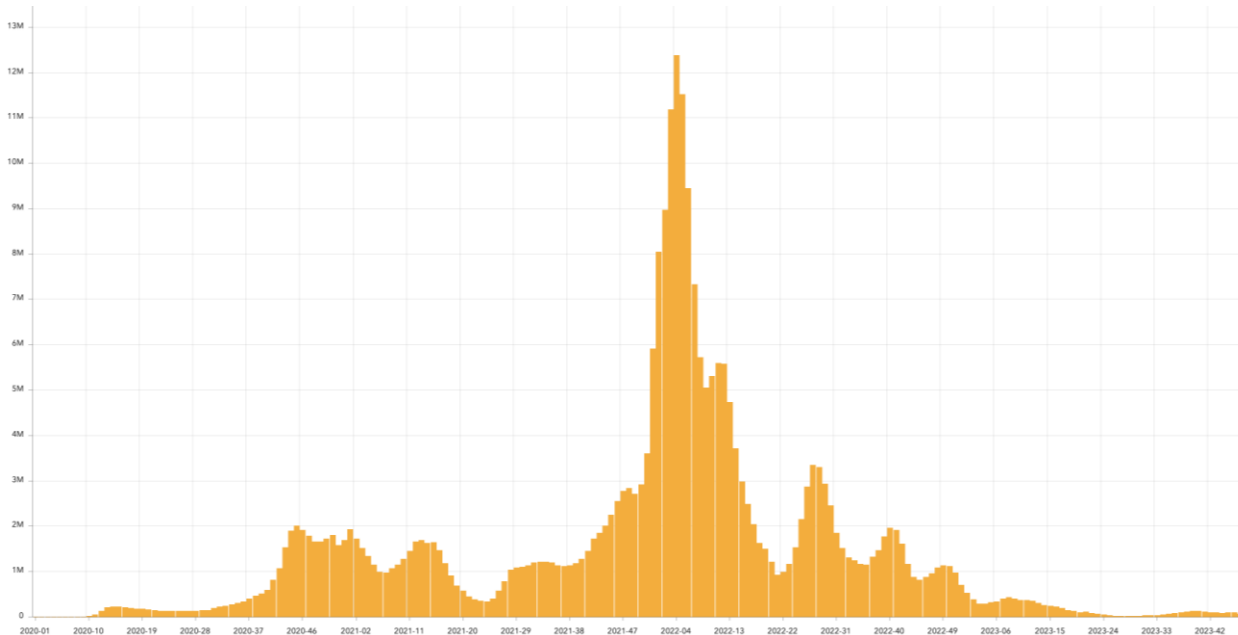
Ad oggi, le forme ricombinanti XBB\* rappresentano il 96.3% delle varianti globalmente circolanti (<https://cov-spectrum.org/explore/World/AllSamples/Past6M>); XBB.1.16 e XBB.2.3, continuano a mostrare una circolazione sostenuta in tutti i continenti, *trend* confermato anche nel nostro Paese, a fronte di una diminuzione della variante BQ.1; da agosto 2023 è risultata in aumento la variante EG.5, discendente di XBB.1.9.2, rappresentando, a fine settembre 2023, la VoI più prevalente in 73 Paesi con il 33.6% delle sequenze depositate in GISAID (20230928\_covid-19\_epi\_update.pdf). La variante FL.1.5.1, definita *Fornax* e descritta in Perù, rappresenta il 6% delle sequenze depositate in GISAID.

Dati del WHO (<https://www.who.int>), aggiornati al 17 dicembre 2023 riportano, tra il 20 novembre e il 17 dicembre 2023, 850 mila nuovi casi di COVID-19 con un incremento del 52% rispetto al mese precedente e oltre 3,000 morti (con decremento dell'8% in comparazione al mese precedente). I casi riportati globalmente sono oltre 770 milioni con oltre 6.8 milioni di decessi (**Figura 9**).



**Figura 9.** Casi da COVID-19 e decessi globali riportati settimanalmente, aggiornati al 17 dicembre 2023. *Fonte* <https://www.who.int/publications/m/item/covid-19-epidemiological-update---22-december-2023>.

In Europa, i casi confermati si attestano oltre 278 milioni con più di 2 milioni di decessi (<https://who.maps.arcgis.com/apps/dashboards/ead3c6475654481ca51c248d52ab9c61>, dati aggiornati al 21 novembre 2023) (**Figura 10**).



**Figura 10.** Numero di nuovi casi di COVID-19 in Europa dal 20 gennaio 2020 al 21 novembre 2023. *Fonte:* <https://who.maps.arcgis.com/apps/dashboards/ead3c6475654481ca51c248d52ab9c61>.

## 1.8 SITUAZIONE IN ITALIA

L'Italia è stato il primo Paese europeo ad essere colpito dalla pandemia. Il 30 gennaio 2020 sono stati confermati i primi due casi di importazione della COVID-19, in due turisti cinesi in viaggio a Roma. Il 17 febbraio 2020 invece è stato confermato il primo caso autoctono a Codogno, in provincia di Lodi. In seguito, diversi studi hanno ipotizzato la circolazione di SARS-CoV-2 già nei mesi precedenti (novembre-dicembre 2019).<sup>30, 31</sup>

L'infezione si è diffusa in breve tempo in tutte le regioni italiane, a partire proprio dal nord Italia. Analisi condotte sulle sequenze ottenute dai campioni dei due turisti cinesi e da quelle dell'epidemia in Lombardia non hanno evidenziato alcuna correlazione. Al contrario, i ceppi virali dei pazienti di Codogno sono risultati legati a un ceppo di SARS-CoV-2 proveniente da Shanghai che ha causato un piccolo focolaio a Monaco di Baviera attorno al 20 gennaio 2020 e che si è poi diffuso in altri paesi europei e oltreoceano. Queste sequenze facevano parte di un *clade* definito europeo e caratterizzato dalla mutazione D614G.<sup>32</sup>

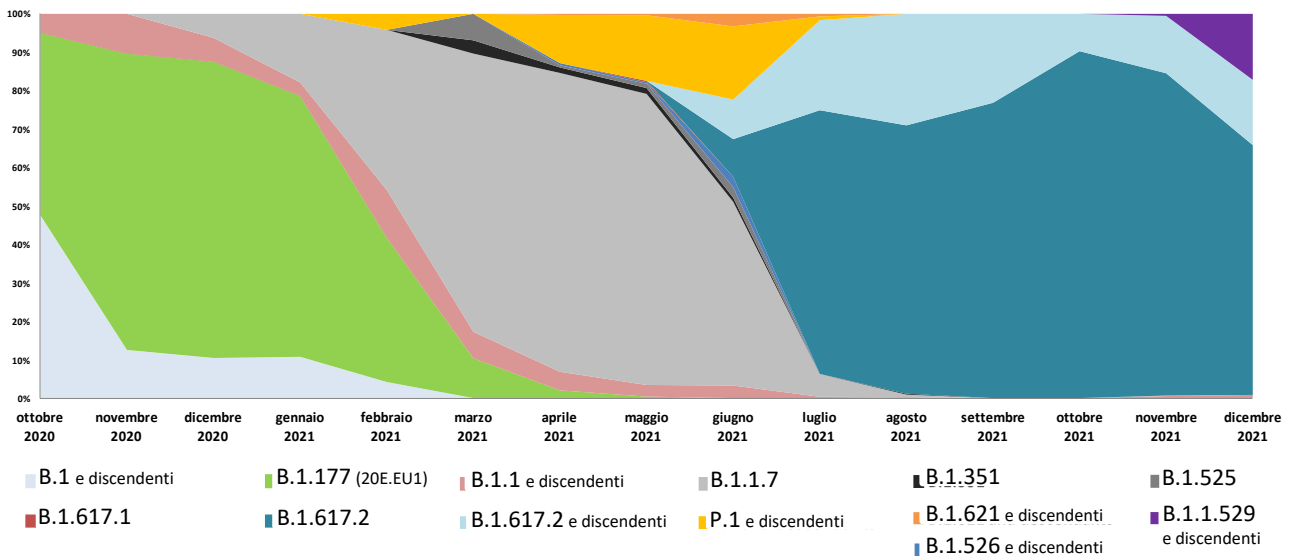
Poche settimane dopo, l'introduzione del *lockdown* (9 marzo 2020) e l'adozione di misure di contenimento quali la quarantena per i viaggiatori provenienti da zone a rischio, hanno portato ad una riduzione del numero di casi nel nostro Paese. Analogamente a quanto accaduto in Europa,



anche nel nostro Paese l'epidemia ha avuto un andamento irregolare, con una prima fase da inizio marzo a fine di maggio 2020, caratterizzata dalla circolazione del ceppo B.1 e discendenti, che ha visto una rapida diffusione di casi e morti prevalentemente nel nord Italia; una fase di transizione, nel periodo estivo fra giugno e metà settembre 2020, caratterizzata da una diffusione contenuta, e una seconda fase a partire dalla fine di settembre 2020 dove i casi sono aumentati esponenzialmente fino all'adozione di nuove misure straordinarie di contenimento (DPCM 3 novembre 2020). Il principale *lineage* circolante in questo periodo è stato il B.1.177 (*clade* 20E.EU1) che rappresentava, insieme ai *lineages* discendenti il ceppo predominante a partire da ottobre 2020 essendo già globalmente diffuso in Europa da giugno 2020.<sup>33</sup>

Il 27 dicembre 2020 in Italia, come in tutta Europa, sono state somministrate le prime dosi di vaccino anti-COVID-19 BNT162b2. A gennaio 2021, poiché la diffusione dell'infezione virale risultava ancora estremamente elevata, veniva prorogato lo stato di emergenza e con esso la reintroduzione di misure restrittive per il suo contenimento. Nel periodo compreso tra la fine di febbraio 2021 e la fine di maggio 2021, la variante Alpha rappresentava oltre l'80% di tutte le sequenze depositate nei *database* pubblici, quali GenBank e GISAID.

Nello studio condotto da *Lai et. al.*,<sup>33</sup> tale variante è stata osservata per la prima volta in Italia in Lombardia a fine dicembre, in linea con l'identificazione dei primi dati retrospettivi provenienti dal Regno Unito e con quelli segnalati dall'Istituto Superiore di Sanità. L'aumento più rilevante della prevalenza di questa variante è stato osservato a partire dalla seconda settimana di gennaio 2021 fino alla fine del mese, raggiungendo oltre il 70% alla fine del mese di febbraio (**Figura 11**). La sua introduzione, notoriamente associata a maggiore trasmissibilità rispetto ad altre varianti e, secondo alcuni studi,<sup>34, 35</sup> ad un aumento del tasso di mortalità, corrisponde al secondo aumento rilevante dei contagi registrati in Italia durante la seconda ondata epidemica.

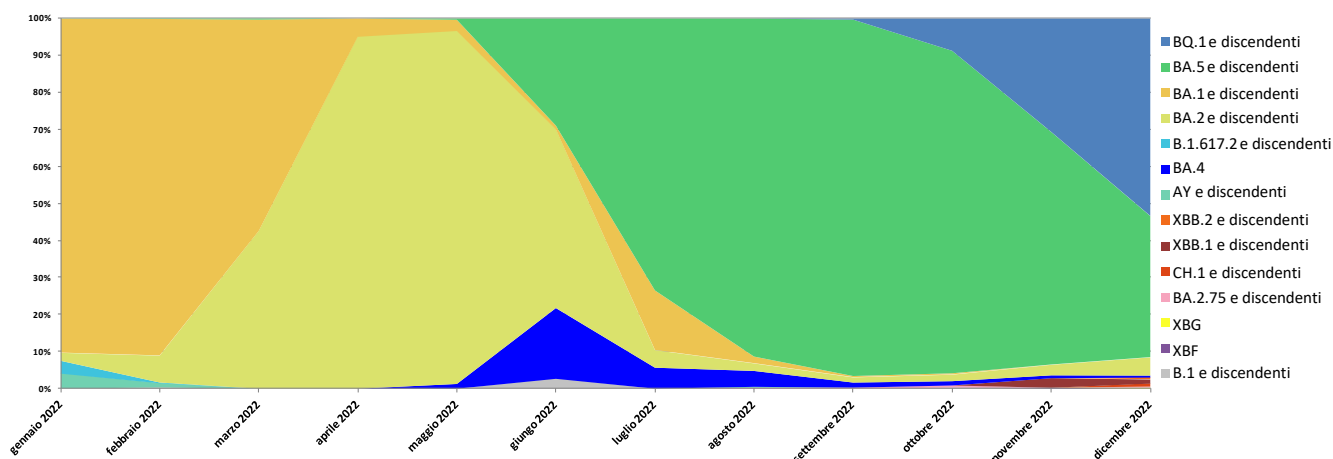


**Figura 11: Distribuzione della prevalenza dei lineages circolanti nel periodo ottobre 2020-aprile 2021. Fonte: modificata da Lai A, 2021 e Lai A, 2022.<sup>33,36</sup>**

A partire da metà giugno del 2021 si osservava in alcune regioni italiane la variante Delta (*lineage* B.1.617.2) che, unitamente ai suoi discendenti, ha rappresentato la principale variante circolante nel nostro Paese fino al mese di dicembre 2021, raggiungendo una prevalenza complessiva di circa l'80% (**Figura 11**).<sup>36</sup> A ottobre 2021, la variante Delta risultava essere quella predominante con una prevalenza di circa il 90%, mentre le altre VoC (Alpha, Beta e Gamma) presentavano ciascuna una prevalenza minore dell'1% ("Prevalenza e distribuzione delle varianti di SARS-CoV-2 di interesse per la sanità pubblica in Italia" ISS, ottobre 2021).

A seguito della diffusione della variante Omicron (B.1.1.529/BA.1) a partire da dicembre 2021, si osservava anche nel nostro Paese un aumento del numero di casi di infezione, caratterizzati nei mesi successivi dalla circolazione di *lineages* derivati (BA.2, BA.3, BA.4 e BA.5) e delle forme ricombinanti inter e intra-VoC.

Dati preliminari relativi all'anno 2022 (**Figura 12**), prodotti dal *network* SCIRE, che vede la collaborazione di numerosi centri clinici e laboratori dislocati su tutto il territorio italiano per la raccolta e la caratterizzazione delle diverse varianti circolanti di SARS-CoV-2, mostrano come nei mesi di gennaio e febbraio 2022, si registravano gli ultimi casi di variante Delta. Nel mese di gennaio, la variante Omicron BA.1 (*clade* 21K) rappresentava il 90.3% delle sequenze caratterizzate nel mese (n=2,482).



**Figura 12. Distribuzione della prevalenza dei *lineages* circolanti nel periodo gennaio 2022-dicembre 2022.**

Nello stesso mese venivano riportati i primi casi di variante Omicron BA.2 e BA.5. Nell'aprile 2022, il 95% delle sequenze apparteneva al *lineage* BA.2 e discendenti mentre a partire da luglio fino a novembre 2022, il *lineage* BA.5 (*clade* 22C) risultava il più prevalente con una proporzione rispettivamente del 73.5%, 91%, 96.4%, 87.2% e 62.9% nei diversi mesi. La variante Omicron BA.4 (*clade* 22A), raggiungeva un picco del 19% delle sequenze totali nel mese di giugno 2022. Il primo caso di variante BQ.1 e discendenti (*clade* 22E) veniva osservato nel mese di settembre, raggiungendo una prevalenza del 50% a dicembre. I primi casi di XBB.1 (*clade* 23A) e XBB.2 (*clade* 22F) e discendenti sono stati osservati a partire dai mesi di novembre e dicembre 2022. Tre casi di variante CH.1 (*clade* 22D) sono stati riportati a dicembre mentre due casi di ricombinanti XBF e XBG sono stati osservati rispettivamente nei mesi di novembre e dicembre.

I dati della Sorveglianza Integrata COVID-19 dell'ISS ([https://www.epicentro.iss.it/coronavirus/bollettino/Bollettino-sorveglianza-integrata-COVID-19\\_13-settembre-2023.pdf](https://www.epicentro.iss.it/coronavirus/bollettino/Bollettino-sorveglianza-integrata-COVID-19_13-settembre-2023.pdf)), aggiornati al 13 settembre 2023, riportano la co-circolazione nell'ultima settimana di agosto, di ceppi virali ricombinanti Omicron XBB, di cui il 13.9% di XBB.1.5, e una circolazione in crescita pari al 43.5%, della variante EG.5, discendente di XBB.1.9.2 e dei suoi *sub-lineages* EG.5.1 e EG.5.1.1.

L'indice di trasmissibilità ( $R_t$ ), basato sui casi con ricovero ospedaliero al 5 settembre 2023 è pari a 1.20 (1.13 – 1.27), in leggero aumento rispetto alla settimana precedente e ancora sopra la soglia epidemica ( $R_t=1.12$  [1.04–1.20], al 29 agosto 2023), con tasso di incidenza settimanale per 100,000 abitanti in aumento, prevalentemente nella fascia di età over 90.

Dall'inizio della pandemia a settembre 2023 in Italia si contano più di 26 milioni di casi e più di 191 mila decessi (<https://www.worldometers.info/coronavirus/country/italy/>).

## 1.9 VACCINI E COPERTURA VACCINALE IN ITALIA

Nonostante il processo per lo sviluppo di vaccini sia lungo e complesso, diverse aziende farmaceutiche si sono sin da subito mobilitate per arrivare, in meno di un anno, allo sviluppo di diversi vaccini per contrastare la pandemia da SARS-CoV-2 (<https://www.epicentro.iss.it/vaccini/>). I vaccini attualmente in commercio contro la COVID-19 sono molteplici e comprendono quelli sviluppati con tecniche convenzionali e altri sviluppati con tecnologie innovative.

**VACCINI A VETTORE VIRALE:** sfruttano dei virus esistenti (generalmente un adenovirus incompetente per la replicazione) per trasportare all'interno della cellula la sequenza genica che codifica per la proteina spike. Il sistema immunitario si attiva contro la proteina e produce degli anticorpi che, qualora il soggetto entrasse a contatto con il virus, lo proteggeranno dall'infezione. I vaccini ottenuti con questa tecnica hanno una buona stabilità, per cui non necessitano di temperature estremamente basse per la conservazione. I vaccini a vettore virale autorizzati da EMA e AIFA sono Vaxzevria (AstraZeneca) e Jcovden (Janssen).

**Vaxzevria (AstraZeneca):** precedentemente chiamato ChadOx1 (*Chimpanzee Adenovirus Oxford 1*) nCoV-19, sviluppato dall'Università di Oxford e AstraZeneca, è stato autorizzato dall'Agenzia Europea del Farmaco (EMA) il 29 gennaio 2021 inizialmente per la somministrazione in soggetti con età pari o superiore ai 18 anni mediante la somministrazione di due dosi a distanza di 4-12 settimane l'una dall'altra. A causa di una maggiore insorgenza di eventi trombotici in donne con età inferiore a 60 anni in seguito alla sua somministrazione, da aprile 2020 questo vaccino è stato somministrato esclusivamente a soggetti con età superiore ai 60 anni. Ai soggetti di età inferiore a 60 anni che avevano già ricevuto la prima dose con questi vaccini è stato consigliato di effettuare una seconda dose con un vaccino a RNA messaggero (mRNA) (vaccinazione eterologa) per il completamento del ciclo vaccinale (<https://www.epicentro.iss.it/vaccini/>). La vaccinazione eterologa è risultata maggiormente efficace determinando una risposta delle cellule T più elevata rispetto all'utilizzo dello stesso vaccino ([www.aifa.gov.it](http://www.aifa.gov.it)).

**Jcovden (Janssen):** (prima chiamato Ad26.COVS.2.S) è un vaccino monovalente sviluppato dalla compagnia farmaceutica *Janssen Pharmaceutical Companies of Johnson & Johnson*, nei Paesi Bassi. Tale vaccino è composto da un vettore ricombinante basato sull'adenovirus umano di tipo 26 incompetente per la replicazione, che codifica per la sequenza completa della glicoproteina spike di SARS-CoV-2 in una conformazione stabilizzata. Dopo la somministrazione, la glicoproteina

S di SARS-CoV-2 viene espressa in maniera transitoria, stimolando sia anticorpi anti-S neutralizzanti che altri anticorpi specifici anti-S funzionali, così come risposte immunitarie cellulari dirette contro l'antigene S.

Inizialmente autorizzato per la somministrazione in soggetti con età superiore a 60 anni, questo vaccino è stato successivamente ampiamente utilizzato per la vaccinazione di soggetti con età pari o superiore a 18 anni. Anche per questi soggetti il completamento del ciclo vaccinale ha previsto la somministrazione di dosi aggiuntive di vaccini a mRNA ([www.aifa.gov.it](http://www.aifa.gov.it)).

**VACCINI A MRNA:** rappresentano i vaccini di più recente sviluppo, agiscono a livello del *citosol* e contengono nella loro formulazione una sequenza di mRNA (RNA messaggero) che codifica per la proteina spike nella conformazione pre-fusione. Questi vaccini hanno il vantaggio di stimolare una maggiore risposta immunitaria e di memoria rispetto a quelli a DNA permettendone la somministrazione in quantità inferiori; mostrano inoltre un miglior profilo di sicurezza. Il segmento di mRNA viene inserito all'interno di nanoparticelle lipidiche che permettono di proteggere e trasportare l'RNA all'interno della cellula. Nonostante l'mRNA del vaccino si degradi pochi giorni dopo la vaccinazione, i linfociti della memoria, che sopravvivono per vari mesi, permettono al sistema immunitario di attivarsi rapidamente nel caso in cui il soggetto vaccinato entri a contatto con il virus. Questa tecnica è stata utilizzata per la produzione dei vaccini Comirnaty (Pfizer) e Spikevax (Moderna) (<https://www.epicentro.iss.it/vaccini/>, [www.aifa.gov.it](http://www.aifa.gov.it)).

**Comirnaty (Pfizer-BioNTech):** Il vaccino Comirnaty (BNT162b2) è stato sviluppato da BioNTech, Fosus Pharma e Pfizer. È stato il primo vaccino ad essere commercializzato contro la COVID-19, autorizzato da EMA il 21 dicembre 2020. È un vaccino che viene somministrato in due dosi a distanza di 21 giorni, inizialmente rivolto ai maggiori di 16 anni, ma dal 28 maggio 2021 approvato anche per la fascia di età 12-15 anni e dal 3 novembre 2021 anche per la fascia 5-11 anni. È ad oggi disponibile una formulazione pediatrica per i bambini di età compresa fra 6 mesi e 4 anni. Questo vaccino richiede una temperatura compresa tra i -90°C e i -60°C per la conservazione e può rimanere tra i 2°C e gli 8°C per 5 giorni. Con l'insorgenza della variante Omicron è stata necessaria la revisione della composizione di questo vaccino, che ha portato a tre nuove formulazioni, denominate Comirnaty Original/Omicron BA.1, Comirnaty Original/Omicron BA.4-5 e Comirnaty Original/Omicron XBB.1.5, somministrati a soggetti con età superiore ai 12 anni che hanno precedentemente ricevuto almeno un ciclo vaccinale primario. Sono disponibili anche formulazioni pediatriche per i bambini di prima infanzia di età pari o superiore a 6 mesi e i bambini di età inferiore a 12 anni mentre la sicurezza e l'efficacia del vaccino nella prima infanzia,

in bambini di età inferiore a 6 mesi, non sono state ancora stabilite. Queste riformulazioni hanno mostrato una superiorità in termini di livello dei titoli neutralizzanti rispetto al vaccino originale ([www.aifa.gov.it](http://www.aifa.gov.it)). Secondo la circolare del Ministero della Salute del 27 settembre 2023 (<https://www.trovanorme.salute.gov.it/norme/renderNormsanPdf?anno=2023&codLeg=96278&parte=1%20&serie=null>), la dose di richiamo con Comirnaty Omicron XBB.1.5, al fine di massimizzare la protezione per la stagione autunno/inverno 2023-2024, si raccomanda a distanza di 6 mesi dall'ultima dose di vaccino anti-COVID-19 ricevuta o dall'ultima infezione, a prescindere dal numero di eventi pregressi (dosi ricevute o diagnosi di infezione). Una singola dose di Comirnaty Omicron XBB 1.5 è inoltre indicata anche per coloro che non sono mai stati vaccinati (ciclo primario). Per i bambini dai 6 mesi ai 4 anni che non hanno completato un ciclo primario di vaccinazione anti-COVID-19 o senza storia di infezione pregressa da SARS-CoV-2, sono invece previste 3 dosi (di cui la seconda a 3 settimane dalla prima e la terza a 8 settimane dalla seconda).

**Spikevax (Moderna):** Il vaccino Spikevax (mRNA-1273) è stato sviluppato dall'azienda farmaceutica Moderna in collaborazione con il *National Institute of Allergy and Infectious Diseases* (NIAID) e autorizzato in Italia da AIFA dal 7 gennaio del 2021. È un vaccino costituito da un mRNA codificante per la proteina spike nella conformazione pre-fusione *full-length* con due sostituzioni amminoacidiche K968P e V987P, e incapsulato in una nanoparticella lipidica (<https://www.epicentro.iss.it/vaccini/>). Il ciclo vaccinale prevede la somministrazione di due dosi a distanza di 28 giorni l'una dall'altra. Inizialmente era rivolto esclusivamente ai maggiori di 18 anni, ma dal 23 luglio 2021 la somministrazione è stata estesa anche alla fascia 12-17 anni. Il vaccino deve essere conservato tra i -25°C/-15°C fino a 7 mesi, ma può essere mantenuto tra i 2°C e gli 8°C per 30 giorni e tra gli 8°C e i 25°C per un massimo di 12 ore. Con l'insorgenza della variante Omicron si sono rese necessarie delle riformulazioni di questo vaccino definite come Bivalent Original/Omicron BA.1, Bivalent Original/Omicron BA.4-5 e Spikevax XBB.1.5. Queste nuove formulazioni sono destinate a soggetti con età pari o superiore ai 6 mesi che hanno precedentemente ricevuto almeno un ciclo vaccinale primario contro COVID-19 ([www.aifa.gov.it](http://www.aifa.gov.it)).

**VACCINI INATTIVATI:** I vaccini inattivati si basano sull'utilizzo di particelle virali o patogeni "inattivati" mediante mezzi chimici o fisici, in grado di stimolare una risposta immunitaria ma non in grado di generare un'infezione. Il primo vaccino inattivato contro SARS-CoV-2 è stato sviluppato dall'Istituto Wuhan in associazione con il *China National Pharmaceutical Group* (Sinopharm). Questo vaccino viene inattivato mediante l'utilizzo di  $\beta$ -propiolattone, che viene propagato in linee cellulari Vero utilizzando l'allume come adiuvante. Un altro vaccino inattivato è CoronaVac,

sviluppato dall'azienda cinese Sinovac Biotech Ltd che ha inattivato il *lineage* B del virus SARS-CoV-2. Questi vaccini sono in uso in diversi Paesi, tra cui Cina, Brasile, Cile, Indonesia e Turchia (<https://www.epicentro.iss.it/vaccini/>). Il vaccino Valneva è un vaccino inattivato e adiuvato che viene somministrato negli adulti con un'età compresa tra i 18 e i 50 anni. È un vaccino approvato per l'Unione Europea che viene somministrato in due iniezioni a distanza di quattro settimane l'una dall'altra ([www.aifa.gov.it](http://www.aifa.gov.it)).

**VACCINI A SUBUNITÀ PROTEICA:** Nei vaccini a subunità proteiche vengono utilizzati veicoli come liposomi, virosomi o nanoparticelle polimeriche in cui sono inserite particelle virali in grado di indurre una risposta immunitaria nell'ospite. Un vaccino generato con questa tecnica è il **Nuvaxovid** (prima chiamato NVX-CoV2373), prodotto dall'Università del Queensland in collaborazione con Novavax, somministrato in soggetti con età pari o superiore ai 12 anni. La vaccinazione prevede un ciclo vaccinale primario di due dosi a distanza di tre settimane l'una dall'altra. ([www.aifa.gov.it](http://www.aifa.gov.it), <https://www.epicentro.iss.it/vaccini/>). Secondo la circolare del Ministero della Salute del 4 dicembre 2023 (<https://www.trovanorme.salute.gov.it/norme/renderNormsanPdf?anno=2023&codLeg=97783&parte=1%20&serie=null>), a fronte dell'avvenuta autorizzazione da parte di EMA ed AIFA della formulazione adattata del vaccino ricombinante adiuvato Nuvaxovid (Omicron XBB.1.5), nella campagna di vaccinazione anti COVID-19 autunnale e invernale 2023-2024 il vaccino Nuvaxovid XBB.1.5 può essere utilizzato in alternativa al vaccino a mRNA Comirnaty XBB.1.5, a partire dai 12 anni di età.

La campagna vaccinale contro la COVID-19 in Italia è stata avviata il 27 dicembre 2020. A fine ottobre 2022, i vaccini utilizzati in Italia nel corso della campagna, autorizzati dall'Agenzia Europea del Farmaco (*European Medicines Agency*, EMA) e dall'Agenzia Italiana del Farmaco (AIFA) erano: Comirnaty (Pfizer-BioNtech), con le formulazioni Original, Original/Omicron BA.1 e Original/Omicron BA.4-5, Spikevax (Moderna) con le formulazioni Original e Original/Omicron BA.1, Vaxzevria (AstraZeneca), COVID-19 Vaccine Janssen (Johnson&Johnson) e Nuvaxovid (Novavax). Fino alla fine di ottobre 2022, i bambini di 0-4 anni erano gli unici a non essere eleggibili per la vaccinazione, mentre nella fascia 5-11 l'unico vaccino autorizzato era Comirnaty con una dose ridotta.

Oltre al ciclo primario con due dosi (una dose per soggetti vaccinati con Janssen), in Italia sono stati approvati due richiami o dosi *booster*: il primo a settembre 2021 e il secondo ad aprile 2022.

In entrambi i casi, il richiamo è stato prima indicato per persone di età  $\geq 80$  anni, per gli ospiti delle residenze sanitarie assistenziali per anziani e per i soggetti di età compresa tra i 60 e gli 79 anni con fattori di rischio per la progressione a COVID-19 severo. Il primo richiamo è stato esteso, a gennaio 2022, a tutti i soggetti a partire dai 12 anni di età. Il secondo richiamo invece è stato esteso l'11 luglio 2022, alle persone con 60 anni o più e a quelli con 12 anni o più che hanno una elevata fragilità motivata da patologie concomitanti/preesistenti. Il 17 ottobre 2022 una ulteriore dose di richiamo con vaccini a mRNA è stata raccomandata a favore delle persone di età  $\geq 80$  anni, degli ospiti delle strutture residenziali per anziani e delle persone di età  $\geq 60$  anni con fragilità motivata da patologie concomitanti/preesistenti. In ogni caso la raccomandazione si rivolge a persone che abbiano superato una infezione da SARS-CoV-2 da oltre 4 mesi.

Al 17 ottobre 2022, l'85.8% della popolazione italiana aveva ricevuto almeno una dose del vaccino (il 77.6% aveva ricevuto l'ultima dose da oltre 180 giorni e solo l'8.3% da meno di 180 giorni), mentre il 14.1% non aveva ricevuto alcuna dose. La copertura vaccinale è inversamente correlata con le fasce di età; tra i vaccinati anche la distanza dall'ultima dose differisce tra le fasce di età. Mentre meno del 3% delle persone nelle fasce di età 12-39 e 40-59 aveva ricevuto l'ultima dose da meno di 180 giorni, questa percentuale aumenta fino al 33.5% nei soggetti con 80 anni o più. Nei bambini di 5-11 anni circa il 65% risulta non vaccinato, l'1% vaccinato da meno di 180 giorni e circa il 34% vaccinato da oltre 180 giorni ([https://www.iss.it/documents/20126/6703853/NT\\_RischioCOVID19+LAST+%281%29.pdf/92°267f8-c8a7-9cf4-dc36-afb7bac5a55c?t=1672323176599](https://www.iss.it/documents/20126/6703853/NT_RischioCOVID19+LAST+%281%29.pdf/92°267f8-c8a7-9cf4-dc36-afb7bac5a55c?t=1672323176599)).

A seguito della nota dell'Agenzia europea per i medicinali (EMA) e del Centro europeo per la prevenzione e il controllo delle malattie (ECDC), pubblicata l'11 luglio 2022, anche in Italia è arrivato il via libera alla seconda dose di richiamo (*second booster*) per tutti gli *over 60* e per tutte le persone, a partire dai 12 anni, con elevata fragilità motivata dalle patologie indicate nella precedente fase della campagna vaccinale, purché sia trascorso un intervallo minimo di almeno 120 giorni dalla prima dose di richiamo o dall'ultima infezione successiva al richiamo (data del test diagnostico positivo).

Complessivamente, al 19 ottobre 2023, il 90.2% della popolazione *over 12* ha completato il ciclo vaccinale primario, mentre l'84.9% ha completato il ciclo vaccinale con dose addizionale o *booster* da almeno 4 mesi. La terza dose o seconda dose *booster* è stata somministrata a una proporzione del 16.9% della popolazione mentre l'8.2% ha ricevuto la terza *booster* (<https://www.governo.it/it/cscovid19/report-vaccini/>).



## 2. PRIMI DATI IN ITALIA

Essendo stata l'Italia uno tra i primi punti di ingresso del virus in Europa e quindi uno dei principali responsabili della sua successiva dispersione nel mondo, l'analisi dell'epidemiologia molecolare di SARS-CoV-2 nel nostro Paese, durante le fasi iniziali della pandemia, è stata di particolare interesse per studiare i primi passi evolutivi del virus fuori dalla Cina e come si sia diffuso nei Paesi occidentali.

Fin dai primi mesi della pandemia da COVID-19 in Italia, un gran numero di centri clinici e laboratori in grado di effettuare la caratterizzazione genomica di SARS-CoV-2, distribuiti in tutte le regioni d'Italia, ha formato la rete collaborativa SCIRE (*SARS-CoV-2-Research Enterprise*) con l'obiettivo di contribuire allo sforzo comune di ottenere il maggior numero possibile di genomi virali completi.

Il lavoro svolto durante il primo anno di dottorato sull'epidemiologia genomica e l'evoluzione di SARS-CoV-2 in un periodo compreso tra febbraio e giugno 2020,<sup>37</sup> ha confermato come già all'inizio della pandemia fossero presenti nel nostro Paese diversi *lineage* e *clade*, con la prevalenza dei *lineages* B, B.1 e B.1.1, corrispondenti ai *clades* 19A, 20A e 20B, caratterizzati da una diversa distribuzione a livello geografico e temporale. Un'elevata prevalenza del *lineage* B.1/20A è stata osservata in molte regioni, soprattutto nel Nord Italia mentre le regioni del Centro-Sud, isole comprese, erano caratterizzate da una maggiore prevalenza dei ceppi B.1.1/20B. Un'alta proporzione dei *lineages* B/19A e B.1.1.1/20D, è stata riscontrata rispettivamente in Veneto e in Piemonte (**Figura 13**).

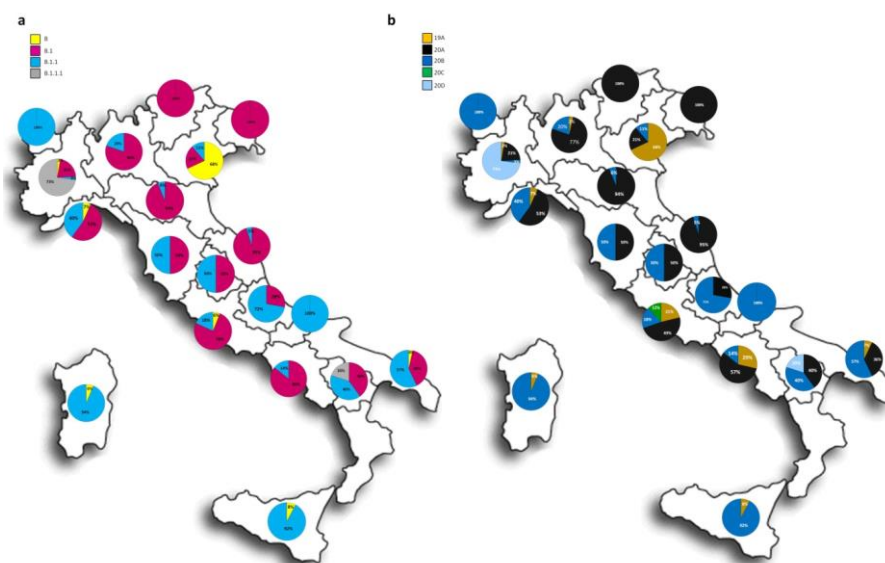
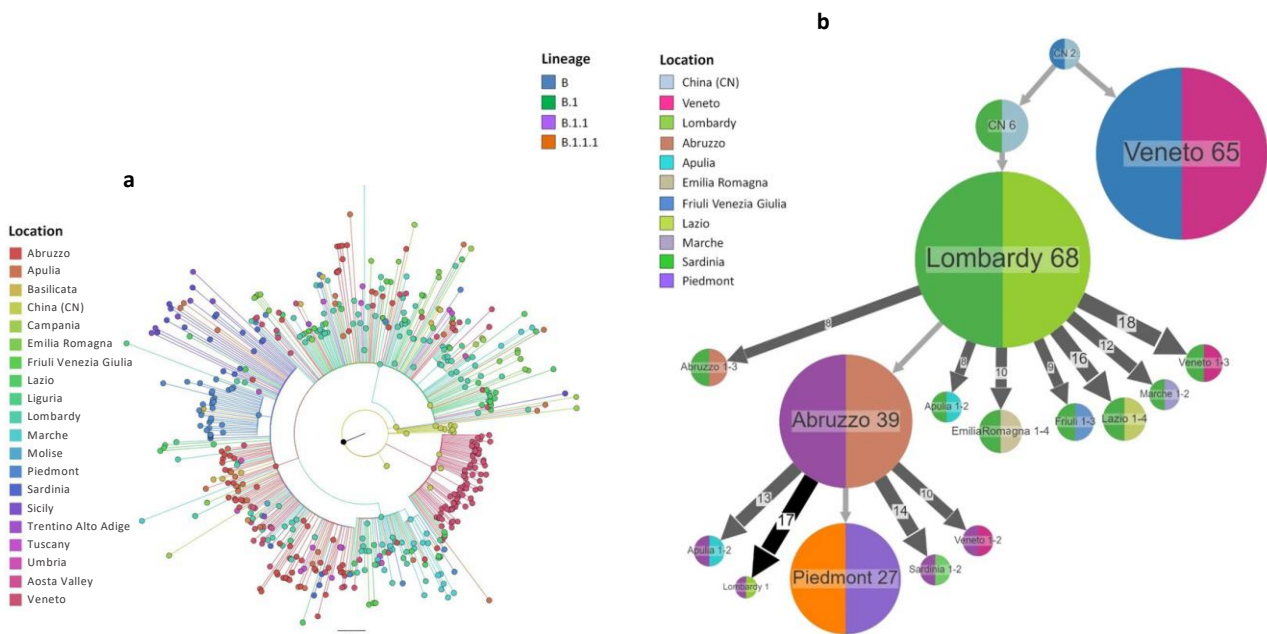


Figura 13. Distribuzione dei differenti *lineages* (a) e *clades* (b) nel periodo compreso tra febbraio e giugno 2020.<sup>37</sup>

Mentre i ceppi B/19A e B.1/20A erano ampiamente prevalenti all’inizio della pandemia, i *lineages* B.1.1/20B e B.1.1.1/20D venivano riscontrati soprattutto nelle fasi successive. La stima del tMRCA dei principali *lineages*, mediante metodi Bayesiani o di *Maximum Likelihood*, ha suggerito un ingresso in Italia precoce dei ceppi B e B.1 a partire dalla seconda metà di gennaio 2020, seguita dalla comparsa dei *lineages* B.1.1 e B.1.1.1, rispettivamente tra la fine di febbraio 2020 e l’inizio di marzo 2020.

L’analisi filogeografica italiana ha mostrato due distinti *patterns* di dispersione di SARS-CoV-2 (**Figura 14**). Nel primo *pathway*, è stato osservato come il *lineage* B, dopo l’introduzione in Veneto, abbia dato origine ad un *cluster* rimasto apparentemente confinato in questa regione, scomparendo entro la prima metà di marzo, in accordo con i dati epidemiologici che mostrano l’effettiva “interruzione” di questo primo focolaio di infezione grazie ai test effettuati, agli efficaci approcci di tracciamento e ad un *lockdown* locale di 14 giorni.<sup>38</sup> Il secondo *pathway* coinvolgeva il *lineage* B.1, inizialmente entrato in Lombardia per poi diffondersi in altre regioni italiane, principalmente dell’Italia centrale (Marche e Abruzzo) e settentrionale (Emilia-Romagna e Veneto). L’Italia centrale, e in modo particolare l’Abruzzo, sembra rappresentare un altro importante centro di dispersione del *lineage* B.1.1, che si sarebbe evoluto dal *lineage* B.1 a partire da metà febbraio. L’introduzione del *lineage* B.1.1.1 sarebbe avvenuta in Piemonte nella seconda metà di marzo 2020 e non sarebbe stata seguita da un’ulteriore diffusione in altre regioni.



**Figura 14.** Ricostruzione del *lineage* B.1 circolante nell’arco temporale febbraio-giugno 2020. La figura mostra l’albero completo (a) e la visualizzazione compressa prodotta da PastML (b). Colori diversi corrispondono a diverse regioni geografiche e *lineages* italiani. I numeri all’interno (o accanto) i cerchi indicano il numero di ceppi assegnati al nodo specifico.<sup>37</sup>

L'analisi dei *cluster* eseguita sul *dataset* internazionale ha indicato la presenza di genomi italiani posizionati vicino alla radice dell'albero, a conferma del ruolo centrale svolto dal nostro Paese nella diffusione dell'epidemia nei primi mesi.

Queste osservazioni sono state ulteriormente confermate dalla ricostruzione filogeografica internazionale che ha nuovamente mostrato due pattern filogeografici ben definiti. Il primo, relativo al *lineage* B, ha mostrato diverse introduzioni indipendenti dalla Cina all'Italia e ad altri Paesi europei, mentre il secondo, relativo al *lineage* B.1 ha evidenziato come, partendo dalla radice dell'albero (Cina), siano emersi indipendentemente due *cluster*, uno italiano e uno europeo. Il *lineage* B.1.1 è apparentemente emerso dal *cluster* B.1 italiano, suggerendo un'evoluzione locale, mentre il *lineage* B.1.1.1 molto probabilmente è emerso dal B.1.1 presente in altri Paesi europei ed è stato introdotto in Piemonte immediatamente a ridosso del *lockdown* di marzo 2020. I dati di questo lavoro<sup>37</sup> hanno suggerito un ruolo centrale dell'Italia nell'esportazione di alcuni *lineages* virali all'inizio dell'epidemia in Europa, mentre successivamente, dopo la metà di marzo, il nostro Paese ha rappresentato un centro di importazione di numerosi *lineages* e *sub-lineages* virali provenienti da diversi Paesi europei.

L'introduzione in Italia della variante D614G, caratterizzata da una maggiore trasmissibilità, e la mancata identificazione della sua circolazione per settimane prima della rilevazione dei primi casi in Italia potrebbe spiegare la rapida diffusione dell'epidemia nel Nord Italia, seguita dalla diffusione ad altre regioni italiane e successivamente al resto d'Europa.

### 3. SCOPO

La continua emergenza di nuove varianti di SARS-CoV-2 ad elevato rischio per la Sanità Pubblica globale ha reso prioritario il loro monitoraggio a livello globale e lo studio delle loro caratteristiche. In questo contesto, il sequenziamento del genoma completo si è dimostrato, fin dall'inizio della pandemia, uno strumento fondamentale per monitorare l'evoluzione del virus e valutarne le implicazioni in termini di diagnosi e di efficacia vaccinale e terapeutica.

La caratterizzazione di genomi interi e la loro condivisione su database globali, come la piattaforma GISAID, che registra a settembre 2023 oltre 16 milioni di sequenze, hanno rappresentato e rappresentano ancora a oggi degli strumenti essenziali per comprendere l'origine e le aree di dispersione del virus e delle sue varianti nella popolazione, attraverso quella che viene definita "sorveglianza genomica". L'analisi filodinamica, basata sulla combinazione delle caratteristiche evolutive, epidemiologiche e immunologiche che influenzano la filogenesi virale, rappresenta ad oggi un importante strumento per la sorveglianza molecolare di un'infezione, in particolare delle infezioni causate da virus emergenti.<sup>39, 40</sup>

In questo contesto lo scopo del progetto di questa tesi di dottorato è stato lo studio dell'epidemiologia genomica dei principali *lineages*/varianti di SARS-CoV-2 circolanti in Italia tra l'estate del 2020 e l'inverno del 2022, attraverso:

1. Campionamento multiregionale esteso di SARS-CoV-2 attraverso collaborazioni esistenti con numerosi centri clinici e laboratori inclusi nel gruppo collaborativo SCIRE (*SARS-CoV-2 Italian Research Enterprise*) e concomitante sequenziamento e caratterizzazione molecolare tramite *Next Generation Sequencing* di genomi interi di SARS-CoV-2 collezionati in Italia tra agosto 2020 e dicembre 2022;
2. Studio della diversità genomica delle principali varianti di SARS-CoV-2 circolanti nel nostro Paese nello stesso periodo quali *lineages* B.1.177 (20E.EU1), B.1.1.7 (variante Alpha) e B.1.167.2 (variante Delta), e le varianti Omicron BA.1, BA.2 e BA.5;
3. Analisi filogenetiche, individuazione dei *cluster* di genomi e caratterizzazione degli stessi in base spazio-temporale nel contesto nazionale ed internazionale;
4. Studio dell'origine temporale, del tasso di evoluzione virale e della dinamica di popolazione delle stesse varianti di SARS-CoV-2.

## 4. MATERIALI E METODI

### 4.1 POPOLAZIONE IN STUDIO

In questo lavoro, che ha visto la collaborazione di oltre 74 centri clinici e laboratori dislocati su tutto il territorio italiano, sono stati caratterizzati 4,894 nuovi genomi completi, collezionati tra agosto 2020 e dicembre 2022, di pazienti positivi a SARS-CoV-2, ospedalizzati o gestiti in regime ambulatoriale, distribuiti sull'intero territorio nazionale. Di questi genomi, 473 sono stati ottenuti presso il Laboratorio di Malattie Infettive dell'Università degli Studi di Milano. Tutti i partecipanti hanno fornito il consenso informato per l'inserimento dei loro dati anonimizzati in un database protetto. Tutti i dati utilizzati in questo studio sono stati precedentemente resi anonimi come richiesto dal Codice Privacy (D.Lgs 196/2003) e dalle autorizzazioni generali rilasciate dal Garante. Lo studio è stato approvato dal Comitato Etico dell'Ospedale Sacco (protocollo n. 47866 del 9 settembre 2020) e condotto in conformità alle linee guida della Buona Pratica Clinica e ai principi della Dichiarazione di Helsinki del 1964.

### 4.2 SEQUENZIAMENTO DEI GENOMI VIRALI

L'estrazione dell'RNA virale è stata condotta a partire da tamponi naso-faringei, lavaggi broncoalveolari (BAL) o surnatanti di coltura di pazienti infetti, mediante il kit QiAmp Viral RNA mini Kit (QIAGEN, Germany). In seguito alla retrotrascrizione mediante il *kit* LunaScript® RT SuperMix (New England BioLabs, United States), l'intero genoma virale è stato amplificato utilizzando una modificazione del protocollo sviluppato da *ARTIC network* che utilizza come *primer* due set di *pool ARTIC nCoV-2019 V3* e *V4 panel* (IDT, United States), ciascuno composto rispettivamente da 98 e 104 *primer*.

A seguito di purificazione dei prodotti di PCR, il sequenziamento dei genomi interi è stato effettuato attraverso l'utilizzo della piattaforma di *Next Generation Sequencing* (NGS) MiSeq System Illumina.

Le sequenze consenso, ottenute dall'assemblaggio delle *reads*, contenute nei file FastQ, rispetto al genoma di riferimento di SARS-CoV-2 (<http://gisaid.org/>, accession ID: EPI\_ISL\_406800), sono state generate utilizzando il *software* Geneious Prime v. 9.1.8 (Biomatters, <http://www.geneious.com>).

La genotipizzazione delle sequenze ottenute è stata effettuata utilizzando Pangolin COVID-19 Lineage Assigner *tool* v. 4.1.1 (<https://pangolin.cog-uk.io/>) e Nextstrain v. 2.5.0

(<https://clades.nextstrain.org/>). Le mutazioni relative alle sequenze italiane sono state identificate utilizzando Nextstrain.

### 4.3 CRITERI DI ALLESTIMENTO DEI DATASET

Per lo studio delle principali varianti circolanti tra l'estate 2020 e l'inverno del 2021 (*clade* 20E.EU1 e varianti Alpha e Delta), sono stati allestiti 3 *datasets* internazionali utilizzando in totale 609 genomi interi caratterizzati dal nostro *network* collaborativo (20E.EU1, n= 269; Alpha, n=164; Delta, n= 176).

Per studiare i principali lineaggi della variante Omicron circolanti nel corso del 2022 (BA.1, BA.2 e BA.5), sono stati allestiti 3 *datasets* nazionali utilizzando 1,658 genomi interi caratterizzati dal *network* collaborativo (BA.1, n= 268; BA.2, n = 677; BA.5, n=713).

A questi 6 *datasets*, sono stati aggiunti genomi italiani ottenuti da GISAID (<https://gisaid.org>) selezionati secondo i seguenti criteri: 10 genomi interi per ciascuna regione italiana e per mese di campionamento durante il periodo di circolazione di ciascuna variante, con un massimo di due sequenze per regione/settimana, escludendo genomi identici e con più del 5% di *gap*. La composizione dei *dataset* e la distribuzione regionale delle sequenze italiane per le varianti in studio sono riassunte nelle **Tabelle 2 e 3**.

**Tabella 2. Composizione dei *dataset* italiani.**

	Variante	Sequenze SCIRE	Sequenze GISAID	Totale
	<b>20E.EU1</b>	269	802	<b>1,071</b>
	<b>Alpha</b>	164	1,159	<b>1,323</b>
	<b>Delta</b>	196	1,319	<b>1,495</b>
Omicron	<b>BA.1</b>	268	612	<b>880</b>
	<b>BA.2</b>	677	950	<b>1,627</b>
	<b>BA.5</b>	713	1,048	<b>1,761</b>

**Tabella 3. Distribuzione in termini di regione delle sequenze italiane incluse nei *datasets*.**

Regioni	20E.EU1 (n=1,071)	Alpha (n=1,323)	Delta (n=1,495)	BA.1 (n=880)	BA.2 (n=1,627)	BA.5 (n=1,761)
Abruzzo	56	80	81	39	46	72
Basilicata	38	61	67	8	22	14
Calabria	12	62	13	13	25	17
Campania	99	141	114	48	73	73
Emilia-Romagna	54	74	82	50	73	77
Friuli-Venezia Giulia	57	60	81	40	64	77
Lazio	90	86	87	105	148	106
Liguria	119	39	148	119	253	184
Lombardia	167	132	110	164	355	426
Marche	7	13	7	10	98	158
Molise	12	41	12	/	/	/
Piemonte	11	55	81	40	63	59
Puglia	67	105	117	37	65	71
Sardegna	64	55	50	/	5	9
Sicilia	43	65	80	50	74	77
Toscana	12	22	58	13	50	60
Trentino Alto Adige	45	62	90	40	59	96
Umbria	41	72	88	49	78	101
Valle d'Aosta	2	23	39	15	7	9
Veneto	75	75	90	40	69	75

Per i genomi internazionali, sono stati selezionati 5 genomi per Paese Europeo e non Europeo e per mese di campionamento, con località nota e data esatta di campionamento, escludendo genomi identici e con più del 5% di gap (**Tabella 4**).

**Tabella 4. Composizione dei *datasets* internazionali.**

	20E.EU1 (n=2,425)	Alpha (n=3,174)	Delta (n=3,248)
<b>IT</b>	1,071	1,323	1,495
EU	1,144	1,470	1,451
<b>no IT</b>			
no EU	210	381	302

IT: sequenze italiane; EU: sequenze europee; noEU: sequenze internazionali.

I genomi inclusi nel *dataset* di *clade* 20E.EU1, avevano una datazione compresa tra agosto 2020 e ottobre 2021, gli isolati della variante Alpha tra novembre 2020 e agosto 2021, mentre i genomi inclusi nel *dataset* della variante Delta datavano tra maggio 2021 e gennaio 2022. Gli isolati inclusi nel *dataset* della variante Omicron BA.1 avevano una datazione compresa tra novembre 2021 e

aprile 2022, gli isolati di variante BA.2 tra gennaio 2021 e dicembre 2022, mentre i genomi inclusi nel *dataset* della variante BA.5 datavano tra aprile 2021 e dicembre 2022.

Per l'allineamento dei genomi con la sequenza riferimento è stato utilizzato il *software* MAFFT (<https://mafft.cbrc.jp/alignment/server/>) che permette l'allineamento automatico di sequenze multiple e il programma BioEdit v. 7.2.6.1 (<http://www.mbio.ncsu.edu/bioedit/bioedit.html>), che consente sia l'allineamento automatico mediante l'algoritmo ClustalW sia un allineamento manuale. Infine, le sequenze sono state tagliate alle estremità per ottenere genomi della stessa lunghezza (29,774 bp).

Per valutare il segnale temporale è stata effettuata per tutti i *datasets* una analisi della regressione *root-to-tip* attraverso il *software* TempEst.<sup>41</sup> Questo programma può analizzare sia alberi contemporanei (dove tutte le sequenze sono state campionate contemporaneamente) sia alberi datati (dove le sequenze sono state campionate in date diverse).

#### 4.4 ANALISI FILOGENETICA

L'analisi filogenetica rappresenta l'unico metodo affidabile per la tipizzazione delle sequenze e permette di esaminare le relazioni genetiche all'interno di un gruppo di sequenze, anche provenienti da aree geografiche diverse, fornendo in questo modo informazioni riguardanti l'origine dell'epidemia e la propagazione del virus in una popolazione.

I modelli di sostituzione nucleotidica permettono il calcolo delle probabilità delle mutazioni nucleotidiche tra i rami di un albero filogenetico. L'uso di un determinato modello di sostituzione può far variare il risultato dell'analisi filogenetica, per questo la scelta del modello di sostituzione nucleotidica è un passaggio essenziale che precede l'analisi filogenetica vera e propria.

La scelta del modello è stata eseguita tramite il *tool* ModelFinder implementato nel programma IQ-TREE v.1.6.12 (<http://www.iqtree.org/>).<sup>42</sup> In questo studio è stato selezionato il modello GTR+F+R4 (*General Time Reversible + empirical base Frequencies + Four number of categories*) per il *clade* 20E.EU1 e per le varianti Alpha e Delta, il modello GTR+F+R3 (*General Time Reversible + empirical base Frequencies + Three number of categories*) per le varianti Omicron BA.1 e BA.2 e il modello GTR+F+R6 (*General Time Reversible + empirical base Frequencies + Six number of categories*) per il *dataset* di Omicron BA.5.

La filogenesi iniziale è stata ricostruita per tutte le varianti in studio utilizzando un approccio di massima verosimiglianza (*Maximun Likelihood-ML*) mediante il programma IQ-TREE, recentemente sviluppato, che rende possibile la conduzione di analisi filogenetiche su vasta scala



in tempi rapidi. L'approccio di massima verosimiglianza ricerca la miglior *likelihood* dell'albero, ovvero la probabilità di ottenere dei dati in base all'albero generato. L'affidabilità dei *clades* osservati è stata stabilita sulla base di valori di *bootstrap* dei nodi interni  $\geq 60\%$  (dopo 1,000 repliche).

L'identificazione dei *cluster* italiani nell'albero ML è stata effettuata mediante il *software* Cluster Picker v.1.2.3 che permette di selezionare, all'interno di un albero filogenetico, gruppi di sequenze potenzialmente correlate tra loro a livello epidemiologico. Sono stati considerati i *cluster* con più di due sequenze, valori di *bootstrap* dei nodi interni maggiori di 60% e una distanza genetica media tra le sequenze inferiore all'1% come *thresholds* per le sequenze dei *datasets* 20E.EU1, Alpha e Delta, e inferiori allo 0.1% per i *datasets* delle varianti Omicron.

Le caratteristiche epidemiologiche dei *cluster* internazionali identificati sono state ulteriormente investigate utilizzando il *software* Cluster Matcher v.1.2.23<sup>43</sup> che consente l'identificazione dei *cluster* che soddisfano determinati criteri. Sono stati selezionati solo i *cluster* contenenti almeno una sequenza italiana e classificati come misti (M) quelli che includevano più di un isolato italiano e sequenze non italiane, come puri italiani (IT) quelli che includevano solamente isolati italiani e come *singleton* (S), i *cluster* contenenti un solo isolato italiano insieme a sequenze non italiane.

#### 4.5 ANALISI FILODINAMICA E STIMA DEL TASSO RIPRODUTTIVO DI BASE

L'analisi filogenetica può essere anche condotta sulla base di un modello temporale definito orologio molecolare (*molecular clock*). Tale modello si basa sul concetto che le mutazioni si accumulano nel tempo in modo pressoché costante permettendo di delineare le relazioni intercorrenti tra le distanze genetiche e il tempo, e permettendo anche di datare gli eventi di divergenza evolutiva, attraverso la stima dell'antenato comune più recente (tMRCA, *time of the Most Recent Common Ancestor*). I modelli più semplici prevedono un orologio molecolare più rigido (*strict clock model*), nel quale si assume che tutti i *lineages* evolvano alla stessa velocità; nei modelli più complessi (*relaxed clock model*) invece le velocità evolutive variano nel tempo e tra i vari *lineages*, fornendo degli intervalli di variabilità intorno ad un valore medio di velocità evolutiva.

Per caratterizzare la storia epidemiologica ed evolutiva delle varianti di SARS-CoV-2 circolanti in Italia nel corso del 2020 e 2021, sono stati selezionati solo i *cluster* comprendenti almeno il 70% di sequenze italiane, considerando i *cluster* maggiori di 10 isolati per *dataset* di 20E.EU1, Alpha e

Delta. *Cluster* italiani caratterizzati da 10 o più isolati sono stati selezionati per l'analisi delle varianti Omicron.

Sono stati utilizzati il modello demografico parametrico *birth-death*, e il modello non parametrico *Bayesian Skyline Plot* (BSP) che consente, suddividendo il periodo di tempo indicato tra la radice dell'albero e il campionamento più recente in 5 intervalli in cui le dimensioni della popolazione si assumono costanti, la ricostruzione del tasso di crescita degli eventi di trasmissione.

Il programma BEAST v.2.7.5,<sup>44</sup> basato sulle teorie della statistica *Bayesiana* che calcola le probabilità a posteriori dell'albero, ovvero la probabilità di ottenere un determinato albero in base ai dati forniti, è stato utilizzato per la costruzione degli alberi datati. Le probabilità a posteriori stimate vengono approssimate mediante l'algoritmo *Markov Chain Monte Carlo* (MCMC); inizialmente viene generato uno *starting tree* casuale che viene poi confrontato con un nuovo albero ottenuto modificando in modo sistematico il precedente e i relativi parametri; il secondo albero viene quindi accettato o scartato sulla base della probabilità a posteriori in uno specifico processo di campionamento. Questo sistema di confronto viene ripetuto a catena fino ad ottenere l'albero con il più alto prodotto di probabilità a posteriore nodale (*posterior probability*, pp) tra tutti quelli generati (*consensus tree* o *maximum clade credibility tree*). In questo modo si fornisce supporto statistico ai valori, attraverso una stima della significatività dei *cluster* osservati nell'albero.

Al fine di delineare la storia epidemiologica ed evolutiva dei diversi *lineages* di SARS-CoV-2 studiati, sono stati utilizzati due modelli, quello della coalescenza e il *birth-death*. Il primo si basa sulla correlazione fra il tempo in cui due *lineages* (o *taxa*) hanno condiviso il medesimo antenato e le dimensioni della popolazione analizzata, permettendo la stima dei cambiamenti della dimensione effettiva della popolazione nel tempo ( $N_e$ ). Per tracciare l'evoluzione dei *lineages*, da cui estrapolare gli eventi di coalescenza, è necessario disegnare uno *skyline plot*, che mostra le dimensioni della popolazione virale con il numero delle infezioni effettive (N) sull'asse delle ordinate e il suo andamento su scala temporale (t) sull'asse delle ascisse. Questo approccio, alla base dei metodi *Bayesiani* e del *software* BEAST, consente di stimare nello stesso momento la topologia dell'albero filogenetico, la dinamica della popolazione effettiva, la velocità evolutiva e il tempo dei più recenti progenitori comuni (tMRCA) corrispondenti ai nodi interni dell'albero, a partire da un allineamento multiplo. L'analisi condotta mediante BEAST si basa su cinque parametri fondamentali:

- *Substitution model*: definisce le differenti velocità evolutive con cui avvengono le sostituzioni lungo i rami dell'albero;
- *Rate model among sites*: definisce la distribuzione delle diverse velocità all'interno dei siti;
- *Rate model among branches*: definisce la distribuzione delle differenti velocità tra i singoli rami, convertendo l'albero, che è in scala temporale, in unità di sostituzioni/sito/anno. Questo modello è usato per valutare il tempo di divergenza della popolazione;
- *Tree*: definisce il modello delle relazioni filogenetiche fra le sequenze;
- *Tree prior*: fornisce i parametri del modello evolutivo e la topologia dell'albero. Sono disponibili diversi modelli parametrici: *constant sites* (1 parametro), *exponential growth* (2 parametri) e *logistic growth* (3 parametri).

L'analisi *Bayesiana* è stata eseguita utilizzando lo stesso modello di sostituzione e orologio molecolare impiegato per le analisi precedentemente descritte. Gli *evolutionary rate* sono stati stimati utilizzando una distribuzione a priori *Log Normal* (media,  $M=8E^{-4}$ ; varianza,  $S=1.25$ ) nello spazio reale utilizzando uno *strict clock model*.

Le analisi MCMC sono state eseguite per 60 milioni di generazioni, con campionamento ogni 3,000. La convergenza è stata valutata attraverso la stima dell'*Effective Sampling Size* (ESS) dopo aver applicato un *burn-in* del 10% attraverso il *software* Tracer v.1.7 (<http://tree.bio.ed.ac.uk/software/tracer/>),<sup>45</sup> accettando ESS pari o superiori a 200. L'incertezza sulle stime ottenute è stata indicata con intervalli di *highest prior density* (HPD) del 95%.

L'albero finale è stato selezionato sulla base del valore massimo di *posterior probability* (pp) dopo aver effettuato un *burn-in* del 10% attraverso il *software* Tree Annotator v.10.4 (incluso nel pacchetto BEAST). Sono state considerate significative le *posterior probability* superiore allo 0.7. Infine, tutti gli alberi sono stati visualizzati e modificati mediante il *software* Figtree v. 1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>).

Lo studio della filodinamica di un'infezione permette di stimare parametri epidemiologici importanti fra cui il tasso di riproduzione di base ( $R_0$ ), che indica il numero di casi secondari attesi da un singolo soggetto infettato al tempo 0, ipotizzando che tutta la popolazione sia suscettibile. L' $R_0$  è stato quindi calcolato tenendo conto del tasso di crescita esponenziale ( $r$ ) utilizzando

l'equazione  $R_0=r*D+1$ , dove D rappresenta la durata media dell'infettività, stimata sulla base di dati disponibili in letteratura.<sup>46, 47</sup>

In una popolazione reale, tuttavia, non essendo costituita da soli soggetti suscettibili a una infezione (vaccinati, infezioni pregresse) è più corretto parlare di tasso di riproduzione effettivo ( $R_e$ ).

Il modello *birth-death skyline*, implementato in Beast 2.7.5, consente di individuare direttamente i cambiamenti in  $R_e$  e altri importanti parametri epidemiologici. La costruzione di un albero *birth-death* prevede infatti la stima dei *priors* di importanti parametri epidemiologici, quali il tasso di riproduzione effettivo ( $R_e$ ), il tasso di rimozione ( $\delta$ ), il tasso di trasmissione ( $\lambda$ ), la probabilità di campionamento ( $\Psi$ )<sup>48</sup> e l'origine dell'epidemia. Dal momento che i campioni sono stati collezionati in un breve periodo temporale, è stato utilizzato un modello "*birth-death skyline serial*".

La stima di tali parametri è stata effettuata utilizzando *priors* caratterizzati da diverse distribuzioni di probabilità, in particolare:

- per l' $R_e$  un *prior* con distribuzione di probabilità *Log Normal*, con una media (M) di 0.0 e una varianza (S) di 1.5 (per il *clade* 20E.EU1, variante Alpha e variante Delta) e 1.8 (per Omicron BA.1, BA.2 e BA.5), che permette di ricercare il valore ottimale di  $R_e$  tra valori minori di 1 e maggiori di 7;
- per il tasso di rimozione (*becoming uninfected*) un *prior normal*, con M=48.8 e S=15 (IC95%: 24.0-73.4) corrispondente ad un periodo di infettività tra i 5.3 e i 19 giorni (media 7.5) in accordo con gli intervalli seriali stimati da *Li et al.*;<sup>49</sup>
- la probabilità di campionamento ( $\rho$ ) è stata valutata assumendo per il *clade* 20E.EU1 e per le varianti Alpha e Delta un medesimo *prior*  $\beta$  ( $\alpha=1.0$ ,  $\beta=3,547$ ) e un differente *prior*  $\beta$  ( $\alpha=1.0$ ,  $\beta=1,500$ ) per le varianti Omicron. I valori  $\alpha$  e  $\beta$  vengono fissati sulla base del numero di genomi disponibili nell'analisi (normalizzando a 1), e il numero di casi attivi di COVID-19 al picco del periodo considerato;
- per l'origine dell'epidemia è stato utilizzato un *prior normal* (M=0.1 e S=0.2) per il *clade* 20E.EU1 e *lognormal* (M=0.1 e S=0.3), per le varianti Alpha, Delta Omicron BA.1, BA.2 e BA.5;
- il tasso di crescita medio è stato calcolato sulla base dei *birth* e *recovery rate* ( $r=\lambda-\delta$ ) e il *doubling time* è stato stimato mediante l'equazione  $doubling\ time=\ln(2)/r$ .

## 5. RISULTATI (1)

### EPIDEMIOLOGIA GENOMICA DELLE PRINCIPALI VARIANTI DI SARS-COV-2 CIRCOLANTI IN ITALIA NELL'ESTATE 2020- INVERNO 2021

#### 5.1 ANALISI DELLE MUTAZIONI NEI GENOMI DI SARS-COV-2 ITALIANI

##### 5.1.1 CLADE 20E.EU1

La comparazione tra i genomi italiani e la sequenza di riferimento ha mostrato la presenza di 10 sostituzioni aminoacidiche in più del 10% degli isolati italiani. La quasi totalità delle sequenze presentava le mutazioni distintive di questo *lineage* quali la A220V (97.4%; n=1,043) nel gene N, A222V (97.8%; n=1,048) nella proteina spike e le sostituzioni V30L e L67F (99.3% ciascuna; n=1,064), presenti rispettivamente nelle regioni ORF10 e ORF14. Le mutazioni D614G nel gene S e P314L nell'ORF1b, caratteristiche del *lineage* B.1, erano presenti in oltre il 90% degli isolati (99.1% and 98.3%, respectively). Mutazioni aggiuntive erano presenti in una proporzione minore, variabile dal 13.9% al 34.6% di genomi: la A262S, P272L, Q675H nel gene S e A3623S nell'ORF1a (**Tabella 5**). Due sequenze mostravano nel gene S la mutazione S98F, identificativa della variante 20A/S:98F, e una sequenza presentava la mutazione D80Y propria della variante 20C/S:80Y.

**Tabella 5. Mutazioni del dataset 20E.EU1. Il simbolo (\*) indica le mutazioni caratteristiche della variante.**

Gene	Mutazione	Totale n=1,071 (%)
ORF 1a	A3623S	214 (20%)
ORF1b	P314L	1,053 (98.3%)
S	A222V*	1,048 (97.8%)
	A262S	352 (32.9%)
	P272L	371 (34.6%)
	D614G	1,061 (99.1%)
	Q675H	149 (13.9)
N	A220V*	1,043 (97.4)
ORF10	V30L*	1,064 (99.3)
ORF14	L67F*	1,064 (99.3)

### 5.1.2 VARIANTE ALPHA

Negli isolati di variante Alpha, sono state riscontrate solo mutazioni e delezioni tipiche di questo *lineage*, come evidenziato nella **Tabella 6**, con l'unica eccezione di due mutazioni consecutive nel gene N, R203K (99.1%, n=1,312) e G204R (96.8%, n=1,281), proprie del *lineage* B.1.1 e dei suoi discendenti.

La mutazione S:E484K, tipica della variante Beta e Gamma, veniva osservata in meno dell'1% delle sequenze analizzate. Lo 0.15% (n=2) delle sequenze presentava la mutazione S:K417N presente nella variante Beta.

**Tabella 6. Mutazioni identificate nella variante Alpha. Il simbolo (\*) indica le mutazioni distintive della variante.**

Gene	Mutazione	Totale n=1,323 (%)
ORF 1a	T1001I*	1,305 (98.6)
	A1708D*	1,305 (98.6)
	I2230T*	1,314 (99.3)
	SGF 3675-3677del*	1,136 (85.9)
S	N501Y*	1,270 (95.9)
	A570D*	1,309 (98.9)
	D614G	1,321 (99.8)
	P681H*	1,280 (96.7)
	T716I*	1,272 (96.1)
	S982A*	1,323 (94.6)
	D1118H*	1,306 (98.7)
	HV 69-70del*	1,233 (93.2)
	Y144del*	1,248 (94.3)
	ORF8	Q27 <sup>#</sup>
R52I*		1,197 (90.5)
Y73C*		1,316 (99.5)
N	D3L*	1,316 (99.4)
	R203K	1,312 (99.1)
	G204R	1,281 (96.8)
	S235F*	1,280 (96.7)

### 5.1.3 VARIANTE DELTA

Nelle sequenze italiane di variante Delta sono state identificate 34 mutazioni e due delezioni in più del 10% delle sequenze analizzate (**Tabella 7**).

**Tabella 7. Mutazioni identificate nella variante Delta. Il simbolo (\*) indica le mutazioni caratteristiche della variante.**

Gene	Mutazione	Totale n=1,495 (%)					
ORF1a	K621N	204 (13.6)	S	T19R*	1,441 (96.4)		
	A1306S	1,293 (86.5)		T95I	635 (42.5)		
	P1640L	195 (13)		G142D*	948 (63.4)		
	P2046L	1,292 (86.4)		R158G*	1,433 (95.9)		
	P2287S	1,287 (86.1)		A222V	191 (12.8)		
	A2529V	167 (11.2)		L452R*	1,458 (97.5)		
	V2930L	1,197 (80.1)		T478K*	1,431 (95.7)		
	A3209V	188 (12.6)		D614G*	1,384 (96.2)		
	T3255I	1,287 (86.1)		P681R*	138 (9.2)		
	T3646A	1,293 (86.5)		D950N*	1,303 (87.2)		
	V3718A	187 (12.5)		E156del*	1,429 (95.6)		
	T3750I	153 (10.2)		F157del*	1,451 (97.1)		
	ORF1b	P314L*		1,442 (96.5)	ORF3a	S26L*	1,486 (99.4)
		G662S		1,484 (99.3)	ORF7a	V82A*	1,355 (90.6)
P1000L*		1,452 (97.2)	T120I*	1,425 (95.3)			
A1918V		1,225 (81.9)	N	Q9L	325 (21.7)		
		D63G*		1,424 (95.3)			
		D377Y*		1,486 (99.4)			
		R203M*		1,488 (99.5)			
			G215C	1,286 (86)			

Sedici mutazioni e tutte le delezioni riscontrate erano caratteristiche del *lineage*. Quattordici delle 16 mutazioni erano presenti in più dell'87% degli isolati (87.2%-99.6%). La mutazione G142D nel gene S è stata riscontrata in più della metà dei casi (63.4%) mentre la mutazione P681R, sempre nel gene S, non raggiungeva il 10% (9.2%). L'unica mutazione caratteristica della variante ma presente in un numero limitato di isolati era la M:I82T (3.5%). Un elevato numero di mutazioni (n=12), non identificative di questa variante, sono state riscontrate nella regione ORF1a.

Le mutazioni N501Y, tipica della variante Alpha, e E484K, presente nella variante Beta e Gamma, erano presenti nello 0.13% (n=2) delle sequenze analizzate. La mutazione E484Q, presente nei *sub-lineages* Delta, B.1.617.1 (variante Kappa) e B.1.617.3, veniva osservata nello 0.33% (n=5) dei casi.

## 5.2 ANALISI FILOGENETICA E DATAZIONE DEI CLUSTER ITALIANI

### 5.2.1 CLADE 20E.EU1

L'analisi filogenetica di ML sul *dataset* 20E.EU1 ha mostrato che 2,104 su 2,425 (87.6%) sequenze venivano incluse in un totale di 255 *cluster*, di cui 26 comprendevano almeno una sequenza italiana, classificati come segue: 6 (23%) italiani puri (IT), 6 (23%) *singleton* (S) e 14 (54%) misti (M) (Tabella 8).

Tabella 8. Tipo e composizione dei *cluster* del *dataset* internazionale del *clade* 20E.EU1.

	cluster	Totale cluster	n sequenze	Origine delle sequenze		
				ITA	EU	noEU
20E.EU1	IT	6	133	133	0	0
	M	14	646	415	181	50
	S	6	138	6	91	41
	<b>Total</b>	<b>26</b>	<b>917</b>	<b>554</b>	<b>272</b>	<b>91</b>

ITA: sequenze italiane; EU: sequenze europee; noEU: sequenze internazionali.

L'analisi *bayesiana* è stata condotta sugli 11 *cluster* italiani caratterizzati da più del 70% di genomi italiani, di cui 5 misti e 6 puri italiani. L'analisi di regressione *root-to-tip* del segnale temporale ha rivelato un'associazione tra distanze genetiche e giorni di campionamento, con un coefficiente di correlazione di 0.44 e un coefficiente di determinazione [R<sup>2</sup>] di 0.2 (Figura 15).

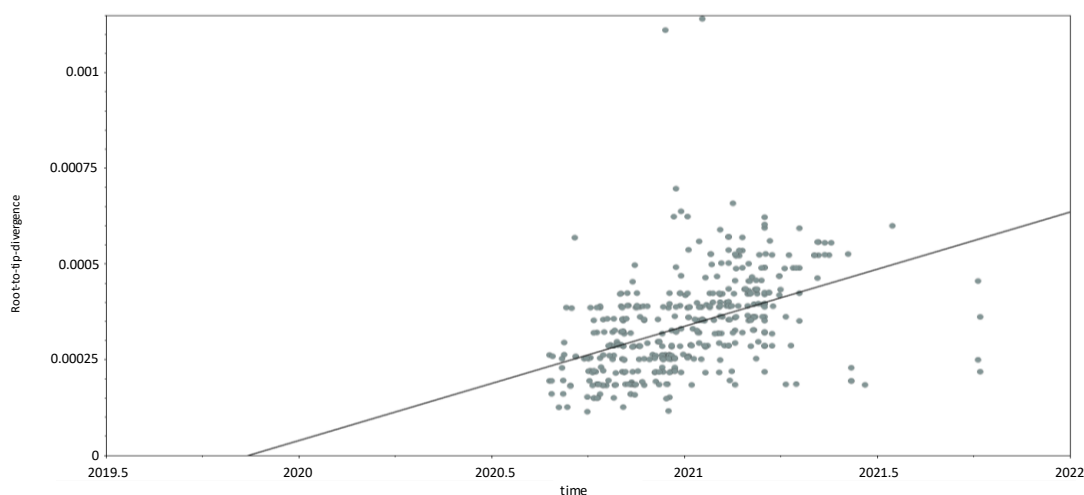


Figura 15. Analisi della regressione *root-to-tip* dei *cluster* italiani del *clade* 20E.EU1.

L'*evolutionary rate* stimato presentava una media di  $3.8 \times 10^{-4}$  sostituzioni/sito/anno (s/s/y) (95%HPD:  $3.35 \times 10^{-4}$ -  $4.39 \times 10^{-4}$ ) e i tMRCAs stimati dei *cluster* datavano da giugno a settembre 2020 (Figura 16, Tabella 9). La dimensione mediana dei *cluster* era di 29 (min 11, max 95) sequenze, con una durata mediana di 11 mesi (*range*: 7-16 mesi). Mentre i *cluster* più precoci, con



tMRCAs datati tra giugno e luglio 2020, mostravano una durata più lunga (9-16 mesi), una maggiore dimensione (con una mediana di 48 genomi;  $p=.01$ ) ed erano prevalentemente misti, i *cluster* più recenti, con tMRCA datati tra agosto e settembre 2020, presentavano una durata più breve (min-max, 7-14 mesi), una più piccola dimensione (mediana di 12 genomi) ed erano prevalentemente italiani.

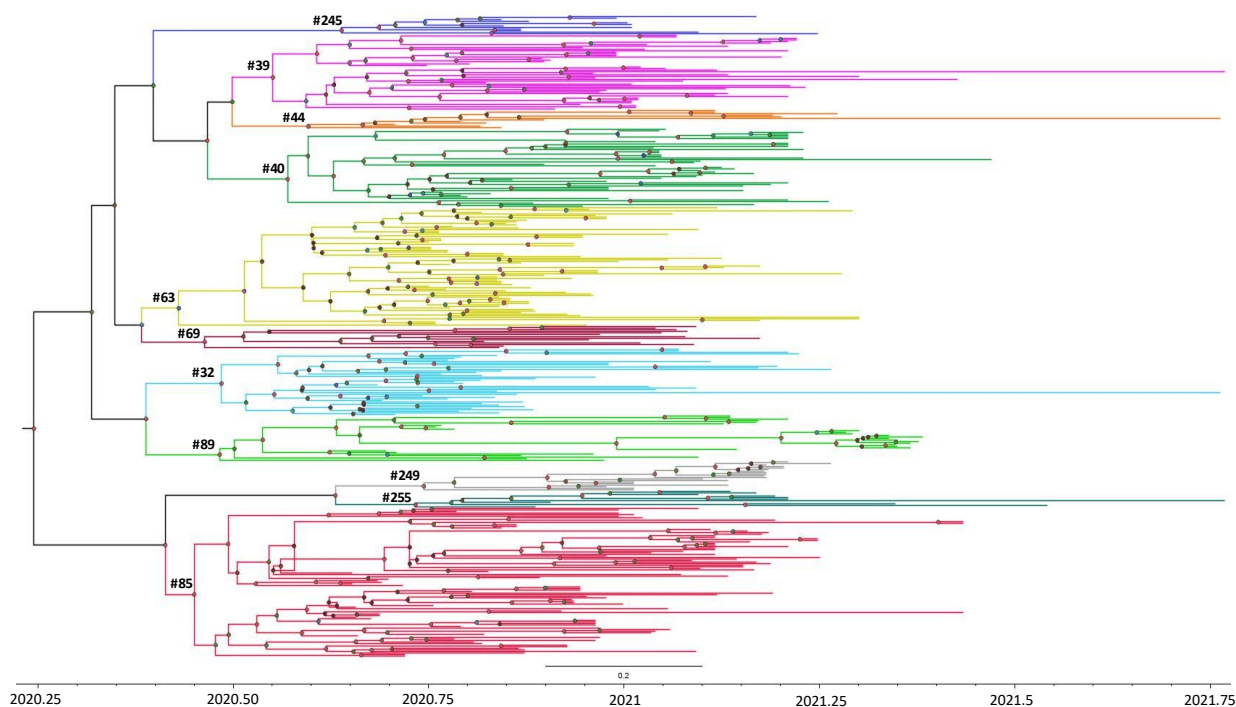


Figura 16. Albero filogenetico *Bayesiano* dei principali *cluster* internazionali del *dataset* di *clade* 20E.EU1. La diversa dimensione dei cerchi rossi sui nodi indica la probabilità a posteriori compresa tra 1 e 0.9. Sono rappresentati con diversi colori i differenti *cluster*. La scala sottostante all'albero rappresenta il tempo espresso in frazioni di anno.

Tabella 9. Stima dei tMRCA dei principali *cluster* internazionali del *dataset* di *clade* 20E.EU1 con relativi intervalli di confidenza, durata temporale e distribuzione geografica delle sequenze italiane.

Cluster ID	Tipo di cluster	n. di sequenze	n. di sequenze italiane	Data	95%HPD <sup>a</sup> Lower	95%HPD <sup>a</sup> Upper	Mesi	Nord Italia % (n)	Italia centrale % (n)	Sud Italia % (n)	Isole maggiori % (n)
<i>radice</i>	/	409	395	30/03/2020	21/02/2020	04/05/2020	19				
85	M	95	88	13/06/2020	19/05/2020	08/07/2020	12	18.2 (16)	14.8 (13)	45.5 (40)	21.6 (19)
69	M	14	13	18/06/2020	30/04/2020	03/08/2020	9	92.3 (12)	/	7.7 (1)	/
89	IT	29	29	25/06/2020	24/05/2020	23/06/2020	11	51.7 (15)	3.4 (1)	44.8 (13)	/
32	M	43	41	26/06/2020	30/05/2020	22/07/2020	16	12.2 (5)	21.9 (9)	63.4 (26)	2.4 /1
63	M	76	74	09/07/2020	04/06/2020	08/08/2020	9	85.1 (63)	1.3 (1)	/	13.5 (10)
39	M	48	46	19/07/2020	24/06/2020	16/08/2020	15	26.1 (12)	10.9 (5)	56.5 (26)	6.5 (3)
40	IT	50	50	26/07/2020	28/06/2020	08/07/2020	11	6 (3)	/	2 (1)	92 (46)
44	IT	12	12	05/08/2020	07/07/2020	03/09/2020	14	8.3 (1)	/	91.7 (11)	/
245	IT	12	12	21/08/2020	17/07/2020	23/09/2020	8	83.3 (10)	/	16.7 (2)	/
255	IT	11	11	25/09/2020	26/08/2020	23/10/2020	13	81.8 (9)	/	18.2 (2)	/
249	IT	19	19	28/09/2020	22/08/2020	05/11/2020	7	/	/	100 (19)	/

<sup>a</sup>HPD: highest posterior density.

Considerando le regioni di campionamento delle sequenze italiane (**Tabella 9**), 5 *cluster* includevano oltre il 50% di sequenze provenienti dal nord Italia, 4 *cluster* dal sud Italia e 1 *cluster* (#40) era caratterizzato per il 92% degli isolati provenienti da Sardegna e Sicilia.

Complessivamente, gli isolati delle isole erano più frequentemente inclusi nei *cluster* rispetto alle sequenze del sud, nord e centro Italia (73.8% vs. 55.3%, 50.6% e 33.3%;  $p=.003$ ).

### 5.2.2 VARIANTE ALPHA

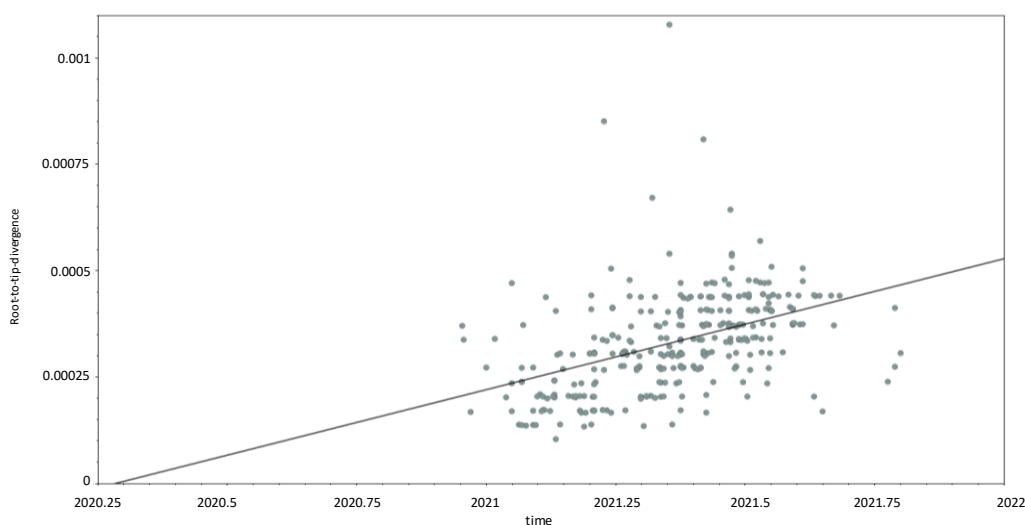
Nel *dataset* di variante Alpha, sono stati identificati 467 *cluster* che comprendevano il 79.1% (2,489/3,147) dei genomi analizzati. Dei 40 *cluster* contenenti almeno 1 sequenza italiana, 14 (60%) erano misti, 15 (37.5%) puri italiani e solo 1 *cluster* (2.5%) era *singleton* (**Tabella 10**).

**Tabella 10. Tipo e composizione dei *cluster* del *dataset* internazionale della variante Alpha.**

	cluster	Totale cluster	n sequenze	Origine delle sequenze		
				ITA	EU	noEU
Alpha	IT	15	235	235	0	0
	M	24	633	245	315	73
	S	1	24	1	22	1
	<b>Total</b>	<b>40</b>	<b>892</b>	<b>481</b>	<b>337</b>	<b>74</b>

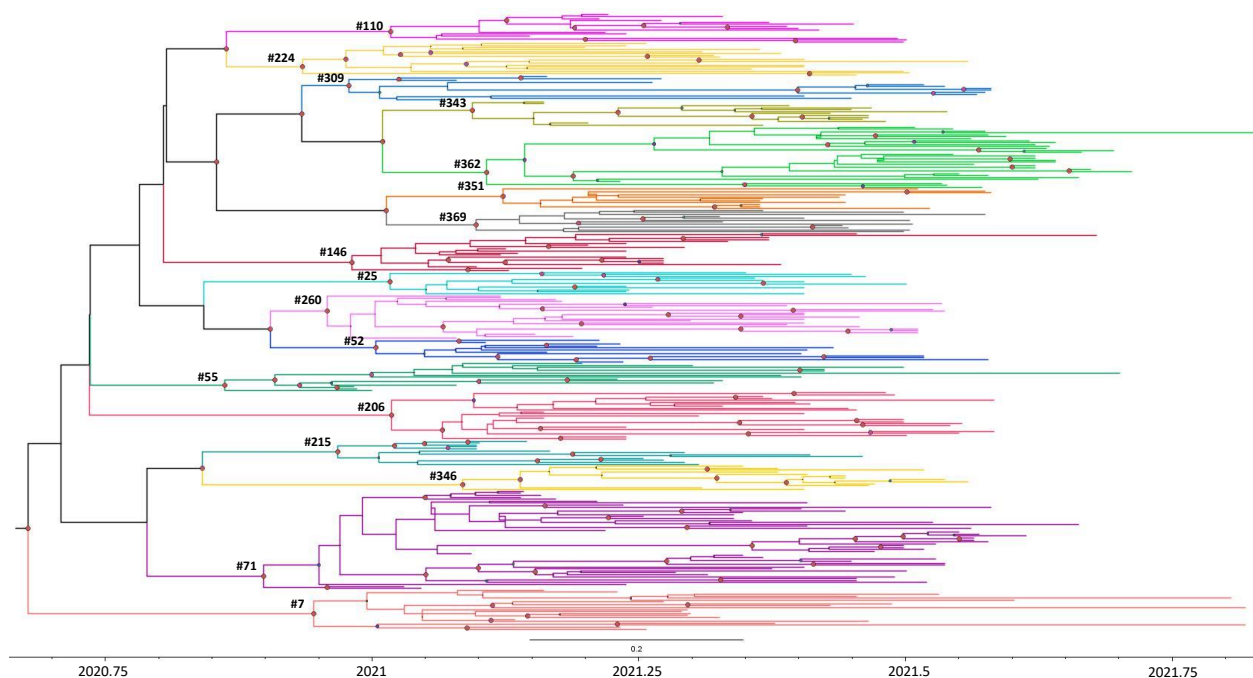
ITA: sequenze italiane; EU: sequenze europee; noEU: sequenze internazionali.

I 17 *cluster* italiani analizzati mediante analisi *Bayesiana*, di cui 15 italiani puri e 2 misti, presentavano un *evolutionary rate* stimato di  $4.87 \times 10^{-4}$  s/s/y (95%HPD:  $4.18 \times 10^{-4}$ - $5.54 \times 10^{-4}$ ). L'analisi di regressione *root-to-tip* del segnale temporale ha mostrato un'associazione tra distanze genetiche e giorni di campionamento, con un coefficiente di correlazione di 0.45 e un coefficiente di determinazione [R<sup>2</sup>] di 0.2 (**Figura 17**).



**Figura 17. Analisi della regressione *root-to-tip* dei *cluster* italiani della variante Alpha.**

Il tMRCA medio dei *cluster* era compreso tra novembre 2020 e febbraio 2021, con il 35.3% (6/17) dei *cluster* con origine a dicembre 2021 (**Figura 18, Tabella 11**). I *cluster* datati novembre-dicembre 2020 includevano una mediana di 18.5 genomi, mentre una mediana di 13 genomi veniva osservata in quelli originati fra gennaio e febbraio. La persistenza dei *cluster* era compresa tra 5 e 10 mesi, con una durata mediana di 7.5 mesi per i più precoci e 5 mesi per i *cluster* più recenti ( $p=.002$ ).



**Figura 18.** Albero filogenetico *Bayesiano* dei principali *cluster* internazionali del *dataset* della variante Alpha. La diversa dimensione dei cerchi rossi sui nodi indica la probabilità a posteriori compresa tra 1 e 0.9. Sono rappresentati con diversi colori i differenti *cluster*. La scala sottostante all'albero rappresenta il tempo espresso in frazioni di anno.

Un'eguale proporzione di *cluster* era caratterizzata da oltre il 50% di sequenze provenienti dal nord e dal sud Italia ( $n=6$ , 35.3%, ciascuna).

Le sequenze delle isole erano più frequentemente incluse nei *cluster* rispetto a quelle dell'Italia settentrionale, centrale e meridionale (53.3% vs. 38.7%, 36.7%, e 29.6%;  $p=.009$ ).

**Tabella 11. Stima dei tMRCA dei principali *cluster* internazionali del *dataset* della variante Alpha con relativi intervalli di confidenza, durata temporale e distribuzione geografica delle sequenze italiane.**

Cluster ID	Tipo di cluster	n. di sequenze	n. di sequenze italiane	Data	95%HPD <sup>a</sup> Lower	95%HPD Upper	Mesi	Nord Italia % (n)	Italia centrale % (n)	Sud Italia % (n)	Isole maggiori % (n)
<b>radice</b>		<b>318</b>	<b>314</b>	<b>05/09/2020</b>	<b>12/08/2020</b>	<b>23/10/2020</b>	<b>13</b>				
55	IT	15	15	10/11/2020	11/10/2020	06/12/2020	9	/	3.7 (1)	93.3 (14)	/
71	M	51	50	25/11/2020	26/10/2020	20/12/2020	9	54 (27)	28 (14)	18 (9)	/
224	IT	17	17	06/12/2020	02/11/2020	08/01/2021	7	64.7 (11)	/	11.8 (2)	23.5 (4)
7	IT	21	21	13/12/2020	03/11/2020	19/01/2021	10	14.3 (3)	9.5 (2)	71.4 (15)	4.6 (1)
260	IT	23	23	16/12/2020	20/11/2020	08/01/2021	7	65.2 (15)	/	30.4 (7)	4.3 (1)
215	IT	13	13	20/12/2020	24/11/2020	11/01/2021	6	76.9 (10)	/	23.1 (3)	/
146	IT	20	20	24/12/2020	25/11/2020	18/01/2021	8	95 (19)	/	5 (1)	/
309	IT	13	13	24/12/2020	02/12/2020	11/01/2021	7	/	46.1 (6)	53.8 (7)	/
52	IT	12	12	01/01/2021	06/12/2020	26/01/2021	6	75 (9)	16.7 (2)	8.3 (1)	/
25	IT	12	12	08/01/2021	06/12/2020	06/02/2021	5	/	/	100 (12)	/
110	IT	15	15	08/01/2021	09/12/2020	06/02/2021	5	33.3 (5)	53.3 (8)	6.7 (1)	6.7 (1)
206	IT	25	25	08/01/2021	09/12/2020	02/02/2021	6	/	/	100 (25)	/
346	IT	13	13	02/02/2021	27/12/2020	07/03/2021	5	/	7.7 (1)	/	92.3 (12)
343	IT	13	13	02/02/2021	11/01/2021	24/02/2021	5	92.3 (12)	/	/	7.7 (1)
369	IT	12	12	06/02/2021	05/01/2021	07/03/2021	5	16.7 (2)	/	83.3 (10)	/
351	IT	11	11	13/02/2021	08/01/2021	25/03/2021	5	90.9 (10)	/	9.1 (1)	/
362	M	32	29	20/02/2021	15/01/2021	07/03/2021	8	6.9 (2)	48.3 (14)	37.9 (11)	6.9 (2)

<sup>a</sup>HPD: highest posterior density.

### 5.2.3 VARIANTE DELTA

L'analisi filogenetica della variante Delta ha mostrato che la quasi totalità degli isolati (2,650/3,248; 81.6%) si raggruppava in 364 *cluster* contenenti più di 3 sequenze; tra questi, 42 erano caratterizzati da almeno una sequenza italiana, di cui 4 (9.5%) *singleton*, 36 misti (85.7%) e 2 puri italiani (4.8%) (Tabella 12).

**Tabella 12. Tipo e composizione dei *cluster* del *dataset* internazionale della variante Delta.**

	cluster	Origine delle sequenze				
		Totale cluster	n sequenze	ITA	EU	noEU
Delta	IT	2	27	27	0	0
	M	36	1,434	737	611	86
	S	4	75	4	14	57
	<b>Total</b>	<b>42</b>	<b>1,536</b>	<b>768</b>	<b>625</b>	<b>143</b>

ITA: sequenze italiane; EU: sequenze europee; noEU: sequenze internazionali.

I 9 *cluster* caratterizzati da più del 70% di isolati italiani contenevano in media 16 isolati (da un minimo di 11 a un massimo di 79 sequenze). L'analisi della regressione *root-to-tip* del segnale

temporale ha mostrato un'associazione tra le distanze genetiche e la data di campionamento, con un coefficiente di correlazione di 0.6 e coefficiente di determinazione ( $R^2$ ) di 0.4 (Figura 19).

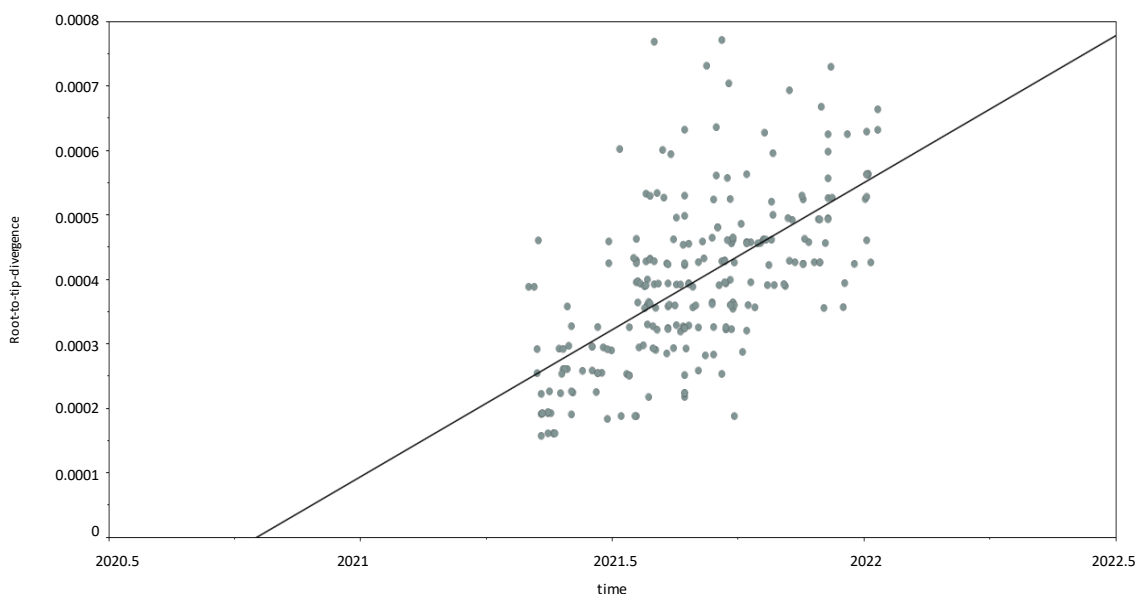


Figura 19. Analisi della regressione *root-to-tip* dei *cluster* italiani della variante Delta.

L'*evolutionary rate* stimato era di  $7.57 \times 10^{-4}$  s/s/a (95%HPD:  $6.65 \times 10^{-4}$ - $8.48 \times 10^{-4}$ ) e la stima dei *tMRCAs* dei *cluster* era compresa tra i mesi di marzo e luglio 2021 (Figura 20, Tabella 13).

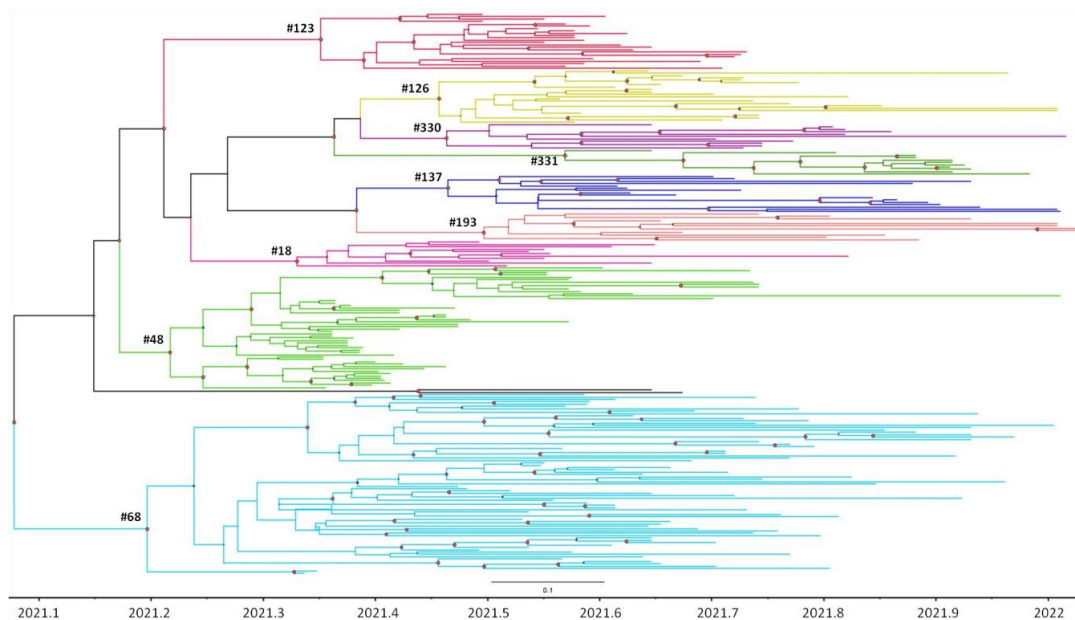


Figura 20. Albero filogenetico *Bayesiano* dei principali *cluster* internazionali del *dataset* della variante Delta. La diversa dimensione dei cerchi rossi sui nodi indica la probabilità a posteriori compresa tra 1 e 0.9. Sono rappresentati con diversi colori i differenti *cluster*. La scala sottostante all'albero rappresenta il tempo espresso in frazioni di anno.

I due *cluster* di maggiori dimensioni (#48 e #68) erano misti, originavano a marzo 2021 e mostravano una durata media di 10 mesi, superiore agli altri *cluster* la cui durata era di 7 mesi ( $p=.03$ ).

Sette *cluster* erano caratterizzati per la maggior parte da sequenze provenienti dal nord Italia (fino a 100%, *cluster* 331) e uno, il #137, includeva unicamente sequenze collezionate nel sud Italia. Il *cluster* 193 conteneva invece un'alta proporzione di sequenze provenienti da nord (44.4%) e centro (55.5%) Italia. Le sequenze delle isole erano più frequentemente incluse nei *cluster* rispetto ai genomi provenienti dal nord e centro Italia (60% vs. 54.2%, 49.4%, 47.9%;  $p=.0002$ ).

**Tabella 13.** Stima dei tMRCA dei principali *cluster* internazionali del *dataset* della variante Delta con relativi intervalli di confidenza, durata temporale e distribuzione geografica delle sequenze italiane.

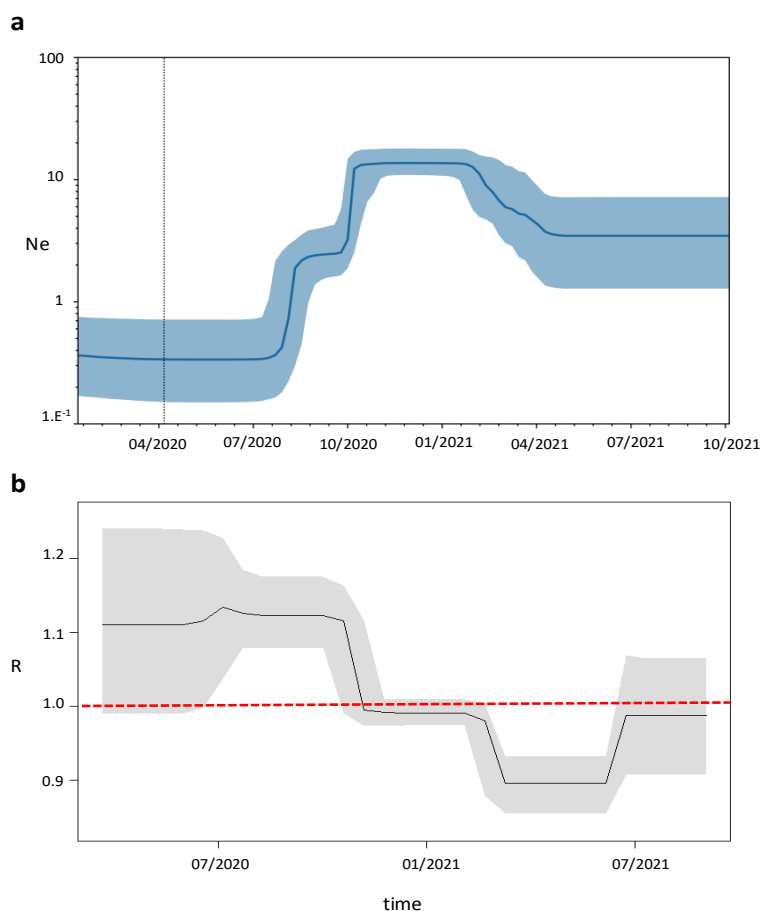
Cluster ID	Tipo di cluster	n. di sequenze	n. di sequenze italiane	Data	95%HPD <sup>a</sup> Lower	95%HPD Upper	Mesi	Nord Italia % (n)	Italia centrale % (n)	Sud Italia % (n)	Isole maggiori % (n)
redice		242	196	03/01/2021	23/12/2020	04/03/2021	12				
68	M	79	58	13/03/2021	11/02/2021	09/04/2021	10	36.2 (21)	17.2 (10)	15.5 (9)	31.3 (18)
48	M	65	52	20/03/2021	26/02/2021	09/04/2021	10	78.9 (41)	1.9 (1)	7.7 (4)	11.5 (6)
18	IT	11	11	30/04/2021	07/04/2021	23/05/2021	6	81.8 (9)	9.1 (1)	9.1 (1)	/
123	M	24	17	08/05/2021	15/04/2021	29/05/2021	4	58.8 (10)	17.6 (3)	23.5 (4)	/
126	M	25	22	16/06/2021	24/05/2021	07/07/2021	7	81.8 (18)	13.6 (3)	4.5 (1)	/
330	M	11	10	18/06/2021	21/05/2021	17/07/2021	7	50 (5)	20 (2)	20 (2)	10 (1)
137	IT	16	16	19/06/2021	24/05/2021	13/07/2021	7	/	/	100 (16)	/
193	M	12	9	30/06/2021	02/06/2021	28/07/2021	7	44.4 (4)	55.5 (5)	/	/
331	M	11	10	27/07/2021	22/06/2021	24/08/2021	5	100 (10)	/	/	/

<sup>a</sup>HPD: highest posterior density.

## 5.3 ANALISI FILODINAMICA DEI CLUSTER ITALIANI

### 5.3.1 CLADE 20E.EU1

L'analisi di *Skyline Plot* delle sequenze 20E.EU1 (**Figura 21, a**), ha mostrato un rapido incremento del numero effettivo di infezioni ( $N_e$ ) nel luglio 2020, seguito da un secondo aumento esponenziale nel mese di ottobre 2020, quando l'infezione raggiungeva una fase di *plateau*. A partire da febbraio 2021, si osservava una diminuzione delle infezioni fino alla stabilizzazione dal mese di aprile.

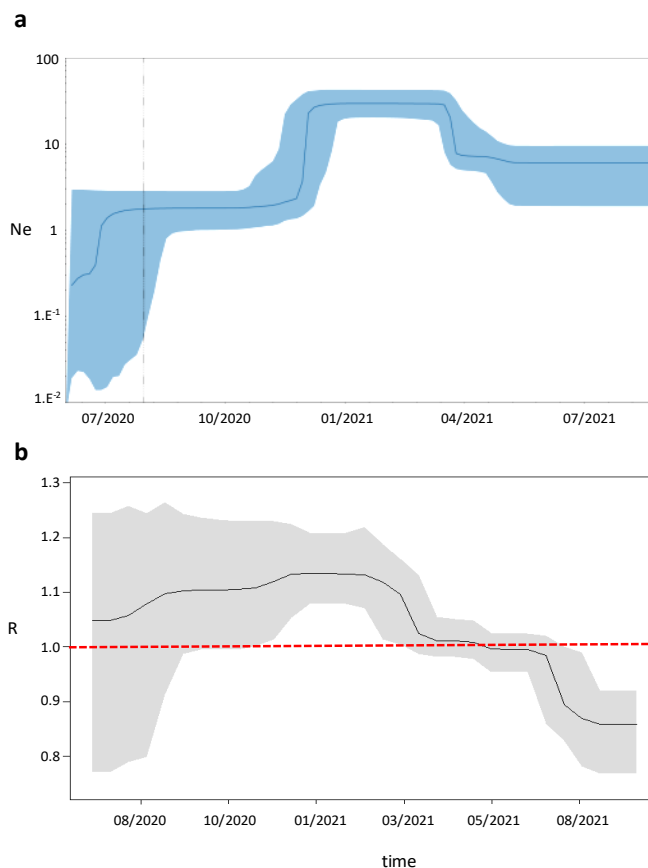


**Figura 21. (a) Bayesian skyline plot del clade 20E.EU1. L'asse y indica la popolazione effettiva ( $N_e$ ), l'asse x mostra il tempo in date. La spessa linea presente nel grafico indica la mediana del valore della stima, mentre l'area blu indica il 95%HPD. (b) Birth-death skyline plot, in relazione al tempo (asse x) e al tasso di riproduzione effettivo ( $R_e$ ) (asse y).**

L'analisi di *birth-death skyline plot* (**Figura 21, b**), coerentemente con la dinamica di infezione, ha mostrato, fin dall'inizio dell'epidemia, valori di  $R_e$  maggiori di 1, raggiungendo un picco di 1.125 (95%HPD: 1.08-1.18) intorno al mese di luglio 2020. A partire dal mese di ottobre 2020 si osservava un rapido decremento dell' $R_e$ , che si assestava a valori intorno all'unità tra i mesi di dicembre 2020 e gennaio 2021, per poi decrescere ulteriormente a valori sotto l'1 in aprile.

### 5.3.2 VARIANTE ALPHA

Lo *skyline* della variante Alpha ha mostrato un incremento esponenziale del numero delle infezioni a dicembre 2020, seguito da una fase di *plateau* nell'inverno 2021. Una rapida diminuzione veniva osservata tra il mese di marzo e aprile 2021, per poi stabilizzarsi nei mesi successivi (**Figura 22, a**).



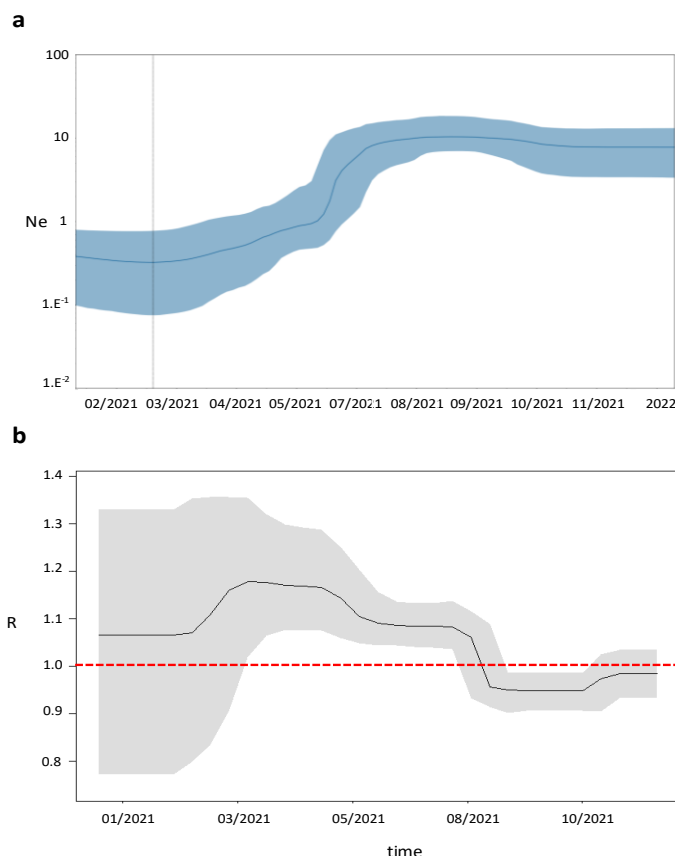
**Figura 22. (a) Bayesian skyline plot della variante Alpha.** L'asse y indica la popolazione effettiva ( $N_e$ ), l'asse x mostra il tempo in date. La spessa linea presente nel grafico indica la mediana del valore della stima, mentre l'area blu indica il 95%HPD. **(b) Birth-death skyline plot,** in relazione al tempo (asse x) e al tasso di riproduzione effettivo ( $R_e$ ) (asse y).

Analogamente, l'analisi di *birth-death skyline plot*, ha mostrato valori di  $R_e$  poco sopra l'unità dall'origine dell'epidemia della variante Alpha nell'autunno 2020, con il raggiungimento del picco a valori di 1.13 (95%HPD:1.01-1.22) nell'inverno 2020/2021. A partire da marzo e aprile 2021 si verificava un decremento di tale valore che raggiungeva valori al di sotto dell'unità dal mese di maggio 2021 (**Figura 22, b**).



### 5.3.3 VARIANTE DELTA

L'analisi di *Skyline Plot* della variante Delta mostrava un aumento del numero delle infezioni nella primavera 2021, tra i mesi di marzo e giugno, seguito da un incremento esponenziale nel luglio 2021. Dal mese di agosto, il numero delle infezioni raggiungeva il *plateau* che persisteva fino alla fine del 2021 (**Figura 23, a**).



**Figura 23. (a) Bayesian skyline plot della variante Delta.** L'asse y indica la popolazione effettiva ( $N_e$ ), l'asse x mostra il tempo in date. La spessa linea presente nel grafico indica la mediana del valore della stima, mentre l'area blu indica il 95%HPD. **(b) Birth-death skyline plot,** in relazione al tempo (asse x) e al tasso di riproduzione effettivo ( $R_e$ ) (asse y).

Analogamente, il valore  $R_e$  (**Figura 23, b**) mostrava un incremento a valori superiori all'unità nella primavera 2021 con il raggiungimento di un picco di 1.17 (95%HPD: 1.1-1.3) nel mese di marzo, seguito da una graduale diminuzione a valori inferiori a 1 nell'estate dello stesso anno (agosto 2021).

## 6. RISULTATI (2)

### EPIDEMIOLOGIA GENOMICA ITALIANA DELLE PRINCIPALI VARIANTI OMICRON CIRCOLANTI IN ITALIA NELL'ANNO 2022

#### 6.1 ANALISI DELLE MUTAZIONI NEI GENOMI DI SARS-CoV-2

##### 6.1.1 OMICRON BA.1

La **Tabella 14** mostra la composizione in termini di *sub-lineage* delle sequenze incluse nel *dataset* italiano della variante Omicron BA.1.

**Tabella 14.** Composizione in termini di *sub-lineages* del *dataset* italiano della variante Omicron BA.1.

Omicron BA.1	Totale n=880 (%)
BA.1.1	290 (32.9)
BA.1.17.2	209 (23.8)
BA.1/BA.1.1.529	160 (18.2)
BA.1.17	74 (8.4)
BA.1.1.1	37 (4.2)
BA.1.15	16 (1.8)
BA.1.18	14 (1.6)
BA.1.1.11	14 (1.6)
BA.1.21.1	12 (1.4)
BA.1.1.14	11 (1.3)
BA.1.20	9 (1.0)
BA.1.21	9 (1.0)
Altri*	25 (2.8)

\**sub-lineages* con una prevalenza inferiore all'1%: BA.1.10, BA.1.13, BA.1.14, BA.1.15.1, BA.1.16, BA.1.19, BA.1.1.10, BA.1.1.13, BA.1.1.15, BA.1.1.18, BA.1.1.3.

La comparazione tra i genomi italiani e la sequenza di riferimento ha mostrato la presenza di 49 sostituzioni aminoacidiche e 7 delezioni in più del 10% degli isolati. In particolare, più di 30 mutazioni erano presenti nella proteina spike. Oltre il 90% delle sequenze presentava le mutazioni distintive di questa variante e dei suoi discendenti (**Tabella 15**). La mutazione V1187I nell'ORF1a, riscontrata nel 32.4% (n=285) dei genomi e caratteristica dei *sub-lineages* BA.1.17 e BA.1.17.2, era presente rispettivamente nel 90.5% (67/74) e nel 97.6% (204/209) di questi isolati. Allo stesso modo, la mutazione R346K nella proteina spike, riscontrata complessivamente nel 39.2% (n=345) delle sequenze, era presente nel 94.8% (n=275) dei genomi BA.1.1 (n=290) e nella quasi totalità (88.4%, n=61) delle sequenze BA.1.1.1 (n=69) e discendenti. La mutazione G446S, nel gene S, tipica

dei *sub-lineages* BA.1.1 e discendenti, è stata riscontrata complessivamente in 665 isolati (75.6%), di cui il 73.3% (n=545) di BA.1.1 e discendenti (n=746). Il 20.5% (n=180) dei genomi riportava la mutazione A701V, caratteristica del *sub-lineage* BA.1.17.2, riscontrata nell'85.6% (179/209) delle sequenze appartenenti a questo *sub-lineage*. In aggiunta, la totalità delle sequenze di *sub-lineages* BA.1.15 e discendenti (17/17), presentavano la mutazione addizionale D343G nella proteina N, distintiva di questo *sub-lineages*.

**Tabella 15. Mutazioni identificate nella variante BA.1. Il simbolo (\*) indica le mutazioni caratteristiche della variante e dei suoi discendenti e in arancione sono indicate le delezioni. In colore azzurro indica le mutazioni proprie dei discendenti BA.1.**

Gene	Mutazione	Totale n=880 (%)		
ORF 1a	K856R*	830 (94.3)	T478K*	495 (56.3)
	<a href="#">V1887I</a>	<a href="#">285 (32.4)</a>	E484A*	511 (58.1)
	<a href="#">S2083del*</a>	<a href="#">863 (98.1)</a>	Q493R*	492 (55.9)
	L2084I*	862 (97.9)	G496S*	475 (54)
	A2710T*	813 (92.4)	Q498R*	446 (50.7)
	T3255I*	850 (96.6)	N501Y*	448 (50.9)
	P3395H*	851 (96.7)	Y505H*	465 (52.8)
	<a href="#">L3674del*</a>	<a href="#">849 (96.5)</a>	T547K*	839 (95.3)
	<a href="#">S3675del*</a>	<a href="#">852 (96.8)</a>	D614G*	812 (92.3)
	<a href="#">G3676del*</a>	<a href="#">851 (96.7)</a>	H655Y*	832 (94.5)
ORF1b	I3758V*	811 (92.2)	N679K*	818 (92.9)
	P314L*	879 (99.9)	P681H*	807 (91.7)
	I1566V*	832 (94.5)	<a href="#">A701V</a>	<a href="#">180 (20.5)</a>
	A67V*	843 (95.8)	N764K*	839 (95.3)
	<a href="#">H69del*</a>	<a href="#">843 (96.5)</a>	D796Y*	799 (90.8)
	<a href="#">V70del*</a>	<a href="#">849 (96.1)</a>	N856K*	851 (96.7)
	T95I*	777 (88.3)	Q954H*	851 (96.7)
	G142del*	814 (92.5)	N969K*	812 (92.3)
	<a href="#">VY 143-144del*</a>	<a href="#">815 (92.6)</a>	L981F*	809 (91.9)
	Y145D*	874 (99.3)	E T9I*	822 (93.4)
S	<a href="#">N211del</a>	<a href="#">787 (89.4)</a>	D3G*	757 (86)
	L212I*	772 (87.7)	M Q19E*	810 (92)
	G339D*	771 (87.6)	A63T*	848 (93.5)
	<a href="#">R346K</a>	<a href="#">345 (39.2)</a>	ORF9b P10S*	877 (99.7)
	S371L*	722 (82)	<a href="#">ENA 27-29del*</a>	<a href="#">880 (100)</a>
	S373P*	754 (85.7)	P13L*	880 (100)
	S375F*	727 (82.6)	<a href="#">ER 31-32del*</a>	<a href="#">880 (100)</a>
	K417N*	668 (75.9)	N <a href="#">S33del*</a>	<a href="#">872 (99.1)</a>
	N440K*	628 (71.4)	R203K*	845 (96)
	<a href="#">G446S</a>	<a href="#">665 (75.6)</a>	G204R*	845 (96)
S477N*	495 (56.3)			

### 6.1.2 OMICRON BA.2

La **Tabella 16** mostra la composizione in termini di *sub-lineage* delle sequenze incluse nel *dataset* italiano della variante Omicron BA.2.

**Tabella 16.** Composizione in termini di *sub-lineages* del *dataset* italiano della variante Omicron BA.2.

Omicron BA.2	Totale n=1,627 (%)
BA.2	1,124 (69.1)
BA.2.9	265 (16.3)
BA.2.12.1	66 (4.1)
BA.2.52	33 (2.0)
BA.2.3	25 (1.5)
BA.2.10	21 (1.3)
BA.2.22	18 (1.1)
BA.2.3.15	18 (1.1)
Altri*	57 (3.5)

\**sub-lineages* con una prevalenza inferiore all'1%: BA.2.10.1, BA.2.12, BA.2.13, BA.2.18, BA.2.23, BA.2.32, BA.2.36, BA.2.37, BA.2.44, BA.2.45, BA.2.49, BA.2.56, BA.2.58, BA.2.65, BA.2.67, BA.2.75.1, BA.2.75.2, BA.2.75.5, BA.2.3.20, BA.2.3.7, BA.2.5, BA.2.9.7.

Nel *dataset* della variante BA.2 e discendenti, sono state riscontrate solo mutazioni e delezioni tipiche di questo *lineage*, in un più dell'80% degli isolati, come evidenziato nella **Tabella 17**. Nella regione ORF3a, la mutazione H78Y, presente complessivamente nel 16.5% (n=268) degli isolati, è stata riscontrata nella maggior parte (88.5%, 234/265) degli isolati di *sub-lineage* BA.2.9. Nella regione ORF1a tre mutazioni, C655R, A2909V e Q3966H osservate in meno del 5% nel *dataset* globale, sono state identificate in quasi il 100% degli isolati di *sub-lineages* BA.2.52.2 (33/33), BA.2.3 (44/44) e BA.2.22 (16/18), dove queste sostituzioni sono caratteristiche. Allo stesso modo, la mutazione S959P nell'ORF1b, con una prevalenza globale dell'1.4%, era presente nel 95.2% (20/21) delle sequenze BA.2.10.

La mutazione L140F in ORF3a, osservata in meno del 3% degli isolati, era presente nella quasi totalità delle sequenze BA.2.3 e discendenti (97.7%, 42/43). Nella proteina S, due mutazioni caratteristiche del *sub-lineage* BA.2.12.1, L452Q e S704L, non raggiungevano il 5% di prevalenza nel *dataset* complessivo, ma venivano identificate rispettivamente nel 94.5% (63/66) e nel 98.5% (65/66) degli isolati di questo *lineage* derivato.

Tabella 17. Mutazioni identificate nella variante BA.2. Il simbolo (\*) indica le mutazioni caratteristiche della variante e dei suoi discendenti e in arancione sono indicate le delezioni. In colore azzurro indica le mutazioni caratteristiche dei discendenti BA.2.

Gene	Mutazione	Totale n=1,627 (%)		
ORF 1a	S135R*	1,603 (98.5)	E484A*	1,506 (92.6)
	T842I*	1,465 (90)	Q493R*	1,417 (87.1)
	G1307S*	1,626 (99.9)	Q498R*	1,240 (76.2)
	L3027F*	1,618 (99.4)	N501Y*	1,277 (78.5)
	T3090I*	1,516 (93.2)	Y505H*	1,296 (79.7)
	L3201F*	1,470 (90.3)	D614G*	1,512 (92.9)
	T3255I*	1,621 (99.6)	H655Y*	1,618 (99.4)
	P3395H*	1,624 (99.8)	N679K*	1,567 (96.3)
	SG 3675-3676del*	1,627 (100)	P681H*	1,594 (98)
	F3677del*	1,626 (99.9)	N764K*	1,607 (98.8)
ORF1b	P314L*	1,626 (99.9)	D796Y*	1,549 (95.2)
	R1315C*	1,619 (99.5)	Q954H*	1,621 (99.6)
	I1566V*	1,553 (95.4)	N969K*	1,566 (96.3)
	T2163I*	1,534 (94.3)	ORF3a	<a href="#">H78Y</a> <a href="#">268 (16.5)</a>
T19I*	1,614 (99.2)	T223I*	1,618 (99.4)	
LPP 24-26del	1,616 (99.3)	E	T9I*	1,537 (94.5)
A27S	1,614 (99.2)	M	Q19E*	1,562 (96)
G142D*	1,547 (95)		A63T*	1,622 (99.7)
V213G*	1,565 (96.2)	ORF6	D61L*	1,523 (93.6)
G339D*	1,333 (81.9)		P10S	1,620 (99.6)
S371F*	1,506 (92.5)	ORF9b	EN 27-28del	1,627 (100)
S373P*	1,518 (93.3)		A29del	1,626 (99.9)
S375F*	1,491 (91.6)		P13L*	1,625 (99.9)
T376A*	1,482 (91.1)		E31del	1,626 (99.9)
D405N*	1,572 (96.6)		R32del	1,627 (100)
R408S*	1,562 (96)	N	S33del	1,610 (98.9)
K417N*	1,564 (96.1)		R203K*	1,625 (99.9)
N440K*	1,391 (85.5)		G204R*	1,624 (99.8)
S477N*	1,397 (85.9)		S413R*	1,549 (95.2)
T478K*	1,426 (87.6)			

### 6.1.3 OMICRON BA.5

La **Tabella 18** mostra la composizione in termini di *lineage* delle sequenze incluse nel *dataset* della variante Omicron BA.5.

**Tabella 18.** Composizione in termini di *sub-lineages* del *dataset* italiano della variante Omicron BA.5.

Omicron BA.5	Totale n=1,761 (%)
BA.5.1	539 (30.6)
BA.5.2	350 (19.9)
BA.5.2.1	330 (18.7)
BA.5.1.10	101 (5.7)
BA.5	49 (2.8)
BA.5.2.20	48 (2.7)
BA.5.1.8	37 (2.1)
BA.5.2.6	32 (1.8)
BA.5.1.5	29 (1.6)
BA.5.1.23	27 (1.5)
BA.5.2.21	26 (1.5)
BA.5.1.30	21 (1.2)
BA.5.1.22	17 (1.0)
Altri*	155 (8.8)

\**sub-lineages* con una prevalenza inferiore all'1%: BA.5.11, BA.5.1.1, BA.5.1.2, BA.5.1.3, BA.5.1.9, BA.5.1.12, BA.5.1.18, BA.5.1.21, BA.5.1.24, BA.5.1.25, BA.5.1.28, BA.5.1.33, BA.5.1.37, BA.5.2.14, BA.5.2.16, BA.5.2.19, BA.5.2.22, BA.5.2.24, BA.5.2.25, BA.5.2.26, BA.5.2.27, BA.5.2.28, BA.5.2.33, BA.5.2.34, BA.5.2.35, BA.5.2.36, BA.5.2.47, BA.5.2.53, BA.5.2.59, BA.5.2.62, BA.5.2.2, BA.5.2.3, BA.5.2.7, BA.5.2.8, BA.5.2.9, BA.5.3, BA.5.3.1, BA.5.5, BA.5.5.1, BA.5.6, BA.5.8, BA.5.9.

Nel *dataset* della variante BA.5 e discendenti sono state identificate 50 mutazioni e 5 delezioni in più del 10% delle sequenze analizzate (**Tabella 19**). Queste sostituzioni sono state identificate in oltre l'80% degli isolati, con l'eccezione della mutazione T1050N, nella regione ORF1b, che pur raggiungendo una prevalenza globale del 20.8% (n=367) era presente nella quasi totalità degli isolati di *sub-lineage* BA.5.2, BA.5.2.2 e discendenti, in cui questa sostituzione è caratteristica, e la mutazione D16G nella regione ORF9b, presente in quasi il 50% (n=844) degli isolati. Questa sostituzione veniva riscontrata prevalentemente nella totalità degli isolati BA.5.2 (n=350) e nel 97.6% (322/330) delle sequenze BA.5.2.1. Nella proteina S, oltre alle mutazioni caratteristiche della variante, il 98.1% (n=1721) delle sequenze presentava la mutazione G142Y.

Sostituzioni globalmente inferiori al 10% ma proprie di differenti *sub-lineages* sono state identificate nelle regioni ORF1a, ORF1b, S ed N. Nella regione ORF1a, le mutazioni S302F, Q556K, K3839R e T4161I, venivano osservate rispettivamente nella totalità delle sequenze BA.5.1.23 (n=27), BA.5.3.1 (n=16), BA.5.1.10 (n=101) e BA.5.1.8 (n=37). Nella regione ORF1b, la mutazione

S1273L è stata riscontrata nella totalità degli isolati BA.5.1.22 (n=17); le mutazioni R346T e P1263Q nel gene S, si trovavano nella totalità degli isolati BA.5.2.6 (n=32) e BA.5.1.5 (n=25). Nella regione N, la mutazione E136D veniva riscontrata nella totalità delle sequenze BA.5.3.1 (n=16).

**Tabella 19. Mutazioni identificate nella variante BA.5. Il simbolo (\*) indica le mutazioni caratteristiche della variante e dei suoi discendenti e in arancione sono indicate le delezioni. In colore azzurro indica le mutazioni caratteristiche dei discendenti BA.5.**

Gene	Mutazione	Totale n=1,761 (%)			
ORF 1a	S135R*	1,753 (99.5)	S	S477N*	1,589 (90.2)
	T842I*	1,652 (93.8)		T478K*	1,657 (94.1)
	G1307S*	1,758 (99.8)		E484A*	1,735 (98.5)
	L3027F*	1,743 (99)		F486V*	1,735 (98.5)
	T3090I*	1,654 (93.9)		Q498R*	1,622 (92.1)
	T3255I*	1,749 (99.3)		N501Y*	1,647 (93.5)
	P3395H*	1,753 (99.5)		Y505H*	1,661 (94.3)
	SGF 3675-3677del*	1,760 (99.9)		D614G*	1,654 (93.9)
	P314L*	1,756 (99.7)		H655Y*	1,759 (99.9)
	<a href="#">T1050N</a>	<a href="#">367 (20.8)</a>		N679K*	1,704 (96.7)
ORF1b	R1315C*	1,754 (99.6)	P681H*	1,709 (97)	
	I1566V*	1,732 (98.4)	N764K*	1,736 (98.6)	
	T2163I*	1,665 (94.5)	D796V*	1,725 (98)	
	T19I*	1,753 (99.5)	Q954H*	1,758 (99.8)	
	L24del*	1,753 (99.5)	N969K*	1,732 (98.3)	
	PP 25-26del*	1,752 (99.5)	ORF3a	T223I*	1,747 (99.2)
	A27S*	1,749 (99.3)	E	T9I*	1,716 (97.4)
	HV 69-70del*	1,749 (99.3)	M	D3N*	1,732 (98.4)
	F140-141del*	1,755 (99.6)		Q19E*	1,685 (95.7)
	S	G142Y*	1,728 (98.1)	A63T*	1,756 (99.7)
V213G*		1,742 (98.9)	P10S	1,724 (97.9)	
G339D*		1,495 (84.9)	<a href="#">D16G</a>	<a href="#">844 (47.9)</a>	
S371F*		1,622 (92.1)	ORF9b	EN 27-28del*	1,745 (99.1)
S373P*		1,640 (93.1)	A29del*	1,628 (92.4)	
S375F*		1,603 (91.0)	P13L*	1,740 (98.8)	
T376A*		1,604 (91.1)	E31del*	1,745 (99.1)	
D405N*		1,724 (97.9)	R32del*	1,744 (99)	
R408S*		1,686 (95.7)	N	S33del*	1,607 (91.2)
K417N*		1,729 (98.2)		R203K*	1,760 (99.9)
N440K*		1,590 (90.3)		G204R*	1,759 (99.9)
L452R*		1,728 (98.1)		S413R*	1,731 (98.3)

## 6.2 ANALISI E DATAZIONE FILOGENETICA DELLE VARIANTI OMICRON

L'analisi filogenetica condotta sulle sequenze italiane della variante Omicron BA.1 ha evidenziato la presenza di 46 *cluster*, caratterizzati da più di due sequenze, che includevano il 29.8% (n=262) del totale delle sequenze analizzate (n=880). La dimensione dei *cluster* andava da un minimo di 3 sequenze a un massimo di 21. Il 34.5% (n=16) dei *cluster* era caratterizzato da sole 3 sequenze e 4 (8.7%) *cluster* includevano più di 10 genomi. Non si notava una diversa tendenza a *clusterizzare* sulla base dell'area di campionamento (nord, sud, centro Italia e isole).

L'analisi della variante Omicron BA.2 ha mostrato la presenza di 111 *cluster*, che includevano il 32% delle sequenze analizzate (521/1,627); il 45.9% dei *cluster* (n=51) era composto da sole 3 sequenze mentre 7 (6.3%) *cluster* includevano più di 10 genomi. Non veniva riscontrata una differente tendenza a *clusterizzare* sulla base dell'area di campionamento.

Il 37.9% (n=668) degli isolati inclusi nel *dataset* della variante Omicron BA.5 (n=1,761) si raggruppava in 135 *cluster*, di cui la metà (49.6%, n=67) conteneva solo 3 isolati. Complessivamente, 10 (7.4%) *cluster* includevano un numero maggiore/uguale a 10 genomi. Le sequenze delle isole si raggruppavano più frequentemente nei *cluster* rispetto agli isolati provenienti da nord, centro e sud Italia (52.3% vs. 37.6%, 35.6%, 37.2%;  $p < .05$ ).

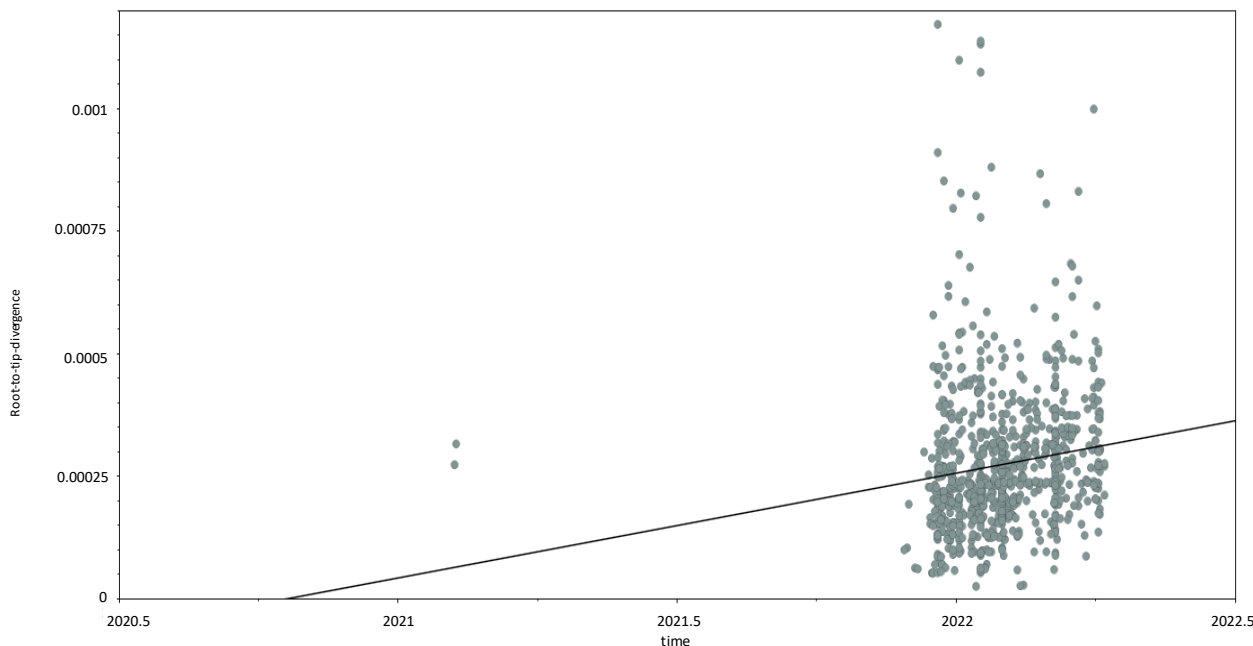
## 6.3 ANALISI BAYESIANA DEI CLUSTER ITALIANI

### 6.3.1 OMICRON BA.1

Dato il limitato numero di *cluster* contenenti più di 10 isolati, l'analisi filodinamica *Bayesiana* è stata condotta su un *dataset* che comprendeva tutte le sequenze incluse nei *cluster* contenenti più di 3 sequenze (n=214).

L'analisi della regressione *root-to-tip* del segnale temporale ha rivelato una relazione lineare fra le distanze genetiche e la data di campionamento, con un coefficiente di correlazione di 0.14 e coefficiente di determinazione (R<sup>2</sup>) di 0.02 (**Figura 24**).





**Figura 24.** Analisi della regressione *root-to-tip* dell'allineamento contenente le 214 sequenze dei *cluster* maggiori di 3 sequenze.

La stima dell'*evolutionary rate* presentava una media di  $4.84 \times 10^{-4}$  s/s/y (95%HPD:  $3.76 \times 10^{-4} - 5.98 \times 10^{-4}$ ).

La radice dell'albero datato è stata stimata corrispondente a fine giugno 2021 (95%HPD: 09/04/2021-17/09/2021). Si distinguevano 12 *cluster* significativi con una datazione compresa tra i mesi di settembre e novembre 2021 (95%HPD: 21/06/2021 e 09/12/2021) (**Figura 25**).

Tali *cluster* contenevano una media di 17.8 genomi (minimo di 4 e massimo di 71) con una persistenza compresa tra 2 e 7 mesi (**Tabella 20**). I *cluster* più precoci, datati settembre 2021, comprendevano un maggior numero di isolati (mediana: 20.5 genomi) ed erano più persistenti (mediana: 7 mesi) rispetto ai *cluster* di ottobre/novembre 2021 (dimensione mediana di 6.5 isolati e una persistenza mediana di 4.5 mesi).

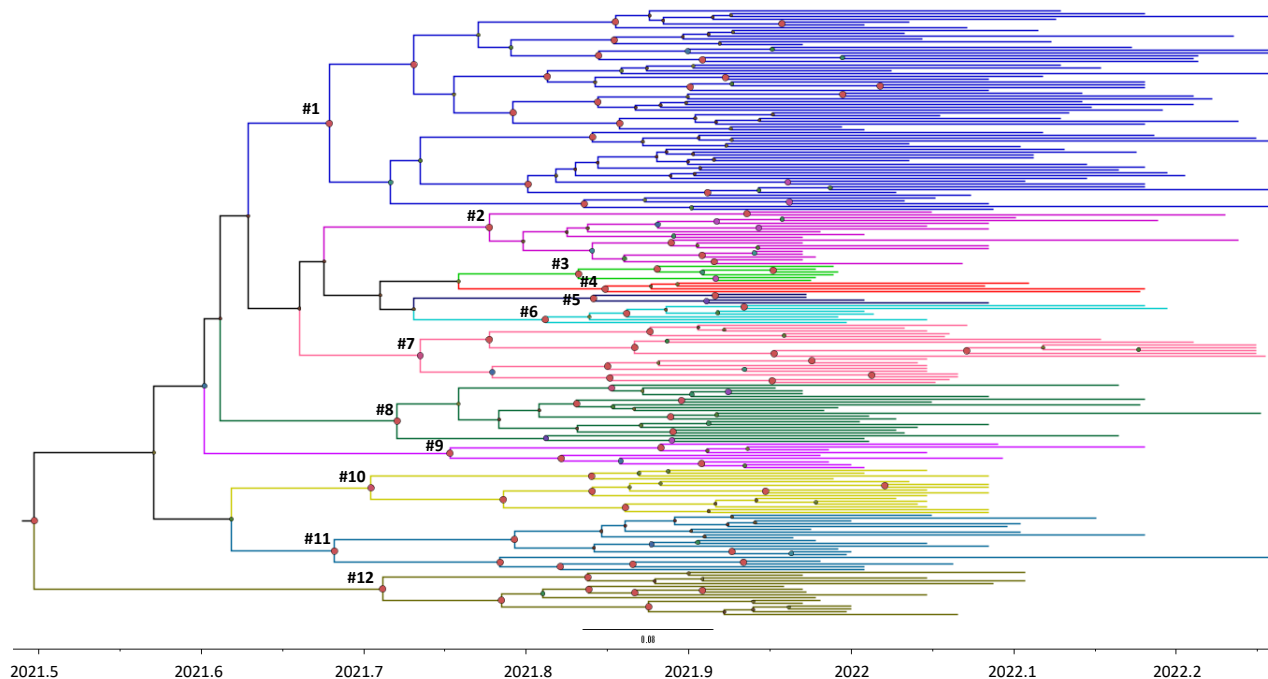


Figura 25. Albero filogenetico *Bayesiano* delle sequenze italiane BA.1 incluse nei *cluster*. La diversa dimensione dei cerchi rossi sui nodi indica la probabilità a posteriori compresa tra 1 e 0.9. Sono rappresentati con diversi colori i differenti *cluster*. La scala sottostante all'albero rappresenta il tempo espresso in frazioni di anno.

La datazione dei *cluster* ottenuta mediante questa analisi è mostrata in **Tabella 20**.

Tabella 20. Stima dei tMRCA dei principali *cluster* con i relativi intervalli di confidenza, data di campionamento del genoma più recente e durata nel tempo di ciascun *cluster*.

	n. di sequenze	$pp^a$	data	95%HPD <sup>b</sup> Lower	95%HPD Upper	Campione più recente	mesi
<b>radice</b>	214	1	30/06/2021	09/04/2021	17/09/2021	05/04/2022	10
#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

<sup>a</sup> $pp$ : posterior probability;

<sup>b</sup>HPD: highest posterior density.

Complessivamente, 3 *cluster* erano caratterizzati da oltre l'80% di isolati provenienti dal nord Italia (#11, #10, #12) (Tabella 21). Altri 3 mostravano una prevalenza (>65%) di sequenze del nord Italia, uno del sud Italia (#3) e uno del centro Italia (#5). I restanti *cluster* includevano un numero simile di isolati provenienti da diverse aree del Paese.

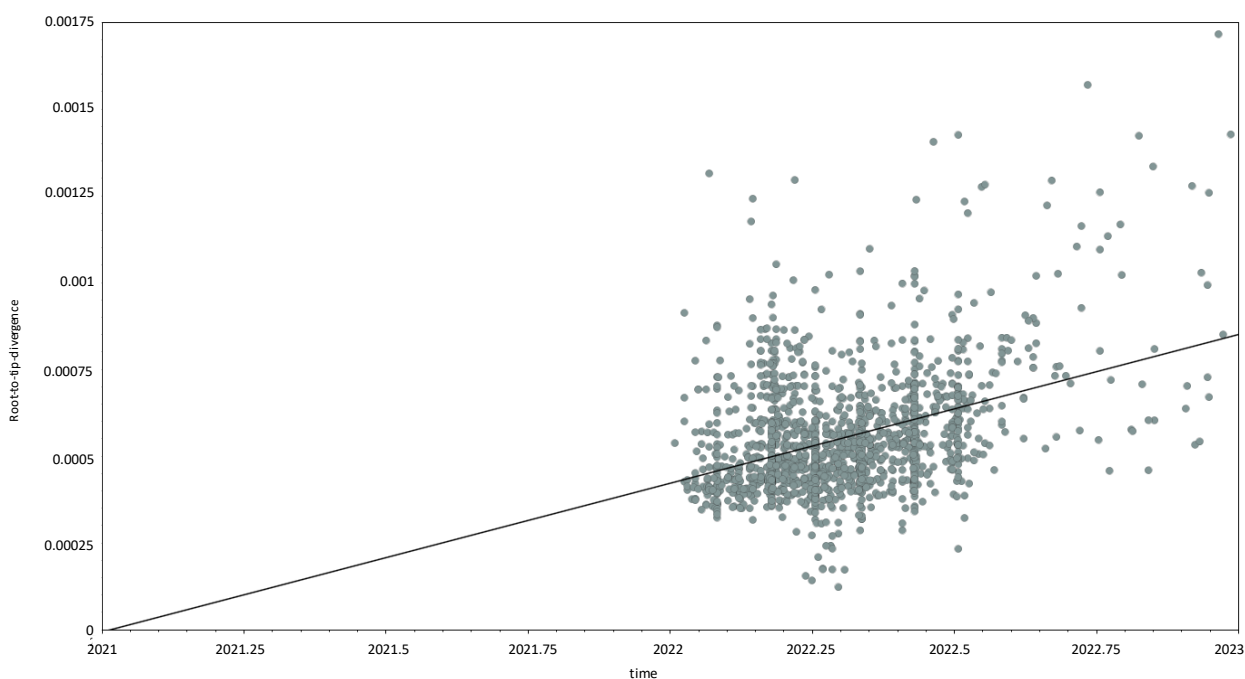
**Tabella 21. Distribuzione regionale delle sequenze incluse in ciascun *cluster*.**

<i>cluster</i>	n. di sequenze	nord Italia n (%)	sud Italia n (%)	centro Italia n (%)	isole n (%)
#1	71	36 (50.7)	11 (15.5)	19 (26.8)	5 (7.0)
#11	20	19 (95)	/	1 (5)	/
#10	16	13 (81.2)	3 (18.8)	/	/
#12	16	15 (93.8)	/	/	1 (6.3)
#8	21	14 (66.7)	1 (4.8)	4 (19)	2 (9.5)
#7	21	6 (28.6)	/	10 (47.6)	5 (23.8)
#9	9	6 (66.7)	1 (11.1)	2 (22.2)	/
#2	19	7 (36.8)	4 (21.1)	8 (42.1)	/
#6	7	1 (14.3)	1 (14.3)	5 (71.4)	/
#3	6	/	4 (66.6)	1 (16.7)	1 (16.7)
#5	4	1 (25)	2 (50)	1 (25)	/
#4	4	3 (75)	1 (25)	/	/

### 6.3.2 OMICRON BA.2

Per la variante Omicron BA.2, l'analisi filodinamica *Bayesiana* è stata condotta su un *dataset* che comprendeva tutte le sequenze incluse nei 7 *cluster* italiani contenenti più di 10 sequenze (#4, #6, #67, #76, #190, #201, #203) per un totale di 111 genomi.

L'analisi della regressione *root-to-tip* del segnale temporale ha rivelato una relazione lineare fra le distanze genetiche e la data di campionamento, con un coefficiente di correlazione di 0.39 e coefficiente di determinazione (R<sup>2</sup>) di 0.154 (**Figura 26**).



**Figura 26.** Analisi della regressione *root-to-tip* dell'allineamento contenente le 111 sequenze dei *cluster* più numerosi.

La stima dell'*evolutionary rate* mostrava una media di  $3.99 \times 10^{-4}$  s/s/y (95%HPD:  $2.70 \times 10^{-4}$ - $5.33 \times 10^{-4}$ ). I *cluster* analizzati contenevano una media di 15.9 genomi (minimo di 10 e massimo di 30), presentavano una datazione compresa tra novembre 2021 e gennaio 2022 (95%HPD: 28/08/2021 e 23/02/2022) e avevano una persistenza compresa tra 5 e 8 mesi (**Figura 27, Tabella 22**).

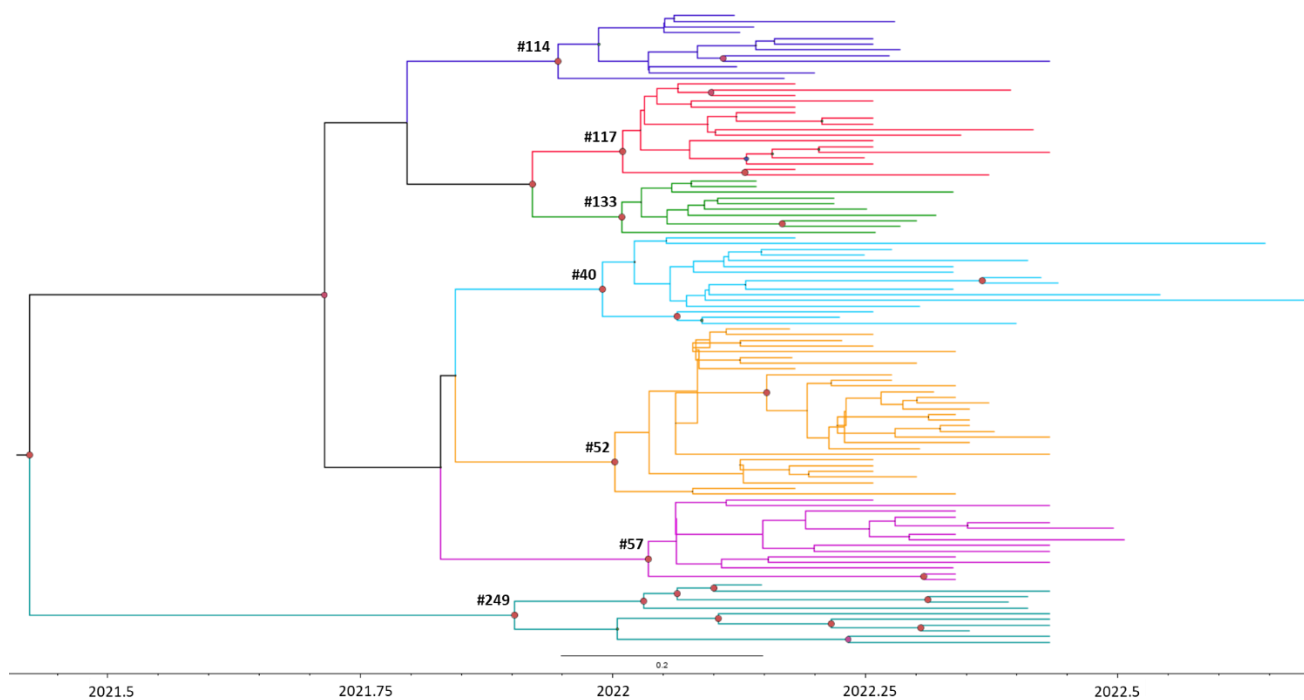


Figura 27. Albero filogenetico *Bayesiano* delle sequenze italiane BA.2 incluse nei *cluster*. La diversa dimensione dei cerchi rossi sui nodi indica la probabilità a posteriori compresa tra 1 e 0.9. Sono rappresentati con diversi colori i differenti *cluster*. La scala sottostante all'albero rappresenta il tempo espresso in frazioni di anno.

Tabella 22. Stima dei tMRCA dei principali *cluster* con i relativi intervalli di confidenza, data di campionamento del genoma più recente e durata nel tempo di ciascun *cluster*.

	n. di sequenze	pp <sup>a</sup>	data	95%HPD <sup>b</sup> Lower	95%HPD Upper	Campione più recente	mesi
<b>radice</b>	111	1	03/06/2021	24/12/2020	09/10/2021	08/09/2022	15
#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
#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

<sup>a</sup>pp: posterior probability;

<sup>b</sup>HPD: highest posterior density.

Complessivamente (**Tabella 23**), 4 *cluster* (#114, #52, #57 e #249) mostravano una elevata prevalenza di genomi isolati nel nord Italia, in particolare i *cluster* #249, #52 e #57 includevano più del 90% di sequenze provenienti da questa area (prevalentemente da Liguria e Lombardia). Due *cluster* (#133 e #117) erano formati da più del 70% di isolati ottenuti dal centro Italia (per la maggior parte rispettivamente dalle Marche e dall'Umbria).

Una proporzione simile di sequenze del sud, centro e nord Italia erano presenti nel *cluster* #40, che risultava essere anche quello più persistente (8 mesi).

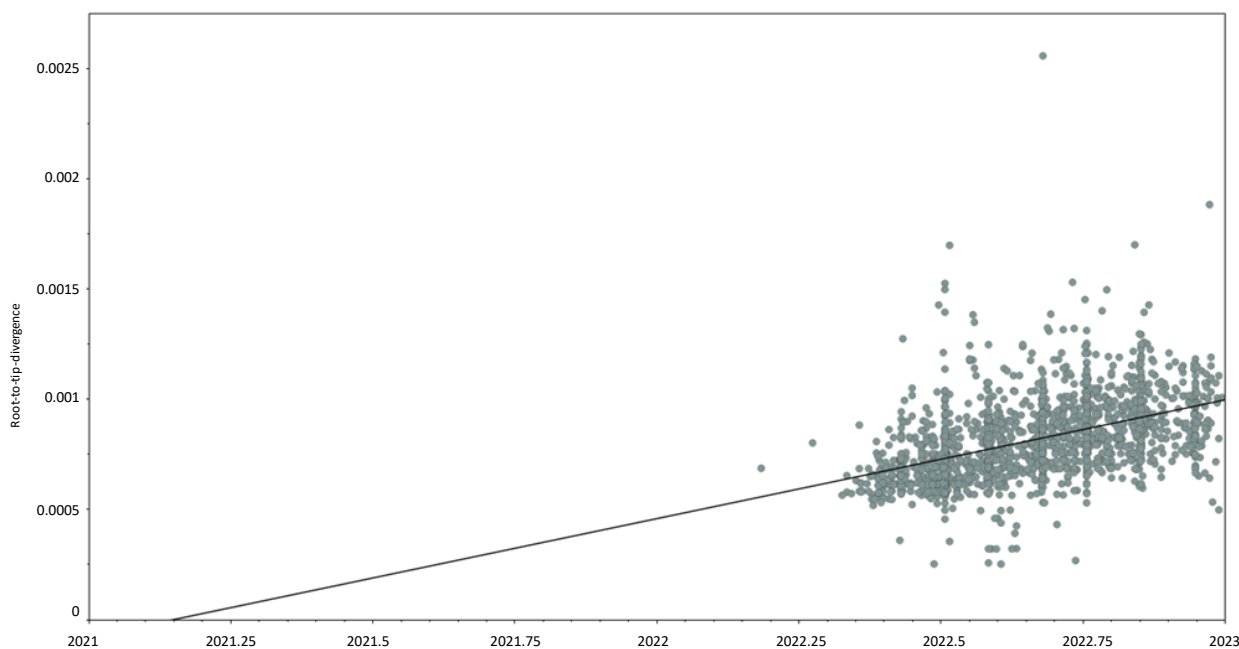
Tabella 23. Distribuzione regionale delle sequenze incluse in ciascun *cluster*.

<i>cluster</i>	n. di sequenze	nord Italia n (%)	sud Italia n (%)	centro Italia n (%)	isole n (%)
#249	11	11 (100)	/	/	/
#114	12	8 (67)	1 (8.3)	1 (8.3)	2 (16.7)
#40	16	4 (25)	6 (37.5)	6 (37.5)	1 (6.3)
#52	30	30 (100)	/	/	/
#133	10	2 (20)	1 (10)	7 (70)	/
#117	17	1 (5.9)	/	15 (88.2)	1 (5.9)
#57	15	14 (93.3)	/	1 (6.7)	/

### 6.3.3 OMICRON BA.5

L'analisi *Bayesiana* della variante Omicron BA.5 è stata condotta sulle 164 sequenze che erano incluse nei 10 *cluster* contenenti più di 10 sequenze.

L'analisi della regressione *root-to-tip* del segnale temporale ha rivelato una relazione lineare fra le distanze genetiche e la data di campionamento, con un coefficiente di correlazione di 0.46 e coefficiente di determinazione (R<sup>2</sup>) di 0.215 (**Figura 28**).



**Figura 28.** Analisi della regressione *root-to-tip* dell'allineamento della variante BA.5 contenente le 164 sequenze dei *cluster* più numerosi.

L'*evolutionary rate* stimato presentava una media di  $4.56 \times 10^{-4}$  s/s/y (95%HPD:  $3.72 \times 10^{-4}$ - $5.44 \times 10^{-4}$ ). I *cluster* mostravano una datazione della radice compresa tra ottobre 2021 e maggio 2022 (95%HPD: 25/07/2021 e 14/07/2022) (**Figura 29**), contenevano una media di 16.4 genomi (minimo 10 e massimo 39) e presentavano una persistenza media di 8.3 mesi (*range*: 5-11 mesi) (**Tabella 24**).

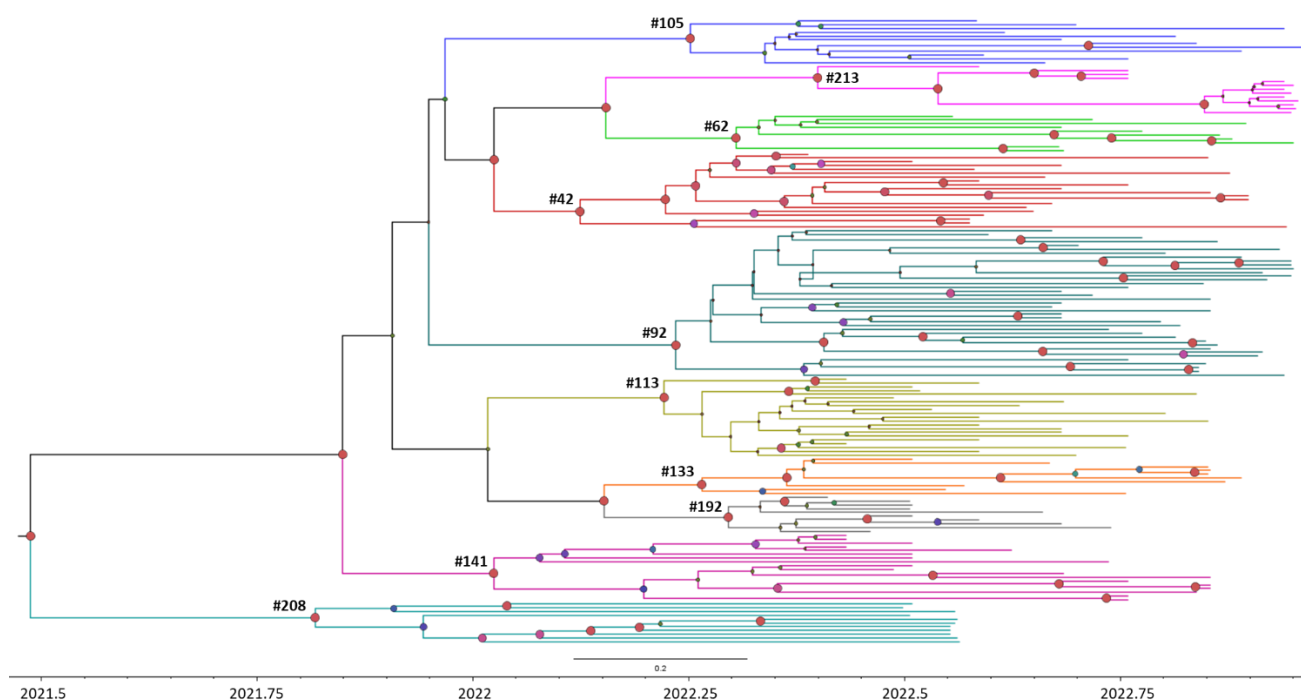


Figura 29. Albero filogenetico *Bayesiano* delle sequenze italiane BA.5 incluse nei *cluster*. La diversa dimensione dei cerchi rossi sui nodi indica la probabilità a posteriori compresa tra 1 e 0.9. Sono rappresentati con diversi colori i differenti *cluster*. La scala sottostante all'albero rappresenta il tempo espresso in frazioni di anno.

Tabella 24. Stima dei tMRCA dei principali *cluster* con i relativi intervalli di confidenza, data di campionamento del genoma più recente e durata nel tempo di ciascun *cluster*.

	n. di sequenze	pp <sup>a</sup>	data	95%HPD <sup>b</sup> Lower	95%HPD Upper	Campione più recente	mesi
<b>radice</b>	164	1	26/06/2021	21/02/2021	17/10/2022	17/12/2022	18
#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

<sup>a</sup>pp: posterior probability.

<sup>b</sup>HPD: highest posterior density.

Il 60% (n=6) dei *cluster* in esame era caratterizzato da isolati provenienti da ogni zona di Italia (nord, sud, centro e isole) (Tabella 25).

Dei 10 *cluster* analizzati, 7 (70%; #42, #62, #105, #133, #141, #208 e #213) erano caratterizzati da oltre il 50% di sequenze provenienti dal nord Italia e uno (#192) presentava una eguale proporzione di sequenze del nord e centro Italia (40%). I restanti *cluster* (#92 e #113) comprendevano sequenze provenienti da tutte le zone di Italia.



Il *cluster* #208 (n=11), il più vecchio, con tMRCA datato 24 ottobre 2021, conteneva l'81.8% (n=9) di sequenze provenienti dal nord Italia, tutte della Lombardia.

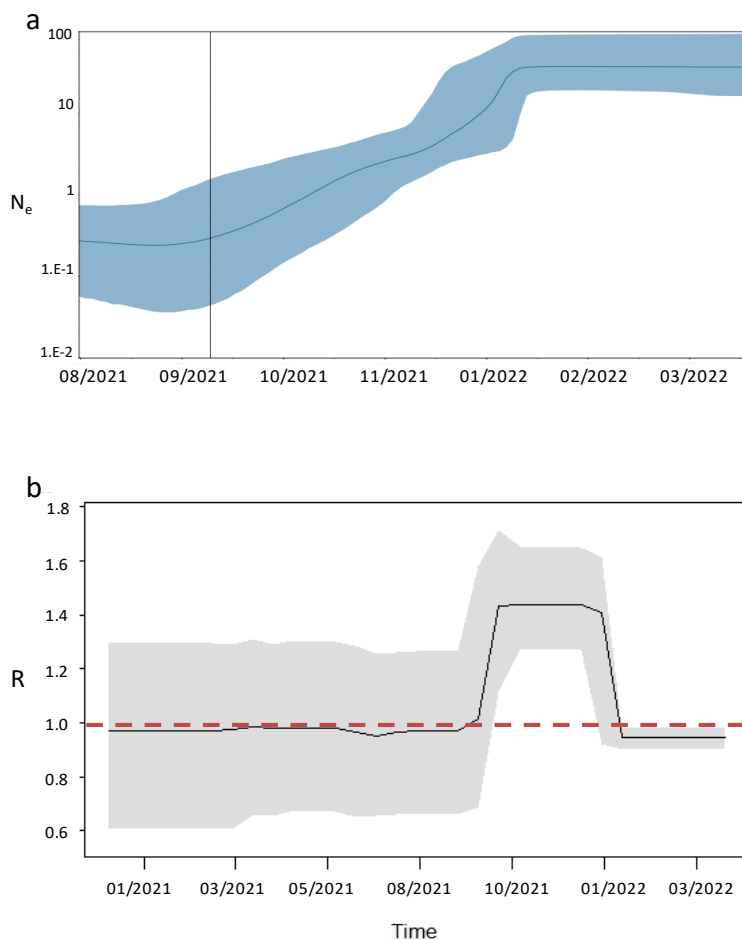
**Tabella 25. Distribuzione regionale delle sequenze incluse in ciascun *cluster*.**

<i>cluster</i>	n. di sequenze	nord Italia n (%)	sud Italia n (%)	centro Italia n (%)	isole n (%)
#208	11	9 (81.8)	2 (18.2)	/	/
#141	18	10 (55.6)	2 (11.1)	5 (27.8)	/
#42	20	11 (55)	5 (25)	3 (15)	1 (5)
#113	21	6 (28.6)	4 (19)	6 (28.6)	5 (23.8)
#92	39	19 (48.7)	8 (20.5)	7 (17.9)	5 (12.8)
#105	12	7 (58.3)	2 (16.7)	2 (16.7)	1 (8.3)
#133	10	6 (60)	1 (10)	2 (20)	1 (10)
#192	10	4 (40)	2 (20)	4 (40)	/
#62	10	6 (60)	1 (10)	2 (20)	1 (10)
#213	13	10 (76.9)	/	3 (23.1)	/

## 6.4 ANALISI FILODINAMICA DEI CLUSTER ITALIANI

### 6.4.1 OMICRON BA.1

Lo *skyline* ottenuto mostrava un iniziale e graduale aumento del numero delle infezioni tra settembre e novembre 2021 seguito da un ulteriore incremento nei mesi di dicembre 2021 e gennaio 2022 quando il numero dei casi si stabilizzava (**Figura 30, a**).



**Figura 30.** (a) *Bayesian skyline plot* della variante BA.1. L'asse y indica la popolazione effettiva ( $N_e$ ), l'asse x mostra il tempo in date. La spessa linea presente nel grafico indica la mediana del valore della stima, mentre l'area blu indica il 95%HPD. (b) *Birth-death skyline plot*, in relazione al tempo (asse x) e al tasso di riproduzione effettivo ( $R_e$ ) (asse y).

L'analisi di *birth-death skyline plot* (**Figura 30, b**) ha mostrato, fin dall'inizio dell'epidemia, un valore di  $R_e$  intorno all'unità (0.98, 95%HPD: 0.67-1.31) che aumentava a partire da settembre 2021 fino a raggiungere un picco di 1.45 (95%HPD: 1.27-1.64) nel mese di ottobre. Da gennaio 2022 si osservava un decremento e stabilizzazione di tale valore attorno all'unità (0.94, 95%HPD: 0.79-0.79).

### 6.4.2 OMICRON BA.2

Lo *skyline* ottenuto mostrava un aumento esponenziale del numero delle infezioni compreso tra gennaio e febbraio 2022 a seguito del quale si osservava una fase di *plateau*, che perdurava per tutto il periodo in esame. (Figura 31, a).

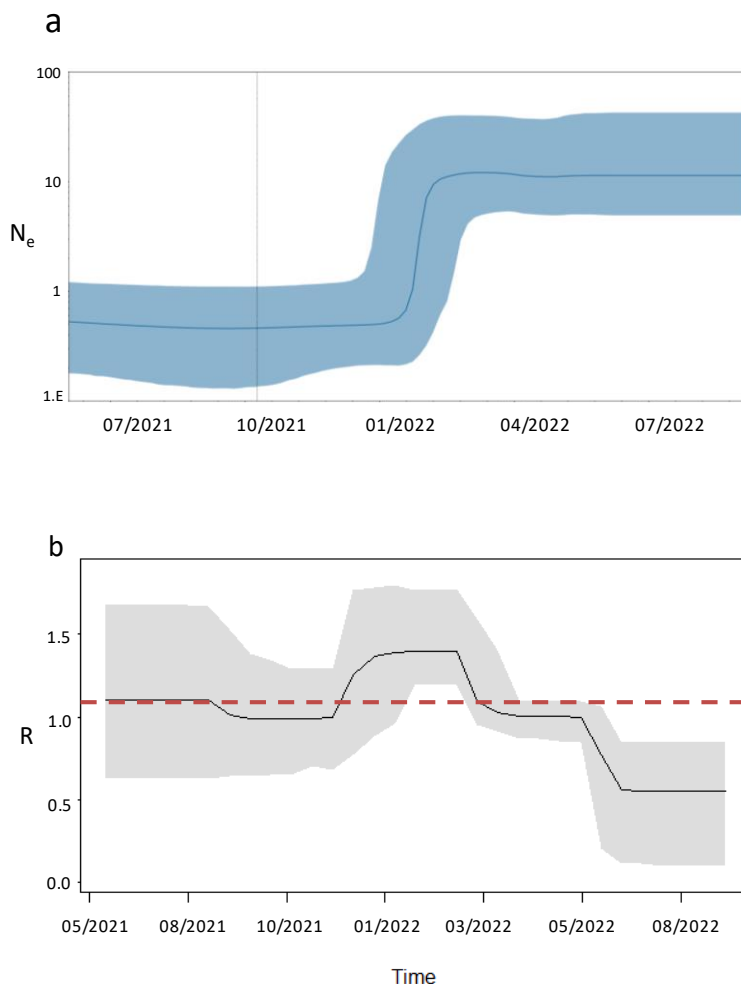
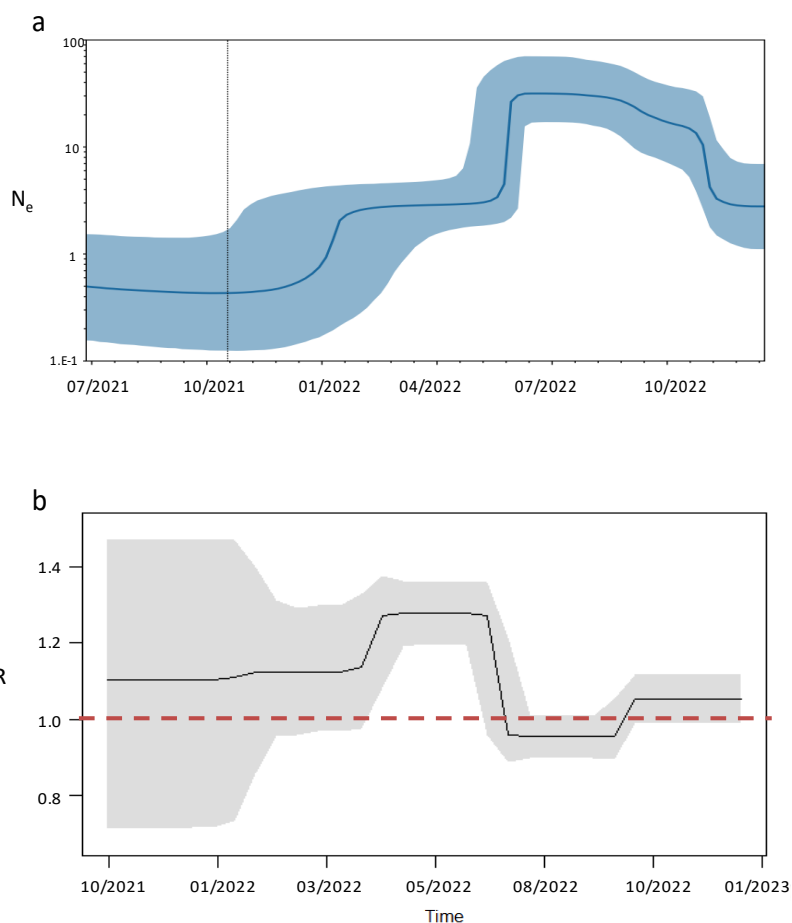


Figura 31. (a) *Bayesian skyline plot* della variante BA.2. L'asse y indica la popolazione effettiva ( $N_e$ ), l'asse x mostra il tempo in date. La spessa linea presente nel grafico indica la mediana del valore della stima, mentre l'area blu indica il 95%HPD. (b) *Birth-death skyline plot*, in relazione al tempo (asse x) e al tasso di riproduzione effettivo ( $R_e$ ) (asse y).

L'analisi di *birth-death skyline plot* (Figura 31, b) mostrava un valore di  $R_e$  poco superiore all'unità all'inizio dell'epidemia (1.11, 95%HPD: 0.63-1.67) che diminuiva intorno a valori di 0.99 (95%HPD: 0.66-1.31) nei mesi compresi tra agosto e novembre 2021. Un aumento di tale valore con raggiungimento del picco di 1.42 (95%HPD: 1.16-1.86) si osservava nei mesi gennaio-febbraio 2022. Un primo decremento del valore di  $R_e$  a 0.98 (95%HPD: 0.84-1.09) si registrava tra febbraio e marzo 2022, seguito da un'ulteriore diminuzione a valori intorno a 0.5 (95%HPD: 0.11-0.84) tra maggio e luglio 2022. Tale valore si manteneva stabile sotto l'unità fino alla fine di agosto 2022.

### 6.4.3 OMICRON BA.5

Lo *skyline* ottenuto dall'analisi filodinamica *Bayesiana* mostrava un primo incremento del numero delle infezioni a gennaio 2021 e un secondo aumento più importante, si natura esponenziale nei mesi di maggio e giugno 2022. A seguito del raggiungimento di una fase di *plateau* ad agosto 2022, si registrava una prima lenta diminuzione del numero dei casi fino ad ottobre 2022 e una seconda, più rilevante, tra ottobre e novembre 2022 (**Figura 32, a**).



**Figura 32.** (a) *Bayesian skyline plot* della variante BA.5. L'asse y indica la popolazione effettiva ( $N_e$ ), l'asse x mostra il tempo in date. La spessa linea presente nel grafico indica la mediana del valore della stima, mentre l'area blu indica il 95%HPD. (b) *Birth-death skyline plot*, in relazione al tempo (asse x) e al tasso di riproduzione effettivo ( $R_e$ ) (asse y).

L'analisi di *birth-death skyline plot* (**Figura 32, b**) ha mostrato un valore di  $R_e$  superiore all'unità (1.10, 95%HPD: 1.71-1.47) sin dall'inizio dell'epidemia, con raggiungimento del picco di 1.28 (95%HPD: 1.19-1.36) nel mese di maggio 2022. A partire da questo periodo si registrava un rapido decremento di tale valore a livelli intorno all'unità (0.95, 95%HPD: 0.89-1.01) nel luglio 2022 per poi risalire e stabilizzarsi a valori di 1.05 (95%HPD: 0.99-1.12) a partire dal mese di settembre.

## 7. DISCUSSIONE

Al mese di ottobre 2023, SARS-CoV-2 ha causato in Italia, uno dei primi Paesi colpiti dalla pandemia, oltre 26 milioni di casi di COVID-19 e più di 192 mila decessi (<https://www.worldometers.info/coronavirus/>).

Il sequenziamento dell'intero genoma di patogeni emergenti rappresenta una componente fondamentale della moderna sorveglianza epidemiologica delle malattie infettive, grazie alle recenti tecnologie di sequenziamento di nuova generazione (NGS) che permettono la rapida produzione di interi genomi di patogeni direttamente da campioni clinici.

La pandemia di COVID-19 ha stimolato sforzi senza precedenti per lo sviluppo di strategie di sorveglianza molecolare in tempo reale, basate sul sequenziamento del genoma virale, come dimostrato dagli oltre 16 milioni di genomi completi di SARS-CoV-2 depositati in una sola delle banche dati dedicate come GISAID ([www.gisaid.org](http://www.gisaid.org)). Questa vasta fonte di dati consente di tracciare le dinamiche evolutive del virus e identificare varianti di potenziale rilevanza clinica per la trasmissibilità e la patogenicità virale.

Nelle fasi iniziali della pandemia COVID-19, l'Italia è stata coinvolta in un'ampia ondata epidemica chiaramente dimostrata dall'elevato numero di ricoveri e decessi, seppure l'effettiva dimensione ed estensione rimangono in parte sconosciute in termini di reale numero di infezioni a causa dell'allora mancanza di un sistema di sorveglianza efficace e alla limitata disponibilità di test diagnostici.<sup>50-52</sup>

Gli studi di epidemiologia genomica hanno dimostrato come, nel corso della prima ondata abbia circolato prevalentemente, in Italia come in tutto il mondo, la variante B.1, caratterizzata dalla mutazione in posizione 614 nella proteina di superficie spike. Le dimensioni della prima ondata epidemica sono diminuite a partire da maggio 2020 e sono rimaste contenute fino al mese di agosto. All'inizio del periodo in esame, il numero dei casi di COVID-19 ha cominciato ad aumentare, probabilmente a seguito della circolazione in Europa del *lineage* 20E a partire dall'estate 2020. Tale ceppo nell'autunno 2020 è risultato responsabile della seconda ondata epidemica in Italia, la più intensa in termini di numero di casi e decessi riportati nell'era pre-Omicron.

A partire da dicembre 2020 è iniziata la circolazione delle varianti di preoccupazione (VoC), varianti di SARS-CoV-2 ampiamente divergenti con numerose mutazioni nella proteina spike, caratterizzate da una elevata trasmissibilità e responsabili quindi delle successive ondate a partire dalla variante

Alpha, nell'inverno 2021, e dalla variante Delta da maggio 2021, responsabile dell'aumento dei casi nell'estate 2021 e della più grande epidemia nell'autunno dello stesso anno, immediatamente prima della diffusione della VoC Omicron che, con i suoi *lineages* derivati, ha dominato il 2022 rappresentando ad oggi l'unica variante in circolazione.

Fin dai primi mesi della pandemia COVID-19 in Italia, un gran numero di centri clinici per la gestione dei pazienti e laboratori in grado di effettuare la caratterizzazione genomica del virus, distribuiti in tutte le regioni d'Italia, ha costituito il *network* collaborativo SCIRE con l'obiettivo di contribuire allo sforzo comune di ottenere e caratterizzare il maggior numero possibile di genomi completi SARS-CoV-2.

Scopo di questo elaborato di tesi è stato lo studio e la descrizione delle dinamiche evolutive ed epidemiologiche dei principali *lineages*/varianti presenti nel nostro Paese nel periodo pre-Omicron compreso tra agosto 2020 e gennaio 2022, e dei principali *lineages* della variante Omicron circolanti nel corso del 2022, mediante filogenesi.

L'analisi filogenetica ci ha permesso di ottenere, per le varianti 20E.EU1, Alpha e Delta, una datazione più precisa dell'ingresso di queste varianti nel nostro Paese.

I genomi analizzati si raggruppavano in oltre 100 *cluster* con un'alta frequenza di *cluster* misti (circa 70%), che comprendevano ceppi circolanti in diverse regioni del mondo, suggerendo l'introduzione multipla di questi *lineages*/varianti in Italia, probabilmente a causa di viaggi internazionali.<sup>37</sup>

L'osservazione di *cluster* precoci di dimensioni maggiori, più persistenti e più frequentemente misti potrebbe essere correlata alle misure di contenimento adottate durante la pandemia, come le restrizioni ai viaggi, che sono state allentate durante l'estate 2020 per poi essere reintrodotte in autunno, con l'arrivo delle varianti di preoccupazione, e gradualmente nuovamente allentate da maggio 2021. I *cluster* italiani puri, che suggeriscono una circolazione locale del virus, risultavano più diffusi nei periodi in cui erano in atto misure restrittive mentre, con l'allentamento delle misure di contenimento, sono diventati sempre meno frequenti. Questo potrebbe anche essere il motivo per cui *cluster* italiani puri sono stati osservati con maggiore frequenza per il *clade* 20E.EU1 e la variante Alpha, rispetto alla variante Delta la cui circolazione in Italia corrispondeva a un periodo di allentamento delle misure restrittive.

L'elevato numero di *cluster* misti ha reso difficile poter stimare sull'intero *dataset* parametri che fanno riferimento a una media di situazioni epidemiologiche molto diverse tra loro. L'analisi dei *cluster* contenenti un'elevata proporzione di genomi italiani ha permesso di isolare dei focolai

epidemici fornendo una stima dei tempi di ingresso e di circolazione locale dei principali *lineages*/varianti in Italia. La maggior parte dei *cluster* 20E.EU1 italiani datavano nell'estate 2020, quelli della variante Alpha principalmente nell'inverno 2020 e 2021, mentre i *cluster* Delta mostravano un'origine stimata tra la primavera e l'estate 2021. È stata osservata anche una parziale segregazione regionale degli isolati; indipendentemente dal *lineage*, i campioni delle isole maggiori si raggruppavano più frequentemente in *cluster* rispetto agli isolati ottenuti in altre parti d'Italia, evidenziando l'importante ruolo della deriva genetica nell'evoluzione del virus.

A sostegno dell'importante ruolo della deriva genetica, la maggior parte delle mutazioni identificate erano tipiche di ogni variante, in particolare per Alpha e per il *clade* 20E.EU1 e non risultavano soggette a pressione selettiva. Al contrario, per la variante Delta, sono state identificate diverse mutazioni aggiuntive, soprattutto nel gene ORF1a, confermando la maggiore eterogeneità genomica di questa variante rispetto alla precedente.

La filodinamica è una disciplina che unisce lo studio dell'epidemiologia con quello dell'evoluzione molecolare degli agenti infettivi. Infatti, per organismi che evolvono rapidamente, come i virus a RNA, i cambiamenti evolutivi si verificano in scala temporale 1:1 con gli eventi epidemiologici, rendendone quindi possibile la ricostruzione sulla base dell'analisi filogenetica.

Nel presente studio le dinamiche di popolazione sono state studiate attraverso due approcci distinti: l'analisi di coalescenza, mediante metodi *Bayesiani (Skyline)*, che stima i cambiamenti nella numerosità della popolazione virale sulla base della teoria della coalescenza, e l'approccio *birth-death*, che a differenza del precedente, tiene in considerazione non solo le "nascite" (cioè i nuovi casi) ma anche le "morti" (i casi "rimossi") e quindi la proporzione di popolazione che viene persa nel corso del campionamento. Mentre il primo approccio consente di valutare i cambiamenti nel tempo (dalla radice dell'albero alle date di campionamento più recenti) e attraverso questi stimare i cambiamenti nelle dimensioni (crescita o stazionarietà o contrazione) ed il tasso di crescita dell'epidemia, il secondo è in grado di fornire direttamente le stime di parametri epidemiologici essenziali come il numero riproduttivo effettivo ( $R_e$ ) dell'infezione.

Per tutte le varianti considerate, è stata osservata una corrispondenza temporale tra l'andamento del numero di infezioni stimato dal grafico *Skyline* e la stima  $R_e$  mediante *birth-death Skyline*. Come è noto, durante una malattia epidemica, il valore di  $R_e$  cambia principalmente in relazione alla diminuzione del numero di soggetti suscettibili. Queste stime concordano con i dati epidemiologici riportati dalla sorveglianza italiana (<https://www.iss.it/coronavirus/>-

/asset\_publisher/1SRKHcCJQ7E/content/faq-sul-calcolo-del-rt), in particolare quelli relativi ai casi ospedalieri e alle stime del parametro  $R_t$ .

Valori più alti di  $R_e$  osservati prima dell'effettiva circolazione di ogni VoC in Italia (estate 2020 per 20E, autunno 2020 per Alpha e marzo 2021 per Delta), risalgono probabilmente all'origine delle epidemie fuori dall'Italia. La variante Delta ha mostrato un picco del valore  $R_e$  superiore a quello delle varianti 20E.EU1 e Alpha, in accordo con altri studi,<sup>53, 54</sup> a causa della maggiore trasmissibilità di questa variante.

Nel complesso, i valori di  $R_e$  stimati erano apparentemente inferiori alle stime di  $R_0$  pubblicate per le diverse varianti (<https://ibz-shiny.ethz.ch/covid-19-re-international/>). Ciò potrebbe essere dovuto alla persistenza in Italia di misure di controllo (come l'allontanamento sociale, l'uso esteso di maschere in pubblico, ecc.) almeno nel primo periodo considerato, e all'attuazione della campagna di vaccinazione iniziata nel nostro Paese alla fine del 2020 e completata da oltre il 75% della popolazione in autunno/inverno 2021 (<https://ourworldindata.org/>). Tuttavia, sono stati presi in considerazione solo alcuni *cluster*, che rappresentano focolai limitati, e l' $R_e$  stimato rappresenta il valore medio dei *cluster* considerati nell'analisi. Questo valore non è quindi rappresentativo dell'intera popolazione italiana, considerando che solo una parte limitata dei casi sono stati campionati in Italia, in particolare prima del 2021.

Un documento pubblicato di recente,<sup>53</sup> sulle stime del valore  $R_e$  basate sui tassi di crescita dei genomi delle VoC depositate in banche dati pubbliche di diversi Paesi del mondo, ha mostrato valori in Italia simili a quelli stimati in questo studio.

La tipica fluttuazione di  $R_e$  osservata durante la pandemia può spiegare il meccanismo di sostituzione della variante precedente, con valore di  $R_e$  al suo minimo (inferiore a 1), con la nuova variante circolante, al suo massimo  $R_e$  ( $R_e > 1$ ).

Dopo la prima segnalazione in Botswana alla fine del 2021, dall'inizio del 2022 la variante Omicron si è rapidamente diffusa in tutto il mondo diventando ad oggi la variante dominante con i suoi *sub-lineages* derivati.<sup>55, 56</sup> La maggiore trasmissibilità, la minore efficacia dei vaccini contro l'infezione, il quadro clinico meno severo, il periodo di incubazione inferiore (4.5-5 giorni per le varianti Alpha e Delta vs. 3.5 giorni per la variante Omicron) e la maggiore efficienza replicativa, hanno reso questa variante un virus molto differente rispetto alle varianti Alpha, Delta e ai *lineages* ancestrali rendendo necessario un aggiornamento dei vaccini attualmente in uso. Inoltre, la presenza in quel periodo di una popolazione in gran parte vaccinata con una o più dosi (oltre l'85% nel nostro Paese; <https://www.lombardianotizie.online/vaccinati-over-80/>) e l'allentamento delle misure di



contenimento, ha cambiato radicalmente l'epidemiologia dell'infezione inducendo a distinguere tra un'era "pre-Omicron" e un'era Omicron.

Tale variante ha sin da subito mostrato un elevato numero di mutazioni rispetto alle varianti apparse in precedenza, con in media 50 mutazioni lungo tutto il genoma di cui più della metà localizzate nella proteina spike e responsabili di una migliore affinità di legame al recettore ACE2 e di una maggiore resistenza ad anticorpi neutralizzanti e monoclonali.<sup>57,58</sup>

Molto rapidamente questa variante si è evoluta dando origine ad una successione di numerosi *lineages* e *sub-lineages* differenti,<sup>26</sup> con peculiari profili mutazionali. Questa eterogeneità è stata riscontrata anche nei *dataset* da noi analizzati, con differenti *sub-lineages* con una frequenza maggiore dell'1% (12 per la variante BA.1, 8 per la variante BA.2 e 13 per BA.5), con i *sub-lineages* BA.1.17.2, BA.1.1, BA.2.9, BA.5.2 e BA.5.2.1 tra i più prevalenti.

Mentre la variante BA.1 si è ad oggi estinta, le varianti BA.2 e BA.5 continuano a circolare con i vari *lineages* derivati e le forme ricombinanti anche grazie alla maggiore trasmissibilità.<sup>58</sup>

Analizzando inoltre il profilo mutazionale delle varianti BA.1, BA.2 e BA.5 in esame, abbiamo riscontrato un elevato numero di mutazioni, localizzate principalmente nel gene S, tutte caratteristiche delle varianti stesse e dei *sub-lineages* derivati.

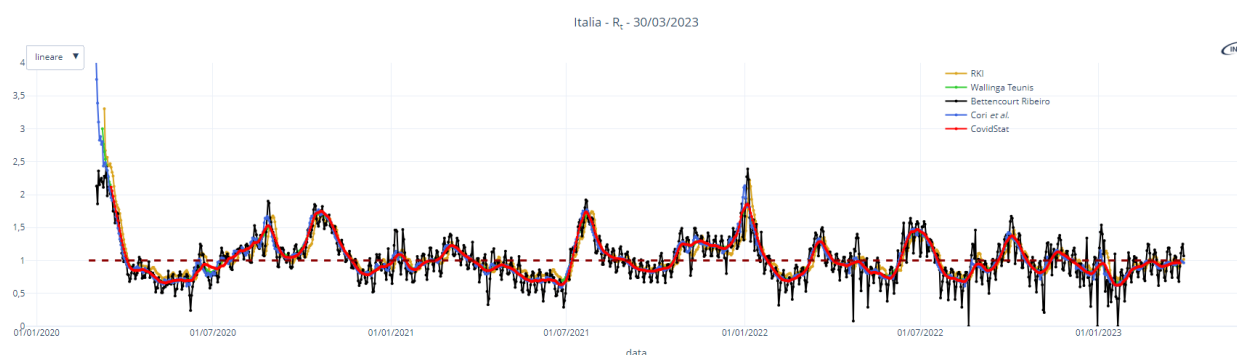
L'elevato numero di mutazioni non solo ha rafforzato l'infettività e la trasmissibilità di queste varianti, ma ha anche compromesso l'immunità vaccinale e/o l'immunità umorale indotta da precedenti infezioni naturali e ha reso inefficaci gli anticorpi monoclonali precedentemente sviluppati. Omicron ha circolato largamente tra i soggetti vaccinati a causa del maggiore potenziale di immuno-*escape* rispetto alla precedente variante Delta.<sup>59, 60</sup>

Le caratteristiche intrinseche della variante e l'assenza di misure di contenimento hanno favorito la circolazione globale<sup>61</sup> e l'evoluzione dei diversi *sub-lineages* di Omicron impedendo l'osservazione di *cluster* di trasmissione consolidati sia su base nazionale che regionale.

Per questo motivo l'analisi filogenetica per queste varianti è stata condotta considerando il solo contesto nazionale in cui abbiamo riscontrato una parziale formazione di *cluster* su base locale principalmente nel caso delle varianti BA.1 e BA.2. Grazie, infatti, a fenomeni di evoluzione neutrale, come l'"effetto fondatore", è stato possibile individuare solo a livello locale (nazionale) dei *cluster* significativi, diversamente da un contesto internazionale in cui, a causa dell'alto grado di correlazione evolutiva tra i ceppi di un'unica variante (BA.1, BA.2 e BA.5), risulta difficoltosa l'identificazione di gruppi di sequenze più affini che condividono un antenato comune recente.<sup>43, 62</sup>

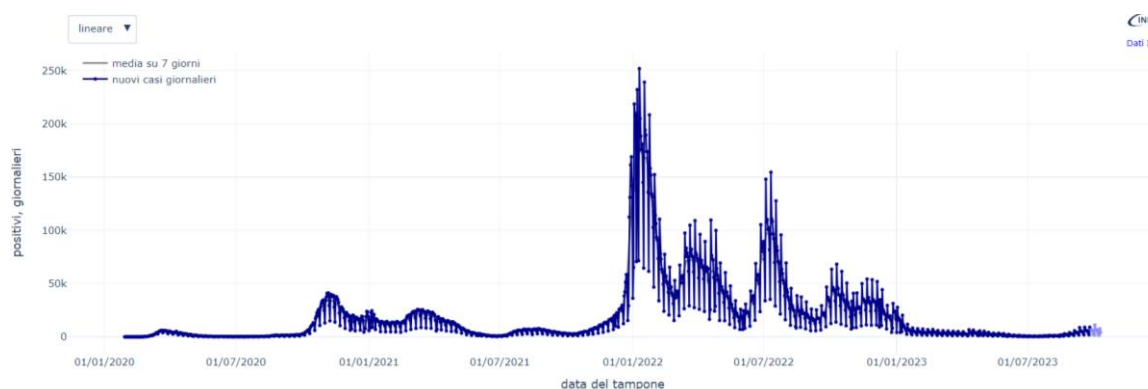
La maggior parte dei *cluster* BA.1 italiani datavano nell'autunno 2021, quelli della variante BA.2 principalmente tra la fine del 2021 e l'inizio del 2022, mentre l'origine stimata dei *cluster* BA.5 copriva un ampio arco temporale, compreso tra ottobre 2021 e maggio 2022.

Per queste varianti sono stati osservati valori di  $R_e$  superiori rispetto a quelli stimati per le varianti precedenti, a conferma della maggiore trasmissibilità di Omicron. Nonostante i dati di letteratura indichino una maggiore trasmissibilità di BA.2 rispetto a BA.1,<sup>23, 63</sup> i valori da noi stimati risultano comparabili. Le nostre stime risultano simili a quelle ufficiali (**Figura 33**, <https://covid19.infn.it/>), con l'eccezione del picco a gennaio 2022, corrispondente alla simultanea circolazione delle varianti Delta e Omicron.



**Figura 33. Tasso di contagio  $R_t$  in Italia. Fonte: <https://covid19.infn.it/>.**

Nonostante il valore medio di  $R_e$  stimato al picco per la variante BA.5 risultava inferiore a quelli stimati per BA.1 e BA.2 (1.28 vs. 1.45 e 1.42, rispettivamente), si osservava una persistenza nel tempo di valori superiori all'1 sin dall'inizio dell'epidemia, che giustificerebbe il picco dei soggetti infetti osservato nell'ondata del 2022 (**Figura 34**).



**Figura 34. Numero di tamponi positivi giornalieri in Italia. Fonte: <https://covid19.infn.it/>.**

Similmente a quanto precedentemente osservato, per tutte le varianti considerate, le nostre analisi mostrano una buona corrispondenza temporale tra l'andamento del numero di infezioni stimato dal grafico *Skyline* e la stima  $R_e$  mediante *birth-death Skyline*.

Gli *evolutionary rate* stimati avevano valori intorno a  $4 \times 10^{-4}$  sostituzioni/sito/anno per tutte le varianti in studio con l'unica eccezione della variante Delta, il cui valore è quasi doppio (**Tabella 35**); questo potrebbe essere dovuto ad una maggiore permanenza di questa variante di quasi un anno e alla generazione di numerosi *lineages* derivati.

**Tabella 35. Tassi evolutivi e relativi intervalli di confidenza delle varianti virali studiate.**

varianti	<i>evolutionary rate (s/s/y)</i>	95% HPD* lower	95%HPD upper
20E.EU1	$3.80 \times 10^{-4}$	$3.35 \times 10^{-4}$	$4.39 \times 10^{-4}$
Alpha	$4.87 \times 10^{-4}$	$4.18 \times 10^{-4}$	$5.54 \times 10^{-4}$
Delta	$7.57 \times 10^{-4}$	$6.65 \times 10^{-4}$	$8.48 \times 10^{-4}$
Omicron	BA.1	$3.76 \times 10^{-4}$	$5.98 \times 10^{-4}$
	BA.2	$2.70 \times 10^{-4}$	$5.33 \times 10^{-4}$
	BA.5	$3.72 \times 10^{-4}$	$5.44 \times 10^{-4}$

\*HPD: *highest posterior density*.

Nel complesso i nostri dati ci hanno permesso un'accurata descrizione delle dinamiche epidemiologiche nel nostro Paese in un vasto periodo, compreso tra agosto 2020 e dicembre 2022, che ha visto un susseguirsi di varianti differenti.

Sebbene questo lavoro sia il frutto della collaborazione di numerosi centri clinici dislocati su tutto il territorio nazionale, alcune regioni risultano meno rappresentate comportando un potenziale *bias* di campionamento che abbiamo cercato di superare accorpendo le regioni in macro-aree.

Un secondo limite è dovuto alla rapidità di circolazione del virus; una volta che una variante si è stabilita in più Paesi, la sua continua diffusione internazionale diventa inevitabile. Nel caso di varianti emergenti, in particolare quelle con maggiore virulenza e immunità decrescente, le misure per controllare o mediare gli effetti della trasmissione del virus dovrebbero essere intraprese ovunque. Inoltre, le restrizioni ai viaggi, soprattutto quelle mirate, sono spesso attuate dopo che il virus ha già di fatto raggiunto altri Paesi,<sup>64, 65</sup> potenzialmente prima della loro identificazione tramite sorveglianza genomica.

## 8. BIBLIOGRAFIA

1. Khan M, Adil SF, Alkhatlan HZ, et al. COVID-19: A Global Challenge with Old History, Epidemiology and Progress So Far. *Molecules (Basel, Switzerland)* 2020; **26**(1).
2. Ochani R, Asad A, Yasmin F, et al. COVID-19 pandemic: from origins to outcomes. A comprehensive review of viral pathogenesis, clinical manifestations, diagnostic evaluation, and management. *Le infezioni in medicina* 2021; **29**(1): 20-36.
3. King AMQ, Lefkowitz EJ, Mushegian AR, et al. Changes to taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2018). *Archives of virology* 2018; **163**(9): 2601-31.
4. Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020; **579**(7798): 270-3.
5. Paraskevis D, Kostaki EG, Magiorkinis G, Panayiotakopoulos G, Sourvinos G, Tsiodras S. Full-genome evolutionary analysis of the novel corona virus (2019-nCoV) rejects the hypothesis of emergence as a result of a recent recombination event. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* 2020; **79**: 104212.
6. Harrison AG, Lin T, Wang P. Mechanisms of SARS-CoV-2 Transmission and Pathogenesis. *Trends in immunology* 2020; **41**(12): 1100-15.
7. Lam TT, Jia N, Zhang YW, et al. Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature* 2020; **583**(7815): 282-5.
8. Zhang T, Wu Q, Zhang Z. Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak. *Current biology : CB* 2020; **30**(8): 1578.
9. Pizzato M, Baraldi C, Boscato Sopotto G, et al. SARS-CoV-2 and the Host Cell: A Tale of Interactions %U <https://www.frontiersin.org/articles/10.3389/fviro.2021.815388>. *Frontiers in Virology* 2022; **1**: %7 %8 2022-January-12 %9 Review %# %! SARS-CoV-2 interactions %\* %<.
10. Redondo N, Zaldívar-López S, Garrido JJ, Montoya M. SARS-CoV-2 Accessory Proteins in Viral Pathogenesis: Knowns and Unknowns. *Frontiers in immunology* 2021; **12**: 708264.
11. Scudellari M. How the coronavirus infects cells - and why Delta is so dangerous. *Nature* 2021; **595**(7869): 640-4.
12. Rambaut A, Holmes EC, O'Toole Á, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature microbiology* 2020; **5**(11): 1403-7.

13. Mehta P, Ravi V, Devi P, et al. Mutational dynamics across VOCs in International travellers and Community transmission underscores importance of Spike-ACE2 interaction. *Microbiological research* 2022; **262**: 127099.
14. Hodcroft EB, Zuber M, Nadeau S, et al. Spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *Nature* 2021; **595**(7869): 707-12.
15. Wang P, Nair MS, Liu L, et al. Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature* 2021; **593**(7857): 130-5.
16. Frampton D, Rampling T, Cross A, et al. Genomic characteristics and clinical effect of the emergent SARS-CoV-2 B.1.1.7 lineage in London, UK: a whole-genome sequencing and hospital-based cohort study. *The Lancet Infectious diseases* 2021; **21**(9): 1246-56.
17. Mukherjee R, Satardekar R. Why are some coronavirus variants more infectious? *Journal of biosciences* 2021; **46**(4).
18. Choi JY, Smith DM. SARS-CoV-2 Variants of Concern. *Yonsei medical journal* 2021; **62**(11): 961-8.
19. Dejnirattisai W, Shaw RH, Supasa P, et al. Reduced neutralisation of SARS-CoV-2 omicron B.1.1.529 variant by post-immunisation serum. *Lancet (London, England)* 2022; **399**(10321): 234-6.
20. Planas D, Veyer D, Baidaliuk A, et al. Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature* 2021; **596**(7871): 276-80.
21. Yamasoba D, Kimura I, Nasser H, et al. Virological characteristics of the SARS-CoV-2 Omicron BA.2 spike. *Cell* 2022; **185**(12): 2103-15 e19.
22. Focosi D, Maggi F. Neutralising antibody escape of SARS-CoV-2 spike protein: Risk assessment for antibody-based Covid-19 therapeutics and vaccines. *Reviews in medical virology* 2021; **31**(6): e2231.
23. 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.
24. Chi X, Yan R, Zhang J, et al. A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. *Science (New York, NY)* 2020; **369**(6504): 650-5.
25. Ou J, Lan W, Wu X, et al. Tracking SARS-CoV-2 Omicron diverse spike gene mutations identifies multiple inter-variant recombination events. *Signal transduction and targeted therapy* 2022; **7**(1): 138.
26. 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): 151-62.

27. Adam DC, Wu P, Wong JY, et al. Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. *Nature medicine* 2020; **26**(11): 1714-9.
28. Billah MA, Miah MM, Khan MN. Reproductive number of coronavirus: A systematic review and meta-analysis based on global level evidence. *PloS one* 2020; **15**(11): e0242128.
29. Salzberger B, Buder F, Lampl B, et al. Epidemiology of SARS-CoV-2. *Infection* 2021; **49**(2): 233-9.
30. Amendola A, Bianchi S, Gori M, et al. Evidence of SARS-CoV-2 RNA in an Oropharyngeal Swab Specimen, Milan, Italy, Early December 2019. *Emerging infectious diseases* 2021; **27**(2): 648-50.
31. Apolone G, Montomoli E, Manenti A, et al. Unexpected detection of SARS-CoV-2 antibodies in the pre-pandemic period in Italy. *Tumori* 2021; **107**(5): 446-51.
32. Lai A, Bergna A, Caucci S, et al. Molecular Tracing of SARS-CoV-2 in Italy in the First Three Months of the Epidemic. *Viruses* 2020; **12**(8).
33. Lai A, Bergna A, Menzo S, et al. Circulating SARS-CoV-2 variants in Italy, October 2020-March 2021. *Virology journal* 2021; **18**(1): 168.
34. Davies NG, Jarvis CI, Edmunds WJ, Jewell NP, Diaz-Ordaz K, Keogh RH. Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7. *Nature* 2021; **593**(7858): 270-4.
35. Twohig KA, Nyberg T, Zaidi A, et al. Hospital admission and emergency care attendance risk for SARS-CoV-2 delta (B.1.617.2) compared with alpha (B.1.1.7) variants of concern: a cohort study. *The Lancet Infectious diseases* 2022; **22**(1): 35-42.
36. 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).
37. 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.
38. Lavezzo E, Franchin E, Ciavarella C, et al. Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo'. *Nature* 2020; **584**(7821): 425-9.
39. Grubaugh ND, Ladner JT, Lemey P, et al. Tracking virus outbreaks in the twenty-first century. *Nature microbiology* 2019; **4**(1): 10-9.
40. Rife BD, Mavian C, Chen X, et al. Phylodynamic applications in 21(st) century global infectious disease research. *Global health research and policy* 2017; **2**: 13.

41. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus evolution* 2016; **2**(1): vew007.
42. 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): 1530-4.
43. Ragonnet-Cronin M, Hodcroft E, Hué S, et al. Automated analysis of phylogenetic clusters. *BMC bioinformatics* 2013; **14**: 317.
44. 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.
45. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic biology* 2018; **67**(5): 901-4.
46. Cereda D, Manica M, Tirani M, et al. The early phase of the COVID-19 epidemic in Lombardy, Italy. *Epidemics* 2021; **37**: 100528.
47. Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers during epidemics. *American journal of epidemiology* 2013; **178**(9): 1505-12.
48. 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): 228-33.
49. 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): 1199-207.
50. Alicandro G, Remuzzi G, La Vecchia C. COVID-19 pandemic and total mortality in the first six months of 2020 in Italy. *La Medicina del lavoro* 2020; **111**(5): 351-3.
51. Remuzzi A, Remuzzi G. COVID-19 and Italy: what next? *Lancet (London, England)* 2020; **395**(10231): 1225-8.
52. Uselli M. The Lombardy region of Italy launches the first investigative COVID-19 commission. *Lancet (London, England)* 2020; **396**(10262): e86-e7.
53. Manathunga SS, Abeyagunawardena IA, Dharmaratne SD. A comparison of transmissibility of SARS-CoV-2 variants of concern. *Virology journal* 2023; **20**(1): 59.
54. Suzuki R, Yamasoba D, Kimura I, et al. Attenuated fusogenicity and pathogenicity of SARS-CoV-2 Omicron variant. *Nature* 2022; **603**(7902): 700-5.

55. Carabelli AM, Peacock TP, Thorne LG, et al. SARS-CoV-2 variant biology: immune escape, transmission and fitness. *Nature reviews Microbiology* 2023; **21**(3): 162-77.
56. Roemer C, Sheward DJ, Hisner R, et al. SARS-CoV-2 evolution in the Omicron era. *Nature microbiology* 2023.
57. 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.
58. 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).
59. Kim YC, Kim B, Son NH, et al. Vaccine Effect on Household Transmission of Omicron and Delta SARS-CoV-2 Variants. *Journal of Korean medical science* 2023; **38**(1): e9.
60. López-Muñoz I, Torrella A, Pérez-Quílez O, et al. SARS-CoV-2 Secondary Attack Rates in Vaccinated and Unvaccinated Household Contacts during Replacement of Delta with Omicron Variant, Spain. *Emerging infectious diseases* 2022; **28**(10): 1999-2008.
61. 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): 3277-90 e16.
62. 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.
63. 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.
64. 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**.
65. 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): 1367-9.