



**UNIVERSITÀ POLITECNICA DELLE MARCHE**

**Dipartimento di Scienze della Vita e dell'Ambiente**

**Corso di Laurea Magistrale in Biologia Molecolare e Applicata (LM-6)**

**ANALISI FILOGENETICA E CARATTERIZZAZIONE  
MOLECOLARE DELLE VARIANTI DI SARS-CoV-2 CIRCOLANTI  
NEL 2023 NELLA REGIONE MARCHE**

**PHYLOGENETIC ANALYSIS AND MOLECULAR  
CHARACTERIZATION OF SARS-CoV-2 VARIANTS  
CIRCULATING IN 2023 IN MARCHE REGION**

Tesi di Laurea Magistrale di:  
Paparoni Silvia

Relatore Chiar.mo. Prof  
Menzo Stefano

Sessione Straordinaria  
A.A. 2022/2023

# Sommario

1. INTRODUZIONE .....	2
1.1 Storia e Tassonomia dei Coronavirus.....	2
1.1.1 Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV).....	5
1.1.2 Middle East Respiratory Syndrome Coronavirus (MERS-CoV).....	6
1.1.3 Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2).....	7
1.2 Genoma Virale e Caratteristiche Strutturali di SARS-CoV-2.....	9
1.2.1 ORF1ab .....	10
1.2.2 Proteina Spike (S) .....	17
1.2.3 Proteina orf3a .....	20
1.2.4 Proteina Envelope (E).....	21
1.2.5 Proteina Membrane (M).....	22
1.2.6 Proteina orf6.....	22
1.2.7 Proteina orf7a .....	23
1.2.8 Proteina orf7b.....	23
1.2.9 Proteina orf8.....	24
1.2.10 Proteina Nucleocapsid (N) .....	24
1.2.11 Proteina orf10 .....	25
1.3 Il Ciclo Replicativo di SARS-CoV-2 .....	26
1.3.1 Attacco al recettore ed entrata nella cellula ospite.....	27
1.3.2 Traduzione delle proteine non strutturali .....	30
1.3.3 Replicazione e trascrizione del genoma virale.....	31
1.3.4 Traduzione delle proteine strutturali e accessorie e assemblaggio del virione	34
1.4 Tropismo di SARS-CoV-2 .....	35
1.5 Patogenesi del COVID-19 .....	36
1.5.1 Attivazione della risposta immunitaria innata dell'ospite verso il SARS-CoV-2	37
1.5.2 Fattori di rischio e complicazioni cliniche del COVID-19 .....	40

1.6	Modalità di trasmissione del SARS-CoV-2 .....	42
1.7	Diagnosi .....	44
1.8	Terapia e Vaccini .....	47
1.9	Varianti virali di SARS-CoV-2 .....	51
1.9.1	Variants of Concern (VOC) .....	55
1.10	Epidemiologia di SARS-CoV-2 .....	59
2.	SCOPO DELLO STUDIO .....	62
3.	MATERIALI E METODI .....	63
3.1	Popolazione oggetto di studio .....	63
3.2	Sistemi diagnostici utilizzati e selezione dei campioni positivi .....	63
3.3	Estrazione automatizzata dell'RNA virale per la genotipizzazione .....	64
3.4	Retrotrascrizione dell'RNA virale per la genotipizzazione .....	65
3.5	Preparazione automatizzata delle librerie di template per il sequenziamento 66	
3.6	Next Generation Sequencing (NGS) .....	69
3.7	Analisi bioinformatica dei dati .....	71
4.	RISULTATI .....	73
4.1	Analisi delle distanze genetiche .....	76
4.2	Analisi filogenetica delle sequenze .....	81
4.3	Analisi del profilo mutazionale .....	83
4.4	Panoramica andamento epidemiologico di SARS-CoV-2 nelle Marche	87
5.	DISCUSSIONE E CONCLUSIONI .....	90
	Fonti Bibliografiche .....	96

# 1.INTRODUZIONE

## 1.1 Storia e Tassonomia dei Coronavirus

Il primo *coronavirus* ad essere scoperto, nel 1931, è stato l'*Avian Infectious Bronchitis Virus* (IBV), un *gammacoronavirus* che infettava le specie aviarie. Successivamente, nel 1966 e nel 1967, vennero descritti i primi *coronavirus* capaci di infettare l'uomo, il HCoV-229E e il HCoV-OC43.

Negli ultimi vent'anni, i *coronavirus* sono spesso stati oggetto di studio in quanto hanno dato origine a salti di specie, infettando l'uomo e determinando l'insorgenza di focolai epidemici di interesse sanitario, come nel caso del SARS-CoV nel 2002-2003, del MERS-CoV nel 2012 e dalla più recente pandemia dovuta al SARS-CoV-2 nel 2019 (Figura 1).<sup>1</sup>

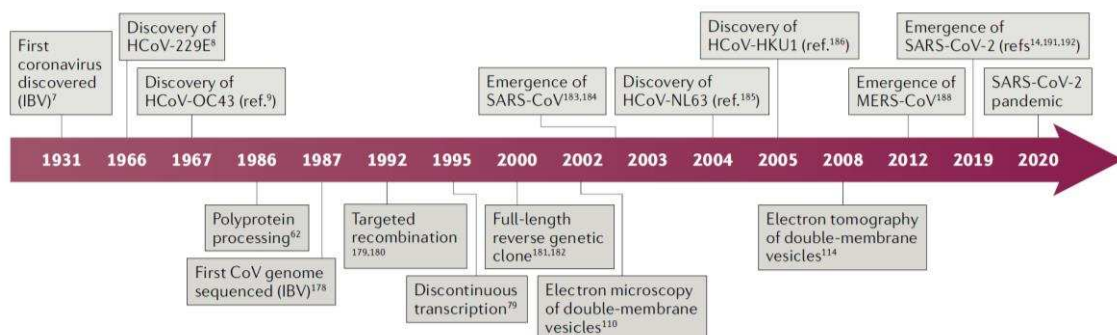
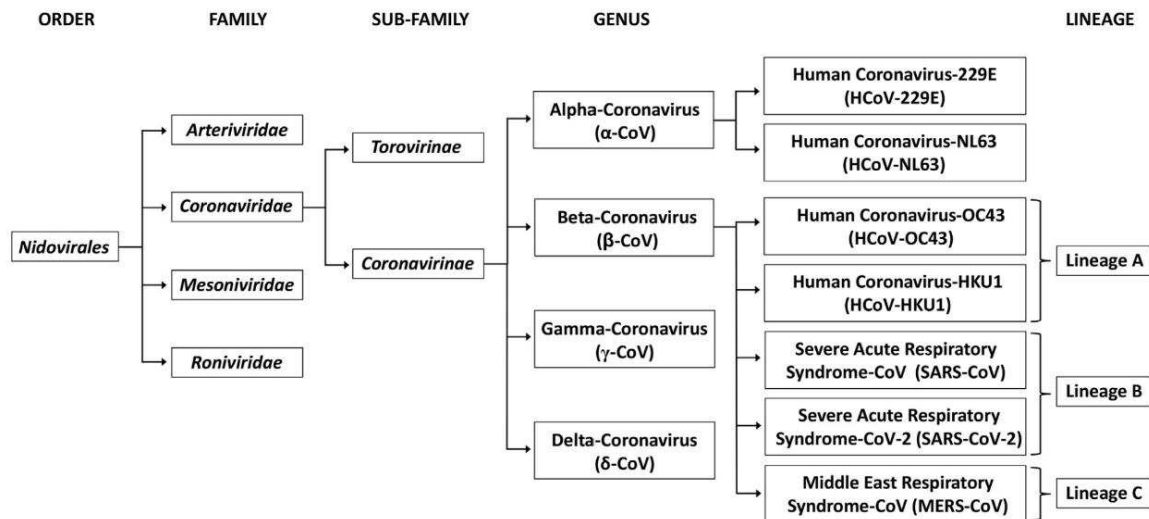


Figura 1 - Cronologia delle scoperte e delle ricerche sui coronavirus. Fonte: V'kovski P et al, 2020.<sup>1</sup>

I *coronavirus* (CoVs) appartengono alla sottofamiglia *Orthocoronavirinae*, famiglia *Coronaviridae*, sottordine *Coronavirineae*, ordine *Nidovirales*. I membri di questo ordine sono virus rivestiti dalle diverse morfologie, tutti caratterizzati da un genoma a RNA a singolo filamento lineare a polarità positiva, con il *cap* al 5' e il 3' poliadenilato.<sup>1,2</sup>

Sulla base di studi filogenetici condotti su regioni genomiche conservate di questi virus, sono stati distinti quattro generi nella sottofamiglia degli *Orthocoronavirinae*: gli *alphacoronavirus*, i *betacoronavirus*, i *gammacoronavirus* e i *deltacoronavirus*. Il genere *betacoronavirus* può essere ulteriormente sottoclassificato in cinque lignaggi: il *lineage A* (sottogenere *Embecovirus*) include HCoV-OC43 e HCoV-HKU1, il *lineage B* (sottogenere *Sarbecovirus*) include SARS-CoV e SARS-CoV-2, il *lineage C* (sottogenere *Merbecovirus*) contiene il *coronavirus* MERS-CoV, il *lineage D* (sottogenere *Nobecovirus*) e il *lineage E* (sottogenere *Hibecovirus*).<sup>3</sup> (Figura 2)



**Figura 2** - Rappresentazione schematica della tassonomia dei Coronaviridae secondo l'International Committee on Taxonomy of Viruses. Fonte: S. Muralidar, S.V. Ambi, S. Sekaran et al., 2020.<sup>13</sup>

Nei generi *alphacoronavirus* e *betacoronavirus* ritroviamo i *coronavirus* che infettano esclusivamente i mammiferi, essere umano compreso, mentre nei *gammacoronavirus* e *deltacoronavirus* ritroviamo virus che hanno un ampio spettro d'ospite, in particolare le specie aviarie.<sup>1</sup>

Fra gli *alphacoronavirus* che infettano l'uomo troviamo gli HCoV-229E e l'HCoV-NL63, mentre nei *betacoronavirus* vi sono il HCoV-HKU1, il HCoV-OC43, il *Middle East Respiratory Syndrome Coronavirus* (MERS-CoV), il *Severe Acute Respiratory Syndrome Coronavirus* (SARS-CoV) e il *Severe Acute Respiratory Syndrome Coronavirus 2* (SARS-CoV-2).

I *coronavirus* HCoV-229E, HCoV-OC43, HCoV-NL63 e HCoV-HKU1 sono generalmente responsabili di infezioni stagionali paucisintomatiche nei soggetti immunocompetenti con blandi sintomi respiratori, come ad esempio il comune raffreddore. Al contrario, SARS-CoV, MERS-CoV e il più recente SARS-CoV-2 si sono dimostrati virus altamente patogeni causando sindromi respiratorie anche di grave entità.<sup>1</sup>

Sulla base delle analisi svolte su un gran numero di sequenze, è noto che tutti i *coronavirus* umani sono di recente origine animale; in particolare, i membri del sottogenere *Sarbecovirus* tendono a subire eventi di ricombinazione che possono dar luogo a nuovi ceppi virali che potenzialmente potrebbero infettare anche l'essere umano.<sup>3</sup>

I pipistrelli, in molti casi, si sono rivelati gli ospiti di lignaggi ancestrali e principali serbatoi naturali di *alfacoronavirus* e *betacoronavirus*.<sup>4</sup> Più di 4.800 sequenze di *coronavirus* sono state riscontrate nei pipistrelli, ossia oltre il 30% di tutti i virus dei pipistrelli sequenziati. Infatti, il serbatoio animale dei HCoV-NL63, HCoV-229E, SARS-CoV, MERS-CoV e SARS-CoV-2 è rappresentato dai pipistrelli, mentre per HCoV-OC43 e HKU1 probabilmente sono i roditori.<sup>5</sup>

### 1.1.1 Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV)

Il SARS-CoV è stato il primo *coronavirus* a rappresentare un problema di sanità pubblica per l'uomo. Questo virus è emerso per la prima volta nel novembre 2002 nella provincia di Guangdong in Cina, determinando l'epidemia da *Severe Acute Respiratory Syndrome*.

I primi casi registrati furono in un Hotel di Hong Kong e in seguito a questi il virus diffuse in oltre 37 paesi nell'arco di nove mesi, portando a circa 8273 casi registrati, di cui 775 con esito fatale.<sup>6</sup>

Il SARS-CoV ha un tasso di mortalità del 9-10% e le manifestazioni cliniche più comuni hanno riguardato principalmente febbre, affaticamento, tosse secca, mialgia e diarrea. In molti casi si riscontrarono anche dispnea, ipossiemia e, negli esiti più gravi, la sindrome da distress respiratorio acuto (ARDS) e la sindrome da disfunzione multiorgano (MODS).<sup>7</sup>

A seguito dell'identificazione dell'agente eziologico di questa sindrome, ossia il SARS-CoV, vennero svolte delle analisi anticorpali anti-SARS-CoV per ricercare la possibile origine di questo *coronavirus*. Da queste analisi emerse la presenza di tali anticorpi negli zibetti della specie *Paguma larvata*, una specie animale frequentemente venduta nei mercati della zona. Successivamente emerse che i ceppi di SARS-CoV trovati negli zibetti del mercato erano stati trasmessi loro da altri animali. Nel 2005, uno studio portò alla scoperta dell'ospite naturale del SARS-CoV, vennero ritrovati *coronavirus* filogeneticamente correlati al SARS-CoV nei pipistrelli *horseshoe*, genere *Rhinolophus*, decretando il ruolo dello zibetto come ospite intermedio.

Molto probabilmente l'origine del SARS-CoV è da ricercarsi in eventi di ricombinazione che hanno interessato ceppi di *coronavirus* filologicamente

correlati in pipistrelli in queste zone. Si ipotizza che i frequenti eventi di ricombinazione che interessano i *coronavirus* e l'ampia diffusione dei pipistrelli possono aver contribuito al salto di specie da quest'ultimi agli zibetti da allevamento e, a seguito di ulteriori mutazioni, la trasmissione del ceppo di SARS-CoV all'uomo.<sup>4</sup>

### **1.1.2 Middle East Respiratory Syndrome Coronavirus (MERS-CoV)**

Il MERS-CoV è un *betacoronavirus* che venne identificato per la prima volta in Arabia Saudita nel 2012 a seguito di un'epidemia di MERS (*Middle East Respiratory Syndrome*) che interessò principalmente paesi medio orientali. Infatti, venne riscontrato in più di 27 paesi fra cui l'Arabia Saudita, gli Emirati Arabi Uniti, il Qatar, l'Austria, il Bangladesh, la Thailandia, l'Indonesia, il Regno Unito e gli USA. Ad oggi sono stati documentati circa 2500 casi globali di infezioni da MERS-CoV, di cui circa il 35% sono risultati fatali.

Le manifestazioni cliniche della MERS sono rappresentate in difficoltà respiratorie, tosse, febbre, sintomi gastrointestinali, in molti casi sfociando in polmonite e progredendo nella sindrome da distress respiratorio acuto. Le complicazioni riscontrate nei casi più gravi, risultate spesso in esiti fatali, riguardano l'insorgenza di insufficienza renale acuta e multiorgano.

Dalle analisi svolte emerse che il MERS-CoV originò nel 2011 in Africa centrale e orientale, trovando il suo serbatoio naturale in varie specie animali, quali pipistrelli, dromedari e nel riccio europeo.<sup>8</sup> Come per il SARS-CoV, si pensa che la genesi del MERS-CoV sia riconducibile a eventi di ricombinazione che hanno interessato *coronavirus* endemici in specie di pipistrelli, in particolare si è riscontrata una forte correlazione con i *coronavirus*



HKU4 e HKU5.<sup>7</sup> Questi eventi hanno permesso al virus di originare nuovi ceppi capaci di infettare una nuova specie animale, in questo caso il dromedario. Successive mutazioni e adattamenti nell'ospite intermedio hanno portato il MERS-CoV ad avere la capacità di attuare un nuovo salto di specie e trasmettersi all'uomo.<sup>8</sup>

### **1.1.3 Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)**

Nel dicembre 2019, nella città di Wuhan, nella provincia di Hubei in Cina, vennero riportati i primi casi di una malattia respiratoria ad eziologia ignota non riconducibile ad alcun patogeno conosciuto. Da indagini successive emerse che i primi casi di infezione furono riconducibili a un'esposizione avvenuta nel *Seafood Market* nella città di Wuhan, conosciuto per la vendita di varie specie di animali vivi. In breve tempo, l'agente responsabile del focolaio venne individuato in un *betacoronavirus*, inizialmente denominato 2019-nCov. Successivamente, il Gruppo di Studio sui Coronavirus (CSG) del *International Committee on Taxonomy of Viruses* assegnò al patogeno emergente il nome *Severe Acute Respiratory Syndrome Coronavirus 2* (SARS-CoV-2), identificandolo come l'agente eziologico responsabile dell'epidemia di COVID-19 (*CoronaVirus Disease-19*).

Il nuovo *coronavirus* diventò presto un'emergenza sanitaria a livello globale, infatti, nel marzo 2020 l'Organizzazione Mondiale della Sanità (OMS) dichiarò lo stato di pandemia.<sup>9</sup>

Da analisi filogenetiche condotte sulla sequenza genomica completa di SARS-CoV-2 emerse un'identità di sequenza del 79,5% con il genoma di SARS-CoV, maggiore rispetto alla sequenza di MERS-CoV (circa il 50% d'identità). Come

per SARS-CoV e MERS-CoV, anche per SARS-CoV-2 si è pensato che il suo *reservoir* animale risieda nei pipistrelli e che è stato diffuso all'uomo tramite un ospite intermedio.

I risultati di varie analisi dimostrarono la forte relazione fra la sequenza genomica di SARS-CoV-2 e quella del *coronavirus* RatG13 che infetta i pipistrelli *Rhinolophus affinis*, con un'identità nucleotidica superiore al 95%. Successivamente si è ricercato un possibile ospite intermedio; si prese in considerazione il pangolino in quanto vennero isolate sequenze di *coronavirus* da pangolini malesi con un'elevata identità di sequenza con il SARS-CoV-2 (91,02%) e con il RatG13 (90,55%).<sup>10,11</sup>

## 1.2 Genoma Virale e Caratteristiche Strutturali di SARS-CoV-2

Il SARS-CoV-2, similmente ad altri *coronavirus*, è un virus dotato di *envelope* il cui genoma è costituito da RNA a singolo filamento positivo non segmentato. I *Coronavirus* si presentano come virus pleomorfi delle dimensioni di circa 80-120 nm. L'*envelope* presenta sulla sua superficie particolari glicoproteine, le *spikes* (S), che conferiscono a questi virus il caratteristico aspetto a corona, da cui il nome della famiglia. Altre glicoproteine costituenti il rivestimento virale sono la proteina M (*Membrane*) e la proteina E (*Envelope*). All'interno dell'*envelope* è presente l'RNA virale associato alla proteina N (*Nucleocapsid*) a formare il nucleocapside.<sup>12</sup> (Figura 3)

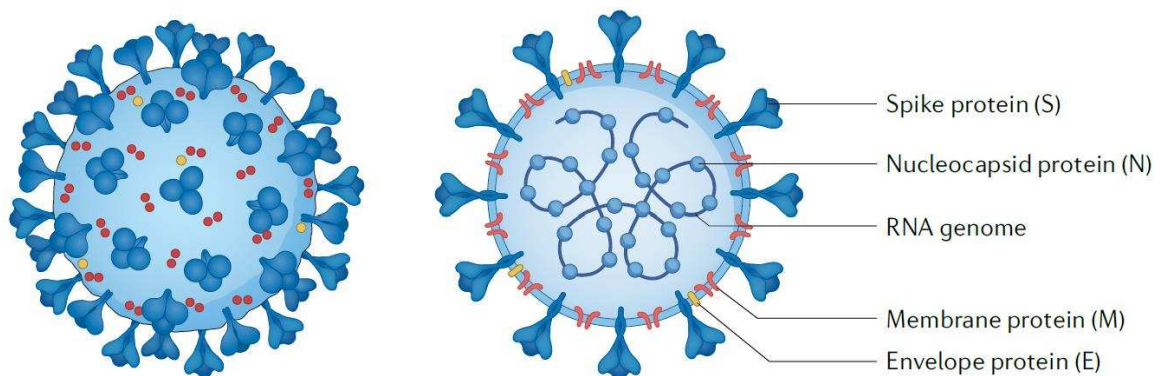


Figura 3 - Rappresentazione schematica del virione di SARS-CoV-2. Fonte: M. M. Lamers e Bart L. Haagmans, 2022.<sup>27</sup>

Il genoma di SARS-CoV-2, costituito da 29,891 nucleotidi e codificante per 9860 aminoacidi, è fra i genomi virali a RNA più grandi conosciuti e, come altri *coronavirus*, presenta il *cap* al 5' e la coda di poliadenilazione al 3'.<sup>13</sup> Sono

presenti due regioni non tradotte, 5'-UTR e 3'-UTR, che fiancheggiano 11 *Open Reading Frames* (ORFs): ORF1ab, ORF2 (*spike*), ORF3a, ORF4 (*envelope*), ORF5 (*membrane*), ORF6, ORF7a, ORF7b, ORF8, ORF9 (nucleocapside) e ORF10 (Figura 4).

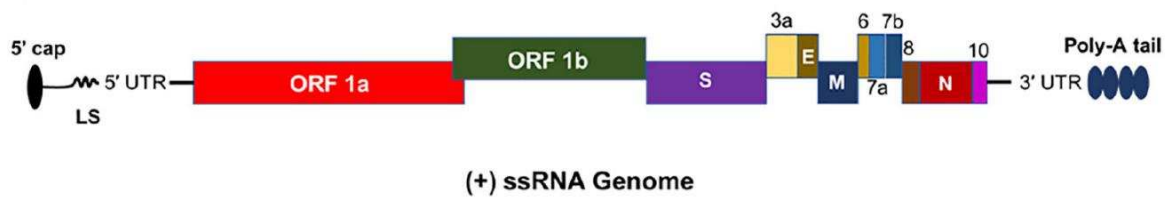
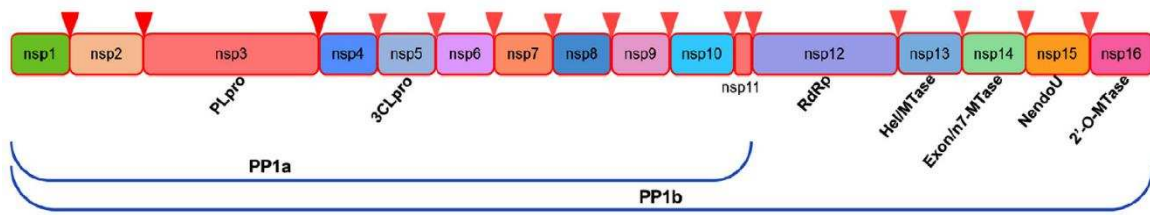


Figura 4 - Rappresentazione schematica del genoma di SARS-CoV-2. Fonte: Naqvi, 2020<sup>19</sup>

I primi due terzi del genoma sono occupati dalla ORF1ab, che codifica 16 proteine non strutturali (nsp 1-16) indispensabili per la replicazione virale. Nella restante parte si trovano i geni strutturali quali S (*Spike*), E (*Envelope*), M (*Membrane*), N (*Nucleocapsid*) e altre ORF accessorie.<sup>14</sup>

### 1.2.1 ORF1ab

ORF1ab occupa i primi due-terzi del genoma nella porzione al 5' e ha una dimensione di circa 22kb. Questo gene comprende due ORF sovrapposte per un nucleotide, l'ORF1a e l'ORF1b, che codificano per due poliproteine, la pp1a di circa 500kDa e la pp1ab di circa 800kDa. Queste poliproteine vengono scisse da proteasi virali specifiche, ossia la PL<sup>pro</sup> e la 3CL<sup>pro</sup>, generando 16 proteine non strutturali (nsp 1-16) importanti per il processo di replicazione e trascrizione virale (Figura 5).



*Figura 5 - Rappresentazione schematica delle proteine non strutturali codificate dalla ORF1ab. Fonte: N. Kirtipal et al, 2020<sup>12</sup>*

Le poliproteine pp1a e pp1ab, durante il processo di maturazione, subiscono modifiche post-traduzionali, perciò non risultano rilevabili durante l'infezione.<sup>3</sup>

### 1.2.1.1 Nsp1 (*Nonstructural Protein 1*)

La nsp1 è la prima proteina ad essere codificata dal gene ORF1ab. È una piccola proteina di circa 179 aminoacidi e, diversamente da altre nsp, non ha un ruolo nella replicazione virale, bensì nella patogenesi, infatti è implicata nella soppressione della risposta immunitaria innata dell'ospite, inibendo l'espressione dell'Interferone di tipo 1 (IFN1) e altre vie di segnalazione antivirale nella cellula.<sup>3,15</sup> Un'altra importante funzione di nsp1 è quella di interferire con la traduzione della cellula ospite legandosi alla subunità 40S del ribosoma e promuovendo la degradazione dell'mRNA cellulare.<sup>15</sup>

### 1.2.1.2 Nsp2 (*Nonstructural Protein 2*)

La nsp2 è una proteina molto conservata di cui non sono ancora ben comprese le funzioni. Questa proteina interagisce con due proteine della cellula ospite, la proibitina 1 (PHB1) e la proibitina 2 (PHB2), che sono implicate in vari processi cellulari, quali la replicazione cellulare, la migrazione cellulare, la

differenziazione cellulare, l'apoptosi e la biogenesi mitocondriale. L'interazione di nsp2 con le PHB1 e PHB2 suggerisce un suo ruolo nell'interferire con questi processi cellulari essenziali.<sup>15</sup>

#### **1.2.1.3 Nsp3 (*Nonstructural Protein 3*)**

La nsp3 è una glicoproteina transmembrana di 215 kDa. Comprende due domini transmembrana, TM1 e TM2, e otto domini molto conservati: i domini *ubiquitin-like 1* (Ubl1) e *ubiquitin-like 2* (Ubl2), un dominio ricco in glutammato anche noto come “regione ipervariabile”, il dominio *papain-like protease* (PL<sup>pro</sup>), l'ectodominio nsp3 (3Ecto) anche noto come “*zinc-finger domain*”, il macrodominio X, CoV-Y e Y1.<sup>3</sup>

Nsp3, con il suo dominio PL<sup>pro</sup>, svolge un importante ruolo nella scissione delle poliproteine in nsp1, nsp2 e nsp3. Un'altra funzione di nsp3 è quella di interagire con altre nsp per formare il complesso di replicazione e trascrizione (RTC), inoltre, altera le modifiche post-traduzionali delle proteine cellulari così da interferire con la risposta immunitaria innata dell'ospite e favorire i processi di infezione del virus.<sup>3,15</sup>

#### **1.2.1.4 Nsp4 (*Nonstructural Protein 4*)**

La nsp4 ha un importante ruolo nel riarrangiamento delle membrane e nella replicazione virale. Infatti, interagendo con le proteine nsp3 e nsp6, porta alla formazione delle *double-membrane vesicles* (DMVs), strutture importanti per la replicazione virale.<sup>15</sup>

#### **1.2.1.5 Nsp5 (*Nonstructural Protein 5*)**

La nsp5, chiamata anche 3CL<sup>pro</sup> o M<sup>pro</sup>, è la principale proteasi di SARS-CoV-2. La sua funzione è quella di promuovere la maturazione di 11 delle 16

poliproteine virali, andando ad agire su 11 siti di taglio. Inoltre, agisce come catalizzatore nella maturazione delle nsp4 e nsp6.<sup>15</sup>

In alcuni *coronavirus*, come nel *Porcine Deltacoronavirus* (PDCoV), è noto come nsp5 interferisca con l'INF di tipo I e la sua via di segnalazione.<sup>3</sup>

#### **1.2.1.6 Nsp6 (*Nonstructural Protein 6*)**

La nsp6 è una proteina transmembrana che, insieme a nsp3 e nsp4, è coinvolta nella formazione della membrana e nell'organizzazione delle DMV. Inoltre, interferisce con la formazione dell'autofagosoma, favorendo l'infezione virale.<sup>3</sup>

#### **1.2.1.7 Nsp7 (*Nonstructural Protein 7*) e Nsp8 (*Nonstructural Protein 8*)**

La nsp7 insieme alla nsp8 sono due cofattori dell'RNA polimerasi RNA-dipendente (RpRd) nsp12. Il loro ruolo risulta essenziale per la replicazione virale, in quanto si ipotizza sia quello di stabilizzare regioni di legame all'RNA della nsp12.<sup>3</sup> Per svolgere questa funzione, nsp8 presenta delle estensioni a  $\alpha$ -elica nelle regioni N-terminali della molecola che sono molto conservate. Queste regioni si localizzano in prossimità del canale di uscita dell'RNA della RdRp e presentano dei residui carichi positivamente che hanno la funzione di interagire e far scivolare l'ibrido di RNA a doppio filamento neosintetizzato, prevenendo una dissociazione prematura dall'nsp12 e aumentando così la processività.<sup>16</sup>

Nsp8, inoltre, agisce da RNA primasi sintetizzando un breve primer che servirà alla RpRd per la polimerizzazione.<sup>3</sup>

#### **1.2.1.8 Nsp9 (*Nonstructural Protein 9*)**

La nsp9 è una proteina legante l'RNA/ssDNA. Questa proteina, interagendo con altre proteine nel complesso RTC, favorisce un'efficiente replicazione e trascrizione del virus.<sup>15</sup>

#### **1.2.1.9 Nsp10 (*Nonstructural Protein 10*)**

La nsp10 svolge un ruolo importante nella sintesi dell'RNA virale e nella maturazione delle poliproteine insieme alla proteasi nsp5. Inoltre, legandosi alle proteine nsp14 e nsp16, partecipa alla formazione del complesso RTC. Il legame nsp10-nsp14 favorisce l'attività esonucleasica di quest'ultimo, mentre il legame nsp10-nsp16 agisce come cofattore per l'attività 2'O-MetilTransferasica della nsp16, regolando così l'aggiunta del 5'cap all'RNA virale.<sup>15</sup>

#### **1.2.1.10 Nsp11 (*Nonstructural Protein 11*)**

La funzione di nsp11 è tutt'ora sconosciuta. Si ipotizza un suo possibile ruolo nella produzione dell'elicasi.<sup>3</sup>

#### **1.2.1.11 Nsp12 (*Nonstructural Protein 12*)**

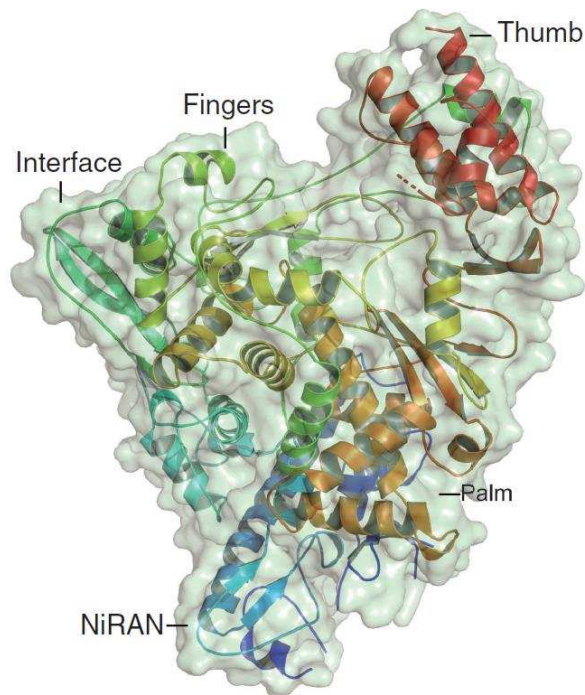
La nsp12 è una proteina di 102kDa molto conservata nei *coronavirus*, consiste in una RNA polimerasi RNA-dipendente (RdRp) e presenta un ruolo essenziale per la formazione del complesso RTC.<sup>3</sup>

L'RdRp presenta diversi domini nella sua struttura: il dominio N-terminale *nidovirus RdRp-associated nucleotidyltransferase* (NiRAN), il dominio della polimerasi C-terminale, che catalizza la formazione dei legami fosfodiesterici tra i ribonucleotidi, e il dominio *interface* che li collega entrambi.



Il dominio polimerasico presenta con una conformazione a “mano destra” dove sono riconoscibili i tre sottodomini *fingers*, *palm* e *thumb* (Figura 6).

Nel dominio N-terminale la proteina assume una struttura secondaria  $\beta$ -*hairpin* che si inserisce fra il dominio NiRAN e il sottodominio *palm* andando a stabilizzare la struttura complessiva della molecola. <sup>16</sup>



**Figura 6** - Rappresentazione della struttura della *nsp12* di SARS-CoV-2. Sono riportati il dominio N-terminal RNA-dependent RNA polymerase associated (NiRAN), il dominio interface e il dominio C-terminale della polimerasi costituito dai sottodomini *fingers*, *palm* e *thumb*. Fonte: Jiahui Wu et al, 2023. <sup>16</sup>

Nsp12 dà inizio alla polimerizzazione interagendo con l'RNA attraverso i sottodomini *thumb* e *fingers*. Il sito attivo della RdRp si compone di sette motivi molto conservati (A-G) che presentano molti residui carichi positivamente, funzionali per il direzionamento dell'RNA attraverso la cavità polimerasica. Nel sottodominio *palm* troviamo cinque di questi motivi (A-E) di cui il motivo C interagisce con l'estremità 3' dell'RNA, mentre i restanti

sottodomini F e G, sono situati nel sottodominio *fingers* e coadiuvano il posizionamento dell'RNA *template*.<sup>10</sup>

La formazione del complesso nsp12-nsp7-nsp8 migliora la processività della polimerasi andando ad abbassare il tasso di dissociazione della nsp12 dall'RNA in sintesi. Nsp7-nsp8 si legano in corrispondenza del sottodominio *thumb*, con una porzione dell'nsp8 legante il dominio *interface*, in prossimità del sottodominio *fingers* e del canale di legame per l'RNA.

In virtù della sua importanza e della sua selettività rispetto alla cellula infettata, l'RdRp è molto studiata come target per farmaci antivirali.<sup>16</sup>

#### **1.2.1.12 Nsp13 (*Nonstructural Protein 13*)**

La nsp13 è una proteina di 66,5 kDa caratterizzata dalla presenza di un dominio legante lo zinco all'N-terminale e dal dominio dell'elicasi al C-terminale.<sup>3</sup>

L'attività di elicasi ha come ruolo quello di svolgere la molecola di RNA a doppio filamento e il legame con nsp12 può favorire questa attività. Un'altra funzione di nsp13 è quella 5'-trifosfatasi, andando quindi ad aggiungere il *cap* al 5' dell'RNA virale, necessario per il riconoscimento traduzionale, per lo splicing e per l'esportazione nucleare dell'RNA.<sup>17</sup>

#### **1.2.1.13 Nsp14 (*Nonstructural Protein 14*)**

La nsp14 è una proteina di 60 kDa dalla duplice funzione: presenta l'attività esoribonucleasica (ExoN) 3'-5' all'N-terminale e l'attività della *S-adenosylmethionine-dependent guanine-N7-methyltransferase* (MTasi) al C-terminale. L'ExoN svolge la funzione di *proofreading*, ossia è coinvolta nella riparazione e del nuovo genoma virale. L'attività di MTasi è importante per il processo di *capping* al 5' dell'RNA virale, che serve a proteggere la molecola dalla degradazione esonucleasica. Infatti, il *cap* al 5' dell'RNA è una

componente essenziale per far sì che il genoma del virus non venga riconosciuto dai meccanismi di difesa della cellula ospite, oltreché per favorire la sua stabilità e la traduzione proteica.<sup>3,17</sup>

#### **1.2.1.14 Nsp15 (*Nonstructural Protein 15*)**

La nsp15 è un'endoribonucleasi virale  $Mn^{2+}$ -dipendente. Essa taglia in modo specifico l'RNA a livello delle poliuridine all'estremità 3' portando alla formazione di prodotti fosfodiesterici ciclici 2'-3'. L'attività endoribonucleasica non risulta essenziale per il processo di sintesi e replicazione dell'acido nucleico virale, tuttavia svolge un ruolo importante nel mediare l'evasione del riconoscimento da parte della cellula ospite del dsRNA del virus.<sup>15</sup>

#### **1.2.1.15 Nsp16 (*Nonstructural Protein 16*)**

La nsp16 è una 2'-O-Metiltrasferasi che, complessandosi con la proteina nsp10, ha un importante ruolo nel processo di *capping* dell'RNA virale.

Il SARS-CoV-2 presenta due metiltrasferasi S-adenosilmetionina-dipendenti, ossia la già citata nsp14 N7-metiltransferasi e la nsp16 2'-O-metiltransferasi, che metilano rispettivamente la guanosina nel sito N7 e il 2'O del ribosio.<sup>3</sup>

### **1.2.2 Proteina Spike (S)**

La proteina che più caratterizza la famiglia dei *Coronaviridae* è la proteina *Spike* (S). La *Spike* è una glicoproteina che si localizza in numerose copie sul pericapside della particella virale conferendo così l'aspetto a "corona" di questi

virus. Questa proteina svolge un importante ruolo nell'ingresso del virus nella cellula ospite, infatti media l'attacco e la fusione della particella virale con la membrana cellulare. Il recettore a cui si lega la proteina *Spike* è ACE2 (*Angiotensin Converting Enzyme 2*).

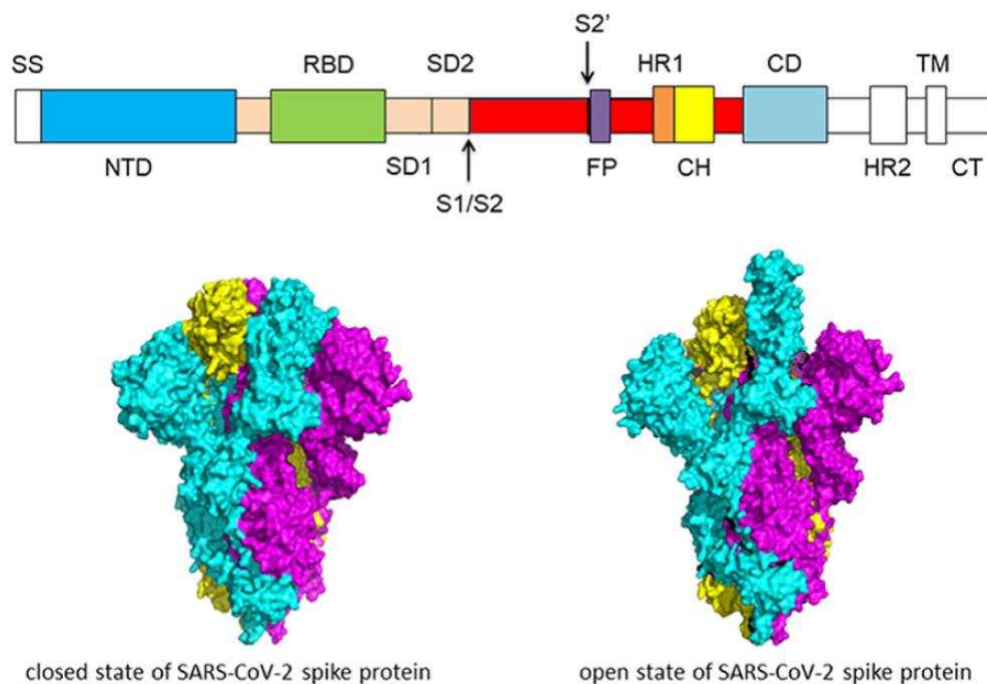
La proteina *Spike*, codificata dal gene S, si organizza a formare un omotrimerico i cui monomeri sono costituiti da due subunità, S1 e S2, associate in modo non-covalente. Durante l'entrata del virus nella cellula ospite, queste due subunità svolgono due funzioni distinte: la subunità S1 lega il recettore cellulare ACE2, mentre la subunità S2 media le successive fasi di ancoraggio e fusione alla membrana cellulare.<sup>18</sup>

La proteina *spike* ha un'organizzazione ben precisa sulla membrana del pericapside: si presenta con una "coda" interna al pericapside, un dominio transmembrana e un ectodominio.<sup>11</sup>

La struttura della subunità S1 si compone di quattro domini: il dominio N-terminale (NTD), costituito soprattutto da quattro  $\beta$ -foglietti e da una serie di *loops*, il dominio di legame al recettore (*Receptor-Binding Domain*, RBD), formato da un  $\beta$ -foglietto antiparallelo centrale di cinque filamenti e da due corte  $\alpha$ -eliche collegate ad esso da dei *loops*, infine due domini C-terminali (CTD1 e CTD2), costituiti soprattutto da strutture a  $\beta$ -foglietto. In particolare, il dominio RBD presenta un sottodominio chiamato "*Receptor-Binding Motif*" (RBM) che interagisce con il recettore ACE2.

Anche la subunità S2 si compone di diversi domini: il *fusion peptide* (FP), *heptad repeat 1* (HR1), *central helix* (CH), *connector domain* (CD), *heptad repeat 2* (HR2), *transmembrane domain* (TM), e *cytoplasmic tail* (CT).

La *spike* si può presentare in due distinte conformazioni, di prefusione e di postfusione, che sono riconducibili ad un rimodellamento della struttura delle due subunità: nella conformazione di prefusione la subunità S2 forma un fascio centrale di nove eliche che sono avvolte dalle subunità S1, mentre i tre RBD del trimero si trovano in posizione apicale nella molecola; nella conformazione di postfusione, invece, si osserva il distacco delle due subunità l'una dall'altra e la S2 si ripiega a formare una struttura allungata e stabile. I tre RBD della molecola, nello stato di prefusione, possono adottare due distinte configurazioni: la conformazione “*up*”, necessaria per il legame con il recettore ACE2, e la conformazione “*down*”, che non permette l'interazione con il recettore.<sup>18</sup> (Figura 7)



**Figura 7** - Rappresentazione schematica della glicoproteina spike di SARS-CoV-2. In altro, sono riportati i vari domini: SS, single sequence; NTD, dominio N-terminale; RBD, receptor-binding domain; SD1, sottodominio 1; SD2, sottodominio 2; S1/S2, sito di taglio della proteasi S1/S2; S2', sito di taglio della proteasi S2'; FP, fusion peptide; HR1, heptad repeat 1; CH, central helix; CD, connector domain; HR2, heptad repeat 2; TM, dominio transmembrana; CT, coda citoplasmatica. Le frecce indicano i siti di taglio delle proteasi. In basso, struttura mediante microscopia crioelettronica (Cryo-EM) della proteina spike, a sinistra nella conformazione “*down*” (closed state) e a destra nella conformazione “*up*” (open state). Fonte: M. Wang et al, 2020.<sup>10</sup>

Nella conformazione “*down*” il dominio RBD di uno dei trimeri assume un particolare ripiegamento verso la struttura centrale di eliche della subunità S2, ciò rende inaccessibile l’RBM al legame con ACE2. Nella configurazione “*up*”, invece, RBD si sposta in modo tale da esporre il sottodominio RBM e renderlo così accessibile per l’interazione con il recettore. Questa fase di transizione da una configurazione “*down*” ad una “*up*” è un passaggio molto importante per il legame della proteina *spike* al recettore ACE2 e quindi per il processo di entrata del virus nella cellula.

Sulla proteina S sono presenti due importanti siti di taglio: il sito S1/S2, costituito da una sequenza di residui multibasici (Arg-Arg-Ala-Arg) situata fra le due subunità della *spike*, e il sito S2’, che precede il dominio FP nella subunità S2.

Il sito S1/S2 viene scisso dalla proteasi transmembrana furina nella cellula produttrice del virus, mentre la scissione del sito S2’ avviene a carico di proteasi della cellula ospite, quali la *Transmembrane Serine Protease 2* (TMPRSS2) o la catepsina L, a seguito del legame del virus con il recettore ACE2.<sup>18</sup>

### **1.2.3 Proteina orf3a**

L’ ORF3a codifica per una proteina transmembrana che si va a localizzare nella membrana cellulare, in quella del reticolo endoplasmatico e nell’apparato del Golgi.

Una delle funzioni della proteina orf3a è quella di sfuggire alla degradazione reticolo endoplasmatico-associata attivando la via di segnalazione *PKR-like ER kinase* (PERK), coinvolto anche nell’induzione dell’apoptosi cellulare.

In alcune linee cellulari, è stato dimostrato che la proteina orf3a induce l'attivazione del Fattore Nucleare kB (NF-kB) e delle *Janus* chinasi (JNK) portando a una sovraregolazione indiretta delle chemochine chiamate RANTES (*Regulated upon activation, Normal T cell expressed and secreted*) e dell'Interleuchina 8 (IL-8).<sup>3</sup>

La proteina orf3a è indirettamente coinvolta nell'attivazione della risposta infiammatoria: l'orf3a svolge la funzione di canale ionico e la sua attività coinvolge l'attivazione dell'inflammasoma NLRP3 (*Nod-like receptor pyrin 3*); attraverso il taglio proteolitico della procaspasi-1 si viene a formare la caspasi-1 che porta all'attivazione dell'Interleuchina-1 $\beta$  (IL-1 $\beta$ ), che è un importante mediatore della risposta infiammatoria, oltre a vari processi cellulari quali la proliferazione cellulare, il differenziamento e l'apoptosi.<sup>17</sup>

#### **1.2.4 Proteina Envelope (E)**

La proteina E è una proteina virale di piccole dimensioni, circa 75 aminoacidi, codificata dal gene E.<sup>19</sup>

La proteina E si localizza nell'envelope ed è caratterizzata da due domini distinti: il dominio idrofobico transmembrana e la coda citoplasmatica.

Questa proteina svolge un importante ruolo nell'assemblaggio e nel rilascio dei virioni. Un'altra funzione nota della proteina E è quella di formare viroporine sulle membrane cellulari così da agire sul trasporto selettivo di ioni, quali il Ca<sup>2+</sup>, attraverso la membrana e favorire il rilascio del virus maturo.<sup>20</sup>

### 1.2.5 Proteina Membrane (M)

La proteina M è la proteina più abbondante presente nella famiglia dei *Coronaviridae*. È una proteina strutturale di membrana di circa 222 aminoacidi caratterizzata dalla presenza di tre domini transmembrana. Questa proteina è codificata dal gene M e, insieme alle proteine E ed S, concorre alla formazione del pericapside virale. La proteina M, inoltre, svolge un importante ruolo nell'assemblaggio della particella virale.<sup>19</sup>

### 1.2.6 Proteina orf6

Il gene ORF6 codifica per un polipeptide di circa 63 aminoacidi caratterizzato da una regione N-terminale anfipatico di 40 aminoacidi e da una regione C-terminale polare. Questa proteina si va a localizzare nella cellula nel reticolo endoplasmatico e nell'apparato di Golgi. La porzione N-terminale presenta una struttura secondaria a  $\alpha$ -elica che si inserisce nel doppio strato delle membrane cellulari. Nella regione C-terminale troviamo due sequenze segnale: la sequenza YSEL, che interagisce con le proteine incorporate negli endosomi, e la coda della molecola a carattere acido, che ha un segnale di esportazione dal RE.<sup>3</sup>

La funzione di orf6 è anche quella di migliorare la replicazione virale andando ad interagire con la nsp8.<sup>17</sup>

L'orf6 ha un importante ruolo nella patogenesi virale del SARS-CoV-2; infatti questa proteina è anche conosciuta per la sua azione antagonista verso l'interferone  $\beta$ , può inoltre interferire con il trasporto reticolo endoplasmatico/apparato del Golgi necessario per la risposta dell'interferone.



La proteina orf6 gioca anche un importante ruolo nell'evasione del sistema immunitario dell'ospite; il suo dominio C-terminale svolge una funzione regolatoria del trasporto nucleare della cellula ospite e della via di segnalazione dell'interferone I.<sup>3</sup>

### **1.2.7 Proteina orf7a**

La ORF7a codifica per una proteina transmembrana di 122 aminoacidi, costituita da una regione N-terminale di 15 aminoacidi con una sequenza segnale, un ectodominio di 81 aminoacidi, un dominio C-terminale transmembrana di 21 aminoacidi e una "coda" citoplasmatica di 5 residui. Questa proteina si va a localizzare nell' *ER-Golgi Intermediate Compartment* (ERGIC).

Orf7a interagisce con altre proteine virali quali M ed E, suggerendo un possibile ruolo nell'assemblaggio del virione durante la replicazione.<sup>3</sup>

### **1.2.8 Proteina orf7b**

La proteina accessoria orf7b consiste in una proteina transmembrana di circa 44 aminoacidi e molto idrofobica. Questa proteina si localizza a livello dell'apparato del Golgi, con la regione N-terminale rivolta verso il lume dell'apparato e la regione C-terminale verso il citoplasma cellulare.

La funzione di orf7a ad oggi non è ancora nota.<sup>3</sup>

### 1.2.9 Proteina orf8

ORF8 nel SARS-CoV-2 codifica per una singola proteina, differentemente dal SARS-CoV che invece ne presenta due, ossia orf8a e orf8b. Nel SARS-CoV è presente una delezione di 29 nucleotidi rispetto alle ORF8 analizzate in specie quali lo zibetto o il pipistrello *horseshoe*. Tale delezione è alla base della divisione in orf8a e orf8b, suggerendo anche un possibile ruolo della ORF8 nella trasmissione interspecie.

Nel SARS-CoV orf8 ha un ruolo nella attivazione degli inflammasomi NLRP3 e nell'innescare percorsi di stress intracellulare, ma questa funzione ad oggi non è ancora stata confermata per la orf8 di SARS-CoV-2.<sup>3</sup>

### 1.2.10 Proteina Nucleocapsid (N)

La proteina nucleocapsidica è codificata dal gene N, ha una dimensione di circa 43-50 kDa e si lega all'RNA virale nel ribonucleocapside. È costituita da tre domini: il dominio N-terminale (NTD), il linker centrale (CL) e il dominio C-terminale (CTD). La funzione di legame all'RNA virale è svolta dal dominio NTD, in quanto presenta il maggior numero di aminoacidi carichi positivamente.

La proteina N ha un'importante funzione nell'assemblaggio del virione, interagendo con la proteina M e con il genoma del virus. Inoltre, ha anche un ruolo nei processi di trascrizione e replicazione dell'RNA virale.<sup>20</sup>

La proteina N è altamente conservata nella famiglia dei *Coronaviridae*, nel SARS-CoV-2 presenta circa il 90% di identità di sequenza con quella di SARS-CoV.<sup>19</sup>

### **1.2.11 Proteina orf10**

La ORF10 di SARS-CoV-2 codifica per una proteina di 38 aminoacidi di cui non sono stati riscontrati omologhi in altri *betacoronavirus*.<sup>3,17</sup>

Ad oggi, le funzioni di questa proteina non si conoscono a pieno.

### 1.3 Il Ciclo Replicativo di SARS-CoV-2

Il ciclo replicativo di SARS-CoV-2 si articola in diverse fasi: l'attacco al recettore cellulare e l'entrata nella cellula ospite, la traduzione delle proteine non strutturali, la replicazione e trascrizione del genoma virale, la traduzione e maturazione delle proteine strutturali e accessorie, infine, l'assemblaggio del virione e il suo rilascio. (Figura 8)

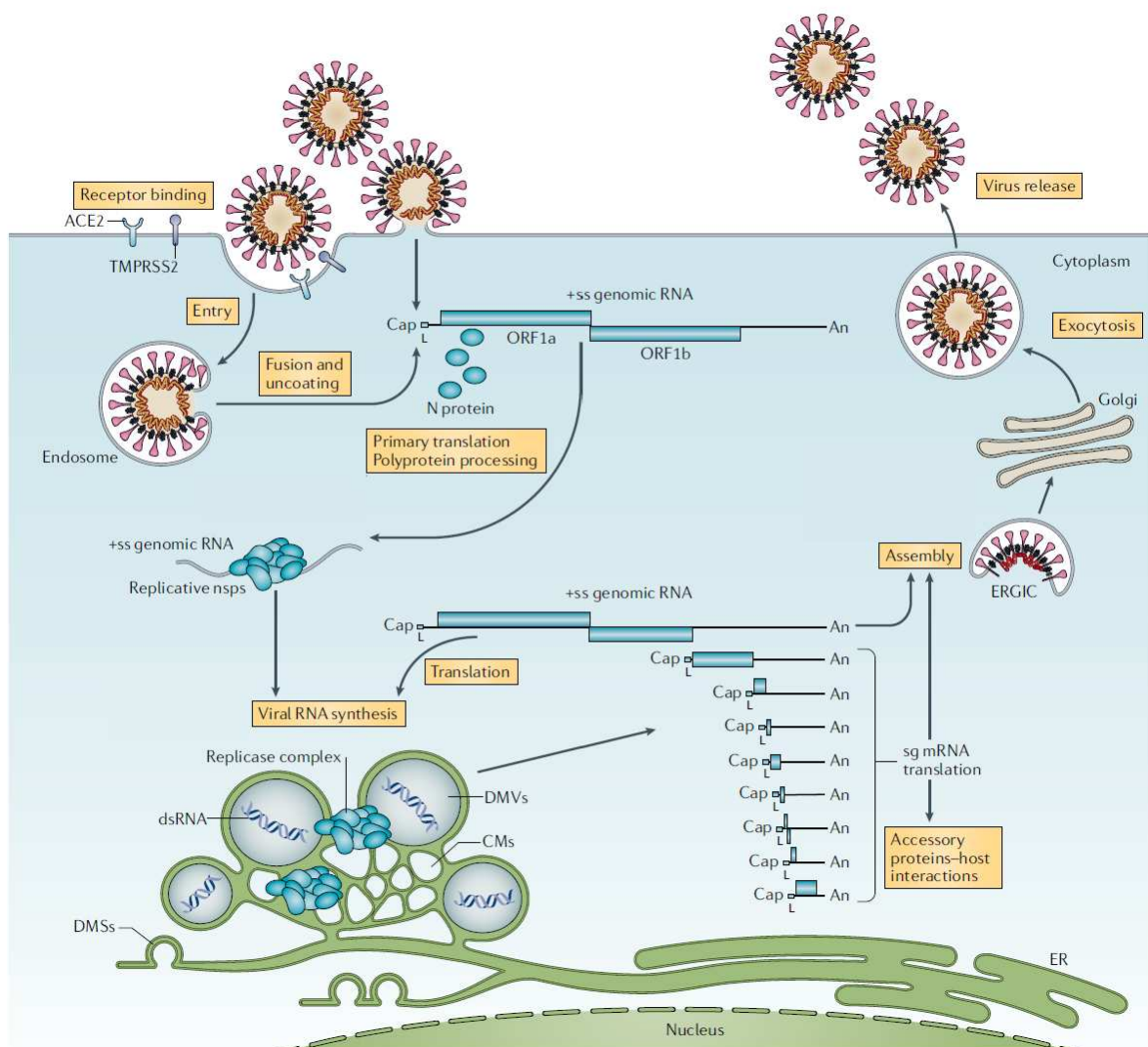


Figura 8 – Fasi del ciclo replicativo di SARS-CoV-2. Fonte: Modificata da P. V'kovski et al, 2020. <sup>1</sup>

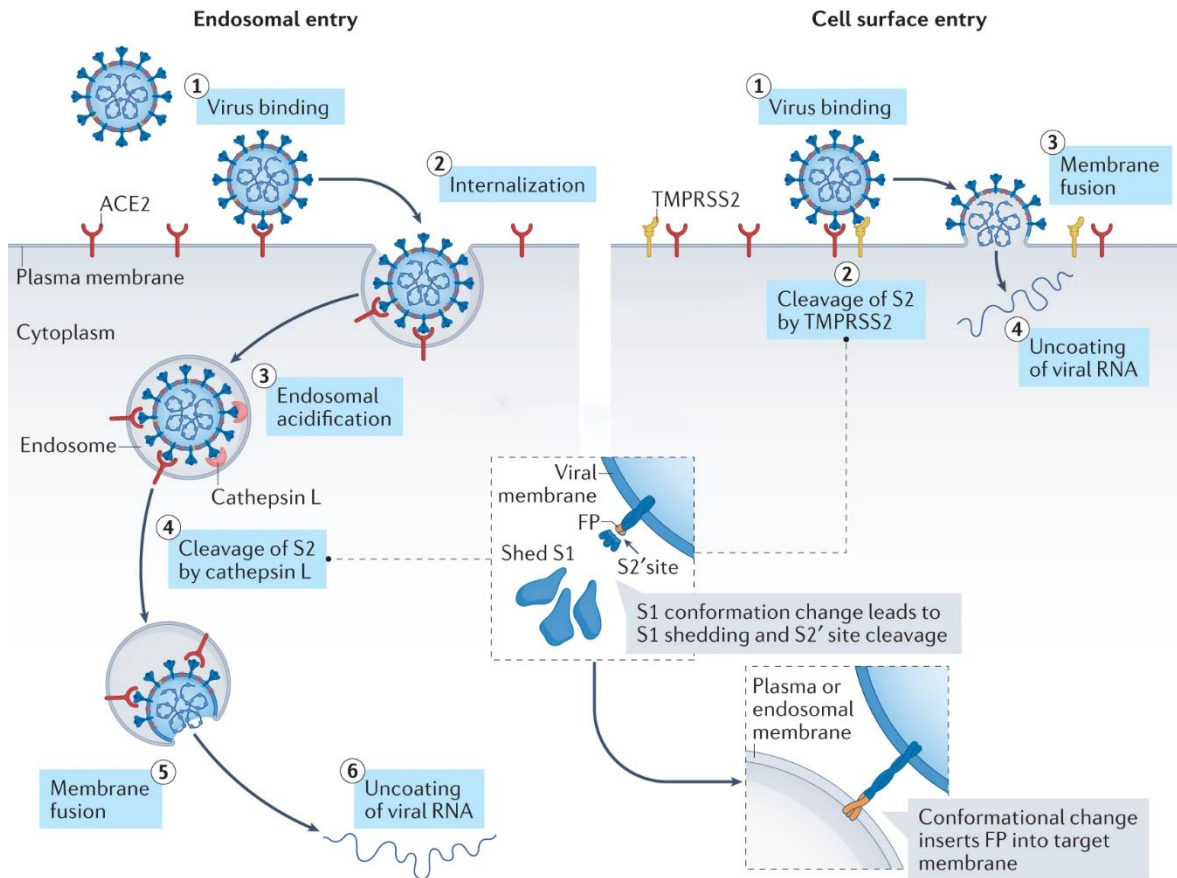
### 1.3.1 Attacco al recettore ed entrata nella cellula ospite

L'ingresso di SARS-CoV-2 nella cellula inizia con l'interazione della glicoproteina *spike* con il recettore target sulla superficie cellulare, l'*Angiotensin Covering Enzyme 2* (ACE2). In particolare, l'interazione con il recettore avviene grazie alla subunità S1 della glicoproteina, dove è sito il dominio RBD.

Per rendere possibile l'attacco del virus sulla superficie cellulare, la proteina *spike* deve subire dei cambiamenti conformazionali da uno stato metastabile a uno stato a minor energia, ossia la proteina virale deve passare dalla conformazione *down* alla conformazione *up*. Questi cambiamenti sono possibili grazie a due eventi di scissione proteolitica a carico di due siti: il sito S1/S2, al confine fra le due subunità della proteina, e al sito S2', nella subunità S2. La scissione del sito S1/S2 è un passaggio essenziale per la successiva scissione del sito S2' e avviene durante la maturazione del virione nella cellula infettata. Invece, il taglio del sito S2' avviene a seguito del legame del virus al recettore ACE2 e coinvolge proteasi cellulari diverse a seconda del meccanismo di entrata.<sup>18</sup>

La modalità di entrata principale è quella mediata dalla proteasi *Transmembrane Serine Protease 2* (TMPRSS2) presente sulla membrana plasmatica della cellula ospite. La via di entrata alternativa, invece, coinvolge le catepsine, in particolare la catepsina L. Quest'ultima avviene nel caso in cui l'infezione virale avvenga in cellule con un basso livello di espressione dell'enzima TMPRSS2. Se la *spike* legata al recettore cellulare non viene tagliata dalla TMPRSS2, il virus viene internalizzato per endocitosi mediata da

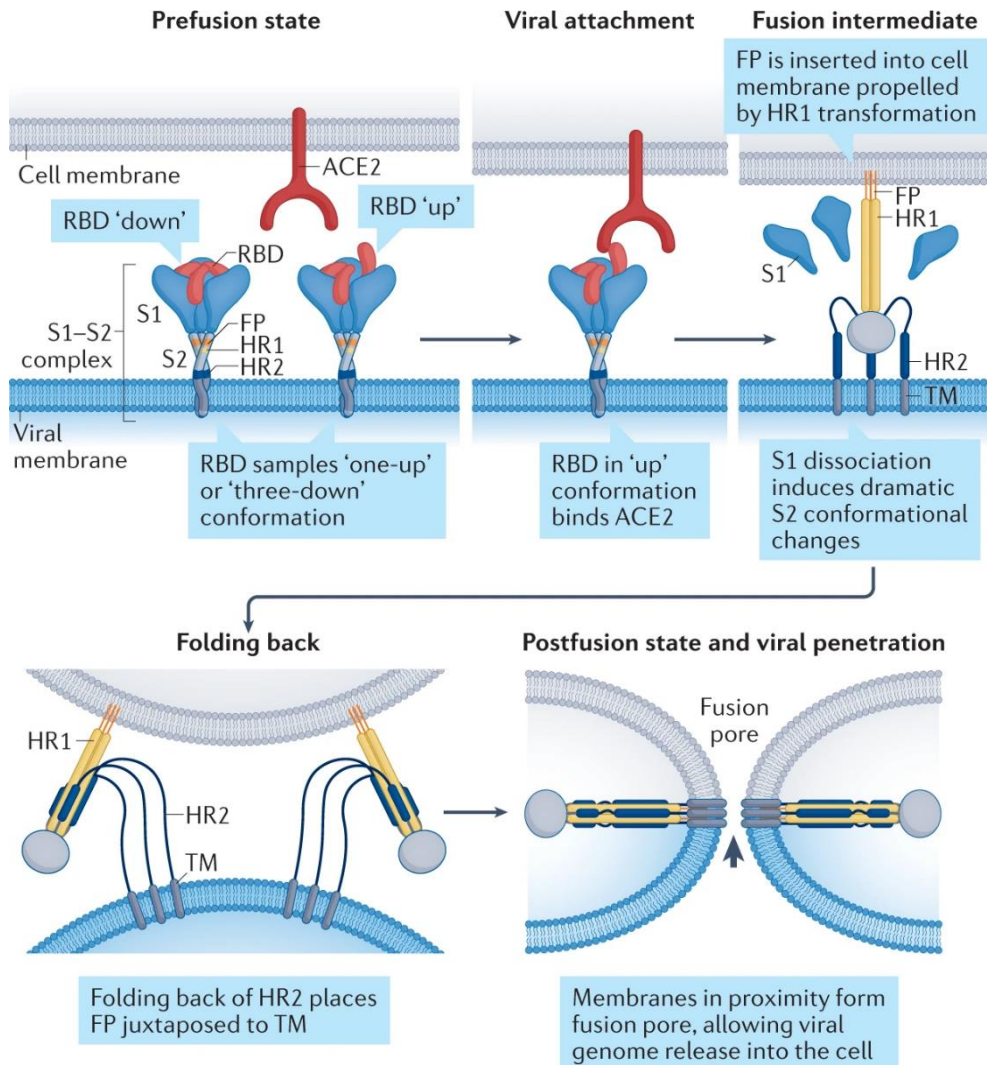
clatrina e, grazie all'ambiente acido dell'endolisosoma, avviene l'azione della catepsina L a livello del sito S2'. (Figura 9)<sup>18</sup>



**Figura 9** - Due meccanismi di entrata nella cellula di SARS-CoV-2. A sinistra, nella via endosomiale di entrata, il taglio al sito S2' è svolto dalla proteasi catepsina L. A destra, nell'entra diretta dalla superficie cellulare, il taglio al sito S2' è svolto dalla proteasi TMPRSS2. Fonte: Modificata da C.B. Jackson et al, 2022.<sup>18</sup>

In entrambe le vie, a seguito del taglio del sito S2', avviene l'esposizione del dominio FP della subunità S2.

La dissociazione delle due subunità della *spike* induce dei drastici cambiamenti conformazionali nella molecola, soprattutto nella subunità S2 che passa da uno stato di prefusione metastabile alla struttura di postfusione stabile. Questa transizione porta il dominio *fusion peptide* (FP) a penetrare nella membrana della cellula bersaglio, guidato dal dominio *heptad repeat 1*, mentre il dominio *heptad repeat 2* si ripiega verso la membrana del virus. Da questa cascata di eventi si ottiene l'avvicinamento delle due membrane, quella virale e quella cellulare, portando così alla loro fusione. (Figura 10) <sup>18</sup>



**Figura 10** - Serie di cambiamenti conformazionali della proteina spike di SARS-CoV-2 che portano all'avvicinamento e alla fusione del rivestimento virale con la membrana cellulare. Fonte: C.B. Jackson et al, 2022. <sup>18</sup>

A seguito della fase di fusione si ha il rilascio dell'RNA virale nel citoplasma cellulare e si prosegue con le successive fasi del ciclo replicativo.<sup>18</sup>

### 1.3.2 Traduzione delle proteine non strutturali

A seguito dell'entrata dell'RNA virale nel citoplasma della cellula ospite e della sua liberazione dalla proteina N, ha inizio la fase di traduzione *cap*-dipendente delle proteine non strutturali. L'RNA di SARS-CoV-2 è a polarità positiva, quindi viene riconosciuto dal macchinario traduzionale della cellula come un mRNA e viene tradotto direttamente dai ribosomi liberi nel citoplasma. In questa fase vengono tradotti i primi due-terzi del genoma virale, ossia l'ORF1a e ORF1b producendo le rispettive poliproteine, pp1a e pp1ab. La traduzione di queste due poliproteine è il risultato della presenza di un *frameshift* di un nucleotide al confine fra le due ORF; nel 20-30% degli eventi traduzionali i ribosomi non raggiungono il codone di stop della ORF1a, ma proseguono la traduzione oltre la sovrapposizione ORF1a/1b portando alla formazione di una poliproteina più lunga, ossia la pp1ab.<sup>2,21</sup>

A favorire la preferenziale traduzione delle proteine virali vi è la nsp1. Infatti, questa proteina è la prima ad essere prodotta e va ad interferire con il processo di traduzione degli mRNA cellulari.

Le due poliproteine virali vengono tagliate dalle proteasi virali M<sup>pro</sup>, taglia la maggior parte delle nsp e corrisponde alla nsp5, e PL<sup>pro</sup>, che costituisce uno dei domini della nsp3. Da questa azione proteolitica si ottengono le 16 nsp del virus.



Le nsp 2-16 prendo parte, sia direttamente che indirettamente, al processo di trascrizione e replicazione del genoma virale. <sup>1</sup>

Le funzioni di tutte le nsp sono trattate più nel dettaglio nel paragrafo 1.2.1.

### 1.3.3 Replicazione e trascrizione del genoma virale

In questa fase, come anche la fase di traduzione proteica e l'assemblaggio del virione, avviene a livello di particolari strutture chiamate "organelli di replicazione", ossia strutture perinucleari a doppia membrana prodotte dalle interazioni delle nsp virali con le membrane del reticolo endoplasmatico. Queste strutture, anche chiamate *Double-Membrane Vesicles* (DMVs), sono importanti in quanto l'RTC svolge la sua funzione di sintesi ancorato alla loro membrana e, inoltre, forniscono agli intermedi di replicazione virale un ambiente meno esposto ai sistemi di difesa cellulari. <sup>1</sup>

Il processo di replicazione del genoma virale inizia con la sintesi di un RNA intermedio a polarità negativa. Il complesso di replicazione e trascrizione, costituito dalle proteine virali nsp12, nsp7 e nsp8, permette l'entrata della molecola di RNA *template* attraverso i motivi A e C del sito attivo della RdRp; il motivo G avvia la sintesi dell'RNA complementare interagendo con il primer, precedentemente prodotto dalla primasi, mentre gli NTP necessari per la polimerizzazione raggiungono il sito di catalitico attraverso il canale che si trova adiacente al sottodominio *palm* della RdRp. L'RNA in uscita è un ibrido a doppio filamento di RNA *template*-prodotto che poi verrà processato per dare RNA a singolo filamento, necessario per la traduzione. <sup>20</sup>

Possiamo suddividere la sintesi dell'RNA di SARS-CoV-2 in replicazione del genoma e trascrizione degli mRNA subgenomici. Infatti, nella replicazione

dell'RNA viene prodotto un filamento a lunghezza intera dell'RNA del virus, chiamato RNA genomico (gRNA); il gRNA viene usato come *template* per la sintesi di ulteriori gRNA, può essere tradotto in poliproteine aggiuntive, infine, viene impacchettato nei virioni che verranno poi rilasciati dalla cellula. Nella trascrizione dell'RNA, invece, vengono prodotti un set annidato di RNA più corti, chiamati mRNA subgenomici (sg-mRNA), che servono per la traduzione delle proteine strutturali e accessorie virali. La sintesi di questi sg-mRNA avviene tramite un meccanismo particolare, la trascrizione discontinua.<sup>21</sup>

La trascrizione virale discontinua che porta alla produzione di molteplici sg-mRNA annidati è una caratteristica distintiva dei *coronavirus* e di molti membri dei *Nidovirales*.<sup>2</sup> A monte delle ORF che codificano per le proteine strutturali, nella porzione 3' del genoma virale, sono presenti brevi elementi di sequenza, le *body transcription regulatory sequences* (TRS-B). Quando il complesso RTC, durante la sintesi dell'RNA a polarità negativa, incontra una sequenza TRS-B interrompe la trascrizione in quel punto e, insieme all'estremità 3' del filamento di RNA nascente, traslocata in una sequenza TRS vicina a una sequenza *leader* (TRS-L), situata a circa 70 nucleotidi dall'estremità 5' del genoma, dove riparte con la sintesi del filamento in sospenso. La trascrizione discontinua richiede l'interazione fra la sequenza TRS-L del gRNA a polarità positiva e il TRS-B dell'RNA a polarità negativa complementare neosintetizzato (anti-TRS-B)<sup>21</sup>.

A seguito della ripresa della sintesi dalla sequenza TRS-L, una copia della sequenza *leader* del filamento negativo viene aggiunta all'RNA nascente per completare la sintesi dell'sgRNA a polarità negativa. Infine, l'sgRNA a polarità negativa appena prodotto viene usato come stampo per la sintesi di sg-mRNA a polarità positiva. Gli sg-mRNA che si ottengono sono sequenze di varia lunghezza, strutturalmente policistronici, ma solo la prima ORF al 5' viene tradotta e quindi risultano funzionalmente monocistronici.<sup>1,21</sup> (Figura 11)

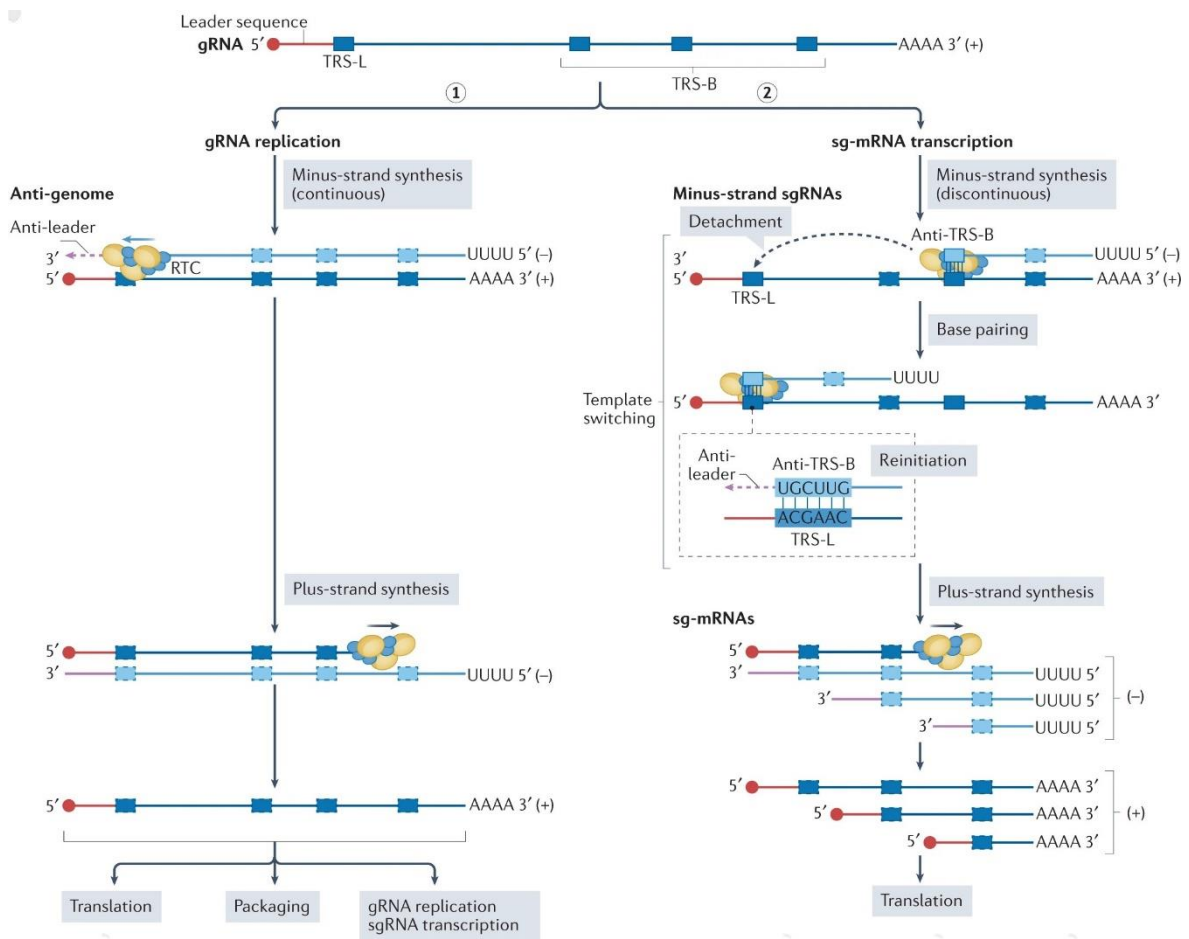


Figura 11 - Rappresentazione schematica della replicazione e trascrizione dell'RNA genomico di SARS-CoV-2. Fonte: B.Malone, 2021.<sup>21</sup>

### 1.3.4 Traduzione delle proteine strutturali e accessorie e assemblaggio del virione

Le proteine strutturali S, M ed E e parte le proteine accessorie vengono tradotte nel reticolo endoplasmatico, mentre la proteina strutturale N viene tradotta dai ribosomi liberi nel citoplasma. Successivamente la proteina N si associa all'RNA genomico neosintetizzato in modo da formare il nucleocapside elicoidale.

Il complesso ribonucleoproteico virale e le proteine S, M ed E si muovono lungo la via secretoria del *ER-Golgi Intermediate Compartment* (ERGIC) dove si assemblano per produrre una particella virale matura. L'assemblaggio del virione avviene grazie a una serie di interazioni proteina-proteina, mediate soprattutto dalla proteina virale M. Anche la proteina E contribuisce all'assemblaggio delle particelle virali inducendo la curvatura della membrana e interagendo con le proteine M, così da prevenire la loro aggregazione. In questa fase si osserva la formazione delle *virus-like particles* (VLPs), ossia le particelle virali che ancora non presentano la proteina *spike*.

Nel compartimento ERGIC il nucleocapside prende contatto con la proteina M sulla membrana delle vescicole del Golgi e per gemmazione si forma l'*envelope*. A questo punto anche la proteina *spike* si inserisce nei virioni così da completarne la maturazione.<sup>15,22,23</sup>

Il virione completamente assemblato viene trasportato verso la superficie della cellula ospite e, infine, rilasciato all'esterno attraverso il meccanismo dell'esocitosi; questo processo è facilitato dalla presenza della proteina virale orf3a che forma un canale ionico sulla membrana utile al rilascio del virus.<sup>20</sup>

## 1.4 Tropismo di SARS-CoV-2

Il SARS-CoV-2, come anche altri *coronavirus*, utilizza il recettore *Angiotensin Converting Enzyme 2* (ACE2) per entrare nella cellula ospite.

ACE2 è un enzima presente sulla membrana cellulare, a cui è ancorato tramite il suo dominio C-terminale e da cui sporge attraverso il suo dominio N-terminale, dove è presente il sito catalitico che è anche il target per l'attacco di SARS-CoV-2.<sup>24</sup> Questo recettore è ampiamente espresso sulle cellule epiteliali di alveoli, trachea, bronchi, ghiandole sierose bronchiali, oltre a monociti e macrofagi alveolari.<sup>25</sup>

Fra le sintomatologie più riscontrate durante l'infezione da SARS-CoV-2 vi è la difficoltà respiratoria, nonostante ACE2 sia poco espresso a livello polmonare, fatta eccezione per le cellule epiteliali alveolari di tipo II.<sup>24</sup>

Il recettore ACE2 è espresso anche in altri distretti anatomici, come quello renale, quello cardiaco-vascolare e quello gastroenterico, permettendo la replicazione del virus in modo significativo.<sup>24</sup> Inoltre, molte cellule umane possiedono recettori accessori ACE2-dipendenti implicati nell'entrata del virus nelle cellule. Fra questi recettori vi sono la furina, la TMPRSS2 e TMPRSS4, la tripsina, le catepsine, la neuropilina-1, glicolipidi con acido sialico, la vimentina, l'eparan solfato e il recettore per la fosfatidilserina.<sup>26</sup>

## 1.5 Patogenesi del COVID-19

Il SARS-CoV-2 è l'agente eziologico responsabile della malattia di *Coronavirus Disease-19* (COVID-19), una condizione che può manifestarsi sia con sintomi di lieve entità o anche progredire con sintomatologie più severe.

Le prime cellule ad essere attaccate da SARS-CoV-2 sono le cellule multiciliate nel tessuto nasofaringeo o nella trachea. Queste cellule esprimono alti livelli di ACE2 e TMPRSS2 sulla loro superficie, favorendo così l'attacco del virus e costituendo la via di entrata del virus nell'organismo. A contrastare la diffusione del virus sono molto importanti le rapide risposte del sistema immunitario innato, in particolare hanno un importante ruolo l'interferone di tipo I e di tipo III. Una bassa o inefficiente risposta dell'interferone ha come conseguenza la disseminazione dell'infezione nell'organismo, in particolare a livello polmonare, e le sintomatologie più gravi del COVID-19.<sup>27</sup>

Le manifestazioni cliniche più comuni del COVID-19 sono tosse, febbre, mialgia, mal di testa e sintomi gastrointestinali, oltre a una condizione di leucopenia e ad alti livelli di citochine nel plasma.<sup>10</sup>

Il COVID-19 può anche sfociare in un decorso più severo, con sintomatologie che possono manifestarsi a circa 1 settimana dall'esordio dei primi sintomi. Le manifestazioni più rilevanti sono l'ipossia, con conseguente desaturazione, e la polmonite bilaterale. Nei casi più gravi la malattia può progredire verso la Sindrome da Distress Respiratorio Acuto (ARDS) che si manifesta con un consistente danno ai polmoni causato da una eccessiva flogosi e la conseguente perdita di aerazione dei tessuti.<sup>10,27</sup>

Alcuni pazienti affetti da COVID-19 manifestano una condizione di iperinflammazione, dovuta a un elevato rilascio di citochine, quali Interleuchina-1 (IL-1), Interleuchina-6 (IL-6), Interleuchina-8 (IL-8) e il

Fattore di Necrosi Tumorale (TNF), oltre ad elevate concentrazioni di marcatori infiammatori, quali D-dimero, ferritina e la Proteina C Reattiva (CRP).<sup>27</sup>

Il COVID-19 può anche portare a complicazioni della coagulazione e trombotiche, e alla Sindrome da Disfunzione Multiorgano (MODS), con possibili danni epatici, renali e cardiaci.

Nei decorsi clinici più critici, soprattutto in soggetti che presentano comorbidità, l'ARDS e la MODS possono risultare in un esito fatale.<sup>10,27</sup> I meccanismi alla base dell'insorgenza di queste due sindromi sono soprattutto a carico della condizione di iperinflamazione sistemica causata dall'infezione, che provoca danni al tessuto polmonare e vascolare, il tutto esacerbato dalla disregolazione del Sistema Renina-Angiotensina-Aldosterone (RAAS).

### **1.5.1 Attivazione della risposta immunitaria innata dell'ospite verso il SARS-CoV-2**

Il SARS-CoV-2 mette in atto delle efficienti strategie di evasione della risposta immunitaria innata. In particolare, attraverso la produzione delle proteine non strutturali, va ad interferire con l'espressione dell'interferone di tipo I, che è una importante molecola antivirale, e favorisce al contempo la produzione di molte citochine proinfiammatorie, quali IL-1, IL-6, IL-8 e TNF- $\alpha$ , implicate nella tempesta di citochine che caratterizza le gravi forme di COVID-19. La soppressione dei meccanismi antivirali consente al SARS-CoV-2 di replicare e propagarsi nell'organismo, il quale nella maggior parte dei casi riesce a contenere l'infezione virale, ma in alcuni soggetti la forte risposta immunitaria innescata dal virus risulta eccessiva e incontrollata, causando così le gravi patologie polmonari e sistemiche del COVID-19.<sup>11</sup>

A livello polmonare, nelle fasi iniziali dell'infezione, il SARS-CoV-2 infetta soprattutto le cellule alveolari di tipo II; il danno virale a queste cellule riduce la produzione di surfactante, obliterando gli alveoli polmonari. In questa fase si riscontra una forte linfopenia dovuta al sequestro linfocitario nelle sedi di infezione.<sup>28</sup>

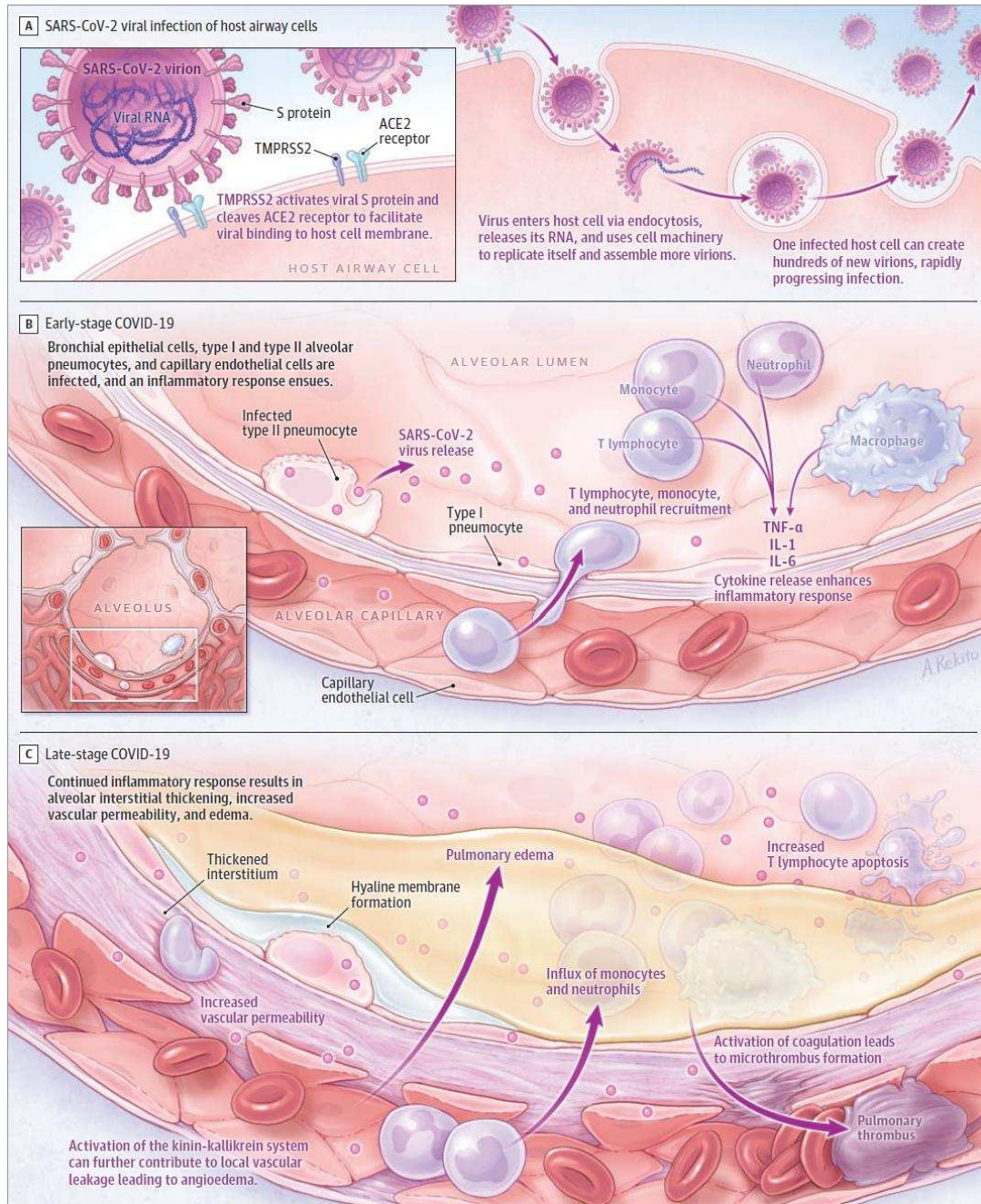
L'infezione virale produce nelle cellule infettate elevate quantità di *Pathogen Associated Molecular Patterns* (PAMPs), costituite dai prodotti della replicazione del virus, e il danno cellulare che ne deriva porta al rilascio di *Damage Associated Molecular Patterns* (DAMPs), quali lipidi ossidati e altri componenti associati all'apoptosi, all'autofagia, alla necroptosi e alla piroptosi. Le PAMPs vengono rilevate dai *Pattern Recognition Receptors* (PRRs) intracellulari e richiamano le cellule del sistema immunitario innato, quali macrofagi, monociti e neutrofili. Sia le PAMPs che le DAMPs portano all'attivazione della risposta proinfiammatoria da parte delle cellule del sistema immunitario che rilasciano grandi quantità di citochine e chemochine, come ad esempio IL-6, IL-8 e *Tumor Necrosis Factor  $\alpha$*  (TNF  $\alpha$ ).<sup>11,29</sup>

Nelle fasi successive dell'infezione, a causa dell'azione dei vari fattori proinfiammatori, si ha un aumento della permeabilità vascolare, utile per il reclutamento delle cellule del sistema immunitario, ma che può compromettere l'integrità della barriera epiteliale-endoteliale, portando anche alla penetrazione di liquidi negli alveoli e provocando dispnea e insufficienza respiratoria.<sup>28,30</sup>

Le cellule *Natural Killer* (NK) hanno un ruolo chiave nel combattere la propagazione virale, ma allo stesso tempo contribuiscono al danno tessutale. Le infiltrazioni interstiziali che si verificano in questa fase dell'infezione sono responsabili della formazione dell'edema polmonare che va ad interferire con i normali scambi di ossigeno tra alveolo e capillare, predisponendo così il



polmone alla Sindrome da Distress Respiratorio Acuto (ARDS).<sup>28,29</sup> (Figura 12)



**Figura 12** - Illustrazione della risposta immunitaria dell'ospite indotta da SARS-CoV-2 a livello del tratto respiratorio inferiore. Sono mostrati (A) l'entrata del virus nella cellula ospite; (B) la fase iniziale di infezione, con il rilascio di citochine infiammatorie da parte delle cellule infettate, dai macrofagi alveolari, dai linfociti T, dai monociti e ai neutrofili in risposta all'elevata carica virale presente; (C) la fase tardiva dell'infezione con comparsa dell'edema polmonare che caratterizza la sindrome da distress respiratorio acuto. Fonte: Wiersinga et al, 2020<sup>28</sup>

Nei casi più gravi di COVID-19, tutti questi fattori, uniti allo stress ossidativo e all'ambiente pro-infiammatorio dovuto allo sbilanciamento del RAAS, portano ad una risposta iperinfiammatoria, una conseguenza dell'eccessiva produzione di molecole infiammatorie anche definita "tempesta di citochine". I neutrofili sono anche implicati nella produzione di *Neutrophil Extracellular Traps* (NET) in risposta al danno endoteliale, alla produzione di ROS e ai diversi stimoli infiammatori. La formazione dei NET serve a contenere la diffusione del virus, ma al contempo innescano le vie della coagulazione e favoriscono così la formazione di trombi.<sup>29</sup>

### **1.5.2 Fattori di rischio e complicazioni cliniche del COVID-19**

Il SARS-CoV-2 può infettare tutti i soggetti suscettibili in una popolazione, ma alcuni fattori incidono negativamente sul decorso della malattia da COVID-19, potenzialmente concorrendo alle sintomatologie più severe. Fra i fattori di rischio individuali vi sono l'invecchiamento e la presenza di comorbidità al momento dell'infezione.

Gli individui appartenenti a fasce di età più avanzate risultano essere maggiormente predisposti alle sintomatologie più severe del COVID-19 e con tassi di mortalità significativamente più alti rispetto ai soggetti più giovani.<sup>30</sup> È stato stimato che persone con età maggiore di 80 anni presentano un tasso di mortalità a causa della malattia 20 volte più alto rispetto a individui con età compresa fra i 50-59 anni. Le motivazioni potrebbero risiedere nella più probabile presenza di malattie croniche associate all'invecchiamento, alla minor efficienza del sistema immunitario innato nel contrastare la replicazione virale.<sup>30,31</sup> Tutto ciò porta il virus a diffondersi nell'organismo in modo

incontrollato e ad una conseguente maggiore risposta proinfiammatoria, aumentando il rischio di morte.<sup>32</sup>

Fra le comorbidità che possono aggravare il decorso della malattia da COVID-19 vi sono l'ipertensione e le malattie cardiovascolari (CVDs). Le cellule endoteliali esprimono il recettore ACE2, quindi possono venir infettate dal virus e subire i danni della replicazione virale e della risposta infiammatoria associata all'infezione; in presenza di una CVD questi effetti deleteri vanno ad accentuare una condizione preesistente.<sup>32</sup>

I soggetti diabetici presentano un tasso di mortalità significativamente più elevato (31%) rispetto ai soggetti che non presentano questa malattia (14%). Il diabete incide sulla minor funzionalità dell'immunità innata, in quanto l'iperglicemia ostacola la produzione dell'INF di tipo I e dell'IL-22, citochine implicate rispettivamente nel contrastare l'infezione virale e nel ridurre l'infiammazione cronica. Il diabete, quindi, predispone maggiormente il polmone ai danni tissutali dovuti allo stress ossidativo e all'infiammazione da COVID-19, aumentando così il tasso di mortalità dei pazienti.<sup>32,33</sup>

Altre comorbidità che influiscono negativamente sul decorso della malattia da COVID-19 riguardano le complicazioni a livello respiratorio, come ad esempio l'asma o la malattia polmonare ostruttiva cronica (COPD), oppure il cancro, in particolare in pazienti che assumono chemioterapici o immunosoppressori o affetti da tumori a carico del sistema immunitario, oppure le malattie autoimmuni e le malattie croniche a livello di fegato o reni.<sup>32,33</sup>

## 1.6 Modalità di trasmissione del SARS-CoV-2

L'impatto antropico sugli habitat naturali di pipistrelli e altri animali selvatici rappresenta un rischio concreto per l'emergenza di nuove malattie infettive per l'uomo. Infatti, la deforestazione, l'allevamento e il commercio di animali selvatici possono facilitare salti di specie e la trasmissione di nuovi virus all'uomo, come nel caso di SARS-CoV, MERS-CoV e SARS-CoV-2.<sup>5</sup> In questo contesto trova la sua origine il SARS-CoV-2, come molti *coronavirus* il suo serbatoio animale risiede nei pipistrelli, ma prima di poter essere trasmesso all'uomo probabilmente ha avuto un ospite intermedio, così come è avvenuto per SARS-CoV e MERS-CoV. Dalle analisi di sequenza svolte sulla *spike* dei *coronavirus* di pangolini malesi, questi presentano un solo amminoacido del *Receptor Binding Domain* (RBD) differente rispetto al SARS-CoV-2, si è quindi ipotizzato il ruolo del pangolino come ospite intermedio. In particolare, potrebbe essere avvenuto un evento di ricombinazione virale tra *coronavirus* di pangolino e *coronavirus* di pipistrello.<sup>34</sup> La trasmissione del nuovo SARS-CoV-2 all'uomo si presume sia avvenuta nel *Seafood Market* nella città di Wuhan, dove vengono venduti molti animali vivi.

La principale via di trasmissione del SARS-CoV-2 interumana, similmente ad altri virus respiratori, è quella diretta attraverso l'aerosol o l'inalazione di *droplets* di persone infette. Infatti, le goccioline respiratorie e gli aerosol permangono in sospensione nell'aria anche per alcune ore e possono rappresentare un significativo veicolo di infezione, specialmente in ambienti chiusi e poco ventilati, come ad esempio nei mezzi pubblici e negli uffici.

Si stima che l'esposizione prolungata e ravvicinata, almeno 15 minuti entro i 2 metri circa di distanza, a una persona infetta sia associata a un rischio più elevato di contrarre l'infezione.<sup>28</sup>

Gli individui infetti presintomatici e asintomatici rappresentano una significativa fonte di trasmissione del SARS-CoV-2, in quanto possono diffondere il virus anche in assenza di sintomatologie che potrebbero identificarne la presenza. <sup>11</sup>

La trasmissione di SARS-CoV-2 può iniziare circa 2-3 giorni prima della comparsa dei sintomi e la carica virale nel tratto respiratorio superiore risulta massima intorno al momento della comparsa dei sintomi. <sup>28</sup>

Gli individui possono presentare una carica virale diversa durante l'infezione da SARS-CoV-2, come risultato della diversa suscettibilità e dello stato di immunizzazione individuale, e ciò può avere un impatto sulla trasmissione del virus. Infatti, sono noti pazienti altamente infettivi, denominati *superspreaders*, che nel corso dell'infezione possono infettare un maggior numero di soggetti suscettibili rispetto a individui con minor carica virale. <sup>35</sup>

L'infezione da SARS-CoV-2 può anche avvenire tramite il contatto con superfici contaminate. A seconda del tipo di materiale su cui si deposita, il virus può mantenere la sua infettività per tempi variabili, ad esempio risulta più stabile sulle superfici lisce, può persistere fino a 4 giorni su vetro e banconote e fino a 7 giorni su superfici impermeabili, come l'acciaio inossidabile e la plastica.

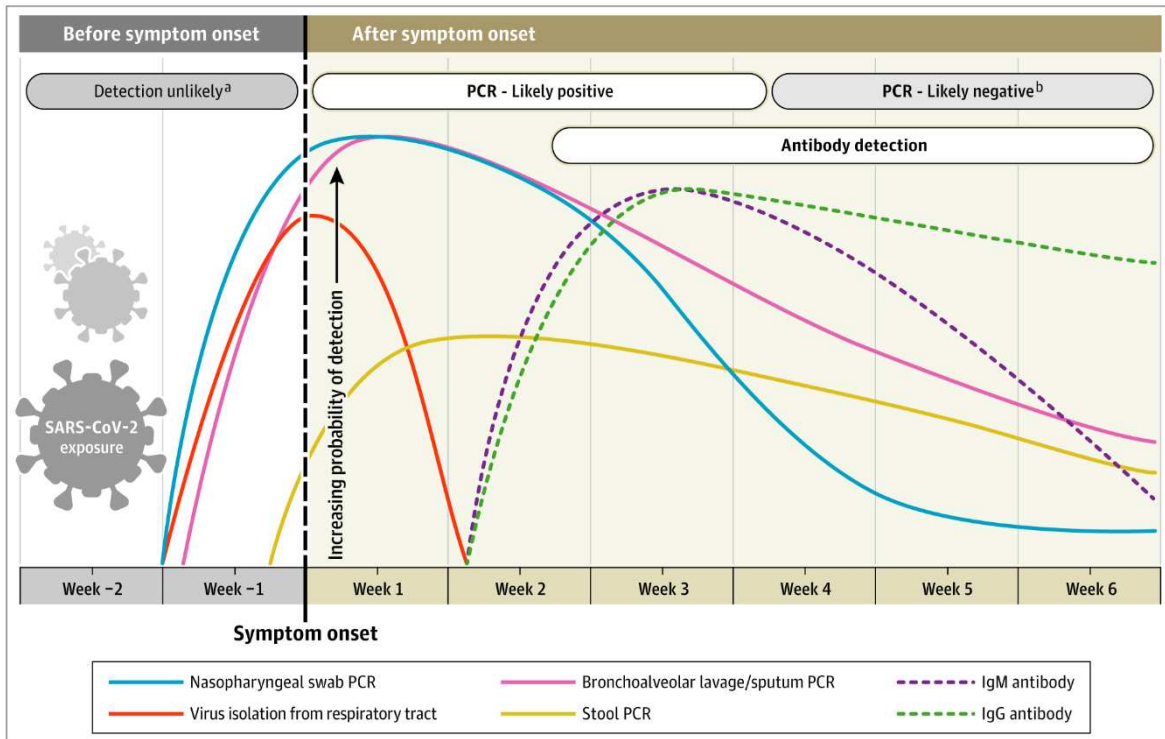
Il SARS-CoV-2 risulta resistente ad un ampio range di pH (3-10), ma è sensibile all'aumentare della temperatura, ai comuni disinfettanti domestici e ai raggi UV. <sup>11</sup>

Le scuole e i locali pubblici rappresentano luoghi in cui possono verificarsi focolai di contagio, ma anche gli ospedali, in quanto queste strutture ospitano spesso un gran numero di persone infette e il virus può circolare facilmente trasmettendosi fra il personale sanitario e i pazienti ricoverati. <sup>11,34</sup>

## 1.7 Diagnosi

La diagnosi dell'infezione da SARS-CoV-2 si basa sulla ricerca di varie molecole virali, quali gli acidi nucleici, come nel caso dei test molecolari, o le proteine, come nei test antigenici.

L'analisi *gold standard* per l'identificazione dell'infezione è di tipo molecolare, mediante la tecnologia *Reverse Transcription - Polymerase Chain Reaction* (RT-PCR), in particolare la *Real Time Polymerase Chain Reaction* (*Real Time-PCR*) o *quantitative PCR* (qPCR) che presenta elevati valori di sensibilità e accuratezza.<sup>36</sup> L'esame è svolto su campioni respiratori, principalmente tamponi nasofaringei. La positività del test varia a seconda del tipo di campione e dalle tempistiche di analisi, con la maggior sensibilità ottenuta da campioni di lavaggi broncoalveolari (93%), seguita da campioni di espettorato (72%), tamponi nasali (63%) e tamponi faringei (32%).<sup>28</sup> La *Real Time PCR* rileva la presenza dell'RNA virale usando come *target* uno o più geni virali a seconda del kit, i più usati sono la *spike* (S), il nucleocapside (N), la RNA polimerasi RNA-dipendente (RdRp), l'*envelope* (E) e la ORF1ab. Nella maggior parte dei soggetti con infezione sintomatica, l'RNA di SARS-CoV-2 risulta rilevabile già dal secondo giorno dopo l'infezione e raggiunge il picco entro la prima settimana dall'insorgenza dei sintomi (Figura 13).<sup>37</sup>



**Figura 13** - Grafico che mostra la stima della variazione nel tempo del rilevamento con test diagnostici dell'infezione da SARS-CoV-2 rispetto all'insorgenza dei sintomi. Fonte: N. Sethuraman et al, 2021. <sup>37</sup>

Altri test per la rilevazione di SARS-CoV-2 sono i test antigenici. Esistono vari kit commerciali, anche in formato immunocromatografico, che analizzano la presenza di due principali antigeni del virus, ossia la proteina *spike* e il nucleocapside. L'attendibilità di questi test e la possibilità di risultati falsi negativi dipendono da vari fattori, quali lo stadio della malattia e la carica virale, ma rappresentano comunque una valida alternativa ai test molecolari in quanto risultano più rapidi e di facile applicabilità al di fuori dei laboratori. <sup>36,38</sup> Infine, ci sono le tecniche basate sulla misurazione degli anticorpi anti-SARS-CoV-2. Queste tecniche non sono applicabili per la diagnosi dell'infezione virale in quanto le risposte anticorpali risultano rilevabili a circa due settimane dall'inizio dell'infezione. Esistono vari test di legame come test di immunofluorescenza, di immunocromatografia, di chemiluminescenza e l'Enzyme-Linked Immunosorbent Assay (ELISA). A seconda dei kit

commerciali, questi test sono mirati verso l'analisi degli anticorpi contro le proteine virali *spike* e nucleocapside e possono misurare sia le immunoglobuline M (IgM), sia le immunoglobuline G (IgG).<sup>36</sup> In particolare, la sieroconversione di IgM e IgG si verifica tra la seconda e la quarta settimana dall'inizio dell'infezione; i livelli di IgM iniziano a diminuire intorno la quinta settimana fino a non essere più rilevabili entro la settima settimana, mentre le IgG persistono oltre la settima settimana.<sup>37</sup>

Tra gli esami sierologici vi è anche il test di neutralizzazione virale che stima, *in vitro*, la capacità neutralizzante degli anticorpi di un soggetto nei confronti del virus. In questo test, il siero, e sue diluizioni scalari, viene incubato con uno *stock* virale e poi aggiunto ad una coltura cellulare di Vero E6 (cellule epiteliali renali di *Clorocebus sp.*). Successivamente ad un'incubazione, si va ad osservare l'effetto citopatico sulle cellule tramite il microscopio ottico e si calcola la diluizione più bassa alla quale non si rileva l'effetto citopatico. L'esame può essere svolto su siero convalescente o post-immunizzazione da vaccino.



## 1.8 Terapia e Vaccini

Per far fronte all'emergenza COVID-19 ad inizio pandemia vennero approvati dall'Agenzia Italiana del Farmaco (AIFA) programmi di uso terapeutico compassionevole di vari farmaci preesistenti a causa della mancanza di piani farmacologici specifici per SARS-CoV-2. Fra questi configuravano la Ribavirina per soluzione inalatoria della società farmaceutica *Bausch Health Companies Inc.* il Solnatide della *APEPTICO Forschung und Entwicklung GmbH*, il Remdesivir della *Gilead Sciences*, il Canakinumab e il Ruxolitinib della *Novartis*.<sup>39</sup>

Attualmente in Italia, i farmaci antivirali autorizzati dall'AIFA sono due: il Veklury (remdesivir) dell'Azienda *Gilead Sciences* e il Paxlovid (nirmatrelvir/ritonavir) dell'Azienda *Pfizer*. Questi due farmaci antivirali agiscono come analoghi nucleotidici andando ad inibire la replicazione di SARS-CoV-2 nelle cellule infettare.

Il Veklury (remdesivir) è un profarmaco analogo dell'adenosina ad ampio spettro e agisce integrandosi nel filamento nascente dell'RNA virale. Viene somministrato per via endovenosa a pazienti con età superiore ai 12 anni e di almeno 40 kg di peso. Questo farmaco inizialmente era rivolto a soggetti presentanti un quadro clinico di polmonite e con necessità di ossigenoterapia supplementare, soprattutto in presenza fattori di rischio predisponenti alla forma severa di COVID-19. Attualmente è anche indicato per adulti non ospedalizzati e non in ossigenoterapia supplementare, ma sempre in presenza di condizioni cliniche predisponenti alla forma severa di COVID-19.

L'altro farmaco approvato in Europa è il Paxlovid (nirmatrelvir/ritonavir). Viene somministrato per via orale e può essere somministrato in adulti che non necessitano di ossigenoterapia supplementare.<sup>36,40</sup>

Altre vie terapeutiche approvate dell' Agenzia Europea per i Medicinali (EMA) per il trattamento dell'infezione di SARS-CoV-2 sono basate sull'uso di anticorpi monoclonali, da soli o in associazione, rivolti contro la proteina *spike*. I farmaci autorizzati sono:

- Il Ronapreve dell'azienda farmaceutica *Regeneron/Roche*, costituito dall'associazione di casirivimab-imdevimab ed è utilizzato per il trattamento e la prevenzione della malattia di COVID-19;
- Il Regkirona dell'azienda farmaceutica *Celltrion Healthcare Hungary Kft*, costituito dal regdanvimab;
- Il Xevudy dell'azienda *GSK*, costituito dal sotrovimab;
- Il Evusheld della *AstraZeneca*, costituito dall'associazione di tixagevimab-cilgavimab ed utilizzato per la profilassi pre-esposizione di COVID-19 e nel trattamento precoce dei pazienti con a rischio di forme gravi della malattia. <sup>41</sup>

Inoltre, l'AIFA raccomanda l'utilizzo di paracetamolo o di Farmaci Antinfiammatori Non Steroidei (FANS) per il trattamento domiciliare del COVID-19.

L'utilizzo di corticosteroidi è indicato soprattutto in soggetti ospedalizzati con malattia grave e non deve essere usato per il trattamento nelle fasi precoci del COVID-19, in quanto può compromettere una buona risposta immunitaria utile a contrastare la replicazione virale.

Infine, l'utilizzo delle eparine a basso peso molecolare è indicato nei pazienti con infezione respiratoria acuta e ridotta mobilità al fine di evitare l'insorgenza di trombi o embolie. <sup>42</sup>

Oltre alla continua ricerca di nuovi trattamenti per contrastare la malattia da COVID-19 e aumentare la sopravvivenza dei pazienti ospedalizzati con sindromi severe, è stata fondamentale la campagna di prevenzione.

Sono state studiate varie tipologie di vaccini per prevenire la malattia grave da SARS-CoV-2 e fra questi troviamo:

- Vaccini a mRNA: si basano sull'RNA messaggero che verrà tradotto dall'organismo nella proteina *spike* di SARS-CoV-2. L'EMA e l'AIFA hanno approvato l'utilizzo di due vaccini a mRNA, il Pfizer mRNA BNT162b2 (Comirnaty) e il COVID-19 Vaccine Moderna mRNA-1273 (Spikevax). Entrambi questi due vaccini veicolano un RNA messaggero modificato con nucleosidi che codifica per la proteina *spike* a lunghezza intera di SARS-CoV-2 stabilizzata nella conformazione pre-fusione e formulata in nanoparticelle lipidiche (LNP). Gli LNP incapsulano e stabilizzano l'mRNA e ne favoriscono l'assorbimento intracellulare. Questi vaccini vengono somministrati per via intramuscolare in due iniezioni iniziali più dosi *booster* e hanno mostrato un'efficacia protettiva superiore al 90%. I vaccini a mRNA presentano diversi vantaggi: sono molto potenti grazie all'azione autoadiuvante dell'RNA, possono essere prodotti rapidamente su larga scala a costi relativamente contenuti, sono considerati molto sicuri in quanto non presentano l'agente patogeno intero e, inoltre, possono venir facilmente aggiornati verso le nuove varianti. Tra gli svantaggi dei vaccini a mRNA vi è la necessità di conservazione a temperature controllate durante la spedizione e lo stoccaggio onde evitarne la degradazione.
- Vaccini vivi attenuati: il virus replica in misura limitata rispetto al *wild-type* ed è geneticamente modificato in modo tale da non causare la malattia.
- Vaccini proteici ricombinanti: possono essere costituiti dalla glicoproteina *spike* a lunghezza intera, dall'RBD o dalle *virus-like*

*particles* (VLP). Le informazioni genetiche codificanti la proteina vengono inserite in un *baculovirus*, utilizzato per l'infezione di colture cellulari in vitro. la produzione della proteina di interesse sarà così a carico delle cellule infettate. La proteina virale viene purificata e successivamente addizionata ad un adiuvante per la produzione del vaccino. L'EMA e l'AIFA hanno autorizzato due vaccini a subunità proteiche, il Nuvaxovid, per soggetti di età pari o superiore a 12 anni, e il VidPrevtyn Beta, per dosi di richiamo in adulti già precedentemente vaccinati con vaccini a mRNA o vettore adenovirale.

- Vaccini a vettore virale: si basano sull'utilizzo di un vettore virale (generalmente un adenovirus) per il trasporto della sequenza codificante la proteina spike del SARS-CoV-2. Il Vaxzevria (ex COVID-19 Vaccine AstraZeneca) e il Jcovden (precedentemente COVID-19 Vaccine Janssen) sono attualmente i vaccini a vettore virale autorizzati da EMA e AIFA per soggetti di età superiore ai 18 anni.
- Vaccini a DNA: sono costituiti da plasmidi modificati in modo da trasportare geni codificanti proteine del virus, generalmente la proteina spike, che verranno prodotte dal soggetto vaccinato.
- Vaccini virali inattivati: il virus viene coltivato in colture cellulari e inattivato chimicamente. Vengono usati in genere per categorie di soggetti a rischio, come gli immunocompromessi, e rappresentano soluzioni molto sicure e di facile trasporto. In Italia, l'EMA e l'AIFA hanno recentemente approvato l'utilizzo di un vaccino a virus inattivato, il vaccino COVID-19 (inattivato, adiuvato) Valneva e viene somministrato negli adulti di età compresa tra 18 e 50 anni. <sup>36,43,44</sup>

## 1.9 Varianti virali di SARS-CoV-2

Tutti i virus, in particolare i virus a RNA, durante la loro replicazione tendono ad accumulare mutazioni. Alcune mutazioni possono risultare vantaggiose per il virus conferendo una maggior trasmissibilità, una maggior patogenicità o la possibilità di evadere la risposta immunitaria dell'ospite. Per queste ragioni, il monitoraggio delle varianti virali in circolazione risulta una strategia efficace per la prevenzione e per stabilire le procedure d'intervento più efficaci nel contesto della sanità pubblica.

SARS-CoV-2 si stima presenti un tasso di mutazione di circa  $1 \cdot 10^{-6}$  -  $2 \cdot 10^{-6}$  mutazioni per nucleotide per ciclo replicativo. Questo tasso di mutazione risulta inferiore rispetto ai valori riscontrati in altri virus a RNA, ma è coerente con la presenza della proteina nsp14 che possiede attività esonucleasica, ovvero di *proofreading*, che riduce il numero di mutazioni introdotte.<sup>45</sup>

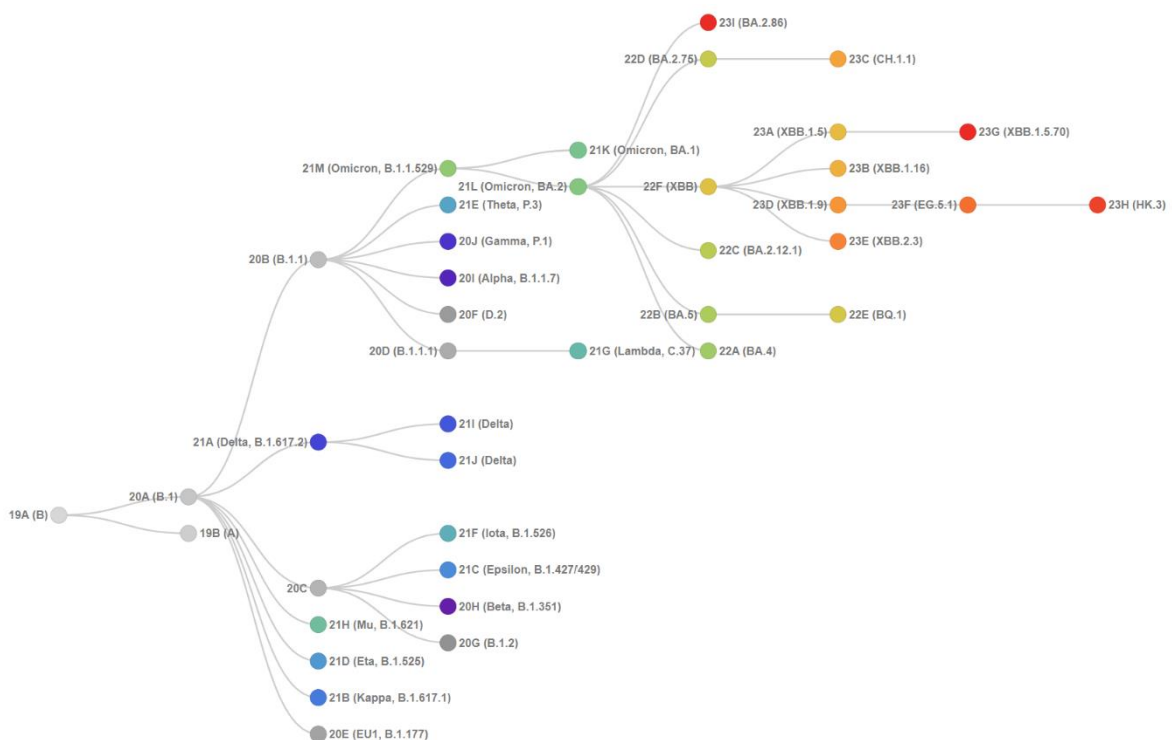
Una forte influenza sul tropismo virale, sulla patogenesi e sulla trasmissibilità del virus è data dalle mutazioni riguardanti la proteina *spike*. In particolare, già nei primi mesi della pandemia, venne registrata la prima mutazione a carico di questa proteina, ossia la sostituzione non-sinonima D614G. La D614G si diffuse a tal punto da risultare in oltre il 74% di tutte le sequenze entro giugno 2020, suggerendo il suo grande vantaggio in termini di fitness per SARS-CoV-2, per poi essere presente in tutte le varianti che si sono originate in seguito. Infatti, questa mutazione ha migliorato la capacità replicativa del virus nel tratto respiratorio superiore, aumentando l'infettività del virione e quindi la sua diffusione nella popolazione.<sup>46</sup>

Inizialmente, le nuove varianti di SARS-CoV-2 venivano classificate secondo la loro origine geografica, ma successivamente questo tipo di denominazione venne sostituita con una combinazione di lettere e numeri secondo la nomenclatura del *database Pangolin*. Questa nomenclatura si basa

sull'assegnazione del *lineage* più probabile a una sequenza genomica di SARS-CoV-2 attraverso il *software* di *Pangolin* (*Phylogenetic Assignment of Named Global Outbreak LINEage*, <https://pangolin.cog-uk.io/>) e viene integrata insieme ad un'unica lettera greca, secondo la denominazione stabilita dall'Organizzazione Mondiale della Sanità (OMS). Altre piattaforme utili per la classificazione delle varianti di SARS-CoV-2 sono *Nextstrain* (<https://nextstrain.org/>), un progetto *open source* che fornisce strumenti di analisi e monitoraggio di vari patogeni, e *GISAID* (*Global Initiative on Sharing All Influenza Data*) un *database* che promuove la condivisione di dati riguardanti vari agenti patogeni.<sup>47,48</sup>

La classificazione di *Pangolin* è attualmente la più usata, si basa sugli studi di Rambaut et al. e suddivide le varianti di SARS-CoV-2 in due lignaggi principali definiti con le lettere A e B. I lignaggi A sono quelli correlati alla variante *Wuhan/WH04/2020*, mentre i lignaggi B sono associati alla variante *Wuhan-Hu-1*, da cui sono derivate tutte le sequenze circolate nel mondo. Le varianti che discendono da questi due lignaggi vengono definite da una lettera seguita da un numero. Per l'assegnazione del *lineage* devono essere rispettate delle condizioni, ossia il nuovo *lineage* deve presentare uno o più nucleotidi differenti rispetto al *lineage* ancestrale, deve comprendere almeno cinque genomi completi per il 95%, i genomi appartenenti allo stesso *lineage* devono mostrare almeno un cambiamento nucleotidico condiviso tra loro e, infine, il valore di *bootstrap* per il nuovo nodo del *lineage* deve essere maggiore del 70%. Una volta identificati questi lignaggi possono essere usati come lignaggi ancestrali per le varianti emergenti, ad esempio dal *lineage* A.1 può derivare il *lineage* A.1.1.1; l'assegnazione dei lignaggi discendenti può avere un massimo di tre sottolivelli oltre il quale verrà assegnata una lettera, ad esempio il *lineage* A.1.1.1.1 diventa C.1.<sup>48,49</sup>

Secondo la classificazione di *Nextstrain*, invece, le varianti di SARS-CoV-2 vengono racchiuse in *clade*. I *clade* si identificano con numeri e lettere, dove i numeri rappresentano l'anno in cui è stata identificata per la prima volta la variante, mentre le lettere vengono assegnate progressivamente secondo l'ordine alfabetico. Per definire un nuovo *clade* la variante deve raggiungere una frequenza globale del 20% in un dato momento e deve presentare almeno due nuove mutazioni rispetto al *clade* progenitore.<sup>48</sup> (Figura 14)



**Figura 14** - Relazioni filogenetiche dei clade di SARS-CoV-2 come definite da Nextstrain. Fonte: <https://covariants.org/variants>.

Secondo l'OMS le varianti di SARS-CoV-2 vengono suddivise in tre gruppi sulla base dell'identificazione del rischio più o meno elevato per la salute pubblica: le VOC (*variants of concern*), le VOI (*variants of interest*) e le VUM (*variants under monitoring*). Le VOC sono quelle varianti che presentano una maggiore

trasmissibilità e virulenza, una maggior gravità dei sintomi della malattia (ad esempio aumentando il tasso di ospedalizzazione e dei decessi) e una ridotta efficienza degli anticorpi neutralizzanti, oltre a una ridotta efficacia terapeutica e della protezione vaccinale. Mentre le VOI sono rappresentate dalle varianti con mutazioni che modificano l'affinità di legame per il recettore, che hanno una ridotta neutralizzazione da parte di anticorpi di precedenti infezioni o da vaccinazione, che hanno una ridotta efficacia terapeutica e di diagnosi.<sup>48</sup>

Fra le VOC vi sono state la variante Alfa, la Beta, la Gamma, la Delta e la Omicron, mentre fra le VOI vi sono le varianti Lambda e Mu.

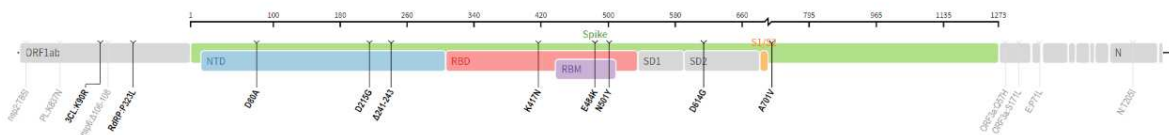
Tutte le VOC e VOI, in seguito ad una eventuale diminuzione della circolazione, a fronte della comparsa di una nuova variante, vengono riclassificate come “*De-escalated variants*”.<sup>50</sup>

Un altro fattore che incide sulla variabilità di SARS-CoV-2 è rappresentato dalla ricombinazione, un meccanismo evolutivo molto comune nei *betacoronavirus* e che avviene in seguito alla coinfezione di due virus geneticamente distinti in un singolo ospite, con la conseguente formazione di un ibrido.<sup>45,47</sup> Le varianti originate da un evento di ricombinazione vengono denominate, secondo la nomenclatura *Pangolin*, con il prefisso “X” seguito dal suffisso secondo i criteri di denominazione usati per altri *lineage*. Di particolare interesse è stata la variante ricombinante XBB (*clade* 22F), identifica intorno la metà del 2022. Questa variante deriva dalla ricombinazione di due varianti Omicron, i *lineages* BJ.1 (BA.2.10.1.1; *clade* 21L) e BM.1.1.1 (BA.2.75.3.1.1.1; *clade* 22D). Verso la fine dell'anno 2022, le varianti XBB e le loro discendenti, quali la XBB.1.5 (*clade* 23A), si diffusero rapidamente in tutto il mondo e destarono particolare interesse per via della loro aumentata trasmissibilità e capacità di eludere la risposta immunitaria dell'ospite.<sup>51,52</sup>



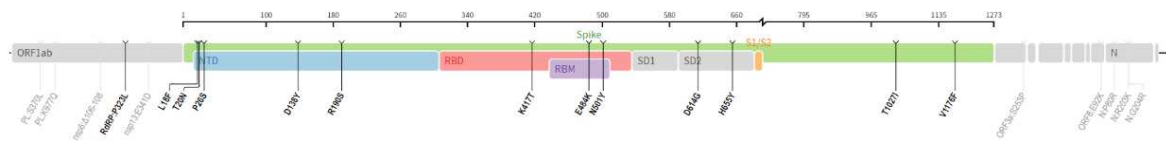


e la A701V che hanno comportato un aumento dell'affinità di legame per il recettore ACE2 (Figura 17). La mutazione E484K, inoltre, è associata a una maggior capacità di evasione di questa variante della risposta immunitaria dell'ospite e a una minor attività neutralizzante da parte dei vaccini. Questa variante ha avuto un forte impatto anche tra soggetti giovani e sani, ed ha anche comportato una diminuzione nell'efficacia protettiva dei vaccini allora in uso.<sup>14,48,53</sup> (Figura 16)



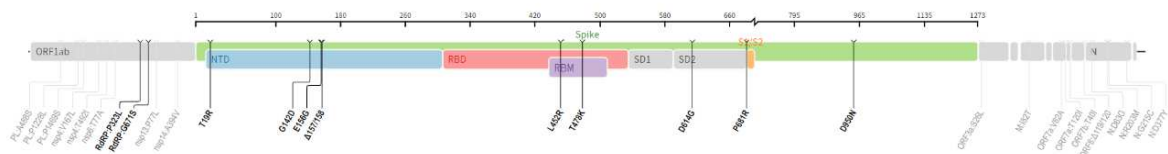
**Figura 16 -** Mutazioni della proteina spike della variante Beta (B.1.351). Fonte: <https://covdb.stanford.edu/variants/beta/>.

La variante Gamma, *lineage* P.1, *clade* 20J/501Y.V3, venne identificata gennaio 2021 in Giappone e in Brasile. Questo *lineage* comprende 10 mutazioni sulla proteina *spike* (di cui due sono la K417N e la N501Y), 8 mutazioni riguardanti l'ORF1ab, 2 mutazioni sono presenti nella ORF8, di cui una corrisponde a un'inserzione, e un'ulteriore mutazione è comparsa nel gene N (Figura 18). La variante Gamma si è dimostrata resistente alla neutralizzazione da parte di plasma convalescente e da vaccinazione e, inoltre, sono stati registrati diversi casi di reinfezione da parte della stessa variante.<sup>14,53</sup> (Figura 17)



**Figura 17** - Mutazioni della proteina spike della variante Gamma (P.1). Fonte: <https://covdb.stanford.edu/variants/gamma/>.

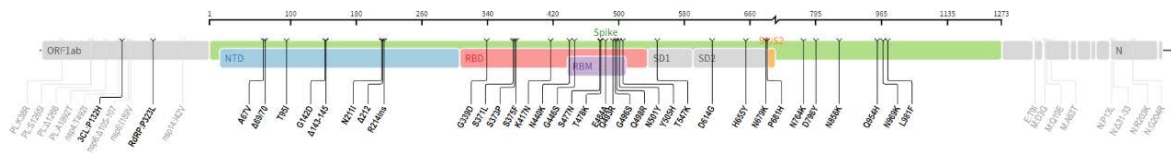
La variante Delta, *lineage* B.1.617.2, *clade* 21A, 21I e 21J, venne identificata in India e si diffuse rapidamente a livello globale nel 2021. Le mutazioni riscontrabili di questa variante sono T19R, G142D, R158G, L452R, T478K, D614G, P681R e D950N, oltre a due delezioni, E156del, F157del, tutte presenti nella proteina spike (Figura 19). In particolare, la mutazione P681R è vicina al sito di taglio della furina ed è implicata nell'aumento della capacità infettiva del virus. Questa variante è risultata altamente trasmissibile e si diffuse sia tra soggetti non vaccinati e vaccinati. La ridotta efficacia delle vaccinazioni è in parte imputabile alle mutazioni L452R e T478K, localizzate nel RBD della proteina *spike*.<sup>14,48,53</sup> (Figura 18)



**Figura 18** - Mutazioni della proteina spike della variante Delta (B.1.617.2). Fonte: <https://covdb.stanford.edu/variants/delta/>.

La variante Omicron, *lineage* B.1.1.529, *clade* 21M, è stata identificata per la prima volta nel novembre 2021 in Sud Africa. Questa variante presenta un elevato numero di mutazioni, 50 in tutto il genoma di cui 26-32 mutazioni nella proteina *spike*. La variante Omicron si è evoluta in vari *sublineages*, BA.1

(*clade* 21K), BA.2 (*clade* 21L), BA.4 (*clade* 22A) e BA.5 (*clade* 22B); BA.1 e BA.2 hanno circolato in modo predominante soprattutto nei primi mesi del 2022, mentre BA.4 e BA.5 hanno rapidamente rimpiazzato le varianti precedenti verso aprile 2022, diventando predominanti nei mesi a seguire. I vari *lineages* della variante Omicron condividono 20 mutazioni identiche sulla proteina *spike* (G142D, G339D, S373P, S375F, K417N, N440K, S477N, T478K, E484A, Q498R, N501Y, Y505H, D614G, H655Y, N679K, P681H, N764K, D796Y, Q954H, N969K). In particolare, le mutazioni N440K, S477N, T478K, E484A, Q493R, N501Y, Q498R e Y505H presenti nell'RBD aumentano l'affinità di legame per il recettore ACE2 umano, incrementando l'infettività delle varianti Omicron (Figura 20). Le mutazioni non sono solo limitate alla proteina *spike*, ma sono presenti anche nella ORF1ab, nella proteina N, nella proteina M e nella proteina E. Le varianti Omicron hanno destato particolare attenzione in quanto presentavano una ridotta attività neutralizzante da parte dei vaccini in commercio al momento della comparsa e, inoltre le terapie basate sugli anticorpi monoclonali (mAb) anti-*spike* risultavano poco efficaci verso la variante BA.2, BA.4 e BA.5. Il numero elevato di mutazioni nella proteina *spike* è stato associato a una più alto tasso di trasmissibilità, di resistenza immunitaria e ad un aumentato del rischio di reinfezione.<sup>48,54,55</sup> (Figura 19)



**Figura 19** - Mutazioni della proteina *spike* della variante Omicron (BA.1). Fonte: [https://covdb.stanford.edu/variants/omicron\\_ba\\_1\\_3/](https://covdb.stanford.edu/variants/omicron_ba_1_3/).

## 1.10 Epidemiologia di SARS-CoV-2

I primi casi di COVID-19 vennero riportati nel dicembre 2019 a Wuhan, nella provincia cinese di Hubei. Vennero registrati diversi contagi in pochi giorni, la maggior parte dei quali correlati a un'esposizione avvenuta nel mercato cittadino dove venivano venduti generi alimentari e animali vivi.

Nei primi mesi del 2020, il SARS-CoV-2 si diffuse inizialmente in Asia e in seguito a livello globale, finché nel marzo 2020 l'OMS dichiarò lo stato di pandemia da COVID-19.

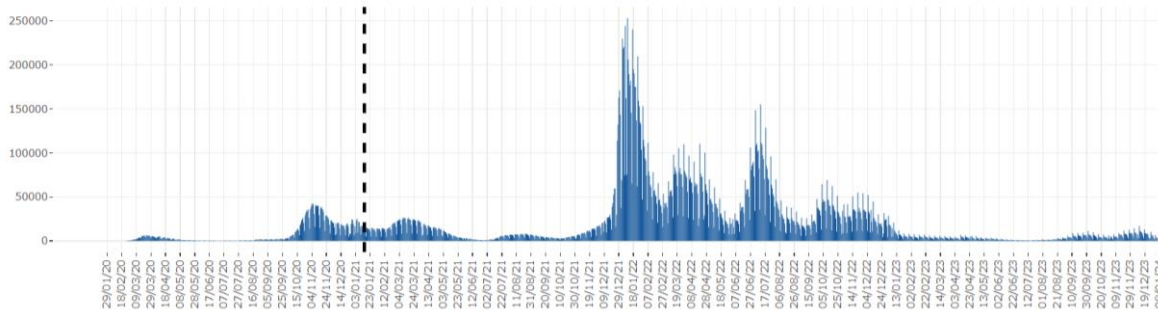
Per far fronte al crescente numero di casi, i vari governi adottarono diversi provvedimenti per il contenimento dei contagi e la prevenzione, fra i quali *lockdown*, distanziamento sociale e l'uso di mascherine facciali.<sup>13,56</sup>

In Italia, i primi casi di COVID-19 vennero registrati a fine febbraio 2020 e rapidamente il virus si diffuse nelle diverse regioni italiane, dando origine alla prima ondata epidemica da SARS-CoV-2. L'Italia è stata fra i primi paesi occidentali ad adottare le misure di *lockdown* per contenere il crescente propagarsi dei contagi.<sup>57</sup>

In questo scenario, il virus si diffuse rapidamente in tutto il mondo e l'elevato numero di casi ha permesso la comparsa di nuove varianti di SARS-CoV-2, più trasmissibili delle precedenti e potenzialmente capaci di eludere la risposta immunitaria dell'ospite. La prima variante ad essere stata identificata fu la Alpha, che si diffuse in Italia fra dicembre 2020 e gennaio 2021, seguita poi dalle altre varianti, fra cui la Delta e la Omicron, riscontrate nel nostro paese rispettivamente nel giugno e nel dicembre 2021 (Figura 20).<sup>56,58</sup>

Il 27 dicembre 2020, ha avuto inizio la campagna vaccinale in Italia, inizialmente rivolta alle persone ad alto rischio di infezione o di malattia severa, quali operatori sanitari e soggetti con età maggiore di 80 anni, poi estesa al resto

della popolazione. Secondo un report dell'ISS, nel gennaio 2022 la copertura vaccinale in Italia era dell'84% e si stima abbia contribuito ad evitare circa 8 milioni di casi, oltre 500.000 ricoveri, più di 55.000 ricoveri in terapia intensiva e circa 150.000 decessi.<sup>59</sup>



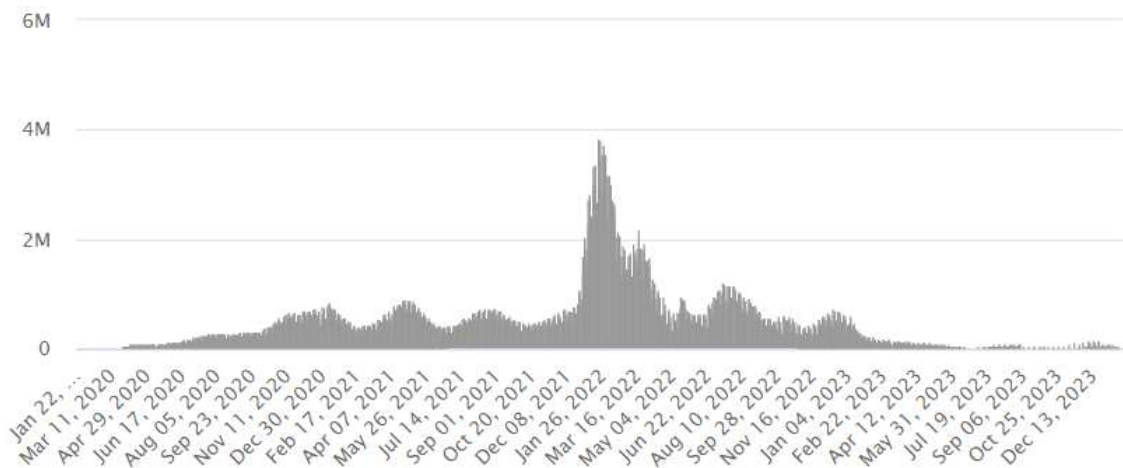
**Figura 20** – Curva epidemica in Italia per incidenza di casi di SARS-CoV-2 diagnosticati, dal gennaio 2020 al gennaio 2024. Fonte: <https://www.epicentro.iss.it/coronavirus/sars-cov-2-dashboard>.

In Italia, ad oggi, l'età media dei casi riscontrati è di 45 anni, mentre non c'è una differenza significativa nell'incidenza nei i due sessi (53,6% nelle femmine e il 46,4% nei maschi).

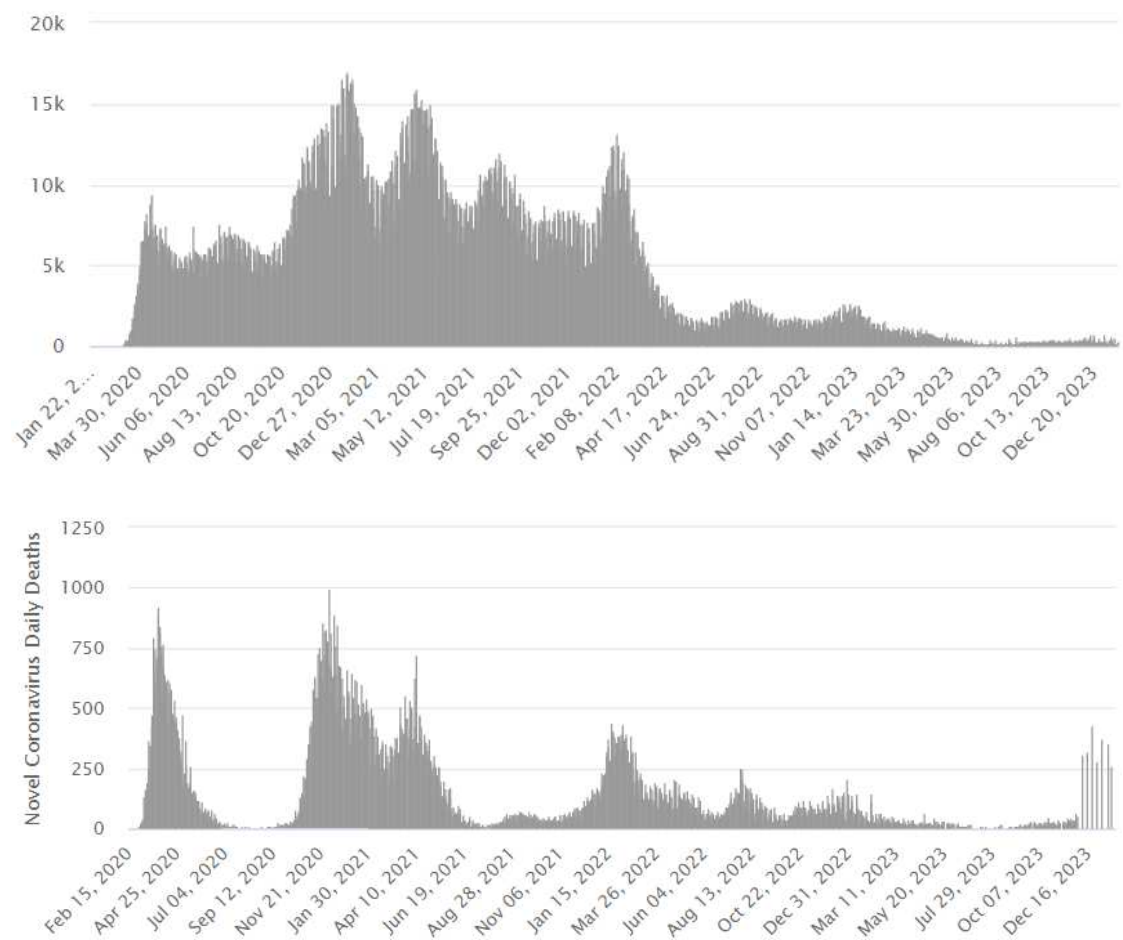
Il tasso di letalità (CFR) in Italia è stimato intorno allo 0,7%, con la maggior percentuale riscontrata fra gli *over 90* (CFR 11,49%), seguita dalla fascia di età di 80-89 anni (CFR 5,82%).

Fra gli operatori sanitari, categoria di persone particolarmente colpita dalla pandemia di COVID-19, i casi segnalati in Italia sono circa 508.000.<sup>60</sup>

Attualmente, dall'inizio dell'emergenza fino agli inizi di gennaio 2024, sono stati registrati più di 701 milioni di casi nel mondo (Figura 21), di cui oltre 26 milioni in Italia, e i decessi registrati hanno superato i 6 milioni, di cui circa 195,000 in Italia (Figura 22).<sup>61</sup>



**Figura 21** - Numero totale di casi giornalieri di COVID-19 registrati nel mondo da inizio pandemia fino a fine 2023. Fonte: <https://www.worldometers.info/coronavirus/>



**Figura 22** - Numero di decessi giornalieri causati dal COVID-19 nel mondo (in alto) e in Italia (in basso) da inizio pandemia fino a fine anno 2023. Fonte: <https://www.worldometers.info/coronavirus/country/italy/>

## 2. SCOPO DELLO STUDIO

Dall'inizio della pandemia di COVID-19 ad oggi, si è assistito alla progressiva comparsa di varianti di SARS-CoV-2 che hanno rappresentato una questione di interesse per la salute pubblica. In particolare, come nel caso delle *Variants Of Concern* (VOC), le mutazioni che hanno accompagnato la circolazione di queste varianti hanno determinato la capacità del virus di trasmettersi e infettare con maggiore efficacia, di aumentare la sua patogenicità, di ridurre l'efficienza delle terapie e di diminuire l'efficacia della protezione vaccinale atta a contenerlo. È quindi risultata chiara la necessità di un monitoraggio costante delle varianti in circolazione in un dato territorio al fronte di controllare e prevenire potenziali rischi sanitari. Per queste ragioni l'Istituto Superiore di Sanità (ISS) ha istituito un programma di monitoraggio mensile delle varianti in circolazione sul territorio italiano. Questo programma coinvolge un elevato numero di laboratori, distribuiti nelle varie regioni del Paese, che si adoperano per il sequenziamento mensile di un adeguato numero di campioni di SARS-CoV-2, al fine di ottenere informazioni di carattere epidemiologico.

Lo scopo di questo studio è stato quello di raccogliere e monitorare informazioni riguardanti le varianti di SARS-CoV-2 circolanti nella regione Marche nell'anno 2023. I dati sono stati raccolti in occasione del programma di monitoraggio mensile istituito dall'ISS, al fine di tracciare un profilo chiaro delle varianti che si sono susseguite e delle loro caratteristiche genetiche.



## **3. MATERIALI E METODI**

### **3.1 Popolazione oggetto di studio**

In questo studio sono stati presi in esame 353 campioni provenienti da pazienti risultati positivi al SARS-CoV-2 mediante analisi di RT-PCR o antigenico nella regione Marche nel periodo compreso fra gennaio 2023 a dicembre 2023, in occasione del monitoraggio mensile richiesto dall'Istituto Superiore di Sanità (ISS). Il lavoro è stato svolto nel laboratorio della SOSD Virologia dell'Azienda Ospedaliero Universitaria delle Marche.

### **3.2 Sistemi diagnostici utilizzati e selezione dei campioni positivi**

I campioni biologici, costituiti da tamponi faringei o da lavaggi broncoalveolari (BAL) provenienti da pazienti con sospetta infezione da SARS-CoV-2, sono stati analizzati mediante test molecolari e antigenici. Per quanto riguarda l'analisi di tipo molecolare, i campioni sono stati sottoposti ad amplificazione genica mediante kit *Alinity m SARS-CoV-2 assay*, indicato per lo strumento ALINITY m (Abbott) e, tra quelli risultati positivi, sono stati scelti i campioni con più alta carica virale. Lo strumento ALINITY m svolge una RT-PCR *Real-Time* utile a rilevare l'eventuale presenza dell'acido nucleico del virus; il kit ha

come bersaglio due geni virali, la RdRp e il gene N, oltre a un terzo gene appartenente al vegetale *Cucurbita pepo* e che ha lo scopo di controllo interno. La presenza dell'RNA di SARS-CoV-2 nel campione porta all'emissione di fluorescenza che viene rilevata dallo strumento. Quest'ultimo restituisce un valore di ciclo soglia (Ct) che corrisponde al numero di cicli di amplificazione in cui si la fluorescenza ha superato una data *threshold*. Il valore Ct è inversamente proporzionale alla quantità di acido nucleico presente nel campione in esame.

L'analisi di tipo antigenico dei campioni biologici è stata svolta mediante lo strumento Cobas e 402 (Roche) e con l'utilizzo del kit *Elecsys SARS-CoV-2 Antigen*. Questo kit è mirato verso il riconoscimento del gene N di SARS-CoV-2, l'analisi è basata sull' *electro-chemiluminescence immunoassay* (ECLIA) il cui principio è quello del *double-antibody sandwich*. Il test è di tipo qualitativo e la luminescenza prodotta viene rilevata dallo strumento in misura della concentrazione antigenica, ma il valore che si ottiene non va inteso come correlazione diretta della concentrazione, in quanto si tratta di un test non calibrato.

### **3.3 Estrazione automatizzata dell'RNA virale per la genotipizzazione**

L'estrazione dell'RNA di SARS-CoV-2 è svolta mediante lo strumento InGenius (ELITech), utilizzando il kit ELITE InGenius<sup>®</sup> SP 200, contenente i reagenti di estrazione. Questa metodica si basa sulla Magtration<sup>®</sup> Technology, dove il campione viene sottoposto a lisi cellulare e gli acidi nucleici vengono adsorbiti a delle biglie magnetiche; queste ultime, tramite separazione

magnetica, hanno lo scopo di trattenere l'acido nucleico durante le successive fasi di lavaggio e purificazione. La procedura è completamente automatizzata, permette l'estrazione in contemporanea di 12 campioni, partendo da un volume iniziale di 200 µl di campione madre per ottenere 50 µl di eluato.

### **3.4 Retrotrascrizione dell'RNA virale per la genotipizzazione**

L'RNA di SARS-CoV-2 è stato retrotrascritto mediante kit *LunaScript*<sup>®</sup> *RT SuperMix* (New England BioLabs). La retrotrascrizione avviene mediante la trascrittasi inversa Luna e con l'utilizzo di *random primers*. La *mastermix*, inoltre, contiene l'inibitore della RNasi murino per impedire la degradazione dell'RNA *template* e un colorante blu con funzione di indicatore visivo. La miscela per la reazione di RT viene preparata diluendo 2,4 µl di *LunaScript*<sup>®</sup> *RT SuperMix* in 3,6 µl di H<sub>2</sub>O per singolo campione. In ogni pozzetto della piastra di lavoro vengono dispensati 6 µl di *mix* RT, a cui successivamente vengono aggiunti 6 µl di campione da retrotrascrivere. La preparazione della miscela di reazione e la dispensazione degli eluati dei campioni vengono svolte entrambe mantenendo la piastra a 4°C mediante l'utilizzo dell'*epMotion*<sup>®</sup> 5075t (Eppendorf), uno strumento per la manipolazione automatizzata di microfluidica. La reazione di retrotrascrizione è stata svolta nel termociclatore *SimpliAmp* (ThermoFisher Scientific) con il seguente profilo termico:

	<i>Temperatura</i>	<i>Durata</i>
<i>Primer annealing</i>	25°C	2 minuti
<i>Reverse transcription</i>	55°C	30 minuti
<i>Heat inactivation</i>	95°C	1 minuto

### 3.5 Preparazione automatizzata delle librerie di template per il sequenziamento

La *library* per il sequenziamento è stata preparata utilizzando lo strumento epMotion® 5075t (Eppendorf), che permette la manipolazione automatizzata delle soluzioni, riducendo le tempistiche e garantendo la riproducibilità della procedura. Il kit utilizzato presso il nostro laboratorio è il *EasySeq™ SARS-CoV-2 WGS Library Prep* (NimaGen) da 96 reazioni, che permette il sequenziamento dell'intero genoma di SARS-CoV-2.

Il kit si compone dei seguenti reattivi:

- Piaste *IDX96-U01* da 96 pozzetti contenenti gli index per il riconoscimento dei campioni;
- *RC-PCR SARS-CoV-2 probe pool A* e *pool B*;
- *PCR 2x Hotstart HiFi Mastermix*;
- *AmpliClean Bead Clean-up* per la purificazione;
- *RC-PCR Probe Dilution Buffer Plus*;
- *RC-PCR Low-TE Buffer*.

Successivamente alla fase di retrotrascrizione dei campioni, si procede con la preparazione delle due *mix* (A e B) necessarie per la reazione di *Reverse*

*Complement Polymerase Chain Reaction (RC-PCR)*. La *mix* per singola reazione viene preparata aggiungendo:

<i>RC-PCR SARS-CoV-2 probe pool A e pool B</i> (sonda <i>panel A</i> per la <i>mix A</i> e sonda <i>panel B</i> per la <i>mix B</i> )	0,2 $\mu$ l
<i>RC-PCR Probe Dilution Buffer Plus</i>	0,8 $\mu$ l
<i>PCR 2x Hotstart HiFi Mastermix</i>	10 $\mu$ l
Acqua	4 $\mu$ l

A questo punto, in ogni pozzetto nelle due piastre, contrassegnate come piastra A e piastra B, vengono dispensati rispettivamente 15  $\mu$ l di mix A e B. Sul fondo di ogni pozzetto sono presenti gli *Unique Dual Indexes (UDI)*, ossia brevi sequenze nucleotidiche di 8 bp che hanno la funzione identificativa del campione, e le sequenze adattatrici P5 e P7 che sono complementari agli oligonucleotidi presenti sulla *flow cell* che serviranno per l'ancoraggio delle sequenze alla stessa.

A questo punto, in ogni pozzetto delle piastre vengono dispensati 5  $\mu$ l di campione e si procede con la reazione di RC-PCR nel termociclatore SimpliAmp thermal cycler (ThermoFisher Scientific) secondo il seguente programma termociclico:

<i>Temperatura</i>	<i>Durata</i>	<i>Tasso di crescita</i>	<i>Cicli</i>
98°C	2 minuti	N/A	1x
98°C 80°C 58°C 72°C	10 secondi 1 secondo 10 minuti 2 minuti	MAX MAX 0,1°C/secondo MAX	1x
95°C 80°C 62°C 72°C	10 secondi 1 secondo 90 minuti 2 minuti	MAX MAX 0,1°C/secondo MAX	2x
95°C 80°C 62°C 72°C	10 secondi 1 secondo 2 minuti 1 minuti	MAX MAX 0,5°C/secondo MAX	40x

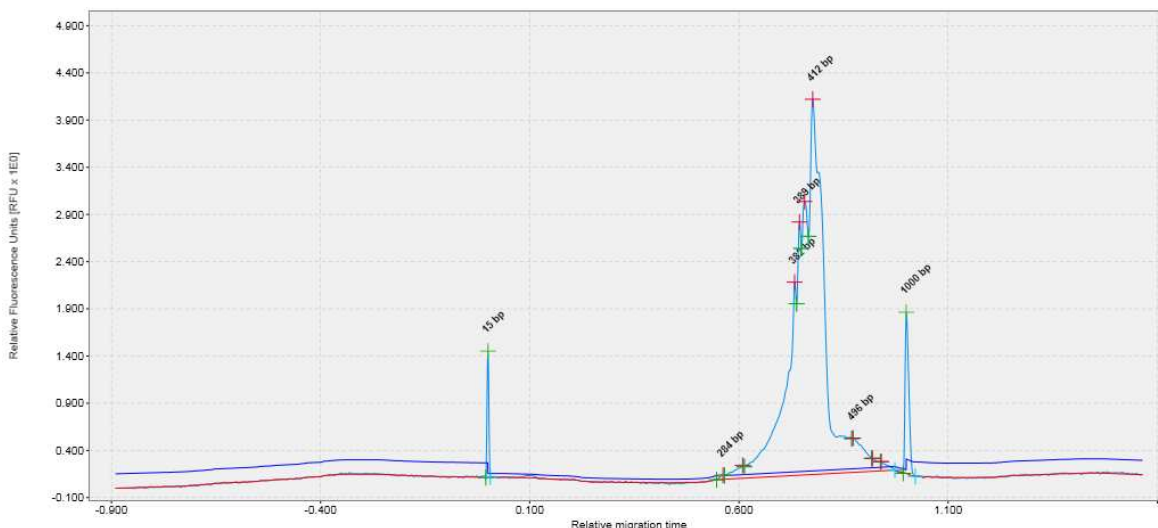
Questa fase richiede circa 6-7 ore e generalmente è effettuata *overnight*.

Successivamente, si procede con il *pooling* di tutti i campioni, aggiungendo 5 µl di ogni pozzetto di entrambe le piastre in due provette da 1,5 ml denominate rispettivamente *pool A* e *pool B*.

Per procedere con la purificazione, vengono spostati 40 µl di entrambi i *pool* all'interno di due pozzetti di una piastra da 96 *wells*. La purificazione viene eseguita in due *step* identici e consequenziali e si basa sull'utilizzo di biglie magnetiche (*AmpliClean Bead Clean-up*), di etanolo al 75% e di *RC-PCR Low-TE Buffer* per l'eluizione finale.

Il *pool A* e il *pool B* purificati vengono quantificati mediante il fluorimetro Qubit (Invitrogen ThermoFisher), utilizzando il kit *Qubit<sup>TM</sup> 1x dsDNA HS Assay*. La quantificazione viene svolta sia sul *tal quale*, sia diluendo la soluzione 1:5 e i valori ottenuti vengono inseriti nel foglio di calcolo *SARS-CoV-2 WGS Calculator* fornito dalla NimaGen, così da poter procedere con la normalizzazione dei due *pool*. Infine, i due *pool* vengono uniti in un'unica provetta ottenendo così la *library* finale di 9 pM.

Per l'analisi della dimensione dei frammenti di cDNA ottenuti è stata utilizzata la piattaforma *QIAxcel Advanced System* (QIAGEN), in cui si visualizza dove ricade il picco elettroforetico che definisce la distribuzione dimensionale dei frammenti amplificati (Figura 23).



**Figura 23** – Esempio di picco elettroforetico ottenuto con il *QIAxcel Advanced System*. Nell'esempio, la maggior parte dei frammenti di cDNA presenti nella library sono delle dimensioni di circa 412 bp.

### 3.6 Next Generation Sequencing (NGS)

Il sequenziamento del genoma di SARS-CoV-2 è stato svolto mediante tecnologia *Next Generation Sequencing* (NGS), che consente il sequenziamento in parallelo di milioni di frammenti di DNA. In particolare, è stata utilizzata la piattaforma *MiSeq System* (Illumina) con il *MiSeq<sup>®</sup> Reagent Kits v2* da 300 cicli, utilizzando il protocollo 2x150 cicli *paired-end*, che permette la lettura del filamento senso e antisenso. La tecnologia Illumina è caratterizzata dall'utilizzo della *flow cell*, un supporto costituito da otto corsie dove si formeranno i *clusters* dei frammenti di DNA. In queste corsie sono presenti le sequenze oligonucleotidiche complementari alle sequenze adattatrici,

legate ai frammenti di DNA durante la preparazione della *library*, così da permetterne l'ancoraggio.

Il processo di sequenziamento inizia con l'amplificazione dei frammenti ancorati alla *flow cell* mediante *bridge amplification*, dove l'estremità libera del filamento di DNA presente sulla *flow cell* forma un arco e si va a ibridare con una sequenza adattatrice vicina presente sulla superficie del supporto. La sequenza adattatrice sul supporto svolge così la funzione di primer e il filamento viene amplificato dalla DNA polimerasi. Il *double-strand DNA* (dsDNA) che si è formato viene denaturato e i *single-strand DNA* (ssDNA) sono sottoposti a numerosi cicli di *bridge amplification*, formando così *clusters* di prodotti di amplificazione.

A seguito della formazione dei *clusters*, vengono eliminati tutti i filamenti *reverse* mediante taglio chimico sull'adattatore corrispondente, così ottenendo *clusters* di soli filamenti *forward*.

Successivamente avviene la fase di sequenziamento vera e propria, mediante la tecnica *Sequencing By Synthesis* (SBS). Nella SBS l'acquisizione della lettura nucleotidica avviene durante la sintesi del filamento di ssDNA; il sistema utilizza un primer complementare alla sequenza adattatrice e la sintesi avviene mediante incorporazione di nucleotidi marcati con quattro fluorofori differenti (uno per ogni base). Quando un nucleotide viene incorporato nel filamento nascente, i nucleotidi non legati vengono lavati via e lo strumento rileva la fluorescenza emessa. Il blocco della sintesi, incorporato nei nucleotidi marcatori, viene rimosso e la reazione è libera di procedere. A fine processo si ottiene una "mappatura" di tutte le fluorescenze ottenute durante il sequenziamento per ogni *cluster* e, mediante algoritmi specifici dello strumento, vengono calcolati i *quality scores*, ossia i punteggi che misurano la probabilità di errore per ogni base e che quindi fornisce un valore



dell'accuratezza della lettura. Ciò che si ottiene dall'elaborazione dei dati ottenuti sono le *reads*, ossia le letture di singole sequenze, mentre la *coverage* è definita dal numero di sequenze che si sovrappongono in ogni determinato segmento del singolo campione. Lo strumento fornisce l'insieme di questi dati in un file di testo FastQ che racchiude tutte le *reads* di un singolo campione.

### 3.7 Analisi bioinformatica dei dati

I FastQ generati dal sequenziamento sono stati analizzati mediante la piattaforma *Italian-COVID-19-Genomic* (I-Co-Gen) di IRIDA-ARIES (<https://irida.iss.it/>). Questa piattaforma rappresenta uno strumento utile fornito dall'Istituto Superiore di Sanità (ISS) per la raccolta e l'analisi delle sequenze di SARS-CoV-2 provenienti da pazienti di tutte le regioni, permettendo così la sorveglianza del virus.

Dopo il caricamento delle *reads* nel sistema, oltre all'assemblaggio, la sequenza viene anche analizzata, ottenendo così informazioni di interesse, quali il *lineage*, il *clade* e la percentuale di non copertura della sequenza. Nel caso di sequenze con un'alta percentuale di non copertura, è stato svolto un ulteriore ri-assemblaggio mediante la piattaforma di analisi della *Stanford University* (<https://covdb.stanford.edu/sierra/sars2/by-patterns/>). Le sequenze consenso in formato fasta sono state generate mediante I-Co-Gen per le successive analisi. L'analisi delle sequenze è stata svolta mediante i *software* Nextclade (<https://clades.nextstrain.org/>) e Pangolin (<https://cov-lineages.org/>), che forniscono, fra le varie informazioni, le mutazioni, i *clade* e i *lineage*.

Per l'allineamento delle sequenze in esame è stato utilizzato il *software* BioEdit v. 7.7 (<https://bioedit.software.informer.com/7.7/>), sia per l'allineamento

manuale sia per quello automatico con l'algoritmo ClustalW, e sono state eliminate le regioni non tradotte al 5' e al 3' (5'-UTR e 3'-UTR).

Le analisi delle distanze genetiche, della pressione selettiva e la costruzione dell'albero filogenetico sono state svolte mediante il *software* MEGA-X (<https://www.megasoftware.net/>).

## 4. RISULTATI

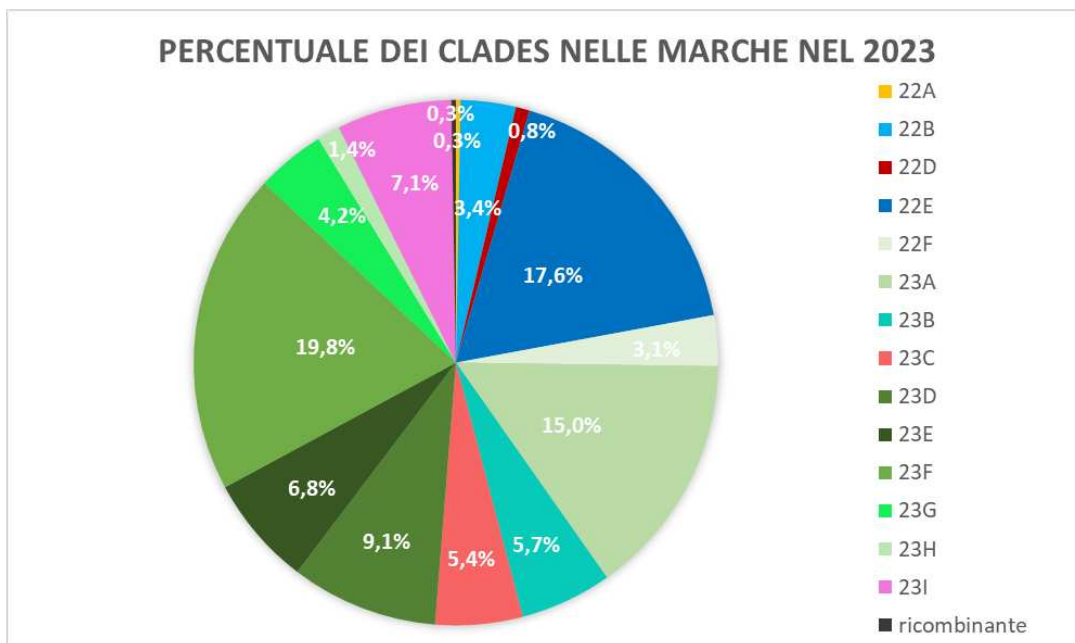
In occasione del monitoraggio mensile di SARS-CoV-2 indetto dall'Istituto Superiore di Sanità, al fine di stimare la prevalenza delle varianti circolanti sul territorio italiano, presso la SOSD Virologia dell'Azienda Ospedaliero Universitaria delle Marche, nell'anno 2023 sono stati sequenziati oltre 400 campioni biologici mediante tecnologia Sanger e *Next Generation Sequencing*. L'analisi epidemiologica si è focalizzata sui genomi interi di SARS-CoV-2, ottenuti mediante sequenziamento *paired-end* sulla piattaforma MiSeq Illumina che al termine di ogni corsa per ogni campione restituisce una coppia di file in formato FastQ (R1 e R2), contenenti tutte le *reads*.

Questi file sono stati caricati sulla piattaforma I-Co-Gen per essere assemblati al fine di ottenere le sequenze consenso in formato Fasta.

I consensi ottenuti sono stati successivamente analizzati mediante gli algoritmi *Nextclade* e *Pangolin*, determinando così i rispettivi lignaggi e *clades*. Inoltre, dalla piattaforma *Nextclade* sono state ottenute anche tutte le informazioni riguardanti le mutazioni associate a ciascuna sequenza.

Per l'analisi sono state selezionate le sequenze con elevata copertura, ottenendo 353 sequenze di SARS-CoV-2 distribuite nei vari mesi.

Nel corso dell'anno 2023 si sono susseguiti 14 *clades*, in proporzioni diverse, come riportato in Figura 24.



*Figura 24 – Grafico della percentuale di ogni clade sul totale delle sequenze campionate nella regione Marche nel 2023.*

Il *clade* 23F (*lineage* EG.5.1) costituisce il 19,8% di tutte le sequenze analizzate nel 2023, costituendo la variante con il maggior numero di sequenze analizzate; a seguire il *clade* 22E (*lineage* BQ.1) che ha rappresentato il 17,6% e il *clade* 23A (*lineage* XBB.1.5) con il 15%.

La maggior parte dei *clades* hanno avuto un impatto minore del 10% sul totale delle sequenze, in particolare i *clades* 22A, 22D e la sequenza ricombinante XBF hanno rappresentato meno dell'1% del totale.

Per quanto riguarda la distribuzione dei vari *clade* nel corso del 2023, fra i mesi di gennaio e febbraio 2023 il *clade* 22E (*lineage* BQ.1) ha circolato in modo prevalente e, in minor misura, il *clade* 22B (*lineage* BA.5) da cui deriva.

Nei primi mesi del 2023 si osserva l'ingresso dei ricombinanti XBB, appartenenti al *clade* 22F e da cui si sono originate diverse varianti che sono diventate le predominanti per tutto il resto dell'anno.

Contemporaneamente, c'è stata una discreta circolazione delle varianti appartenenti al *clade* 23C (*lineage* CH.1.1), derivante dal *clade* 22D (*lineage* BA.2.75), seppur non siano mai diventate le prevalenti.

La variante ricombinante appartenente al *clade* 23B (XBB.1.16), denominata anche *Arcturus*, discende dalla linea delle 22F (XBB) ed è comparsa in aprile, circolando per il resto dell'anno seppur in minor misura rispetto ad altre ricombinanti, quali i *clade* 23A (*lineage* XBB.1.5) e 23D (*lineage* XBB.1.9).

La variante *Eris* appartenente al *clade* 23F (*lineage* EG.5.1) e le sue derivate, fra cui la 23H (*lineage* HK.3), sono state le principali varianti a circolare nella seconda metà dell'anno 2023, fino alla comparsa della variante BA.2.86 appartenente al *clade* 23I fra novembre e dicembre 2023 (Figura 25).

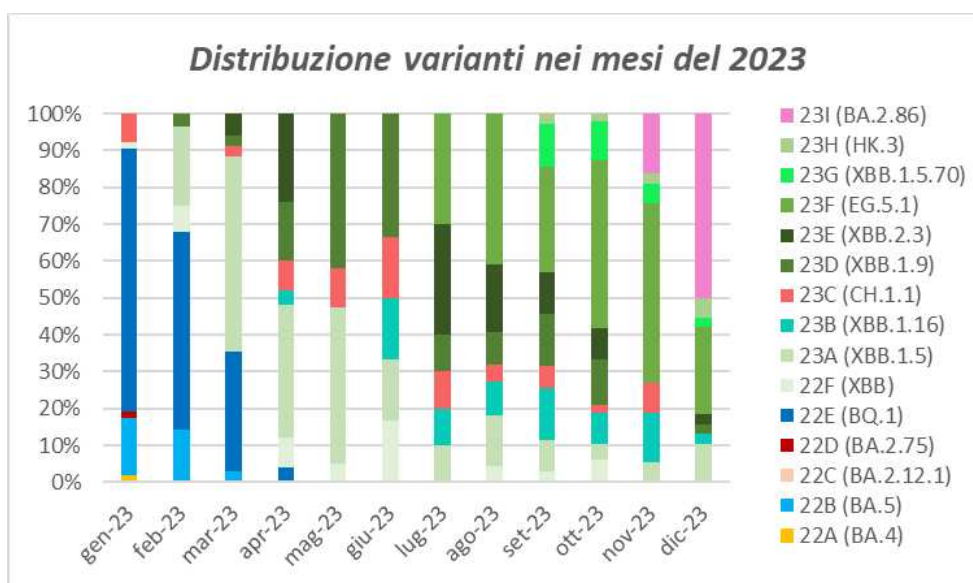
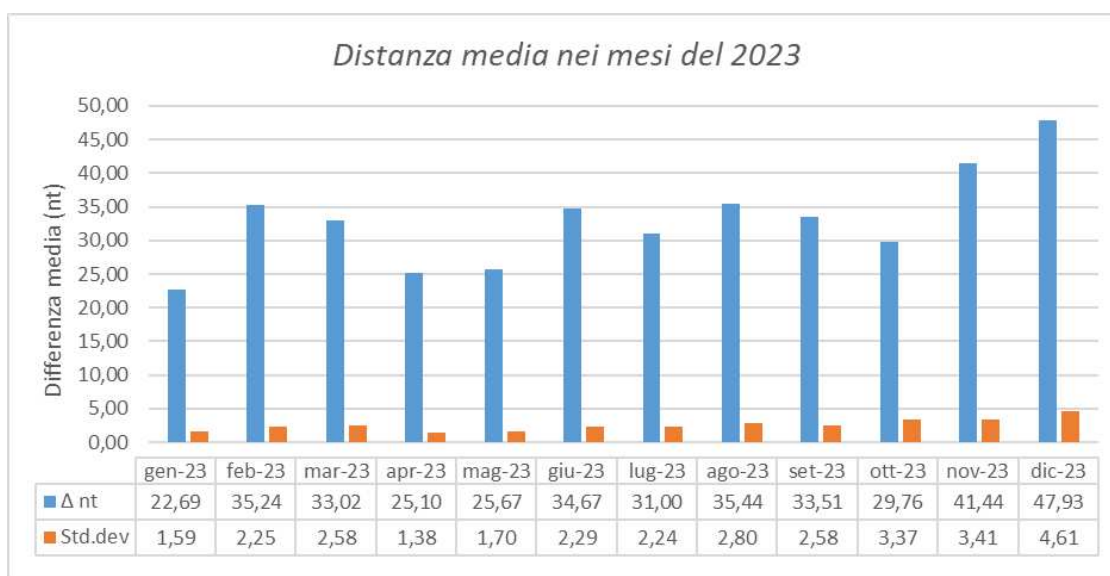


Figura 25 – Grafico della distribuzione percentuale dei vari clades nei mesi del 2023.

## 4.1 Analisi delle distanze genetiche

L'analisi delle distanze genetiche è stata svolta mediante il *software* MEGA-X; da questa analisi si è evidenziato che la *Overall Mean Distance*, ossia la distanza media di tutte le sequenze, calcolata utilizzando il metodo *No of differences*, che stima il numero di siti nucleotidici che differiscono tra le sequenze, è risultata essere di 44.57 nucleotidi con una deviazione standard di 2.92.

Raggruppando le sequenze su base mensile, sono state analizzate le distanze nucleotidiche presenti all'interno di ogni mese ed emerge come la differenza media all'interno di ogni mese sia diversa, incrementata in particolar modo nei mesi di novembre e dicembre (Figura 26).



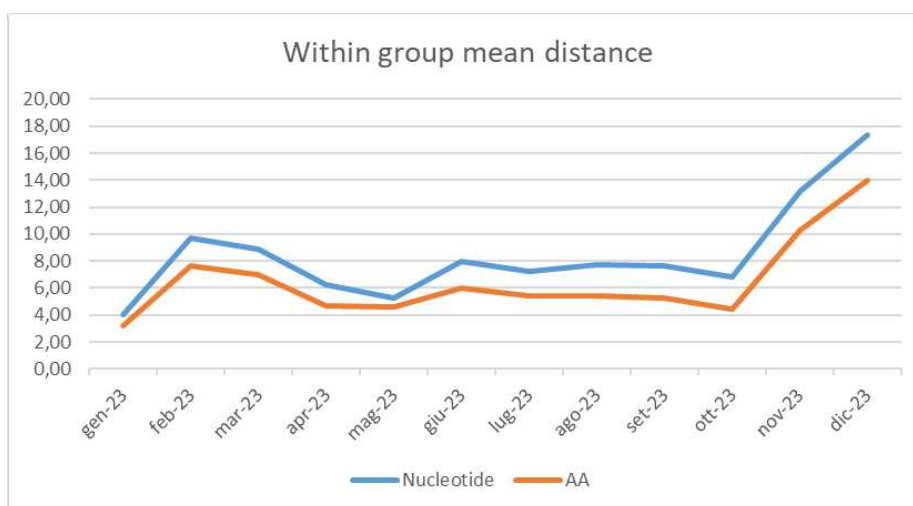
*Figura 26 – Grafico della distanza nucleotidica media nei mesi del 2023. Errore standard calcolato con bootstrap di 100.*

La variante appartenente al *clade* 23I (BA.2.86), anche denominata *Pirola*, si caratterizza per un'elevata divergenza nucleotidica rispetto alle altre varianti circolanti, presentando oltre 50 mutazioni nella sequenza del gene S.

La comparsa della BA.2.86 giustifica la maggior distanza genetica osservata nei mesi di novembre e dicembre, analogamente a quanto osservato agli inizi dell'anno con la comparsa delle varianti ricombinanti.

Tuttavia, nei vari mesi, la divergenza nucleotidica che si riscontra dai dati è anche dipendente dalla circolazione di vari *clades*, ognuno caratterizzato da un profilo di mutazioni caratteristico.

Ponendo l'attenzione sulla distanza intragruppo del solo gene S nei vari mesi del 2023, si può notare la maggior divergenza nucleotidica e amminoacidica in concomitanza dei mesi di febbraio-marzo e novembre-dicembre, ossia intervalli di tempo in cui sono emerse le varianti ricombinanti XBB e la variante BA.2.86 rispettivamente (Figura 27).



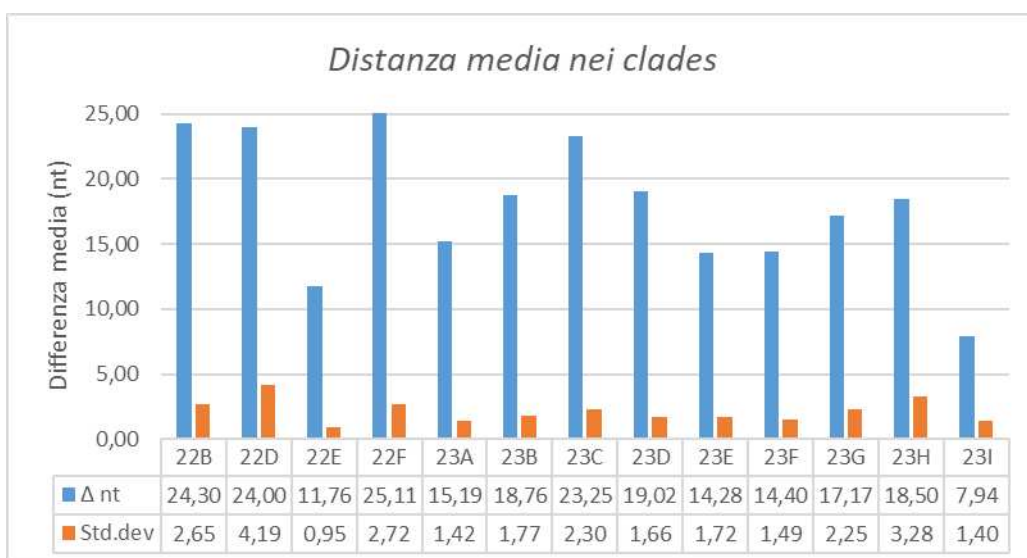
**Figura 27** – Grafico della distanza nucleotidica e amminoacidica media riferita al gene S nei mesi del 2023. Errore standard calcolato con bootstrap di 100

Successivamente sono state calcolate le distanze nucleotidiche all'interno di ogni *clade*. Ciò che emerge è una maggiore divergenza nucleotidica all'interno del *clade* 22B (*lineage* BA.5), in quanto, essendo un *clade* che racchiude tutti i discendenti della Omicron 5 ad eccezione della BQ.1, risulta essere più eterogeneo rispetto ad altri.

Questo si osserva anche nel *clade* 22D (*lineage* BA.2.75) e nel *clade* 23C (*lineage* CH.1.1), derivante dal precedente.

Una maggiore divergenza nucleotidica rispetto agli altri *clade* è riscontrata anche nel *clade* 22F (XBB), a cui appartengono le varianti che si sono originate dalla ricombinazione del *lineage* BJ.1 (*sublineage* BA.2.10.1) con il *lineage* BM.1.1.1 (*sublineage* BA.2.75.3).

La minor eterogeneità, invece, è riscontrata nel *clade* 23I (*lineage* BA.2.86), in quanto rappresenta un *clade* costituito di recente, perciò presenta un minor numero di sottovarianti e quindi minor variabilità nucleotidica al suo interno (Figura 27).



**Figura 27** – Grafico della distanza nucleotidica media nei clades riscontrati nel 2023. Errore standard calcolato con bootstrap di 100.

In seguito, le sequenze sono state tagliate e allineate per ogni gene e *Open Reading Frame* (ORF) di SARS-CoV-2 e sono state calcolate le distanze in termini nucleotidici, amminoacidici, di mutazioni sinonime e non sinonime per poi stimare la pressione selettiva specifica per ogni gene (Tabella 1).



Overall mean distance N° of differences - BS100									
	Nucleotide	s.e.	AA	s.e.	SYN	s.e.	NON SYN	s.e	dN/dS
nsp1	0,81	0,48	0,60	0,42	1,60E-03	4,00E-04	1,50E-03	1,10E-03	9,38E-01
nsp2	2,18	0,56	1,52	0,44	1,50E-03	3,00E-04	1,10E-03	3,00E-04	7,33E-01
nsp3	5,09	0,80	2,71	0,56	1,80E-03	4,00E-04	6,20E-04	1,40E-04	3,44E-01
nsp4	1,29	0,53	0,94	0,47	1,50E-03	4,00E-04	2,00E-04	1,00E-04	1,33E-01
nsp5	0,41	0,11	0,11	0,05	1,50E-03	5,00E-04	1,70E-04	7,00E-05	1,13E-01
nsp6	1,28	0,45	0,95	0,52	1,80E-03	4,00E-04	1,50E-03	7,00E-04	8,33E-01
nsp7	0,15	0,11	0,02	0,013	2,50E-03	2,10E-03	2,00E-04	1,00E-04	8,00E-02
nsp8	0,7	0,38	0,27	0,12	3,30E-03	2,50E-03	6,00E-04	3,00E-04	1,82E-01
nsp9	0,84	0,5	0,50	0,49	7,00E-03	4,00E-03	4,10E-03	4,30E-03	5,86E-01
nsp10	0,11	0,04	0,04	0,02	7,00E-04	4,00E-04	1,00E-04	1,00E-04	1,43E-01
nsp11	0,006	0,005	0,006	0,005	0,00E+00	0,00E+00	2,00E-04	2,00E-04	
nsp12	3,08	0,84	1,14	0,5	3,10E-03	9,00E-04	5,00E-04	2,00E-04	1,61E-01
nsp13	2,55	0,89	1,22	0,53	3,20E-03	1,60E-03	1,00E-03	5,00E-04	3,13E-01
nsp14	1,7	0,58	0,44	0,16	3,70E-03	1,70E-03	4,00E-04	1,00E-04	1,08E-01
nsp15	0,42	0,09	0,33	0,08	4,00E-04	1,00E-04	5,00E-04	1,00E-04	1,25E+00
nsp16	0,41	0,13	0,28	0,13	6,00E-04	2,00E-04	4,00E-04	2,00E-04	6,67E-01
Gene S	12,99	1,77	10,14	1,57	2,40E-03	6,00E-04	4,40E-03	8,00E-04	1,83E+00
ORF 3a	1,49	0,41	0,00	0,00	4,60E-03	2,70E-03	1,10E-03	2,00E-04	2,39E-01
Gene E	0,52	0,36	0,40	0,33	1,50E-03	9,00E-04	3,00E-03	2,90E-03	2,00E+00
Gene M	1,56	0,64	0,75	0,49	5,40E-03	2,50E-03	1,60E-03	1,00E-03	2,96E-01
ORF 6	1,43	0,57	0,39	0,31	1,90E-02	1,00E-02	5,05E-03	4,51E-03	2,66E-01
ORF 7a	0,78	0,38	0,24	0,14	6,00E-03	5,00E-03	1,00E-03	5,00E-04	1,67E-01
ORF 7b	0,15	0,11	0,13	0,08	2,20E-03	2,30E-03	1,80E-03	1,10E-03	8,18E-01
ORF 8	0,94	0,48	0,28	0,10	2,10E-03	8,00E-04	1,10E-03	4,00E-04	5,24E-01
Gene N	2,40	0,68	0,72	0,29	4,90E-03	1,80E-03	1,00E-03	4,00E-04	2,04E-01
ORF 10	0,40	0,32	0,38	0,38	2,30E-04	2,40E-04	4,70E-03	4,80E-03	2,04E+01

*Tabella 1 – Overall mean distance nucleotidica, amminoacidica, delle mutazioni sinonime e non sinonime di tutti i geni di SARS-CoV-2. Errore standard calcolato con bootstrap di 100.*

Le mutazioni sinonime (dS), ossia che non implicano un cambiamento nell'amminoacido relativo al codone mutato, e le mutazioni non sinonime (dN), che comportano un cambiamento amminoacidico, vengono calcolate per stimare il rapporto dN/dS. Il valore ottenuto da tale rapporto, quando è maggiore di 1, indica che c'è stata una selezione positiva nel gene in esame, ovvero che tale gene sia predisposto all'accumulo di mutazioni non sinonime rispetto alle mutazioni neutre.

Dai risultati ottenuti si riscontrano valori maggiori di 1 per la ORF della proteina non strutturale 15, per il gene S, per il gene E e per la ORF10.

Tuttavia, la maggior parte delle mutazioni sono essenzialmente sinonime, non determinando perciò un cambiamento a livello amminoacidico.

Per quanto riguarda il gene S, essendo la proteina *spike* quella deputata all'attacco del virus alla cellula ospite, è noto per essere una regione sottoposta a costante selezione positiva, infatti, mutazioni favorevoli possono riguardare anche l'aumento dell'affinità per il recettore cellulare.

Inoltre, il gene S è uno dei maggiori target per i vaccini e gli anticorpi prodotti da infezioni pregresse, e tale pressione selettiva positiva è in relazione con la capacità elusiva delle risposte umorali riscontrata nelle varianti emergenti.

Una possibile spiegazione per i valori ottenuti dall'analisi del gene E può riguardare la sua vicinanza spaziale alla proteina *spike* nel rivestimento virale e del suo ruolo nell'assemblaggio del virione maturo. È possibile quindi ipotizzare che i due geni S ed E co-evolvano al fine di mantenere le corrette interazioni elettrostatiche che ci sono sulla superficie del virione.

La proteina nsp15 è un'endoribonucleasi ed è generalmente una proteina molto conservata; la comparsa di mutazioni non sinonime possono influire sulla attività catalitica di questo enzima.

La proteina ORF10 rappresenta una piccola proteina di circa 38 amminoacidi di cui, tutt'oggi, non sono ben chiare le funzioni; perciò, risulta difficile stimare le implicazioni biologiche di una selezione in atto in questa sequenza.

Alcuni geni, al contrario, si mostrano molto conservati, in quanto mutazioni in essi potrebbero essere deleterie per la *fitness* virale, e solo pochi cambiamenti amminoacidici si fissano in tali regioni del genoma. Esempi sono il Gene N, che rappresenta il gene più conservato nella famiglia dei *coronavirus*, ma anche quei geni implicati nella replicazione, come la nsp12, ossia la polimerasi virale, e la nsp14, ovvero quello con attività di correttore di bozze del genoma di nuova sintesi.

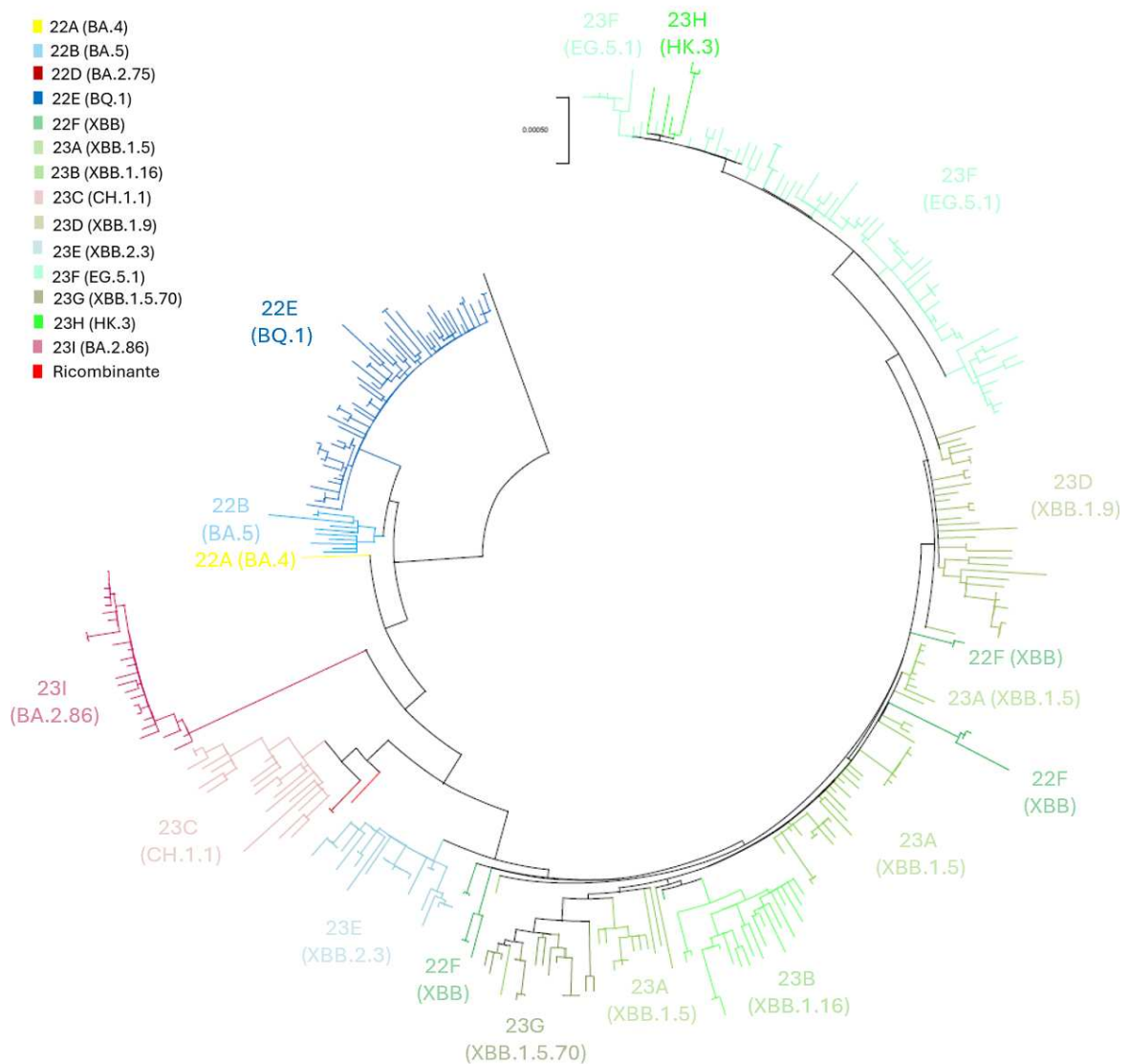
## 4.2 Analisi filogenetica delle sequenze

Per comprendere meglio le relazioni evolutive che intercorrono fra le varie sequenze, è stato modellizzato un albero filogenetico. A tale scopo il modello di sostituzione applicato è stato il *Kimura 2-parameter*, che tiene conto dei differenti tassi di sostituzione delle transizioni e delle trasversioni. Il metodo statistico scelto per l'analisi filogenetica è stato il *Maximum Likelihood*, che si basa sulla stima dell'albero che ha la massima probabilità all'interno di un set di dati, ossia il valore di "massima verosimiglianza", con un *bootstrap* di 100. Per la costruzione dell'albero filogenetico è stata utilizzata la sequenza di riferimento di SARS-CoV-2 *Wuhan-Hu-1* (NC\_045512) su cui è stata posta la radice dell'albero (Figura 28).

Dall'analisi emerge come le sequenze risultano ripartite in due gruppi iniziali, ossia le sequenze derivanti dalle Omicron BA.2 e quelle derivanti dalle Omicron BA.5.

Per quanto riguarda quest'ultime, vi appartengono i *clades* 22B (BA.5) e 22E (BQ.1) che, infatti, si trovano vicini nell'albero filogenetico, comprensibilmente per il fatto che il *lineage* BQ.1 discende dal clade 22B.

Le restanti sequenze si raggruppano uniformemente, ma è possibile evidenziare i *clades* direttamente derivati da BA.2, come le BA.2.75 e i suoi derivati, e la più recente BA.2.86; mentre i ricombinanti di BA.2, le XBB, si generano da un unico nodo dell'albero, che a sua volta si divide in altri nodi che raggruppano sequenze di *lineages* progenitori e derivati, o fa posizionare vicino sequenze di *lineages* correlati.



*Figura 28 – Albero filogenetico costruito con analisi Maximum Likelihood, metodo Kimura 2-parameter e Bootstrap 100.*

È interessante notare il legame filogenetico che c'è tra il *cluster* di sequenze appartenenti ai *clades* 22D (BA.2.75) e 23C (CH.1.1) e il *cluster* costituito dalle varianti ricombinanti del *clade* 22F (XBB) e le sue discendenti. Infatti, il *clade* 23C ha origine a partire dal *clade* 22D, mentre il *clade* 22F deriva da un evento di ricombinazione fra una BJ.1 (*lineage* BA.2.10.1.1) e una BM.1.1.1 (*lineage* BA.2.75.3.1.1.1), due varianti appartenenti rispettivamente al *clade* 21L

(BA.2), da cui hanno avuto origine tutte le varianti attualmente in circolazione, e al clade 22D (BA.2.75).

Tra le sequenze in esame è presente anche un'altra sequenza ricombinante, XBF, non associata al *clade* 22F che si posiziona nell'albero adiacente ai *clades* 22D e 23C.

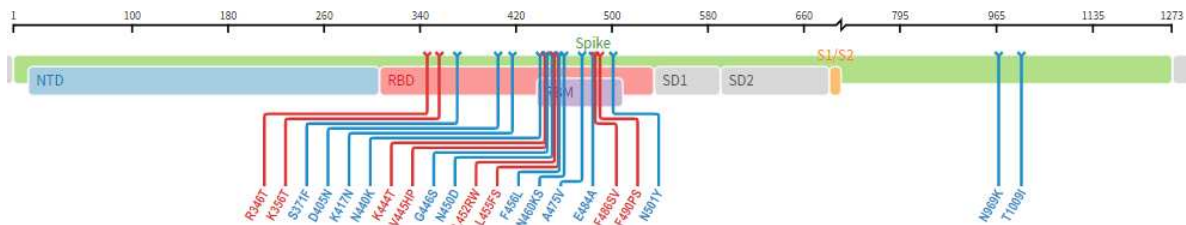
Infine, la nuova variante in circolazione, appartenente al clade 23I (BA.2.86), si presenta più filogeneticamente vicina all'antenato comune rispetto ai clade sopracitati.

### 4.3 Analisi del profilo mutazionale

Successivamente sono state esaminate più nel dettaglio le mutazioni riscontrate nelle varie sequenze.

Le sequenze sono state analizzate mediante il *software Nextstrain* e le mutazioni sono state suddivise per ogni *clade*; ai fini dell'analisi sono state prese in considerazione le mutazioni con una rappresentanza maggiore del 5% all'interno del *clade* e sono state analizzate mediante il programma di analisi della *Stanford University* (<https://covdb.stanford.edu/sierra/sars2/by-patterns/>). Per quanto riguarda le mutazioni riscontrate nella proteina *spike*, nei *clades* che hanno circolato nel 2023 nella regione Marche sono state riscontrate diverse mutazioni che possono conferire al virus resistenza verso determinate terapie a base di anticorpi monoclonali (mAb) o anche una maggior affinità verso il recettore cellulare *Angiotensin-converting enzyme 2* (ACE2). In particolare, 27 mutazioni si sono dimostrate conferire al virus resistenza verso mAb, ovvero R346T, K356T, S371F, D405N, K417N, N440K, K444T, V445HP, G446S, N450D, L452RW, L455FS, F456L, N460KS, A475V, E484A, F486SV,

F490PS, N501Y, N969K e T1009I, di cui alcune mutazioni hanno una prevalenza a livello globale di meno del 0,01% (Figura 29).



**Figura 29** – Mutazioni nella proteina spike associate alla resistenza verso gli anticorpi monoclonali. in rosso sono riportate le mutazioni presenti in meno dello 0,01% a livello globale. (<https://covdb.stanford.edu/sierra/sars2/by-patterns/>)

Analogamente a quanto riscontrato in letteratura, la maggior parte delle mutazioni ricadono nel *Receptor Binding Domain* (RBD) nella subunità S1, ossia il dominio responsabile del legame al recettore cellulare, e in particolare nel *Receptor Binding Motif* (RBM) al suo interno.

Le mutazioni presenti nell'RBD della proteina spike sono state confrontate fra tutti i *clades* circolanti per stimare la loro frequenza (Figura 30).

Sono state selezionate solo le mutazioni con una frequenza maggiore dell'1%. È da notare come molte delle mutazioni di interesse clinico presenti nell'RBD siano ampiamente diffuse nella maggioranza dei *clades*, come R346T, S371F, D405N, K417N, N440K, G446S, N460K, E484A, F490S e N501Y, indice che queste mutazioni una volta comparse, si sono fissate e sono rimaste nella maggioranza delle varianti discendenti.

Sono stati evidenziati diversi *hotspot* di mutazione, ovvero punti del genoma che mostrano un cambio amminoacidico nelle diverse varianti, come i siti 445, 452 e 455, tuttavia l'amminoacido risultante non è sempre lo stesso, come ad esempio nell'ultimo *clade* emerso nel 2023, il 23I, sono presenti le mutazioni

V445H, L452W e L455S, che negli altri *clades* sono occupati da un diverso amminoacido, V445P, L452R e L455F.

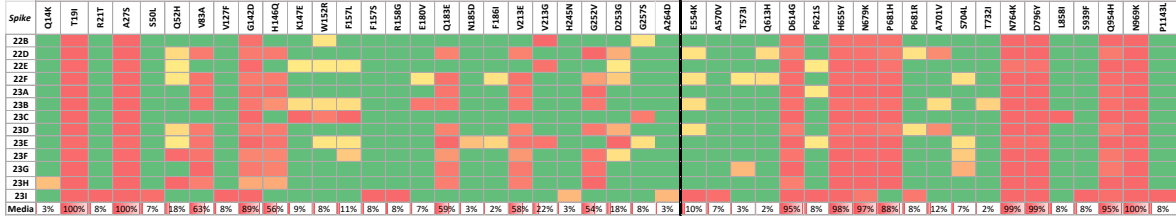
Le restanti mutazioni sono distribuite con varia frequenza nei vari *clades*.

RBD	22B	22D	22E	22F	23A	23B	23C	23D	23E	23F	23G	23H	23I	Media
S:G339D	Red	Green	Red	Red	Green	Red	Red	Red	Green	Green	Green	Green	Green	45%
S:G339H	Red	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	38%
S:R346T	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Yellow	Red	Green	84%
S:K356T	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	6%
S:L368I	Green	Yellow	Green	Orange	Orange	Orange	Green	Orange	Orange	Yellow	Yellow	Yellow	Green	28%
S:S371F	Red	Yellow	Red	Orange	Orange	Orange	Green	Orange	Red	Yellow	Yellow	Yellow	Green	53%
S:S373P	Red	Yellow	Red	Orange	Orange	Orange	Green	Orange	Red	Yellow	Yellow	Yellow	Green	53%
S:S375F	Red	Yellow	Red	Orange	Orange	Orange	Green	Orange	Red	Yellow	Yellow	Yellow	Green	52%
S:T376A	Red	Yellow	Red	Orange	Orange	Orange	Green	Orange	Red	Yellow	Yellow	Yellow	Green	52%
S:R403K	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	6%
S:D405N	Red	Yellow	Red	Orange	Orange	Orange	Green	Orange	Red	Yellow	Yellow	Yellow	Green	52%
S:R408S	Red	Yellow	Red	Orange	Orange	Orange	Green	Orange	Red	Yellow	Yellow	Yellow	Green	52%
S:K417N	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	96%
S:N440K	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	100%
S:K444T	Green	Green	Red	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	17%
S:V445H	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	8%
S:V445P	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	68%
S:G446S	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	85%
S:N450D	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	8%
S:L452M	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Green	3%
S:L452R	Red	Red	Red	Green	Green	Green	Red	Green	Red	Yellow	Green	Green	Green	34%
S:L452W	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	8%
S:F456L	Red	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	7%
S:L455F	Green	Green	Green	Yellow	Yellow	Red	Green	Yellow	Red	Red	Red	Red	Green	27%
S:L455S	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	7%
S:F456L	Green	Green	Green	Yellow	Yellow	Red	Green	Yellow	Red	Red	Red	Red	Green	42%
S:N460K	Yellow	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	94%
S:A475V	Green	Green	Green	Green	Yellow	Green	Green	Green	Green	Green	Green	Green	Green	3%
S:S477N	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	56%
S:T478K	Orange	Green	Orange	Orange	Green	Green	Green	Orange	Orange	Orange	Orange	Orange	Orange	37%
S:T478R	Orange	Green	Orange	Orange	Green	Green	Green	Orange	Orange	Orange	Orange	Orange	Orange	15%
S:N481K	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	2%
S:E484A	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	92%
S:F486P	Green	Red	Green	Red	Red	Red	Red	Red	Red	Red	Red	Red	Yellow	65%
S:F486S	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	9%
S:F486V	Red	Green	Red	Green	Green	Green	Green	Red	Green	Green	Green	Green	Green	22%
S:F490S	Green	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	61%
S:Q498R	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	100%
S:N501Y	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	100%
S:Y505H	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	100%
S:V511I	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Green	Green	3%
S:P521S	Green	Green	Green	Yellow	Green	Green	Green	Red	Green	Green	Green	Green	Green	12%
S:P521T	Green	Green	Green	Green	Green	Yellow	Green	Green	Green	Green	Green	Green	Green	2%

Figura 30 – Heatmap delle mutazioni nel Receptor Binding Domain (RBD) della proteina spike, colorate in base alla frequenza all'interno di ogni clade.

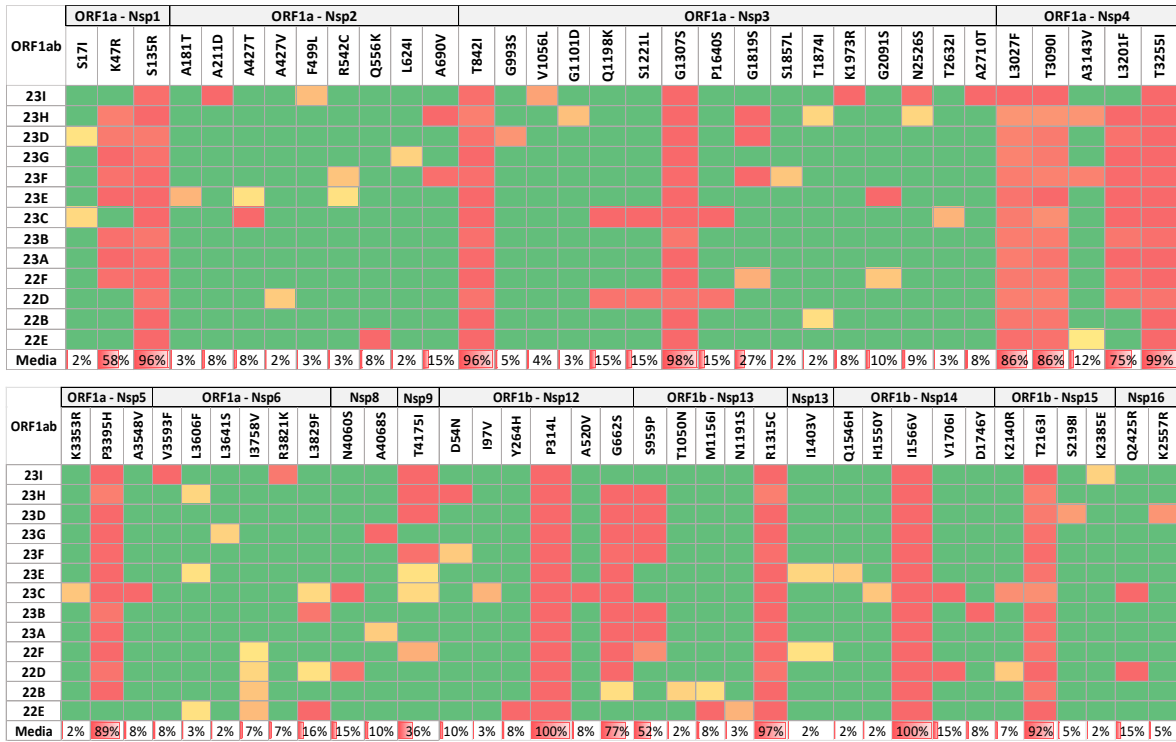
Analogamente, nelle restanti porzioni del gene S sono presenti mutazioni largamente condivise fra i *clades* (Figura 31). Molte di queste mutazioni riducono la suscettibilità alla neutralizzazione da parte degli anticorpi, come ad esempio le mutazioni G142D o la D253G. Di particolare rilevanza sono le

mutazioni che ricadono in siti di interesse, tra queste ci sono le mutazioni N679K e P681HR che si localizzano in prossimità del sito di taglio S1/S2 e che possono apportare dei vantaggi per la replicazione e il tropismo virale.



*Figura 31 – Heatmap delle mutazioni riscontrate nel gene S, colorate in base alla frequenza all'interno di ogni clade. La linea nera indica la collocazione dell'RBD.*

Sono state analizzate anche le mutazioni riguardanti la ORF1ab. In questa regione del genoma vengono codificate molte proteine di fondamentale importanza per la replicazione virale; quindi, la comparsa di mutazioni in questi geni possono notevolmente influenzare la *fitness* virale.



*Figura 32 – Heatmap delle mutazioni riscontrate nella ORF1ab, colorate in base alla frequenza all'interno di ogni clade.*



Da questa analisi sono emerse diverse mutazioni presentanti una frequenza nei vari *clades* maggiore del 90%, quali S135R, T842I, G1307S, L3027F, T3090I, T3255I, P314L, R1315C, I1566V, T2163I.

Di particolare importanza ci sono le mutazioni che ricadono nel gene Nsp12, codificante la RNA polimerasi RNA-dipendente (RdRp); in virtù del suo ruolo nella replicazione del genoma virale, questa regione risulta molto conservata, ma nel tempo alcune mutazioni si sono fissate nelle diverse varianti, quali la P314L e, soprattutto nelle varianti ricombinanti e nel *clade* 23C, la G662S.

Analogamente, le mutazioni che ricadono in geni essenziali per la produzione delle proteine virali sono oggetto di analisi, ossia la Nsp3 (PL<sup>pro</sup>) e la Nsp5 (M<sup>pro</sup>), ovvero le proteasi responsabili del taglio delle poliproteine pp1a e pp1ab; in questi geni, le mutazioni con maggior frequenza fra i *clades* sono la T842I e la G1307S per quanto riguarda la Nsp3 e la P3395H per la Nsp5.

#### **4.4 Andamento epidemiologico di SARS-CoV-2 nelle Marche**

Per avere una panoramica complessiva dell'andamento delle varianti di SARS-CoV-2 nel corso della pandemia nella regione Marche, sono state unite le sequenze raccolte nel territorio nell'anno 2023 a quelle dei precedenti anni.

Il grafico mostra il susseguirsi delle varianti a partire da febbraio 2020 fino a dicembre 2023 (Figura 31).

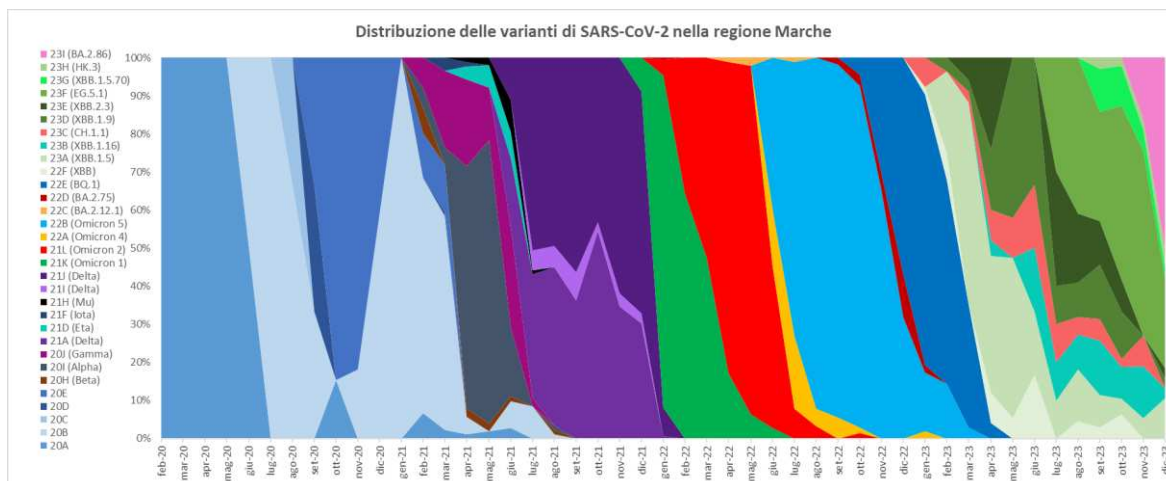


Figura 31 – Distribuzione dei vari clades di SARS-CoV-2 nella regione Marche tra febbraio 2020 e dicembre 2023.

Nei primi mesi del 2020, in un contesto di *lockdown* per l'Italia, la variante predominante è stata quella appartenente al *clade* 20A (*lineage* B.1), ossia la discendente diretta del ceppo B, *Wuhan-Hu-1*. In particolare, dal *clade* 20A hanno avuto origine diverse varianti che hanno circolato nel territorio per molti mesi, come ad esempio il *clade* 20B (*lineage* B.1.1) e la e la 20E (*lineage* B.1.177) nei mesi tra maggio 2020 e agosto 2021.

A partire da inizio 2021 sono emerse nella regione Marche le prime due varianti classificate al tempo come *Variants Of Concern* (VOC) dall' Organizzazione Mondiale della Sanità (OMS), ossia le varianti Alpha (*clade* 20I, *lineage* B.1.1.7) e Gamma (*clade* 20J, *lineage* P.1), entrambe discendenti dal *clade* 20B e che sono circolate fino ad agosto 2021. Altra VOC, la Beta (*clade* 20H, *lineage* B.1.351), ha circolato nello stesso periodo, seppur in minor misura.

Nel mese di giugno 2021 è emersa un'altra VOC, la variante Delta (*clade* 21A, *lineage* B.1.617.2), che ha avuto origine dal *clade* 20A e che insieme ai suoi *clades* discendenti, 21I e 21J, hanno soppiantato i precedenti *clades*, diventando le predominanti fino a gennaio 2022. È interessante notare come i

tre *clades* Delta nella regione hanno circolato nello stesso periodo, senza che l'uno predominasse sull'altro.

Al contrario, le varianti Omicron registrate nelle Marche per la prima volta a fine 2021, si sono soppiantate l'un l'altra, con dapprima la comparsa dal *clade* 21K (*lineage* BA.1), poi rimpiazzata dalla Omicron 21L (*lineage* BA.2) e dagli altri *clades*, tra cui il 22B (*lineage* BA.5) che è risultato predominante da luglio 2022 fino agli inizi dell'anno 2023.

All'inizio del 2023 sono comparse le prime varianti ricombinanti, originate dalla ricombinazione di due varianti Omicron, i *lineages* BJ.1 (BA.2.10.1.1; *clade* 21L) e BM.1.1.1 (BA.2.75.3.1.1.1; *clade* 22D); a partire dal *clade* 22F (XBB) si sono originati quattro *clades* distinti che hanno circolato simultaneamente nella regione, ossia il 23A (*lineage* XBB.1.5), il 23B (*lineage* XBB.1.16), il 23D (*lineage* XBB.1.9) e il 23E (*lineage* XBB.2.3), e hanno a loro volta dato origine a nuovi *clades*, continuando a coesistere nello stesso tempo.

Predominante nella prima metà del 2023 è stato il *clade* 23A (*lineage* XBB.1.5) che ha poi circolato per tutto il resto dell'anno, originando a sua volta il *clade* 23G (*lineage* XBB.1.5.70) intorno settembre 2023.

Un altro *clade* molto rappresentato nel 2023 è stato il 23F (*lineage* EG.5.1), discendente dal *clade* 23D e che ha dato origine al 23H (*lineage* HK.3).

## 5. DISCUSSIONE E CONCLUSIONI

La pandemia di COVID-19 ha rappresentato una grande sfida a livello globale per la salute pubblica. L'agente eziologico responsabile, ovvero il SARS-CoV-2, nel corso del tempo ha accumulato mutazioni che in molti casi hanno conferito un vantaggio evolutivo al virus; in particolare, dal 2020 ad oggi si sono susseguite diverse varianti virali scaturite dall'acquisizione di tali mutazioni che hanno fornito al virus maggior trasmissibilità, di resistenza alle terapie, ma soprattutto la capacità di evadere le difese immunitarie maturate in seguito a precedenti infezioni o dalla vaccinazione.

In virtù di tali caratteristiche virali è chiaro come l'attuazione di programmi di monitoraggio territoriali delle varianti in circolazione risulta necessaria per comprendere a pieno l'andamento dell'epidemia e poter così disporre di strumenti utili per contrastarla, come ad esempio l'adozione di terapie più efficaci o vaccini aggiornati verso nuove varianti.

In occasione del monitoraggio mensile delle varianti di SARS-CoV-2 richiesto dall'Istituto Superiore di Sanità (ISS) sul territorio italiano, nella regione Marche nell'anno 2023 si è registrata la circolazione di 14 *clades* differenti, situazione che ricalca la circolazione di tali varianti nel resto d'Europa, senza che si sia evidenziata una peculiarità italiana.

A predominare nei primi mesi dell'anno è stato il *clade* 22E (*lineage* BQ.1), discendente del *clade* 22B, ossia il *lineage* BA.5 che ha prevalso nell'anno 2022, ma che nel 2023 ha circolato solo nei primi mesi dell'anno.

Nel 2023 c'è stata la comparsa delle prime varianti ricombinanti di SARS-CoV-2, a partire dal clade 22F (*lineage* XBB) e dai *clades* suoi discendenti (23A, 23B, 23D, 23E, 23F, 23G, 23H).

Le diverse varianti ricombinanti hanno soppiantato i *clades* precedenti e hanno circolato nel territorio marchigiano simultaneamente per la maggior parte dell'anno, senza eradicarsi l'un l'altra. Infatti, nella seconda metà dell'anno, per diversi mesi hanno co-circolato fino a otto varianti ricombinanti contemporaneamente.

In minor misura, a circolare al contempo delle altre varianti, c'è stato il *clade* 23C (*lineage* CH.1.1), derivante dal *clade* 22D (*lineage* BA.2.75).

Di particolare attenzione vi è il *clade* 23I (*lineage* BA.2.86) che fa la sua comparsa a novembre 2023, ma che nel giro di un solo mese ha prevalso sulle altre varianti in circolazione. Attualmente è ancora la variante maggiormente campionata nelle Marche ed è di particolare interesse in quanto presenta oltre 50 mutazioni nella sola proteina *spike* rispetto al clade progenitore, ossia il *clade* 21L (BA.2).

Indagando sulle distanze in termini nucleotidici, sia per quanto riguarda le differenze mensili e sia intraclade, è emerso come le maggiori distanze nucleotidiche siano presenti soprattutto nei mesi di novembre e dicembre 2023 e siano minori nei primi mesi dell'anno; invece, le maggiori divergenze intraclade sono riscontrabili nei *clades* 22B (BA.5), 22D (BA.2.75) e 22F (XBB), mentre le minori risiedono nei *clades* 22E (BQ.1) e 23I (BA.2.86).

La spiegazione per le distanze nucleotidiche intra-mese è dovuta al numero di varianti diverse circolanti in ogni mese. Infatti, nel mese di gennaio predominano le varianti riferite alle Omicron BA.5 che quindi non si discostano molto l'un l'altra; tra febbraio e marzo inizia a salire la divergenza in quanto entrano in gioco le varianti ricombinanti del *clade* 22F (XBB) e le sue

derivante. Mentre negli ultimi mesi del 2023 emerge la nuova variante del *clade* 23I (BA.2.86), che presenta una forte differenza in termini di nucleotidi rispetto alle varianti in circolazione nello stesso mese (per lo più ricombinanti XBB), portando così ad un aumento della divergenza nel gruppo in esame.

Per quanto riguarda le differenze nucleotidiche intraclade, è da notare come i *clades* con minore divergenza interna sono rappresentati per lo più da varianti nuove o omogenee, come nel caso della 23I (BA.2.86), mentre la maggior divergenza riguarda *clades* con una maggiore eterogeneità di *lineages* al loro interno, come il 22B (BA.5), il 22D (BA.2.75) e il 22F (XBB).

Dall'analisi delle distanze nucleotidiche e amminoacidiche dei singoli geni, è stata confermata la già nota tendenza ad accumulare mutazioni amminoacidiche nel gene della proteina *spike* e a differenziarsi rispetto le altre proteine del virus. Infatti, il gene S è sottoposto a una costante selezione positiva, in quanto è il bersaglio dell'attività neutralizzante degli anticorpi prodotti sia in seguito a vaccinazione che a infezione naturale. A destare più interesse sono le mutazioni che ricadono nel *Receptor Binding Domain* (RBD), in quanto possono conferire al virus una maggiore affinità per il recettore cellulare *Angiotensin-Converting Enzyme 2* (ACE2), oltre alla resistenza verso specifici anticorpi monoclonali, spesso usati per il trattamento del COVID-19. Altre proteine virali a presentare un maggior rapporto di mutazioni non sinonime rispetto alle sinonime, indicando quindi una selezione positiva nel rispettivo gene, sono la proteina *Envelope* (E), la proteina non strutturale 15 (nsp15) e la ORF10.

La proteina E è una piccola proteina strutturale che si colloca in vicinanza della proteina *spike* nel rivestimento di SARS-CoV-2, oltre ad essere un'importante viroporina implicata nella maturazione del virione. Si può ipotizzare che le mutazioni a carico di questa proteina siano funzionali per il mantenimento delle

corrette interazioni elettrostatiche con la vicina proteina *spike* sulla superficie virale.

La proteina nsp15 è un'endoribonucleasi e ha un ruolo nell'evadere i sistemi di riconoscimento del genoma virale da parte della cellula ospite. In virtù di questa sua funzione, la comparsa di mutazioni non sinonime nella sequenza amminoacidica possono comportare un vantaggio indiretto nella replicazione virale.

La ORF10 è una breve sequenza nucleotidica presente al 3' del genoma virale e codifica per una piccola proteina di circa 38 amminoacidi. Tutt'oggi non sono ben chiare le funzioni di questa proteina; perciò, sono necessari ulteriori studi per poter comprendere le implicazioni e capirne l'evoluzione.

È possibile che le mutazioni in entrambi questi geni, della nsp15 e della ORF10, rappresentino il perfezionamento del salto di specie all'uomo.

Le relazioni filogenetiche che intercorrono fra le sequenze in esame hanno messo in evidenza la divisione fra i *clades* 22B (BA.5) e 22E (BQ.1), derivanti dalle BA.5, e i restanti *clades* derivanti dalle BA.2. Di quest'ultimi i *clusters* direttamente discendenti dalle Omicron BA.2 sono rappresentati dalle varianti BA.2.75, che nel 2023 hanno circolato limitatamente nel mese di gennaio, e le BA.2.86, che invece fanno la loro comparsa negli ultimi mesi dell'anno. Sempre derivanti dalle Omicron BA.2, ci sono i *clusters* formati dalle varianti ricombinanti, XBB, che presentando una forte similarità, si suddividono in tanti nodi dove si vengono a trovare sequenze appartenenti a *lineages* differenti, ma affini.

Ai fini del monitoraggio delle varianti di SARS-CoV-2 è importante ricercare il significato biologico delle mutazioni riscontrate nei vari *clades*. Dall'analisi è emerso come molte mutazioni di interesse clinico sono localizzate nella proteina *spike*, la maggior parte concentrate nella regione del *Receptor Binding*

*Domain* (RBD) della molecola. Tali mutazioni per lo più conferiscono una minor suscettibilità verso specifici anticorpi monoclonali, spesso usati come trattamento terapeutico.

È stata anche riscontrata l'acquisizione di un maggior numero di mutazioni nell'RBD nella variante BA.2.86 e ciò potrebbe spiegare l'elevata diffusione di questa variante dalla sua comparsa.

Si ipotizza che alla base della grande diversità genetica della variante BA.2.86, rispetto agli altri *lineages*, sia dovuta ad una co-circolazione del virus in altre specie animali, mediante un evento di zoonosi inversa. Questa ipotesi è stata inizialmente suggerita per spiegare la comparsa della variante Omicron, emersa verso la fine del 2021. Così come la BA.2.86, la variante Omicron emerse con un numero di mutazioni molto elevato rispetto alle varianti precedenti, ossia 50 mutazioni in più di cui 30 nella proteina *spike*.

Infatti, è nota la capacità di SARS-CoV-2 di infettare anche diverse specie animali e ciò può rappresentare un'opportunità per il virus di andare incontro ad eventi di ricombinazione o di acquisire nuove mutazioni vantaggiose, permettendo la comparsa di nuove varianti che possono essere ri-trasmesse all'uomo a seguito di un successivo salto di specie.<sup>62,63</sup>

Differentemente dalla proteina *spike*, altre porzioni del genoma di SARS-CoV-2 sono molto più conservate, soprattutto se implicate in funzioni essenziali per il virus, come la replicazione e la maturazione virale; pertanto, la fissazione di nuove mutazioni in questi geni risulta ridotta, tranne per quelle in grado di apportare un vantaggio evolutivo al virus.

In conclusione, la continua comparsa di nuove varianti di SARS-CoV-2 porta con sé la necessità di un monitoraggio costante di questo virus nel territorio al fine di predisporre misure di prevenzione e di contenimento del virus.



La comprensione delle caratteristiche genetiche che possono potenzialmente avere un impatto sulla trasmissibilità del virus, sulla capacità elusiva delle risposte immunitarie o sulla resistenza ai farmaci in uso, è alla base per lo sviluppo di nuovi trattamenti più efficaci o sulla progettazione di nuovi vaccini.

# Fonti Bibliografiche

1. V'kovski, P., Kratzel, A., Steiner, S., Stalder, H. & Thiel, V. Coronavirus biology and replication: implications for SARS-CoV-2. *Nature Reviews Microbiology* vol. 19 155–170 Preprint at <https://doi.org/10.1038/s41579-020-00468-6> (2021).
2. *785 Nidovirales Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*. (2012).
3. Parlikar, A. *et al.* Understanding genomic diversity, pan-genome, and evolution of SARS-CoV-2. *PeerJ* **2020**, (2020).
4. Cui, J., Li, F. & Shi, Z. L. Origin and evolution of pathogenic coronaviruses. *Nature Reviews Microbiology* vol. 17 181–192 Preprint at <https://doi.org/10.1038/s41579-018-0118-9> (2019).
5. Ruiz-Aravena, M. *et al.* Ecology, evolution and spillover of coronaviruses from bats. *Nature Reviews Microbiology* vol. 20 299–314 Preprint at <https://doi.org/10.1038/s41579-021-00652-2> (2022).
6. Su, S. *et al.* Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses. *Trends in Microbiology* vol. 24 490–502 Preprint at <https://doi.org/10.1016/j.tim.2016.03.003> (2016).
7. Rabaan, A. A. *et al.* SARS-CoV-2, SARS-CoV, and MERS-CoV: A Comparative Overview.
8. Rabaan, A. A. *et al.* MERS-CoV: epidemiology, molecular dynamics, therapeutics, and future challenges. *Annals of Clinical Microbiology and Antimicrobials* vol. 20 Preprint at <https://doi.org/10.1186/s12941-020-00414-7> (2021).
9. Ochani, R. K. *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, n* vol. 20 (2021).
10. Wang, M. Y. *et al.* SARS-CoV-2: Structure, Biology, and Structure-Based Therapeutics Development. *Frontiers in Cellular and Infection Microbiology* vol. 10 Preprint at <https://doi.org/10.3389/fcimb.2020.587269> (2020).
11. Farrag, M. A. *et al.* SARS-CoV-2: An overview of virus genetics, transmission, and immunopathogenesis. *International Journal of Environmental Research and Public Health* vol. 18 Preprint at <https://doi.org/10.3390/ijerph18126312> (2021).
12. Kirtipal, N., Bharadwaj, S. & Kang, S. G. From SARS to SARS-CoV-2, insights on structure, pathogenicity and immunity aspects of pandemic human coronaviruses. *Infection, Genetics and Evolution* vol. 85 Preprint at <https://doi.org/10.1016/j.meegid.2020.104502> (2020).
13. Muralidar, S., Ambi, S. V., Sekaran, S. & Krishnan, U. M. The emergence of COVID-19 as a global pandemic: Understanding the epidemiology, immune response and potential

- therapeutic targets of SARS-CoV-2. *Biochimie* vol. 179 85–100 Preprint at <https://doi.org/10.1016/j.biochi.2020.09.018> (2020).
14. Chavda, V. P., Patel, A. B. & Vaghasiya, D. D. SARS-CoV-2 variants and vulnerability at the global level. *Journal of Medical Virology* vol. 94 2986–3005 Preprint at <https://doi.org/10.1002/jmv.27717> (2022).
  15. Haque, S. K. M., Ashwaq, O., Sarief, A. & Azad John Mohamed, A. K. A comprehensive review about SARS-CoV-2. *Future Virology* vol. 15 625–648 Preprint at <https://doi.org/10.2217/fvl-2020-0124> (2020).
  16. Wu, J. *et al.* SARS-CoV-2 RNA-dependent RNA polymerase as a target for high-throughput drug screening. *Future Virology* vol. 18 51–62 Preprint at <https://doi.org/10.2217/fvl-2021-0335> (2023).
  17. Yoshimoto, F. K. The Proteins of Severe Acute Respiratory Syndrome Coronavirus-2 (SARS CoV-2 or n-COV19), the Cause of COVID-19. *Protein Journal* vol. 39 198–216 Preprint at <https://doi.org/10.1007/s10930-020-09901-4> (2020).
  18. Jackson, C. B., Farzan, M., Chen, B. & Choe, H. Mechanisms of SARS-CoV-2 entry into cells. *Nature Reviews Molecular Cell Biology* vol. 23 3–20 Preprint at <https://doi.org/10.1038/s41580-021-00418-x> (2022).
  19. Naqvi, A. A. T. *et al.* Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. *Biochimica et Biophysica Acta - Molecular Basis of Disease* vol. 1866 Preprint at <https://doi.org/10.1016/j.bbadis.2020.165878> (2020).
  20. Kadam, S. B., Sukhramani, G. S., Bishnoi, P., Pable, A. A. & Barvkar, V. T. SARS-CoV-2, the pandemic coronavirus: Molecular and structural insights. *Journal of Basic Microbiology* vol. 61 180–202 Preprint at <https://doi.org/10.1002/jobm.202000537> (2021).
  21. Malone, B., Urakova, N., Snijder, E. J. & Campbell, E. A. Structures and functions of coronavirus replication–transcription complexes and their relevance for SARS-CoV-2 drug design. *Nature Reviews Molecular Cell Biology* vol. 23 21–39 Preprint at <https://doi.org/10.1038/s41580-021-00432-z> (2022).
  22. Mihaescu, G. *et al.* SARS-CoV-2: From structure to pathology, host immune response and therapeutic management. *Microorganisms* vol. 8 1–28 Preprint at <https://doi.org/10.3390/microorganisms8101468> (2020).
  23. Malik, Y. A. *Properties of Coronavirus and SARS-CoV-2.*
  24. Scialo, F. *et al.* ACE2: The Major Cell Entry Receptor for SARS-CoV-2. *Lung* vol. 198 867–877 Preprint at <https://doi.org/10.1007/s00408-020-00408-4> (2020).
  25. Ortiz-Prado, E. *et al.* Clinical, molecular, and epidemiological characterization of the SARS-CoV-2 virus and the Coronavirus Disease 2019 (COVID-19), a comprehensive literature review. *Diagnostic Microbiology and Infectious Disease* vol. 98 Preprint at <https://doi.org/10.1016/j.diagmicrobio.2020.115094> (2020).

26. Lim, S., Zhang, M. & Chang, T. L. ACE2-Independent Alternative Receptors for SARS-CoV-2. *Viruses* vol. 14 Preprint at <https://doi.org/10.3390/v14112535> (2022).
27. Lamers, M. M. & Haagmans, B. L. SARS-CoV-2 pathogenesis. *Nature Reviews Microbiology* vol. 20 270–284 Preprint at <https://doi.org/10.1038/s41579-022-00713-0> (2022).
28. Wiersinga, W. J., Rhodes, A., Cheng, A. C., Peacock, S. J. & Prescott, H. C. Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus Disease 2019 (COVID-19): A Review. *JAMA - Journal of the American Medical Association* vol. 324 782–793 Preprint at <https://doi.org/10.1001/jama.2020.12839> (2020).
29. Amor, S., Fernández Blanco, L. & Baker, D. Innate immunity during SARS-CoV-2: evasion strategies and activation trigger hypoxia and vascular damage. *Clin Exp Immunol* **202**, 193–209 (2020).
30. Costa, L. B. *et al.* Insights on SARS-CoV-2 Molecular Interactions With the Renin-Angiotensin System. *Frontiers in Cell and Developmental Biology* vol. 8 Preprint at <https://doi.org/10.3389/fcell.2020.559841> (2020).
31. Martynowicz, H., Jodkowska, A., Poręba, R., Mazur, G. & Więckiewicz, M. Demographic, clinical, laboratory, and genetic risk factors associated with covid-19 severity in adults: A narrative review. *Dent Med Probl* **58**, 115–121 (2021).
32. Risk Factors for COVID-19.
33. Gupta, A., Marzook, H. & Ahmad, F. Comorbidities and clinical complications associated with SARS-CoV-2 infection: an overview. *Clinical and Experimental Medicine* vol. 23 313–331 Preprint at <https://doi.org/10.1007/s10238-022-00821-4> (2023).
34. Sharma, A., Ahmad Farouk, I. & Lal, S. K. Covid-19: A review on the novel coronavirus disease evolution, transmission, detection, control and prevention. *Viruses* vol. 13 Preprint at <https://doi.org/10.3390/v13020202> (2021).
35. Puhach, O., Meyer, B. & Eckerle, I. SARS-CoV-2 viral load and shedding kinetics. *Nature Reviews Microbiology* vol. 21 147–161 Preprint at <https://doi.org/10.1038/s41579-022-00822-w> (2023).
36. Fernandes, Q. *et al.* Emerging COVID-19 variants and their impact on SARS-CoV-2 diagnosis, therapeutics and vaccines. *Annals of Medicine* vol. 54 524–540 Preprint at <https://doi.org/10.1080/07853890.2022.2031274> (2022).
37. Sethuraman, N., Jeremiah, S. S. & Ryo, A. Interpreting Diagnostic Tests for SARS-CoV-2. *JAMA - Journal of the American Medical Association* vol. 323 2249–2251 Preprint at <https://doi.org/10.1001/jama.2020.8259> (2020).
38. Filchakova, O. *et al.* Review of COVID-19 testing and diagnostic methods. *Talanta* vol. 244 Preprint at <https://doi.org/10.1016/j.talanta.2022.123409> (2022).
39. AIFA - Agenzia Italiana del Farmaco. Programmi di uso compassionevole – COVID-19. <https://www.aifa.gov.it/programmi-di-uso-compassionevole-covid-19>.

40. AIFA - Agenzia Italiana del Farmaco. Uso degli antivirali per COVID-19. <https://www.aifa.gov.it/uso-degli-antivirali-orali-per-covid-19>.
41. AIFA - Agenzia Italiana del Farmaco. Uso degli anticorpi monoclonali per COVID-19. <https://www.aifa.gov.it/uso-degli-anticorpi-monoclonali>.
42. *RACCOMANDAZIONI AIFA SUI FARMACI per La Gestione Domiciliare Di COVID-19*. <https://www.ema.europa.eu/en/documents/product->.
43. EpiCentro - L'epidemiologia per la sanità pubblica. I vaccini anti COVID-19 . <https://www.epicentro.iss.it/vaccini/covid-19#writers>.
44. AIFA - Agenzia Italiana del Farmaco. Vaccini COVID-19. <https://www.aifa.gov.it/vaccini-covid-19>.
45. Markov, P. V. *et al.* The evolution of SARS-CoV-2. *Nature Reviews Microbiology* vol. 21 361–379 Preprint at <https://doi.org/10.1038/s41579-023-00878-2> (2023).
46. Plante, J. A. *et al.* Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* **592**, 116–121 (2021).
47. Lai, A., Bergna, A., Ventura, C. Della & Zehender, G. Genomic epidemiology and phylogenetics applied to the study of SARS-CoV-2 pandemic. *New Microbiologica* **46**,.
48. Flores-Vega, V. R. *et al.* SARS-CoV-2: Evolution and Emergence of New Viral Variants. *Viruses* vol. 14 Preprint at <https://doi.org/10.3390/v14040653> (2022).
49. Rambaut, A. *et al.* A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol* **5**, 1403–1407 (2020).
50. European Centre for Disease Prevention and Control. Surveillance and disease data on COVID-19.
51. Tamura, T. *et al.* Virological characteristics of the SARS-CoV-2 XBB variant derived from recombination of two Omicron subvariants. *Nat Commun* **14**, (2023).
52. 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* vol. 20 151–162 Preprint at <https://doi.org/10.7150/ijms.79116> (2023).
53. CoVariants. Overview of Variants/Mutations.
54. Focosi, D., McConnell, S. & Casadevall, A. The Omicron variant of concern: Diversification and convergent evolution in spike protein, and escape from anti-Spike monoclonal antibodies. *Drug Resistance Updates* vol. 65 Preprint at <https://doi.org/10.1016/j.drug.2022.100882> (2022).
55. Zhou, Y., Zhi, H. & Teng, Y. The outbreak of SARS-CoV-2 Omicron lineages, immune escape, and vaccine effectivity. *Journal of Medical Virology* vol. 95 Preprint at <https://doi.org/10.1002/jmv.28138> (2023).

56. Koelle, K., Martin, M. A., Antia, R., Lopman, B. & Dean, N. E. The changing epidemiology of SARS-CoV-2. *Science* vol. 375 1116–1121 Preprint at <https://doi.org/10.1126/science.abm4915> (2022).
57. BBC. Coronavirus: Italy extends emergency measures nationwide.
58. Lai, A. *et al.* Epidemiological and Clinical Features of SARS-CoV-2 Variants Circulating between April–December 2021 in Italy. *Viruses* **14**, (2022).
59. Tecnica, N., Bella, A., Boros, S., Bressi, M. & Del Manso, M. *Infezioni Da SARS-CoV-2, Ricoveri e Decessi Associati a COVID-19 Direttamente Evitati Dalla Vaccinazione Italia, 27 Dicembre 2020-31 Gennaio 2022 Gruppo Di Lavoro ISS e Ministero Della Salute 'Sorveglianza Vaccini COVID-19' ISS. Emanuela BALOCCHINI.*
60. EpiCentro - L'epidemiologia per la sanità pubblica. Dati della Sorveglianza integrata COVID-19 in Italia. <https://www.epicentro.iss.it/coronavirus/sars-cov-2-dashboard>.
61. worldometers. COVID-19 CORONAVIRUS PANDEMIC. <https://www.worldometers.info/>.
62. Pourbagher-Shahri, A. M., Mohammadi, G., Ghazavi, H. & Forouzanfar, F. Susceptibility of domestic and companion animals to SARS-CoV-2: a comprehensive review. *Trop Anim Health Prod* **55**, (2023).
63. Thakur, V. & Ratho, R. K. OMICRON (B.1.1.529): A new SARS-CoV-2 variant of concern mounting worldwide fear. *Journal of Medical Virology* vol. 94 1821–1824 Preprint at <https://doi.org/10.1002/jmv.27541> (2022).