



UNIVERSITA' POLITECNICA DELLE MARCHE
DIPARTIMENTO DI SCIENZE DELLA VITA E DELL'AMBIENTE

Corso di Laurea
Scienze Biologiche

La variabilità genomica all'interno delle sequenze del Coronavirus

Genomic variance of the 2019-nCoV Coronavirus

Relazione finale di
Di Lorenzo Diletta

Docente Referente
Chiar.mo Prof.
Marco Barucca

Sessione Autunnale Dicembre 2020
A.A 2019\2020

Genomic variance of the 2019-nCoV coronavirus

Carmine Ceraolo | Federico M. Giorgi 

RIASSUNTO

In questa relazione è proposto un impiego della variabilità genetica, nello studio delle sequenze genomiche del Coronavirus. In particolar modo, l'analisi filogenetica e l'analisi delle componenti principali, sono coinvolte nel processo di determinazione evolutiva e di somiglianza genomica tra il Coronavirus e i così detti Bat-Coronavirus, come conferma di un'origine zoonotica del Coronavirus.

Nello specifico, i genomi condividono un'identità di sequenza superiore al 99%¹ un dato, che può portare alla realizzazione di un'eventuale terapia farmacologica tale da risultare efficace per tutti.

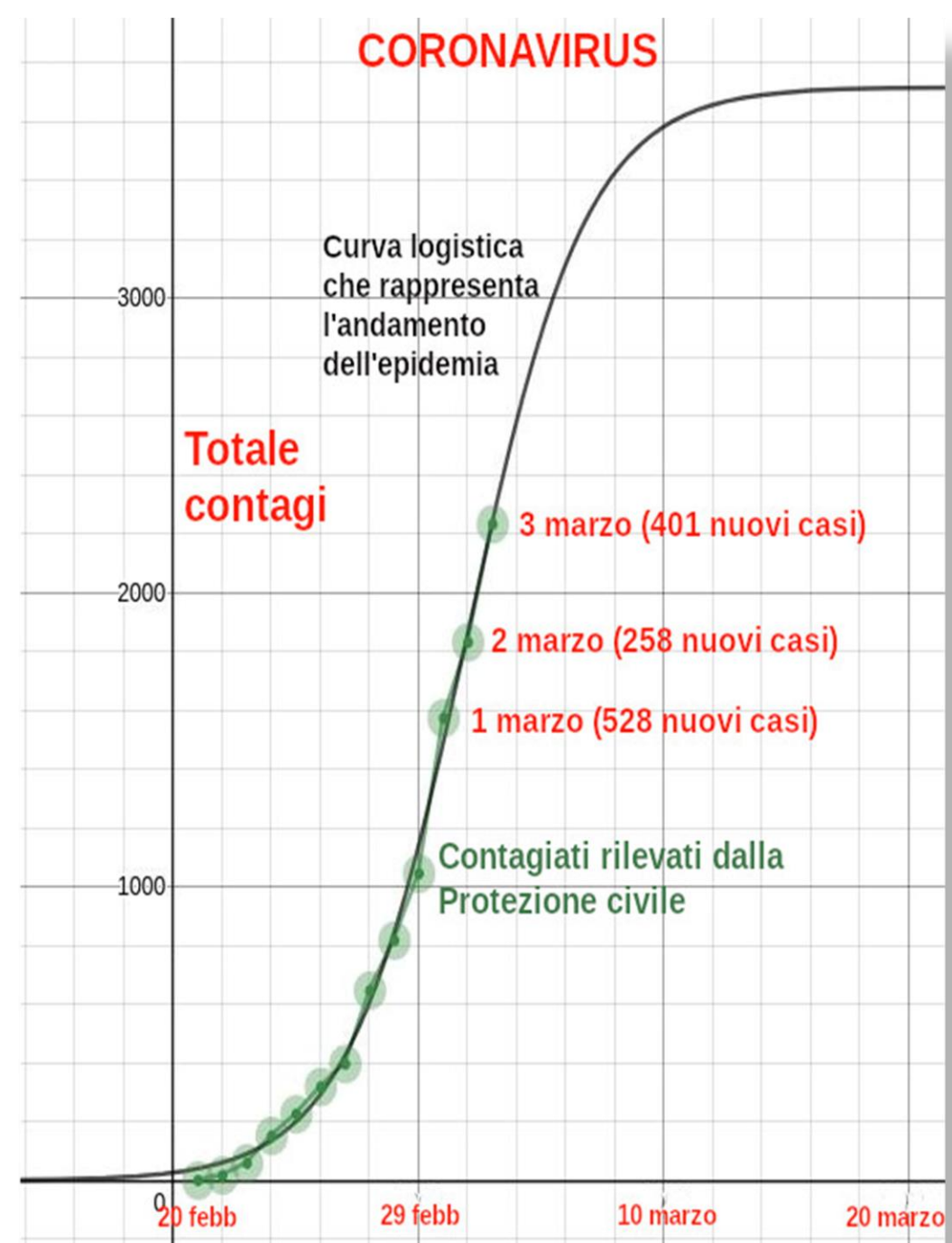
Infine, nonostante la bassa eterogeneità del genoma virale è presente una regione di elevata variabilità nelle proteine del virus, rappresentata da due sottotipi virali. Questi differiscono per un singolo aminoacido, in grado di cambiare sequenza e struttura nella proteina accessoria ORF-8, componente propria del virus.

INTRODUZIONE

È quasi un anno che un nuovo Coronavirus, chiamato Sars-CoV-2 rappresenta il principale agente causale della malattia respiratoria, denominata Covid-19.

Si è diffuso rapidamente in tutto il mondo, tanto che l'Organizzazione mondiale della sanità, ha definito la diffusione del Covid-19 non più una epidemia confinata ad alcune zone geografiche, ma una pandemia diffusa in tutto il pianeta. ²

Sono stati infatti registrati 14.441 casi di polmonite, provenienti dalla provincia di Wuhan, in Cina, seguiti dai 315 decessi in data 2 febbraio 2020, e più di 118 casi gravi in altri paesi. ³



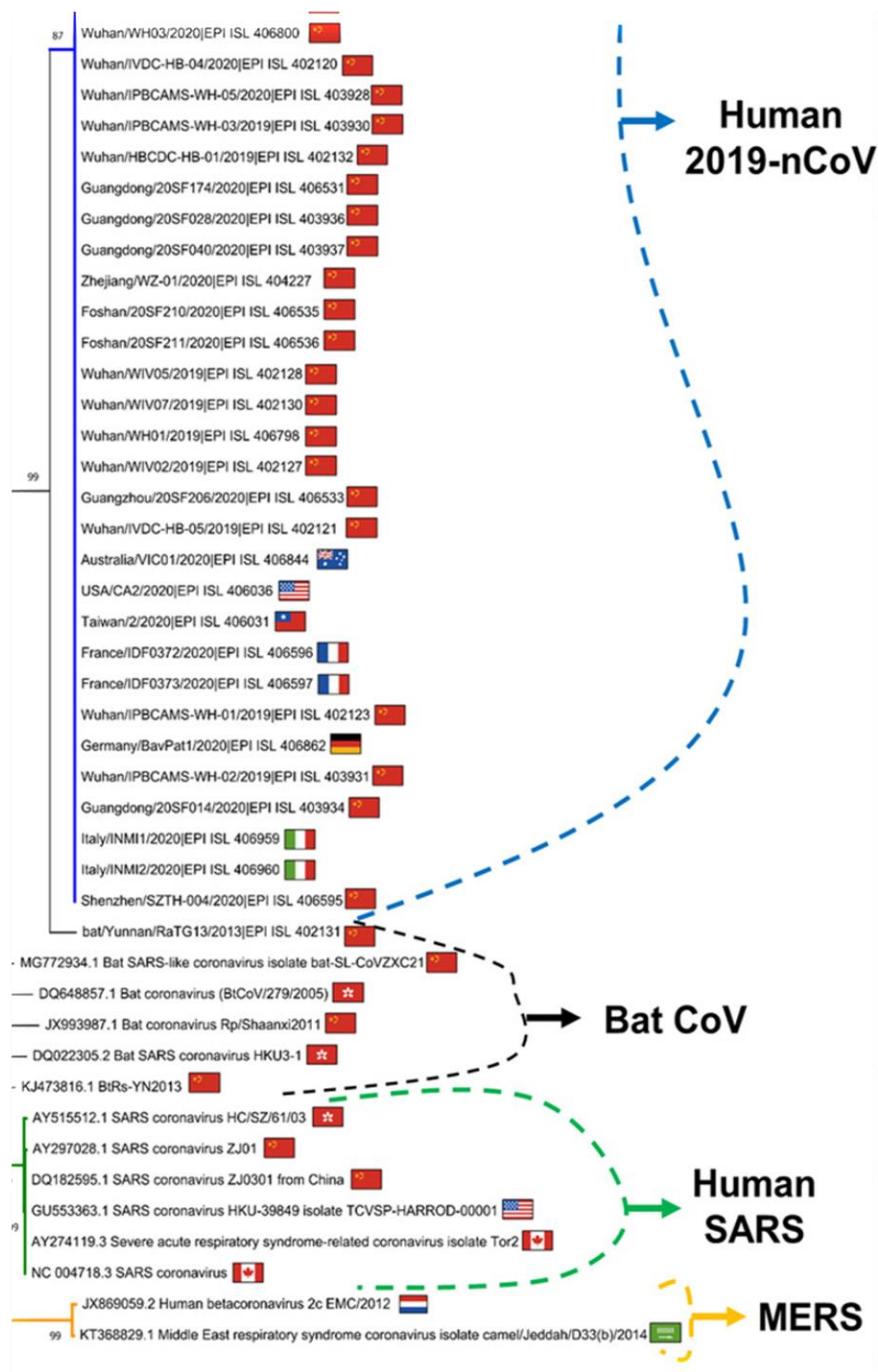
ANALISI FILOGENETICA

Prevede la realizzazione di un albero filogenetico,⁴ che abbia come scopo quello di comparare i genomi di alcuni virus isolati, ed evidenziare anche, come unico filo conduttore, le possibili relazioni evolutive tra queste specie virali, alla base del concetto di affinità e discendenza.

Sono distinte le seguenti sequenze genomiche:

- 53 sequenze genomiche complete del Coronavirus umano
- 1 sequenza genomica, riferita all'isolato caso di Wuhan, Cina
- 2 sequenze genomiche riferite ai soli 2 casi italiani, ottenute dall'ospedale Spallanzani di Roma.

Confrontate con sequenze genomiche relative al Coronavirus del Pipistrello (Bat CoV), al SARS-CoV e al MERS-CoV.



Esistono, pertanto delle relazioni evolutive fra queste sequenze genomiche, infatti :

- 1) Tutti i genomi umani del Coronavirus sembrano essere molto simili fra loro, nonostante le diverse posizioni di campionamento, quali Inghilterra, Cina, Australia, Italia, e Germania.
- 2) Tra i Bat Coronavirus, vi è un virus che infetta i pipistrelli della specie *Rhinolophus affinis* e che con il Coronavirus umano condivide una identità di sequenza del 96%. Questo conferma l'idea di un trasferimento zoonotico dai pipistrelli agli essere umani.
- 3) Mentre i genomi SARS e MERS sono i più geneticamente distinti dalle altre sequenze.

ANALISI DI DIVERGENZE GENOMICHE

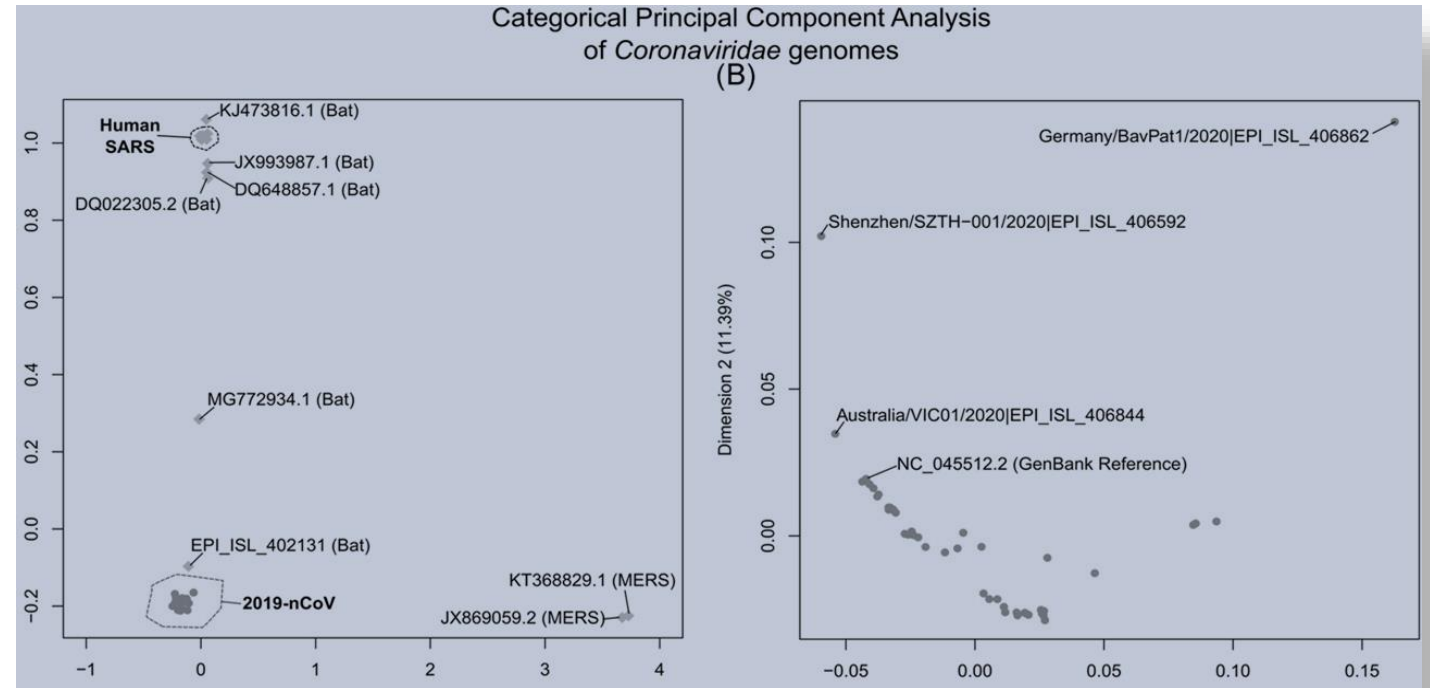
È basata su un metodo utile per classificare le sequenze genomiche, chiamato CATPCA. ⁵

E' un'analisi riguardante le componenti principali dei genomi dei Coronavirus, distinte in categorie, e descrive ed evidenzia i punti di coordinate che presentano la variabilità genomica più alta.

Come mostra la figura, sono rappresentate le sequenze genomiche del Coronavirus, per le quali si evidenzia una certa variabilità interna, riferite ai genomi :

1) EPI_ISL_406862 (raccolto in Germania)

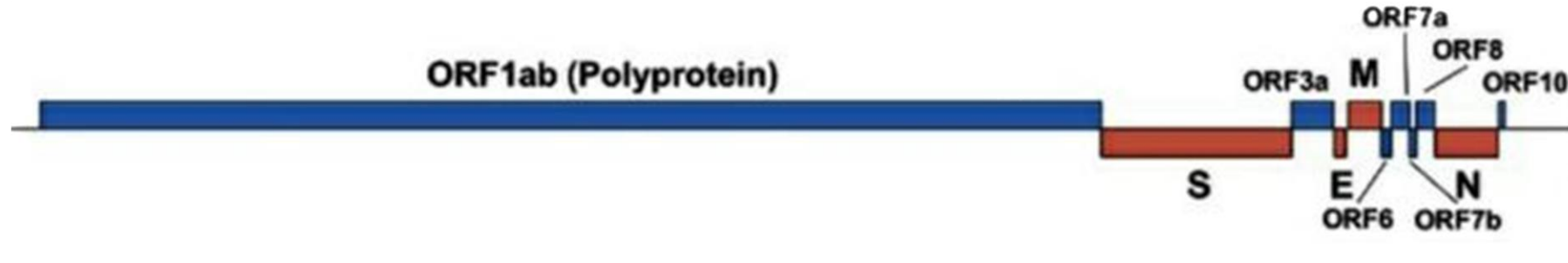
2) EPI_ISL_406592 (raccolti a Shenzhen, in Cina).



STIMA DELLA VARIABILITÀ GENOMICA

Prevede un'attenta analisi finalizzata a stabilire quali possibili variazioni possono presentarsi all'interno di una determinata sequenza genomica.

La sequenza del nuovo Coronavirus, mostra la stessa disposizione genomica della specie SARS.⁶



Come mostra la figura, il gene che codifica per una proteina, ORF1ab, presenta all'estremità 5', quattro principali geni strutturali che codificano per le proteine strutturali, quali:

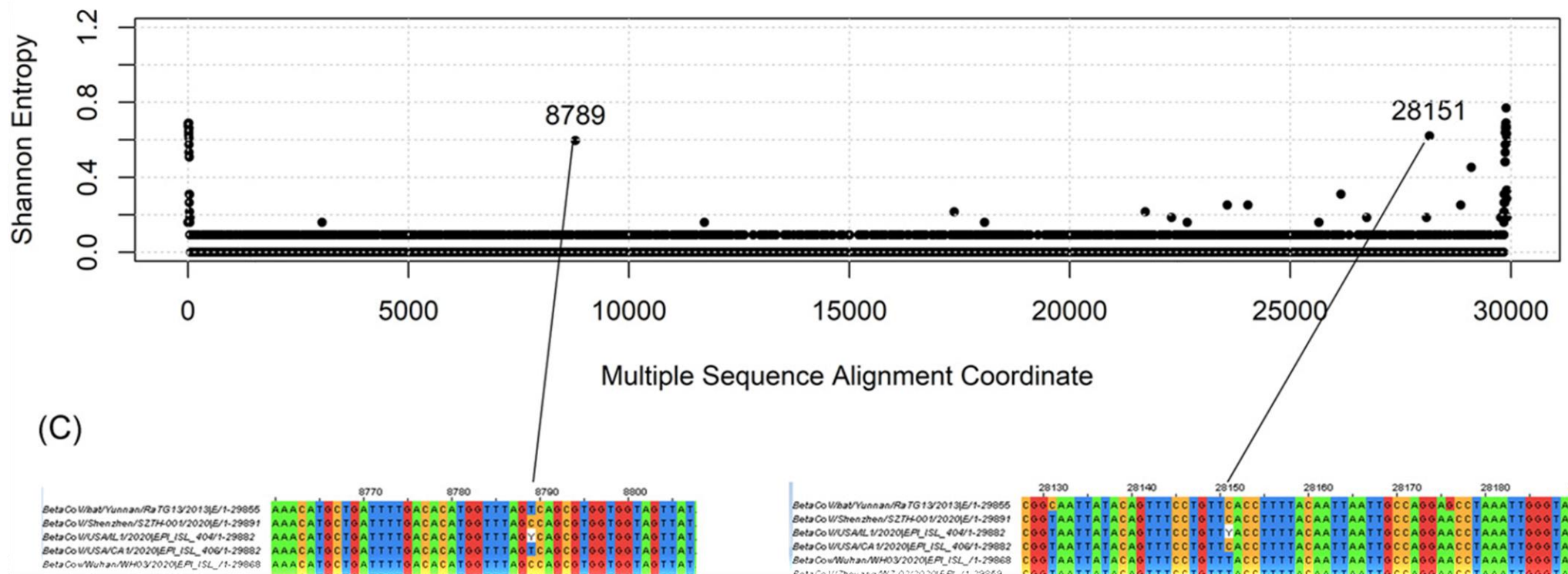
- 1) la proteina Spike (S);
- 2) la proteina dell'involucro o envelope (E);
- 3) la proteina (M) della membrana;
- 4) la proteina (N) rappresentante il nucleocapside virale.

Il gene è formato anche da altre sei proteine accessorie come ORF3a, ORF6, ORF7a e b, ORF8 e ORF10.

ENTROPIA DI SHANNON

È utilizzata come metodo per misurare la variabilità delle sequenze genomiche, in base alla loro posizione all'interno del gene di riferimento.⁷

Le coordinate di posizione risultate con il più alto valore di entropia sono due, come mostra la figura: 8789 e 28151.



8789 : non determina differenze a livello fenotipico nei diversi ceppi.

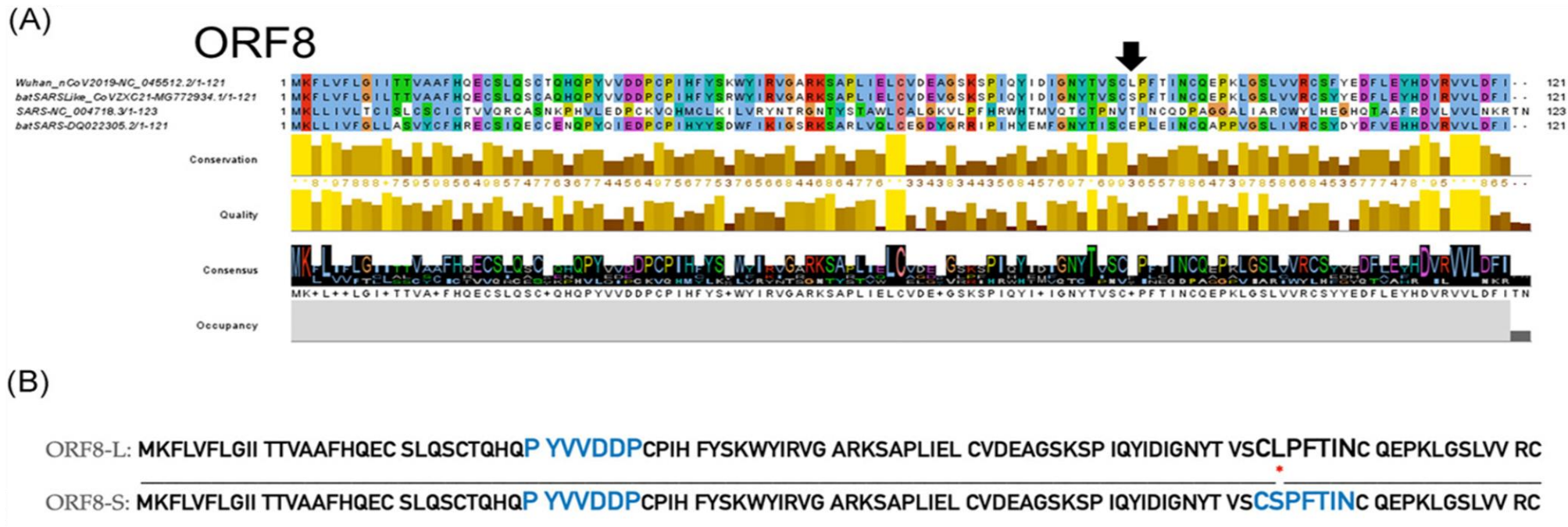
28151 : è presente nel gene della proteina ORF-8, causando un cambiamento amminoacidico tra la SERINA e la LEUCINA che può a sua volta influenzare la conformazione del peptide, dato che la serina è un amminoacido polare e la leucina non polare, quindi la sua struttura terziaria, dalla quale dipende in modo essenziale l'attività biologica della proteina stessa.

ANALISI DELLE ISOFORME : ORF8-L \ ORF8-S

Prevede l'utilizzo di un metodo basato sull' algoritmo di Russell-Linding,⁸ con lo scopo di evidenziare quali sequenze genomiche differiscono tra loro, in base al loro disordine strutturale.

In particolare è l' isoforma ORF-S, che induce un cambiamento strutturale nella proteina virale.

Come è mostrato in figura, l' allineamento delle due isoforme ORF8-L e ORF8-S, evidenzia come queste possano variare nella posizione dell' amminoacido 84 (indicato in figura con un punto rosso).



ANALISI DI CONSERVAZIONE PROTEICA

Prevede un confronto analitico, riguardo tutte le proteine codificate sia dal Coronavirus, sia dai virus appartenenti alla classe Beta, al fine di evidenziare quali tra queste si conservano nel corso dell'evoluzione.

Si nota, infatti un'elevata conservazione nelle proteine strutturali, soprattutto la proteina E, e la proteina M, del Bat-CoV rispetto alle proteine accessorie come ORF-8, rappresentate da un basso livello di conservazione.

Risulta mediamente, che il Coronavirus condivide circa il 91,1% delle sequenze proteiche con la sequenza MG772934 del virus del pipistrello, il 79,7% con il Bat-virus, e il 77% con il genoma di riferimento per SARS.

Gene	Pipistrello MG772934.1	Bat DQ022305.2	SARS NC_004718,3
ORF1ab (poliproteina)	95.15%	85.78%	86.12%
S (Spike)	80.32%	76.04%	75.96%
E (envelope)	100%	94.74%	94.74%
M (Membrana)	98.65%	90.99%	90.54%
ORF8	94.21%	57.02%	30.16%
N (Nucleocapside)	94.27%	89.55%	90.52%

CONCLUSIONI

Alla luce di quanto dettosi, riguardo la possibile variabilità genomica, si può rispondere che vi è un alto livello di conservazione genetica all'interno dei genomi del Coronavirus sequenziati, nonostante siano presenti due sequenze nucleotidiche con alta variabilità.

Queste analisi, mostrano dunque una netta origine evolutiva del Coronavirus in questione, dagli altri Coronavirus appartenenti al raggruppamento Beta, in modo particolare i Bat-coronavirus, SARS-CoV e MERS-CoV.

In conclusione, questi risultati, permettono maggiormente di porre in modo più dettagliato la base per una possibile ed efficace ricerca, implicata nel trattamento dell'infezione data da SARS-CoV-2.

BIBLIOGRAFIA

1. Lu R, Zhao X, Li J, et al. *La Lancetta*.
2. Coordinamento scientifico a cura del Centro Nazionale per la Prevenzione delle malattie e la Promozione della Salute, CNAPPS – ISS.
3. Shu Y, McCauley J. Iniziativa globale sulla condivisione di tutti i dati sull'influenza, dalla visione alla realtà.
4. Studio di Carmine Ceraolo e Federico M. Giorgi sulle possibili relazioni evolutive di sequenze genomiche dei Coronavirus.
5. Le S, Joss J, Husson F. FactoMineR: un pacchetto R per l'analisi multivariata. *J Stat Softw*. 2008.
6. Narayanan K, Huang C, Makino S. Proteine accessorie coronavirus SARS. *Virus Res*. 2008.
7. Manaresi E, Conti I, Bua G, Bonvicini F, Gallinella G. Un genoma sintetico Parvovirus B19: caratteristiche di sequenza e competenza funzionale. *Virologia*. 2017.
8. Linding R. GlobPlot: esplorare sequenze proteiche per globularità e disordine. *Acidi nucleici Res*. 2003.