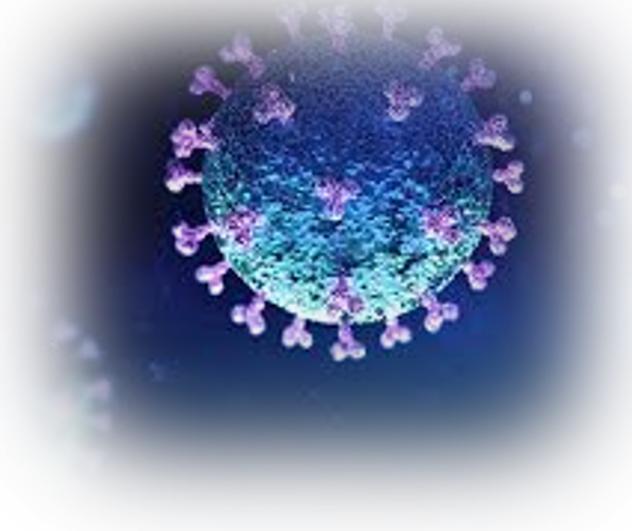


La circolazione dei genotipi di SARS-COV-2 in Sicilia: i ceppi importati con i migranti

S. Reale



Popolazione oggetto di studio



Il 21 giugno, una nave di soccorso di un'organizzazione non governativa (ONG) ha salvato 210 migranti vicino al confine libico ed è arrivata in Italia a Porto Empedocle, nel sud della Sicilia.

Dei 210 migranti, 68 (32,3%) erano bambini o adolescenti. L'età media era di 24 anni. Uno dei migranti, che presentava febbre e sintomi respiratori, era in cura per la tubercolosi e trasferito in ospedale, il suo tampone rinofaringeo è risultato positivo al test molecolare SARS-CoV-2.

Sono state attivate, quindi, le misure di quarantena e, dopo lo screening molecolare dei tamponi rinofaringei dei restanti 209 migranti, 28 (13 uomini, 9 donne e 6 bambini) sono risultati positivi (13,4%).

Di questi, 8 provenivano dal **Camerun**, 5 dalla **Guinea Conakry**, 3 dal **iMal**, 2 dalla **Costa d'Avorio**, **Sierra Leone** e **Somalia**, mentre 1 rispettivamente da **Nigeria**, **Togo**, **Senegal**, **Ghana**, **Liberia** e **Bangladesh**.

Nessuno di tali migranti ha presentato o sviluppato segni o sintomi indicativi di COVID-19 durante il periodo di quarantena.

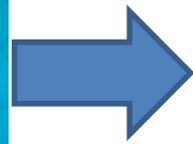
Abbiamo studiato i polimorfismi del genoma virale di SARS-CoV-2 ottenuto da campioni rinofaringei di migranti in arrivo in Sicilia attraverso il Mar Mediterraneo, seguendo la rotta libica, e ospitati in centri di accoglienza dedicati.



Sequenziamento genoma di SARS-CoV-2

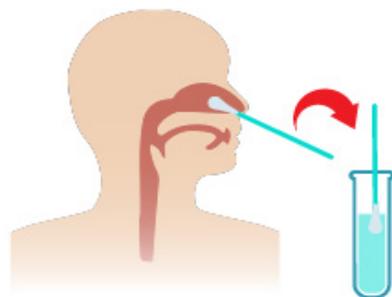


18 CAMPIONI

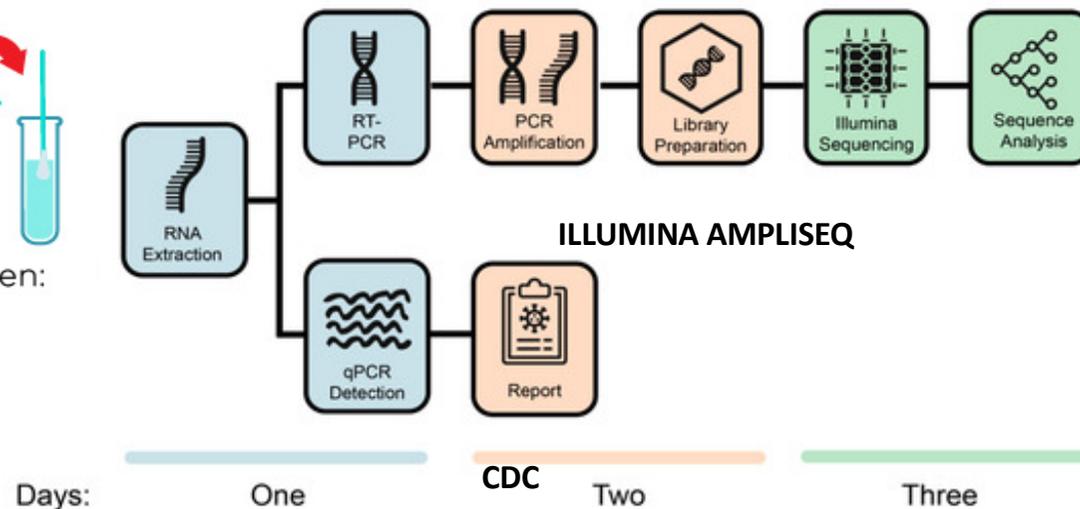


PROTOCOL

Il sequenziamento dell'intero genoma virale è stato condotto in accordo con la procedura di sequenziamento NGS (Next Generation Sequencing) per ampliconi, Ampliseq (Illumina), previa retrotrascrizione del RNA estratto. Le librerie di ampliconi sono state quindi amplificate impiegando il cocktail di primers della Thermo, purificate, quantificate e sequenziate su piattaforma MiSeq (Illumina).

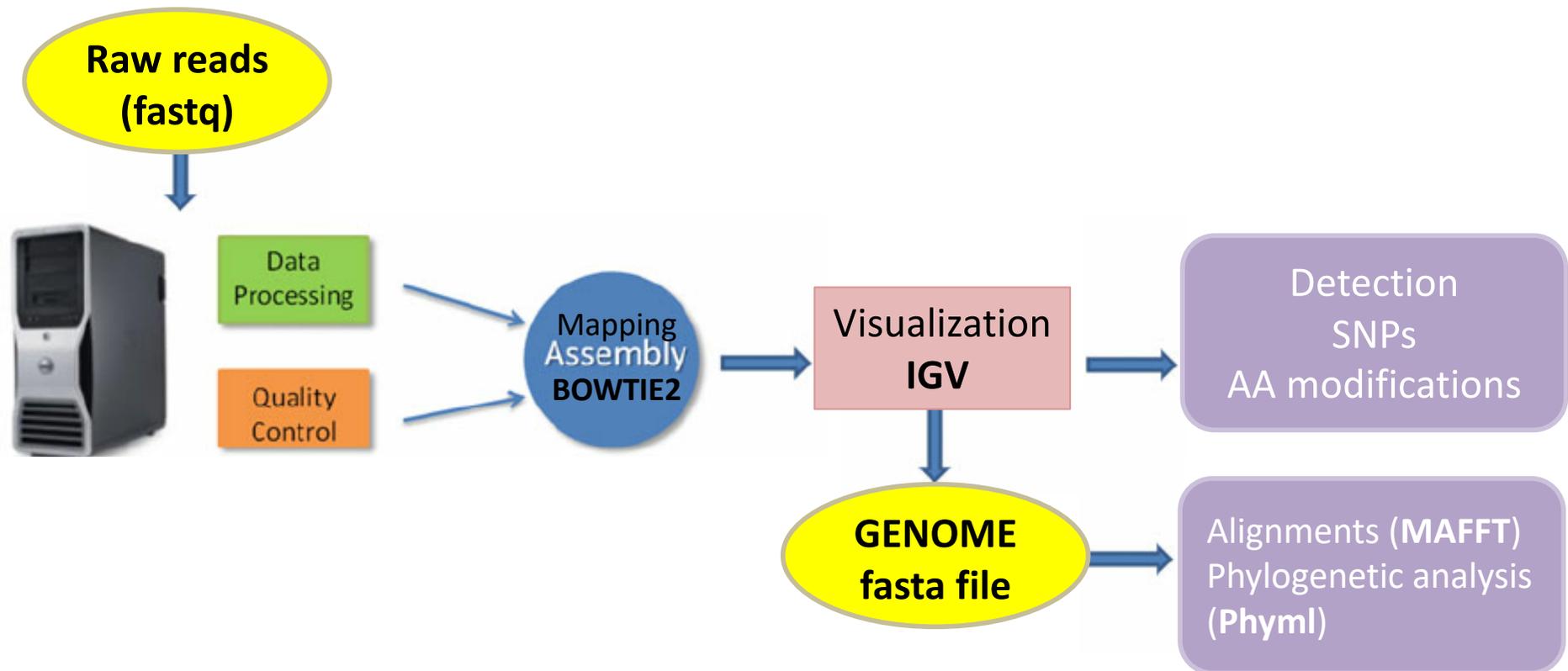


Obtain Specimen:
Swab.



I file fastq ottenuti dal sequenziamento sono stati filtrati secondo parametri di qualità e le reads mappate con il software **Bowtie2**, rispetto al genoma virale di riferimento di Wuhan-Hu-1 (GenBank: NC 045512.2), per ottenere sequenze genomiche complete del virus.

I dati genomici sono stati visualizzati dal software **IGV 2.8.0** al fine di studiare i polimorfismi a singolo nucleotide (SNP). Gli amminoacidi (AA) potenzialmente determinanti variabilità nelle proteine derivate rispetto al riferimento di Wuhan, sono stati studiati mediante ispezione visiva degli allineamenti.



Analisi filogenetica

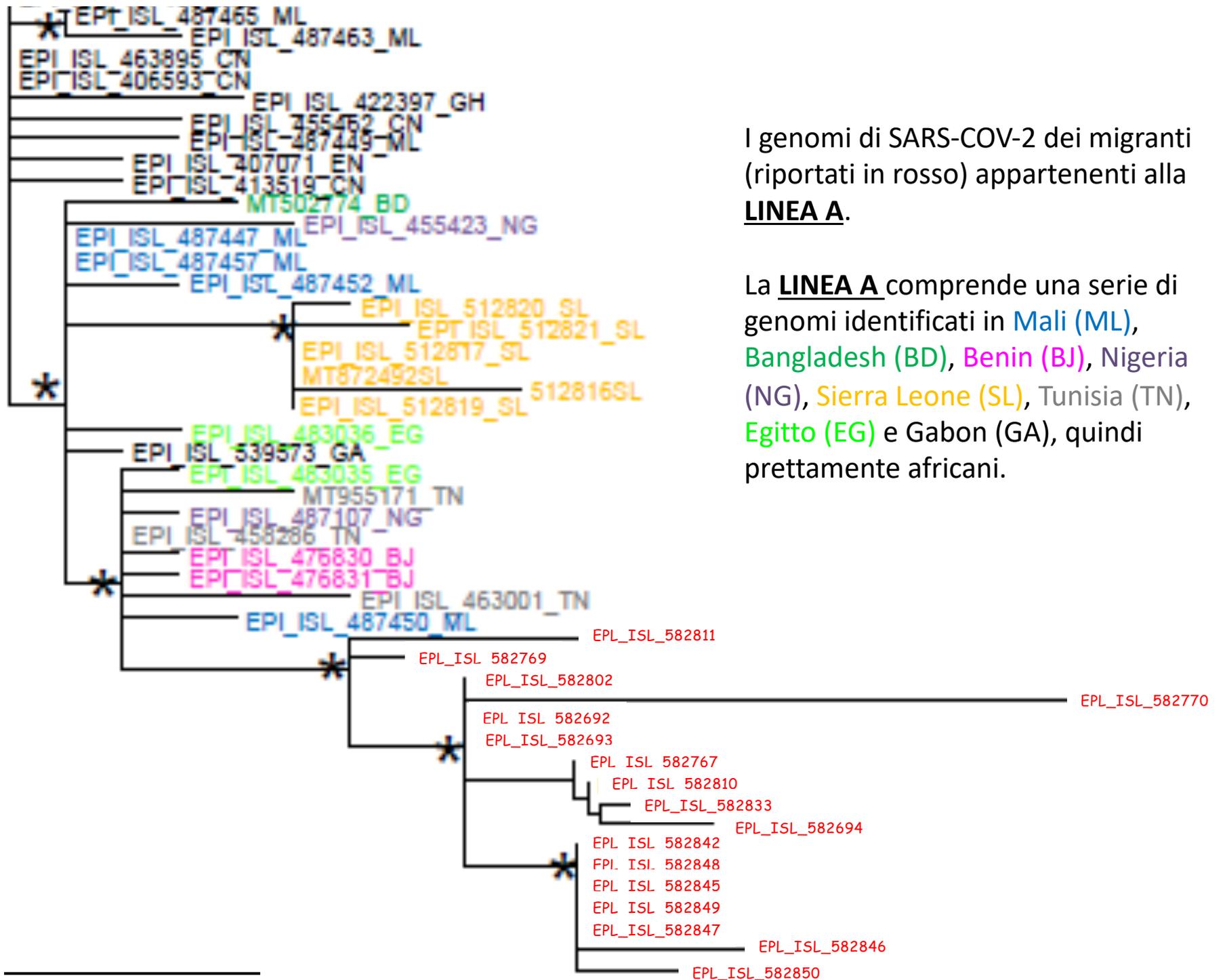
Per valutare la distanza genetica delle linee di virus circolanti nelle popolazioni migranti oggetto di studio, i genomi sono stati analizzati in un contesto filogenetico insieme a genomi completi SARS-CoV-2 di diversi paesi, recuperati dalle banche dati internazionali **GISAID** e GenBank, includendo anche la sequenza di riferimento dell'isolato **Wuhan-Hu-1**.

L'allineamento delle sequenze nucleotidiche è stato eseguito utilizzando MAFFT v.7, sulla piattaforma Galaxy, ed è stato modificato manualmente tramite il programma Bioedit.

I diciotto genomi SARS-CoV-2 provenienti dai migranti sono stati classificati attraverso il "**Pangolin COVID-19 Lineage Assigner**".

Gli alberi filogenetici sono stati creati tramite il metodo di massima verosimiglianza (ML), ed elaborati con **Phyml v3.0**. con algoritmi di tipo bayesiano di aLRT (aBayes).

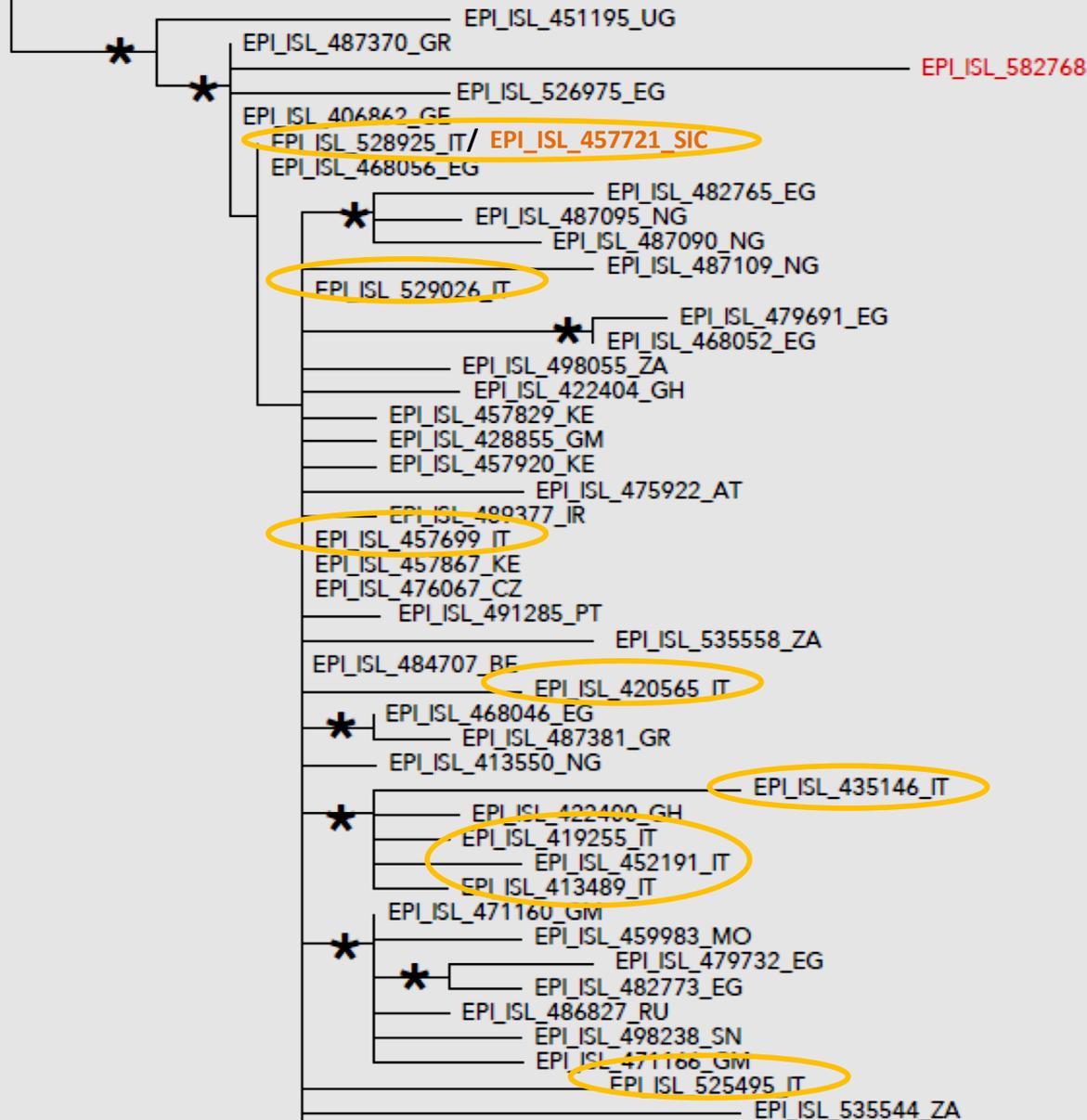
L'analisi filogenetica dei dati genomici ha mostrato che la maggior parte delle sequenze virali ottenute dai migranti (17/18) si sono localizzate nella LINEA A, mentre una sola sequenza, EPI_ISL_582768, si è separata da tutte le altre, localizzandosi nella LINEA B-B.1.



I genomi di SARS-COV-2 dei migranti (riportati in rosso) appartenenti alla **LINEA A**.

La **LINEA A** comprende una serie di genomi identificati in Mali (ML), Bangladesh (BD), Benin (BJ), Nigeria (NG), Sierra Leone (SL), Tunisia (TN), Egitto (EG) e Gabon (GA), quindi prettamente africani.

2.0E-4



Unico genoma di SARS-COV-2 dei migranti (riportato in rosso) appartenente alla **LINEA B-B.1**.

La **LINEA B-B.1** comprende genomi di varie provenienze geografiche tra cui l'Egitto (EG) e l'Uganda (UG) ma quelli più affini risultano provenienti dall'Europa: **Italia (IT)**, Grecia (GR) e Germania (GE).

EPI_ISL_457721_SIC
Genoma isolato in Sicilia

Mutazioni

Più in dettaglio, 14 SNP sono stati identificati nel 100% dei genomi localizzati nella LINEA A (in rosso nella tabella), cinque dei quali non hanno comportato un cambiamento di sequenze amminoacidiche.

Questi SNP hanno determinato una mutazione degli aminoacidi di *nsp3*, nella RNA polimerasi RNA-dipendente, in *ORF8*, e due mutazioni nella fosfoproteina nucleocapsidica.

Due SNP sono stati identificati nel 50% delle sequenze osservate, generando mutazioni in *nsp3* e in *nsp8*.

Sette SNP sono stati identificati solo in un genoma (7%, $n = 1/14$) e quattro di questi non hanno determinato cambiamenti di AA. Tra quelli che coinvolgono variazioni di AA, abbiamo trovato due sostituzioni in *nsp2* ed una nella proteina *spike*.

Dei 14 SNP osservati nel 100% dei campioni, 11 sono risultati strettamente caratterizzanti le sequenze provenienti dai migranti.

Infatti, abbiamo trovato solo tre SNP in comune, a diverse percentuali, con i genomi dei paesi africani presenti nella LINEA A. In particolare, gli SNP nelle posizioni nucleotidiche 361 (38%, $n = 8/21$), 8782 (95%, $n = 20/21$), 22468 (95%, $n = 20/21$).

Mutazioni rispetto al genoma di Wuhan-Hu-1

nucleotide position in Acc. Num. NC_045512	nt modif	observations	gene	product and nt position in each respective protein sequence	AA in NC_045512	AA mutation identified
361	A > G	100%	ORF1ab	nsp1 - 96 nt	32G	32G - no aa change
950	T > C	7%	ORF1ab	nsp 2 - nt 145	49Y	49H
1427	C > T	7%	ORF1ab	nsp 2 - nt 622	208H	208Y
3656	C > T	50%	ORF1ab	nsp 3- nt 937	313L	313F
5903	G > A	100%	ORF1ab	nsp3 - nt 3184	1062V	1062I
7011	C > T	100%	ORF1ab	nsp3 - nt 4292	1431A	1431V
7554	T > C	100%	ORF1ab	nsp3 - nt 4835	1612V	1612A
8327	C > T	100%	ORF1ab	nsp 3 - nt 5608	1870L	1870F
8782	C > T	100%	ORF1ab	nsp4 - nt 228	76S	76S - no aa change
8881	G > A	7%	ORF1ab	nsp 4 - nt 327	109T	109T- no aa change
12225	C > T	50%	ORF1ab	nsp8 - nt 134	45A	45V
14053	C > T	7%	ORF1ab	RNA-dependent RNA polymerase - nt 613	205L	205 L- no aa change
14064	T > C	7%	ORF1ab	RNA-dependent RNA polymerase - nt 624	208D	208D - no aa change
14177	C > T	100%	ORF1ab	RNA-dependent RNA polymerase - nt 737	246T	246I
22468	G > T	100%	S	spike - nt 906	302T	302T - no aa change
23695	T > A	100%	S	spike - nt 2133	711S	711S - no aa change
24588	C > T	7%	S	spike- nt 3026	1009T	1009I
28144	T > C	100%	ORF 8	ORF 8 protein- nt 251	84L	84S
28870	A > G	7%	N	nucleocapsid phosphoprotein - nt 597	199P	199P - no aa change
28878	G > A	100%	N	nucleocapsid phosphoprotein - nt 605	202S	202N
29100	G > A	100%	N	nucleocapsid phosphoprotein - nt 827	276R	276K
29449	G > T	100%	N	nucleocapsid phosphoprotein - nt 1176	392V	392V - no aa change
29742	G > A	100%	3' UTR - stem loop	/	/	/

Il genoma da noi registrato su GISAID come EPI_ISL_582768, e che si separa dal clade dei genomi africani, ha rivelato, rispetto al genoma di Wuhan-Hu-1, 5 SNP che non hanno determinato modifiche AA, mentre altri 4 SNP che hanno determinato il cambiamento di AA nelle posizioni 21M e 216F in *nsp3*, la mutazione 277S in *nsp6* e la mutazione 614G nella *spike*.

Nucleotide position in Acc. Num. NC_045512	nt modif	gene	product and nt position in each respective protein sequence	AA in NC_045512	AA mutation identified
241	C > T	5' UTR	/	/	/
475	C > T	ORF1ab	nsp1 - nt 210	70F	70F - no aa change
2780	G > A	ORF1ab	nsp 3- nt 61	21V	21M
3037	C > T	ORF1ab	nsp3 - nt 318	106F	106F- no aa change
3365	C > T	ORF1ab	nsp 3 - nt 646	216L	216F
7210	A > G	ORF1ab	nsp3 - nt 4491	1497K	1497K - no aa change
11801	G > A	ORF1ab	nsp6- nt 829	277G	277S
22882	T > C	S	spike- nt 1320	440N	440N- no aa change
23403	A > G	S	spike - nt 1841	614D	614G
26936	C > T	M	membrane glycoprotein - nt 414	138L	138L - no aa change

Tuttavia questo campione ha mostrato un clustering unico tra i genomi della linea B-B.1, con un profilo di mutazione caratteristico, e più affine a genomi europei o di paesi africani, che però riportavano una storia di viaggi in Europa (Yang et al., 2020).

Conclusioni

L'analisi filogenetica ha sistematicamente collocato i genomi dei migranti, ad eccezione di uno, in prossimità di genomi virali africani (LINEA A) identificati in Mali, Bangladesh, Benin, Nigeria, Sierra Leone, Tunisia, Egitto e Gabon. Tali genomi, per quanto diversi, mostrano una coerenza geografica. Nel complesso, sembra che i migranti, nella maggior parte dei casi, abbiano acquisito l'infezione da SARS-CoV-2 nel paese di origine. Tuttavia, non è possibile escludere la trasmissione del SARS-CoV-2 durante il viaggio o negli hotspot libici sovraffollati e/o in imbarcazioni illegali.

Diverse ipotesi, invece, potrebbero essere a sostegno dell'origine dell'infezione dell'unico genoma appartenente alla LINEA B-B.1. Il genoma EPI_ISL_582768, si è localizzato in un clade misto a genomi provenienti da vari paesi mostrando un profilo di mutazione caratteristico, ma più affine a genomi di paesi europei tra cui l'Italia, ed in particolare anche un genoma con provenienza siciliana (EPI_ISL_457721_SIC), e di alcuni paesi africani che però riportavano una storia di viaggi in Europa.

Questi risultati evidenziano l'importanza di un'ampia sorveglianza genomica di SARS-CoV-2 tra i richiedenti asilo che arrivano in Italia, attraverso il cancello siciliano.

Monitorare gli "hotspot" di accoglienza dei migranti, può fornire un'opportunità per migliorare la conoscenza sulla dinamica SARS-CoV-2 nelle aree geografiche "trascurate" e sulle rotte migratorie al fine di migliorare i programmi di prevenzione e controllo per le popolazioni vulnerabili.

Entrambi i profili di mutazione osservati, sia per i campioni della LINEA A che della LINEA B-B.1, sono risultati caratteristici. La variabilità genetica dovuta alla presenza di SNP associati alle diverse importanti proteine codificanti, necessita di essere attentamente monitorata in quanto ruolo cruciale nell'evoluzione di SARS-CoV-2. In particolare, le mutazioni nel gene spike e nella RNA polimerasi RNA-dipendente possono avere un ruolo come bersaglio per la progettazione del vaccino e il trattamento antivirale.