

Riccardo Palmieri 807445

Simone Saleri

Thomas Pellegrini 807087

## INTRODUZIONE

E' stato sviluppato un tool per produrre le variazioni delle sequenze scaricate rispetto alla reference, in formato JSON.

Le sequenze che sono state trattate appartengono a Paesi selezionati in base alla similarità dei dati demografici che li caratterizzano e a eventuali scelte nella gestione dell'emergenza che in qualche modo si discostano dalle misure adottate dalla stragrande maggioranza degli altri Paesi.

Sequenze scelte:

Si è cercato di utilizzare sequenze raccolte in tempi più vicini possibile tra loro:

Corea del Sud: [MT304474](#) **Collection Date:** 2020-02-27

Svezia: [MT093571](#) **Collection Date:** 2020-02-07

Francia: [MT320538](#) **Collection Date:**2020-03

Italia: [MT077125](#) **Collection Date:** 2020-01-31

Spagna: [MT292574](#) **Collection Date:** 2020-03-02

Germania:[MT358640.1](#) **Collection Date** 2020-02

Nel caso di sequenze prese da GISAID si è preferito allineare le sequenze collezionate nel mese di Aprile (file **alligned-sequences.json** appartenente ad **alligned.fasta** allegati) provenienti dagli stati da noi considerati.

I Paesi che sono stati selezionati in base alla similarità dei dati demografici sono:

- Italia
- Francia
- Spagna
- Germania

Nella tabella nella pagina successiva si possono notare diverse somiglianze tra i dati demografici da noi raccolti, tra cui Età media, Obesità negli adulti, posti letto ospedalieri, spese per sanità, ecc.

I Paesi, invece, che sono stati selezionati in base alle diverse misure prese per affrontare la pandemia sono:

- Corea del Sud
- Svezia

Sia la Svezia che la Corea del Sud non hanno adottato misure molto stringenti, non è mai stato imposto infatti nessun tipo di lockdown (inteso come divieto di spostamenti non inerenti a motivi di lavoro/salute/assoluta necessità), ma solo la chiusura di alcuni luoghi pubblici, divieti di assembramento, obbligo di indossare una mascherina e controllo della temperatura all'ingresso degli esercizi commerciali o pubblici.

## Dati demografici e sanitari dei paesi considerati

NB: Non sono stati presi dati aggiornati durante la pandemia in quanto sono stati effettuati dei potenziamenti a livello sanitario, di conseguenza è stato scelto di utilizzare dati raccolti negli anni trascorsi da parte di “the world factbook” di competenza della CIA ed [indexmundi.com](http://indexmundi.com)

	Italia	Francia	Germania	Spagna	Corea del Sud	Svezia
Età media	44.91	40.81	44,36	42.63	33.90	41.20
Aspettativa di vita	83.24	82.52	80,99	83.33	82.5	82.2
Obesità (prevalenza media adulti)	19.9% (2016)	21.6%	22.3%	23.8%	4.7%	20.6%
Inquinamento (concentrazione PM2.5 µg/m³)	17	12	12	10	24.0	7.4
Posti terapia intensiva ICU-CCB/100.000 abitanti	12.5	11.6	29.2	9.7	10.6	5.8
Spese per sanità (% PIL)	5-8	11.5	11.1	9	7,4	11
Densità medici (medici / 1,000 abitanti)	4	3.23	4.21	4.07	2,37	5,4

Posti letto ospedalieri (posti letto / 1000 abitanti)	2.5	6.5	8.3	3	11.5	2.4
Densità di popolazione (Abitanti per chilometro quadrato)	199.82	101	232	92	515,62	22.3
Data di inizio lockdown totale (se avvenuto)	8 marzo 2020	16 marzo 2020	15 marzo 2020	14/03/2020	non attivato	non attivato

**Casi di Covid-19 accertati al momento della selezione degli stati da considerare (24/04/2020)**

(fonte worldometers)

	Italia	Germania	Francia	Spagna	Corea del Sud	Svezia
Casi totali (accertati)	192994	152438	122577	219764	10718	17567
Morti totali (accertati)	25969	5500	22245	22524	240	2152
Test totali effettuati (se comunicati)	1147850	non comunicati	non comunicati	non comunicati	595161	non comunicati
Case Fatality Rate (morti totali/casi totali)	0,13	0,03	0.18	0.10	0.022	0.12

## Formato dell'Output

Il formato per la rappresentazione delle variazioni da noi scelto è il JSON, in quanto comprensibile, molto utilizzato e semplice da serializzare /deserializzare. E' stata creata una soluzione C# per deserializzare i file in input e confrontare le sequenze.

### Dati raccolti tramite l'utilizzo del nostro tool relativi alle sequenze con dati demografici simili da noi scelte:

E' stato scelto di avere come reference la prima sequenza Cinese (LOCUS NC\_045512 e hCoV-19/Wuhan/IPBCAMS-WH-01/2019|EPI\_ISL\_402123|2019-12-24/1-29899 per Gisaid), facendo attenzione a scegliere sequenze che provenissero da organi o apparati il più possibile simili (es. Oronasopharinx o nasopharyngeal aspirate quando citati)

Il nostro software SequencesParser prende in input le sequenze allineate con Clustal Omega in formato JSON (aprendo il file **aligned.fasta** con JALVIEW e facendo FILE->Output to Textbox-> JSON e salvandolo come **aligned-sequences.json**), il file di gene annotation in formato json ("**gene-annotation.json**") e costruisce un altro file dello stesso formato confrontando le differenze tra la sequenza di reference, nel nostro caso hCoV-19/Wuhan/IPBCAMS-WH-01/2019|EPI\_ISL\_402123|2019-12-24/1-29899 presa da GISAID e NC\_045512/1-29903 presa da NCBI, e le sequenze allineate con la stessa sequenza reference (presente nel file JSON di input).

Il formato costruito sarà del tipo:

```
"DifferenceLists":[  
  
  {  
  
    "Seq1":{" ←Reference  
  
    "Name": "nomesequenza",  
  
    "Start": "posizione iniziale",  
  
    "End": "posizione finale",  
  
    "Seq": "sequenza in formato fasta",
```

```

"Order": "ordine nel file json di input"

},

"Seq2": { ← Altra sequenza

"Seq1": { ← Reference

"Name": "nomesequenza",

"Start": "posizione iniziale",

"End": "posizione finale",

"Seq": "sequenza in formato fasta",

"Order": "ordine nel file json di input"

},

"Differences": [

{

"Position": "indice in cui è stato trovato il primo carattere",

"Newletter": "Nuovi caratteri trovati",

"Oldletter": "Caratteri presenti nella sequenza reference",

"Protein": "nome del gene in cui è compresa quella variazione"

},

{

"Position": "indice in cui è stato trovato il primo carattere",

"Newletter": "Nuovi caratteri trovati",

"Oldletter": "Caratteri presenti nella sequenza reference",

"Protein": "nome del gene in cui è compresa quella variazione"

},

{...}

]

}

},

{...},

{...}

]

}

```