

# Come la condivisione dei dati sta accelerando la ricerca sull'epidemia

I progressi nel *sequenziamento genetico* hanno permesso agli scienziati di rintracciare e monitorare la pandemia di COVID-19 più velocemente di qualsiasi epidemia precedente. (*sequenziamento genetico* vuol dire “leggere” l'ordine in cui sono disposte lungo il DNA le *lettere* del codice genetico).

Quando il nuovo coronavirus (noto come SARS-CoV-2) è stato identificato in Cina a gennaio, gli scienziati di tutto il mondo hanno pubblicato online in pochi giorni l'intera composizione genetica del virus. In confronto, durante l'epidemia della SARS nel 2003, ci sono voluti quasi tre mesi.

I progressi nella tecnologia hanno ridotto significativamente il costo del sequenziamento genico e le apparecchiature sono ora abbastanza piccole da stare nel palmo di una mano. Ciò ha semplificato la sequenza di numerosi campioni in tutto il mondo.

*“Dalle sequenze si può vedere come si diffonde il virus, la velocità con cui si sta diffondendo e stimare il numero di persone infette. Man mano che otteniamo sempre più sequenze, i numeri sono sempre più precisi”*, ha affermato la professoressa Anne-Mieke Vandamme di KU Leuven, Belgio.

Il sequenziamento di nuova generazione, o NGS, può generare enormi quantità di dati e la sfida diventa trovare modi per analizzarli correttamente.

Nel 2015, il Prof. Vandamme ha guidato un progetto chiamato VIROGENESIS per sviluppare nuovi strumenti per aiutare ad analizzare e interpretare i dati che derivano dal sequenziamento, in particolare per i laboratori che non erano

abituati a trattare sofisticate analisi genetiche. *“Quando stavamo realizzando il progetto, c'erano solo laboratori di ricerca che avevano NGS. Ora tutti hanno NGS”*. ha detto.

Uno degli strumenti sviluppati, chiamato Genome Detective, può prendere i dati grezzi dalla macchina di sequenziamento, filtrare i risultati dai non-virus, mettere insieme il genoma e usarli per identificare il virus. Non si basa su ipotesi o ipotesi precedenti, quindi può anche identificare virus che non sono stati mai conosciuti prima. Il sistema è stato usato per confermare il primo caso di COVID-19 in Belgio, identificandolo come un coronavirus correlato alla SARS.

Il potere del sequenziamento genico deriva dal confronto dei risultati tra diversi casi. *“E' stato fantastico vedere il livello di collaborazione a livello internazionale. C'è molta più condivisione online di dati e sequenze rispetto al passato perché abbiamo molti più strumenti di condivisione online disponibili”*. afferma il Prof. Vandamme.

Uno di questi strumenti online è NextStrain, una risorsa che utilizza i dati del genoma per monitorare in tempo reale l'evoluzione di organismi patogeni come i virus. Ha monitorato diversi focolai tra cui Zika, Ebola e Dengue ed è stato persino utilizzato per informare l'OMS sull'influenza stagionale.

I lavori di ricerca richiedono in genere mesi per essere pubblicati. La necessità di condividere rapidamente le informazioni ha incoraggiato una maggiore condivisione di bozze di documenti che non sono ancora stati sottoposti ad una *peer review* (revisione paritaria).

*“La spinta verso la scienza aperta, gli open data e articoli in versione pre stampa hanno davvero cambiato il modo in cui viviamo il discorso scientifico per questa epidemia rispetto a quelle precedenti”*, ha affermato il professor Richard Neher, dell'Università di Basilea, in Svizzera, che guida il progetto

NextStrain.

NextStrain ha già oltre 700 genomi del nuovo coronavirus, che può utilizzare per tracciare l'epidemia rilevando nuove mutazioni nel virus. Le mutazioni non influenzano necessariamente il comportamento del virus, ma possono fungere da firma genetica per collegare i casi correlati. Come rintracciare i tuoi antenati attraverso un test del DNA, un virus sequenziato a Madrid, ad esempio, potrebbe avere mutazioni che suggeriscono che sia nato da un focolaio in Italia.

NextStrain pubblica un rapporto settimanale sulla situazione che analizza queste tendenze. Il team è stato in grado di stimare che l'epidemia in Iran potrebbe essere stata introdotta da una sola persona, mentre almeno quattro diverse introduzioni erano responsabili dell'epidemia nel Regno Unito, a partire dal 13 marzo. *“Il sequenziamento genetico diventerà ancora più importante perché quando inizieremo a reprimere la pandemia, ci dirà quante catene di trasmissione sono ancora in circolazione e se il virus viene trasportato da un luogo a un altro. Mano che il virus continuerà a diffondersi, accumulerà una maggiore diversità genetica e ci fornirà maggiori informazioni su come il virus viene trasmesso.”*

Nonostante il progetto genetico del nuovo coronavirus sia prontamente disponibile, non ci dice ancora molto su come differisce dagli altri coronavirus. Gran parte di ciò che sappiamo è emerso dai dati su come si è diffuso nella popolazione. Ora è chiaro quanto sia diverso dai precedenti focolai di coronavirus, come SARS e MERS.

*“Erano certamente molto meno facili da trasmettere e presentavano anche una caratteristica molto diversa in quanto solo poche persone erano asintomatiche. Una delle molte sfide che stiamo affrontando adesso è che le persone che hanno solo sintomi molto lievi sono state sostanziali nella trasmissione di questo virus. È molto più difficile da controllare perché*

*devi convincere qualcuno che è sostanzialmente sano a prendere le distanze dagli altri” ha affermato il Prof. Neher.*

Tuttavia, non è chiaro perché sia *così*. I tratti del virus, come la sua infettività e gravità, sono guidati dalle sue proteine *che* sono responsabili dell’invasione delle nostre cellule e della replicazione del genoma virale.

*“Oggi il sequenziamento di un genoma è piuttosto rapido, ma per le proteine è diverso”,* ha affermato la dott.ssa Charlotte Uetrecht, dell’Istituto Heinrich Pette, Leibniz Institute for Experimental Virology, Germania. Studia le proteine *del* coronavirus attraverso un progetto chiamato SPOCKS MS .

*“Il mio laboratorio sta producendo le proteine (del nuovo coronavirus) in questo momento. Quindi vogliamo vedere se si comportano allo stesso modo (come altri coronavirus). Di solito abbiamo bisogno di produrre le proteine e purificarle in una certa misura in modo da poterle guardare. Quindi è molto più laborioso del sequenziamento.”*

Anche piccole modifiche alle proteine *virali* possono influenzare in modo significativo il modo in cui interagiscono tra loro. Il dottor Uetrecht studia queste associazioni, che sono cruciali per la replicazione del virus. *“Sappiamo un po’ come appare, ma non capiamo davvero quale delle proteine debba associare per produrre un nuovo genoma”,* ha detto.

Sebbene la comprensione di questi processi possa fornire nuovi obiettivi per i farmaci antivirali, la dott.ssa Uetrecht afferma che storicamente c’è stato scarso interesse nello studio dei coronavirus poiché finora hanno avuto un impatto relativamente scarso.

Il numero dei casi era basso per SARS e MERS e l’interesse è diminuito dopo gli scoppi. *“I coronavirus che causano il raffreddore comune non erano (considerati) pericolosi.”*

*“Non ci sono state molte ricerche sui coronavirus, fino alla SARS. Conosco alcune persone che hanno lavorato ai coronavirus dagli anni '90 e non sono state molto apprezzate – hanno avuto difficoltà a ottenere finanziamenti. Era considerato un virus noioso e irrilevante”.*

*(Traduzione dell'articolo di Ian Le Guillou su Horizon Research and Innovation della Commissione Europea)*