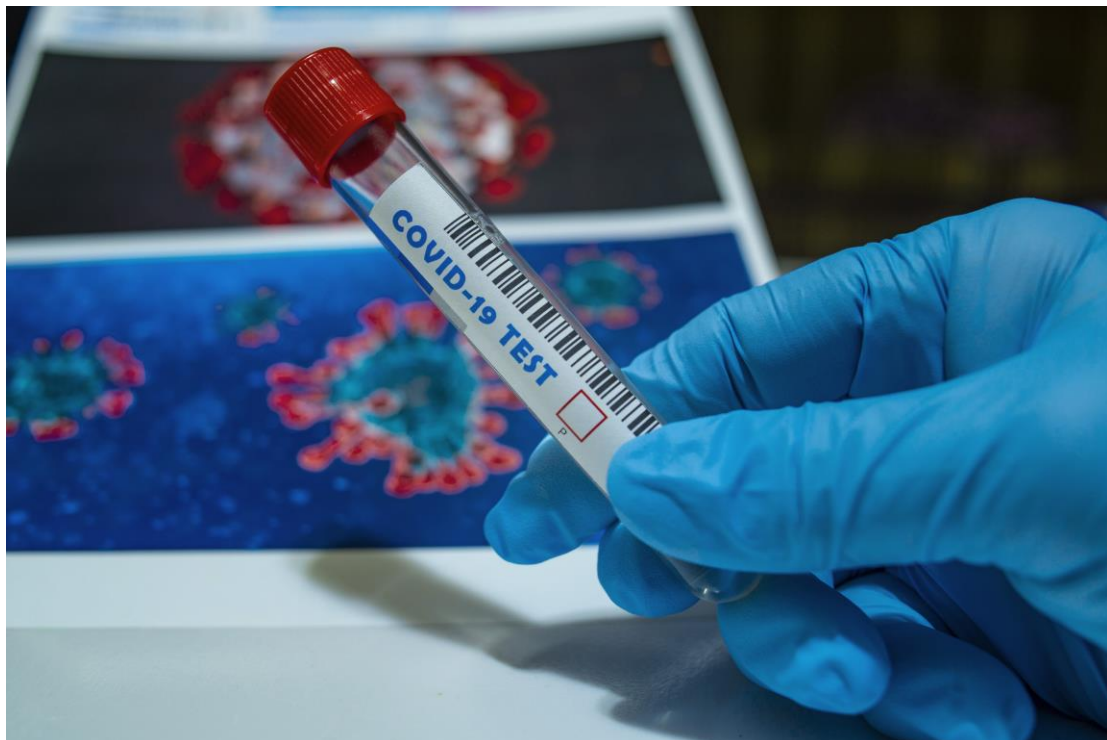


València, 2 d'abril de 2025

Troben errors en les seqüències genètiques del coronavirus incloses en la major base de dades mundial

- Un estudi liderat pel CSIC descobreix 'artefactes' en les seqüències amb reparació de les deleccions del virus que provoca la COVID-19, la qual cosa afecta la infecció i a la resposta a vacunes
- Moltes de les seqüències amb reparació de mutacions en la proteïna *spike* del virus, la clau per a infectar les cèl·lules humanes, es van deure a errors en el processament de dades



Després de realitzar un minut en la base de dades del SARS-CoV-2 més usada en la pandèmia van descobrir errors./ Pixabay

Un equip multidisciplinari liderat per l'Institut de Biologia Integrativa de Sistemes (I2SysBio), centre mixt del Consell Superior d'Investigacions Científiques (CSIC) i la Universitat de València (UV), acaba de publicar un estudi que descobreix una nova perspectiva sobre la capacitat del virus SARS-CoV-2 per a mutar i infectar als humans. Mitjançant una revisió de la base de dades genètics del virus més utilitzada durant la pandèmia, l'equip d'investigació va trobar 'falsos positius' en la seua capacitat de

reparació de les deleccions, un procés que restaura seccions del genoma viral que afecta la capacitat del virus per a replicar-se o evadir el sistema immunològic de l'hoste. En el treball, publicat a la revista *Virus Evolution*, participen investigadors de l'Institut de Biomedicina de València (IBV) del CSIC i de l'Institut d'Investigació Sanitària La Fe (IIS-La Fe).

El treball liderat pel grup de Patogenòmica de l'I2SysBio en col·laboració amb el grup de Biologia Viral del mateix institut ofereix una perspectiva innovadora sobre uns certs canvis genètics rars en la proteïna *spike* del SARS-CoV-2, la 'clau' que utilitza el coronavirus per a infectar les nostres cèl·lules. La investigació es va centrar en els anomenats esdeveniments de reparació de deleccions en aquesta proteïna, en els quals el virus aparenta corregir el seu genoma.

Després de realitzar un minat de dades massiu en la base de dades de genomes del virus SARS-CoV-2 més usada en la pandèmia, anomenada GISAID, van descobrir que diversos de les troballes inicials es devien probablement a errors introduïts pel processament de dades en grans bases de dades genètiques. Els mètodes informàtics emprats per a analitzar milions de seqüències virals poden induir equivocacions, generant la impressió que el virus repara les seues deleccions amb major regularitat. En comparar aquestes dades ja processades amb la informació obtinguda directament de la seqüenciació de genomes (lectures de seqüenciació), l'equip ha aconseguit obtindre una visió més realista dels canvis genètics que pateix el virus.

Menys del 60% d'esdeveniments de reparació confirmats

“Utilitzant el repositori de seqüències genètiques GISAID estimem una freqüència molt alta d'aquests esdeveniments de reparació de deleccions que s'espera que siguin rars”, explica **Mireia Coscollá Devís**, investigadora del CSIC que lidera l'estudi. “Ens vam adonar que les seqüències de la base de dades de GISAID estan processades per cada laboratori de manera diferent i contenia molts falsos positius per a aquesta mena de marcadors. Així, encara que en determinats casos vam poder confirmar que es tractava d'un fenomen real, en la majoria eren conseqüència del processament de les seqüències”, revela.

Així, “vam veure que menys del 60 per cent dels esdeveniments de reparació de deleccions es podia confirmar. Encara que no hem pogut quantificar-ho exactament per a tots, podem comprar les proporcions del marcador en diverses bases de dades, i veiem que la diferència és de 5 a 51 vegades menys freqüent del que apareixia en les bases de dades processades”, calcula la investigadora del CSIC.

Encara que aquests esdeveniments de reparació són poc comuns, l'estudi evidencia que, quan ocorren, poden afectar de manera subtil al comportament del virus. “Per exemple, unes certes reparacions poden modificar la forma en què el virus ingressa a les cèl·lules o influir en la resposta als anticossos generats per la vacunació”, assegura Coscollá, alguna cosa que l'equip d'investigació va demostrar mitjançant experiments *in vitro*.

Intercanvi de dades genòmiques de patògens

Així, “la nostra investigació posa en relleu la importància d'examinar detingudament les dades genètiques per a evitar conclusions errònies”, remarca la investigadora del CSIC. L'Organització Mundial de la Salut (OMS) recomana una política d'intercanvi de dades genòmiques de patògens per a protegir la salut pública. No obstant això, a Espanya no existeix una recopilació central de dades de seqüències de patògens humans, animals i ambientals, com tampoc existeix una política per a l'intercanvi de dades anonimitzades entre institucions de salut i científiques. Això dificulta el seguiment i la resposta a les malalties infeccioses incloent el seguiment de la resistència als antimicrobians, destaquen els investigadors.

El treball ha estat finançat pel Ministeri de Ciència, Innovació i Universitats i per la Unió Europea amb fons NextGenerationEU/PRTR a través de la PTI+ Salut Global del CSIC. Ademés, està secundat per la Generalitat Valenciana i el Fons Social Europeu a través de l'ajuda CIACIF/2022/333. El treball computacional es va realitzar en Garnatxa, el clúster de computació d'alt rendiment (HPC) de l'Institut de Biologia Integrativa de Sistemes.

Referència:

Miguel Álvarez-Herrera, Paula Ruiz-Rodriguez, Beatriz Navarro-Domínguez, Joao Zulaica, Brayan Grau, María Alma Bracho, Manuel Guerreiro, Cristóbal Aguilar-Gallardo, Fernando González-Candelas, Iñaki Comas, Ron Geller, Mireia Coscollá, ***Genome data artifacts and functional studies of deletion repair in the BA.1 SARS-CoV-2 spike protein***, *Virus Evolution*, 2025; <https://doi.org/10.1093/ve/veaf015>