

Variabilitat i seqüenciació massiva de virus. El SARS-CoV-2 com a exemple

Cristina Andrés,¹ David Taberero,^{2,3} Tomás Pumarola,^{1,5} Andrés Antón^{1,5} i Josep Quer^{2,4,6}

¹ Departament de Microbiologia, Vall d'Hebron Institut de Recerca (VHIR), Vall d'Hebron Barcelona Hospital Campus

² Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Instituto de Salud Carlos III

³ Unitat de Patologia Hepàtica, Departaments de Bioquímica i Microbiologia, Vall d'Hebron Barcelona Hospital Campus

⁴ Malalties Hepàtiques - Hepatitis Virals, Vall d'Hebron Institut de Recerca (VHIR), Vall d'Hebron Barcelona Hospital Campus

⁵ Departament de Microbiologia, Universitat Autònoma de Barcelona (UAB)

⁶ Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona (UAB)

Correspondència: Andrés Antón. Departament de Microbiologia, Vall d'Hebron Institut de Recerca, Vall d'Hebron Barcelona Hospital Campus. Passeig de la Vall d'Hebron, 119-129. 08035 Barcelona. Adreça electrònica: aanton@vhebron.net.

Josep Quer. Malalties Hepàtiques. Vall d'Hebron Institut de Recerca, Vall d'Hebron Barcelona Hospital Campus. Passeig de la Vall d'Hebron, 119-129. 08035 Barcelona. Tel.: +34 934 894 034. Adreça electrònica: josep.quer@vhir.org.

DOI: 10.2436/20.1501.02.217

ISSN (ed. impresa): 0212-3037

ISSN (ed. digital): 2013-9802

<http://revistes.iec.cat/index.php/TSCB>

Rebut: 01/02/2022

Acceptat: 14/03/2022

Resum

Una de les coses que hem après en els darrers dos anys, concretament amb la pandèmia del SARS-CoV-2, és que no podem emmurallar els virus per a evitar-ne la propagació, i la manera més eficaç de lluitar-hi en contra és amb una detecció i una caracterització genètica ràpides. Les pandèmies causades per virus han ocorregut, n'estem patint una i tard o d'hora en tornarem a patir. En els darrers anys, el risc de noves pandèmies s'ha incrementat degut a factors que les afavoreixen, com ara la destrucció d'espais naturals, la facilitat per a fer viatges de llarga distància en qüestió d'hores, el canvi climàtic i l'augment de concentracions humanes i activitats socials en un món cada vegada més globalitzat. En aquest sentit, la seqüenciació massiva o seqüenciació de nova generació (NGS, *next generation sequencing*) ha demostrat ser una eina molt poderosa que ha ajudat en el diagnòstic, en el desenvolupament de vacunes, en la identificació i el seguiment de variants i a conèixer millor la biologia del virus. La pandèmia del SARS-CoV-2 ha permès demostrar la utilitat de l'NGS i la necessitat de reforçar la xarxa de laboratoris de seqüenciació. Com més preparats estiguem, més ràpida i eficaç serà la resposta contra una possible amenaça.

Paraules clau: variabilitat, virus, seqüenciació massiva, NGS, diagnòstic.

Introducció

Les pandèmies causades per virus han ocorregut, n'estem patint ara mateix i tard o d'hora en tornarem a patir. La raó principal la trobem en la gran capacitat que tenen els virus de generar variabilitat i seleccionar mutacions que, per exemple, els permeten saltar d'una espècie a una altra i causar una zoonosi (transmissió d'un animal a un humà). Els virus amb més capacitat de variar són els que tenen el genoma en forma d'RNA, com els virus de les hepatitis A, C i E; el virus del Zika (ZIKV); el virus del chikungunya; el virus de la immunodeficiència humana (HIV); el virus de Crimea-Congo; el virus de l'Ebola; el virus de Marburg (EVM); el virus de la influència; els virus entè-

rics; el virus de la febre groga; el virus del xarmpió; el virus de les galteres o els enterovirus, i els coronavirus (CoV) estacionals (NL63, OC43, HKU1 o 229E), així com el SARS-CoV, el MERS-CoV i el SARS-CoV-2. Dins el grup de virus altament variables s'inclouen els que durant el cicle de replicació passen per una fase d'RNA, com el virus de l'hepatitis B (Domingo, 2020).

En els darrers anys, el risc de noves pandèmies s'ha incrementat degut a factors que les afavoreixen, com la destrucció d'espais naturals, que ha facilitat el contacte directe amb animals salvatges, i que moltes vegades són reservoris de virus fins ara desconeguts; la facilitat per a fer viatges de llarga distància en

Variability and massive sequencing of viruses. SARS-CoV-2 as an example

Abstract

One of the things that we have learned in the last two years, specifically from the SARS-CoV-2 pandemic, is that we cannot wall in viruses to prevent their propagation, and that the most effective way to combat them is to ensure their rapid detection and genetic characterisation. Pandemics caused by viruses have broken out in the past, we are currently experiencing such a pandemic, and sooner or later we will surely be struck by another one. The risk of new pandemics has grown in recent years, due to factors favouring their appearance such as the destruction of natural areas, the ease of long-distance travel in a matter of hours, climate change, and the increase of human concentrations and social activities in a constantly more globalised world. In this respect, massive or next-generation sequencing (NGS) has proven to be a very powerful tool that has been of help in diagnosis and in the development of vaccines, the identification and monitoring of variants, and the study of the biology of the virus. The SARS-CoV-2 pandemic has allowed to demonstrate the utility of NGS as well as the need to strengthen the network of sequencing laboratories. The better we are prepared, the quicker and more effective will be our response to a possible threat.

Keywords: variability, virus, massive sequencing, NGS, diagnosis.

qüestió d'hores; el canvi climàtic, que ha facilitat la distribució de vectors virals on abans no estaven presents, i, a més a més, l'augment de concentracions humanes i activitats socials en un món cada vegada més globalitzat (Wearn *et al.*, 2012; Gibb *et al.*, 2020; Tollefson, 2020).

Una de les coses que hem après en els darrers dos anys, concretament amb la pandèmia del SARS-CoV-2, és que no podem emmurallar els virus per a evitar-ne la propagació, i que la manera més eficaç de lluitar-hi en contra és amb una detecció i una caracterització genètica ràpides. En aquest sentit, la seqüenciació massiva o seqüenciació de nova generació (NGS, *next generation sequencing*) ha demostrat ser

una eina molt poderosa, ja que ha permès desvelar en temps rècord la seqüència del genoma viral. A partir de les primeres seqüències, s'han pogut desenvolupar assajos moleculars basats en amplificació d'àcids nucleics (per exemple, per al disseny dels encebadors i sondes) per a la detecció del virus; caracteritzar la regió més immunògena (l'espícula o *spike*) per a desenvolupar vacunes basades en mRNA o de DNA (Amanat i Krammer, 2020; Ruiz-Rodriguez *et al.*, 2021; Wouters *et al.*, 2021); identificar i fer el seguiment de les noves variants identificant les mutacions associades, i estudiar-ne l'evolució i la variabilitat; detectar i monitorar brots nosocomials i a la comunitat, i conèixer millor la biologia del virus (identificar-ne les debilitats) per al disseny de compostos amb activitat antiviral.

Un tret comú entre els virus amb aquesta gran capacitat per a generar variabilitat és que no podem parlar d'un virus com una única seqüència, sinó com una població de genomes. I és en aquesta població de genomes o quasiespècies on podem descriure un reservori de genomes minoritaris que podrien ser seleccionats davant una pressió selectiva, perquè són aquestes variants minoritàries portadores d'una característica que li permeten un avantatge biològic. Per a descriure en detall aquesta població de genomes, és a dir, la seva variabilitat, l'eina més adient és, sens dubte, la tecnologia d'NGS.

1. Conceptes bàsics per a entendre la variabilitat dels virus RNA

Els virus amb el genoma d'RNA i els que durant el seu cicle de replicació passen per una fase d'RNA basen el seu èxit evolutiu en la gran capacitat que tenen de generar variabilitat. Això és degut al fet que tots fan servir RNA-polimerases RNA-dependents que, en la gran majoria de casos, no poden corregir aquells errors que de manera natural s'acumularien durant la còpia de l'RNA genòmic. Una excepció en són els CoV, que, com veurem més endavant, presenten una proteïna amb propietats correctores d'error.

Per a poder entendre la variabilitat viral, ens cal definir dos conceptes (taula 1): la taxa de mutació (MR, de l'anglès *mutation rate*) i la taxa d'evolució (ER, de l'anglès *evolution rate*), també coneguda com a *taxa de fixació de mutacions (rate of accumulation [fixation] of mutations)*, a més a més del concepte de *quasiespècies* (taula 2).

La majoria de virus RNA tenen una gran capacitat de variar degut a la manca de proteïnes de lectura i correcció d'errors (*proofread-*

↓ Taula 1. Definició de conceptes clau per a entendre la variabilitat viral. Elaboració pròpia.

Concepte	Definició	Exemples
1a. Taxa de mutació (mutation rate, MR)	Quantifica el nombre d'errors (substitucions) que ha introduït l'RNA-polimerasa RNA-dependent per cada nucleòtid copiat i en cada cicle de replicació (substitucions / nucleòtid / cicle de replicació). <i>Un virus com el de l'hepatitis C (VHC), que consta d'un genoma d'RNA de 10.000 nucleòtids amb una MR de 10^{-3}-10^{-4}, vol dir que cada cop que un genoma és copiat, el nou genoma que es genera tindrà entre 1 i 10 mutacions aleatòries al llarg del nou genoma, respectivament.</i> Aquesta taxa és un valor que depèn del procés de còpia de l'RNA viral, de la presència o absència de mecanismes de lectura i correcció d'errors, i del fet que hi hagi manca o desequilibri en els quatre nucleòtids ATGC, entre altres necessitats bioquímiques. Per tant, la taxa de mutació (MR) és la conseqüència d'un procés bioquímic.	<i>Virus RNA: 10^{-3}-10^{-5}</i> <i>Retrovirus: 10^{-3}-10^{-6}</i> <i>Virus DNA: 10^{-3}-10^{-8}</i> <i>DNA cel·lular: 10^{-9}-10^{-11}</i> <i>Un SARS-CoV-2 amb un genoma de 29.903 nucleòtids i una MR estimada de 10^{-6}-10^{-7} vol dir que de cada 300 genomes (virions) que analitzem, un tindrà una mutació situada a l'atzar en el genoma.</i>
1b. Taxa d'evolució (evolution rate, ER) o taxa de fixació de mutacions (rate of accumulation [fixation] of mutations)	Mesura del nombre de substitucions per nucleòtid i per any (substitucions/nt/any). L'ER mesura com el virus està evolucionant en la natura, i està directament afectada per l'eficàcia de transmissió, o facilitat que té el virus per a infectar i generar noves partícules virals. És a dir, un virus amb una taxa de mutació molt alta, però que no es pugui transmetre (per exemple, perquè la persona infectada estigui confinada), tindrà una taxa d'evolució nul·la; en canvi, un virus amb una taxa de mutació baixa, però que es transmeti molt fàcilment i amb capacitat per a generar un gran nombre de partícules virals (càrrega viral elevada), tindrà una taxa d'evolució alta.	<i>Virus respiratoris: 10^{-3}-10^{-4}</i> <i>Virus de l'hepatitis: 10^{-3}-$7,9$-10^{-5}</i> <i>HCV: $1,5 \cdot 10^{-3}$ subst/nt/any</i> <i>El SARS-CoV-2 té una ER estimada de $8,4 \cdot 10^{-4}$ molt semblant als altres virus respiratoris i altres virus RNA.</i>

↓ Taula 2. Definició de quasiespècies. Elaboració pròpia.

Concepte	Definició	Rellevància clínica
Quasiespècies	Els virus RNA amb les taxes de mutació més elevades són els organismes amb més capacitat de generar diversitat. La conseqüència d'aquesta alta capacitat de variació és l'aparició continuada de variants. Per tant, si seqüenciem el genoma dels virus (població viral) que detectem en un òrgan o teixit d'un pacient infectat, no trobarem una sola seqüència, sinó una barreja complexa de genomes que s'assemblen molt, però presenten diferències. Tota aquesta població viral l'anomenem <i>quasiespècies</i> . Una quasiespècie està composta per una barreja de genomes diferents, però que s'assemblen molt (tenen unes poques diferències a escala de seqüència l'una respecte de l'altra). La composició d'una quasiespècie està sotmesa a canvis continuats (és una estructura molt dinàmica) mentre el virus està replicant i depèn dels principis de l'evolució darwiniana: — Cada partícula viral pot alhora replicar i, per tant, generar noves variants. — Els virus que componen aquesta quasiespècie estan sotmesos a selecció competitiva. — Els membres de la quasiespècie poden interaccionar i cooperar.	Quins avantatges biològics té per al virus aquesta gran variabilitat? Li facilita: — Adquirir canvis antigènics i així escapar de la resposta adquirida per vacunació o infecció natural. — Seleccionar variants amb mutacions associades a resistència a antivirals, que a més poden ser transmeses. — Adquirir un mecanisme d'evasió de la resposta immunitària induint fenòmens de tolerància perifèrica. — Afavorir una virulència o transmissibilitat superiors.

ing activity en anglès), però, en canvi, els virus de la família Coronaviridae, entre els quals el SARS-CoV-2, tenen un mecanisme de correcció d'errors que fa que tinguin taxes de mutació mil vegades inferiors que altres virus RNA. Aquest mecanisme de correcció d'errors està basat en l'expressió de la proteïna no estructural nsp14 (*non-structural protein 14* en anglès),

que té la funció de lectura de l'RNA que sintetitza l'RNA-polimerasa RNA-dependent (*RNA-dependent RNA polymerase* o nsp12) i, comparant-la amb la seqüència motlle, detecta si hi ha un error i el corregeix amb la seva activitat 3' → 5' exoribonucleasa (ExoN). Això fa que la còpia d'un genoma del SARS-CoV-2 tingui alta fidelitat i en redueix la taxa de mu-

tacions, o el que és el mateix, el nombre de substitucions en el nou genoma resultant (Eckerle *et al.*, 2010; Smith i Denison, 2013; Gribble *et al.*, 2020; Rambaut *et al.*, 2020). Ara bé, des de la seva circulació inicial entre la població humana el desembre de 2019, el virus no ha deixat d'evolucionar i en van apareixent variants noves dia rere dia, algunes amb impacte a escala mundial, com són les variants de preocupació (VOC, de l'anglès *variants of concern*). Es defineixen com a VOC perquè les mutacions que incorporen poden suposar canvis que afectin significativament la salut pública de manera global, com, per exemple, un increment de la transmissibilitat, la virulència i la gravetat de la malaltia, o una reducció de la capacitat de detecció de la variant, l'efectivitat de vacunes o els tractaments antivirals. El febrer de 2022 es consideren VOC les variants alfa (B.1.1.7), beta (B.1.351), gamma (P.1), delta (B.1.617.2) i ara òmicron (B.1.1.529).

Com s'explica que el SARS-CoV-2 tingui una taxa de mutació baixa, mil vegades inferior a altres virus RNA, i en canvi s'observi una aparició continuada de variants noves arreu del món?

És obvi que el SARS-CoV-2 està evolucionant, i alguns dels seus canvis permeten que s'adapti cada cop millor a la població humana, amb una aparició continuada de variants. El febrer de 2022, s'havien comptabilitzat 22 clades (19A,B, 20A-J, 21A,C,D,F-J,L-M) segons la classificació establerta per Nextstrain (Nextstrain, 2021), i 1.728 llinatges (O'Toole *et al.*, 2022) seguint la classificació PANGOLIN (Phylogenetic Assignment of Named Global Outbreak Lineages) (Rambaut *et al.*, 2020). Bàsicament, estem parlant d'una nomenclatura dinàmica basada en l'anàlisi de similitud de seqüències per inferència filogenètica (Rambaut *et al.*, 2020).

Però, tal com hem comentat abans, la baixa taxa de mutació del SARS-CoV-2 es veu compensada per una elevada grandària de la població viral que cada persona infectada pot presentar, especialment en el pic de màxima càrrega viral, de fins a un total de 10^9 a 10^{11} virions, i que a més es veu afavorida per una gran capacitat de transmissió entre persones. Això permet que milions de persones s'estiguin infectant on el virus està replicant i, al mateix temps, canviant de manera continuada. Tot i el gran nombre de morts, el SARS-CoV-2 és causa d'una d'infecció de baixa gravetat de

mostrada per un nombre elevat d'individus infectats de manera asimptomàtica o lleu, fet que ajuda a facilitar-ne la transmissió, ja que com més silencios és (asimptomàtic), més fàcil és que una persona infectada pugui transmetre el virus a una persona no infectada, en no adoptar mesures per a evitar aquesta infecció, com és l'ús de mascaretes o reduir el contacte social, entre altres mesures no farmacològiques.

A l'inici de les pandèmies, amb l'absència d'immunitat preexistent a la població humana, el virus es va estendre sense límits i sense necessitat de canviar, fet que es va traduir en una similaritat de seqüència molt alta (99,9%) entre els virus detectats i caracteritzats per seqüenciació arreu del món (Rambaut *et al.*, 2020). A mesura que el virus s'ha anat distribuïnt mundialment i la pandèmia ha anat progressant, amb una incidència elevada degut a la seva gran capacitat per a la transmissió de persona a persona, el virus ha anat canviant, de manera que hem pogut observar cocirculació de diferents variants, tot i que només unes poques s'han seleccionat, segurament pel fet de presentar un avantatge biològic. Aquest avantatge biològic pot entendre's com una capacitat més gran de transmissió o d'infecció en una població cada vegada més immunitzada, sobretot si parlem dels països rics, que, a hores d'ara, presenten taxes elevades de cobertura vacunal. En la majoria de pandèmies prèvies, en les quals no es tenien recursos sanitaris, de diagnòstic ni de control (Rodríguez-Frías *et al.*, 2021a), s'observa que, a mesura que avança la pandèmia, es van seleccionant noves variants més transmissibles però menys agressives i el virus pot esdevenir un virus endèmic i estacional en el cas de virus respiratoris i entèrics. Una variant que causi una malaltia greu es detecta més ràpidament, fet que permet adoptar mesures d'aïllament per al pacient i evitar, així, que infecti altres persones. Contràriament, un virus silencios, és a dir, una variant que causi una infecció asimptomàtica, no és detectat perquè els malalts no necessiten atenció sanitària i, per tant, els infectats poden transmetre el virus sense ser-ne conscients, cosa que en facilita l'expansió i dona més probabilitats al virus de persistir en la població i seguir canviant.

Els virus tenen una capacitat elevada de generar variabilitat, però també és cert que tenen constreyniments, és a dir, limitacions en la variació sense perdre capacitat de transmissió o d'infecció. La seqüenciació massiva o NGS no sols ens permet detectar mutacions minori-

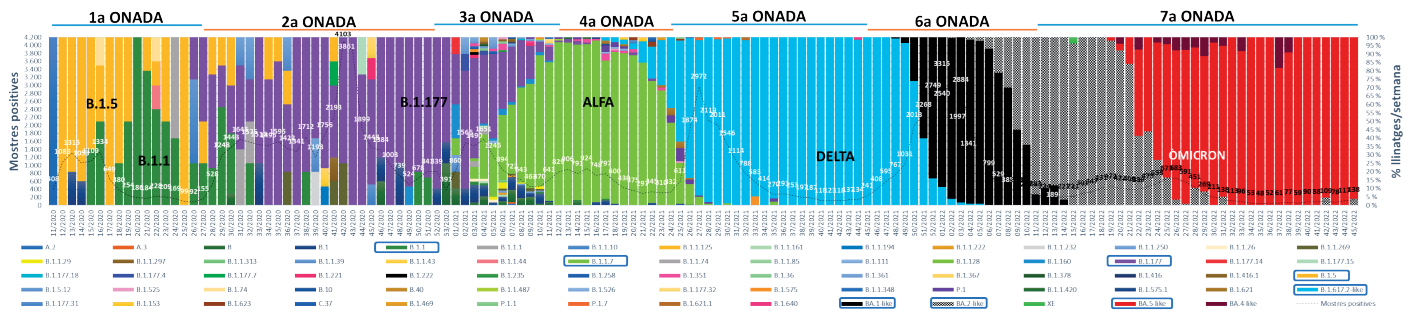
tàries, sinó també les regions genòmiques més conservades. En el cas del SARS-CoV-2, si som capaços d'identificar quins són aquests límits que té el virus per a variar, quines són les proteïnes més sensibles a aquests canvis aleatoris (regions del genoma més conservades) i quines són les mutacions que li donen un avantatge selectiu, podrem desenvolupar eines de vacunació i tractament eficaces que puguin veure's menys compromeses en el cas de virus capaços de generar variants d'escapament, i que ens permetran a mitjà termini arribar al control de la pandèmia. Aquesta informació de variabilitat i evolució de seqüència completa del genoma viral i de diversitat només la pot aportar l'NGS.

De totes maneres, hem d'ésser conscients que el virus s'està transmetent a escala mundial, i que les variants estan emergint arreu, però les capacitats de seqüenciació no són les mateixes a tots els continents. Cal reforçar aquestes capacitats de seqüenciació a tots els països per a detectar-ne variants noves, que, tot i tenir un origen llunyà, tard o d'hora ens poden arribar i ser causa de problemes al nostre territori. Així ha passat amb la variant delta (amb origen a l'Índia) i òmicron (amb origen a Sud-àfrica i Botswana).

2. Origen i evolució de les variants del SARS-CoV-2 a Catalunya

2.1. Variants dominants de la primera a la sisena onada pandèmica

L'NGS del genoma complet del virus ha estat clau per a fer la seqüenciació completa del genoma viral i així fer un seguiment setmana a setmana de les variants del SARS-CoV-2 en cocirculació a partir de mostres aleatoris entre els virus detectats, per a detectar possibles variants que eren importades d'altres països, o bé per a fer estudis selectius, especialment, dels casos sospitosos de reinfecció o fallades vacunals (infeccions irruptives). En aquest sentit, a Catalunya la variant predominant (figura 1) durant la primera onada de la pandèmia va ser la B.1.5 (51%), que va dominar des de març fins a maig de 2020, quan la variant B.1.1 (33%) va augmentar significativament, coincidint amb l'inici de la desescalada. Aquest virus, junt amb altres virus pertanyents també al clade G de GISAID, es caracteritzaven per ser portadors de la mutació D614G (canvi de l'aa àcid aspàrtic/D per Glicina/G a la posició 614) a l'espícula, la qual n'afavoreix la transmissibilitat (Korber *et al.*, 2020) i es va fixar en el genoma viral, que encara és present en la majoria



↑ Figura 1. Distribució setmana a setmana dels llinatges des de març de 2020. S'han encerclat els llinatges (variants) predominants en cada onada pandèmica. Els valors numèrics indiquen el nombre de casos positius (incidència) atesos a l'atenció primària de la ciutat de Barcelona més els ingressats a l'hospital (planta i UCI. Elaboració pròpia.

de virus en circulació actualment. No obstant això, el final del confinament i la flexibilització de la mobilitat i l'organització de les activitats d'estiu van coincidir amb la detecció de la variant B.1.177, que primer es va detectar a Casp (Saragossa), Alcanyís (Terol) i en una localitat holandesa el juny de 2020, i que va ser escampada pels temporers d'Osca i Lleida (López *et al.*, 2021). Més tard, els anglesos la van anomenar *variant de les vacances*, ja que va esdevenir predominant durant tota la segona onada fins a finals de 2020 a mig Europa (Alm *et al.*, 2020; Hodcroft *et al.*, 2021). Aquesta variant tenia com a mutació principal la substitució de l'aminoàcid alanina per valina a la posició 222 (A222V) de la proteïna de l'espícula.

El 14 de desembre de 2020, el Regne Unit va declarar l'augment d'incidència d'una nova variant a l'est i al sud-est d'Anglaterra i a l'àrea metropolitana de Londres. Aquesta variant es va caracteritzar per tenir vint-i-tres mutacions diferents, vuit de les quals en el gen de l'espícula. De les vint-i-tres mutacions, catorze donen lloc a canvi d'aminoàcid (mutacions no sinònimes), tres delecions (del) i sis mutacions sinònimes (no canvia l'aminoàcid). Les mutacions més importants a l'espícula, i que defineixen aquesta variant, eren delecio H69-V70 (del69/70), delY144, N501Y, A570D, D614G, P681H, T716I, S982A i D1118H. Aquesta variant, pertanyent al llinatge PANGOLIN B.1.1.7, a principis de 2021 va ser considerada per l'Organització Mundial de la Salut (OMS) variant de preocupació (VOC-202012/01), anomenada *alfa*, i a Catalunya i a la resta de països va substituir progressivament els virus pertanyents al llinatge B.1.177, que s'havia detectat a Europa amb la mutació A222V (Hodcroft *et al.*, 2021), de manera que es va convertir en dominant durant el final de la tercera onada (55%). A principis de la quarta onada

pandèmica la seva prevalença ja era del 86% (figura 1).

La selecció de noves variants amb més capacitat de replicació i transmissió (més *fitness* biològic) pot haver sorgit de manera independent en diferents parts del món i per diferents mecanismes, encara que a vegades hi pot haver semblança entre elles. Una de les variants que ha tingut més impacte i que ha desplaçat completament la variant alfa és la variant delta, pertanyent al llinatge B.1.617.2, que es va detectar per primer cop a l'Índia, i que va ser considerada VOC el 31 de maig de 2021. Aquesta variant estava definida per les mutacions T19R, E156G, delF157/R158, L452R, T478K, D614G, P681R i D950N a l'espícula, a més de múltiples canvis distribuïts en el seu genoma. Tal com va passar abans, aquesta variant va entrar a la nostra població substituint completament les altres variants fins a arribar a estar present en el 100% de les mostres seqüenciades (Andrés *et al.*, 2022) i dominant completament la cinquena onada, però va ser desplaçada per la variant òmicron a mitjan desembre de 2021, a partir del començament de la sisena onada, de la qual és, en gran manera, responsable. A mesura que s'ha anat distribuint arreu del món, fins a finals de 2021, aquesta variant delta s'ha anat diferenciant en més de 200 llinatges i sublinatges designats com a AY, adquirint noves mutacions i delecions, no només a l'espícula, sinó també a altres gens del virus (Shu i McCauley, 2017).

2.2. Òmicron

El 9 de novembre de 2021 es va confirmar el primer cas d'una variant que, més tard, es va classificar com a pertanyent al llinatge B.1.1.529 en el Lancet Laboratories de Johannesburg, Sud-àfrica. Aquesta variant va ser batejada per l'OMS com a *variant òmicron* i en

pocs dies va passar a ser considerada VOC pel Centre Europeu de Prevenció i Control de Malalties (European Centre for Disease Prevention and Control, ECDC) (ECDC, 2022). Encara que els primers casos van ser detectats a Botswana i després a Sud-àfrica, els experts suggereixen que la variant va sorgir cap a principis d'octubre en algun lloc que es desconeix. Però és interessant esmentar que es va observar un fort augment de la càrrega viral en aigües residuals a l'àrea de Pretòria cap a finals d'octubre o principis de novembre, i ja se sap que l'anàlisi en aigües residuals és un bon indicador predictiu i primerenc d'alerta (Carcereny *et al.*, 2021; NICD, 2022). La variant òmicron ha estat la causa d'una nova onada pandèmica amb una elevada taxa d'incidència, primer als països del sud d'Àfrica i més tard a tot el món, encara que la població hagués estat prèviament molt exposada a aquest virus i hagués adquirit una bona immunitat natural, o bé hagués adquirit aquesta immunitat per campanyes de vacunació massiva que han permès arribar a taxes elevades de cobertura vacunal. Durant aquesta darrera onada, la variant òmicron ha estat capaç de desplaçar la delta com a variant dominant, tal com la delta va fer amb la variant alfa. És evident que aquesta variant presenta avantatges selectius respecte a les altres. La seqüenciació del seu genoma complet ha permès identificar que només en el gen de l'espícula, aquesta variant té seixanta mutacions respecte de la de Wuhan (ref. MN908947.3 GenBank), amb trenta substitucions no sinònimes, tres delecions (del) curtes i una inserció (ins): A67V, DEL69-70, T95I, G142D, del143-145, del211-212, ins214EPE, G339D, S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493K, G496S, Q498R, N501Y, Y505H, T547K, D614G, H655Y, N679K,

P681H, N764K, D796Y, N856K, Q954H, N969K, L981F. Algunes d'aquestes mutacions estan relacionades amb l'increment de la transmissibilitat per l'augment d'unió del virus al receptor ACE2 i/o per l'escapament a la resposta immunitària; moltes d'aquestes mutacions, a més, han estat observades de manera individual en altres variants (Zahradník *et al.*, 2021). Per exemple, aquesta variant és portadora de les mutacions S477N, E484A, Q498R i N501Y (aquesta última comuna entre les variants alfa, beta i gamma), mutacions clau del domini d'unió al receptor (RBD, de l'anglès *receptor binding domain*) per a unir-se a ACE2. També conté mutacions de l'espícula observades en altres variants com la del69-70 (escapament a resposta immunitària detectada en alfa), del143-145 (delta), K417N (beta), T478K (delta), H655Y (gamma), N679K (gamma) o P681H (alfa en el lloc de tall de la Furina). A més, aquesta variant presenta altres canvis en diferents regions genòmiques (Kim *et al.*, 2020): K38R, V1069I, del1265, L1266I i A1892T en la proteïna no estructural 3 (nsp3); T492I en nsp4; P132H en nsp5; del105-107 i A189V en nsp6; P323L en nps12; I42V en nsp14; així com T9I en la proteïna estructural (E); D3G, Q19E i A63T en la proteïna M; P13L, del31-33, R203K i G204R en la proteïna N.

El que hom espera, per experiència d'altres pandèmies, és que el virus evolucioni cap a variants més transmissibles però més silencioses i esdevingui un virus estacional com els coronavirus estacionals (OC43, NL63, HKU1 o 229E). I és per aquesta evasió immunològica que, de fet, la majoria de les hospitalitzacions i morts s'estan donant en pacients no vacunats, i s'ha observat un increment de les infeccions en persones amb vacunació completa (infeccions irruptives o *breakthrough infections*), encara que sense evolucionar a malaltia greu. És ben sabut que les vacunes contra els virus respiratoris, com en el cas del virus de la grip, no protegeixen de la infecció, però sí de la malaltia greu i l'hospitalització per complicacions clíniques. Cal esmentar que, degut al gran nombre de mutacions a la proteïna de l'espícula, la majoria dels anticossos monoclonals per al tractament de la malaltia en els primers estadis de la infecció no serveixen ja per a òmicron perquè aquests virus han demostrat *in vitro* ser-hi resistents (Chen *et al.*, 2022).

L'aparició sobtada d'òmicron ha obert les portes a especular sobre el seu origen, i s'han proposat diverses opcions del que hauria pogut passar:

1. D'una manera similar a com es va suggerir per a la variant alfa (B.1.1.7) (Haseltine, 2021), aquestes variants amb gran nombre de mutacions podrien tenir l'origen en pacients amb una resposta immunitària compromesa (immunosuprimits). En aquests pacients, la infecció pot passar a ser una infecció de llarga durada, ja que són pacients que pateixen algun tipus d'immunosupressió, de manera que la seva resposta immunitària no és prou efectiva per a eliminar el virus. Quan això passa, el virus pot continuar infectant i replicant-se, produint noves partícules virals i acumulant mutacions (indels, mutacions puntuals...) durant el procés de replicació viral, normalment sota la pressió selectiva d'un tractament amb activitat antiviral que pot, a més a més, condicionar l'emergència de variants amb cert grau de resistència. Per exemple, si apareix una mutació que li permeti escapar a un tractament en curs amb anticossos monoclonals o d'un antiviral, especialment en monoteràpia, aquest mutant pot ser seleccionat i el virus pot adquirir resistència.

Per què hi ha sospites que òmicron hauria pogut sorgir al sud de l'Àfrica? A Sud-àfrica trobem un ambient propici a la selecció de noves variants. Per una banda, tenim una alta taxa d'incidència de noves infeccions, i per l'altra, una població amb taxes d'immunosupressió molt altes degudes a la prevalença elevada de la infecció per VIH (KFF, 2021; UNAIDS, 2022); són pacients amb una resposta immunitària feble, i això fa factible la selecció de noves variants en aquest escenari, tal com hem esmentat anteriorment.

2. Una altra hipòtesi és que òmicron hagi aparegut en un país on no es faci un seguiment de les variants, i des d'allà s'hagi transmès de manera silenciosa i hagi arribat a Sud-àfrica i d'aquí a la resta del món. En aquest sentit, la seqüenciació a tot el continent africà és molt inferior a la que es duu a terme a països desenvolupats i les variants intermèdies d'òmicron haurien pogut passar desapercebudes. De totes maneres, no hi ha encara una explicació per a saber per què no s'han detectat variants intermèdies prèvies a òmicron. Caldria, a més, que en aquesta regió on podria haver sorgit l'òmicron el virus estigués sotmès a una pressió selectiva i a infeccions persistents.

3. També hauria pogut ser per zoonosi reversa (transmissió d'humà a animal i tornada a humà). Hi ha consens en la comunitat científica de l'origen zoonòtic del SARS-CoV-2 (Andersen *et al.*, 2020; Calisher *et al.*, 2021; World Health Organization, 2021; Zhou *et al.*, 2020;

Wacharapluesadee *et al.*, 2021). Es proposa que l'origen devia ser en un virus de ratpenat (el més proper fins ara és el del RaTG13), des del qual va passar a un hoste intermediari, on va evolucionar a partir de múltiples salts des de l'animal salvatge fins a l'humà (zoonosi) i vice-versa (zooantroponosi), fins que va aparèixer el SARS-CoV-2 (Sikkema i Koopmans, 2021; Jia *et al.*, 2021). La seqüenciació massiva va ser essencial per a demostrar que els brots del SARS-CoV-2 a les granges de visons de Dinamarca i dels Països Baixos van ser causats per humans, que el virus va poder evolucionar per a adaptar-se a infectar de manera eficient els visons, i de nou va reinfectar els humans ja lleugerament evolucionat. Tot i que s'ha demostrat transmissió d'home a animals salvatges com lleons en els zològics, i també en l'àmbit domèstic, com ara en gossos i gats, aquestes infeccions han dut a punts i final, és a dir, no s'ha pogut traçar que des d'aquests animals infectats el virus infectés altres animals o tornés als humans, i per tant, fins ara, no hi ha proves d'una transmissió eficient en aquests animals. Per tant, ara per ara, amb les variants actuals, es descarta que aquests animals tinguin un paper destacat en la pandèmia. De totes maneres, cal ser prudents i evitar la infecció massiva d'animals per a no facilitar l'adaptació viral a aquests nous hostes, cosa que afavoriria l'emergència de variants amb efectes impredecibles (Jia *et al.*, 2021). Aquesta hipòtesi, que explicaria l'origen de la variant òmicron, té algunes consideracions que la posen en dubte i calen més dades per a demostrar-la. Els nostres estudis (Andrés *et al.*, 2020) demostren que el virus ha patit un procés d'adaptació a l'humà a mesura que ha infectat la població. Les variants de preocupació alfa, beta i delta presenten freqüències de delecions que generen virus defectius més baixes que les variants a l'inici de la pandèmia B.1.5, B.1.1 o B.1.177, la qual cosa pot anar acompanyada de càrregues virals més elevades (observat en la variant delta) i, en conseqüència, suposar un dany tissular més important en el tracte respiratori (Tegally *et al.*, 2020 i 2021; Campbell *et al.*, 2021). Un virus que s'està adaptant a l'humà augmenta la seva capacitat de replicació i transmissió (també anomenada *fitness*), que va unida a la seva eficiència de transmissió. El salt a una altra espècie animal suposa un canvi d'ambient, i el virus ha de readaptar-se a les noves condicions, la qual cosa té un cost de *fitness* (Domingo, 2016). Perquè el virus s'adapti, necessita un cert temps per a adquirir mutacions que li permetran ser millor en el nou hostatger a costa de perdre *fit-*

ness en l'hostatger original (l'humà). Si la hipòtesi d'origen animal d'òmicron fos certa, caldria temps per a explicar-la, la qual cosa implicaria que el virus s'hauria adaptat temporalment a infectar un hoste no humà i que en el moment de tornar a infectar l'humà aquesta nova variant hauria necessitat temps per a recuperar *fitness* per a arribar a tenir l'avantatge selectiu respecte de la variant delta i la resta de variants. De fet, els estudis epidemiològics i filogenètics mostren més semblança d'òmicron amb variants de mitjan 2020 que respecte d'alfa, beta o delta.

Ara per ara, la discussió sobre l'origen d'òmicron està oberta i no es pot descartar cap de les tres possibilitats, ni tampoc cap que resulti de la seva combinació.

3. Què és l'NGS i què hem après de la pandèmia del SARS-CoV-2?

La tècnica de seqüenciació de nova generació o NGS consisteix en la seqüenciació de milions de fragments de DNA en una mateixa carretera de seqüenciació. Altres maneres d'anomenar aquesta metodologia són *seqüenciació massiva*, *seqüenciació profunda*, *seqüenciació massiva en paral·lel* i *seqüenciació d'alt rendiment*, entre d'altres, totes igualment vàlides. La sigla NGS (*next generation sequencing*) es va emprar per primer cop l'any 2000 i encara és acceptada com a oposició a la metodologia de seqüenciació clàssica coneguda com a *seqüenciació de Sanger*, basada en una reacció química de síntesi i l'ús de dideoxinucleòtids trifosfat (ddNTP) com a terminadors de la reacció de còpia d'un DNA i lectura per electroforesi capil·lar. Amb la metodologia de seqüenciació de Sanger, obtenim una sola seqüència per capil·lar i, per tant, si tenim una barreja de genomes virals en la mostra que hem d'estudiar, la seqüència que obtindrem serà la seqüència consens, és a dir, que a cada posició de nucleòtids veurem el nucleòtid o nucleòtids més representats, però que no és fidel a tota la diversitat genètica present en la mostra. Per tant, no es podran observar totes les variants al·lèliques que hi són presents, especialment les minoritàries. De fet, llegint els pics de l'electroferograma d'una carrera (*run*) de seqüenciació de Sanger, sí que es poden intuir mutacions, però només les presents per sobre d'un 20 %. En canvi, a més de permetre incrementar el rendiment dels processos de seqüenciació, per a estudiar genomes complets i per a un gran nombre de mostres, l'NGS permet identificar i quantificar genomes virals individuals en una barreja complexa, ja sigui

estudiant el fragment d'un genoma en profunditat o inclús el genoma sencer per assemblatge de les seqüències que resulten d'aquest procés.

Així, l'NGS s'ha convertit en una eina clau per a la identificació i caracterització ràpides de nous agents virals. La detecció ràpida d'un agent infeccios permet implementar mesures de control en el moment que hi ha pocs casos, de manera que augmenta l'efectivitat de les accions quan es produeixen brots o l'emergència de nous virus. Qualsevol endarreriment en la identificació o la no identificació d'un patogen suposa una reducció de l'eficàcia de la resposta. En el cas de la pandèmia del SARS-CoV-2, l'NGS va permetre seqüenciar en poques setmanes el genoma complet d'aquest nou virus dels primers casos diagnosticats a finals de desembre de 2019; la seqüència genètica va estar disponible públicament al banc de dades del Centre Nacional per a la Informació Biotecnològica (NCBI) el 17 de gener de 2020 (MN908947.3). Disposar d'aquesta seqüència va permetre un gran avenç en la lluita contra la pandèmia.

La seqüenciació del SARS-CoV-2 ha ajudat a caracteritzar les regions clau del genoma viral per a:

a) Desenvolupar vacunes altament efectives basades en plataformes d'mRNA/DNA en un temps rècord de sis mesos (World Health Organization, 2021). Cal destacar que es van trigar trenta-quatre anys (1954-1988) per a desenvolupar una vacuna per a la varicel·la, quinze per al virus del papil·loma humà (1991-2006), nou per al xarampió (1954-1963), set per a la primera vacuna de la poliomièlitis (1948-1955) i quatre per a les galteres (1963-1967).

b) Desenvolupar solucions diagnòstiques qualitatives i quantitatives basades en PCR per a confirmar la presència o absència del virus en mostres respiratòries i no respiratòries (Chan *et al.*, 2020; Corman *et al.*, 2020).

c) Detectar mutacions rellevants associades a un escapament a la resposta immunitària adquirida per la vacunació o infecció natural, resistència als antivirals o als tractaments basats en anticossos monoclonals o plasma convalescent (Ruiz-Rodríguez *et al.*, 2021).

d) Estudiar la composició de les quasispècies presents en els teixits d'un pacient (Piñana *et al.*, 2021).

e) Fer el seguiment de l'evolució natural del virus per a identificar la circulació de noves variants i associar-ne la prevalença a la situació epidemiològica del moment. A més, ha

permès seqüenciar milions de genomes i compartir-los a la plataforma GISAID (Global Initiative on Sharing All Influenza Data) (Elbe i Buckland-Merrett, 2017; Shu i McCauley, 2017; John Hopkins University [JHU], 2022), que ha servit com a font de dades per a altres webs que permetien visualitzar-ne l'evolució i la distribució:

— Nextstrain: <https://nextstrain.org/ncov/gisaid/global>;

— Outbrak.info: <https://virological.org/t/outbreak-info-sars-cov-2-mutation-situation-reports/629>;

— CoV-lineages: https://cov-lineages.org/lineage_list.html.

Tot el monitoratge de variants, llinatges i subllinatges en circulació es duu a terme principalment mitjançant la seqüenciació de genoma complet, tasca clau per al seguiment de l'evolució de la pandèmia i la presa de decisions per les institucions de salut pública.

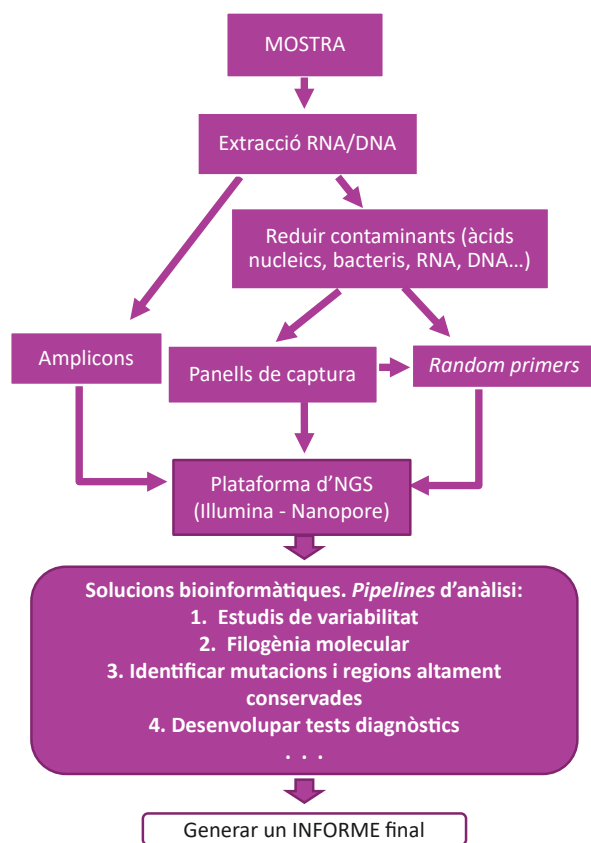
f) Estudiar com canvia genèticament el virus al llarg del temps, quines mutacions poden anar associades a una transmissibilitat i/o gravetat més elevades, l'origen de les noves variants, la seva ruta i/o eficàcia de transmissió, i dur a terme estudis de les reinfeccions, dels escapaments vacunals i de la infecció persistent (Hodcroft *et al.*, 2021; Ruiz-Rodríguez *et al.*, 2021; Andrés *et al.*, 2022).

4. Ús de l'NGS per a identificar i caracteritzar virus/patògens emergents, reemergents o nous

Es poden seguir diferents estratègies metodològiques per a la seqüenciació de genomes virals, depenent de si en som o no coneixedors del genoma (encebadors a l'atzar o *random primers* en anglès), o bé si disposem d'un mínim d'informació (enriquitment per amplicons o panells de captura). Aquestes estratègies estan esquematitzades en la figura 2.

4.1. Seqüenciació de genoma complet emprant encebadors a l'atzar

Pel que fa a la seqüenciació de genoma complet (WGS, de l'anglès *whole-genome sequencing*), els encebadors a l'atzar (o *random primers*) més emprats són els que tenen sis nucleòtids de llargària (hexàmers), de cadena simple i inclouen totes les possibles combinacions de les quatre bases nucleotídiques (A, C, T i G), de manera que s'obté una barreja de 4.096 (4⁶) hexàmers diferents. Com que tots els hexàmers possibles estan presents, es poden unir a qualsevol fragment d'RNA o DNA i actuar com a iniciadors de còpia del material genò-



↑ Figura 2. Esquema general de les metodologies per a la seqüenciació de genomes virals. Elaboració pròpia.

mic i permetre'n l'amplificació i seqüenciació posterior.

Avantatges:

- No cal tenir informació del genoma.
- Permet identificar tots els patògens de qualsevol mostra (plasma, sèrum, femtes, aigües residuals, etc.), cosa que permet fer estudis metagenòmics.

- El cost dels encebadors per còpia i amplificació és molt baix.

Desavantatges:

- Els genomes minoritaris sempre queden infrarepresentats i moltes vegades no es detecten; per tant, és una tècnica que té una baixa sensibilitat.

- Cal tenir una quantitat elevada del material genòmic de l'agent que volem detectar, ja que la proporció de seqüències de l'agent viral infeccios que volem detectar sempre és molt inferior.

- El material genòmic de l'hostatger o de microorganismes presents és, de fet, un contaminant, i la seva presència, a més de reduir l'eficiència de la seqüenciació, suposa també un cost computacional. Una possible solució és dur a terme mètodes d'enriquiment, que en-

careixen el cost de la tècnica (Fernandez-Cassi *et al.*, 2018; Chrzastek *et al.*, 2017).

4.2. Captura de seqüències, recuperació de fragments i NGS

Un mètode d'enriquiment del patògen que hom vol seqüenciar es basa en l'ús de panells de captura i consisteix a emprar petites sondes complementàries al genoma de l'agent patògen.

Avantatges:

- Podem enriquir el genoma de l'agent patògen en una barreja complexa de genomes.

- La seqüenciació del genoma complet té un cost baix perquè no cal una gran profunditat de seqüenciació per a obtenir les seqüències del virus que podrem assemblar per a la seva detecció i caracterització.

- Les variants minoritàries es preserven i permeten obtenir una imatge més fidel de la variabilitat de la mostra.

Desavantatges:

- Per a dissenyar les sondes de captura cal conèixer la seqüència del microorganisme.

- No permet seqüenciar nous patògens, però sí patògens molt relacionats genètica-

ment en compartir regions altament conservades.

- Requereix gran expertesa en la preparació de la mostra com a pas previ a la seqüenciació.

- És molt dependent de la profunditat de seqüenciació.

4.3. Ús d'amplicons

Consisteix en el disseny d'encebadors específics que permetin amplificar un segment del genoma (*targeted sequencing*), o tot el genoma complet (WGS) emprant encebadors de fragments encavalcats. Aquesta estratègia ha estat la més emprada en el nostre laboratori (Gregori *et al.*, 2014; Quer *et al.*, 2015; Rodríguez-Frías *et al.*, 2017; Chen *et al.*, 2020) i el mètode escollit per a la seqüenciació del SARS-CoV-2 amb la metodologia ARTIC (Quick, 2020).

Avantatges:

- És el mètode més sensible i el que ofereix més confiança per a estudis de variabilitat.

- Permet amplificar específicament un agent infeccios enmig d'una barreja complexa de genomes minimitzant l'efecte del fons (*background*).

- Permet obtenir una bona cobertura del genoma (*coverage*), inclús a baixes concentracions del patògen en la barreja.

- Permet detectar variants presents a molt baixa freqüència.

Desavantatges:

- Cal conèixer la seqüència del microorganisme que volem detectar i dissenyar-ne els encebadors específics.

- És un mètode laboriós.

- És altament dependent de la qualitat dels encebadors, de la seva especificitat. Hi pot haver canvis en la regió de l'encebador, per exemple, per deriva genètica natural del virus, que poden crear un biaix en l'amplificació i seqüenciació, o inclús fer que no sigui possible.

- Cal emprar un nombre gran de cicles d'amplificació i, per tant, cal tenir en compte que poden aparèixer mutacions artefactuals generades en aquest procés de laboratori.

La seqüenciació massiva és, doncs, una de les eines més poderoses que tenim per a la identificació i caracterització genètica de virus emergents, reemergents o nous. El risc de noves pandèmies existeix, i la història ens demostra que l'amenaça és real. En aquesta cursa, els virus, amb la seva enorme capacitat de generar variabilitat, ens porten un gran avantatge. Enfront d'això, la tecnologia actual no permet preveure ni quan ni com apareixerà la propera

pandèmia. El que sí que podem fer és estar cada cop més ben preparats, i una de les eines clau que tenim és l'NGS. Cal crear xarxes de vigilància interconnectades, reforçar els centres de seqüenciació actualment en actiu creant nous nodes i millorar les metodologies de seqüenciació on encara hi ha recorregut per fer. L'eficàcia en la resposta a una pandèmia és

més alta com més ràpida sigui la detecció, identificació i caracterització genètiques de l'agent infecciós.

Agraïments

Volem donar les gràcies als finançadors dels nostres projectes: Pla Estratègic de Recerca i Innovació en Salut (PERIS) de la Direcció Ge-

neral de Recerca i Innovació en Salut (DGRIS), Generalitat de Catalunya; Centro para el Desarrollo Tecnológico Industrial (CDTI) del Ministerio d'Economia i Competitivitat, núm. IDI-20200297; Beca de Gilead GLD21_00006, i Beca PI19/00301 Instituto de Salud Carlos III, cofinançada per European Regional Development Fund (ERDF).

Bibliografia

- ALM, E. [et al.] (2020). «Geographical and temporal distribution of SARS-CoV-2 clades in the WHO European Region, January to June 2020». *Eurosurveillance*, 25 (32). DOI: 10.2807/1560-7917.ES.2020.25.32.2001410.
- AMANAT, F.; KRAMMER, F. (2020). «SARS-CoV-2 vaccines: Status report». *Immunity*, 52(4): 583-589. DOI: 10.1016/j.immuni.2020.03.007.
- ANDERSEN, K. G. [et al.] (2020). «The proximal origin of SARS-CoV-2». *Nature Medicine* [Estats Units], 26: 450-452. DOI: 10.1038/s41591-020-0820-9.
- ANDRÉS, C. [et al.] (2020). «Naturally occurring SARS-CoV-2 gene deletions close to the spike S1/S2 cleavage site in the viral quasispecies of COVID-19 patients». *Emerging Microbes & Infections* [Estats Units], 9 (1): 1900-1911. DOI: 10.1080/22221751.2020.1806735.
- (2022). «A year living with SARS-CoV-2: An epidemiological overview of viral lineage circulation by whole-genome sequencing in Barcelona city (Catalonia, Spain)». *Emerging Microbes & Infections* [Estats Units], 11 (1): 172-181. DOI: 10.1080/22221751.2021.2011617.
- CALISHER, C. H. [et al.] (2021). «Science, not speculation, is essential to determine how SARS-CoV-2 reached humans». *Lancet* [Londres, Anglaterra], 398 (10296): 209-211. DOI: 10.1016/S0140-6736(21)01419-7.
- CAMPBELL, F. [et al.] (2021). «Increased transmissibility and global spread of SARS-CoV-2 variants of concern as at June 2021». *Eurosurveillance*, 26 (24): 1-6. DOI: 10.2807/1560-7917.es.2021.26.24.2100509.
- CARCCERENY, A. [et al.] (2021). «Monitoring emergence of the SARS-CoV-2 B.1.1.7 variant through the Spanish national SARS-CoV-2 wastewater surveillance system (VATar COVID-19)». *Environmental Science & Technology*, 55 (17): 11756-11766. DOI: 10.1021/acs.est.1c03589.
- CHAN, J. F.-W. [et al.] (2020). «Improved molecular diagnosis of COVID-19 by the novel, highly sensitive and specific COVID-19-RdRp/Hel real-time reverse transcription-PCR assay validated in vitro and with clinical specimens». *Journal of Clinical Microbiology*, 58 (5). DOI: 10.1128/JCM.00310-20.
- CHEN, J. [et al.] (2022). «Omicron variant (B.1.1.529): Infectivity, vaccine breakthrough, and antibody resistance». *Journal of Chemical Information and Modeling*, 62 (2): 412-422. DOI: 10.1021/acs.jcim.1c01451.
- CHEN, Q. [et al.] (2020). «Deep-sequencing reveals broad subtype-specific HCV resistance mutations associated with treatment failure». *Antiviral Research* [Elsevier B. V.], 174: 104694. DOI: 10.1016/j.antiviral.2019.104694.
- CHYZASTEK, K. [et al.] (2017). «Use of sequence-independent, single-primer-amplification (SISPA) for rapid detection, identification, and characterization of avian RNA viruses». *Virology*, 509: 159-166. DOI: 10.1016/j.virol.2017.06.019.
- CORMAN, V. M. [et al.] (2020). «Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR». *Eurosurveillance*, 25 (3). DOI: 10.2807/1560-7917.ES.2020.25.3.2000045.
- DOMINGO, E. (2016). *Virus as populations: Composition, complexity, dynamics, and biological implications*. 1a ed. Amsterdam: Academic Press. 412 p.
- (2020). «Molecular basis of genetic variation of viruses: error-prone replication». A: *Virus as populations*. 2a ed. Londres: Academic Press, 35-71. DOI: 10.1016/B978-0-12-816331-3.00002-7.
- ECKERLE, L. D. [et al.] (2010). «Infidelity of SARS-CoV Nsp14-exonuclease mutant virus replication is revealed by complete genome sequencing». *PLoS Pathogens* [Estats Units], 6 (5): e1000896. DOI: 10.1371/journal.ppat.1000896.
- ELBE, S.; BUCKLAND-MERRETT, G. (2017). «Data, disease and diplomacy: GISAID's innovative contribution to global health». *Global Challenges* [Hoboken, NJ], 1 (1): 33-46. DOI: 10.1002/gch2.1018.
- EUROPEAN CENTRE FOR DISEASE PREVENTION AND CONTROL (ECDC) (2022). *Assessment of the further spread and potential impact of the SARS-CoV-2 Omicron variant of concern in the EU/EEA, 19th update*. Etcocm: ECDC.
- FERNANDEZ-CASSI, X. [et al.] (2018). «Viral concentration and amplification from human serum samples prior to application of next-generation sequencing analysis». *Methods in Molecular Biology* [Clifton, N. J., Estats Units], 1838: 173-188. DOI: 10.1007/978-1-4939-8682-8_13.
- GIBB, R. [et al.] (2020). «Zoonotic host diversity increases in human-dominated ecosystems». *Nature* [Anglaterra], 584 (7821): 398-402. DOI: 10.1038/s41586-020-2562-8.
- GREGORI, J. [et al.] (2014). «Inference with viral quasispecies diversity indices: Clonal and NGS approaches». *Bioinformatics*, 30 (8): 1104-1111. DOI: 10.1093/bioinformatics/btt768.
- GRIBBLE, J. [et al.] (2020). «The coronavirus proofreading exonuclease mediates extensive viral recombination». *BioRxiv*. Cold Spring Harbor Laboratory. DOI: 10.1101/2020.04.23.057786.
- HASELTINE, W. A. (2021). «Persistently infected Covid-19 patients: A potential source for new variants». *Forbes* [en línia]. <https://www.forbes.com/sites/williamhaseltine/2021/02/16/persistently-infected-covid-19-patients-a-potential-source-for-new-variants/?sh=771761de208a> [Consulta: 14 març 2022].
- HODCROFT, E. B. [et al.] (2021). «Spread of a SARS-CoV-2 variant through Europe in the summer of 2020». *Nature* [Anglaterra], 595 (7869): 707-712. DOI: 10.1038/s41586-021-03677-y.
- JIA, P. [et al.] (2021). «New approaches to anticipate the risk of reverse zoonosis». *Trends in Ecology & Evolution*, 36 (7): 580-590. DOI: 10.1016/j.tree.2021.03.012.
- JOHN HOPKINS UNIVERSITY (JHU) (2022). *COVID-19 dashboard by the Center for Systems Science and Engineering (CSSE) at John Hopkins University (JHU)* [en línia]. Baltimore, Estats Units. <https://coronavirus.jhu.edu/map.html> [Consulta: 14 març 2022].
- KFF (2021). «The global HIV/AIDS epidemic» [en línia]. <https://www.kff.org/global-health-policy/fact-sheet/the-global-hiv-aids-epidemic/> [Consulta: 14 març 2022].
- KIM, D. [et al.] (2020). «The architecture of SARS-CoV-2 transcriptome». *Cell*, 181 (4): 914-921.e10. DOI: 10.1016/j.cell.2020.04.011.
- KORBER, B. [et al.] (2020). «Tracking changes in SARS-CoV-2 spike: Evidence that D614G increases infectivity of the COVID-19 virus». *Cell*, 182 (4): 812-827.e19. DOI: 10.1016/j.cell.2020.06.043.
- LÓPEZ, M. G. [et al.] (2021). «The first wave of the COVID-19 epidemic in Spain was associated with early introductions and fast spread of a dominating genetic variant». *Nature Genetics*, 53 (10): 1405-1414. DOI: 10.1038/s41588-021-00936-6.
- MALLAPATY, S. (2022). «The hunt for the origins of Omicron». *Nature*, 602 (3 febrer): 26-28.
- NEXTSTRAIN (2021). *Genomic epidemiology of SARS-CoV-2* [en línia]. <https://nextstrain.org/ncov/global> [Consulta: 14 març 2022].
- NICD (2022). *Wastewater-based epidemiology for SARS-CoV-2 in South Africa* [en línia]. <https://www.nicd.ac.za/diseases-a-z-index/disease-index-covid-19/surveillance-reports/weekly-reports/wastewater-based-epidemiology-for-sars-cov-2-in-south-africa/> [Consulta: 4 desembre 2022].
- O'TOOLE, A. [et al.] (2022). *SARS-CoV-2 lineages* [en línia]. <https://cov-lineages.org/lineage_list.html> [Consulta: 14 març 2022].
- PIÑANA, M. [et al.] (2021). «Viral populations of SARS-CoV-2 in upper respiratory tract, placenta, amniotic fluid and umbilical cord blood support viral replication in placenta». *Clin. Microbiol. Infect.*, 27 (10): 1542-1544. DOI: 10.1016/j.cmi.2021.07.008.
- QUER, J. [et al.] (2015). «High-resolution hepatitis C virus subtyping using NS5B deep sequencing and phylogeny, an alternative to current methods». *Journal of Clinical Microbiology*, 53 (1): 219-226. DOI: 10.1128/JCM.02093-14.
- QUICK, J. (2020). *nCoV-2019 sequencing protocol* [en línia]. <https://www.protocols.io/view/ncov-2019-sequencing-protocol-v3-locost-bp2l6n26rgqe/v3> [Consulta: 19 setembre 2022].
- RAMBAUT, A. [et al.] (2020). «A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology». *Nature Microbiology* [Anglaterra], 5 (11): 1403-1407. DOI: 10.1038/s41564-020-0770-5.
- RODRÍGUEZ-FRÍAS, F. [et al.] (2017). «High HCV subtype heterogeneity in a chronically infected general population revealed by high-resolution hepatitis C virus subtyping». *Clinical Microbiology and Infection*, 23 (10): 775.e1-775.e6. DOI: 10.1016/j.cmi.2017.02.007.
- (2021a). «Microorganisms as shapers of human civilization, from pandemics to even our genomes: Villains or friends? A historical approach». *Microorganisms*, 9 (12). DOI: 10.3390/microorganisms9122518.
- (2021b). «Evolutionary and phenotypic characterization of two spike mutations in European lineage 20E of SARS-CoV-2». *MBio*, 12 (6): e0231521. DOI: 10.1128/mBio.02315-21.
- ROGER, F. [et al.] (2021). *Origin of the Covid-19 virus: The trail of mink farming* [en línia]. <https://theconversation.com/origin-of-the-covid-19-virus-the-trail-of-mink-farming-155989> [Consulta: 14 març 2022].
- RUIZ-RODRÍGUEZ, P. [et al.] (2021). «Evolutionary and phenotypic characterization of two spike mutations in European Lineage 20E of SARS-CoV-2». *MBio*, 12 (6): e02315-21. DOI: 10.1128/mBio.02315-21.
- SHU, Y.; MCCAULEY, J. (2017). «GISAID: Global initiative on sharing all influenza data - from vision to reality». *Eurosurveillance*: 30494. DOI: 10.2807/1560-7917.ES.2017.22.13.30494.
- SIKKEMA, R. S.; KOOPMANS, M. P. G. (2021). «Preparing for emerging zoonotic viruses». A: *Encyclopedia of virology* [en línia]. Elsevier, 256-266. DOI: 10.1016/B978-0-12-814515-9.00150-8.
- SMITH, E. C.; DENISON, M. R. (2013). «Coronaviruses as DNA wannabes: A new model for the regulation of RNA virus replication fidelity». *PLoS Pathogens*, 9 (12): e1003760. DOI: 10.1371/journal.ppat.1003760.
- TEGALLY, H. [et al.] (2020). «Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple

- spike mutations in South Africa». *MedRxiv*, 2020.12.21.20248640.
- TEGALLY, H. [et al.] (2021). «Detection of a SARS-CoV-2 variant of concern in South Africa». *Nature*, 592 (7854): 438-443. DOI: 10.1038/s41586-021-03402-9.
- TOLLEFSON, J. (2020). «Why deforestation and extinctions make pandemics more likely». *Nature* [Anglaterra], 584: 175-176. DOI: 10.1038/d41586-020-02341-1.
- UNAIDS (2022). *Global data on HIV epidemiology and response* [en línia]. <<https://aidsinfo.unaids.org/>> [Consulta: 14 març 2022].
- WACHARAPLUESADEE, S. [et al.] (2021). «Evidence for SARS-CoV-2 related coronaviruses circulating in bats and pangolins in Southeast Asia». *Nature Communications*, 12 (1): 972. DOI: 10.1038/s41467-021-21240-1.
- WEARN, O. R. [et al.] (2012). «Extinction debt and windows of conservation opportunity in the Brazilian Amazon». *Science* [Nova York, Estats Units], 337 (6091): 228-232. DOI: 10.1126/science.1219013.
- WORLD HEALTH ORGANIZATION (2021). *WHO-convened global study of origins of SARS-CoV-2: China Part* [en línia]. Ginebra, Suïssa: WHO. <<https://www.who.int/publications/i/item/who-convened-global-study-of-origins-of-sars-cov-2-china-part>> [Consulta: 14 març 2022].
- WOUTERS, O. J. [et al.] (2021). «Challenges in ensuring global access to COVID-19 vaccines: production, affordability, allocation, and deployment». *Lancet* [Londres, Anglaterra], 397 (10278): 1023-1034. DOI: 10.1016/S0140-6736(21)00306-8.
- ZAHRADNÍK, J. [et al.] (2021). «SARS-CoV-2 variant prediction and antiviral drug design are enabled by RBD in vitro evolution». *Nature Microbiology* [Anglaterra], 6 (9): 1188-1198. DOI: 10.1038/s41564-021-00954-4.
- ZHOU, H. [et al.] (2020). «A novel bat coronavirus closely related to SARS-CoV-2 contains natural insertions at the S1/S2 cleavage site of the spike protein». *Current Biology*, 30 (11): 2196-2203.e3. DOI: 10.1016/j.cub.2020.05.023.