

Genética y evolución en la pandemia de COVID-19

Fernando González Candelas

Unidad Mixta Infección y Salud Pública FISABIO-Universidad de Valencia, Instituto de Biología Integrativa de Sistemas, I2SysBio (CSIC-UV). Avda. Cataluña, 21. E-46020. Valencia
CIBER en Epidemiología y Salud Pública

Resumen

La pandemia ocasionada por el coronavirus SARS-CoV-2 ha alterado la vida, la economía, las relaciones sociales y personales en nuestro planeta y, desgraciadamente, ha producido más de 5 millones de muertes (datos a noviembre de 2021). Por otra parte, ha puesto de manifiesto el papel esencial que tiene la investigación biomédica y lo dependiente que es nuestra sociedad de muchos trabajos científicos que solían pasar inadvertidos. Además de mostrar el poder que se deriva de la acción coordinada y cooperativa de grupos de investigación, la pandemia también ha puesto un foco en disciplinas que, en gran medida, han sido ignoradas hasta el momento. La aparición de variantes de preocupación e interés y su relación con los incrementos en la incidencia de las infecciones en diferentes olas o la posibilidad de que representen escapes vacunales ha llevado a la OMS a establecer la vigilancia genómica como una de las herramientas fundamentales para conocer y controlar la expansión del virus. En este artículo, vamos a analizar algunos aspectos fundamentales de la evolución y genética de las poblaciones del virus y cómo nos pueden ayudar a interpretar adecuadamente los datos derivados de los análisis de secuencias completas del genoma viral, método ideal para realizar dicha vigilancia.

Palabras clave

Escape vacunal, SARS-CoV-2, selección natural, transmisión, variantes, vigilancia genómica.

El SARS-CoV-2 es un betacoronavirus (subgénero *Sarbecovirus*), un virus de ARN, de cadena sencilla y polaridad positiva (lo que le permite expresar directamente su mensaje genético cuando infecta una célula) que comparte estas y otras características con otros miembros de esa misma familia, como el SARS-CoV, el virus causante de infecciones respiratorias agudas que en 2002-2003 afectó a unas 10000 personas con una letalidad de aproximadamente el 10 %. Gracias a los estudios realizados en varios laboratorios a consecuencia de ese brote de SARS (y también de los casos de MERS desde 2012), el conocimiento inicial del nuevo coronavirus pudo avanzar rápidamente (Wu *et al.*, 2020). Si apenas transcurrieron 10 días desde que se sospechó que varios casos de neumonía grave en la ciudad china de Wuhan podían deberse a un virus desconocido hasta que se hizo pública la primera secuencia completa de su genoma, en cuestión de horas se dispuso de pruebas diagnósticas para el nuevo virus (basadas en RT-qPCR sobre regiones específicas de ese genoma) y se empezaron a diseñar lo que unos meses más tarde nos han suministrado como vacunas de ARNm, una novedad farmacológica acelerada por la necesidad de controlar una pandemia prácticamente desbocada.

Desde esa primera secuencia de un genoma hasta la actualidad, se han obtenido y hecho públicas las secuencias de más de 5 millones de genomas virales (<https://www.gisaid.org>), proporcionando una información sobre la variabilidad genómica de este virus que no tiene equivalente para ningún otro organismo. Explorar esta información ha supuesto un reto a muchos niveles, desde la informática hasta la biología molecular, la biotecnología o la biología evolutiva. A estos niveles se les añaden dos componentes externos al virus, pero esenciales para entender su evolución: el huésped y el ambiente, así como las interacciones que se establecen entre ellos y con el propio virus (**Fig. 1**).

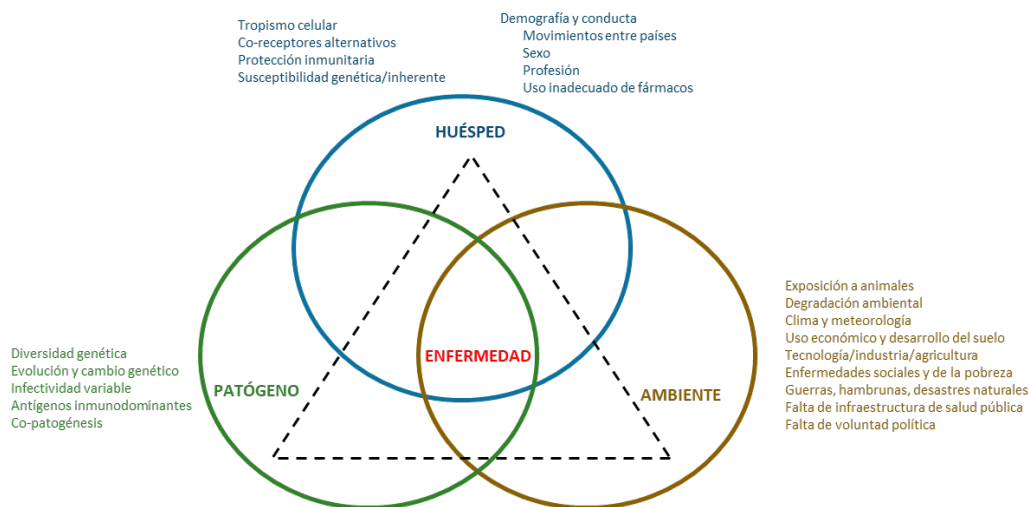


Figura 1. Perspectiva multifactorial de las enfermedades infecciosas.

Tomemos, por ejemplo, el efecto que tiene la mutación sobre la evolución del virus. Como sabemos, la mutación es la fuente última de toda variación genética. Una de las características de los coronavirus es que poseen una cierta capacidad de corrección de errores en el proceso de replicación de su material hereditario, lo que hace que su tasa de mutación sea inferior a la de otros virus ARN (del orden de $10^{-4} - 10^{-5}$ sustituciones/sitio/replicación), casi dos órdenes de magnitud inferior a la del VIH (virus de la inmunodeficiencia humana) o el VHC (virus de la hepatitis C). Esa tasa de mutación se traduce en una tasa de evolución de aproximadamente dos sustituciones al mes cuando se analiza la acumulación de diferencias en un linaje viral a lo largo del tiempo, comparado con el virus del que deriva (Duchêne *et al.*, 2020).

Sin embargo, ese valor es un promedio y tiene una gran varianza, lo que implica que algunos linajes acumularán muchas más mutaciones y otros muchas menos que ese valor. Pero, además, ese valor se ha obtenido a partir de datos de virus obtenidos de pacientes en las primeras etapas sintomáticas de su infección, cuando se toman las muestras para el diagnóstico que acaban siendo secuenciadas y depositadas en las bases de datos. De ahí son tomadas por los grupos que estudian la evolución y epidemiología molecular del virus, obteniendo las estimaciones de tasa de evolución ya indicadas. No obstante, hay una fracción de infecciones que se prolongan en el tiempo, de forma per-

sistente, especialmente en casos de personas inmunodeprimidas y que son sometidos a tratamientos, algunos de ellos basados en anticuerpos mono- o policlonales frente al SARS-CoV-2. En varios de estos pacientes se ha observado una acumulación de mutaciones con una tasa mucho más elevada. Es la consecuencia de la acción continuada de la selección natural sobre la población viral infectiva, que tiene que hacer frente a la presión de esos fármacos. El resultado es que aquellas mutaciones que favorecen el escape a la neutralización por esos anticuerpos usados en el tratamiento son favorecidas y se acumulan rápidamente, arrastrando con ellas otras mutaciones que pueden haberse producido en los mismos genomas y que no tienen efecto sobre la eficacia del virus. De hecho, se ha comprobado que muchas de esas mutaciones coinciden con las observadas en las variantes de preocupación (VOCs, por *Variants of Concern*) y de interés (VOIs) y que han ocasionado alarmas ante la posibilidad de escape a la inmunidad conferida tanto por las vacunas como por la infección natural.

Si bien la mutación es la principal fuente de variabilidad genética, sobre esta actúan diversos procesos evolutivos, casi siempre de forma concurrente, aunque con intensidad variable, lo que dificulta enormemente la realización de predicciones precisas. Hemos comentado la acción de la selección intrapaciente en los casos de infección persistente, pero la selección natural también actúa en los casos de infección aguda, la más habitual en la COVID-19, en la que se resuelve en unas dos semanas. En estos casos, y con independencia de la gravedad que pueda ocasionar la infección en las personas afectadas, la persistencia de la población del virus depende de que pueda encontrar un nuevo huésped en el que proseguir su ciclo infectivo. Por tanto, si hay una característica del virus que puede ser favorecida por la selección natural es su mayor transmisibilidad.

Este objetivo puede cubrirse de varias formas: menor tiempo hasta alcanzar una cantidad de virus infectivos en las vías respiratorias, mayor carga viral en las mismas, mayor persistencia en el ambiente, etc. Muchas de estas propiedades dependen de una proteína viral concreta, la espícula (proteína S o “spike”), sobre la que se han concentrado los esfuerzos de muchos investigadores. La presión selectiva sobre este carácter ha ido modificándose a lo largo de la pandemia, y seguramente continuará haciéndolo mientras el virus permanezca en nuestras poblaciones.

En las etapas iniciales, prácticamente la totalidad de la población humana era susceptible a la infección por el SARS-CoV-2. En esas circunstancias, la presión selectiva era relativamente pequeña y la mayoría de las mutaciones que aparecían en la espícula se comportaban neutralmente, es decir, no poseían una ventaja sobre otras variantes presentes en la población. Una excepción clara la representó la mutación D614G, la sustitución del aspártico de la posición 614 de la proteína de la espícula por una glicina que, con una ligera ventaja en la transmisión, llegó a desplazar casi por completo a la variante original (linaje A de Wuhan). Esta mutación apareció en varias ocasiones antes de alcanzar esa gran frecuencia, lo que resalta la importancia del entorno en que aparece una mutación para determinar su destino final, a pesar de la posible ventaja intrínseca que posee (Korber *et al.*, 2020). En este caso, la mutación que llegó a casi fijarse en la población mundial apareció en Italia, a finales de febrero de 2020, coincidiendo con el inicio de la expansión de la pandemia en ese país y, al poco en el resto de Europa.

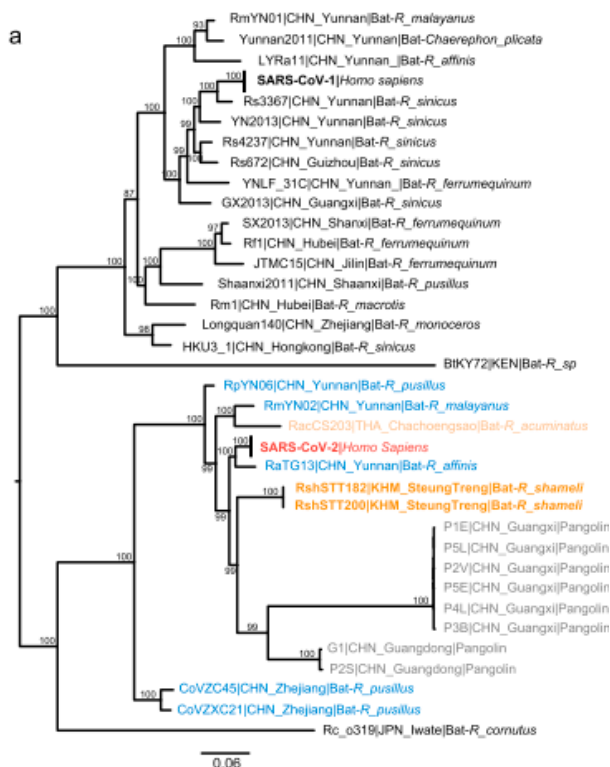
La ventaja selectiva de la mutación D614G se vio notablemente incrementada por su asociación con una rápida expansión epidemiológica, que favoreció su implantación en otros países y su gran predominancia en ellos. Una excepción a este patrón general lo representó España, donde las primeras introducciones del virus, de las que se derivaron gran parte de los casos iniciales de COVID-19 en nuestro país, se debieron a virus que no incorporaban esa mutación. En consecuencia, en las primeras semanas de la pandemia, nuestro país se caracterizó por una elevada frecuencia de la variante original (D614) que sólo fue reemplazada por la mutante (G614) tras las medidas de confinamiento que redujeron drásticamente la circulación del virus (López *et al.*, 2021).

Una evidencia directa del poderoso efecto que ejerce la situación epidemiológica sobre la distribución de la variación genética del virus, con independencia de cuál sea esta, la proporcionó la expansión de otra mutación de la espícula, A222V, en varios países europeos durante el verano de 2020. Esta mutación está asociada al linaje B.1.177, también conocido como 20E (EU1), y fue detectada originalmente en el Noreste de España a mediados de junio de 2020, vinculada a brotes de trabajadores temporeros agrícolas. En poco tiempo estos brotes iniciales se extendieron a otras regiones de nuestro país y en julio era la variante mayoritaria, con más del 80 % de las secuencias correspondiendo a la misma. En esas fechas, las autoridades levantaron las restricciones de movilidad internacional y se permitió la entrada de turistas en nuestro territorio. Aunque se mantenían algunas restricciones y recomendaciones, tanto en España como en otros países, las vacaciones estivales actuaron como catalizadoras de las interacciones personales, con una bajada en el cumplimiento de esas medidas de control. El resultado fue un aumento progresivo de la incidencia no sólo en nuestro país sino, también, en otros muchos países europeos. En algunos, como el Reino Unido, Suiza, Noruega o Irlanda, la variante B.1.177 pasó a ser también dominante. Nuestros análisis (Hodcroft *et al.*, 2021) demostraron una relación directa entre la movilidad a y desde España de estos países y el aumento en la frecuencia de esta variante. Sin embargo, los experimentos realizados en el laboratorio no lograron demostrar que la mutación A222V estuviese asociada a una mayor transmisibilidad del virus. El aumento de su frecuencia, en realidad de todo el linaje, no se debía, como en el caso de la variante alfa, a una mayor transmisibilidad sino al efecto de crecimiento epidemiológico independiente de la naturaleza del virus. Ello sin que hubiese diferencias sustanciales en la dinámica de aumento cuando se comparó con la de la variante alfa, por ejemplo. De hecho, en otros países europeos, con los que el intercambio de viajeros con España también fue frecuente, no se observó el aumento de esta variante, a pesar de haberse detectado en los mismos. El ejemplo más claro lo proporciona Francia, con frecuente movimiento de personas con nuestro país, y en el que se estableció en esas fechas otra variante (20A.EU2), extendida también a Bélgica. Tampoco se ha demostrado una mayor transmisibilidad de esta variante y, en competencia con la 20E (EU1), gozó de la ventaja que proporciona la anticipación, estar antes en el lugar y momento en que empezaron a aumentar los contagios.

Si la fuente última de variabilidad genética es la mutación, este no es el único proceso responsable de la variación que observamos en cualquier especie. Además de la selección, la deriva o la migración, la recombinación juega un papel importante

en la evolución de casi todos los organismos, SARS-CoV-2 incluido. Debemos aclarar que la recombinación en virus y bacterias no es estrictamente equivalente a la recombinación en organismos diploides con reproducción sexual, en la que se produce recombinación entre cromátidas hermanas en la primera etapa de la generación de gametos. En esos microorganismos, los diferentes procesos que originan recombinación lo hacen de forma asimétrica, en los que un individuo recibe material genético de otro, pero no a la inversa: no es un intercambio, es una transferencia. De hecho, la diferencia entre recombinación y transferencia génica horizontal se centra en la distancia (evolutiva) entre donante y receptor del material transferido, lo que afecta a la frecuencia y el posible destino evolutivo de ese material.

La recombinación es un proceso habitual en la evolución de los coronavirus. Si bien no se ha podido determinar su implicación directa e inmediata en la aparición del SARS-CoV-2, la documentación de su presencia en etapas evolutivas anteriores (Boni *et al.*, 2020) y en la actualidad (Gutierrez *et al.*, 2021; Jackson *et al.*, 2021) hacen que no sea una mera conjetura invocar su papel en el origen de este virus. La obtención de la primera secuencia genómica del SARS-CoV-2 en enero de 2020 permitió un análisis inmediato de su relación con otros coronavirus ya secuenciados hasta ese momento. Además de determinar que se trataba de una nueva especie viral, los primeros análisis mostraron que los virus más estrechamente relacionados con él se encontraban en distintas especies de murciélagos, todos ellos muestreados en China: RaTG13 en un ejemplar de *Rhinolophus affinis* secuenciado en 2013, RmYN02 secuenciado en 2019 procedente de un individuo de *R. malayanus*, a los que se unió, ya en 2020, RpYN06 procedente de *R. pusillus*. Un estudio reciente (Delaune *et al.*, 2021) presenta la secuencia genómica de dos aislados de Camboya que se encuentran aún más próximos a SARS-CoV-2 que los anteriores (menos RaTG13), denominados RshSTT182 y RshSTT200 (Fig. 2).



presenta la secuencia genómica de dos aislados de Camboya que se encuentran aún más próximos a SARS-CoV-2 que los anteriores (menos RaTG13), denominados RshSTT182 y RshSTT200 (Fig. 2).

Figura 2. Análisis filogenético de SARS-CoV-2 y otros sarbecovirus junto con su contexto geográfico y biológico, indicado por la especie y el país (azul: China; naranja claro: Tailandia; naranja: Camboya) en que fueron identificados. El árbol se obtuvo a partir de secuencias de genomas completos por el método de máxima verosimilitud (Delaune *et al.*, 2021).

El análisis con el genoma completo no muestra la diversidad de diferentes contribuciones al genoma de SARS-CoV-2 y de sus ancestros desde otros sarbecovirus. Para apreciarlo, es necesario un análisis más detallado de la historia filogenética de distintas porciones del genoma del virus, que estos autores han obtenido mediante el método de ventanas deslizantes. El resultado de su análisis (**Fig. 3**) muestra que ha habido varios eventos de recombinación con la consecuencia de que distintas porciones del genoma del SARS-CoV-2 están más estrechamente emparentadas con virus diferentes, si bien todos ellos incluidos entre los identificados hasta el momento.

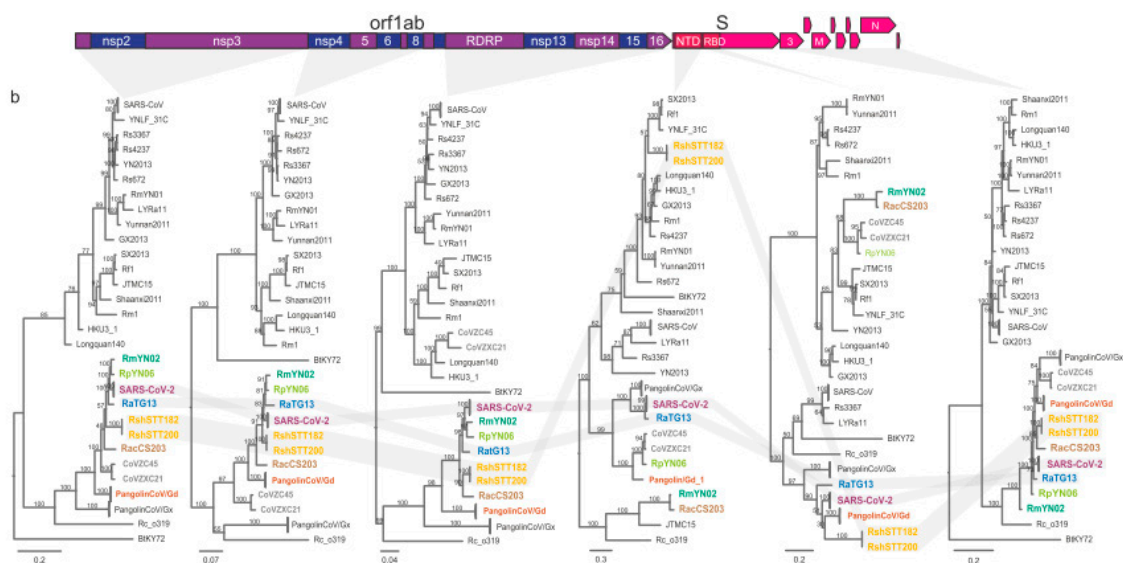


Figura 3. Árbol filogenético de diferentes regiones genómicas. Se indican los valores de soporte por *bootstrap* (1000 réplicas) de las ramas. Las secuencias de SARS-CoV y SARS-CoV-2 se han colapsado para facilitar la claridad (Delaune *et al.*, 2021).

Nuestro conocimiento de la variabilidad de los sarbecovirus es, a pesar del interés que su estudio ha despertado en los meses precedentes, aún muy limitado y es posible que nuevas expediciones de muestreo revelen ancestros aún más próximos en algún animal salvaje. Sin embargo, la hipótesis científicamente más aceptada de salto desde un hospedador natural, casi con seguridad algún murciélago, a nuestra especie a través de algún intermediario será extremadamente difícil de comprobar. A diferencia de otros saltos zoonóticos (virus ébola, VIH, etc.), el SARS-CoV-2 parece haber saltado sólo en una ocasión a la especie humana y ya en ella se ha diversificado y dado lugar a la expansión pandémica. La posibilidad de que la preadaptación de sitios clave en la secuencia genómica del virus, como el punto de corte de la furina o la estructura del dominio de unión al principal receptor en las células humanas, la enzima convertidora de la angiotensina 2 (ACE-2), se haya producido en esa supuesta especie intermediaria o ya en la propia especie humana queda pendiente de comprobar, al menos por el momento. Incluso podemos considerar la posibilidad de que se haya producido en un hospedador natural y que

la recombinación diese como resultado un virus ya adaptado para infectar y transmitirse en nuestra especie (Khadka *et al.*, 2021).

En los escasos dos años transcurridos desde el inicio de la pandemia hemos tenido numerosas oportunidades para observar la acción de la evolución. En cierto modo, que un virus sea capaz de evolucionar con rapidez no es novedoso, pues cambios similares se habían observado previamente. Sin embargo, en mi opinión, la principal diferencia con esos casos radica en la trascendencia que la evolución del SARS-CoV-2 ha tenido en la opinión pública, en los medios de comunicación, incluso en la adopción de medidas de control de la transmisión de nuevas variantes y mutaciones. La gente, de muy diferentes procedencias, edades o niveles culturales, se pregunta cómo es que la ciencia no puede prever la aparición de estos mutantes y, por tanto, prevenir sus consecuencias. Ya hemos comentado sobre los procesos que conducen a la expansión de estas mutaciones y de algunos linajes. Posiblemente, es mucho más difícil transmitir la idea de que la evolución de un organismo no es predecible salvo en las condiciones controladas del laboratorio y para algunos casos sencillos, y ni siquiera esto se cumple siempre, como se demostró en el LTEE (*Long Term Evolution Experiment*) con *E. coli* de R.E. Lenski (Blount *et al.*, 2012).

¿Qué podemos predecir sobre el futuro evolutivo del SARS-CoV-2? Desgraciadamente, poco, y lo poco que podemos afirmar no permite adoptar medidas de prevención que nos sirvan para paliar los daños potenciales. El virus va a seguir mutando. Mientras se replique, tanto dentro de un paciente infectado como cuando pase a infectar a una nueva persona, las mutaciones seguirán produciéndose. Algunas cambiarán alguna característica del virus, las menos, pero la gran mayoría de esas nuevas mutaciones desaparecerán bien por la acción de la selección natural, bien por casualidad en el cuello de botella que se produce en cada transmisión. Este cuello de botella disminuye tremendamente la efectividad de la selección natural, incluso cuando ésta actúa favoreciendo una nueva mutación, pues los cuellos de deriva reducen el tamaño efectivo poblacional de forma que se necesitan coeficientes de selección muy elevados para que la selección actúe eficazmente (Ohta, 1973). Debemos insistir en que baja probabilidad no implica imposibilidad y que, como hemos visto, ocasionalmente algunas de esas mutaciones, especialmente las que aportan alguna ventaja al virus, aumentarán su frecuencia en la población, arrastrando con ellas otras mutaciones, posiblemente neutrales, presentes en el genoma en que aparecieron y con el que tuvieron éxito evolutivo. ¿Qué tipo de ventaja pueden aportar al virus? Pues depende del entorno en que aparezcan. En este caso, el entorno es la combinación de inmunidad de la población y comportamiento individual y social. En Biología evolutiva es conocido que la mayor eficiencia de la selección se produce hasta el momento en que el organismo llega a reproducirse. En el caso de los virus, podemos asimilar la reproducción a la transmisión a un nuevo hospedador. Por eso, hasta el momento y con seguridad en el futuro, aquellas mutaciones y variantes que faciliten la transmisión son las favorecidas por la selección natural. Se pueden comparar las dinámicas de las variantes preocupantes (VOCs) alfa

y delta, caracterizadas por mayor transmisibilidad, con las de beta y gamma, que recibieron ese calificativo por la posibilidad de que escapasen a la acción neutralizante de algunos anticuerpos (**Fig. 4**).

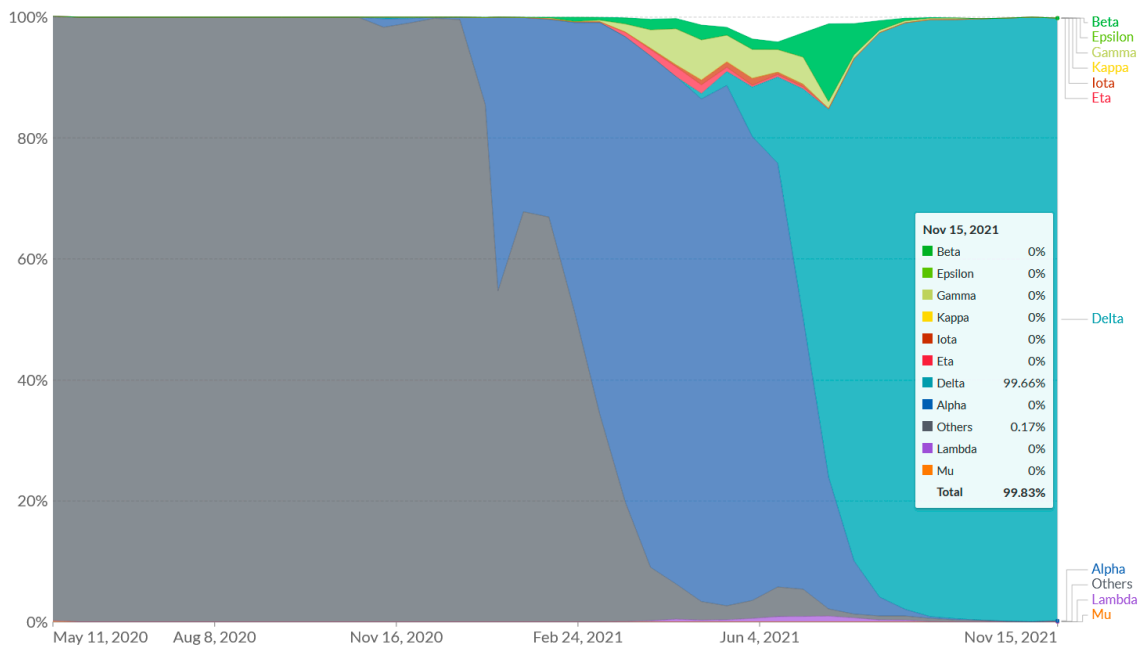


Figura 4. Dinámica de las frecuencias relativas de las VOCs y VOIs entre las secuencias de SARS-CoV-2 analizadas en España y depositadas en GISAID (<https://www.gisaid.org>) y analizadas por Covariants.org.

Mientras haya una parte sustancial de la población no inmunizada, la mayor parte de las transmisiones se producirán sobre esa fracción de la población. Debemos aclarar que la inmunidad que previene la transmisión no es la que confieren las vacunas actuales, sino la que producirán vacunas administradas nasalmente y que actúan sobre las mucosas nasales que representan el primer punto de la infección en las personas. La inmunidad en las mucosas tiene componentes diferentes a la inmunidad humoral, que es la que activan las vacunas inyectadas. Por tanto, a efectos de transmisión, la gran mayoría de la población no está inmunizada, aunque el nivel de protección frente a la infección y la subsiguiente transmisión es mayor en personas vacunadas que en las que no lo han sido. Las vacunas que hemos recibido este último año actúan protegiéndonos de infecciones graves, que conducen a hospitalizaciones, ingresos en UCI y desenlaces fatales. Por tanto, es posible que sigamos observando la aparición de alguna mutación con mejor transmisibilidad que las variantes actuales. Sin embargo, no podemos predecir ni cuál, ni dónde, ni cuándo, ni siquiera con plena certeza si llegará a aparecer.

¿Y si conseguimos una inmunidad que nos proteja de la infección? Pues lo que hemos comentado en el párrafo anterior, también aplica en esta situación. Cabría la posibilidad, remota pero no nula, de que aparezcan mutaciones que eludan esa inmunidad protectora frente a la infección y

también frente a la gravedad de la infección. Retornaríamos entonces a una situación similar a la del inicio de la pandemia, con gran parte de la población sin inmunidad efectiva. Solución: volver a vacunar usando cepas parecidas a las circulantes y que restauren las inmunidades perdidas. Pero esta estrategia es la que ya usamos con la vacunación anual contra la gripe, aunque el nivel al que deberíamos alcanzarla en la población es sustancialmente mayor que frente a ésta. Por ello, la mejor estrategia en estos momentos pasa por disminuir al máximo el tamaño de la población viral, disminuyendo la producción de esas mutaciones. Esto sólo se puede lograr con una vacunación universal, en todos los países y no sólo en los privilegiados, pues no importa dónde se produce una mutación ya que su capacidad de alcanzar cualquier rincón del globo es, como hemos podido comprobar, casi total.

El reciente anuncio de la disponibilidad de dos antivirales específicos para el tratamiento de la COVID-19, con alta eficacia para evitar gravedad de la misma si se administran en los días inmediatos al diagnóstico, plantea otra posible vía de aparición de mutaciones indeseables en el genoma del virus. Estos medicamentos son de distinta naturaleza y modo de acción: Molnupiravir, de la farmacéutica Merck (<https://www.merck.com>), es un agente mutágeno específico para el ARN, mientras que PAXLOVID, de Pfizer, actúa como inhibidor de la proteasa viral (<https://www.pfizer.com>). La historia nos demuestra que los virus de ARN tienen una gran capacidad para encontrar mutaciones que disminuyan o eliminen la actividad de diferentes fármacos antivirales y el SARS-CoV-2 no es previsible que represente una excepción. No obstante, debemos considerar no sólo la posibilidad de que aparezcan esas mutaciones, sino de que se extiendan en la población y representen una amenaza para la eficacia de esos medicamentos en su aplicación a la población necesitada. Una administración adecuada y controlada ayudaría a limitar esa indeseable situación.

Por tanto, una vigilancia genómica sobre la variación genética del virus y unas medidas adecuadas para disminuir el riesgo de aparición de esas mutaciones, a través de la extensión universal de la vacunación y la administración juiciosa de los fármacos antivirales, parecen medidas adecuadas para lograr el objetivo de reducir a niveles mínimos la circulación de virus capaces de infectar y, caso de hacerlo, conducir a gravedad en las infecciones producidas. El resto de medidas necesarias para ello está en nuestras manos.

Bibliografía

- Blount, Z.D., Barrick, J.E., Davidson, C.J. y Lenski, R.E. 2012. Genomic analysis of a key innovation in an experimental *Escherichia coli* population. *Nature*, 489:513–518.
- Boni, M.F., Lemey, P., Jiang, X., Lam, T.T.-Y., Perry, B.W. *et al.* 2020. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nat Microbiol*, 5:1408–1417.
- Delaune, D., Hull, V., Karlsson, E., Hassanin, A., Ou, T.P. *et al.* 2021. A novel SARS-CoV-2 related coronavirus in bats from Cambodia. *Nat Commun*, 12:1–7.

- Duchêne, S., Featherstone, L., Haritopoulou-Sinanidou, M., Rambaut, A. *et al.* 2020. Temporal signal and the phylodynamic threshold of SARS-CoV-2. *Virus Evol.* 6:veaa061.
- Gutierrez, B., Castelan, H., da Silva, C., Jackson, B., Fleishon, S. *et al.* 2021. Emergence and widespread circulation of a recombinant SARS-CoV-2 lineage in North America. *medRxiv*, 2021.11.19.21266601.
- Hodcroft, E.B., Zuber, M., ..., González-Candelas, F., Stadler, T., Neher, R.A., 2020. Spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *Nature*, 595:707–712.
- Jackson, B., Boni, M.F., Bull, M.J., Colleran, A., Colquhoun, R.M. *et al.* 2021. Generation and transmission of inter-lineage recombinants in the SARS-CoV-2 pandemic. *Cell*, 184:5179–5188.
- Khadka, B. y Gupta, R. 2021. Conserved molecular signatures in the spike protein provide evidence indicating the origin of SARS-CoV-2 and a Pangolin-CoV (MP789) by recombination(s) between specific lineages of Sarbecoviruses. *PeerJ*, 9:e12434.
- Korber, B. Fischer, W.M., Gnanakaran, S., Yoon, H., Theiler, J. *et al.* 2020. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell*, 182:812–827.
- López, M., Chiner-Oms, A., García de Viedma, D., Ruiz-Rodríguez, P., Bracho, M.A., ..., González-Candelas, F., Comas, I. 2021. The first wave of the COVID-19 epidemic in Spain was associated with early introductions and fast spread of a dominating genetic variant. *Nat Genet*, 153:1405–1414.
- Ohta, T. 1973. Slightly deleterious mutant substitutions in evolution. *Nature*, 246:96–98.
- Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W. *et al.* 2020. A new coronavirus associated with human respiratory disease in China. *Nature*, 579:265–269.