

Dinámica poblacional de virus y su papel en la pandemia COVID-19

Esteban Domingo
edomingo@cbm.csic.es

Académico Numerario de la RAC
Centro de Biología Molecular
Severo Ochoa (CSIC-UAM)

Celia Perales
cperales@cbm.csic.es

Instituto de Investigación
Sanitaria Fundación Jiménez Díaz
Centro Nacional de Biotecnología
(CSIC)

Palabras clave: virus RNA, mutación, cuasiespecies.



RESUMEN:

Los virus son elementos genéticos ubicuos e infecciosos caracterizados por una gran diversidad estructural y funcional. Muestran dos de las propiedades que caracterizan la vida: replicación y capacidad de evolución, y ambas quedan condicionadas a expresar su programa genético y completar su ciclo infectivo dentro de una célula hospedadora. Fuera de la célula, un virus es un agregado macromolecular complejo pero inerte. La variación genética de los virus, en forma de mutación y recombinación, es inherente a su mecanismo de replicación y, además, constituye el motor para su adaptación a nuevos ambientes. En este artículo describimos las bases moleculares y los patrones temporales de la evolución de los virus, que también son aplicables al nuevo coronavirus SARS-CoV-2 y a la pandemia de COVID-19.

ABSTRACT:

Viruses are ubiquitous and infectious genetic elements that display great structural and functional diversity. They exhibit two of the properties of life: replication and evolution, both contingent upon the expression of their genetic program and completion of their infectious cycle inside a host cell. Outside the cell, viruses are complex, albeit inert, macromolecular aggregates. Genetic variation of viruses, through mutation and recombination is inherent to their mechanism of replication, and it is also the driver of their adaptation to new environments. In the present article we describe the molecular bases and the temporal patterns of virus evolution, that also apply to SARS-CoV-2, and its associated COVID-19 pandemic.



1. INTRODUCCIÓN. VARIOS NIVELES DE DIVERSIDAD BIOLÓGICA

Vivimos inmersos en una biosfera que, aun bajo la amenaza del cambio climático que acelera la extinción de especies y disminuye la diversidad dentro de cada especie, percibimos como extraordinariamente diversa. Ello ha sido la consecuencia de aplicar nuevos métodos de muestreo general a nivel genómico, lo que constituye la nueva disciplina denominada meta-genómica, consistente en el análisis de los distintos tipos de genomas identificados en una misma muestra biológica o medioambiental.

La introducción del concepto de biodiversidad, inicialmente referida principalmente al número de especies que comparten un cierto espacio medioambiental, se remonta a mediados del siglo pasado, con contribuciones muy importantes de R. Margalef, E. O. Wilson y T. Lovejoy, entre otros. Varios estudios destacaron la biodiversidad como característica compartida por todas las formas de vida. Así, por ejemplo, el índice de diversidad de Margalef permitió estimaciones comparativas a partir del contejo de número de individuos de las distintas especies presentes en una muestra. Estos importantes desarrollos tempranos en el campo de la ecología ejercieron una notable influencia en biología general. Con particular relevancia para el presente artículo, se sitúa su impacto en el campo de la virología, como herramienta para cuantificar resultados de análisis de genomas mediante métodos de secuenciación masiva (que aquí abreviaremos con UDS, de "ultra deep sequencing"). Sirva como ejemplo la introducción por J. Gregori, como parte de los estudios realizados en el Hospital Vall d'Hebrón de Barcelona con el virus de la hepatitis C, de índices de diversidad empleados en ecología (números de Hill, índices de Simpson y de Gini-Simpson, entropías de Rao y de Shannon, etc.) (Gregori *et al.*, 2016) para cuantificar la diversidad de las poblaciones de virus que colonizan un individuo infectado.

En las investigaciones sobre diversidad de virus que nos ocupan en el presente artículo, nos referimos no tanto a la coexistencia de distintas especies de virus en una muestra dada ("especie" en el sentido de una cercanía a nivel de genoma más una serie de características compartidas, tal como propone el Comité Internacional sobre Taxonomía de Virus o ICTV, de "International Committee on Taxonomy of Viruses") sino fundamentalmente a las variaciones entre los genomas individuales de un mismo virus. El concepto paralelo en el mundo celular sería la variación entre individuos de una misma especie, ya sea animal, vegetal o microbiana o incluso variaciones entre células que pueblan un mismo tejido, crecientemente documentadas mediante análisis de células individuales ("single cell biology"). Cada vez conocemos más diferencias genéticas, epigenéticas y funcionales entre las entidades individuales que constituye un colectivo biológico. Ello está significando una notable

ruptura conceptual respecto a lo vigente hasta mediados del siglo XX, cuando la mutación genética se consideraba un evento infrecuente y con escasa influencia en procesos evolutivos contemporáneos.

Los virus no han sido una excepción a este cambio de paradigma. Un ejemplo ilustrativo lo ofrece la expectativa que se creó en virología clínica con la incorporación de las técnicas de ingeniería genética (también denominadas de ADN recombinante) al clonaje y expresión de genes víricos. Durante la década de los años 1970 y 1980 se consideró que la administración como vacuna de proteínas víricas desencadenantes de una respuesta inmune, obtenidas mediante ingeniería genética, sería una solución general para la prevención de enfermedades víricas. Para el caso del virus de la gripe, la principal proteína inmunogénica (inductora de anticuerpos protectores) es la hemaglutinina, una glicoproteína insertada en la superficie de las partículas de virus, de la cual se han descrito 18 formas principales distintas. Según esa visión, un repositorio con genes codificantes de esas hemaglutininas permitiría que en cada temporada de gripe se eligiera el gen clonado correspondiente al virus circulante y se expresara para formular la vacuna efectiva. No solo para este caso, sino que también para otros virus patógenos, las empresas biotecnológicas del momento se lanzaron a clonar y expresar aquellas proteínas víricas que pudieran ofrecer vacunas "a la carta". Aunque todo ello aparezca como extraordinariamente ingenuo en el año 2022, todavía **persiste una tendencia a tratar de combatir la complejidad del mundo vírico (del mundo microbiano en general) con la simplicidad ofrecida por las técnicas moleculares**. Precisamente la diversidad existente dentro de una misma población de virus es la principal dificultad para un control efectivo de las enfermedades víricas (Domingo y Perales, 2019a).

2. LOS VIRUS, PARADIGMA DE DIVERSIDAD ESTRUCTURAL Y FUNCIONAL

Los virus, considerados como el conjunto biológico que denominamos "virosfera" son estructuralmente y funcionalmente diversos (**Figura 1**). Las partículas víricas transmisibles, denominadas "viriones" tienen un tamaño variable que, según el grupo de virus, es entre 100 y 5.000 veces menor que el de una célula de mamífero típica. El material genético de los virus consiste en una o varias moléculas de ARN o de ADN, linear o circular, de banda simple o de banda doble. Los genomas víricos de ARN caracterizados hasta el momento, tienen una longitud máxima de unos 30.000 nucleótidos, mientras que los ADN genómicos de los virus pueden alcanzar el millón de pares de bases. Según el tipo de ácido nucleico (ARN o ADN) que participa en el circuito replicativo (no en la expresión de genes que siempre es ARN) se distinguen

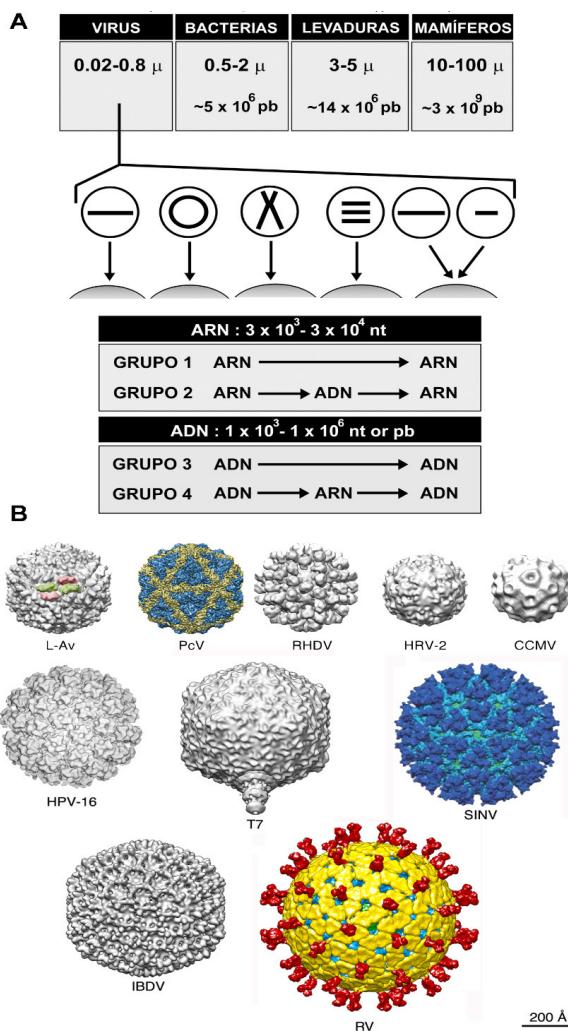


Figura 1. Características de las partículas víricas y de su material genético. (A) Arriba: tamaño comparativo de virus y células (en micras) y de su material genético (en pares de bases, pb). Formas del material genético de los virus, representadas mediante trazos gruesos dentro de una partícula simplificada mediante un círculo; las flechas apuntan a una célula a infectar. Los cuadros centrales indican la longitud de genomas víricos (en nucleótidos, nt o pares de bases, pb) y una clasificación basada en el tipo de ácido nucleico que participa en su circuito replicativo. (B) Ejemplos de partículas víricas esféricas reconstruidas a partir de datos de crio-microscopía electrónica y de difracción de rayos X por cristales de virus. L-Av, virus de *Saccharomyces cerevisiae*; PcV, virus de *Penicillium chrysogenum*; RHDV, virus de la enfermedad hemorrágica del conejo; HRV-2, rinovirus humano tipo 2; CCMV, virus del moteado clorótico del caupí; HPV-16, virus papiloma tipo 16; T7, cabeza del bacteriófago T7; SINV, virus Sindbis; IBDV, virus de la bursitis infecciosa; RV, rotavirus. Figura compuesta a partir de otras publicadas en Domingo, E. Virus as Populations, Elsevier, 2020, con permiso.

cuatro grupos de virus. Grupo 1: aquellos en los que interviene solo ARN, que resultan ser los más abundantes (ejemplos: virus de la gripe, resfriado común, sarampión, hepatitis C o coronavirus). Grupo 2: aquellos con ARN pero con un ADN intermedio (ejemplo: los retrovirus como el virus del SIDA). Grupo 3: aquellos en los que interviene solo ADN (ejemplos: virus herpes, viruela o peste porcina africana). Grupo 4: aquellos con ARN pero con un ADN intermedio (ejemplo: el virus de la hepatitis B). En relación con conceptos de variabilidad genética que tratamos en apartados posteriores, cabe adelantar que **aquellos grupos de virus en los que interviene ARN**

en cualquier punto de su ciclo replicativo, son los más variables. Les asignamos la propiedad de replicarse con propensión a cometer errores (introducción de mutaciones) (**Figura 1A**) (Domingo, 2020).

Las partículas de virus pueden ser desnudas (formadas por proteínas ensambladas con simetría icosaédrica o helicoidal) o incluir una envuelta lipídica en la que se insertan glicoproteínas protuberantes (**Figura 1B**). Los dominios expuestos de las proteínas de la superficie contienen determinantes antigenicos (denominados epítopos B) que desencadenan la producción de anticuerpos protectores en el organismo infectado o vacunado. Cuando por mutación del material genético se producen cambios de aminoácido en los sitios antigenicos de los virus circulantes, éstos pueden escapar a la acción neutralizante (inactivante) de los anticuerpos previamente generados por infección o vacunación. Tanto las proteínas presentes en los viriones (denominadas "proteínas estructurales") como aquellas que participan en su ciclo infeccioso intracelular (denominadas "proteínas no estructurales") pueden contener otro tipo de determinantes antigenicos denominados epítopos T. Esta clase de epítopos es la responsable de desencadenar en el hospedador la respuesta inmune celular (que incluye las denominadas "células citotóxicas" porque pueden reconocer y eliminar a células infectadas por virus). La combinación de la respuesta de anticuerpos y la respuesta inmune celular suele conferir a los organismos protección frente a reinfecciones por el mismo virus.

La estabilidad antigenica –término con el que se describe que una respuesta inmune inducida por un virus siga siendo activa frente a variantes genéticas del mismo virus– es extraordinariamente dispar entre virus. Las bases moleculares de tal disparidad solo se conocen parcialmente, pero se dan ejemplos particularmente espectaculares. Así, se ha identificado un solo serotipo del virus de la hepatitis A humana mientras que se han descrito 160 serotipos de rinovirus humanos [virus responsable del resfriado común (uno de ellos mostrado en la **Figura 1B**)], aun ambos formando parte de la misma familia denominada *Picornaviridae* (de "virus pequeño"). Pertenencia a una misma familia representa que poseen una dotación genética similar (número y función biológica de los genes que codifican) y también una estructura general comparable de sus partículas. Ello no implica ninguna semejanza respecto a las enfermedades que producen (Domingo, 2020).

Los virus como colectivo biológico pueden infectar a una gran variedad de tipos celulares y organismos unicelulares y multicelulares, incluyendo bacterias, animales y plantas. Se han encontrado en todo tipo de ambientes, como ha sido establecido mediante varias prospecciones metagenómicas. El número total de partículas víricas en la



biosfera se ha estimado en 10^{31} a 10^{32} , valor que excede en diez veces al número total de células. Según algunos estudios, la vida celular que conocemos y de la que formamos parte fue construida gracias al transporte de material genético promovido por elementos genéticos móviles, entre los cuales se hallaban los ancestros de los virus actuales. No es sorprendente que los virus, como albañiles de la edificación del mundo celular durante unos 2.000 millones de años, hayan persistido hasta nuestros días (Domingo y Perales, 2019b). Las enfermedades que producen son muy probablemente un efecto colateral de la función troncal para la que fueron seleccionados: la formación y diversificación de las células. Ello se hace muy patente con los virus de plantas, la mayoría de los cuales no producen enfermedad y en algunos casos proporcionan rasgos beneficiosos (por ejemplo, resistencia a factores de estrés) a las plantas que los hospedan.

3. LA VARIABILIDAD DE LOS VIRUS Y SU IMPACTO PARA EL CONTROL DE LAS ENFERMEDADES

Según algunas teorías del origen de virus y células en nuestro planeta, ya desde tiempos remotos debió establecerse lo que recientemente se ha llamado una "carrera armamentística" (aunque sería más apropiado denominarle "influencias selectivas mutuas") entre las células para superar perturbaciones producidas por virus y éstos para sobreponerse a barreras impuestas por las células. La respuesta inmune y los mecanismos moleculares de escape por parte de los virus, a lo que hemos aludido en el apartado anterior, serían herencia de interacciones ancestrales entre virus y células. Si así fuera, la capacidad de variación de los virus, que estamos descubriendo actualmente, sería un rasgo consustancial con su propia existencia. De hecho, varias líneas de investigación apoyan el concepto de que los errores de replicación ya se daban en el mundo RNA primitivo, libre de proteínas. Los errores probablemente tuvieron un papel activo en el origen de la vida, entendido como la aparición y selección de las primeras formas replicantes, a partir de las cuales se organizaron formas más complejas (Eigen, 1993, 2013; Eigen y Schuster, 1979). Según este punto de vista, somos la herencia de un cúmulo de errores.

Sea cual fuere el origen evolutivo de la variación genética de los virus, este rasgo se ha mostrado fundamental para su comportamiento y supervivencia, así como para su potencial de causar enfermedad. Ello se evidenció a lo largo del siglo XX con numerosos estudios tanto con sistemas modelo simplificados de infección por virus no patogénicos (evolución experimental en cultivos celulares), como con virus de gran importancia en salud pública, en su doble faceta clínica y de laboratorio. Ilustraremos el impacto de la variación de virus con tres

ejemplos de grandísima relevancia histórica: (i) poliovirus, el agente causante de lo que a mitad del siglo XX fue la tan temida poliomielitis ("parálisis infantil"); (ii) el virus de la gripe humana tipo A, el responsable habitual de la gripe que todos hemos padecido; y (iii) el virus de la inmunodeficiencia humana tipo 1 (VIH-1), causante del SIDA, o síndrome de la inmunodeficiencia adquirida, la gran pandemia entre los siglos XX y XXI.

En el caso de poliovirus, la constatación más dramática fue que por efecto de mutación y recombinación (formación de genomas mosaico a partir de dos genomas parentales distintos), la vacuna viva atenuada de Sabin, que se empezó a administrar a mediados del siglo XX, podía modificarse durante su multiplicación en el individuo vacunado, dando lugar a formas virulentas que causaban poliomielitis. Ello se dio en aproximadamente uno de cada 500.000 vacunados o en sus contactos. Quizá fue una de las primeras observaciones que alimentaron el nacimiento de movimientos "anti-vacunas", sin reparar en los imponentes beneficios de la vacunación frente a la no-vacunación.

En el caso del virus de la gripe, la situación es todavía más compleja ya que su dotación genética está formada por ocho segmentos distintos de ARN, cada uno de los cuales puede ser sustituido por otro proveniente de otra cepa del virus, incluidos virus de origen animal (como, por ejemplo, aves). Este proceso de intercambio se denomina reordenamiento génico y es el responsable de la introducción de genes que codifican proteínas antigenicas como son la hemaglutinina (de la que hemos hablado anteriormente) y la neuraminidasa, provenientes de virus de animales en virus humanos. Ello se traduce en **"saltos" antigenicos**, una de cuyas consecuencias es la pérdida de eficacia de las vacunas en uso. Asimismo, cambios de aminoácido debido a mutaciones puntuales (que afectan a un nucleótido, sin necesidad de recombinación o reordenamientos) son los responsables de la llamada **"deriva antigenica"** que también puede alterar gradualmente el perfil antigenico de los virus circulantes. Esta alteración antigenica subyace a una eficacia limitada de las vacunas de virus inactivado que se emplean actualmente para prevenir la gripe y también a la necesidad de actualización periódica de la vacuna.

El control del SIDA representó un desafío dramático ya que pronto se observó que la variación genética del VIH-1 en los individuos infectados permitía su escape continuado a la respuesta inmune ya de por sí debilitada. Ello es una de las razones por las que no se ha desarrollado una vacuna efectiva para prevenir el SIDA a pesar de ingentes inversiones en talento y dinero. La variación de este virus, tanto por mutación como por recombinación, ha tenido otro tipo de consecuencias funestas. Las mutaciones en genes diana de agentes antirretrovirales empleados en

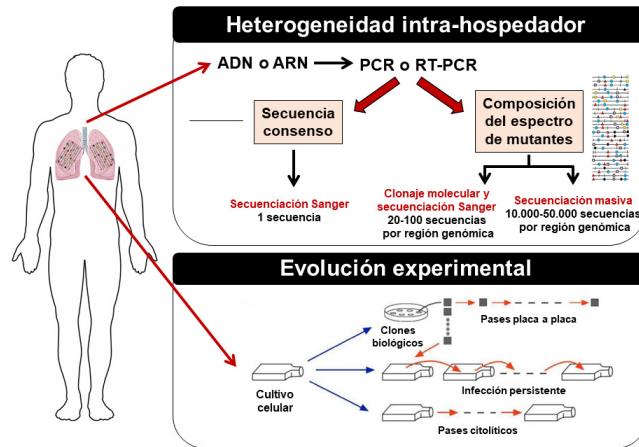


Figura 2. Metodología para el estudio de la diversidad y evolución de virus. Muestras de ácido nucleico viral de un individuo infectado se amplifican para determinar la secuencia consenso o la composición del espectro de mutantes (cuadro superior; a la derecha se dibuja un alineamiento de genomas con mutaciones representadas mediante símbolos). Los virus que crecen en cultivos celulares pueden someterse a diseños de evolución experimental para estudiar mecanismos de adaptación y evolución (cuadro inferior).

monoterapia (administración de un solo agente antiviral) fueron responsables de fallos terapéuticos y del fallecimiento de millones de personas infectadas en la década de los 1980's y comienzos de los 1990's. El problema se redujo muy considerablemente cuando se implementaron las terapias altamente agresivas, denominadas "TARGA" (de "tratamiento antirretroviral de gran actividad"), consistentes en la administración de tres o más agentes antirretrovirales dirigidos a dianas distintas del virus, generalmente a la polimerasa, proteasa e integrasa, todas ellas enzimas esenciales para el virus. Ello redujo espectacularmente la mortalidad y convirtió la infección en crónica y controlable. La razón de los beneficios del TARGA se sustenta en que la probabilidad de que la misma molécula genómica del VIH-1 mute en tres (o más) sitios distintos, que sería lo requerido para resistir a los tres agentes administrados, es muy baja. De hecho, dada la tasa de error del virus, el número de partículas víricas necesario para que esté representado un genoma con las tres mutaciones necesarias, excede en órdenes de magnitud al tamaño poblacional del virus en un individuo infectado. **Las terapias de combinación son las empleadas comúnmente en la actualidad, no solamente para el SIDA sino también para otras enfermedades crónicas como es el caso de la hepatitis C, con éxitos muy destacables. Tales terapias reducen notablemente la selección de mutantes de escape a los agentes administrados y constituyen uno de los grandes éxitos de la farmacología antivírica actual** (Perales *et al.*, 2017).

Estos tres ejemplos ilustran algunas de las implicaciones de la variación de los virus para adaptarse a nuevos ambientes y superar barreras impuestas por la práctica médica. Cabe destacar que los tres mecanismos moleculares implicados (mutación, recombinación y reordenamientos genómicos) son compatibles entre sí.

Obviamente los reordenamientos quedan restringidos al caso de virus con genoma segmentado, particularmente abundantes entre los virus de plantas. En cambio, las mutaciones puntuales (transiciones y transversiones) son una fuente universal de variación genética, y no se conocen sistemas biológicos en los que no se produzcan. La recombinación fue bien establecida a lo largo del siglo pasado para ADN celular y vírico, pero no así para virus ARN, incluso con dudas sobre su existencia. Actualmente la recombinación en virus ARN se percibe como muy extendida, aunque se acepta que su frecuencia e impacto biológico varía según la familia de virus. Curiosamente, varios de los virus con gran impacto en salud pública, notablemente picornavirus, coronavírus y retrovirus, muestran elevadas frecuencias de recombinación, probablemente estrechamente ligadas a su mecanismo de replicación (Domingo, 2020).

4. SECUENCIAS CONSENSO Y VELOCIDADES DE EVOLUCIÓN

Dos grupos principales de métodos se emplean actualmente para cuantificar la diversidad de los virus y explorar sus mecanismos de variación a nivel virológico: secuenciación de nucleótidos y diseños de evolución experimental (Figura 2). La determinación de la secuencia de los nucleótidos del material genético de los virus se cimenta en la posibilidad de amplificación de pequeñas cantidades de ácido nucleico mediante PCR (de "polymerase chain reaction" o reacción en cadena de polimerasas termoestables) para ADN y de RT-PCR para ARN (RT indica la copia del ARN en ADN mediante la enzima retrotranscriptasa). Los métodos de secuenciación clásicos, conocidos como "secuenciación Sanger" (como homenaje al gran bioquímico F. Sanger, merecedor dos veces del Premio Nobel de Química, uno por la determinación de



secuencias de aminoácido en proteínas en 1958 y otro por la secuenciación de nucleótidos en 1980) dan información sobre la secuencia denominada "consenso". Se trata de un promedio de todas las secuencias presentes en la muestra analizada, de manera que a cada posición del genoma le corresponde el nucleótido más abundante en esa posición. La secuencia consenso es la que habitualmente sirve para comparar distintos aislados de un mismo virus o virus distintos. Revela información importante para la identificación y clasificación de los virus por parte de la ICTV. La secuencia consenso permite derivar árboles filogenéticos que definen relaciones de parentesco entre los genomas comparados. De esta manera sabemos, por ejemplo, que el coronavirus SARS-CoV-2 (SARS responde a "síndrome respiratorio agudo y severo") se relaciona con otros coronavirus humanos que causan patologías leves y que difiere muchísimo de los virus de la gripe humana, a pesar de tratarse todos ellos de virus respiratorios.

La comparación de secuencias consenso es también la principal herramienta para medir la velocidad de evolución de los virus durante su expansión en la naturaleza, particularmente durante brotes de enfermedad, epidemias o pandemias (epidemias de alcance mundial). La velocidad de evolución se expresa mediante la proporción de sitios genómicos que varían por unidad de tiempo. Una conclusión significativa de estos estudios ha sido que los virus ARN acumulan mutaciones a velocidades que son casi un millón de veces mayores que las estimadas para la evolución de los organismos diferenciados a los que infectan (Holland *et al.*, 1982). Uno de cada diez mil nucleótidos del material genético de esos virus cambia cada año. Una comparación ilustrativa la proporcionó un estudio de epidemiología molecular del virus de la fiebre aftosa realizado hace cuatro décadas en nuestro laboratorio. Este virus es otro miembro de la familia *Picornaviridae* que produce la glosopeda del ganado, con un gran impacto económico a nivel mundial. En el estudio se observaron cuatro diferencias de aminoácido al comparar la secuencia de una proteína de la partícula del virus de dos aislados separados por un año de evolución, en el transcurso de un brote de enfermedad ocurrido en Catalunya en la década de los años 1970. Pues bien, una diferencia de tres aminoácidos es la que se da en una porción de longitud equivalente de la proteína citocromo c del caballo y del hombre. Se estima que los dos linajes se separaron en el Cretácico, hace unos 80 millones de años, para iniciar la línea evolutiva de los hipomorfos (équidos) y de los primates (simios y humanos). Este y otros ejemplos ilustran que los virus pueden alcanzar velocidades de evolución del orden de millones de veces mayores que las estimadas para los organismos diferenciados. También el SARS-CoV-2 evoluciona a grandes velocidades que, al menos de momento, son comparables a las calculadas para el resto de virus ARN.

Los diseños de evolución experimental (**Figura 2**) son los que permiten controlar variables [temperatura, tipo de célula, estirpe de virus, infección persistente (sin muerte celular) o citolítica (con muerte celular), presencia o ausencia de drogas o anticuerpos, etc.] y examinar cómo evoluciona el virus frente ante esas diferencias. Mucho de lo que hemos avanzado en el conocimiento de los virus a nivel poblacional se basa en diseños como los esbozados en la **Figura 2**.

5. ESPECTROS DE MUTANTES E INSUFICIENCIA DE LA SECUENCIA CONSENSO

¿Cuál es el origen de las diferencias en velocidad de evolución entre la virosfera de ARN y el mundo de organismos ADN? Para abordarlo debemos diseccionar a nivel poblacional cómo son realmente los virus dentro de las células de los organismos infectados. Al hacerlo empleando UDS, observamos que **el virus no replica con una secuencia definida, sino que incluye una distribución de genomas mutados que se denominan espectros o nubes de mutantes. Multitud de nubes coexisten en distintos compartimentos (células, tejidos, órganos) de un individuo infectado** (**Figura 3**) (Domingo y Perales, 2019a; Domingo, 2020).

A ello le denominamos **estructura en cuasiespecies de los virus ARN** (columna de genomas dibujada a la derecha de la **Figura 3**), cuyo origen como concepto explicamos en el apartado siguiente. Esta constatación de extremada heterogeneidad introduce de inmediato la cuestión de la representatividad de la secuencia consenso (línea horizontal al final de la columna de genomas de la **Figura 3**) que nos ha ofrecido clásicamente la visión de cómo es un virus a nivel molecular y que todavía hoy llena las bases de datos sobre genomas víricos y los libros de texto de virología. Ahora nos damos cuenta de que la secuencia consenso es no solamente una simplificación de una realidad mucho más compleja, sino una abstracción, dado que con altísima probabilidad esa secuencia no estará representada entre los genomas individuales que componen la población (compruébese este punto en la **Figura 3**). Como escribió M. Eigen, el valor de la secuencia consenso de un conjunto de genomas es equivalente al que tiene un número de teléfono que se obtiene situando en cada una de las posiciones (nueve para los números de teléfono de Madrid) el número más frecuente en esa posición según la guía telefónica. El "consenso" de números de teléfono puede no coincidir con ninguno de los teléfonos individuales (Eigen, 2013). El 91 inicial (por repetirse en todos los números) nos permite saber que nos hallamos ante un(a) ciudadano(a) de Madrid, pero no identificamos ni un nombre, ni una dirección postal, y por tanto no tenemos acceso a conocer ni su modo de ser, ni sus aficiones. Algo parecido ocurre con los virus:



una secuencia consenso permite reconocer el grupo al que pertenece, pero no varias de sus propiedades biológicas relevantes.

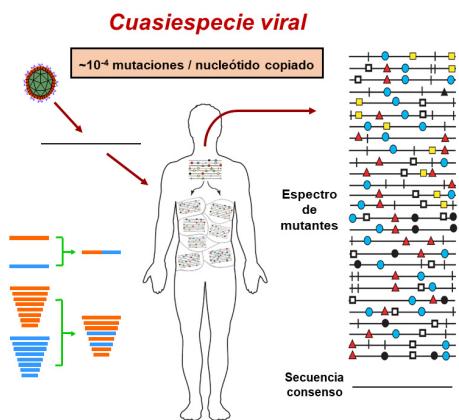


Figura 3. Formación de cuasiespecies víricas y mecanismos de evolución. La infección por un solo genoma vírico da lugar a espectros de mutantes debido a replicación con generación de copias erróneas (primera ecuación de la Figura 4). Las líneas horizontales representan genomas y los símbolos sobre las líneas indican mutaciones. Los espectros de mutantes están compartimentados en el individuo infectado. En la ampliación de la derecha se ilustra que la secuencia consenso puede no existir en la población y ser idéntica al genoma del virus infectante. En la parte inferior izquierda se dibuja esquemáticamente un ejemplo de recombinación (formación de un genoma mosaico) y de reordenamiento en un genoma segmentado.

Cada vez tenemos más evidencia de que ciertas facetas del comportamiento de un virus dependen del conjunto de genomas distintos y no solamente de individuos concretos. Mediante interacciones entre componentes de un mismo espectro de mutantes, puede modularse el comportamiento del conjunto (de la Torre y Holland, 1990; Domingo y Schuster, 2016). Ello, además de acrecentar la insuficiencia de la secuencia consenso para entender el comportamiento de un virus, nos acerca a una de las premisas básicas de la teoría de cuasiespecies: que **un conjunto de genomas actúa como una unidad de selección** (Domingo, 2020). Este punto merece ser ampliado.

6. ENCUENTRO DE LA VIROLOGÍA EXPERIMENTAL CON LA BIOFÍSICA TEÓRICA: CUASIESPECIES

El concepto de cuasiespecies tuvo dos orígenes, uno teórico y otro experimental, ambos acaecidos durante la década de los años 1970 y sin conexión entre los grupos participantes. El origen teórico se debe al trabajo de M. Eigen y P. Schuster en la Universidad de Göttingen (Alemania) cuando formularon matemáticamente un modelo del origen de la vida consistente en replicones sencillos (de ARN o de alguna molécula similar) cuya característica central era que durante su replicación mutaban con alta frecuencia. Ello les permitía auto-organizarse y aumentar su eficacia replicativa mediante la sustitución de unas distribuciones de mutantes por otras. El término "cuasiespecie" no se

refiere a una modificación o limitación del concepto de especie tal como lo entendemos en biología, sino a una cualificación de la especie química que era el replicón primitivo (un conjunto de genomas parecidos salpicados de mutaciones) (Eigen, 1993, 2013; Eigen y Schuster, 1979). Las dos fórmulas básicas de la teoría de cuasiespecies describen la **replicación de genomas con introducción de mutaciones y la limitación de la complejidad de información heredable impuesta por la tasa de error** (Figura 4). Ambas fórmulas tienen varias implicaciones para entender la adaptación de los virus y diseñar nuevas maneras de controlarlos.

Fórmulas básicas de la teoría de cuasiespecies

$$\frac{dx_i}{dt} = (A_i Q_i - D_i)x_i + \sum_{k=1, k \neq i}^n W_{ik} x_k - \Phi_i$$

$$v < v_{\max} = \frac{\ln \sigma_0}{1 - \bar{q}} = \frac{\ln \sigma_0}{\bar{p}} \quad \text{and} \quad \bar{p} < \bar{p}_{\max} = \frac{\ln \sigma_0}{v}$$

Figura 4. Las dos fórmulas básicas de la teoría de cuasiespecies. La primera ecuación indica replicación del genoma i (variación de su frecuencia en función del tiempo) con producción de errores $[x_i, x_k]$, concentración de k ; A_i , replicación de i ; Q_i , precisión de la replicación de i ; D_i , degradación de i ; $W_{ik} x_k$, síntesis de i a partir de k ; Φ_i , flujo de i para salir del conjunto replicativo]. Las segundas ecuaciones definen los parámetros que sitúan el umbral de error [v , cantidad de información (que podemos aproximar al número de nucleótidos del genoma); σ_0 , superioridad de la secuencia maestra (o dominante) respecto al conjunto; \bar{q} , fidelidad de copia promedio; \bar{p} , tasa de error promedio.

Para la formulación de la teoría de cuasiespecies, Eigen y Schuster se basaron en conceptos de teoría de información y en estudios experimentales realizados por el grupo de S. Spiegelman en New York en la década de los 1960, acerca de replicación enzimática en el tubo de ensayo de ARN de un virus bacteriano (bacteriófago) denominado $Q\beta$. Los estudios de Spiegelman revelaron que también en el laboratorio la replicación de moléculas ARN mostraba características darwinianas: mutación y selección (Mills *et al.*, 1967). En cambio, Eigen y Schuster no conocían otra línea de trabajo experimental, también con el bacteriófago $Q\beta$ y que se estaba realizando en la Universidad de Zürich (Suiza), en el laboratorio de C. Weissmann y que representó el origen experimental del concepto de cuasiespecies (Domingo *et al.*, 1978). El grupo de Zürich puso a punto un método de mutagénesis dirigida usando el bacteriófago $Q\beta$. De hecho, el procedimiento tuvo un gran éxito y representó el comienzo de lo que se denomina "genética inversa", consistente en producir a voluntad mutaciones en sitios preseleccionados y analizar sus consecuencias (al revés de lo que se hacía en genética clásica). Estos estudios también permitieron calcular la primera tasa de error para un virus ARN (Batschelet *et al.*, 1976).

Para este desarrollo fue necesario comparar mapas de oligonucleótidos del genoma de numerosas poblaciones

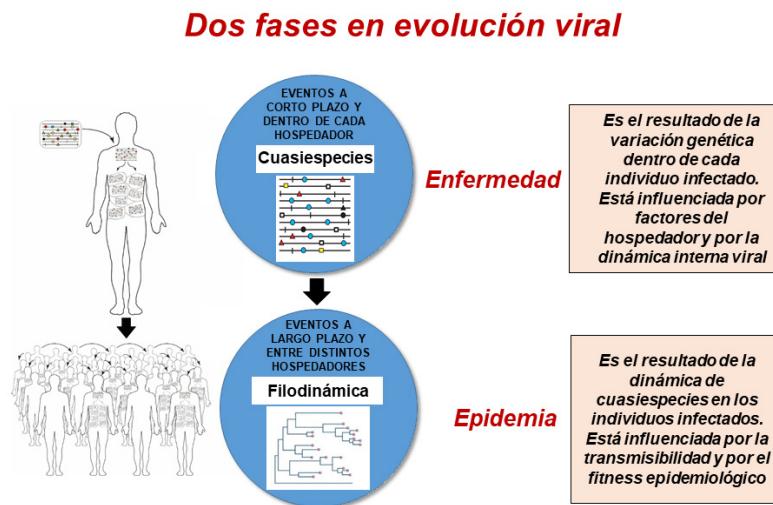


Figura 5. Las dos fases en la evolución de un virus. En la parte superior se indica la primera fase de evolución a corto plazo, o dentro de cada individuo infectado, que está relacionada con mecanismos de enfermedad. En la parte inferior se indica la evolución a largo plazo, relacionada con la expansión epidémica de las enfermedades víricas.

del bacteriófago, dado que todavía no se disponía de métodos eficientes de secuenciación de nucleótidos. Los mapas detectaron mutaciones y la conclusión de varias comparaciones de clones víricos fue que el bacteriófago Q β realmente consistía en una nube de mutantes, como anticipaba la teoría de cuasiespecies. Fue algo sorprendente para la época dado que a la mutación se le atribuía un papel secundario en los procesos evolutivos. Se dice que, al enterarse de los resultados de Zúrich durante un seminario en Suiza en el año 1978, Eigen exclamó: “¡Cuasiespecies en realidad!”. Ello fue el inicio de una fructífera colaboración entre dos disciplinas aparentemente distantes pero que, en su intersección, han modificado nuestra percepción de la dinámica evolutiva de los virus. Si en el año 1980 alguien hubiera presentado un proyecto proponiendo el estudio del origen de la vida para entender y combatir enfermedades víricas le hubiera sido denegado con solemne rotundidad. En cambio, eso es lo que realmente se desarrolló durante los siguientes cuarenta años.

7. DOS ETAPAS EN LA EVOLUCIÓN DE LOS VIRUS

La formación de espectros de mutantes que siguen una dinámica de cuasiespecies (es decir, una continua variación genética en las células infectadas, competición entre las variantes producidas y selección de las más adaptadas en cada microambiente) es relevante para comprender las primeras etapas de una infección. En cambio, etapas posteriores como son la transmisión del virus desde un individuo infectado a otro susceptible y la expansión de un virus a nivel epidemiológico se rigen por otras reglas y procedimientos para su cuantificación

(Holmes, 2009; Geoghegan y Holmes, 2018). Ahora bien, en cada infección individual de las que componen un brote epidémico se producen espectros de mutantes de los cuales surgirá la variante (o conjunto de variantes) que colonizarán individuos susceptibles. Es decir, **en la evolución de los virus se distinguen dos fases. La primera consiste en la formación de espectros de mutantes en cada hospedador infectado y la segunda abarca la expansión del virus entre la población de hospedadores susceptibles** (Figura 5). Ambas fases operan también en el caso de SARS-CoV-2 y su expansión para producir la COVID-19.

La primera fase, cuya palabra clave es “cuasiespecies” tiene mucho que ver con mecanismos de enfermedad. En efecto, la enfermedad vírica puede considerarse en buena medida el resultado de interacciones virus-célula favorecidas por adaptación local del virus a ciertos tejidos y órganos (Rueca *et al.*, 2020). En esta adaptación resulta clave la selección de subpoblaciones de los espectros de mutantes, de ahí su relevancia en la patogenia vírica. Incluso en casos en los que la principal causa de enfermedad grave es una respuesta inmune descontrolada, ésta es resultado de interacción de virus con células de las que depende la modulación de una respuesta inmune (protectora frente a exacerbada y patológica). Además, en algunos casos se ha relacionado la complejidad de los espectros de mutantes con la capacidad de producir enfermedad o de respuesta a tratamientos antivirales, puntos que actualmente se están estudiando con mucho interés.

La palabra clave para la segunda fase de evolución de los virus es “filodinámica” y se relaciona con la expansión epidémica del virus (Geoghegan y Holmes, 2018). En esta fase intervienen parámetros como transmisibilidad, estabilidad del virus durante su fase de estancia fuera

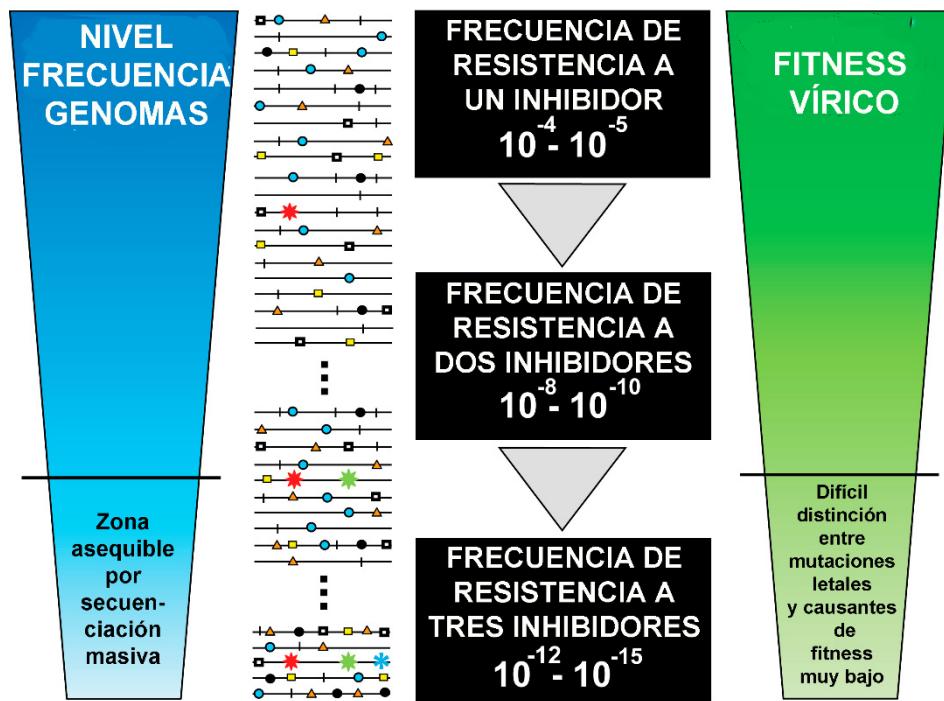


Figura 6. Gradiente de frecuencias de los genomas que pueblan un espectro de mutantes. La frecuencia refleja la capacidad replicativa (fitness) de los distintos genomas que de modo dinámico surgen y compiten en la población. Los virus portadores de dos o más mutaciones necesarias para conferir resistencia a un inhibidor de la multiplicación del virus se hallan a frecuencias decrecientes en la misma población. Figura modificada a partir de una publicada en Domingo, E. Virus as Populations, Elsevier, 2020, con permiso.

del organismo infectado, etc. Se visualiza mediante árboles filogenéticos, con los que se establecen relaciones de parentesco entre las distintas variantes que surgen durante la expansión del virus en la naturaleza. También los espectros de mutantes pueden jugar un papel importante (mediante producción al azar de variantes que resultarán más transmisibles o estables, etc.) pero el éxito epidemiológico depende muchos otros factores. Nosotros propusimos hace años el término **“fitness epidemiológico”** para representar el conjunto de factores que finalmente hacen que un brote de enfermedad desaparezca prácticamente sin intervención humana (como fue la enfermedad respiratoria asociada al coronavirus SARS-CoV-1) o que se propague sostenidamente para producir una de las mayores catástrofes sanitarias que se conocen (como es la COVID-19 causada por el coronavirus SARS-CoV-2) (Domingo, 2020). Dicho de modo conciso: ambas fases evolutivas están relacionadas, pero mientras la primera es muy relevante para entender la enfermedad, la segunda lo es principalmente por sus implicaciones en salud pública.

8. IMPLICACIONES FUNCIONALES Y MÉDICAS DE LAS NUBES DE MUTANTES

De nuevo nos preguntamos: ¿por qué son importantes los espectros de mutantes y resulta insuficiente determinar y trabajar solamente con secuencias consenso? Son varias las razones, que abarcan desde entender el comportamiento

de los virus como poblaciones a consecuencias para el control de enfermedades víricas. Las esbozamos a continuación.

Una nube de mutantes no tiene una distribución homogénea de variantes por varias razones. Primero porque cada nueva variante que surge debido a la introducción de mutaciones que ocurre fundamentalmente al azar (decimos que se trata de un acontecimiento *“estocástico”*) puede modificar la capacidad replicativa de ese genoma en concreto. Sabemos por datos de UDS obtenidos con muchos virus que hay un rango de frecuencias de genomas que es paralelo a una distribución de capacidades replicativas (o fitness). Esta distribución explica también la creciente menor probabilidad de hallar en una población de virus un mutante resistente a uno, dos o tres inhibidores administrados simultáneamente como terapia (**Figura 6**). La creciente capacidad de detección de mutaciones muy minoritarias introduce dudas acerca de si son mutaciones letales o causantes de un gran descenso del fitness del virus.

Por otra parte, cabe insistir en que una nube de mutantes no es un mero agregado de genomas cuya frecuencia queda determinada por un proceso de mutación-selección, sino que, además, se establecen interacciones internas entre sus componentes. Tales interacciones pueden ser positivas (de complementación o cooperación) o negativas (de interferencia). Las interacciones positivas permiten la supervivencia de genomas que, si estuvieran en condiciones de aislamiento, estarían condenados a

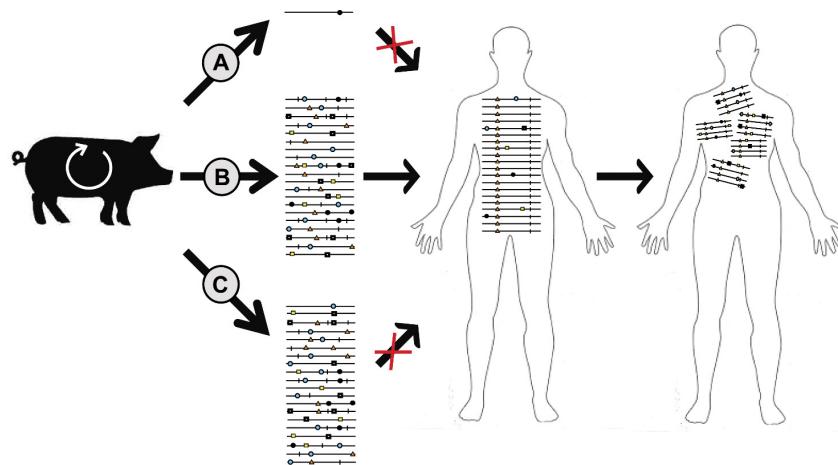


Figura 7. Implicación de espectros de mutantes en el origen zoonótico de enfermedades víricas humanas. Muchas de las enfermedades emergentes humanas tienen un origen zoonótico y se deben a múltiples factores. Uno de ellos puede ser que genomas del espectro de mutantes que se multiplica en un animal sean aptos para establecerse en un humano, iniciando un proceso de adaptación y expansión. Figura modificada a partir de una publicada en Domingo, E., *Virus as Populations*, Elsevier, 2020, con permiso.

la extinción. Las interacciones negativas hacen que un genoma que en condiciones de aislamiento tendría una gran capacidad de multiplicación quede suprimido por el conjunto que le rodea. Este punto fue demostrado por primera vez en 1990 por J.C. de la Torre y J.J. Holland mediante experimentos en los que un clon de un virus denominado virus de la estomatitis vesicular que mostraba gran poder de multiplicación, resultaba inhibido por una nube de mutantes del mismo virus que exhibía mucha menor capacidad replicativa. Asimismo, las vacunas de virus atenuado empleadas para prevenir la poliomielitis contienen una baja proporción de virus virulento que queda suprimido por el conjunto. Solamente cuando la proporción excede de cierto valor, entonces la vacuna causa enfermedad y no es válida para su administración. Las consecuencias derivadas de interacciones dentro de un espectro de mutante deben distinguirse del efecto que pueden tener dos o más mutaciones localizadas en un mismo genoma individual. Tal conjunción de mutaciones puede aumentar o disminuir la capacidad replicativa del genoma individual, lo que se conoce como epistasia positiva o negativa, respectivamente.

Los efectos supresivos de los espectros de mutantes hacen que entre sus componentes haya genomas individuales cuyo comportamiento si se encontraran aislados del resto, diferiría del comportamiento mostrado por el conjunto. Ello convierte a los cuellos de botella poblacionales (reducciones extremas del tamaño poblacional) en importantes actores de la evolución de los virus. Un cuello de botella —como los que a menudo ocurren cuando un virus coloniza un nuevo órgano en un individuo o es transmitido entre individuos— separa a uno o a unos pocos genomas de una población del conjunto que los albergaba. De esta manera, se modifican posibles influencias ejercidas por el conjunto. Las variaciones de tamaño poblacional convierten a las siguientes fases de la evolución en episodios renovados en los que nuevos genomas desarrollarán cuasiespecies

distintas, en una permanente "lotería" de variaciones genéticas y sus derivadas fenotípicas. Ello tiene también consecuencias para la emergencia de virus, por ejemplo, el origen zoonótico de nuevas enfermedades, condicionado al azar de contactos entre especies biológicas y a que una distribución de mutantes contenga algún mutante apto para establecerse en un nuevo organismo (**Figura 7**).

La composición de un espectro de mutantes, incluso en un solo individuo infectado es fluida y cambiante, de manera análoga a la composición en pasajeros de un aeropuerto internacional, que es distinta segundo a segundo y que nunca se repetirá en las miradas de segundos de su historia de funcionamiento como aeropuerto.

9. NUEVAS ESTRATEGIAS ANTIVIRALES. MUTAGÉNESIS LETAL

La consideración de las poblaciones víricas como nubes dinámicas de mutantes nos ha forzado a pensar nuevos métodos para controlar las enfermedades que producen. Algunos ya nos parecen obvios y los hemos comentado en un apartado anterior: no debemos administrar monoterapia con antivirales ni tratar de prevenir una enfermedad mediante vacunación con un péptido sintético (Perales *et al.*, 2017; Domingo, 2020). Ambos diseños conducirán a la selección de mutantes de virus resistentes a las barreras impuestas. Debemos aplicar terapias de combinación y vacunas complejas que estimulen al sistema inmune para inducir una variada respuesta de anticuerpos y celular para conferir protección.

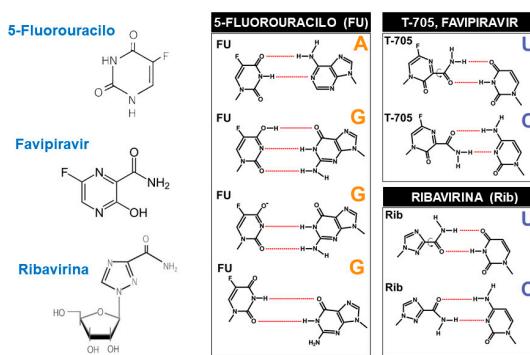
La teoría de cuasiespecies también abrió un nuevo campo de acción: inducir extinción de virus por la razón que se sintetiza matemáticamente en la segunda ecuación mostrada en la **Figura 4** (Holland *et al.*, 1990; Loeb *et al.*,



1999; Crotty *et al.*, 2000; Arias *et al.*, 2014; Gallego *et al.*, 2019). El concepto básico es que para cualquier sistema replicativo (considerado en términos abstractos) hay un máximo en el número de errores tolerados en su material genético para que la información pueda heredarse de modo estable. Ese número máximo permitido depende de la cantidad de información a mantener. El genoma humano acepta menos errores que un virus ARN porque transmite más información. No es sorprendente que el genoma humano codifique unas 100 proteínas implicadas de alguna manera en mantener la integridad de su material genético. Los genomas DNA celulares incluyen actividades de corrección de errores para mantener la integridad funcional (Reha-Krantz *et al.*, 1998, entre otros muchos estudios). En cambio, los virus ARN no hacen nada (o poco en algunos casos) para corregir errores ya que los explota en su propio beneficio en forma de agilidad adaptativa. Algunos coronavirus codifican una actividad exonucleasa capaz de corregir parte de los errores que se producen durante la elongación del RNA vírico. No se conoce todavía si tal actividad reduce la tasa de error del SARS-CoV-2 (revisado en Domingo *et al.*, 2021a).

Este concepto puede expresarse mediante la existencia de un **umbral de error, o una tasa de error por encima de la cual la información contenida en un sistema replicativo desaparece. Se dice que el sistema entra en catástrofe de error.** Aunque los virus RNA toleren mutaciones y se aprovechen de ellas, no escapan a la existencia también para ellos del umbral de error. Este es el principio de la mutagénesis letal de virus, consistente en provocar la extinción de virus por exceso de mutaciones y que se está erigiendo como una estrategia antiviral efectiva para muchos virus ARN, incluido el SARS-CoV-2 con la droga molnupiravir. La faceta experimental de esta nueva estrategia antiviral empezó en el laboratorio de J.J. Holland en la Universidad de California San Diego. Con el virus de la poliomielitis y de la estomatitis vesicular se demostró que un aumento de la tasa de error mediada por agentes mutagénicos comprometía la capacidad replicativa de los virus. Unos años más tarde, el grupo de L.A. Loeb demostró el efecto inhibitorio de un análogo mutagénico de pirimidina sobre el VIH-1 y acuñó el nombre de mutagénesis letal para el proceso de pérdida de infectividad debida a un exceso de mutaciones. Los estudios se extendieron a muchos virus y actualmente se están empleando e investigando análogos de base o de nucleósido (convertidos intracelularmente en sus correspondientes nucleótidos que se incorporan por las polimerasas víricas) para inducir mutagénesis letal. Su actividad se fundamenta principalmente en el apareamiento erróneo con las bases estándar que conduce a la introducción de mutaciones que perturban las funciones que el virus necesita para completar su ciclo de infección (**Figura 8A**).

A Bases químicas de la inducción de mutaciones



B Transiciones clave hacia la extinción de virus

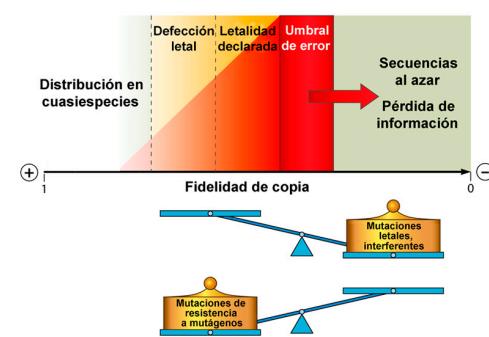


Figura 8. Mutagénesis letal de virus. (A) Los análogos empleados en mutagénesis letal ejercen su actividad mutagénica mediante apareamientos ambiguos con nucleótidos estándar. Se muestran algunos de los apareamientos posibles del 5-fluorouracilo (FU), favipiravir (T-705) y ribavirina (Rib). Otros análogos empleados actualmente en terapia, incluida para COVID-19, actúan de manera similar. (B) Relación entre teoría de cuasiespecies y mutagénesis letal. La teoría de cuasiespecies predice la existencia de un umbral de error cuando la fidelidad de un sistema replicativo (eje horizontal) supera un cierto valor (segunda ecuación de la Figura 4). Se han distinguido dos etapas (defeción letal o interferencia por genomas defectivos) y letalidad declarada (masiva acumulación de mutaciones en los genomas víricos) que preceden a la violación del umbral de error y la extinción del virus. Figura compuesta a partir de otras publicadas en Domingo, E. Virus as Populations, Elsevier, 2020, con permiso.

Una combinación de estudios experimentales y predicciones teóricas ha desvelado que al menos dos etapas sucesivas parcialmente solapantes preceden a la violación del umbral de error cuando se aumenta la tasa de error de un virus por la presencia de agentes mutagénicos. La primera etapa se denomina de **defeción letal y se debe a la formación de genomas defectivos que interfieren con la replicación de los genomas no mutados con los que coexisten en la población.** Ello se logra con un modesto incremento de la tasa de error. Le sigue una etapa de **letalidad declarada en la que, debido a un aumento del número de mutaciones, se da el colapso funcional del sistema, en nuestro caso del virus.** Tal como se indica en el eje horizontal de la **Figura 8B**, la bajada de la fidelidad de copia hasta cruzar el umbral de error conduce a la extinción del virus. Sabemos que hay factores que favorecen esta transición y otros que la impiden, notablemente la selección de mutantes de virus resistentes a los agentes mutagénicos, lo que, de nuevo, impone terapias de combinación.



La mutagénesis letal se ha empleado con éxito para extinguir numerosos virus ARN tanto en experimentos en cultivos celulares como en ensayos con organismos infectados *in vivo*. Cabe destacar que la ribavirina, un agente antiviral que se administra a humanos desde la década de los años 1970 (pensando que era un mero inhibidor de la multiplicación de varios virus ARN), se descubrió en el año 2000 que actuaba mediante mutagénesis letal. (Crotty et al. 2000) Hasta qué punto éste y otros análogos manifiestan su actividad antiviral debido a mutagénesis letal u a otras propiedades de los mismos compuestos es todavía una cuestión abierta. En favor de que la mutagénesis letal es al menos parcialmente responsable de su eficacia es que la transición hacia la extinción va acompañada de aumentos de la frecuencia de mutación del virus, tal como predicen las observaciones experimentales y modelos teóricos (compárense las **Figuras 4 y 8B**).

El interés en la mutagénesis letal se ha reforzado con motivo de la COVID-19, fundamentalmente porque drogas que se están mostrando efectivas, como el molnupiravir ejercen su acción beneficiosa al menos en parte mediante mutagénesis letal. La gran enseñanza de la situación que estamos viviendo es que un concepto emanado de la biofísica teórica como es la teoría de cuasiespecies, está encontrando una aplicación para luchar contra enfermedades víricas graves. La ciencia básica es fuente de conocimiento y solo con conocimiento se pueden diseñar aplicaciones eficaces.

CONCLUSIÓN GENERAL

Los virus ARN replican en forma de espectros o nubes de mutantes que siguen una dinámica de cuasiespecies en los organismos infectados (Domingo et al., 2021b). Esta observación anticipó la gran diversidad del mundo biológico en general, desvelada por nuevos métodos de secuenciación masiva y análisis meta-genómicos. La dinámica de cuasiespecies introduce un nuevo punto de vista sobre los virus según la cual su comportamiento depende de la acción colectiva de conjuntos de genomas. Muy lejos queda el concepto de "tipo salvaje" (el "wild type" de los textos de genética del siglo pasado) para definir y caracterizar virus.

La dinámica de cuasiespecies ha interpretado la capacidad adaptativa de los virus como una sustitución de unas subpoblaciones dominantes por otras que son incipientes en los espectros de mutantes y que permiten una respuesta a las demandas ambientales. Ello tiene implicaciones para entender las enfermedades que producen los virus y diseñar modos más efectivos para combatirlas. Los tratamientos de combinación, las vacunas capaces de desencadenar una amplia respuesta inmune y la mutagénesis letal son un

legado práctico de haber captado las consecuencias de la dinámica de cuasiespecies. Como dijeron J. J. Holland y sus colaboradores hace cuarenta años, todavía falta mucho para entender las consecuencias de tener un mundo ARN dinámico insertado en una biosfera presidida fundamentalmente por un mundo celular basado en ADN. El desafío es tratar de seguir entendiendo (Holland et al., 1982).

BIBLIOGRAFÍA

- Arias, A, Thorne, L, Goodfellow, I (2014) Favipiravir elicits antiviral mutagenesis during virus replication *in vivo*. **Elife** 3: e03679
- Batschelet E, Domingo E, Weissmann C (1976) The proportion of revertant and mutant phage in a growing population, as a function of mutation and growth rate. **Gene** 1(1):27-32.
- Crotty S, Maag D, Arnold JJ, Zhong W, Lau JY, Hong Z, Andino R, Cameron CE (2000) The broad-spectrum antiviral ribonucleoside ribavirin is an RNA virus mutagen. **Nat Med.** 6(12):1375-9.
- de la Torre JC, Holland JJ (1990) RNA virus quasispecies populations can suppress vastly superior mutant progeny. **J. Virol.** 64(12):6278-81.
- Domingo E, Sabo D, Taniguchi T, Weissmann C (1978) Nucleotide sequence heterogeneity of an RNA phage population. **Cell** 13(4):735-44.
- Domingo, E. and Schuster, P (2016) Quasispecies: from theory to experimental systems. **Current Topics in Microbiology and Immunology** Vol. 392. Springer.
- Domingo E, and Perales C (2019a) Viral quasispecies. **PLoS Genet.** 15(10): e1008271.
- Domingo, E and Perales, C (2019b) Virus evolution. In: Schmidt, Thomas M. (ed.) **Encyclopedia of Microbiology**. 4th edition. vol. 4, pp. 558-568. UK: Elsevier.
- Domingo E (2020) Virus as populations. Elsevier, Amsterdam.
- Domingo, E, García-Crespo, C, Lobo-Vega, R, Perales C (2021a) Mutation rates, mutation frequencies, and proofreading-repair activities in RNA virus genetics. **Viruses** 13(9):1882.
- Domingo E, García-Crespo C, Perales C (2021b) Historical Perspective on the Discovery of the Quasispecies Concept. **Annu. Rev. Virol.** 8(1):51-72.
- Eigen M and Schuster P (1979) The hypercycle. A principle of natural self-organization. Springer, Berlin.
- Eigen M (1993) Viral quasispecies. **Sci. Am.** 269(1):42-9.
- Eigen M (2013) From strange simplicity to complex



familiarity. Oxford University Press, Oxford.

Gallego I, Soria ME, Gregori J, de Ávila AI, García-Crespo C, Moreno E, Gadea I, Esteban J, Fernández-Roblas R, Esteban JI, Gómez J, Quer J, Domingo E, Perales C (2019) Synergistic lethal mutagenesis of hepatitis C virus. *Antimicrob. Agents Chemother.* 63(12):e01653-19.

Geoghegan JL, Holmes EC (2018) Evolutionary Virology at 40. *Genetics* 210(4):1151-1162.

Gregori J, Perales C, Rodriguez-Frias F, Esteban JI, Quer J, Domingo E (2016) Viral quasispecies complexity measures. *Virology* 493:227-37.

Holland J, Spindler K, Horodyski F, Grabau E, Nichol S, VandePol S (1982) Rapid evolution of RNA genomes. *Science* 215(4540):1577-85.

Holland JJ, Domingo E, de la Torre JC, Steinhauer DA (1990) Mutation frequencies at defined single codon sites in vesicular stomatitis virus and poliovirus can be increased only slightly by chemical mutagenesis. *J. Virol.* 64(8):3960-2.

Holmes EC (2009) The evolution and emergence of RNA viruses. *Oxford Series in Ecology and Evolution*. Oxford University Press: New York.

Loeb, LA, Essigmann, JM, Kazazi, F, Zhang, J, Rose, KD and Mullins, JL (1999) Lethal mutagenesis of HIV with mutagenic nucleoside analogs. *Proc. Natl. Acad. Sci. USA.* 96: 1492-1497.

Mills DR, Peterson RL, Spiegelman S (1967) An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. *Proc. Natl. Acad. Sci. USA.* 58(1):217-24.

Perales, C., Ortega-Prieto, A.M., Beach, N.M., Sheldon, J., Menéndez-Arias, L. and E. Domingo (2017) Quasispecies and Drug Resistance. In: *Handbook of Antimicrobial Resistance*. Gotte, M., Berghuis, A., Matlashewski, G., Wainberg, M. and Sheppard, D. (eds). Springer Science+Business Media, New York, pp. 123-147.

Reha-Krantz, LJ, Marquez, LA, Elisseeva, E, Baker, RP, Bloom, LB, Dunford, HB, Goodman, MF (1998) The proofreading pathway of bacteriophage T4 DNA polymerase. *J. Biol. Chem.* 273(36):22969-76.

Rueca, M, Bartolini, B, Gruber CEM, Piralla A, Baldanti F, Giombini E, Messina F, Marchioni L, Ippolito G, Di Caro A, Capobianchi MR (2020) Compartmentalized Replication of SARS-CoV-2 in Upper vs. Lower Respiratory Tract Assessed by Whole Genome Quasispecies Analysis. *Microorganisms* 8(9):1302.