El origen del SARS-CoV-2 y la Navaja de Ockham

ANTONI ROMEU Y ENRIC OLLÉ :: 31/05/2021

La Navaja de Ockham es un principio metodológico y filosófico, que se debe al fraile franciscano inglés Guillermo de Ockham de la Edad Media, según el cual “en igualdad de condiciones, la explicación más sencilla suele ser la más probable”. La expresión inglesa *keep it simple, stupid* se relaciona con este principio.

El problema del origen del SARS-CoV-2, también hay que abordarlo con mentalidad de genética forense. Los resultados basados en marcadores genéticos siempre se expresan en términos estocásticos o probabilísticos. En genética no hay certezas absolutas, pero sí evidencias lo altamente probables que permiten dictar sentencias de culpabilidad o determinar relaciones de paternidad y/o maternidad. El origen del SARS-CoV-2, desde el punto de vista genético, no se podrá demostrar como un teorema matemático. Pero entre dos teorías en igualdad de condiciones, que tienen las mismas consecuencia, la explicación más sencilla será la correcta: Navaja de Ockham.

Como ya hemos planteado en otros artículos (1,2), el origen del SARS-CoV-2 se puede reducir al origen o adquisición de una característica molecular en su proteína S, el llamado “sitio de escisión de furina”, que facilita enormemente la interacción del virus con las células humanas, y su gran capacidad de infección (pandemia). Técnicamente, es una “ganancia de función”: ha habido una inserción en el gen S del virus que ha provocado que la proteína S haya ganado más capacidad de infectar. El SARS-CoV-2 pertenece al grupo de los Betacoronavirus (Sarbecovirus). Sorprende que sea la única especie **conocida de Sarbecovirus** que tenga dicho sitio de furina. De momento, nadie sabe cómo y cuando le ha llegado. Esto genera controversia sobre su origen.

Querido lector, antes de seguir con el tema, es importante revisar algunos conceptos básicos de biología. En la materia viva, las proteínas son macromoléculas que hacen las funciones (la proteína S del virus, hace la función de facilitar la infección). Estructuralmente, las proteínas son cadenas de otras moléculas pequeñas llamadas “aminoácidos”. El alfabeto del lenguaje de las proteínas está constituido por 20 aminoácidos (20 letras). Por otra parte, el ADN codifica las proteínas. El ADN también es una macromolécula constituida por una cadena de otras moléculas pequeñas llamadas “nucleótidos”. El alfabeto del lenguaje del ADN está constituido por 4 nucleótidos (4 letras). El “código genético” traduce la información codificada en un alfabeto de 4 letras (A,C,G,T), en otro de 20 letras.

El código genético usa tres unidades del ADN, para especificar cada unidad de aminoácido de una proteína. Cuando se leen en grupos de 3, los 4 tipos diferentes de unidades de ADN pueden especificar 4 x 4 x 4 o 64 tripletes diferentes, o codones como se les llaman. Dado que solo hay 20 aminoácidos, hay tripletes más que suficientes para todos, lo que permite que algunos aminoácidos sean especificados por más de un triplete. Hay aminoácidos codificados por un único triplete, y hay aminoácidos codificados hasta por seis tripletes. Por

ejemplo, el aminoácido “arginina” (cuyo símbolo es R), está codificado por cualquiera de los 6 tripletes: AGG, AGA, CGA, CGC, CGG, CGT.

Cada organismo (también cada virus) tiene que fabricar sus propias proteínas (en biología, las proteínas de una especie no sirven para otra). En la síntesis de proteínas, cada aminoácido no utilizan por igual todos los tripletes que lo codifican. Por ejemplo, si un aminoácido está codificado por sólo un triplete, no hay opción; pero, si está codificado por 2 tripletes, ambos no se utilizan al 50%. Si un aminoácido está codificado por más de 2 tripletes, cada triplete es utilizado en una determinada proporción. En toda especie (incluyendo virus), siempre hay unos tripletes mayoritarios y minoritarios. En este sentido, una huella de identidad (*fingerprint*) de una especie es la tabla de frecuencias del uso de tripletes del ADN en la síntesis de sus proteínas.

Volviendo al SARS-CoV-2, el sitio de furina se caracteriza por la inserción de una secuencia de 4 aminoácidos (PRRA), que corresponde con la inserción de 12 nucleótidos (3 x 4). En el SARS-CoV-2, el deblete RR del sitio de furina está codificado por los tripletes CGG-CGG. Hay que decir que el sitio de escisión de furina, con el doblete RR es muy común en el mundo de los virus (incluyendo los Coronavirus, pero exceptuando los Betacoronavirus). La recombinación con otros virus, es la explicación más natural que se ha dado para la adquisición de dicho sitio en el SARS-CoV-2. Para analizar la verosimilitud de dicha recombinación, nosotros rastreamos las bases de datos (3), obteniendo como resultado que no hay ningún doblete RR de sitios de furina de otros virus que esté codificado por los tripletes CGG-CGG. Observamos que en estos RR, el triplete AGA es mayoritario. En toda recombinación, siempre hay un donador y un aceptor de material genético; y el código del donador pasa al aceptor. Si el SARS-CoV-2 ha adquirido el sitio de furina por recombinación con otro virus, a partir de la información disponible en las bases de datos genómicas, no se puede saber cuál a ha podido ser.

Con estos resultados, tuvimos interés en determinar el patrón de uso de los tripletes de arginina del SARS-CoV-2 (3). Estudiando la composición de todas sus proteínas, el resultado es: AGG (13%), AGA (45%), CGA (5%), CGC (10%), CGG (3%), CGT (24%). El triplete AGA es el mayoritario, y curiosamente, CGG es el minoritario. En el caso concreto de la proteína S, de las 42 argininas que tiene, 20 están codificadas por AGA, y solo 2 por CGG. Sorprende que esos 2 (CGG-CGG) sean los que codifican el doblete RR del sitio de furina.

En este sentido, los tripletes de arginina son verdaderos marcadores genéticos. En un estudio de genética forense, tal como se puede enmarcar el origen del SARS-CoV-2, el poder disponer de marcadores genéticos útiles es también una “lotería”. En ciencia, la suerte también juega. En este caso, el hecho de que el aminoácido clave del sitio de furina del virus pandémico esté codificado por 6 tripletes permite abordar la “trazabilidad” de su origen. En la materia viva, otros aminoácidos están codificados por 2 tripletes. Si este fuese el caso del aminoácido arginina, una trazabilidad vía código genético sería muy difícil. En este supuesto, una recombinación como mecanismo de adquisición del sitio de furina estaría aceptada. Es más, el origen natural del virus no sería objeto de debate: no habría caso!

Como se ha dicho, cada especie tiene su propio patrón de frecuencia de uso de los tripletes del ADN. Por lo que se refiere al aminoácido arginina, y en *Homo sapiens,* la frecuencia de

uso de tripletes es (4): AGG (20%), AGA (20%), CGA (11%), CGC (19%), CGG (21%), CGT (9%). Este patrón en el genoma humano es muy distinto del patrón en el genoma del virus. En la especie humana, el triplete CGG es el mayoritario (21%) y en el virus es el minoritario (3%). Sorprende que en el SARS-CoV-2, el doblete RR del sitio de furina esté codificado por la secuencia CGG-CGG del triplete mayoritario del genoma humano.

La teoría que sostiene un origen natural del SARS-CoV-2, debe contemplar que hayan ocurrido en el tiempo dos procesos independientes: (*i*) la inserción por mutación o recombinación de una secuencia de 12 nucleótidos en un sitio estratégico del gen S, codificando el sitio furina; y (*ii*) que la secuencia insertada en el virus contenga el triplete de arginina mayoritario del genoma humano. Dado que ambos procesos son independientes, todo sugiere una probabilidad baja.

La teoría que sostiene un origen de laboratorio del SARS-CoV-2, contempla que la inserción del sitio de furina en el genoma del virus ha sido de forma controlada, seleccionando la secuencia y el lugar de inserción. Considerando las aplicaciones de la ingeniería genética, en la actualidad se insertan (se clonan) genes de ciertas proteínas humanas en microorganismos adecuados para su fabricación comercial. Un ejemplo típico es la producción de insulina que se obtiene a partir de la levadura *Sacharomyces cerevisae*. En estos procesos, el gen humano que se inserta se optimiza con los tripletes mayoritarios del microorganismo receptor. En el caso del sitio de escisión de furina, da la impresión de que haya sido clonado en el genoma de un virus, dando lugar al SARS-CoV-2. Precisamente por el hecho de que el aminoácido arginina tiene 6 tripletes en el código genético universal, se vislumbra este indicio de clonación.

Este es el estado del arte sobre el origen del SARS-CoV-2. Probablemente, el principio de la Navaja de Ockham ayudará esclarecer la controversia.

Referencias

Antoni Romeu, Enric Ollé. COVID-19: descifrando el origen. La Haine, 10/11/2020. Acceso, 31/05/2021. https://www.lahaine.org/mm_ss_mundo.php/covid-19-descifrando-el-origen.

Antoni Romeu, Enric Ollé. SARS-CoV-2: doce nucleótidos sin piedad . La Haine, 17/02/2021. Acceso, 31/05/2021. https://www.lahaine.org/mm_ss_est_esp.php/sars-cov-2-doce-nucleotidos.

Romeu, A.R.; Ollé, E. SARS-CoV-2 and the Secret of the Furin Site. Preprints 2021, 2021020264 (doi: 10.20944/preprints202102.0264.v1). Acceso, 31/05/2021. https://www.preprints.org/manuscript/202102.0264/v1

GenScript Codon Usage Frequency Table(chart) Tool. Acceso, 31/05/2021. https://www.genscript.com/tools/codon-frequency-table.

Antonio Romeu1 y Enric Ollé2

1

: Químico. Catedrático de Bioquímica y Biología Molecular de la Universidad Rovira i Virgili.

Tarragona. España. Correo electrónico: antonioramon.romeu@iubilo.urv.cat

2

: Veterinario, Bioquímico. Profesor Asociado del Departamento de Bioquímica y Biotecnología de

la Universidad Rovira i Virgili. Tarragona. España. Correo electrónico: enric.olle@urv.cat

_____

https://www.lahaine.org/mm_ss_est_esp.php/el-origen-del-sars-cov