Artículo

Revista Ciencia y Naturaleza

Revelando fenotipos moleculares a través de la evolución del código genético

Ciencia al Instante

El código genético, ese diccionario molecular que traduce la información de los genes en proteínas, no apareció completo de un solo golpe. La vida no surgió instantáneamente, sino que la evolución de los organismos está marcada por grandes transiciones moleculares, y una de las más importantes fue la emergencia del código genético. Un modelo desarrollado por investigadores de la UNAM propone que, partiendo de un código primordial mínimo basado en RNA, dos rutas evolutivas paralelas —Ex1 y Ex2— condujeron al código estándar moderno. Hoy, las biomoléculas actuales aún conservan huellas de ese viaje desde la Tierra primitiva hasta los umbrales del LUCA.

Artículo

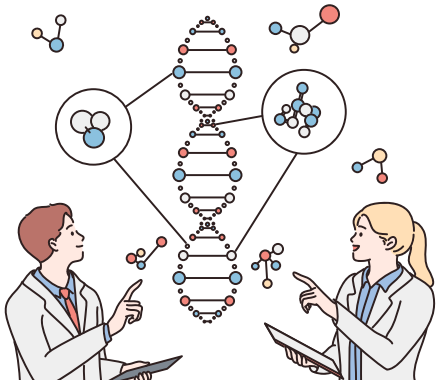
Revelando fenotipos moleculares a través de la evolución del código genético

Un modelo evolutivo del código genético revela trayectorias paralelas, en proteínas y en RNA, desde el paisaje molecular proto-biótico hasta los umbrales del LUCA.

Cómo citar este artículo: Palacios-Pérez M, José MV. 2026. Revelando fenotipos moleculares a través de la evolución del código genético. Revista Ciencia y Naturaleza (1235).

La Tierra temprana y los bloques moleculares de la vida

La vida no surgió instantáneamente, sino que la evolución de los organismos está marcada por grandes transiciones moleculares, improntas que les transformaron radicalmente. Una de las transformaciones más importantes fue la emergencia del código genético; este código funciona como un diccionario en el que están inscritas las instrucciones para producir los grandes bloques para la vida o biomoléculas: las proteínas –formadas por aminoácidos– y los ácidos nucleicos (DNA y RNA) –formados por nucleótidos.



Es importante no confundir el código genético con la información almacenada en el DNA. Esa información es el genoma, el 'texto' en sí. El código genético son las reglas que permiten leer ese texto y traducirlo en otras biomoléculas.

Actualmente, las tareas de cada biomolécula están claramente definidas: el DNA es el estable almacén de la información genética será descifrada siguiendo las reglas del código genético para producir biomoléculas nuevas; las proteínas llevan a cabo funciones catalíticas, regulan procesos, transportan sustancias y dan estructura a las células; el RNA también regula algunos procesos celulares y además es intermediario entre el DNA y la proteínas. Empero, la historia de la vida en la Tierra primigenia pudo ser aún más sorprendente e interesante.

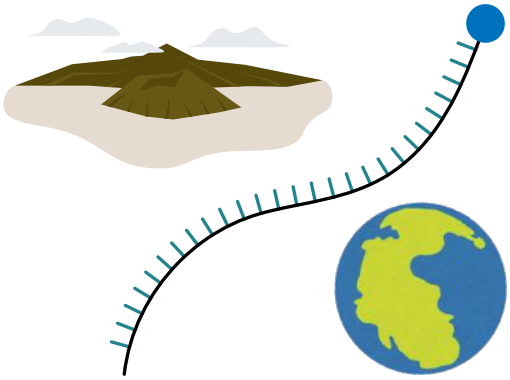
La vida no surgió de golpe: se construyó a partir de transiciones moleculares profundas.

En la década de 1980, los equipos de Thomas Cech y de Syndey Altman descubrieron independientemente que el RNA posee actividades tanto catalíticas (como una enzima proteica) como genéticas (como el DNA).

Esto llevó a la postulación de que el RNA fue la biomolécula primordial; en 1986 Walter Gilbert acuñó el término “Mundo del RNA” para referirse a tal etapa evolutiva. Sabemos que en la actualidad el código genético está cifrado como un diccionario de DNA y que las moléculas funcionales son proteínas y RNA; sin embargo, el diccionario primigenio, el primer código genético, pudo haber sido sorprendentemente pequeño e iniciar como una molécula autorreplicante.

Por otro lado, en 1953 Stanley Miller publicó un artículo revolucionario sobre la obtención experimental de aminoácidos sencillos, recreando en el laboratorio las condiciones de la Tierra primitiva. Miller retomó ideas previas de Aleksandr Ivanovich Oparin, quien en las décadas de 1920 y 1930 había propuesto que los compuestos orgánicos pudieron formarse bajo esas mismas condiciones.

El primer código genético pudo haber sido sorprendentemente simple



Uniendo entonces todos los conceptos enunciados, llegamos a una idea fascinante: las primeras biomoléculas autorreplicantes eran cadenas de RNA capaces de plegarse sobre sí mismas porque su patrón seguía la forma RNY. Dichas cadenas formaron eventualmente el código genético primordial (PGC, por sus siglas en inglés). El PGC podía decodificarse en los aminoácidos más simples y pequeños, aquellos formados a partir de los materiales y condiciones disponibles en la Tierra primitiva.

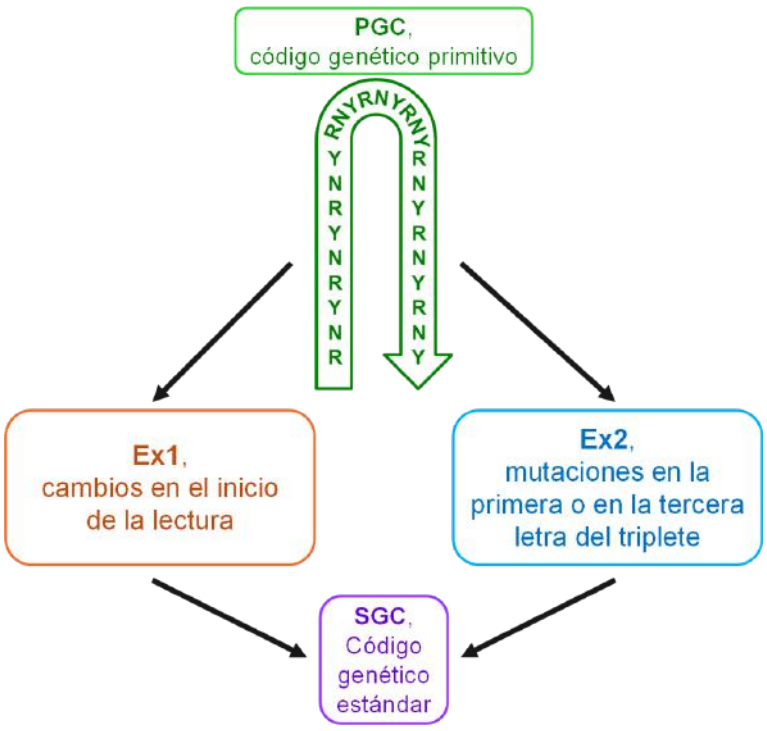
Cuando el diccionario comenzó a expandirse

Sin embargo, los tripletes que siguen el patrón RNY son solamente 16 y se corresponden con apenas 8 aminoácidos. En cambio, el código genético moderno, que comparte la gran mayoría de organismos actuales, tiene 64 tripletes que corresponden a 20 aminoácidos.

Cómo entonces llegamos al código genético estándar (SGC, por sus siglas en inglés) moderno a partir de cadenas de RNA con sólo un patrón ...RNYRNYRNY...? Más de 40 años pasaron para que surgieran modelos de evolución del código genético que tomaron como idea semilla la postulación de Eigen. Uno de los grupos que trabaja en este problema ha sido encabezado por Marco V. José, de la UNAM.

Su equipo demostró que a partir del código primordial es posible llegar al estándar moderno siguiendo dos rutas distintas. Cada una de las rutas implica cambios diferentes a partir del mismo código primordial. La primera ruta, el Código Genético Extendido 1 (Ex1) implica cambios en el punto de inicio de lectura del RNA; la segunda ruta, el Código Genético Extendido 2 (Ex2), implica un tipo particular de mutación en la primera o en la tercera letra del triplete. Eventualmente ambas rutas se complementan entre sí como se ve en la Figura 1.

Figura 1. Esquema de la expansión del código genético primordial. El RNA primigenio con patrón RNY se pliega sobre sí mismo y da origen a dos rutas de expansión: el Código Genético Extendido 1 (Ex1) –por cambios en el inicio de lectura– y el Código Genético Extendido 2 (Ex2) –por mutaciones en la primera o tercera letra del triplete. Ambas rutas convergen en el código genético estándar (SGC) moderno.



Un modelo evolutivo del código genético revela trayectorias paralelas, en proteínas y en RNA, desde el paisaje molecular proto-biótico –el mundo químico en el umbral de la vida– hasta el pre-LUCA o FUCA.

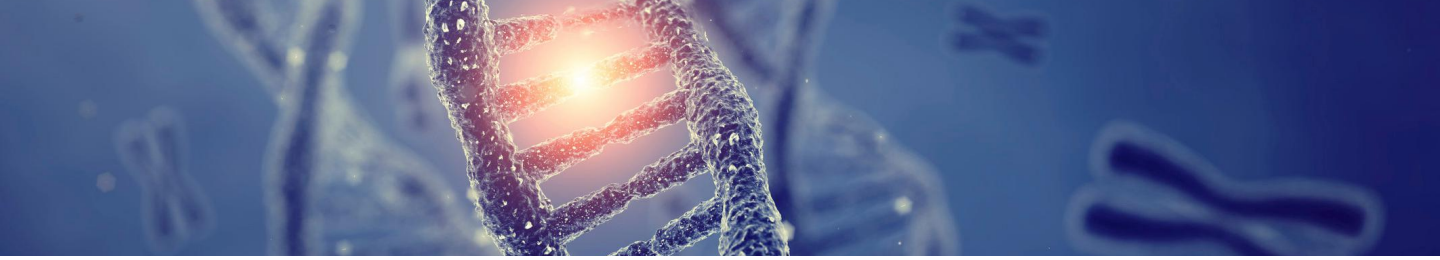
Rastros moleculares de un código en expansión

El Último Ancestro Común Universal (LUCA) es el último ancestro de todos los seres vivos que conocemos y si éste ya poseía un código genético moderno, entonces el primordial y los dos extendidos debieron formarse antes de la emergencia de LUCA y al proto-organismo anterior a LUCA se le puede llamar “Primer Ancestro Común Universal (FUCA)”.

Los códigos genéticos —antiguo, intermedios o el estándar moderno— pueden concebirse como el genotipo de cada etapa evolutiva, es decir, como las instrucciones de cada momento. Bajo esta perspectiva, las proteínas y los RNA que cada código produjo constituyen su fenotipo molecular, las moléculas resultantes de ejecutar esas instrucciones. Surge entonces una pregunta natural: si el código genético se expandió siguiendo rutas bien definidas, ¿conservan las biomoléculas actuales rasgos que reflejen las distintas etapas evolutivas del código genético que les dio origen?

¿Pueden las biomoléculas modernas conservar vestigios de los primeros códigos genéticos?

En los trabajos que presentamos a continuación, en los que participamos como parte de este grupo, exploramos dicha pregunta desde el fenotipo molecular, analizando RNA y proteínas a la luz del modelo descrito de evolución del código genético. Es importante resaltar que las moléculas actuales de RNA y las proteínas no pertenecen en su totalidad a una sola etapa del código genético; por el contrario, cada molécula contiene regiones que pueden rastrearse hasta distintos momentos de esa expansión evolutiva.



A los conjuntos de biomoléculas más antiguas que podemos reconstruir teóricamente los denominamos usando el prefijo alemán 'Ur-', que significa primigenio o ancestral: el Ur-RNAoma para la colección de RNAs ancestrales y el Ur-proteoma para la colección de proteínas ancestrales.

Las etapas tempranas del código genético: un marco para su exploración molecular

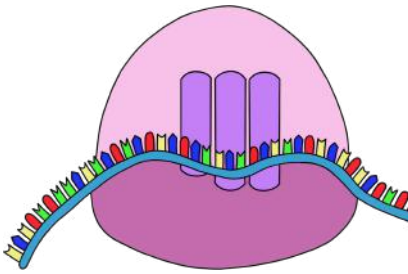
El código genético no apareció completo desde el inicio. Antes de existir un sistema capaz de traducir proteínas, moléculas de RNA con patrones simples RNY pudieron plegarse y adquirir funciones básicas. Con el tiempo, estas asociaciones se estabilizaron en un código primigenio (PGC) que comenzó a expandirse. Dos rutas evolutivas complementarias (Ex1 y Ex2) permitieron añadir complejidad, hasta consolidar el código genético casi universal de los seres vivos actuales.

El Ur-RNAoma: plegamientos ancestrales formados por RNY

Desde hace varias décadas, estudios experimentales han propuesto que el RNA primigenio debió plegarse como si fueran mini-hélices u horquillas –conocidas coloquialmente como “pasadores”, en México–, capaces de conferir cierta estabilidad estructural. Sin embargo, esos experimentos no evaluaron si los plegamientos contenían patrones específicos de secuencia. Si aceptamos que el RNA ancestral seguía mayoritariamente un patrón RNY, resulta entonces natural preguntarse si fragmentos de RNA actuales que conservan exclusivamente este tipo de tripletes presentan propiedades estructurales particulares al estar aisladas de su contexto. Buscamos sistemáticamente en moléculas modernas de RNA, regiones formadas sólo por tripletes RNY. Encontramos que estas secuencias tienden a plegarse preferentemente en horquillas y mini-hélices, tal como se había postulado.

Esto sugiriere que ciertas conformaciones estructurales simples pudieron emerger de manera espontánea, a partir de la composición misma del RNA primigenio. Existen moléculas particularmente importantes para la evolución temprana de la vida, éstas son los RNA de transferencia (tRNA). Los tRNA actúan como adaptadores físicos del código genético, pues transportan cada aminoácido al ribosoma –la fábrica molecular encargada de ensamblar las proteínas– para hacerlo, siguen las instrucciones de un RNA mensajero (mRNA), que reconoce una porción del tRNA llamada anticodón.

Encontramos que la región del anticodón es la porción más antigua evolutivamente del tRNA, pues está formada por tripletes RNY, lo que desafía los modelos tradicionales que sitúan el origen del tRNA en su extremo aceptor de aminoácidos. Estos resultados, sin embargo, no implican todavía la existencia de un código genético propiamente dicho.



Es cierto que los tripletes RNY codifican hoy en día para aminoácidos prebióticos, pero un código genético funcional requiere de una maquinaria capaz unir aminoácidos entre sí, como el Centro Peptidil Transferasa (PTC), así como elementos para posicionar un ribosoma rudimentario. Ninguna de estas estructuras clave está formadas por tripletes RNY, lo que sugiere que el Ur-RNAoma representó una etapa previa a la instauración del código genético como sistema de traducción.

El Ur-proteoma: propiedades estructurales de los primeros péptidos

Aunque aún no comprendemos con total claridad todos los pasos intermedios, la evidencia disponible sugiere que el patrón RNY terminó por establecerse como el código genético primordial (PGC).

Nuestra evidencia indica que los primeros péptidos codificados por este patrón, compuestos sólo por aminoácidos prebióticamente plausibles, poseían propiedades funcionales notables. En particular, estos péptidos ancestrales formaban módulos capaces de unir y estabilizar cofactores –pequeñas moléculas indispensables para la actividad de muchas proteínas modernas y que se consideran auténticos fósiles moleculares.

En este contexto, proponemos que los péptidos ancestrales codificados por tripletes RNY actuaron como plataformas mínimas para la unión y estabilización de estos cofactores, facilitando así la transición de sistemas prebióticos –químicos, pero con potencial de vida– a proto-biológicos –organizados con las primeras características reconocibles de los seres vivos.

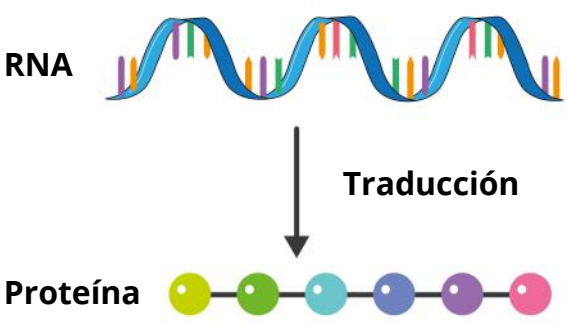
Las regiones peptídicas responsables de esta función pueden considerarse el primordio de las proteínas modernas y, por extensión, el origen de familias proteicas completas. Por ello, denominamos a estos módulos ancestrales como sitios de unión y estabilización de cofactores (CSBSs, por sus siglas en inglés) y a su conjunto funcional el “bindoma” (por el préstamo del inglés “binding”).

Adicionalmente, fuimos capaces de modelar estructuralmente ciertos péptidos que habían sido propuestos previamente como universalmente conservados y, por tanto, presentes en el Último Ancestro Común Universal (LUCA). Nuestros análisis sugieren, sin embargo, que estos módulos podrían ser aún más antiguos, situándose en una etapa evolutiva previa al propio LUCA y reforzando la noción de un Ur-proteoma funcional anterior a las proteínas modernas.

El RNA antes de LUCA: las huellas del Ex1 y Ex2

Al identificar, en moléculas de RNA de organismos actuales, las regiones codificadas por tripletes pertenecientes a uno u otro código genético extendido (Ex1 y Ex2), encontramos que prácticamente todas las moléculas de RNA modernas estaban ya en gran medida conformadas antes de la emergencia del LUCA. Nuestros resultados sugieren que, en la etapa evolutiva previa asociada a FUCA, las secuencias de RNA codificadas por ambos extendidos tendieron a complementarse entre sí, más que a competir, a diferencia de lo que caracteriza a la evolución Darwiniana clásica.

En este contexto, la traducción de proteínas ya habría sido posible, pues prácticamente todos los tipos de RNA funcionales estaban casi completamente codificados por tripletes pertenecientes a ambos códigos extendidos.



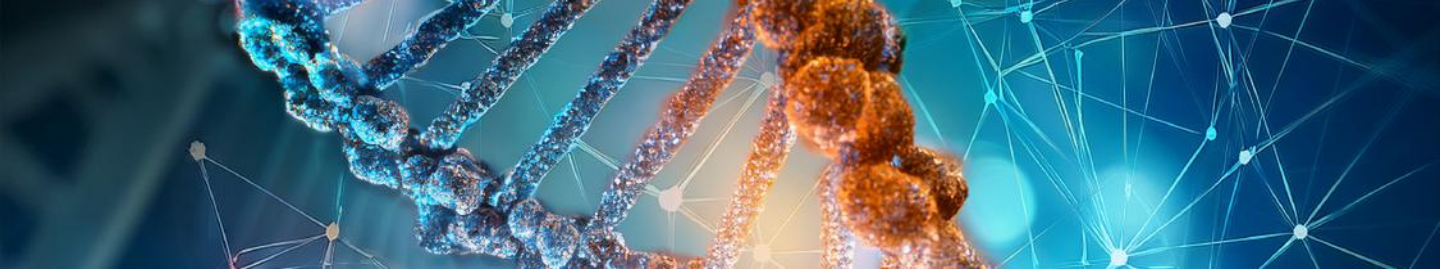
Si bien estas moléculas no eran idénticas a sus contrapartes modernas, muchas ya mostraban configuraciones sorprendentemente similares a las actuales. El hallazgo más revelador fue que, en casi todas estas moléculas, los tripletes del tipo Ex2 codifican predominantemente las regiones más importantes para la traducción de proteínas. Un claro ejemplo es el PTC, que ya habíamos mencionado, así como elementos esenciales para posicionar el ribosoma. Todos ellos fácilmente identificables mediante tripletes Ex2, pero no mediante Ex1 que parecen sólo completar las estructuras moleculares establecidas. La única excepción notable son los tRNA que se comportan de manera radicalmente distinta.

En los tRNA la huella evolutiva del código genético se manifiesta principalmente a través del aminoácido y del anticodón, más que del tipo de triplete (Ex1 vs Ex2) que los codifica. Al analizar la cronología de los tRNA, observamos que aquellos más antiguos —con anticodones ancestrales y asociados a aminoácidos considerados prebióticos— se encontraban ya casi completamente formados antes del LUCA. Por el contrario, los tRNA que cargan aminoácidos incorporados más tardíamente al código genético muestran una recuperación parcial, independientemente de si sus secuencias corresponden a tripletes Ex1 o Ex2. En este sentido, cuanto más temprana fue la incorporación de un aminoácido y de sus codones al código genético, mayor es la proporción de tRNA reconstruidos a partir de ambos extendidos, con una inclusión frecuente de la región del anticodón.

Huellas del Ex1 y Ex2 en las proteínas: hacia la funcionalidad moderna

Al analizar las proteínas que contienen regiones codificadas por tripletes pertenecientes a ambos extendidos, observamos que se distribuyen ampliamente en diversas rutas metabólicas. Destacan aquellas asociadas con el procesamiento de ácidos nucleicos, con rutas metabólicas ancestrales, así como proteínas implicadas en el transporte transmembranal de moléculas diversas.

Por ejemplo, un grupo de trabajo independiente propone la vía de Wood-Ljungdahl como la ruta metabólica más plausible para LUCA, aunque dicho estudio no identifica las enzimas clave de este metabolismo dentro de su propio enfoque; en contraste, en nuestro análisis fue posible identificar todas las enzimas clave de dicho metabolismo como parte del proteoma asociado a Ex1 y Ex2.

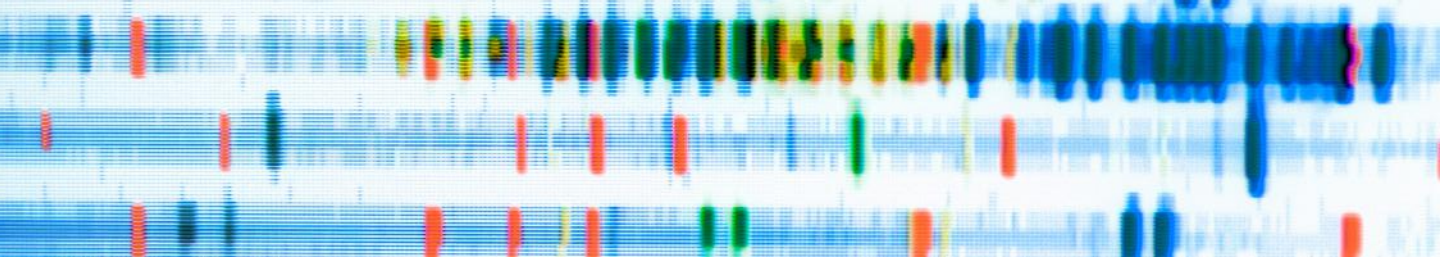


Asimismo, las proteínas incluidas en nuestras listas coinciden con aquellas identificadas por otros grupos como proteínas antiguas, obtenidas mediante enfoques metodológicos muy distintos. De manera particular, observamos que, en casi cada proteína identificada, las regiones de la molécula codificadas por los tripletes asociados a cada extendido resultan complementarias entre sí.

Mientras que los tripletes del Ex2 muestran una cobertura casi completa de la molécula, los tripletes del Ex1 cubren únicamente regiones que no son codificadas por los primeros. En este sentido, las porciones no cubiertas por un tipo de triplete sí lo son por el otro, dando lugar a un patrón de codificación cooperativo. Así, el Ex1 parece desempeñar un papel de ajuste fino, completando los huecos que deja el Ex2, lo que sugiere un mecanismo de codificación complementario entre ambos.

De tripletes ancestrales a sistemas funcionales: una lectura integrada

A la luz de estos resultados y en consonancia con aportes de la biomatemática evolutiva, proponemos un escenario in silico (computacionalmente) en el que los conjuntos ancestrales de RNA y proteínas pueden reconstruirse como componentes contemporáneos de una misma etapa evolutiva previa al LUCA (Figura 2). Este escenario se sitúa antes de la consolidación del DNA como soporte genético dominante y en un periodo en el que las proteínas comenzaban a adquirir un papel estructural y catalítico central. Aunque las hipótesis aquí planteadas deberán ser contrastadas mediante experimentación in vitro, el estudio permite delinear una posible trayectoria para el establecimiento progresivo del código genético.



Las biomoléculas actuales pueden ser archivos vivos de la historia del código genético

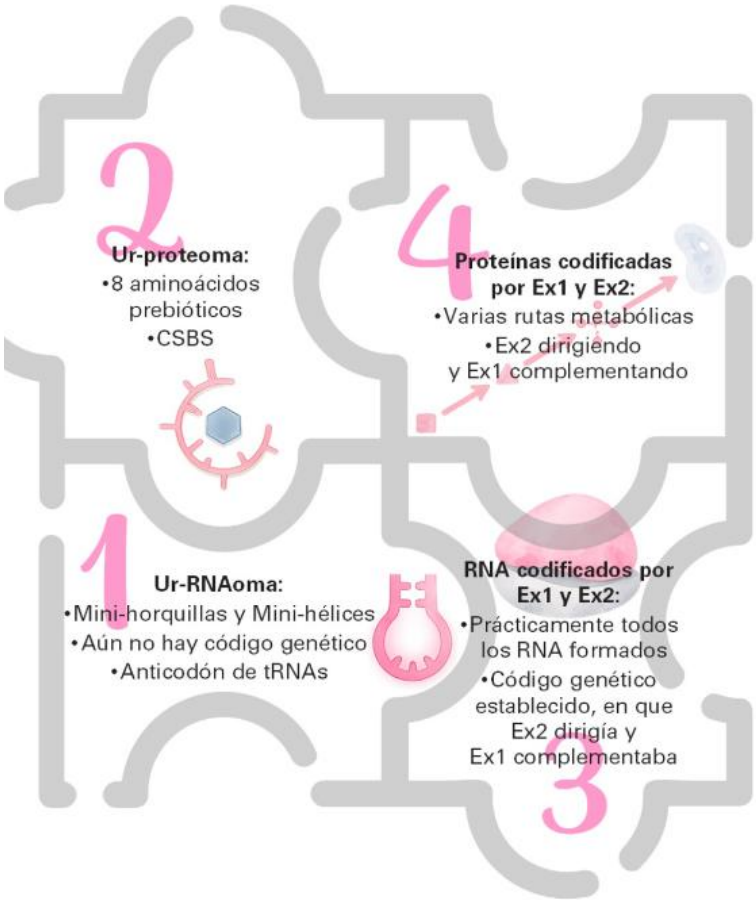


Figura 2. Síntesis de las etapas evolutivas del código genético. Cada pieza representa un conjunto de biomoléculas asociado a una etapa de expansión del código genético, desde el RNA primigenio (Ur-RNAoma) hasta las proteínas codificadas por los códigos genéticos extendidos, en el periodo previo al Último Ancestro Común Universal (LUCA).

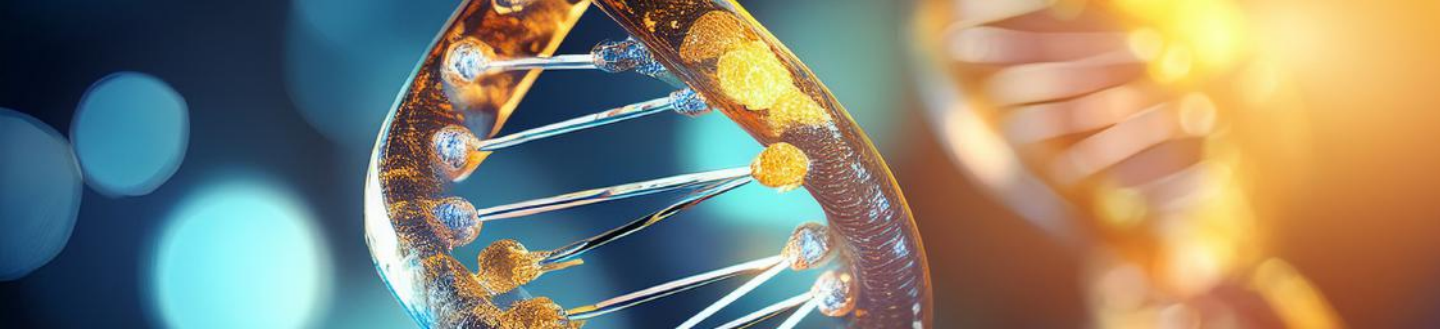
Desde esta perspectiva, el establecimiento del código genético puede concebirse como un proceso gradual. En una etapa temprana, cadenas de RNA con patrón RNY habrían existido como entidades estructurales capaces de plegarse –a modo de pasadores moleculares– sin constituir aún un código genético propiamente dicho, en ausencia del PTC. Posteriormente, sin que los mecanismos precisos sean aún conocidos, el patrón RNY se habría fijado como un código genético primordial (PGC), permitiendo la síntesis de los primeros péptidos a partir de aminoácidos prebióticos.

En este sentido, la célebre afirmación de Theodosius Dobzhansky —que nada en biología tiene sentido si no es a la luz de la evolución— adquiere aquí un matiz particularmente literal. El código genético, lejos de haber surgido como un sistema acabado, aparece como el resultado de una historia de ensamblaje progresivo, en la que estructuras de RNA, péptidos ancestrales y sistemas de codificación parciales se fueron integrando de manera complementaria. Sólo al reconstruir esa trayectoria, previa al LUCA, es posible comprender por qué el código genético moderno posee la arquitectura, la redundancia y la coherencia funcional que hoy observamos.

"Nada en biología tiene sentido, excepto a la luz de la evolución"
Thedosius Dobzhansky

Para Consulta

- Eigen M, Gardiner W, Schuster P, Winkler-Oswatitsch, R. 1981. The Origin of Genetic Information. Scientific American 244: 88–119.
- Miller SL. 1953. A Production of Amino Acids Under Possible Primitive Earth Conditions. Science 117: 528–529. [Link]
- José MV, Govezensky T, García JA, Bobadilla JR. 2009. On the evolution of the standard genetic code: vestiges of critical scale invariance from the RNA world in current prokaryote genomes. PLoS ONE 4: e4340. [Link]
- Palacios-Pérez M, José MV. 2023. A Proposal of the Ur-RNAome. Genes 14: 2158. [Link]
- Palacios-Pérez M, Andrade-Díaz F, José MV. 2018. A Proposal of the Ur-proteome. Origin of Life and Evolution of Biospheres 48: 245–258. [Link]
- Palacios-Pérez M, José MV. 2024. A Proposal for the RNAome at the Dawn of the Last Universal Common Ancestor. Genes 15: 1195. [Link]
- Palacios-Pérez M, José MV. 2019. The evolution of proteome: From the primeval to the very dawn of LUCA. Biosystems 181: 1–10. [Link]
- Prosdocimi F, José MV, de Farias ST. 2019. The First Universal Common Ancestor (FUCA) as the Earliest Ancestor of LUCA's (Last UCA) Lineage. 43–54. In: Pontarotti P (Ed.). Evolution, Origin of Life, Concepts and Methods. Springer International Publishing Cham. [Link]



Agradecimientos

MPP ha sido financiada por SECHTI (antes CONAHCYT, CVU 694877); también recibió becas de viaje otorgadas por NoRCEL y Sohan Jheeta para participar en diversas conferencias internacionales en las que se han presentado los trabajos originales aquí mencionados.
MVJ ha recibido apoyo financiero de DGAPA PAPIIT-IN200823, UNAM, México.

Crédito de imágenes en orden de aparición: Science Photo Library, sciencephoto, TrueCreatives, Vectorium, Thanapipat Kulmuangdoan, Wilvie's Images, Vesvostd, irasutoya, Getty Images, Pro500, Anastasiia Torianyk's Images, berkay08, Vidhi Vora's Images, Katsiaryna Hatsak, natrot, berkay08, Jaturawutthichai's Images, HL12 Studio, Nobi_Prizue, Pravokrugulnik. Créditos de figuras: Las figuras fueron elaboradas por los autores en PowerPoint (Microsoft®); la Figura 2 incorpora además imágenes generadas con apoyo de la IA generativa Gemini de Google [Link]. Durante la preparación de este artículo de divulgación los autores utilizaron ChatGPT de OpenAI [Link] y posteriormente Claude de Anthropic [Link] como apoyo para mejorar la fluidez, coherencia y legibilidad de una parte del texto; no obstante, las ideas presentadas son originales de los autores. Los autores revisaron cuidadosamente el escrito completo, realizaron las correcciones pertinentes y asumen plena responsabilidad por el contenido.

Dra. Verónica Reyes-Galindo
Co-Editora Invitada.
Primer Congreso Mexicano de Evolución

Diseño de publicación: Sofia Paz

Miryam Palacios Pérez

Laboratorio de Biología Teórica, Departamento de Inmunología, Instituto de Investigaciones Biomédicas, UNAM. Investigadora postdoctoral becada por SECIHTI (antes CONAHCyT). Candidata en el SNII. Presidenta en "The NoRCEL Institute" y responsable de su capítulo latinoamericano. Tiene experiencia en evolución estructural de biomoléculas tempranas y modernas, así como en microbiota intestinal humana.

contacto: mir.pape@iibiomedicas.unam.mx

Marco V. José

Laboratorio de Biología Teórica, Departamento de Inmunología, Instituto de Investigaciones Biomédicas, UNAM. Investigador Titular C de TC. Nivel 3 en el SNII. Tiene amplia experiencia en el modelado matemático de diversos fenómenos biológicos como: frecuencia cardiaca; enfermedades infecciosas y epidemias; unión ligando-receptor; evolución genómica de virus, bacterias, arqueas y primates; redes de interacción en diversos tipos de cáncer; modelado matemático de microbiotas.

contacto: marcojose@iibiomedicas.unam.mx