

# VOLUMEN 2

## ORÍGENES, (CO)EVOLUCIÓN, DIVERSIDAD Y SÍNTESIS DE LA VIDA

### **Coordinadores del tema**

Paola Bovolenta, Miguel Manzanares  
y Javier Buceta

DESAFÍOS CIENTÍFICOS DEL CSIC: RUMBO AL 2030

Desafíos coordinados por:

Jesús Marco de Lucas y M. Victoria Moreno-Arribas

VOLUMEN 2

ORÍGENES,  
(CO)EVOLUCIÓN,  
DIVERSIDAD Y  
SÍNTESIS DE LA VIDA

Este es un artículo de acceso abierto distribuido bajo los términos de la licencia de uso y distribución Creative Commons Reconocimiento 4.0 Internacional (CC BY 4.0).  
Más información sobre esta licencia en <https://creativecommons.org/licenses/by/4.0/>

Las noticias, los asertos y las opiniones contenidos en esta obra son de la exclusiva responsabilidad del autor o autores. La editorial, por su parte, solo se hace responsable del interés científico de sus publicaciones.

*Catálogo de publicaciones de la Administración General del Estado:*  
<https://cpage.mpr.gob.es>

EDITORIAL CSIC:  
<http://editorial.csic.es> (correo: [publ@csic.es](mailto:publ@csic.es))



© CSIC  
© de cada texto, sus autores  
© de las ilustraciones, las fuentes mencionadas

ISBN Vol. 2: 978-84-00-11035-2  
ISBN O.C.: 978-84-00-11008-6  
e-ISBN Vol. 2: 978-84-00-11036-9  
e-ISBN O.C.: 978-84-00-11010-9  
NIPO: 833-22-114-6  
e-NIPO: 833-22-115-1  
DL: M-15686-2022

Diseño y maquetación: gráfica futura

DESAFÍOS CIENTÍFICOS DEL CSIC: RUMBO AL 2030

# VOLUMEN 2

## ORÍGENES, (CO)EVOLUCIÓN, DIVERSIDAD Y SÍNTESIS DE LA VIDA

**Coordinadores del tema**

Paola Bovolenta, Miguel Manzanares  
y Javier Buceta

## **DESAFÍOS CIENTÍFICOS DEL CSIC: RUMBO AL 2030**

¿Cuáles son los principales desafíos científicos de la primera mitad del siglo XXI?  
¿Podemos establecer las prioridades para el futuro? ¿Cómo debe abordarlos la comunidad científica?

Este libro presenta las reflexiones del Consejo Superior de Investigaciones Científicas (CSIC) sobre 14 temas estratégicos establecidos en función de su impacto científico e importancia social.

Se abordan cuestiones fundamentales como el origen de la vida, la exploración del universo, la inteligencia artificial, el desarrollo de energías limpias, seguras y eficientes o la comprensión del funcionamiento del cerebro. El documento identifica desafíos complejos en áreas como la salud y las ciencias sociales, y los temas estratégicos seleccionados abarcan tanto cuestiones básicas como posibles aplicaciones del conocimiento. En este análisis han participado cerca de 1100 investigadores de más de 100 centros del CSIC y otras instituciones (organismos públicos de investigación, universidades, etc.). Todos coinciden en la necesidad de adoptar un enfoque multidisciplinario y en fomentar la investigación colaborativa que permita poner en marcha proyectos ambiciosos centrados en temas concretos.

Estos 14 *libros blancos*, concebidos como marco de referencia para el desarrollo de la estrategia científica de la institución, permitirán conocer la investigación que se realiza actualmente en el CSIC y, al mismo tiempo, construir una visión global de lo que serán los principales desafíos científicos en la próxima década.

### **VOLÚMENES QUE COMPONEN LA OBRA**

- 1 *Nuevos fundamentos para una sociedad global sostenible*
- 2 *Orígenes, (co)evolución, diversidad y síntesis de la vida*
- 3 *Genoma y epigenética*
- 4 *Desafíos en el ámbito de la biomedicina y la salud*
- 5 *Cerebro, mente y comportamiento*
- 6 *Producción primaria sostenible*
- 7 *Impactos del cambio global*
- 8 *Energía limpia, segura y eficiente*
- 9 *Comprensión de los elementos básicos del universo, su estructura y evolución*
- 10 *Información digital y compleja*
- 11 *Inteligencia artificial, robótica y ciencia de los datos*
- 12 *¿Nuestro futuro? Espacio, colonización y exploración*
- 13 *Desafíos de las ciencias del mar para el 2030*
- 14 *Dynamic Earth: Explorando el pasado, preparando el futuro*

## **Desafíos científicos del CSIC: rumbo al 2030**

### **Desafíos coordinados por:**

Jesús Marco de Lucas y M.<sup>a</sup> Victoria Moreno-Arribas

### **Volumen 2**

### ***Orígenes, (co)evolución, diversidad y síntesis de la vida***

#### **Coordinadores del tema**

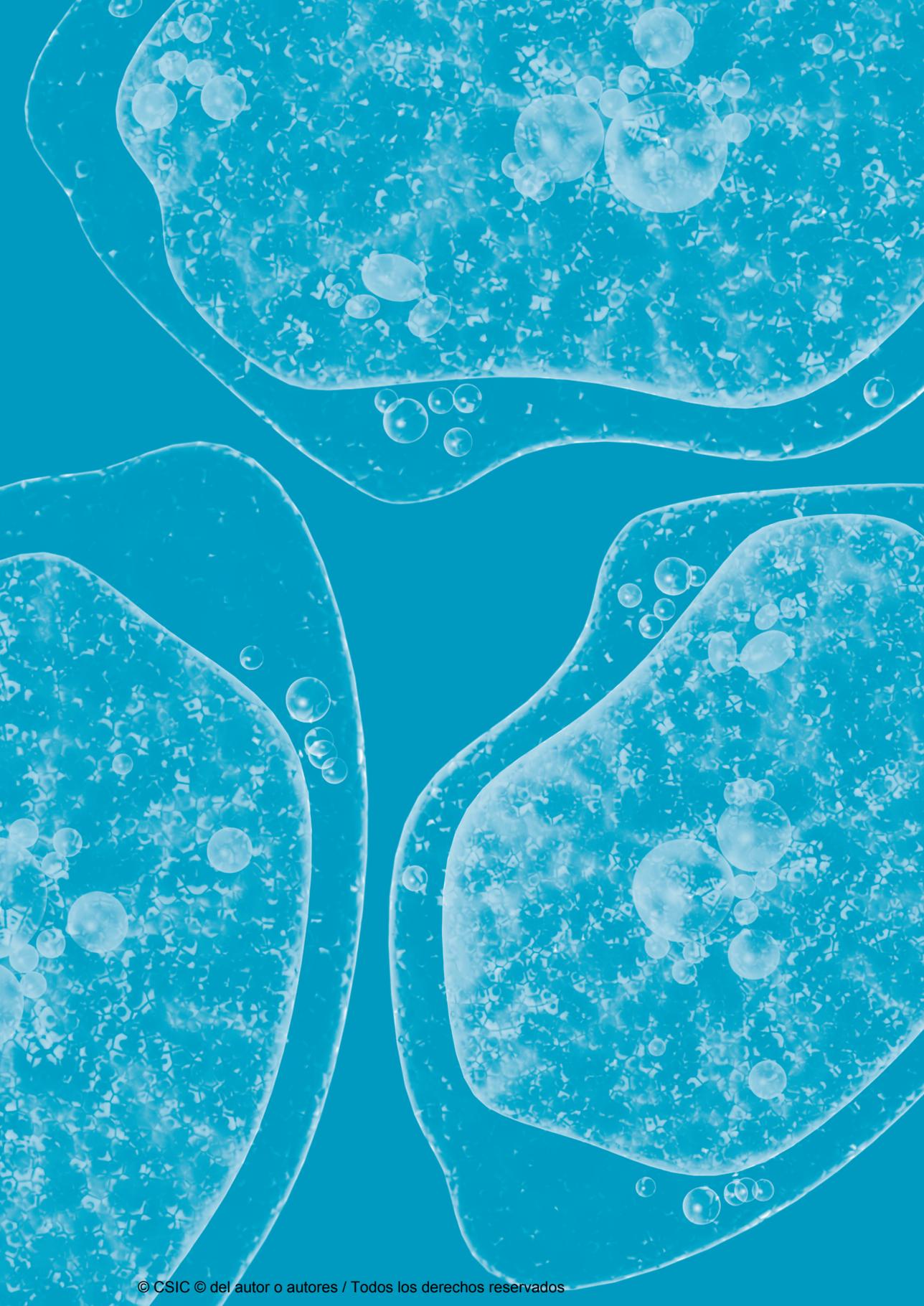
Paola Bovolenta (CBM-SO), Miguel Manzanares (CBM-SO) y Javier Buceta (I2SysBio CSIC-UV)

#### **Coordinadores de desafíos**

Carlos Briones (CAB, CSIC/INTA); Izaskun Jiménez-Serra (CAB, CSIC/INTA); José M. Valpuesta (CNB); Santiago Ramón-Maiques (CBMSO); Rafael Zardoya (MNCN-CSIC); Ana Riesgo (MNCN-CSIC); Fernando Casares (CABD, CSIC-UPO); Ignacio Maeso (CABD, CSIC-UPO); Sergi Valverde (IBE); Saúl Ares (CNB); Carles Lalueza-Fox (IBE); Ignacio de la Torre (IH-CCHS); Santiago F. Elena (I2SysBio); Iñaki Comas (IBV); Germán Rivas (CIB-MS) y Eva García (ICP)

#### **Centros participantes**

Centro de Astrobiología (CAB, CSIC/INTA)  
Centro Andaluz de Biología del Desarrollo (CABD-CSIC)  
Centro de Biotecnología y Genómica de las Plantas (CBGP)  
Centro de Biología Molecular Severo Ochoa (CBM-SO)  
Centro de Investigaciones Biológicas Margarita Salas (CIB-MS)  
Centro de Investigación Cardiovascular (CIC)  
Centro de Investigaciones sobre Desertificación (CIDE-CSIC)  
Centro Nacional de Biotecnología (CNB)  
Estación Biológica de Doñana (EBD-CSIC)  
Estación Experimental de Zonas Áridas (EEZA-CSIC)  
Instituto de Biología Integrativa de Sistemas (I2SysBio CSIC-UV)  
Instituto Andaluz de Ciencias de la Tierra (IACT, CSIC-UG)  
Instituto Botánico de Barcelona (IBB-CSIC)  
Instituto de Biomedicina y Biotecnología de Cantabria (IBBTEC CSIC-UNICAN)  
Instituto de Biología Evolutiva (IBE)  
Instituto Biofísica (IBF, CSIC-UPV-EHU)  
Instituto de Ciencias del Mar (ICM-CSIC)  
Instituto de Ciencia de Materiales de Madrid (ICMM)  
Instituto de Filosofía (IFS-CCHS)  
Instituto de Historia (IH-CCHS)  
Instituto de Investigaciones Marinas (IIM)  
Instituto Mediterráneo de Estudios Avanzados (IMEDEA-CSIC)  
Instituto de Neurociencias (IN, CSIC-UMH)  
Instituto de Nanociencia de Aragón (INA)  
Instituto de Parasitología y Biomedicina López-Neyra (IPBLN)  
Instituto de Productos Naturales y Agrobiología (IPNA-CSIC)  
Instituto de Química Física Rocasolano (IQFR-CSIC)  
Instituto de Recursos Naturales y Agrobiología de Sevilla (IRNAS-CSIC)  
Museo Nacional de Ciencias Naturales (MNCN-CSIC)  
Real Jardín Botánico (RJB-CSIC)  
Universidad Autónoma de Madrid (UAM)  
Universidad de Alcalá de Henares (UAH)  
Universidad de Extremadura (UEX)



- 8 **RESUMEN EJECUTIVO**  
ORÍGENES, (CO)EVOLUCIÓN, DIVERSIDAD Y SÍNTESIS DE LA VIDA  
**Coordinadores** Paola Bovolenta, Miguel Manzanares and Javier Buceta
- 21 **DESAFÍO 1**  
LOS ORÍGENES DE LA VIDA. DE LA QUÍMICA A LA BIOLOGÍA  
**Coordinadores** Carlos Briones y Izaskun Jiménez-Serra
- 60 **DESAFÍO 2**  
BASES ESTRUCTURALES DE LA VIDA Y EVOLUCIÓN DE  
LA COMPLEJIDAD MACRO-MOLECULAR  
**Coordinadores** José M. Valpuesta y Santiago Ramón-Maiques
- 80 **DESAFÍO 3**  
EL ÁRBOL DE LA VIDA: ENTRELAZANDO LA GENÓMICA  
Y LA EVOLUCIÓN  
**Coordinadores** Rafael Zardoya y Ana Riesgo
- 104 **DESAFÍO 4**  
LA GÉNESIS DEL FENOTIPO  
**Coordinadores** Fernando Casares y Ignacio Maeso
- 122 **DESAFÍO 5**  
BIOLOGÍA DE SISTEMAS EVOLUTIVA  
**Coordinadores** Sergi Valverde y Saúl Ares
- 138 **DESAFÍO 6**  
EVOLUCIÓN SOCIAL Y HUMANA  
**Coordinadores** Carles Lalueza-Fox y Ignacio de la Torre
- 160 **DESAFÍO 7**  
EVOLUCIÓN DE LA SALUD Y LAS ENFERMEDADES  
**Coordinadores** Santiago F. Elena y Iñaki Comas
- 188 **DESAFÍO 8**  
VIDA SINTÉTICA  
**Coordinadores** Germán Rivas y Eva García

## RESUMEN

Cómo surgió la vida en la Tierra y cómo se diversificó después en las diferentes formas de vida existentes en la actualidad son las preguntas sin respuesta que se tratarán en este volumen. Estas preguntas ahondan en el pasado profundo de nuestro planeta, donde la biología se entremezcla con la geología y la química, para explorar el origen de la vida y comprender su evolución, ya que «nothing makes sense in biology except in the light of evolution» (nada tiene sentido en biología si no es a la luz de la evolución) (Dobzhansky, 1964). Los ocho desafíos que componen este volumen resumen nuestros conocimientos actuales y las futuras líneas de investigación tocando diversos aspectos del estudio de la evolución, que puede considerarse una disciplina fundamental de las ciencias de la vida. El volumen discute las teorías recientes sobre cómo surgieron, se organizaron y adquirieron su estructura las primeras moléculas, lo que permitió las primeras formas de vida. También intenta explicar cómo esta vida ha cambiado a lo largo del tiempo, lo que ha dado lugar, a partir de bases moleculares muy similares, a una inmensa diversidad biológica, así como comprender cuál es la relación filogenética entre las diversas formas de vida. En el volumen, se analiza, además, la evolución humana, su relación con el medioambiente y sus implicaciones en la salud humana y en la sociedad. Para cerrar el círculo, el volumen trata la posibilidad de diseñar nuevas máquinas biológicas, con la consiguiente creación de un prototipo celular a partir de sus componentes, y si este conocimiento puede aplicarse para mejorar nuestro ecosistema. El CSIC, con una eficaz coordinación entre sus tres grandes áreas de conocimiento, puede convertirse en un referente internacional de la investigación en este campo.

## PALABRAS CLAVE

árbol de la vida  
arqueología de la evolución humana  
astroquímica bioingeniería  
biología del desarrollo biología estructural  
biología de sistemas  
biología de sistemas evolutiva  
biología sintética célula mínima  
coevolución huésped-patógeno  
fenotipo evolutivo filogenómica  
genómica evolutiva  
mapa genotipo-fenotipo  
medicina darwiniana metabolismo mínimo  
mundo del ARN química prebiótica  
paleoantropología paleogenómica  
protocolos sistemas complejos

# ORÍGENES, (CO)EVOLUCIÓN, DIVERSIDAD Y SÍNTESIS DE LA VIDA

## Coordinadores del tema

Paola Bovolenta, Miguel Manzanares y Javier Buceta

## RESUMEN EJECUTIVO

Algunas de las mayores incógnitas de la ciencia moderna tienen que ver con cómo surgió la vida en la Tierra y cómo, a partir de ahí, se diversificó en las múltiples formas de vida actuales. Estas preguntas ahondan en el pasado profundo de nuestro planeta, donde la biología se entremezcla con la geología y la química, para explorar el origen de la vida. Además, al aprender cómo cambian los sistemas biológicos a lo largo del tiempo, podemos diseñar nuevas máquinas biológicas basadas en este conocimiento para llevar a cabo cometidos aun no cumplidos. El principal asunto que se aborda en este tema es la evolución, teniendo en cuenta que «nothing makes sense in biology except in the light of evolution» (nada tiene sentido en biología si no es a la luz de la evolución) (Dobzhansky, 1964). Comprender la evolución nos dará pistas sobre el origen de la vida y sobre los mecanismos moleculares concretos que actúan en los seres vivos y cómo cambian con el tiempo. De este modo, estaremos preparados para intentar conjugar estos mecanismos de nuevas formas, con lo que allanaremos el camino a la biología sintética. La evolución es el principio general más importante y uno de los pocos —si no el único— de la biología.

En este volumen, se exploran y discuten diversos aspectos relacionados con el estudio de la evolución, considerada como una disciplina fundamental y central en las ciencias de la vida, que debería calar en las diferentes áreas de investigación en los próximos años. Los ocho desafíos que componen este volumen resumen nuestros conocimientos actuales y las futuras líneas de

investigación en aspectos diferentes pero relacionados con esta cuestión primordial.

El primer desafío analiza el origen y la etapa inicial de evolución de la vida con el objetivo de explicar la síntesis de los componentes bioquímicos y su polimerización para originar los biopolímeros funcionales y portadores de información responsables de la bioquímica de la vida, así como su acoplamiento en sistemas autónomos con capacidades evolutivas. Entender los fundamentos de la vida y su diversidad evolutiva requiere también descifrar la estructura tridimensional y la naturaleza dinámica de todas las macromoléculas, cómo se ensamblan y funcionan de manera coordinada, oportuna y precisa. El segundo desafío aborda este problema y analiza cómo este conocimiento podría favorecer nuestra comprensión y tratamiento de las enfermedades, y permitirnos explotar los procesos biológicos con fines biotecnológicos y para diseñar nuevas entidades biológicas.

Una vez creada la vida, esta ha evolucionado para dar lugar a una inmensa diversidad biológica, cuya proporción desconocemos realmente. El tercer desafío analiza cómo la aplicación generalizada de técnicas genómicas de alto rendimiento permitirá reconstruir el *Árbol de la Vida* e identificar las dianas genómicas de la selección natural, proporcionando claves fundamentales para entender la génesis de esta diversidad. Sin embargo, ¿se reflejan todos los cambios genómicos en los cambios fenotípicos? El cuarto desafío pretende responder a esta pregunta analizando cómo se genera un fenotipo. Se trata de una cuestión fundamental en biología, con implicaciones prácticas para la salud humana, la producción de alimentos o el cambio climático, que también afecta a varias áreas de la ingeniería y las ciencias sociales. Este mismo problema puede analizarse con un enfoque sistémico. Así, el quinto desafío tiene como objetivo generar una comprensión mecanicista y evolutiva de los mapas genotipo-fenotipo a múltiples escalas utilizando una combinación de enfoques matemáticos, moleculares y celulares.

Dentro del escenario evolutivo, la evolución humana siempre ha suscitado una atención particular. El sexto desafío busca comprender los procesos de adaptación social y biológica que tuvieron lugar a lo largo de la historia evolutiva humana. Estos procesos tienen dimensiones moleculares, genéticas, conductuales, sociales y anatómicas, cuya comprensión solo puede resolverse con nuevos enfoques multidisciplinarios. Estrechamente ligada a nuestra curiosidad por la evolución humana está nuestra necesidad de entender por qué enfermamos y, más en general, por qué enferman los seres vivos. Las

enfermedades son el resultado de alteraciones de la homeostásis, causadas por perturbaciones endógenas (por ejemplo, enfermedades hereditarias o cáncer) o exógenas (por ejemplo, infecciones o intoxicaciones). El séptimo desafío examina las enfermedades como el resultado de procesos coevolutivos, aportando una perspectiva que ofrece una mayor capacidad de predicción — por ejemplo, en la evolución de las pandemias— y que puede permitir que se combatan las enfermedades con mayor eficacia.

El octavo desafío, que cierra el círculo y enlaza de nuevo con el primero, estudia la posibilidad de ensamblar una unidad vital mínima con funcionalidad programable. Ser capaz de construir una célula sintética a partir de sus componentes esenciales debería contribuir a nuestra comprensión de los principios básicos de la vida, proporcionando herramientas para soluciones novedosas a problemas medioambientales y biomédicos.

El CSIC se encuentra en una posición única para afrontar el desafío de entender la evolución con todas sus implicaciones y ramificaciones, y mediante una coordinación eficaz entre sus tres grandes áreas de conocimiento, puede convertirse en un referente internacional de la investigación en este campo.

## INTRODUCCIÓN

Algunas de las mayores incógnitas de la ciencia moderna tienen que ver con cómo surgió la vida en la Tierra y cómo, a partir de ahí, se diversificó en las múltiples formas de vida actuales. Estas preguntas ahondan en el pasado profundo de nuestro planeta, donde la biología se entremezcla con la geología y la química, para explorar el origen de la vida. Además, al aprender cómo cambian los sistemas biológicos a lo largo del tiempo, podemos diseñar nuevas máquinas biológicas basadas en este conocimiento para llevar a cabo cometidos aun no cumplidos.

El principal asunto que se aborda en este tema es la evolución. No por mucho repetirse deja de ser cierta la afirmación de Theodosius Dobzhansky (1964): «Nothing makes sense in biology except in the light of evolution» (nada tiene sentido en biología si no es a la luz de la evolución). Comprender la evolución nos dará pistas sobre el origen de la vida y sobre los mecanismos moleculares concretos que actúan en los seres vivos y cómo cambian con el tiempo. De este modo, estaremos preparados para intentar conjugar estos mecanismos de nuevas formas, con lo que allanaremos el camino a la biología sintética. La evolución es el principio general más importante y uno de los pocos —si no el único— de la biología.

En este tema, analizamos y debatimos varios aspectos relacionados con el estudio de la evolución, como disciplina fundamental y central en las ciencias de la vida, que debe calar en las diferentes áreas de investigación en los próximos años.

*Orígenes, árboles y la génesis del fenotipo.* El origen y la etapa inicial de evolución de la vida es uno de los temas científicos más desafiantes, ya que pretende explicar la síntesis de unidades estructurales bioquímicas y su polimerización para dar lugar a los biopolímeros funcionales e informativos responsables de la bioquímica de la vida. Se trata de un proceso complejo que permitió la transición de la química a la biología en la Tierra, quizás incluso más allá de nuestro planeta. Múltiples mecanismos pudieron haber contribuido a la biogénesis, incluidas las reacciones sintéticas en entornos astrofísicos, la llegada de materia orgánica de meteoritos y núcleos cometarios, así como los procesos geofísicos y geoquímicos que tuvieron lugar en la Tierra primitiva, hace unos cuatro mil millones de años. El descubrimiento de más de doscientas moléculas pequeñas en el espacio interestelar y en cuerpos extraterrestres sugiere que los procesos prebióticos básicos son omnipresentes en el universo.

La combinación de compuestos en diferentes entornos de la Tierra primitiva dio lugar a un número creciente de biomoléculas, que pudieron llegar a acoplarse en sistemas autónomos. Estos representarían los primeros organismos autorreproducibles y evolucionables que deberían haber sido capaces de mantener sus componentes moleculares unidos y distinguirse de su entorno; mantenerse alejados del equilibrio termodinámico captando recursos energéticos y materiales del entorno; y transmitir información heredable a su progenie. Así pues, los compartimentos de membrana, las maquinarias metabólicas y los mecanismos de replicación deben de haberse originado y combinado en la transición de sistemas químicos complejos (aunque todavía dirigidos termodinámicamente) a sistemas protobiológicos y, en última instancia, a organismos vivos (controlados de forma cinética y espacial).

Entender cómo se unen y cambian estos componentes es también un aspecto fundamental de los estudios evolutivos modernos. Para ello es necesario descifrar la estructura tridimensional y la naturaleza dinámica de todas las macromoléculas subyacentes a los procesos vivos, y cómo se ensamblan y funcionan de forma coordinada, oportuna y precisa. Abordar este desafío nos permitirá comprender y tratar las enfermedades, aprovechar los procesos biológicos con fines biotecnológicos y diseñar sintéticamente nuevas entidades biológicas. Las máquinas macromoleculares, formadas principalmente por proteínas y ácidos nucleicos, se han ido perfeccionando con la evolución, volviéndose más

complejas, reguladas de forma sofisticada e integradas en vías operativas específicas. En el caso de las proteínas, las principales moléculas operativas de la vida, sus funciones dependen en gran medida de su forma tridimensional y su dinámica. Por tanto, es necesario conocer estos procesos a nivel atómico para comprender en profundidad el mecanismo de acción molecular, incluida su dinámica, lo que nos permitiría encontrar soluciones para su mal funcionamiento y, en última instancia, crear nuevas actividades para nuestro beneficio en biomedicina y biotecnología. En el caso de los ácidos nucleicos, el panorama estructural es mucho más diverso de lo que se pensaba. Los ADN no canónicos están surgiendo como estructuras clave en una gran variedad de procesos biológicos, como la transcripción del genoma, la reparación o el mantenimiento de los telómeros. Además, los transcritos de ARN, incluidos los ARN pequeños y largos no codificantes, parecen regular casi todos los pasos de la expresión genética y tienen amplias repercusiones en el desarrollo y la enfermedad.

Paradójicamente, la función de un número cada vez mayor de proteínas, regiones proteicas y también de ARN parece residir en su capacidad de permanecer desestructurados. Estas macromoléculas intrínsecamente desordenadas están implicadas, entre otras cosas, en la promoción de cambios en el estado físico del medio celular (separación de fases líquido-líquido), lo que permite la formación de compartimentos celulares sin membrana con múltiples fines, como el almacenamiento transitorio o las funciones de respuesta al estrés.

Otro nivel de complejidad es el análisis de cómo, dónde y cuándo se ensamblan y actúan de forma combinada estas máquinas macromoleculares. La célula puede considerarse una fábrica con múltiples compartimentos, a veces de corta duración, cuyo contenido macromolecular depende de las necesidades de la célula y debe controlarse cuidadosamente. Esta disposición subcelular de las macromoléculas y sus correspondientes funciones asociadas —denominada *sociología molecular* de la célula— tiene una finalidad que es preciso reconocer. Resulta especialmente interesante el papel de las membranas no solo como barreras para separar los componentes celulares, sino también como regiones de agrupación para funciones específicas en las que las proteínas de membrana tienen papeles clave. Un análisis estructural exhaustivo y comparativo de las proteínas, el ARN y los complejos macromoleculares que actúan en procesos similares en diferentes organismos proporcionará información esencial para reconstruir la historia y la evolución de la vida. En última instancia, este conocimiento nos permitirá diseñar nuevos objetos y entidades biológicas, así como aprovechar la biología sintética.

Un aspecto fundamental en los estudios evolutivos consiste en desentrañar la historia que hay detrás de la diversificación de la vida y poder organizar a los seres vivos en un árbol de la vida estructurado y con sentido, donde se establezcan las relaciones entre las especies. La llegada de la secuenciación de alto rendimiento permite ensamblar genomas a nivel cromosómico, caracterizar transcriptomas unicelulares y determinar modificaciones epigenómicas. Una vez que se aplique ampliamente a la diversidad de los organismos vivos, se logrará la reconstrucción del Árbol de la Vida y la identificación de las dianas genómicas de la selección natural. Los principales esfuerzos se centrarán en la obtención de muestras de las fronteras bióticas, en el tratamiento de genomas gigantes y de proporciones importantes de elementos repetitivos, en la identificación de tipos de homología y ploidía, en la detección de peculiaridades genómicas de selección, en la inferencia de funciones de genes candidatos y en la recopilación e incorporación de metadatos asociados a la historia natural, geológicos, ecológicos y ambientales a largo plazo en un marco filogenético. A largo plazo, deberíamos ser capaces de catalogar la biodiversidad, desvelar los mecanismos subyacentes a la adaptación evolutiva y dirigir nuestros esfuerzos de conservación basándonos en estas pruebas.

Conocidos los componentes celulares básicos y su aparición durante la evolución, la siguiente cuestión abierta estriba en comprender cómo estos sistemas moleculares y celulares básicos alcanzan una organización de nivel superior como individuos. Aquí tenemos que comprender el desarrollo, el mantenimiento y el declive de los seres vivos, es decir, la génesis del fenotipo. Un enfoque multidisciplinario basado en el marco de la evolución debe abordar cómo se almacena la información biológica y cómo se mantiene, se hereda y cambia, además de cómo esta información codificada impulsa la aparición del fenotipo, es decir, de todos los rasgos y procesos perceptibles de los sistemas vivos. También debemos entender hasta qué punto es predecible el fenotipo, con el fin de evaluar nuestra capacidad de ingeniería de los sistemas vivos. Por último, es primordial abordar cómo evolucionan los nuevos fenotipos y cuáles son las escalas temporales y espaciales de los cambios fenotípicos evolutivos.

*Implicaciones sociales, medioambientales y sanitarias de la evolución humana.* Las cuestiones que se plantean al estudiar la evolución adquieren un interés adicional cuando se aplican al ser humano. La evolución no solo puede ayudarnos a entender el origen de nuestra especie y las pautas históricas de migración y colonización de diversos hábitats en el mundo, sino que también

proporciona una nueva mirada a otros problemas aparentemente más lejanos, como la salud y la enfermedad.

Uno de los mayores desafíos en el estudio de la evolución humana es comprender los procesos adaptativos sociales y biológicos que tuvieron lugar a lo largo de nuestra historia como especie. En un marco temporal más profundo, la continua obtención de genomas completos de múltiples especies que ofrecen información sobre la filogenia de nuestro linaje permite estudiar con detalle cuándo y cómo aparecieron algunos rasgos definitorios del ser humano a lo largo de la evolución. La aplicación de las técnicas *ómicas* al estudio del pasado permite ahora a los investigadores abordar una serie de cuestiones que antes eran objeto casi exclusivo de disciplinas de las Humanidades, como la Historia y la Arqueología, al investigar las migraciones; o de la Medicina, como la Genómica Médica, al investigar la arquitectura genética de rasgos y enfermedades complejas. En un marco temporal más profundo, la continua obtención de genomas completos de múltiples especies que ofrecen información sobre la filogenia de nuestro linaje permite estudiar con detalle cuándo y cómo aparecieron algunos rasgos definitorios del ser humano a lo largo de la evolución.

En cuanto al estudio de la evolución cultural y del comportamiento humano, abordado principalmente mediante los estudios arqueológicos, los desafíos multidisciplinares deberían intentar redefinir el papel de la tecnología en la configuración de las sociedades del pasado, pero también en la transformación del entorno, así como la naturaleza de las interacciones entre los agentes bióticos y abióticos a lo largo de la evolución de nuestro género. Nuestro objetivo es identificar patrones, más que eventos, en el curso de la evolución conductual y cultural de nuestra especie.

Hasta qué punto los factores evolutivos bióticos frente a los abióticos dictan trayectorias divergentes en el pasado humano es una de las cuestiones principales en el estudio del comportamiento humano del pasado. Desde el papel del clima en las primeras adaptaciones humanas hasta su relevancia en la aparición de la producción de alimentos, el interés secular por la influencia de las causas abióticas en la evolución social está dando lugar a un cambio de perspectiva. Los nuevos grandes desafíos de la evolución social deberían poner de relieve, por ejemplo, la importancia de la estructura de las comunidades de mamíferos en la conformación de los primeros comportamientos humanos, o el impacto de las acciones humanas sobre la fauna y la flora y, en los últimos tiempos, incluso sobre el clima.

Aunque a menudo se descuida, el campo emergente de la medicina evolutiva está aportando nuevas respuestas a preguntas planteadas desde hace tiempo. Las enfermedades son el resultado de la perturbación de la homeostasis fisiológica de los organismos y pueden ser endógenas (por ejemplo, enfermedades hereditarias, limitaciones del desarrollo o cáncer) o exógenas (por ejemplo, infecciones o intoxicaciones). En la medida en que los propios organismos, y la forma en que interactúan con sus entornos bióticos y abióticos, son el resultado de fuerzas evolutivas, sus enfermedades son también el resultado de complejos procesos coevolutivos. Al incorporar una perspectiva evolutiva, entenderíamos y combatiríamos mejor las enfermedades. Esto puede aplicarse no solo al ser humano, sino también a los animales de granja y a las plagas de los cultivos. Algunas áreas importantes de investigación son el estudio de los orígenes y la propagación de nuevos agentes infecciosos, la aparición de microorganismos multirresistentes o la dinámica de las células cancerosas y el crecimiento de los tumores. Asimismo, el estudio de la evolución de las comunidades microbianas será de gran importancia, a la luz del papel del microbioma en múltiples aspectos de la salud y la enfermedad. Además, comprender el origen de patologías multigénicas como la diabetes, la obesidad, las enfermedades cardiovasculares, las enfermedades autoinmunes y, a largo plazo, el envejecimiento, será sin duda instructivo para diseñar nuevos enfoques terapéuticos.

Todos estos procesos se rigen por las mismas fuerzas básicas y universales de la evolución: la mutación, la deriva genética, la migración, la selección natural y las compensaciones adaptativas. Cada vez está más claro que las soluciones a estos complejos problemas implican potencialmente todos los niveles de organización, desde las moléculas hasta las poblaciones.

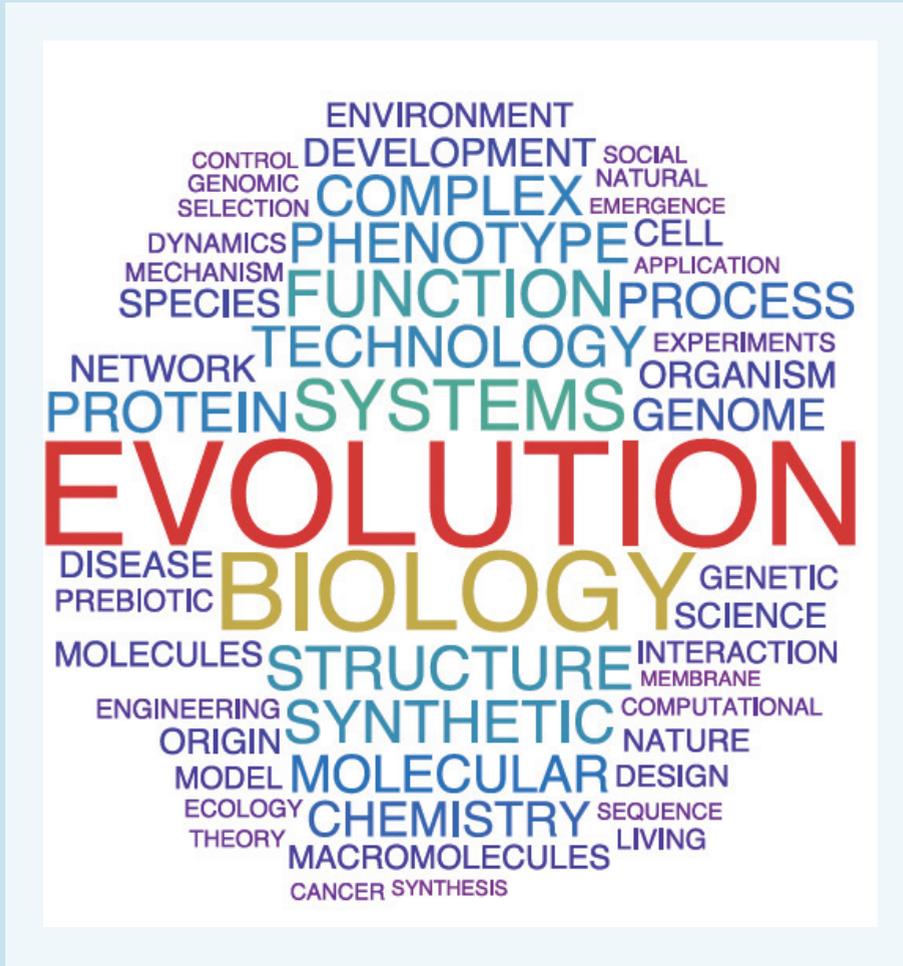
*La vida futura: ¿dominar la evolución?* Como ya hemos dicho, la evolución es un proceso complejo y a varios niveles que actúa a escalas de tiempo largas. Por tanto, es evidente que solo puede entenderse desde una perspectiva sistémica. Aunque disponemos de una gran cantidad de datos experimentales, el principal desafío consiste en desarrollar modelos y marcos teóricos para entender los resultados empíricos y plantear preguntas experimentales mejor enfocadas. Una de las primeras cuestiones sin resolver, como se ha descrito anteriormente, es cómo surgen los fenotipos a partir de los genotipos. ¿Pueden construirse funciones biológicas complejas a partir de módulos más simples? ¿Cómo surgen los circuitos reguladores? ¿Cuáles son los límites del diseño de módulos funcionales robustos y transportables? Las respuestas a estas preguntas evaluarán la validez de los enfoques reduccionistas, frente a

la consideración de la innovación como un fenómeno emergente, que surge de propiedades distribuidas en forma de red. En un marco más amplio, deberíamos preocuparnos por el origen mecanicista de las transiciones evolutivas y por el papel que desempeñan las fuerzas externas en contraposición a los fenómenos contingentes o estocásticos en su generación.

La propia evolución puede considerarse una herramienta para la biología sintética, ya que la selección evolutiva dirigida es una forma de alcanzar las funciones deseadas. Uno de los principales objetivos se refiere a integrar el diseño en la exploración de espacios fenotípicos impulsada por la selección. Los avances dependerán de la creación de grandes plataformas donde la evolución experimental informada por el diseño y la teoría pueda proceder de forma masiva. La investigación en las fronteras entre la ciencia evolutiva, el clima y la ecología aborda no solo cómo afectarán estos organismos al ecosistema, sino también cómo afectará el ecosistema a los organismos en un entorno que cambia rápidamente.

Por supuesto, las cuestiones planteadas con anterioridad repercuten de forma directa en los objetivos de la biología sintética, que consisten en ensamblar una unidad viva mínima con funcionalidad programable. La consecución de este gran desafío, la construcción de una célula sintética totalmente nueva, contribuirá a nuestra comprensión de los principios básicos de la vida y de su origen a partir de componentes sin vida. También proporcionará soluciones novedosas a problemas ambientales y sanitarios pendientes.

Todavía no entendemos cómo estas piezas interactúan de forma coordinada para desarrollar las funciones celulares. En esta línea, la generación de vida a partir de los componentes moleculares existentes en la Tierra primitiva es uno de los grandes enigmas de la vida y, por tanto, un gran desafío científico. La biología sintética ofrece nuevas estrategias para su resolución. Desde una perspectiva fundamental, la integración de módulos moleculares que den lugar a células sintéticas funcionales ayudará a desvelar los límites de la vida. A este respecto, podemos prever que la fusión de la biología sintética con la evolución molecular y celular puede llevarnos a la síntesis de células vivas desde cero, cuya función se ajustará mediante mecanismos evolutivos controlados. Otro enfoque novedoso que tiene en cuenta los principios evolutivos y la integración de unidades independientes más pequeñas es el de la robótica de enjambre. En pocas palabras, entender cómo la evolución produjo sistemas vivos como los que nos rodean, nos permitirá diseñar y generar sistemas vivos artificiales desde cero.



**Environment Development:** Medioambiente; **Development:** Entorno; **Control:** Control; **Genomic Selection:** Genómica; **Selection:** Selección; **Social:** Social; **Natural:** Natural; **Emergence:** Emergencia; **Complex:** Complejo; **Phenotype:** Fenotipo; **Dynamics:** Dinámica; **Mechanism:** Mecanismo; **Cell:** Célula; **Application:** Aplicación; **Species:** Especies; **Function:** Función; **Process:** Proceso; **Technology:** Tecnología; **Experiments:** Experimentos; **Organism:** Organismo; **Network:** Red; **Protein:** Proteína; **Systems:** Sistemas; **Genome:** Genoma; **Evolution:** Evolución; **Disease:** Enfermedad; **Prebiotic:** Prebiótico; **Biology:** Biología; **Genetic Science:** Ciencia Genética; **Molecules:** Moléculas; **Structure:** Estructura; **Interaction:** Interacción; **Membrane:** Membrana; **Engineering:** Ingeniería; **Origin:** Origen; **Model:** Modelo; **Ecology:** Ecología; **Theory:** Teoría; **Synthetic:** Sintético; **Molecular:** Molecular; **Chemistry:** Química; **Macromolecules:** Macromoléculas; **Cancer:** Cáncer; **Synthesis:** Síntesis; **Computational:** Computacional; **Nature:** Natural; **Design:** Diseño; **Sequence:** Secuencia; **Living:** Vida

*Un análisis cuantitativo de un esfuerzo colectivo.* Un enfoque cuantitativo mediante un análisis de nube de palabras (fig. 1), que identificó los términos más representados en la colección de los ocho capítulos que siguen, permite extraer algunas conclusiones generales sobre los conceptos más importantes destacados aquí. Como es de suponer, la evolución en biología es el tema principal de estos capítulos, lo que justifica la cita de Theodosius Dobzhansky que utilizamos para comenzar esta introducción («Nothing makes sense in biology except in the light of evolution» [nada tiene sentido en biología si no es a la luz de la evolución]).

Curiosamente, tras este tema principal, podemos ver con claridad los fundamentos y enfoques que pueden ayudar a entender la evolución en un sentido amplio. Así, estos capítulos o desafíos (C) han revelado que el enfoque probablemente más válido para entender el *problema de la evolución* debe ser por fuerza integrador y en múltiples escalas: desde el genoma hasta el organismo completo, pasando por sus relaciones ecológicas y sociales, e incluyendo todas las escalas intermedias (por ejemplo, moléculas, células, etc.). Esta idea queda plasmada y subrayada por el término *Sistemas* en la nube de palabras. Las razones que justifican este punto de vista metodológico son la complejidad y la intrincada red de interacciones entre componentes a diferentes escalas, que lleva a las relaciones entre estructura (en un sentido amplio) y función. Una segunda conclusión hace referencia a las herramientas necesarias para poner en práctica estos avances del campo: se necesitan nuevas tecnologías, enfoques computacionales y de ingeniería, así como marcos teóricos capaces de proponer modelos y mecanismos cuantitativos, en combinación con los enfoques de la biología genómica y sintética. Una tercera conclusión alude a las posibles aplicaciones derivadas de estos estudios. Por tanto, una comprensión profunda de la *evolución biológica* por estos medios es el camino para adquirir conocimientos y proponer soluciones sobre las enfermedades (por ejemplo, el cáncer), sus posibles impulsores ambientales y, finalmente, arrojar luz sobre el origen del fenómeno emergente que llamamos *vida*.

*Evolución en el CSIC.* Como se ha señalado anteriormente y se explora con más detalle en las páginas siguientes, solo el esfuerzo colectivo entre los investigadores pertenecientes a los tres dominios en los que se divide actualmente el CSIC —Vida, Materia y Sociedad— puede conducir a una mejor comprensión de la vida, su origen, su diversidad y su larga historia de coevolución, entendiéndose con este término que la evolución de cada organismo individual depende de la del entorno que lo rodea, incluidos otros organismos, aunque la

escala temporal pueda ser diferente. El CSIC se encuentra en una posición única entre los organismos de investigación de ámbito nacional para abordar el desafío de entender la evolución con todas sus implicaciones y ramificaciones. Asimismo, cuenta con centros de investigación especializados dedicados al estudio de la biodiversidad viva y fósil, múltiples institutos donde el estudio de los mecanismos moleculares que actúan durante la evolución es un objetivo principal, y centros específicos para el estudio de la evolución genómica o la biología de sistemas, así como otros que abordan las diferentes ramas de las matemáticas, la física, la química, las ciencias sociales y las humanidades. Sin embargo, esta organización fragmentada debería evolucionar al mismo ritmo que la investigación. Los mayores desafíos solo pueden abordarse desde una perspectiva multidisciplinaria. La participación de investigadores pertenecientes a los tres grandes dominios del CSIC en la composición de este tema es un claro reflejo de esta tendencia. Así, el CSIC cuenta con las herramientas, la mano de obra, el conocimiento y, lo que es más importante, ha identificado colectivamente los desafíos más destacados e inexplorados para las próximas décadas en el marco de este tema. El gran desafío que hay que abordar ahora es el de coordinar de forma eficaz la actividad de un gran número de investigadores con diferente formación y experiencia, crear los escenarios adecuados para su actividad, fomentar sin reparos los enfoques emergentes, aumentar la masa crítica, educar a los científicos del futuro y darles libertad e independencia para que exploren sus propias ideas desde el principio, y ofrecerles el apoyo financiero necesario para que sean competitivos a escala internacional.

La proximidad física y el intercambio continuo de ideas siguen siendo el mejor caldo de cultivo para la investigación innovadora. La creación de un centro dinámico de Biología Evolutiva en el que se desarrollen las actividades aquí propuestas debería figurar, pues, entre los objetivos a medio plazo de este gran desafío.



## DESAFÍO 1

### RESUMEN

El origen y la etapa inicial de evolución de la vida es uno de los temas científicos más desafiantes e interdisciplinarios. Su objetivo consiste en explicar la síntesis de unidades estructurales bioquímicas y su polimerización para dar lugar a los biopolímeros funcionales y portadores de información responsables de la bioquímica de la vida, así como su acoplamiento en sistemas autónomos con capacidad de evolución abierta. El CSIC tiene el potencial de contribuir activamente a este campo gracias a los muy relevantes trabajos que ya están desarrollando varios de sus grupos de investigación, a las sólidas complementariedades detectadas y a las sinergias que seguramente surgirán entre ellos, así como con sus colaboradores de universidades españolas, otros OPI y socios internacionales.

### PALABRAS CLAVE

astroquímica

quiralidad

sistemas complejos

Tierra primitiva

evolución

mapa genotipo-fenotipo

evolución *in vitro*

medio interestelar

LUCA

metabolismo

ARN no codificante

orígenes de la vida

química prebiótica

química de sistemas prebióticos

protocélula

replicación

mundo del ARN

ciencia de superficies

biología sintética

vesículas

viroides

virus

# LOS ORÍGENES DE LA VIDA. DE LA QUÍMICA A LA BIOLOGÍA

## Coordinadores

Carlos Briones, (CAB, CSIC/INTA)

Izaskun Jiménez-Serra, (CAB, CSIC/  
INTA)

## Investigadores y centros de investigación participantes:

Jacobo Aguirre, (CAB, CSIC/INTA)

Alfredo Berzal-Herranz, (IPBLN)

Pedro Cintas, (Univ. Extremadura)

Emilio J. Cocinero, (IBF,  
CSIC-UPV-EHU)

Elena R. Alonso, (IBF,  
CSIC-UPV-EHU)

Andrés de la Escosura, (UAM)

Juan Manuel García-Ruiz, (IACT,  
CSIC-UG)

Jordi Gómez, (IPBLN)

David Hochberg, (CAB, CSIC/INTA)

Marta Ruiz, (CAB, CSIC/INTA)

Eva Mateo-Martí, (CAB, CSIC/INTA)

Susanna Manrubia, (CNB)

José Ángel Martín-Gago, (ICMM)

César Menor-Salván, (Univ. Alcalá de  
Henares)

Juli Peretó, (I2SYSBIO, CSIC-UV)

Germán Rivas, (CIB)

Kepa Ruiz-Mirazo, (IBF,  
CSIC-UPV-EHU)

## RESUMEN EJECUTIVO

El origen de la vida es un proceso complejo que permitió la transición de la química a la biología en la Tierra, quizás incluso más allá de nuestro planeta. Múltiples mecanismos pudieron haber contribuido a la biogénesis, incluidas las reacciones sintéticas en entornos astrofísicos, la llegada de materia orgánica de meteoritos y núcleos cometarios, y los procesos geofísicos y geoquímicos que tuvieron lugar en la Tierra primitiva, hace unos cuatro mil millones de años. El descubrimiento de más de doscientas moléculas pequeñas en el espacio interestelar y en cuerpos extraterrestres sugiere que los procesos prebióticos básicos son omnipresentes en el universo. De hecho, se han encontrado compuestos de bajo peso molecular en asteroides (cuerpos parentales de meteoritos) y cometas: luego, estos cuerpos menores pudieron haber contribuido de forma decisiva al enriquecimiento del llamado *caldo primordial* con materia orgánica y moléculas que contienen CHONPS. La combinación de compuestos endógenos y exógenos en diferentes entornos de la Tierra primitiva dio lugar a un número creciente de biomoléculas, que pudieron interactuar a través de procesos de autoensamblaje y actividades catalíticas

que actúan en solución acuosa e interfaces reactivas (incluidas agua-hielo y agua-lípido), así como en superficies minerales y metálicas.

Las cuestiones abiertas en el campo de la química prebiótica incluyen lo siguiente: la síntesis abiótica de las moléculas orgánicas necesarias en condiciones fisicoquímicas plausibles; el origen de la homquiralidad de los biomoléculas actuales (por ejemplo, azúcares, nucleótidos y aminoácidos); el autoensamblaje y la polimerización de ciertos biomoléculas en polímeros más largos, como el ARN o los péptidos, en medios homogéneos o heterogéneos; la formación de vesículas de base anfifílica con suficiente permeabilidad; el surgimiento de funciones bioquímicas en microambientes compartimentados; el origen de las actividades metabólicas básicas y de la información heredable; la asociación supramolecular de las moléculas complejas generadas; el establecimiento de un código genético y el origen de los ribosomas, así como la disociación progresiva genotipo/fenotipo. Los primeros organismos celulares autorreproducibles y evolucionables deberían haber sido capaces de: i) mantener sus componentes moleculares unidos y distinguirse de su entorno; ii) mantenerse alejados del equilibrio termodinámico captando recursos energéticos y materiales del entorno; y iii) transmitir información heredable a su progenie. Así pues, los compartimentos de membrana, las maquinarias metabólicas y los mecanismos de replicación deben de haberse originado y combinado en la transición de sistemas químicos complejos (aunque todavía dirigidos termodinámicamente) a sistemas protobiológicos y, en última instancia, a organismos vivos (controlados de forma cinética y espacial).

La química prebiótica y la bioquímica se han ocupado de aspectos parciales de un problema tan complejo. Sin embargo, los mecanismos fisicoquímicos implicados en la formación de esos subsistemas infrabiológicos, cuando se aplicaron y desarrollaron por separado, resultaron a menudo incompatibles. Como se discute más adelante, niveles más altos de heterogeneidad molecular, junto con el uso de enfoques basados en la química de sistemas, proporcionan un escenario más realista de los orígenes de la vida. En este marco, la controversia tradicional entre la replicación (es decir, el modelo del mundo del ARN) y el metabolismo (autotrófico o heterotrófico) puede sustituirse por un escenario en el que las moléculas clave (incluidos los monómeros y oligómeros, los catalizadores metálicos y minerales o las interfaces reactivas con medios de base acuosa) pudieran coevolucionar desde el principio, formando redes de interacción mixtas y prebioquímicas.

Las implicaciones de este nuevo enfoque se describirán en este capítulo, tanto en lo que respecta a las unidades individuales (como sistemas automantendidos con una organización interna) como en relación con su dinámica evolutiva colectiva y a largo plazo (basada en los procesos de competencia, colaboración y selección entre esos individuos, incluidas las redes protoecológicas y las relaciones sintróficas que deben haber establecido). Este enfoque interdisciplinario pueden llevarlo a cabo los grupos del CSIC y sus socios en este desafío científico.

## 1. INTRODUCCIÓN Y DESCRIPCIÓN GENERAL

### **De la astroquímica a la astrobioquímica: la búsqueda de moléculas prebióticas en el espacio**

El espacio entre las estrellas no está vacío, sino que contiene un gran número y diversidad de átomos, núcleos de alta energía, partículas de polvo de tamaño (sub)micrométrico y moléculas. El estudio de su abundancia y de las reacciones que se producen en el espacio, y su interacción con la radiación, se denomina Astroquímica. Se trata de una nueva disciplina a medio camino entre la Astrofísica y la Química que comenzó con la detección de las primeras moléculas en el espacio (los radicales diatómicos CH y CN y el ion molecular CH<sup>+</sup>) a finales de la década de 1930 con telescopios ópticos. Hubo que esperar hasta finales de los años sesenta para descubrir otras especies moleculares como el OH, el H<sub>2</sub>O y el NH<sub>3</sub> en longitudes de onda centimétricas gracias al desarrollo de los radiotelescopios. Cincuenta años más tarde, se han registrado hasta la fecha más de 200 especies moleculares en el espacio, lo que revela que *nuestro universo es molecular* (véase la Cologne Database for Molecular Spectroscopy, CDMS, para consultar un inventario actualizado de las especies moleculares detectadas en el espacio).

La mayoría de estas moléculas son orgánicas y contienen abundantes elementos interestelares como H, C, N y O. Se encuentran en diferentes entornos (es decir, nubes oscuras frías, núcleos calientes, envolturas circunestelares alrededor de estrellas evolucionadas, etc.) mediante el análisis de la emisión de líneas moleculares en sondeos espectroscópicos realizados en longitudes de onda centimétricas, milimétricas y submilimétricas. Estos sondeos están tremendamente saturados de líneas, conocidas como *malezas astrofísicas*, que pertenecen a las numerosas moléculas presentes en el medio interestelar (MIE).

Una fracción significativa de las especies detectadas en el espacio corresponde a moléculas grandes y pesadas —lo que, en este contexto, significa que tienen más de cinco átomos en sus estructuras— denominadas *moléculas orgánicas complejas* (MOC). La formación y supervivencia de estas grandes moléculas resulta algo sorprendente. El MIE y el espacio que rodea a las estrellas (el medio circunestelar o MCE) son entornos adversos para las moléculas en general, y para las MOC en particular, porque están expuestos a fenómenos energéticos como la radiación UV/rayos X, los rayos cósmicos o las ondas de choque.

Entre las MOC detectadas en el MIE, hay un subconjunto de especies relevantes para la química prebiótica y para las teorías sobre los orígenes de la vida (Patel *et al.*, 2015). Entre ellas se encuentran la cianamida ( $\text{NH}_2\text{CN}$ ), el cianoacetileno ( $\text{HC}_3\text{N}$ ), el glicolaldehído ( $\text{HOCH}_2\text{CHO}$ ), la formamida ( $\text{NH}_2\text{CHO}$ ) o la urea ( $\text{NH}_2\text{CONH}_2$ ), que hoy en día se detectan de forma habitual en el MIE (por ejemplo, Belloche *et al.*, 2008; Jorgensen *et al.*, 2012; Jiménez-Serra *et al.*, 2020). A pesar de este enorme progreso, las unidades estructurales de la vida terrestre, como los aminoácidos (por ejemplo, la glicina), las nucleobases (como la adenina) o los azúcares relativamente complejos (por ejemplo, la ribosa), siguen sin aparecer en el MIE. Sin embargo, la presencia en el espacio de aminoácidos o monosacáridos precursores (Cocinero *et al.*, 2012; Haykal *et al.*, 2013; Peña *et al.*, 2013; Alonso *et al.*, 2019; Calabrese *et al.*, 2020), y de sus moléculas prebióticas (Alonso *et al.*, 2016), es de suma importancia para entender los procesos químicos que pudieran llevar a la formación de las unidades estructurales de la vida, haciendo que su detección en el MIE sea casi tan importante como la de los propios aminoácidos o monosacáridos.

### **Reacciones geoquímicas y nichos geológicos para el origen de la vida**

El propio nombre que los geólogos han dado a los primeros 500 millones de años de la Tierra, el eón hádico (nombre griego que hace referencia al inframundo) se debe a que se creía que en sus primeros años el planeta era un infierno, un lugar inhóspito de condiciones extremas, de intensa radiación ultravioleta (UV), alta temperatura, sin agua líquida e innumerables volcanes y mares de magma. Sin embargo, esta opinión ha cambiado drásticamente en los últimos años gracias a la información deducida de algunos extraños cristales de circón conservados. La Tierra y los planetas y lunas similares a la Tierra presentaban las condiciones químicas y físicas necesarias para que la vida floreciera mucho antes de lo que se pensaba, es decir, hace unos 4400 millones de años (Wilde *et al.*, 2001). Esto es importante cuando se consideran las

escalas de tiempo de los diferentes procesos implicados en la química prebiótica.

Desvelar cómo era un planeta sin vida en aquella época es una de las tareas ineludibles en el campo de los orígenes de la vida. El agua se condensó en la superficie del planeta poco después de la solidificación de la primera corteza ultramáfica. La interacción impulsada por la temperatura entre el agua y los minerales ultramáficos (serpentinización) creó inevitablemente una hidrosfera alcalina y reducida, y una atmósfera rica en metano. En estas condiciones fisicoquímicas, el planeta debió de ser una fábrica global de compuestos orgánicos simples y complejos (García-Ruiz *et al.*, 2020). La cuestión de por qué no hay pruebas de la existencia de vida en otros lugares puede deberse entonces a las dificultades para sintetizar los microorganismos más simples, y no tanto a las unidades estructurales químicas iniciales, como se explica a continuación.

### 1.3. Química prebiótica

La química prebiótica abarca los procesos físico-químicos que ocurren dentro de un determinado entorno planetario, desde su formación hasta la aparición de los primeros sistemas autorreplicantes sobre los que puede comenzar a actuar el proceso darwiniano (Eschenmoser, 2007; Ruiz-Mirazo *et al.*, 2014; Pross, 2016). Algunas de las principales cuestiones abiertas son: i) el origen abiótico de las moléculas orgánicas necesarias (el principal campo de la química prebiótica); ii) el origen de la homquiralidad de todos los compuestos bioorgánicos actuales, ya que actualmente se acepta que es imposible construir ningún biopolímero funcional utilizando mezclas de compuestos enantioméricos; iii) la polimerización de monómeros, el autoensamblaje molecular y la reactividad en superficies; iv) la asociación supramolecular de las moléculas complejas generadas y la formación de vesículas, en el marco de la química de sistemas prebióticos (véase más adelante).

La hipótesis comúnmente aceptada para explicar el origen de la vida es que las moléculas inorgánicas simples (por ejemplo,  $\text{CH}_4$ ,  $\text{CO}_2$ ,  $\text{N}_2$ ,  $\text{NH}_3$ , u otras moléculas de un solo C como el HCN, la formamida o la urea) reaccionaron para formar compuestos orgánicos más complejos (Marín-Yaseli *et al.*, 2016; Mompeán *et al.*, 2019) o incluso sustancias oligoméricas/poliméricas (Mann, 2013; Ruiz-Bermejo *et al.*, 2019) que sufrieron reacciones posteriores para dar lugar a moléculas biológicamente funcionales. Estos compuestos se autoorganizan y autoensamblan para aumentar la organización a nivel molecular.

Dicha sucesión de procesos químicos se considera la vía más probable para la aparición de los primeros organismos vivos. Por lo tanto, el primer paso para verificar esta teoría sería comprobar cómo los componentes de las proteínas y los ácidos nucleicos, o los constituyentes de otros polímeros protobiológicos, pudieran formarse en condiciones abióticas. Entre los escenarios fisicoquímicos relevantes para que la química prebiótica funcionara, tanto los *pequeños estanques calientes* como los sistemas hidrotermales pudieron haber desempeñado un papel importante. Además, los productos orgánicos e inorgánicos de la superficie oceánica pudieron transmitirse en aerosoles a la atmósfera, donde se expusieron a la radiación y a las descargas eléctricas en forma de rayos, con lo que contribuyeron así al origen de las moléculas orgánicas en condiciones abióticas.

Además, entender cómo surgió una quiralidad *única* en la Tierra (es decir, los L-aminoácidos y D-azúcares homoquirales en los biopolímeros) durante la evolución molecular es una característica decisiva de cualquier propuesta científicamente coherente. La teoría de la evolución química, por consiguiente, no solo debe dar cuenta de la transición de la química prebiótica (es decir, un conjunto diverso de unidades estructurales aquirales o racémicas relativamente simples) a la protobiología (entidades supramoleculares como las proteínas, los polisacáridos, los ácidos nucleicos y los lípidos en bicapa, capaces de interactuar con otros, con autoorganización y evolución darwiniana) (Breslow y Cheng, 2009; Mann, 2013; Krishnamurthy, 2017), sino que también debe proporcionar respuestas a cuestiones pendientes como: i) ¿cuándo surgió la homoquiralidad biológica?; ii) ¿por qué existe un sesgo quiral para los L-aminoácidos y los D-azúcares?; iii) ¿cómo se formaron los aminoácidos y los azúcares quiralizados no racémicos a partir de moléculas aquirales simples en ausencia de catalizadores asimétricos?

Las respuestas a estas preguntas requieren la aplicación de la química física, la física química y la termodinámica del no equilibrio para comprender la ruptura de la simetría quiral y la amplificación quiral en sistemas de reacción modelo plausibles desde el punto de vista prebiótico. Esto implica el estudio de las reacciones autocatalíticas, los sistemas químicos autorreplicantes y los procesos de ruptura de la simetría quiral, como precursores del origen de la homoquiralidad biológica (Ribó *et al.*, 2017). Igualmente importantes son las condiciones impuestas por la termodinámica de los sistemas alejados del equilibrio y, especialmente, los aspectos entrópicos de la ruptura de la simetría de reflexión molecular y la aplicación de los criterios de producción de máxima/

mínima entropía como principios de selección para estas redes químicas no lineales (Hochberg y Ribó, 2019; Hochberg y Cintas, 2020).

Las interacciones y la reactividad entre las biomoléculas y las superficies minerales son también piezas importantes en el rompecabezas de la química prebiótica, ya que los minerales pueden adsorber y concentrar dichas biomoléculas actuando como catalizadores de las reacciones bioquímicas. Además de la teoría ampliamente aceptada del *caldo primordial* o *sopa prebiótica*, Huber y Wachtershauser propusieron la teoría del *mundo de hierro y azufre*, que afirma que las primeras reacciones que condujeron a la formación de los aminoácidos no se produjeron en una solución madre en los océanos, sino en la superficie de los minerales (como la pirita), ya que facilitan la polimerización prebiótica y las reacciones químicas entre las moléculas orgánicas (Huber y Wachtershauser, 1998). Las superficies minerales pudieran permitir casi cualquier tipo de catálisis, incluso las de baja especificidad y eficiencia. Sin embargo, hay que estudiar con detalle las condiciones ambientales que favorecen la adsorción de minerales moleculares, las reacciones que pueden intervenir en un sistema tan heterogéneo y la naturaleza precisa de las interacciones entre minerales y moléculas. Las técnicas de la ciencia de superficies son herramientas fiables para abordar el estudio de los procesos moleculares en las superficies de los minerales catalíticos (Sánchez-Arenillas y Mateo-Martí 2016; Gálvez-Martínez *et al.*, 2019; Mateo-Martí *et al.*, 2019). Las simulaciones de los procesos de polimerización utilizando moléculas pequeñas y sencillas, especialmente en condiciones que se creen plausibles en la Tierra primitiva y en otros planetas y lunas, meteoritos o cometas, son de interés intrínseco para entender la evolución química hacia la biología (Lavado *et al.*, 2018; Ruiz-Bermejo *et al.*, 2019).

En paralelo, la síntesis sobre la superficie ha surgido recientemente como un enfoque prometedor para obtener nanoestructuras carbonosas o moléculas complejas con precisión atómica. Tanto los polímeros unidimensionales como las intrincadas estructuras moleculares se han sintetizado directamente sobre superficies (Martín-Gago *et al.*, 2011; Méndez *et al.*, 2011). Este método se basa en la comprensión precisa de las interacciones entre las moléculas adsorbidas y la superficie de algunos materiales, que catalizan reacciones covalentes de acoplamiento. Uno de los méritos de la síntesis en superficie se centra en la capacidad de abrir nuevas vías de reacción que no son posibles mediante la química orgánica estándar. Este novedoso enfoque pudiera proporcionar una vía adicional, complementaria a la química húmeda sintética

propuesta, para explicar, por ejemplo, la polimerización abiótica de ácidos nucleicos a partir de nucleótidos, con la necesidad de su activación química previa o sin ella (Ferris, 2006). Esto tiene importantes implicaciones en el origen del mundo del ARN.

#### **1.4. El camino hacia un mundo del ARN**

El descubrimiento de que ciertas moléculas de ARN específicamente plegadas realizan funciones catalíticas (que actúan como enzimas de ARN o *ribozimas*) puso sobre el tablero que el ARN puede combinar genotipo y fenotipo en una sola entidad molecular. Este notable hallazgo reforzó la hipótesis de que el mundo actual de los organismos vivos, basado en el flujo de información genética ADN→ARN→proteínas, fue precedido por una época en la que el ARN era el único material genético y catalizador (Atkins *et al.*, 2011; Ruiz-Mirazo *et al.*, 2014).

Por ello, en las últimas décadas, los estudios experimentales de las raíces químicas de los sistemas bioquímicos han estado dominados por una perspectiva centrada en el ARN. En estos estudios, la construcción del alfabeto molecular del ARN se ha centrado principalmente en la química del cianuro de hidrógeno (HCN) o de la formamida (CHO-NH<sub>2</sub>, un producto de hidratación del HCN), como posibles precursores prebióticos de los ácidos nucleicos (Oró y Kimball, 1961). Sin embargo, quedan por resolver cuestiones esenciales como estas: i) ¿cómo se formaron las unidades estructurales de los biopolímeros (nucleótidos y aminoácidos)?; ii) ¿cómo se formaron los primeros nucleósidos y cuál fue su destino en el entorno prebiótico de la Tierra?; iii) ¿cómo se incorporaron los fosfatos u otros grupos enlazadores plausibles a la evolución química (el llamado *problema de los fosfatos*)?; iv) ¿cómo se formaron los protobiopolímeros que dieron lugar a organizaciones supramoleculares que permitieron la aparición de un sistema capaz de evolucionar de forma darwiniana?; y v) ¿cuáles son los escenarios plausibles en los que la interacción entre la química prebiótica y la geoquímica local permitió la formación y evolución química de los biopolímeros?

En este campo, uno de los mayores desafíos consisten en comprender cómo se produjo la polimerización de las primeras estructuras de proto-ARN, que condujo a la organización supramolecular, la selección y la evolución hacia agregados macromoleculares con funciones bioquímicas esenciales, como los protorribosomas. El origen del código genético y de los ribosomas completos permitió que la información codificada en el ARN se tradujera en proteínas,

con lo que desvinculó así el fenotipo del genotipo y se desencadenó la amplia exploración del espacio funcional (incluidas las actividades catalíticas necesarias para establecer un protometabolismo) mediante un número creciente de familias de proteínas (Ruiz-Mirazo *et al.*, 2017).

Sin embargo, la información que codifica las proteínas solo representa una pequeña parte de la carga genética de los organismos vivos actuales. De hecho, se ha identificado una cantidad abrumadora de ARN funcionales (denominados *ARN no codificantes* o ARNnc) en las células actuales, lo que demuestra que la plasticidad funcional del ARN ha perdurado hasta la vida moderna. Como ejemplo relevante, el ARN cataliza la formación de enlaces peptídicos en los ribosomas y proporciona el andamiaje para la transferencia de información del ARN a las proteínas, lo que respalda la hipótesis de que el ARN precedió a las proteínas en la evolución, aunque los péptidos —y varias moléculas de bajo peso molecular— pudieron haber modulado y complementado las propiedades catalíticas del ARN (Ruiz-Mirazo *et al.*, 2014).

Es más, los ARNnc representan una pieza fundamental en el proceso de regulación de la expresión génica en la mayoría de los organismos vivos (Atkins *et al.*, 2011). Asimismo, se han producido moléculas de ARN con diferentes capacidades a partir de poblaciones complejas mediante experimentos de selección/evolución *in vitro*, lo que dio lugar a ribozimas naturales tanto artificiales como de ingeniería (Puerta-Fernández, 2003; Joyce, 2004), así como a aptámeros con las actividades de unión a dianas deseadas (Briones, 2015) que muestran una creciente aplicabilidad en biotecnología y biomedicina (por ejemplo, Sánchez-Luque *et al.*, 2014; Moreno *et al.*, 2019a). Esta metodología imita los procesos evolutivos naturales en un tubo de ensayo, en condiciones controladas experimentalmente y ha ampliado en gran medida la versatilidad funcional del ARN y su papel en el origen de la vida (Joyce y Szostak, 2018).

La función del ARN es una consecuencia intrínseca de su plegado tridimensional en solución, que a su vez depende del mapa degenerado secuencia-estructura del ARN (diferentes secuencias pueden plegarse en la misma estructura secundaria y terciaria, como se verá más adelante). Además, algunos dominios estructurales de ARN, incorporados a través de la evolución en moléculas de ARN mucho más grandes y versátiles (que pueden contener regiones codificantes y no codificantes) pueden realizar algunas funciones específicas. En cambio, moléculas de ARN cortas estructuradas de forma independiente pueden haberse unido y desempeñar un nuevo papel catalítico o regulador (Briones *et al.*, 2009). Los genomas de ARN virales son buenos

ejemplos de esa versatilidad funcional, puesta a punto a través de la evolución. Codifican dominios de ARN estructurales/funcionales que establecen una estrecha comunicación entre ellos para dar lugar a una red de interacciones ARN-ARN compleja y dinámica, el llamado *interactoma del ARN*, que asegura la finalización del ciclo viral dentro de la célula huésped, así como la adaptación del virus al entorno celular (Romero-López y Berzal-Herranz, 2020).

Se han utilizado diferentes métodos bioquímicos y biofísicos para buscar elementos de ARN antiguos que hayan conservado su estructura secundaria o terciaria específica en los ARN genómicos virales y los ARNm celulares. Este enfoque se ha definido como *arqueológico*, ya que puede descubrir patrones evolutivos ocultos a través de una estrategia no filogenética y no representativa (Ariza-Mateos *et al.*, 2019). En particular, se han encontrado elementos similares al ARNt en posiciones estructurales o funcionalmente relevantes tanto en el ARN viral como en ciertos ARNm examinados. La concatenación basada en la ligadura (Briones *et al.*, 2009) de estos motivos ARN (entre otros) pudieron haber ocurrido en los grupos de ARN ya presentes en el mundo del ARN (Witzany, 2020). La amplia alteración de las secuencias de nucleótidos que con toda probabilidad desencadenó la transición de los predecesores de los ARN codificantes a los primeros ARNm plenamente funcionales (lo que no ocurrió en la construcción escalonada de los ARNn) dificulta la identificación basada en la filogenética de los elementos de ARN que pudieron haber estado activos antes de la llegada de la síntesis de proteínas. Por tanto, el método arqueológico es una forma de profundizar en la versatilidad estructural/funcional de aquellos elementos de ARN que tuvieron que adaptarse —perdiendo su función original— a las presiones selectivas que actuaron en la evolución desde el mundo del ARN a un mundo ARN-proteína, en el que se favoreció la capacidad codificante de los ARNm progresivamente más largos (Ariza-Mateos *et al.*, 2019).

### **1.5. La química de sistemas prebióticos como nuevo enfoque en la investigación de los orígenes de la vida**

La química de sistemas representa una nueva estrategia de investigación en las ciencias moleculares y abarca una visión holística de los sistemas químicos complejos, entendidos como conjuntos de moléculas diversas interconectadas mediante procesos de transformación o autoensamblaje. Este campo está llamado a proporcionar conocimientos clave para la resolución de importantes cuestiones científicas abiertas, como la forma en que podrían surgir entidades vivas a partir de la materia inerte (en relación con la investigación

sobre los orígenes de la vida) y si es posible diseñar procesos y materiales artificiales similares a la vida (en relación con el campo de la biología sintética y la nanotecnología). Con este objetivo, resulta habitual estudiar la fusión de diferentes unidades estructurales biológicas en sistemas sintéticos, con propiedades derivadas de la combinación de sus componentes biomoleculares. Normalmente, este enfoque se basa en la autoorganización o en los procesos químicos que son dinámicos por naturaleza, así como en observar de cerca los efectos quirales de esos componentes para alcanzar morfologías específicas. El objetivo a largo plazo de esta estrategia consiste en crear redes y conjuntos químicos con propiedades emergentes características de la vida (Morales-Reina *et al.*, 2020).

En el contexto de la investigación sobre los orígenes de la vida, el concepto de evolución química es fundamental, ya que abarca los mecanismos fisicoquímicos plausibles por los que pudieron haberse ensamblado las primeras protocélulas vivas. Históricamente, este término comenzó a utilizarse poco después de que se dieran los primeros pasos en el campo de la química prebiótica, aunque su uso ha cobrado una energía renovada en los últimos años, gracias a la aparición del campo de la química de sistemas prebióticos (Ruiz-Mirazo *et al.*, 2014). La opinión general es que, para entender la transición de la materia inanimada a los organismos vivos, hay que abarcar la complejidad a nivel químico. Actualmente suponemos que las primeras entidades vivientes deben haber estado compuestas, al menos, por un compartimento protocelular, un protogenoma y una red metabólica autocatalítica capaz de respaldar al sistema con moléculas de energía y sustrato. Además, la dinámica de replicación de estos tres subsistemas debe haberse acoplado para la reproducción eficaz del sistema en su conjunto. Tales requisitos implican un gran nivel de complejidad, tanto en lo que respecta a la estructura molecular de los componentes de la protocélula como a su dinámica de interacción, cuyo establecimiento parece muy improbable en ausencia de una fuerza motriz evolutiva (de la Escosura, 2019). De hecho, el desarrollo de poblaciones de protocélulas altamente dinámicas e integradas (en lugar de complejas redes de reacción en solución en masa, conjuntos de moléculas replicantes artificialmente evolucionables, o incluso estas mismas moléculas replicantes encapsuladas en compartimentos pasivos) proporciona el marco evolutivo más apropiado para abordar el difícil problema de cómo la química prebiótica tendió un puente hacia los organismos vivos completos (Shirt-Ediss *et al.*, 2017).

## 1.6. Aparición de la función e innovación molecular en un mundo precelular

Una de las principales dificultades en el camino desde los compuestos químicos simples hasta la primera célula consiste en la aparición de moléculas funcionales, que son ladrillos básicos que pueden, mediante un proceso de ensayo y error, permitir la formación de conjuntos más complejos dotados de un repertorio funcional más amplio. Los cinco conceptos esenciales en la aparición de las primeras entidades moleculares autorreplicantes pueden resumirse de la siguiente manera:

- a. El sesgo fenotípico refleja la redundancia del mapa genotipo-fenotipo. Las secuencias, en forma de polímeros aleatorios, pueden explorar uniformemente el espacio de los genotipos; sin embargo, los fenotipos, entendidos aquí como estructuras moleculares (y posibles funciones), no se muestrean del mismo modo. El hecho de que ciertas estructuras sean más probables que otras (por ejemplo, las secuencias de ARN cortas y abiertas al azar se pliegan en estructuras tipo horquilla; si son circulares, adoptan preferentemente pliegues de tipo varilla) puede estar detrás de la aparición de funciones como la ligadura (Briones *et al.*, 2009) o de replicadores de tipo viroide (Catalán *et al.*, 2019).
- a. La elevada dimensionalidad de los espacios de secuencias conlleva la existencia de redes de genotipos cuasineutrales y astronómicamente grandes que garantizan una navegación que no cuesta nada y una exploración eficiente de nuevas funciones moleculares (Manrubia *et al.*, 2020). Un efecto secundario del sesgo fenotípico es que la gran mayoría de las secuencias se mapean en fenotipos grandes y frecuentes: estos fenotipos abundantes son parsimoniosamente alcanzables si la evolución comienza en cualquier otro fenotipo viable. Por lo tanto, una vez que se encuentra por casualidad una función básica, la topología no trivial del mapa genotipo-fenotipo casi asegura la aparición de nuevas funcionalidades.
- b. Sin embargo, la aparición de una función no implica que esta se fije. Para que una nueva función resulte destacada, debe existir un nicho ecológico vacío. En ocasiones, el nicho molecular puede estar implícitamente disponible (como pudiera ser el caso de la ligadura, la escisión o la replicación una vez que los polímeros funcionales están presentes), mientras que en otras puede estar ausente y aparecer como una nueva capa de complejidad que depende de funciones previamente fijadas (como en la construcción del nicho). Un ejemplo pudiera ser los parásitos moleculares, incluidos los virus y los viroides: replicadores

egoístas que se aprovechan de la maquinaria de autorreplicación, pero no le devuelven nada al sistema. Una vez que los parásitos están en su lugar, los nuevos parásitos pueden tener dificultades para desplazar a los colonizadores iniciales y oportunistas. Las funciones que aumentan la complejidad, o la capacidad del sistema para autorreplicarse, están sujetas a una dinámica similar. El resultado de estos procesos por orden de llegada en la ecología molecular son los conocidos en la evolución como *accidentes congelados*: soluciones subóptimas cuya única ventaja es haber llegado primero. La llegada de la función está profundamente relacionada con el sesgo fenotípico (Schaper y Louis, 2014).

- b. Un mecanismo a menudo olvidado que favorece la adaptación y la aparición de nuevas funciones es la promiscuidad molecular. Esto resulta especialmente cierto en el caso del ARN, ya que una misma secuencia puede adoptar de forma fácil diferentes conformaciones (y, posiblemente, diferentes funciones) cuando el entorno físico-químico cambia. Algunos ejemplos son los interruptores de ARN, que pudieran ser un mecanismo importante en las primeras etapas de la innovación (Schultes y Bartel, 2000). Las funciones secundarias son una fuente potencial de innovación y, al mismo tiempo, confieren robustez al sistema al realizar, en caso necesario, la función primaria de una molécula diferente. Esto también puede entenderse como un caso de mimetismo molecular, una característica explotada por ciertos viroides que se disfrazan de moléculas de doble cadena de ADN para replicarse por medio de una ADN polimerasa.
- c. Un último mecanismo se basa en las interacciones sinérgicas entre las moléculas funcionales. Resulta difícil exagerar la importancia de la transferencia génica horizontal (TGH) cuando se habla de la construcción del genoma, así como el papel que la modularidad pudo haber desempeñado en el origen de las funciones complejas (Briones *et al.*, 2009). Sin embargo, una situación más simple, que no requiere la asociación estable de genes o módulos, pudiera surgir cuando múltiples funciones están cocirculando en un entorno abierto. Un ejemplo paradigmático de estas asociaciones transitorias pudieran ser los virus multipartitos existentes y las asociaciones virus-satélite (Lucía-Sanz y Manrubia, 2017). En el contexto de los orígenes de la vida, estas asociaciones pudieron haber sido mucho más frecuentes y laxas, ya que la arquitectura jerárquica de múltiples capas que posteriormente se fomentó con la aparición de los consorcios de microorganismos, los metazoos y los ecosistemas en general aún no estaba en marcha.

### **1.7. El papel de la teoría de la complejidad en la investigación de los orígenes de la vida**

Aunque la aplicación de las herramientas asociadas a la teoría de la complejidad ha sido muy fructífera en otros contextos, su potencialidad para desvelar las primeras etapas de la vida no se ha evaluado lo suficiente. Esta novedosa visión requiere un enfoque interdisciplinario, en el que las herramientas teóricas tomadas de la combinación de la teoría de redes complejas y la teoría de juegos se refuerzan con trabajos numéricos y experimentos *in vitro*.

Ya se ha desarrollado un amplio trabajo informático en el contexto de la *vida digital* (Wilke y Adami, 2002). El objetivo consistía en modelar el modo en que diferentes organismos digitales luchan por los recursos, mutan y se adaptan, imitando así la vida real. Esta perspectiva ha sido especialmente útil para arrojar luz sobre lo que Stephen Jay Gould denominó *reproducir la cinta de la vida* repetidamente para averiguar qué tipo de biosfera se obtendría en cada caso. Cuestiones como si la vida es una consecuencia inevitable de la interacción química entre los ingredientes del caldo primordial, o si es un acontecimiento extremadamente raro, pueden beneficiarse de este enfoque. Sin embargo, la mayoría de los trabajos realizados han sido estrictamente numéricos y se han basado en plataformas informáticas en las que los organismos digitales compiten por recursos como la memoria y el tiempo de la CPU.

Por otra parte, la combinación de la teoría de las redes complejas, la teoría de los juegos y la vida digital sigue estando enormemente inexplorada, tanto desde el punto de vista teórico como numérico. La medida en que este enfoque combinado puede ampliar nuestro conocimiento y comprensión de los orígenes de la vida (no solo en la Tierra, sino en un entorno genérico) cumpliendo los requisitos básicos para su aparición es una cuestión abierta.

### **1.8. Pistas de la biología sintética**

La biología sintética implica el diseño y la síntesis racionales de sistemas complejos basados en la biología o inspirados en ella que muestran funciones útiles, incluso las que no existen en la naturaleza (Ausländer *et al.*, 2017). Este innovador enfoque permitiría encontrar un segundo ejemplo de vida, mediante la síntesis química de una célula artificial (Solé, 2016). En este sentido, la célula viva se considera una fábrica altamente compleja equipada con máquinas químicas que realizan una variedad de tareas, las cuales pueden diseñarse para originar sistemas novedosos con una funcionalidad programable. Uno de los principales objetivos de la biología sintética es identificar una

configuración mínima de una célula biológica capaz de cumplir sus tareas más importantes: crecimiento, replicación, metabolismo y producción de energía. En esta línea, la pregunta sobre los orígenes de la vida se aborda mediante un enfoque experimental ascendente (complementario al enfoque descendente, que modifica las células preexistentes y actuales) que implica la ingeniería inversa de una función biológica concreta utilizando un conjunto mínimo de componentes moleculares naturales, modificados o sintéticos en ausencia de células (Fletcher 2016; Schwille *et al.*, 2018; Yewdall *et al.*, 2018.).

La construcción de tales células mínimas o artificiales mediante un diseño dirigido contribuirá a completar nuestra comprensión de cómo las unidades estructurales de las células modernas actúan de forma colectiva para la transición a un sistema vivo. Dado que un requisito básico para el origen de la vida consiste en la compartimentación, se han realizado esfuerzos considerables para establecer compartimentos de tamaño celular en forma de gotas y vesículas lipídicas controlables, gracias a una variedad de tecnologías como la microfluídica (Supramanian *et al.*, 2019). También se han logrado avances significativos en la reconstitución ascendente y la comprensión cuantitativa de las maquinarias celulares esenciales (incluido el procesamiento del ADN, la división celular, los patrones proteicos espaciales autoorganizados o los sistemas de expresión génica sin células) en contenedores similares a las células (Schwille *et al.*, 2018; véase el Desafío 8).

Todos estos rápidos avances, junto con los de la química de sistemas prebióticos y la nanotecnología, nos permiten abordar el problema de la integración de sistemas individuales básicos hasta un nivel de complejidad (actualmente desconocido) en el que puedan surgir propiedades similares a la vida (Schwille, 2017; Stano, 2018). Para superar esta barrera de complejidad, los estudios teóricos y computacionales serán esenciales para guiar el progreso experimental. Las ciencias de la ingeniería también deben participar, ya que tienen mucho éxito en la disección de sistemas complejos en un conjunto controlable de módulos funcionales. La aplicación de estos enfoques a la complejidad celular permitirá recopilar los elementos básicos que pueden reunir las funciones bioquímicas esenciales (Fletcher, 2016). En el campo de los orígenes de la vida, el principal desafío será comprender cómo los módulos funcionales se autoorganizan en el tiempo y el espacio para que surjan propiedades autónomas similares a las de las células.

## 2. IMPACTO EN EL PANORAMA DE LA CIENCIA BÁSICA Y EN LAS APLICACIONES POTENCIALES

### 2.1. Aspectos básicos y filosóficos que abordar

Una lista mínima de cuestiones filosóficas que merecen atención en cualquier plan de investigación sobre los orígenes de la vida (Mariscal *et al.*, 2019) incluye las siguientes: i) la naturaleza de la vida y la búsqueda de una definición de vida; ii) el poder explicativo de la investigación sobre los orígenes de la vida, en la que se superponen los programas de investigación universal (es decir, válido para todos los orígenes), histórico (descripciones del origen de la vida en la Tierra) y sintético (formas posibles de originar la vida), todos ellos con interesantes proyecciones científicas; iii) las estrategias de dicha investigación, que suelen considerarse descendentes (infiriendo de la vida actual al último ancestro común universal, LUCA) o ascendentes (partiendo de la ausencia de vida y averiguando cómo se inició), que se enfrentan a diferentes problemas epistemológicos y requieren distintos compromisos filosóficos; iv) el debate sobre si fue primero el metabolismo o la replicación, que algunos enfoques sistémicos han puesto en duda (véase más adelante); v) la naturaleza de la evolución antes de LUCA, que era ciertamente diferente de la evolución contemporánea, es decir, el inicio de la selección; vi) la naturaleza de las entidades anteriores a LUCA, que a veces se consideran comunidades sueltas; vii) los desafíos de los orígenes de la vida que son comunes a las ciencias multidisciplinares: programas de investigación que compiten entre sí, estándares de evidencia diversos y comunicación entre divisiones disciplinares; y viii) el desarrollo de nuevas teorías o herramientas, que ofrecen oportunidades para nuevas vías de investigación, pero que también pueden limitar otras.

### 2.2. Búsqueda de los componentes clave de la vida en el espacio interestelar

Para profundizar en las conexiones entre el repertorio molecular en el espacio interestelar y los orígenes de la vida, hace poco se ha iniciado un enfoque novedoso, donde el problema se contempla desde el punto de vista de un (bio) químico en lugar de un astrofísico. Se centra en la búsqueda en el MIE de las moléculas orgánicas precursoras necesarias para el origen de las nucleobases y los ribonucleótidos completos que pudieron haber dado lugar a un mundo del ARN (Powner *et al.*, 2009; Patel *et al.*, 2015; Becker *et al.*, 2019). Esta estrategia ya ha proporcionado las primeras detecciones en el espacio de reactivos y productos clave dentro de este esquema, como el glicolonitrilo (OHCH<sub>2</sub>CN;

Zeng *et al.*, 2019), la cianometanimina (HNCHCN; Rivilla *et al.*, 2019) y la urea (Jiménez-Serra *et al.*, 2020).

Sin embargo, esto es solo el punto de partida. Los esquemas químicos prebióticos también incluyen otras especies como los ácidos carboxílicos, los ácidos grasos, los alcoholes grasos, los aminoácidos, los azúcares y los hidrocarburos aromáticos policíclicos (HAP). Una gran parte de estos compuestos se ha detectado en meteoritos (por ejemplo, la ribosa; Furukawa *et al.*, 2019) y en cometas (por ejemplo, la glicina; Altwegg *et al.*, 2016). Curiosamente, también se han identificado nucleobases en condritas carbonáceas (Martins *et al.*, 2008; Callahan *et al.*, 2011). Esto implica que estas moléculas relativamente grandes, o sus precursores, pudieron haberse formado originalmente en el espacio interestelar y haber llegado después a la Tierra por los impactos de meteoritos y núcleos cometarios.

Por tanto, el objetivo del grupo de Medio Interestelar y Circunestelar del CAB para las próximas décadas consiste en poblar el *esquema de química prebiótica interestelar* para determinar qué compuestos orgánicos podrían formarse ya en el medio interestelar. Con ese objetivo, el grupo ha desarrollado nuevas herramientas de análisis de datos astronómicos (el paquete MADRID DATA CUBE ANALYSIS, MADCUBA) y configuraciones experimentales (la cámara de ultraalto vacío ISAC), no solo para buscar estas moléculas, sino también para tratar de entender sus vías de formación en el espacio (Muñoz-Caro *et al.*, 2002; véase C9.6). Estos esfuerzos se combinarán con la modelización astroquímica de materia orgánica compleja en el medio interestelar (Quenard *et al.*, 2018).

Para lograr este objetivo, la espectroscopia desempeña un papel fundamental. La identificación de una determinada molécula requiere la comparación directa de las frecuencias (líneas de rotación molecular) observadas mediante radiotelescopios, con las experimentales obtenidas en el laboratorio. Los métodos computacionales mejoran día a día, pero no son capaces de predecir las frecuencias de una molécula con la exactitud necesaria y, por lo tanto, no queda más remedio que caracterizar la molécula en el laboratorio. El grupo de Espectroscopia Molecular de la Universidad del País Vasco (UPV-EHU) y el Instituto Biofísica (CSIC/UPV-EHU) han diseñado y construido instrumentación en la región centimétrica de microondas. El objetivo pasa por combinar diferentes herramientas espectroscópicas en el dominio del tiempo y la frecuencia en longitudes de onda centimétricas, milimétricas y submilimétricas, para proporcionar el conjunto preciso de constantes espectroscópicas que

podieran utilizarse para buscar estas especies en el medio interestelar. Este trabajo será esencial para la explotación científica de futuras instrumentaciones como el radiotelescopio Square Kilometre Array (SKA), en el que España (y más concretamente, el grupo de Medio Interestelar y Circunestelar del CAB) está muy involucrada (grupo de trabajo Cradle of Life Science Working Group).

### **2.3. El papel de las superficies minerales en el origen de la vida**

Tanto las razones termodinámicas como las cinéticas sostienen la propuesta de que los minerales desempeñaron un papel clave en las rutas químicas y protobiológicas hacia los primeros organismos vivos. Se ha propuesto que la montmorillonita (Ferris, 2006), así como muchos otros minerales (incluidos zeolitas, sulfuros, óxido de hierro y sílice) aceleran la síntesis de las moléculas necesarias para la vida tal y como la conocemos, así como las reacciones de polimerización que dieron lugar al ARN y a los péptidos (Saladino *et al.*, 2019).

Se han realizado trabajos experimentales que respalden estas afirmaciones, pero aún se echa en falta una buena comprensión de las razones estructurales de por qué las superficies minerales catalizan reacciones relevantes para la química prebiótica y el metabolismo primitivo y cómo lo hacen (Hazen, 2010). Este desafío requerirá que mineralogistas, químicos orgánicos, científicos de materiales, bioquímicos, químicos de catálisis y microbiólogos, trabajen de forma sinérgica en las siguientes tareas: i) buscar atajos para las rutas químicas hacia la polimerización, la compartimentación, la protección UV y la formación de moléculas complejas a la larga capaces de replicarse, como los polipéptidos, los ácidos nucleicos u otras moléculas no *inventadas* por la vida; ii) buscar procesos químicos autocatalíticos que puedan proporcionar energía para desencadenar reacciones que presenten altas barreras cinéticas; iii) comprender la capacidad demostrada de la sílice para adsorber compuestos orgánicos y catalizar sus reacciones. Esto, junto con la exploración de otros minerales de hierro y magnesio, abrirá nuevas vías para el descubrimiento de la historia prebiológica y protobiológica de este planeta.

En este marco, la búsqueda de los restos más antiguos de vida es también crucial para establecer el límite inferior de edad del planeta sin vida y las condiciones geoquímicas que desencadenaron la aparición de esta y la evolución darwiniana (véase el Desafío 3). Por desgracia, esto no es en absoluto una tarea fácil (Javaux, 2019). Desde los primeros estudios micropaleontológicos sobre los restos de vida más antiguos, era evidente la dificultad de evaluar con

rigor la biogenicidad, es decir, si un elemento ha sido formado por la vida o no. Ni las firmas isotópicas ni la materia orgánica en sí misma son una biofirma inequívoca, ya que puede sintetizarse abióticamente en muchos entornos diferentes. Los microfósiles putativos y las estructuras sedimentarias de origen microbiano, como los estromatolitos, tampoco son fiables, ya que pueden ser el resultado de procesos abióticos. Se ha demostrado experimentalmente la capacidad de los procesos físicos y químicos puramente inorgánicos para imitar los restos celulares de la vida primitiva, por lo que la morfología por sí misma no puede utilizarse como criterio de biogenicidad (García-Ruiz *et al.*, 2003). El problema de la detección de vida primitiva aquí o en otros lugares debe abordarse con una búsqueda imparcial, es decir, es preciso buscar patrones minerales extraños (ya sean morfológicos, físicos o químicos) y, luego, aplicar un protocolo que contenga criterios de biogenicidad y no biogenicidad. Por último, el tema inexplorado de la tafonomía bacteriana debería desarrollarse en paralelo a la investigación de la autoorganización mineral abiótica.

#### **2.4. Objetivos científicos en el ámbito de la química prebiótica**

La investigación en curso en el campo de la química prebiótica incluye la influencia de las condiciones ambientales físico-químicas en los procesos primarios que generan compuestos orgánicos complejos a partir de moléculas simples en las fases gaseosa y líquida, así como en los hielos. Se estudiarán diferentes fuentes de energía, incluida una amplia gama de temperaturas, plasmas fríos y calientes, fotólisis UV, choques por impacto y vaporización. Además, la interacción de biomoléculas en superficies metálicas y minerales informará sobre la estabilidad de las moléculas adsorbidas en diferentes condiciones ambientales, los procesos de autoensamblaje y la catálisis heterogénea en relación con el origen de la vida. Esto pudiera dar pistas sobre la aparición de sistemas alejados del equilibrio basados en los ciclos oscilantes de Krebs, lo que pudiera allanar el camino hacia un protometabolismo.

De particular importancia es la elucidación de la unidireccionalidad de los biomonómeros (por ejemplo, aminoácidos y nucleótidos) y de los polímeros (por ejemplo, péptidos o ácidos nucleicos) y si la ruptura de la simetría se produjo por fuerzas físicas (capaces de causar pequeños, pero significativos, desequilibrios enantioméricos en orgánicos quirales formados en medios extra-terrestres) o tuvo lugar después de la polimerización (Cintas, 2016; Hochberg y Ribó, 2019; Hochberg y Cintas, 2020). De hecho, la asimetría molecular representa una biofirma de la vida (Glavin *et al.*, 2020) que motiva las

investigaciones de los mecanismos abióticos y sitúa la vida en una perspectiva más amplia. Por lo tanto, es preciso estudiar la selección enantiomérica impulsada por las moléculas quirales adsorbidas en las superficies y la desencadenada por la cristalización de moléculas no quirales (por ejemplo, el clorato de sodio), generalmente siguiendo vías autocatalíticas, que son dominantes en nuestro planeta debido a la existencia de una hidrosfera (Viedma *et al.*, 2015).

Una cuestión fundamental en este marco estriba en desarrollar los aspectos teóricos de la química prebiótica y el origen de la quiralidad molecular. Uno de los principales objetivos será el estudio de los sistemas de reacción autocatalítica, los sistemas químicos autorreplicantes y los sistemas de ruptura de la simetría quiral, como precursores del origen de la homoquiralidad biológica. La meta se centra en analizar las redes de reacción relevantes, utilizando la física, la fisicoquímica y las simulaciones numéricas para estudiar sus propiedades de estabilidad dinámica (teoría de perturbaciones), sus propiedades críticas (mediante el grupo de renormalización dinámico), el papel del ruido intrínseco limitado por la difusión (métodos estocásticos) y su evolución espacial y temporal (simulaciones numéricas). Otro rumbo prometedor se refiere a la explotación de los modelos astroquímicos para las reacciones en el medio interestelar teniendo en cuenta los procesos físicos naturales, como la fotoquímica quiral, que pueden influir en la formación de pequeños excesos enantioméricos en moléculas orgánicas quirales simples encontradas en el polvo cósmico.

## **2.5. Profundización en el papel del ARN en el origen y la evolución temprana de la vida**

Hay dos escuelas de pensamiento principales sobre el origen del ARN: el ARN es el producto de reacciones prebióticas secuenciales que dieron lugar a los ribonucleótidos (Sutherland, 2016) o el resultado de la evolución química/prebiológica a partir de otros polímeros genéticos (Hud *et al.*, 2013; Higgs y Lehman, 2015). Aunque se han producido avances clave en este campo (Powner *et al.*, 2009; Patel *et al.*, 2015; Becker *et al.*, 2019), la demostración de una síntesis prebiótica directa del ARN se enfrenta a numerosos desafíos, lo que refuerza la hipótesis de que la disponibilidad de información heredable puede haber comenzado con un polímero proto-ARN ancestral. Encontrar un ancestro plausible del ARN que pudo haber resultado del autoensamblaje molecular conciliaría la química prebiótica y la geoquímica encontrada en los entornos planetarios (Ruiz-Mirazo *et al.*, 2014). Una posibilidad se basa en la

introducción de bases no canónicas que encajan bien en la estructura de un ácido nucleico (como el ácido barbitúrico, un análogo prebiótico del uracilo, Menor-Salván *et al.*, 2009; Menor-Salván y Marin-Yaseli, 2013) y forman nucleósidos de forma espontánea, superando los desafíos de la síntesis prebiótica directa de nucleósidos canónicos.

Cuando se habla de los orígenes de la vida, los autores modernos suelen hacer referencia a la famosa carta que Charles R. Darwin envió a Joseph D. Hooker en 1871, en la que reflexionaba sobre un «warm little pond, with all sorts of ammonia and phosphoric salts... that a protein compound was chemically formed ready to undergo still more complex changes» (pequeño estanque cálido, con toda clase de amoníaco y sales fosfóricas... en el que un compuesto proteico se formó químicamente, listo para sufrir cambios aún más complejos). Utilizando las propiedades físicas y químicas de la urea, una molécula clave en la química prebiótica, el modelo del estanque cálido combina en un solo escenario la geoquímica de la Tierra arcaica/hádica, la química única de la atmósfera en esta época y la química de los fosfatos. El carácter problemático de los fosfatos, la forma predominante de fósforo, se debe a su bajísima solubilidad y a su incapacidad para formar organofosfatos con facilidad. Esto dificulta la propuesta de un modelo geoquímico significativo y da lugar al problema de la fosforilación en el campo de los orígenes de la vida. El cianuro prebiótico pudiera ser el precursor de sustancias químicas prebióticas clave, como la urea y el formiato de amonio. Además, puede promover directamente la síntesis de nucleótidos mediante la movilización del fosfato de los minerales. Por lo tanto, si asumimos la presencia de minerales de cianuro, urea y fosfato en la Tierra primitiva, cabría la posibilidad de que la química y los minerales prebióticos establecieran un sistema en el que la formación de nucleobases, nucleósidos y la fosforilación prebiótica pudiera haberse realizado fácilmente en un escenario sencillo y en consonancia con la geoquímica y la geología planetaria (Burcar *et al.*, 2016; Burcar *et al.*, 2019; Burcar y Menor-Salván, 2020).

Además, para profundizar en el conocimiento del papel del ARN en el origen de la vida sería necesaria una combinación de análisis bioquímicos, biofísicos, mutacionales y filogenéticos, junto con el uso de novedosas herramientas computacionales, en el contexto de diferentes modelos biológicos (Briones *et al.*, 2009; Atkins *et al.*, 2011; Romero-López y Berzal-Herranz, 2013; Takahashi *et al.*, 2016; Moreno *et al.*, 2019b). Es preciso abordar la comprensión de la función de la plétora de ARNnc actuales para rastrear algunos de los procesos biológicos clave de la vida. Además, conocer la diversidad funcional de los ARN y

sus mecanismos de acción, así como la arquitectura de su interactoma (es decir, la relación entre la estructura y la función biológica en los ARN virales y celulares), debería esclarecer el papel de este biopolímero en el origen y la evolución temprana de la vida (Atkins *et al.*, 2011; Ariza-Mateos *et al.*, 2019). El desarrollo adicional de nuevas herramientas moleculares mediante estrategias de selección *in vitro* también pudo ayudar a dilucidar la función de los elementos estructurales contenidos en moléculas de ARN celular.

Una evaluación cuantitativa de las cuestiones relacionadas con la aparición de la función compleja y la innovación molecular en un mundo precelular requerirá una importante investigación transdisciplinaria. Hay que desarrollar análisis matemáticos del mapa genotipo-fenotipo-función, nuevos enfoques computacionales que permitan una caracterización no sesgada de los espacios de secuencias y experimentos adaptados para responder a preguntas específicas (Manrubia *et al.*, 2020; véase el Desafío 5).

## **2.6. La aplicación de la teoría de la complejidad en el campo del origen de la vida**

Esta novedosa línea de investigación pretende analizar la vida como un proceso que surge espontáneamente de la compleja interacción (competencia/cooperación) entre sistemas en red que representan las estructuras bioquímicas existentes en la Tierra primitiva. La investigación se beneficiará de la comprensión actual de cómo se organizan, interactúan y evolucionan los sistemas en red, obtenida gracias a la aplicación durante más de una década de la teoría de redes complejas a sistemas reales (Boccaletti *et al.*, 2006). De hecho, en muchos casos, se ha demostrado que una red es una red de redes y, aunque se ha estudiado a fondo la detección de módulos dentro de una red, la influencia de las interconexiones de las redes en sus procesos dinámicos aún supone un gran desafío. Por ejemplo, estas interacciones son cruciales en procesos como la propagación de enfermedades, la transmisión de conocimientos e incluso la difusión de rumores (Aguirre *et al.*, 2013). La interacción entre los ladrillos de construcción de la vida no es una excepción. Por ello, desarrollar una teoría que describa completamente la competencia y la cooperación entre redes biológicas o químicas, haciendo uso de una sinergia entre la teoría de redes y la de juegos (Iranzo *et al.*, 2016), tanto en equilibrio como sin él (Buldú *et al.*, 2019), se convierte en un requisito necesario para entender los primeros pasos de la vida. Además, cada una de las muchas redes bioquímicas ya descritas en la bibliografía (redes proteicas, genéticas, metabólicas o de genotipos neutros, entre otras), tiene peculiaridades y parece difícil —si no imposible— extraer

comportamientos y propiedades generales. Para lograr el objetivo, también será fundamental realizar un análisis exhaustivo de cuáles de estas redes pudieran aplicarse —y cómo— para describir la transición de la química a la biología, o la necesidad de desarrollar redes *ad hoc* que permitan describir mejor la idiosincrasia del mundo prebiótico.

## 2.7. Desafíos en el campo de la química de sistemas prebióticos

Actualmente existen varias líneas de trabajo experimental y computacional, que se centran en la construcción de conjuntos protocelulares mediante la integración de los tres subsistemas básicos de la vida celular a través de diferentes tipos de procesos fisicoquímicos. Las investigaciones sobre la autorreplificación, las redes autocatalíticas y los compartimentos autorreproductores son las más importantes, pero todas ellas se enfrentan a varios problemas y limitaciones inherentes. En primer lugar, la mayor parte del trabajo se ha realizado con componentes moleculares tomados de organismos vivos existentes (por ejemplo, fosfolípidos, péptidos, oligonucleótidos, etc.), asumiendo que hubieran estado disponibles en la Tierra prebiótica. Este es un enfoque útil para estudiar los modelos funcionales de las primeras protocélulas, pero es muy improbable que esas biomoléculas pudieran haberse producido espontáneamente mediante procesos no evolutivos (reacciones de condensación, polimerizaciones aleatorias de aminoácidos o nucleótidos, etc.), en cantidades suficientes y con la estructura/secuencia adecuada para ejercer su función. Además, una fuerte limitación radica en la dificultad de integrar los complejos comportamientos dinámicos de cada subsistema por separado (de la Escosura, 2019). Con el fin de superar estas limitaciones, muchos científicos se han dado cuenta de que para desentrañar las principales vías y mecanismos de la evolución química será necesario investigar conjuntos moleculares que sean capaces de regular la producción de sus propios ingredientes a partir de las unidades estructurales más simples (por ejemplo, ácidos grasos, azúcares simples, aminoácidos, nucleobases, etc.). Esto implica grandes desafíos tanto en los estudios sintéticos y supramoleculares que se ocupan de las redes de reacción orgánicas e inorgánicas, como en la química analítica, debido a las mezclas desordenadas extremadamente complejas en las que intervienen cientos de reacciones dinámicas que actúan en paralelo. Recientemente, se han dado pasos importantes en el desafío de conectar la química orgánica sintética y la bioquímica relacionada con las *vías metabólicas centrales* (Keller *et al.*, 2016; Coggins y Powner 2017; Muchowska *et al.*, 2017; 2019; Springsteen *et al.*, 2018), que implican una notable diversidad de especies y en ausencia de catalizadores enzimáticos. Sin embargo, las actuales analíticas estándar de

recogida de moléculas (resonancia magnética nuclear, cromatografía con detección, espectrometría de masas, etc.) no parecen suficientes para abordar tal complejidad, por lo que tendrán que complementarse con métodos de cribado de alto rendimiento, adaptados de las disciplinas biológicas *ómicas* (genómica, proteómica, metabolómica, etc.) y la modelización computacional para desentrañar los comportamientos emergentes dentro de esas mezclas (Ruiz-Mirazo *et al.*, 2014).

Un segundo aspecto clave se refiere a la capacidad de evolución. Se trata de la capacidad de los sistemas vivos para procesar información heredable, codificada en el ADN y utilizada para indicar el funcionamiento de la célula. El panorama general que está surgiendo muestra que el procesamiento de la información en las células se produce a través de una jerarquía de genes que regulan la actividad de otros genes por medio de un interactoma complejo y de redes metabólicas. Existe un carácter semiótico implícito en esta forma de tratar la información, basada en moléculas funcionales que actúan como signos para lograr la autorregulación de toda la red (de la Escosura, 2019; Ariza-Mateos *et al.*, 2019). A diferencia de las células, los sistemas químicos no son capaces de procesar información. Sin embargo, deben haber precedido a los organismos biológicos y haber evolucionado hacia ellos. Por lo tanto, deben haber existido conjuntos moleculares prebióticos que pudieran procesar de alguna manera la información, con el fin de regular sus propias reacciones constitutivas y los procesos de organización supramolecular. Por tanto, un gran desafío para el futuro próximo consiste en investigar si existen principios universales que impregnen la forma en que el procesamiento de la información determina una complejidad dinámica cada vez mayor del mundo material, en especial en sus manifestaciones vivas y, posiblemente, hacia ellas. Esta perspectiva también plantea la cuestión de si existe una relación causal entre el procesamiento de la información y la evolución, no solo en los organismos vivos, sino también en las etapas abióticas de la evolución química (de la Escosura *et al.*, 2015; Ruiz-Mirazo *et al.*, 2017). Además, teniendo en cuenta que las poblaciones de protocélulas deben considerarse como las unidades propias de la evolución prebiótica, hay que pensar en tres desafíos experimentales dirigidos a construir sistemas de protocélulas: acoplar las redes de reacción química con la dinámica de las vesículas, encontrar las condiciones y los mecanismos para una integración funcional mínima y caracterizar la dinámica evolutiva de las protocélulas predarwinianas (Shirt-Ediss *et al.*, 2017).

## 2.8. Hacia las primeras protocélulas

La química de los sistemas prebióticos está superando las controversias clásicas en torno al origen de la vida, basadas en dicotomías como la aparición temprana frente a la tardía de los compartimentos, los modelos de metabolismo primero frente a los de genética primero o los escenarios autótrofos frente a los heterótrofos (Peretó, 2005; Ruiz-Mirazo *et al.*, 2014). Los resultados experimentales en química de sistemas prebióticos sugieren que el protometabolismo en la Tierra arcaica generó un pequeño conjunto de biomonómeros que prefiguraron la bioquímica actual (Peretó, 2019; véanse el Desafío 5 y el Desafío 8).

No obstante, existen importantes desafíos a la hora de aplicar sistemas experimentales con niveles de complejidad cada vez mayores. 1) Necesitamos saber más sobre el acoplamiento dinámico de las vesículas lipídicas con mezclas químicas heterogéneas de pequeñas moléculas y polímeros (Murillo-Sánchez *et al.*, 2016), así como su evolución experimental hacia sistemas más robustos que simulen protocélulas con comportamiento predarwiniano o darwiniano. Esto incluiría la exploración de los procesos evolutivos en el contexto de experimentos de población/selección de protocélulas (Budin y Szostak, 2011; Adamala y Szostak, 2013) o incluso la aparición de las primeras relaciones protoecológicas o sintróficas entre protocélulas, como se argumenta recientemente (Ruiz-Mirazo *et al.*, 2020). 2) La fusión de la química de sistemas con las metodologías de evolución experimental es un requisito indispensable si queremos cerrar la brecha entre el protometabolismo y las protocélulas funcionalmente integradas (Ruiz-Mirazo *et al.*, 2017; Shirt-Ediss *et al.*, 2017). En resumen, las soluciones difícilmente accesibles por el diseño racional pueden encontrarse mediante la propia dinámica del sistema. Por lo tanto, la fusión de tecnologías de laboratorio, computacionales, robóticas, microfluídicas y de biología sintética es obligatoria para explorar en profundidad la transición de mezclas químicas simples a sistemas con comportamientos complejos (Scharf *et al.*, 2015). 3) Algunos aspectos, sin embargo, todavía requieren esfuerzos de investigación significativos y persistentes, incluida la búsqueda de las rutas de aparición de redes autocatalíticas de reacciones que exhiben una dinámica de relevancia prebiótica.

## 3. PUNTOS CLAVE DEL DESAFÍO

El origen de la vida es un campo multidisciplinario en el que intervienen la física, la química, la biología y la geología, con implicaciones relevantes para las

ciencias sociales y las humanidades. El enfoque que aquí se propone combina los conocimientos de diferentes grupos del CSIC. Esto debería posicionar a nuestra institución no solo como líder nacional en este campo, sino también en una situación privilegiada para competir y colaborar con otras instituciones internacionales líderes. Dentro de este marco, los puntos clave que deben abordarse son los siguientes:

- Buscar en el medio interestelar las moléculas orgánicas precursoras necesarias para el origen de la vida e intentar comprender sus vías de formación en el espacio interestelar.
- Análisis de los patrones minerales (morfológicos, físicos o químicos) presentes en algunos de los restos de vida más antiguos y la aplicación de un protocolo con criterios de biogenicidad y no biogenicidad.
- Investigación de rutas químicas hacia la síntesis de biomoléculas, polimerización de péptidos y ácidos nucleicos, así como autoensamblaje molecular en superficies minerales o metálicas.
- Diseño y realización de experimentos de química prebiótica que generen compuestos orgánicos complejos a partir de moléculas simples en fase gaseosa y líquida, así como en hielos, utilizando fuentes de energía como diferentes temperaturas, plasmas fríos y calientes, fotólisis UV, choques por impacto y vaporización.
- Estudio de la selección enantiomérica impulsada por moléculas quirales adsorbidas en superficies, así como la provocada por la cristalización de moléculas no quirales.
- Desarrollo de los aspectos teóricos de la química prebiótica y el origen de la homociralidad molecular.
- Investigación de cómo el fosfato y otros grupos enlazadores plausibles se incorporaron a la evolución química, en particular al ARN.
- Profundizar en las relaciones secuencia-estructura-función-evolución en moléculas y poblaciones de ARN.
- Análisis experimental de la estructura y función de los ARNnc, así como de sus patrones de interacción en ARN virales y celulares seleccionados.
- Para llevar a cabo experimentos de evolución *in vitro* mediante combinaciones moleculares, dentro del marco de la química de sistemas prebióticos.
- Investigación del autoensamblaje de moléculas complejas y la formación de vesículas en diferentes escenarios experimentales, así como el acoplamiento de las redes de reacción química con la dinámica de las vesículas.

- Integración de tecnologías experimentales, computacionales y de biología sintética para explorar la transición de mezclas químicas simples a sistemas dotados de comportamientos complejos.
- Identificar los principios universales que permiten que el procesamiento de la información determine el aumento de la complejidad dinámica del mundo viviente.
- Uso de nuevos enfoques computacionales que permiten una caracterización no sesgada de los espacios de secuencias y el análisis del mapa genotipo-fenotipo.
- Desarrollo de una teoría para describir la competencia y la cooperación entre redes químicas o biológicas, haciendo uso de las teorías de redes y de juegos, tanto en equilibrio como sin él.

## DESAFÍO 1 | BIBLIOGRAFÍA

- Adamala, K., Szostak, J.W. (2013). Competition between model protocells driven by an encapsulated catalyst. *Nat. Chem.* 5, 495-501.
- Aguirre, J., Papo, D., Buldú, J.M. (2013). Successful strategies for competing networks. *Nat. Phys.* 9, 230-234.
- Alonso, E.R., Kolesníková, L., Tercero, B. *et al.* (2016). Millimeter Wave Spectrum and Astronomical Search for Vinyl Formate. *The Astrophysical Journal* 832, 42.
- Alonso, E.R., McGuire, B.A., Kolesníková, L. *et al.* (2019). The Laboratory Millimeter and Submillimeter Rotational Spectrum of Lactaldehyde and an Astronomical Search in Sgr B2(N), Orion-KL, and NGC 6334I. *The Astrophysical Journal* 883, 18.
- Altwegg, K., Balsiger, H., Bar-Nun, A. *et al.* (2016). Prebiotic chemicals-amino acid and phosphorus-in the coma of comet 67P/Churyumov-Gerasimenko. *Science Adv.* 2, 5.
- Ariza-Mateos, A., Briones, C., Perales, C., Domingo, E., Gómez, J. (2019). The archaeology of coding RNA. *Annals of the New York Academy of Sciences* 1447, 119-134.
- Atkins, J.F., Gesteland, R.F., Cech, T.R., (Ed.). (2011). *RNA worlds: from life's origins to diversity in gene regulation*. CSHL Press, NY, US.
- Ausländer, S., Ausländer, D., Fussenegger, M. (2017). Synthetic biology- the synthesis of biology. *Angew. Chem. Int. Ed. Engl.* 56, 6396-6419.
- Becker, S., Feldmann, J., Wiedemann, S. *et al.* (2019). Unified prebiotically plausible synthesis of pyrimidine and purine RNA ribonucleotides. *Science* 366, 76-82.
- Belloche, A., Menten, K.M., Comito, C. *et al.* (2008). Detection of amino acetonitrile in Sgr B2(N). *Astronomy & Astrophysics* 482, 179B.
- Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, D.U. (2006). Complex networks: structure and dynamics. *Phys. Rep.* 424, 175-308.
- Breslow, R., Cheng, Z-L. (2009). On the origin of terrestrial homochirality for nucleosides and amino acids, *Proc. Natl. Acad. Sci. USA* 106, 9144-9146.
- Briones, C. (2015). Aptamer. *Encyclopedia of Astrobiology*, 2.<sup>a</sup> ed. M. Gargaud y W.M. Irvine, Eds., Springer, Heidelberg. Pp. 112-113.
- Briones, C., Stich, M., Manrubia, S.C. (2009). The dawn of the RNA world: Towards functional complexity through ligation of random RNA oligomers. *RNA* 15, 743-749.
- Budin, I., Szostak, J.W. (2011). Physical effects underlying the transition from primitive to modern cell membranes. *Proc. Natl. Acad. Sci. USA* 108, 5249-5254.
- Buldú, J.M., Pablo-Martí, F., Aguirre, J. (2019). Taming out-of-equilibrium dynamics on interconnected networks. *Nat. Commun.* 10, 5314.
- Burcar, B., Castañeda, A., Lago, J. *et al.* (2019). A Stark Contrast to Modern Earth: Phosphate Mineral Transformation and Nucleoside Phosphorylation in an Iron- and Cyanide-Rich Early Earth Scenario. *Angew. Chemie Int. Ed.* 58, 16981-16987.
- Burcar, B., Menor-Salván, C. (2020). The multiple roles of urea in chemical evolution. *Chem. Rev.*, en prensa.
- Burcar, B., Pasek, M., Gull, M. *et al.* (2016). Darwin's Warm Little Pond: A One-Pot Reaction for Prebiotic Phosphorylation and the Mobilization of Phosphate from Minerals in a Urea-Based Solvent. *Angew. Chemie Int. Ed.* 55, 13249-13253.
- Calabrese, C., Uriarte, I., Insausti, A. *et al.* (2020). *Observation of the Unbiased Conformers of Putative DNA-Scaffold Ribosugars*. ACS Central Science 0.
- Callahan, M.P., Smith, K.E., Cleaves, H.J. *et al.* (2011). Carbonaceous meteorites contain a wide range of extraterrestrial nucleobases. *Proc. Natl. Acad. Sci.* 108, 13995-13998.
- Catalán, P., Elena, S.F., Cuesta, J.A., Manrubia, S. (2019). Parsimonious scenario for the emergence of viroid-like replicons de novo. *Viruses* 11, 425.
- Cintas, P. (2016). Homochirogenesis and the emergence of life-like structures. *Chirality in Supramolecular Assemblies-Causes and Consequences* (Ed.: R. Keene), John Wiley & Sons, NY, Ch. 2, pp. 44-64.
- Cocinero, E.J., Lesarri, A., Écija, P. *et al.* (2012). Ribose Found in the Gas Phase. *Angew. Chem. Int. Ed.* 51, 3119.

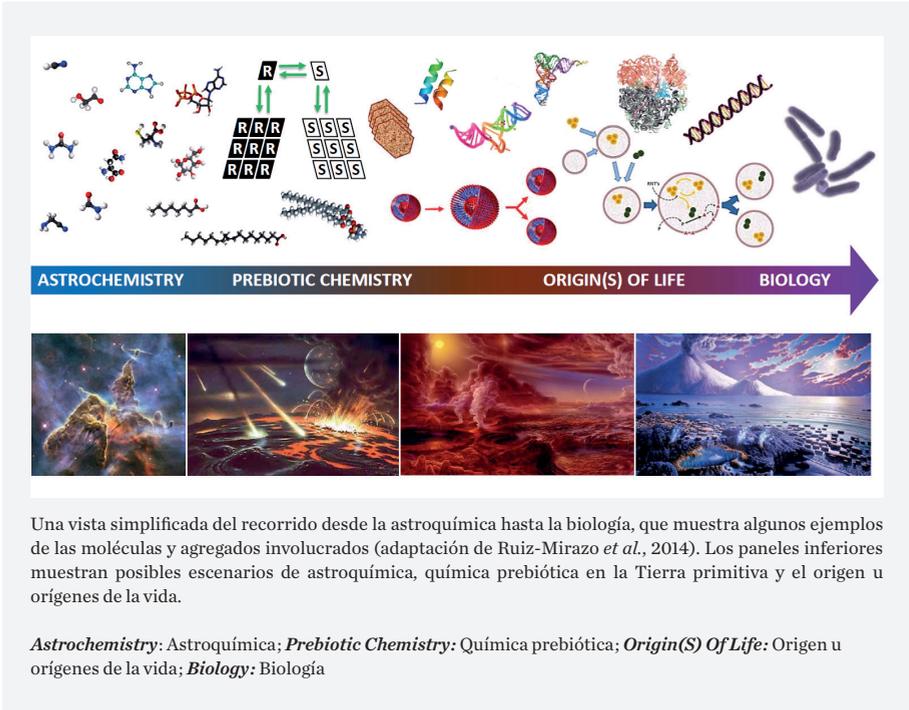
- Coggins, A.J., Powner, M.W. (2017).** Prebiotic synthesis of phosphoenol pyruvate by alpha-phosphorylation-controlled triose glycolysis. *Nat. Chem.* 9, 310-317.
- Cologne Database for Molecular Spectroscopy (CDMS): <https://cdms.astro.uni-koeln.de/classic/molecules>
- De la Escosura, A. (2019).** The informational substrate of chemical evolution: implications for abiogenesis. *Life* 9, 66.
- De la Escosura, A., Briones, C., Ruiz-Mirazo, K. (2015).** The systems perspective at the crossroad between chemistry and biology. *J. Theor. Biol.* 381, 11-22.
- Eschenmoser, A. (2007).** The search for the chemistry of life's origins. *Tetrahedron* 63, 12821-12844.
- Ferris, J.P. (2006).** Montmorillonite-catalysed formation of RNA oligomers: the possible role of catalysis in the origins of life. *Phil. Trans. R. Soc. B* 361, 1777-1786.
- Fletcher, D.A. (2016).** Bottom-up biology: Harnessing engineering to understand nature. *Dev. Cell.* 38, 587-589
- Furukawa, Y., Chikaraishi, Y., Ohkouchi, N. et al. (2019).** Extraterrestrial ribose and other sugars in primitive meteorites. *PNAS* 116, 49
- Gálvez-Martínez, S., Escamilla-Roa, E., Zorzano, M.P., Mateo-Martí, E. (2019).** Defects on a pyrite(100) surface produce chemical evolution of glycine under inert conditions: experimental and theoretical approaches. *Phys. Chem. Chem. Phys.* 21, 24535.
- García-Ruiz, J.M., Melero-García, E., Hyde, S.T. (2009).** Morphogenesis of self-assembled nanocrystalline materials of barium carbonate and silica. *Science* 323, 362-365.
- García-Ruiz, J.M., Van Zuilen, M.A., Bach, W. (2020).** Mineral self-organization on a lifeless planet. *Physics of Life Reviews*, en prensa.
- Glavin, D.P., Burton, A.S., Elsila, J.E., Aponte, J.C., Dworkin, J.P. (2020).** The search for chiral asymmetry as a potential biosignature in our solar system. *Chem. Rev.*, en prensa.
- Haykal, I., Margulès, L., Huet, T.R. et al. (2013).** The CM-, MM-, and SUB-MM-WAVE spectrum of allyl isocyanide and radioastronomical observations in Orion KL and the SgrB2 line surveys. *The Astrophysical Journal* 777, 120.
- Hazen, R.M., Sverjensky, D.A. (2010).** Mineral surfaces, geochemical complexities, and the origins of life. *Cold Spring Harb Perspect Biol.* 2, a002162.
- Higgs, P.G., Lehman, N. (2015).** The RNA World: Molecular Cooperation at the Origins of Life. *Nat. Rev. Genet.* 16, 7-17.
- Hochberg, D., Cintas, P. (2020).** Does pressure break mirror-image symmetry? A perspective and new insights. *Chem. Phys. Chem.* 21, 633-642.
- Hochberg, D., Ribó, J.M. (2019).** Entropic Analysis of Mirror Symmetry Breaking in Chiral Hypercycles. *Life* 9, 28.
- Huber, C., Wächtershauser, G. (1998).** Peptides by activation of amino acids with CO on (Ni,Fe)S surfaces: implications for the origin of life. *Science* 281, 670-672.
- Hud, N.V., Cafferty, B.J., Krishnamurthy, R., Williams, L.D. (2013).** The Origin of RNA and "My Grandfather's Axe". *Chem. Biol.* 20, 466-474.
- Iranzo, J., Buldú, J.M., Aguirre, J. (2016).** Competition among networks highlights the power of the weak. *Nat. Commun.* 7, 13273.
- Javaux, E. (2019).** Challenges in evidencing the earliest traces of life. *Nature* 572, 451-460.
- Jiménez-Serra, I., Martín-Pintado, J., Rivilla, V.M. et al. (2020).** CAB: Towards the RNA-world in the interstellar medium – detection of urea, and search of 2-amino-oxazole and simple sugars. *Astrobiology* 20, 7.
- Jorgensen, J., van der Wiel, M., Coutens, A. et al. (2012).** Detection of the Simplest Sugar, Glycolaldehyde, in a Solar-type Protostar with ALMA. *Astronomy & Astrophysics* 595, A117
- Joyce, G.F., Szostak, J.W. (2018).** protocells and RNA Self-Replication. *Cold Spring Harb. Perspect. Biol.* 10, a034801.
- Joyce, G.F. (2004).** Directed evolution of nucleic acid enzymes. *Annu. Rev. Biochem.* 73, 791-836.
- Keller, M.A., Zylstra, A., Castro, C., Turchyn, A.V., Griffin, J.L., Ralser, M. (2016).** Conditional iron and pH-dependent activity of a non-enzymatic glycolysis and pentose phosphate pathway. *Sci. Adv.* 2, e1501235.
- Krishnamurthy, R. (2017).** Giving rise to Life: Transition from Prebiotic Chemistry to Protobiology. *Acc. Chem. Res.* 50, 455-459.

- Lavado, N., García de la Concepción, J., Babiano, R., Cintas, P. (2018). Formation of cyanamide-glyoxal oligomers in aqueous environments relevant to primeval and astrochemical scenarios: a spectroscopic and theoretical study. *Chem. Eur. J.* 24, 4069-4085.
- Lucía-Sanz, A., Manrubia, S. (2017). Multipartite viruses: adaptive trick or evolutionary treat? *Systems Biology and Applications* 3, 34.
- Mann, S. (2013). The Origins of Life: Old Problems, New Chemistries. *Angew. Chem. Int. Ed.* 52, 155-162.
- Manrubia, S., Cuesta, J.A., Aguirre, J. *et al.* (2020). From genotypes to organisms: State-of-the-art and perspectives of a cornerstone in evolutionary dynamics. *Journal of the Royal Society Interface*, enviado.
- Marín-Yaseli, M.R., González-Toril, E., Mompeán, C., Ruiz-Bermejo, M. (2016). The role of the aqueous aerosols in the “Glyoxylate Scenario”: An experimental approach. *Chem. Eur. J.* 22, 12785-12799.
- Mariscal, C., Barahona, A., Aubert-Kato, N. *et al.* (2019). Hidden Concepts in the History and Philosophy of Origins-of-Life Studies: a Workshop Report. *Orig. Life Evol. Biosph.* 49, 111-145.
- Martín-Gago, J.A. (2011). Polycyclic aromatics: On-surface molecular engineering. *Nature Chemistry* 3, 11-12.
- Martins, Z., Botta, O., Fogel, M.L. (2008). Extraterrestrial nucleobases in the Murchison meteorite. *Earth Planet. Sci. Lett.* 270, 130-136.
- Mateo-Martí, E., Gálvez-Martínez, S., Gil-Lozano, C., Zorzano, M.P. (2019). Pyrite-induced UV-photocatalytic abiotic nitrogen fixation: implications for early atmospheres and Life. *Scientific Reports* 9, 15311.
- Méndez, J., López, M., Martín-Gago, J. (2011). On-surface synthesis of cyclic organic molecules. *Chemical Society Reviews* 40, 4578.
- Menor-Salván, C., Marín-Yaseli, M.R. (2012). Prebiotic Chemistry in Eutectic Solutions at the Water-Ice Matrix. *Chem. Soc. Rev.* 41, 5404.
- Menor-Salván, C., Ruiz-Bermejo, M., Guzmán, M.I., Osuna-Esteban, S., Veintemillas-Verdaguer, S. (2009). Synthesis of Pyrimidines and Triazines in Ice: Implications for the Prebiotic Chemistry of Nucleobases. *Chem. - A Eur. J.* 15, 4411-4418.
- Mompeán, C., Marín-Yaseli, M.R., Espigares, P., González-Toril, E., Zorzano, M.P., Ruiz-Bermejo, M. (2019). Prebiotic chemistry in neutral/reduced-alkaline gas-liquid interfaces. *Scientific Reports* 9, 1916.
- Morales-Reina, S., Giri, C., Leclercq, M. *et al.* (2020). Programmed recognition between complementary dinucleolipids to control the self-assembly of lipidic amphiphiles. *Chem. Eur. J.* 26, 1082-1090.
- Moreno, M., Fernández-Algar, M., Fernández-Chamorro, J., Ramajo, J., Martínez-Salas, E., Briones, C. (2019a). A combined ELONA-(RT) qPCR approach for characterizing DNA and RNA aptamers selected against PCBP-2. *Molecules* 24, 1213.
- Moreno, M., Vázquez, L., López-Carrasco, M.A., Martín-Gago, J.A., Flores, R., Briones, C. (2019b). Direct visualization of the native structure of viroid RNA at single-molecule resolution by atomic force microscopy. *RNA Biology* 16, 295-308.
- Muchowska, K.B., Varma, S.J., Chevallot-Beroux, E., Lethuillier-Karl, L., Li, G., Moran, J. (2017). Metals promote sequences of the reverse Krebs cycle. *Nat. Ecol. Evol.* 1, 1716-1721.
- Muchowska, K.B., Varma, S.J., Moran, J. (2019). Synthesis and breakdown of universal metabolic precursors promoted by iron. *Nature* 569, 104-107.
- Muñoz-Caro, G., Meierhenrich, U.J., Schutte, W.A. *et al.* (2002). Amino acids from ultraviolet irradiation of interstellar ice analogues. *Nature* 416, 403.
- Murillo-Sánchez, S., Beaufils, D., González-Mañas, J.M., Pascal, R., Ruiz-Mirazo, K. (2016). Fatty acids' double role in the prebiotic formation of a hydrophobic dipeptide. *Chem. Sci.* 7, 3406-3414.
- Oro, J., Kimball, A.P. (1961). Synthesis of Purines under Possible Primitive Earth I. Adenine from Hydrogen Cyanide. *Arch. Biochem. Biophys.* 94, 217-227.
- Patel, B.H., Percivalle, C., Ritson, D.J. *et al.* (2015). Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. *Nature Chemistry* 7, 301.
- Peña, I., Cocinero, E.J., Cabezas, C. *et al.* (2013). Six Pyranoside Forms of Free 2-Deoxy-D-Ribose. *Angewandte Chemie - International Edition* 52, 11840-11845.

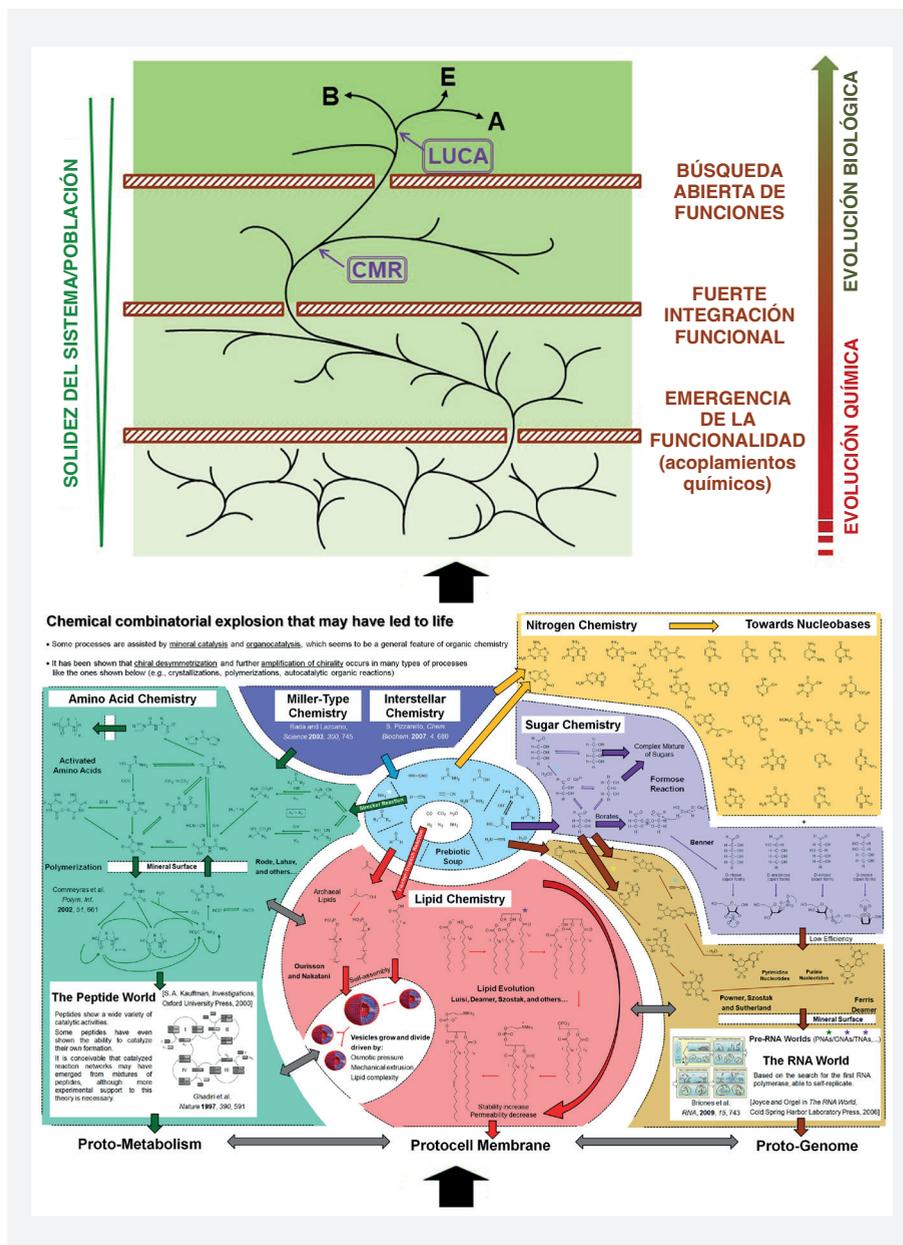
- Peretó, J. (2005). Controversies on the origin of life. *Int. Microbiol.* 8, 23-31.
- Peretó, J. (2019). Prebiotic chemistry that led to life. En: *Handbook of Astrobiology* (V Kolb ed.) CRC Press.
- Powner, M.W., Gerland, B., Sutherland, J.D. (2009). Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* 459, 239-242.
- Pross, A. (2016). *What is Life? How Chemistry Becomes Biology*, Oxford University Press: Oxford.
- Puerta-Fernández, C., Romero-López, C., Barroso-del Jesús, A., Berzal-Herranz, A. (2003). Ribozymes: recent advances in the development of RNA tools. *FEMS Microbiology Reviews* 27, 75-97.
- Ribó, J.M., Crusats, J., El-Hachemi, Z., Moyano, A., Hochberg, D. (2017). Spontaneous mirror symmetry breaking in heterocatalytically coupled enantioselective replicators. *Chem. Sci.* 8, 763-769.
- Rivilla, V.M., Martín-Pintado, J., Jiménez-Serra, I. *et al.* (2019). Abundant Z-cyano-methanimine in the interstellar medium: paving the way to the synthesis of adenine. *Monthly Notices of the Royal Astronomical Society* 483, L114-L119.
- Romero-López, C., Berzal-Herranz, A. (2013). Unmasking the information encoded as structural motifs of viral RNA genomes: a potential antiviral target. *Rev. Med. Virol.* 23, 340-354.
- Romero-López, C., Berzal-Herranz, A. (2020). The role of the RNA-RNA interactome in the hepatitis C virus life cycle. *Int. J. Mol. Sci.* 21, 1479.
- Ruiz-Bermejo, M., de la Fuente, J.L., Carretero-González, J., García-Fernández, L., Aguilar, M.R. (2019). A comparative study on HCN polymers synthesized by polymerization of NH<sub>4</sub>CN or diaminomaleonitrile in aqueous media: new perspectives for prebiotic chemistry and materials science. *Chem. Eur. J.* 25, 11437-11455.
- Ruiz-Bermejo, M., de la Fuente, J.L., Carretero-González, J., García-Fernández, L., Aguilar, M.R. (2019). A comparative study of HCN polymers synthesized from NH<sub>4</sub>CN or DAMN polymerization in aqueous media: New perspectives for prebiotic chemistry and material science. *Chem. Eur. J.* 25, 11437-11455.
- Ruiz-Mirazo, K., Briones, C., de la Escosura, A. (2014). Prebiotic systems chemistry: new perspectives for the origins of life. *Chem. Rev.* 114, 285-366.
- Ruiz-Mirazo, K., Briones, C., de la Escosura, A. (2017). Chemical roots of biological evolution: the origins of life as a process of development of autonomous functional systems. *Open Biology* 7, 170050.
- Ruiz-Mirazo, K., Shirt-Ediss, B., Escribano-Cabeza, M., Moreno, A. (2020). The construction of biological 'inter-identity' as the outcome of a complex process of protocol development in prebiotic evolution. *Frontiers in Physiology (Systems Biology)*, en prensa.
- Saladino, R., Di Mauro, E., García-Ruiz, J.M. (2019). A Universal Geochemical Scenario for Formamide Condensation and Prebiotic Chemistry. *Chemistry—A European Journal.* 25, 3181-3189.
- Sánchez-Arenillas, M., Mateo-Martí, E. (2016). Pyrite surface environment drives molecular adsorption: cystine on pyrite(100) investigated by X-ray photoemission spectroscopy and low energy electron diffraction. *Phys. Chem. Chem. Phys.* 18, 27219.
- Sánchez-Luque, F.J., Stich, M., Manrubia, S., Briones, C., Berzal-Herranz, A. (2014). Efficient HIV-1 inhibition by a 16 nt-long RNA aptamer designed by combining in vitro selection and in silico optimisation strategies. *Scientific Reports* 4, 6242.
- Schaper, S., Louis, A.A. (2014). The Arrival of the Frequent: How Bias in Genotype-Phenotype Maps Can Steer Populations to Local Optima. *PLoS ONE* 9, e86635.
- Scharf, C., Virgo, N., Cleaves, H.J. 2.<sup>a</sup> *et al.* (2015). A Strategy for Origins of Life Research. *Astrobiology* 15, 1031-1042.
- Schultes, E.A., Bartel, D.P. (2000). One sequence, two ribozymes: implications for the emergence of new ribozyme folds. *Science* 289, 448-452.
- Schwille, P. (2017). How simple could life be? *Angew Chem. Int. Ed. Engl.* 56, 10998-11002.
- Schwille, P., Spatz, J., Landfester, K. (2018). MaxSynBio: Avenues towards creating cells from the bottom up. *Angew Chem. Int. Ed. Engl.* 57, 13382-13392.

- Shirt-Ediss, B., Murillo-Sánchez, S., Ruiz-Mirazo, S. (2017).** Framing major prebiotic transitions as stages of protocell development: three challenges for origins-of-life research. *Beilstein J. Org. Chem.* 13, 1388-1395.
- Solé, R. (2016).** Synthetic transitions: Towards a new synthesis. *Phil. Trans. R. Soc. B: Biol. Sci.* 371, 20150438.
- Springsteen, G., Yerabolu, J.R., Nelson, J., Rhea, C.J., Krishnamurthy, R. (2018).** Linked cycles of oxidative decarboxylation of glyoxylate as protometabolic analogs of the citric acid cycle. *Nat. Commun.* 9, 1-8.
- Stano, P. (2018).** Is research on “synthetic cells” moving to the next level? *Life (Basel)* 9(1), 3.
- Supramaniam, P., Ces, O., Salehi-Reyhani, A. (2019).** Microfluidics for artificial life: Techniques for bottom-up synthetic biology. *Micromachines (Base)* 10, 299.
- Sutherland, J.D. (2016).** The Origin of Life - Out of the Blue. *Angew. Chemie Int. Ed.* 55, 104-121.
- Takahashi, M.K., Watters, K.E., Gasper, P.M., Abbott, T.R., Carlson, P.D., Chen, A.A., Lucks, J.B. (2016).** Using in-cell SHAPE-Seq and simulations to probe structure-function design principles of RNA transcriptional regulators. *RNA* 22, 920-933.
- Viedma, C., Coquerel, G., Cintas, P. (2015).** Crystallization of chiral molecules. En *Handbook of Crystal Growth* (Ed.: T. Nishinaga), Elsevier, Amsterdam, Vol. 1B, Ch. 22, pp. 952-1002.
- Wilde, S.A., Valley, J.W., Peck, W.H., Graham, C.M. (2001).** Evidence from detrital zircons for the existence of continental crust and oceans on the Earth 4.4 Gyr ago. *Nature* 409, 175-178.
- Wilke, C.O., Adami, C. (2002).** The biology of digital organisms. *Trends in Ecology and Evolution* 17, 528-532.
- Witzany, G. (2020).** What is Life? *Front. in Astron. Space Sci.* 7, 7.
- Yewdall, N.A., Mason, A.F., van Hest, J.C.M. (2018).** The hallmarks of living systems: towards creating artificial cells. *Interface Focus.* 8, 20180023.
- Zeng, S., Quenard, D., Jimenez-Serra, I. et al. (2019).** First detection of the pre-biotic molecule glycolonitrile (HOCH<sub>2</sub>CN) in the interstellar medium. *Monthly Notices of the Royal Astronomical Society* 484, L43Z.

## RESUMEN PARA EL PÚBLICO GENERAL

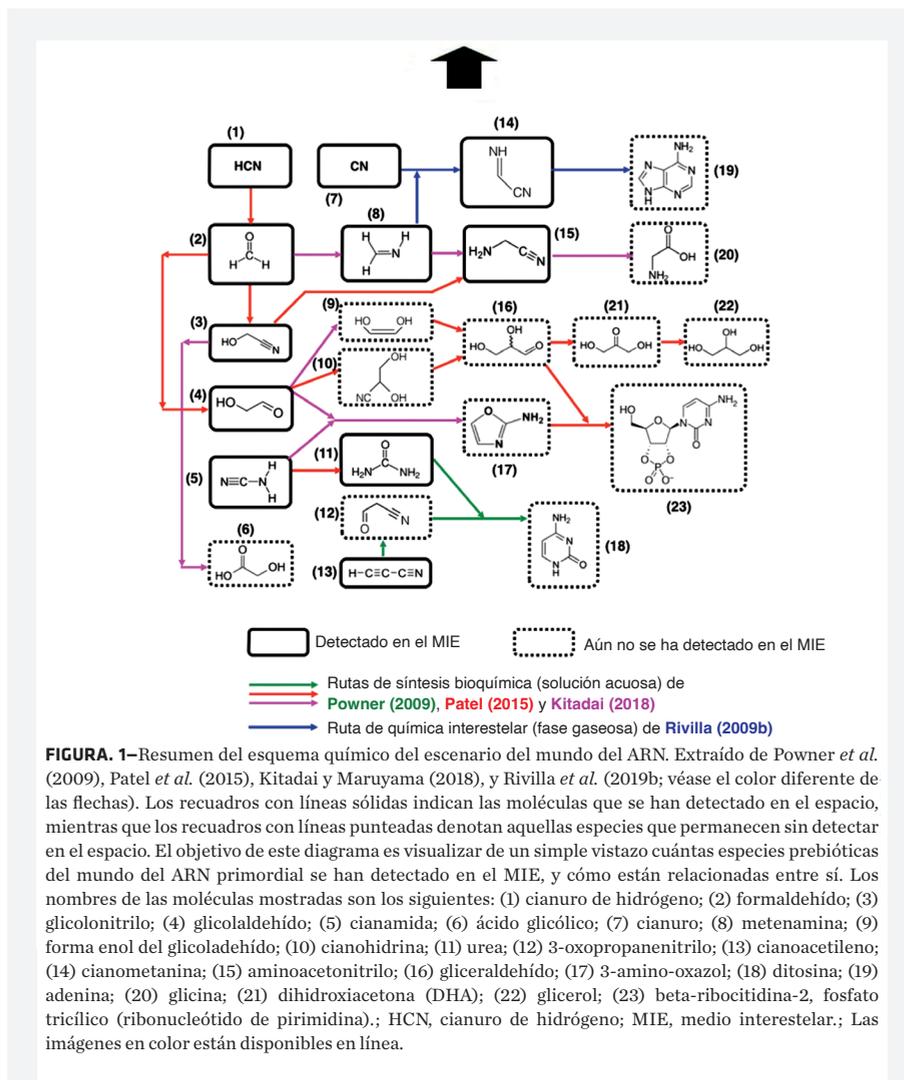


## RESUMEN PARA EXPERTOS



*Chemical combinatorial explosion that may have led to life*: Explosión química combinatoria que puede haber originado la vida; *Some processes are assisted by mineral catalysis and organocatalysis, which seems to be a general feature of organic chemistry*: Algunos procesos son asistidos por la catálisis mineral y la organocatálisis, que parece ser una característica general de la química orgánica; *It has been shown that*

*chiral desymmetrization and further amplification of chirality occurs in many types of processes like the ones shown below (e.g. crystallizations, polymerizations, autocatalytic organic reactions):* Se ha demostrado que la desimetrización quiral y la posterior amplificación de la quiralidad se producen en muchos tipos de procesos como los que se muestran a continuación (por ejemplo, cristalizaciones, polimerizaciones, reacciones orgánicas autocatalíticas, etc.); **Amino Acid Chemistry:** Química de los aminoácidos; **Activated Amino Acids:** Aminoácidos activados; **Polymerization:** Polimerización; **Mineral Surface:** Superficie mineral; **Rode, Lahav, and others...:** Rode, Lahav, et al.; **Commeyras et al., Polym. Int., 2002, 51, 661:** Commeyras et al., *Polym. Int.*, 2002, 51, 661; **The Peptide World. Peptides show a wide variety of catalytic activities. Some peptides have even shown the ability to catalyze their own formation. It is conceivable that catalyzed reaction networks may have emerged from mixtures of peptides, although more experimental support to this theory is necessary.;** [S. A. Kauffman, *Investigations*, Oxford University Press, 2000]; **Ghadiri et al. Nature. 1997, 390, 591:** El mundo de los péptidos; Los péptidos muestran una gran variedad de actividades catalíticas.; Algunos péptidos han mostrado incluso la capacidad de catalizar su propia formación.; Es concebible que las redes de reacciones catalizadas puedan haber surgido a partir de combinaciones de péptidos, aunque se precisa más respaldo experimental para esta teoría.; [S. A. Kauffman, *Investigations*, Oxford University Press, 2000]; **Ghadiri et al. Nature. 1997, 390, 591;** **Proto-Metabolism:** Protometabolismo; **Miller-Type Chemistry. Bada and Lazcano, Science. 2003, 300, 745:** Química de Miller, Bada and Lazcano, *Science*. 2003, 300, 745; **Interstellar Chemistry;** S. Pizzarello, *Chem. Biochem.* 2007, 4, 690; Química interestelar. S. Pizzarello, *Chem. Biochem.* 2007, 4, 690; **Strecker Reaction:** Reacción de Strecker; **Prebiotic Soup:** Sopa prebiótica; **Fischer-Tropsch Reaction:** Reacción Fischer-Tropsch; **Lipid Chemistry:** Química de los lípidos; **Archeal Lipids:** Lípidos arcaicos; **Ourisson and Nakatani:** Ourisson y Nakatani; **Self-assembly:** Autoensamblaje; **Vesicles grow and divide driven by: Osmotic pressure;** **Mechanical extrusion;** **Lipid complexity:** Las vesículas crecen y se dividen impulsadas por: La presión osmótica, extrusión mecánica, complejidad de los lípidos; **Lipid Evolution:Luisi, Deamer, Szostak, and others...:** Evolución de los lípidos. Luisi, Deamer, Szostak, et al.; **Stability increase. Permeability decrease:** Aumento de la estabilidad. Disminución de la permeabilidad; **Protocell Membrane:** Membrana de la protocélula; **Nitrogen Chemistry:** Química del nitrógeno; **Towards Nucleobases:** Hacia las nucleobases; **Sugar Chemistry:** Química de los azúcares; **Complex Mixture of Sugars:** Combinación compleja de azúcares; **Formose Reaction:** Reacción de la formosa; **Borates:** Boratos; **Benner:** Benner; **D-ribose (open form):** D-ribosa (forma abierta); **D-arabinose (open form):** D-arabinosa (forma abierta); **D-xylose (open form):** D-xilosa (forma abierta); **D-lyxose (open form):** D-lixosa (forma abierta); **Low Efficiency:** Baja eficiencia; **Pyrimidine Nucleotides:** Nucleótidos de pirimidina; **Purine Nucleotides:** Nucleótidos de purina; **Powner, Szostak and Sutherland:** Powner, Szostak y Sutherland; **Ferris Deamer:** Ferris Deamer; **Mineral Surface:** Superficie mineral; **Pre-RNA Worlds (PNAs/GNAs/TNAs,...). The RNA World. Based on the search for the first RNA polymerase, able to self-replicate. [Joyce and Orgel in The RNA World, Cold Spring Harbor Laboratory Press, 2006];** **Briones et al. RNA. 2009, 15, 743:** Mundos anteriores al ARN (PNAs/GNAs/TNAs,...). El mundo del ARN. Basado en la búsqueda del primer ARN polimerasa capaz de autorreplicarse. [Joyce y Orgel en *The RNA World*, Cold Spring Harbor Laboratory Press, 2006]. **Briones et al. RNA. 2009, 15, 743;** **Proto-Genome:** Protogenoma



Representación esquemática de la relación entre los campos científicos de la astroquímica (panel inferior: Jiménez-Serra *et al.*, 2020) y la química prebiótica (panel central: de la Escosura, Briones y Ruiz-Mirazo, inédito), con ejemplos de las moléculas y los procesos implicados. El panel superior representa la transición de la evolución química a la biológica, en la que el acoplamiento compartimento-metabolismo-replicación (CMR) se identifica como un paso clave en el origen de la vida (Ruiz-Mirazo *et al.*, 2017).



## DESAFÍO 2

### RESUMEN

Entender las bases de la vida y su diversidad evolutiva requiere descifrar la estructura tridimensional y la naturaleza dinámica de todas las macromoléculas que subyacen a los procesos vivos, y cómo se acoplan y funcionan de manera coordinada, oportuna y precisa. Abordar este desafío nos permitirá comprender y tratar las enfermedades, aprovechar los procesos biológicos con fines biotecnológicos y diseñar sintéticamente nuevas entidades biológicas.

### PALABRAS CLAVE

- Proteínas
- ácidos nucleicos
- complejos macromoleculares
- proteínas de membrana
- macromoléculas intrínsecamente desordenadas
- estructura tridimensional
- bases moleculares dinámica de la enfermedad
- diseño estructural de medicamentos
- criomicroscopía electrónica
- microscopía correlativa
- criotomografía electrónica
- cristalografía de rayos X
- resonancia magnética nuclear
- métodos de molécula única
- proteómica

# BASES ESTRUCTURALES DE LA VIDA Y EVOLUCIÓN DE LA COMPLEJIDAD MACRO- MOLECULAR

**Coordinadores**

José M. Valpuesta  
(CNB)  
Santiago Ramón-Maiques  
(CBMSO)

**Investigadores y centros de  
investigación participantes**

José M. Carazo  
(CNB)  
Carlos González  
(IQFR-CSIC)  
Juan A. Hermoso  
(IQFR-CSIC)  
Fernando Moreno-Herrero  
(CNB)  
Carmen San Martín  
(CNB)

**RESUMEN EJECUTIVO**

La vida depende de un sinnúmero de procesos llevados a cabo por máquinas macromoleculares, formadas principalmente por proteínas y ácidos nucleicos. Dichas máquinas se han ido perfeccionando con la evolución, volviéndose más complejas, reguladas de forma sofisticada e integradas en vías operativas específicas. Entender las bases de la vida y su diversidad evolutiva requiere descifrar la naturaleza atómica de estas macromoléculas, para explicar cómo se sintetizan y ensamblan, y cómo funcionan de forma coordinada, puntual y precisa. En el caso de las proteínas, las principales moléculas operativas de la vida, sus funciones dependen en gran medida de su forma tridimensional y su dinámica. Por tanto, es necesario conocer estos procesos a nivel atómico para comprender en profundidad el mecanismo de acción molecular, incluida su dinámica, lo que nos permitiría encontrar soluciones para su mal funcionamiento y, en última instancia, crear nuevas actividades para nuestro beneficio en biomedicina y biotecnología. En el caso de los ácidos nucleicos, el panorama estructural es mucho más diverso de lo que se pensaba. Los ADN no regulares están surgiendo como estructuras clave en una gran variedad de procesos biológicos, como la transcripción del genoma, la reparación o el mantenimiento de los telómeros. Además, los transcritos de ARN, incluidos los ARN

pequeños y largos no codificantes, parecen regular casi todos los pasos de la expresión genética y tienen amplias repercusiones en el desarrollo y la enfermedad. Paradójicamente, la función de un número cada vez mayor de proteínas, regiones proteicas y también de ARN parece residir en su capacidad de permanecer desestructurados. Estas macromoléculas intrínsecamente desordenadas están implicadas, entre otras cosas, en el fomento de cambios en el estado físico del medio celular (separación de fases), lo que permite la formación de compartimentos celulares sin membrana con múltiples fines, como el almacenamiento transitorio o las funciones de respuesta al estrés. Los estudios futuros debieran intentar comprender qué es lo que desencadena la formación de estos nuevos cuerpos sin membrana y cuál es la relación con su entorno acuoso.

Otro nivel de complejidad es el análisis de cómo, dónde y cuándo se ensamblan y actúan de forma combinada estas máquinas macromoleculares. La célula puede considerarse una fábrica con múltiples compartimentos, a veces de corta duración, cuyo contenido macromolecular depende de las necesidades de la célula y debe controlarse cuidadosamente. Esta disposición subcelular de las macromoléculas y sus correspondientes funciones asociadas —denominada *sociología molecular* de la célula— tiene una finalidad que es preciso reconocer. Resulta especialmente interesante el papel de las membranas no solo como barreras para separar funciones, sino también como puntos de concentración de funciones específicas en las que las proteínas de membrana tienen papeles clave.

Los conocimientos aportados por la biología estructural y las técnicas biofísicas arrojarán luz sobre los cambios evolutivos que, en general, han conducido a las macromoléculas hacia un orden de mayor complejidad, lo que en la mayoría de los casos se traduce en un ajuste más preciso y un mayor control de sus actividades. Un análisis estructural exhaustivo y comparativo de las proteínas, el ARN y los complejos macromoleculares que actúan en procesos similares en diferentes organismos proporcionará información esencial para reconstruir la historia y la evolución de la vida. En última instancia, este conocimiento nos permitirá diseñar nuevos objetos y entidades biológicas, así como aprovechar la biología sintética.

## 1. INTRODUCCIÓN Y DESCRIPCIÓN GENERAL

La vida tal y como la conocemos, con su extraordinaria diversidad de formas, su comportamiento aparentemente intencionado y su capacidad de crecer y reproducirse depende de una gran cantidad de procesos llevados a cabo por máquinas macromoleculares formadas en su mayoría por proteínas y ácidos nucleicos. La vida surgió cuando un conjunto mínimo de estas macromoléculas funcionales se ensambló y, de alguna manera, se compartimentó utilizando las primeras estructuras similares a las membranas, en sistemas autopertuantes. Dichas máquinas macromoleculares se han ido perfeccionando con la evolución, volviéndose más complejas, reguladas de forma sofisticada e integradas en vías operativas específicas. Entender las bases de la vida y su diversidad evolutiva requiere descifrar la naturaleza atómica de estas macromoléculas, explicar cómo se sintetizan y ensamblan, y cómo funcionan de manera coordinada, oportuna y precisa. Este objetivo solo puede lograrse mediante la participación conjunta de grupos de investigación que abarcan una amplia gama de disciplinas. El CSIC se encuentra en una posición privilegiada para abordar este ambicioso desafío gracias a una próspera y coordinada comunidad de investigadores cualificados centrados en diferentes aspectos de la estructura y la función de las macromoléculas.

Las proteínas son las principales moléculas operativas de la vida y desempeñan funciones clave que dependen en gran medida de su forma tridimensional y su dinámica. Para llevar a cabo su actividad, las proteínas emprenden procesos de reconocimiento molecular con otras biomoléculas, en las que las interacciones no covalentes son decisivas. Así pues, es necesario conocer estos procesos a nivel atómico para: 1) comprender en profundidad el mecanismo de acción molecular, incluida su dinámica; 2) encontrar soluciones para su mal funcionamiento; y 3) crear nuevas actividades para nuestro beneficio en biomedicina y biotecnología. Nuestra comunidad científica cuenta con numerosos grupos dedicados a estudiar el vínculo entre la estructura y la función de las proteínas, así como la forma en que estas adquieren su estructura tridimensional final —a veces ayudadas por otras macromoléculas—, y cómo la energía química aviva los cambios conformacionales y mecánicos para ejercer su acción. Muchas enfermedades tienen su origen en proteínas defectuosas con funciones alteradas o propiedades estructurales deterioradas, por lo que nuestra investigación tiene una clara orientación biomédica. Además, las proteínas son los objetivos habituales de los tratamientos médicos, ya sea para restaurar funciones o para combatir enfermedades infecciosas o tumores, y

también pueden manipularse con fines biotecnológicos. Por ello, la obtención de información estructural de las proteínas resulta clave para un mejor diseño de fármacos con aplicaciones terapéuticas o biotecnológicas. Lo mismo puede decirse de los ácidos nucleicos. El panorama de las estructuras del ADN es mucho más diverso de lo que se pensaba hace algunos años, y los ADN no regulares están surgiendo como estructuras clave en una variedad de procesos biológicos, como la replicación del genoma, la transcripción y la reparación, el mantenimiento de los telómeros, etc. Por otra parte, los transcritos de ARN, incluidos los ARN pequeños y largos no codificantes, parecen regular casi todos los pasos de la expresión génica y tienen amplias repercusiones en el desarrollo y la enfermedad (Djebali *et al.*, 2012). El entendimiento de la estructura de las proteínas y los ácidos nucleicos ha hecho posible, por ejemplo, diseñar innovadoras herramientas de modificación del genoma como CRISPR/Cas. A pesar de los continuos esfuerzos de investigación durante décadas, estamos lejos de comprender muchas de las reglas generales que determinan cómo las biomacromoléculas adquieren su forma final. Además, la determinación estructural experimental es una necesidad básica. Por supuesto, la experimentación y el análisis computacional van de la mano para ayudarnos a descifrar las bases estructurales y dinámicas que hay detrás de las interacciones biomoleculares. Este conocimiento nos permitirá diseñar y sintetizar nuevos catalizadores, andamios o fármacos *a la carta* con mayor eficacia, estabilidad o nuevas finalidades, para intervenir o emular procesos biológicos, con lo que nos acercaremos así al desafío de construir vida artificial (sintética).

Paradójicamente, la función de un número cada vez mayor de proteínas, regiones proteicas y también de ARN parece residir en su capacidad de permanecer desestructurados. Tenemos conocimientos muy limitados sobre estas macromoléculas intrínsecamente desordenadas, que pudieran adquirir conformaciones definidas al interactuar con otros componentes celulares o promover cambios en el estado físico del medio celular (separación de fases), lo que permite la formación de condensados biomoleculares —o compartimentos celulares sin membrana— con múltiples propósitos, como el almacenamiento transitorio o funciones de respuesta al estrés. Los estudios futuros deberán abordar el análisis minucioso (*in vitro* e *in vivo*) de cómo las macromoléculas desordenadas introducen orden, qué desencadena la formación de estos nuevos cuerpos sin membrana y cuál es su relación con su entorno acuoso. A pesar de los avances en la resolución de estructuras tridimensionales de proteínas aisladas o ensamblajes, estamos lejos de comprender el número y la versatilidad de los complejos macromoleculares en la célula, y

cuándo, dónde y cómo se ensamblan y actúan de forma combinada. La célula puede considerarse una fábrica con múltiples compartimentos, a veces de corta duración, cuyo contenido macromolecular depende de las necesidades de la célula y debe controlarse cuidadosamente. Se necesita un conocimiento preciso de la homeostasis de las proteínas y los ácidos nucleicos —el fino equilibrio entre la síntesis, el plegado, el mal plegado y la degradación— para extraer algunas reglas generales sobre cuándo y dónde están presentes las macromoléculas. Luego, estas macromoléculas se asocian en complejos más o menos dinámicos y transitorios y con toda probabilidad interactúan con otros conjuntos macromoleculares libres o anclados a membranas, ADN u otras estructuras celulares, para ejecutar multitud de funciones celulares que no se producen al azar. Esta disposición subcelular de las macromoléculas y sus correspondientes funciones asociadas —apodada *sociología molecular* de la célula (Robinson *et al.*, 2007)— tiene una finalidad que es preciso reconocer. Para comprender la disposición dinámica de las estructuras subcelulares y su relación con las distintas funciones y los complejos macromoleculares que las ejecutan, será necesario un enfoque integrador que combine numerosas técnicas y el desarrollo de otras nuevas. Resulta especialmente interesante el papel de las membranas y las paredes celulares como puntos de concentración de funciones específicas, aparte de la separación física que imponen. Además, las proteínas de membrana son moléculas clave en las comunicaciones celulares, desde la transducción de señales hasta el transporte de iones, metabolitos y otras moléculas, y protegen a los organismos vivos de factores tóxicos. Sin embargo, la estructura de las proteínas integrales de membrana y la naturaleza de las interacciones entre membranas y lípidos siguen siendo poco conocidas, a pesar de su clara importancia no solo para entender la propia existencia de la célula, sino también como objetivos evidentes para el diseño de fármacos. Las proteínas de membrana se encuentran entre los tipos de proteínas más difíciles de trabajar, desde la expresión y purificación hasta su análisis por parte de casi cualquier tecnología, lo que crea la paradoja de que somos los que menos sabemos sobre uno de los tipos de proteínas más importantes (Cheng, 2018).

Los conocimientos aportados por la biología estructural y las técnicas biofísicas arrojarán luz sobre los cambios evolutivos que, en general, han conducido a las macromoléculas hacia un orden de mayor complejidad, lo que en la mayoría de los casos se traduce en un ajuste más preciso y un mayor control de sus actividades. Por lo que sabemos, esto se suele conseguir mediante la inserción de nuevos dominios funcionales —en la mayoría de los casos

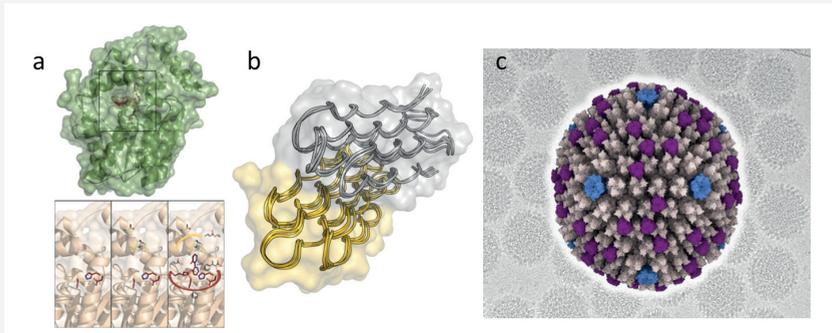
procedentes de otras proteínas—, la incorporación de nuevas subunidades a los complejos y a través de las interacciones estables (o la mayoría de las veces de corta duración) de los diferentes complejos que forman parte de un determinado proceso, con el fin de hacerlo más eficaz y controlable. Un análisis estructural exhaustivo y comparativo de las proteínas, el ARN y los complejos macromoleculares que actúan en procesos similares en diferentes organismos proporcionará información esencial para reconstruir la historia y la evolución de la vida. En última instancia, este conocimiento nos permitirá diseñar nuevos objetos y entidades biológicas, así como aprovechar la biología sintética.

## 2. IMPACTO EN EL PANORAMA DE LA CIENCIA BÁSICA Y EN LAS APLICACIONES POTENCIALES

Desde el descubrimiento de la estructura tridimensional del ADN, la capacidad de dilucidar la estructura de las macromoléculas biológicas ha causado un amplio y profundo impacto en la biología, lo que proporciona una comprensión exhaustiva de las reacciones y los procesos que son fundamentales para la vida.

La importancia de este campo de investigación reside, en parte, en la constatación de que las estructuras macromoleculares plegadas de manera incorrecta y defectuosas son responsables de muchas enfermedades. Se reconoció que la primera *enfermedad molecular*, la anemia falciforme, es el resultado de una única mutación que hace que la hemoglobina se polimerice en filamentos que distorsionan y destruyen los glóbulos rojos. Desde entonces, se sabe que muchas enfermedades moleculares, como el cáncer o los errores innatos del metabolismo, son consecuencia de mutaciones que alteran la estructura y la función de las proteínas. En solo dos décadas, el campo de la biología estructural se ha ampliado desde el estudio de las proteínas solubles a las proteínas de membrana, desde los dominios aislados a los multidominios y los grandes complejos macromoleculares, y desde los retratos fijos a las visiones dinámicas de los comportamientos macromoleculares (Palamini *et al.*, 2016). Este conocimiento ha contribuido de manera fundamental a la comprensión de las enfermedades humanas, tanto comunes como raras, a los mecanismos de las infecciones patógenas y al descubrimiento de remedios para estos males. Las técnicas estructurales están bien establecidas en el sector farmacéutico para el descubrimiento de fármacos, con un impacto reconocido en la identificación de objetivos tanto ortostéricos como alostéricos, y en el diseño racional

**FIGURA 1—**a) Estructura cristalina de la enzima éster hidrolasa eh1ab1 del metagenoma del lago de Arreo (Alonso *et al.*, 2019). Se trata de la primera prueba de una plurizima con dos sitios activos biológicos diseñados, como enfoque para dotar de nuevas propiedades a los andamios enzimáticos. El objetivo consiste en la generación de biocatalizadores de alta eficacia para procesos industriales más sostenibles y con un enorme potencial en muchas otras aplicaciones terapéuticas y de diagnóstico en biomedicina. b) Estructura en solución del dímero de la proteína anticongelante de *pulga de la nieve* obtenida a partir de datos de RMN. Cada monómero se construye sobre seis hélices de poliprolina II, lo que crea una estructura plana y muy rígida que posiblemente será un excelente material nuevo para aplicaciones de biotecnología y biología sintética (Treviño *et al.*, 2018) c) Estructura del adenovirus reptiliano a una resolución casi atómica mediante crioME de partícula única. Imagen facilitada por el Dr. R. Marabini.



y la optimización de fármacos de moléculas pequeñas y proteínas terapéuticas. De los 210 nuevos medicamentos aprobados por la Administración de Alimentos y Fármacos (FDA, por sus siglas en inglés) de los Estados Unidos entre 2010 y 2016, 184 se basaron en la biología estructural (88 %) (Westbrook y Burley, 2019). En combinación con las potentes tecnologías genómicas y proteómicas, los futuros avances en biología estructural, sobre todo en el estudio de las proteínas de membrana y de los grandes conjuntos moleculares flexibles, permitirán comprender mejor las patologías (por ejemplo, las mutaciones de resistencia a la quimioterapia o las bacterias resistentes a los antibióticos), descubrir objetivos más complejos, vías susceptibles a ser moduladas por fármacos aún desconocidas y nuevos modos de actuación que impulsarán el desarrollo de fármacos novedosos y ayudarán a adaptar las terapias en la medicina personalizada.

Además del interés médico, el conocimiento estructural de las macromoléculas y nuestra capacidad para modificarlas, modular su actividad o sintetizarlas a voluntad ofrece ilimitadas aplicaciones biotecnológicas. La modificación estructural de las proteínas para aumentar su eficacia y estabilidad es una práctica habitual en los procesos industriales, así como en la mejora de los

cultivos o la ingeniería microbiana. La modificación de macromoléculas para producir dispositivos nanométricos autoensamblados (nanomáquinas) con propiedades particulares y estructura controlada es también un campo innovador y apasionante en la frontera entre la biotecnología y la biología sintética para generar nuevos materiales, sensores o fármacos *inteligentes* (Schwarz *et al.*, 2017).

### 3. PUNTOS CLAVE DEL DESAFÍO

Hay varios desafíos que deben abordarse para comprender adecuadamente cómo se originan, ensamblan, funcionan, interactúan y evolucionan las moléculas biológicas.

#### 3.1. Plegado y homeostasis de macromoléculas biológicas

Las proteínas son moléculas grandes y complejas, esenciales para la inmensa mayoría de los procesos vitales. Lo que las proteínas pueden hacer depende de su forma, de cómo se ensamblan, se mueven y cambian. A lo largo de las últimas cinco décadas, diferentes técnicas experimentales —por ejemplo, la cristalografía de rayos X (fig. 1a), la resonancia magnética nuclear o RMN (fig. 1b) y la criomicroscopía electrónica o criome (fig. 1c)— han revelado las estructuras tridimensionales de un gran número de proteínas, y aún quedan por determinar muchas más. Estos conocimientos han puesto de manifiesto la delicada conexión entre estructura y función que controla todos los procesos vitales, y han arrojado luz sobre los mecanismos de cómo las proteínas adquieren su forma única.

La capacidad de predecir la forma de una proteína a partir de su secuencia lineal de aminoácidos se conoce como el *problema del plegado de proteínas*. Recientes avances fundamentales en los ámbitos experimental y computacional han despertado el interés por sacar partido a este importante desafío (Senior *et al.*, 2020), ampliando el problema del plegado a proteínas de tamaño creciente, de naturaleza desordenada (véase más adelante) o capaces de adoptar más de una forma nativa (por ejemplo, morfeínas y proteínas metamórficas). Asimismo, es esencial tener en cuenta que la comprensión de cómo se construyen y mantienen las estructuras tridimensionales requiere comprender una gran variedad de mecanismos moleculares que controlan todos los aspectos de la homeostasis de las proteínas, desde la síntesis y el plegado incipiente, pasando por el mal plegado y la agregación, hasta la degradación final. Además de aumentar los conocimientos básicos fundamentales, la consecución

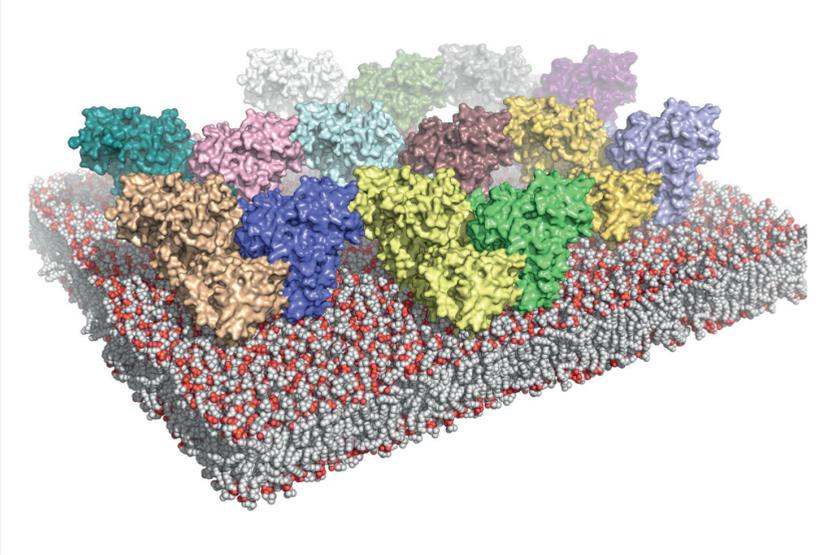
de este desafío tendrá un profundo impacto médico. El mal plegado de las proteínas inducido por el estrés, las mutaciones o los daños provoca trastornos. Una serie de ensamblajes de proteínas mal plegadas y más ordenadas (por ejemplo, agregados, cuerpos de inclusión, agresomas y gránulos de estrés) están relacionados con enfermedades neurodegenerativas (como el alzhéimer, el párkinson, la enfermedad de Huntington o la esclerosis lateral amiotrófica) y con el envejecimiento. La predicción y comprensión precisas del plegado de las proteínas impulsará la investigación de cualquier proceso biológico y guiará el desarrollo de nuevas estrategias terapéuticas.

Nos enfrentamos a un desafío similar en la comprensión de las moléculas de ARN. Durante décadas se ha descuidado el repertorio de dominios estructurales y funcionales del ARN, el cual permanece en su mayoría sin caracterizar. La gran flexibilidad del ARN dificulta la predicción de la estructura y la inferencia funcional, especialmente cuando las moléculas son largas. La comprensión de los andamios bidimensionales de ARN y su ensamblaje estable o transitorio en una estructura terciaria tridimensional, o en varias, es obligatoria para describir procesos biológicos fundamentales con impacto en la medicina y la biotecnología. Los nuevos conocimientos sobre la complejidad y versatilidad del plegado de las proteínas y los ácidos nucleicos abrirán oportunidades para el diseño *de novo* de macromoléculas sintéticas con formas específicas y funciones particulares de relevancia médica y tecnológica (véanse el Desafío 7 y el Desafío 8).

### **3.2. Dinámica estructural y funcional de las máquinas moleculares**

Para entender el funcionamiento, necesitamos estudiar la estructura atómica de las macromoléculas en acción (fig. 1). La biología estructural, con la relevante contribución de la cristalografía de rayos X y la RMN, ha dilucidado las estructuras tridimensionales de un vasto número de macromoléculas (por ejemplo, proteínas, ADN, ARN, etc.) y sus complejos, lo que revela los detalles de los mecanismos catalíticos y reguladores. Sin embargo, nuestra visión actual del espacio macromolecular de la célula es muy limitada, y la forma y organización de muchas macromoléculas, incluida la gran mayoría de las proteínas de membrana (véase más adelante), siguen siendo desconocidas. En los últimos años, los avances tecnológicos, como la criomicroscopía electrónica de alta resolución y los procesos automatizados, han acelerado la investigación estructural de grandes conjuntos macromoleculares. Prevemos que en las próximas décadas, una combinación de enfoques computacionales, experimentales y estructurales logrará el desafío definitivo de obtener informa-

**FIGURA 2**—Polimerización de la proteína quinasa de adhesión focal (FAK) en la membrana celular, donde se activa al interactuar con otras proteínas. La FAK regula la adhesión, la migración y la supervivencia de las células. La comprensión de esta maquinaria asociada a la membrana guiará el diseño de nuevas estrategias para combatir la invasión y la metástasis de las células cancerosas. Imagen facilitada por el Dr. D. Lietha (CIB-CSIC).



ción tridimensional de todas las proteínas y ARN codificados en un genoma.

Se espera que, en los próximos años, el uso de las metodologías estructurales más avanzadas —principalmente la criomicroscopía electrónica y la cristalografía de rayos X—, junto con nuestra capacidad para producir y aislar macromoléculas, aporten un gran número de estructuras nuevas, que representen modelos fijos o, en el mejor de los casos, un número limitado de estados conformacionales poblados. Sin embargo, ir más allá de la imagen estática plantea otro gran desafío, que abarca cuestiones clave pendientes que deberíamos tratar de resolver:

- a. ¿Cuál es la dinámica interna y la flexibilidad conformacional de una macromolécula?
- b. ¿Cuáles son los estados conformacionales activos e inactivos y sus energías relativas?
- c. ¿Qué barreras hay que cruzar para cambiar de estado y cuál es la cinética de estas transformaciones?

- d. ¿Cómo alteran el panorama estructural las mutaciones, los ligandos y otros elementos del entorno celular?
- e. ¿Cómo podemos utilizar este conocimiento para modelar las propiedades macromoleculares?

La comprensión del comportamiento dinámico de las estructuras macromoleculares implicará el desarrollo de técnicas experimentales con resolución espacio-temporal, incluidos la RMN, la espectroscopia fluorescente, los métodos biofísicos de una sola molécula, como las pinzas ópticas y magnéticas (Miller *et al.*, 2018), la microscopia electrónica (ME) ultrarrápida, la cristalografía de rayos X con lapso de tiempo y resolución temporal, y la mayor potencia y robustez de la dinámica molecular y los enfoques computacionales de la estructura. La integración de todos los métodos experimentales y computacionales disponibles en enfoques estructurales híbridos será obligatoria para: i) dilucidar la función y la disfunción de las proteínas con dominios bien estructurados y regiones desordenadas; ii) comprender cómo interactúan las macromoléculas biológicas con el agua, los iones, los lípidos y los efectores de moléculas pequeñas en solución o en la membrana; iii) entender las consecuencias de las modificaciones y mutaciones; y iv) desarrollar nuevos fármacos terapéuticos para combatir, por ejemplo, las bacterias resistentes a los antibióticos, los patógenos emergentes o los tumores resistentes a la quimioterapia.

### 3.3. Proteínas de membrana: plegado, estructura y dinámica

El estudio de las proteínas de membrana es uno de los mayores desafíos en el ámbito de la biología estructural y merece una mención especial. Aproximadamente un tercio de nuestra información genética codifica proteínas (por ejemplo, enzimas, transportadores, canales iónicos, receptores de señalización o transductores de energía) que ejercen su función integrados en las bicapas lipídicas que delimitan los compartimentos celulares. Estas proteínas constituyen objetivos primordiales para los fármacos terapéuticos porque desempeñan funciones esenciales en la célula, como el control del flujo de información y materiales o la mediación de la acción hormonal y los impulsos nerviosos (fig. 2). Dado que las proteínas de membrana están implicadas en muchas enfermedades diferentes (por ejemplo, cardiopatías, alzhéimer, fibrosis quística, etc.), es de vital importancia que se caractericen sus estructuras para que salgan a la luz nuevas estrategias terapéuticas y tratamientos. Sin embargo, de las 8000 proteínas de membrana conocidas aproximadamente que se encuentran en las células humanas, no más de 300 tienen su

estructura atómica determinada, debido sobre todo a la dificultad de producir y aislar estas macromoléculas en condiciones solubles.

Por ello, uno de los principales obstáculos consistirá en optimizar o desarrollar nuevas técnicas para diseñar y producir proteínas de membrana que puedan caracterizarse posteriormente mediante métodos estructurales. Algunas proteínas de membrana son sintetizadas por el ribosoma, al igual que otras proteínas solubles y, luego, tienen que dirigirse a diferentes ubicaciones de la membrana dentro de una célula, mientras que otras siguen estrategias específicas de traslación y translocación. ¿Cómo hace frente la célula a las exigencias únicas y a veces contradictorias que se plantean a las proteínas de membrana en cuanto a su plegado, translocación y estabilidad? ¿Cómo se transportan estas proteínas a la membrana? ¿Cómo dan forma las membranas a la estructura de las proteínas? ¿Cómo interactúan los lípidos específicos y modulan la actividad de las proteínas? ¿Cómo interactúan y se ensamblan las proteínas de membrana con otras proteínas solubles o unidas a la membrana? ¿Cómo se produce la transmisión de señales a través de la membrana? Por lo tanto, las proteínas de membrana presentan tres importantes desafíos para los que la biología estructural pudiera ser de suma importancia:

- a.** Comprender la biogénesis, la maduración y el tráfico de las proteínas de membrana (por ejemplo, la biología estructural pudiera contribuir informando sobre los estados intermedios en la biogénesis o las proteínas implicadas en la translocación y el transporte de los orgánulos).
- b.** Caracterizar la relación estructura-función de las proteínas de membrana. Una cuestión derivada relevante consiste en comprender el papel de los lípidos específicos en la estructura y función de las proteínas de membrana. En este sentido, el aislamiento de las proteínas de membrana con sus propios lípidos evitando el uso de detergentes pudiera resultar crucial.
- c.** Comprender la dinámica de las proteínas de membrana y las interacciones transitorias entre ellas. Lo importante en esta dinámica es entender tanto la dinámica interna de la propia proteína de membrana como la dinámica externa, es decir, su movilidad a través de la membrana. La biología estructural puede hacer contribuciones importantes mediante la caracterización de complejos transitorios, como un receptor unido a un canal o un receptor unido a un

transportador. La caracterización de estos oligómeros pudiera ser de una gran importancia para su validación *in vivo* (por ejemplo, mediante mutagénesis de sitio dirigido) y para el desarrollo de nuevos fármacos (por ejemplo, mediante el diseño específico de fármacos que modulen las interacciones entre proteínas).

### **3.4. Separación de fases líquido-líquido y formación de condensados biomoleculares o cuerpos sin membrana**

Contrariamente a lo que se pensaba, la organización intracelular no se basa tan solo en la formación de compartimentos basados en la membrana. Existen otros compartimentos que dependen de las transiciones de fase líquido-líquido, un proceso por el que un grupo de biomoléculas —proteínas o ARN— cambian de un estado físico a otro que se separa del entorno acuoso general, lo que desencadena la formación de estructuras transitorias sin membrana. Entre estos orgánulos sin membrana, los más conocidos son el centrosoma, los viroplasmas y los cuerpos P en el citoplasma, o los cuerpos de Cajal y el nucléolo en el núcleo. Estos compartimentos que aún no se comprenden bien sirven para múltiples propósitos, como el almacenamiento, las funciones localizadas o el ensamblaje de virus (Heinrich *et al.*, 2018). El objetivo en este caso concreto radica en analizar la estructura fina (*in vitro* e *in vivo*) de los cuerpos sin membrana, para entender qué desencadena su formación y cuál es su relación con su entorno acuoso. Un grupo cada vez más numeroso de proteínas con parte de la secuencia o su totalidad en un estado naturalmente desordenado tiene que ver con la formación de cuerpos sin membrana. No se sabe mucho sobre estas proteínas intrínsecamente desordenadas (IDP por sus siglas en inglés), por lo que hay que utilizar diferentes técnicas para estudiar cómo interactúan entre ellas o con proteínas estructuradas y si adquieren una estructura secundaria al interactuar.

### **3.5. Ensamblaje de redes macromoleculares**

Las macromoléculas rara vez actúan como entidades únicas. La mayoría de los procesos vitales, como la señalización celular, las vías metabólicas, la respuesta inmunitaria o la replicación y transmisión del material genético, dependen de la asociación, a menudo transitoria, de complejos macromoleculares. Para comprender realmente las bases estructurales de la vida nos enfrentamos al desafío de describir estos complejos macromoleculares y los procesos inherentes que controlan su ensamblaje y desensamblaje. ¿Cuáles son las interacciones estables y transitorias entre las macromoléculas y cómo influyen en sus funciones? ¿Cómo cambian estos *interactomas* en el tiempo,

con el tipo de célula (por ejemplo, neurona frente a hepatocito o quiescente frente a proliferante/tumoral) o las condiciones ambientales (por ejemplo, estado metabólico, estrés, patologías, etc.)? ¿Cómo se pueden aislar o estudiar *in vivo*? ¿Existen patrones comunes para la interacción de las proteínas? ¿Podemos predecir las interfaces de reconocimiento? Este desafío requiere impulsar las fronteras tecnológicas de los métodos híbridos estructurales integrados, en combinación con técnicas de imagen celular de superresolución, métodos biofísicos novedosos y datos *ómicos* masivos.

El esclarecimiento de la organización dinámica de las redes de macromoléculas arrojará luz sobre procesos biológicos centrales, como los siguientes: i) el ensamblaje-desensamblaje de los virus; ii) la arquitectura y regulación de la cromatina; iii) la intrincada complejidad de las paredes celulares en plantas o bacterias; iv) la organización, función y capacidad de ser modulados por fármacos de los conjuntos macromoleculares anclados a la membrana; v) los eventos de señalización desencadenados por mensajeros secundarios, como el  $\text{Ca}^{2+}$ , y hormonas; o vi) el reconocimiento de sustancias químicas naturales o fármacos por parte de los receptores de membrana, entre otros.

### **3.6. Observación *in cellulo* de máquinas moleculares en funcionamiento**

El estudio estructural de las macromoléculas de la vida debe tener en cuenta la complejidad del entorno y su impacto en la dinámica estructural de cualquier sistema molecular. Además, algunas macromoléculas son funcionales —o incluso existen— solo en su nicho nativo y determinados conjuntos supramoleculares no pueden extraerse de la célula. Por ello, es necesario mejorar las técnicas actuales y desarrollar nuevas herramientas para estudiar las estructuras macromoleculares en su contexto nativo. Necesitamos métodos no invasivos que proporcionen una visión molecular detallada del entorno intracelular, de la aglomeración molecular y de la distribución y disposición relativa de los conjuntos supramoleculares, incluidos los de la membrana. En resumen, queremos adentrarnos en el espacio macromolecular de la célula para seguir investigando cómo se regulan las acciones.

Este desafío requerirá la elaboración de modelos a distintas escalas, mediante la integración de la criotomografía electrónica de alta resolución y, en general, la criomicroscopía electrónica de resolución temporal, la función molecular *in vitro*, la microscopía de luz correlacionada y la microscopía electrónica, y la información estructural y dinámica acumulada de

componentes macromoleculares aislados con datos *ómicos* celulares (Rout y Sali, 2019). Con toda esta información —que es un tipo de problema complejo de datos masivos, que implica cantidades masivas de información heterogénea— será posible dar un enorme salto cualitativo en nuestra capacidad para reconstruir los interactomas celulares y arrojar luz sobre las vías celulares interconectadas. En las últimas décadas, se han producido increíbles avances a la hora de descifrar las estructuras de las macromoléculas gracias a la biología molecular recombinante y a los métodos estructurales *clásicos*. Sin embargo, hay mucho más que aprender para la próxima generación de enfoques de la biología estructural, partiendo de la ruptura con la simplificación e idealización de la célula como una máquina compuesta de elementos discretos, y la suposición de que al definir la estructura/función de todos y cada uno de estos elementos será posible reconstruir la función del sistema en su conjunto. Nuestros conocimientos actuales solo nos permiten captar la organización molecular enredada, y aún desconocemos en gran medida el número y la complejidad de las interacciones, reacciones y condiciones dentro del medio celular, y cómo modulan las funciones macromoleculares conocidas y las aún por descubrir. En las próximas décadas, utilizaremos los datos experimentales estructurales, dinámicos y topológicos para aplicar la inteligencia artificial con la intención de inferir interacciones moleculares, interconectar vías celulares y construir modelos de sistemas biológicos (véase el Desafío 5).

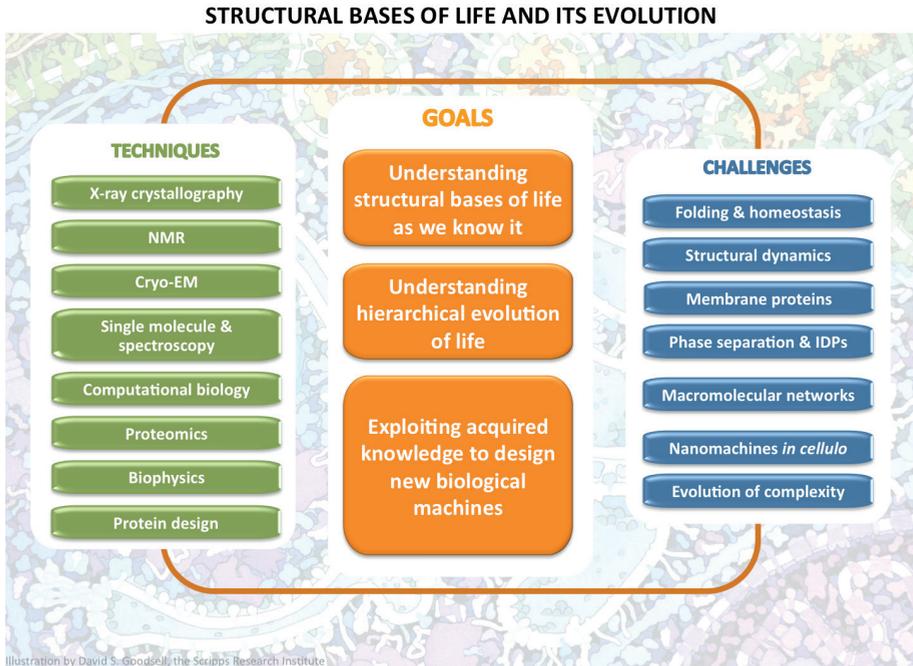
### 3.7. Comprensión de la complejidad de la evolución macromolecular

La evolución de la complejidad de la vida es poco conocida. Aunque admitimos que ha habido un aumento de la complejidad a lo largo de la historia de la vida, el origen de esta complejidad y el propio concepto son difíciles de definir. La biología moderna ha demostrado una profunda relación en los genes, los mecanismos moleculares básicos y las redes de interacción macromolecular entre organismos muy distantes. El análisis estructural comparativo de la maquinaria responsable de procesos similares entre especies nos ayudará a reconstruir la historia y la evolución de la vida, proporcionándonos un marco explicativo que nos permita entender cómo las macromoléculas evolucionaron hacia sistemas más controlables y sofisticados (Bamford *et al.*, 2005; Beck *et al.*, 2018). Este conocimiento, a su vez, puede utilizarse a la larga para diseñar vías enzimáticas sintéticas a medida, módulos de señalización, nanomateriales de autoensamblaje o la construcción de una célula sintética a partir de sus componentes moleculares constituyentes que pudieran incluir nuevos módulos funcionales (véase el Desafío 8).

## DESAFÍO 2 BIBLIOGRAFÍA

- Albert, S., Schaffer, M., Beck, F., Mosalaganti, S., Asano, S., Thomas, H.F., Plitzko, J.M., Beck, M., Baumeister, W. y Engel, B.D. (2017). Proteasomes tether to two distinct sites at the nuclear pore complex. *Proceedings of the National Academy of Sciences* 114, 13726.
- Alonso, S., Santiago, G., Cea-Rama, I., Fernández-López, L., Coscolín, C., Modregger, J., Ressmann, A.K., Martínez-Martínez, M., Marrero, H., Bargiela, R., Pita, M., González-Alfonso, J.L., Briand, M.L., Rojo, D., Barbas, C., Plou, F.J., Golyshin, P.N., Shahgaldian, P., Sanz-Aparicio, J., Guallar, V. y Ferrer, M. (2019). Genetically engineered proteins with two active sites for enhanced biocatalysis and synergistic chemo- and biocatalysis. *Nature Catalysis* 3, 319-328. doi:10.1038/s41929-019-0394-4.
- Asano, S., Fukuda, Y., Beck, F., Aufderheide, A., Förster, F., Danev, R. y Baumeister, W. (2015). A molecular census of 26S proteasomes in intact neurons. *Science* 347, 439.
- Bamford, D.H., Grimes, J.M. y Stuart, D.I. (2005). What does structure tell us about virus evolution? *Curr. Opin. Struct. Biol.* 15 655-663.
- Beck, M., Mosalaganti, S. y Kosinski, J. (2018). From the resolution revolution to evolution: structural insights into the evolutionary relationships between vesicle coats and the nuclear pore. *Current Opinion in Structural Biology* 52, 32-40.
- Cheng, Y. (2018). Membrane protein structural biology in the era of single particle cryo-EM. *Curr. Opin. Struct. Biol.* 52 58-63.
- Djebali, S., et al. (2012). Landscape of transcription in human cells. *Nature* 489(7414), 101-108.
- Efremov, R.G., Gatsogiannis, C. y Raunser, S. (2017). Lipid Nanodiscs as a Tool for High-Resolution Structure Determination of Membrane Proteins by Single-Particle Cryo-EM. *Methods Enzymol.* 594, 1-30.
- Frank, J. (2018). New Opportunities Created by Single-Particle Cryo-EM: The Mapping of Conformational Space. *Biochemistry* 57, 888.
- Gilbert, J.A. y Dupont, C.L. (2010). Microbial Metagenomics: Beyond the Genome. *Annual Review of Marine Science* 3, 347-371.
- Heinrich, B.S., Maliga, Z., Stein, D.A., Hyman, A.A. y Whelan, S.P.J. (2018). Phase Transitions Drive the Formation of Vesicular Stomatitis Virus Replication Compartments. *mBio* 9, e02290-02217.
- Miller, H., Zhou, Z., Shepherd, H., Wollman, A.J.M. y Leake, M.C. (2018). Single-molecule Techniques in Biophysics: A Review of the Progress in Methods and Applications. *Rep. Prog. Phys.* 81(2), 024601.
- Murray, D.T., Kato, M., Lin, Y., Thurber, K.R., Hung, I., McKnight, S.L. y Tycko, R. (2017). Structure of FUS Protein Fibrils and Its Relevance to Self-Assembly and Phase Separation of Low-Complexity Domains. *Cell*, 171(3), 615-627, e16. <https://doi.org/10.1016/j.cell.2017.08.048>
- Palamini, M., Canciani, A. y Forneris, F. (2016). Identifying and Visualizing Macromolecular Flexibility in Structural Biology. *Front Mol. Biosci.* 3, 47.
- Parmar, M., Rawson, S., Scarff, C.A., Goldman, A., Dafforn, T.R., Muench, S.P. y Postis, V.L.G. (2018). Using a SMALP platform to determine a sub-nm single particle cryo-EM membrane protein structure. *Biochim. Biophys. Acta Biomembr.* 1860, 378-383.
- Pfeffer, S. y Mahamid, J. (2018). Unravelling molecular complexity in structural cell biology. *Current Opinion in Structural Biology* 52, 111-118.
- Robinson, C.V., Sali, A. y Baumeister, W. (2007). The molecular sociology of the cell. *Nature* 450, 973-982.
- Rout, M.P. y Sali, A. (2019). Principles for Integrative Structural Biology Studies. *Cell*, 177, 1384-1403.
- Schwarz, B., Uchida, M. y Douglas, T. (2017). Biomedical and Catalytic Opportunities of Virus-Like Particles in Nanotechnology. In Kielian, M., Mettenleiter, T.C. y Roossinck, M.J. (eds.), *Advances in Virus Research. Academic Press*, Vol. 97, pp. 1-60.
- Senior, A.W., Evans, R., Jumper, J., Kirkpatrick, J. et al. (2020). Improved protein structure prediction using potentials from deep learning. *Nature* 577, 706-710.
- Treviño, M.A., Pantoja-Uceda, D., Menéndez, M., Gomez, M. V., Mompeán, M. y Laurents, D.V. (2018). The Singular NMR Fingerprint of a Polyproline II Helical Bundle. *Journal of the American Chemical Society* 140, 16988-17000.
- Westbrook, J.D. y Burley, S.K. (2019). How Structural Biologists and the Protein Data Bank Contributed to Recent FDA New Drugs approval. *Structure* 27, 211-217.

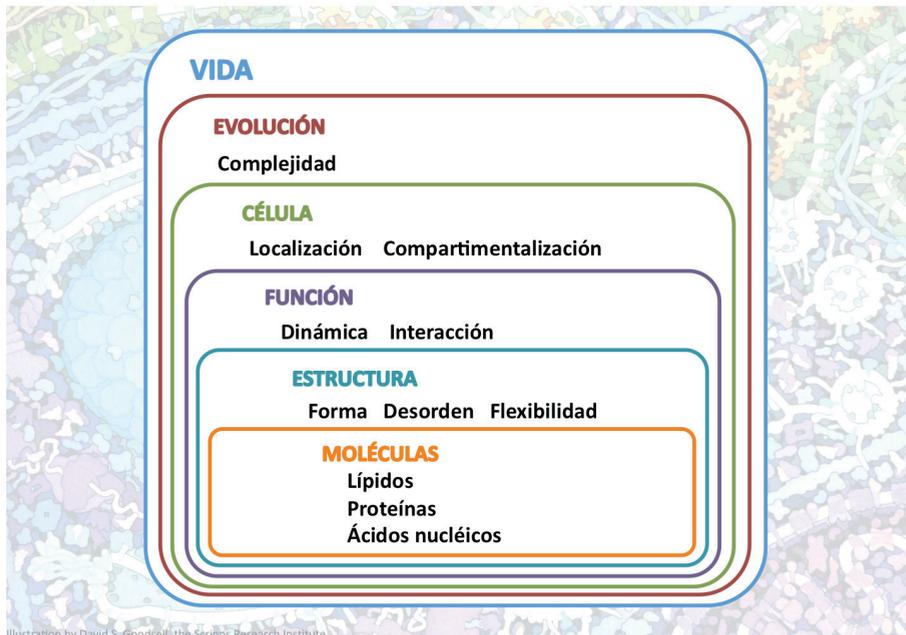
## RESUMEN PARA EXPERTOS



**STRUCTURAL BASES OF LIFE AND ITS EVOLUTION:** BASES ESTRUCTURALES DE LA VIDA Y SU EVOLUCIÓN; **TECHNIQUES:** TÉCNICAS; *X-ray crystallography, NMR, Cryo-EM, Single molecule and spectroscopy, Computational biology, Proteomics, Biophysics, Protein design:* Cristalografía de rayos X, RMN, CrioME (criomicroscopía electrónica), Molécula única y espectroscopía, Biología computacional, Proteómica, Biofísica, Diseño de proteínas; **GOALS:** OBJETIVOS; *Understanding structural bases of life as we know it:* Comprender las bases estructurales de la vida tal y como la conocemos; *Understanding hierarchical evolution of life:* Comprender la evolución jerárquica de la vida; *Exploiting acquired knowledge to design new biological machines:* Aprovechar los conocimientos adquiridos para diseñar nuevas máquinas biológicas; **CHALLENGES:** DESAFÍOS; *Folding & homeostasis, Structural dynamics, Membrane proteins, Phase separation & IDPs, Macromolecular networks, Nanomachines in cellulo, Evolution of complexity:* Plegado y homeostasis, Dinámica estructural, Proteínas de membrana, Separación de fases e IDP, Redes macromoleculares, Nanomáquinas *in cellulo*, Evolución de la complejidad

## RESUMEN PARA EL PÚBLICO GENERAL

### BASES ESTRUCTURALES DE LA VIDA Y SU EVOLUCIÓN





## DESAFÍO 3

### RESUMEN

La biología evolutiva trata de comprender cómo se origina y mantiene la diversidad biológica. La secuenciación de alto rendimiento permite ensamblar genomas a nivel cromosómico, caracterizar transcriptomas a partir de una única célula y determinar modificaciones epigenómicas. Una vez que se aplique de forma generalizada a la diversidad de los organismos vivos, se logrará la reconstrucción del Árbol de la Vida y la identificación de las dianas genómicas de la selección natural.

### PALABRAS CLAVE

<i>filogenómica</i>	árbol de la vida	radiaciones
metagenómica	fronteras bióticas	
genómica evolutiva	método comparativo	
adaptación	homología	
genomas de referencia		
taxones desatendidos		

# EL ÁRBOL DE LA VIDA: ENTRELAZANDO LA GENÓMICA Y LA EVOLUCIÓN

**Coordinadores**

Rafael Zardoya, (MNCN-CSIC)

Ana Riesgo, (MNCN-CSIC)

**Investigadores y centros de investigación participantes**

Silvia G. Acinas, (ICM-CSIC)

Paula Arribas, (IPNA-CSIC)

Damien P. Devos, (CABD-CSIC)

Rosa Fernández, (IBE-CSIC)

José M. Gomez-Reyes, (EEZA-CSIC)

Juan M. Gonzalez-Grau, (IRNAS-CSIC)

Jesús Lozano, (IBE-CSIC)

Borja Milá, (MNCN-CSIC)

Joaquín Ortego, (EBD-CSIC)

Jaume Pellicer, (IBB-CSIC)

Sergio Pérez-Ortega, (RJB-CSIC)

Ramon Rosselló-Mora, (IMEDEA-CSIC)

Gerard Talavera, (IBE-CSIC)

Miguel Verdú, (CIDE-CSIC)

**RESUMEN EJECUTIVO**

La mejora continua de la secuenciación de alto rendimiento abre la posibilidad real de completar con alta calidad los genomas de referencia para todas las especies vivas. Esto permitirá reconstruir un Árbol de la Vida (AdlV) robusto utilizando la filogenómica y mejorar nuestra comprensión de los determinantes genómicos responsables del origen y la diversificación de la vida mediante la genómica evolutiva. Se trata de una tarea ardua no exenta de grandes desafíos que requieren una fuerte colaboración en red y recursos informáticos dedicados a la gestión y análisis de grandes datos. Los principales esfuerzos se centrarán en la obtención de muestras (taxones desatendidos y microbios no cultivados) de las fronteras bióticas, el tratamiento de los genomas gigantes y de proporciones importantes de elementos repetitivos, la identificación de tipos de homología y de ploidía, la detección de las señas genómicas de la selección, la inferencia de las funciones de los genes candidatos y la obtención e incorporación de metadatos a largo plazo sobre historia natural, geología, ecología y medio ambiente en un marco filogenético. Ciertamente, el CSIC se encuentra en una posición privilegiada para abordar tal tarea, ya que cuenta con expertos de renombre que trabajan en muchos linajes microbianos, animales, vegetales y fúngicos, realizando investigaciones punteras en

filogenómica y genómica evolutiva. Mediante el establecimiento de un programa a largo plazo bajo estos auspicios, el CSIC debería ser capaz de contribuir a la catalogación de la biodiversidad, la comprensión del origen de las especies, la revelación de los mecanismos que subyacen a la adaptación evolutiva, la mejora en la conservación de la naturaleza y el descubrimiento en las especies cercanas dentro del AdlV de numerosos metabolitos y fármacos naturales útiles, que son el producto de millones de años de evolución y selección, contribuyendo así al bienestar humano y a un mejor conocimiento del cambio global en la Tierra.

## 1. INTRODUCCIÓN Y DESCRIPCIÓN GENERAL

La teoría de la Evolución de Darwin es un principio unificador de la biología, que establece la selección natural como el principal mecanismo que sustenta el origen y el mantenimiento de la diversidad biológica. Durante más de un siglo, los biólogos evolutivos han estado documentando los innumerables caminos explorados por la selección natural para generar la biodiversidad, a fin de inferir las leyes generales de la evolución. La comprensión de la evolución requiere un enfoque a múltiples niveles para determinar el ensamblaje y la función de los ecosistemas, las interacciones ecológicas y la base genómica de la adaptación. Encontrar sistemas (u organismos) modelo apropiados para estudiar la evolución no es fácil, ya que se trata de un proceso gradual que requiere muchas generaciones para hacerse evidente. Una posibilidad es recurrir a la selección artificial realizada por los seres humanos en taxones, bien con tasas de evolución rápidas (virus y bacterias), o que fueron domesticados en tiempos históricos. Una alternativa es centrarse en los casos en que la selección natural bien aceleró las tasas de diversificación (radiaciones adaptativas) o culminó en soluciones convergentes.

La máxima prioridad de la genómica evolutiva para los próximos años es completar las secuencias de los genomas de todos los organismos vivos a fin de poder delimitar las especies, reconstruir el Árbol de la Vida (AdlV) y realizar análisis comparativos destinados a identificar los determinantes genómicos de la adaptación (Richards, 2015). Se trata de una tarea ambiciosa (hay al menos 1,5 millones de eucariotas descritos y entre 1 y 10 millones de Archaea y Bacteria por describir; Yarza *et al.*, 2014) que debería ser cada vez más factible a medida que las nuevas tecnologías de secuenciación, las herramientas bioinformáticas y los recursos informáticos mejoren más allá de lo que se dispone actualmente (Lewin *et al.*, 2018). Por ejemplo, no fue posible

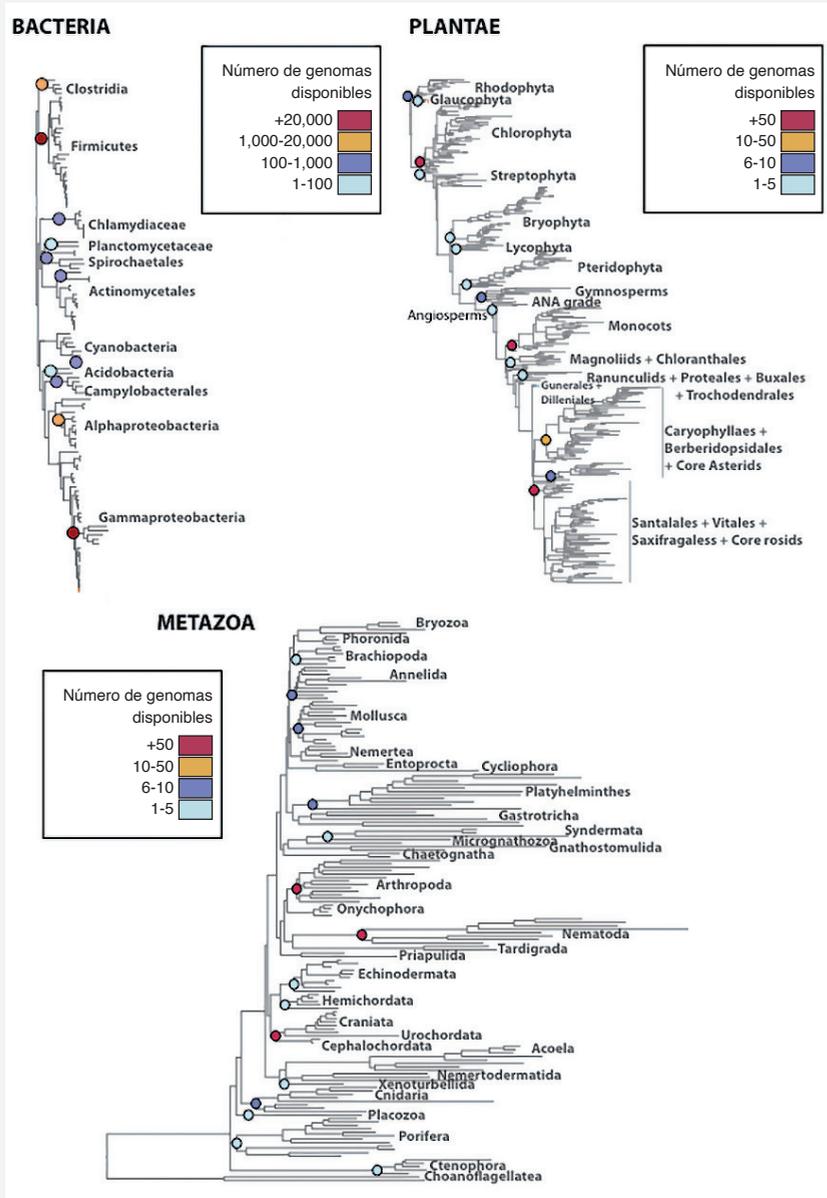
desbloquear el acceso a los genomas microbianos sin necesidad de purificarlos hasta que se elaboró una aproximación metagenómica (por ejemplo, Almeida *et al.*, 2019).

Tras la finalización del genoma humano, y en menos de 20 años, la genómica evolutiva ha experimentado un impulso sin precedentes gracias al continuo perfeccionamiento de las tecnologías de secuenciación de alto rendimiento, que han aumentado constantemente el rendimiento de las lecturas, han reducido considerablemente los costes y han mejorado la calidad general de los resultados (Goodwin *et al.*, 2016). En la actualidad, la secuenciación basada en lecturas cortas se utiliza ampliamente para obtener transcriptomas, resequestrar genomas o generar miles de marcadores filogenéticos mediante la técnica de enriquecimiento de fragmentos diana. Además, actualmente es posible y cada vez más asequible obtener genomas de referencia a nivel cromosómico mediante la combinación de lecturas largas (Amarashinge *et al.*, 2020) y tecnologías que capturan información sobre la cromatina.

A la espera de contar con genomas de referencia de alta calidad que abarquen toda la biodiversidad, se han diseñado una plétora de otros recursos genómicos para poder abordar cuestiones evolutivas específicas de forma rentable. Estas tecnologías siguen siendo útiles y su eficiencia mejorará en los años venideros. Por ejemplo, los mitogenomas se han utilizado ampliamente para reconstruir árboles filogenéticos robustos (por debajo del nivel taxonómico de orden) durante muchos años, y ahora pueden secuenciarse de forma rutinaria en un conjunto de especies a la vez y posteriormente ensamblarse y separarse en las secuencias correspondientes a cada especie presente en la mezcla (la llamada «metagenómica mitocondrial»; Arribas *et al.*, 2019) u obtenerse como subproductos de proyectos de transcriptomas (Plese *et al.*, 2019).

La aplicación de la secuenciación de ARN a los estudios de genómica evolutiva también ha florecido durante el último decenio, ya que los transcriptomas constituyen una buena representación de todo el repertorio genético de un organismo, pero son más baratos de secuenciar y mucho más rápidos de ensamblar, anotar y analizar que un genoma. La filotranscriptómica se ha utilizado con éxito para reconstruir grandes filogenias de forma robusta (por ejemplo, Laumer *et al.*, 2019), pero una de sus limitaciones es la necesidad de disponer de especímenes frescos, ya que el ARN se degrada fácilmente si no se conserva adecuadamente. Para eludir este problema, se han desarrollado varias técnicas de filtrado del genoma con el objetivo de secuenciar con alto rendimiento genes específicos a partir del ADN. Estas técnicas han allanado

**FIGURA 1**—Número de genomas secuenciados en tres clados seleccionados del AdIV. Las filogenias de animales y plantas se han modificado a partir de Laumer *et al.* 2019 y Chen *et al.* 2019, respectivamente.



el camino a la «museómica», la secuenciación de alto rendimiento de muestras conservadas en museos y herbarios, desbloqueando así una nueva fuente de material precioso (series tipo, especímenes raros y recientemente extinguidos) clave para iluminar el AdlV (Trevisan *et al.*, 2019).

El manejo de esos datos de secuencias generadas masivamente ha requerido el desarrollo de recursos computacionales apropiados, que han florecido rápidamente a lo largo de los años. La cornucopia de herramientas bioinformáticas ahora disponibles para ensamblar genomas y transcriptomas permite a los diferentes grupos de investigación construir rutinas informáticas adaptadas a sus necesidades específicas. También existe una amplia variedad de paquetes de programas informáticos para la anotación automatizada, aunque este paso todavía produce muchas secuencias codificantes de proteínas «hipotéticas» debido al carácter incompleto de las bases de datos de genes actualmente disponibles. Las metodologías filogenéticas también se están adaptando fácilmente al uso de grandes conjuntos de datos de secuencias. Por una parte, los genes codificadores de proteínas universales (comunes a todos los organismos) se seleccionan sobre la base de su divergencia evolutiva relativa para maximizar la señal filogenética (Parks *et al.*, 2018). Por otra parte, la filogenómica está cambiando de la concatenación de todos los datos al uso de enfoques basados en la coalescencia que infieren árboles a partir de cada uno de los genes por separado, obteniendo así una visión más compleja de la historia evolutiva de los organismos (Bravo *et al.*, 2019). Finalmente, en los últimos años se han desarrollado numerosos métodos estadísticos para poder profundizar en el conocimiento de los patrones evolutivos usando el método comparativo. Estos métodos utilizan marcos filogenéticos no sólo para reconstruir el pasado (por ejemplo, cómo evolucionó un rasgo, con qué rapidez se diversificó un clado) sino también para informar sobre el presente (por ejemplo, cuántas especies microbianas no cultivadas existen) y para predecir el futuro (por ejemplo, cómo una especie exótica invadirá un ecosistema, cómo un parásito cambiará de huésped).

Una aproximación genómica a la diversidad biológica mundial requiere la colaboración de las comunidades de investigadores a nivel internacional, dado el gran número de taxones a analizar. Los esfuerzos hasta la fecha se han concentrado en los vertebrados, mientras que la gran mayoría del AdlV aguarda atención (Fig. 1). La colaboración en materia de investigación debería funcionar a distintos niveles, desde la coordinación de las labores de muestreo hasta el intercambio de recursos informáticos y rutinas en la nube.

El establecimiento de estándares comunes de alta calidad debería ser otra ventaja de la creación de asociaciones en red. Los esfuerzos de tipo consorcio internacional impulsados por investigadores que trabajan en un taxón determinado deberían complementarse con programas más amplios financiados por instituciones o gobiernos.

Así como la secuenciación del genoma humano produjo un cambio de paradigma en la investigación biomédica, la secuenciación de los genomas de referencia que representan la diversidad de la vida en la Tierra debería provocar una revolución en la biología evolutiva en los próximos años (Richards, 2015). Será posible hacer frente a muchos de los desafíos actuales en este campo, entre ellos la reconstrucción de relaciones filogenéticas robustas, la mejora en la determinación de las relaciones ortólogas y parálogas, la caracterización de las tasas y el modo de evolución de las familias de genes, la comprensión de las dinámicas genómicas, la identificación de los determinantes genómicos de la selección natural, la detección exhaustiva de las innovaciones evolutivas, el reconocimiento de las conexiones causales entre el genotipo y el fenotipo, la caracterización de las regiones y funciones genómicas que responden a los cambios ambientales, etc. En conjunto, la disponibilidad de los genomas de referencia debería sentar las bases para mejorar globalmente la investigación biológica (y la conservación) de grupos anteriormente desatendidos, ya que muchas de las últimas tecnologías sólo son aplicables si se dispone de datos genómicos (Richards, 2015). Una aproximación genómica al AdIV también tiene importantes resultados aplicados y, por ejemplo, debería acelerar el descubrimiento en especies cercanas dentro del AdIV de una variedad de metabolitos y drogas naturales altamente eficientes y específicos, que son producto de millones de años de evolución y selección (véase también Desafío 2 y Desafío C 7).

## 2. IMPACTO EN EL PANORAMA DE LA CIENCIA BÁSICA Y SUS POSIBLES APLICACIONES

La obtención completa de los genomas de referencia que representan la diversidad de la vida debería tener un impacto duradero en la biología básica y aplicada. Un primer resultado importante será la plena resolución del AdIV, que es esencial para cualquier investigación posterior de biología comparada (véase también el Desafío 3 y el Desafío 6). La resolución del AdIV generará un enorme conocimiento no sólo sobre los patrones y las relaciones internas de la biodiversidad, sino también sobre la complejidad y función de los

ecosistemas, y ayudará a descubrir las leyes fundamentales que rigen los procesos evolutivos (véase también el Desafío 4). Este conocimiento debería permitir la conservación de la biodiversidad y maximizar los beneficios para la sociedad y el bienestar humano (Lewin *et al.*, 2018).

Pero todavía hay muchos taxones desconocidos que deben incorporarse al AdIV, así como muchas relaciones internas en los árboles que son recalcitrantes y deben resolverse. La microbiota (Archaea, Bacteria y eucariotas unicelulares; nótese que los virus no pueden ser incluidos en el AdIV, ver Moreira y López-García, 2009) de ambientes extremos o altamente inaccesibles constituye una gran parte de lo desconocido en el AdIV. Una prospección extensa de la diversidad del ARNr 16S indicó que la mayor parte de la diversidad microbiana conocida surge de la exploración de entornos altamente redundantes, mientras que todos los sistemas naturales aún inexplorados constituyen una fuente de novedad continua (Yarza *et al.*, 2014). El número de especies microbianas actualmente clasificadas está seriamente subestimado porque muchas de ellas nunca han sido aisladas en cultivo puro. A este respecto, aceptar que una secuencia de ADN se convierta en material tipo abriría la puerta para clasificar los genomas ensamblados de metagenomas (GEM) y los genomas amplificados a partir de una sola célula (GAS), lo que mejoraría la estima de la diversidad real de especies del mundo microbiano (Rossello-Mora *et al.*, 2020).

Una razón para descubrir la diversidad microbiana desconocida es ampliar los conocimientos generales que se han obtenido a partir de organismos modelo como *Escherichia coli* o *Saccharomyces cerevisiae*. A medida que se identifican más especies microbianas, se hace evidente que la diversidad existente de la biología molecular, celular y funcional en la naturaleza va mucho más allá de lo que se había aprendido de los organismos modelo (por ejemplo, Rivas-Marín *et al.*, 2016). Otra razón importante para impulsar el descubrimiento de especies microbianas es ayudar a discernir la contribución relativa de Archaea y Bacteria al origen endosimbiótico de Eukarya (véase también el Desafío 1). Por ejemplo, recientemente se identificó un nuevo grupo natural de Archaea, el filo Lokiarchaea, mediante el análisis metagenómico de los sedimentos marinos profundos. Los genomas de estos organismos contenían un gran número de características previamente consideradas específicas de eucariotas, aunque las células carecían de la organización celular típica de eucariotas (Imachi *et al.*, 2020). Además, varias filogenias colocaron de forma consistente a los eucariotas dentro del filo Lokiarchaea, de manera que el debate

permanece abierto. Por último, una razón aplicada para promover una amplia prospección del mundo microbiano es que es probable que los microbios sean la fuente de muchas nuevas rutas metabólicas de importancia en los ciclos globales y de muchos productos interesantes de utilidad biotecnológica que podrían detectarse a medida que se desarrollen herramientas bioinformáticas más eficientes.

Dentro de los eucariotas, el estudio de los orígenes de la multicelularidad concentrará muchos esfuerzos en las próximas décadas. La transición evolutiva a la multicelularidad compleja se ha producido de forma independiente en plantas, animales, hongos, algas (verdes, marrones y rojas) y algunos mohos mucilaginosos. La incorporación de los parientes unicelulares en los estudios filogenéticos ha sido de suma importancia para conocer mejor los orígenes de la multicelularidad. Las transiciones a la multicelularidad compleja parecen requerir la cooptación de genes ya presentes en las formas unicelulares ancestrales, que ya eran organismos complejos con componentes de matriz extracelular e intrincadas vías de señalización (por ejemplo, Sebé-Pedrós *et al.*, 2017). La disponibilidad prevista de genomas de referencia de calidad de más linajes unicelulares, junto con la aplicación de tecnologías de edición de genomas (por ejemplo, CRISPR) impulsará sin duda alguna los estudios sobre este tema.

Una fuente relevante de nodos no resueltos en el AdIV está constituida por taxones altamente diversificados originados mediante radiaciones rápidas, que son habituales en diferentes niveles taxonómicos. Además de la dificultad de inferir relaciones filogenéticas debido a que las ramas internas son intrínsecamente cortas, el reto ha consistido en determinar si el polimorfismo compartido entre los taxones de la radiación se debe a una divergencia reciente y a una separación incompleta de los linajes, o si es en parte el resultado de la introgresión y el flujo de genes durante la especiación. Los ensamblajes de genomas completos no sólo proporcionarán una mayor potencia para resolver definitivamente las relaciones filogenéticas en las radiaciones rápidas, sino también para entender el papel de la hibridación debida a la introgresión en la promoción, no en la prevención, de la especiación (Stryjewski y Sorenson, 2017). En este sentido, la genómica comparada es y será ampliada enormemente a nivel de población (es decir, la genómica de poblaciones) con el fin de comprender el papel relativo de la selección natural, la deriva genética, la migración, la introgresión, la separación incompleta de linajes y la demografía en la diversificación de las especies. El continuo de la especiación podrá

estudiarse exhaustivamente con una resolución sin precedentes, desde la fragmentación de la población y la divergencia ecológica hasta la división de linajes y la formación de especies. Esto es de suma importancia para comprender cómo interactúan los organismos con los componentes bióticos y abióticos de la heterogeneidad del paisaje, lo que tiene importantes repercusiones para prever sus futuras respuestas al cambio global.

Conectar el genotipo, el fenotipo y el medio ambiente (véase también el Desafío 4) sigue siendo un gran desafío que se beneficiará de los estudios de genómica evolutiva (Edelaar *et al.*, 2017). La plasticidad fenotípica, la capacidad de un genotipo de producir fenotipos diferentes cuando se expone a diferentes entornos, es una característica omnipresente en la vida. Puede tener importantes consecuencias evolutivas y ecológicas que afectan a las interacciones bióticas y a los nichos ecológicos, así como configuran la coexistencia de las especies y la estructura y dinámica de las redes ecológicas (por ejemplo, Sexton *et al.*, 2017). Sin embargo, el papel de la plasticidad fenotípica en la adaptación y la contribución de los cambios epigenómicos son temas que siguen estando en gran medida sin explorar y tendrán especial relevancia en los próximos años debido a la presión antropogénica (pérdida de hábitat, calentamiento global, especies invasoras, tolerancia a los contaminantes, etc.).

Del mismo modo, la genómica de la adaptación es también un tema floreciente, alimentado por la creciente disponibilidad de genomas de alta calidad de una amplia gama de organismos. La amplia aplicación de los estudios de asociación genómica (EAG) como los que ya se han realizado para comprender la diversidad de las razas de perros (Plassais *et al.*, 2019), será fundamental para asociar las variantes genéticas de las poblaciones a los rasgos fenotípicos que se están seleccionando, identificar las regiones genómicas específicas que intervienen en la restricción del flujo de genes entre las poblaciones, comprender la importancia relativa de los rasgos poligénicos bajo la influencia de muchos loci y aquellos controlados por unos pocos loci de gran efecto, así como evaluar los efectos pleiotrópicos de genes únicos en diferentes rasgos (por ejemplo, Morris *et al.* 2019). Así, un enfoque genómico de la adaptación y la especiación comparando sistemas modelo evolutivos apropiados (por ejemplo, radiaciones adaptativas y/o casos de fenotipos convergentes) ayudará a discernir la importancia relativa de las regiones genómicas reguladoras frente a las codificadoras como dianas de la selección natural, de las innovaciones evolutivas frente a los múltiples cambios acumulativos, de los genes huérfanos frente a las familias de genes, etc.

En los niveles más altos de la organización biológica, el uso de filogenias ha ampliado nuestra comprensión de las comunidades ecológicas, siendo hoy en día un enfoque estándar en los estudios de muchas disciplinas ecológicas como la biología de la conservación, la ecología de comunidades, la biogeografía y la macroecología (Srivastava *et al.*, 2012). La diversidad filogenética puede afectar el funcionamiento de los ecosistemas tan intensamente como la diversidad taxonómica o la funcional. El análisis filogenético basado en rasgos de las redes ecológicas surge como una forma novedosa de incorporar la historia evolutiva de los gremios que están interactuando para poder comprender cómo se ensamblan. Pero la mayor disponibilidad de datos genómicos para muchos taxones debería beneficiar definitivamente a las políticas, la planificación y la gestión de la conservación. Los metacódigos de barras y la metagenómica mitocondrial explotan todo el potencial que ofrece la secuenciación de alto rendimiento para detectar e identificar en cualquier lugar miles de especies de forma simultánea a partir de muestras masivas de organismos o del ADN ambiental (Deiner *et al.*, 2017). Estas herramientas se aplican al estudio de múltiples cuestiones sobre los patrones espaciales y temporales de la biodiversidad, así como a la conservación y gestión de la biodiversidad. De hecho, se ha propuesto su combinación con la tecnología de observación de la Tierra como la forma más prometedora y eficaz de vigilar los impactos de la gestión en la biodiversidad, en sus funciones y servicios (Bush *et al.*, 2017). El hecho de proporcionar un contexto filogenómico a los conjuntos masivos de datos generados por la metagenómica a nivel de comunidad, y en particular a aquellos procedentes de las denominadas «fronteras bióticas», abre una ventana a nuevos análisis, incluido el estudio de la filobetadiversidad, la dinámica de la diversificación o las redes de co-ocurrencia (por ejemplo, Goberna *et al.*, 2019).

### 3. PUNTOS CLAVE DEL DESAFÍO

*Esfuerzos de muestreo a escala mundial:* El muestreo de representantes de las diferentes especies vivas es y será probablemente uno de los problemas más críticos a resolver. El acceso a los taxones sobre el terreno es generalmente difícil y en los peores casos puede ser particularmente costoso en entornos extremos como las profundidades marinas, o arriesgado en regiones políticamente inestables. Hay importantes ecosistemas que apenas han sido explorados y su diversidad es en gran medida desconocida. Entre ellos, las «fronteras bióticas» definitivas se encuentran probablemente dentro del mundo microbiano, así como la mesofauna del suelo y los sedimentos de las

profundidades marinas. En la actualidad, los investigadores realizan en su mayoría muestreos de campo de forma individual sin mayor coordinación. Por consiguiente, existe una necesidad urgente de organizar grandes expediciones multidisciplinarias y colaborativas, concentradas en los lugares críticos de la biodiversidad y las fronteras bióticas. Ejemplos a seguir son las expediciones de Nuestro Planeta Revisado, Tara Oceans y Malaspina, centradas en diversos entornos marinos (por ejemplo, Acinas *et al.*, 2019). Asimismo, el uso de sitios de vigilancia a largo plazo de la biodiversidad para caracterizar la diversidad genómica a escala comunitaria podría desempeñar un papel importante en la adquisición de muestras (por ejemplo, los Observatorios Genómicos; Davies *et al.*, 2012).

La toma de muestras para posteriores análisis genómicos y transcriptómicos necesita metodologías de preservación que garanticen la obtención de ADN de alto peso molecular y ARN intacto, respectivamente. Esto incluye un manejo apropiado en el campo (incluidas las laboriosas disecciones de tejidos) y una preservación adecuada en las colecciones de acuerdo con protocolos estandarizados, que aún no se han aplicado ampliamente. Aunque existen códigos de barras automatizados basados en secuencias para la identificación de especies bien conocidas, la identificación inequívoca de especies poco conocidas, crípticas y desconocidas requiere, en última instancia, sólidas colecciones de referencia de material tipo y la labor dedicada de taxónomos experimentados, que a menudo no existen para grupos desatendidos y muy diversificados. Cuestiones importantes a la espera de coordinación son la necesidad de especímenes de referencia (muestras representativas depositadas y almacenadas en colecciones) y metadatos curados, así como de protocolos de procesamiento e intercambio de datos. En este sentido, se están realizando esfuerzos organizados para secuenciar Bacteria y Archaea (el proyecto del Microbioma Terrestre; Gilbert *et al.*, 2014) y Eukarya (el proyecto del BioGenoma Terrestre; Lewin *et al.*, 2018). Estas iniciativas mundiales utilizan un formato impulsado por la taxonomía, para el que son esenciales las contribuciones de los museos de historia natural, los jardines botánicos, los parques zoológicos y los grandes acuarios. Para acelerar la toma de muestras, estas iniciativas proponen capitalizar el floreciente movimiento de la ciencia ciudadana (alimentado por Internet y los medios sociales) y las nuevas tecnologías robóticas autónomas (Lewin *et al.*, 2018).

*Tamaño del genoma, contenido repetitivo y ploidía.* A pesar de los avances en las tecnologías de secuenciación, hay algunos genomas que, debido a su gran

tamaño, alto contenido de elementos repetitivos, y/o poliploidía, siguen siendo un gran desafío en términos de ensamblaje y anotación. El tamaño del genoma desempeña un papel fundamental como impulsor de la evolución, dadas sus implicaciones en la biología de los organismos (por ejemplo, Pellicer *et al.*, 2018), y es un rasgo fundamental que debe tenerse en cuenta a la hora de diseñar un proyecto de secuenciación, ya que proporciona información esencial para estimar los costes generales, los recursos necesarios y los inconvenientes previstos. Por ejemplo, los ensamblajes de los genomas gigantes del pez pulmonado jaspeado (*Protopterus aethiopicus*, 1C=129,90 Gb) y del lirio (*Paris japonica*, 1C=148,80 Gb) son grandes retos y requerirán el desarrollo de nuevas tecnologías y herramientas informáticas.

La dinámica del tamaño del genoma está regulada principalmente por la frecuencia relativa de la amplificación frente a la supresión del ADN repetitivo y/o la incidencia de la poliploidía. Los elementos repetitivos (elementos transponibles y repeticiones en tándem) constituyen una fracción significativa de los genomas animales y vegetales. Dado el carácter ubicuo de los elementos transponibles, pueden alterar la expresión de los genes mediante inserciones, activar respuestas al estrés que permitan adaptaciones genéticas o influir en la reestructuración cromosómica, entre otros procesos. A medida que los elementos repetitivos se acumulan en el genoma a través del tiempo, es más probable que sufran una erosión que dé lugar a un panorama general de repeticiones degradadas, a menudo conocidas como la «materia oscura» del genoma (Maumus y Quesneville, 2014). Por lo tanto, los genomas jóvenes tendrían perfiles de repetición más homogéneos, mientras que en los genomas gigantes la fracción repetitiva del genoma mostraría una proporción sustancial de elementos no caracterizados y probablemente difuntos. Descifrar la estructura, la función y la dinámica de la materia oscura de los genomas será uno de los principales desafíos de la genómica evolutiva en los próximos años.

Los genomas resultantes de eventos de poliploidía y/o duplicación del genoma completo (DGC) son particularmente interesantes para los estudios evolutivos, pero también son difíciles de ensamblar y anotar debido a sus grandes tamaños e importantes niveles de paralogía. La poliploidía se ha asociado con frecuencia a episodios de hibridación ancestral, y se ha estudiado en gran medida en las plantas debido a sus consecuencias a nivel genómico y fenotípico. El aumento de los datos transcriptómicos y genómicos que se están obteniendo en los últimos años ha puesto de manifiesto que la poliploidía ha sido un fenómeno recurrente en la evolución de las plantas (Landis *et al.*, 2018) y

que también se produjo en las primeras fases de la diversificación de los vertebrados. Tanto en las plantas como en los animales, la poliploidía suele contrarrestarse con reestructuraciones genómicas que dan lugar a la pérdida de una gran fracción del genoma duplicado. Los genes duplicados retenidos pueden adquirir nuevas funciones que den lugar a nuevas formas de adaptación. Sin embargo, el establecimiento de un vínculo entre esos procesos y las explosiones de diversificación mediante el aumento de nuevas adquisiciones fenotípicas ha demostrado ser complejo (Landis *et al.*, 2018) y constituye una interesante línea de investigación para el futuro próximo. Por último, es importante señalar que los seres humanos, a través de la domesticación, han seleccionado desde hace mucho tiempo poliploides para mejorar la acuicultura y los sistemas agrícolas (por ejemplo, fresas, algodón, salmón). La dinámica evolutiva de los genomas poliploides domesticados y sus consecuencias adaptativas son y serán temas de investigación fascinantes con resultados aplicados.

*Asignación de homología.* La homología, o la similitud debida a la ascendencia compartida, es un concepto central en la biología evolutiva. Identificar la homología es clave para comprender lo que ha sido retenido por la selección, y lo que ha cambiado en la estructura y/o función durante la evolución. Dos genes pueden ser homólogos si han surgido por especiación (es decir, ortólogos) pero también si han surgido por duplicación (es decir, parálogos). Las búsquedas de similitud de secuencias no pueden distinguir los genes ortólogos de los parálogos, y ambos de los genes funcionalmente convergentes. La única forma de evaluar la homología es mediante la reconstrucción de árboles filogenéticos, y esto es particularmente difícil cuando se analiza el conjunto completo de genes de un genoma. En el último decenio se ha desarrollado un número notable de algoritmos para inferir los tipos de homología. Una consecuencia indeseable inmediata es que las asignaciones de homología no son en la mayoría de los casos comparables. En el último decenio se ha producido una explosión de secuencias de genomas y transcriptomas de organismos no modelo, pero a menudo, esos conjuntos de datos están incompletos y contienen errores e isoformas no resueltas. Estos pueden violar gravemente los supuestos que subyacen a algunos métodos de inferencia de homología. Por lo tanto, se espera que a medida que se consigan completar más genomas de alta calidad, la determinación de la homología sea más fiable, lo cual es fundamental, ya que la genómica evolutiva necesita distinguir los diferentes tipos de homología, y la reconstrucción del AdIV sólo puede basarse en genes ortólogos.

Un problema relacionado es la identificación exacta de los genes huérfanos, es decir, los genes restringidos a un taxón que no poseen homólogos en ningún otro linaje (véase también el Desafío 2). Algunos linajes animales pueden tener hasta un 30% de genes huérfanos en sus genomas (Fernández y Gabaldón, 2020). Los genes huérfanos pueden surgir de la duplicación, el reordenamiento (incluidas la fusión y la fisión) y una posterior divergencia rápida, pero también de la evolución *de novo* de regiones no codificantes, incluida la traducción de péptidos de evolución neutra (Rödelsperger *et al.*, 2019). Los genes huérfanos también podrían ser el resultado de la pérdida de genes en los linajes troncales durante la evolución. Por ejemplo, recientemente se ha descrito una pérdida masiva de genes en todos los linajes de animales (Fernández y Gabaldón, 2020), y algunos genes que permanecían en clados restringidos podrían haberse convertido entonces en huérfanos. La rápida acumulación de secuencias genómicas casi completas de taxones raros, desatendidos o previamente difíciles de secuenciar en las bases de datos, modificará potencialmente las predicciones sobre el número de genes huérfanos en la mayoría de los linajes.

*Transferencia horizontal de genes.* Uno de los mayores desafíos de la década actual es la evaluación del alcance de la transferencia horizontal de genes (THG) que se produce entre Archaea y Bacteria en sus entornos naturales, y cómo eso afectaría a nuestra visión de la diversidad real de estos taxones. Se ha propuesto que la evolución de los taxones de procariotas no puede describirse completamente sin la THG (Palmer *et al.*, 2019), y que los intercambios genéticos son tan abrumadores que desdibujarían el AdIV al menos para estos taxones. Esto último puede considerarse demasiado extremo, ya que la filogenia más reciente basada en casi 100k genomas reflejaba bastante bien los árboles reconstruidos a partir del gen del ARNr 16S, que se supone se hereda sólo verticalmente (Parks *et al.*, 2018). En la actualidad, existen diferentes plataformas como la base de datos MiGA (Rodríguez-R *et al.*, 2019), que adquieren y organizan genomas de alta calidad a partir de nuevos aislamientos microbianos, así como GEM de muestras ambientales y de microbiomas de especies. Por lo tanto, esto proporcionará acceso a conjuntos de datos estadísticamente sólidos tanto para reconstruir el AdIV de Archaea y Bacteria como para rastrear la THG. En los eucariotas, la detección y validación de la THG han demostrado ser mucho más complejas, y a menudo requieren una alta cobertura de los genomas así como estrictos controles de la contaminación bacteriana. Dicho esto, hay muchos ejemplos de THG en eucariotas (Husnik y McCutcheon, 2018). En los eucariotas, cuando se transfiere un gen completo

de Archaea o Bacteria, se puede conservar la función del gen, ampliando los complementos funcionales del organismo, incluidas las mejoras nutricionales, el suministro de toxinas, la adaptación a entornos extremos y la protección contra los patógenos de Archaea o Bacteria. No cabe duda de que a medida que mejore la calidad de los genomas y se secuencien más taxones desatendidos, se revelará finalmente el alcance y la diversidad de la THG en los eucariotas.

*La discordia entre árboles.* Los análisis filogenéticos basados en los diferentes genes de un genoma o un transcriptoma dan como resultado árboles de genes, que pueden ser diferentes entre sí y pueden alejarse del árbol de especies (Degnan y Rosenberg, 2009). El estudio de la discordancia entre los árboles puede proporcionar información útil sobre el tamaño efectivo de la población de los antepasados, las tasas de divergencia de las especies y una información comparativa sobre la evolución de los diferentes genes a lo largo del tiempo. Dos procesos evolutivos no exclusivos explican la discordancia entre los árboles de genes: la separación incompleta de los linajes (SIL) y la hibridación por introgresión. La SIL o coalescencia profunda se produce cuando el polimorfismo génico intraespecífico permanece en el tiempo después de los eventos de especiación. Se trata de un fenómeno generalizado, que se prevé es mayor cuando los tamaños efectivos de las poblaciones ancestrales son grandes, así como en los casos de divergencia rápida de especies. La introgresión posterior a la especiación, o la incorporación de material genético de un linaje o deme al acervo genético de otro mediante hibridación y retrocruzamiento, es también un fenómeno comúnmente distribuido a lo largo de todo el AdIV (Mallet *et al.*, 2016).

La discriminación exacta de los factores que conducen a la discordancia entre árboles es uno de los principales desafíos de la filogenómica. Dado que la SIL y la introgresión entre taxones estrechamente relacionados pueden producir árboles filogenéticos discordantes similares, la distinción entre ambos procesos es compleja y requiere desarrollar aproximaciones estadísticas apropiadas que utilicen datos de todo el genoma para conjuntos reducidos de taxones (Durand *et al.*, 2011). El desarrollo de estas pruebas estadísticas está aún en sus comienzos y su mejora (por ejemplo, mediante la inferencia de la dirección del flujo de genes en conjuntos grandes de taxones) representa un campo abierto y activo.

Más allá de la inferencia filogenética, la SIL y la introgresión son de gran importancia para comprender los procesos evolutivos que promueven o limitan

la divergencia de las especies. Esto es fundamental para una delimitación precisa de las especies y, por lo tanto, para evaluar adecuadamente la biodiversidad y gestionar las unidades de conservación. La mayor parte de las investigaciones sobre la SIL y la introgresión se han realizado bajo escenarios simulados o conjuntos de datos empíricos relativamente pequeños. Sin embargo, en los próximos años se producirá un florecimiento sin precedentes de los estudios de genómica de poblaciones gracias a la creciente posibilidad de secuenciar los genomas a nivel de población, incluida la obtención de la fase genómica (es decir, la distinción de alelos), lo que ofrece un escenario prometededor para estudiar la SIL y la introgresión a una escala más amplia. Esto implica el uso de métodos filogenéticos basados en la coalescencia (Bravo *et al.*, 2019), que experimentan actualmente un desarrollo activo. Las futuras implementaciones que impliquen la coestimación de la filogenia y el tiempo de coalescencia a escala genómica podrían solventar en gran medida los actuales problemas (por ejemplo, el error sistemático). Esto todavía no es posible y es necesario que sea viable.

*Incorporación de datos fósiles y análisis de relojes moleculares.* Los taxones extintos representados en el registro fósil permiten comprender la evolución escalonada de los caracteres y los planes corporales, restringir mejor los estados de caracteres ancestrales e inferir la cronología de los principales acontecimientos de diversificación. A este respecto, se han generado matrices morfológicas a gran escala, incluso incluyendo rangos estratigráficos o eventos biogeográficos. La incorporación de datos paleontológicos en las filogenias de los taxones vivos tiene un efecto notable en la inferencia de la topología (Koch y Parry, 2019), pero no es sencillo. Durante el último decenio se han desarrollado una gran cantidad de métodos de evidencia total que permiten la estimación y datación simultáneas de las relaciones entre taxones vivos y fósiles utilizando datos moleculares y morfológicos, respectivamente (por ejemplo, Ronquist *et al.*, 2012). Si bien existen modelos realistas de evolución para las particiones moleculares, los modelos de evolución morfológica aún no están plenamente desarrollados y este campo está aún en su infancia. Es importante señalar que la carga computacional de estos enfoques es prohibitiva. En los próximos años se producirá una revolución a medida que los enfoques de pruebas de evidencia total resuelvan los problemas mencionados y entren en la era genómica.

Un árbol filogenético representa tanto las relaciones entre los taxones (topología) como su divergencia relativa respecto de los antepasados comunes más

recientes (longitud de las ramas). Esta última puede utilizarse para inferir la escala temporal evolutiva para el origen y las divisiones de los linajes, y así inferir sobre cuáles fueron las circunstancias (por ejemplo, climáticas, geológicas) que rodearon los procesos de diversificación. La estrategia más utilizada para datar un árbol filogenético consiste en transformar las longitudes de las ramas en tiempo mediante la calibración de ciertos nodos del árbol utilizando la edad de los taxones fósiles (también podrían utilizarse los acontecimientos geológicos del pasado); es el llamado análisis del reloj molecular. Este método ha dado lugar al establecimiento de marcos temporales para muchos linajes en el AdIV (Blair-Hedges *et al.*, 2015). Las divergencias en la datación dependen en última instancia de una justificación filogenética robusta y de una edad geológica exacta de los fósiles. En los últimos años se ha hecho un esfuerzo considerable por reunir datos de fósiles en catálogos curados como la base de datos de paleobiología (PaleoDB), lo que facilitaría la posibilidad de añadir edades a los nodos al mismo ritmo que se reconstruye completamente el AdIV.

*Recursos computacionales.* En general, el número de conjuntos de datos genómicos secuenciados en el último decenio (incluidos los mitogenomas, los clorogenomas, los genomas nucleares, los transcriptomas y los conjuntos de datos de enriquecimiento de fragmentos diana) representa probablemente unos cuantos miles (Fig. 1). Este número aumentará exponencialmente en los años venideros, a medida que la secuenciación de alto rendimiento se haga más barata y esté ampliamente disponible. Por lo tanto, los principales desafíos previstos en esta nueva era de grandes datos están relacionados con su almacenamiento y análisis. Por ejemplo, los datos en bruto de un solo transcriptoma pueden ocupar unos pocos gigabytes, mientras que un genoma necesita medio terabyte de espacio de almacenamiento. Los clústeres de computación no se conciben como unidades de almacenamiento. Por lo tanto, los científicos se enfrentan al reto de encontrar un espacio de almacenamiento adecuado y asequible para mantener sus datos (incluidas las copias de seguridad) a largo plazo, y tenerlos fácilmente accesibles a través de la nube.

El análisis de datos de alto rendimiento será otro gran desafío en los próximos años, ya que el tiempo de computación es un factor limitante crítico. Actualmente, se necesitan meses de computación para completar el ensamblaje y la anotación de grandes genomas. Asimismo, ahora es factible obtener grandes conjuntos de datos filogenómicos (incluidos cientos de taxones y miles de genes) que exigen importantes cargas de tiempo de cálculo. Por ejemplo, un

análisis filogenómico del AdlV animal que contiene 201 especies y 422 grupos ortólogos requirió 1,5 años (realizado en paralelo en 64 núcleos en un clúster de computación con uno de los modelos de mezcla de sustitución de aminoácidos más complejos; Laumer *et al.*, 2019). Además, esto se aplica también a los modelos espaciales explícitos de paisaje y filogeográficos, los análisis de calibración temporal, los análisis genómicos de poblaciones, etc. Por lo tanto, se necesita una transformación revolucionaria en el poder analítico.

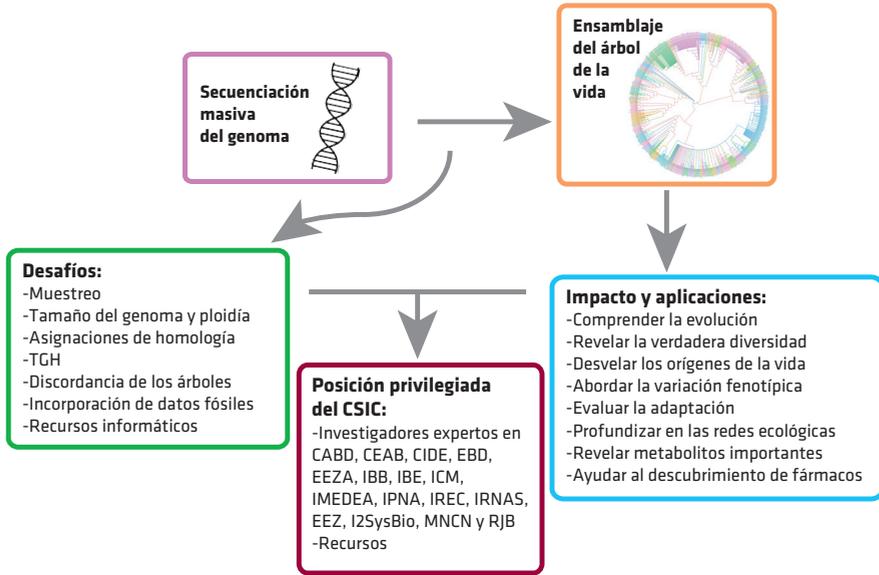
## DESAFÍO 3 BIBLIOGRAFÍA

- Acinas, S. G., Sánchez, P., Salazar, P. G., Cornejo-Castillo, F. M., Sebastián, M., Logares, R., Sunagawa, S., Hingamp, P., Ogata, H., Lima-Mendez, G., Roux, S., González, J. M., Arrieta, J. M., Alam, I. S., Kamau, A., Bowler, C., Raes, J., Pesant, S., Bork, P., Agustí, S., Gojobori, T., Bajic, V., Vaqué, D., Sullivan, M. B., Pedrós-Alió, C., Massana, R., Duarte, C. M. and Gasol, J. M. (2019). Metabolic Architecture of the Deep Ocean Microbiome. *bioRxiv*, P. 635680.
- Arribas, P., Andújar, C., Moraza, M.L., Linard, B., Emerson, B.C. and Vogler, A.P. (2020). Mitochondrial metagenomics reveals the ancient origin and phylodiversity of soil mites and provides a phylogeny of the Acari. *Molecular Biology and Evolution* 37, 683-694.
- Amarasinghe, S.L., Su, S., Dong, X., Zappia, L., Ritchie, M.E. and Gouil, Q. (2020) Opportunities and challenges in long-read sequencing data analysis. *Genome Biology* 21, 30.
- Almeida, A., Mitchell, A.L., Boland, M., Forster, S.C., Gloor, G.B., Tarkowska, A., Lawley, T.D. and Finn, R.D. (2019) A new genomic blueprint of the human gut microbiota. *Nature*, 568, 499.
- Blair-Hedges, S., Marin, S., Suleski, M., Paymer, M. and Kumar, S. (2015) Tree of Life reveals clock-like speciation and diversification. *Molecular Biology and Evolution* 32, 835-845.
- Bravo, G.A., Antonelli, A., Bacon, C.D., Bartoszek, K., Blom, M.P., Huynh, S., Jones, G., Knowles, L.L., Lamichaney, S., Marcussen, T. and Morlon, H. (2019) Embracing heterogeneity: coalescing the Tree of Life and the future of phylogenomics. *PeerJ* 7: e6399.
- Bush, A., Sollmann, R., Wilting, A., Bohmann, K., Cole, B., Balzter, H., Martius, C., Zlinszky, A., Calvignac-Spencer, S., Cobbold, C.A. and Dawson, T.P. (2017) Connecting Earth observation to high-throughput biodiversity data. *Nature Ecology and Evolution* 1, 176.
- Chen, J., Hao, Z., Guang, X. Zhao, C., Wang, P., et al., (2019) *Liriodendron* genome sheds light on angiosperm phylogeny and species-pair differentiation. *Nature Plants* 5, 18-25.
- Davies, N., Meyer, C., Gilbert, J. A., Amaral-Zettler, L., Deck, J., Bicak, M., Rocca-Serra, P., Assunta-Sansone, S., Willis, K. and Field, D. (2012) A call for an international network of genomic observatories (GOs). *GigaScience* 1, 2047-217X-1-5.
- Degnan J.H., and Rosenberg N.A. (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology and Evolution* 24, 332-340.
- Deiner, K., Bik, H.M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S., Bista, I., Lodge, D.M., De Vere, N. and Pfrender, M.E. (2017) Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology* 26, 5872-5895.
- Durand, E.Y., Patterson, N., Reich, D. and Slatkin, M. (2011) Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution* 28, 2239-2252.
- Edelaar, P., Jovani, R. and Gomez-Mestre, I. (2017) Should I change or should I go? Phenotypic plasticity and matching habitat choice in the adaptation to environmental heterogeneity. *The American Naturalist* 190, 506-520.
- Fernández, R. and Gabaldón, T. (2020) Gene gain and loss across the metazoan tree of life. *Nature Ecology & Evolution* 4, 524-533.
- Gilbert, J. A., Jansson, J. K., and Knight, R. (2014) The Earth Microbiome project: successes and aspirations. *BMC Biology* 12(1): 69.
- Goberna, M., Montesinos-Navarro, A., Valiente-Banuet, A., Colin, Y., Gómez-Fernández, A., Donat, S., Navarro-Cano, J.A. and Verdú, M. (2019) Incorporating phylogenetic metrics to microbial co-occurrence networks based on amplicon sequences to discern community assembly processes. *Molecular Ecology Resources* 19, 1552-1564.
- Goodwin, S., McPherson, J. and McCombie, W. (2016) Coming of age: ten years of next-generation sequencing technologies. *Nature Review Genetics* 17: 333-351.
- Husnik, F. and McCutcheon, J.P. (2018) Functional horizontal gene transfer from bacteria to eukaryotes. *Nature Reviews Microbiology* 16, 67.

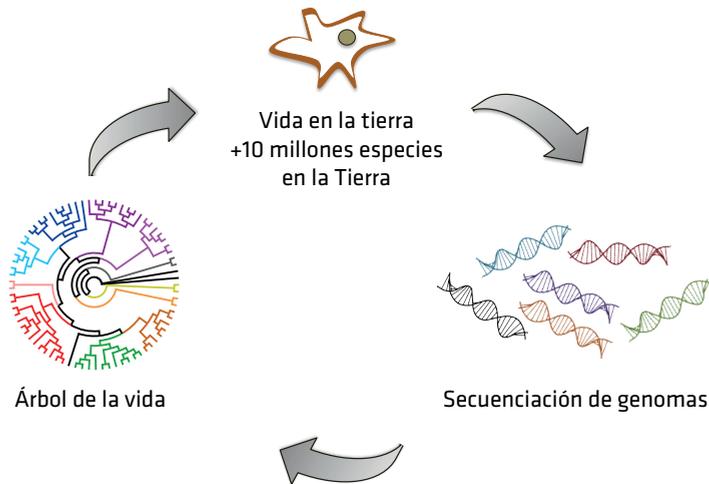
- Imachi, H., Nobu, M.K., Nakahara, N., Morono, Y., Ogawara, M., Takaki, Y., Takano, Y., Uematsu, K., Ikuta, T., Ito, M. and Matsui, Y. (2020)** Isolation of an archaeon at the prokaryote–eukaryote interface. *Nature* 577, 519–525.
- Koch, N. M. and Parry, L. A. (2019)** Death is on our side: paleontological data drastically modify phylogenetic hypotheses. *BioRxiv*, 723882.
- Landis, J. B., Soltis, D. E., Li, Z., Marx, H. E., Barker, M. S., Tank, D. C. and Soltis, P. S. (2018)** Impact of whole-genome duplication events on diversification rates in angiosperms. *American Journal of Botany* 105, 348–363.
- Laumer, C.E., Fernández, R., Lemer, S., Combosch, D., Kocot, K.M., Riesgo, A., Andrade, S.C., Sterrer, W., Sørensen, M.V. and Giribet, G. (2019)** Revisiting metazoan phylogeny with genomic sampling of all phyla. *Proceedings of the Royal Society B* 286, 20190831.
- Lewin, H.A., Robinson, G.E., Kress, W.J., Baker, W.J., Coddington, J., Crandall, K.A., Durbin, R., Edwards, S.V., Forest, F., Gilbert, M.T.P. and Goldstein, M.M. (2018)** Earth BioGenome Project: Sequencing life for the future of life. *Proceedings of the National Academy of Sciences USA* 115, 4325–4333.
- Mallet, J., Besansky, N., and Hahn, M. (2015)** How reticulated are species? *Bioessays* 38, 140–149.
- Maumus, F., and Quesneville, H. (2014)** Deep investigation of *Arabidopsis thaliana* junk DNA reveals a continuum between repetitive elements and genomic dark matter. *PLoS One* 9, e94101.
- Morris, J., Navarro, N., Rastas, P., Rawlins, L.D., Sammy, J., Mallet, J. and Dasmahapatra, K.K., (2019)** The genetic architecture of adaptation: convergence and pleiotropy in *Heliconius* wing pattern evolution. *Heredity* 123, 138–152.
- Moreira, D. and López-García, P. (2009)** Ten reasons to exclude viruses from the tree of life. *Nature Reviews Microbiology* 7, 306–311.
- Palmer, M., Venter, S. N., Coetzee, M. P. and Steenkamp, E. T. (2019)** Prokaryotic species are sui generis evolutionary units. *Systematic and Applied Microbiology* 42, 145–158.
- Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarszewski, A., Chaumeil, P.A. and Hugenholtz, P. (2018)** A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature Biotechnology* 36, 996–1004.
- Pellicer, J., Hidalgo, O., Dodsworth, S. and Leitch, I. J. (2018)** Genome size diversity and its impact on the evolution of land plants. *Genes* 9, 88.
- Plassais, J., Kim, J., Davis, B.W., Karyadi, D.M., Hogan, A.N., Harris, A.C., Decker, B., Heidi G. Parker, H.G. and Ostrander, E.A. (2019)** Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. *Nature Communications* 10, 1489.
- Plese, B., Rossi, M. E., Kenny, N. J., Taboada, S., Koutsouveli, V. and Riesgo, A. (2019)** Trimitomics: an efficient pipeline for mitochondrial assembly from transcriptomic reads in nonmodel species. *Molecular Ecology Resources* 19, 1230–1239.
- Richards, S. (2015)** It's more than stamp collecting: How genome sequencing can unify biological research. *Trends in Genetics* 31, 411–421.
- Rivas-Marín, E., Canosa, I., and Devos, D. P. (2016)** Evolutionary cell biology of division mode in the bacterial Planctomycetes-Verrucomicrobia-Chlamydiae superphylum. *Frontiers in Microbiology*, 7, 1964.
- Rödelsperger, C., Prabh, N. and Sommer, R.J. (2019)** New gene origin and deep taxon phylogenomics: opportunities and challenges. *Trends in Genetics* 35, 914–922.
- Rodríguez-R, L. M., Gunturu, S., Harvey, W. T., Rosselló-Mora, R., Tiedje, J. M., Cole, J. R. and Konstantinidis, K. T. (2018)** The Microbial Genome Atlas (MiGA) webserver: taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. *Nucleic Acids Research* 46, W282–W288.
- Ronquist, F., Klopfstein, S., Vilhelmsen, L., Schulmeister, S., Murray, D. L. and Rasnitsyn, A. P. (2012)** A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. *Systematic Biology* 61, 973–999.

- Rosselló-Mora, R., Konstantinidis, K.T., Sutcliffe, I. and Whitman, W. (2020)** Opinion: Response to concerns about the use of DNA sequences as types in the nomenclature of prokaryotes. *Systematic and Applied Microbiology* 43, 126070.
- Sebé-Pedrós, A., Degnan, B.M., and Ruiz-Trillo, I. (2017)** The origin of Metazoa: a unicellular perspective. *Nature Reviews Genetics* 18, 498-512.
- Sexton, J.P., Montiel, J., Shay, J.E., Stephens, M.R. and Slatyer, R.A. (2017)** Evolution of ecological niche breadth. *Annual Review of Ecology, Evolution, and Systematics* 48, 183-206.
- Srivastava, D.S., Cadotte, M.W., MacDonald, A.A.M., Marushia, R.G. and Mirochnick, N. (2012)** Phylogenetic diversity and the functioning of ecosystems. *Ecology letters* 15, 637-648.
- Stryjewski, K.F. and Sorenson, M.D. (2017)** Mosaic genome evolution in a recent and rapid avian radiation. *Nature Ecology and Evolution* 1, 1912-1922.
- Trevisan, B., Alcantara, D.M., Machado, D.J., Marques, F.P. and Lahr, D.J. (2019)** Genome skimming is a low-cost and robust strategy to assemble complete mitochondrial genomes from ethanol preserved specimens in biodiversity studies. *PeerJ* 7, e7543.
- Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F.O., Ludwig, W., Schleifer, K.H., Whitman, W.B., Euzéby, J., Amann, R. and Rosselló-Móra, R. (2014)** Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nature Reviews Microbiology* 12, 635-645.

## RESUMEN PARA EXPERTOS



## RESUMEN PARA EL PÚBLICO GENERAL





## DESAFÍO 4

### RESUMEN

Entender el desarrollo, mantenimiento, declive y evolución, es decir, la *génesis*, del *fenotipo* es una cuestión fundamental en biología, con implicaciones prácticas para la salud humana, la producción de alimentos o el cambio climático, y que también afecta a diversas áreas de la Ingeniería y las Ciencias Sociales. Se han encontrado desafíos específicos en relación con esta cuestión y se traza una ambiciosa estrategia para posicionar al CSIC como institución líder mundial en la resolución de dichos desafíos.

### PALABRAS CLAVE

fenotipo    codificación    plasticidad

evolución    biología del desarrollo

sistemas complejos

modelación matemática y computacional

teoría biológica    salud    ingeniería

# LA GÉNESIS DEL FENOTIPO

**Coordinadores**

Fernando Casares  
(CABD, CSIC-UPO)

Ignacio Maeso  
(CABD, CSIC-UPO)

**Investigadores y centros de  
investigación participantes**

Josefa González  
(IBE, CSIC-UPF)

Iván Gómez-Mestre  
(EBD-CSIC)

Ángela Nieto  
(IN, CSIC-UMH)

Juan F. Poyatos  
(CNB-CSIC)

## RESUMEN EJECUTIVO

La biología se ha dividido tradicionalmente en varias subdisciplinas enfocadas en diferentes aspectos y escalas biológicas con el objetivo último de comprender cómo la materia viva genera, replica y evoluciona en forma y función. Sin embargo, cuestiones tan complejas requieren, en última instancia, enfoques multidisciplinares que aprovechen los conocimientos generados por las distintas disciplinas y los integren para proponer nuevas preguntas que permitan ampliar las fronteras del conocimiento.

Se han identificado cuatro temas que, por su complejidad, deberían beneficiarse de dicho enfoque multidisciplinario: (i) cómo se almacena la información biológica y cómo se mantiene, hereda y cambia esta información; (ii) cómo esta información codificada impulsa la aparición del fenotipo, es decir, de todos los rasgos y procesos perceptibles de los sistemas vivos; (iii) hasta qué punto es predecible el fenotipo y cuál es nuestra capacidad de ingeniería de los sistemas vivos; y (iv) cómo evolucionan los nuevos fenotipos y cuáles son las escalas temporales y espaciales de los cambios fenotípicos evolutivos.

Responder a estas preguntas debería permitir comprender el desarrollo, el mantenimiento, el declive y la evolución de los fenotipos. Estas cuestiones de investigación básica tienen implicaciones de gran alcance no solo para la salud humana, la producción de alimentos o el cambio climático, sino también para diversas áreas de la ingeniería y las ciencias sociales.

Por último, aunque el CSIC tiene potencial como institución de investigación líder en la resolución de estos desafíos, esto exigirá un cambio importante en la organización, función y gestión de sus centros actuales. Una transformación de esta magnitud solo podría llevarse a cabo mediante la creación de un nuevo instituto de investigación multidisciplinario.

## 1. INTRODUCCIÓN Y DESCRIPCIÓN GENERAL

El modo en que la materia viva genera, replica y evoluciona la forma y la función en todos los niveles de la organización biológica es una de las cuestiones centrales de la biología (Thompson, 1945), que, sin embargo, está lejos de tener respuesta. La materia viva se caracteriza por un alto grado de interdependencia entre sus componentes, así como entre estos y el entorno. Asimismo, posee un fuerte carácter dinámico, ya que experimenta ciclos de desarrollo y reproducción a corto plazo y cambios a largo plazo a través de escalas de tiempo geológico (evolución) que, a su vez, se basan en las variaciones heredables del desarrollo. La vida es un **proceso dinámico y complejo** (Goodwin y Sole, 2002; von Bertalanffy, 1968). Para examinar esta complejidad, la biología se ramificó en varias subdisciplinas que se centraron en diferentes aspectos y escalas biológicas. En particular, e independientemente de que se discuta la estructura formativa de las proteínas o de los ácidos nucleicos, los diferentes tipos de células, la arquitectura de los tejidos y la homeostasis, la alometría de las partes del cuerpo o el repertorio conductual de los organismos (todos ellos atributos del organismo y parte de su *fenotipo*), todas las disciplinas biológicas abordan cuestiones comparables.

Entre ellas, se incluyen el análisis de los numerosos elementos que pueden almacenar información biológica, cómo varían, cómo estos elementos interactúan y contribuyen a definir el fenotipo, y cómo evoluciona todo este sistema. El fenotipo de un organismo es lo que determina su éxito relativo con respecto a otros individuos conespecíficos (es decir, su *aptitud* o «fitness») y cómo interactúa con otros organismos para acabar dando forma a la red de sus interacciones ecológicas. A pesar de su papel fundamental, nuestros conocimientos sobre la aparición del fenotipo son todavía muy limitados. Explicar la génesis, el mantenimiento y el declive de los fenotipos representa, de este modo, el objetivo último de la investigación biológica en el siglo XXI. El CSIC aspira a desempeñar un papel de liderazgo en ese nuevo campo de la biología. Además, los principios que se establecerán siguiendo este desafío podrían tener implicaciones en muchas otras disciplinas científicas, desde la ingeniería

hasta las ciencias sociales. Por ejemplo, muchos aspectos de la robótica acaban enfrentándose al problema de la definición del fenotipo o al diseño de procedimientos de adaptación de este a condiciones fluctuantes. Del mismo modo, muchas estructuras sociales, desde las empresas hasta las ciudades, experimentan el problema de cómo la información codificada emerge dinámicamente en función del conjunto.

La genética ha sido una disciplina puntera en la búsqueda de algunas de estas cuestiones, con el descubrimiento de que los principales elementos que almacenan la información biológica son los *genes*, entendidos como secuencias específicas de ADN que contribuyen al *genotipo*. La variación de la secuencia del ADN es una fuente importante de cambio evolutivo y adopta muchas formas, desde los polimorfismos de nucleótido único (SNP), pasando por las variaciones en el número de copias (CNV), como inserciones, deleciones, duplicaciones genómicas y reordenamientos genómicos, hasta la incorporación de ADN extraño, por ejemplo, a través de la hibridación de especies (introgresión) o las invasiones de elementos transponibles y otros fenómenos de transferencia lateral de genes. Incluso dentro de este marco aparentemente sencillo, quedan muchos interrogantes sobre la aparición del fenotipo. Los genes interactúan entre sí durante el desarrollo y en los procesos fisiológicos a múltiples niveles (epistasia) (Lehner, 2011). Además, hay que reconocer que la función de la mayoría de los genes no es única y que pueden influir en varios rasgos fenotípicos distintos (pleiotropía) (Saltz *et al.*, 2017). A menudo, el efecto de los genes en el fenotipo depende del entorno (interacciones gen-ambiente) (Eguchi *et al.*, 2019), y los cambios inducidos por el entorno en la expresión genética pueden dar lugar a nuevos fenotipos e innovaciones evolutivas (Moczek *et al.*, 2011; West-Eberhard, 2005). Así, la combinación de diferentes «drivers» (impulsores) de entrada (internos y externos) a lo largo de la vida de un individuo influirá en el fenotipo y, especialmente relevante para los seres humanos, en la calidad de vida y en las expectativas de envejecimiento saludable (Connallon y Hall, 2018; Lehner, 2013). Actualmente se reconoce que la experiencia de los progenitores puede arrastrar consecuencias a la prole (efectos transgeneracionales). Por lo tanto, el fenotipo resultante en cada individuo también podría afectar a la población futura (Jablonka y Lamb, 2005). Estas ideas ayudan a reconocer la complejidad de las cuestiones planteadas en este desafío. **Se han identificado cuatro temas de investigación** que deberían permitir a los investigadores del CSIC ampliar sus conocimientos sobre la génesis del fenotipo.

## 2. IMPACTO EN EL PANORAMA DE LA CIENCIA BÁSICA Y EN LAS APLICACIONES POTENCIALES

*El primer tema* corresponde a **cómo esta se mantiene, hereda y cambia**. Esta *codificación* ya no está vinculada a una *lista de genes*. De hecho, la idea misma de genotipo está en tela de juicio, ya que está claro que la estructura primaria del ADN (es decir, su secuencia lineal) no es la única plantilla para almacenar información biológica. Por ejemplo, además de la variación en la estructura primaria del ADN (tanto del ADN nuclear como del de los orgánulos), las modificaciones epigenéticas de diversa índole, incluidas las modificaciones post-traduccionales de histona y el ARN regulador no codificante, contribuyen a la base heredable de nuestros fenotipos, al menos en escalas temporales cortas (Harvey *et al.*, 2018; Skvortsova *et al.*, 2018).

Hay varios factores adicionales que contribuyen a la información biológica, como los efectos maternos en forma de orgánulos subcelulares, las hormonas y el suministro de nutrientes en la yema de huevo, los efectos intergeneracionales y transgeneracionales de los padres, etc. ¿Qué elementos se deben tener en cuenta entonces para captar plenamente la codificación del fenotipo? ¿Existe una jerarquía de codificación natural que ayude a almacenar fragmentos fundamentales de la información fenotípica fijada? ¿Existen estrategias particulares de codificación que favorezcan la variación cualitativa y cuantitativa? La resolución de las preguntas de este tema conducirá a una redefinición del *genotipo* que tendrá importantes repercusiones en el futuro de la genética.

*El segundo tema* va desde las preguntas sobre el almacenamiento de la información **hasta cómo esta información codificada impulsa la aparición del fenotipo** (la *interpretación*) y cómo las condiciones ambientales modulan dicho impulso. ¿Qué tipo de mecanismos regulan la aparición de fenotipos en los sistemas biológicos? Es probable que no se pueda observar toda la información codificada en todos los individuos o en muchos de los entornos que estos individuos pueden ocupar. Si este es el caso, hay que abordar la gama de fenotipos que puede mostrar un organismo. ¿Cuáles son los límites de los fenotipos que puede desarrollar un organismo? ¿Hasta qué punto el entorno es un inductor de fenotipos además de un filtro selectivo? ¿Y cuáles son las escalas temporales en las que el medioambiente ejerce sus efectos? ¿Cómo interactúan los distintos factores que conforman el fenotipo a lo largo del desarrollo? Estas preguntas conducen, por supuesto, a problemas más específicos: cómo se toleran las mutaciones deletéreas en las células (Eyre-Walker y Keightley, 2007), cómo cambian varios rasgos simultáneamente (Cheverud, 1988),

si es modular el mapeo desde la codificación hasta el fenotipo (Wagner *et al.*, 2007); cómo las condiciones ambientales ajustan la expresión de la información (Arnold *et al.*, 2008); cuáles son los costes y los límites asociados a la plasticidad fenotípica (Murren *et al.*, 2015); si existe una relación jerárquica entre las fuentes de variación fenotípica que contribuyen a ella (Phillips y Arnold, 1999), y cómo varían los fenotipos a lo largo del ciclo de vida del individuo y cómo degeneran durante la enfermedad y el envejecimiento (Bahar *et al.*, 2006), por nombrar algunos. Todos estos factores, internos y externos al organismo, contribuyen a determinar cómo se crea y mantiene el fenotipo, todo ello en una red muy compleja de interacciones que a menudo actúan de forma sinérgica y con retroalimentación, lo que da lugar, por tanto, a una dinámica no lineal. Por consiguiente, este segundo tema explorará de manera significativa cómo *envejece* el fenotipo (Metcalf y Alonso-Álvarez, 2010; Wiley y Campisi, 2016) y cómo se desvía de un fenotipo normal y saludable un estado de enfermedad.

*El tercer tema aborda la previsibilidad del fenotipo y la capacidad de producir mediante ingeniería de sistemas vivos para producir fenotipos de diseño.* Si se conoce mejor la génesis, el mantenimiento y el declive de los fenotipos, ¿será posible predecir el fenotipo a partir del conocimiento del entorno y de un genotipo completo (definido más allá de la secuencia de ADN actual)? ¿Pueden diseñarse nuevos fenotipos y producirlos por ingeniería? Se podría anticipar que este es el caso y asumir dos posibles estrategias. En la primera, se puede producir una descripción fenomenológica de los fenotipos biológicos. Para ello, se deben identificar las variables adecuadas para representar los fenotipos, es decir, reducir la complejidad (Anderson, 1972). Sin embargo, no basta con la identificación de las variables, ya que es necesario integrarlas en un sistema dinámico, lo que requerirá nuevos enfoques de modelación matemática y computacional, por ejemplo, incorporando técnicas modernas de asimilación de datos, que conectan los modelos y los datos experimentales para comprender mejor la dinámica. Una segunda estrategia podría requerir el desarrollo de sofisticadas herramientas de inteligencia artificial que ayuden a identificar toda la información codificada que influye en la expresión del fenotipo, de modo que puedan identificarse patrones (Topol, 2019). Pudiera ser también que la capacidad de diseñar sistemas biológicos sea viable, pero sin una comprensión fundamental clara (por ejemplo, los aviones vuelan, pero sin una explicación precisa de por qué lo hacen las aves). Además, pudiera ser que existan limitaciones intrínsecas en la capacidad de predecir fenotipos (Waddington, 1957). Una extensión de esta investigación conducirá a la

ingeniería de fenotipos *in vitro*. Gracias a nuevas técnicas que permitan el cultivo, la diferenciación y la edición genética de células madre pluripotentes, se podrían implementar fenotipos diseñados para la mejora, la reparación, la sustitución de órganos o incluso el desarrollo de nuevos *biobots* con capacidades programadas e incorporadas (Kriegman *et al.*, 2020), lo que también podría incluir detalles sobre la biomecánica de dichas organizaciones celulares. Este tema llevará la investigación a aplicaciones biomédicas para la reparación de tejidos dañados, enfermos o envejecidos, la mejora de órganos existentes y el diseño de fenotipos novedosos.

*El cuarto y último tema se centra en la evolución.* Explicar la génesis del fenotipo significa entender cómo el grado de cambio en los elementos que codifican la información se conecta con la variación de los fenotipos sobre los que actúa la selección, y cómo surgen los fenotipos novedosos. Con anterioridad, se ha argumentado que el mapeo desde la información almacenada hasta el fenotipo puede restringir parte de esta variabilidad, pero se desconocen qué mecanismos contribuyen a las propiedades que definen dicha asignación y cómo estos evolucionan (Uller *et al.*, 2018). ¿La arquitectura de los sistemas biológicos promueve la evolución del fenotipo (Watson *et al.*, 2016)? ¿Es predecible esta evolución? También es necesario conocer bien las escalas espaciales y temporales de la dinámica evolutiva (Catullo *et al.*, 2019; Segar *et al.*, 2020), así como la manera en que la evolución pudiera transformar la codificación en el mapa del fenotipo (Martin *et al.*, 2015). Por el contrario, ¿existen peculiaridades especiales de este mapa que estimulen la evolución de arquitecturas complejas? Este mapa no será un mapa estático de *fenotipos finales*, sino más bien dinámico en el que cada fenotipo se considerará como una *trayectoria* temporal.

*Impactos más amplios.* Hay una serie de áreas en las que este programa de investigación tendrá un impacto significativo:

#### a. *Biología y salud humanas*

**1.1 Medicina personalizada.** Comprender la descodificación de los «drivers» del fenotipo permitirá, en el caso de las enfermedades humanas, utilizar los componentes genéticos y no genéticos específicos de cada paciente para entender las especificidades de cada patología y proponer intervenciones personalizadas para dirigir el fenotipo enfermo hacia uno sano.

**1.2 Reparación de tejidos.** La comprensión de los mecanismos que regulan la homeostasis fenotípica ayudará a diseñar mejores estrategias para promover la reparación y regeneración de los tejidos.

- 1.3 Ingeniería de tejidos y órganos.** La comprensión de los mecanismos que controlan las operaciones de desarrollo, combinada con estudios computacionales y sistemas *in vitro*, permitirá el diseño y la síntesis controlados de tejidos y órganos a medida para su sustitución. Para ello, serán esenciales nuevas estrategias de modelación matemática y computacional.
- 1.4 Calidad de vida y envejecimiento saludable.** Todos los conocimientos obtenidos a partir de lo anterior pueden tener un gran impacto en el diseño de mejores estrategias terapéuticas para estados patológicos. Y, lo que es más importante, también contribuirán a la promoción de la calidad de vida y, en particular, del envejecimiento saludable.
- b. Ecología y cambio global.** Al identificar los «drivers» del fenotipo y su efecto en el desarrollo del organismo, se pueden predecir los efectos del cambio ecológico (incluida la temperatura) en los organismos y, por tanto, anticipar las consecuencias del cambio global. Comprender cómo afectan estos cambios al fenotipo también debería permitir anticipar qué especies son más vulnerables y cuáles más capaces de hacer frente al cambio ambiental.
- c. Producción animal y vegetal.** La producción agrícola de productos animales y vegetales se beneficiará de una mejor comprensión de la generación de algunos de los rasgos con valor comercial, nutricional o biomédico, así como de una mejor comprensión de la ingeniería necesaria para mejorar estos rasgos o generar otros nuevos.
- d. Exobiología.** En la búsqueda de vida extraterrestre, la comprensión del principio de almacenamiento, codificación y decodificación de la información necesaria para la creación, mantenimiento y evolución de la vida terrestre podría ayudar a buscar principios organizativos o informativos similares en otros lugares, independientemente de su implementación física.
- e. Impactos transdisciplinarios.** Varias disciplinas contemporáneas investigan la compleja dinámica de muchas partes que interactúan, con el objetivo de predecir y diseñar su comportamiento, entre ellas la robótica (por ejemplo, el comportamiento de la robótica de enjambres), las ciencias sociales (por ejemplo, las ciudades inteligentes, los sistemas de organización social o las redes sociales) y la ingeniería (por ejemplo, las redes energéticas inteligentes o las redes de información). Los principios derivados del estudio de los «drivers» del fenotipo, su decodificación y los plazos de acción pueden ayudar a establecer analogías útiles en esos campos. De forma recíproca, el estudio de la

organización biológica y de la aparición y variación del fenotipo debería beneficiarse del intercambio de estos otros ámbitos. La integración a gran escala de las interacciones mutuas y constantes entre los sistemas biológicos y el entorno (tan complejo como las ciudades, los sistemas de asistencia sanitaria, los factores culturales o el desarrollo tecnológico) dará una visión global de cómo surge y cambia el fenotipo.

- f. Impacto social y ético y políticas públicas.** Entre los diferentes marcos teóricos y conceptos desarrollados dentro de las ciencias de la vida, los asociados al presente desafío (evolución, mapas genotipo-fenotipo, etc.) son probablemente los que tienen un impacto más fuerte y duradero en la sociología, la filosofía e incluso la política y las creencias religiosas, como la profunda transformación cultural que desencadenó Darwin con la publicación de *El origen de las especies*. Lo mismo ocurre con muchos de los avances tecnológicos asociados a este desafío (bioingeniería, CRISPR, medicina personalizada, etc.). No obstante, este alto impacto social tiene un coste y existen varios ejemplos históricos en los que, debido a esfuerzos insuficientes o inadecuados en la comunicación de la ciencia o a casos de mala conducta dentro o fuera de la comunidad científica, el uso indebido de este impacto ha acarreado drásticas consecuencias en la sociedad y en el diseño de políticas públicas (desde la eugenesia durante el siglo XX hasta el uso del determinismo genético como justificación de las políticas educativas y la desigualdad de ingresos en la actualidad). Así, los futuros conocimientos que se adquieran en los próximos 50 años acerca de la génesis del fenotipo deberán ir acompañados de esfuerzos interdisciplinarios con los científicos sociales (como los del Instituto de Políticas y Bienes Públicos del CSIC) para prevenir estos problemas y fomentar un impacto beneficioso en la sociedad. No solo cambiarán nuestras ideas filosóficas. La posibilidad de predecir el curso del desarrollo humano o la capacidad de diseñar órganos u organismos (incluidos nosotros mismos) tendrán un enorme impacto en la organización y evolución de nuestras sociedades, y exigirán grandes debates políticos, filosóficos y éticos.

### 3. PUNTOS CLAVE DEL DESAFÍO

*Desvelar todo el repertorio de «drivers» del fenotipo.* En los últimos 70 años, nuestro conocimiento de los mapas genotipo-fenotipo se ha transformado por completo desde el descubrimiento del ADN como material genético. Esta visión centrada en el ácido nucleico ha sido increíblemente productiva pero,

para comprender plenamente cómo se generan y evolucionan los fenotipos, será necesaria una perspectiva más amplia en el futuro. Cada día se describen más formas en las que los sistemas vivos almacenan la información para la creación y el mantenimiento del fenotipo del organismo, desde las modificaciones postraduccionales de la histona hasta los priones y la polaridad celular (Ciabrelli *et al.*, 2017; Duempelmann *et al.*, 2019; Harvey *et al.*, 2020; Jung *et al.*, 2017; Klosin *et al.*, 2017; Lev *et al.*, 2019; Manzano-López *et al.*, 2019; Posner *et al.*, 2019; Shirokawa y Shimada, 2016), muchas de las cuales no pudieron investigarse antes con más detalle por limitaciones técnicas. ¿Pueden entenderse plenamente los fenotipos si nos centramos exclusivamente en el ADN? ¿Cuántas otras estructuras biológicas se han pasado por alto que pudieran tener el potencial de actuar como factores hereditarios? ¿Tienen propiedades en común? ¿Cuál es su contribución relativa al fenotipo en comparación con el ADN? ¿Operan todos los sistemas de memoria sin ADN a las mismas escalas temporales y de organización? ¿Han desarrollado diferentes linajes de forma independiente distintos sistemas hereditarios adicionales? ¿Pueden diseñarse nuevos sistemas sintéticos?

*Pruebas experimentales de la jerarquía de «drivers» del fenotipo.* Se utilizarán diferentes sistemas de estudio susceptibles de experimentación para comprender cómo los diferentes «drivers» de la variación fenotípica, como las variantes genéticas, las contribuciones maternas o paternas, la inducción ambiental y las interacciones bióticas, contribuyen cada uno de ellos de forma adicional o sinérgica a la generación de la variación fenotípica. Al centrarse en el análisis de un único rasgo relevante para todos los organismos (por ejemplo, la respuesta al estrés), debería ser posible comprender cómo los diferentes organismos han encontrado soluciones para los mismos desafíos. Este conocimiento de las diferentes estrategias disponibles en la naturaleza pudiera servir para diseñar algunas de estas soluciones en los organismos que nos interesen (por ejemplo, los seres humanos en el contexto de las enfermedades o los animales y plantas en el contexto de la producción de alimentos).

*Definir el alcance de los fenotipos.* Los organismos biológicos constituyen la base sobre la que siempre se han planteado las relaciones genotipo-fenotipo. Sin embargo, las definiciones y los límites tradicionales de los organismos biológicos se encuentran actualmente en tela de juicio. Ahora se están definiendo conceptos más inclusivos de los organismos biológicos, como el de *holobionte* (Roughgarden *et al.*, 2017), en un intento de dar cuenta de la complejidad de los sistemas multiespecíficos co-construidos, y es probable que en el

futuro se desarrollen conceptos aún más amplios. ¿Cómo se pueden describir los fenotipos de estos sistemas supraorgánicos? ¿Es posible aplicar el concepto de fenotipo a escalas mucho más amplias de lo que se ha hecho hasta ahora? En caso afirmativo, ¿cómo se generan o desarrollan estos fenotipos supraorgánicos? ¿Cómo evolucionan? ¿Dónde y cómo se codifican estos fenotipos? ¿Qué tipo de enfoques experimentales deberían desarrollarse para estudiarlos?

*Predicción del fenotipo.* Un posible enfoque para anticipar los fenotipos se centra en el uso de grandes conjuntos de datos que incluyan características ómicas, parámetros ecológicos, morfometría (derivada del uso de técnicas avanzadas del trata de imágenes), etc. Estos conjuntos de datos se analizarán mediante técnicas de «deep learning» y otras herramientas de inteligencia artificial (IA) con el objetivo de predecir el fenotipo. Estas técnicas de IA podrán utilizarse para definir *leyes* de mapeo. También debería contribuir a validar nuestra confianza en estos enfoques en contextos biomédicos relevantes.

*La homeostasis y el declive del fenotipo.* Una extensión del estudio de la aparición del fenotipo con importantes implicaciones para la salud humana se refiere a cómo se mantiene este fenotipo en la salud y cómo degenera en la enfermedad y el envejecimiento. Este programa de investigación examina la escala de tiempo de la vida del individuo (el marco temporal del ciclo vital) e incluye el análisis específico de todos los factores que contribuyen al mantenimiento de una función saludable y sus cambios a medida que envejece el individuo. ¿Qué se sabe sobre las variaciones normales? ¿Dónde está el límite entre los fenotipos fisiológicos y patológicos? Dada la conocida reactivación de los programas de desarrollo en la enfermedad en adultos (Nieto, 2013), ¿de qué manera se puede utilizar este conocimiento adquirido sobre la aparición y el mantenimiento del fenotipo para comprender la patología? El conocimiento de los mecanismos que subyacen al mantenimiento del fenotipo y la homeostasis de los tejidos debería ayudar a comprender los que están detrás de su deterioro o pérdida durante la enfermedad y el envejecimiento. Se dedicarán grandes esfuerzos a entender cómo otros organismos envejecen y reparan y regeneran sus heridas, o de qué manera son menos susceptibles a ciertas enfermedades que los humanos. Estos conocimientos se utilizarán en proyectos destinados a mejorar la salud humana.

*Los requisitos de la autoorganización.* Los sistemas biológicos complejos, como la materia activa en la corteza celular, el embrión en fase temprana o la retina, implican una serie de partes que interactúan entre sí. Este desafío

determinará cuáles son los requisitos para la autoorganización de los fenotipos biológicos a diferentes escalas. Por ejemplo, ¿existen límites inferiores (y superiores) al número de células necesarias para organizarse en una estructura similar a la de un embrión en fase temprana, como un embrioide, en el que se rompe la simetría y surge una organización básica de la capa germinal?

*Caracterización 4D del desarrollo a resolución unicelular.* Se analizará la dinámica de procesos de desarrollo específicos (como *operaciones de desarrollo* modulares) en el espacio y el tiempo a una resolución unicelular, incluidas la transcriptómica, la proteómica y la metabolómica, y midiendo los costes energéticos y los flujos de información. Para lograr este objetivo, será necesario el desarrollo de **técnicas de tratamiento de imágenes** capaces de captar el comportamiento dinámico de muchas células, además de **la detección multiplexada** de moléculas y **modelado computacional**. El uso de estos modelos será doble: por un lado, podrán describir todas las interacciones y determinar los parámetros de forma cuantitativa. Por otro lado, se utilizarán para derivar las propiedades globales que rigen la dinámica del proceso de desarrollo. Los metaanálisis de datos permitirán comparar estados normales y enfermos, o bien diferentes especies que muestren fenotipos distintos.

*Biología sintética del desarrollo y evolución.* El objetivo consistirá en sintetizar órganos con composición celular y características morfofuncionales diseñadas. Este desafío requerirá (1) definir las principales *operaciones* de desarrollo (como la división celular, el plegado de tejidos, la migración celular colectiva); (2) obtener una comprensión a nivel de sistema de estas operaciones de desarrollo, incluidos los parámetros que pueden modular el comportamiento de cada una de ellas (por ejemplo, dos especies pueden plegar un neuroepitelio para desarrollar un órgano, pero el pliegue puede ser más profundo en una de las dos especies); (3) la selección evolutiva *in silico* de fenotipos que combinen las operaciones de desarrollo (evolución *in silico*/computacional); y (4) la implementación biológica de estos fenotipos (ya sea *in vitro* o mediante la modificación genética de especies específicas) mediante la ingeniería biológica, ya sea mediante la modificación de organismos existentes, mediante organoides o mediante el diseño y la síntesis *de novo*.

*Examinar la amplitud de la diversidad fenotípica sin perder de vista el desafío.* Abordar este desafío correctamente dependerá de la definición de objetivos y sistemas experimentales claros. Para ello, será esencial contar con una doble estrategia que sea al mismo tiempo *centrífuga* y *centrípeta*. Se necesitará una

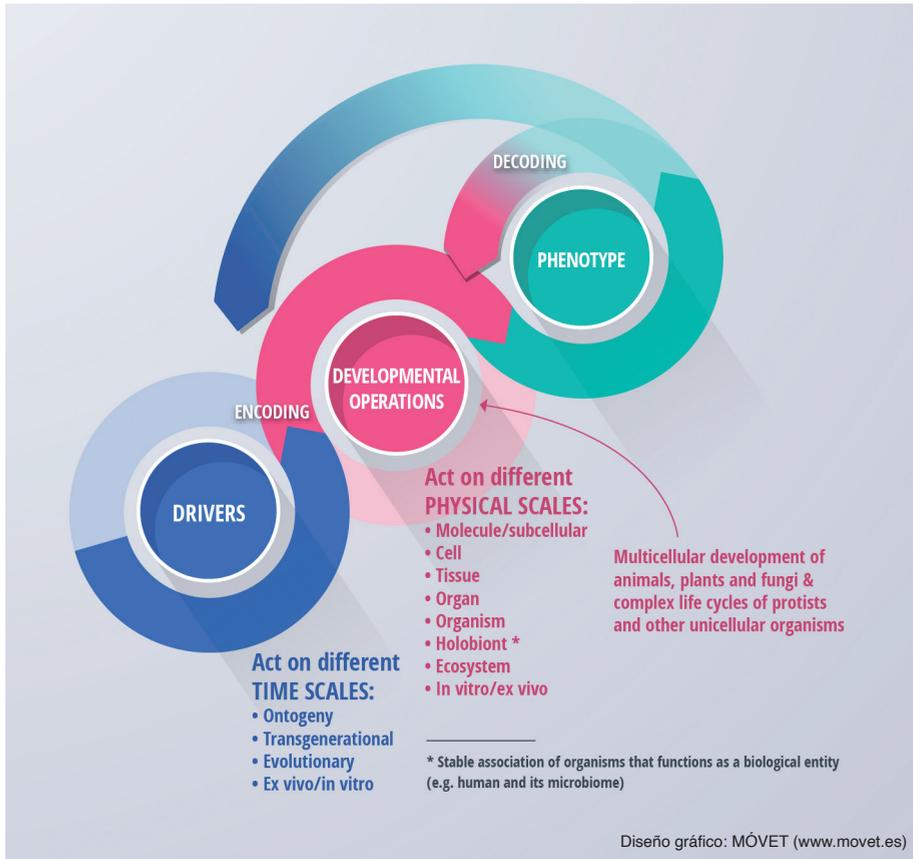
estrategia centrífuga para ampliar y diversificar la lista extremadamente corta de organismos utilizados actualmente como sistemas experimentales. Características como la plasticidad fenotípica, los tiempos de generación, las capacidades de regeneración, la distribución biogeográfica y la complejidad del ciclo de vida, entre otras, son muy diferentes según la especie y, por tanto, no todas serán igualmente abordables para tratar los diferentes temas y cuestiones de *la génesis del desafío del fenotipo*. Y lo que es más importante: hasta ahora, solo se han descrito ciertos fenómenos biológicos de especial relevancia para este desafío (por ejemplo, los priones adaptativos o la eliminación del ADN/cromosoma) en un puñado de especies (Harvey *et al.*, 2018; Smith, 2018). Por lo tanto, para tener una imagen completa, será importante estudiar una amplia gama de organismos que sea más representativa de toda la diversidad biológica, incluidas las especies que actualmente no se pueden cultivar en el laboratorio, pero que deberían ser susceptibles de experimentación en un futuro próximo. Una estrategia centrípeta será igualmente importante, ya que, para lograr una buena comprensión de la génesis del fenotipo, se necesitarán una verdadera integración y coordinación de enfoques multidisciplinarios. Así, para poner en marcha este tipo de acciones combinadas, se deberán diseñar proyectos específicos que centralicen estos esfuerzos multidisciplinarios centrándose en conjuntos particulares de órganos, organismos o procesos modelo.

## DESAFÍO 4 | BIBLIOGRAFÍA

- Anderson, P.W. (1972). More is different. *Science* 177, 393-396.
- Arnold, S.J., Burger, R., Hohenlohe, P.A., Ajie, B.C., Jones, A.G. (2008). Understanding the evolution and stability of the G-matrix. *Evolution* 62, 2451-2461.
- Bahar, R., Hartmann, C.H., Rodríguez, K.A., Denny, A.D. *et al.* (2006). Increased cell-to-cell variation in gene expression in ageing mouse heart. *Nature* 441, 1011-1014.
- Catullo, R.A., Llewelyn, J., Phillips, B.L., Moritz, C.C. (2019). The Potential for Rapid Evolution under Anthropogenic Climate Change. *Curr Biol* 29, R996-R1007.
- Cheverud, J.M. (1988). A Comparison of Genetic and Phenotypic Correlations. *Evolution* 42, 958-968.
- Ciabrelli, F., Comoglio, F., Fellous, S., Bonev, B., Ninova, M. *et al.*, (2017). Stable Polycomb-dependent transgenerational inheritance of chromatin states in *Drosophila*. *Nat. Genet.* 49, 876-886.
- Connallon, T., Hall, M.D. (2018). Genetic constraints on adaptation: a theoretical primer for the genomics era. *Annals of the New York Academy of Sciences* 1422, 65-87.
- Duempelmann, L., Mohn, F., Shimada, Y., Oberti, D., Andriollo, A., Lochs, S., Buhler, M. (2019). Inheritance of a Phenotypically Neutral Epimutation Evokes Gene Silencing in Later Generations. *Mol. Cell.* 74, 534-541 e534.
- Eguchi, Y., Bilollikar, G., Geiler-Samerotte, K. (2019). Why and how to study genetic changes with context-dependent effects. *Curr. Opin. Genet. Dev.* 58-59, 95-102.
- Eyre-Walker, A., Keightley, P.D. (2007). The distribution of fitness effects of new mutations. *Nat. Rev. Genet.* 8, 610-618.
- Goodwin, B., Sole, R. (2002). *Signs of Life: How complexity pervades biology*. Basic Books.
- Harvey, Z.H., Chakravarty, A.K., Futia, R.A., Jarosz, D.F. (2020). A Prion Epigenetic Switch Establishes an Active Chromatin State. *Cell* 180, 928-940 e914.
- Harvey, Z.H., Chen, Y., Jarosz, D.F. (2018). Protein-Based Inheritance: Epigenetics beyond the Chromosome. *Mol. Cell.* 69, 195-202.
- Jablonka, E., Lamb, M.J. (2005). *Evolution in Four Dimensions. Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*. The MIT Press.
- Jung, Y.H., Sauria, M.E.G., Lyu, X., Cheema, M.S., Ausio, J., Taylor, J., Corces, V.G. (2017). Chromatin States in Mouse Sperm Correlate with Embryonic and Adult Regulatory Landscapes. *Cell reports* 18, 1366-1382.
- Klosin, A., Casas, E., Hidalgo-Carcedo, C., Vavouri, T., Lehner, B. (2017). Transgenerational transmission of environmental information in *C. elegans*. *Science* 356, 320-323.
- Kriegman, S., Blackiston, D., Levin, M., Bongard, J. (2020). A scalable pipeline for designing reconfigurable organisms. *Proc. Natl. Acad. Sci. USA* 117, 1853-1859.
- Lehner, B. (2011). Molecular mechanisms of epistasis within and between genes. *Trends Genet.* 27, 323-331.
- Lehner, B. (2013). Genotype to phenotype: lessons from model organisms for human genetics. *Nature Reviews Genetics* 14 168-178.
- Lev, I., Toker, I.A., Mor, Y., Nitzan, A., Weintraub, G., Antonova, O. *et al.* (2019). Germ Granules Govern Small RNA Inheritance. *Curr. Biol.* 29, 2880-2891 e2884.
- Manzano-López, J., Matellán, L., Álvarez-Llamas, A., Blanco-Mira, J.C., Monje-Casas, F. (2019). Asymmetric inheritance of spindle microtubule-organizing centres preserves replicative lifespan. *Nat. Cell. Biol.* 21, 952-965.
- Martin, L.B., Ghalambor, C.K., Woods, H.A. (2015). *Integrative Organismal Biology*. Wiley Blackwell.
- Metcalfe, N.B., Alonso-Álvarez, C. (2010). Oxidative stress as a life-history constraint: the role of reactive oxygen species in shaping phenotypes from conception to death. *Functional Ecology* 24, 984-996.
- Moczek, A.P., Sultan, S., Foster, S., Ledon-Rettig, C. *et al.* (2011). The role of developmental plasticity in evolutionary innovation. *Proc. Biol. Sci.* 278, 2705-2713.
- Murren, C.J., Auld, J.R., Callahan, H., Ghalambor, C.K., Handelsman, C.A. *et al.* (2015). Constraints on the evolution of phenotypic plasticity: limits and costs of phenotype and plasticity. *Heredity (Edinb)* 115, 293-301.

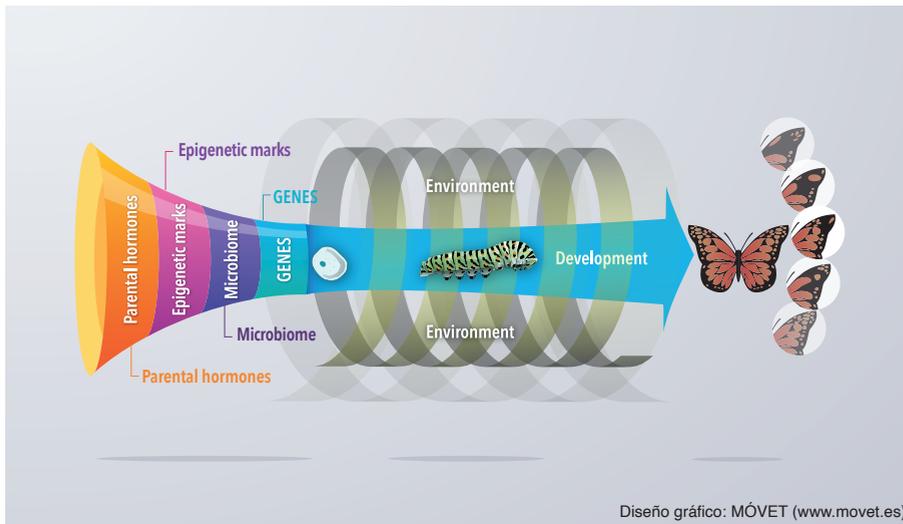
- Nieto, M.A. (2013).** Epithelial plasticity: a common theme in embryonic and cancer cells. *Science* 342, 1234850.
- Phillips, P.C., Arnold, S.J. (1999).** Hierarchical Comparison of Genetic Variance-Covariance Matrices. I. Using the Flury Hierarchy. *Evolution* 53, 1506-1515.
- Posner, R., Toker, I.A., Antonova, O., Star, E. et al. (2019).** Neuronal Small RNAs Control Behavior Transgenerationally. *Cell* 177, 1814-1826 e1815.
- Roughgarden, J., Gilbert, S.F., Rosenberg, E., Zilber-Rosenberg, I., Lloyd, E.A. (2017).** Holobionts as Units of Selection and a Model of Their Population Dynamics and Evolution. *Biological Theory* 13, 44-65.
- Saltz, J.B., Hessel, F.C., Kelly, M.W. (2017).** Trait Correlations in the Genomics Era. *Trends Ecol. Evol.* 32, 279-290.
- Segar, S.T., Fayle, T.M., Srivastava, D.S., Lewinsohn, T.M., Lewis, O.T., Novotny, V., Kitching, R.L., Maunsell, S.C. (2020).** The Role of Evolution in Shaping Ecological Networks. *Trends Ecol. Evol.* 35, 454-466.
- Shirokawa, Y., Shimada, M. (2016).** Cytoplasmic inheritance of parent-offspring cell structure in the clonal diatom *Cyclotella meneghiniana*. *Proc. Biol. Sci.* 283, 20161632
- Skvortsova, K., Iovino, N., Bogdanovic, O. (2018).** Functions and mechanisms of epigenetic inheritance in animals. *Nat. Rev. Mol. Cell. Biol.* 19, 774-790.
- Smith, J.J. (2018).** Programmed DNA Elimination: Keeping Germline Genes in their Place. *Curr. Biol.* 28, R601-R603.
- Thompson, D.W. (1945).** *On growth and form.* University Press, Cambridge.
- Topol, E. (2019).** *Deep Medicine. How artificial intelligence can make healthcare human again.* Basic Books.
- Uller, T., Moczek, A.P., Watson, R.A., Brakefield, P.M., Laland, K.N. (2018).** Developmental Bias and Evolution: A Regulatory Network Perspective. *Genetics* 209, 949-966.
- Von Bertalanffy, L. (1968).** *General System Theory: Foundations, Development, Applications.* George Braziller, New York.
- Waddington, C.H. (1957).** *The strategy of the genes: a discussion of some aspects of theoretical biology.* Allen & Unwin.
- Wagner, G.P., Pavlicev, M., Cheverud, J.M. (2007).** The road to modularity. *Nat. Rev. Genet.* 8, 921-931.
- Watson, R.A., Mills, R., Buckley, C.L. et al. (2016).** Evolutionary Connectionism: Algorithmic Principles Underlying the Evolution of Biological Organisation in Evo-Devo, Evo-Eco and Evolutionary Transitions. *Evol. Biol.* 43, 553-581.
- West-Eberhard, M.J. (2005).** Developmental plasticity and the origin of species differences. *Proc. Natl. Acad. Sci. USA* 102(Suppl 1), 6543-6549.
- Wiley, C.D. Campisi, J. (2016).** From Ancient Pathways to Aging Cells-Connecting Metabolism and Cellular Senescence. *Cell metabolism* 23, 1013-1021.

## RESUMEN PARA EXPERTOS



*decoding*: descodificación; *phenotype*: fenotipo; *developmental operations*: operaciones de desarrollo; *encoding*: codificación; *drivers*: impulsores; *Act on different PHYSICAL SCALES*: Actuación sobre diferentes ESCALAS FÍSICAS; *Molecule/subcellular*: Molécula/subcelular; *Cell*: Célula; *Tissue*: Tejido; *Organ*: Órgano; *Organism*: Organismo; *Holobiont \**: Holobionte\*; *Ecosystem*: Ecosistema; *In vitro/ex vivo*: In vitro/ex vivo; *Multicellular development of animals, plants and fungi & complex life cycles of protists and other unicellular organisms*: Desarrollo multicelular de animales, plantas, hongos y ciclos vitales complejos de protistas y otros organismos unicelulares; *Act on DIFFERENT TIME SCALES*: Actuación sobre diferentes ESCALAS TEMPORALES; *Ontogeny*: Ontogenia; *Transgenerational*: Transgeneracional; *Evolutionary*: Evolutiva; *Ex vivo/in vitro*: Ex vivo/in vitro; \* *Stable association of organisms that functions as a biological entity (e.g. human and its microbiome)*: \* Asociación estable de organismos que funciona como una entidad biológica (por ejemplo, el ser humano y su microbioma)

## RESUMEN PARA EL PÚBLICO GENERAL



Diseño gráfico: MÓVET ([www.movet.es](http://www.movet.es))

*Epigenetic marks*: Marcas epigenéticas; *GENES*: GENES; *Parental hormones*: Hormonas parentales; *Microbiome*: Microbioma; *Environment*: Entorno; *Development*: Desarrollo



## DESAFÍO 5

### RESUMEN

La biología de sistemas evolutiva es un enfoque sistémico de la biología que pretende generar una comprensión mecanicista y evolutiva de los mapas genotipo-fenotipo a varias escalas. Al combinar enfoques matemáticos, moleculares y celulares de la evolución, la biología de sistemas evolutiva mejorará la capacidad de predicción y explicación de la síntesis evolutiva moderna.

### PALABRAS CLAVE

sistemas complejos

evolución

mapa genotipo-fenotipo

biología sintética biología de sistemas biología teórica

# BIOLOGÍA DE SISTEMAS EVOLUTIVA

## Coordinadores

Sergi Valverde  
(IBE)  
Saúl Ares  
(CNB)

## Investigadores y centros de investigación participantes

Eva Balsa-Canto  
(IIM)  
Julio R. Banga  
(IIM)  
Santiago F. Elena  
(I2SysBio)  
Jaime Iranzo  
(CBGP)  
Susanna C. Manrubia  
(CNB)

## RESUMEN EJECUTIVO

«Nothing in Biology makes sense except in the light of Evolution» (Nada tiene sentido en biología si no es a la luz de la evolución)  
Theodosius Dobzhansky (1973).

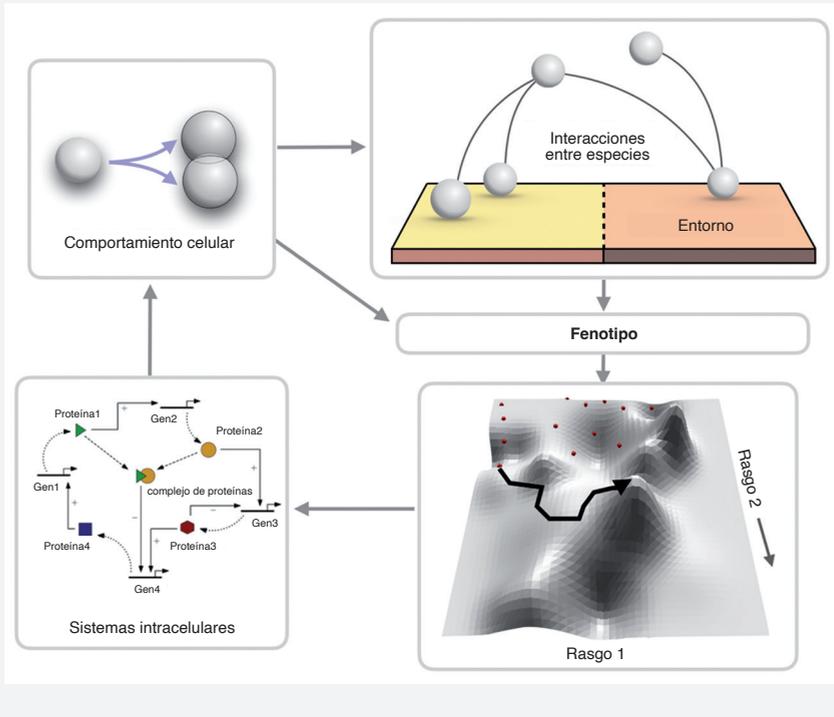
La evolución es un proceso complejo de varios niveles que opera a largas escalas de tiempo y que solo se puede entender desde una perspectiva sistémica. Aunque disponemos de una gran cantidad de datos experimentales, el principal desafío consiste en desarrollar modelos y marcos teóricos para entender los resultados empíricos y plantear preguntas experimentales mejor enfocadas. Una de las primeras cuestiones sin resolver estriba en cómo surgen los fenotipos a partir de los genotipos. Se carece de una idea completa y la modelación se limita a la dinámica de las moléculas —por ejemplo, el ARN o el plegado de proteínas— o a la aparición de interacciones moleculares simples, como motivos reguladores. Un conocimiento más exhaustivo de la estructura de los espacios genotípicos pudiera conducir, entre otras cosas, a una cuantificación de los papeles relativos que desempeñan los mecanismos neutrales y adaptativos. Este programa de investigación tendrá que investigar los efectos de la innovación evolutiva. ¿Pueden construirse funciones biológicas complejas a partir de módulos anteriores más simples? ¿Cómo surgen los circuitos reguladores? ¿Cuáles son los límites del diseño de módulos funcionales robustos y transportables? Las respuestas a estas preguntas evaluarán la validez de los enfoques reduccionistas, frente a la consideración de la

innovación como un fenómeno emergente, que surge de propiedades distribuidas en forma de red. En un marco más amplio, deberíamos preocuparnos por el origen mecanicista de las transiciones evolutivas y por el papel que desempeñan las fuerzas externas en contraposición a los fenómenos contingentes o estocásticos en su generación. La propia evolución puede considerarse una herramienta para la biología sintética, ya que la selección evolutiva dirigida es una forma de alcanzar las funciones deseadas. Uno de los principales objetivos se refiere a integrar el diseño en la exploración de espacios fenotípicos impulsada por la selección. Los avances dependerán de la creación de grandes plataformas evolutivas donde la evolución experimental informada por el diseño y la teoría pueda proceder de forma masiva. La exploración de los posibles fenotipos de los organismos diseñados mediante ingeniería a gran escala, a diferencia de la simple adición o supresión de un gen, exige el conocimiento urgente de la plasticidad y adaptabilidad de los nuevos organismos y sus rasgos de interés. Un eventual uso comercial de estos organismos exige más investigación en las fronteras entre la ciencia evolutiva, el clima y la ecología: ¿cómo afectarán estos organismos al ecosistema?, ¿cómo afectará el ecosistema a los organismos en un entorno que cambia con tanta rapidez?

## 1. INTRODUCCIÓN Y DESCRIPCIÓN GENERAL

Esta cuestión es el núcleo de la biología de sistemas evolutiva, una nueva área de investigación interdisciplinaria que reconstruye paisajes de aptitud para predecir (1) la aptitud de los organismos individuales bajo cambios de estado y ambientales y (2) las trayectorias evolutivas de las poblaciones a través de los paisajes (Medina, 2005; Soyer y O'Malley, 2013; Loewe, 2016). La biología de sistemas evolutiva pertenece a una tendencia más amplia en las ciencias biológicas, es decir, la síntesis evolutiva moderna, que ha estado creando una visión coherente de la evolución desde la década de 1920 (Mayr y Provine, 1998). Se trata de un campo emergente que toma conocimientos actualmente dispersos en muchos campos de investigación, desde la genética de poblaciones y la bioquímica hasta la ecología. La biología de sistemas evolutiva reconoce que la evolución es un proceso complejo de varios niveles que opera a largas escalas temporales y que solo se puede entender desde una perspectiva verdaderamente sistémica. Los esfuerzos anteriores intentaron explicar los procesos evolutivos en el contexto más estrecho de los genes individuales y las estructuras proteicas. Sin embargo, los organismos complejos no pueden reducirse al funcionamiento de sus componentes de forma aislada. La biología de sistemas evolutiva aborda este objetivo modelando los fenotipos como

**FIGURA 1**—El objetivo de la biología de sistemas evolutiva se centra en comprender y predecir los mapas genotipo-fenotipo en los sistemas biológicos. La evolución es un proceso de varios niveles que opera en una amplia gama de escalas temporales. La biología de sistemas abordó esta cuestión centrándose en los subsistemas intracelulares (por ejemplo, las redes de regulación genética) y pasando por alto la evolución del organismo. La biología de sistemas evolutiva amplía la visión de la biología de sistemas al incluir los factores ecológicos y evolutivos de la complejidad de los organismos. Las redes celulares determinan las interacciones de las especies con su entorno y con otras especies. Las interacciones ecológicas entre las especies son responsables de la aptitud de los organismos. Los procesos evolutivos (por ejemplo, la deriva neutra y la adaptación) mueven a las poblaciones de organismos en paisajes de aptitud dinámicos cambiando las características de las redes intracelulares.



el resultado de la evolución de subsistemas intracelulares, por ejemplo, de señalización, regulación y metabolismo, que interactúan entre sí. La biología de sistemas evolutiva pretende sintetizar y, en última instancia, predecir las complejas interacciones a varias escalas entre los procesos evolutivos y las propiedades sistémicas (fig. 1).

La biología de sistemas evolutiva también se ha visto impulsada por la adopción generalizada de métodos cuantitativos y computacionales en las ciencias biológicas. Un claro exponente es la biología de sistemas o el estudio de cómo

se organizan los organismos combinando datos experimentales con técnicas de modelación matemática y análisis asistido por ordenador (Ideker *et al.*, 2001; Kitano, 2002). Tradicionalmente, la biología de sistemas se ha enfocado en los aspectos de modelación de los sistemas biológicos, es decir, *cómo* se implementan los sistemas a nivel molecular (véase Boogerd *et al.*, 2007). Inevitablemente, si se pasa por alto el componente evolutivo (*por qué* los sistemas biológicos tienen esas características), no se obtendrá una explicación total de las funciones biológicas. La combinación entre la biología evolutiva y la biología de sistemas permite conocer mejor las características biológicas complejas. El aspecto evolutivo es el que unifica la gran diversidad de métodos empleados por los investigadores de la biología de sistemas evolutiva. Al integrar las preguntas del *cómo* y el *por qué* bajo el mismo marco, podemos entender mejor cómo funcionan los sistemas biológicos y por qué lo hacen de esa manera concreta.

La biología de sistemas evolutiva abarca una integración de enfoques teóricos, empíricos y computacionales. Un componente clave es la síntesis de los *principios de diseño* universales en biología (Poyatos, 2012; Katsnelson *et al.*, 2018). Lograr este objetivo a largo plazo depende fundamentalmente de la identificación (y caracterización) de los principios universales que se aplican a grandes dominios de la vida. Por ejemplo, hemos desarrollado sofisticadas herramientas empíricas y computacionales que permiten el estudio detallado de las funciones biológicas. Sin embargo, se desconoce hasta qué punto la dinámica y la organización observadas en una especie concreta, por ejemplo, cómo permiten las redes de señalización la quimiotaxis en *Escherichia coli* o la dinámica de la osmorregulación en *Saccharomyces cerevisiae* (Alon *et al.*, 1999; Klipp *et al.*, 2005) pueden explicar también las respuestas fisiológicas en otras especies. Ignorar los estados intermedios de la variación a nivel de población no puede explicar completamente la evolución de la complejidad biológica (Lynch, 2007). Las características complejas de los sistemas biológicos pueden ser adaptativas (Kashtan y Alon, 2005), neutrales (Wagner, 2003; Solé y Valverde, 2006) o una combinación entre procesos funcionales y no adaptativos (Wagner, 2008). En este contexto, los métodos evolutivos pueden favorecer la generalización de las propiedades del sistema, como, por ejemplo, los patrones de red, más allá de cualquier modelo biológico específico.

## 2. IMPACTO EN EL PANORAMA DE LA CIENCIA BÁSICA Y EN LAS APLICACIONES POTENCIALES

La propia evolución puede considerarse una herramienta para la biología sintética, ya que la selección evolutiva dirigida es una forma de generar y optimizar las funciones deseadas (Arnold, 2018). Uno de los principales objetivos se refiere a integrar el diseño en la exploración de espacios fenotípicos impulsada por la selección. Los avances dependerán de la creación de grandes plataformas evolutivas donde la evolución experimental informada por el diseño y la teoría pueda proceder de forma masiva. La exploración de los posibles fenotipos de los organismos diseñados mediante ingeniería a gran escala, a diferencia de la simple adición o supresión de un gen, exige el conocimiento exhaustivo y fiable de la plasticidad y adaptabilidad de los nuevos organismos y sus rasgos de interés. Un eventual uso comercial de estos organismos exige más investigación en las fronteras entre la ciencia evolutiva, la climatología, la agricultura y la ecología: cómo afectarán estos organismos al ecosistema, así como el análisis y la predicción de la aparición de nuevos patógenos y su potencial propagación epidémica.

### 2.1 Modelación y simulación asistida por ordenador de sistemas biológicos

La interacción de la modelación matemática con los experimentos es uno de los elementos centrales de la biología de sistemas. La creación de modelos es, por tanto, un paso clave, lo que implica la ingeniería inversa (es decir, la inferencia y la identificación) de las ecuaciones que describen el biosistema, así como su calibración con los datos existentes (Klipp *et al.*, 2005). Normalmente, estos modelos pueden ser mecanicistas (Stelling, 2004) o estar basados en datos (como en el aprendizaje automático y la estadística; Kell y Oliver, 2004). Aunque estos últimos pueden resultar útiles en muchas aplicaciones, los modelos mecanicistas suelen proporcionar un mejor marco para destilar conocimiento y comprensión de los datos. A pesar de los numerosos avances obtenidos durante las dos últimas décadas en la creación de modelos en la biología de sistemas, la mayoría de estos modelos carece de la perspectiva evolutiva. Por ello, un marco teórico y computacional general para la modelación a varias escalas en la biología de sistemas evolutiva sigue siendo uno de los principales objetivos.

El verdadero poder de los modelos matemáticos se libera cuando se explotan mediante métodos para su simulación, análisis, optimización y control

asistidos por ordenador (Wolkenhauser y Mesarović, 2005; Sontag, 2005). Aunque estas herramientas se han utilizado ampliamente en áreas como la ingeniería de biosistemas y de bioprocesos (Park *et al.*, 2008), todavía nos falta el componente evolutivo integrado a través de diferentes escalas. Por lo tanto, otro de los principales objetivos sería explotar los modelos de biología de sistemas evolutiva como núcleo de la ingeniería de sistemas de procesos en las bioindustrias (por ejemplo, la biotecnología agroalimentaria e industrial), especialmente en áreas como la ingeniería metabólica. Una posible vía para lograr este objetivo pudiera ser la integración de los enfoques evolutivos de teoría de juegos con la modelación a varias escalas. El mismo enfoque puede adaptarse y ampliarse a problemas clave en ámbitos como la ingeniería medioambiental (por ejemplo, la dinámica de las comunidades microbianas en la biorremediación) y la medicina (por ejemplo, la evolución del cáncer o múltiples infecciones microbianas).

## **2.2 Un primer paso hacia la comprensión de la complejidad ecológica: modelación de comunidades microbianas complejas**

En las últimas décadas se han desarrollado enfoques de modelación para describir las comunidades celulares: desde modelos continuos hasta modelos individuales de tejidos y biofilms; desde modelos ecológicos fenomenológicos hasta aproximaciones más mecanicistas a escala del genoma (Zomorodi y Segre, 2016). Sin embargo, su alcance se limita principalmente a sistemas *simples* (cultivos individuales o conjuntos) en condiciones ambientales extracelulares controladas o en estado estacionario. Además, apenas se tienen en cuenta los aspectos evolutivos. La integración de modelos dinámicos multi-específicos que incorporan mecanismos ecológicos y evolutivos (Valverde *et al.*, 2020) es necesaria para aprovechar plenamente el potencial de los sistemas multicelulares para responder a preguntas y aplicaciones biológicas.

Los cultivos mixtos microbianos diseñados de forma óptima a partir de modelos permitirán realizar tareas metabólicamente complejas mediante la división del trabajo o la cooperación experimentalmente evolucionada (Harcombe, 2010; Hays *et al.*, 2015; Thommes *et al.*, 2019). Como resultado, los bioprocesos serán más eficientes, productivos y estables que los actuales bioprocesos de una especie y permitirán una gama más amplia de productos.

Se podrían utilizar enfoques similares para diseñar el microbioma con el fin de generar posibles terapias contra enfermedades metabólicas, inflamatorias

e inmunológicas, entre otras. La dimensión evolutiva resulta fundamental para comprender la dinámica de las comunidades en las que coexisten microbios y virus. En estos sistemas, la rápida coevolución de los virus y los huéspedes microbianos (por ejemplo, a través de la adquisición de espaciadores en las matrices CRISPR) no puede desvincularse de los procesos ecológicos, lo que afecta directamente al desarrollo de terapias con fagos y otras estrategias de ingeniería del microbioma basadas en virus.

### **2.3 Biología sintética: diseño humano de construcciones no biológicas**

En la biología sintética, nuestro objetivo es utilizar los principios de ingeniería del diseño racional para modificar organismos o para construir nuevos bioartefactos (Purnick y Weiss, 2009; Cameron *et al.*, 2014; Schwille *et al.*, 2018). Al igual que ocurre con la biología de sistemas, la biología sintética también puede explotar los enfoques basados en modelos como guía para el diseño, el análisis, la optimización y el control de los sistemas genéticos (Marchisio *et al.*, 2009). Además de los recientes avances en la estandarización y caracterización de piezas genéticas (McLaughlin *et al.*, 2018), en los últimos años también hemos sido testigos de importantes avances en la aplicación de la microfluídica, el aprendizaje automático y la automatización aplicada a la biología sintética (Melin y Quake, 2007; Nielsen *et al.*, 2016; Aoki *et al.*, 2019; Carbonell *et al.*, 2019). También estamos ya cerca de disponer de lenguajes de programación para diseñar circuitos computacionales en células vivas (Nielsen *et al.*, 2016). Estos enfoques abren nuevas vías para aplicaciones importantes, como los biosensores (Gupta *et al.*, 2019), los medicamentos bioterpéuticos (Ozdemir *et al.*, 2018), la ingeniería metabólica (Keasling, 2012), la biofabricación (Chen *et al.*, 2020) y la biorremediación (de Lorenzo *et al.*, 2018).

Sin embargo, con la excepción de los esfuerzos en el campo de la evolución dirigida (Arnold, 2018), todavía carecemos de un enfoque verdaderamente evolutivo para la biología sintética. En particular, aunque los métodos asistidos por ordenador pueden ayudarnos con el diseño de piezas sintéticas de forma similar a como se hace, por ejemplo, en la electrónica, se deben tener en cuenta la imprevisibilidad y la complejidad creadas por la variabilidad y la evolucionabilidad del comportamiento celular (Kwok, 2010) para lograr la bioingeniería prevista. Por último, los enfoques sin células (Hodgman y Jewett, 2012) nos ayudarán a comprender mejor el diseño de los biosistemas naturales evolucionados, así como a permitir un nuevo tipo de biofabricación con más libertad de diseño y mejor control.

### 3. PUNTOS CLAVE DEL DESAFÍO

#### 3.1 Marco teórico de varias escalas

Aunque disponemos de una gran cantidad de datos experimentales, uno de los principales desafíos consiste en desarrollar modelos y marcos teóricos que expliquen estos resultados y plantear preguntas experimentales mejor enfocadas. Actualmente carecemos de un marco de modelación a varias escalas que capte las características esenciales de los sistemas biológicos, desde el genoma hasta el metabolismo. Al mismo tiempo, necesitamos modelos que sean lo suficientemente sencillos como para proporcionar respuestas y conocimientos útiles. Una parte esencial de los futuros desarrollos en la biología de sistemas evolutiva es la creación de una jerarquía de herramientas *in silico* y modelos computacionales que puedan guiar los estudios experimentales.

#### 3.2 Evolución de las novedades

Los ecosistemas son sistemas dinámicos complejos y extremadamente no lineales (May y Leonard, 1975; Clark y Luis, 2020). La competencia, la cooperación o la dinámica víctima-explotador son interacciones dependientes de la densidad que inducen efectos no lineales en la dinámica de la población. Esto resulta especialmente relevante cuando se trata de las respuestas de los ecosistemas a las perturbaciones externas, en particular de origen antropogénico (Lade *et al.*, 2020). Se ha conjeturado que estas no linealidades dan lugar a cambios bruscos en la composición del ecosistema, también conocidos como *puntos de inflexión*, que se están convirtiendo en un importante tema de investigación en ecología (Berdugo *et al.*, 2020). Además, en la biología de sistemas, se ha informado de un punto de inflexión que provoca la extinción de la población de levaduras (Dai *et al.*, 2012).

Los cambios repentinos podrían estar asociados a grandes modificaciones. Lo curioso es que los estudios sugieren que las alteraciones suaves del entorno también podrían ser responsables de estos cambios drásticos. Los análisis teóricos de los sistemas dinámicos han interpretado estas transiciones como saltos a lo largo de la optimización evolutiva, donde los largos periodos sin cambios evidencian la presencia de altas barreras de aptitud que la población no puede superar fácilmente (Huynen *et al.*, 1996). Este programa de investigación tendrá que investigar las causas de estos cambios repentinos en la composición genética de las poblaciones. Explicar el origen mecanicista de las transiciones evolutivas y el papel que desempeñan los forzamientos externos

frente a los fenómenos contingentes o estocásticos en su generación ayudará a comprender los efectos de las novedades evolutivas (y, con suerte, a predecirlas) (Wagner y Lynch, 2010).

### 3.3 Modelación y simulación de sistemas biológicos

Los sistemas biológicos se componen de genes que codifican la maquinaria molecular que proporciona las funciones básicas de la vida. Las funciones biológicas rara vez pueden reducirse a componentes específicos de forma aislada. En cambio, son el resultado de varias interacciones entre diferentes componentes. Por ejemplo, las redes de interacciones reguladoras especifican cómo se expresan los genes y ambas operan en múltiples niveles jerárquicos de organización. Cuantificar las características estructurales de los sistemas biológicos y artificiales ha sido el objetivo de la ciencia de redes (una rama de la física estadística y los sistemas complejos desarrollada durante los últimos 25 años). La biología de sistemas ha confiado en los resultados de la ciencia de las redes, pero todavía no comprendemos el origen de las regularidades estructurales en las redes biológicas, ni cómo dan forma a la función y la evolución. Esto resulta especialmente relevante en uno de los principales problemas abiertos en la biología de sistemas, a saber, cómo definir un enfoque robusto para la ingeniería inversa y la identificación de sistemas en sistemas biológicos (Villaverde y Banga, 2014). Lo que se necesita es la cooperación de los científicos de la red con los investigadores de biología de sistemas evolutiva para desarrollar una teoría evolutiva de la complejidad biológica rigurosa y biológicamente realista.

Los modelos a escala del genoma y los principios de optimalidad han demostrado su potencial de aplicación en la generación de mapeos genoma-fenotipo en la biología de sistemas evolutiva, por ejemplo, en la predicción de resultados fenotípicos de la evolución adaptativa a corto plazo o en el análisis de la viabilidad de las cepas mutantes (Palsson, 2015). Sin embargo, para desarrollar todo su potencial para la biología de sistemas evolutiva, se debe mejorar la capacidad de predicción de estos modelos de diferentes maneras. Está claro que la suposición de un crecimiento óptimo en los modelos metabólicos a escala del genoma no resulta adecuada para mutantes ni es válida en muchas condiciones ambientales (variables en el tiempo). Por lo tanto, hay que explorar objetivos evolutivos alternativos o compensaciones o enfoques de teoría de juegos. Todavía se están realizando ampliaciones que incluyen estructuras de proteínas, modelos integrados de metabolismo y expresión de proteínas (O'Brien *et al.*, 2015), así como marcos de modelación híbridos que

incorporen información explícita sobre las tasas de reacción. Los modelos a escala genómica también se han ampliado de poblaciones individuales de células a comunidades microbianas simples (Harcombe *et al.*, 2014); se requieren más desarrollos para describir poblaciones complejas multiespecíficas y entornos cambiantes (en el tiempo y el espacio).

### **3.4. Estructura de los paisajes de aptitud**

Para predecir la trayectoria evolutiva de una especie a otra, primero tenemos que entender el espacio de configuración subyacente, es decir, necesitamos un *mapa* en el que podamos ubicar a todas las especies intermedias posibles, para cada condición ambiental (véase el apéndice C). Por desgracia, los paisajes de aptitud son enormes, viven en espacios de muy alta dimensionalidad y son difíciles de visualizar. La metáfora tradicional del paisaje de aptitud se basa en una perspectiva gradualista sobre el cambio y la evolución del organismo. La evolución se ha descrito como la difusión de poblaciones en un paisaje relativamente suave, siempre ascendiendo hacia regiones de alta aptitud, que al final quedan atrapadas en picos de montaña posiblemente separados por valles profundos de menor aptitud (Wright, 1931). Gracias a los avances conceptuales en nuestra comprensión de la estructura molecular de las poblaciones, ahora sabemos que esta imagen se encuentra claramente incompleta y que pasa por alto ingredientes importantes. La estructura de los paisajes de aptitud no es lisa ni continua sino accidentada (Kauffman y Levin, 1987). En cambio, la estructura del espacio de los genotipos es una red de redes (o red multicapa) cuyos nodos (genomas) son mutuamente accesibles mediante mutaciones. En el paisaje, podemos encontrar regiones con discontinuidades pronunciadas (que señalan la presencia de mutaciones letales y deletéreas) y conexiones de largo alcance entre regiones distantes del paisaje. La biología de sistemas evolutiva nos ayudará a comprender (y caracterizar) las propiedades del paisaje para obtener modelos realistas de paisajes de aptitud adaptativos (en particular, mediante el desarrollo de modelos de mapas genotipo-fenotipo en diferentes niveles de la jerarquía biológica, véase más adelante) y a validar estos modelos con la reconstrucción de paisajes empíricos mediante herramientas de red.

### **3.5. Nuevas construcciones de mapas genotipo-fenotipo (GP)**

La función de mapeo entre las instrucciones codificadas en los genotipos y las estructuras y funciones de los fenotipos resulta fundamental para todos los aspectos de la biología. Dada su importancia, los mapas GP han atraído mucha atención tanto de teóricos como de experimentadores (de Visser y Krug,

2014; Ahnert, 2017). Como resultado de estos ejercicios, se han derivado una serie de propiedades universales compartidas por la mayoría de los mapas GP. Estas propiedades son estructurales en el sentido de que dependen de la distribución de los fenotipos en la red de genotipos. Estas propiedades incluyen la redundancia de genotipos (muchos codifican para el mismo fenotipo), una distribución muy poco uniforme del número de genotipos por fenotipo que da lugar a una gran robustez fenotípica y la capacidad de explorar el paisaje de forma eficaz, con lo que se llega a fenotipos muy distantes a través de un número bastante limitado de cambios genotípicos. A pesar de estos avances, seguimos echando en falta una descripción teórica coherente de los mapas GP que explique por qué surgen todas estas propiedades (Manrubia *et al.*, 2020). Enfoques novedosos como la modelación de los mapas GP como una red de redes en la que los diferentes nodos en el nivel de red genotípica corresponden a redes de genotipos neutrales y estos nodos se mapean en redes similares en el espacio fenotípico (Aguirre *et al.*, 2018) o la metáfora del paisaje marino, en la que la topología del paisaje fluctúa como resultado de los cambios en la composición biótica y abiótica (Mustonen y Lässig, 2009) parecen prometedores. Todavía hay preguntas sin resolver. Por nombrar dos de las más relevantes: a) si el mapa GP como objeto evoluciona (Manrubia *et al.*, 2020) y, si lo hace, por selección natural o por procesos neutros; b) cómo se ajusta el mapa GP a los cambios en el tamaño del genoma; en otras palabras, cuáles son las consecuencias de que la red genotípica (de redes) crezca o se reduzca con el tiempo evolutivo (véase también el Desafío 4).

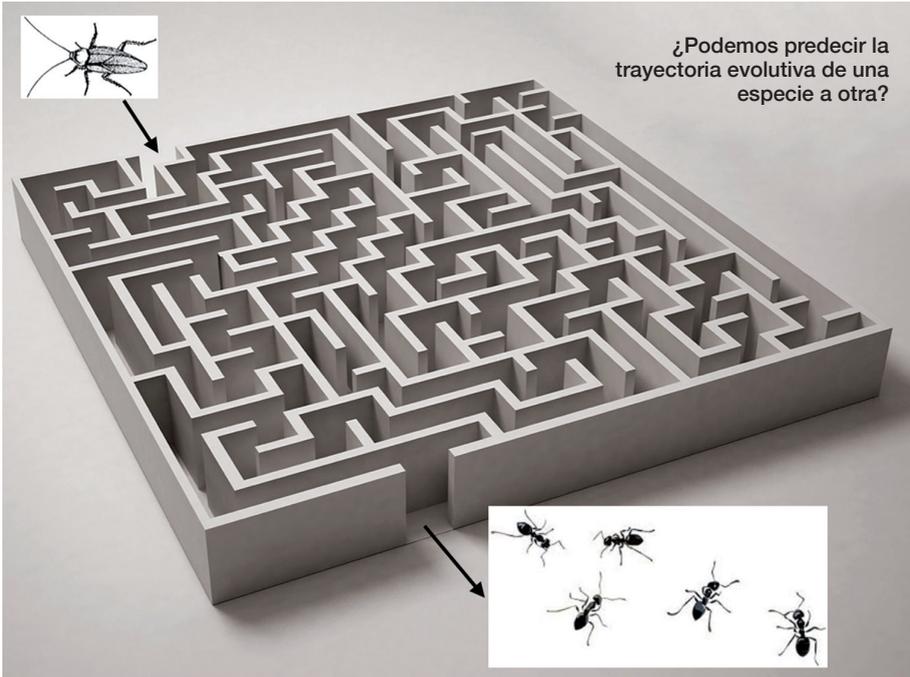
## DESAFÍO 5 BIBLIOGRAFÍA

- Aguirre, J. et al. (2018).** On the networked architecture of genotype spaces and its critical effects on molecular evolution. *Open Biology* 8, 180069.
- Alon, U. et al. (1999).** Robustness in bacterial chemotaxis. *Nature* 397 168-171.
- Aoki, S.K. et al. (2019).** A universal biomolecular integral feedback controller for robust perfect adaptation. *Nature* 570 533-537.
- Ahnert, S.E. (2017).** Structural properties of genotype-phenotype maps. *Journal of the Royal Society Interface* 14, 20170275.
- Arnold, F.H. (2018).** Directed evolution: bringing new chemistry to life. *Angewandte Chemie International Edition* 57, 4143-4148.
- Berdugo, M. et al. (2020).** Global ecosystem thresholds driven by aridity. *Science* 367, 787-790.
- Boogerdt, F.C. et al. (eds.) (2007).** *Systems Biology: Philosophical Foundations*. Amsterdam: Elsevier
- Cameron, D.E., Bashor, C.J., Collins, J.J. (2014).** A brief history of synthetic biology. *Nature Reviews Microbiology* 12, 381-390.
- Carbonell, P., Radivojevic, T., García Martín, H. (2019).** Opportunities at the intersection of synthetic biology, machine learning, and automation. *ACS Synthetic Biology* 8, 1474-1477.
- Chen, Y. et al. (2020).** Systems and synthetic biology tools for advanced bioproduction hosts. *Current Opinion in Biotechnology* 64, 101-109.
- Clark, T.J., Luis, A.D. (2020).** Nonlinear population dynamics are ubiquitous in animals. *Nature Ecology and Evolution* 4, 75-81.
- Dai, L. et al. (2012).** Generic indicators for loss of resilience before a tipping point leading to population collapse. *Science* 336, 1175-1177.
- De Lorenzo, V. et al. (2018).** The power of synthetic biology for bioproduction, remediation and pollution control. *EMBO Reports* 19, e45658.
- De Visser, J.A.G.M., Krug, J. (2014).** Empirical fitness landscapes and the predictability of evolution. *Nature Review Genetics* 15, 480-490.
- Gorter, F.A., Manhart, M., Ackermann, M. (2020).** Understanding the evolution of interspecies interactions in microbial communities. *Philosophical Transactions of the Royal Society B: Biological Sciences* 375, 20190256.
- Gupta, N., et al. (2019).** Cell-based biosensors: Recent trends, challenges and future perspectives. *Biosensors and Bioelectronics* 141, 111435.
- Harcombe, W.R. (2010).** Novel cooperation experimentally evolved between species. *Evolution* 64, 2166-2172.
- Harcombe, W.R. et al. (2014).** Metabolic resource allocation in individual microbes determines ecosystem interactions and spatial dynamics. *Cell Reports* 7, 1104-1115.
- Hays, S.G. et al. (2015).** Better together: engineering and application of microbial symbioses. *Current Opinion in Biotechnology* 36, 40-49.
- Hodgman, C.E., Jewett, M.C. (2012).** Cell-free synthetic biology: thinking outside the cell. *Metabolic Engineering* 14, 261-269.
- Huynen, M.A., Stadler, P.F., Fontana, W. (1996).** Smoothness within ruggedness: the role of neutrality in adaptation. *Proceedings of the National Academy of Sciences of the USA* 93, 397-401.
- Ideker, T., Galitski, T., Hood, L. (2001).** A new approach to decoding life: systems biology. *Annual Reviews in Genomics and Human Genetics* 2, 343-372.
- Kashtan, N., Alon, U. (2005).** Spontaneous evolution of modularity and network motifs. *Proceedings of the National Academy of Sciences of the USA* 102, 13773-13778.
- Katsnelson, M.I., Wolf, Y.I., Koonin, E.V. (2018).** Towards physical principles of biological evolution. *Physica Scripta* 93, 043001.
- Kauffman, S., Levin, S. (1987).** Towards a general theory of adaptive walks on rugged landscapes. *Journal of Theoretical Biology* 128, 11-45.
- Keasling, J.D. (2012).** Synthetic biology and the development of tools for metabolic engineering. *Metabolic Engineering* 14, 189-195.

- Kell, D.B., Oliver, S.G. (2004).** Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. *BioEssays* 26, 99-105.
- Kitano, H. (2002).** Systems Biology: A Brief Overview. *Science* 295, 1662-1664.
- Klipp, E. et al. (2005).** *Systems biology in practice: concepts, implementation and application*. John Wiley & Sons.
- Klipp, E. et al. (2005).** Integrative model of the response of yeast to osmotic shock. *Nature Biotechnology* 23, 975-982.
- Kwok, R. (2010).** Five hard truths for synthetic biology: can engineering approaches tame the complexity of living systems? *Nature* 463, 288-291.
- Lade, S.J. et al. (2020).** Human impacts on planetary boundaries amplified by Earth system interactions. *Nature Sustainability* 3, 119-128.
- Loewe, L. (2016).** Systems in evolutionary systems biology. *Encyclopedia of Evolutionary Biology* 4, 297-318.
- Lynch, M. (2007).** The evolution of genetic networks by non-adaptive processes. *Nature Reviews Genetics* 8, 803-813.
- Manrubia, S. et al. (2020).** From genotypes to organisms: State-of-the-art and perspectives of a cornerstone in evolutionary dynamics. *arXiv*, 2002.00363v1.
- Marchisio, M.A., Stelling, J. (2009).** Computational design tools for synthetic biology. *Current Opinion in Biotechnology* 20, 479-485.
- May, R.M., Leonard, W.J. (1975).** Nonlinear aspects of competition between three species. *SIAM Journal of Applied Mathematics* 29, 243-253.
- Mayr, E., Provine, W.B. (eds) (1998).** *The evolutionary synthesis: perspectives on the unification of Biology*. Harvard University Press
- McLaughlin, J.A. et al. (2018).** SynBioHub: a standards-enabled design repository for synthetic biology. *ACS Synthetic Biology* 7, 682-688.
- Medina, M. (2005).** Genomes, phylogeny, and evolutionary systems biology. *Proceedings of the National Academy of Sciences of the USA* 102, 6630-6635.
- Melin, J., Quake, S.R. (2007).** Microfluidic large-scale integration: the evolution of design rules for biological automation. *Annual Reviews Biophysics and Biomolecular Structure* 36, 213-231.
- Mustonen, V., Lässig, M. (2009).** From fitness landscapes to seascape: non-equilibrium dynamics of selection and adaptation. *Trends in Genetics* 25, 111-119.
- Nielsen, A.A. et al. (2016).** Genetic circuit design automation. *Science* 352, aac7341.
- O'Brien, E.J., Monk, J.M., Palsson, B.O. (2015).** Using genome-scale models to predict biological capabilities. *Cell* 161, 971-987.
- Ozdemir, T. et al. (2018).** Synthetic biology and engineered live biotherapeutics: toward increasing system complexity. *Cell Systems* 7, 5-16.
- Park, J.H., et al. (2008).** Application of systems biology for bioprocess development. *Trends in Biotechnology* 26, 404-412.
- Poyatos J.F. (2012).** On the search for design principles in biological systems. *Advances in experimental medicine and biology* 751, 183-193. [https://doi.org/10.1007/978-1-4614-3567-9\\_9](https://doi.org/10.1007/978-1-4614-3567-9_9).
- Purnick, P.E., Weiss, R. (2009).** The second wave of synthetic biology: from modules to systems. *Nature Reviews Molecular Cell Biology* 10, 410-422.
- Schwille, P., et al. (2018).** MaxSynBio: avenues towards creating cells from the bottom up. *Angewandte Chemie International Edition* 57, 13382-13392.
- Solé, R.V., Valverde, S. (2006).** Are network motifs the spandrels of cellular complexity? *Trends in Ecology and Evolution* 21, 419-422.
- Sontag, E.D. (2005).** Molecular systems biology and control. *European Journal of Control* 11, 396-435.
- Soyer, S.O., O'Malley, M.A. (2013).** Evolutionary systems biology: what it is and why it matters. *BioEssays* 35, 696-705.
- Stelling, J. (2004).** Mathematical models in microbial systems biology. *Current Opinion in Microbiology* 7, 513-518.
- Thommes, M., et al. (2019).** Designing metabolic division of labor in microbial communities. *mSystems* 4, e00263-18.

- Valverde, S., Vidiella, B., Montañez, R., Sacristan, S., Fraile, A., y García-Arenal, F. (2020).** Coexistence of nestedness and modularity in host-pathogen infection networks. *Nature Ecology and Evolution* 4, 568-577.
- Villaverde, A.F., Banga, J.R. (2014).** Reverse engineering and identification in systems biology: Strategies, perspectives and challenges. *Journal of the Royal Society Interface* 11, 20130505.
- Wagner, A. (2003).** Does selection mold molecular networks? *Science Signaling* 202, pe41.
- Wagner A. (2008).** Neutralism and selectionism: a network-based reconciliation. *Nature Review Genetics* 9, 965-974.
- Wagner, G. P. y Lynch, J. L. (2010).** *Evolutionary novelties* 20, R48-R52.
- Wolkenhauer, O., Mesarović, M. (2005).** Feedback dynamics and cell function: why systems biology is called systems biology. *Molecular BioSystems* 1, 14-16.
- Wright, S. (1931).** Evolution in mendelian populations. *Genetics* 16, 97-159.
- Zomorodi, A.R., Segrè, D. (2016).** Synthetic ecology of microbes: mathematical models and applications. *Journal of Molecular Biology* 428, 837-861.

## RESUMEN PARA EXPERTOS Y EL PÚBLICO GENERAL



## DESAFÍO 6

### RESUMEN

Uno de los mayores desafíos en el estudio de la evolución humana tiene que ver con comprender los procesos adaptativos sociales y biológicos que tuvieron lugar a lo largo de nuestra historia evolutiva. Este desafío presenta dimensiones moleculares, genéticas, conductuales, sociales y morfológicas anatómicas, y requiere nuevos enfoques multidisciplinares y tecnológicos.

### PALABRAS CLAVE

paleogenómica paleoproteómica  
paleoepigenómica genoma/fenoma  
datos masivos organoides cerebro  
anatomía digital imágenes en 3D  
bases de datos virtuales paleoantropología  
factores bióticos/abióticos tafonomía  
arqueología de la evolución humana

# EVOLUCIÓN SOCIAL Y HUMANA

## Coordinadores

Carles Lalueza-Fox  
(IBE)

Ignacio de la Torre  
(IH-CCHS)

## Investigadores y centros de investigación participantes

Markus Bastir  
(MNCN)

Víctor Borrell  
(INA)

Pedro Díaz del Río  
(IH-CCHS)

Arcadi Navarro  
(IBE)

Tomàs Marquès-Bonet  
(IBE)

Juan Manuel Vicent  
(IH-CCHS)

## RESUMEN EJECUTIVO

Comprender la evolución social y biológica del linaje humano es una tarea compleja que requerirá la integración de diferentes campos científicos y de sus enfoques más innovadores y vanguardistas. En un marco temporal más profundo, la disponibilidad de genomas completos de múltiples especies que ofrece información sobre la filogenia de nuestro linaje permite estudiar con detalle cuándo y cómo aparecieron algunos de los rasgos definitorios del ser humano durante la evolución. La aplicación de las técnicas ómicas, tanto con especies existentes como extintas (paleogenómica, paleoepigenómica, la emergente paleoproteómica, junto con los datos de la genómica comparativa y la genómica evolutiva), al estudio del pasado permite ahora a los investigadores abordar una serie de cuestiones que antes eran estudiadas casi exclusivamente por disciplinas de las humanidades, como la historia y la arqueología, al investigar las antiguas migraciones. Los esfuerzos multidisciplinarios que implican estos desafíos requerirán también la gestión segura de las grandes cantidades de información que se están generando (lo que se conoce como *macrodatos* o *datos masivos*), desde los biobancos actualmente en expansión hasta los resultados de proyectos de investigación concretos que deben ponerse a disposición de todos los investigadores según los principios FAIR (por su siglas en inglés, de las iniciales traducidas como *Encontrables, Accesibles, Interoperables y Reutilizables*). Al mismo tiempo, surgirán nuevas cuestiones éticas que habrá que abordar.

En cuanto al estudio de la evolución cultural y del comportamiento humano, abordado principalmente mediante los estudios arqueológicos, los desafíos disciplinarios deberían intentar redefinir el papel de la tecnología en la configuración de las sociedades del pasado, pero también en la transformación del entorno, así como la naturaleza de las interacciones entre los agentes bióticos y abióticos a lo largo de la evolución de nuestro género. Dichos estudios deben adoptar enfoques de datos masivos y contar con un alcance internacional, a fin de superar las perspectivas locales. Su objetivo también consiste en identificar patrones, más que eventos, en el curso de la evolución conductual y cultural de nuestra especie. Así, cualquier caracterización de la cultura humana incluirá necesariamente la tecnología como elemento definitorio y concepto clave para explicar la evolución social en nuestro linaje ancestral y trayectoria histórica. A pesar de su papel central en los mecanismos explicativos de la evolución biológica humana y del cambio social, la mera definición de la tecnología, su potencial singularidad en nuestro género y su función como motor de las transformaciones culturales son conceptos actualmente cuestionados y constituyen desafíos fundamentales en las disciplinas relacionadas con la reconstrucción de la evolución social del pasado.

Una de las cuestiones principales en el estudio del comportamiento humano del pasado es hasta qué punto los factores evolutivos bióticos frente a los abióticos dictan las trayectorias divergentes en el pasado humano. Desde el papel del clima en las primeras adaptaciones humanas hasta su relevancia en la aparición de la producción de alimentos, el interés secular por la influencia de las causas abióticas en la evolución social está dando lugar a un cambio de perspectiva. Los nuevos grandes desafíos de la evolución social deberían poner de relieve, por ejemplo, la importancia de la estructura de las comunidades de mamíferos en la conformación de los primeros comportamientos humanos, o el impacto de las acciones humanas sobre la fauna y la flora y, en los últimos tiempos, incluso sobre el clima. Los conjuntos de datos primarios y las fuentes de datos masivos deben integrarse en grandes conjuntos de información como infraestructuras de datos espaciales susceptibles de análisis espacial y accesibles a los colaboradores interconectados de varios campos disciplinares e instituciones, a los que se pueden aplicar perspectivas a escala continental.

## 1. INTRODUCCIÓN Y DESCRIPCIÓN GENERAL

El estudio de la evolución humana a partir de la divergencia con el linaje de nuestro pariente vivo más cercano, el chimpancé, se ha basado principalmente, hasta hace poco, en el análisis de las evidencias arqueológicas y paleontológicas. Sin embargo, tras la secuenciación de los genomas del ser humano y del chimpancé (2001-2006), se ha avanzado poco en la comprensión de las bases funcionales de las diferencias genéticas halladas entre ambas especies. Esto puede atribuirse, en parte, a los seis millones de años transcurridos en cada linaje evolutivo, así como a la complejidad de los procesos implicados y a las limitaciones inherentes al registro fósil. Parece claro que hace falta una nueva perspectiva para este problema científico, junto con nuevos enfoques metodológicos y técnicos.

La rápida evolución de las técnicas digitales ha supuesto un impulso para disciplinas como la paleontología y la arqueología, y las nuevas perspectivas de análisis espacial y las iniciativas de datos masivos permiten realizar estudios a gran escala de los patrones evolutivos humanos bajo nuevas premisas teóricas. Estos avances en disciplinas ya consolidadas e implicadas en los estudios de la evolución humana, junto a las revolucionarias perspectivas introducidas por las disciplinas derivadas de la genética, presentan una oportunidad única para promover una verdadera visión multidisciplinaria del estudio del pasado, para lo cual el CSIC debería asumir el liderazgo en el ámbito internacional.

## 2. IMPACTO EN EL PANORAMA DE LA CIENCIA BÁSICA Y EN LAS APLICACIONES POTENCIALES

Los resultados obtenidos mediante la caracterización genética de los homínidos arcaicos (incluidos los neandertales y los denisovanos) para reconstruir los acontecimientos migratorios y de mezcla del pasado también incluyen pruebas sobre la demografía o la estructura social del pasado que pueden correlacionarse con las observaciones morfológicas y arqueológicas extraídas del registro fósil. Por supuesto, este enfoque requiere una mayor integración con otras disciplinas y una verdadera visión multidisciplinaria del estudio del pasado. La aplicación de diferentes enfoques *ómicos*, algunos de ellos aún en fases tempranas de desarrollo, como la recuperación de péptidos antiguos (*paleoproteómica*), promete impulsar significativamente el campo de la evolución humana en los próximos años.

Paralelamente, estos emergentes conjuntos de datos genéticos antiguos permitirán el cribado directo de las variantes genéticas adaptativas que han aumentado a lo largo del tiempo hasta alcanzar las frecuencias actuales. Así, cabe la posibilidad de estudiar en detalle en un futuro próximo, directamente en el espacio y el tiempo, los procesos selectivos que pudieran tener implicaciones biomédicas. No obstante, estos genomas son indicativos de eventos selectivos que tuvieron lugar en los últimos cientos de miles de años de la evolución humana.

Se ha demostrado que algunos rasgos evolutivamente relevantes (por ejemplo, el sentido de la equidad o la capacidad de resolver problemas) no son exclusivos de nuestra especie, sino que pueden haber sufrido una evolución convergente en varias ocasiones. Es importante destacar que algunos rasgos de relevancia biomédica y ecológica, como el tamaño y la función del cerebro, la duración de la vida, la dieta o la fertilidad, son especialmente adecuados para un enfoque filogenómico. Los humanos, por ejemplo, se diferencian notablemente de la mayoría de los mamíferos por una serie de características, como tener el cerebro más grande de la naturaleza (en relación con el tamaño del cuerpo) con una superficie notablemente mayor debido a su plegado específico. El modo en que se logra esta complejidad y diversidad sigue siendo en gran medida desconocido y solo se puede tratar a fondo si se llevan a cabo exhaustivos estudios filogenéticos sobre la encefalización en todo el linaje de los mamíferos, incluido el estudio de cualquier proceso adaptativo pertinente. De nuevo, estas preguntas ofrecen la oportunidad de realizar estudios evolutivos que iluminen la evolución humana y tengan un impacto biomédico relevante.

Paralelamente a los avances en los enfoques *ómicos* y sus requisitos multidisciplinarios, la arqueología evolutiva está a la vanguardia de los avances disciplinarios en las ciencias sociales debido a su continua interacción con las ciencias naturales y a su dependencia del progreso de las técnicas, desde la física y la química, para desarrollar nuevos enfoques analíticos en la datación y la caracterización de los conjuntos arqueológicos.

Un compendio de los grandes desafíos del siglo XXI para la arqueología (Kin-tigh *et al.*, 2014) reconoce que las cuestiones clave de la disciplina ya no tienen que ver con lo más antiguo, lo más longevo o lo exclusivo. En todo caso, la profundidad de edad cada vez mayor para la aparición y dispersión de los seres humanos prehistóricos solo exacerba las complejidades de los primeros datos arqueológicos. A medida que las diferencias entre los yacimientos

conocidos se amplían en el tiempo y en el espacio, resulta crucial ir más allá de los debates sobre la primera apariencia y centrarse en los patrones biogeográficos y culturales implicados. Los sustanciales esfuerzos de trabajo de campo en las últimas décadas han producido extensos conjuntos de datos arqueológicos, y ahora es el momento de la investigación sintética a gran escala en la evolución cultural (Brewer *et al.*, 2017).

En cuanto a la transferencia de resultados y aplicaciones potenciales: 1) Los avances en nuestra comprensión de la naturaleza de la movilidad y las migraciones en la Prehistoria proporcionarán datos valiosos para cuestionar las interpretaciones simplistas de las migraciones en el pasado que, a veces, se utilizan (erróneamente) para construir narrativas artificiales de las identidades nacionales. 2) Las infraestructuras de datos espaciales son el futuro del intercambio de datos en las ciencias sociales (como ya es habitual en las ciencias biológicas) y servirán tanto para impulsar la colaboración científica internacional como para promover la difusión de los resultados académicos al público. 3) Una unidad de campo especializada puede promover la colaboración con el sector privado mediante su participación en proyectos civiles relacionados con yacimientos arqueológicos relevantes para este desafío.

### 3. PUNTOS CLAVE DEL DESAFÍO

#### 3.1 Aplicación de la *multiómica* al pasado reciente

*Paleogenómica.* La aparición de la paleogenómica o arqueogenómica humana (la recuperación y el análisis de los genomas humanos del pasado), impulsada por el desarrollo de las tecnologías de secuenciación de última generación (NGS, por sus siglas en inglés) a partir de 2005, ha proporcionado una nueva y poderosa fuente de información que puede utilizarse para poner a prueba las hipótesis anteriores sobre la evolución humana y las migraciones pasadas. En el campo de la historia de la humanidad, la paleogenómica ha aportado respuestas a antiguos debates sobre las relaciones evolutivas de nuestra especie y algunos linajes antiguos como los neandertales y los denisovanos, además de informar sobre otras facetas importantes de la experiencia humana, como las migraciones pasadas, la estructura social o los movimientos de la población en función del sexo. La interacción entre arqueólogos, antropólogos e investigadores del emergente campo de la paleogenómica ha carecido siempre de esfuerzos de colaboración. Estos desacuerdos entre los distintos campos pueden explicarse, en parte, por algunas limitaciones asociadas a los marcadores genéticos que se solían emplear hace unos años (por ejemplo, el

ADN mitocondrial) antes de la llegada de las tecnologías de NGS, pero también por la forma en que los genetistas han abordado el estudio del material arqueológico. Por lo tanto, la posibilidad de trabajar con genomas antiguos es una oportunidad para crear un enfoque más integrador del estudio del pasado en el que participen diferentes institutos de investigación centrados en distintas disciplinas (Racimo *et al.*, 2020).

Las mejoras en la bioinformática y la inferencia genética de las poblaciones han servido para extraer una información inestimable de estos genomas, incluidos los patrones de crecimiento y contracción de la población, el cruce entre grupos de parentesco lejano y las pruebas de la selección natural que opera en los loci fenotípicamente importantes. Algunas de las limitaciones tradicionales del ADN antiguo se han resuelto, en gran medida, con la aplicación de las tecnologías de NGS, mientras que otras han ido emergiendo. Y lo que es más importante, la NGS no depende de la amplificación por reacción en cadena de la polimerasa (PCR) de las moléculas cortas antiguas mediante cebadores. El ADN endógeno tiende a degradarse a fragmentos más cortos con el tiempo y las condiciones ambientales (principalmente la temperatura, pero también la humedad y el pH del suelo), esta posibilidad significa que ahora se tiene acceso a una fracción mucho mayor de ADN endógeno que con el uso de la PCR. Otra ventaja clave de la NGS es que permite utilizar los patrones de degradación en los extremos de las lecturas (un aumento de la sustitución de C a T en el extremo 5' y de las sustituciones de G a A en el extremo 3') para discriminar entre los contaminantes de ADN modernos y el ADN endógeno. Con una cobertura suficiente y herramientas de imputación, ahora es posible atribuir secuencias específicas a los daños o a la contaminación. Por lo tanto, aunque no se controle la contaminación *a priori* en un lugar concreto, puede controlarse *a posteriori*, lo que permite obtener estimaciones a partir del ADNmt de alta cobertura o de las heterogeneidades en los cromosomas sexuales. En estos genomas antiguos, serían aceptables estimaciones de contaminación por debajo del 1-2 %.

Sin embargo, una característica crítica ahora es la eficiencia de una muestra particular en la secuenciación de escopeta (es decir, la proporción de ADN endógeno frente al ADN contaminante ambiental). Esto no era importante con el enfoque basado en la PCR, ya que se diseñaron y utilizaron cebadores específicos para recuperar las secuencias de ADN de interés, independientemente del contenido ambiental de la muestra. Al resumir los datos de diferentes estudios, se observa que las muestras convencionales de huesos o dientes

procedentes de un entorno europeo templado suelen tener una eficacia inferior o raramente superior al 5-10 %. En el sur de Europa, esta cifra suele oscilar entre el 0,1 y <2 %. Las muestras con un contenido endógeno superior al 50 %, como Ötzi, proceden de contextos arqueológicos inusuales y condiciones tafonómicas únicas, pero probablemente constituyen una rareza en la Europa prehistórica. Por lo tanto, una estrategia de muestreo adecuada resulta crucial para estos nuevos enfoques paleogenómicos. Otro avance ha sido el reciente descubrimiento de que los huesos petrosos (una región muy dura del hueso temporal) son los mejores contenedores de ADN.

A pesar de los problemas metodológicos mencionados, para la mayoría de los análisis genéticos de poblaciones no se requieren secuencias del genoma completo, y un conjunto de varios cientos de miles de posiciones informativas de un solo nucleótido o SNP puede permitir análisis de alta resolución. Cada vez está más claro que a la mayoría de las muestras antiguas no se las puede someter a la secuenciación de escopeta, aunque tienen suficiente contenido de ADN para capturarlo para el genotipado de SNP. Si bien resulta difícil calcular cifras precisas, pudiera ser que aproximadamente una de cada diez o veinte muestras tenga la posibilidad de someterse a una secuenciación de escopeta, mientras que en el mismo periodo se pudieran genotipar los SNP de alrededor de una de cada dos o tres muestras.

A pesar de las limitaciones técnicas, los estudios paleogenómicos están descifrando los complejos patrones evolutivos de los linajes humanos y mostrando múltiples eventos de mezcla en diferentes momentos y regiones, además de proporcionar información sobre las adaptaciones a las condiciones ambientales, las migraciones pasadas, las tendencias demográficas y las estructuras sociales. Es probable que estos estudios se conviertan en la opción convencional en unos 10 años más, pero ahora mismo siguen teniendo un gran potencial científico.

*Paleoproteómica.* La paleogenómica constituye un valioso recurso de información de especies extinguidas y poblaciones del pasado. Sin embargo, hay un plazo claro más allá del cual no se espera que sobreviva ningún ADN. Actualmente se desconoce este límite (en parte porque depende mucho de las condiciones térmicas del entorno), pero no parece ir más allá de medio millón de años en condiciones templadas o quizá hasta un millón de años en condiciones extremadamente favorables (de tipo polar). Ahora mismo, el registro absoluto en condiciones templadas corresponde al yacimiento de la Sima de los Huesos situada en Atapuerca (fechaado hace unos 430 000 años), mientras que el de

condiciones favorables corresponde a un hueso de caballo de hace unos 700 000 años encontrado en Alaska. En entornos cálidos, incluidos los tropicales o subtropicales, el ADN solo puede sobrevivir unas decenas de miles de años.

En cambio, las proteínas antiguas constituyen una fuente de información molecular más duradera. Algunos estudios recientes han informado de la supervivencia de proteínas hasta el rango de 2 millones de años, especialmente en el esmalte dental, que parece conservar alrededor de media docena de proteínas implicadas de forma precisa en la formación de este tejido. Aunque estas proteínas son, por definición, muy estables en la evolución, estas secuencias peptídicas tan antiguas pueden utilizarse de forma fiable para crear árboles moleculares que reconstruyan las relaciones evolutivas entre especies existentes y extintas.

En última instancia, la secuenciación de paleoproteínas puede hacer que las reconstrucciones evolutivas moleculares se remonten mucho más atrás en el tiempo que el análisis del ADN antiguo. Esto resulta especialmente interesante porque puede proporcionar información filogenética en el marco temporal crítico de unos 2-3 millones de años (como mínimo) que vio la aparición de nuestro propio género, *Homo*, y su primera dispersión fuera de África, así como la aparición en el registro fósil de algunos linajes de *Australopithecus*, incluidas las llamadas formas *robustas*. Ahora mismo se ha aplicado con éxito a un rinoceronte extinto (*Stephanorhinus*) del yacimiento de Dmanisi, en Georgia (datado en 1,77 millones de años), al extinto *Gigantopithecus* de China (datado en 1,9 millones de años) (Welker *et al.*, 2019) y a los restos paradigmáticos de *Homo antecessor* de Atapuerca (datados en 800 000 años) (Welker *et al.*, 2020). Si la técnica puede retroceder aún más (actualmente se desconocen los límites temporales), es posible que también se puedan analizar los homínidos antiguos (y quizá los del Mioceno).

Por lo tanto, es probable que, en los próximos 10-15 años, el campo de la evolución humana (y, en general, la paleontología animal) se vea muy afectado por los avances de la paleoproteómica, lo que también exigirá profesionales cualificados con una formación que combine el análisis computacional y filogenético, la experiencia en espectrometría de masas y la formación paleoantropológica.

*Paleoepigenómica.* Además de la genómica comparativa tradicional entre humanos y simios o entre humanos y neandertales (o denisovanos), una fuente adicional de información para comprender lo que pueden constituir los

rasgos humanos exclusivos es el estudio de los paleoepigenomas. Se puede plantear la hipótesis de que, además de la comparación de la secuencia genómica, muchos rasgos genómicos específicos de los humanos con efecto fenotípico pueden identificarse comparando los programas de regulación de los genes, ya que se considera que las diferencias en la regulación de los genes son una fuerza evolutiva importante que explica las diferencias fenotípicas entre especies estrechamente relacionadas. En consecuencia, para entender las diferencias fenotípicas entre los distintos grupos humanos, resulta fundamental comprender las formas en que difieren en sus programas de regulación.

La metilación del ADN se produce casi exclusivamente en las citosinas en el contexto de los dinucleótidos CpG. La metilación del ADN de los potenciadores es una característica clave de la actividad de los genes en los mamíferos y está inversamente correlacionada con el nivel de expresión. Esta información puede extraerse de genomas antiguos de alta calidad (es decir, genomas de alta cobertura, aunque es probable que los futuros algoritmos permitan trabajar también con genomas de baja cobertura) y pudiera ayudar a entender cómo se regulaban estos genomas. El enfoque se basa en los procesos naturales de degradación del ADN antiguo, en los que las citosinas metiladas y no metiladas se desaminan con el tiempo en timinas y uracilos, respectivamente.

Una forma de estudiar las diferencias en los patrones epigenómicos consiste en identificar las regiones que están metiladas de forma diferencial en humanos, humanos arcaicos y grandes simios. Las posibles diferencias encontradas en los patrones de metilación pudieran probarse posteriormente en líneas celulares y organoides humanos. La tecnología de edición del genoma CRISPR/Cas9 puede utilizarse para generar versiones arcaicas de los genes, así como dCas9 fusionado con enzimas de desmetilación del ADN para alterar específicamente los patrones de metilación del ADN y probar su efecto en el contexto de los organoides humanos.

La paleoepigenómica es un campo emergente que probablemente proporcionará importantes avances en la evolución y la diversidad humanas en los próximos 10-15 años, en especial a medida que se disponga de más genomas antiguos de alta calidad. Un estudio reciente ha sido capaz de predecir, con sorprendente precisión, la morfología de un cráneo denisovano solo a partir del análisis epigenómico del genoma denisovano de alta cobertura (Gokhman *et al.*, 2019). Cabe mencionar también que ahora mismo solo hay cinco epigenomas de antiguos *H. sapiens* analizados, uno de los cuales corresponde a España (el genoma del Mesolítico de La Braña 1, en León).

### **3.2 Enfoque filogenético comparativo de especies distantes para estudiar la evolución humana**

#### *Análisis filogenético del genoma-fenoma*

Para entender los procesos biológicos poligénicos y complejos, como los que pueden ser exclusivos de los humanos, es posible adoptar una estrategia filogenética comparativa aprovechando los datos *multiómicos*. Algunos cambios genómicos (por ejemplo, los cambios de aminoácidos, las tasas de evolución de las proteínas o los estados de multicopia de los genes, por nombrar solo algunos) pueden asociarse a cambios en los rasgos ecológicos, antropométricos o biomédicos de interés (incluidos, entre otros, el aumento o la reducción de la vida y el tamaño del cerebro). Estos cambios pueden comprenderse mejor si se analiza un contexto evolutivo amplio, es decir, si se estudia una amplia selección de especies de vertebrados, mamíferos y primates, cada vez más disponibles, que abarquen nodos filogenéticos de gran valor informativo, en lugar de centrarse únicamente en los humanos y sus parientes vivos más cercanos. Con este enfoque filogenético comparativo, es posible identificar correlaciones moleculares evolutivas (por ejemplo, mutaciones paralelas) entre la variación genética en un contexto amplio y rasgos humanos específicos.

Para este objetivo, ya se dispone de técnicas actuales como el método de mínimos cuadrados generalizados filogenéticos (PGLS), el análisis de rutas filogenéticas, el estudio filogenético de asociación del genoma completo (GWAS) o los métodos bayesianos, y solo queda afinarlas, adaptándolas a cuestiones como la ordenación incompleta de los linajes (para poder utilizar el árbol adecuado para cada segmento del genoma) o la inclusión, además de la variación de nucleótidos, de cambios regulatorios o de grandes cambios como las duplicaciones segmentarias.

*Evolución del cerebro humano.* El cerebro humano es el resultado de intrincados procesos de desarrollo que se desarrollan a lo largo de décadas, durante las cuales distintos tipos de células maduran a medida que cambian sus identidades moleculares, morfológicas y funcionales a través de la precisa regulación espacio-temporal de su transcriptoma. Identificar los rasgos específicos del desarrollo neurológico humano y cómo aparecieron durante la evolución resulta fundamental para entender la evolución de las características y capacidades distintivas del cerebro humano. Entre estos rasgos acentuados (pero no exclusivos), se incluyen el aumento multiplicado por mil del tamaño cortical en comparación con el ratón, o el triple aumento en comparación con

nuestros parientes más cercanos, los chimpancés, así como nuestras capacidades cognitivas de nivel superior (como el pensamiento abstracto, el lenguaje sintáctico-gramatical o la memoria episódica), propiedades estructurales y hodológicas específicas, y trastornos cerebrales específicos.

Cada vez hay más pruebas de que la desregulación de los eventos transcripcionales y reguladores clave en el neurodesarrollo tiene consecuencias directas y profundas en la función cerebral, o que favorece en gran medida el riesgo de trastornos neuropsiquiátricos. Para revelar los eventos moleculares que subyacen a la evolución de las características del cerebro humano y su papel en la enfermedad, es necesario identificar variantes específicas que se producen en elementos funcionales del genoma, además de realizar pruebas experimentales. Sin embargo, encontrar las variantes relevantes que subyacen a los cambios fenotípicos en el cerebro en desarrollo y maduro constituye una difícil tarea, ya que exige la identificación de relativamente pocas variantes significativas entre miles de variantes neutras, lo que supone *una aguja en un pajar*. Una primera aproximación para reducir el espacio de búsqueda pudiera consistir en seguir un enfoque *evo-devo* (del inglés *evolutionary developmental biology*, o biología evolutiva del desarrollo), comparando la regulación genómica del desarrollo del cerebro a través de la filogenia de los mamíferos para identificar las señales de selección positiva o negativa que han surgido durante la evolución del cerebro que conduce a los humanos. Ya se ha demostrado que el trabajo con especies de especial valor filogenético estratégico, como el ratón, el hurón y el macaco, resulta extremadamente potente para identificar y comprender los determinantes genómicos de la evolución del cerebro humano. Centrarse en los genes que se expresan en el cerebro y que muestran señales de asociación con rasgos neuropsiquiátricos (GWAS/variantes raras) o señales de selección positiva que han surgido desde la separación de nuestros parientes más cercanos puede tener un valor único para la identificación de las características específicas humanas del desarrollo y la evolución del cerebro.

Uno de los principales obstáculos para estudiar el desarrollo del cerebro en los seres humanos y nuestros parientes más cercanos (los chimpancés) es la falta de muestras. El descubrimiento de las células madre pluripotentes inducidas (iPSC, por sus siglas en inglés) y su capacidad para diferenciarse en cualquier tipo de célula alberga un gran potencial para investigar el desarrollo en condiciones de laboratorio en una cantidad, en principio, ilimitada de muestras y a una escala temporal fina. Por otra parte, la comprensión de los

mecanismos evolutivos que influyen en el desarrollo del cerebro y que conducen a la adquisición de los rasgos humanos puede lograrse con éxito estudiando especies de no primates con posiciones estratégicas en la filogenia y que muestran rasgos cerebrales únicos. Entre ellos se encuentran el ratón (roedor con un cerebro pequeño y liso), el hurón (carnívoro con un cerebro mediano y plegado), el tití (primate con un cerebro liso y pequeño), el macaco (primate con un cerebro grande y plegado) y el ser humano. Los estudios genómicos y de desarrollo comparativos entre estas especies ya han demostrado el extraordinario potencial de los estudios evo-devo —o biología evolutiva del desarrollo— para la identificación de las trayectorias evolutivas del desarrollo del cerebro de los mamíferos que conducen al humano. Y, lo que es más importante, desde la llegada de la edición del genoma mediada por CRISPR, se están realizando manipulaciones genéticas del desarrollo del cerebro en algunas de las especies modelo (sobre todo, en el ratón) *in vivo* en el embrión en desarrollo, lo que tiene un valor inigualable en comparación con cualquier enfoque *in vitro*.

Un segundo obstáculo importante es el desafío que supone comparar los tipos de células y su organización a escala fina entre especies. Los recientes avances en la elaboración de perfiles unicelulares en paralelo y de forma masiva (transcriptómica, epigenómica, proteómica) están permitiendo el cotejo y la comparación de tipos de células homólogas entre especies. Los primeros análisis comparativos de células individuales en cerebros en desarrollo indican una conservación de amplias poblaciones celulares entre humanos y ratones, pero diferencias dramáticas en las proporciones de los tipos de células, su distribución laminar y sus morfologías. Esta tecnología en rápida evolución pronto permitirá el análisis simultáneo de múltiples características en células individuales con una resolución mucho más fina, en mayor número y a mayor velocidad, lo que, combinado con los métodos analíticos que mejoren rápidamente, resulta muy prometedor para los estudios de la evolución de los mecanismos celulares en el desarrollo del cerebro. Por lo tanto, los campos de la biología evolutiva y neurodesarrollo (con nuevos y refinados modelos animales experimentales y herramientas de edición del genoma) están listos para combinarlos con la intención de poner a prueba las hipótesis sobre los cambios moleculares que tuvieron lugar en la evolución del cerebro humano en desarrollo, que condujeron a su mayor tamaño y complejidad arquitectónica y, a partir de ahí, explicar el mal funcionamiento observado en los estados de enfermedad.

Un desafío científico apremiante es dilucidar las bases genéticas de la evolución de las características específicas del linaje del neurodesarrollo y la enfermedad en los seres humanos, con respecto tanto a los primates no humanos más cercanos como a otras especies de mamíferos más distantes en pasos clave del proceso evolutivo. Para probar eficazmente esta cuestión será necesario utilizar sistemas *in vitro* próximos al desarrollo del cerebro fetal humano, así como modelos animales no convencionales que presenten características relevantes en esta cuestión. El desarrollo de líneas de células madre pluripotentes inducidas y organoides cerebrales de humanos y otros grandes simios son prometedores enfoques *in vitro* para desentrañar los procesos funcionales del desarrollo temprano del cerebro que conducen a rasgos específicos del ser humano. La edición del genoma de estas líneas celulares iPSC para sustituir las secuencias endógenas por las de especies o grupos estratégicamente relevantes y producir después organoides cerebrales para el análisis funcional del desarrollo del cerebro es muy prometedora para la próxima generación de descubrimientos en este campo. Por ejemplo, la edición de iPSC humanas con secuencias de neandertales y denisovanos, o de iPSC de chimpancés con secuencias de humanos modernos, puede ayudar a descifrar las diferencias metabólicas y fisiológicas entre especies. Estos enfoques requieren el acceso a células vivas de grandes simios, para su transformación en iPSC, y también permitirán desentrañar las redes moleculares y génicas que subyacen a las diferencias entre humanos y grandes simios en otros tejidos. En cuanto a la elucidación de los primeros pasos de la evolución del cerebro de los mamíferos que conducen al humano, esto implicará el uso de especies no convencionales susceptibles de manipulación genética experimental en el embrión en desarrollo, así como de relevancia estratégica para esta cuestión debido a su posición clave a través de la filogenia humana a largo plazo. Entre las especies en las que se ha logrado la transgénesis transitoria y estable de genes que afectan al desarrollo del cerebro y que, por tanto, se perfilan como modelos animales de próxima generación para estudiar el desarrollo y la función del cerebro, se encuentran el macaco, el tití y el hurón.

En resumen, la comprensión de la evolución del cerebro humano requerirá i) un análisis bioinformático *in silico* y multiómico que aproveche múltiples tipos de datos procedentes de múltiples fuentes (células individuales frente a tejidos en masa; datos transcriptómicos, de organización de la cromatina y proteómicos; escaneos de selección y estudios de asociación) para priorizar las variantes genéticas evolutivamente relevantes que afecten a los genes candidatos relacionados con los fenotipos del neurodesarrollo específicos de cada

linaje; ii) el uso de herramientas de edición del genoma en modelos animales no convencionales pero estratégicos, así como en células madre pluripotentes inducidas (iPSC) diferenciadas en linajes neuronales y organoides cerebrales, para comprobar las alteraciones de los parámetros citológicos (por ejemplo, morfología celular, composición celular y neurogénesis) y moleculares (por ejemplo, transcriptoma y epigenoma). Por ejemplo, se pudiera introducir una variante específica humana en iPSC de chimpancé para comprobar si puede cambiar, y en qué medida, las redes moleculares y los fenotipos a un estado más parecido al humano. Este enfoque es el más adecuado para estudiar los primeros procesos de desarrollo y el único posible ahora que la experimentación con individuos simios está, con razón, prohibida en investigaciones.

### **3.3 Avances en paleoantropología para el estudio de la evolución humana**

Si bien los temas anteriores hacen referencia principalmente a las innovaciones en materia de biología molecular y genética de los datos extraídos de los restos fósiles de los homínidos, a nivel fenotípico también se están produciendo importantes avances en el estudio de la evolución humana.

*Anatomía digital, imágenes 3D avanzadas y bases de datos virtuales.* Las mejoras metodológicas de los últimos tiempos han llevado los estudios cuantitativos de la morfología fósil de los homínidos a nuevos niveles. Esto se ha visto facilitado por los avances en las técnicas de obtención de imágenes digitales de alta resolución, tanto a nivel interno (volumen) mediante el escaneo industrial avanzado de microtomografía o nanotomografía como a nivel externo (superficie). La aplicación de estos métodos de imagen a los fósiles de homínidos ha mejorado la posibilidad de cuantificar las características internas y externas de la morfología de los tejidos duros, lo que ha permitido comprender mejor los antiguos procesos biológicos relacionados con el crecimiento, la morfogénesis y, en general, la ontogenia. Por ejemplo, la aplicación de métodos de alta radiación a los dientes fosilizados ha permitido comprender el cambio de las tasas de crecimiento no solo de la dentición, sino también de la historia del organismo y de la vida en las especies recientes de homínidos, así como de los fósiles craneodentales altamente mineralizados en el origen de los géneros *Homo*, *Australopithecus* y *Paranthropus*.

A otro nivel, las imágenes avanzadas en 3D también han dado lugar a bases de datos digitales más fácilmente accesibles de restos fósiles digitalizados. A este

respecto, los análisis en curso de la nueva especie de homínido *Homo naledi* (cuevas Rising Star, Sudáfrica) han proporcionado un ejemplo pionero en la mejora de la investigación paleoantropológica al facilitar el acceso abierto y gratuito a los modelos digitales en 3D para la investigación anatómica comparativa o la impresión en 3D o las excavaciones en línea con distribuciones en directo a través de Internet entre otros, lo que fomenta la democratización de la investigación paleoantropológica, que siempre se ha caracterizado por un acceso limitado a las instituciones privilegiadas.

Cada vez son más las instituciones de investigación que siguen estas tendencias participando en redes europeas y mundiales de actividades para la recopilación de digitalización de especies de homínidos y primates fósiles y existentes relevantes para el estudio de la evolución humana. Garantizar la participación actual y futura del CSIC en estas redes de colecciones digitales (DISSCo, SYNTHESYS, etc.) resultará crucial para mantener su posición en los dominios fenotípicos de los estudios evolutivos humanos.

*Morfometría 3D de alta densidad y resolución.* Los especímenes fósiles digitales a disposición de los laboratorios morfológicos virtuales siguen cada vez más las nuevas tendencias en materia de análisis morfométrico y modelación. Tras el desarrollo pionero de los métodos morfométricos geométricos para la cuantificación de las geometrías de las curvas y las superficies de los objetos anatómicos virtuales en 3D, los esfuerzos actuales se centran principalmente en el desarrollo de métodos para el análisis de las mediciones en 3D de alta densidad (HD) de cobertura y de métodos para el registro de toda la superficie mediante técnicas sin puntos de referencia para el análisis y el modelado sofisticados de las formas.

Estos nuevos métodos morfométricos no solo permiten realizar potentes análisis cuantitativos y visuales sino también reconstrucciones cuantitativas basadas en hipótesis de restos fósiles anatómicamente incompletos (es decir, fragmentados). Las reconstrucciones cuantitativas basadas en hipótesis de las estructuras corporales de los homínidos fósiles son fundamentales para analizar la función y, por lo tanto, la evaluación de su posible valor adaptativo, que cada vez es más fácil de comparar con las pruebas genéticas en contextos funcionales. Además, la reconstrucción de las estructuras corporales también es importante para modelar los procesos de crecimiento de las estructuras que permiten estudiar la interacción entre los organismos homínidos fósiles y sus ecosistemas (eco-evo-devo). Por ejemplo, la morfometría 3D de alta definición de las reconstrucciones de cráneos de homínidos fósiles se ha

combinado recientemente con datos masivos sobre imágenes cerebrales y secuencias de genomas antiguos para proporcionar un modelo de la evolución y aparición de las funciones cerebrales y la morfología del cráneo de los humanos modernos. Otra vía de desarrollo futuro es la mencionada relación entre los datos epigenómicos y su potencial para la reconstrucción morfológica de los perfiles esqueléticos de especies de homínidos como los denisovanos.

Además, es probable que las funciones corporales se estudien con mayor eficacia utilizando la morfometría de alta definición aplicada a la anatomía virtual de los homínidos digitales existentes y extintos. Las simulaciones de funciones corporales como la locomoción o la respiración en estructuras 3D reconstruidas y sistemas de órganos basados en la modelación informática de procesos 4D en tiempo real relevantes para la evolución humana tienen el potencial no solo de producir un conocimiento paleoantropológico más preciso, sino también de extenderlo al uso en la investigación biomédica aplicada, el deporte y otras ciencias relacionadas con los seres humanos vivos. Debido a la necesidad de extraer la mayor cantidad de información posible de los escasos restos fósiles de homínidos, la investigación paleoantropológica ha interactuado siempre con la tecnología biomédica para mejorar los métodos analíticos, además de los constantes esfuerzos en el campo para proporcionar un registro fósil más completo para los enfoques multidisciplinarios de estudio de los procesos en la evolución humana.

### **3.4 El papel que desempeña la tecnología en la evolución humana**

*La tecnología como motor de la evolución social.* Definida como un conjunto de dispositivos materiales e informativos, el objetivo de la tecnología en el contexto de la evolución humana se centra en optimizar el intercambio de energía con el entorno (White, 1946). El éxito adaptativo humano se basa, en última instancia, en la eficiencia termodinámica de las sociedades y en su capacidad para progresar mediante la introducción acumulativa de nuevos dispositivos tecnológicos que compensen el aumento demográfico y los cambios en las condiciones ambientales o de subsistencia. Por lo tanto, la interacción entre el cambio tecnológico y el social es una característica primordial del proceso evolutivo humano y es un tema de alta prioridad en la arqueología prehistórica como disciplina. En los próximos años habrá que afrontar tres desafíos concretos: 1) La naturaleza de los sistemas tecnológicos como dispositivos termodinámicos y la manera en que la trayectoria evolutiva de las sociedades está conformada por procesos de producción alternativos. 2) El impacto de la tecnología en los sistemas sociales: la interacción entre la tecnología y la

organización social, que es bidireccional, ya que la tecnología da forma a los resultados termodinámicos del trabajo humano y determina aspectos básicos de la organización social, como la división del trabajo y el control de los medios de producción, entre otros. 3) El papel que desempeñan la transmisión y el cambio tecnológico en el proceso de evolución social; los conocimientos y las habilidades técnicas no están determinados genéticamente en los seres humanos y se transfieren a través de procesos de aprendizaje regulados socialmente, que deben estudiarse desde la perspectiva de los sistemas sociales en los que están inmersos.

*La singularidad de la cultura humana y la aparición de la tecnología en nuestro pasado evolutivo.* La idea clásica de las primeras tecnologías humanas supone que la fabricación de herramientas de piedra a partir de hace unos tres millones de años se centró en los bordes cortantes, principalmente a través de láminas obtenidas mediante talla lítica (Harmand *et al.*, 2016). Sin embargo, algunos autores (por ejemplo, Haslam *et al.*, 2009) han destacado el papel fundamental y más amplio de otras actividades con uso de herramientas aparte de la talla, que pudieran haber sido las más importantes en las primeras etapas de la tecnología humana. Esto tiene importantes implicaciones evolutivas, ya que algunos primates no humanos actuales utilizan herramientas en diversas tareas. Los paralelismos entre las tecnologías de percusión de los primates y los primeros yacimientos arqueológicos están produciendo resultados pioneros en los últimos años (por ejemplo, Proffitt *et al.*, 2016) y es necesario seguir explorando para salvar las diferencias entre campos como la primatología y la arqueología, fomentando las colaboraciones interdisciplinarias destinadas a sentar las bases para una comprensión de enfoque múltiple de los fundamentos evolutivos del comportamiento tecnológico humano. Las preguntas que impulsan esta investigación son las siguientes: ¿Existen motivos ecológicos que impulsen los patrones tecnológicos variables detectados en el registro arqueológico y primatológico temprano? ¿En qué se parecen las herramientas arqueológicas y los objetos líticos no fabricados por el hombre? ¿Representan las mismas actividades? ¿Existen paralelismos que permitan comparar las herramientas de piedra de los primates con las de los primeros arqueólogos?

### **3.5. Las respuestas humanas a los factores abióticos y bióticos y el papel que desempeña el clima en la evolución del comportamiento humano**

En los estudios sobre la evolución humana, existe una abrumadora preferencia por las interpretaciones abióticas de las adaptaciones y la evolución de los homínidos, desde los modelos que destacan el papel que desempeña el clima en la configuración de la especiación (deMenocal, 2011) hasta los que ven los cambios de especies en el contexto del cambio climático. Sin embargo, aunque la influencia de los mecanismos abióticos en las distribuciones biogeográficas de los primeros humanos es evidente, no explican de forma satisfactoria los patrones observados en el registro arqueológico y paleontológico. Estudios recientes sobre los cambios de la fauna han subrayado la importancia de los factores continuos (y probablemente bióticos) en la modulación del cambio de la fauna (Bibi y Kiessling, 2015), lo que sugiere que las comunidades de mamíferos pudieran ser más robustas de lo esperado, y que las propiedades de las redes alimentarias están desvinculadas de los cambios climáticos del Pleistoceno. Cada vez hay más pruebas de que los factores climáticos no explican por sí solos la dinámica biogeográfica de los homínidos, y las perspectivas de investigación pioneras están explorando modelos en los que las interacciones bióticas (por ejemplo, la competencia), que se sabe que son esenciales para configurar la biodiversidad y los patrones biogeográficos, pueden haber desempeñado un papel fundamental en la configuración de los primeros comportamientos adaptativos de los homínidos. Este campo emergente presenta una serie de desafíos, entre los que se incluyen: 1) la exploración de hipótesis contrapuestas sobre el carácter y la influencia de las fluctuaciones ambientales en las comunidades de mamíferos del Pleistoceno mediante la elaboración de modelos computacionales; 2) la calibración de los datos geoquímicos de los estudios paleoclimáticos y de paleovegetación en las secuencias arqueológicas y paleontológicas para comprobar los mapas climáticos y ayudar a interpretar los resultados de los modelos; y 3) la creación de bases de datos georreferenciadas de mamíferos y homínidos modernos y fósiles que deberían combinarse con las capas climáticas para perfeccionar los modelos biogeográficos de distribución de las especies en el Viejo Mundo.

### **3.6 Repositorios de datos masivos e infraestructuras de datos espaciales (IDE)**

La investigación sobre la ocupación humana premoderna del Viejo Mundo ya no puede compartimentarse de forma regional, y cualquier estudio exhaustivo se debe basar en un análisis comparativo de las vías evolutivas y culturales

alternativas observadas en cada zona. En los últimos años se ha reconocido la necesidad de establecer similitudes y diferencias entre las primeras secuencias arqueológicas, pero los esfuerzos se han centrado en revisiones bibliográficas generales de las pruebas de cada región o en comparaciones de atributos concretos de artefactos arqueológicos específicos. La investigación sintética defendida por Kintigh *et al.* (2014) para la arqueología es ahora posible y muy necesaria para entender adecuadamente los patrones ecológicos y culturales de los primeros humanos. No obstante, esta investigación debe basarse en el análisis sistemático de conjuntos de datos cuantitativos (superando así la fase actual de comparaciones de revisiones bibliográficas superficiales) e incluir la modelación computacional. Los importantes conjuntos de datos producidos en los últimos años suponen una oportunidad única para comprender las pautas que estructuraron la ecología conductual de los humanos premodernos durante la primera colonización del Viejo Mundo, pero también un desafío debido a la disparidad de archivos, las formas tradicionales de investigación y la falta de comparaciones sistemáticas intrarregionales e interregionales. Estos desafíos y oportunidades excepcionales deben abarcar objetivos de gran alcance para comprender las trayectorias evolutivas alternativas adoptadas por los homínidos que compartían un fondo biológico y cultural global (es decir, eran humanos premodernos que utilizaban soluciones tecnológicas de la Edad de Piedra temprana), así como abordar cuestiones de investigación clave como las siguientes: 1) ¿Cómo se identifican las oleadas migratorias durante el Pleistoceno y cuántas fueron? 2) ¿Cuál es la continuidad de la ocupación humana en los presuntos centros evolutivos en contraposición al registro discontinuo de otras regiones del Viejo Mundo? 3) ¿Pueden los riesgos y desafíos a los que se enfrentan los humanos en cada región explicar la variabilidad de las adaptaciones de los homínidos?

## DESAFÍO 6 BIBLIOGRAFÍA

- Bibi, F. y Kiessling, W. (2015).** Continuous evolutionary change in Plio-Pleistocene mammals of eastern Africa. *Proceedings of the National Academy of Sciences* 112, 10623-10628.
- Brewer, J. et al. (2017).** Grand challenges for the study of cultural evolution. *Nature Ecology & Evolution* 1, 0070.
- De Menocal, P.B. (2011).** Climate and Human Evolution. *Science* 331, 540-542.
- Gokhman, D., Mishol, N., de Manuel, M., de Juan, D., Shuqrun, J., Meshorer, E., Marques-Bonet, T., Rak, Y., Carmel, L. (2019).** Reconstructing Denisovan anatomy using DNA methylation maps. *Cell* 179, 182-192.
- Harmand, S. et al. (2015).** 3.3-million-year-old stone tools from Lomekwi 3, West Turkana, Kenya. *Nature* 521, 310-315.
- Haslam, M. et al. (2009).** Primate archaeology. *Nature* 460, 339-344.
- Kintigh, K. W. et al. (2014).** Grand challenges for archaeology. *Proceedings of the National Academy of Sciences USA* 111, 879-880.
- Proffitt, T. et al. (2016).** Wild monkeys flake stone tools. *Nature* 539, 85-88.
- Racimo, F., Sikora, M., Vander Linden, M., Schroeder, H., Lalueza-Fox. (2020).** Beyond broad strokes: sociocultural insights from the study of ancient genomes. *Nature Review Genetics* 21, 355-366. doi: 10.1038/s41575-020-0218-z.
- White, Leslie A. (1946).** *The Science of Culture. A study on man and civilization.* Grove Press Books and Farrar, Strauss Giroux. New York.
- Welker, F. et al. (2019).** Dental enamel proteome shows that *Gigantopithecus* was an early divergent pongine. *Nature* 576, 262-265.
- Welker, F. et al. (2020).** The dental enamel proteome of *Homo antecessor*. *Nature* 580, 235-238. <https://doi.org/10.1038/s41586-020-2153-8>.

## RESUMEN PARA EXPERTOS

### D 2.6 - EVOLUCIÓN SOCIAL Y HUMANA

Comprendiendo las claves sociales y biológicas en el linaje evolutivo humano

#### Desafíos

- Reconstruir el árbol filogenético de los homínidos y las dispersiones humanas.
- Revelar adaptaciones humanas.
- Evolución genética.
- Evolución del cerebro humano.
- Evolución de la sociabilidad, comportamiento y tecnología humanos.
- Explorar la diversidad (genética y morfológica) y enfermedad humanas.

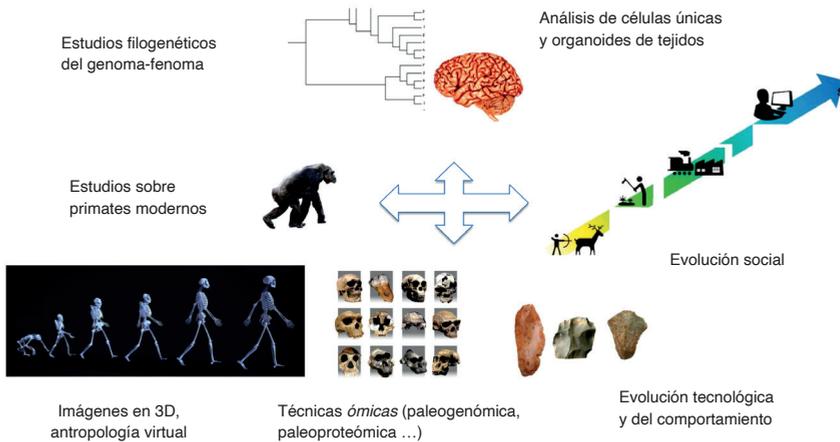
#### Paquetes de trabajo

- Aplicación y desarrollo de técnicas “ómicas” (paleogenómica, paleoproteómica, paleoepigenómica).
- Análisis filogenéticos genoma-fenoma.
- Perfilado de célula única.
- Organoides.
- Imágenes 3D en alta resolución.
- Repositorios Big Data e Infraestructuras de Datos Espaciales.

**Centros CSIC:** IBE; IH-CCHS; MNCN; INA; CNB; IMF; IACT; INCIPIT; IAM; EBD

**Tecnología:** gestión de Big Data (en genética, fósiles 3D y base de datos espacial de arqueología); secuenciación de alto rendimiento; espectrometría de masas; reconstrucciones peptídicas y paleoambientales; imágenes avanzadas en 3D; CRISPR; diferenciación de tejidos.

## RESUMEN PARA EL PÚBLICO GENERAL



## DESAFÍO 7

### RESUMEN

Las enfermedades son el resultado de la alteración de la homeostasis fisiológica de los organismos. Las perturbaciones pueden ser endógenas (por ejemplo, enfermedades hereditarias, limitaciones del desarrollo o cáncer) o exógenas (como infecciones o intoxicaciones). En la medida en que los propios organismos y su forma de interactuar con sus entornos bióticos y abióticos se deben a fuerzas evolutivas, sus enfermedades también son el resultado de complejos procesos (co)evolutivos. Hoy en día, al incorporar una perspectiva evolutiva, es posible que se produzca una mejor comprensión y se ayude a combatir las enfermedades.

### PALABRAS CLAVE

resistencia a antibióticos

cáncer como sistema evolutivo

patógenos emergentes de la medicina darwiniana

evolucionabilidad

coevolución huésped-patógeno

Una Salud

evolución de los patógenos

# EVOLUCIÓN DE LA SALUD Y LAS ENFERMEDADES

## Coordinadores

Santiago F. Elena  
(I2SysBio)

Iñaki Comas  
(IBV)

## Investigadores y centros de investigación participantes

Jesús Blázquez  
(CNB)

Jordi Gómez  
(IPBLN)

Elena Gómez-Díaz  
(IPBLN)

Jesús Pérez-Losada  
(CIC)

Francisco Sobrino  
(CBM-SO)

## RESUMEN EJECUTIVO

En este capítulo se describe cómo la incorporación de un pensamiento evolutivo generaría nuevos conocimientos y proporcionaría nuevas perspectivas sobre las enfermedades de los seres humanos, los animales de granja y las plantas. Se necesita urgentemente un cambio en el paradigma conceptual reduccionista para abordar problemas complejos como el origen, la difusión y la propagación de nuevos agentes infecciosos, el origen y la propagación de microbios resistentes a los fármacos antibióticos, el papel que desempeña la microbiota en la prevención de las invasiones de microbios patógenos, la dinámica de las células cancerosas y el crecimiento de los tumores, el origen de afecciones médicas como la diabetes, la obesidad, las enfermedades cardiovasculares, el asma, las alergias y las enfermedades autoinmunes, el envejecimiento, la preeclampsia y la menopausia. Todos estos procesos se rigen por las mismas fuerzas básicas y universales de la evolución: la mutación, la deriva genética, la migración, la selección natural y las compensaciones adaptativas. Cada vez está más claro que las soluciones a estos complejos problemas implican potencialmente todos los niveles de organización, desde las moléculas hasta las poblaciones. La síntesis evolutiva moderna, o teoría neodarwinista de la evolución, proporciona los marcos conceptuales, teóricos y matemáticos que, junto con la genómica, la bioinformática y una perspectiva de biología de sistemas integradora, brindarán una comprensión holística. Se podrían diseñar tratamientos quimioterapéuticos más seguros y duraderos,

estrategias de vacunación más eficaces que tengan en cuenta toda la biosfera y un uso más racional de los medicamentos teniendo en cuenta el inevitable componente evolutivo de los organismos vivos.

## 1. INTRODUCCIÓN Y DESCRIPCIÓN GENERAL

La forma en que los seres vivos responden a la enfermedad, ya sea infecciosa o se deba a un fallo en la homeostasis celular, es el resultado de un proceso de (co)evolución. Todo desarrollo racional de nuevas terapias en el futuro debería tener en cuenta este componente evolutivo. El ejemplo más evidente de la importancia de la biología evolutiva es el de las enfermedades infecciosas. Los seres humanos, los animales de granja y los cultivos vegetales están continuamente expuestos a patógenos que ejercen fuertes presiones de selección sobre la evolución de los genomas, lo que hace evolucionar complejos mecanismos de defensa. En respuesta a ello, la presión inmunitaria o la acción de los fármacos antibióticos sobre las poblaciones de patógenos da lugar a la evolución de contradefensas y variantes resistentes capaces de infectar y reproducirse. Esta *carrera armamentística* lleva a la coevolución entre ambos actores. Otro ejemplo paradigmático de proceso ecoevolutivo y enfermedad es el cáncer. En muchos sentidos, las células tumorales se comportan como parásitos de su huésped, se replican mucho más rápido y generan muchos más errores que las células normales, a las que desplazan de su entorno. La caracterización y la comprensión de la diversidad genética de susceptibilidad a las enfermedades que existen en diferentes poblaciones de huéspedes se ha convertido en el principal tema de muchos estudios en los últimos años. No es posible entender por qué ciertos individuos son más resistentes que otros sin conocer la historia evolutiva de estas poblaciones. La razón por la que los alelos de susceptibilidad a las enfermedades se han mantenido en altas frecuencias no puede entenderse sin conceptos evolutivos como las compensaciones de aptitud.

La vida ha demostrado una gran flexibilidad y puede encontrarse en cualquier entorno: hielo perpetuo, agua hirviendo, pH extremo, alta salinidad o presión abrumadora. Por lo tanto, cupiera esperar que los seres vivos puedan desarrollar mecanismos para resistir a cualquier fármaco que los humanos usen contra ellos. La resistencia constituye, por tanto, una respuesta evolutiva natural a la exposición a los fármacos. Por un lado, mantener la estabilidad de la información genética resulta vital para la perpetuación de las especies. Por ello, los organismos celulares desarrollaron un mecanismo de replicación del ADN

para minimizar los errores y proteger y reparar el ADN. La mayoría de los virus, por el contrario, no efectúan codificación para tales mecanismos de corrección de errores, ya que su estrategia de vida se basa en una replicación rápida pero de baja fidelidad. En ambos casos, los errores se producen durante la replicación del genoma. Por otro lado, la evolución adaptativa, que se produce a través de la variación hereditaria y la selección, permite a los organismos adaptarse a nuevos entornos o a condiciones adversas. Sin cambio genético, no hay nuevos genes o alelos y, por tanto, no hay adaptación. Por ello, los seres vivos tuvieron que enfrentarse al dilema de cambiar para adaptarse o permanecer estables y enfrentarse a la posibilidad de la extinción. La naturaleza ha seleccionado aquellos organismos que poseen una tasa de mutación que constituye un compromiso entre la adaptabilidad y la resistencia.

Las enfermedades infecciosas representan el 20 % de todas las enfermedades humanas ([www.who.int/whr/1996/media\\_centre/press\\_release/en/](http://www.who.int/whr/1996/media_centre/press_release/en/)). Los patógenos han acompañado al ser humano desde su origen, hace unos 200 000 años. La viruela, la peste bubónica, la gripe, la esquistosomiasis, la tuberculosis y la malaria han diezmado poblaciones enteras y han moldeado la evolución humana (Domínguez-Andrés y Netea, 2019). Aunque la ciencia produce cada día nuevos y extraordinarios resultados, todavía se sabe poco sobre las bases moleculares de las enfermedades infecciosas humanas. Como indicó la Organización Mundial de la Salud (OMS) en su último informe: «in the context of globalization, climate change and population growth, the risk and impact of emerging and re-emerging infectious diseases is extreme, and there is an urgent need to find new strategies for their eradication» (en el contexto de la globalización, el cambio climático y el crecimiento de la población, el riesgo y el impacto de las enfermedades infecciosas emergentes y reemergentes son extremos, por lo que urge encontrar nuevas estrategias para su erradicación) (Engelbrecht *et al.*, 2017).

Hoy en día, el cáncer ha desplazado el foco de atención de las enfermedades infecciosas. El cáncer constituye un problema crítico en todo el mundo. Sin embargo, se necesita un marco conceptual adecuado para responder a preguntas como por qué se produce el cáncer, por qué resulta tan alto el riesgo de padecerlo en la vida, por qué ha aumentado el riesgo y por qué cuando, tras la introducción de un fármaco, se produce una regresión tumoral seguida, casi inevitablemente, de recurrencia y resistencia. El cáncer es un proceso darwiniano, algo que debe tenerse en cuenta a la hora de comprenderlo y controlarlo.

Aquí destacaremos algunos casos en los que el pensamiento evolutivo aporta una nueva perspectiva a la enfermedad. La teoría evolutiva proporciona un marco conceptual común para entender procesos que, de otro modo, no estarían relacionados, como la evolución de la virulencia de los patógenos, la resistencia de los patógenos y el cáncer a las terapias farmacológicas, y cómo la propia historia evolutiva del ser humano ha moldeado su susceptibilidad a la enfermedad.

*Enfermedades infecciosas: evolución de la virulencia y de las interacciones huésped-patógeno.* Uno de los mayores desafíos de la infectología consiste en comprender la variación en el daño causado por el parásito (virulencia) y su capacidad de transmisión (infectividad), así como en la susceptibilidad y resistencia del huésped a la infección. Las pruebas actuales demuestran que las interacciones y la dinámica intrahuésped e interhuésped dan forma al fenotipo de la infección y resultan clave en la evolución de nuevas estrategias para la gestión de la virulencia y la infectividad, así como para la resistencia del huésped (Ewald, 1994). Dentro del huésped, la supervivencia del parásito depende principalmente de la evasión de su sistema inmunitario, y existen múltiples formas de hacerlo: estrategias para penetrar y multiplicarse dentro de diferentes células y tejidos, variar la superficie y modular las proteínas, o suprimir la respuesta inmunitaria. Para optimizar la transmisión entre huéspedes, los estadios transmisibles del parásito deben estar presentes en el momento o tejido adecuado. Para ello, hace falta un control muy preciso del ciclo celular. Sin embargo, los parásitos y los huéspedes experimentan una enorme variación en sus entornos externos e internos que repercuten en su aptitud, por lo que los ajustes en el fenotipo deben producirse con rapidez. La solución está en dotar al sistema de plasticidad fenotípica: la capacidad de un genotipo para expresar diferentes fenotipos en un entorno heterogéneo (Taylor *et al.*, 2006; Gómez-Díaz *et al.*, 2012; Westneat *et al.*, 2019). En este contexto, entender mejor los mecanismos moleculares que subyacen a la variación fenotípica en la virulencia o transmisión del parásito y la susceptibilidad o resistencia del huésped resulta clave para evitar la (re)aparición de enfermedades infecciosas y el fracaso de las intervenciones de erradicación existentes (Louhi *et al.*, 2012).

*El papel que desempeña la variación antigénica en los patógenos.* Los virus, las bacterias, los hongos y los protozoos afrontan desafíos similares al infectar a un huésped susceptible. Deben evitar la eliminación mecánica, adherirse con éxito a su tejido o célula preferida, evitar que el sistema inmunitario los

reconozca y evitar su eliminación por parte de diversos componentes del sistema inmunitario. Para sobrevivir y desencadenar las infecciones y, de este modo, asegurar la transmisión, han desarrollado estrategias similares. Una de estas estrategias es la variación antigénica: la capacidad de un patógeno de alterar las proteínas de la superficie o los hidratos de carbono, lo que evita la respuesta inmunitaria. El término abarca tanto la *variación de fase*, que constituye la activación o desactivación de un gen, y la verdadera *variación antigénica*, que es la expresión de formas alternativas de una proteína.

Se han descrito diversos mecanismos genéticos y epigenéticos que subyacen a la capacidad de producir fenotipos variables. Los mecanismos genéticos incluyen la mutación y la recombinación, e implican un cambio en la secuencia genómica de un gen codificador de antígenos, lo que da lugar a cambios en la estructura secundaria de la proteína o de sus elementos reguladores, que alteran el nivel de expresión. Estos cambios se producen en tasas relativamente bajas (excepto en los virus de ARN) y son irreversibles y heredables. Los mecanismos epigenéticos implican, por el contrario, cambios en la expresión genética que son independientes o se producen sin alterar la secuencia genómica primaria. Estos cambios son estables y potencialmente heredables. A diferencia de los eventos genéticos, pueden ocurrir en una generación y son reversibles e inducibles por estímulos internos o externos.

La variación de fase se describió por primera vez en bacterias como el fenómeno de cambio entre dos fenotipos alternativos entre las células de una población clonal. Un mecanismo habitual consiste en la regulación transcripcional, que puede ser de tipo genético o epigenético. *Borrelia* spp., *Salmonella* spp., *Treponema pallidum* son ejemplos de variación de fase que dependen de la recombinación homóloga del ADN (Vink *et al.*, 2012). También se ha descrito la regulación transcripcional epigenética, por ejemplo, en *Escherichia coli* donde la expresión de los pili, que está bajo el control del operón pap, está regulada por la metilación del ADN dependiente del entorno (Braaten *et al.*, 1991). Del mismo modo, en *Candida glabrata* la variación de fase del gen *Epa6* se induce como respuesta directa al nivel de NAD<sup>+</sup> encontrado en la orina (Mundy y Cormack, 2009). La variación de fase también puede ser el resultado de la regulación de la traducción a través de una variedad de mecanismos que incluyen el desajuste de la cadena de deslizamiento, la disociación temprana del ribosoma y la inestabilidad del ARNm.

Algunos microbios han desarrollado una estrategia más compleja de variación antigénica que se basa en grandes familias de genes multicopia e

hipervariables, en las que cada copia individual del gen codifica una proteína de superficie antigénicamente distinta. En *Plasmodium* spp. y *Trypanosoma* spp., la variación antigénica implica el cambio transcripcional entre los miembros de la *var* (~60 genes) y el *vsg* (~1000 genes), lo que hace que aparezcan parásitos con diferentes características antigénicas y fenotípicas en diferentes momentos dentro de una población. El cambio se produce bajo control genético, a través de la recombinación homóloga, y epigenético, en el que intervienen diferentes mecanismos: la propagación de la heterocromatina y la acetilación de las histonas, la accesibilidad de la cromatina y el posicionamiento nuclear, y factores de transcripción específicos (Maree y Patterton, 2015; Gómez-Díaz *et al.*, 2017; Ruiz *et al.*, 2018). En las bacterias, la variación antigénica a través de la recombinación homóloga se ha descrito para *Borrelia* spp. y *Neisseria* spp., y en los hongos para *Candida* spp. (Cahoon y Seifert, 2011).

*La genética y la epigenética de la susceptibilidad o resistencia a la infección.* El resultado de la interacción entre los parásitos y sus huéspedes no solo depende de la variabilidad y la capacidad de evolución del patógeno, sino también de la variabilidad espacial y temporal de los genes de susceptibilidad del huésped.

La base de muchas enfermedades infecciosas está asociada a la variación de los genomas de los huéspedes y, en este caso, se puede heredar ese rasgo de susceptibilidad a la enfermedad. Un ejemplo es el rasgo de anemia falciforme, que se asocia a una menor susceptibilidad a la malaria. Su prevalencia en la población se considera el resultado de la presión selectiva que los parásitos de la malaria han ejercido sobre las poblaciones humanas a lo largo del tiempo evolutivo (Karlsson *et al.*, 2014). Los estudios de asociación del genoma completo (GWAS, por sus siglas en inglés) se han utilizado ampliamente para caracterizar los marcadores genéticos asociados al riesgo de enfermedad, informar sobre su progresión y, por tanto, ayudar a elegir el tratamiento más adecuado (Craig, 2008). Como resultado de estos estudios, se ha asociado la variación del antígeno leucocitario humano (ALH) a la susceptibilidad o la resistencia a la malaria, la tuberculosis, la lepra, el SIDA y la persistencia de la hepatitis viral. La variación del promotor del factor de necrosis tumoral  $\alpha$ , así como las variantes alélicas del receptor de la vitamina D, parecen estar asociadas a una susceptibilidad diferencial a varias enfermedades infecciosas (Cooke y Hill, 2001). En el contexto de la actual pandemia de la COVID-19, se están llevando a cabo investigaciones para desvelar las bases genéticas de la susceptibilidad al SARS-CoV-2 que pueden explicar las diferencias en los

resultados clínicos y entre países. Las variaciones en el *ACE2*, *TMPRSS2* y los genes del ALH se han propuesto como probables factores de riesgo (Sungnak *et al.*, 2020). En esta y en futuras pandemias o epidemias, los GWAS y la genómica comparativa resultarán fundamentales para conocer las interacciones genéticas entre el huésped y el patógeno, así como un requisito previo para desarrollar vacunas y terapias eficaces.

Asimismo, hay cada vez más pruebas de que las asociaciones genotipo-genotipo entre el huésped y el patógeno son relevantes en diferentes aspectos de la infección. Por ejemplo, la transmisibilidad en la tuberculosis parece depender de combinaciones específicas de huésped y antecedentes genéticos (Gagneux *et al.*, 2006). Lo mismo ocurre con los resultados clínicos en *Helicobacter pylori*, donde el cáncer gástrico resulta más probable en ciertas combinaciones de antecedentes genéticos humanos y bacterianos (Kodaman *et al.*, 2014). En ambos casos, estas diferencias están relacionadas con la alteración de la coevolución a largo plazo de combinaciones específicas de genotipos humanos y bacterianos, lo que se denomina *asociaciones simpáticas*. Algunos de los loci implicados en el patógeno ahora pueden explorarse mediante enfoques de GWAS, dado el gran número de secuencias de patógenos disponibles (Falush, 2016). Sin embargo, para identificar LOS loci relevantes que puedan informar sobre el desarrollo de vacunas de terapias dirigidas al huésped, los GWAS deben evolucionar hacia la comparación de genoma a genoma donde se incorpore tanto la variación del huésped como la de la bacteria. Este enfoque ha comenzado a aplicarse en los virus (Fellay y Pedergnana, 2019) y, probablemente, será una piedra angular para descifrar los loci de interacción entre el huésped y el patógeno para una serie de enfermedades infecciosas.

Además del componente genético, los factores ambientales también influyen en la etiología de la enfermedad, el llamado *triángulo de la enfermedad* (Scholthof, 2007). La interacción del individuo con el entorno, y a lo largo del tiempo, da lugar a una gran variabilidad en las formas en que se presenta la enfermedad, el fenotipo de la infección, con un impacto en las intervenciones de salud pública. La tuberculosis constituye un clásico ejemplo de enfermedad de la pobreza en la que los determinantes socioeconómicos desempeñan un papel importante. La inmunosenescencia en los seres humanos se caracteriza por una disminución de la respuesta innata y adaptativa, lo que aumenta el riesgo de enfermedades infecciosas. También existe una variabilidad estacional en la susceptibilidad a las enfermedades autoinmunes e infecciosas (Castle, 2000), probablemente relacionada con la variación estacional de la

exposición al sol y el color de la piel, con consecuencias en micronutrientes como la vitamina D.

Cuando se habla de la variabilidad de la enfermedad impulsada por el medioambiente, una consideración importante son los mecanismos moleculares subyacentes, ya que esta variabilidad, *a priori*, no es genética. Se ha propuesto que los mecanismos epigenéticos vinculan los cambios ambientales y el fenotipo mediante la modulación de la expresión genética (Klemm *et al.*, 2019). Comprender el papel que desempeña la epigenética en las enfermedades se ha convertido en una prioridad de la investigación biomédica actual. De hecho, los cambios en el estado de metilación y la acetilación o metilación de histonas de múltiples genes implicados en las vías de señalización inmunológica y de ataque inmunológico se han asociado a la susceptibilidad y la progresión de la enfermedad para los patógenos vegetales virales y bacterianos (Wang *et al.*, 2019), la malaria, el SIDA, la hepatitis B y la tuberculosis (Grabiec y Potempa, 2018), aunque los estudios siguen siendo escasos. El estudio de cómo se pueden modificar o revertir estos factores de riesgo epigenéticos se conoce como *terapia epigenética* (Dumitrescu, 2018). El conocimiento de cuál es el impacto de estos cambios a largo plazo (heredabilidad), su causalidad y las aplicaciones clínicas de las terapias epigenéticas constituye una desafiante investigación en este campo.

*Enfermedades infecciosas emergentes.* La devastadora pandemia del SARS-CoV-2 a la que se enfrenta actualmente el mundo pone de manifiesto el tremendo riesgo y la importancia que las enfermedades infecciosas emergentes y reemergentes suponen para la humanidad. La rapidísima propagación del SARS-CoV-2, su alta capacidad de contagio y el número relativamente alto de casos graves de la COVID-19 que requieren cuidados intensivos están sofocando los sistemas sanitarios y, a corto plazo, la actividad industrial y la economía a escala nacional y regional. Esta situación resulta aún más amenazante para los países en desarrollo que carecen de infraestructuras y recursos sanitarios básicos y en los que las pandemias coincidirán con otras como las de la tuberculosis, la malaria o el VIH-1. La pandemia del SARS-CoV-2 representa el último y paradigmático episodio de una serie de epidemias humanas causadas por virus y bacterias emergentes que se han producido en las últimas décadas. Entre ellas, la enfermedad de Lyme, la enfermedad enterohemorrágica *E. coli* O157:H7, el virus chikungunya, el virus del dengue, el hantavirus, el VIH-1, el virus del Nilo occidental, el virus del Zika y los coronavirus SARS-CoV y MERS-CoV, entre muchos otros, incluida la epidemia de ébola,

aún activa en África central. En la mayoría de estos brotes, el virus emergente aparece como resultado de un salto de las especies silvestres o domésticas a los humanos que ha llevado a acuñar la expresión *Una Salud* para hacer frente a este nuevo escenario, que es concomitante con el calentamiento global y otras agresiones del ser humano a la naturaleza y a la fauna (Zinsstag *et al.*, 2018). Siendo coherentes con el concepto *Una Salud*, no solo se deben tener en cuenta los virus que afectan directamente a los seres humanos, sino también cualquier otro patógeno que afecte al bienestar y la producción animales, así como a la salud vegetal. Tres ejemplos que ilustran esta enrevesada red de interconexiones, cuya complejidad apenas se vislumbra, son la observación de que el SARS-CoV-2 puede transmitirse del ser humano a los tigres y a los gatos domésticos, la actuación de los animales como reserva de la resistencia a los antibióticos vinculada a la cría de animales y la existencia de un cáncer no viral transmisible por clones que afecta a los demonios de Tasmania.

*Evolución de las resistencias.* En 2016, las infecciones por bacterias resistentes a los antibióticos causaron 700 000 muertes y pérdidas económicas devastadoras. En 2050, las muertes por infecciones resistentes a los antibióticos ascenderán a 10 millones, de modo que superará al cáncer como principal causa de mortalidad en el mundo. Además, las resistencias a virus, hongos y parásitos también suponen un número importante de empeoramiento de los resultados clínicos y de muertes. La incidencia y el tipo de infecciones resistentes son muy diferentes entre los países desarrollados y los países en vías de desarrollo. Mientras que en los países de ingresos altos los principales problemas están relacionados con la aparición de superbacterias en entornos nosocomiales, en los países de ingresos bajos y medios (LMIC, por sus siglas en inglés) la transmisión en la comunidad constituye un importante desafío para la salud pública. Por ejemplo, cada año mueren 250 000 personas a causa de la tuberculosis multi-resistente y se prevén tres millones para 2050. Es importante destacar que los datos genómicos y epidemiológicos, combinados con el poder del análisis evolutivo, están permitiendo comprender, e incluso anticipar, cómo surge y se propaga la resistencia con una resolución que antes no estaba disponible.

La resistencia a cualquier molécula se desarrolla mediante los mecanismos genéticos básicos de mutación y recombinación (que crean y combinan determinantes de resistencia). En el caso de los patógenos bacterianos, también hay que tener en cuenta su capacidad de transferencia horizontal de determinantes de resistencia (entre miembros de la misma especie o de diferentes especies, incluso filogenéticamente distantes) (MacLean y San Millán, 2019).

Conocer estos mecanismos y la forma en que se ven afectados por los desafíos farmacológicos resulta fundamental para comprender el proceso básico de adquisición de la resistencia y el desarrollo de nuevos objetivos tanto para los nuevos fármacos como para las pruebas de diagnóstico. En consecuencia, se espera que, al reducir el impacto de estos mecanismos básicos, se minimice el desarrollo y la diseminación de la resistencia. El uso generalizado de los medicamentos cambió la práctica de la medicina, lo que permitió el desarrollo de las prácticas médicas modernas. Sin embargo, la aparición y la propagación de resistencias a los fármacos en las últimas décadas están amenazando la salud mundial.

Muchas resistencias se adquieren por transferencia horizontal de genes resistentes. En muchos casos, incluidos los de las bacterias, los hongos y las células humanas, la resistencia a los fármacos se genera por mutaciones que aumentan el eflujo del fármaco o disminuyen la unión a la diana del fármaco, la importación de este o la modificación del profármaco. Las mutaciones de resistencia o sus efectos se pueden mejorar mediante la recombinación como, por ejemplo, la ampliación del espectro de enzimas de resistencia. Por lo tanto, las tasas de hipermutación e hiperrecombinación serán beneficiosas para volverse más resistentes y aptos. Se han encontrado clones hipermutadores en todos los ámbitos de la vida, como bacterias, hongos, parásitos y células cancerosas malignas (Blázquez, 2003). Los virus tienen, por lo general, una elevada tasa de mutación, por lo que se comportan efectivamente como hipermutadores. Además, algunos agentes antifecciosos no solo actúan como selectores de alelos resistentes o hipervariables, sino que pueden provocar la inestabilidad genética del patógeno y, por lo tanto, influir en la evolución y la propagación de los determinantes de resistencia de diferentes maneras, entre las que se incluyen la potenciación de la mutación, la recombinación y la transferencia génica horizontal. Además, la mayoría de las políticas de salud pública se han basado en el supuesto de que la resistencia conlleva un coste de adecuación en ausencia del fármaco, por lo que la mayoría de las acciones de salud pública para una serie de enfermedades infecciosas se destinan al tratamiento individual. Sin embargo, en *Mycobacterium tuberculosis* hay ahora pruebas claras que vinculan la evolución experimental y los datos clínicos de que el coste de aptitud puede compensarse haciendo que las cepas resistentes sean tan transmisibles como las susceptibles (Comas *et al.*, 2011). Los estudios de epidemiología molecular también están caracterizando la propagación de la resistencia en los entornos nosocomiales y comunitarios para una serie de patógenos y determinantes de resistencia.

Sin embargo, actualmente estamos lejos de comprender los complejos factores y sus interacciones que influyen en el nivel de resistencia alcanzado. Estos factores incluyen, entre otros, las interacciones del patógeno con los fármacos y los huéspedes, la variabilidad genética intrínseca e inducida del patógeno, la corresponsabilidad a fármacos no relacionados, la aparición de clones resistentes con éxito que no suponen un coste de aptitud (debido a la presencia de otras mutaciones epistáticas) y las tasas de transmisión de los patógenos entre los seres humanos, los animales y el medioambiente. También son importantes otros factores públicos, como las tasas de vacunación, los sistemas de atención sanitaria y la densidad y movilidad de la población. Por lo tanto, la resistencia supone un problema multifactorial que se debe abordar desde diferentes disciplinas, como la medicina, la genética, la microbiología, la epidemiología, la química, la ecología y la sociología, pero siempre desde la perspectiva de la evolución, ya que finalmente la resistencia es la consecuencia inevitable del proceso evolutivo de la vida. No tener en cuenta esta premisa conllevará una lucha encarnizada contra la evolución microbiana, ya que, hasta ahora, ningún fármaco se ha librado de la resistencia.

Por lo tanto, urge actuar para evitar volver a la era preantibiótica. Entre las medidas, una mejor comprensión de los mecanismos evolutivos y los impulsores de la resistencia, el desarrollo de nuevos fármacos a prueba de evolución y la aplicación de métodos para aumentar la eficacia del tratamiento basados en conceptos evolutivos resultan fundamentales para hacer frente a este problema mundial. Los virus, bacterias, hongos, parásitos y células humanas siempre encontrarán una solución para sobrevivir a la presión de los medicamentos. El problema puede mejorarse o evitar que se agrave, pero la experiencia demuestra que no es posible encontrar un medicamento o dolencia para el que la vida microbiana no encuentre una solución.

*El cáncer como proceso evolutivo.* El principal problema del cáncer estriba en la heterogeneidad clonal, es decir, la presencia y aparición continua de diferentes grupos o clones de células tumorales, con nuevas características genéticas y fenotípicas, como la capacidad de metastatizar y resistir a los tratamientos. La heterogeneidad clonal de los tumores proviene de la evolución clonal, en la que algunas células dan lugar a otras con nuevas características fenotípicas (Greaves, 2015).

La evolución se caracteriza por largos períodos de estabilidad interrumpidos por cambios explosivos puntuales (Gould y Eldridge, 1977). Este último sería también el caso del cáncer: explosiones de expansión clonal secundarias a la

acción de mutaciones de alto impacto seleccionadas por cambios de presión ambiental. Mientras tanto, habría múltiples clones coexistiendo y compitiendo entre sí en paralelo (Greaves, 2015).

La unidad de selección evolutiva, responsable del inicio del tumor y de la formación de nuevos clones, son las células madre con capacidad de autorrenovación. Por lo tanto, son esenciales para el mantenimiento de los tejidos normales y cancerosos. Cualquier célula tumoral puede adquirir mutaciones, pero solo las células madre podrían propagarse. Las células madre serían las encargadas de generar nuevos clones y de las metástasis, ya que también se requeriría la capacidad de autorrenovación para mantener un clon a distancia, como en el tumor primario, e incluso la resistencia a la quimioterapia (Greaves, 2013). De todas estas mutaciones, solo unas pocas inducen funciones esenciales para el desarrollo del tumor (mutaciones conductoras) y generan nuevos clones. Así, solo el 3 % de las mutaciones de un tumor serían conductoras, el resto serían mutaciones pasajeras no sometidas a selección. Las mutaciones conductoras dependen del contexto. Pueden resultar ventajosas en algunos entornos tumorales pero no en otros. Además, las ventajas de la mutación pueden ser epistáticas con otras mutaciones. Las mutaciones conductoras pueden ser diferentes en distintos clones, por lo que la terapia dirigida puede no ser óptima para tratar todo el tumor.

La complejidad del cáncer va más allá de las células tumorales locales, ya que el cáncer no constituye un proceso autónomo de las células, sino un tejido que crece en un contexto sistémico. Para crecer, las células cancerosas necesitan interactuar con el estroma y con compartimentos lejanos relacionados con la fisiología del organismo, necesitan nutrientes, oxígeno, hormonas como la insulina, etc. Por lo tanto, las presiones de selección para la aparición de nuevos clones tumorales provienen no solo del compartimento del parénquima tumoral, sino también de los otros compartimentos, estromal y a larga distancia. En estos compartimentos, varias estructuras y funciones se comportan como fenotipos intermedios o subfenotipos de cáncer a nivel sistémico, tisular, celular y molecular. Por último, está el nivel genético, con las mutaciones en las células tumorales, y el fondo genético del organismo que regula todos los niveles de subfenotipos. Los subfenotipos tienen interacciones complejas y retroalimentaciones negativas que estabilizan el cáncer como sistema y lo hacen resistente a nuevos desequilibrios. Toda esta organización da lugar a múltiples conexiones que se pueden estudiar desde la perspectiva de la biología de sistemas. Así, el cáncer constituye un sistema evolutivo y adaptativo que

suele evolucionar en la dirección de una robustez más significativa, es decir, de una mayor capacidad de soportar desafíos o estrés, como la evasión de terapias.

*La evolución humana, las compensaciones evolutivas y la incidencia del cáncer.* Cuando aparecieron los organismos multicelulares hace 600 000 millones de años, surgieron mecanismos para mantener la integridad de los tejidos. De este modo, en esa época, aparecieron los primeros genes supresores de tumores. Los mecanismos de reparación del ADN son una respuesta adaptativa para prevenir el aumento del riesgo de cáncer. Son un rasgo heredado y, por tanto, se han generado y mejorado a lo largo de la evolución. La capacidad de reproducción en los seres humanos disminuye notablemente a partir de los 55 años. Todas las adaptaciones fisiológicas están diseñadas para actuar eficazmente antes de la vejez. Llegar a la vejez es un hecho relativamente reciente, consecuencia de las mejoras médicas y de la salud pública. Sin embargo, los genes los seleccionó una presión que ya no existe en los estilos de vida de hoy en día. Por lo tanto, habría una disociación entre los genes y el entorno real, ya que los primeros estarían preparados principalmente para responder al entorno paleolítico con una actividad física considerable, una alimentación intermitente, una resistencia a las hambrunas, al frío, etc. (Tooby y Cosmides, 1990). Entre estas disociaciones gen-medioambiente también estarían los sistemas de reparación del ADN. Con la llegada del envejecimiento, la capacidad de reparación del ADN empieza a fallar y se produce un aumento progresivo de la incidencia del cáncer con la edad.

Otros cambios en la incidencia del cáncer provienen de los cambios en el estilo de vida en comparación con el paleolítico. Es importante destacar que este hecho ofrece la posibilidad de prevención mediante estrategias que intentan imitar esos procesos ambientales. La visión darwiniana del cáncer ofrece una mejor comprensión de la enfermedad, así como nuevas alternativas de prevención y enfoques terapéuticos que promoverán una medicina más personalizada. Así surge la medicina evolutiva, que aborda la cuestión de cómo el pasado evolutivo ha configurado la vulnerabilidad a enfermedades como el cáncer, la diabetes tipo II, la artrosis, etc., tanto como especie como en lo que respecta a la diferente susceptibilidad entre individuos (Williams y Nesse, 1991).

*Malas adaptaciones, limitaciones evolutivas y enfermedades.* Hay muchos ejemplos que ilustran el modo en que la historia evolutiva de los seres humanos determina la prevalencia de ciertas enfermedades en las poblaciones humanas.

A continuación, y a título ilustrativo, se expondrán tres casos bien estudiados. En primer lugar, las mujeres de culturas sin anticoncepción y con intervalos de nacimiento normales seguidos de unos tres años de lactancia materna tenían unos 100 ciclos menstruales durante toda su vida. Sin embargo, en el mundo occidental moderno, las mujeres pueden tener hasta 400 ciclos. En consecuencia, experimentan un número mucho mayor de divisiones celulares, lo que las sitúa en un mayor riesgo de cáncer de mama (Eaton *et al.*, 1994). Las mujeres que dan a luz jóvenes y que pasan la mayor parte de su periodo reproductivo embarazadas o en amenorrea lactacional tienen tasas de cáncer de mama significativamente más bajas (Strassmann, 1999). La menopausia no era un fenómeno relevante para los primates y los ancestros *Homo* debido a su menor esperanza de vida. Ahora bien, se trata de un problema real: la reducción de los niveles de estrógeno provoca el desarrollo de la osteoporosis y el debilitamiento de los huesos. Se deberían diseñar mejores tratamientos anticonceptivos teniendo en cuenta estas compensaciones evolutivas. De hecho, hay varias estrategias en marcha que quieren imitar el efecto protector de los embarazos tempranos y repetidos. En cuanto a la existencia de formas alélicas que favorecen la aparición del cáncer de mama, la prevalencia relativamente alta de mutaciones en el gen *BRCA1* se ha asociado a un aumento de la fertilidad en sus portadoras, causa de un fenómeno de pleiotropía antagónica. Probablemente, este hecho también ocurriera en otros tumores (Smith *et al.*, 2013).

Los seres humanos evolucionaron en un entorno en el que las infecciones eran omnipresentes y, la mayor parte del tiempo, la mayoría de las personas eran portadoras de todo tipo de parásitos. Algunos de estos parásitos regulan a la baja las respuestas inmunitarias del huésped para mejorar su supervivencia. Al hacerlo, redujeron la susceptibilidad a las enfermedades autoinmunes al disminuir la producción total de anticuerpos. Los ambientes antisépticos mantienen a las sociedades libres de muchos parásitos, pero las enfermedades autoinmunes son mucho más comunes ahora. Por ejemplo, los niños gaboneses con esquistosomiasis tienen menos reacciones alérgicas a los ácaros del polvo; asimismo, la incidencia del asma es mucho menor en los adultos infectados por nematodos que viven en los LMIC (Wilson y Maizels, 2004). La selección puede tardar cientos de generaciones en equilibrar los mecanismos de detección de los sistemas inmunitarios con la limpieza de los entornos modernos.

Otro problema evolutivamente interesante estriba en el conocido conflicto entre madre e hijo. Mientras que las madres están igualmente interesadas en

el éxito de cada uno de sus hijos, el feto, por desgracia, tiene intereses evolutivos diferentes a los de su madre con respecto a sus hermanos. El feto manipula el estado fisiológico de la madre a través de la placenta. La preeclampsia y la diabetes de la madre son subproductos de este conflicto evolutivo (Trivers, 1974).

*Un cambio de paradigma: el enfoque «Una Salud».* El ser humano constituye probablemente una de las mayores fuerzas evolutivas del planeta en la actualidad (Palumbi, 2001). Una de las pruebas más dramáticas de la influencia del ser humano en los procesos evolutivos se encuentra en la rápida selección de la resistencia en los patógenos microbianos (Davies, 2007). Los efectos antropogénicos tienen un impacto significativo en todos los componentes de la biosfera. En el medioambiente se libera una cantidad considerable de agentes antinfeciosos de uso humano y veterinario. Estas moléculas no solo están seleccionando microorganismos resistentes, sino que es probable que afecten a la capacidad de evolución de las bacterias, lo que tiene consecuencias imprevisibles para el equilibrio de la biosfera. Debido a la necesidad de adaptarse rápidamente a los desafíos antinfeciosos, los microorganismos con alta capacidad de evolución podrían haberse seleccionado por un efecto de selección de segundo orden (Gillings y Stokes, 2012). Además, como se ha indicado anteriormente, algunos agentes antinfeciosos pueden aumentar directamente la evolucionabilidad microbiana (Blázquez *et al.*, 2018). En resumen, es probable que la presencia de agentes antinfeciosos a escala mundial esté cambiando el ritmo de la evolución microbiana.

*Desafíos éticos del enfoque Una Salud.* La mayoría de los aspectos mencionados tienen importantes implicaciones éticas relacionadas con tres áreas principales: dónde se investiga, cómo se hace y quién lo hace.

*¿Dónde?* Hoy más que nunca, entender el origen y la aparición de los organismos resulta clave para identificar los determinantes genéticos de los rasgos fenotípicos, como la susceptibilidad a las enfermedades en los seres humanos o la virulencia y la resistencia a los fármacos en los patógenos. En este sentido, hay un conjunto emergente de trabajos en la encrucijada entre la genómica, la arqueología y la sociología, en los que se investigan las raíces de las enfermedades y cómo han cambiado a lo largo del tiempo. Esto es posible gracias a la llegada de las tecnologías de secuenciación del ADN antiguo. Estas tecnologías han permitido a los científicos acceder a los genomas completos de huéspedes y microbios de restos humanos antiguos. Por ejemplo, los análisis de ADN humano antiguo han descubierto una plétora de *Homo* que coexistieron

y se cruzaron con el *Homo sapiens*. Asimismo, se ha podido recuperar la secuencia del genoma de patógenos procedentes de restos humanos que datan de hace cientos o miles de años, como *M. tuberculosis*, *Yersinia pestis*, *Vibrio cholerae*, el virus Influenza A H1N1, VIH-1 o *Plasmodium falciparum* (Bennett y Baker, 2019). Estas técnicas permiten comprender ahora las causas de plagas bien conocidas. La mayoría de estos patógenos siguen contribuyendo considerablemente a la morbilidad humana y constituyen importantes problemas sanitarios a escala mundial. Es importante destacar que estos extraordinarios avances en la genómica del ADN antiguo con potencial para informar sobre la salud humana solo son posibles gracias a equipos multidisciplinares de antropólogos y sociólogos capaces de identificar las causas probables de la muerte. Sin embargo, a veces, la carrera por el material de ADN antiguo ha dejado atrás a las comunidades y países donde se obtienen las muestras.

En general, siempre ha habido un sesgo en la representación de los LMIC en la investigación y especialmente en los temas de salud global. Por ejemplo, la integración de los LMIC resulta clave para comprender la aparición de patógenos globales con origen en esas regiones, ya que no se pueden ignorar las cuestiones culturales y sociales para entender el panorama general. Un reflejo de la cuestión constituye el escaso papel que los LMIC han desempeñado hasta ahora en la investigación genómica humana. La mayoría de los análisis de poblaciones humanas a nivel genético destinados a buscar, por ejemplo, variantes patógenas, se han basado en individuos del norte mundial, lo que también se ha traducido en herramientas, como las matrices de tipificación de SNP, muy sesgadas por la diversidad de estas poblaciones. Esto resulta paradójico, ya que se sabe que África constituye un foco de diversidad genética humana importante para entender no solo la evolución humana sino también los determinantes genéticos de la salud humana. Como en el caso del ADN antiguo, resulta necesario pasar a una investigación más ética y socialmente responsable. La clave será participar en la investigación y en la comunicación de los resultados a las comunidades afectadas. En el caso de la genómica humana, nuevas iniciativas como H3Africa pretenden corregir el sesgo potenciando a los países africanos en la genómica aplicada a la salud humana. No solo abarca la investigación dirigida por investigadores africanos sino también la creación de capacidades y tecnologías *in situ*.

*¿Cómo?* Uno de los avances científicos más importantes de la biología molecular reciente consiste en la aplicación de los enfoques de edición de genes CRISPR/Cas9 como una nueva estrategia para combatir las enfermedades

infecciosas y parasitarias (Soppe y Lebbink, 2017). Este avance plantea un importante debate ético y moral a la hora de liberar organismos modificados genéticamente en la naturaleza, en especial en el caso de los sistemas de genética dirigida que pretenden suprimir poblaciones naturales o incluso especies, por ejemplo, los mosquitos transmisores de enfermedades (Kyrou *et al.*, 2018). Se espera que los riesgos de bioseguridad sean bajos, ya que estos sistemas son muy específicos y la tecnología está bien desarrollada. No obstante, existen riesgos ecológicos. Los sistemas naturales son complejos y lo que funciona en condiciones confinadas de laboratorio puede no funcionar como se espera en la naturaleza. Es difícil predecir cómo reaccionará el ecosistema a esta introducción y cuáles serán los efectos secundarios a corto y largo plazo. Sin embargo, los riesgos sociales son muy elevados. Existe un movimiento mundial contra los organismos modificados genéticamente y una desconfianza generalizada en los expertos y las instituciones. El debate público y el conflicto en torno a esta tecnología son más importantes en los países en desarrollo. En estas regiones, los beneficios potenciales de curar enfermedades importantes como la malaria o el dengue se priorizan sobre los riesgos éticos/sociales y ecológicos. Resulta esencial que la comunidad científica examine adecuadamente los aspectos técnicos y la metodología, y que realice pruebas piloto sobre el terreno, procedimientos de contención y control, así como programas de participación de la comunidad.

*¿Quién?* Se trata de un problema multifacético que depende de la inversión en investigación que hacen los distintos países y, lo que resulta más preocupante, de las cuestiones de desigualdad de género. En un mundo ideal, las mujeres y los hombres tienen las mismas oportunidades de entrar y progresar en todas las disciplinas científicas sin sesgos ni prejuicios. Sin embargo, la realidad es que las mujeres representan, en el mejor de los casos, menos del 30 % de los puestos de investigación en todo el mundo. Y lo que es más importante, carecen de una masa crítica de representación en todas las áreas de CTIM (ciencia, tecnología, ingeniería y matemáticas), incluidos los campos relacionados con la salud mundial y las enfermedades infecciosas. Es decir, a pesar de que las mujeres científicas están liderando investigaciones pioneras en todo el mundo, siguen estando muy poco representadas en las citas, en el personal editorial de las revistas científicas, como ponentes invitadas en los congresos científicos y como miembros de los comités de toma de decisiones, y muchas más mujeres que hombres abandonarán la ciencia a medida que avancen en su carrera (English *et al.*, 2020). La misma situación se da en España según el último informe de la Comisión de Mujeres y Ciencia del CSIC, y el del

Ministerio de Ciencia e Innovación. Un ejemplo claro es la composición de la PTI Salud Global, donde las mujeres coordinan solo el 34 % de las subáreas. Del mismo modo, las mujeres solo lideran el 32 % de los desafíos de este libro. La solución al techo de cristal y al desequilibrio de género en la ciencia resulta compleja y requerirá una reforma de los sistemas educativo, social y científico. Sin embargo, ya es hora de que se produzca este cambio transformador y las instituciones deben tomar medidas urgentes.

## 2. IMPACTO EN EL PANORAMA DE LA CIENCIA BÁSICA Y EN LAS APLICACIONES POTENCIALES

*Un enfoque evolutivo para minimizar el impacto de los patógenos emergentes.* Para prevenir las enfermedades infecciosas emergentes, resulta esencial anticipar el rápido potencial de adaptación de estos patógenos mediante la caracterización exhaustiva de la diversidad genómica o epigenómica existente, así como desvelar los mecanismos de regulación de la variación antigénica.

Varias áreas de investigación que ya eran relevantes para la identificación, la comprensión y el control de las enfermedades emergentes se convierten ahora en cruciales para que las sociedades modernas se den cuenta y prevengan no solo el SARS-CoV-2, sino cualquier patógeno relacionado o no que surja tarde o temprano. Las principales áreas de investigación que se deben potenciar en este escenario, siempre bajo la consideración de promover enfoques multidisciplinares y polivalentes, así como la investigación aplicada y traslacional, son los siguientes:

- Estructura y biología moleculares de patógenos emergentes y reemergentes ya conocidos (viroides, virus, bacterias, parásitos, hongos), para obtener información relevante para su detección y control. Los estudios deben abordar los mecanismos que conducen a la aparición en la población natural de patógenos humanos nuevos o alterados con mayor virulencia, variaciones antigénicas, resistencia a los fármacos o transmisibilidad o infectividad modificadas.
- Factores ecológicos y medioambientales que influyen en la aparición y distribución de las enfermedades, como la influencia de los cambios

medioambientales naturales, los provocados directamente por el hombre o los inducidos por el clima en la aparición de patógenos; los efectos de las alteraciones en la densidad y distribución de la población de huéspedes o vectores en las enfermedades. De hecho, se han registrado casos humanos del virus del dengue en la costa mediterránea, y su mosquito vector está presente en zonas de España. Más preocupante aún resulta la corta cadena de transmisión de la fiebre hemorrágica de Crimea-Congo que tuvo lugar en Castilla y León hace unos años, donde la garrapata era el vector común en la fauna silvestre.

- Desarrollo de nuevas herramientas de diagnóstico basadas en la detección tanto de ácidos nucleicos como de anticuerpos específicos utilizando, en la medida de lo posible, enfoques versátiles, rápidos, automatizables y fácilmente adaptables para garantizar la detección y vigilancia de patógenos, como pruebas baratas y fáciles de usar en los puntos de atención, especialmente en los LMIC. Asimismo, se deben apoyar los estudios epidemiológicos de alta calidad y hacer especial hincapié en la optimización del uso de las bases de datos nacionales e internacionales, así como en el fomento de la incorporación de datos medioambientales y de la fauna silvestre.
- Desarrollo de la vigilancia genómica local y mundial en tiempo real de los patógenos asociados a los avances de la sanidad electrónica, ya que los datos genómicos permiten simultáneamente el diagnóstico, la vigilancia y la epidemiología. Uso de datos genómicos en tiempo real junto con los análisis evolutivos para informar sobre las intervenciones en escenarios comunitarios, nosocomiales y epidémicos.
- Identificación de antivirales y otros compuestos antimicrobianos y antiparasitarios mediante el uso de bibliotecas de cribado de compuestos ya conocidos, así como teniendo en cuenta nuevos compuestos en investigación y anticuerpos monoclonales. Además de los conocimientos básicos y aplicados resultantes de estas investigaciones, la disponibilidad de una amplia variedad de familias de fármacos será relevante para combatir los agentes infecciosos emergentes.
- Las resistencias virales y a otros patógenos ya no se pueden atribuir a la mala suerte. Según el informe O'Neill (amr-review.org), probablemente la evaluación científica más completa sobre el impacto presente y futuro de la resistencia a los antibióticos, se estima que en 2050 se perderán diez millones de vidas al año debido a una infección resistente a los antimicrobianos, lo que superaría al cáncer. Incluso hoy la cifra resulta asombrosa: 700 000 al año. Más de una cuarta parte de esas muertes

están y estarán relacionadas con formas de tuberculosis resistentes a múltiples fármacos, mientras que el resto se debe principalmente al aumento de superbacterias como el *Staphylococcus aureus* resistente a la meticilina (SARM). Además, la resistencia de los parásitos al tratamiento, como en el caso de la resistencia a la artemisinina en la malaria, contribuye de forma importante a la morbilidad, principalmente en los LMIC.

- En el caso de los virus, las propiedades de la estructura de los genomas de ARN, en forma de colectivos moleculares que interactúan entre sí de muchas maneras diferentes (cuasiespecies), y su dinámica evolutiva, permiten predecir la aparición de mutantes resistentes a un mismo fármaco. Una de las propiedades del colectivo es la memoria y los tratamientos con un solo fármaco solo pueden servir para agotarla como posibilidad de futuro. Este conocimiento debería guiar en el futuro no solo la forma de administrar los fármacos antivirales y seguir su efecto, sino también la búsqueda de nuevas estrategias terapéuticas basadas en este conocimiento. Por ejemplo, considerando las estructuras virales de ARN altamente conservadas como potenciales dianas farmacológicas o empleando mezclas de fármacos que lleven al virus a la extinción por acumulación de mutaciones.
- Desarrollo de nuevas vacunas profilácticas y terapéuticas versátiles y fácilmente adaptables, ya sea a base de virus inactivados, ARN, ADN, subunidades de virus o virus atenuados por ingeniería, así como una nueva tecnología de administración de vacunas, que han de prevenir y tratar enfermedades, con un impacto sustancial en la salud humana y animal. Se debe prestar atención a la evaluación de su eficacia y espectro antigénico, al riesgo de selección de mutantes resistentes y a la simplificación de la adaptación de las vacunas para hacer frente a las nuevas variantes de los virus.

*Diseño de vacunas y otras terapias a prueba de evolución.* La vacunación ha reducido en gran medida la carga de las enfermedades infecciosas. Sin embargo, las actuales vacunas autorizadas, cuya acción se basa casi exclusivamente en los anticuerpos, protegen contra los agentes patógenos con escasa variabilidad antigénica. Los patógenos antigénicamente variables representan, de hecho, la mayor carga de enfermedades infecciosas en la actualidad y su capacidad para escapar de la inmunidad natural del huésped socava todos los esfuerzos terapéuticos. La realidad es que el éxito de las vacunas contra estos patógenos ha sido limitado, con un 40-50 % de eficacia en el mejor de los casos, una protección que tiende a disminuir con el tiempo.

Ha habido dos estrategias principales: (i) patógenos vivos atenuados. Estas vacunas imitan la infección natural, pero de forma debilitada y no patógena. (ii) Toxinas derivadas, preparados de subunidades, cócteles de carbohidratos o vacunas conjugadas. Esta estrategia supone que las secuencias conservadas derivadas del patógeno son objetivos inmunológicos prometedores; sin embargo, en la mayoría de los casos, no han logrado inducir ni mantener la inmunidad protectora. La suposición de que todas las respuestas inmunitarias son beneficiosas para el huésped se está cuestionando en el caso de varias enfermedades infecciosas. La identificación del lugar donde reside la inmunidad protectora constituye un importante desafío en las enfermedades infecciosas. Un ejemplo claro y contraintuitivo está en *M. tuberculosis*, cuyos antígenos están muy conservados y no muestran ningún signo de evasión inmunitaria. La opinión que está surgiendo es que se necesitan algunas respuestas inmunitarias para la inmunopatología necesaria para la transmisión de los bacilos (Comas *et al.*, 2010).

Un problema adicional para la vacunación y el desarrollo de fármacos en patógenos de rápida evolución estriba en la diversidad y variabilidad de las cepas y fenotipos en las infecciones naturales, que en la mayoría de los casos permanecen sin caracterizar. Se ha demostrado que esta constituye la causa de la eficacia limitada temporal y espacialmente de numerosas vacunas. Además, la mayoría de los estudios de vacunación se centra en cepas de referencia que se han mantenido y pasado en serie en los laboratorios durante décadas (mediante condiciones de medios no estandarizados) y que no son representativas del potencial de adaptación del patógeno. Otro problema se encuentra en el uso de modelos animales no apropiados para el desarrollo preclínico de vacunas (Gerds *et al.*, 2015).

Los desafíos consisten en encontrar nuevas estrategias dirigidas específicamente a la variabilidad genética o antigénica, y que sean representativas del proceso natural de diversidad o infección. Una de las estrategias consiste en diseñar inmunógenos específicos para la enfermedad, la población y el paciente que reflejen o incorporen los complejos y rápidamente cambiantes paisajes de epítomos o antígenos de las dianas de las vacunas. El fundamento sostiene que la activación de diversos grupos de células inmunes B y T requiere que los inmunógenos de la vacuna lleven una diversidad antigénica que refleje las infecciones naturales (Servín-Blanco *et al.*, 2016). Otro enfoque podría consistir en el desarrollo de moléculas que interfieran con las proteínas o secuencias reguladoras que controlan la expresión de la variante con el fin de mejorar la eficacia de la vacuna. En comparación con los objetivos tradicionales de las vacunas, estos reguladores maestros son probablemente más robustos y menos mutagénicos o recombinantes.

*Presiones ambientales y oportunidades oncoterapéuticas.* El cáncer, como sistema adaptativo, adquiere varias funciones que le permiten sobrevivir. Estas funciones coinciden con las denominadas señas de identidad del cáncer (Hannahan y Weinberg, 2011), y todas ellas favorecen, directa o indirectamente, la expansión clonal. Normalmente, las estrategias de tratamiento atacan algunas de estas características, como la proliferación, la angiogénesis o la inestabilidad genómica. La visión del cáncer como un proceso evolutivo abre la posibilidad de intentar modificar las presiones de selección medioambiental que determinan la evolución de los tumores. Muchas de estas presiones de selección son desconocidas, como las que empujan a las células cancerosas a hacer metástasis. Se ha relacionado con la hipoxia tumoral, pero se necesitan más estudios para aclarar este punto. Algunas oportunidades terapéuticas serían las siguientes: (i) La presión selectiva que determina la resistencia a la quimioterapia es la propia quimioterapia. Esta presión no resulta fácil de eliminar. Sin embargo, debe ir acompañada de un mejor conocimiento de las redes de señalización y de los genes que interactúan en un contexto que convierte la mutación en motor. Conocer las otras mutaciones que participan con el comportamiento de la mutación conductora (epistasis) y dirigirse a ellas pudiera suponer una excelente oportunidad terapéutica. Se trata del concepto genético de letalidad sintética, en el que la combinación de mutaciones en múltiples genes provoca la muerte celular, y proporciona un marco para diseñar nuevos enfoques terapéuticos contra el cáncer. (ii) La dinámica de las células madre tumorales que salen del ciclo celular en condiciones de hipoxia aumenta su resistencia a la quimioterapia. Sería factible alterar esta dinámica y volver a introducir las células madre tumorales en el ciclo celular de manera específica para evitar su resistencia a la quimioterapia. (iii) El paradigma terapéutico que modifica la selección de la presión del entorno tumoral se lleva a cabo con las nuevas estrategias de inmunoterapia con un alto impacto en la terapia del cáncer. El sistema inmunitario se infiltra en los tumores, como en los tejidos cicatrizados, lo que activa el estroma, y permite la nueva angiogénesis y, por tanto, la llegada de oxígeno y nutrientes y el crecimiento de las células tumorales (Coussens y Werb, 2002). El sistema inmunitario también elimina las células tumorales que son antigénicamente muy diferentes de las normales (inmunoección). Así, existe una presión selectiva para que el sistema inmunitario tolere las células tumorales antigénicamente. De esta forma, se explica por qué el cáncer puede surgir en condiciones de inmunocompetencia. Los nuevos tratamientos de inmunoterapia, y especialmente con anticuerpos anti-PD1/PDL1, disminuyen la tolerancia del sistema inmunitario a los tejidos tumorales y son rechazados (Havel *et al.*, 2019).

### 3. PUNTOS CLAVE DEL DESAFÍO

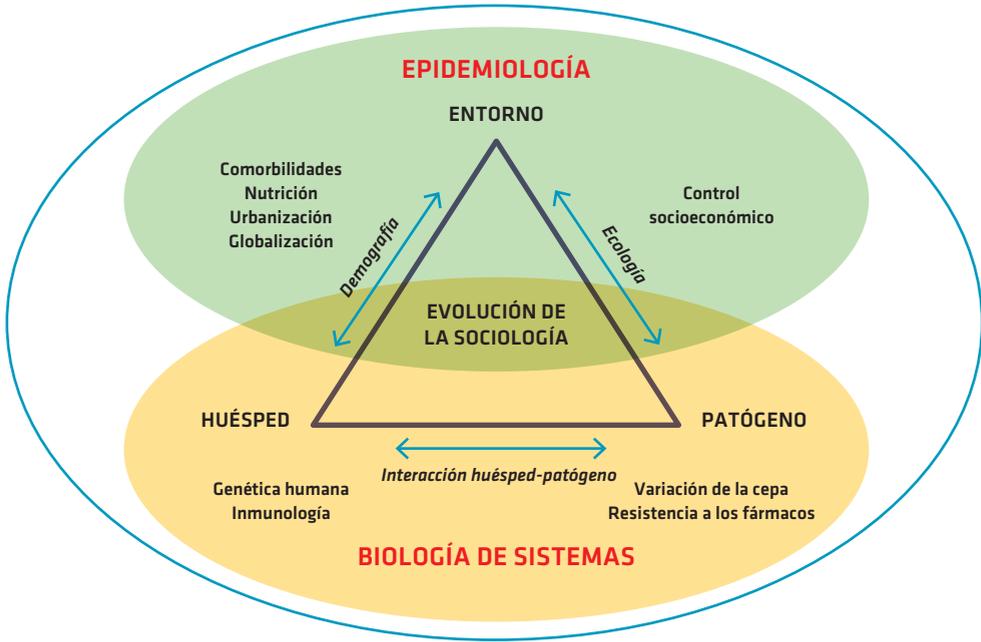
- Identificación de los determinantes patógenos de la virulencia y la propagación y el equilibrio entre ambos.
- Integración de la investigación experimental sobre la evolución de los patógenos con las observaciones clínicas y epidemiológicas.
- Identificación y caracterización de los mecanismos y vías genéticas básicas que rigen la evolución y la diseminación de la resistencia. Refuerzo del conocimiento de las funciones de las moléculas antinfeciosas en la biosfera y su efecto en la evolución microbiana y el desarrollo de la resistencia.
- Búsqueda de mecanismos para evitar el vertido de agentes antinfeciosos al medioambiente o inactivarlos para prevenir sus efectos en la evolución de la resistencia.
- Desarrollo de estrategias a prueba de evolución contra los patógenos microbianos, incluida la búsqueda de fármacos que interfieran con la variación genética y la transferencia horizontal.
- Consolidación de lo que ya se conoce como medicina evolutiva o medicina darwiniana.
- Metagenómica y metavirómica ambientales para obtener una imagen realista de los patógenos presentes en diversos ecosistemas (por ejemplo, aguas urbanas, dulces y costeras, agroecosistemas, áreas silvestres).
- Impulso de agendas de investigación que integren la investigación básica, aplicada y clínica.
- Incorporación de los avances más recientes de las técnicas *ómicas* (en particular, la genómica unicelular) para comprender la interacción entre los patógenos y las células del huésped en diferentes niveles de complejidad biológica (desde las células hasta las poblaciones).
- Integración de la modelización epidemiológica y las herramientas de análisis evolutivo, incluida la epidemiología genómica en tiempo real, para caracterizar los parámetros epidémicos básicos de los patógenos.
- Desarrollo de nuevos métodos matemáticos y estadísticos para integrar la información multinivel en los modelos de predicción.
- Incorporación de enfoques computacionales de biología de sistemas para comprender mejor la relación huésped-patógeno y la evolución del cáncer.
- Atracción de los inmunólogos clínicos a esta área de investigación.
- Incorporación de los entomólogos que trabajan en los vectores de transmisión de los patógenos zoonóticos.

## DESAFÍO 7 | BIBLIOGRAFÍA

- Bennett, R.J., Baker, K.S. (2019).** Looking backward to move forward: The utility of sequencing historical bacterial genomes. *Journal of Clinical Microbiology* 57, e00100-19.
- Blázquez, J. (2003).** Hypermutation as a factor contributing to the acquisition of antimicrobial resistance. *Clinical Infectious Diseases* 37, 1201-1209.
- Blázquez, J., Rodríguez-Beltrán, J., Matic, I. (2018).** Antibiotic-induced genetic variation: how it arises and how it can be prevented. *Annual Reviews Microbiology* 8, 209-230.
- Braaten, B.A. et al. (1991).** Evidence for methylation-blocking factor (mbf) locus involved in pap pilus expression and phase variation in *Escherichia coli*. *J. Bacteriology* 173, 1789-1800.
- Cahoon, L.A., Seifert, H.S. (2011).** Focusing homologous recombination: pilin antigenic variation in the pathogenic *Neisseria*. *Molecular Microbiology* 81, 1136-1143.
- Castle, S.C. (2000).** Clinical relevance of age-related immune dysfunction. *Clinical Infectious Diseases* 31, 578-585.
- Cooke, G.S., Hill, A.V.S. (2001).** Genetics of susceptibility to human infectious disease. *Nature Reviews Genetics* 2, 967-977.
- Coussens, L.M., Werb, Z. (2002).** Inflammation and cancer. *Nature* 420 860-867.
- Craig, J. (2008).** Complex diseases: research and applications. *Nature Education* 1, 184.
- Comas, I. et al. (2012).** Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. *Nature Genetics* 44, 106-110.
- Davies, J. (2007).** Microbes have the last word. *EMBO Reports* 8, 616-621.
- Domínguez-Andrés, A., Netea, M.G. (2019).** Impact of historic migrations and evolutionary processes on human immunity. *Trends in Immunology* 40, 1105-1119.
- Dumitrescu, R.G. (2018).** Early epigenetic markers for precision medicine. *Methods in Molecular Biology* 1856, 3-17.
- Eaton, S.B. et al. (1994).** Women's reproductive cancers in an evolutionary context. *Quarterly Reviews in Biology* 69, 353-367.
- Engebretsen, E., Heggen, K., Ottersen, O.P. (2017).** The sustainable development goals: ambiguities of accountability. *Lancet* 389, 365.
- English, E.D., Power, B.J., Gómez-Díaz, E. (2020).** Building parasitology communities to promote gender equality. *Trends in Parasitology*, en impresión.
- Falush, D. (2016).** Bacterial genomics: microbial GWAS coming of age. *Nature Microbiology* 26, 16059.
- Ewald, P.W. (1994).** *Evolution of infectious diseases*. Oxford, Oxford University Press.
- Gagneux, S. et al. (2006).** Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences of the USA* 103, 2869-2873.
- Gerdt, V. et al. (2015).** Large animal models for vaccine development and testing. *ILAR Journal* 56, 53-62.
- Gillings, M.R., Stokes, H.W. (2012).** Are humans increasing bacterial evolvability? *Trends in Ecology and Evolution* 27, 346-352.
- Gómez-Díaz, E. et al. (2012).** Epigenetics of host-pathogen interactions: the road ahead and the road behind. *PLoS Pathogens* 8, e1003007.
- Gómez-Díaz, E. et al. (2017).** Epigenetic regulation of *Plasmodium falciparum* clonally variant gene expression during development in *Anopheles gambiae*. *Scientific Reports* 7, 40655.
- Gould, S.J., Eldridge N. (1977).** Punctuated equilibria: the tempo and mode of evolution reconsidered. *Paleobiology* 3, 115-151.
- Grabiec, A.M., Potempa, J. (2018).** Epigenetic regulation in bacterial infections: targeting histone deacetylases. *Critical Reviews in Microbiology* 44, 336-250.
- Greaves, M. (2013).** Cancer stem cells as 'units of selection'. *Evolutionary Applications* 6 102-108.
- Greaves, M. (2015).** Evolutionary determinants of cancer. *Cancer Discovery* 5, 806-820.
- Hanahan D. Weinberg, R.A. (2011).** Hallmarks of cancer: the next generation. *Cell* 144, 646-674.
- Havel, J.J., Chowell, D., Chan, T.A. (2019).** The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. *Nature Reviews Cancer* 19, 133-150.

- Karlsson, E.K., Kwiatkowski, D.P., Sabeti, P.C. (2014).** Natural selection and infectious disease in human populations. *Nature Reviews Genetics* 15, 379-393.
- Klemm, S.L., Shipony, Z., Greenleaf, W.F. (2019).** Chromatin accessibility and the regulatory epigenome. *Nature Reviews Genetics* 20, 207-220.
- Kodaman, N. et al. (2014).** Human and *Helicobacter pylori* coevolution shapes the risk of gastric disease. *Proceedings of the National Academy of Sciences of the USA* 111, 1455-1460.
- Kyrou, K. et al. (2018).** A CRISPR/Cas9 gene drive targeting doublesex causes complete population suppression in caged *Anopheles gambiae* mosquitoes. *Nature Biotechnology* 36, 1062-1066.
- Louhi, K.R., et al. (2013).** Genotypic and phenotypic variation in transmission traits of a complex life cycle parasite. *Ecology and Evolution* 3, 2116-2127.
- MacLean, RC, San Millán, A. (2019).** The evolution of antibiotic resistance. *Science* 365, 1082-1083.
- Maare, J.P., Patterson, H.G. (2014).** The epigenome of *Trypanosoma brucei*: a regulatory interface to an unconventional transcriptional machine. *Biochimica et Biophysica Acta* 1839, 743-750.
- Wilson, M.S., Maizels, R.M. (2004).** Regulation of allergy and autoimmunity in helminth infection. *Clinical Reviews in Allergy and Immunology* 26, 35-50.
- Mundy, R.D., Cormack, B. (2009).** Expression of *Candida glabrata* adhesins following exposure to chemical preservatives. *Journal of Infectious Diseases* 199, 1891-1898.
- Palumbi, S.R. (2001).** Humans as the world's greatest evolutionary force. *Science* 293, 1786-1179.
- Fellay, J., Pedergnana, V. (2020).** Exploring the interactions between the human and viral genomes. *Human Genetics* 139, 777-781
- Ruiz, J.L. et al. (2018).** Characterization of the accessible genome in the human malaria parasite *Plasmodium falciparum*. *Nucleic Acids Research* 18, 9414-9431.
- Scholthof, K.G. (2007).** The disease triangle: pathogens, the environment and society. *Nature Reviews Microbiology* 5, 152-156.
- Servín-Blanco, R. et al. (2016).** Antigenic variability: obstacles on the road to vaccines against traditionally difficult targets. *Human Vaccines and Immunotherapeutics* 12, 2640-2648.
- Smith, K.R., Hanson, H.A., Hollingshaus, M.S. (2013).** BRCA1, and BRCA2 mutations and female fertility. *Current Opinion in Obstetrics and Gynecology* 25, 207-213.
- Soppe, J.A., Lebbink, R.J. (2017).** Antiviral goes viral: harnessing CRISPR/Cas9 to combat viruses in humans. *Trends in Microbiology* 25, 833-850.
- Strassmann, B.I. (1999).** Mensual cycling and breast cancer: an evolutionary perspective. *Journal of Women's Health* 8, 193-202.
- Sungnak, W., et al. (2020).** SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nature Medicine* 26, 681-687.
- Taylor, P.D., et al. (2006).** The evolutionary consequences or plasticity in host-pathogen interactions. *Theoretical Population Biology* 69, 323-331.
- Tooby, J., Cosmides, L. (1990).** The past explains the present. Emotional adaptations and the structure of ancestral environments. *Ethology and Sociobiology* 11, 375-424.
- Trivers, R.L. (1974).** Parent-offspring conflict. *American Zoologist* 14, 249-264.
- Vink, C., Rudenko, G., Seifert, H.S. (2012).** Microbial antigenic variation mediated by homologous DNA recombination. *FEMS Microbiology Reviews* 36, 917-948.
- Wang, C. et al. (2019).** Epigenetics in the plant-virus interaction. *Plant Cell Reports* 38, 1031-1038.
- Westneat, D.F. et al. (2019).** Causes and consequences of phenotypic plasticity in complex environments. *Trends in Ecology and Evolution* 34, 555-568.
- Williams, G.C., Nesse, R.M. (1991).** The dawn of Darwinian medicine. *Quarterly Reviews in Biology* 66, 1-22.
- Zinsstag, J. et al. (2018).** Climate change and One Health. *FEMS Microbiology Letters* 365, fny085.

## RESUMEN PARA EXPERTOS Y EL PÚBLICO GENERAL





## DESAFÍO 8

### RESUMEN

Uno de los principales objetivos de la biología sintética consiste en ensamblar una unidad de vida mínima con funcionalidad programable. La consecución de este gran desafío —crear una célula sintética desde cero— nos ayudará a comprender los principios básicos de la vida y su aparición a partir de componentes sin vida; asimismo, proporcionará las herramientas necesarias para encontrar soluciones novedosas a problemas medioambientales y sanitarios pendientes.

### PALABRAS CLAVE

Biología sintética (ascendente, descendente, *in silico*)

célula mínima

protocélula

autoorganización biológica

bioingeniería

reconstrucción de máquinas celulares

sistemas sin células

metabolismo mínimo

# VIDA SINTÉTICA

**Coordinadores**

Germán Rivas (CIB-MS)

Eva García (ICP)

**Investigadores y centros de investigación participantes**

Jorge Barriuso (CIB-MS)

Fernando de la Cruz (IBBTEC  
CSIC-UNICAN)

Rafael Giraldo (CNB)

Sonsoles Martín-Santamaría  
(CIB-MS)

Miguel A. Peñalva (CIB-MS)

Manuel Porcar (I2SysBio CSIC-UV)

Jesús Rey (IFS-CCHS)

## RESUMEN EJECUTIVO

¿Es posible crear una célula sintética desde cero? La ciencia moderna ha dedicado importantes esfuerzos a desvelar los principios básicos de la vida. Esta investigación está contribuyendo a un profundo conocimiento de las piezas que componen las máquinas macromoleculares que operan en la célula. Sin embargo, seguimos sin entender cómo estas piezas interactúan de forma coordinada para desarrollar las funciones celulares. Por consiguiente, la generación de vida a partir de los componentes moleculares que existieron en la Tierra primitiva es uno de los grandes enigmas sin resolver y, por tanto, un gran desafío científico. La biología sintética ofrece nuevas estrategias para su resolución. Desde una perspectiva fundamental, la integración de módulos moleculares que den lugar a células sintéticas funcionales ayudará a desvelar los límites de la vida. A este respecto, podemos prever que la fusión de la biología sintética con la evolución molecular y celular puede acabar en la síntesis de células vivas desde cero, cuyas funciones se ajustarán mediante mecanismos evolutivos controlados. Además de responder a preguntas sobre los principios operativos básicos de la vida, la realización de células sintéticas llevará a una nueva revolución tecnológica sin precedentes. Entender cómo funciona la célula viva reconstruyéndola a partir de sus componentes esenciales abrirá nuevos horizontes de aplicación en medicina y biotecnología. El diseño de sistemas de prueba optimizados para el descubrimiento de nuevos fármacos y materiales biodegradables son solo algunos ejemplos. Por consiguiente, los

conocimientos y tecnologías generados durante el proceso de construcción de células sintéticas contribuirán a un mundo más sano y sostenible. No obstante, la investigación sobre la vida sintética debe abordar los desafíos de la sociedad, incluidos los aspectos éticos y filosóficos de estas investigaciones. Asimismo, requiere la formación de futuros científicos en formas novedosas de explorar los sistemas vivos con la aplicación de la ingeniería para comprender y dominar la complejidad biológica.

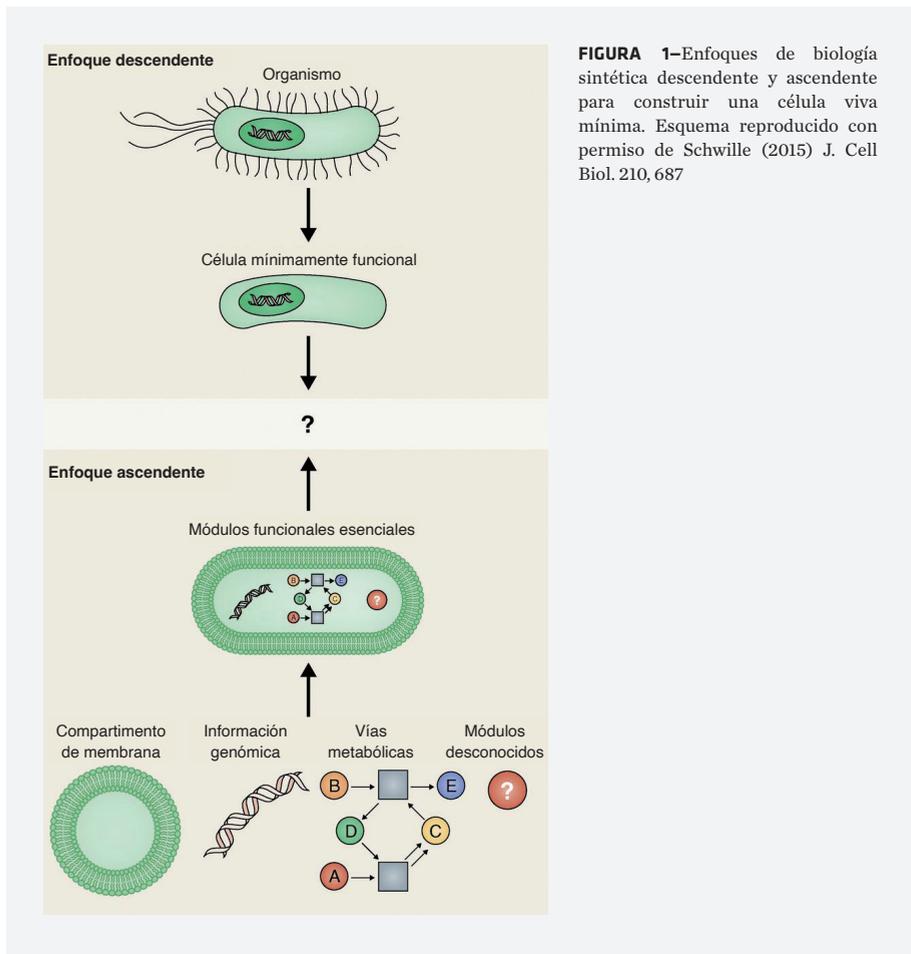
La búsqueda de células sintéticas es un esfuerzo mundial en el que Europa desempeña un papel importante. Recientemente se ha puesto en marcha una iniciativa europea de células sintéticas (European Synthetic Cell) para diseñar una célula mínima a partir de sus componentes moleculares, con el objetivo de superar este desafío en las próximas dos décadas. El CSIC acoge a un importante número de investigadores de primera línea de diversas disciplinas, que actualmente trabajan de forma independiente en diferentes aspectos de la biología sintética y áreas relacionadas. Integrar sus esfuerzos en un programa intramural situaría al CSIC entre los principales centros europeos de investigación en vida sintética y podría mantener una posición destacada en la vanguardia de este gran desafío.

## 1. INTRODUCCIÓN Y DESCRIPCIÓN GENERAL

### 1.1. La búsqueda de la vida sintética

Uno de los grandes desafíos científicos e intelectuales de este siglo consiste en la creación de una célula sintética a partir de sus componentes moleculares constitutivos, con la inclusión de nuevos módulos funcionales. A pesar de nuestros amplios conocimientos sobre los bloques moleculares que componen las células actuales, no entendemos cómo funcionan estos bloques colectivos para llevar a cabo la transición a los sistemas vivos. Saber cómo construir una célula sintética a partir de componentes sin vida ayudará a responder a la pregunta fundamental de cómo funciona la vida. Asimismo, proporcionará una visión exhaustiva de cómo la vida puede surgir espontáneamente a partir de sus componentes no vivos y permitirá interconectar sistemas biológicos con materiales no vivos (Bayley, 2019; Beales *et al.*, 2018; Porcar y Peretó, 2016).

Superar este desafío también permitirá diseñar plataformas de células artificiales y células reprogramadas hacia biofábricas de nueva generación para abrir soluciones únicas a problemas medioambientales, energéticos y



**FIGURA 1**—Enfoques de biología sintética descendente y ascendente para construir una célula viva mínima. Esquema reproducido con permiso de Schwille (2015) *J. Cell Biol.* 210, 687

sanitarios. La comprensión profunda (y el control) de la vida celular para crear células sintéticas desde cero también planteará importantes cuestiones filosóficas, éticas y sociales sobre el impacto de estas tecnologías, lo que abre perspectivas relevantes para el diálogo entre ciencia y humanidades, y entre ciencia y sociedad (Schwille *et al.*, 2018)

## 1.2. Biología sintética

El diseño y la síntesis de nuevos sistemas biológicos (o bioinspirados) que despliegan funciones útiles, incluso las que no existen en la naturaleza, se han acelerado en las últimas décadas y han madurado hasta convertirse en biología sintética. Esta disciplina introduce la perspectiva de la ingeniería para

estudiar los procesos biológicos (Ausländer *et al.*, 2017). En esta óptica, la célula puede considerarse una intrincada fábrica compuesta por dispositivos y máquinas que realizan varias tareas y que se pueden diseñar para producir sistemas con funcionalidad programable.

Muchos logros de la biología sintética han sido el resultado de enfoques descendentes en los que se han modificado células preexistentes (Ausländer *et al.*, 2017). Estos avances han permitido utilizar la ingeniería de circuitos genéticos, módulos biológicos y vías sintéticas para reprogramar organismos y producir fármacos, productos medioambientalmente sostenibles, etc. Sin embargo, en muchos casos, no se conocen bien las propiedades funcionales de las células rediseñadas ni su control, lo que frena los avances. Estas limitaciones básicas en nuestra comprensión han impulsado un enfoque complementario ascendente de la biología sintética. Dicho enfoque tiene como objetivo rediseñar y reconstruir partes, dispositivos y sistemas biológicos con niveles crecientes de complejidad hacia un andamio mínimo similar al de una célula (fig. 1: Bottom-up biology. *Nature*. 2018; 563, 171).

### 1.3 Células mínimas y protocélulas

La posibilidad de producir células mínimas o sintéticas forma parte de los programas de investigación desde hace muchos años (por ejemplo, la obtención de sangre artificial con fines médicos). En esta línea, uno de los principales objetivos de la biología sintética consiste en identificar la configuración mínima que sustenta una célula biológica. ¿Cuál es el conjunto mínimo de proteínas y genes que necesita una célula viva para realizar sus tareas esenciales, como la autorreplicación, el metabolismo y la respuesta a las señales del entorno?

La creación de estas células sintéticas proporcionará valiosos conocimientos sobre los procesos de autoorganización que llevaron a las primeras células a evolucionar a partir de elementos no vivos (Beales *et al.*, 2018). En la actualidad, la comunidad científica considera que uno de los requisitos esenciales para el inicio de la vida fue la compartimentación en espacios definidos de los componentes bioquímicos primordiales. Por ello, uno de los objetivos de los programas de investigación en vida sintética se centra en la optimización de las tecnologías para la producción de compartimentos semipermeables, ya sea en forma de microgotas, vesículas lipídicas o cápsulas basadas en proteínas o ácidos nucleicos.

Los genomas mínimos que contienen un número reducido de genes pueden dar lugar a células viables, como han demostrado el grupo de Craig Venter y el proyecto Synthetic Yeast Genome Project Sc 2.0 (Hutchison *et al.*, 2016;

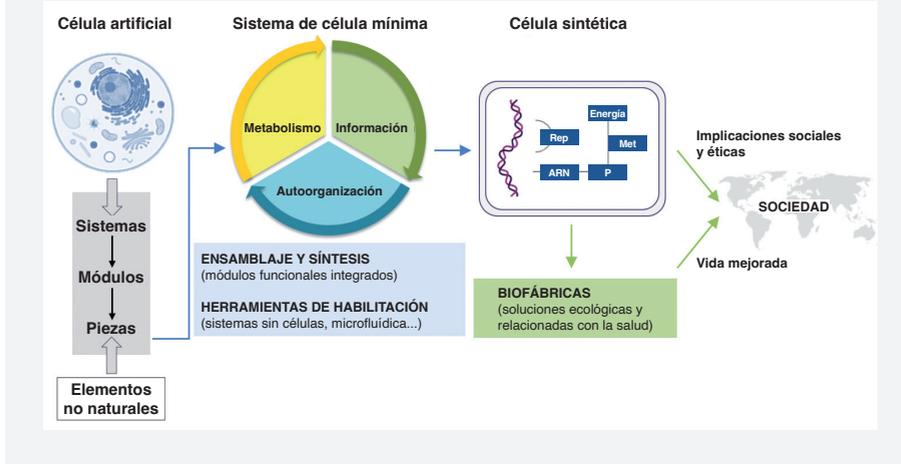
Shao *et al.*, 2018). En estos enfoques descendentes, la creación de una célula mínima se basa en la eliminación selectiva de componentes de sus genomas naturales, con la importante limitación de que se desconocen las propiedades funcionales de un número significativo de genes. Por lo tanto, este enfoque deja abiertas muchas preguntas esenciales sobre el funcionamiento de estas células. De ahí que, a causa de estas complejidades, todavía no haya sido posible diseñar y crear una forma de vida simplificada mediante estrategias ascendentes, partiendo de un número limitado de bloques de construcción químicos o bioquímicos. Aunque nuestra comprensión fundamental de los componentes individuales de los sistemas vivos ha mejorado significativamente, la creación de un conjunto mínimo de componentes del que puedan surgir propiedades similares a la vida sigue suponiendo un gran desafío.

En las últimas décadas, la evolución dirigida, un método que reproduce en el tubo de ensayo los principales procesos de la evolución natural, ha sido fundamental para diseñar proteínas *ad hoc* para aplicaciones biotecnológicas (Arnold, 2018). La implementación de estos mecanismos evolutivos debería afinar sistemas enteros similares a la vida y bloques de construcción de la misma manera que la evolución natural afinó los organismos naturales. La fusión de la biología sintética con la evolución controlada acabará generando células sintéticas con capacidades funcionales sin precedentes y proporcionará conocimientos fundamentales sobre las bases moleculares de la adaptación evolutiva (de Lorenzo, 2018)

#### **1.4 Biología ascendente**

En los últimos años, los avances en la comprensión bioquímica y biofísica de los bloques de construcción individuales y sus interacciones en diferentes niveles de complejidad, incluidos los sistemas de membranas, han impulsado un progreso considerable en la reconstitución ascendente de las máquinas celulares esenciales (por ejemplo, el procesamiento del ADN, el ensamblaje del citoesqueleto, la división celular, etc.) a partir de organismos unicelulares simples y sistemas de mamíferos (Bayley, 2019; Göpfrich *et al.*, 2018). Muchos de estos logros fueron el resultado de estudios de biología fundamental basados en la ingeniería inversa de una función biológica concreta utilizando un conjunto mínimo de componentes moleculares, con el propósito de *comprender construyendo* (Fletcher, 2016). Estos esfuerzos también mostraron las limitaciones y desafíos para reconstruir (y comprender) funciones biológicas más complejas, como aquellas en las que las máquinas moleculares individuales o las células actúan de forma colectiva (Nature 563,188-189 [2018]).

**FIGURA 2**—Hoja de ruta en la investigación de células sintéticas. *Información*: Replicación del ADN, circuitos de genes, etc. *Autoorganización*: ensamblaje del citoesqueleto, división celular, forma y crecimiento, etc. *Metabolismo*: redes de enzimas, producción de energía, etc. (véase 3.0 y 4.1 para más detalles)



Asimismo, se han producido considerables avances en la química de sistemas y la nanotecnología, lo que permite el diseño y la síntesis de sistemas moleculares complejos con propiedades similares a las de la vida para construir andamios y contenedores bioinspirados para la funcionalización precisa de las proteínas (Yeates, 2017). Por otro lado, los avances en ingeniería molecular permitieron el diseño de mutantes específicos o quimeras de proteínas con origen, estructura y función diversos. Entre ellos se encuentran, por ejemplo, las variantes antiguas reconstruidas de las enzimas que pueden evolucionar por evolución dirigida (Alcalde, 2017). Estos desarrollos han tenido lugar al mismo tiempo que los avances tecnológicos complementarios en la tomografía crioeléctrica de alta resolución, la ingeniería genómica ultraeficiente (por ejemplo, la tecnología CRISPR) y las sofisticadas herramientas de imagen molecular.

En conjunto, estos avances facilitarán la integración de los sistemas moleculares individuales en módulos funcionales hacia la construcción de una célula sintética de abajo arriba. Estos enfoques también pueden llevar a generar dispositivos artificiales y bioinspirados que pudieran utilizarse para resolver problemas pendientes relacionados con el medioambiente y la salud (fig. 2).

## 2. IMPACTO EN EL PANORAMA DE LA CIENCIA BÁSICA Y EN LAS APLICACIONES POTENCIALES

La ciencia moderna ha dedicado importantes esfuerzos a desvelar los principios básicos de la vida. Esta investigación está contribuyendo a un profundo conocimiento de las piezas que componen las máquinas moleculares que operan en la célula (como las responsables de la replicación del ADN, la síntesis de proteínas o la propia división celular). Sin embargo, pese a estos avances, seguimos sin entender cómo se coordinan estas piezas para interactuar entre ellas y desarrollar las funciones celulares. La creación de una célula sintética tendrá un gran impacto en la sociedad por su potencial para responder a preguntas pendientes como las siguientes: ¿Qué define y cómo funciona la vida? ¿Cómo empezó la vida en la Tierra? ¿Es la forma de vida actual la única posible?

Además de responder a preguntas sobre los principios básicos del funcionamiento de la vida, el diseño y la creación de una célula sintética también proporcionará nuevas herramientas para ayudar a afrontar los desafíos medioambientales y sanitarios mundiales. Un gran número de sectores de la biotecnología verde, las energías alternativas, la salud, la alimentación y los biomateriales se beneficiarán probablemente de las tecnologías desarrolladas en los programas de investigación relacionados con la generación de células sintéticas. Algunos ejemplos son el diseño de sistemas optimizados para la producción de materiales respetuosos con el medioambiente en la industria de alta tecnología, nuevos biocombustibles y polímeros biodegradables, y las mejoras dirigidas a la biorremediación y al descubrimiento de fármacos.

En particular, dado que los sistemas celulares sintéticos no necesitan depender exclusivamente de las biomoléculas naturales ni de los procesos naturales de la vida, el uso de enfoques alternativos ofrecerá posibilidades de aplicaciones que pueden ir más allá de la tecnología de biología sintética prevista en la actualidad. Se trata de portadores de información (códigos genéticos), aminoácidos y enzimas sintéticos, y el posible desarrollo de sistemas vivos ortogonales (que no interfieran con los sistemas vivos naturales), entre otros.

### 3. PUNTOS CLAVE DEL DESAFÍO

Definir el nivel de complejidad necesario para que surjan propiedades similares a las de la vida e integrar módulos en el tiempo y el espacio para construir sistemas autónomos son los principales desafíos que hay que superar para desarrollar tecnologías celulares sintéticas robustas. Para superar dichos desafíos, se requiere un enfoque integrador y de colaboración en el que participen científicos de un amplio abanico de disciplinas (desde las ciencias biológicas y físico-químicas hasta la ingeniería, pasando por las ciencias sociales y las humanidades). En este sentido, las ciencias de la ingeniería son capaces de diseccionar sistemas complejos en un conjunto manejable de módulos funcionales fundamentales. La aplicación de enfoques de ingeniería (incluida la estandarización de protocolos y piezas) a las células vivas ayudará a crear un conjunto de elementos cruciales (por ejemplo, proteínas, lípidos) para recapitular funciones críticas de la vida, como el procesamiento de la información, la generación de energía y el metabolismo, la compartimentación, el crecimiento y la división, etc. Estas estrategias permitirán primero comprender y controlar las propiedades funcionales de las formas de vida simples para luego aplicarlas al dominio de las características de los organismos multicelulares.

#### 3.1. Montaje: reconstitución de módulos moleculares funcionales biomiméticos

Una de las primeras fases de la búsqueda de células sintéticas se centra en la disección sistemática y la reconstitución ascendente de los rasgos esenciales que caracterizan a los sistemas vivos frente a los no vivos (por ejemplo, el procesamiento de la información genética, el metabolismo, el crecimiento y la división, y la evolución funcional). Esta etapa de ensamblaje constituye un gran desafío experimental debido a la naturaleza altamente dinámica de la mayoría de estas máquinas, que, en muchos casos, están asociadas a sistemas de membrana (Rivas *et al.*, 2014; Fletcher, 2016). Los sistemas macromoleculares funcionales desarrollados en esta fase proporcionarán los módulos necesarios para su posterior integración en células artificiales (véase 3.2). Asimismo, liderarán la optimización de tecnologías y enfoques para aplicaciones medioambientales, biotecnológicas y sanitarias (véase 3.4; fig. 2). Para superar este desafío, los expertos en biología, química, física e ingeniería deben trabajar juntos y explorar varias estrategias. En particular, los estudios de genética cuantitativa y biología celular de las células naturales serán muy relevantes, ya que ayudarán a optimizar el diseño de estos sistemas. Del mismo

modo, será crucial integrar en el proceso experimental los conocimientos derivados de la investigación sobre el origen de la vida y las condiciones potenciales identificadas que permiten la aparición de la vida.

### **3.2. Síntesis: integración de módulos funcionales en células artificiales y naturales**

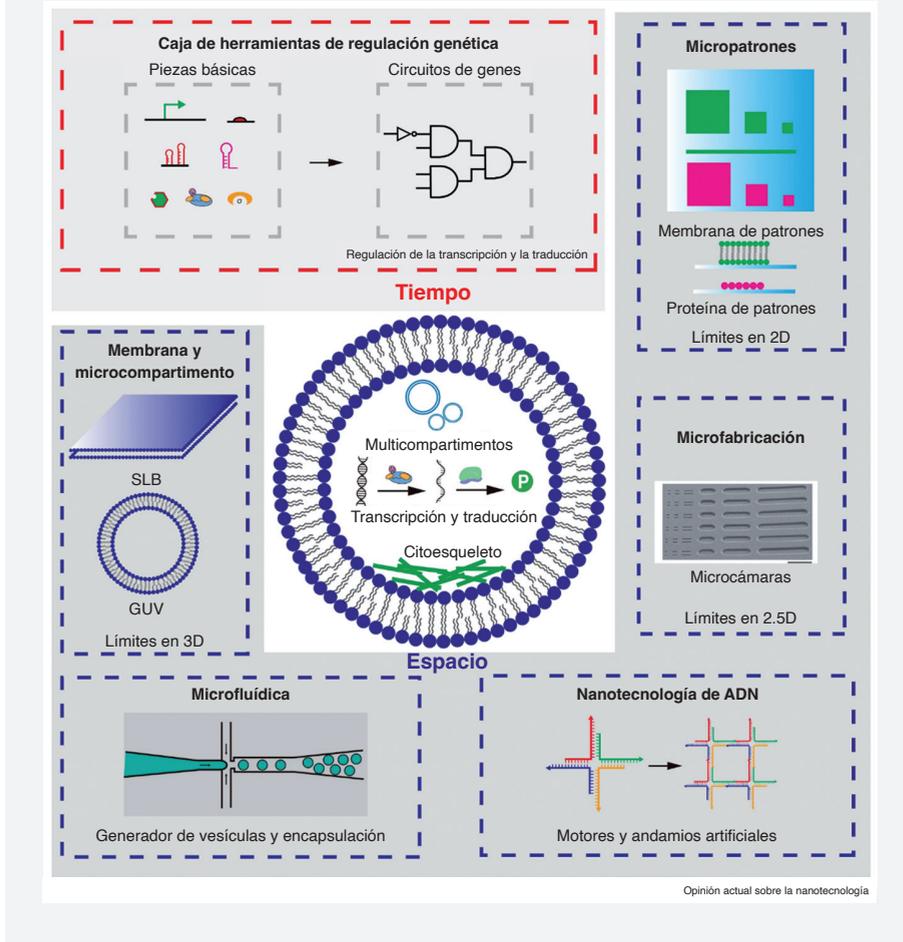
Uno de los desafíos más significativos para el desarrollo exitoso de las células sintéticas consistirá en la integración, mediante un enfoque modular, de las máquinas macromoleculares esenciales antes mencionadas en sistemas que finalmente conducirán a células artificiales de replicación autónoma. Debido a su complejidad intrínseca, este objetivo aún está en ciernes (Bayley, 2019; Beales *et al.*, 2018). La optimización del diseño experimental requiere la combinación de enfoques físico-químicos y ascendentes, con enfoques complementarios de biología sintética descendente impulsados por las modernas tecnologías de ingeniería genómica, así como herramientas de modelación e ingeniería de primera línea. La comprensión cuantitativa de las reglas básicas que rigen la organización funcional de las formas más sencillas de vida celular permitirá entonces explorar diferentes características de la célula, relevantes para la formación de tejidos y organismos multicelulares (por ejemplo, la comunicación y la diferenciación celular, entre otras; Göpflich *et al.*, 2018).

### **3.3. Herramientas: tecnologías que permiten la investigación con células sintéticas**

El diseño y la producción de células sintéticas requerirá el desarrollo de herramientas experimentales innovadoras e integradoras (fig. 3).

*3.3a. Generación y control de compartimentos parecidos a los celulares mediante microfluidos.* Los compartimentos biomiméticos son esenciales para reproducir los procesos de la vida celular en sistemas confinados. La microfluídica ofrece una potente tecnología para fabricar eficazmente un gran número de compartimentos homogéneos del tamaño de una célula, como gotas de agua en aceite y vesículas gigantes. Muchos programas de investigación relacionados con la microfluídica y las células sintéticas se concentran en la fabricación de 1) compartimentos y reactores en un chip, como novedosos dispositivos biotecnológicos y biomédicos, y 2) compartimentos multicelulares que contienen orgánulos artificiales que serán fundamentales para la construcción de órganos y tejidos artificiales (Götfried *et al.*, 2018; Sato y Takinoue, 2019). Los esfuerzos de investigación se centrarán en resolver los siguientes desafíos tecnológicos clave: la optimización de los ensayos para medir con precisión la

**FIGURA 3**—Aspectos, herramientas y tecnologías moleculares que permiten la investigación con células sintéticas. Esquema reproducido con permiso de Jia y Schwille (2019) *Curr Opin Biotechnol* 60:179



eficacia de la encapsulación de biomoléculas dentro de los contenedores, el control de la forma de las gotitas o vesículas y la reducción del tamaño del compartimento hasta la escala micrométrica.

*3.3b. Sistemas sin células para la producción y el descubrimiento de proteínas.* La purificación de proteínas constituye uno de los cuellos de botella para sintetizar piezas funcionales que se utilizarán en módulos de sistemas vivos mínimos. La síntesis de proteínas sin células (CFPS, por sus siglas en inglés) es una

estrategia alternativa atractiva para la producción de proteínas, especialmente para aquellas que son más difíciles de purificar, como las proteínas integrales de membrana o las proteínas solubles que contienen modificaciones cotraduccionales y postraduccionales funcionalmente relevantes (Schwille *et al.*, 2018). Los recientes avances en microfluídica permiten encapsular los CFPS en una variedad de microcompartimentos, lo que ha impulsado el uso de estos sistemas para desarrollar imitaciones celulares. Sin embargo, la mayoría de los CFPS comercialmente relevantes se basan en extractos celulares, cuya composición no se conoce con los detalles necesarios. Así, los avances en los programas de investigación de células sintéticas permitirán un diseño mucho más racional de estos sistemas sin células, de manera que se puedan optimizar individualmente o diseñar para fines específicos (por ejemplo, el diseño de proteínas *de novo*, ya sea impulsado por la modelación *in silico* o por la evolución *in vitro*).

*3.3c. Bloques de construcción naturales de ingeniería.* La biología descendente se basa en la caja de herramientas de la biología sintética (herramientas para la expresión genética, edición o escritura del genoma completo, ingeniería de proteínas, optimización de vías, ingeniería metabólica, etc.) para proporcionar un conjunto robusto de instrumentos para la ingeniería eficiente de bloques de construcción naturales. Estos bloques de ingeniería acabarán transfiriéndose a chasis heterólogos y generando elementos funcionales ortogonales, lo que allanará el camino para el desarrollo de chasis avanzados como biofábricas. Estas piezas diseñadas añadirán funcionalidades novedosas no directamente acopladas a los procesos celulares. En conjunto, estos bloques de construcción naturales diseñados representan una plataforma modular sintonizable para programar células sintéticas con características únicas que se comportan de manera controlada y bien definida.

*3.3d. Nanotecnología de ADN.* En los últimos años, la nanotecnología del ADN ha desarrollado varios módulos para las células sintéticas que constan de arquitecturas y dispositivos de ADN plegado (por ejemplo, *origami* de ADN) fabricados mediante diseño asistido por ordenador en combinación con la gran caja de herramientas para la funcionalización selectiva de sitios del ADN (Göppfrich *et al.*, 2018). Estas estructuras de ADN serán herramientas versátiles para ensamblar componentes de forma programable o para construir partes de un sistema artificial. En este sentido, el diseño de módulos funcionales de ADN basados en plásmidos capaces de replicarse y evolucionar dentro de células mínimas resultará fundamental para seguir avanzando en la investigación de la vida sintética.

### **3.4. Biofábricas: explotación de células sintéticas para una vida mejor**

Los desafíos que se plantean en la investigación con células sintéticas también deben abordar los que permiten ofrecer avances tecnológicos, especialmente para dar soluciones a los problemas medioambientales, industriales y sanitarios pendientes, como se expone a continuación.

*3.4a. Biorreactores sintéticos para soluciones ecológicas.* La biotecnología actual ya tiene mucho éxito en el uso de organismos unicelulares simples para producir compuestos con valor añadido para el medioambiente (por ejemplo, biocombustible o productos químicos verdes) y para biorremediación de la contaminación del suelo. Sin embargo, nuestra limitada comprensión de los principios celulares está obstaculizando los avances. El desarrollo de células sintéticas sentará las bases para el diseño de nuevos biorreactores que se puedan diseñar y controlar para mejorar las soluciones ecológicas (Ausländer *et al.*, 2017). Por ejemplo, las células sintéticas especializadas permitirán optimizar sus vías metabólicas y encapsular las reacciones biocatalíticas para mejorar la producción de biocombustibles y productos químicos verdes, lo que facilita la optimización de las condiciones de trabajo y el rendimiento de la producción en un entorno controlado. Estos avances acabarán dando lugar al desarrollo de la próxima generación de biorrefinerías. En una cuestión relacionada, los cloroplastos artificiales basados en células sintéticas optimizarán la conversión de la luz y el almacenamiento de energía (Miller *et al.*, 2020). Por último, las células sintéticas que combinan las ventajas de cada una de las células *naturales* utilizadas para la biorremediación ofrecerán una solución eficaz y económicamente viable para la descontaminación del suelo.

*3.4b. Tecnología celular sintética para la salud.* Las células sintéticas serán útiles para probar nuevos fármacos, para estudiar cómo interactúa un fármaco específico con una vía determinada de la enfermedad o para optimizar la eficacia de los fármacos existentes. El uso de contenedores parecidos a los celulares para la administración dirigida y controlada de fármacos o compuestos en el cuerpo humano es un objetivo tecnológico desde hace mucho tiempo. Sin embargo, existen barreras fundamentales para dirigir estos envases a los tejidos humanos deseados y provocar la liberación del compuesto activo de forma controlada. Idealmente, es posible imaginar el diseño de células artificiales autosuficientes que se controlan para producir y liberar un compuesto específico solo después de llegar al tejido deseado en el cuerpo. A largo plazo, la tecnología de las células sintéticas pudiera ser decisiva para las

aplicaciones en la salud humana. Los sistemas moleculares funcionales bioinspirados, como los fabricados mediante rutas de síntesis química, inspirarán nuevas clases de materiales activos, como los materiales activos (poliméricos) (Budding y van Hest, 2017) o incluso tejidos artificiales que pudieran aplicarse, por ejemplo, en la medicina regenerativa. En principio, estos *nuevos materiales* optimizados deberían poder autoorganizarse de forma activa, crecer e interactuar funcionalmente con el cuerpo humano. Las aplicaciones previstas pudieran consistir en la sustitución de células enfermas del cuerpo por otras artificiales y en la generación de órganos artificiales y tejidos sintéticos. Estos programas de investigación están íntimamente relacionados con la robótica molecular (Sato y Takinoue, 2019).

### **3.5. Sociedad: implicaciones de la investigación sobre la vida sintética**

La participación de las ciencias sociales y las humanidades en la investigación con células sintéticas goza de suma importancia. La generación de *vida artificial* implica una evaluación ética y filosófica constante y concurrente para dar cabida a posibles cambios de paradigma en nuestra concepción de la materia viva, así como para garantizar que la posible implantación tecnológica proporcione beneficios en lugar de amenazar a la sociedad. El debate ético y social sobre la materia viva frente a la no viva y los problemas de seguridad relacionados con las biotecnologías emergentes son algunos ejemplos (Schwille *et al.*, 2018; Bauer y Bogner, 2020).

La perspectiva histórica y el enfoque de referencia también desempeñarán un papel importante en la comprensión de las implicaciones de la investigación sobre la vida de las células sintéticas para la sociedad. Por ejemplo, la Conferencia de Asilomar sobre el ADN recombinante, celebrada en 1975, supuso la primera ocasión en la que la comunidad científica abordó los problemas de bioseguridad de la tecnología del ADN recombinante y abrió la investigación científica a los debates públicos. Los organismos transgénicos constituyen también ejemplos paradigmáticos del papel crítico que pueden desempeñar en la sociedad la evaluación y la comunicación del valor de la investigación biológica y las aplicaciones de la biotecnología. Los transgénicos (y, en particular los alimentos manipulados genéticamente) han generado un profundo debate social, alarma y movilización social. Estas reacciones suelen producirse sin un análisis riguroso ni un conocimiento detallado de los fundamentos biológicos, los intereses, los valores (creencias), los aspectos éticos, los beneficios y los riesgos. La investigación con células sintéticas debería aprender de estas y otras experiencias pasadas (por ejemplo, la tecnología de Internet) a

fin de generar el marco necesario para expandirse de manera éticamente responsable. Así pues, el desarrollo de la ciencia y la tecnología debe producirse en concierto con la educación de la próxima generación de estudiantes, el apoyo técnico y la sociedad en general. Este esfuerzo garantizará una comprensión precisa de los avances científicos resultantes del desarrollo y el uso de células sintéticas. De este modo, las actividades de difusión resultarán fundamentales para acercar la biología sintética al público en general.

## DESAFÍO 8 | BIBLIOGRAFÍA

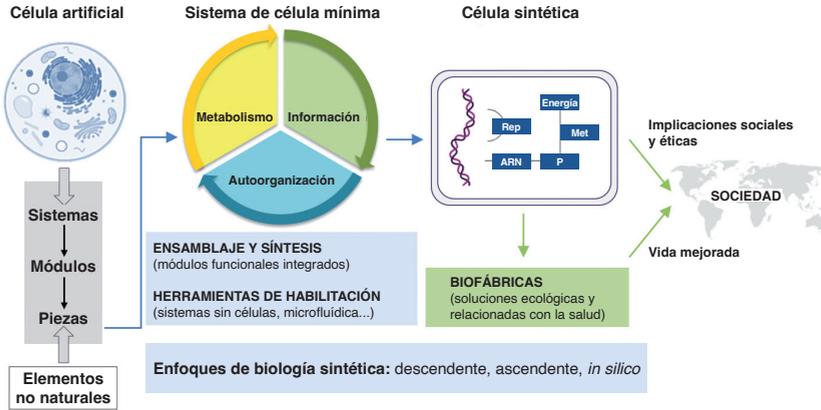
- Alcalde, M. (2017).** When directed evolution met ancestral enzyme resurrection. *Microbial Biotechnology* 10, 22-24.
- Arnold, F.H. (2018).** Directed Evolution: Bringing New Chemistry to Life. *Angew Chem. Int. Ed. Engl.* 57, 4143-4148.
- Ausländer, S., Ausländer, D. y Fussenegger, M. (2017).** Synthetic Biology-The Synthesis of Biology. *Angew Chem. Int. Ed. Engl.* 56, 6396-6419.
- Bauer, A. y Bogner, A. (2020).** Let's (not) talk about synthetic biology: Framing an emerging technology in public and stakeholder dialogues. *Pub. Underst. Sci.* 29(1), 492-507. doi. org/10.1177/0963662520907255.
- Bayley, H. (2019).** Building blocks for cells and tissues. *Emerg. Top. Life Sci.* 3, 433-667
- Beales, P.A., Ciani, B. y Mann, S. (2018).** The artificial cell: biology-inspired compartmentalization of chemical function. *Roy. Soc. Interface Focus* 8, 20180046. doi. org/10.1098/rsfs.2018.0046
- Buddingh, B.C. y van Hest, J.C.M. (2017).** Artificial cells: Synthetic compartments with life-like functionality and adaptivity. *Acc. Chem. Res.* 50, 769-777.
- De Lorenzo, V. (2018).** Evolutionary tinkering vs. rational engineering in the times of synthetic biology. *Life Sci. Soc. Policy* 14, 18.
- Fletcher, D.A. (2016).** Bottom-up biology: Harnessing engineering to understand Nature. *Dev Cell.* 38, 587-589.
- Göpfrich, K., Platzman, I. y Spatz, J.P. (2018).** Mastering complexity: towards bottom-up construction of multifunctional eukaryotic synthetic cells. *Trends Biotechnol.* 36, 938-951.
- Hutchison, C.A. 3rd, Chuang, R.Y., Noskov, V.N. et al. (2016).** Design and synthesis of a minimal bacterial genome. *Science* 351(6280), aad6253.
- Miller, T.E., Beneyton, T., Schwander, T. et al. (2020).** Light-powered CO<sub>2</sub> fixation in a chloroplast mimic with natural and synthetic parts. *Science* 368, 649-654.
- Porcar, M. y Peretó, J. (2016).** Nature versus design: synthetic biology or how to build a biological non-machine. *Integr. Biol. (Camb)* 8, 451-455.
- Rivas, G., Vogel, S.K. y Schwille, P. (2014).** Reconstitution of cytoskeletal protein assemblies for large-scale membrane transformation. *Curr. Opin. Chem. Biol.* 22, 18-26.
- Sato, Y., Takinoue, M. (2019).** Creation of artificial cell-like structures promoted by microfluidics technologies. *Micromachines* 10, 216; doi:10.3390/mi10040216
- Shao, Y., Lu, N., Wu, Z. et al. (2018).** Creating a functional single-chromosome yeast. *Nature* 560, 331-335.
- Schwille, P., Spatz, J., Landfester, K. et al. (2018).** MaxSynBio: Avenues towards creating cells from the bottom up. *Angew Chem. Int. Ed. Engl.* 57, 13382-13392.
- Yeates, T.O. (2017).** Geometric principles for designing highly symmetric self-assembling protein nanomaterials. *Annu. Rev. Biophys.* 46, 23-42.

RESUMEN PARA EXPERTOS

# CSyCell – Constructing Synthetic Cells

Dominio de las capacidades intrínsecas de los sistemas biológicos:

- para comprender los principios básicos de la vida y su aparición
- para proporcionar soluciones novedosas a los problemas medioambientales y sanitarios

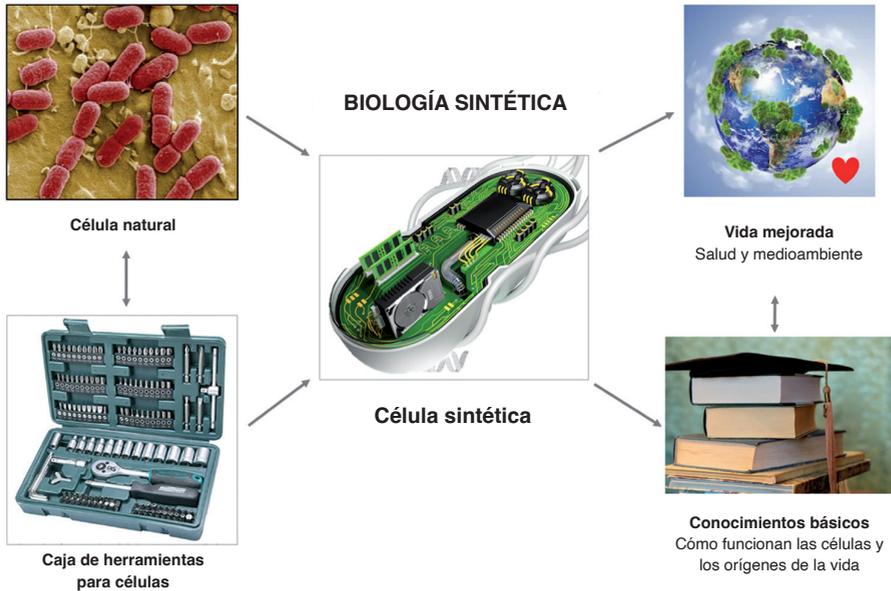


SynCell Europe: MaxSynBio, BaSyC (Países Bajos), Max-Planck/Bristol Minimal Biology, FABRICELL (Reino Unido)



Centros del CSIC: CIB, CNB, I2SysBio Valencia, IBBTEC-Cantabria, CMBSO, CAB-INTA, ICP, IQFR, IFS-CCHS, Inst. Biofísica – Bilbao...

RESUMEN PARA EL PÚBLICO GENERAL





Cómo apareció la vida en la Tierra y cómo luego se diversificó son las preguntas sin respuesta discutidas en este volumen. Estas preguntas ahondan en el pasado profundo de nuestro planeta, donde la biología se entremezcla con la geología y la química, para explorar el origen de la vida y comprender su evolución, ya que “nada tiene sentido en biología sin un enfoque evolutivo” (Dobzhansky, 1964). Los ocho desafíos que componen este volumen resumen nuestro conocimiento actual y plantean las futuras direcciones de investigación, tocando diferentes aspectos del estudio de la evolución. El volumen discute teorías recientes sobre cómo surgieron las primeras moléculas, se organizaron y adquirieron su estructura, lo que permitió la aparición de las primeras formas de vida. También intenta explicar cómo la vida ha ido cambiando con el tiempo, dando lugar, a partir de unas bases moleculares similares, a una inmensa diversidad biológica, y a entender cuál es la relación filogenética entre todas las diferentes formas de vida. Se analiza además la evolución de los seres humanos, su relación con el medio ambiente, y las implicaciones en la salud humana y la sociedad. Cerrando el círculo, el volumen discute la posibilidad de diseñar nuevas máquinas biológicas, para crear células prototipo a partir de sus componentes y cómo este conocimiento se puede aplicar para mejorar nuestro ecosistema. Con una coordinación eficaz entre las tres grandes áreas de conocimiento implicadas, el CSIC puede convertirse en un referente internacional en la investigación en este campo.