

## Cuando los discos duros fueron reemplazados por bacterias: futuras aplicaciones de la ingeniería genética

Paula Bellés Sancho

La fotografía familiar tomada de tus últimas vacaciones en el pueblo de los abuelos, el correo spam de aquella página web a la que te subscribiste y ahora no recuerdas de qué era, o el trabajo de investigación que hiciste en la ESO sobre los Reyes Católicos son solo tres granos de arena que forman la gran duna del desierto de la generación de la información. Cada día se generan millones de datos que son procesados, analizados y almacenados, dispuestos al alcance de prácticamente cualquier persona, y todo esto, gracias a Internet, uno de los pilares más robustos sobre los que se ha generado la sociedad digitalizada en la que vivimos. En un estudio publicado en la revista *Science* (Hilbert, 2011) el año 2011, se cuantificó la cantidad de información generada y almacenada en el mundo hasta el año 2007 y el resultado fue bastante impactante: 290 exabytes, o lo que es lo mismo, 290 millones de terabytes. Difícil de visualizar, ¿verdad? Pero poniéndonos en situación, si cualquier disco duro de precio razonable que puedas adquirir en una tienda de electrónica tiene una capacidad de aproximadamente 1 terabyte, empezamos a percatarnos de la magnitud del asunto. Toda esta información necesita ser guardada en dispositivos o unidades de almacenamiento, aparatos que leen y escriben estos datos formando lo que conocemos como “memoria”, tanto de forma interna en un ordenador como de forma externa o *flash*. Ahora, todo el mundo utiliza los *pendrives*, pero no siempre ha sido así y, al igual que todo en esta vida, es el resultado de la evolución, hasta el momento, de los dispositivos de almacenamiento. Si la sociedad avanza, la tecnología tiene que avanzar, y viceversa. Ya hemos empezado a olvidar los CDs, DVDs e incluso las tarjetas SD, por no hablar de los VHS o de aquellos discos magnéticos o “disquetes” en los que solo cabían un par de documentos Word y no precisamente de gran tamaño. La tecnología se mueve a pasos agigantados y, como consecuencia, estos dispositivos quedan obsoletos con el paso de los años, por lo que es de esperar, que el futuro de estos aparatos tan corrientes como lo son a día de hoy un *pendrive* o una *nanoSD*, vayan a terminar de la misma forma, en el olvido del desuso.

Para hacer frente a esta situación en la que nos encontramos, científicos de todo el mundo se encuentran trabajando juntos en soluciones que se alejan de todo concepto tecnológico que se conocía hasta ahora, inmersos en proyectos interdisciplinares que mezclan la biología (concretamente, el ADN) con la informática. La capacidad inventiva de estos investigadores cruza las fronteras de la imaginación: que una imagen o una película pueda estar guardada en una bacteria es una idea un tanto descabellada que cuesta de creer, casi sacada de un guion de ciencia ficción. Pero, aunque parezca difícil, está empezando a materializarse. Los grandes avances en el ámbito computacional y la ingeniería genética han permitido que, por ejemplo, los sonetos de Shakespeare hayan sido traducidos al código del ADN e introducidos en una bacteria, rompiendo todos los esquemas sobre el almacenamiento de datos, pero basándose en uno de los pilares de la biología: el material genético como mecanismo para guardar

información de forma segura y durante un largo periodo de tiempo. El ADN es un concepto que todo el mundo sabe a qué hace referencia, pero no todos llegan a entender o saben exactamente qué es. Por lo tanto, para entender como estos científicos llegaron a desarrollar esta disruptiva idea, primero hablaremos sobre el ácido desoxirribonucleico o ADN.

## ¿Qué es el ADN?

El ADN, o ácido desoxirribonucleico, es una molécula química presente en todo organismo vivo, desde microorganismos, hongos, plantas hasta animales. Todo ser viviente está formado por células, la mínima unidad funcional con vida, y cada una de estas piezas que completan nuestro organismo contiene el mismo ADN, confiriéndole su característica identidad. La función principal de esta molécula es guardar toda la información necesaria para que dicha célula pueda realizar sus funciones y, por ello, es necesario que esta información pase a sus descendientes.

Seguramente, al leer o escuchar “ADN”, todos visualizamos la famosa imagen de la doble hélice que aparece en cualquier libro de biología, en anuncios o en una noticia del telediario, pero es posible que nadie nos haya explicado cual es el significado de esta estructura. Para comprender como funciona el ADN, hemos de tener presente que esta molécula es el conjunto de diversos elementos que, ordenados de una forma determinada, se leen como si de un código se tratara. Por ejemplo, el código morse es un sistema de comunicación en el que, basándose en la combinación de distintas señales emitidas de forma intermitente (puntos o señales cortas, y rayas o señales largas), se transmite un mensaje. Tres puntos consecutivos representan la letra “S”, y si seguido de estos tres añadimos tres señales largas y luego otra vez tres cortas, obtenemos “S.O.S.”, la señal de socorro. Comparativamente, la molécula de ADN está formada por 4 “señales” diferentes o, con su nombre científicamente correcto, “bases químicas”: adenina, guanina, citosina y timina, abreviadas como A, G, C y T, respectivamente. Estos elementos formarían lo que, en la imagen de la doble hélice, son los peldaños en horizontal, y su combinación sucesiva da lugar al código de la información que la célula interpretará y traducirá.

No obstante, para que esta información pueda ser copiada y transmitida a los descendientes cada vez que la célula lo necesite, esta estructura funciona como si de una cremallera se tratara. En realidad, la molécula de ADN contiene la información por duplicado y cada una de estas copias, una contrapuesta y complementaria a la otra, es lo que llamamos “hebra”, lo que vendría siendo una de las dos cadenas que conforman la estructura en forma de hélice. Por ello, cuando nos referimos a las “bases”, en realidad siempre hablamos de “pares de bases”, ya que cada base química es complementaria a una de las otras tres (A - T por un lado, y G - C por el otro). Así, cuando la célula empieza a multiplicarse, esta cremallera se abre y empieza a copiarse complementariamente (donde hay una C, la complementaria tendrá una G y así sucesivamente) hasta tener, por completo, dos cadenas de ADN iguales con la misma información en cada una de las hebras.

Desde que Watson y Crick, a mediados del siglo pasado, publicaron la estructura del ADN gracias a la famosa imagen de difracción de rayos X del ADN obtenida por Rosalind Franklin, el conocimiento y desarrollo de la genética molecular ha crecido exponencialmente hasta

nuestros tiempos. Actualmente, es posible conocer por completo el código genético de una célula e incluso de un ser humano gracias a la mejora en las técnicas de secuenciación, aunque no fue así desde el inicio. Hace quince años, secuenciar el genoma de un humano o cualquier organismo costaba varios millones de dólares, pero gracias al desarrollo de estas tecnologías, estamos llegando al punto de poder secuenciar un genoma por el módico precio de mil dólares y, a este ritmo, la tendencia va a la baja. Sin embargo, el gran motor que ha impulsado el desarrollo de esas técnicas de análisis y secuenciación fue, sin duda alguna, el Proyecto Genoma Humano (Shendure, 2017), iniciado en 1990 que tenía como objetivo fundamental determinar la secuencia de pares de bases que componen el ADN del genoma humano desde un punto de vista físico y funcional, con tal de utilizar esta información para aplicaciones médicas como el diagnóstico de enfermedades genéticas. Con un presupuesto de 3 mil millones de dólares y la participación de diferentes países, en 2004 se publicó la versión completa de nuestro genoma, dejando como legado no solo esta valiosa información. El desarrollo de este proyecto supuso un cambio a mejora de la producción de datos a gran escala y el desarrollo de recursos computacionales potentes y robustos, impulsando la transición de la investigación biomédica hacia el concepto que conocemos actualmente.

### **Pero ¿cómo surge la idea de introducir nuestros datos en el ADN de una bacteria?**

Partiendo del concepto de que el ADN es utilizado por la célula para guardar una elevada densidad de información necesaria y que, por su vital importancia, esta información tiene que mantenerse durante largos periodos de tiempo, un grupo de científicos tuvo la idea de utilizar esta característica de la biología para solucionar uno de los rompecabezas que más preocupan a los investigadores hoy en día: el almacenamiento de la información. Como mencionábamos al principio del artículo, cada día generamos millones de datos dentro de cualquier ámbito de la sociedad y, en los centros de investigación, es uno de los campos en los que la generación de esta información nunca cesa.

Era mediados de febrero del 2011 cuando Nick Goldman, líder de un grupo de investigadores del *European Bioinformatics Institute* (EBI), comentando con algunos compañeros bioinformáticos el problema que suponía guardar las secuencias de genomas y otros datos de todo el mundo, empezaron a bromear sobre la disparatada idea de utilizar el problema en sí (el ADN) como solución. Era cierto que almacenar datos en el ADN sería un proceso patéticamente lento comparado con la escala de microsegundos que se tarda en leer o escribir bits en un chip. Se necesitarían mucho tiempo para poder codificar toda la información en una cadena de ADN y, aún más, recuperar esta información usando las técnicas de secuenciación. Sin embargo, también era cierto que todo el genoma humano cabe perfectamente dentro de una célula invisible para el ojo humano. Así que, con la idea en mente, Goldman y Ewan Birney, uno de los otros visionarios, empezaron a involucrarse y sacar adelante el proyecto.

Uno de los primeros problemas con los que se enfrentaron fue la síntesis del ADN y la secuenciación. Estas técnicas, a pesar de lo avanzadas que son, no aseguran que la información sea correcta al 100%. Durante estos procesos, la maquinaria utilizada puede introducir un error cada 100 bases aproximadamente, algo que, para utilizar el ADN con este fin, se tiene

que minimizar al máximo, ya que queremos que nuestra información permanezca tal y como es, sin que evolucione o cambie, como bromea Goldman en algunas entrevistas que circulan por Internet. Pero, como todo en esta vida, encontraron una la solución utilizando un sistema que permitía codificar bits en pares de bases de forma que les permitía detectar los errores y repararlos. Así, dos años más tarde, anunciaron el gran logro: consiguieron codificar cinco archivos diferentes en el ADN (Goldman, 2013). Para aquel entonces, otro grupo liderado por el biólogo George Church en la universidad de Harvard en Cambridge, Massachusetts, había conseguido de forma independiente, demostrar la capacidad de almacenar información en el ADN codificando su propio libro "Regenesis" dentro de una bacteria haciendo 90 mil millones de copias (Church, 2014). Este hecho solo fue el pistoletazo de salida y el año pasado, un grupo de científicos liderado por Seth Shipman (Ledford, 2017), consiguieron almacenar uno de los primeros cortometrajes de la historia del cine dentro de una bacteria y luego leerla. Lo que impulsó a Shipman en este caso, no fue el problema físico del almacenamiento de información, sino más bien, explorar la posibilidad de hacer grabaciones a nivel celular que le mostraran cómo las células en el cerebro adquieren diferentes identidades.

## Desde los píxeles al ADN y viceversa

Conseguir codificar cualquiera de los tipos de archivos nombrados anteriormente no fue tarea fácil. Para conseguirlo, la idea general fue asociar el código binario (sistema numérico de dos dígitos o bits (0 y 1) utilizado en informática y telecomunicaciones para el almacenamiento de datos) al código genético (los pares de bases) (Extance, 2016).

En el proyecto de Shipman (Shipman, 2017) empezaron introduciendo una sola imagen en blanco y negro y reconstruyéndola desde la bacteria (Ilustración 1). Para ello, se asoció una sombra a cada píxel de la imagen a cada par de base de ADN formando pequeños fragmentos de material genético. Al inicio de cada uno de estos fragmentos, se asociaba un código de barras que posicionaba cada grupo de píxeles en la imagen y así, la totalidad de todos los fragmentos formaba el código de ADN de toda la imagen. Una vez sintetizadas todas las secuencias, estas se introdujeron en el genoma de una bacteria intestinal común, concretamente la especie *Escherichia coli* o *E. coli*, de forma ordenada para poder saber el orden de cada uno de los códigos de barras utilizando la tecnología de CRISPR, metodología que ha revolucionado la ingeniería genética actual.

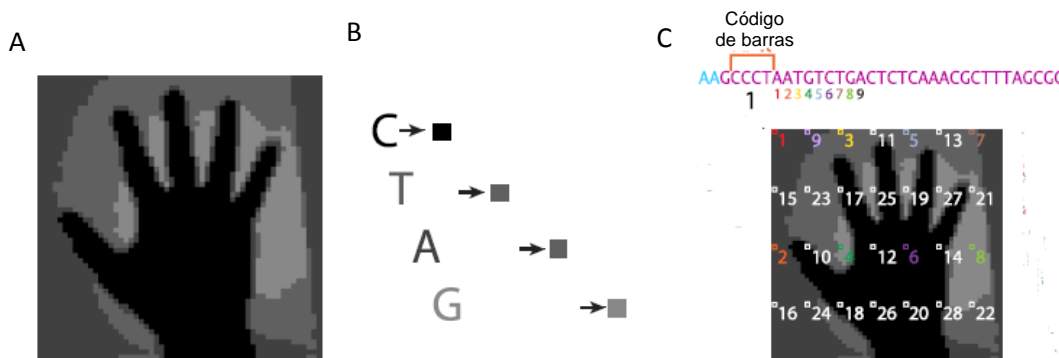


Ilustración 1. Esquema visual del proyecto de Shipman: (A) Imagen inicial; (B) asociación de cada sombra a una base química y (C) esquema de la posición de cada secuencia con su código de barras. Modificada de Shipman et al., 2017

CRISPR se trata de una nueva técnica de edición de genes que ha revolucionado el mundo de la genética molecular. Este método, en realidad, es una adaptación de un mecanismo de defensa que utilizan algunas bacterias contra los virus, parecido al sistema inmune de los animales. Cuando un virus quiere infectar una bacteria, éste introduce su material genético dentro del microorganismo para poder vivir a su costa, reproducirse y poderse transmitir dentro de esa población bacteriana. Sin embargo, se ha visto que estas bacterias, cuando son infectadas por primera vez, cortan ciertos fragmentos del ADN del virus y lo guardan junto con su material genético. De esta forma, cuando un virus de la misma especie vuelve a infectar esta bacteria, ella ya ha generado una especie de memoria inmunológica que le permite evitar que ese virus la vuelva a infectar. A todo este mecanismo es al que se le conoce como CRISPR (Ledford, Five big mysteries about CRISPR's origins, 2017).

En el ámbito de la biología, cuando hablamos de la "edición de genes", nos referimos al proceso de introducir en un genoma, una secuencia génica que inicialmente no formaba parte de este genoma. En otras palabras, al hecho de "cortar-y-pegar" el ADN. En la investigación se trabaja con este concepto desde hace años ya que permite tener modelos experimentales "a la carta", es decir, con las características genéticas que se requieren para cada tipo de experimento. No obstante, esto siempre se ha llevado a cabo mediante otros procesos que son mucho más costosos, caros y que han tardado mucho tiempo en perfeccionarse. CRISPR ha abierto las puertas a un mundo nuevo en la edición de genes, ampliando su utilización a otros horizontes: terapia génica en humanos para poder curar enfermedades genéticas, cultivos transgénicos más resistentes a plagas que no necesitan de tratamientos químicos o "engineered ecosystems" en los que los vectores de enfermedades, como los mosquitos o garrapatas, son modificados para evitar su transmisión, son algunos de los ejemplos en los que la tecnología CRISPR puede jugar un papel muy importante debido a su gran potencial (Ledford, 2015).

Volviendo al proceso, una vez introducidas todas nuestras secuencias dentro del genoma bacteriano, las bacterias empezaron a reproducirse y se almacenaron durante un periodo de tiempo. Fue entonces cuando se secuenció el genoma de estas *E. coli* mediante potentes secuenciadores, poniendo todos los fragmentos en orden y convirtiéndolos en formato digital, habiendo aplicado durante este proceso métodos que permitían corregir errores comparando cada fragmento con otras copias. Así, consiguieron con éxito su primer propósito y utilizaron el mismo método para introducir la película. El cortometraje fue también guardado en el genoma de la bacteria preservándose intacta con cada una de las nuevas generaciones de la progenie (Ilustración 2).

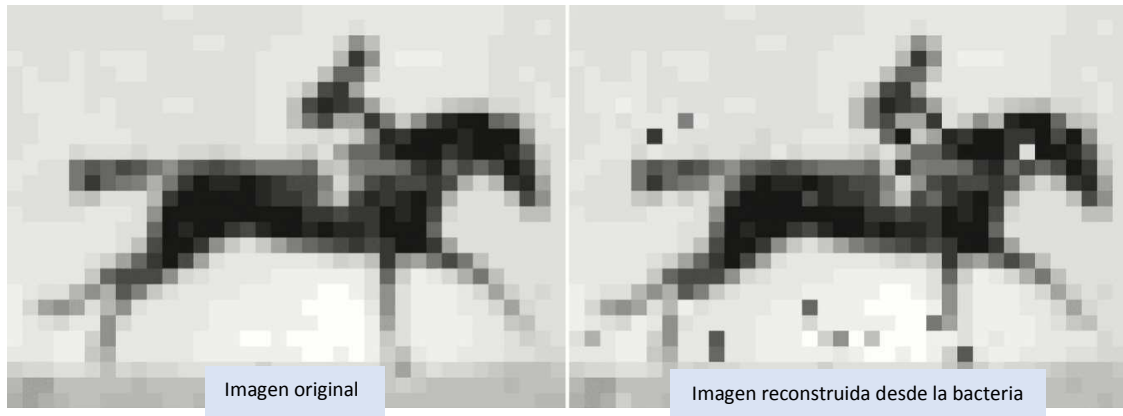


Ilustración 2. Imagen de la película *Human and Animal Locomotion* del fotógrafo Eadweard Muybridge de 1887, filmografía seleccionada por Shipman para realizar el proyecto. Modificada de Ledford, *Lights, camera, CRISPR: Biologists use gene editing to store movies in DNA*. 2017.

### ¿Solo utilizarse para almacenar nuestra información?

“La biología no es simple información escrita, es hacer algo con ella” dijo el doctor Richard Feynman en una clase universitaria en 1959, y es que la naturaleza nos proporciona una gran cantidad de información y conocimiento, pero en nuestras manos está como la utilizamos y sacamos provecho. Con creatividad y perspectiva, podemos utilizar esta ciencia para solventar problemas que conciernen a la sociedad en la que vivimos. Existen otras preocupaciones en las que el desarrollo de este tipo de tecnología podría ser útil en el ámbito de la biomedicina.

Uno de los usos en los que Church ya soñaba es en utilizar las bacterias que conviven en nuestro cuerpo como si fueran las “cajas negras” de los aviones, grabando lo que ocurre en nuestro organismo y, en el momento en el que surge algún problema, acceder a estas grabaciones para poderlo diagnosticar y dar una solución específica y concreta. Birney también ha hablado en algunas conferencias sobre esta idea. Actualmente, con el desarrollo de la genética y la publicación del genoma humano, ha surgido la farmacogenética, un campo que complementa la farmacología con la genética. Muchas veces, los tratamientos con fármacos no son efectivos para la misma enfermedad en diferentes personas, como ocurre con el caso del cáncer, y esto es debido a nuestro propio genoma. Cada persona tiene su material genético único y por ello, nuestra genética juega un papel muy importante en el desarrollo de ciertas enfermedades. Por lo tanto, para tratarlas de forma efectiva, el desarrollo de esta idea futurista sería de gran utilidad proponiendo tratamientos personalizados para cada individuo, aumentando la eficiencia y disminuyendo así los efectos secundarios que puedan derivarse.

**Bibliografía:**

- Church, G. M., Elowitz, M. B., Smolke, C. D., Voigt, C. A., & Weiss, R. (2014). Realizing the potential of synthetic biology. *Nature Reviews Molecular Cell Biology*, 15(4), 289.
- Extance, A. (2016). How DNA could store all the world's data. *Nature*, 537(7618).
- Goldman, N., Bertone, P., Chen, S., Dessimoz, C., LeProust, E. M., Sipos, B., & Birney, E. (2013). Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature*, 494(7435), 77.
- Hilbert, M., & López, P. (2011). The world's technological capacity to store, communicate, and compute information. *science*, 332(6025), 60-65.
- Ledford, H. (2015). CRISPR, the disruptor. *Nature*, 522(7554), 20.
- Ledford, H. (2017). Five big mysteries about CRISPR's origins. *Nature News*, 541(7637), 280.
- Ledford, H. (2017). Lights, camera, CRISPR: Biologists use gene editing to store movies in DNA. *Nature News*.
- Shendure, J., Balasubramanian, S., Church, G. M., Gilbert, W., Rogers, J., Schloss, J. A., & Waterston, R. H. (2017). DNA sequencing at 40: past, present and future. *Nature*, 550(7676), 345.
- Shipman, S. L., Nivala, J., Macklis, J. D., & Church, G. M. (2017). CRISPR–Cas encoding of a digital movie into the genomes of a population of living bacteria. *Nature*, 547(7663), 345.