

L'ADN, el disc dur del futur



L'adenina, guanina, timina i citosina, els blocs fonamentals de l'ADN, poden servir per codificar els zeros i uns dels arxius digitals.

L'any 2012 es va aconseguir per primer cop emmagatzemar informació digital en fragments d'ADN. En l'era del Big Data, en què la quantitat d'informació generada augmenta a gambades, tot indica que els centres de dades acabaran sent espais amb recipients d'ADN. És molt més resistent que qualsevol format electrònic i pot emmagatzemar la informació en molt menys espai. Amb una nova tècnica anunciada a principis de mes, tota la informació que s'ha generat des de l'inici de la computació es podria ficar dins de dues camionetes.

Entrevista d'Àstrid Bierge

Acid desoxiribonucleic. A escola tots vam haver de memoritzar l'enrevessat nom del manual d'instruccions que ens fa i ens defineix. L'ADN és la molècula que codifica la informació genètica de tots els éssers vius, més els virus. Segons com estiguin organitzats els seus quatre blocs fonamentals, o bases —adenina, guanina, timina i citosina—, la criatura que en sorgeix és una o una altra. Però l'ADN no és només el *software* més complex i sofisticat que coneguem, també dóna mil voltes a tots els *hardwares* de la tecnologia humana. Extremadament compacte, és capaç d'emmagatzemar una quantitat ingent de dades en un volum d'espai irrisori i, a més, pot durar milers d'anys sense patir ni una rascada. Si aquest maquinari serveix per emmagatzemar la vida, què ens impedeix utilitzar-lo per guardar altres tipus d'informació? Llibre en mà, res. I a la pràctica, millor o pitjor, ja es fa.

El 2012 es va aconseguir per primer cop utilitzar molècules d'ADN per codificar i guardar informació en format digital. En un article publicat a la revista *Science*, George Church i el seu equip de la Universitat de Harvard van anunciar que havien pogut guardar un llibre, onze imatges JPG i un programa JavaScript. Van utilitzar l'alfabet de quatre lletres —A, G, T, C— per codificar els zeros i uns dels arxius digitals. Cal apuntar que hi ha tècniques per sintetitzar ADN artificial, és a dir, que qualsevol persona pot dissenyar un fragment d'ADN en què les lletres estiguin organitzades d'una determinada manera i un laboratori s'encarrega de fabricar aquesta molècula concreta.

Aquest estudi va establir les bases de l'anomenada memòria d'àcid nucleic, però el sistema de codificació que van utilitzar aquests científics era molt ineficient. Per cada gram d'ADN, la tècnica només permetia guardar 1,28 petabytes d'informació. Una xifra molt inferior a la capacitat d'aquest *hardware* biològic, que, segons els científics, té un límit teò-

ric d'emmagatzematge d'1,8 bits d'informació per nucleòtid —una de les quatre lletres més una molècula de sucre i una d'àcid fosfòric.

A partir d'aquest estudi, altres grups científics van començar a desenvolupar tècniques per millorar la fita de Church. El 2013, uns investigadors de l'Institut Europeu de Bioinformàtica van codificar una fotografia de la seu d'aquest organisme, un arxiu en què es descrivia la tècnica emprada en aquest estudi, els 154 sonets de Shakespeare i un fragment en àudio del discurs "Jo tinc un somni". El fragment d'ADN que contenia aquests arxius era tan insignificantment petit que Nick Goldman, cap del projecte, en rebre el tub d'assaig amb l'ADN que havien programat, va pensar que era buit. En realitat, hi havia al voltant de cinc milions de bits d'informació. Tots aquests arxius van poder ser llegits i reproduïts amb una precisió que oscil·lava entre el 99,99% i el 100%. Això vol dir que, un cop la informació nucleica que havien emmagatzemat va ser traduïda al llenguatge digital, es podia escoltar la veu de Martin Luther King a la perfecció.

Any d'inflexió

En l'últim any aquest camp d'estudi ha fet un salt de gegant. Fins fa poc, tots els mètodes utilitzats tenien un gran des-

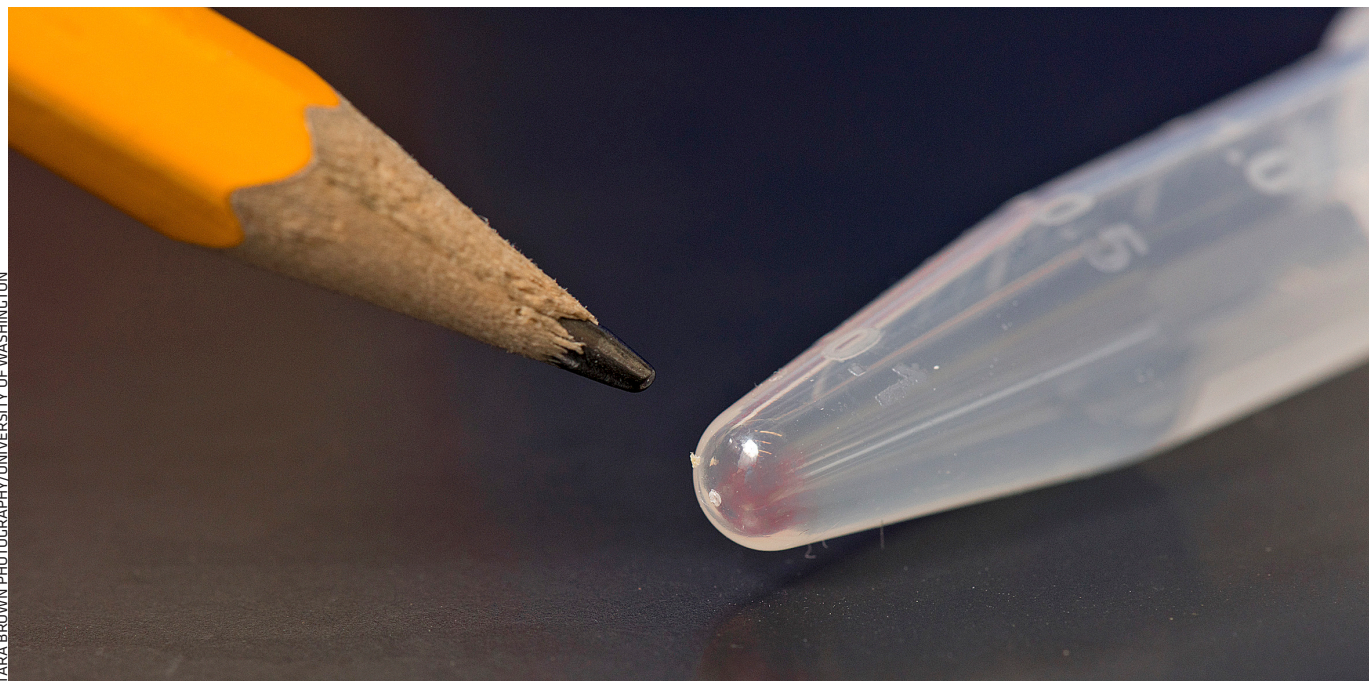
avantatge: calia seqüenciar tota la tira d'ADN per poder llegir un dels diversos arxius codificats. Si n'hi ha pocs, el problema no és greu, les tècniques per llegir ADN són cada cop més ràpides i barates, però si aquest format d'emmagatzematge ha de substituir els actuals, com molts auguren, el fet d'haver de descodificar tots els arxius per poder llegir-ne un de concret no resulta gens pràctic.

L'abril passat, un grup format per investigadors de la Universitat de Washington i per Microsoft Research van idear una tècnica que permet accedir aleatòriament a qualsevol dels arxius codificats. Amb aquest mètode, cada arxiu contingut en l'ADN té un codi que funciona com una etiqueta i que permet destriar-lo dels altres. El grup també va marcar un rècord emmagatzemant 200 megabytes d'informació i a més ho va fer aprofitant molt més la capacitat de la molècula. Amb la seva tècnica, van dir, tot l'internet accessible podria cabre en una capsa de sabates.

A principis de març va haver-hi un altre punt d'inflexió. Yaniv Erlich, enginyer informàtic de la Universitat de Colúmbia, i Dina Zielinski, del Centre Genòmic de Nova York, van anunciar un sistema que augmenta en un 60% la densitat d'emmagatzematge respec-

te dels mètodes previs. Concretament, van poder codificar 215 petabytes —mil bilions de bytes— per gram d'ADN, una xifra que representa un 85% del límit teòric. Un disc dur d'un terabyte pesa uns 150 grams. Doncs bé, amb la nova tècnica, en un sol gram d'ADN es podrien guardar 215.000 cops més d'informació. *Science* utilitza una analogia molt gràfica: amb aquesta densitat es podria emmagatzemar tota la informació generada pels humans en un parell de camionetes plenes d'ADN.

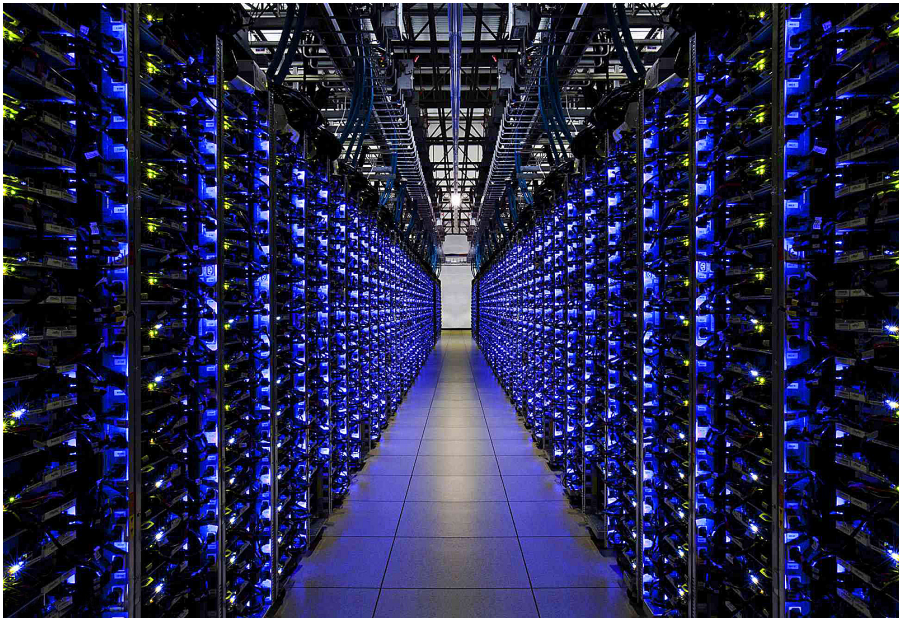
És molt impressionant, però malauradament encara no és viable substituir els *racks* amb discs durs dels centres de dades per recipients amb ADN. El problema principal, com sempre, és el preu. Llegir ADN i fer-ne còpies és ràpid i no és excessivament car. El que costa és fabricar-ne la primera mostra. Crear ADN sintètic és lent i car. Cada cop menys però encara massa lent i massa car per generar-lo a gran escala. Perquè es facin una idea, sintetitzar els dos megabytes d'informació dels arxius que es van codificar en aquest darrer projecte, va costar set mil dòlars i es va trigar dues setmanes. Segons va dir Zielinski en un xat de preguntes i respostes, en uns deu anys el cost del procés ja serà viable i l'ADN podrà començar a competir amb els suports electrònics. →



TARA BROWN PHOTOGRAPHY/UNIVERSITY OF WASHINGTON

Tota la informació de 600 mòbils intel·ligents (uns 10.000 gigabytes), pot ser emmagatzemada en la brossa d'ADN que hi ha en aquest tub d'assaig.

Empreses i hospitals tenen grans centres de dades per emmagatzemar la informació que genera la seva activitat. Quan sigui econòmicament viable guardar dades en molècules d'ADN, els magatzems d'informació seran molt més petits i molt més barats de mantenir.



amb la radiació ultraviolada —per això cal mantenir-lo en un lloc fosc—, no es veu afectat pel camp magnètic.

Hi ha un altre avantatge clau. Aquest mètode d'emmagatzematge, a diferència de la resta, mai no quedarà obsolet. Actualment seria molt complicat poder trobar un suport capaç de llegir, per exemple, un disquet o un vídeo VHS. En canvi, sempre hi haurà interès per llegir seqüències d'ADN. Per tant, sempre hi haurà màquines capaces de fer-ho. Mentre siguem de naturalesa biològica, el format ADN no passarà mai de moda.

Més enllà de la memòria, fa més de vint anys que s'estudien altres possibles aplicacions computacionals de l'àcid desoxiribonucleic. Amb aquesta molècula ja s'han pogut fer, per exemple, versions senzilles de portes lògiques i de circuits, que són els pilars de la computació. La programació amb ADN, realitzada per primer cop l'any 1994 per Leonard Adleman, ha demostrat que es poden utilitzar les reaccions químiques de l'ADN per resoldre problemes matemàtics complexos. La revolució científica i tecnològica que va desencadenar el Projecte Genoma Humà, doncs, no només pot servir per entendre'ns, arreglar-nos i modificar-nos sinó també perquè els humans puguem aprofitar el llenguatge de la vida per fer les nostres pròpies creacions. ●

→ **Eternament modern**

En els últims dos anys, destaca *Science*, s'ha creat més informació que la generada al llarg de la història de la informàtica. Amb l'Internet de les Coses la informació augmentarà de manera exponencial i serà imprescindible optimitzar els sistemes d'emmagatzematge. L'ADN és probablement el millor candidat per fer front a aquest repte.

Com hem dit, pot encabir molta informació en un espai molt reduït i està

fet a prova de bomba. Si es conserva en un lloc fred, fosc i humit, pot durar més, molt més, que els suports actuals. Pensem en un CD, per exemple, que es fa malbé en uns quants anys. Fins i tot les cintes magnètiques s'espantlen en 15 o 20 anys. Per contra, s'han trobat mostres d'ADN en bon estat de mamuts morts fa 20.000 anys i fins i tot s'han pogut llegir seqüències genètiques d'una espècie de cavall de fa aproximadament 700.000 anys. A més, si bé l'ADN pot fer-se malbé

Avancem junts



Port de Barcelona

Port de Barcelona, el primer *hub* logístic del sud d'Europa.

www.portdebarcelona.cat
