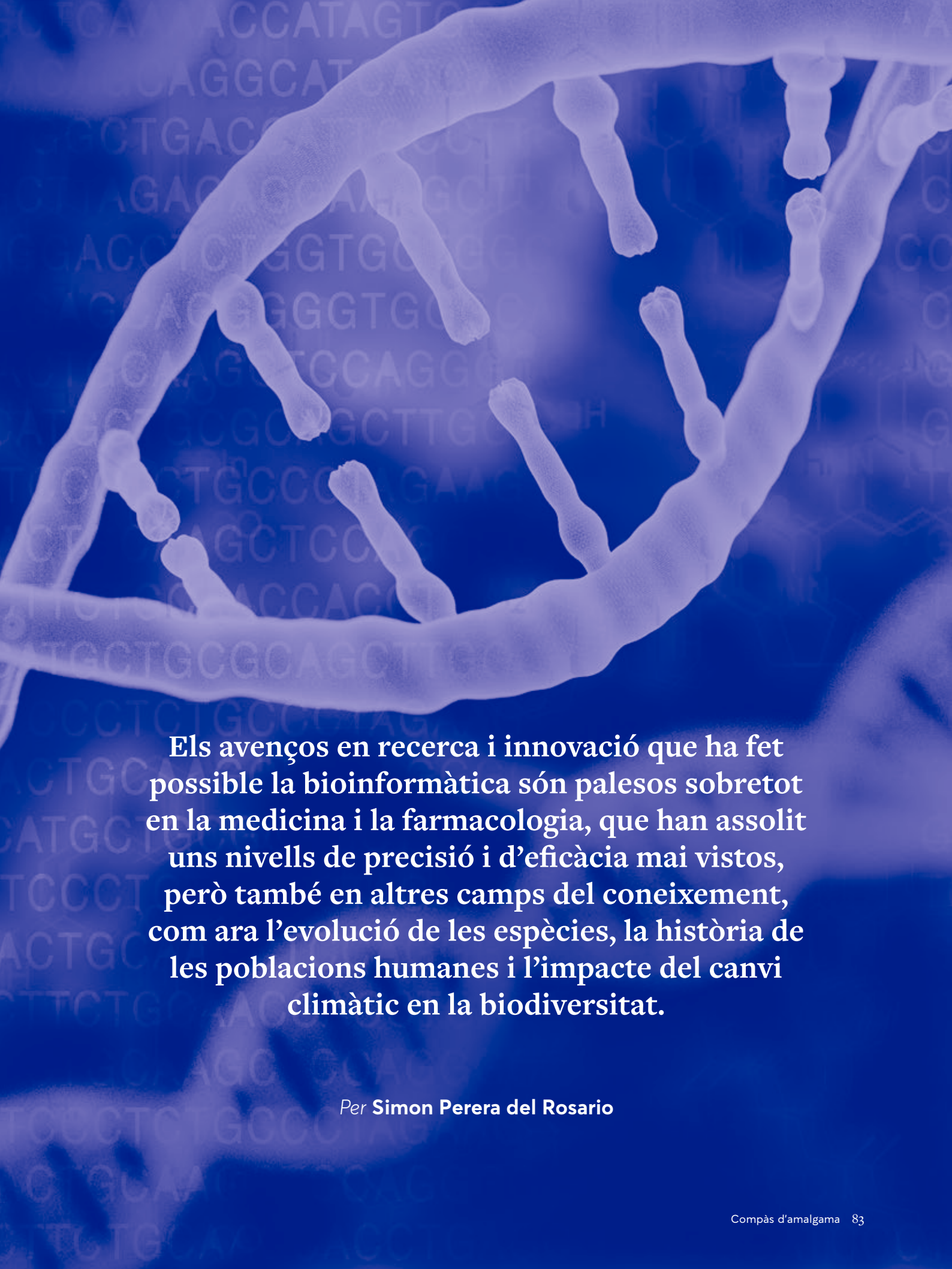


# Bioinformàtica: llegint el llibre de la vida per viure més i millor



**Els avenços en recerca i innovació que ha fet possible la bioinformàtica són palesos sobretot en la medicina i la farmacologia, que han assolit uns nivells de precisió i d'eficàcia mai vistos, però també en altres camps del coneixement, com ara l'evolució de les espècies, la història de les poblacions humanes i l'impacte del canvi climàtic en la biodiversitat.**

*Per* **Simon Perera del Rosario**



## «La bioinformàtica uneix l'essència de dues disciplines: és part d'ambdues sense ser-ne ben bé cap.»

**Hermafrodit era un déu grec**, fill d'Afrodita i Hermes. És únic en la mitologia grega per tenir caràcters sexuals tant femenins com masculins, arran de la seva unió miraculosa amb la nàiada Salmacis. No només era un símbol del matrimoni, sinó que se'l considerava un déu de la fertilitat.

Com Hermafrodit, la bioinformàtica uneix l'essència de dues disciplines: és part d'ambdues sense ser-ne ben bé cap. Són, és clar, la biologia i la informàtica. Quan parlem de bioinformàtica, fem referència a aplicar els coneixements informàtics per crear eines que ens permetin analitzar les dades de tipus biològic. Computació aplicada a la biologia.

La biologia i la informàtica estudien entitats extraordinàriament complexes creades, paradoxalment, a partir d'*alfabets* molt senzills. Pensem en els milions de dispositius interconnectats que ens permeten accedir en mil·lisegons a una quantitat d'informació incommensurable, o a les xarxes tròfiques que connecten de manera complexa milions d'organismes en els ecosistemes terrestres. La informàtica fa tot això partint de dues unitats bàsiques d'informació: el 0 i l'1. La biologia ho fa amb quatre *lletres*, que simbolitzen els quatre tipus de molècula que formen l'ADN: A, C, G i T.

Cadascuna de les nostres cèl·lules no sexuals conté quaranta-sis cromosomes, és a dir, *macromolècules* d'ADN, formades per una successió d'aquestes *lletres*: ...AGCA-TAAC... (Les cèl·lules sexuals, oòcits-òvuls i espermatozous, en tenen la meitat cadascuna, i arriben a quaranta-sis quan s'ajunten en la fecundació.) El *llibre* de la nostra informació genètica, el nostre *genoma*, té tres mil milions de lletres a cada cèl·lula. El canvi d'una lletra per una altra pot no tenir cap efecte, o pot provocar efectes tan devastadors com la mort o la malaltia. Les possibilitats intermèdies són incomptables. D'una cèl·lula a una altra —o d'una persona a una altra— poden variar no només lletres, sinó *paraules* (grups de tres lletres), frases, paràgrafs o pàgines senceres, amb conseqüències de tot tipus. Els genomes de dues persones qualssevol es dife-

rencien, de mitjana, en un 0,5%: vint milions de lletres de diferència.

Per *fer una persona*, les instruccions genètiques emmagatzemades a l'ADN es transcriuen en un alfabet lleugerament diferent, l'ARN, una molècula relacionada amb l'ADN i probablement ancestral a totes les formes de vida que coneixem. L'ARN es tradueix en proteïnes, que són la major part de molècules efectores de la cèl·lula, és a dir: molècules que fan coses. Les proteïnes actuen sobre tots els components del nostre organisme, entre els quals hi ha les molècules més senzilles o *metabòlits*, que també es veuen afectats directament per la interacció amb l'entorn que es dona en activitats com l'alimentació o la secreció. A més, en les últimes dècades hem après que la informació genètica pot ser regulada per mecanismes més o menys heretables, d'una manera que anomenem epigenètica. En tots aquests nivells poden produir-se errors de còpia o variacions a partir de la informació genètica, que afegeixen una gran complexitat molecular als maons que fan les nostres cèl·lules, teixits i òrgans.

La quantitat d'informació biològica que hi ha en una sola persona és aclaparadora. Al principi, cada lletra de la seqüència d'ADN s'havia de *llegir* manualment. Fa alguns anys, estudiar la seqüència d'un gen podia ser la feina d'una tesi doctoral, o de tota una vida. Els mètodes de seqüenciament han millorat en paral·lel a la capacitat i els mètodes de computació. L'any 2001 va poder llegir-se tot el genoma humà, fruit d'un projecte que havia durat més de deu anys i havia costat més de cinc mil milions de dòlars. Avui, seqüenciar el genoma d'una persona costa al voltant de mil euros.

Tot plegat, aquest *big data* biològic necessita la computació per dissenyar eines específiques que permetin emmagatzemar, tractar i analitzar les dades. D'això parlem quan diem bioinformàtica. I atès que les preguntes que poden plantejar-se des de la biologia són molt diferents, la bioinformàtica ens permet atènyer i aplicar coneixements variats, que van des del coneixement de la nostra pròpia història fins al desenvolupament de medicaments i vacunes,

passant per la conservació dels ecosistemes o l'anàlisi forense. La bioinformàtica és, en definitiva, el motor de gran part de la recerca i la innovació que es fan en disciplines relacionades amb la biologia, com ara la medicina.

\*

Quan un organisme es reproduïx, la seva descendència rep una còpia del seu ADN, sola o en combinació amb l'ADN d'un altre organisme (en el cas de la reproducció sexual, com és la de l'ésser humà). Les proteïnes que copien l'ADN són molt fidels a l'original, però fan un error cada deu mil o cent mil lletres (a banda d'errors més greus que poden donar-se per «accidents moleculars» durant la còpia de l'ADN i la divisió cel·lular). Així, cada cèl·lula porta milers de lletres diferents de les seves cèl·lules mare.

Comparant l'ADN de les cèl·lules filles, podem esbrinar com era l'ADN original. Us convido a fer un exercici amb paper i llapis: si l'ADN dels descendents és ...AATA..., ...ACAA... i ...AAAG..., quina és la seqüència més probable en l'ADN ancestral? La resposta és ...AAAA..., ja que la lletra A es troba, per cada posició variant, en la majoria de les seqüències descendents.

Comparant les seqüències d'espècies, poblacions o individus diferents, podem descobrir la relació que tenen entre si i, per tant, respondre a preguntes sobre l'evolució de les espècies, sobre la història de les poblacions humanes o sobre l'origen dels individus. Així, hem arribat a descobrir que tots els éssers vius provenen d'un únic ancestre comú, LUCA (*last universal common ancestor*); o que l'*Homo sapiens* (la nostra espècie) s'ha hibridat amb l'*Homo neanderthalensis*, els humans de Denisova, i fins i tot poblacions humanes que encara no coneixem i de les quals només tenim constància gràcies a aquestes anàlisis bioinformàtiques. Aquestes mateixes tècniques de comparació de seqüències són les que fem servir rutinàriament en les anàlisis de paternitat o en les proves d'ADN forenses (a partir del famós *pèl* o una altra mostra biològica de l'investigat).

\*

La bioinformàtica no només ens aporta coneixement: fa dècades que és una de les peces clau en el descobriment i desenvolupament de fàrmacs, i és un dels pilars fonamentals de la medicina actual i la medicina del futur, el que anomenem medicina 4P: predictiva, preventiva, personalitzada (o de precisió) i participativa.

Podem dir amb seguretat que li devem milions de vides. En molts casos, ha permès identificar els gens causants (de forma total o parcial) d'una malaltia, la qual cosa ha permès identificar molècules adequades per modificar la funció patogènica d'aquests gens; és a dir, nous medicaments, gràcies a tècniques com l'acoblament molecular o la dinàmica molecular. Altres vegades ens ha fet possible entendre per què un fàrmac útil per a algunes persones no ho era per a d'altres, i posar-hi remei. Des de fa alguns anys, alguns càncers ja no es tracten segons les seves manifestacions, sinó segons les seves causes moleculars, identificades i diagnosticades gràcies a la bioinformàtica. Aquest camp es coneix com a farmacogenòmica. Mitjançant els estudis de relació quantitativa estructura-activitat (QSAR), també podem predir tota una sèrie de característiques dels compostos a partir de la seva estructura molecular: propietats físicoquímiques, biològiques, farmacològiques, toxicològiques i ecotoxicològiques que ens permeten estalviar costos, temps i assajos animals durant la recerca per trobar nous fàrmacs. La bioinformàtica també aprofita els fàrmacs ja existents per tractar noves malalties (reposicionament de fàrmacs), i moltes altres aplicacions mèdiques.

Amb la medicina 4P, la bioinformàtica permet predir el risc de cada persona a tenir una determinada malaltia. Per exemple, sabem que les persones amb una sèrie de versions específiques en gens del cromosoma 3 són les que tenen més risc de patir una versió greu, amb hospitalització, de COVID-19 (medicina predictiva). També podem detectar-ne els casos positius amb tests cada cop més senzills, que permeten prendre mesures preventives que impedeixen l'expansió del virus. Atès que part de la gravetat de la malaltia COVID-19 es deu a una resposta immune inadequada en algunes persones, una de les lí-

## «Llegir el llibre de la vida, això que ens permet la bioinformàtica, ens fa créixer com a humanitat.»

nies de recerca contra el virus se centra a buscar els fàrmacs més adequats per a cada persona, segons les seves característiques genètiques o fisiològiques, com ja es fa amb èxit fa anys en altres malalties com el càncer (medicina personalitzada o de precisió).

Per acabar, la bioinformàtica és participativa en almenys dos aspectes. D'una banda, per poder trobar informació aplicable de manera personalitzada cal estudiar les variants gèniques menys freqüents i, per tant, la col·laboració de *donants* que comparteixin les seves dades amb els investigadors (regits per principis bioètics com la confidencialitat, l'anonimització, la beneficència, etc.). D'altra banda, posa a disposició de les persones eines que els permeten conèixer de primera mà els seus condicionants biològics (com els graus de risc genètic) i actuar en conseqüència.

\*

La bioinformàtica no beneficia només els humans. Les seves aplicacions van més enllà, i permeten guiar els esforços de conservació de les espècies, de manera que aquestes tinguin més possibilitats de sobreviure als perills d'extinció. Per exemple, el Projecte BioGenoma de la Terra estudia els genomes de tota la biodiversitat eucariota del nostre planeta per conèixer l'impacte del canvi climàtic sobre la biodiversitat, promoure la conservació de les espècies i ecosistemes en perill d'extinció, entre d'altres.

Fins i tot es parla de *desextinció*: tornar espècies extintes a la vida. Això es va aconseguir per primer cop al món (tot i que amb resultats limitats) amb la cabra pirinenca, i s'estudia fer-ho amb altres espècies com el mamut llanut o l'ur.

\*

La història d'Hermafrodit, llegida amb els ulls contemporanis, té els seus grisos: els déus en van canviar el cos en contra de la seva voluntat. La bioinformàtica, com Hermafrodit, té com un dels seus grans reptes la col·laboració entre disciplines (biologia, biomedicina, biotecnologia..., computació, tecnologies de la informació, informàtica mèdica...), entre ciutadans (les dades biològiques personals són el seu *combustible*), entre entitats, regions i països (per tenir sistemes interoperables de recollida de dades, que evitin malgastar temps i recursos), etc. A casa nostra, hi treballen centenars de professionals en entitats com Bioinformàtics Barcelona, l'Institut Nacional de Bioinformàtica o l'Institut Europeu de Bioinformàtica.

Els gens van ser descoberts l'any 1866 i oblidats fins a l'any 1900. L'estructura de l'ADN va trigar cinquanta-tres anys més a ser descoberta. En aquests setanta anys, la bioinformàtica ha estat un dels motors de l'avenç de la medicina i ens ha fornint un profund coneixement sobre la història dels éssers vius (i sobre nosaltres mateixos). Avui en dia, noves formacions en bioinformàtica, com el grau i el doctorat, s'implanten cada any, i els centres de recerca i hospitals demanen bioinformàtics i tècnics en bioinformàtica. Si llegir ens fa créixer com a persones, llegir el *llibre de la vida*, això que ens permet la bioinformàtica, ens fa créixer com a humanitat. ●