

# L'evolució humana. L'evidència genòmica

David Comas i Francesc Calafell

Institut de Biologia Evolutiva (CSIC-UPF) i Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona

Correspondència: [francesc.calafell@upf.edu](mailto:francesc.calafell@upf.edu)

DOI: 10.2436/20.1501.02.202  
ISSN (ed. impresa): 0212-3037  
ISSN (ed. digital): 2013-9802  
<http://revistes.iec.cat/index.php/TSCB>  
Rebut: 05/01/2021  
Acceptat: 21/01/2021

## Resum

A partir de Darwin, entenem el procés que genera la diversitat biològica, inclosa la humana, com un arbre genealògic que relliga tots els individus i totes les espècies. Tenint present aquest marc d'anàlisi, i comparant els genomes a diferents escales —entre individus de la mateixa o diferents poblacions o espècies—, podem reconstruir aquest procés, és a dir, conèixer la història dels humans com a espècie o com a conjunt de poblacions. Veurem com, més enllà del moment actual, l'anàlisi del DNA antic ha permès una perspectiva diacrònica en què ja no veiem només les puntes de les branques, sinó les branques senceres. Il·lustrarem aquesta perspectiva amb les relacions verticals i horitzontals entre humans, neandertals i denissovens —i altres homínids que poden haver contribuït als nostres genomes actuals. Dins de la nostra espècie, donarem exemples de la història més profunda, com l'origen africà de la humanitat, però també de la més recent, com la reconstrucció de la diàspora gitana.

**Paraules clau:** humans anatòmicament moderns, neandertals, denissovens, introgressió, genomes.

## L'arbre genealògic de la humanitat

Sovint es resumeix el pinyol de la teoria darwiniana de l'evolució com la suma de dues idees: l'herència amb modificació més la supervivència diferencial segons l'herència rebuda. En aquest capítol veurem com la primera idea —la combinació de la genealogia i la mutació— ens ha permès entendre l'origen dels humans com a espècie i la seva ulterior evolució, mentre que remetem amablement el lector al text d'Elena Bosch en aquest mateix monogràfic, on explora com la segona idea, la selecció natural llegida en el genoma, ens dona pistes d'allò que ens fa humans, o almenys d'allò que ens ha permès sobreviure fins al present.

El 1837, al quadern B de la sèrie sobre la transmutació de les espècies, Darwin va dibuixar l'esquema bàsic de qualsevol arbre filogenètic (Figura 1), com a relació genealògica entre espècies, en una imatge que ha esdevingut icònica però que també ens va llegar un marc sobre com pensar en l'origen de les espècies i en l'acumulació de la variació que contenen. És sobre aquesta idea que es basa la filogenètica actual.

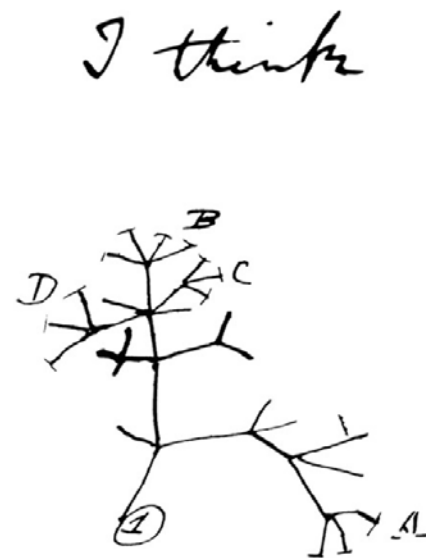
<sup>1</sup> Evidentment, hi ha una excepció flagrant en la transferència genètica horitzontal, que es pot donar a moltes escales, entre espècies llunyanes com en bacteris, o en situacions que ens poden fer replantejar el mateix marc de referència d'allò que considerem espècies diferents, com és el cas de la introgressió d'humans arcaics, que tractarem en aquest capítol.

## Abstract

Since Darwin, we understand the process that generates biological diversity, including human diversity, as a genealogical tree that links all individuals and species. Within this analytical framework, and by comparing genomes at different scales (between individuals of the same or different populations or species) we can reconstruct this process, that is, we can understand the history of humans as a species. We will see how, beyond the present, the analysis of ancient DNA has allowed a diachronic perspective in which we no longer see only the tips of the branches, but the whole branches. We will illustrate this perspective with the vertical and horizontal relationships between humans, Neanderthals, and Denisovans (and other hominins that may have contributed to our genomes). Within our species, we will provide examples of the deepest history, such as the African origin of humankind, but also of the most recent, such as the reconstruction of the Roma diaspora.

**Keywords:** anatomically modern humans, Neanderthals, Denisovans, introgression, genomes.

Qualsevol genoma de qualsevol ésser viu és el resultat de l'acumulació de la variació al llarg de la història genealògica que porta des de l'origen de la vida fins al present. Però el camí que porta des de l'inici fins a cada organisme és propi i específic de cada individu: és la seva genealogia. És a partir d'aquest arbre que s'acumula la variació genòmica<sup>1</sup> en les seves múltiples formes: mutacions de canvi de nucleòtid, o de canvis de la mateixa estructura del genoma en escales que van de les poques bases als cromosomes sencers, o de salts d'elements que són, o havien estat, o semblen virus. Cadascun d'aquests esdeveniments planta una bandera en una branca de l'arbre: tots els descendents d'aquella branca portaran aquest nou estat en els seus genomes —tret que es doni la mutació que reverteixi el genoma al seu estat original, cosa que és molt poc probable. La comparació de diferents genomes ens permet trobar aquests senyals compartits i reconstruir l'arbre genealògic del nostre objecte d'estudi.



↑ Figura 1. Esquema d'arbre filogenètic que Darwin va dibuixar el 1837 al quadern B de la sèrie sobre la transmutació de les espècies. Per cert, aquest quadern es troba desaparegut, possiblement robat, des del 2001.

Però els arbres són també un magnífic exemple de fractals, una de les formes bàsiques que trobem en els éssers vius, tal com va sistematitzar i entendre l'enyorat Jorge Wagensberg. Una característica única de les fractals és la seva invariància d'escala, cosa que també s'aplica als arbres filogenètics. Imaginem una família que busca un dels seus membres entre els enterrats en les fosses comunes de la Guerra Civil: caldrà comparar els seus genomes —o una petita part— amb el d'un cadàver prometedor per verificar si la genealogia genètica que els relliga amb el cadàver coincideix amb la genealogia familiar. D'alguna manera, encara que calgui adaptar les tècniques, la mateixa lògica s'empra per decidir quin arbre explica millor les relacions més bàsiques de la vida, com el que pugui relligar aquesta família d'*Homo sapiens* amb les malves, els crisantems, els fongs i els bacteris que han contribuït a descompondre el cadàver o els arqueus que podien haver viscut en l'intestí del finat. Dedicarem aquest capítol, però, als arbres que relliguen els humans entre ells i amb els seus parents morts més propers, neandertals i denissovens.

Però l'acumulació de la diversitat ens dona més que la forma de l'arbre. Gràcies al fet que coneixem la velocitat amb què s'acumula aquesta variació, podem donar a aquest arbre una escala temporal que la mateixa forma de l'arbre, com hem vist, no ens dona. Disposem, doncs, d'un rellotge molecular, per bé que potser funcionaria millor, vista la seva diversitat, la comparació amb un conjunt de metrònoms que marquen ritmes molt diferents. Sabem posar el rellotge molecular a l'hora de diferents maneres, segons el cas. Bàsicament, podem observar la mutació en directe en microorganismes, comptant mutacions i dividint entre el temps transcorregut; fins i tot, podem adaptar aquesta estratègia per a certes formes de variació en el genoma humà, en què podem veure i comptar mutacions si observem centenars o milers de meiosis, sigui comparant pares i fills, sigui analitzant pedigrís ben documentats com els d'Islàndia o Flandes. Però la majoria de mutacions són massa lentes per ser observades en directe; llavors, cal recórrer a la paleontologia per poder-ne calibrar el rellotge molecular. Si el registre fòssil ens proporciona una data més o menys clara per a la separació entre dues espècies,

llavors el nombre de canvis acumulats entre ambdues dividit pel temps transcorregut ens dona la taxa de substitució<sup>2</sup>.

En el moment de desenvolupament tecnològic actual, es poden generar un gran nombre de genomes humans complets amb un esforç econòmic raonable. Disposem de milers de genomes seqüenciats amb una alta qualitat, amb projectes que assegurin una representació geogràfica adequada de l'espècie humana. A més, disposem d'encara més informació gràcies a estratègies que ens donen una aproximació a la diversitat nucleotídica, com són el genotipat d'*arrays* d'SNPs (polimorfismes d'un nucleòtid), en què s'interroguen fins a 1-2 milions de posicions escampades per tot el genoma, o la seqüenciació de l'exoma, en què es produeix la seqüència de la gran majoria de les regions del genoma que codifiquen per a proteïnes. El resultat és que disposem de moltíssims genomes caracteritzats d'una manera o altra. Això ens ha permès passar de les anàlisis basades en un sol locus —un sol gen, el cromosoma Y, el DNA mitocondrial—, o fins i tot en una sola posició, a entendre la variació simultàniament en tot el genoma. A partir de la comparació de molts genomes d'origens poblacionals diversos som capaços de reconèixer seqüències parcials (haplotips) més freqüents en unes poblacions que en d'altres. Això ens permet quantificar el flux gènic entre poblacions, o per a un cert individu, «pintar-ne» els cromosomes com un mosaic d'origens poblacionals diferents. I més enllà del flux gènic entre poblacions actuals, també som capaços de reconèixer regions del genoma introgressades d'humans arcaics, com explicarem més endavant. Per tant, som capaços també de desplegar l'arbre genealògic dels humans en l'espai —en diem *filogeografia*—, i de reconèixer cadascun dels nostres genomes com un palimpsest de múltiples orígens i connexions.

L'altra gran pota que sustenta la comprensió actual de l'evolució humana és la disponibilitat d'un gran nombre de genomes<sup>3</sup> antics. Aquesta ha estat una gesta tecnològica increïble que, amb gran dificultat, ens ha permès allò que no fa pas tants anys ens semblava impossible: disposar de genomes antics d'humans i dels nostres parents més propers, amb una cobertura geogràfica que abasta els cinc continents. Val a dir, però, que,

sigui per la facilitat de conservació del DNA en climes més freds, sigui per l'interès eurocèntric o la disponibilitat d'antropòlegs, arqueòlegs i genetistes sobre el terreny, Europa està sobrerrepresentada també en el camp del DNA antic. Si entenem per DNA antic aquell provinent d'un difunt, disposem d'una àmplia representació temporal, des de fa 50.000 anys fins a èpoques històriques ben recents. Per tant, ara és possible fer una reconstrucció diacrònica de l'evolució humana, com fan Olalde *et al.* (2019) en el cas de la península Ibèrica, des del mesolític fins als musulmans. Per tant, per bé que l'observació de la diversitat contemporània ha permès reconstruir la història de les poblacions humanes, ara aquesta aproximació es pot considerar com una font d'emissió d'hipòtesis que sovint, però no sempre, podem verificar acudint al DNA dels nostres avantpassats.

Més enllà de la nostra espècie en el sentit més estricte, la disponibilitat de genomes de neandertals i de denissovens ha revelat inapel·lablement l'arbre genealògic que ens lliga, com també ens ha descobert la contribució lateral o introgressió dels altres genomes humans cap al nostre. Però, sens dubte, la capacitat que tenim ara de situar genomes reals, no inferits, en les branques interiors dels arbres filogenètics és una prova espectacular i fefaent de l'evolució humana.

### El tangled bank de l'evolució humana

En aquesta secció veurem com vam poder reconstruir l'arbre recent de l'evolució humana, i com vam descobrir que, en realitat, l'evolució humana recent és més un arbust de branques entrecruades que no pas una successió de fissions que donin lloc a branques estanques entre elles. Tant la primera seqüència del DNA mitocondrial d'un neandertal que es va publicar, la d'un fòssil de Vindija (Croàcia) de fa 38.000 anys (Green *et al.*, 2008), com el primer genoma sencer, el d'una mostra d'Altai (Sibèria) de fa 50.000 anys (Green *et al.*, 2010; Prüfer *et al.*, 2014) van donar uns resultats clars: els llinatges dels neandertals i els humans anatòmicament moderns van divergir fa uns 500.000 anys. Aquesta és la darrera bifurcació de l'arbre de la vida que porta cap a nosaltres.

Però ja des de l'obtenció del primer esborrany de genoma de neandertal, els autors

<sup>2</sup> Que no és el mateix que la taxa de mutació, ja que hi pot haver hagut mutacions paral·leles, o mutacions revertides, que no donen canvis observables en comparar dues espècies.

<sup>3</sup> Utilitzem aquí *genomes* en el sentit de seqüències del genoma humà parcialment caracteritzades, ja sigui seqüenciant a baixa cobertura, ja sigui capturant milers d'SNPs al llarg de tot el genoma.

van notar un fet inesperat: els neandertals compartien més al·lels derivats amb els europeus i els asiàtics actuals que no pas amb els africans. En comparar les seqüències humanes, de neandertals i de ximpanzé, sovint europeus, asiàtics i neandertals portaven el mateix al·lel, que era diferent del d'africans i ximpanzés. Podria ser que hi hagués mutacions que revertissin a l'estat ancestral, o que s'haguessin donat en paral·lel en la branca neandertal i en la branca que porta a europeus i asiàtics. Però no hi ha cap raó per la qual aquestes reversions o mutacions paral·leles no es donessin també, i en la mateixa mesura, en africans. Els autors proposen una explicació diferent: hi va haver *introgressió* entre neandertals i humans anatòmicament moderns; els avantpassats dels europeus i els asiàtics actuals i els neandertals es van trobar, es van encreuar i van tenir descendència fèrtil fa uns 50.000 anys, probablement al Pròxim Orient, on hi ha evidència arqueològica de la presència d'humans anatòmicament moderns i de neandertals en marcs temporals propers.

La introgressió provoca una excepció en un model filogenètic rígid: segons quines regions del genoma ens mirem, obtindrem arbres filogenètics diferents. Tanmateix, es tracta d'excepcions, i, a partir de la majoria del genoma, podem obtenir un arbre de consens clar, que és, precisament, allò que ens permet detectar aquestes divergències que interpretem com a resultat de la introgressió. Però, en si mateixa, la introgressió neandertal no és cap raresa, i en altres grups d'espècies properes<sup>4</sup> se n'han descrit força exemples, com entre ximpanzés i bonobos (De Manuel *et al.*, 2016). Vegeu Gokcumen (2019) per a una revisió sobre introgressió entre arcaics i humans.

Les seqüències introgressades pateixen una sèrie de processos que n'alteren el destí. D'entrada, la recombinació fa que els haplotips introgressats s'escurcin progressivament, en recombinar-se amb seqüències d'humans anatòmicament moderns. Aquest procés es pot usar per datar la introgressió i per distingir possibles esdeveniments d'introgressió independents. Per exemple, donada la proporció de DNA neandertal que conté i la llargada dels fragments, s'estima que unes restes trobades a Peștera cu Oase (Romania), i que morfològicament semblen intermèdies, corresponen a un humà que tenia un rebesavi

neandertal, mentre que la resta dels seus avantpassats eren anatòmicament moderns.

Però el factor més important per a la pervivència de les seqüències introgressades és, com Darwin sabia molt bé, l'efecte sobre la supervivència dels portadors. Aquests efectes poden abastar tota la gamma de possibilitats, des de ser incompatibles amb la vida a donar un avantatge selectiu, passant per la neutralitat pura. Així, no s'han trobat ni cromosomes Y ni genomes mitocondrials (mtDNA) neandertals en humans. La manca de recombinació i la reduïda grandària efectiva poblacional d'aquestes regions genòmiques en poden explicar la pèrdua purament per deriva, o per qualsevol mala adaptació en qualsevol locus de l'Y o del mtDNA. Tanmateix, en mapar les regions introgressades neandertals en humans se n'observa un dèficit significatiu al cromosoma X, cosa que suggeriria algun tipus d'incompatibilitat reproductiva parcial entre humans i neandertals. En general, les seqüències neandertals són menys freqüents en les regions codificants i reguladores del genoma humà, i s'infereix que la introgressió podria haver alterat les interaccions epistàtiques existents dins de cada espècie. La variació introgressada neutra, per definició, ha seguit el mateix camí que la variació pròpiament humana, canviant de freqüència, o fins i tot extingint-se, amb la mateixa probabilitat. Però hi ha exemples també de variació introgressada que va resultar adaptativa en els humans; l'exemple més espectacular i que veurem més endavant, però, prové dels denissovans. La majoria de casos d'introgressió adaptativa es refereix al sistema immunitari (haplotips d'HLA) i a la interacció amb patògens com virus d'RNA, incloent-hi SARS-CoV-2!

### L'humà sense rostre

La capacitat de descriure la diversitat genètica existent en un genoma provinent d'una resta antiga i de comparar-la en un context filogenètic —és a dir, dins d'un arbre genealògic— ha permès el descobriment d'un nou grup d'humans, els denissovans. La cova de Denissova, a Sibèria, conté algunes restes òssies que anatòmicament són neandertals, i d'altres que són massa petites per identificar-les amb certesa. La seqüenciació del mtDNA extret d'una falange datada fa 50.000-80.000 anys va proporcionar una sorpresa majúscula:

no entrava dintre de la variació neandertal coneguda, sinó que havia divergit dels neandertals i els humans fa un milió d'anys (Krause *et al.*, 2010). L'anàlisi del genoma nuclear complet, que es va publicar poc després (Reich *et al.*, 2010), va precisar que els denissovans són filogenèticament més propers als neandertals, dels quals es van separar fa uns 600.000 anys, mentre que els avantpassats dels humans es van separar de la branca neandertal-denissovana fa 800.000 anys (i no fa 500.000 anys com es va deduir en comparar només humans i neandertals). En el mateix article, es confirmava genèticament com a denissovà un queixal trobat a la mateixa cova. Donada l'escassetat de restes òssies disponibles, la definició de què és un denissovà es basava primàriament en el seu genoma; per primer cop, es definia una espècie<sup>5</sup> de mamífer no a partir de la morfologia o del comportament, sinó del seu genoma. I aquest ha estat l'únic criteri de què disposem per identificar restes denissovanes; més enllà d'altres fragments d'ossos de la mateixa cova de Denissova, l'únic espècimen denissovà trobat és la mandíbula de Xiahe, al Tibet, per bé que la identificació es basa no en el mateix genoma sinó en la composició proteica del col·lagen extret de la mandíbula, i en la presència de seqüències denissovanes de DNA en mostres ambientals de la cova de Xiahe (vegeu l'article de María Martínón en aquest mateix monogràfic). Els patrons de metil·lació en un conjunt de gens relacionats amb el desenvolupament i la morfologia craniofacial han proporcionat més informació sobre l'aspecte dels denissovans que no pas les escasses restes òssies que se'ls poden atribuir.

Els denissovans també es van encreuar amb els humans i amb els neandertals; en el darrer cas, un dels ossos trobats a Denissova corresponia a una noia de tretze anys, el genoma de la qual indica un pare denissovà i una mare neandertal. En el cas dels humans, la presència de seqüències denissovanes és major al sud-est asiàtic i especialment entre els melanesis; és possible que els primers humans anatòmicament moderns trobessin els denissovans en la seva colonització del sud-est asiàtic i de Sahul. Devem als denissovans un exemple espectacular d'introgressió adaptativa: els tibetans actuals porten un haplotip d'origen denissovà al gen EPAS1 —un factor de transcripció induïble per la hipòxia— que els permet l'adaptació a l'altura.

<sup>4</sup> De fet, la introgressió fa que el mateix concepte d'espècie esdevingui més vague i imprecís.

<sup>5</sup> Svante Pääbo, que va dirigir aquesta recerca, prefereix no parlar d'«espècie» sinó de «grup d'hominins».

## Fantasmes del passat

Acabem de veure com, a partir de molt poc material biològic, podem inferir la presència d'una nova espècie. Hem anat un pas més enllà: en diversos casos, reconeixem humans arcaics a través de la petjada que han deixat en els nostres genomes, sense que puguem atribuir aquests genomes arcaics a cap resta biològica concreta. I tot plegat és possible gràcies al marc teòric que ens va llegar Darwin, bàsicament reconstruint els camins filogenètics que han seguit els nostres genomes. En efecte, podem proposar models hipotètics per a la reconstrucció de la història que ha generat la diversitat humana. Aquests models inclouen la seqüència de fissions i fusions (introgressió, flux gènic), els temps en què van ocórrer aquests esdeveniments i la grandària de les poblacions implicades. Podem proposar models alternatius, que, per exemple, incloguin o no la presència i la contribució d'humans arcaics desconeguts («poblacions fantasma»). Per tal de decidir entre diferents models, els implementem en forma de simulacions i comptem quantes vegades ens donen resultats raonablement semblants a la diversitat actual, descrita a través d'un conjunt d'estadístics. Aquell model que més sovint s'acosti a la diversitat actual serà més probablement aquell que s'esdevingué en realitat. Aquesta aproximació, anomenada computació bayesiana aproximada (ABC), depèn d'una munió de paràmetres i implica prendre un gran nombre de decisions a priori arbitràries; per això, Mondal *et al.* (2019) van acoblar *deep learning* a ABC i van inferir que una població arcaica desconeguda i diferent de neandertals i denissovans va contribuir als genomes actuals de les poblacions de l'Àsia meridional. Cal remarcar, però, que ABC ens pot dir si el model A és millor que el model B, però no garanteix que el model A sigui correcte; ens trobem limitats per la nostra imaginació i per allò que creiem versemblant a l'hora d'escatir quins processos i quines espècies, tangibles o fantasmes, poden haver-hi estat implicats.

També s'interpreta com a senyal d'introgressió arcaica la presència d'haplotips molt divergents respecte de la resta del genoma, encara que aquests haplotips no s'hagin descrit en cap genoma antic conegut.

Aquest mètode va ser emprat també en alguns dels estudis que detecten introgressió arcaica d'origen desconegut en els africans actuals, amb un temps de divergència situat entre 0,5 i 2 milions d'anys (Lorente-Galdos *et al.*, 2019; Durvasula i Sankararaman, 2020). Cal tenir en compte, però, diversos factors de confusió possibles, com l'acció de la selecció equilibradora o l'estructuració poblacional del grup d'humans que, a l'Àfrica, va originar els humans moderns. Aquests factors podrien també explicar parcialment la presència de seqüències divergents en els africans actuals.

## Respostes sobre l'evolució humana inscrites al genoma

Deixant de banda les relacions filogenètiques dels humans amb altres espècies, fins i tot les més properes i extintes com les exposades anteriorment de neandertals i denissovans, l'evolució dels humans ha deixat empremtes al nostre genoma com a resultat de processos demogràfics<sup>6</sup>. Podem reconstruir migracions, barreges de poblacions, aïllaments i expansions de grups humans mitjançant l'anàlisi poblacional i la comparació de genomes d'individus humans actuals i fins i tot incloent-hi dades d'humans extints. Aquest tipus d'anàlisi ens permet la reconstrucció de l'evolució dels humans, de la nostra història, a diferents escales temporals. Podem refer la història més recent d'un individu, de les seves relacions en les darreres generacions, mitjançant l'anàlisi centrada en disciplines com la genètica forense o la genealogia. Però també podem reconstruir esdeveniments molt més antics i globals, com l'origen de la nostra espècie o el poblament de grans territoris i continents. En una escala temporal intermèdia entre els anteriors exemples, podem esclarir quins han estat els processos demogràfics que han patit grups de poblacions concretes i com s'han colonitzat regions geogràfiques locals. Per tal d'exemplificar alguns d'aquests processos, presentem a continuació un conjunt de casos que ens han permès reconstruir la història humana a diferents escales temporals amb l'ajuda de dades provinents d'altres disciplines, com la paleontologia, la lingüística o l'arqueologia, entre moltes d'altres. Com veurem, l'anàlisi dels genomes ens permet confirmar o rebutjar

hipòtesis sobre la història recent i antiga de les poblacions humanes, o proposar-ne de noves.

## L'origen africà de l'espècie humana

Des de finals del segle xx hi ha hagut un ampli consens des de diferents disciplines sobre l'origen africà i recent de la nostra espècie. Les dades paleoantropològiques, arqueològiques, genètiques i fins i tot lingüístiques confirmen l'origen dels humans a l'Àfrica fa uns 200.000 anys. Les primeres evidències genètiques de l'origen africà de la humanitat provenen de l'anàlisi del mtDNA i mostren que els genomes mitocondrials més diferenciats i basals de les poblacions actuals són els presents en poblacions africanes, i que la resta de genomes són una petita fracció de la diversitat trobada a l'Àfrica. L'origen africà de les poblacions humanes actuals es va confirmar també amb l'anàlisi dels llinatges paternals, mitjançant seqüències del cromosoma Y. De totes maneres, cal tenir en compte que tant els llinatges de mtDNA i el cromosoma Y (genomes uniparentals) representen només una fracció petita dels nostres genomes i estan més afectats pels efectes de la deriva genètica que la resta del genoma (autosòmic) a causa de la seva menor grandària efectiva. Per tant, cal analitzar altres parts del genoma per poder aportar conclusions sobre els nostres orígens que seran tant més robustes com més gran sigui la fracció del genoma analitzada. En aquest sentit, totes les anàlisis genètiques autosòmiques, des de l'anàlisi de pocs marcadors fins a la seqüència completa de genomes (Auton *et al.*, 2015; Mallick *et al.*, 2016), mostren més diversitat en poblacions de l'Àfrica degut al nostre origen africà, però també pel fet que les poblacions africanes han mantingut una mida efectiva més gran que els ha permès acumular més variació. El fet que la nostra espècie sorgís a l'Àfrica i durant milers d'anys es diversifiqués pel continent africà, abans de colonitzar la resta del planeta, va permetre que s'acumulessin més variants genètiques en comparació a la resta de poblacions humanes. Els humans no africans representen un subconjunt de la diversitat genètica de les poblacions de l'Àfrica a causa del coll d'ampolla —reducció dràstica d'individus— que van patir durant la sortida de l'Àfrica fa entre 45.000 i 60.000 anys<sup>7</sup>.

<sup>6</sup> Tal com hem comentat en l'inici d'aquest article, esdeveniments adaptatius que també deixen empremtes al genoma es tracten al treball d'Elena Bosch en aquest monogràfic.

<sup>7</sup> Aquesta sortida de l'Àfrica de la nostra espècie es coneix amb el nom d'*Out of Africa* (OOA) en referència a la novel·la homònima publicada el 1937 per l'escriptora danesa Karen Blixen. Una bona revisió de les dades genètiques que confirmen aquesta sortida de l'Àfrica dels humans es pot trobar a Henn, Cavalli-Sforza i Feldman (2012). Vegeu també l'estat actual del tema en l'article de Maria Martín en aquest monogràfic.

Malgrat el consens de l'origen africà dels humans, les dades genètiques actuals no permeten afirmar de manera contundent quina va ser la regió geogràfica dins el continent on va aparèixer la nostra espècie. Hi ha dades genòmiques que apunten a l'est del continent (Tishkoff *et al.*, 2009) o a poblacions de llengües khoisanides del sud (Henn *et al.*, 2011), i fins i tot hi ha dades uniparentals que suggereixen la presència de llinatges basals a l'oest de l'Àfrica (Mendez *et al.*, 2013; Lipson *et al.*, 2020). El fet que les poblacions actuals no siguin representatives de la zona geogràfica on viuen actualment, degut a múltiples migracions durant desenes de mil·lennis i a la possible desaparició de poblacions humanes, dificulta la localització precisa de l'origen dels humans. Aquesta manca de certesa en l'origen local dins de l'Àfrica podria donar suport a les hipòtesis d'un possible origen multiregional de la nostra espècie dins el continent africà. La compilació de més dades genòmiques, incloent-hi genomes complets d'individus actuals i extints, i l'anàlisi de models demogràfics complexos ens permetrà refinar i contrastar les diferents hipòtesis sobre l'origen local dels humans dins l'Àfrica.

### La diàspora oblidada: l'origen del poble romaní

Les dades genòmiques, conjuntament amb les dades aportades per altres disciplines, han permès reconstruir de manera acurada la història i la demografia dels humans en l'àmbit continental i macrogeogràfic. Dins l'àmbit intracontinental, hi ha migracions humanes i diàspores que han estat genèticament documentades, com el cas de les expansions dels pobles bantu a l'Àfrica subsahariana, la diàspora jueva, o l'expansió de llengües tupí-guaraní a l'Amèrica del Sud. De totes maneres, hi ha múltiples qüestions a escala geogràfica més local i de poblacions específiques que encara resten desconegudes i que les dades genòmiques estan ajudant a resoldre. Una d'aquestes històries locals és la diàspora del poble romaní, altrament conegut com a gitano, que ha focalitzat part de la nostra recerca. Malgrat que són la minoria ètnica transnacional més nombrosa en el continent europeu<sup>8</sup>, hi ha hagut pocs estudis genètics centrats en l'origen del poble gitano. Els

registres històrics mostren que el poble gitano ha estat marginat i socialment perseguit en diferents moments i regions d'Europa. Les dades lingüístiques suggereixen un origen dels gitanos al subcontinent indi, ja que l'idioma romaní pertany a la branca indoària de la família lingüística indoiraniana, emparentat amb llengües com el sànscrit, el nepalès, el gujarati, l'hindi o el panjabi. Els primers estudis genètics que apuntaven al subcontinent indi com l'origen del poble gitano van ser aquells centrats en l'anàlisi de malalties mendelianes característiques de gitanos, on es van descriure mutacions compartides amb poblacions actuals de l'Índia i el Pakistan (Kalaydjieva, Gresham i Calafell, 2001).

L'anàlisi de genomes uniparentals va mostrar que la composició genètica del poble romaní és extremadament heterogènia, amb una barreja de llinatges provinents del subcontinent indi i llinatges típicament europeus, alguns dels quals són llinatges amb un reduït nombre de fundadors. Aquesta composició genètica és compatible amb una diàspora dels avantpassats del poble gitano des del subcontinent indi fins a Europa, caracteritzada per múltiples efectes fundadors, colls d'ampolla poblacionals i flux genètic diferencial entre homes i dones amb poblacions europees (Gresham *et al.*, 2001; Martínez-Cruz *et al.*, 2016; García-Fernández *et al.*, 2020). Tal com hem esmentat anteriorment, els llinatges uniparentals representen una petita part del nostre genoma, però l'anàlisi de marcadors autosòmics i genomes complets ha confirmat i refinat les dades uniparentals mitjançant l'aplicació de models demogràfics complexos (Mendizabal *et al.*, 2012; Font-Porterias *et al.*, 2019; Bianco *et al.*, 2020). L'origen del poble gitano segons les dades autosòmiques es pot localitzar al nord-oest de l'Índia, al voltant de l'actual Panjab, fa uns 1.500 anys, quan els avantpassats dels gitanos actuals van patir un fort coll d'ampolla i una migració que va travessar l'Orient Mitjà fins a arribar a Europa fa uns 900 anys. A partir de llavors es van dispersar ràpidament pel continent europeu barrejant-se amb les poblacions locals de manera heterogènia depenent del grup romaní. Per exemple, les dades genòmiques mostren més barreja amb població europea local en els

grups gitanos de la península Ibèrica que dels Balcans.

Malgrat els esforços per fer visible el poble romaní dins el paisatge genètic europeu, encara disposem de pocs estudis i poques dades, i donada l'heterogeneïtat dels grups gitanos, molta de la seva diversitat genètica i de la seva història encara resta desconeguda, especialment aquells petits grups no estudiats o els components de la segona gran diàspora dels gitanos cap al continent americà.

### La prehistòria d'Europa reconstruïda a partir del DNA antic

Els exemples exposats anteriorment reconstrueixen la història de les poblacions humanes a través de l'estudi genètic de les poblacions actuals. Analitzant la diversitat genètica que observem en el moment actual intentem reconstruir com ha estat el passat. Dit d'una altra manera, fem una mirada al passat des del present. Però gràcies als avenços tecnològics en la recuperació de material biològic antic, podem reconstruir la història passada d'una manera diferent, observant directament com eren les poblacions del passat mitjançant l'anàlisi genètica d'individus antics. Aquesta aproximació s'ha generalitzat en la darrera dècada i actualment disposem de moltes dades de DNA antic de poblacions humanes, encara que tècnicament és molt costós i depèn de l'estat de conservació de la mostra a analitzar, ja que el DNA es degrada amb facilitat.

En la cursa per a l'obtenció de DNA d'humans antics a Europa, la primera fita la va posar el grup de Svante Pääbo, que va recuperar el mtDNA d'un individu momificat als Alps tirolesos d'uns cinc mil anys d'antiguitat i de qui posteriorment es va poder seqüenciar tot el genoma i fer una anàlisi poblacional (Handt *et al.*, 1994). Des de llavors hom ha aconseguit recuperar DNA de centenars de restes antigues al continent europeu de diferents èpoques: caçadors-recol·lectors del paleolític, primers agricultors i ramaders neolítics, individus de l'edat dels metalls, fins a individus de la història més recent d'Europa, inclosos personatges històrics com Ricard III d'Anglaterra o el darrer tsar de Rússia (Ivanov *et al.*, 1996; King *et al.*, 2014). Les primeres compilacions de

<sup>8</sup>No hi ha estadístiques acurades del nombre de gitanos a Europa, i les aproximacions fetes apunten a una població al voltant dels 10 milions.

dades genòmiques de diferents èpoques a Europa i la comparació amb dades genòmiques d'europaus actuals van mostrar que les poblacions actuals són el resultat de barreges i reemplaçaments poblacionals que han tingut lloc des de la primera ocupació del continent per poblacions de caçadors-recol·lectors provinents de l'Orient Mitjà. Aquests estudis genètics ajuden a resoldre l'antic debat suscitat a través de diferents disciplines científiques en què es qüestionava el paper dels primers caçadors-recol·lectors paleolítics envers els agricultors i ramaders neolítics en la composició actual dels europeus. Els partidaris d'una substitució i extinció dels individus paleolítics s'enfrontaven als partidaris d'un reemplaçament cultural sense recanvi demogràfic. Les dades genòmiques antigues a Europa apunten a un model híbrid on les poblacions paleolítiques i neolítiques es barrejaren, i fins i tot poblacions posteriorment arribades de les estepes eurasiàtiques contribuïren al bagatge genètic europeu actual d'una manera molt heterogènia.

Però l'estudi del DNA antic en europeus ens va proporcionar una sorpresa: més enllà de la dicotomia entre mesolítics locals i neolítics nous, entre els avantpassats europeus actuals trobem, en la mateixa proporció que els dos anteriors, individus provinents de l'estepa eurasiàtica, els descendents dels quals arriben a l'Europa central i occidental amb l'edat del bronze. En particular, el recanvi genètic va ser molt més intens en els cromosomes Y: entre el 50% i gairebé el 100% dels cromosomes Y dels europeus occidentals pertanyen a un llinatge, R1b-M269, que no es troba en el registre del DNA antic abans de l'edat de bronze i que prové clarament de l'est del continent.

Actualment els estudis de DNA antic a Europa tracten aspectes més locals, com per exemple els moviments poblacionals a la península Ibèrica en els darrers vuit mil anys mitjançant l'anàlisi de centenars de genomes antics, la influència de l'agricultura al sud-est d'Europa, la composició genètica i les migracions a l'antiga Roma imperial o la

influència de les poblacions del Mediterrani oriental en les illes més occidentals, per posar-ne alguns exemples (Mathieson *et al.*, 2018; Antonio *et al.*, 2019; Fernandes *et al.*, 2020). Aquests estudis mostren que la disponibilitat de més genomes antics permetrà en un futur pròxim resoldre processos històrics cada cop més recents a Europa i a altres regions geogràfiques.

Gràcies a l'obra de Darwin, disposem de les eines per reconstruir la nostra evolució com a espècie, tant en relació amb els homínins més propers com dintre dels humans. Però, com mostra l'índex d'aquesta obra que esteu llegint, la reconstrucció del nostre passat és una tasca complexa on cada disciplina aporta una tessella d'aquest mosaic i on caldrà un esforç col·lectiu per fer-les encaixar.

## Bibliografia

- ANTONIO, M. L., *et al.* (2019). «Ancient Rome: A genetic crossroads of Europe and the Mediterranean». *Science* 366: 708–714.
- AUTON, A., *et al.* (2015). «A global reference for human genetic variation». *Nature* 526: 68–74.
- BIANCO, E., *et al.* (2020). «Recent Common Origin, Reduced Population Size, and Marked Admixture Have Shaped European Roma Genomes». *Mol. Biol. Evol.* 37: 3175–3187.
- DE MANUEL, M., *et al.* (2016). «Chimpanzee genomic diversity reveals ancient admixture with bonobos». *Science* 354: 477–481.
- DURVASULA, A.; SANKARARAMAN, S. (2020). «Recovering signals of ghost archaic introgression in African populations». *Sci. Adv.* 6: eaax5097.
- FERNANDES, D. M., *et al.* (2020). «The spread of steppe and Iranian-related ancestry in the islands of the western Mediterranean». *Nat. Ecol. Evol.* 4: 334–345.
- FONT-PORTERIAS, N., *et al.* (2019). «European Roma groups show complex West Eurasian admixture footprints and a common South Asian genetic origin». *PLoS Genet.* 15: e1008417.
- GARCÍA-FERNÁNDEZ, C., *et al.* (2020). «Sex-biased patterns shaped the genetic history of Roma». *Sci. Rep.* 10: 14464.
- GOKCUMEN, O. (2019). «Archaic hominin introgression into modern human genomes». *Am. J. Phys. Anthropol.* 171: 60–73.
- GREEN, R. E., *et al.* (2008). «A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing». *Cell* 134: 416–426.
- GREEN, R. E., *et al.* (2010). «A draft sequence of the Neandertal genome». *Science* 328: 710–722.
- GRESHAM, D., *et al.* (2001). «Origins and divergence of the Roma (gypsies)». *Am. J. Hum. Genet.* 69: 1314–1331.
- HAMMER, M. F., *et al.* (2011). «Genetic evidence for archaic admixture in Africa». *Proc. Natl. Acad. Sci.* 108: 15123–15128.
- HANDT, O., *et al.* (1994). «Molecular genetic analysis of the Tyrolean Ice Man». *Science* 264: 1775–1778.
- HENN, B. M., *et al.* (2011). «Hunter-gatherer genomic diversity suggests a southern African origin for modern humans». *Proc. Natl. Acad. Sci.* 108: 5154–5162.
- HENN, B. M., *et al.* (2012). «The great human expansion». *Proc. Natl. Acad. Sci.* 109: 17758–64.
- HSIEH, P. H., *et al.* (2016). «Model-based analyses of whole-genome data reveal a complex evolutionary history involving archaic introgression in Central African Pygmies». *Genome Res.* 26: 291–300.
- IVANOV, P. L., *et al.* (1996). «Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II». *Nat. Genet.* 12: 417–420.
- KALAYDJIEVA, L., *et al.* (2001). «Genetic studies of the Roma (Gypsies): a review». *BMC Med. Genet.* 2: 5.
- KING, T. E., *et al.* (2014). «Identification of the remains of King Richard III». *Nat. Commun.* 5: 1–8.
- KRAUSE, J., *et al.* (2010). «The complete mitochondrial DNA genome of an unknown hominin from southern Siberia». *Nature* 464: 894–897.
- LIPSON, M., *et al.* (2020). «Ancient West African foragers in the context of African population history». *Nature* 577: 665–670.
- LORENTE-GALDOS, B., *et al.* (2019). «Whole-genome sequence analysis of a Pan African set of samples reveals archaic gene flow from an extinct basal population of modern humans into sub-Saharan populations». *Genome Biol.* 20: 1–15.
- MALLICK, S., *et al.* (2016). «The Simons Genome Diversity Project: 300 genomes from 142 diverse populations». *Nature* 538: 201–206.
- MARTÍNEZ-CRUZ, B., *et al.* (2016). «Origins, admixture and founder lineages in European Roma». *Eur. J. Hum. Genet.* 24: 937–943.
- MATHIESON, I., *et al.* (2018). «The genomic history of southeastern Europe». *Nature* 555: 197–203.
- MENDEZ, F. L., *et al.* (2013). «An African American paternal lineage adds an extremely ancient root to the human Y chromosome phylogenetic tree». *Am. J. Hum. Genet.* 92: 454–459.
- MENDIZABAL, I., *et al.* (2012). «Reconstructing the population history of European Romani from genome-wide data». *Curr. Biol.* 22: 2342–2349.
- MONDAL, M., *et al.* (2019). «Approximate Bayesian computation with deep learning supports a third archaic introgression in Asia and Oceania». *Nat. Commun.* 10: 1–9.
- OLALDE, I., *et al.* (2019). «The genomic history of the Iberian Peninsula over the past 8000 years». *Science* 363: 1230–1234.
- PRÜFER, K., *et al.* (2014). «The complete genome sequence of a Neanderthal from the Altai Mountains». *Nature* 505: 43–49.
- REICH, D., *et al.* (2010). «Genetic history of an archaic hominin group from Denisova Cave in Siberia». *Nature* 468: 1053–1060.
- SIKORA, M., *et al.* (2014). «Population Genomic Analysis of Ancient and Modern Genomes Yields New Insights into the Genetic Ancestry of the Tyrolean Iceman and the Genetic Structure of Europe». *PLoS Genet.* 10: e1004353.
- TISHKOFF, S. A., *et al.* (2009). «The genetic structure and history of Africans and African Americans». *Science* 324: 1035–1044.
- XU, D., *et al.* (2017). «Archaic Hominin Introgression in Africa Contributes to Functional Salivary MUC7 Genetic Variation». *Mol. Biol. Evol.* 34: 2704–2715.