

# Les adaptacions que ens han fet humans. El genoma

Elena Bosch

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

Correspondència: [elena.bosch@upf.edu](mailto:elena.bosch@upf.edu)

DOI: 10.2436/20.1501.02.204

ISSN (ed. impresa): 0212-3037

ISSN (ed. digital): 2013-9802

<http://revistes.iec.cat/index.php/TSCB>

Rebut: 14/01/2021

Acceptat: 16/03/2021

## Resum

Identificar què ens fa humans i comprendre les bases genètiques de la nostra singularitat són qüestions d'interès i debat continuat en biologia. Mitjançant l'anàlisi de dades genòmiques podem detectar la petjada de la selecció natural adaptativa (positiva) en el genoma amb l'objectiu d'identificar les adaptacions del nostre passat més o menys recent i els trets fenotípics resultants en les poblacions actuals i que ens defineixen com a humans. Tot i que la genòmica ens ha permès reconèixer algunes d'aquestes adaptacions a partir del patró de variació esperat sota el model de l'escombrat selectiu clàssic o dur, el camp comença a abordar noves aproximacions. Per exemple, mètodes per identificar adaptacions complexes o poligèniques i estratègies multidisciplinàries per comprendre la relació entre les empremtes genòmiques detectades, les variants funcionals subjacents, el fenotip adaptatiu que aquestes determinen i la pressió selectiva que ha conduït cada adaptació.

**Paraules clau:** adaptació humana, selecció natural, selecció positiva, escombrat selectiu, selecció poligènica.

## Abstract

Identifying what makes us human and understanding the genetic basis of our singularity are questions of interest and ongoing debate in biology. Through the analysis of genomic data we can detect the footprint of adaptive (positive) natural selection in our genome to reveal the specific adaptations that occurred in our past and that account for phenotypic adaptive traits still present in current human populations and that define us as humans. Although genomics has allowed evolutionary biologists to recognize some of these adaptations by scanning for the expected pattern of variation under the classical or hard selective model, the field starts to address interesting new approaches. For example, methods to identify complex or polygenic adaptations and multidisciplinary strategies to understand the relationship between detected genomic fingerprints, the underlying functional variants, the adaptive phenotype that these determine, and the selective pressure that has led each adaptation.

**Keywords:** human adaptation, natural selection, positive selection, selective sweep, polygenic selection.

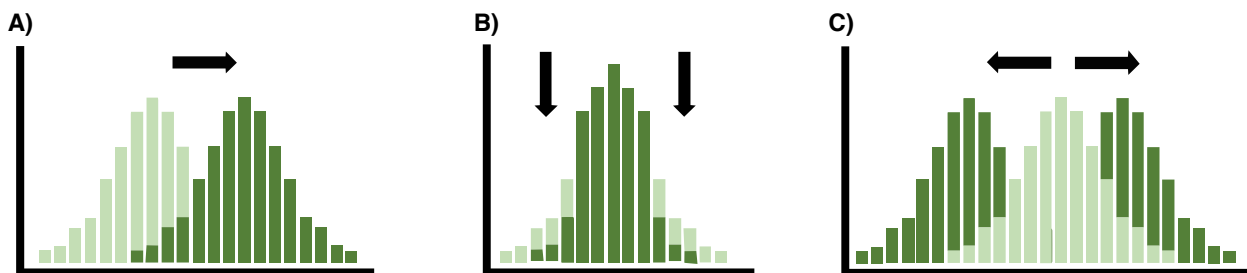
## Introducció

La selecció natural fa referència al procés gradual no aleatori d'acumulació de característiques beneficioses i heretables que observem generació rere generació en les poblacions com a resultat de l'existència de variació heretable en l'èxit reproductiu i/o en la capacitat de supervivència dels individus. El concepte, proposat simultàniament per

Charles Darwin i Alfred R. Wallace (Darwin i Wallace, 1858), és un dels principals mecanismes que explica el canvi evolutiu, però sobretot és l'única força evolutiva que facilita l'adaptació de les poblacions i espècies als seus ambients i entorns canvians.

Gràcies als avenços de la genòmica dels darrers anys i a la recent abundància de dades genòmiques en diferents poblacions humanes,

podem localitzar amb relativa facilitat quines regions específiques del nostre genoma mostren patrons de variació compatibles amb diferents tipus de selecció natural (Figura 1). Podem fer aquesta cerca sense necessitar cap mena de suposició a priori respecte de quins gens o funcions biològiques han facilitat la nostra adaptació a diferents ambients. Tanmateix, a partir d'aquestes regions



↑ Figura 1. Tipus de selecció natural. A) Selecció direccional (positiva o darwiniana). Tipus de selecció en la qual s'afavoreixen els individus de la població que presenten les característiques d'un dels extrems de la distribució fenotípica d'un caràcter biològic. A continuació, la distribució fenotípica de la població es mou cap a l'extrem de la distribució amb els fenotips més afavorits. B) Selecció estabilitzadora (purificadora o negativa). Els extrems de la distribució d'un fenotip concret són seleccionats en contra en afavorir un rang concret de valors en la distribució. Per tant, la distribució fenotípica s'estreny. C) Selecció disruptiva o balancejadora. Tipus de selecció en la qual s'afavoreixen els individus d'ambdós extrems de la distribució fenotípica. En conseqüència, els individus amb les característiques fenotípiques intermèdies desapareixen i la distribució es divideix en dos (o més) pics que representaran els fenotips divergents afavorits a la vegada en un determinat ambient.

candidates inicialment identificades amb patrons de variació compatibles amb l'acció de la selecció natural, caldrà destriar les possibles variants adaptatives subjacents, el fenotip adaptatiu que determinen, així com la pressió selectiva que ha dirigit cada adaptació. En aquesta revisió es presentaran les principals estratègies per identificar l'empremta de la selecció natural adaptativa —també anomenada selecció darwiniana, positiva o direccional (Figura 1A)—, alguns exemples concrets d'adaptacions humanes ben caracteritzades identificades amb aquests mètodes, i les principals limitacions en la interpretació i comprensió d'aquestes empremtes.

En funció d'en quin marc temporal de la nostra història evolutiva vulguem interrogar l'acció de la selecció natural, emprarem estratègies basades en la comparació de dades genòmiques entre el genoma humà i el d'altres espècies de primats, entre diferents poblacions humanes —aquest és un nivell en el qual els mètodes són més robusts—, o fins i tot entre determinades poblacions humanes i homínids extints, gràcies a la recent disponibilitat del genoma de diferents individus neandertals i denissovans. D'aquesta manera, podrem inferir adaptacions específiques de la nostra espècie, esdeveniments selectius restringits geogràficament en determinades poblacions humanes, o fins i tot introgressions adaptatives rebudes d'humans arcaics.

### Comparacions interespecífiques

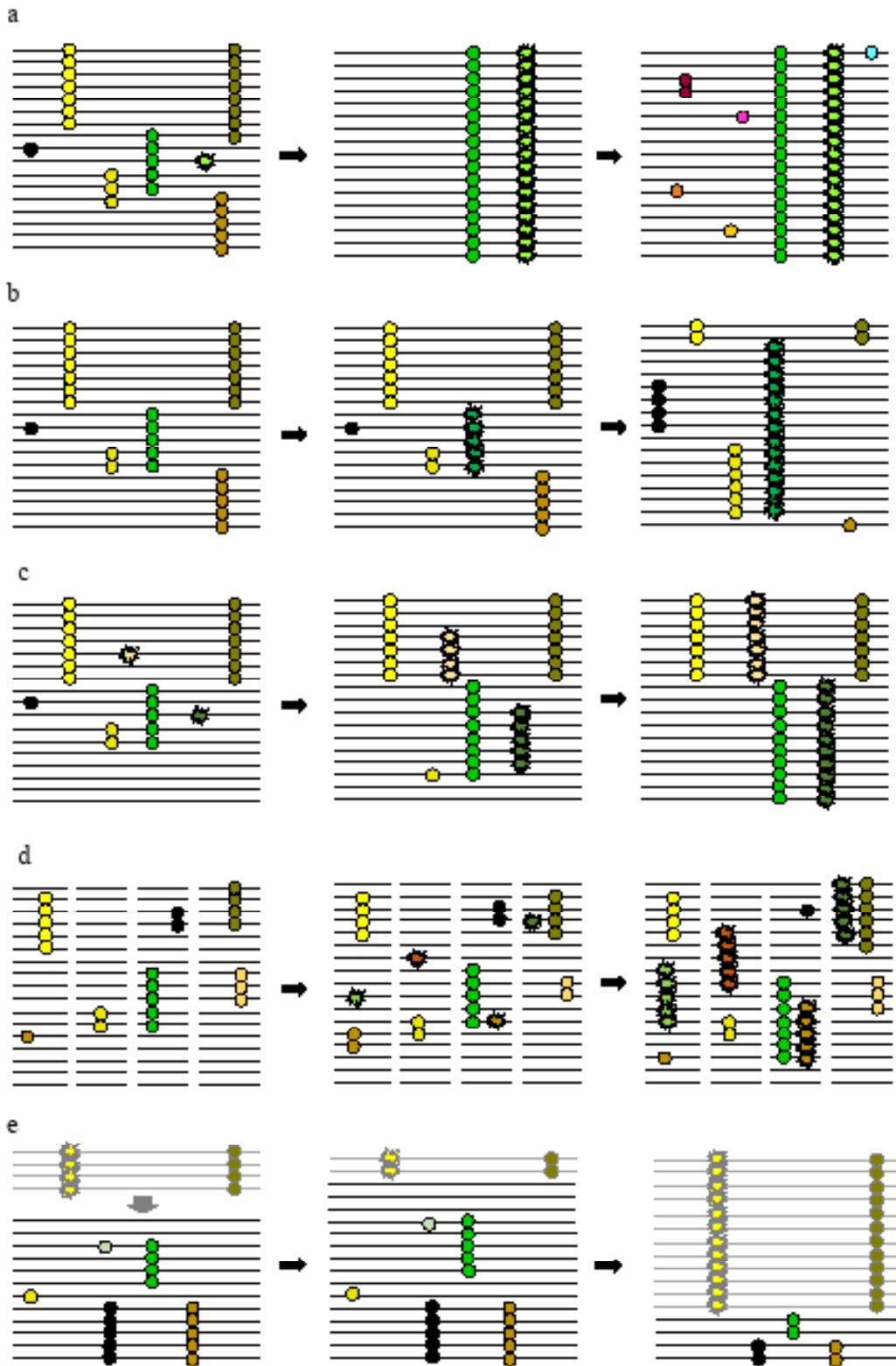
Els tests de selecció basats en la comparació entre espècies permeten identificar substitucions i canvis del genoma específics del nostre llinatge i que potencialment poden haver contribuït a la nostra divergència. Permeten explorar, per tant, adaptacions que varen ocórrer fa milions d'anys (Ma) —per exemple, uns 6 Ma si comparem el genoma humà amb el del ximpanzé— i que ens caracteritzen com a espècie. En aquest cas, una de les estratègies més emprades per identificar l'empremta de la selecció positiva entre espècies és la comparació de les taxes de substitució dels nucleòtids del DNA. Donada una regió de DNA codificadora per una proteïna, podem comparar la proporció de canvis no sinònims —canvis de nucleòtids del DNA que impliquen un canvi d'aminoàcid de la proteïna— respecte de la proporció de canvis sinònims —canvis de nucleòtids que no impliquen un canvi d'aminoàcid— entre els genomes de dues espècies per calcular la ràtio

$Ka/Ks$ , també anomenada segons el context  $dN/dS$  o  $w$ . Suposant que els canvis sinònims són neutres —no estan sotmesos a la selecció natural—, la proporció ens permetria identificar la possible acumulació de mutacions perjudicials o beneficioses, segons el cas, en una regió codificant. En molts casos, les mutacions que canvien un aminoàcid no són ben rebudes per la selecció natural, ja que comprometen el correcte funcionament de la proteïna corresponent; per tant, la major part de gens del nostre genoma presenten una ràtio  $Ka/Ks$  força inferior a 1. Valors superiors a 1 són excepcionals i es donen quan s'ha afavorit l'acumulació de canvis d'aminoàcid en una proteïna perquè aquests canvis podrien representar novetats evolutives i ser, per tant, potencialment adaptatius en un dels llinatges comparats. D'altra banda, els mètodes es poden adaptar per explorar la ràtio  $Ka/Ks$  en determinats llinatges o branques en un arbre filogenètic definit inicialment, així com en determinats codons o regions codificants d'una proteïna (Yang i Nielsen, 2002; Zhang *et al.*, 2005). En general, si bé aquests tipus d'anàlisis tenen una potència limitada per detectar la selecció en gens individuals, sí que són prou potents per identificar quines funcions biològiques han canviat més significativament en el nostre llinatge respecte del ximpanzé o altres primats no humans.

La gran similitud entre les seqüències dels genomes humà i de ximpanzé, i el fet que les nostres proteïnes són pràcticament idèntiques, ha portat a explorar altres tipus de variació entre els nostres genomes per identificar altres possibles innovacions biològiques en el nostre llinatge. En aquest cas, l'ús de tècniques d'hibridació comparativa amb l'ús de sondes de cDNA (DNA complementari) en *arrays* (Marques-Bonet *et al.*, 2009) i més recentment, també, la quantificació de diferències de cobertura —*coverage* o nombre de lectures úniques que inclouen un determinat nucleòtid a la seqüència de DNA reconstruïda— al llarg dels genomes després de la seva seqüenciació massiva paral·lela (Kronenberg *et al.*, 2018), han permès identificar duplicacions i expansions gèniques específiques del llinatge humà que podrien representar novetats evolutives del nostre llinatge, especialment quan aquests canvis s'han pogut relacionar amb canvis de funció particulars o patrons d'expressió diferencials reals entre les espècies comparades.

### Tests basats en la variació intraespecífica

Els tests basats exclusivament en la variació intraespecífica del nostre genoma (o polimorfisme humà) abastaran períodes de temps molt més pròxims de la nostra història evolutiva i, segons el cas, permetran interrogar adaptacions de fa poques desenes de milers d'anys fins a adaptacions més antigues, incloent diferents períodes clau de la nostra història evolutiva, com són l'expansió humana fora de l'Àfrica i la subsegüent colonització de nous territoris amb condicions climàtiques molt diferents a la sabana ancestral, o la introducció de l'agricultura i la ramaderia com a manera de subsistència en el neolític (Sabeti *et al.*, 2006). En aquest cas, la major part d'escombrats de tot el genoma (*genome scans*) per detectar selecció positiva realitzats en humans s'han centrat a detectar les empremtes que esperariem sota l'anomenat model d'escombrat selectiu dur. El model assumeix l'aparició d'un nou al·lel en una determinada regió del nostre genoma que determinarà un avantatge selectiu als seus portadors i, consegüentment, generació rere generació, augmentarà ràpidament la seva freqüència cap a la fixació, arrossegant amb ell el conjunt d'al·lells immediatament flanquejants en la posició o regió on ha sorgit per mutació i eliminant les combinacions d'al·lells presents en la resta de cromosomes de la població no portadors de l'al·lel adaptatiu (Figura 2a). Disposem de diferents grups de tests estadístics que permeten capturar diferents aspectes del patró de variació resultant d'aquest escombrat selectiu clàssic o dur: (i) mètodes basats en l'espectre de freqüències, que permeten detectar la reducció de la diversitat, un excés d'al·lells derivats d'alta freqüència i/o una tendència cap a un excés d'al·lells rars; (ii) mètodes basats en l'extensió del desequilibri de lligament que detecten haplotips inusualment llargs en relació amb les seves freqüències a la població; i (iii) mètodes basats en la diferenciació entre poblacions que capturen variants al·lèliques amb alts nivells de diferenciació poblacional i que, per tant, resulten altament informatius quan els esdeveniments selectius i consegüents adaptacions es troben restringits geogràficament.



† Figura 2. Empecentes moleculars de la selecció natural positiva o darwiniana. a) Escombat clàssic o dur. b) Escombat suau a partir de variació neutra preexistent en la població que esdevé adaptativa per un canvi ambiental. c) Escombat suau a partir de l'aparició de més d'una variant adaptativa en un mateix locus. d) Selecció poligènica o selecció en un caràcter complex. e) Introgessió adaptativa.

## i) Tests basats en l'espectre de freqüències

En una regió genòmica sota selecció, sota el model d'un escombrat clàssic, en fixar-se la variant adaptativa i arrossegar aquesta la variació immediatament flanquejant de la posició on ha aparegut per mutació, esperarem trobar una reducció dràstica de la seva diversitat en comparació a altres regions del genoma evolucionant sota neutralitat, les quals presentaran polimorfismes més o menys antics —i freqüents— en la mateixa població. Hi ha una pèrdua de diversitat entorn de la regió seleccionada. Amb el temps, apareixeran nous al·lells per mutació en la regió afectada per l'escombrat selectiu, però aquests nous al·lells, en ser molt probablement neutres, tardaran molt temps a augmentar la seva freqüència. En el patró de variació de les regions sota selecció positiva, esperem trobar, per tant, una empremta característica de baixa diversitat global i un cert excés d'al·lells rars un cop la mutació comença a introduir noves mutacions. Disposem de moltes proves estadístiques que permeten detectar aquest tipus de desviació en l'espectre de freqüències respecte de neutralitat. Entre les més emprades trobaríem, en aquest cas, la  $D$  de Tajima (Tajima, 1989) i altres tests relacionats, com la  $D$  de Fu i Li (Fu i Li, 1993) i la  $F$  de Fu (Fu, 1997).

Els al·lells derivats, és a dir, els al·lells no ancestrals, sorgeixen per mutació i generalment solen tenir freqüències al·lèliques més baixes que els al·lells ancestrals. En un escombrat selectiu, però, els al·lells derivats lligats a l'al·lel beneficiós poden ser arrossegats a alta freqüència en l'anomenat efecte de l'autoestopista (*hitchhiking effect*). Com que molts d'aquests al·lells derivats no arribaran a la fixació completa —com a resultat d'un escombrat incomplet o d'alguna recombinació entre aquests i l'al·lel seleccionat durant l'escombrat selectiu—, la selecció positiva ens pot crear una nova característica en el patró de variació de les regions sota selecció consistent en un excés d'al·lells derivats a alta freqüència respecte de neutralitat. Les aproximacions estadístiques més emprades per detectar aquest excés són el test de la  $H$  de Fay i Wu (Fay i Wu, 2000) o la cerca directa d'una proporció elevada d'SNPs (polimorfismes d'un sol nucleòtid) amb un determinat llindar de freqüència per al·lel derivat, com, per exemple, superior al 80% (Walsh *et al.*, 2006).

## ii) Tests basats en la diferenciació poblacional

En humans, els nivells de diferenciació poblacional estan determinats majoritàriament

per la deriva genètica i la migració, però l'adaptació local, aquella específica d'una determinada àrea geogràfica, pot accentuar els nivells de diferenciació en determinats loci del nostre genoma. Per exemple, l'al·lel  $O$  del sistema sanguini *Duffy* es troba pràcticament absent fora de l'Àfrica, mentre que el mateix al·lel presenta elevades freqüències en moltes poblacions africanes on s'ha seleccionat de manera local perquè confereix resistència a la malària. Altres casos coneguts d'adaptacions genètiques locals identificades a partir de diferenciacions poblacionals inusuals inclouen diferents variants que determinen una pigmentació de pell més clara fora de l'Àfrica, així com altres variants genètiques que confereixen protecció a la malària a l'Àfrica. Podem quantificar aquesta diferència de freqüències al·lèliques tan extrema observada en el sistema *Duffy* mitjançant l'estadístic  $F_{ST}$  —la mesura més correntment emprada per mesurar les diferències entre poblacions— calculat entre poblacions africanes i no africanes i utilitzant com a referència tots els SNPs del genoma.

Altres estratègies més refinades, com l'estadístic PBS (*Population Branch Statistic*) (Yi *et al.*, 2010) utilitzen els valors  $F_{ST}$  obtinguts entre tres poblacions per identificar quines posicions genòmiques presenten diferenciacions extremes al llarg de cada una de les branques del corresponent arbre de tres poblacions. Aquest disseny ha resultat especialment informatiu per identificar loci implicats en adaptacions locals a condicions extremes, com el clima àrtic o l'elevada altitud en comparar dues poblacions de divergència relativament recent en el temps, però vivint sota pressions selectives locals clarament diferents —fred extrem o hipòxia, en els anteriors casos, en comparació a condicions normals— amb una tercera població de divergència més antiga (Yi *et al.*, 2010; Fumagalli *et al.*, 2015; Amorim *et al.*, 2017).

## iii) Tests basats en l'extensió dels haplotips o desequilibri de lligament

Els escombrats selectius creen una empremta distintiva en els patrons de desequilibri de lligament que no esperariem sota neutralitat. Sota selecció positiva, un nou al·lel és afavorit i augmenta la seva freqüència amb suficient rapidesa sense que, en principi, la recombinació pugui trencar l'associació inicial d'aquest nou al·lel afavorit amb la resta d'al·lells dels loci veïns en el cromosoma ancestral on va aparèixer per mutació. En aquest cas, la

selecció positiva recent donarà al·lells a elevades freqüències que presenten, a la vegada, associacions de llarg abast —o desequilibri de lligament— amb altres al·lells presents en el mateix cromosoma. Detectarem aquestes associacions de llarg abast en forma d'un haplotip o determinada combinació d'al·lells —on originàriament va aparèixer l'al·lel seleccionat— sempre que la recombinació no actuï de forma substancial, cosa que acabarà passant amb el temps. Altrament, com que els al·lells neutres necessitaran molt de temps per arribar a alta freqüència per deriva, si és el cas, no presentaran mai simultàniament elevada freqüència i associacions a llargues distàncies. D'aquesta manera, disposem de diferents tests per a la detecció d'esdeveniments de selecció positiva recents —menys de 30.000 anys—, que exploren la relació entre la freqüència dels al·lells i l'extensió del desequilibri de lligament comparant directament la longitud dels haplotips associats a diferents al·lells de cada regió candidata o l'extensió de l'homozigositat haplotípica, entre d'altres.

Entre aquests darrers estadístics, destaquen l'*iHS* (*integrated Haplotype Score*) i l'*XP-EHH* (*Cross-Population Extended Haplotype Homozygosity*). Mentre que l'*iHS* permet comparar l'extensió d'homozigositat haplotípica entre els al·lells d'un determinat SNP en una població i és especialment informatiu per identificar variants sota selecció positiva que encara no han arribat a la fixació, és a dir, detectar escombrats selectius incomplets o en curs (Voight *et al.*, 2006), l'*XP-EHH* compara l'extensió de l'homozigositat haplotípica entre poblacions i serveix per detectar variants recentment seleccionades després d'un escombrat selectiu complet —o quasi complet— en una població concreta (Sabeti *et al.*, 2007).

## Verificant la significació estadística

Els tests de neutralitat i estadístics que acabem de veure permeten, en principi, distingir el patró de variació genètica esperat sota selecció natural positiva respecte de neutralitat. Tanmateix, la història demogràfica particular de la població —com els colls d'ampolla i les expansions demogràfiques— pot crear alguns patrons de variació similars als de la selecció positiva i complicar en certs casos la inferència correcta dels esdeveniments de selecció, especialment quan s'estudia un nombre reduït de loci. Com que els esdeveniments demogràfics afectaran per igual a tot el conjunt

del genoma i, en canvi, els esdeveniments d'adaptació només afectaran el patró de variació de les regions sota selecció, hom espera poder identificar casos de selecció emprant dades a escala genòmica i determinant quines regions del genoma presenten desviacions respecte la neutralitat més extremes mitjançant qualsevol dels estadístics descrits. A partir de la distribució empírica dels valors d'aquests estadístics al llarg de tot el genoma, sigui calculats per SNP o posició polimòrfica, o mitjançant finestres de menor o major longitud i més o menys superposades, podem fàcilment identificar així quins loci o regions genòmiques presenten valors extrems en cada distribució i considerar-los com els candidats més probables a haver patit fortes pressions de selecció (Akey, 2009). Tot i que en aquest tipus d'aproximació podem obtenir falsos positius, la posterior identificació de variants funcionals en aquestes regions candidates i la seva relació amb un fenotip adaptatiu i una pressió selectiva subjacent podrà ajudar a identificar convincentment els casos adaptatius reals. A la vegada, cal ser igualment conscients que segurament obtindrem molts falsos negatius per manca de potència en focalitzar-nos només en les empremtes més extremes dels escombrats selectius clàssics.

Mitjançant simulacions podríem millorar la identificació d'aquests possibles valors extrems o *outliers* en aplicar un model neutre i/o considerar diferents models de selecció incorporant, segons el cas, demografies específiques per a les poblacions estudiades, el mapa de recombinació del genoma i/o possibles biaixos en la variació estudiada, ja sigui per l'ús de dades de genotipatge d'SNPs o per emprar dades de seqüenciació amb baixa cobertura. Sens dubte, si la variació neutra del genoma es trobés perfectament modelada en aquests escenaris simulats, podríem simplement identificar les variants que són més externes d'aquesta distribució neutra com a possibles candidats de selecció, o fins i tot podríem provar si coincideixen amb un règim de selecció simulat determinat donats una certa demografia i un coeficient de selecció. Així i tot, desconexem sovint la forma real de la distribució neutra i com ha estat afectada per la pròpia història demogràfica de les diferents poblacions humanes. D'altra banda, i especialment per a poblacions amb històries complexes de barreja en el seu passat, quan hom disposa d'informació genòmica de diverses poblacions, l'ús de gràfics de mescla

(*admixture*, és a dir, de mescla) ha demostrat ser també una estratègia força efectiva per identificar esdeveniments de selecció forts i poder-los associar a canvis de freqüències al·lèliques a determinades branques del seu corresponent arbre (Refoyo-Martínez *et al.*, 2019). De manera similar, l'estimació de genealogies a escala genòmica per a milers de mostres a la vegada es pot emprar per estimar la velocitat de dispersió d'un determinat llinatge en relació amb altres llinatges alternatius en un determinat període de temps i investigar diferents formes de selecció (Speidel *et al.*, 2019).

Finalment, altres aproximacions empen diferents estratègies compostes per combinar el conjunt de les diferents empremtes específiques que els diferents estadístics de selecció detecten amb l'objectiu d'obtenir més sensibilitat en l'anàlisi de selecció, identificar les variants concretes seleccionades en cada cas o fins i tot classificar les regions detectades sota selecció, segons hagin estat afectades per escombrats selectius recents/antics i incomplets/complets en diferents poblacions (Grossman *et al.*, 2010; Pybus *et al.*, 2015; Szpak *et al.*, 2018). En principi, com que cada estadístic capta diferents aspectes de l'escombratge selectiu, qualsevol estratègia composta sempre ajudarà a evitar falsos positius particulars d'algun estadístic i guanyar més potència estadística.

### Més enllà de l'escombrat selectiu clàssic

Tot i que la majoria d'estudis de selecció positiva en humans s'han centrat a captar les empremtes de l'escombrat selectiu clàssic, altres models de selecció, com ara la selecció poligènica —on múltiples loci contribueixen a un determinat fenotip adaptatiu— i els anomenats escombrats selectius suaus —quan diferents variants noves en un loci concret del genoma són seleccionades a la vegada o quan la selecció afavoreix una variant polimòrfica preexistent en la població que esdevé adaptativa per un canvi d'ambient—, poden haver contribuït igualment a l'adaptació humana (Pritchard *et al.*, 2010). En aquests casos, les empremtes que apareixen en el patró de variació de les regions sota selecció són clarament més subtils i difícils de capturar amb els mètodes estàndard desenvolupats per detectar escombrats durs clàssics (Figura 2).

En els escombrats suaus, com que les variants afavorides estaran presents en diferents haplotips o tipus de cromosomes

(Figura 2b i c), en augmentar aquestes la seva freqüència i fixar-se, no detectarem empremtes de selecció tan clares com en un escombrat selectiu clàssic, on una única variant que acaba d'aparèixer per mutació en un únic haplotip és seleccionada. En aquest cas, estadístics com *nSL* (Ferrer-Admetlla *et al.*, 2014), H12 i H2/H1 (Garud i Rosenberg, 2015) o SDS (Field *et al.*, 2016) han estat proposats com a estratègies que millorarien la detecció d'aquest tipus d'empremtes més dèbils. De manera semblant a l'*iHS*, l'estadístic *nSL* fou dissenyat per capturar l'augment d'homozigositat haplotípica que la selecció positiva crea en afavorir una variant encara segregant en la població. En aquest cas, l'*nSL* mesura l'extensió de l'homozigositat a partir del nombre de llocs segregants amb al·lèls compartits entre els haplotips que acompanyen cada al·lel ancestral i derivat d'una determinada posició polimòrfica i no necessita un mapa genètic (Ferrer-Admetlla *et al.*, 2014). D'altra banda, els estadístics H12 i H2/H1, basats també en l'homozigositat haplotípica, permeten considerar segons el cas tots els haplotips de la població, l'haplotip més abundant o els dos haplotips més freqüents presents a la mostra amb l'objectiu de capturar i distingir, gràcies a la seva combinació, tant escombrats durs —on esperem un únic haplotip majoritari afavorit— com suaus —amb dos o més haplotips seleccionats a la vegada (Garud i Rosenberg, 2015). L'estadístic SDS (Singleton Density Score; Field *et al.*, 2016) empra dades de seqüència del genoma complet per inferir canvis recents en les freqüències al·lèliques d'SNPs a partir de la distorsió que la selecció recent fa sobre les genealogies dels haplotips mostrejats, la qual cosa dona lloc a branques terminals més curtes en el cas dels al·lèls afavorits. El mètode permet detectar tant escombrats clàssics com casos de selecció positiva recent a partir de variació neutra prèviament preexistent.

Pel que fa al cas de la selecció poligènica (Figura 2d), les empremtes resultants en el nostre genoma poden implicar només canvis subtils en les freqüències al·lèliques de múltiples loci implicats en determinades funcions o caràcters adaptatius. Com que la majoria dels fenotips humans tenen probablement una arquitectura genètica complexa i no són el resultat només d'una o poques variants d'efecte fort, s'estan dissenyant igualment altres aproximacions i estratègies per poder reconèixer l'empremta molecular de l'adaptació poligènica. Tot i que



les empremtes de la selecció en cada un dels gens implicats en aquests fenotips complexos sovint poden no ser suficients per destacar de manera individual en el context de tot el genoma, un enfocament prou eficient per capturar aquest tipus d'escenari consisteix a comparar l'evidència d'adaptació en tot el conjunt de gens relacionats amb determinats fenotips o caràcters complexos a la vegada (Daub *et al.*, 2013). Tanmateix, no sempre es disposa d'un coneixement complet de tots els gens implicats en un fenotip complex, ni és fàcil modelar-ne els efectes, a vegades realment petits, que alguns d'aquests loci tenen en aquests caràcters complexos. Especialment en poblacions natives, generalment no considerades en els grans estudis d'associació o consorcis de GWAS, l'ús de puntuacions de risc poligènic (*polygenic risk score*) basades en poblacions de referència per modelar els efectes de cada un d'aquests loci podria ser problemàtica. Altres aproximacions vàlides per detectar els patrons adaptatius més subtils esperats en el cas de la selecció poligènica se centren a correlacionar canvis de freqüències al·lèliques en les poblacions humanes amb diferents característiques de l'entorn —variables ambientals com la temperatura, la humitat, la radiació solar o la disponibilitat de nutrients, entre d'altres. Si bé és important corregir per l'estructura genètica de les poblacions incloses

en l'anàlisi, les possibles associacions ambientals identificades permeten, en emprant dades de seqüenciació, relacionar directament variants genètiques concretes amb una pressió selectiva causal. Aquesta estratègia ha permès identificar amb èxit diferents polimorfismes funcionals humans associats a la tolerància al fred, la radiació i la pigmentació i a altres diferències climàtiques (Hancock *et al.*, 2011).

Finalment, la introgressió amb una altra espècie podria, en casos excepcionals, facilitar la nostra adaptació a partir d'una variant adaptativa preexistent en una altra espècie (Figura 2e). Hi ha diferents aproximacions estadístiques específiques per detectar aquests casos, les quals han permès descriure alguns exemples concrets de variants adaptatives que diferents poblacions humanes haurien adquirit a partir de la introgressió amb neandertals o amb denissovans (Racimo *et al.*, 2015); entre d'altres, en els receptors *TLR6-TLR1-TLR10* implicats en el reconeixement de patògens externs en poblacions europees (Dannemann *et al.*, 2016) o en el gen *EPAS1* en poblacions tibetanes, on facilitaria la seva adaptació a la hipòxia (Huerta-Sánchez *et al.*, 2014).

**Limitacions i futur dels estudis de selecció**

Independentment del mode de selecció que operi, un cop identificada qualsevol empremta

de la selecció positiva en el genoma caldrà comprendre la seva rellevància biològica com a cas d'adaptació genètica humana. Finalment, es tracta ara d'anar del genotip al fenotip. Això inclou no només desxifrar l'acció molecular i el fenotip subjacent de les variants adaptatives seleccionades en cada cas, sinó identificar també la pressió selectiva en joc. Tot i que el nombre de casos completament caracteritzats encara és limitat (Rees *et al.*, 2020), han permès identificar i comprendre adaptacions humanes que sens dubte han contribuït a la nostra supervivència i èxit colonitzador com a espècie —vegeu-ne exemples concrets relacionats amb la nostra defensa enfront de patògens, la dieta i la disponibilitat de recursos, la radiació ultraviolada i l'altitud a la Taula 1. El gran coll d'ampolla dels estudis de selecció positiva no és la identificació en si de les regions candidates, sinó l'enorme dificultat que tenim per entendre, sense cap hipòtesi prèvia o disseny experimental amb diferents pressions selectives, la relació entre els genotips de les regions candidates i els seus possibles fenotips adaptatius.

Els primers exemples d'adaptacions humanes caracteritzades amb èxit van començar amb el coneixement previ d'un fenotip que s'intuïa adaptatiu, per al qual posteriorment es van identificar proves genètiques per a la selecció, generalment

↓ Taula 1. Exemples d'adaptacions genètiques en humans.

ADAPTACIÓ	GENS	POBLACIÓ/REGIÓ GEOGRÀFICA
Llenguatge	<i>FOXP2</i>	Llinatge humà
Malària	<i>HBB, HBA, GYP A, GYP B, GYP C, G6PD, FY</i>	Àfrica subsahariana
Tripanosoma	<i>APOLI</i>	Àfrica
Bacteris flagel·lats	<i>TLR5</i>	Àfrica
Reposta immunitària general	<i>TLR1-TLR6-TLR10, grup HLA, CD5</i>	Múltiples poblacions
Radiació UV	<i>SLC24A5, SLC45A2, OCA2, MCRI, TYRPI, DCT</i>	Europea, Asiàtica
Morfologia del cabell	<i>EDAR</i>	Asiàtica
Altitud	<i>APAS1, ELGN1, PPARA, VAV3, CBARA1, ARNT2, ELGINI, EDNRA, NOS2A, BRINP3</i>	Tibetans Altiplà d'Etiòpia Població andina
Persistència de la lactasa	<i>LCT</i>	Europea i africanes
Dieta - Zinc	<i>SLC30A, SLC39A8</i>	Àsia de l'est i Àfrica
Dieta - Seleni	<i>DI2, GPX1, GPX3, CELF1, SPS2, SEPSEC3</i>	Xinesa
Dieta - Ferro	<i>HFE</i>	Europeus
Dieta rica en greixos	<i>FADS</i>	Inuits, nadius americans
Toxicitat arsènica	<i>AS3MT</i>	Argentina

dirigides a gens candidats plausibles (Vitti *et al.*, 2013). La cerca de les empremtes de selecció a escala genòmica, però, té el gran potencial de permetre inferir noves hipòtesis per testar nous casos de selecció sense cap biaix o suposició prèvia respecte de quines regions genòmiques podrien facilitar l'adaptació en cada context selectiu. Tanmateix, hem de ser conscients que la majoria d'aquestes cerques proporcionen sovint només llistes de regions candidates d'adaptació sense indicació de la possible variació funcional o dels fenotips adaptatius concrets que aquestes regions han facilitat. La seva interpretació i comprensió generalment no és directa i cal realitzar estudis funcionals posteriors per validar les possibles variants i fenotips adaptatius en relació amb determinades pressions selectives. És més, la majoria de les empremtes detectades solen abastar grans regions del genoma, amb múltiples variants funcionals, i molts gens diferents sovint amb més d'una funció cada un d'ells i que, per tant, poden potencialment afectar molts fenotips diferents. Per il·lustrar la dificultat a l'hora de destriar i identificar les variants i els fenotips adaptatius darrere una forta empremta de selecció podem fixar-nos en el treball de Wang *et al.* (2020), en què s'investiga una de les empremtes més fortes de selecció positiva detectades en humans: el locus 2p21.3. El locus conté el gen de la lactasa, i els senyals de selecció trobats en poblacions europees s'han associat a l'al·lel derivat d'una variant reguladora que permet la persistència de la lactasa en edat adulta i, per tant, la digestió de la llet. En aquest cas clàssicament s'inferia que, en aquelles poblacions on culturalment s'empra la llet fresca com a recurs alimentari, els portadors de la variant que facilita la persistència de la lactasa haurien vist incrementada la seva supervivència en episodis de penúria. En investigar en detall el contingut genòmic de tota la regió sota selecció, així com alguns dels fenotips associats a la regió en altres espècies, els autors identificaren un microRNA (miR-128-1) en el mateix haplotip afavorit en les poblacions europees, directament implicat en processos metabòlics de despesa energètica que podria igualment representar una adaptació a períodes de fam en el nostre passat.

Els *scans* genòmics de selecció basats en la comparació de poblacions que viuen en ambients diferenciats, o inclús en condicions extremes, han permès amb cert èxit inferir una relació causal més directa entre les possibles

forces selectives considerades a priori i les regions candidates identificades en el genoma. Aquest tipus de disseny aplicat a diverses poblacions humanes que viuen a gran altitud a les muntanyes del Tibet, els Andes o Etiòpia ha permès identificar, per exemple, diferents adaptacions genètiques humanes, algunes convergents, relacionades amb la hipòxia (Yi *et al.*, 2010; Huerta-Sánchez *et al.*, 2013; Borda *et al.*, 2020). En altres casos, la combinació de dades genètiques de tot el genoma amb dades fenotípiques relacionades amb la hipòtesi adaptativa proporciona també una estratègia força prometedora a l'hora d'intentar identificar empremtes de selecció relacionades amb fenotips adaptatius particulars, com, per exemple, la baixa estatura en poblacions pigmees de l'Àfrica (Jarvis *et al.*, 2012; Mendizabal *et al.*, 2012) o de les illes Andaman (Mondal *et al.*, 2016).

Sens dubte, el nombre creixent d'anotacions funcionals a les regions codificadores i reguladores del genoma humà, la disponibilitat creixent d'estudis d'associació a tot el genoma per a diferents fenotips complexos, l'adopció de noves estratègies experimentals per interrogar la relació entre el genotip i el fenotip i la incorporació d'altres models de selecció permetrà, no només millorar la comprensió de les diferents adaptacions ocorregudes en la nostra història evolutiva i que, per tant, ens defineixen com a humans, sinó identificar a la vegada variants funcionals en el nostre genoma amb potencial rellevància biomèdica i certa especificitat poblacional.

### Agraïments

Aquesta revisió ha estat possible gràcies al suport de la Direcció General de Recerca, Generalitat de Catalunya (2017SGR702), el Ministerio de Ciencia e Innovación i l'Agència Estatal de Investigació (AEI) mitjançant el projecte PID2019-110933GB-I00/AEI/10.13039/501100011033 i la Unidad de Excelencia María de Maeztu, finançada per l'AEI (CEX2018-000792-M).

## Bibliografia

- AKEY, J. M. (2009). «Constructing genomic maps of positive selection in humans: Where do we go from here?». *Genome Res.* 19: 711–722.
- AMORIM, C. E. G., *et al.* (2017). «Genetic signature of natural selection in first Americans». *Proc. Natl. Acad. Sci.* 114: 2195–2199.
- BORDA, V., *et al.* (2020). «The genetic structure and adaptation of Andean highlanders and Amazonians are influenced by the interplay between geography and culture». *Proc. Natl. Acad. Sci.* 117: 32557–32565.
- DANNEMANN, M., *et al.* (2016). «Introgression of Neandertal- and Denisovan-like haplotypes contributes to adaptive variation in human Toll-like receptors». *Am. J. Hum. Genet.* 98: 22–33.
- DARWIN, C.; WALLACE, A. (1858). «On the Tendency of Species to form Varieties; and on the Perpetuation of Varieties and Species by Natural Means of Selection». *J. Proc. Linn. Soc. London. Zool.* 3: 45–62. [versió catalana: *La lluita per la vida*. Trad.: J. Peretó. València: PUV, 2008, 31–64].
- DAUB, J. T., *et al.* (2013). «Evidence for polygenic adaptation to pathogens in the human genome». *Mol. Biol. Evol.* 30: 1544–1558.
- FAY, J. C.; WU, C. (2000). «Hitchhiking under positive Darwinian selection». *Genetics* 155: 1405–1413.
- FERRER-ADMETLLA, A., *et al.* (2014). «On detecting incomplete soft or hard selective sweeps using haplotype structure». *Mol. Biol. Evol.* 31: 1275–1291.
- FIELD, Y., *et al.* (2016). «Detection of human adaptation during the past 2000 years». *Science* 354: 760–764.
- FU, Y. X.; LI, W. H. (1993). «Statistical tests of neutrality of mutations». *Genetics* 133: 693–709.
- FU, Y. X. (1997). «Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection». *Genetics* 147: 915–925.
- FUMAGALLI, M., *et al.* (2015). «Greenlandic Inuit show genetic signatures of diet and climate adaptation». *Science* 349: 1343–1347.
- GARUD, N. R.; ROSENBERG, N. A. (2015). «Enhancing the mathematical properties of new haplotype homozygosity statistics for the detection of selective sweeps». *Theor. Popul. Biol.* 102: 94–101.
- GROSSMAN, S. R., *et al.* (2010). «A composite of multiple signals distinguishes causal variants in regions of positive selection». *Science* 327: 883–886.
- HANCOCK, A. M., *et al.* (2011). «Adaptations to climate-mediated selective pressures in humans». *PLoS Genet.* 7: e1001375.
- HUERTA-SÁNCHEZ, E., *et al.* (2013). «Genetic Signatures Reveal High-Altitude Adaptation in a Set of Ethiopian Populations». *Mol. Biol. Evol.* 30: 1877–1888.
- HUERTA-SÁNCHEZ, E., *et al.* (2014). «Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA». *Nature* 512: 194–197.
- JARVIS, J. P., *et al.* (2012). «Patterns of ancestry, signatures of natural selection, and genetic association with stature in Western African pygmies». *PLoS Genet.* 8: e1002641.
- KRONENBERG, Z. N., *et al.* (2018). «High-resolution comparative analysis of great ape genomes». *Science* 360: eaar6343.
- MARQUES-BONET, T., *et al.* (2009). «A burst of segmental duplications in the genome of the African great ape ancestor». *Nature* 457: 877–881.
- MENDIZABAL, I., *et al.* (2012). «Adaptive evolution of loci covarying with the human African Pygmy phenotype». *Hum. Genet.* 131: 1305–1317.
- MONDAL, M., *et al.* (2016). «Genomic analysis of Andamanese provides insights into ancient human migration into Asia and adaptation». *Nat. Genet.* 48: 1066–1070.
- PRITCHARD, J. K., *et al.* (2010). «The Genetics of Human Adaptation: Hard Sweeps, Soft Sweeps, and Polygenic Adaptation». *Curr. Biol.* 20: R208–R215.
- PYBUS, M., *et al.* (2015). «Hierarchical boosting: A machine-learning framework to detect and classify hard selective sweeps in human populations». *Bioinformatics* 31: 3946–3952.
- RACIMO, F., *et al.* (2015). «Evidence for archaic adaptive introgression in humans». *Nat. Rev. Genet.* 16: 359–371.
- REES, J. S., *et al.* (2020). «The Genomics of Human Local Adaptation». *Trends Genet.* 36: 415–428.
- REFOYO-MARTÍNEZ, A., *et al.* (2019). «Identifying loci under positive selection in complex population histories». *Genome Res.* 29: 1506–1520.
- SABETI, P. C., *et al.* (2006). «Positive natural selection in the human lineage». *Science* 312: 1614–1620.
- SABETI, P. C., *et al.* (2007). «Genome-wide detection and characterization of positive selection in human populations». *Nature* 449: 913–918.
- SPEIDEL, L., *et al.* (2019). «A method for genome-wide genealogy estimation for thousands of samples». *Nat. Genet.* 51: 1321–1329.
- SZPAK, M., *et al.* (2018). «FineMAV: Prioritizing candidate genetic variants driving local adaptations in human populations». *Genome Biol.* 19: 5.
- TAJIMA, F. (1989). «Statistical method for testing the neutral mutation hypothesis by DNA polymorphism». *Genetics* 123: 585–595.
- VITTI, J. J., *et al.* (2013). «Detecting natural selection in genomic data». *Annu. Rev. Genet.* 47: 97–120.
- VOIGHT, B. F., *et al.* (2006). «A map of recent positive selection in the human genome». *PLoS Biol.* 4: e72.
- WALSH, E. C., *et al.* (2006). «Searching for signals of evolutionary selection in 168 genes related to immune function». *Hum. Genet.* 119: 92–102.
- WANG, L., *et al.* (2020). «A MicroRNA Linking Human Positive Selection and Metabolic Disorders». *Cell* 183: 684–701.
- YANG, Z.; NIELSENT, R. (2002). «Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages». *Mol. Biol. Evol.* 19: 908–917.
- YI, X., *et al.* (2010). «Sequencing of 50 human exomes reveals adaptation to high altitude». *Science* 329: 75–78.
- ZHANG, J., *et al.* (2005). «Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level». *Mol. Biol. Evol.* 22: 2472–2477.