

TAXES EVOLUTIVES I EVOLUCIÓ DE GENS I GENOMES

LLUÍS SERRA¹ I ANTONIO FONTDEVILA²

¹ *Departament de Genètica, Facultat de Biologia, Universitat de Barcelona.*

² *Departament de Genètica i Microbiologia, Facultat de Biociències, Universitat Autònoma de Barcelona.*

Adreça per a la correspondència: Lluís Serra. Departament de Genètica, Facultat de Biologia, Universitat de Barcelona. Diagonal, 645. 08028 Barcelona.
Adreça electrònica: lserra@ub.edu.

RESUM

S'introdueix el concepte de *taxa evolutiva* i es discuteix el seu significat, així com també la seva aplicació per determinar l'impacte potencial que poden tenir les alteracions mediambientals provocades per l'acció humana. S'analitzen també els canvis evolutius en seqüències de nucleòtids o d'aminoàcids i la dinàmica de les taxes de substitució. Això permet introduir el concepte de *rellotge molecular* i caracteritzar la teoria neutralista com una referència estàndard per interpretar les taxes evolutives. S'explica també la utilització de seqüències d'àcids nucleics o de proteïnes per establir les relacions evolutives entre les espècies, amb un exemple de construcció d'un arbre filogenètic i d'un mètode estadístic per determinar-ne la fiabilitat. Tot seguit es consideren els canvis genètics que ens fan diferents del ximpanzé, el nostre parent més proper, i els mecanismes d'evolució de la grandària del genoma mitjançant duplicacions intragèniques, redistribució d'exons, duplicacions de gens complets i duplicacions genòmiques. Finalment, es considera el significat evolutiu dels elements genètics transposables.

Paraules clau: taxa evolutiva, rellotge molecular i neutralisme, filogènia molecular, evolució de gens i genomes.

EVOLUTIONARY RATES AND EVOLUTION OF GENES AND GENOMES

SUMMARY

We introduce the concept of evolutionary rate and discuss its significance and application in the study of systematic human-induced alterations of natural habitats. The evolutionary changes in nucleotide or amino acid sequences and the dynamics of substitution rates are analysed, introducing the concept of molecular clock and the neutral theory as a null hypothesis to interpret evolutionary rates. The use of nucleotide or amino acid se-

quences to establish evolutionary relationships between species is also discussed, giving an example of the construction of a phylogenetic tree and a statistical method to determine tree reliability. We then proceed to consider those genetic changes that make us different from the chimpanzee, our closest relative, and to discuss the mechanisms responsible of genome-size evolution, in particular intragenic duplication and gene elongation, exon shuffling, gene duplication, and genome duplication. Finally, we discuss the evolutionary significance of transposable genetic elements.

Key words: evolutionary rate, the molecular clock and the neutral theory, molecular phylogeny, evolution of genes and genomes.

TAXES EVOLUTIVES

Les taxes evolutives s'utilitzen per descriure la dinàmica dels canvis en una línia filètica, al llarg de les generacions. Els canvis poden ser en el mateix genoma o en l'expressió fenotípica de fenòmens genètics subjacents. Simpson (1944) va ser el primer a fer una anàlisi comparativa de les taxes d'evolució en diferents grups taxonòmics i Haldane (1949) va desenvolupar algunes mesures quantitatives per estimar les taxes d'evolució. Una aplicació de les taxes evolutives és la determinació de l'impacte potencial que poden tenir les alteracions mediambientals provocades per l'acció humana (Balanyà *et al.*, 2006). En particular, és important valorar si les poblacions o les espècies poden respondre a pressions selectives molt variables de manera suficientment ràpida per evitar l'extinció. L'anàlisi de les taxes evolutives en estudis recents permet concloure que l'evolució fins ara considerada «ràpida» pot ser la norma i no l'excepció. Sembla que quan les poblacions o espècies s'exposen a ambients variables, com succeeix en les colonitzacions o les invasions, l'evolució és més ràpida que en situacions més estabilitzades i en mesures fetes en períodes més llargs.

En els estudis microevolutius interessa conèixer quina és la velocitat d'evolució a la natura i si les trajectòries evolutives són predictibles o idiosincràtiques (Huey *et al.*, 2000). Obtenir aquesta informació és fona-

mental per poder predir les respostes evolutives en ambients naturals i alterats per l'home. Sovint, les taxes microevolutives s'estimen mesurant els canvis fenotípics en poblacions locals, al llarg del temps. La predictibilitat de l'evolució s'estima determinant si poblacions replicades presenten respostes convergents. Les espècies recentment introduïdes, que colonitzen ràpidament extenses àrees geogràfiques anàlogues a les originals, són excel·lents rèpliques poblacionals, que ofereixen oportunitats úniques per mesurar les taxes microevolutives i la predictibilitat dels canvis a escala geogràfica. Si aquestes espècies desenvolupen clines (canvis sistemàtics de les freqüències de marcadors genètics o morfològics seguint gradients ambientals) que convergeixen amb les existents en les poblacions ancestrals, això s'interpreta com una prova de l'existència d'una evolució ràpida i predictable. En un estudi realitzat amb l'espècie *Drosophila subobscura* (Huey *et al.*, 2000), s'ha determinat la taxa i la predictibilitat dels canvis microevolutius, a escala intercontinental. En les poblacions ancestrals europees de l'espècie, la longitud de l'ala augmenta amb la latitud. En les poblacions colonitzadores d'Amèrica del Nord després de dues dècades va evolucionar una clina per a aquests caràcters, convergent amb la clina ancestral. La taxa d'evolució morfològica a escala intercontinental ha estat de les més ràpides descrites per a caràcters morfològics, fins i tot en comparació de les taxes

mesurades en poblacions locals. El fet que diferents seccions de l'ala siguin les que determinin principalment les clines a Europa i Amèrica del Nord demostra que l'evolució de la variabilitat geogràfica de la longitud de l'ala ha estat predictable, però la manera com això s'ha produït ha estat, en certa mesura, contingent. Han estat descrits altres casos d'evolució ràpida, entre els quals cal esmentar, com a més conegut, el dels canvis en la grandària del bec dels pinsans de Darwin després de períodes de sequera a les illes Galápagos (Grant i Grant, 1995). Tots aquests estudis donen suport al fet que la selecció natural pot explicar els canvis morfològics ràpids que observem en el procés evolutiu. Malgrat tot, les espècies que viuen en ambients molt estables durant llargs períodes de temps presenten taxes d'evolució fenotípica molt baixes. Com a exemple, podríem esmentar el cas del cranc *Limulus polyphemus*, un fòssil vivent, el qual té un fenotip molt semblant al dels exemplars fòssils de fa més de dos-cents milions d'anys. En l'altre extrem tindriem les notables diferències existents entre plantes cultivades i animals domèstics i els seus avantpassats salvatges, fruit de la selecció artificial practicada pels milloradors.

S'ha observat, però, que hi ha una relació inversa entre la taxa d'evolució i l'interval de temps en el qual s'ha mesurat aquesta taxa: els casos d'evolució més ràpida observats tendeixen a correspondre a intervals més curts de temps que els casos d'evolució més lenta. Aquest fet té una explicació senzilla: si la direcció de l'evolució fluctua, la taxa d'evolució mesurada en intervals de temps curts és més gran que la taxa mesurada en intervals més llargs, ja que els canvis que es produeixen en els subintervals es compensen. Per exemple, s'ha demostrat que el bec dels pinsans de les Galápagos va evolucionar en el sentit de fer-se més gros en èpoques d'escassetat, i més petit en èpoques d'abundor. Al llarg del temps, doncs,

aquests canvis es compensen i poden dificultar la detecció dels canvis morfològics ràpids produïts per la selecció.

Si considerem ara els mètodes d'estimació de les taxes evolutives en períodes més llargs, podríem analitzar, per exemple, l'estima de la taxa evolutiva de la domesticació del blat de moro (*Zea mays*), a partir del seu avantpassat *teosinte* (*Z. parviglumis*). En primer lloc, cal estimar el temps des de la divergència a partir d'un avantpassat comú, que en aquest cas és d'uns 7.500 anys, tal i com ho indiquen les dades arqueològiques de Mesoamèrica. La taxa evolutiva del canvi fenotípic es pot obtenir comparant els valors d'un o més caràcters d'interès. Aquests caràcters, però, poden dependre de les condicions ambientals, i per això és important que, en el nostre exemple, les plantes es cultivin en les mateixes condicions. Aleshores es pot calcular la taxa de canvi d'un fenotip com la diferència del valor mitjà del caràcter en les dues espècies dividit pel temps de divergència. Ara bé, cal observar que aquest càlcul pot ser enganyós, ja que l'avantpassat comú podria tenir un valor del caràcter que no fos intermedi al de les dues espècies. De fet, en l'exemple, l'avantpassat comú era més semblant al *teosinte* actual que no pas al blat de moro; la major part del canvi fenotípic s'hauria produït com a conseqüència de la ràpida evolució del blat de moro, en el procés de la seva domesticació.

CANVIS EVOLUTIUS EN SEQÜÈNCIES DE NUCLEÒTIDS O D'AMINOÀCIDS

Una substitució consisteix en el reemplaçament d'un aminoàcid o d'un nucleòtid per un altre en una població. Les substitucions determinen diferències fixades entre llinatges evolutius; per exemple, en els humans trobem l'aminoàcid Met (M) en la posició 12 del citocrom C, mentre que en els cavalls hi

ha l'aminoàcid Gln (Q) en la mateixa posició. No totes les substitucions de nucleòtids produeixen substitucions d'aminoàcids, a causa de la redundància del codi genètic. Al DNA cal distingir entre diferències silencioses o sinònimes, que no canvien la seqüència d'aminoàcids de la proteïna, i les diferències no sinònimes, que sí que la canvien. Quan es comparen dues seqüències és fàcil estimar el nombre de diferències que hi ha entre aquestes. Per exemple, els humans i el macaco rhesus difereixen en quatre dels cent quaranta-un aminoàcids de l' α -globina. Entre els humans i les vaques hi ha disset diferències. Tanmateix, el nombre de diferències no és necessàriament el mateix que el nombre total de substitucions des de la divergència. Això és degut al fet que un determinat lloc (aminoacídic o nucleotídic) pot haver estat substituït més d'una vegada al llarg del temps evolutiu. Per tant, el nombre de diferències observades és una estima per defecte del nombre de substitucions que s'espera que s'hagin produït. Com més temps fa que dues seqüències han divergit independentment, és més probable que un lloc determinat hagi estat substituït més d'una vegada i, per tant, que el nombre de diferències observades sigui inferior al de substitucions que s'han produït. Hi ha diferents mètodes per estimar el nombre de substitucions esperades (ocorregudes) a partir del de diferències observades.

Els estudis de les taxes evolutives de diferents nivells d'organització han conduït a conclusions contradictòries i, com a conseqüència, a una controvèrsia entre diferents doctrines de l'evolució. Moltes d'aquestes contradiccions són el resultat d'utilitzar diferents variables per mesurar les taxes d'evolució. L'evolució molecular de variants neutres, amb una taxa estable i uniforme, es contraposa amb l'evolució adaptativa no uniforme del fenotip.

TAXES DE SUBSTITUCIÓ I RELLOTGE MOLECULAR

Quan parlem de taxes d'evolució molecular ens referim a taxes de substitució, és a dir, al nombre esperat de substitucions de nucleòtids o d'aminoàcids per lloc i per unitat de temps. En comparar dues seqüències, s'utilitza el símbol p per indicar la proporció observada de diferències (que és equivalent al nombre observat de diferències per lloc), el símbol K per indicar el nombre estimat de substitucions per lloc i el símbol k per indicar la taxa de substitució. Normalment, les taxes de substitució s'expressen com el nombre estimat de substitucions per lloc i per any. Si T és el temps, en anys, transcorregut des de la divergència a partir d'un avantpassat comú, la taxa de substitució s'expressa com $k = K/2T$. Des de la divergència, en cada línia han transcorregut T anys d'evolució; per tant, el total d'anys d'evolució en el conjunt de les dues línies és $2T$. Per a un parell d'espècies es poden comptar el nombre de diferències d'aminoàcids en una proteïna determinada i obtenir també el temps de divergència entre aquestes espècies a partir del registre fòssil. Això s'ha fet en el cas de l' α -globina en vuit espècies de vertebrats (vegeu la figura 1). Quan els valors de K es representen gràficament, en funció del temps de divergència, s'obté una distribució de punts que s'ajusta molt bé a una línia recta (recta de regressió). El pendent d'aquesta recta (K/T) és una estimació del doble de la taxa de substitució d'aminoàcids per any (k). En altres proteïnes també s'han obtingut resultats semblants. Això va portar Zuckerkandl i Pauling (1965) a proposar que, per a una proteïna determinada, la taxa de substitució d'aminoàcids és aproximadament constant durant milions d'anys. Aquests autors varen utilitzar la denominació *rellotge molecular evolutiu* per descriure aquesta constància de la taxa de substitució.

L'any 1968 Kimura va proposar que la majoria de substitucions a nivell molecular eren el resultat de la deriva genètica i no de la selecció natural, la qual cosa implica que la majoria de substitucions no afecten l'aptitud (*fitness*) dels genotips, són neutres. El suport d'aquesta proposta el tenim en dues observacions: la constància de les taxes d'evolució (k) aminoacídica o nucleotídica en diferents llinatges i l'abundància de variabilitat molecular, ambdues observacions totalment incompatibles amb una evolució darwinista per selecció natural. Estudis posteriors, més acurats, de les taxes de substitució, han obligat a perfilar més aquesta proposta que ha conduït a una teoria neutra de l'evolució molecular (neutralisme), que comprèn no solament l'estudi de les mutacions neutres sinó també les quasi neutres, i combina en la seva formalització la selecció i la deriva genètica.

El neutralisme (Kimura, 1983) s'utilitza com a referència estàndard per interpretar les taxes evolutives; de fet, ens ha proporcionat una hipòtesi nul·la contra la qual podem detectar la presència o no de selecció natural i ha conduït al concepte de *rellotge molecular*. La teoria de la genètica de poblacions demostra que la taxa evolutiva esperada de canvi genètic en el cas de mutacions neutres depèn només de la taxa en la qual es produeixen aquestes mutacions, i no de la grandària de la població o de la selecció natural. Aquest resultat es pot justificar de la manera següent: si μ és la taxa de mutació neutra i N la grandària efectiva de la població, en cada generació es produiran $2N\mu$ mutacions neutres noves en una població d'organismes diploides. Com són neutres, cadascuna d'aquestes mutacions no té una probabilitat major o menor de ser substituïda que qualsevol dels $2N$ al·lels presents en la població, és a dir, la probabilitat de fixació és $1/2N$; això implica que la taxa de substitució de les mutacions neutres és $2N\mu/2N = \mu$. En altres paraules, la taxa d'evolució

molecular, en el cas de mutacions neutres, es comporta com un rellotge estocàstic, que fa «tic-tac» amb la taxa μ .

Les taxes de substitució d'aminoàcids són molt diferents segons el gen considerat. Els fibrinopèptids, per exemple, evolucionen amb una taxa nou vegades més ràpida que la ubiquitina. Segons els neutralistes, aquesta variació de la taxa en diferents proteïnes s'explica per les diferències en les restriccions selectives i no pas per selecció positiva. Intentarem aclarir aquest punt. Com més restringit és un gen, des d'un punt de vista funcional, més gran és la proporció de mutacions deletèries que pot tenir i, per tant, n'hi ha menys que siguin neutres, la qual cosa farà que la taxa de substitució nucleotídica observada sigui menor. En el cas de gens menys restringits, hi haurà menys mutacions deletèries i més mutacions neutres, i això vol dir que la taxa de substitució s'incrementarà. Per tant, les diferències en la restricció funcional —és a dir, en la intensitat de la selecció negativa— determinaran diferències en les taxes de substitució neutra. Els pseudogens —còpies inactives de gens funcionals— són un cas extrem de la relació entre restricció funcional i taxa de substitució. En no tenir cap restricció funcional, i d'acord amb la teoria neutralista, aquests pseudogens tenen unes taxes de substitució molt altes. Els fibrinopèptids, que semblen no tenir cap funció, presenten també una taxa de substitució alta i representen un cas menys extrem de molècula poc restringida. En l'altre extrem, les histones són molècules molt restringides que tenen una de les taxes de substitució més baixes conegudes (nou-cents vegades menor que els fibrinopèptids), a causa de la gran especificitat funcional i estructural del complex DNA-histona.

En el cas de les posicions nucleotídiques de les regions codificadores dintre dels gens funcionals, la qüestió no és tan clara. A causa de la degeneració del codi genètic,

les mutacions en diferents posicions dels codons tenen una probabilitat diferent d'alterar l'aminoàcid codificat; per tant, les tres posicions del codó difereixen en la seva restricció funcional. En els gens de les globines, per exemple, les terceres posicions dels codons, la substitució de les quals no comporta en molts casos un canvi de l'amino-

àcid codificat (són sinònimes), evolucionen tres vegades més ràpidament que les dues primeres posicions. S'ha demostrat que, en general, els llocs sinònims (silenciosos) evolucionen més ràpidament que els no sinònims i aquests últims tenen taxes molt més variables a causa de la selecció.

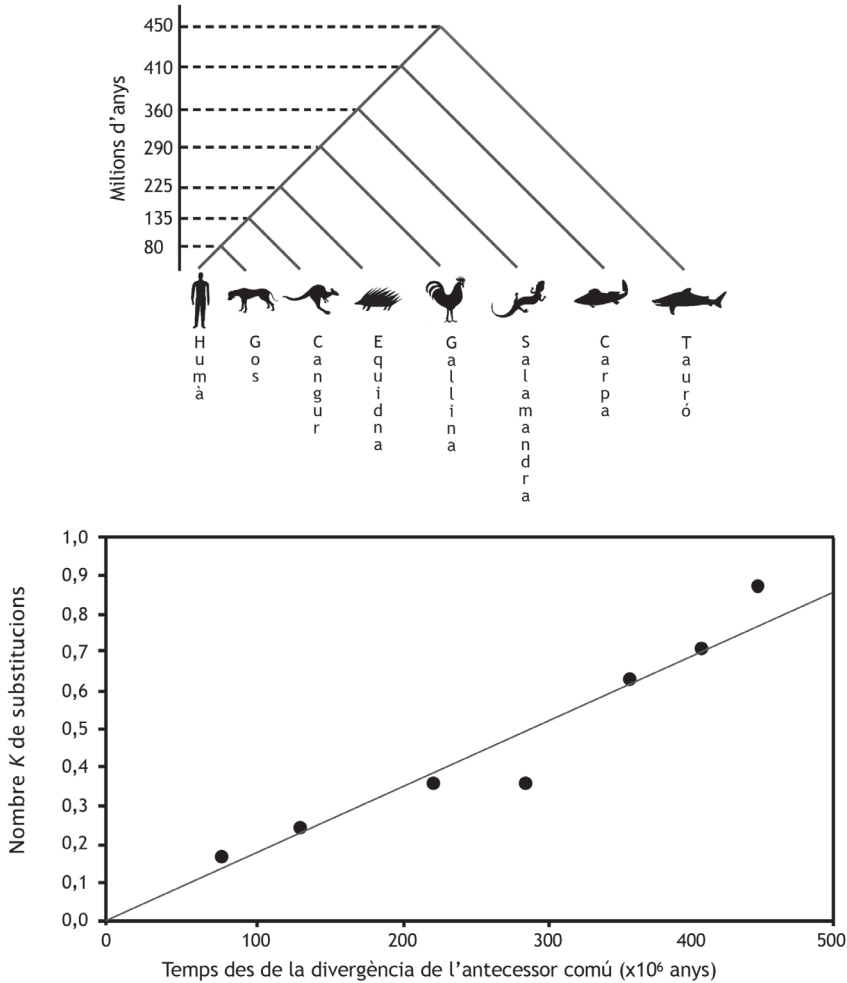


FIGURA 1. El rellotge molecular. En la figura es representa un arbre (cladograma) de vuit espècies en què el temps de divergència està calibrat segons el nombre de substitucions d'aminoàcids en l' α -globina des de l'avantpassat comú. La regressió entre el temps i el nombre de substitucions és significativa i indica la constància de la taxa de substitució molecular, la qual s'estima com la meitat del coeficient de regressió (modificat de Fontdevila i Moya, 2003).

FILOGÈNIA MOLECULAR I GENÒMICA COMPARADA

La filogènia molecular utilitza les seqüències d'àcids nucleics o de proteïnes per establir relacions evolutives entre les espècies. Algunes filogènies, en estar basades en poques seqüències, poden ser enganyoses i donar resultats diferents en funció de les seqüències comparades. Per exemple, la comparació de proteïnes implicades en la transcripció, la traducció, la replicació del DNA i la reparació, destaca les semblances entre les arquees i els eucariotes, mentre que les comparacions de les proteïnes implicades en el metabolisme celular agrupen les arquees amb els bacteris. Els projectes genòmics han permès canviar tot aquest panorama. En poder fer comparacions de seqüències genòmiques completes, s'ha obtingut una visió global de les proximitats evolutives, la qual cosa ha motivat el naixement d'una ciència nova, la genòmica comparada. Aquestes dades també ens permeten entendre com els genomes actuals es relacionen entre si, la qual cosa permet identificar seqüències conservades importants.

Si dues seqüències o més presenten un grau de semblança suficient (homologia de seqüències), es pot suposar que procedeixen d'una seqüència ancestral comuna. Els alineaments de seqüències s'utilitzen per obtenir índexs i estadístics que permeten descriure el grau de relació entre les seqüències. Comparar seqüències de la mateixa longitud és relativament senzill, si presenten un cert grau d'homologia. Però, molt sovint, les seqüències de DNA que es comparen han sofert delecions o insercions, i això fa que calguin mètodes matemàtics i programes d'ordinador per fer els alineaments. Un cop s'han obtingut els alineaments, es pot construir un arbre evolutiu.

Hi ha diferents mètodes. Molts utilitzen el de les matrius de distàncies. A la figura 2 es

mostra el procediment anomenat UPGMA per al conjunt de les quatre seqüències (S_1 - S_4) de la taula 1, en què els guions indiquen la presència del mateix nucleòtid que la seqüència S_1 .

El primer pas de l'algorisme consisteix a calcular la distància evolutiva entre tots els parells de seqüències de la base de dades i disposar-les en una matriu. Aquesta distància evolutiva es pot expressar com el nombre de diferències nucleotídiques (el cas del nostre exemple de la figura 2) o de substitucions d'aminoàcids entre les dues seqüències, o el nombre de diferències nucleotídiques o aminoacídiques per lloc. El pas següent consisteix a unir els dos membres del parell de seqüències que tenen la distància (D) més petita, per exemple, S_1 i S_2 en el nostre cas, amb un node arrel entre aquestes situat a una distància $D/2$. Després d'aquest primer agrupament, S_1 i S_2 es consideren com una sola seqüència, i es construeix una nova matriu de distàncies. El procés es repeteix fins que s'han col·locat en l'arbre totes les seqüències. Hi ha d'altres mètodes per construir arbres, com el de la màxima parsimònia o el de la màxima versemblança. Aquests altres mètodes consideren tots els arbres possibles que podrien explicar les relacions observades entre les seqüències i en seleccionen aquells que requereixen el menor nombre de passos. Un cop s'ha obtingut un arbre evolutiu, s'utilitzen mètodes estadístics per determinar-ne la fiabilitat. D'aquests, el més popular és el mètode de *bootstrap*, un tipus de simulació de Monte Carlo. El *bootstrap* és una tècnica estadística que permet determinar empíricament la variabilitat d'un estimador. Consisteix a mostrejar repetidament, amb reemplaçament, les dades d'una mostra experimental i obtenir mostres fictícies de la mateixa grandària que l'original. Vegem-ne ara un exemple concret, utilitzant seqüències de nucleòtids. Considerem l'alineament de cinc seqüències i deu nucleòtids (vegeu la taula 2).

S'escull a l'atzar un lloc (una columna) i s'utilitza com el primer lloc del pseudoalineament. Després, s'escull un altre lloc a l'atzar i s'utilitza com a segon lloc i el procés continua fins que el pseudoalineament conté el mateix nombre de llocs que l'alineament original. Cal advertir que el mostreig

de l'alineament original es fa amb reemplaçament, la qual cosa significa que el mateix lloc (columna) es pot incloure més d'una vegada en el pseudoalineament, o pot ser que no aparegui. Un possible pseudoalineament podria ser el de la taula 3. A partir d'aquest pseudoalineament, es

	S ₁	S ₂	S ₃
S ₂	D ₁₂ = ⑤	—	—
S ₃	D ₁₃ = ⑧	D ₂₃ = ⑩	—
S ₄	D ₁₄ = ⑭	D ₂₄ = ⑯	D ₃₄ = ⑦

	S ₁ S ₂	S ₃
S ₃	D ₍₁₂₎₃ = 1/2 (D ₁₃ + D ₂₃) = ⑨	—
S ₄	D ₍₁₂₎₄ = 1/2 (D ₁₄ + D ₂₄) = ⑮	D ₃₄ = ⑦

$$D_{(12)(34)} = \frac{1}{2} [D_{(12)3} + D_{(12)4}] = ⑫$$

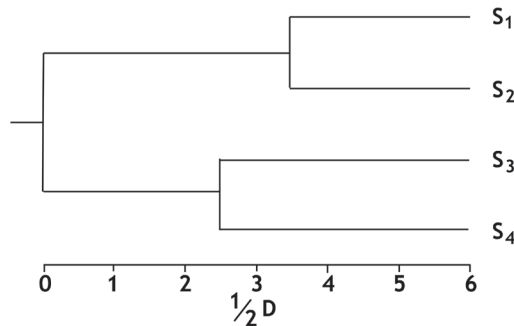


FIGURA 2. El mètode UPGMA de reconstrucció filogenètica. La matriu primera (a dalt) dona les distàncies D en diferències nucleotídiques entre les quatre seqüències de l'exemple del text. La segona matriu conté totes les distàncies una vegada s'han agrupat els taxons S₁ i S₂ en un sol taxó. Les distàncies entre S₁-S₂ i els altres taxons es calculen com la mitjana de les distàncies entre cada taxó original (S₁; S₂) i l'altre taxó. Finalment, s'agrupen els taxons S₃ i S₄ (els de distància més petita en la segona matriu) en un únic taxó i es calcula la distància entre S₁-S₂ i S₃-S₄. La construcció de l'arbre es fa situant els nodes arrel de cada clade a la meitat de la distància entre els dos taxons que formen el clade.

TAULA 1

S ₁	A	C	G	T	C	G	A	T	T	C	C	G	T	A	A	T	C	T	G	A	A	C
S ₂	-	-	-	A	-	A	T	-	-	-	-	-	-	C	-	-	T	-	-	-	-	-
S ₃	T	-	-	C	-	-	-	A	A	-	-	T	-	T	-	C	A	-	-	-	-	-
S ₄	T	G	C	G	-	-	-	A	A	-	T	T	-	T	-	C	A	A	-	T	-	T

TAULA 2

1	2	3	4	5	6	7	8	9	0
A	A	T	G	G	G	A	T	T	T
C	C	C	G	G	G	G	C	C	G
A	A	T	G	G	A	G	A	T	T
T	A	T	G	G	C	G	G	T	T
T	G	T	G	G	G	C	A	T	T

TAULA 3

4	8	3	6	0	2	4	9	5	1
G	T	T	G	T	A	G	T	G	A
G	C	C	G	G	C	G	C	G	C
G	A	T	A	T	A	G	T	G	A
G	G	T	C	T	A	G	T	G	T
G	A	T	G	T	G	G	T	G	T

construeix un arbre utilitzant el mateix mètode que l'emprat en la construcció de l'arbre original. Aleshores es compara l'arbre original amb l'arbre nou. Per a cada clade (conjunt d'espècies que s'originen a partir d'un node arrel) existent en l'arbre original, s'assigna el valor 1 si el clade també és en l'arbre nou, i el valor 0 si no hi és. Tot aquest procés constitueix una mostra *bootstrap*. S'emmagatzema la puntuació obtinguda per cada clade i s'inicia el següent cicle de *bootstrap* (Hall, 2001). Generalment es fan de cent a mil repeticions *bootstrap* per estimar la fiabilitat de l'arbre obtingut. Al final, s'imprimeix un arbre en què es mostra, per a cada clade, el nombre de vegades (o fracció) que el clade estava present en les repeticions *bootstrap*. Quan aquest valor és, per exemple, del 95 % o superior, podem confiar en la fiabilitat del clade, però no pas quan és inferior. Els valors de *bootstrap* entre noranta-cinc i cent in-

diquen un nivell de confiança alt del clade corresponent.

COMPLEXITAT ENFRONT DIVERGÈNCIA

Una característica fonamental del genoma és la seva repetibilitat, producte de la multiplicació total o parcial de seqüències, que genera grandàries genòmiques moltes vegades no correlacionades amb el nombre de gens codificants. La raó bàsica és que gran part d'organismes tenen una elevada proporció de DNA, entre el 30 i el 99 %, que no codifica. Així, els lliris tenen una mida genòmica de noranta gigabases (una gigabase equival a mil milions de bases) i uns cinquanta mil gens codificants, mentre que els humans tenim uns trenta mil gens i una mida de només 3,3 gigabases. Aquesta desproporció entre nombre de gens i la grandària del genoma (valor C) es coneix com la *paradoxa del valor C*. Tampoc no hi ha una relació directa entre el nombre de gens i la complexitat de l'organisme. Els humans ens arreglem amb uns trenta mil gens, mentre que alguns peixos i moltes plantes en tenen més de cinquanta mil. Aquestes i altres paradoxes aparents són el resultat de l'evolució del genoma i han quedat paleses en els estudis de genòmica comparada.

La comparació de genomes complets ha fet palès un fet sorprenent: els diferents animals comparteixen essencialment els mateixos gens. Amb molt poques excepcions, a cada gen humà li correspon un gen equivalent en el ratolí. Això vol dir que no s'han «inventat» gens nous durant els cin-

quanta milions d'anys transcorreguts des de la divergència de les dues espècies. Anàlogament, l'home i el peix globus es varen separar fa uns quatre-cents milions d'anys, però els dos genomes contenen el mateix nombre de gens i la majoria d'aquests —més de les tres quartes parts— es poden alinear sense dificultat. La conservació genètica observada en els vertebrats també es troba en altres cordats, com *Ciona*, el qual conté la meitat dels gens presents en els vertebrats, dels quals es va separar fa més de cinc-cents milions d'anys. Quasi dos terços dels gens estructurals de *Ciona* tenen un equivalent en els vertebrats. L'increment en el nombre de gens dels vertebrats és degut, principalment, a duplicacions de gens que ja existien en *Ciona*. Aquest grau tan alt de conservació dels gens existents en diferents animals ha posat de manifest la importància dels canvis en l'expressió gènica com a mecanisme generador de diversitat evolutiva. Aquesta es pot generar mitjançant l'expressió d'un conjunt fixat de gens en diferents patrons. És important indicar, però, que aquest fenomen que trobem en els animals no es dona en el cas dels bacteris. Els bacteris varen aparèixer fa uns tres mil milions d'anys, mentre que els animals ho varen fer fa uns cinc-cents milions d'anys. La ràpida evolució dels bacteris, conjuntament amb la seva llarga història evolutiva, ha donat lloc a diferents formes de metabolisme, la qual cosa fa que puguin viure en ambients molt diversos i en condicions molt extremes. Existeix una enorme variació, tant en el nombre com en els tipus de gens dels diferents genomes bacterians. Els bacteris més senzills, com *Mycoplasma*, contenen només uns cinc-cents gens, mentre que els més sofisticats com *Streptomyces* contenen uns set mil gens. Aquest interval de variació tan gran contrasta amb l'observat en els animals (tan sols el doble). El contingut genètic també és molt divergent, fins i tot en espècies bacterianes molt properes. Per

exemple, *Stephococcus* i *Escherichia coli* varen divergir fa uns cinquanta milions d'anys, a partir d'un avantpassat comú, la qual cosa és comparable al temps de divergència entre l'home i el ratolí. Tanmateix, ambdós bacteris comparteixen només el 75 % dels gens estructurals (aquells que codifiquen proteïnes). La resta són únics, sense cap equivalent en l'altra espècie.

La comparació entre genomes demostra que la densitat de gens difereix molt entre organismes, per la qual cosa la grandària del genoma no és indicativa del nombre de gens. Això, en part, és el resultat de l'aparició dels introns en els eucariotes i la tendència a acumular DNA repetitiu dintre dels introns i en les regions intergèniques, a mesura que els genomes esdevenen més complexos. Per exemple, el 45 % del genoma humà està format per repeticions d'elements mòbils, mentre que aquest valor en el ratolí baixa al 37 % i els valors equivalents en *Drosophila* o *Caenorhabditis* són molt inferiors. Un 80 % de les proteïnes del ratolí tenen els seus ortòlegs, estrictament 1:1, en el genoma humà, amb una identitat de seqüència que oscil·la entre el 70 i el 100 %. Sembla que les proteïnes que evolucionen més ràpidament estan implicades en el sistema immunitari o de defensa, per exemple els gens del complex major d'histocompatibilitat (MHC), o en processos de reproducció. Per a algunes d'aquestes proteïnes s'ha identificat l'existència d'una selecció positiva. Moltes proteïnes humanes i del ratolí pertanyen a famílies gèniques que han experimentat una expansió diferencial en almenys un dels dos genomes, fet que determina que no es trobi una estricta relació 1:1, encara que, quan varia el nombre de gens en una família gènica, pot ser difícil identificar els ortòlegs autèntics, sobretot si hi ha una divergència de seqüències important.

La complexitat genètica d'un organisme no es pot extrapolar directament a partir del nombre de gens. Per exemple, l'em-

palpament alternatiu, mitjançant el qual en els transcrits del mateix gen pot haver-hi representades diferents combinacions d'exons, pot incrementar molt la complexitat proteica. Aquest mecanisme es produeix en la majoria de gens humans. Això vol dir que els trenta mil gens que s'estima que té el genoma humà poden codificar un nombre molt més alt de proteïnes. Un altre fenomen que també pot contribuir a generar més complexitat en el genoma és el cas de l'anomenat *quimerisme en tàndem*: dos gens adjacents situats en la mateixa orientació, que normalment es transcriuen independentment, a vegades es poden transcriure simultàniament en una única seqüència de RNA (transcrit primari). El producte de l'empalmament d'aquest transcrit primari codifica una proteïna que està determinada per exons dels dos gens. Aquest fenomen és diferent del cas dels operons policistrònics dels procariotes, en què un sol transcrit es tradueix en diferents proteïnes, però sense produir-ne de quimèriques.

Quins canvis genètics ens fan diferents del ximpanzé, el nostre parent més proper? Els científics s'han fet aquesta pregunta durant dècades, i la publicació de l'esborrany del genoma del ximpanzé (*Pan troglodytes*) el setembre de 2005 ha significat un pas endavant envers l'obtenció d'una resposta. Una manera de determinar quines són les diferències importants entre els dos genomes és identificar els canvis evolutius específics dels humans. Una altra és investigar l'existència d'empremtes (signatures) de la selecció natural positiva en les seqüències dels dos genomes. Les dades noves mostren que el genoma humà i el del ximpanzé difereixen tan sols en un 1,2 % en termes de substitucions nucleotídiques. La divergència entre les dues seqüències varia en funció de la regió genòmica analitzada a causa, segurament, de variacions locals de les taxes de mutació, de les constriccions selectives i de la taxa de recombinació en-

tre parells de cromosomes durant la divisió cel·lular. La divergència màxima correspon al cromosoma Y i la mínima al cromosoma X. Independentment del fet que quan es quantifiquen les divergències entre les seqüències se solen utilitzar les substitucions nucleotídiques, una gran proporció de les diferències entre el genoma humà i el del ximpanzé és deguda a insercions/deleccions (*indels*) i a duplicacions recents de segments de DNA (el 3 % i el 2,7 %, respectivament). Més d'una tercera part dels *indels* són deguts a seqüències repetides i, aproximadament, una quarta part a elements transposables. A més a més d'aquestes diferències, aproximadament la meitat dels gens de les regions duplicades específiques de l'ésser humà presenten una taxa d'expressió diferencial en comparació amb el ximpanzé, i la majoria estan sobreexpressats.

En resum, malgrat que els dos genomes són molt semblants, hi ha uns trenta-cinc milions de diferències nucleotídiques, cinc milions d'*indels* i molts reordenaments cromosòmics que cal tenir en compte. És probable que la major part d'aquest canvis no tinguin un efecte biològic significatiu; per tant, la identificació de les diferències genòmiques que determinen les característiques «humanes», com la gran capacitat cranial, la posició bípeda i el desenvolupament del cervell, és una tasca molt feixuga. A causa del poc temps transcorregut des de la separació d'ambdues espècies, és probable que les diferències fenotípiques existents entre l'home i el ximpanzé siguin degudes a poques mutacions en les regions cis-reguladores, la combinació de les quals generi efectes importants.

EVOLUCIÓ DE GENS I GENOMES

La influència de la genètica en la teoria de l'evolució ha estat permanent durant tot el segle xx. Actualment, la comparació de

genomes complets ha esdevingut el nucli fonamental d'una disciplina nova, la genòmica comparada, la qual ha permès d'obtenir noves dades significatives referents a la posició de les espècies en l'arbre de la vida. En particular, ja hem vist que les dades comparatives permeten d'obtenir informació rellevant sobre la ubicació vertadera de l'ésser humà, sobretot quan es compara la gran semblança del genoma humà amb els d'altres primats.

Avui sabem que els gens i les proteïnes dels eucariotes superiors són més grans i més complexos que els dels organismes més senzills. Els gens dels eucariotes solen tenir grans introns, la mida mitjana dels quals reflecteix normalment la complexitat del genoma. Les seqüències codificadores dels gens dels eucariotes són també típicament més llargues que les dels gens dels procarïotes. Aquesta complexitat és el resultat de l'oportunitat de l'evolució sobre diferents mecanismes genòmics, entre els quals destaquen les duplicacions i les transposicions que permeten l'expansió i la diversificació de les seqüències codificadores.

Els gens complexos poden evolucionar per duplicació intragènica, sobretot per la duplicació d'exons, mitjançant diferents mecanismes, com per exemple per entrecreuament desigual entre cromàtides germanes o no germanes. Hi ha moltes observacions que afavoreixen aquest procés. Per exemple, alguns gens codifiquen polipèptids, les seqüències dels quals estan formades per dominis repetits, amb una homologia de seqüència molt alta entre aquests; per exemple, la involucrina té cinquanta-nou còpies d'un domini, de les quals les trenta-nou còpies centrals tenen una homologia de seqüència molt elevada. Sabem que un 10 % dels gens humans, de *Drosophila melanogaster* i de *Caenorhabditis elegans* (un nematode) tenen exons duplicats. Això té alguns avantatges: els dominis repetits poden millorar l'aptitud, sobretot en el cas de les proteïnes

que tenen una funció estructural important. Moltes vegades, després de la duplicació intragènica es produeix una divergència nucleotídica entre les diferents unitats repetides, la qual cosa pot permetre adquirir funcions diferents, encara que relacionades. També es pot generar diversitat mitjançant l'empalmament alternatiu. Aquest mecanisme selecciona diferents conjunts d'exons d'un grup d'exons duplicats per donar lloc a diversos productes gènics finals. Com que els exons duplicats presenten petites diferències entre si, això dóna lloc a formes proteïques lleugerament diferents (isoformes).

En els organismes pluricel·lulars moltes proteïnes contenen dominis que es troben també en d'altres proteïnes. La fibronectina, una proteïna de la matriu extracel·lular, conté molts dominis repetits. Un dels dominis repetits, l'anomenat *domini del tipus 1 de la fibronectina*, es troba també en l'activador del plasminogen tissular. Aquest últim també conté altres dominis que es troben en un precursor del factor de creixement epidèrmic i en la prourocinasa. Aquestes i altres observacions han suggerit la possibilitat de l'anomenada *redistribució d'exons (exon shuffling)* entre gens: exons o grups d'exons que codifiquen dominis complets són copiats i inserits en altres gens.

La redistribució d'exons està afavorida per la capacitat de recombinació entre introns, ja sigui perquè són idèntics o pel fet que contenen insercions de seqüències repetitives homòlogues. La gran majoria de proteïnes que han evolucionat per redistribució d'exons es troben en les línies dels metazous; en altres organismes com procarïotes, molts protistes, fongs i plantes, les proteïnes modulars són menys freqüents. La raó principal es deu a l'abundància de grans introns en els gens dels metazous que afavoreixen la recombinació i el muntatge (redistribució) d'exons. Les proves d'evolució de proteïnes per muntatge d'exons són molt abundants.

Un altre mecanisme que s'ha proposat per explicar la redistribució d'exons és el de la retrotransposició. La retrotransposició és el procés mitjançant el qual se sintetitza una còpia de DNA (cDNA) d'un transcrit de RNA i aquesta còpia s'insereix en una nova localització cromosòmica. Aquest procés ha estat molt important en l'evolució del genoma humà, més d'un 40 % del qual està format per seqüències repetides derivades per retrotransposició (retroelements). Un petit percentatge d'aquestes seqüències encara es transposen activament. La maquinària de la retrotransposició ha contribuït significativament a l'evolució dels gens, copiant exons i transposant-los d'una localització genòmica a una altra. La retrotransposició de seqüències gèniques ha permès remodelar els genomes complexos.

La duplicació de gens complets ha estat un mecanisme fonamental per a l'evolució dels organismes pluricel·lulars. La duplicació (vegeu la figura 3) pot donar lloc a dues còpies idèntiques del gen, incloent-hi tant les regions reguladores no codificadores, situades en el mateix cromosoma del gen (regions cis-reguladores), com les seqüències codificadores. Quan es duplica un gen, una de les còpies pot divergir ràpidament

a causa de l'absència de selecció negativa perquè la funció està garantida per l'altra còpia. Aquesta còpia pot acumular mutacions i degenerar en un pseudogèn. En alguns casos, però, la còpia pot produir un nou producte funcional avantatjós i conservar-se per selecció positiva. En altres casos les mutacions en les seqüències cis-reguladores de la còpia poden produir canvis ràpids en l'expressió del gen, la qual cosa fa que el producte gènec es pugui expressar en teixits diferents o en etapes diferents del desenvolupament.

Els gens homòlegs —gens semblants que tenen un mateix origen evolutiu procedents d'un avantpassat comú— poden ser ortòlegs o paràlegs (vegeu la figura 4). Els gens ortòlegs es troben en diferents espècies i procedeixen per descendència d'un gen en un avantpassat comú. Els gens paràlegs són el resultat d'una duplicació. Les duplicacions reiterades generen conjunts de gens amb funcions relacionades que s'anomenen *famílies gèniques*. Per exemple, en l'ésser humà les famílies de les α i β globines s'han originat per duplicacions seqüencials en tàndem. Les globines es divideixen en quatre classes funcionals: les hemoglobines i la mioglobina, que han estat molt estudiades,

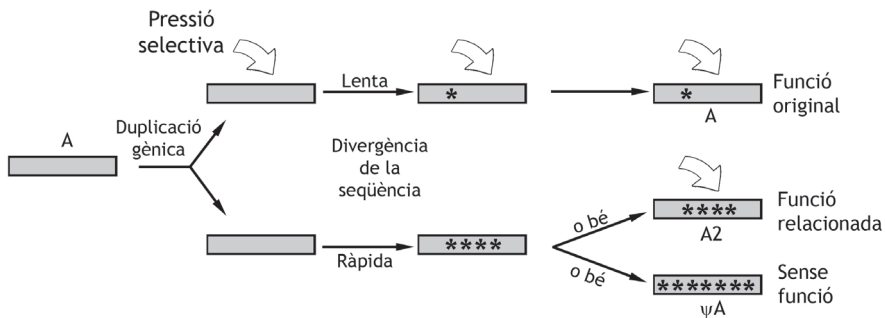


FIGURA 3. La duplicació gènica. La duplicació d'un gen (A) genera dues còpies gèniques, al principi equivalents, però que amb el temps divergeixen per acumulació de mutacions. És palès que la pressió selectiva (selecció negativa) ha d'actuar només sobre una de les còpies per mantenir la funció original del gen. L'altra còpia, en absència de selecció, acumularà mutacions (asteriscs) ràpidament, que si són deletèries produiran un gen no funcional (pseudogèn ψA). Tanmateix, en alguns casos, les mutacions poden generar una còpia (A2) amb funcions noves o amb diferents patrons d'expressió que sigui selectivament afavorida (modificat de Strachan i Read, 2004).

i la neuroglobina i la citoglobina, descrites més recentment. El grau d'homologia de seqüència entre les diferents globines, i el fet que s'hagin conservat en general les posicions intròniques, indica que les quatre famílies s'han originat a partir d'un gen ancestral mitjançant una sèrie de duplicacions que es varen produir en un període comprès entre fa vuit-cents i quatre-cents cinquanta milions d'anys. Més tard, els gens ancestrals de l' α -globina i la β -globina experimentaren duplicacions noves, algunes de les quals són força recents. Per exemple, els dos gens humans de l' α -globina (*HBA1* i *HBA2*) codifiquen productes idèntics, i els productes codificats pels gens de la γ -glo-

bina (*HBG1* i *HBG2*) difereixen en un únic aminoàcid. Algunes duplicacions han donat lloc també a pseudogens.

Quan es produeixen duplicacions en tàndem, es dupliquen tant les seqüències codificadores com les cis-reguladores. La divergència posterior de les seqüències cis-reguladores pot produir una expressió gènica diferent, tant en l'espai (per exemple, en diferents teixits) com en el temps (en diferents etapes del desenvolupament). La superfamília de les globines n'és un exemple. Les duplicacions del gen ancestral de les globines varen generar còpies que varen divergir respecte a les seqüències cis-reguladores i, això va determinar que els productes gènics s'expressessin en diferents teixits (per exemple, en la sang, en el teixit muscular o en el nerviós). Cada còpia es va adaptar al seu ambient nou, la qual cosa va determinar una divergència considerable de les seqüències i va donar lloc al gen ancestral de l' α -globina i la β -globina –cadascuna mantenint, però, la capacitat d'unir-se a l'oxigen. Posteriorment, varen aparèixer diferents varietats de globines, que s'expressaven en etapes diferents del desenvolupament. Per exemple, les cadenes ϵ , ζ i γ estan especialment adaptades a unir-se a l'oxigen en els ambients menys oxigenats de les primeres etapes del desenvolupament (embrió i fetus), mentre que les cadenes α i β són més aptes en els ambients dels teixits adults.

Tal com va proposar Ohno (1970), les duplicacions genòmiques produïdes per la poliploidia són un sistema eficaç per incrementar la grandària i la versatilitat del genoma. Permeten disposar simultàniament de còpies de tots els gens, i eviten els problemes de dosis gèniques diferencials. Moltes espècies són poliploides o han esdevingut poliploides degenerats, com la majoria de plantes fanerògames, molts peixos, i llevats, entre d'altres. Ohno també va proposar que

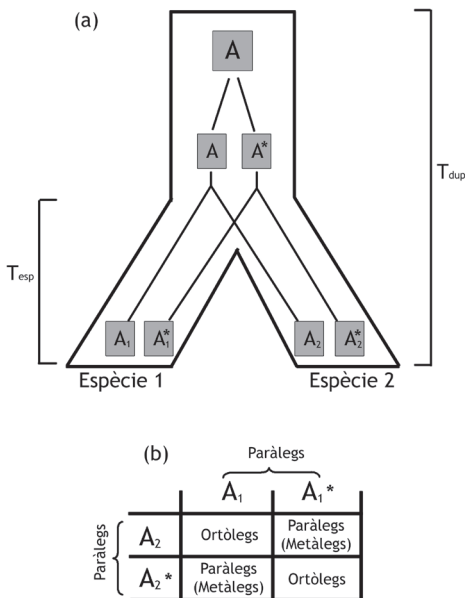


FIGURA 4. a) Esquema de duplicació d'un gen *A* en una espècie ancestral que posteriorment experimenta un procés d'especiació. En b) es defineixen les relacions d'homologia. Els gens evolucionats per duplicació s'anomenen *paràlegs*, els quals poden estar en la mateixa espècie o en espècies diferents (paràlegs metàlegs). Els *ortòlegs*, producte d'una especiació, es troben sempre en espècies diferents. T_{dup} : temps de duplicació; T_{esp} : temps d'especiació (modificat de Fontdevila i Moya, 2003).

en el decurs de l'evolució dels vertebrats es varen produir dues (o tres) duplicacions del genoma. En particular, el genoma humà seria, doncs, un paleopoliploide: un antic poliploide que més tard va esdevenir diploide mitjançant la divergència de les seqüències dels cromosomes duplicats. Aquesta hipòtesi d'Ohno ha tingut el suport de dues línies de recerca. En primer lloc, els estudis de genòmica comparada: la comparació dels mapes genètics dels mamífers ha confirmat que aquests contenen segments grans amb l'ordre genètic molt conservat (sintènics), els quals serien romanents d'una poliploïdia ancestral. En segon lloc, el descobriment que els quatre complexos *Hox* dels mamífers, que tenen una funció fonamental en el desenvolupament, haurien evolucionat per quadruplicació d'un complex prototípic semblant al de *Drosophila*.

La forma més acceptada de la hipòtesi —anomenada *hipòtesi 2R*— és que es varen produir dues duplicacions genòmiques en la línia evolutiva dels vertebrats: una immediatament abans i una altra immediatament després de la separació de la línia de les llampreses de mar (Wolfe, 2001). L'anomenada *regla del 4:1*, segons la qual els gens dels invertebrats, com *Drosophila*, haurien de tenir quatre paràlegs complets en els vertebrats, s'interpreta sovint com un corollari de la hipòtesi 2R, encara que això depèn del grau en què els gens s'hagin eliminat (per acumulació de canvis mutacionals) després de cada duplicació genòmica. La comparació de seqüències dels genomes de *Drosophila* i l'ésser humà ha demostrat que menys del 5 % de les famílies genètiques homòlogues segueixen la regla del 4:1 (Dehal i Boore, 2005). Això és degut principalment a la pèrdua de la funció de molts gens duplicats que esdevenen pseudogens, la qual cosa ha anat esborrant l'empremta de les duplicacions genòmiques.

SIGNIFICAT EVOLUTIU DELS ELEMENTS TRANSPOSABLES: DNA PORQUERIA?

Només l'1 per cent del genoma humà és codificant; la pregunta obvia és: per a què serveix el 99 % restant? El mateix es pot preguntar del genoma de la majoria de mamífers. La fracció no codificant «invisible» del genoma seria equivalent a la matèria fosca del cosmos de la qual ens parlen els físics. Aquest DNA no codificant podria contenir seqüències de regulació fonamentals per explicar la complexitat dels organismes. En humans, s'accepta que només del 2 al 3 % d'aquesta «matèria fosca» conté seqüències que controlen l'ús dels gens. La resta està formada en gran part per seqüències repetitives, entre les quals destaca, per la seva abundància, una classe de DNA dispers que és mòbil, formada pels elements transposables (ET). Aquest DNA no codificant ha estat qualificat genèricament de *porqueria* (*junk*) com a resultat de la seva acumulació en el procés evolutiu. Mentre que és possible que part del DNA porqueria sigui una realitat, la fracció d'ET presenta unes característiques especials que mereixen tenir-se en compte abans de descartar-ne el valor evolutiu.

Hi ha dues classes fonamentals d'ET: els de la classe I, anomenats *retroelements*, transcriuen una còpia de RNA que posteriorment retrotranscriuen a una seqüència de DNA la qual s'insereix en un lloc nou (transposició), d'una manera semblant com ho fan els retrovirus, com el VIH; i els de la classe II, de transposició directa mitjançant excisió-inserció de seqüències de DNA. A causa de la seva capacitat d'envair el genoma, els ET s'han considerat com a paràsits egoistes dels quals el genoma s'ha de defensar mitjançant mecanismes, alguns dels quals, com la metilació, són evolutivament molt antics. Malgrat tot, la seva abundància, que en humans pot arri-

bar a més del 50 % del genoma, mereix ser explicada.

Els ET produeixen mutacions. La gran majoria (més del 80 %) de mutacions en *Drosophila* i una quantitat important (un 15 %) en mamífers es deuen a la inserció d'ET. Una inserció en una zona codificant del gen altera l'expressió del gen i sovint l'inactiva, per la qual cosa és eliminat per selecció negativa i no es pot mantenir en les poblacions. Malgrat tot, tenim evidències que almenys un 4 % del gens humans, és a dir, uns mil dos-cents gens, contenen ET. Però on possiblement la inserció d'un ET té conseqüències evolutivament més importants és en les zones reguladores del gen, perquè les seqüències reguladores mateixes dels ET poden canviar l'expressió del gen. Un exemple d'aquest tipus de mutació reguladora es dona en els gens de la família

de l'amilasa, que en mamífers s'expressen generalment al pàncrees. Per contra, en humans, el gen *Amy1*, d'aquesta família, conté un retrotransposó que és responsable de la seva expressió a les glàndules salivals. Molts dels ET inserits en regions importants, com la regió HLA-DR del complex major d'histocompatibilitat en humans, podrien tenir també funcions reguladores (vegeu la figura 5). Moltes vegades la seqüència responsable del canvi regulador és només una part de l'ET. El gen *oncomodulin* de la rata pertany a la família de les albúmines i s'expressa a la placenta. La seqüència responsable d'aquesta especificitat de teixit està continguda en un LTR (una repetició terminal del retroelement) inserit en posició 5' del gen. Hi ha molts altres exemples, en els quals només una petita part d'un ET és responsable dels canvis reguladors; és

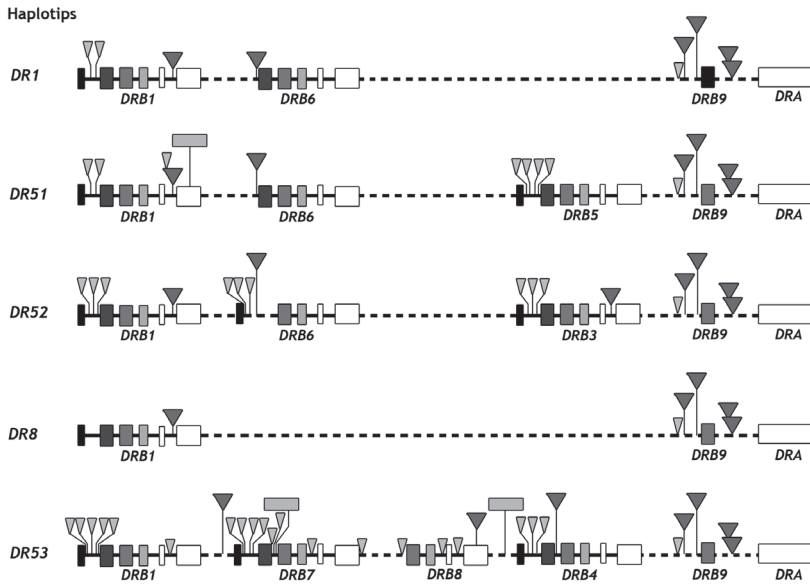


FIGURA 5. Descripció gràfica de cinc haplotips DR de la regió de gens antigens leucocitaris humans (HLA) que codifiquen les molècules que «presenten» els antigens en la resposta immunitària. Els exons es mostren com quadrilàters en diferents tons del gris. Diferents retroelements es troben inserits dintre i entre els gens amb una alta densitat (un per cada kb) i generen empalmaments alternatius i regulació transcripcional, la qual cosa augmenta la plasticitat evolutiva de la resposta immunitària. Alu (triangles isòsceles); LINE (triangles equilàters foscos); retrovirus endògens, HERV (rectangles ombrejats) (Anderson *et al.*, 1998).

precisament aquesta part la que es manté per selecció natural.

La dinàmica dels ET té un abast evolutiu molt important en la reorganització del genoma. La transcripció i l'excisió de l'element no són perfectes i de vegades arrosseguen segments genòmics flanquejants que són transposats a altres llocs del genoma (transducció). Aquests segments poden contenir seqüències codificants o reguladores capaces de produir o induir funcions gèniques noves en els llocs d'inserció. Un altre mecanisme important de reorganització genòmica consisteix en l'aparellament entre seqüències homòlogues de dues còpies d'ET inserides en diferents llocs del genoma (aparellament ectòpic). Aquest aparellament pot afectar grans segments cromosòmics o seqüències curtes de DNA. En el primer cas el resultat és la producció de reordenacions cromosòmiques de gran importància evolutiva, com inversions, translocacions, duplicacions i delecions. En *Drosophila* hi ha proves que els punts de trencament d'inversions de poblacions naturals contenen ET i que els ET són capaços d'incrementar la producció d'inversions en experiments de laboratori. En el segon cas les reordenacions es donen en gens o fins i tot entre exons, la qual cosa dona una gran capacitat evolutiva per produir funcions gèniques noves. Ja hem explicat que molts gens estan construïts empalmant peces d'altres gens, moltes de les quals estan repetides, i el paper dels ET en aquest procés de recombinació és important en alguns casos.

La capacitat dels ET de replicar-se i envair el genoma va ser per a alguns evolucionistes la prova que el DNA mòbil era un exemple clar de DNA egoista, semblant al dels virus o altres paràsits, sense cap valor evolutiu per al genoma. Aquesta idea, encara bastant estesa entre la comunitat científica, està canviant a mesura que les dades noves ens indiquen el seu paper en l'evolució

del genoma. La hipòtesi egoista visualitza la interacció entre ET i genoma com una carrera d'armaments evolutiva en la qual el genoma evoluciona amb mecanismes de defensa (selecció negativa) i els ET responen amb mecanismes d'invasió (taxa de transposició), de manera que la distribució d'ET que observem és el resultat de l'equilibri entre ambdues forces. És molt possible que originalment els ET tinguessin un comportament totalment paràsit, però això no impedeix que posteriorment els ET coevolucionessin amb el genoma, el qual s'hauria apropiat d'algunes de les seves funcions en un procés de domesticació, anàleg a una endosimbiosi. Les observacions que hem descrit més amunt sobre el paper regulador i reorganitzador dels ET en el genoma suggereixen aquest enfocament. Encara que McClintock, la Premi Nobel descobridora dels ET, ja va indicar fa més de quaranta anys el seu paper regulador en el genoma, hem hagut d'esperar a l'era genòmica per descobrir les petjades, sovint molt febles i alterades, que els ET han anat deixant arreu del genoma en la seva llarga caminada evolutiva de domesticació. Aquesta cascada d'informació que proporcionen les tècniques genètiques i bioinformàtiques actuals està sense analitzar en gran part i el futur ens donarà, sens dubte, més dades sorprenents. Però, pel que fa a la nostra visió del genoma, aquesta és ben diferent de la que teníem fa pocs anys. El genoma és mòbil, si entenem per *mòbil* que la seva dinàmica està condicionada per un munt de transposicions que són domesticades i que han estat produint-se des del principi de la vida. Aquesta dinàmica era molt difícil de preveure en l'època de Darwin, però sense cap dubte el mateix Darwin n'hauria quedat molt fascinat.

BIBLIOGRAFIA

- ANDERSON, G.; SVENSSON, A.; SETTERBLAD, N.; RASK, L. (1998). «Retroelements in the human MHC class II region». *Trends in Genetics*, 14: 109-114.
- BALANYÀ, J.; OLLER, J. M.; HUEY, R. B.; GILCHRIST, G. W.; SERRA, L. (2006). «Global genetic change tracks global climate warming in *Drosophila subobscura*». *Science*, 313: 1773-1775.
- DEHAL, P.; BOORE, J. L. (2005). «Two rounds of whole genome duplication in the ancestral vertebrate». *PLOS Biology*, 3(10): e314.
- FONTDEVILA, A.; MOYA, A. (2003). *Evolución: origen, adaptación y divergencia de las especies*. Madrid: Síntesis.
- GRANT, P. R.; GRANT, B. R. (1995). «Predicting micro-evolutionary responses to directional selection on heritable variation». *Evolution*, 49: 241-251.
- HALDANE, J. B. S. (1949). «Suggestions as to the quantitative measurement of rates of evolution». *Evolution*, 3: 51-56.
- HALL, B. G. (2001). *Phylogenetic trees made easy*. Sinauer Associates.
- HUEY, R. B.; GILCHRIST, G. W.; CARLSON, M. L.; BARRIGAN, D.; SERRA, L. (2000). «Rapid evolution of a geographic cline in size in an introduced fly». *Science*, 287: 308-309.
- KIMURA, M. (1983). *The neutral theory of molecular evolution*. Cambridge: Cambridge University Press.
- OHNO, S. (1970). *Evolution by gene duplication*. Londres: George Allen and Unwin.
- SIMPSON, G. G. (1944). *Tempo and mode in evolution*. Nova York: Columbia University Press.
- STRACHAN, T.; READ, A. P. (2004). *Molecular Human Genetics*. Londres: Garland Science.
- WOLFE, K. H. (2001). «Yesterday's polyploids and the mystery of diploidization». *Nat. Rev. Genet.*, 2: 333-341.
- ZUCKERKANDL, E.; PAULING, L. (1965). «Evolutionary divergence and convergence in proteins». A: BRYSON, V.; VOGEL, H. J. [ed.]. *Evolving genes and proteins*. Nova York: Academic Press, 97-166.