

**Elements transposables de tipus non-LTR
als ascidis, amfioxos i àgnats**

Jon Permanyer i Ugartemendia

Programa de Doctorat del Departament de Genètica
Facultat de Biologia
Bienni 2002-2004

Elements transposables de tipus non-LTR als ascidis, amfioxos i àgnats

Memòria presentada per
Jon Permanyer i Ugartemendia

Per optar al grau de
Doctor

Tesi doctoral sota la direcció del **Dr. Ricard Albalat Rodríguez** i
la **Dra. Roser González Duarte**
al Departament de Genètica de la Facultat de Biologia de la Universitat de
Barcelona

Els directors

L'autor

Dr. Ricard Albalat

Dra. Roser González

Jon Permanyer

Barcelona, Novembre 2006

Objectius

L'objectiu general d'aquesta Tesi Doctoral ha estat la caracterització genòmica de retrotransposons non-LTR a cordats basals per tal d'evaluar la importància d'aquests elements en l'evolució genòmica del llinatge dels vertebrats. Per a fer-ho ens vàrem proposar els següents objectius:

1. Crivellar *in silico* el genoma de *Ciona intestinalis v1.0* per tal d'identificar els tipus de retrotransposons non-LTR que presenta i determinar mitjançant aproximacions experimentals i *in silico*, el nombre de còpies, l'entorn genòmic i l'estat de metilació dels diferents tipus descrits.
2. Utilitzar la base de dades genòmica de *Branchiostoma floridae* per tal d'engalzar seqüències per a formar *contigs* on identificar els diferents tipus de retrotransposons non-LTR que presenta el seu genoma d'aquesta espècia a partir de diverses aproximacions. Així mateix, poder determinar el nombre de còpies d'aquests elements.
3. Caracteritzar retrotransposons non-LTR a l'àgnat *Myxine glutinosa* mitjançant metodologies de cerca específica a partir del disseny d'encebadors degenerats i el crivellatge d'una genoteca genòmica.

Resum Global

La presència de retrotransposons non-LTR en un genoma és una característica inherent a quasi tots els organismes. El tipus i nombre d'elements és característic de cada espècie i comprèn un ventall molt ampli que oscil·la des de l'absència d'aquests elements a *S. cerevisiae* fins a situacions com l'observada a mamífers com *H. Sapiens*, on bàsicament existeixen dos únics tipus tan freqüents que han esdevingut una de les fraccions més importants del genoma. Dins d'aquesta variabilitat també s'han descrit situacions intermèdies, com el cas de *D. melanogaster*, amb molts tipus diferents d'elements però tots ells poc abundants. Tot i ser poc abundants en alguns casos, la seva presència juga un paper important en l'evolució ja que els TEs interactuen amb l'entorn genòmic on es troben i incrementen el potencial evolutiu de l'hoste, encara que també es sabut que poden tenir efectes deleteris. Els estudis comparatius entre aquests elements en els genomes preduplicatius d'ascidis i amfioxos així com en el llinatge dels àgnats ens han permès valorar el seu impacte en l'evolució genòmica durant la transició dels cordats invertebrats als vertebrats.

1. CiLINEs, BfLINEs i MgLINE

L'anàlisi *in silico* dels projectes genoma de *Ciona intestinalis* i *Branchiostoma floridae* i l'ús de metodologies *in vitro* de cerca específica d'elements en el genoma de *Myxine glutinosa* ens han permès caracteritzar retrotransposons non-LTR en aquests organismes. Cal tenir en compte que els mètodes de cerca actuals presenten una limitació molt important ja que tots es basen en l'homologia a elements prèviament descrits. Això implica que l'existència d'elements divergents i poc abundants pot passar desapercebuda. Un cas extrem el trobem quan aquests presenten estructures no descrites fins al moment. Tal i com va succeir en els elements de DNA anomenats *helitrons* que no varen ser descoberts fins fa pocs anys, ja que tot i ser elements autònoms, de mida gran (~15 Kb) presents com a mínim en animals i plantes, no contenien la informació esperada segons els elements descrits fins al moment en altres TEs i, per tant, no eren identificats com a tals en les cerques de DNA repetitiu (Kapitonov 2001). Aquesta limitació també pot afectar a les aproximacions experimentals de cerca utilitzades en aquesta Tesi, ja que el disseny d'oligonucleòtids degenerats es fonamenta en la comparació de dominis conservats; per tant, els elements divergents no s'haurien detectat a partir de la nostra cerca. A més, s'ha descrit molta diversitat entre elements i la presència de dominis clarament conservats en els TEs és escassa i per tant la cerca de TEs mitjançant PCR queda limitada a elements amb dominis conservats, com la majoria d'elements autònoms. Aquesta conservació només es produeix entre els dominis que tenen un constrenyiment evolutiu que, en el cas dels retrotransposons non-LTR, sols correspon als dominis transcriptasa inversa i al de l'endonucleasa. En canvi, les regions entre aquests dominis i les dels extrems dels elements són molt divergents i, per tant, només es poden

definir mitjançant la comparació de diferents còpies d'un tipus concret en el seu hoste.

En aquest treball hem caracteritzat 5 tipus de retrotransposons non-LTR a l'ascidi anomenats CiLINEs (CiI, CiL1, CiL2, CiLOA i CiR2), 6 a l'amfiox o BfLINEs (BfCR1, BfI, BfL1, BfL2, BfNeSL i BfRTE) i un a la mixina (MgLINE) (Figura 7). En tots ells hem descrit la pauta de lectura oberta que codifica per a la transcriptasa inversa, característica dels retroelements i imprescindible per a poder determinar a quin clade pertanyen. La caracterització dels dominis 0, 1, 2, 2a, 3, 4, 5, 6, 7, 8 i 9 d'aquesta pauta ens ha permès establir en la majoria de casos la relació filogenètica d'aquests elements amb els elements d'altres organismes (Figura 8). Els nodes descrits amb un menor suport filogenètic (*bootstrap* < 70%) corresponen a elements de clades poc caracteritzats o que són font de controvèrsia ja que segons alguns autors (Malik 1999) inclouen elements que pertanyerien a més d'un clade. Aquest és el cas del clade I on els elements s'agrupen amb un suport filogenètic baix. Tanmateix, els elements CiI i BfI presenten homologia alta amb diferents membres del clade i l'ordenació i tipus dels dominis addicionals que presenten permet definir-los com a elements I. A més a més, per al clade I s'ha proposat una subdivisió en els clades I i Ingi (Craig 2001). Si analitzem l'adscripció dels elements de l'ascidi i l'amfiox en aquest nou context observem com el valor *bootstrap* del nou grup I supera el 70%, confirmant el resultat basat en la homologia blast i l'estructura dels seus dominis és similar a altres elements I. De forma similar, l'element BfNeSL tot i que el suport no es prou robust (*bootstrap* del 64%), s'ha inclòs dins del clade NeSL ja que les comparacions BLAST li otorguen similitud amb altres elements NeSL descrits. Probablement en el futur aquest grup d'elements es subdividirà en més d'un clade (Malik 2000) com ha succeït en d'altres grups. La situació contrària també es produeix, els elements L2 i CR1, que estan molt relacionats i dels quals certs autors en discuteixen l'existència com a clades separats, ja que, tot i estar diferenciats filogenèticament, l'estructura que presenten és molt similar (Malik 2000). La classificació dels retrotransposons non-LTR en clades és una convenció laxa que pretenia estructurar racionalment els elements descrits. No obstant això, el creixent nombre de clades ha fet que es qüestionari aquest sistema de classificació ja que cada cop és més complex i no representa ni l'origen ni la relació entre els diferents tipus existents. Així, per evitar la complexitat creixent en la classificació d'aquests elements, s'ha proposat la divisió en 5 grups que englobarien els clades prèviament definits en funció de la relació filogenètica i el tipus i ordre dels dominis enzimàtics que presenten (Craig 2001). Si tenim en compte aquesta nova classificació, observem que com a l'ascidi hi ha elements que corresponen a 4 grups, 5 a l'amfiox i 1 a la mixina.

La baixa processivitat atribuïda al domini amb activitat transcriptasa inversa explica perquè la majoria de retrotransposons non-LTR es troben presents com a còpies truncades per 5' tal i com s'observa en el genoma humà

amb un centenar de còpies funcionals de l'element L1 entre el milió de còpies existents (Kazazian 2004). Les estratègies emprades per a identificar els CiLINEs, BfLINEs i MgLINE no ens ha permès definir elements sencers o íntegres (*full length*), per a tots ells manca determinar les duplicacions del lloc d'inserció o TSDs que delimiten la majoria d'aquests elements. A més a més, pels *young* LINEs, s'han descrit tots els dominis esperats de l'ORF2 tot i que manca la caracterització completa de l'APE de l'element Bfl. Ara bé, no s'ha pogut caracteritzar l'ORF1 ni els dominis predits d'unió a àcids nucleics. L'ORF1 és difícil d'identificar ja que té un grau de conservació baix entre els elements de diferents organismes i tan sols es pot definir en comparar còpies íntegres d'un mateix tipus d'element dins d'una mateixa espècie. En el nostre estudi, hem identificat també elements corresponents al grup dels *old* LINEs, els menys freqüents en els genomes seqüenciats. CiR2 representa l'element més ben caracteritzat en aquest treball ja que se n'ha determinat tots els dominis així com la seva localització genòmica en els gens de l'RNA 28S. Tot i això, no podem determinar si es tracta d'un element íntegre o no ja que no s'ha pogut establir les TSD. En contraposició a aquest element, trobem BfNeSL on només hem pogut determinar el domini RT.

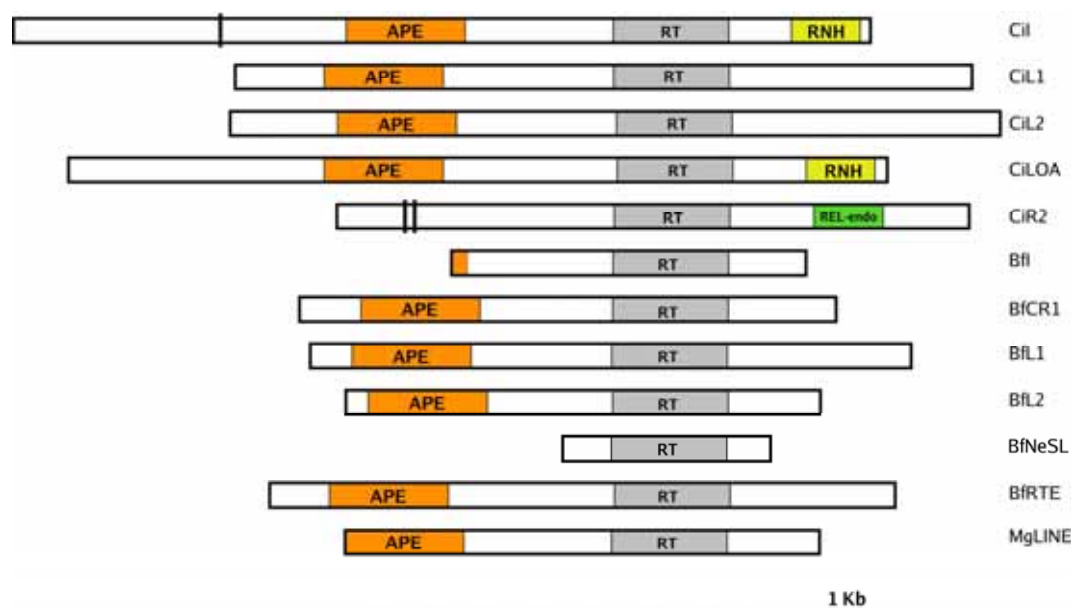


Figura 7: CiLINEs, BfLINEs i MgLINE caracteritzats a *C. intestinalis*, *B. floridae* i *M. glutinosa* respectivament. Les caixes indiquen els dominis codificants caracteritzats per a les activitats enzimàtiques assenyalades en les caixes. APE: endonucleasa apurínica/apirimidínica, REL-endo: endonucleasa similar a enzims de restricció, RNH: RNAsa H i RT: transcriptasa inversa. Les barres verticals indiquen dominis d'unió a àcids nucleics.

Malgrat no hem identificat elements sencers en les espècies estudiades, és molt possible que els genomes d'aquests organismes allotgin còpies íntegres. Cal recordar que els CiLINEs i BfLINEs s'han caracteritzat a partir de les dades de seqüenciació de la fracció eucromàtica, on el DNA repetitiu està poc

representat. L'engalament del DNA repetitiu, especialment abundant en les regions heterocromàtiques, és computacionalment molt complex i, freqüentment no degudament caracteritzat en els projectes de seqüenciació dels genomes. Per tant, roman sense establir l'existència d'elements, tant íntegres com truncats en la fracció heterocromàtica dels genomes, on degut la naturalesa silenciosa d'aquesta fracció genòmica hagi probablement afavorit la inserció d'elements mòbils. Finalment, malgrat els esforços realitzats, l'element MgLINE caracteritzat no és íntegre. En aquest sentit, la caracterització de noves còpies permetrà la futura descripció completa d'aquest element en aquesta espècie.

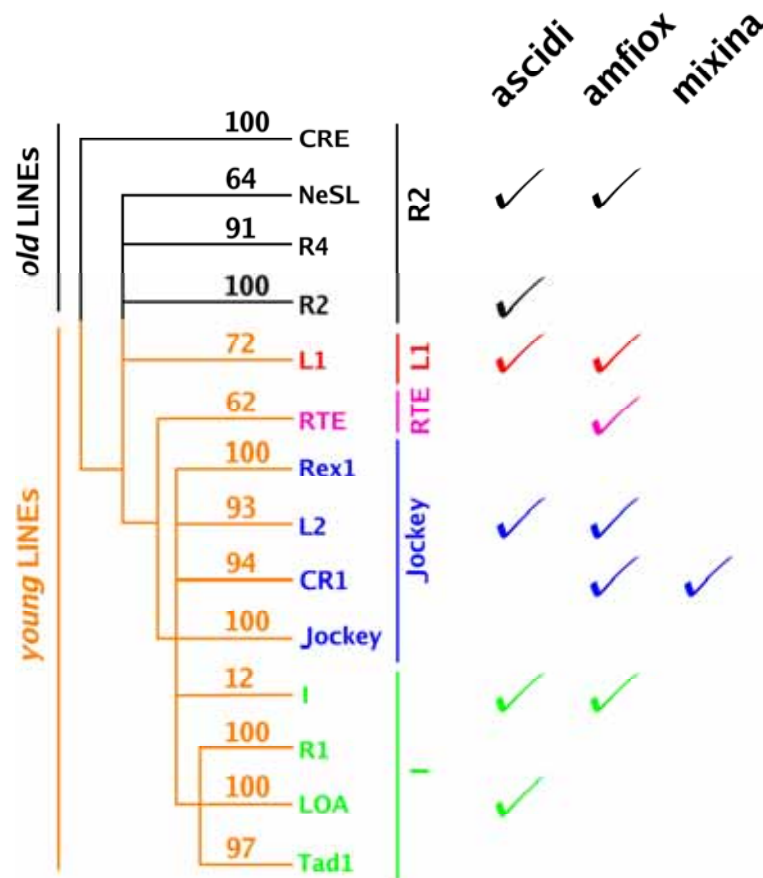


Figura 8: Arbre filogenètic on s'analitzen els elements CiLINEs, BfLINEs i MgLINE amb les seqüències dels dominis RT de 90 elements corresponents als 14 clades descrits. S'indica el valor de *bootstrap* assolit per a cada clade. A l'esquerra de l'arbre s'indica la gran divisió en *old/young* LINEs i a la dreta la classificació en 5 grans grups. També s'indiquen els tipus d'elements descrits a l'ascidi, l'amfiox i la mixina.

Un treball posterior a les cerques realitzades en el genoma de *Ciona intestinalis* basat en la mateixes metodologies (Kojima 2004) ha permès identificar elements de tipus L1 (elements L1Ci-A, L1Ci-B i L1Ci-C), NeSL (YURE-Ci) i R2 (R2Ci-A, R2Ci-B i R2Ci-C) en aquesta espècie. Els elements descrits corresponen als proposats per a l'ascidi en aquesta Tesi a excepció de

l'element YURE-Ci que correspon al clade NeSL. Aquests resultats reforcen l'elecció de l'estratègia utilitzada que ha permès determinar la presència de la majoria de tipus existents i confirmen la mateixa limitació a l'hora de determinar elements sencers. L'element YURE-Ci correspon a un grup d'old LINEs amb un grau de caracterització baix i amb elements molt divergents; això explica les diferències en el resultat de la cerca, probablement degudes a la utilització de diferents esquers i bases de dades per realitzar les cerques.

Els elements de tipus R2 presenten com a característica diferencial addicional respecte d'altres elements que utilitzen l'rRNA 28S com a diana preferent d'inserció. Emprant oligonucleòtids degenerates, s'han amplificat membres d'aquest clade en artròpodes, equinoderms, àgnats i gnatostomats (Kojima and Fujiwara 2005). Les dades d'aquesta Tesi i les dels autors esmentats posen de manifest que els retrotransposons non-LTR de tipus R2, històricament definits com a elements exclusius d'artròpodes (Malik 1999), no presenten cap limitació d'hoste. Aquesta característica definida per als elements R2 sembla que es pot estendre a tots els clades d'elements non-LTR ja que *a priori* per a cap tipus d'element no trobem limitacions funcionals o estructurals que justifiquin la presència d'un determinat clade en un determinat organisme.

De fet, el nombre i tipus d'elements presents en un organisme a dia d'avui és un reflex de l'evolució del seu llinatge i el reflex de l'èxit aconseguit en la colonització del genoma hoste en detriment d'uns altres. En canvi, el ventall d'hostes que presenta cada clade és proporcional al moment en què es va originar l'element i a fenòmens de transferència horitzontal. Els estudis filogenètics dels retrotransposons non-LTR suggereixen que la majoria d'ells s'originaren fa uns 600 milions d'anys, abans de l'era Càmbrica (Malik 1999), malgrat que poguem trobar elements més antics com els dels clades RTE que es van originar abans de la diversificació dels eucariotes i que avui els trobem a animals, plantes i algues marrons (Zupunski 2001); o més joves com els del clade L2 que tot i aparèixer abans de la radiació dels vertebrats, només presenta elements actius als llinatges de peixos (Lovsin 2001). A més de la transferència vertical, la història dels TEs també depèn dels fenòmens de transmissió horitzontal (Mizrokhi 1990). Aquests fenòmens són extremadament rars fins al punt que alguns autors en discuteixen l'existència (Malik 1999). La demostració d'aquests és clarament complexa però existeixen evidències indirectes que permeten intuir-los. L'evidència més clara, tot i que no definitiva, la trobem en realitzar anàlisis filogenètiques d'elements d'un mateix clade descrits en diferents organismes i observar com els elements que s'han transferit horitzontalment no respecten la filogènia de les espècies acceptada per als hostes que els contenen. Aquest fenomen el vàrem observar en analitzar la relació filogenètica de l'element BfCR1 amb altres elements del clade i advertir com l'element SR1 d'*Schistosoma mansoni* s'enbrancava amb els elements CR1 de vertebrats i no pas amb els de protòstoms. Donat que S.

mansoni és un paràsit de vertebrats i per tant hi estableix un contacte íntim, seria plausible que l'element SR1 provingués d'un element CR1 de vertebrats per transferència horitzontal.

2. Càrrega i distribució genòmica

Els TEs participen de manera activa en les característiques genòmiques d'un organisme. La magnitud de la fracció de TEs en un genoma determina en gran mesura la seva mida i, a més, no només és important el nombre de còpies present sinó que també cal considerar-ne la distribució, ja que els seus efectes varien en funció de l'entorn on es troben i de la seva capacitat de mobilitzar-se. Els retrotransposons non-LTR que hem descrit tant a l'ascidi com a l'amfiox es troben presents amb un nombre de còpies molt baix en comparació amb els seus homòlegs de vertebrats. Les estimes realitzades determinen <150 còpies per genoma haploid en ambdós genomes. Aquestes còpies predites per als seus genomes representen tant còpies senceres i per tant teòricament funcionals com truncades i per tant no funcionals. De fet, podria ser que el nombre de còpies fos major, ja que les estimes presenten biaixos metodològics. Ara bé, la congruència entre les 2 aproximacions realitzades i les dades d'altres autors (Simmen, Leitgeb et al. 1999) indiquen que els biaixos són mínims.

Si assumim que totes les còpies dels retrotransposons non-LTR són íntegres i que tenen una mida mitja de 5 Kb, observem com la càrrega genòmica que representa la fracció d'aquests elements no supera l'1% del genoma haploid (0,42% per a *Ciona intestinalis* i 0,13% per a *Branchiostoma floridae*) (Taula 2). Aquestes estimes correspondrien al llindar màxim a nivell genòmic ja que, tot i que les dades dels projectes genoma no inclouen la fracció heterocromàtica, suposar que totes les còpies són íntegres és una sobreestimació. Aquesta queda palesa quan s'analitzen les dades d'altres projectes genoma on s'ha determinat el tipus i nombre de TEs; així, al genoma humà trobem que si considerem tant els nucleòtids corresponents a retrotransposons non-LTR com el nombre de còpies d'aquests observem que la mida mitjana és de ~500 nucleòtids, deu vegades menys que la de l'element íntegre (Lander 2001). En el genoma de la mixina, on hem descrit un element de tipus CR1 que es troba present en unes 23000 còpies per genoma haploid, l'estima de la càrrega indica que representen com a màxim el 6% del genoma. Això suggereix que els genomes altament duplicats, com el dels àgnats, tenen una fracció genòmica corresponent a TEs significativa que condiciona la mida total del genoma. L'elevat nombre de còpies observat per a l'element MgLINE suggereix a més a més que molt probablement existeixen còpies senceres. Tot i no poder establir quina mida total té la fracció corresponent als retrotransposons non-LTR, sembla que el comportament d'aquesta seria típica de vertebrats. L'elevada presència d'aquest element en el genoma de la mixina el converteix en una bona eina per a estudiar un genoma, *a priori*, altament complex.

	<i>C. elegans</i>	<i>D. melanogaster</i>	<i>C. intestinalis</i>	<i>B. floridae</i>	<i>M. glutinosa</i>	<i>T. rubripes</i>	<i>R. norvegicus</i>	<i>H. sapiens</i>
CRE						2000		
I		67	9	3				
Jockey		392						
L1			22	32		500	597000	904000
L2			24	35		6500	48000	408000
L3/CR1	1000			25	23000		11000	55000
LOA		18	69					
NeSL	110		6	6		30		
R1		130						
R2		3-60	13					
R4						1000		
Rex1						2000		
RTE	15			42		2300		
Tad1								
Clades	3	5	6	6	> 1	7	3	3
Grups	3	3	4	5	1	5	2	2
Còpies	1125	667	<150	<150	> 23000	14330	657000	0
Càrrega	<5	<3	<1	<1	> 6	1.3	23.1	21.05
Mida	$9,7 \cdot 10^7$	$1,6 \cdot 10^8$	$1,8 \cdot 10^8$	$5,8 \cdot 10^8$	$3,9 \cdot 10^9$	$4 \cdot 10^8$	$2,9 \cdot 10^9$	$3,2 \cdot 10^9$

Taula 2: Retrotransposons non-LTR a protòstoms i deuteròstoms. S'indica el nombre de clades presents (Clades), el nombre de grups presents (Grups) el nombre de còpies per genoma haploid per a cada clade (Còpies), la càrrega que representa per al genoma en % (Càrrega) i la mida del genoma (Mida). Les dades utilitzades es troben a: Waterston, 1995, Vieira 1999, Permanyer 2003, Permanyer 2006, Aparicio 2002, Gibbs 2004 i Lander 2001.

Un component cabdal dels retrotransposons non-LTR a l'hora de determinar l'abast de la colonització és el tipus d'endonucleasa (Zingler 2005). Els elements que presenten una endonucleasa de tipus REL-endo s'inserten de manera quasi específica en regions de gens presents en múltiples còpies com els rDNAs. Per altra banda, els elements que presenten una endonucleasa APE no tenen una diana concreta d'inserció, fet que ha permès un major èxit colonitzador per part dels *young* LINEs, que codifiquen per una endonucleasa d'aquest tipus. Això queda reflectit en els genomes de l'ascidi i de l'amfiox on els elements més abundants són els *young* LINEs. En canvi, els elements descrits amb REL-endo, o *old* LINEs, són els menys abundants. Per als elements del clade I s'ha proposat a *D. melanogaster* un mecanisme d'autoregulació anomenat cossupressió que controla el nombre de còpies mitjançant la

repressió transcripcional dels elements en funció del nombre de còpies existents (Jensen 1999). Tot i que només s'ha comprovat a *Drosophila* i sembla ser exclusiu d'aquest clade, aquest mecanisme podria justificar els baixos valors que hem determinat en els elements Cil i Bfl en comparació als altres *young* LINEs i inferiors als dels *old* LINEs.

Tot i que per a molts organismes s'ha descrit que els TEs tendeixen a acumular-se en regions heterocromàtiques, tal i com s'ha determinat per al genoma d'*A. gambiae* (Holt 2002), en el cas dels cordats invertebrats no és possible que el seu genoma suporti una càrrega tan elevada com la que s'observa als vertebrats gràcies a la fracció heterocromàtica. Les estimacions experimentals, que analitzen tot el genoma, realitzades a l'ascidi i l'amfiox, no difereixen de les realitzades *in silico* indicant així que la fracció heterocromàtica no es troba enriquida amb aquests elements i per tant, almenys en l'ascidi, el nombre de còpies senceres serà extremadament baix. L'existència d'un projecte genoma amb les seqüències engalzades en *contigs*, com el genoma de *C. intestinalis*, ha permès l'anàlisi de la densitat gènica de les regions on es troben els CiLINEs, mitjançant l'estudi de les regions flanquejants. Aquests estudis ens han permès mostrar com aquests elements s'inserten en regions on la densitat gènica és significativament menor que la mitjana del genoma (1gen/15,8Kb vs 1gen/7,5Kb). Com que el genoma de l'ascidi és a nivell de complement gènic qualitativament similar al de l'amfiox, és esperable que aquesta distribució també la trobem en l'amfiox. Tot i no poder analitzar les regions flanquejants a totes les còpies de l'element MgLINE, el genoma de la mixina és *a priori* més proper qualitativament al genoma dels vertebrats que no pas al dels cordats invertebrats i per tant s'espera que la distribució dels elements transposables sigui similar a la que observem en genomes com el de l'home. De fet, les ~30 Kb flanquejants a les 3 còpies seqüenciades no contenen cap gen i, per tant, l'element MgLINE es trobi probablement en regions on la densitat gènica és molt baixa a causa o a conseqüència de la inserció d'aquest element. Això s'observa per a l'element L1 del genoma humà on moltes de les còpies estimades per a aquest es troben en regions heterocromàtiques que molt difícilment seran seqüenciades (revisat a (Graham 2006).

La disponibilitat de la seqüència genòmica d'un ampli ventall d'organismes ha permès determinar una relació empírica entre la mida del genoma i la fracció d'aquest que correspon a TEs (Kidwell 2002). Així, es possible establir *a priori* la càrrega mòbil de qualsevol genoma en funció de la seva mida. Si comparem els valors reals i els que aquesta relació prediu, observem com aquesta relació s'ajusta molt bé per als genomes grans però no per als petits. Aquest esbiaix indica que la mida dels genomes grans és proporcional als TEs que presenta però no en els genomes petits ja que en aquests probablement existeixen altres factors més importants a l'hora de determinar la mida del genoma. Quan s'aplica aquesta relació als animals estudiats en aquest treball

s'estima una càrrega mòbil del 5,4% a l'ascidi, 38,1% a l'amfiox i 50,6% a la mixina. Aquestes estimes indiquen que el genoma dels àgnats tindrien un genoma similar al d'altres vertebrats on un 50% o més del seu genoma correspon a TEs. També mostra com el genoma de l'ascidi és similar al d'altres invertebrats com *C. elegans* ja que té un genoma petit amb una fracció mòbil discreta, fet que es va evidenciar quan es va publicar la seqüència genòmica completa (Dehal 2002). L'estima realitzada per a l'amfiox és la més sorprenent de totes ja que els valors obtinguts són similars als observats a vertebrats però no són congruents amb els treballs previs (Permanyer 2006, Holland 2006 i Osborne 2006) que indiquen en tots els casos una fracció mòbil més similar a la de l'ascidi. Per tant, si la relació establerta entre la mida del genoma i la fracció que correspon a tots els TEs és certa, el genoma de l'amfiox hauria d'estar clarament enriquit en TEs no descrits fins al moment en aquest organisme. Finalment, cal destacar que a mesura que es vagin seqüenciant nous organismes i se'n determini la fracció mòbil, aquesta relació empírica millorarà i probablement permetrà establir clarament el paper dels TEs a l'hora de determinar la mida genòmica total (figura 9).

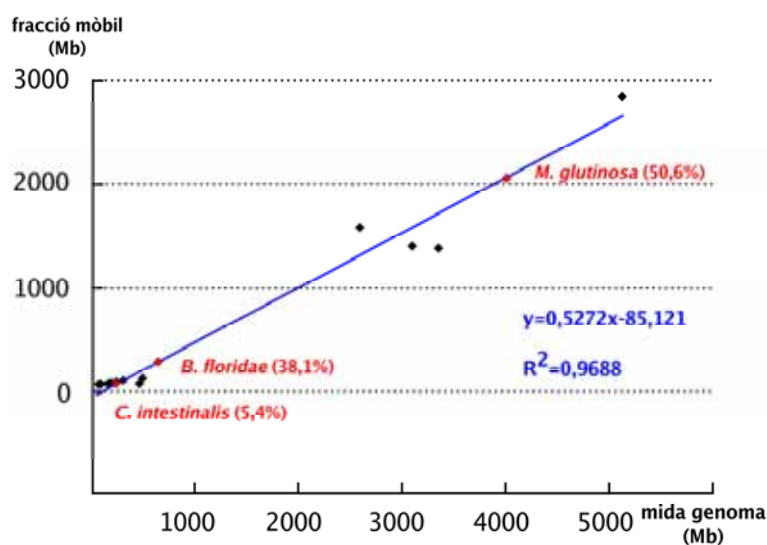


Figura 9: Representació gràfica de la relació empírica entre la mida del genoma i la fracció mòbil que presenta. Els rombes negres indiquen els valors dels genomes de *S. cerevisiae* (Kim 1998), *D. discoideum* (Glockner 20001), *C. elegans* (Waterston 1995), *A. thaliana* (Initiative 2000), *D. melanogaster* (Vieira 1999), *A. gambiae* (Holt 2002), *F. rubripes* (Aparicio 2002), *O. sativa* (Turcotte 2001), *Z. mays* (SanMiguel 1996), *H. sapiens* (Lander 2001), *M. musculus* (Waterston 2002) i *H. vulgare* (Kumar 1999). Els rombes vermells indiquen les estimes realitzades per a aquest treball. En blau s'indica la recta resultant de la regressió lineal realitzada.

Per tant, semblaria que en els organismes cordats no existeix un patró genòmic comú. Així, el genoma d'ascidis i amfioxos tindrien una fracció de TEs molt petita similar a la descrita en genomes de protòstoms, amb 6 dels 14 clades descrits i si els agrupem en els grans grups proposats corresponen a 4 grups en l'ascidi i 5 en l'amfiox. Són elements poc freqüents, i insertats en regions de baixa densitat gènica, on els *young* LINES serien majoritaris. El genoma de la mixina, en contraposició, presentaria ja les característiques descrites per als genomes de vertebrats amb pocs clades però molt abundants i preferentment en regions de molt baixa densitat gènica. Cal destacar que els genomes dels organismes no vertebrats allotgen un menor nombre de retrotransposons non-LTR però que aquests pertanyen a la majoria de grans grups proposats. Per altra banda, els genomes dels vertebrats presenten una fracció corresponent a retrotransposons non-LTR molt significativa però en la que s'han tolerat molt pocs tipus d'elements.

3. Mobilitat

Una part important dels efectes, deleteris o beneficiosos, dels TEs són conseqüència de la seva capacitat de moure's. La mobilitat d'aquests elements no depèn de l'autonomia o integritat de l'element ja que en molts tipus d'elements s'ha vist com es produeix la mobilització en *trans* de còpies defectives o fins i tot d'altres tipus d'elements com els SINEs que es mobilitzen gràcies als retrotransposons non-LTR. Existeixen aproximacions experimentals establertes per als elements L1s humans (Moran 1996) o els R2 de *Bombyx mori* (Luan 1995) que permeten estudiar el mecanisme de transposició, la preferència pel lloc d'inserció o qualsevol aspecte molecular de la transposició d'aquests elements. Aquestes metodologies no permeten, però, establir la mobilitat dels elements autònoms d'un determinat hoste. De totes maneres, existeixen evidències indirectes que poden indicar l'activitat recent d'aquests elements. Així doncs, en un *Southern* genòmic amb DNA de diferents individus hibridat amb una sonda DNA corresponent a un TE, la presència de patrons de bandes diferents entre individus, suggereix la inserció de l'element en diferents llocs genòmics fruit d'una recent activitat. Una evidència similar la podem obtenir fent el mateix *Southern* genòmic i utilitzant com a sonda el fragment de DNA corresponent a la regió del punt d'inserció; això permet valorar la inserció d'un TE en un punt determinat de forma més concreta. A *B. floridae* existeixen individus on els dos al·lels del gen BfPS presenten l'element BfCR1 a l'intró 1, individus que només tenen l'element en un dels 2 al·lels i individus, tot i que en menor mesura, que no presenten l'element en cap al·lel del gen BfPS. Això representa una evidència indirecta de mobilitat per a l'element BfCR1 probablement deguda a una activitat en *trans*, ja que la còpia descrita al gen BfPS és una còpia truncada i per tant necessita de l'acció d'altres elements

per a moure's. Una segona evidència indirecta de mobilitat la trobem quan analitzem els canvis nucleotídics de diferents còpies. Un nombre baix de canvis són idicantius característics d'elements que han estat actius almenys fins fa poc temps i la deriva gènica encara no ha fet divergir les diferents còpies (Sanchez-Gracia 2005). Finalment, la detecció de la transcripció d'aquests elements és una de les millors evidències de la seva activitat, tot i que sempre s'ha de verificar que correspongui a transcripció legítima i no espúria. En aquest sentit es va realitzar l'anàlisi de l'expressió mitjançant RT-PCR de l'element MgLINE ja que *a priori*, en ser més abundant que els CiLINEs i BfLINEs hauria de donar uns nivells d'expressió majors. Aquests assajos però mai varen donar resultats positius, fet congruent amb dades d'altres autors que mostren com a mamífers o *Drosophila melanogaster* els retrotransposons non-LTR tenen uns nivells d'expressió molt restringits tant en el temps com en l'espai (Bourc'his 2004, Kalmykova 2005). Tot i això, la presència d'ESTs corresponents a CiLINEs i BfLINEs, les evidències de mobilització en trans de l'element BfCR1 i els pocs canvis nucleotídics observats entre les 3 còpies seqüenciades de l'element MgLINE i EbLINE, la còpia descrita al mixinoideu *Eptatretus burgeri* indicarien que aquests elements presenten còpies actives que, a més, en alguns casos podrien mobilitzar les còpies truncades. El grau d'activitat per als elements de l'ascidi i l'amfiox és molt baix; el nombre de còpies que hem determinat indicaria que l'activitat és molt limitada en contraposició a l'element MgLINE que seria o hauria estat molt més actiu.

4. Mecanismes de control

Els TEs són quasi invisibles genèticament ja que tot i que poden ser presents amb moltes còpies per genoma haploid, l'elevada presència de còpies no funcionals i la baixa taxa de transposició fa que els efectes que s'observen d'una generació a una altra siguin quasi menyspreables; tot el contrari succeeix amb els efectes a llarg termini ja que en aquest sentit representen potents moduladors evolutius. De l'ampli ventall de dades generades entorn als tipus de TEs i els seus efectes destaca més la sorprenent estabilitat genòmica observada que no pas un desgavell constant provocat per aquests elements. Aquesta observació planteja l'existència de mecanismes de control.

Als anys 90 es va proposar la hipòtesi que indicava com a bona candidata la metilació del DNA com a repressora de la transposició als vertebrats, i per tant es proposà aquest sistema com a mecanisme de control dels TEs (*defense model*) o com a reductor dels nivells d'expressió gènica (*noise reduction*) (Bennetzen 1994, Bird 1995 i Yoder 1997). Per aquest motiu hem estudiat l'estat de metilació dels elements caracteritzats en aquest treball. L'anàlisi del patró de metilació dels elements Ci1, CiL1, CiL2, CiLOA de *C. intestinalis* i BfCR1 de *B. floridae* mostra que aquests elements es troben a la fracció no metilada del genoma d'aquests organismes. L'estat de metilació de l'element

CiR2 és incert ja que mai s'ha arribat a resoldre el *Southern* per manca de senyal, probablement com a conseqüència del baix nombre de còpies. Per altra banda, l'element MgLINE el trobem a la fracció metilada del genoma de *M. glutinosa*. Aquests resultats, congruents amb els resultats parcials que havien generat altres autors per a l'ascidi (Simmen 1999), verifiquen que la metilació dels TEs és una innovació dels vertebrats. Ara bé, la metilació del DNA no és exclusiva de vertebrats tot i que en altres organismes presenta altres funcions directa o indirectament relacionades amb la regulació de l'expressió gènica. A àfids s'ha descrit l'activació de certs gens per metilació i a coccidis s'ha vist que activa la transcripció d'al·lels amb *imprinting* (revisat a Field 2004). Per tant, sembla que invertebrats i vertebrats utilitzarien la metilació com a marcador epigenètic per a reclutar determinades proteïnes que determinarien en cada cas la funció associada a la metilació. En aquest context, la metilació seria una etiqueta per a diferenciar diferents compartiments genòmics i en el cas dels vertebrats s'hauria adoptat la metilació per a identificar regions que han de ser transcripcionalment silenciades (Mandrioli 2004). Ara bé, si la metilació no està involucrada en el control dels TEs en els cordats basals, com es controlen aquests elements? Alguns autors han suggerit l'existència d'una relació entre la compactació del genoma i la reducció tant del nombre de còpies com de la diversitat d'elements (Abrusan 2006, Volff 2004). Aquest fet indicaria la presència de mecanismes que o bé reprimeixen l'expressió -i per tant la transposició- o bé d'altres que facilitin l'eliminació. Aquesta relació no sembla que es compleixi per a alguns dels genomes caracteritzats a data d'avui. El genoma de *T. rubripes*, tot i ser extremadament compacte, té una gran diversitat d'elements (Volff 2003) amb una elevada taxa de recanvi (Aparicio 2002). En canvi, el genoma humà presenta un grau de compactació molt baix i té una diversitat d'elements més aviat petita (Lander 2001). Per tant, tot i que no podem menysprear la grandària del genoma com a factor modulador dels TEs, serien necessaris altres mecanismes de control. La naturalesa d'aquesta maquinària és font de controvèrsia però s'està assolint un consens sobre l'existència de mecanismes de detecció de l'expressió gènica i el reconeixement i processament de seqüències duplicades. Els diferents mecanismes proposats s'han agrupat sota el nom d'*RNA silencing* i ha estat descrit a protistes (Wu-Scharf 2000), fongs (Rossignol 1994, Schramke 2003), plantes (Waterhouse 2001, Lippman 2003) i animals (Bingham 1997). Es planteja un sistema de silenciament transcripcional (*Transcriptional Gene Silencing*) i post-transcripcional (*Post-Transcriptional Gene Silencing*) on hi interactuen els sistemes d'RNAi i els de silenciament de la cromatina, com l'acetilació i la metilació, per tal de silenciar seqüències duplicades (figura 11). Aquest sistema, proposat per a tots els organismes incloent l'home (revisat a Horman 2006) i recentment demostrat a *Drosophila* (Saito 2006), es basa en la detecció de trànscrips estructuralment anòmals (shRNAs), dsRNAs o seqüències duplicades i la generació de siRNAs mitjançant la proteïna Dicer. Els siRNAs

participarien en la degradació del transcrit anòmal (silenciament post-transcripcional) i en presència de RISC reclutarien la maquinària de silenciament de la cromatina (principalment Swi6 i acetilases i metiltransferases d'histones, silenciament transcripcional). El reestructurament de la cromatina activaria una RNA polimerasa dirigida per RNA (RdRP) assegurant així la producció contínua de dsRNAs i siRNAs homòlegs a la regió silenciada (revisat a Buchon 2006). Amb subtils diferències entre organismes a nivell de les proteïnes implicades, aquest sistema és aplicable tant a deuteròstoms com a protòstoms. Als cordats basals, almenys *C. intestinalis* i *B. floridae* no metilen les seqüències duplicades com els retrotransposons non-LTR i per tant cal establir quins són els mecanismes encarregats per a marcar-les i reprimir-les. De fet, la mida i taxa de recombinació d'un genoma, les diferències en la maquinària de silenciament i la inespecificitat de certs siRNAs (Jackson 2004) són els principals factors que modulen tant el nombre com la diversitat d'elements presents en un genoma (Abrusan 2006). Així, els genomes petits presenten una elevada taxa de recombinació que comporta la presència de pocs TEs molt diversos, per evitar ser eliminats com observem en els genomes de l'ascidi i l'amfiox. En contraposició, els genomes grans tenen una menor recombinació, fet que permet l'augment del nombre de còpies amb una menor diversitat. Aquesta situació podria ser aplicable al genoma de la mixina. Per tant, és possible que l'existència massiva dels TEs en determinats genomes, com els dels mamífers, sigui conseqüència del mecanisme de detecció de l'expressió gènica i el reconeixement i processament de seqüències duplicades. Aquest mecanisme, però, no sembla ser específic per al control dels TEs però gràcies a aquest els TEs s'haurien expandit ja que es trobarien sota control i per tant s'haurien minimitzat els possibles efectes deleteris.

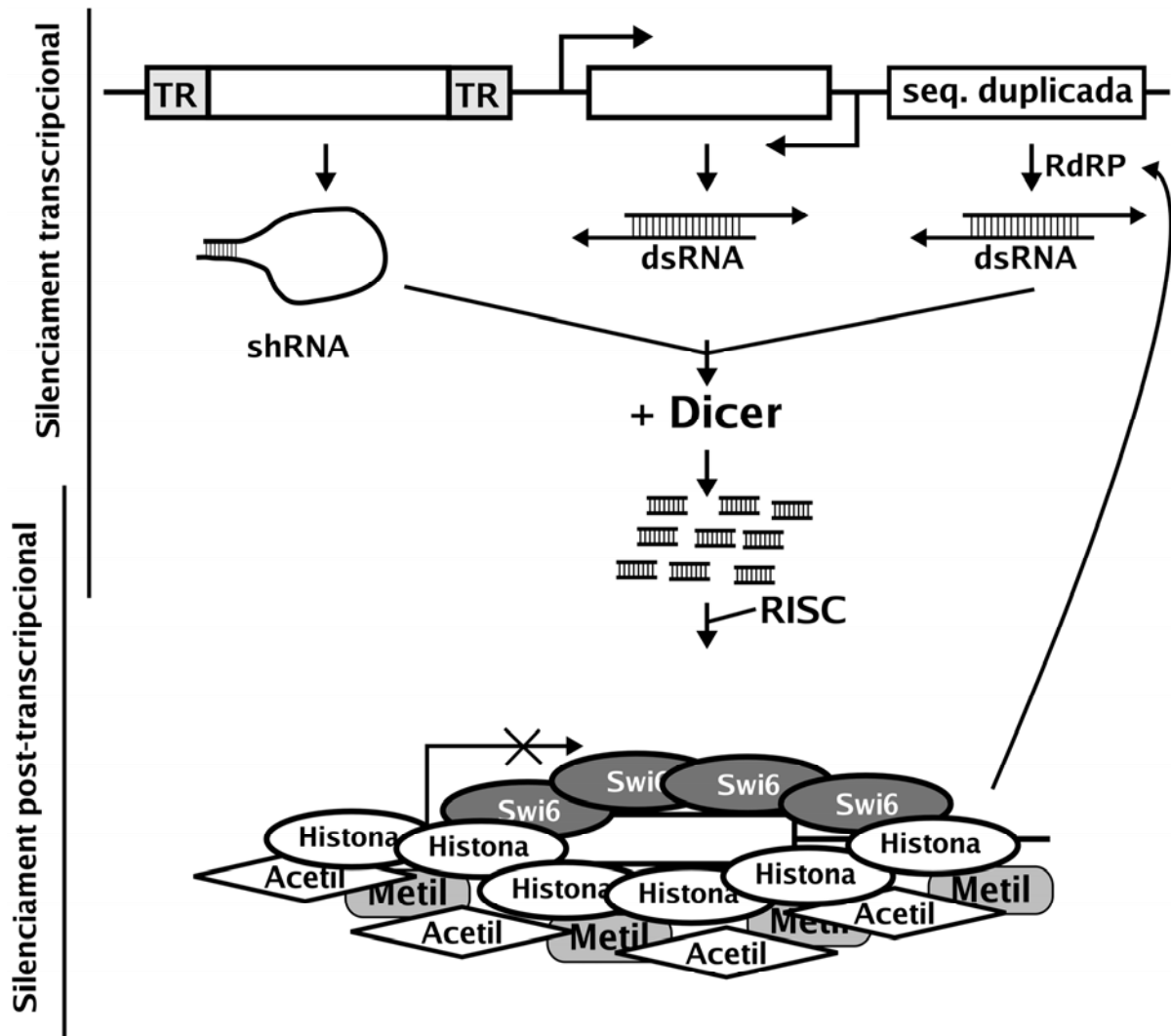


Figura 10: Esquema dels mecanismes d'RNA silencing per al silenciament transcripcional i post-transcripcional. La transcripció de TEs amb repeticions terminals (TR) pot generar shRNA i la transcripció il·legítima o l'acció de la RdRP pot generar dsRNAs. Ambdós RNAs activarien a Dicer que els degradaria (silenciament transcripcional) mitjançant la generació de siRNAs. Aquests també participen en el reclutament de la maquinària de silenciament post-transcripcional de la cromatina on hi intervé RISC. El silenciament de la cromatina es produeix mitjançant l'acetilació i metilació d'histones i la participació de Swi6. A més, la modificació de la cromatina activa la RdRP que assegura el silenciament de la regió mitjançant la síntesi de dsRNAs.

5. Duplicacions genòmiques i expressió dels TEs als cordats

Les dades recollides en aquesta Tesi i la informació derivada de projectes genoma així com descripcions puntuals de determinats TEs permet tenir una visió general sobre determinats aspectes de l'evolució genòmica dels cordats. En general, els cordats no vertebrats presenten genomes qualitativament equiparables que, en comparació amb els vertebrats, tenen una mida petita (genoma haploid <600 Mb) amb relativament pocs gens (a l'ascidi se n'han predit ~16000 gens), una elevada densitat gènica (a l'ascidi s'ha determinat 1 gen/7'5 Kb) i pocs TEs (a l'ascidi s'estima <5%). A més, el contingut GC dels genomes d'aquests organismes és baix i la seva distribució és homogènia (de Luca di Roseto 2002), fet que implica l'absència d'isocores. Aquestes característiques són similars a les descrites per als genomes caracteritzats de protòstoms (Dehal 2002) i permet suggerir que el genoma de l'ancestre dels cordats probablement presentaria aquestes característiques (figura 11).

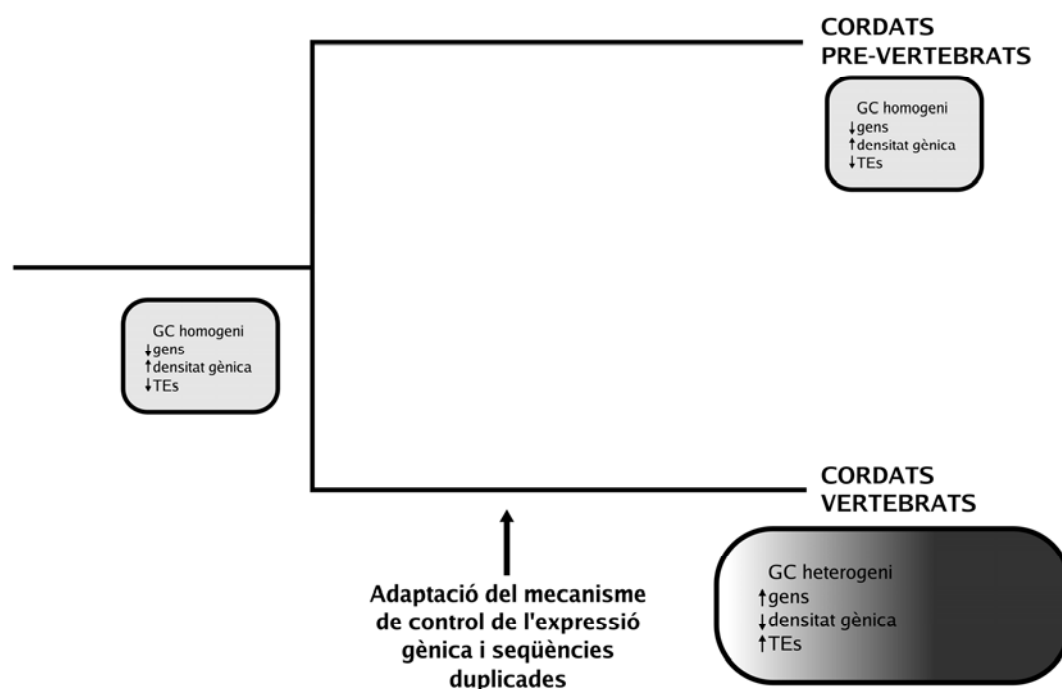


Figura 11: Evolució genòmica dels cordats. La millora en el mecanisme de control de l'expressió gènica i seqüències duplicades hauria permès les duplicacions genòmiques. Ambdós fenòmens són cabdals per a permetre l'expansió de TEs que va remodelar el genoma dels vertebrats.

Al 1970, Ohno proposà que les duplicacions genòmiques podien ser la font de material necessària per incrementar la complexitat genòmica que presenten els vertebrats (Ohno 1970), ja que observà que les diferències tant de mida com de complexitat entre els cordats basals i els vertebrats podrien justificar-se mitjançant fenòmens de duplicacions genòmiques. Avui, gràcies a les dades recollides, podem afirmar que aquestes duplicacions han estat cabdals per assolir la complexitat actual dels vertebrats però no ha estat l'únic fenomen important que ha succeït. Les duplicacions genòmiques s'haurien tolerat gràcies a l'existència dels mecanismes de detecció i control de l'expressió gènica (Craig 2001), en el qual es va adoptar la metilació com a etiqueta per a diferenciar seqüències que s'havien de silenciar, i evitar així, que les duplicacions genòmiques no esdevinguessin un desajust fatal de dosi. Aquest fet va permetre les grans duplicacions genòmiques que van provocar un augment en el complement gènic i els espais intergènics. Aquests canvis van permetre que els genomes duplicats fossin més tolerants a la càrrega massiva de TEs principalment per tres motius. Primer, les duplicacions genòmiques redueixen el risc de perdre una funció gènica com a conseqüència de la inserció d'un TE, ja que es disposa de més d'una còpia per a cada gen. Segon, les duplicacions genòmiques afecten tant als gens com a les regions intergèniques i per tant la duplicació genòmica aportaria noves regions on insertar-se sense interferir amb els gens. I tercer, la mateixa maquinària que hauria permès les duplicacions controlaria els TEs, minimitzant-ne l'expressió i reduint la recombinació entre còpies mitjançant la heterocromatització de la regió. Aquestes noves condicions haurien permès l'augment dels TEs en l'ancestre dels vertebrats. Així, els genomes recentment duplicats i amb un augment dels TEs van patir els seus efectes en tots els sentits, tant els deleteris com els beneficiosos, augmentant-ne el potencial evolutiu mitjançant la generació de punts calents de recombinació, aportant el mecanisme de *exon/genomic shuffling* i aportant una font de motius idonis per a ser utilitzats com a elements reguladors transcripcionals, senyals de poliadenilació i dominis proteics entre altres. D'aquesta manera podrien haver evolucionat els genomes dels vertebrats que són qualitativament oposats als que trobem en els cordats no vertebrats i probablement en l'ancestre dels cordats. Els genomes dels vertebrats són altament complexos i es caracteritzen per un contingut GC molt heterogeni, un complement gènic diferencialment distribuït en les diferents isocores i clarament superior als seus ancestres invertebrats, una elevada taxa d'*splicings* alternatius, una baixa densitat gènica i una fracció mòbil molt significativa que es troba principalment en regions heterocromàtiques (Lander 2001). Aquests genomes tenen un funcionament molt farragós però l'èxit evolutiu d'aquests sembla evident i probablement en aquest context l'expansió dels TEs haurien esdevingut més un avantatge que no pas un llast.

Conclusions

1.- L'anàlisi *in silico* dels projectes genoma de *Ciona intestinalis* i *Branchiostoma floridae* i l'ús de metodologies experimentals de cerca específica d'elements en el genoma de *Myxine glutinosa* ens han permès caracteritzar 5 retrotransposons non-LTR en l'ascidi, 6 en l'amfiox i un a la mixina. Els elements han estat classificats segons la seva afiliació filogenètica, l'homologia a elements coneguts i la presència dels dominis estructurals de cada clade. Els nous membres descrits pertanyen als clades I (CiI i BfI), **CR1** (BfCR1 i MgLINE), **L1** (CiL1 i BfL1), **L2** (CiL2 i BfL2), **LOA** (CiLOA), **NeSL** (BfNeSL), **R2** (CiR2) i **RTE** (BfRTE).

2.- El nombre de còpies dels elements non-LTR de l'ascidi i l'amfiox és baix en comparació als seus homòlegs de vertebrats. L'element MgLINE és molt més abundant, tot i que no arriba als valors descrits pels elements més prolífics dels vertebrats. L'abundància relativa entre els diferents tipus mostra que els *young* LINEs són més abundants que els *old* LINEs, probablement a causa d'un marge molt més ampli de possibles dianes d'inserció que li que li confereix l'activitat endonucleàsica dels primers.

3.- La metilació de l'entorn genòmic de l'element MgLINE, en contraposició a l'absència de metilació en els elements de l'ascidi i l'amfiox, suggereix que la possible regulació per metilació de l'activitat dels TEs està restringida als vertebrats. Descartada la metilació com a mecanisme de control en els cordats no vertebrats, mecanismes com la cossupressió o els agrupats sota el nom d'*RNA silencing* podrien ser els responsables d'aquest control.

4.- La capacitat de mobilització dels retrotransposons non-LTR és difícil de determinar, però la presència d'ESTs dels CiLINEs i BfLINEs, les evidències indirectes de mobilització de l'element BfCR1 i el nombre limitat de canvis nucleotídics entre les diferents còpies de l'element MgLINE i de l'element EbLINE d'*Eptatretus burgeri*, suggereixen una recent activitat. En aquest cas, la hipòtesis més probable és que els genomes dels hostes contenen almenys alguns elements íntegres i funcionals.

5.- L'evolució del llinatge dels vertebrats ha estat marcat per un augment del complement genètic mitjançant duplicacions genòmiques totals o parcials. Reordenacions que haurien estat tolerades per innovacions en els mecanismes de control de l'expressió genètica de les seqüències duplicades. Aquest fenomen hauria permès l'expansió en el número de còpies dels TEs, augmentant el potencial evolutiu dels vertebrats primitius.

Bibliografia

A

- Abrusan, G. and H. J. Krambeck (2006). "Competition may determine the diversity of transposable elements." *Theor Popul Biol* **70**(3): 364-75.
- Adams, M. D., S. E. Celniker, et al. (2000). "The genome sequence of *Drosophila melanogaster*." *Science* **287**(5461): 2185-95.
- Agrawal, A., Q. M. Eastman, et al. (1998). "Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system." *Nature* **394**(6695): 744-51.
- Aparicio, S., J. Chapman, et al. (2002). "Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*." *Science* **297**(5585): 1301-10.

B

- Bailey, J. A., G. Liu, et al. (2003). "An Alu transposition model for the origin and expansion of human segmental duplications." *Am J Hum Genet* **73**(4): 823-34.
- Bennetzen, J. L., K. Schrick, et al. (1994). "Active maize genes are unmodified and flanked by diverse classes of modified, highly repetitive DNA." *Genome* **37**(4): 565-76.
- Best, S., P. Le Tissier, et al. (1996). "Positional cloning of the mouse retrovirus restriction gene Fv1." *Nature* **382**(6594): 826-9.
- Bingham, P. M. (1997). "Cosuppression comes to the animals." *Cell* **90**(3): 385-7.
- Bird, A. P. (1995). "Gene number, noise reduction and biological complexity." *Trends Genet* **11**(3): 94-100.
- Boeke, J. D. and V. G. Corces (1989). "Transcription and reverse transcription of retrotransposons." *Annu Rev Microbiol* **43**: 403-34.
- Boeke, J. D., D. J. Garfinkel, et al. (1985). "Ty elements transpose through an RNA intermediate." *Cell* **40**(3): 491-500.
- Bourc'his, D. and T. H. Bestor (2004). "Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L." *Nature* **431**(7004): 96-9.
- Brenner, S., G. Elgar, et al. (1993). "Characterization of the pufferfish (*Fugu*) genome as a compact model vertebrate genome." *Nature* **366**(6452): 265-8.
- Brett, D., J. Hanke, et al. (2000). "EST comparison indicates 38% of human mRNAs contain possible alternative splice forms." *FEBS Lett* **474**(1): 83-6.
- Brookfield, J. F. (2005). "The ecology of the genome - mobile DNA elements and their hosts." *Nat Rev Genet* **6**(2): 128-36.
- Buchon, N. and C. Vaury (2006). "RNAi: a defensive RNA-silencing against viruses and transposable elements." *Heredity* **96**(2): 195-202.
- Burke, W. D., C. C. Calalang, et al. (1987). "The site-specific ribosomal insertion element type II of *Bombyx mori* (R2Bm) contains the coding sequence for a reverse transcriptase-like enzyme." *Mol Cell Biol* **7**(6): 2221-30.
- Burke, W. D., H. S. Malik, et al. (1999). "The domain structure and retrotransposition mechanism of R2 elements are conserved throughout arthropods." *Mol Biol Evol* **16**(4): 502-11.

C

- Canestro, C., R. Albalat, et al. (2003). "Isolation and characterization of the first non-autonomous transposable element in amphioxus, ATE-1." Gene **318**: 69-73.
- Canestro, C., L. Hjelmqvist, et al. (2000). "Amphioxus alcohol dehydrogenase is a class 3 form of single type and of structural conservation but with unique developmental expression." Eur J Biochem **267**(22): 6511-8.
- Capy, P., G. Gasperi, et al. (2000). "Stress and transposable elements: co-evolution or useful parasites?" Heredity **85** (Pt 2): 101-6.
- Cohen, J. B., D. Liebermann, et al. (1985). "Tsp transposons: a heterogeneous family of mobile sequences in the genome of the sea urchin *Strongylocentrotus purpuratus*." Mol Cell Biol **5**(10): 2814-25.
- Corbo, J. C., A. Di Gregorio, et al. (2001). "The ascidian as a model organism in developmental and evolutionary biology." Cell **106**(5): 535-8.
- Craig, N. L. (2001). Mobile DNA II. Washington, D.C., ASM Press.
- Curcio, M. J. and K. M. Derbyshire (2003). "The outs and ins of transposition: from mu to kangaroo." Nat Rev Mol Cell Biol **4**(11): 865-77.

D

- de Luca di Roseto, G., G. Bucciarelli, et al. (2002). "An analysis of the genome of *Ciona intestinalis*." Gene **295**(2): 311-6.
- Dehal, P., Y. Satou, et al. (2002). "The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins." Science **298**(5601): 2157-67.
- Deininger, P. L., J. V. Moran, et al. (2003). "Mobile elements and mammalian genome evolution." Curr Opin Genet Dev **13**(6): 651-8.
- Delsuc, F., H. Brinkmann, et al. (2006). "Tunicates and not cephalochordates are the closest living relatives of vertebrates." Nature **439**(7079): 965-8.

E

- Eickbush, T. H. (1992). "Transposing without ends: the non-LTR retrotransposable elements." New Biol **4**(5): 430-40.
- Eickbush, T. H. and A. V. Furano (2002). "Fruit flies and humans respond differently to retrotransposons." Curr Opin Genet Dev **12**(6): 669-74.

F

- Fawcett, D. H., C. K. Lister, et al. (1986). "Transposable elements controlling I-R hybrid dysgenesis in *D. melanogaster* are similar to mammalian LINEs." Cell **47**(6): 1007-15.
- Field, L. M., F. Lyko, et al. (2004). "DNA methylation in insects." Insect Mol Biol **13**(2): 109-15.

- Finnegan, D. J. (1989). "Eukaryotic transposable elements and genome evolution." Trends Genet **5**(4): 103-7.
- Furlong, R. F. and P. W. Holland (2002). "Were vertebrates octoploid?" Philos Trans R Soc Lond B Biol Sci **357**(1420): 531-44.

G

- Garrett, J. E., D. S. Knutzon, et al. (1989). "Composite transposable elements in the *Xenopus laevis* genome." Mol Cell Biol **9**(7): 3018-27.
- Gibbs, R. A., G. M. Weinstock, et al. (2004). "Genome sequence of the Brown Norway rat yields insights into mammalian evolution." Nature **428**(6982): 493-521.
- Glockner, G., K. Szafranski, et al. (2001). "The complex repeats of *Dictyostelium discoideum*." Genome Res **11**(4): 585-94.
- Goodier, J. L., E. M. Ostertag, et al. (2000). "Transduction of 3'-flanking sequences is common in L1 retrotransposition." Hum Mol Genet **9**(4): 653-7.
- Graham, T. and S. Boissinot (2006). "The genomic distribution of L1 elements: the role of insertion bias and natural selection." J Biomed Biotechnol **2006**(1): 75327.
- Greenwood, A. D., C. Leib-Mosch, et al. (2005). "Abyss1: a novel L2-like non-LTR retroelement of the snakelocks anemone (*Anemonia sulcata*)." Cytogenet Genome Res **110**(1-4): 553-8.

H

- Halanych, K. M. (2004). "The new view of animal phylogeny." Annual Review of Ecology, Evolution, and Systematics **35**(1): 229-256.
- Harendza, C. J. and L. F. Johnson (1990). "Polyadenylation signal of the mouse thymidylate synthase gene was created by insertion of an L1 repetitive element downstream of the open reading frame." Proc Natl Acad Sci U S A **87**(7): 2531-5.
- Hattori, M., S. Kuhara, et al. (1986). "L1 family of repetitive DNA sequences in primates may be derived from a sequence encoding a reverse transcriptase-related protein." Nature **321**(6070): 625-8.
- Heierhorst, J., K. Lederis, et al. (1992). "Presence of a member of the Tc1-like transposon family from nematodes and *Drosophila* within the vasotocin gene of a primitive vertebrate, the Pacific hagfish *Eptatretus stouti*." Proc Natl Acad Sci U S A **89**(15): 6798-802.
- Hillier, L. W., W. Miller, et al. (2004). "Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution." Nature **432**(7018): 695-716.
- Hohjoh, H. and M. F. Singer (1997). "Sequence-specific single-strand RNA binding protein encoded by the human LINE-1 retrotransposon." Embo J **16**(19): 6034-43.
- Holt, R. A., G. M. Subramanian, et al. (2002). "The genome sequence of the malaria mosquito *Anopheles gambiae*." Science **298**(5591): 129-49.
- Holland, L. Z. (2006). "A SINE in the genome of the cephalochordate amphioxus is an Alu element." Int J Biol Sci **2**(2): 61-5.

- Holland, L. Z., V. Laudet, et al. (2004). "The chordate amphioxus: an emerging model organism for developmental biology." *Cell Mol Life Sci* **61**(18): 2290-308.
- Holland, P. W. and J. Garcia-Fernandez (1996). "Hox genes and chordate evolution." *Dev Biol* **173**(2): 382-95.
- Holland, P. W., J. Garcia-Fernandez, et al. (1994). "Gene duplications and the origins of vertebrate development." *Dev Suppl*: 125-33.
- Horman, S. R., P. Svoboda, et al. (2006). "The potential regulation of l1 mobility by RNA interference." *J Biomed Biotechnol* **2006**(1): 32713.
- Hull, R. (2001). "Classifying reverse transcribing elements: a proposal and a challenge to the ICTV. International Committee on Taxonomy of Viruses." *Arch Virol* **146**(11): 2255-61.

I

- Initiative, T. A. G. (2000). "Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*." *Nature* **408**(6814): 796-815.

J

- Jabbari, K. and G. Bernardi (2004). "Body temperature and evolutionary genomics of vertebrates: a lesson from the genomes of *Takifugu rubripes* and *Tetraodon nigroviridis*." *Gene* **333**: 179-81.
- Jackson, A. L. and P. S. Linsley (2004). "Noise amidst the silence: off-target effects of siRNAs?" *Trends Genet* **20**(11): 521-4.
- Jaillon, O., J. M. Aury, et al. (2004). "Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype." *Nature* **431**(7011): 946-57.
- Jensen, S., M. P. Gassama, et al. (1999). "Taming of transposable elements by homology-dependent gene silencing." *Nat Genet* **21**(2): 209-12.

K

- Kajikawa, M., K. Ohshima, et al. (1997). "Determination of the entire sequence of turtle CR1: the first open reading frame of the turtle CR1 element encodes a protein with a novel zinc finger motif." *Mol Biol Evol* **14**(12): 1206-17.
- Kajikawa, M. and N. Okada (2002). "LINEs mobilize SINES in the eel through a shared 3' sequence." *Cell* **111**(3): 433-44.
- Kalmykova, A. I., M. S. Klenov, et al. (2005). "Argonaute protein PIWI controls mobilization of retrotransposons in the *Drosophila* male germline." *Nucleic Acids Res* **33**(6): 2052-9.
- Kaminker, J. S., C. M. Bergman, et al. (2002). "The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective." *Genome Biol* **3**(12): RESEARCH0084.
- Kapitonov, V. V. and J. Jurka (2001). "Rolling-circle transposons in eukaryotes." *Proc Natl Acad Sci U S A* **98**(15): 8714-9.
- Kazazian, H. H., Jr. (2004). "Mobile elements: drivers of genome evolution." *Science* **303**(5664): 1626-32.

- Kidwell, M. G. (2002). "Transposable elements and the evolution of genome size in eukaryotes." Genetica **115**(1): 49-63.
- Kim, J. M., S. Vanguri, et al. (1998). "Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence." Genome Res **8**(5): 464-78.
- Kimmel, B. E., O. K. ole-MoiYoi, et al. (1987). "Ingi, a 5.2-kb dispersed sequence element from *Trypanosoma brucei* that carries half of a smaller mobile element at either end and has homology with mammalian LINES." Mol Cell Biol **7**(4): 1465-75.
- Kojima, K. K. and H. Fujiwara (2004). "Cross-genome screening of novel sequence-specific non-LTR retrotransposons: various multicopy RNA genes and microsatellites are selected as targets." Mol Biol Evol **21**(2): 207-17.
- Kojima, K. K. and H. Fujiwara (2005). "Long-term inheritance of the 28S rDNA-specific retrotransposon R2." Mol Biol Evol **22**(11): 2157-65.
- Kooter, J. M., M. A. Matzke, et al. (1999). "Listening to the silent genes: transgene silencing, gene regulation and pathogen control." Trends Plant Sci **4**(9): 340-347.
- Kordis, D. and F. Gubensek (1998). "The Bov-B lines found in *Vipera ammodytes* toxic PLA2 genes are widespread in snake genomes." Toxicon **36**(11): 1585-90.
- Kreahling, J. and B. R. Graveley (2004). "The origins and implications of Aluternative splicing." Trends Genet **20**(1): 1-4.
- Kumar, A. and J. L. Bennetzen (1999). "Plant retrotransposons." Annu Rev Genet **33**: 479-532.

L

- Lander, E. S., L. M. Linton, et al. (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.
- Li, X., W. A. Scaringe, et al. (2001). "Frequency of recent retrotransposition events in the human factor IX gene." Hum Mutat **17**(6): 511-9.
- Lindblad-Toh, K., C. M. Wade, et al. (2005). "Genome sequence, comparative analysis and haplotype structure of the domestic dog." Nature **438**(7069): 803-19.
- Lippman, Z., B. May, et al. (2003). "Distinct Mechanisms Determine Transposon Inheritance and Methylation via Small Interfering RNA and Histone Modification." PLoS Biol **1**(3): E67.
- Loeb, D. D., R. W. Padgett, et al. (1986). "The sequence of a large L1Md element reveals a tandemly repeated 5' end and several features found in retrotransposons." Mol Cell Biol **6**(1): 168-82.
- Long, M., W. Wang, et al. (1999). "Origin of new genes and source for N-terminal domain of the chimerical gene, jingwei, in *Drosophila*." Gene **238**(1): 135-41.
- Long, Q., C. Bengra, et al. (1998). "A long terminal repeat of the human endogenous retrovirus ERV-9 is located in the 5' boundary area of the human beta-globin locus control region." Genomics **54**(3): 542-55.
- Lovsin, N., F. Gubensek, et al. (2001). "Evolutionary dynamics in a novel L2 clade of non-LTR retrotransposons in Deuterostomia." Mol Biol Evol **18**(12): 2213-24.
- Luan, D. D. and T. H. Eickbush (1995). "RNA template requirements for target DNA-primed reverse transcription by the R2 retrotransposable element." Mol Cell Biol **15**(7): 3882-91.

- Lundin, L. G. (1993). "Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and the house mouse." Genomics **16**(1): 1-19.
- Lyko, F., B. H. Ramsahoye, et al. (2000). "DNA methylation in *Drosophila melanogaster*." Nature **408**(6812): 538-40.

M

- Makalowski, W. (2000). "Genomic scrap yard: how genomes utilize all that junk." Gene **259**(1-2): 61-7.
- Makalowski, W., G. A. Mitchell, et al. (1994). "Alu sequences in the coding regions of mRNA: a source of protein variability." Trends Genet **10**(6): 188-93.
- Malik, H. S., W. D. Burke, et al. (1999). "The age and evolution of non-LTR retrotransposable elements." Mol Biol Evol **16**(6): 793-805.
- Malik, H. S. and T. H. Eickbush (1999). "Modular evolution of the integrase domain in the Ty3/Gypsy class of LTR retrotransposons." J Virol **73**(6): 5186-90.
- Malik, H. S., S. Henikoff, et al. (2000). "Poised for contagion: evolutionary origins of the infectious abilities of invertebrate retroviruses." Genome Res **10**(9): 1307-18.
- Mandrioli, M. (2004). "Epigenetic tinkering and evolution: is there any continuity in the role of cytosine methylation from invertebrates to vertebrates?" Cell Mol Life Sci **61**(19-20): 2425-7.
- Marracci, S., R. Batistoni, et al. (1996). "Gypsy/Ty3-like elements in the genome of the terrestrial Salamander hydromantes (Amphibia, Urodela)." J Mol Evol **43**(6): 584-93.
- McDonald, J. (1995). "Transposable elements: possible catalysts of organismic evolution." Trends Ecol Evol **10**: 123-126.
- Miller, W. J., S. Hagemann, et al. (1992). "P-element homologous sequences are tandemly repeated in the genome of *Drosophila guanche*." Proc Natl Acad Sci U S A **89**(9): 4018-22.
- Mita, K., M. Kasahara, et al. (2004). "The genome sequence of silkworm, *Bombyx mori*." DNA Res **11**(1): 27-35.
- Mizrokhi, L. J., S. G. Georgieva, et al. (1988). "jockey, a mobile *Drosophila* element similar to mammalian LINEs, is transcribed from the internal promoter by RNA polymerase II." Cell **54**(5): 685-91.
- Mizrokhi, L. J. and A. M. Mazo (1990). "Evidence for horizontal transmission of the mobile element jockey between distant *Drosophila* species." Proc Natl Acad Sci U S A **87**(23): 9216-20.
- Moran, J. V., R. J. DeBerardinis, et al. (1999). "Exon shuffling by L1 retrotransposition." Science **283**(5407): 1530-4.
- Moran, J. V., S. E. Holmes, et al. (1996). "High frequency retrotransposition in cultured mammalian cells." Cell **87**(5): 917-27.
- Mount, S. M. and G. M. Rubin (1985). "Complete nucleotide sequence of the *Drosophila* transposable element copia: homology between copia and retroviral proteins." Mol Cell Biol **5**(7): 1630-8.

N

- Nekrutenko, A. and W. H. Li (2001). "Transposable elements are found in a large number of human protein-coding genes." Trends Genet **17**(11): 619-21.
- Nisson, P. E., R. J. Hickey, et al. (1988). "Identification of a repeated sequence in the genome of the sea urchin which is transcribed by RNA polymerase III and contains the features of a retroposon." Nucleic Acids Res **16**(4): 1431-52.
- Norris, J., D. Fan, et al. (1995). "Identification of a new subclass of Alu DNA repeats which can function as estrogen receptor-dependent transcriptional enhancers." J Biol Chem **270**(39): 22777-82.

O

- Ohno, S. (1970). Evolution by gene duplication. Heidelberg, Springer-Verlag.
- Osborne, P. W., G. N. Luke, et al. (2006). "Identification and characterisation of five novel miniature inverted-repeat transposable elements (MITEs) in amphioxus (*Branchiostoma floridae*)." Int J Biol Sci **2**(2): 54-60.
- Ostertag, E. M., J. L. Goodier, et al. (2003). "SVA elements are nonautonomous retrotransposons that cause disease in humans." Am J Hum Genet **73**(6): 1444-51.

P

- Permanyer, J., Gonzalez-Duarte et al. (2003). "The non-LTR retrotransposons in *Ciona intestinalis* : new insights into the evolution of chordate genomes" Genome Biol **4**(11) :R73
- Permanyer, J., R. Albalat, et al. (2006). "Getting closer to a pre-vertebrate genome: the non-LTR retrotransposons of *Branchiostoma floridae*." Int J Biol Sci **2**(2): 48-53.
- Pickeral, O. K., W. Makalowski, et al. (2000). "Frequent human genomic DNA transduction driven by LINE-1 retrotransposition." Genome Res **10**(4): 411-5.

R

- Ray, D. A., D. J. Hedges, et al. (2005). "Chompy: an infestation of MITE-like repetitive elements in the crocodylian genome." Gene **362**: 1-10.
- Regev, A., M. Lamb, et al. (1998). "The Role of DNA Methylation in Invertebrates: Developmental Regulation or Genome Defense? ." Mol Biol Evol **15**(7): 880-891.
- Roeder, G. and Pink (1983). Mobile genetic elements. J. Shapiro, Academic Press: 299-328.
- Rossignol, J. L. and G. Faugeron (1994). "Gene inactivation triggered by recognition between DNA repeats." Experientia **50**(3): 307-17.
- Roy-Engel, A. M., M. L. Carroll, et al. (2002). "Non-traditional Alu evolution and primate genomic diversity." J Mol Biol **316**(5): 1033-40.

S

- Saito, K., K. M. Nishida, et al. (2006). "Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome." Genes Dev **20**(16): 2214-22.
- Sanchez-Gracia, A., X. Maside, et al. (2005). "High rate of horizontal transfer of transposable elements in *Drosophila*." Trends Genet **21**(4): 200-3.
- SanMiguel, P., A. Tikhonov, et al. (1996). "Nested retrotransposons in the intergenic regions of the maize genome." Science **274**(5288): 765-8.
- Schaeffer, B. (1987). "Deuterostome Monophyly and Phylogeny." Evolutionary Biology **21**: 179-235.
- Schramke, V. and R. Allshire (2003). "Hairpin RNAs and retrotransposon LTRs effect RNAi and chromatin-based gene silencing." Science **301**(5636): 1069-74.
- Schwarz-Sommer, Z., L. Leclercq, et al. (1987). "Cin4, an insert altering the structure of the A1 gene in *Zea mays*, exhibits properties of nonviral retrotransposons." Embo J **6**(13): 3873-3880.
- Selker, E. U. (1997). "Epigenetic phenomena in filamentous fungi: useful paradigms or repeat-induced confusion?" Trends Genet **13**(8): 296-301.
- Simmen, M. W. and A. Bird (2000). "Sequence analysis of transposable elements in the sea squirt, *Ciona intestinalis*." Mol Biol Evol **17**(11): 1685-94.
- Simmen, M. W., S. Leitgeb, et al. (1998). "Gene number in an invertebrate chordate, *Ciona intestinalis*." Proc Natl Acad Sci U S A **95**(8): 4437-40.
- Simmen, M. W., S. Leitgeb, et al. (1999). "Nonmethylated transposable elements and methylated genes in a chordate genome." Science **283**(5405): 1164-7.
- Singer, M. F. and J. Skowronski (1985). "Making sense out of LINEs: long interspersed repeat sequences in mammalian genomes." Trends Biochem Sci **10**: 119-122.
- Sinzelle, L., N. Pollet, et al. (2005). "Characterization of multiple lineages of Tc1-like elements within the genome of the amphibian *Xenopus tropicalis*." Gene **349**: 187-96.
- Sirijovski, N., C. Woolnough, et al. (2005). "NfCR1, the first non-LTR retrotransposon characterized in the Australian lungfish genome, *Neoceratodus forsteri*, shows similarities to CR1-like elements." J Exp Zool B Mol Dev Evol **304**(1): 40-9.
- Skrabaneck, L. and K. H. Wolfe (1998). "Eukaryote genome duplication - where's the evidence?" Curr Opin Genet Dev **8**(6): 694-700.
- Springer, M. S., E. H. Davidson, et al. (1991). "Retroviral-like element in a marine invertebrate." Proc Natl Acad Sci U S A **88**(19): 8401-4.

T

- Takezaki, N., F. Figueroa, et al. (2003). "Molecular phylogeny of early vertebrates: monophyly of the agnathans as revealed by sequences of 35 genes." Mol Biol Evol **20**(2): 287-92.
- Traverse, K. L. and M. L. Pardue (1988). "A spontaneously opened ring chromosome of *Drosophila melanogaster* has acquired He-T DNA sequences at both new telomeres." Proc Natl Acad Sci U S A **85**(21): 8116-20.

- Tu, Z., J. Isoe, et al. (1998). "Structural, genomic, and phylogenetic analysis of Lian, a novel family of non-LTR retrotransposons in the yellow fever mosquito, *Aedes aegypti*." Mol Biol Evol **15**(7): 837-53.
- Turcotte, K., S. Srinivasan, et al. (2001). "Survey of transposable elements from rice genomic sequences." Plant J **25**(2): 169-79.

V

- van de Lagemaat, L. N., J. R. Landry, et al. (2003). "Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions." Trends Genet **19**(10): 530-6.
- Van Dellen, K., J. Field, et al. (2002). "LINEs and SINE-like elements of the protist *Entamoeba histolytica*." Gene **297**(1-2): 229-39.
- Vansant, G. and W. F. Reynolds (1995). "The consensus sequence of a major Alu subfamily contains a functional retinoic acid response element." Proc Natl Acad Sci U S A **92**(18): 8229-33.
- Vieira, C., D. Lepetit, et al. (1999). "Wake up of transposable elements following *Drosophila simulans* worldwide colonization." Mol Biol Evol **16**(9): 1251-5.
- Vinogradov, A. E. (2003). "Selfish DNA is maladaptive: evidence from the plant Red List." Trends Genet **19**(11): 609-14.
- Volff, J. N., L. Bouneau, et al. (2003). "Diversity of retrotransposable elements in compact pufferfish genomes." Trends Genet **19**(12): 674-8.
- Volff, J. N., C. Korting, et al. (2001). "Evolution and discontinuous distribution of Rex3 retrotransposons in fish." Mol Biol Evol **18**(3): 427-31.
- Volff, J. N., C. Korting, et al. (2000). "Multiple lineages of the non-LTR retrotransposon Rex1 with varying success in invading fish genomes." Mol Biol Evol **17**(11): 1673-84.
- Volff, J. N., H. Lehrach, et al. (2004). "Retroelement Dynamics and a Novel Type of Chordate Retrovirus-Like Elements in the Miniature Genome of the Tunicate *Oikopleura dioica*." Mol Biol Evol.

W

- Waterhouse, P. M., M. B. Wang, et al. (2001). "Gene silencing as an adaptive defence against viruses." Nature **411**(6839): 834-42.
- Waterston, R. and J. Sulston (1995). "The genome of *Caenorhabditis elegans*." Proc Natl Acad Sci U S A **92**(24): 10836-40.
- Waterston, R. H., K. Lindblad-Toh, et al. (2002). "Initial sequencing and comparative analysis of the mouse genome." Nature **420**(6915): 520-62.
- Wu-Scharf, D., B. Jeong, et al. (2000). "Transgene and transposon silencing in *Chlamydomonas reinhardtii* by a DEAH-box RNA helicase." Science **290**(5494): 1159-62.

Y

Yoder, J. A., C. P. Walsh, et al. (1997). "Cytosine methylation and the ecology of intragenomic parasites." Trends Genet **13**(8): 335-40.

Z

Zhou, Y. H., J. B. Zheng, et al. (2002). "Novel PAX6 binding sites in the human genome and the role of repetitive elements in the evolution of gene regulation." Genome Res **12**(11): 1716-22.

Zimmerly, S., H. Guo, et al. (1995). "Group II intron mobility occurs by target DNA-primed reverse transcription." Cell **82**(4): 545-54.

Zingler, N., O. Weichenrieder, et al. (2005). "APE-type non-LTR retrotransposons: determinants involved in target site recognition." Cytogenet Genome Res **110**(1-4): 250-68.

Zupunski, V., F. Gubensek, et al. (2001). "Evolutionary dynamics and evolutionary history in the RTE clade of non-LTR retrotransposons." Mol Biol Evol **18**(10): 1849-63.