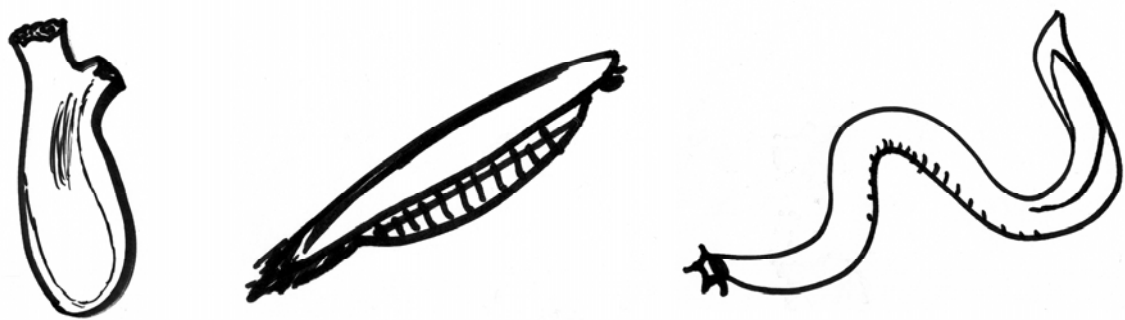




Elements transposables de tipus non-LTR als ascidis, amfioxos i àgnats

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

γνώθι σεαυτόν
Coneix-te tu mateix
Oracle de Delfos

Πάντα χορεύει καὶ οὐδὲν μένει
Tot flueix, res és quiet
Heràclit d'Efes (544 aC-484 aC)

Aquestes paraules que ara escric són probablement la part més agraïda d'escriure de tota la memòria que ve a continuació. Agraïdes no és, en cap cas, sinònim de senzilles ni de simples ja que plasmar a totes les persones a les quals els hi haig d'agrair quelcom és alguna cosa semblant a una odissea.

Primerament voldria agrair la confiança que han dipositat en la meva feina els directors d'aquesta Tesi. Roser, gràcies per haver confiat en un ganàpia que encara no havia acabat la carrera ara farà uns set anys, per il·lusionar-me en els moments durs i per compartir la passió pel meravellós món dels transposons. Ricard, que haig de dir de la persona que va ensenyar-me a fer quasi tot el que ser fer, simplement gràcies.

En segon lloc voldria fer un ampli agraïment a tota la gent que m'ha ajudat tant en la meva formació tant en el dia a dia o perquè no, en el nit a nit. En primer terme vull donar tot el meu agraïment als meus companys de grup. Amb ells he après moltes coses en molts àmbits, hi ha gent que ja no hi és però això no vol dir que els haguem d'oblidar. El Cristian i la Diana van ser els seniors que hem vigilaven (i també manaven) en els meus inicis. La Laura i la Neus, mereixen un capítol separat per elles soles, han estat companyes però no només això, han estat les meves amigues a les quals confiar els problemes que tothom pateix al llarg d'una Tesi però més important encara, han estat les meves *Marilyns* particulars. Una imatge val més que mil paraules i el que jo he vist no hi ha paraules que ho expliquin. No vull oblidar-me la Mari, estrictament no forma part del grup però sempre ha estat la meva companya de *poiata*. Mari, gracias por ayudar siempre y en cualquier situación, gracias por la vidilla que me has dado criticándome y por lo bien que nos lo hemos pasado durante siete años compartiendo *poyata*.

Si sortim del grup, la llista de gent seria també interminable i segur que hi haurà persones que no agrairé però que s'ho mereixen. El Jordi Domènech, *alter* para los amigos, gràcies per fer-me riure, filosofar i veure les coses des d'una perspectiva simplement diferent. També haig de recordar a la gent d'humana tant els que hi són com els que han marxat. Miquel gràcies per acompanyar-me durant murant més de 20 anys; l'Eva, l'AnnaDi, la Sara, la Kika... L'Olga desde el primer dia ja em va marcar el camí a seguir a cop de crit però sempre disposada a ajudar. Al Paps, Solana i Víctor, veïns de l'antic laboratori, tantes estones compartides mereixen un record (a la planta 0 i amb un cafè o en un futbolín amb un cubata). Tampoc m'oblido ni dels Epis del Jordi (Salvatore, Senda, Champi, Nacho i Manu), els de l'Emili (Kike, Mette, Marta i Loli) ni els rafaelitos (Solana *again*, Capullot, Chema i Cisco) Que dir-ne dels drosophileros, especialment en Manel que ha estat company de múltiples fatigues. També vull incloure certs titis que m'han donat suport o m'han ajudat a trencar-me les banyes, Gemma, Jordi, Rafa i Lluís entre altres. També es mereixen un record especial els amics que tinc per Bioquímica, les nenes de TEC i en especial l'Adela i el Josep Lluís.

Portes enfora de la facultat tinc bons amics que han comprès que l'amistat no cal mantenir-la dia a dia ni mes a mes. Gràcies Toni, Laura, Edo, Elvi...

Finalment vull dedicar un sentit agraïment a la família. Pare, ama, no hi ha res

en aquest projecte personal que no sigui fruit del suport i ajut constant que he rebut sempre per part vostra; moltes gràcies per tot. Als meus germans (Iñaki i Ander), també enrolats en la recerca, gràcies pel suport i ànims en els vostres projectes, al final sempre surt el sol. Es mereix un record especial la meva àvia que tot i no entendre res del que faig sempre m'ha mostrat el seu apreci. També es mereixen el meu agraïment la família de Donosti que ha vist com en els darrers anys les meves visites minvaven tot i l'estima que sempre els hi he tingut a tots (Mayte, Iñaki, Mikel i Leyre).

And last but not least, Sandra, gràcies per ser com ets, gràcies per estar al meu costat, gràcies per estimular-me, gràcies per esperar-me, gràcies per fer-me riure, gràcies per fer-me ser com sóc, gràcies per... necessitaria pàgines per agrair-te tantes i tantes coses que, crec que només amb una ho puc resumir: GRÀCIES.

INTRODUCCIÓ	1
1. ELEMENTS TRANSPOSABLES	1
1.1 <i>Definició</i>	1
1.2 <i>Classificació dels Tes</i>	2
1.3 <i>Retrotransposons non-LTR</i>	3
1.4 <i>Elements evolutivament relacionats amb els retrotransposons non-LTR</i>	7
2. TES COM A COMPONENTS DELS GENOMES ANIMALS	9
2.1 <i>Cerca d'elements transposables</i>	9
2.2 <i>TEs en el genoma dels protòstoms</i>	11
2.3 <i>TEs en el genoma dels deuteròstoms</i>	12
3. CONTRIBUCIÓ DELS ELEMENTS TRANSPOSABLES EN ELS GENOMES	17
3.1 <i>Conseqüències de la presència de TEs</i>	17
3.2 <i>Mecanismes de control</i>	19
4. ELS CORDATS EN LA ESCALA EVOLUTIVA	21
OBJECTIUS	27
RESUM GLOBAL	31
1. CILINES, BFLINES I MGLINE	31
2. CÀRREGA I DISTRIBUCIÓ GENÒMICA	36
3. MOBILITAT	40
4. MECANISMES DE CONTROL	41
5. DUPLICACIONS GENÒMIQUES I EXPRESSIÓ DELS TES ALS CORDATS	45
CONCLUSIONS.....	49
BIBLIOGRAFIA.....	53
PUBLICACIONS	65
CAPÍTOL 1: RETROTRANSPOSONS NON-LTR A CIONA INTESTINALIS	65
CAPÍTOL 2: RETROTRANSPOSONS NON-LTR A BRANCHIOSTOMA FLORIDAE	77
CAPÍTOL 3: RETROTRANSPOSONS NON-LTR A MYXINE GLUTINOSA.....	93
APÈNDIX	105

Introducció

1 Elements transposables

1.1 Definició

transposar *v tr* invertir l'ordre en què està posada una cosa respecte a d'altres.

Gran Diccionari de la llengua catalana

transposition *n* the movement of a DNA sequence to another position on the same or a different DNA molecule by replication of the sequence and insertion of the copy either at random or at some preferred but not homologous site (usually without loss of the original sequence), which occurs in both bacteria and eukaryotic cells, the transposable sequences being known generally as transposable elements or transposons, and which variously results in deletions, duplications, and changes in gene expression at the site of insertion

Henderson's Dictionary of Biological Terms

transposable element *n* any of a wide variety of genetic elements (DNA sequences) present both in eukaryotic and prokaryotic genomes that possess the property of inserting themselves by transposition (*q.v.*) at various non-homologous regions on the chromosomes and other DNAs in the bacterial cell or eukaryotic nucleus. Such transpositions can result in the inactivation of genes into which the transposable elements insert or changes in the expression of the nearby genes. Permanent insertions into genes in germline cells results in heritable mutations.

Henderson's Dictionary of Biological Terms

transposon is a DNA sequence able to insert itself at a new location in the genome (whithout any sequence relationship with the target locus).

Genes VIII

Totes les definicions coincideixen que els elements transposables són un grup heterogeni de fragments de DNA que poden moure's -transposar-se- d'una localització genòmica a una altra.

Al llarg del segle XX, la visió que teníem dels genomes va canviar dràsticament; passant d'una visió estable, quasi fixa, a considerar el genoma com a un sistema d'emmagatzemar informació dinàmic i flexible. A la primera meitat del segle passat, els genomes eren considerats com una cadena de gens que es trobaven ordenats en els cromosomes, susceptibles d'incorporar canvis graduals, mitjançant les mutacions a l'atzar, que es produïen a un ritme constant. Però, gràcies als treballs pioners sobre "*les irregularitats incorregibles del color del blat de moro*" de Barbara McClintock a la dècada dels 40 (Figura 1), les primeres evidències sobre la flexibilitat dels genomes van sortir a la llum quan aquesta investigadora va demostrar que els elements mòbils *Ds* i *Ac* eren responsables de la alteració en el color del gra de blat de moro. Aquesta nova visió dels genomes es va anar consolidant a mesura que es

van recollir noves dades sobre altres elements en diferents organismes; a la dècada dels 60 i dels 70, els descobriments dels DNAs mòbils a procariotes van esdevenir cabdals i un cop es va disposar de les eines moleculars per a la caracterització d'aquests elements a eucariotes a principis de la dècada dels 80, es va veure que els elements transposables (TEs, de l'anglès *Transposable Element*) eren components universals del genoma de tots els organismes vius.

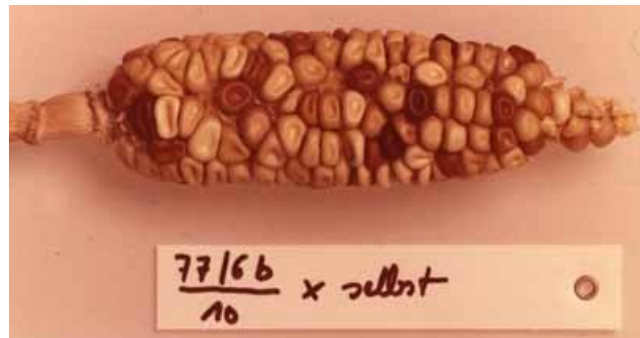


Figura 1: Panotxa de blat de moro dels estudis de B. McClintock on s'observa una coloració anòmla en els diferents grans.

1.2 Classificació dels TEs

Els estudis posteriors a la seva descoberta i l'important impacte dels projectes de seqüenciació genòmica ha generat una enorme quantitat d'informació d'aquests elements que fa necessari un sistema de classificació per als diferents tipus descrits. La millor classificació seria aquella que tingués en compte la relació filogenètica entre els elements. Això, però, és difícil ja que les filogènies es realitzen a partir de fragments, o mòduls, dels TEs i això implica que la filogènia d'una part és representativa de tot l'element. Els TEs estan formats per un conjunt de mòduls amb orígens i filogènies diferents (Malik 2000) i per tant l'anàlisi d'un mòdul pot no ser representativa de l'evolució de l'element ja que els diferents mòduls poden tenir històries evolutives diferents. Des del descobriment dels TEs s'ha evidenciat que són molt més variats i abundants del que es pensava inicialment i que s'aniran descobrint nous tipus d'elements a mesura que es vagin estudiant. Per això, recentment, s'han proposat noves maneres de classificar-los, taxonòmicament (Hull 2001) o en funció de les proteïnes implicades en la transposició (Curcio 2003). Ara bé, no existeix cap sistema que classifiqui clarament tots els tipus d'elements descrits ni que determini la relació filogenètica que hi ha entre ells. Per aquest motiu, la classificació més utilitzada és la més simple, la proposada per D. Finnegan (Finnegan 1989) en 2 grans classes, classe I i classe II, en funció del seu mecanisme de transposició. Dins de cada classe i en funció de la seva capacitat de mobilitzar-se, els elements transposables s'agrupen també en elements autònoms -aquells que poden transposar-se per ells mateixos ja que codifiquen per als enzims necessaris per a fer-ho- i no autònoms -aquells que

necessiten de l'acció en *trans* d'activitats enzimàtiques d'altres elements-. Aquesta darrera divisió no s'ha de confondre amb la divisió entre elements funcionals i no funcionals ja que els elements no autònoms poden ser funcionals gràcies a l'acció en *trans* d'elements autònoms i els elements autònoms poden ser no funcionals per acumulació, entre altres, de mutacions o delecions.

Els elements de classe I, o retroelements, es caracteritzen per un mecanisme de transposició en el qual hi participa un intermediari d'RNA que es retrotranscriu per a ser integrat. Aquesta classe agrupa als elements autònoms amb LTR (*Long Terminal Repeat* o repetició terminal llarga) o retrotransposons LTR, els elements sense LTR o retrotransposons non-LTR i als elements no autònoms de tipus SINEs (*Short Interspersed Nuclear Element* o element nuclear dispers curt). Alguns autors consideren que els retrovirus també s'agrupen en aquesta classe ja que estructuralment són extremadament similars als retrotransposons LTR (Hull 2001); altres consideren que el terme retrotransposons non-LTR agrupa tant els elements autònoms com els no autònoms. Alguns elements característics d'aquesta classe són els *Ty* (*Transposon Yeast*) com a retrotransposons LTR de llevats, l'exitós L1 del genoma de molts vertebrats superiors dins els retrotransposons non-LTR i les seqüències Alu com a SINE majoritari del genoma humà.

Els elements transposables de classe II, transposons *sensu strictu* o transposons de DNA, agrupen elements que es transposen directament, sense intermediaris d'RNA. Aquesta classe representa un grup molt heterogeni com els elements de tipus MITE, FB o TIR amb mecanismes de transposició molt diferents. Alguns elements característics d'aquesta classe són: els elements *Ac/Ds* de *Zea mays*, els elements P de *Drosophila* o els Tc1/Mariner presents a molts animals.

1.3 Retrotransposons non-LTR

Els genomes dels eucariotes superiors estan farcits d'elements transposables, alguns actius, però la majoria presents com a velles còpies degenerades i truncades. L'estudi dels genomes dels animals ha evidenciat que malgrat haver-hi diferents tipus d'elements transposables, els retroelements són els que han colonitzat més exitosament els seus genomes (Figura 2). El mecanisme de transposició dels retroelements, tot i que diferent per a cada tipus, té en comú el fet de no escindir la còpia original per a poder transposar-se. Això implica que el nombre de còpies presents va augmentant a mesura que l'element es va transposant fins arribar a un equilibri dinàmic entre la creació de noves còpies i l'eliminació de les existents que acaba determinant l'abundància de cada tipus (revisat a Brookfield 2005).

Els retrotransposons non-LTR -també anomenats elements LINE (*Long Interspersed Nuclear Element* o element nuclear dispers llarg), LINE-like o retrotransposons poly(A) o retroposons- representen un dels retroelements amb

més èxit en els vertebrats, amb $1.4 \cdot 10^6$ còpies per genoma haploid estimades en l'home (~21% del genoma). Són autònoms i tenen una mida d'entre 4000 i 8000 pb. Tot i ser presents en la majoria d'eucariotes, la seva descoberta va ser tardana degut a la imprecisa integració i a la seva absència a *Saccharomyces cerevisiae*. La presència, en certs casos, d'una cua d'adenines va fer pensar que es tractava d'un retroelement però l'absència de seqüències terminals va portar a pensar que es tractava d'elements no autònoms. Això va portar a classificar-los com a DNA repetitiu del grup dels pseudogens processats i els SINEs, anomenant-los LINEs per analogia (Singer 1985). Posteriorment, amb la descoberta a primats i ratolí de la presència de pautes de lectura amb homologia a transcriptases inverses, es postulà que aquests elements eren de fet un nou tipus d'elements autònoms (Hattori 1986 i Loeb 1986) i es van anar caracteritzant a diferents organismes (Burke 1987, Fawcett 1986, Kimmel 1987 i Schwarz-Sommer 1987).

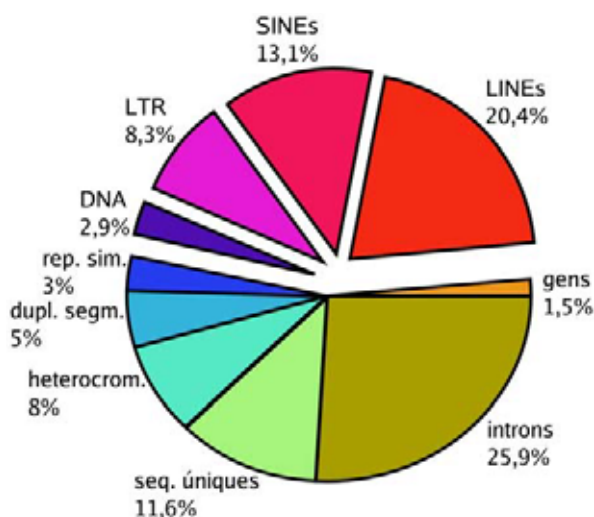


Figura 2: El genoma humà es caracteritza per la presència de 2 tipus de TEs majoritaris, els retrotransposons non-LTR o LINEs i els SINEs. Aquests 2 grups, representats principalment pel LINE L1 i pel SINE Alu, representen el 33,5% del genoma humà. Si considerem la resta d'elements transposables aquesta fracció representa quasi el 45% del genoma humà (Lander 2001). LTR: retrotransposons LTR, DNA: transposons de DNA, rep.sim.: repeticions simples, dupl. segm.: duplicacions segmentals, heterocrom.: heterocromatina, seq. úniques: seqüències de còpia única.

Estructuralment, els retrotransposons non-LTR es caracteritzen per la presència d'una o dues pautes de lectura segons el tipus d'elements, ORF1 i ORF2 (revisat a Malik 1999). Tots ells presenten almenys la ORF2 que codifica per a una proteïna amb diferents dominis enzimàtics entre els que sempre trobem una activitat transcriptasa inversa (RT, de l'anglès *Reverse Transcriptase*) i una endonucleasa, i optativament, dominis addicionals d'unió

a àcids nucleics o amb activitat ribonucleasa H (RNH). S'ha postulat que la primera pauta de lectura, ORF1, podria codificar per una proteïna que retornaria el transcrit al nucli un cop traduït o bé permetria la unió entre la proteïna de la ORF2 al RNA per ser retrotranscrit (*cis-binding*) (Hohjoh 1997).

Aquests elements es transposen mitjançant el model TPRT (*Target-Primed Reverse Transcription* o transcripció inversa encebada a la diana) proposada inicialment per als elements R2 (Luan 1995). Aquest model explica per què existeixen còpies truncades per 5', com es generen els diferents tipus de duplicacions del lloc d'inserció per un mateix tipus d'element i la integració en absència d'una integrasa. La transposició d'aquests elements comença per la transcripció de l'RNA promoguda per un promotor intern (Mizrokhi 1988) encara que també existeixen casos de transcripció a partir de promotors no propis (Eickbush 1992). El transcrit és usat tant per a ser traduït i produir les proteïnes necessàries per a la transposició com per a fer de motllo de la transcriptasa inversa. Un cop transcrit, l'mRNA i les proteïnes traduïdes tornen al nucli. L'activitat endonucleasa de la proteïna ORF2 de l'element genera un tall en una de les cadenes, en el lloc on s'insertarà la nova còpia generant així un grup 3' hidroxil lliure. Un cop s'ha generat el tall, la transcriptasa inversa s'uneix a l'extrem 3' de l'RNA que serà retrotranscrit i el posiciona a prop d'on s'integrarà el nou element però sense unir-lo al DNA. La interacció de l'RNA al DNA és inestable i implica que la transcriptasa inversa pot necessitar diversos intents abans d'iniciar amb èxit la retrotranscripció; aquesta inestabilitat genera repeticions simples a l'extrem 3' del nou element (Boeke 1989). La baixa processivitat de la RT és un fet característic d'aquests elements i explica la presència d'un elevat nombre d'elements truncats per 5'. Durant el procés de retrotranscripció es produeix el tall de la segona cadena del DNA hoste, aquest però no és en el mateix punt que el primer i això generarà les duplicacions en tàndem que normalment flanquegen aquests elements ja que es genera un extrem protuberant que s'haurà de reparar i generarà així les repeticions típiques del punt d'inserció o TSD (*Target Site Duplications* o duplicacions del lloc d'inserció). Per a completar el procés cal degradar l'RNA motllo i, tot i que alguns retrotransposons non-LTR contenen un domini amb activitat RNAsaH, la majoria no el tenen. Es creu que en absència del domini RNAsaH, aquesta tasca la realitzen enzims de l'hoste. Finalment, la síntesi de la segona cadena no està resolt en el model ja que encara no se sap si és duta a terme per la mateixa transcriptasa inversa o hi participa la maquinària de reparació del DNA de l'hoste.

Els estudis filogenètics dels retrotransposons non-LTR s'han fet comparant l'únic domini que tots ells comparteixen, el domini RT. L'anàlisi estructural d'aquest domini mostra com la proteïna es pot plegar en una estructura de "mà dreta" com la descrita per a la transcriptasa inversa de retrovirus (Burke 1999) tot i que conté 2 segments addicionals a la "mà" i una regió d'uns 100 residus que correspon al "polze" (Malik 1999) que no es troben en les RT retrovirals. La

seqüència corresponent a RT que pot ser usada per als anàlisis filogenètics s'extén uns 440 residus que comprenen 11 dominis conservats anomenats 0, 1, 2, 2a, 3, 4, 5, 6, 7, 8 i 9. Els estudis filogenètics basats en aquests dominis han definit 14 clades de retrotransposons non-LTR (Figura 3a) que reben el nom de l'element més ben caracteritzat dins de cada grup; els clades descrits són CR1, CRE, I, Jockey, L1, L2, LOA, NeSL, R1, R2, R4, Rex1, RTE i Tad1. Ara bé, a mesura que es van caracteritzant nous elements en noves espècies, aquesta classificació en clades va augmentant en nombre sense que es resolgui com i quan es van originar tots aquests clades. És per això que darrerament s'ha redefinit la classificació tenint en compte la relació filogenètica basada en el domini RT i la presència o absència i ordenació d'alres dominis. Amb aquesta nova classificació es defineixen 5 grans grups: R2 (agrupa CRE, NeSL, R2 i R4), L1, RTE, I (I, LOA, R1 i Tad1) i Jockey (CR1, Jockey, L2 i Rex1) (Tu 1998) (Figura 3b).

Aquests elements també s'han classificat en funció del tipus d'endonucleasa que presenten ja que hi ha elements que contenen endonucleases similars a enzims de restricció (*restriction enzyme-like endonuclease* o REL-endo) i endonucleases apuríniques/apirimidíniques (*apurynic/apirymidinic endonuclease* o APE). Aquesta classificació és alhora filogenètica i funcional. Així, el grup dels anomenats LINEs vells (*old* LINEs), probablement els evolutivament més antics, presenten una única pauta de lectura que, a més de la RT, codifica per a una REL-endo que fa talls al DNA en dianes específiques. Els LINEs joves (*young* LINEs), originats possiblement a partir dels elements anteriors, agrupen els elements més abundants, tant en tipus com en nombre. Es troben estructurats en 2 pautes de lectura oberta i els defineix una endonucleasa APE que genera talls en el DNA de manera inespecífica que va permetre la inserció més generalitzada en el genoma hoste i hauria facilitat la diversificació dels llinatges. Un d'aquests llinatges va adquirir també un domini RNH i va generar els actuals clades I, LOA, R1 i Tad1, tots ells corresponent al grup I (Figura 3c).

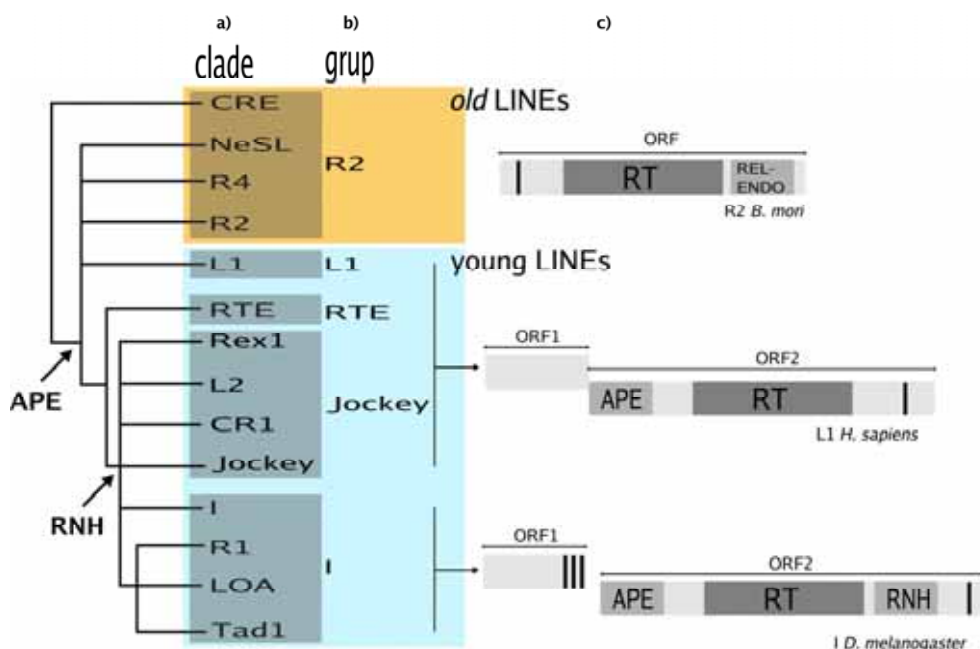


Figura 3: Tots els retrotransposons non-LTR s'han agrupat en almenys 14 clades filogenètics (a) que alhora es poden agrupar en 5 grans grups (b) en funció de la relació filogenètica basada en la RT i la presència i ordre d'altres dominis. Existeix també una classificació funcional i filogenètica en funció del tipus d'endonucleasa (c). Els *old* LINEs, representats per l'element R2 de *B. mori* que es troba estructurat en una única ORF. Els *young* LINEs presenten dues ORFs que poden solapar entre elles. APE: endonucleasa apurínica/apirimidinica, REL-endo: endonucleasa similar a enzims de restricció, RNH: RNAsaH, RT: transcriptasa inversa. Les barres negres verticals indiquen dominis d'unió a àcids nucleics.

1.4 Elements evolutivament relacionats amb els retrotransposons non-LTR

Els retroelements inclouen no només els retrotransposons LTR i non-LTR sinó que també els retrons d'eubacteris, els plàsmids mitocondrials de *Neurospora crassa*, les telomerases eucariotes i els introns de tipus II de mitocondris i eubacteris. Aquests darrers són considerats representants de l'estructura primigènia a partir de la qual s'han originat la resta de retroelements. Tots els retroelements es caracteritzen per la presència d'un domini enzimàtic amb homologia a RT que s'ha trobat a eubacteris i eucariotes però no a arquees, pel que es creu que els retroelements es van originar als eubacteris d'on varen passar als eucariotes tot i que l'opció contrària -un origen eucariòtic i una dispersió cap als eubacteris- és possible, però difícil d'explicar (Zimmerly 1995 i Malik 1999).

Els retrotransposons non-LTR semblen haver participat en l'origen dels retrotransposons LTR. Els retrotransposons LTR contenen en els extrems una estructura amb repeticions que els és característica: són elements autònoms de 6000-11000 pb que codifiquen per a la maquinària de transposició i propagació

(gens *gag*, *env* i *pol*). Els retrotransposons LTR varen ser els primers retroelements descrits a eucariotes (Boeke 1985 i Mount 1985). L'estructura d'aquests elements i els dominis per als quals codifiquen són molt similars als retrovirus i estan estructurats en 1, 2 o 3 pautes de lectura, en molts casos amb solapament parcial entre pautes, segons al tipus d'element. En un primer moment es pensà que els retrotransposons LTR tenien un origen molt antic però l'absència d'aquests elements tant a eubacteris com a arquees no recolzaria aquesta hipòtesi. Recentment s'ha proposat que els retrotransposons LTR s'originaren a partir d'una fusió entre un element de tipus non-LTR i un transposó de DNA. Del primer van retenir els dominis RT i RNH i del segon n'adquiriren la integrasa. Fins i tot s'ha suggerit que l'element non-LTR ancestral podria ser del grup I ja que és el que presenta tant la ORF1 més similar a *gag* com un domini RNAsaH. Aquest element híbrid va haver de canviar molt per a assolir l'estructura dels retrotransposons LTR actuals ja que va haver d'incorporar entre altres funcions, l'ús de petits RNAs com a encebadors de la transcripció inversa i adquirir les LTR per a permetre i promoure el cicle replicatiu. Finalment, el domini amb activitat proteàsica que permet separar el complex transcriptasa inversa/RNAsaH-integrasa s'hauria incorporat a partir d'un gen amb aquesta activitat d'algun hoste eucariòtic (Craig 2001).

Un segon tipus de retroelements evolutivament i funcionalment relacionats amb els retrotransposons non-LTR són els SINEs. S'anomenen els paràsits dels paràsits, ja que no són elements autònoms on els seus 80 a 500 pb no codifiquen per a cap proteïna que permeti la seva mobilització. L'origen dels SINEs no és clar però la relació amb els retrotransposons non-LTR és evident ja que, com s'ha descrit pel retrotransposó non-LTR UnaL2 i el SINE UnaSINE1 d'*Anguilla japonica* (Kajikawa 2002), ambdós elements comparteixen l'extrem 3', necessari per al reconeixement per part de la RT que el retrotranscriurà.

En resum, tot i ser un model encara poc fonamentat és plausible que els retrotransposons non-LTR s'originessin a partir dels introns de tipus II i que, a la seva vegada, haguessin participat en l'origen dels retrotransposons LTR i en l'expansió dels SINEs en els genomes actuals.

2 TEs com a components dels genomes animals

Els éssers vius es poden agrupar en diferents grups en funció de les característiques que comparteixen i els fan diferents de la resta. El regne animal es divideix en grans grups o fílums (Figura 4) que alhora es subdivideixen en grups més petits fins a arribar a les diferents espècies existents. Malgrat que entre aquests grups hi ha grans diferències a nivell morfològic la seqüenciació de genomes sencers de diversos representants d'aquests grups ha posat de manifest una conservació significativa del complement gènic. La genòmica ha representat una revolució en la manera de generar i interpretar les dades, i l'automatització metodològica ha permès que es desenvolupin més de 2000 projectes genòmics a data d'avui, tant de DNA genòmic com d'ESTs (<http://genomesonline.org/>). D'aquests, més de 600 projectes, corresponen a organismes eucariotes. Entre els projectes eucariotes trobem més de 40 genomes finalitzats dels quals trobem diversos animals com *Anopheles gambiae*, *Apis mellifera*, *Bombyx mori*, *Caenorhabditis briggsae*, *C. elegans*, *Canis familiaris*, *Ciona intestinalis*, *Danio rerio*, *Drosophila melanogaster*, *Homo sapiens*, *Mus musculus*, *Strongylocentrotus purpuratus*, *Takifugu rubripes*, *Tetraodon nigroviridis* i *Rattus norvegicus*. Tot i l'enorme quantitat de projectes iniciats, encara manca acabar i interpretar els resultats d'organismes cabdals de la història animal o d'especial interès per l'home. Els resultats obtinguts en aquests projectes han confirmat que la paradoxa del valor C, que expressa que el nombre de gens no es correlaciona necessàriament amb la mida del genoma, es justifica per la presència diferencial de DNA repetitiu com els elements transposables, ja que representa una fracció molt important dels genomes d'eucariotes superiors.

2.1 Cerca d'elements transposables

La caracterització de TEs ha estat generalment acompanyada per la troballa a l'atzar d'un element determinat. Existeixen, però, mètodes per a la cerca de TEs, tot i que la majoria de metodologies es basen en la caracterització de DNA repetitiu sense diferenciar-ne el tipus. Així, un cop aïllada la fracció de DNA repetitiu, només manca la seva caracterització en la qual molt segurament identificarem TEs. Una segona estratègia per a la cerca de TEs ha estat la seqüenciació de fragments cromosòmics grans a partir de BACs, YACs o còsmids; els TEs són tan abundants que *a priori* s'espera trobar una representació d'aquests en seqüenciar un fragment gran de DNA genòmic. Aquesta tècnica va permetre la caracterització dels 4 primers TEs descrits a *Ciona intestinalis* (Simmen 1999). Aquesta aproximació metodològica té

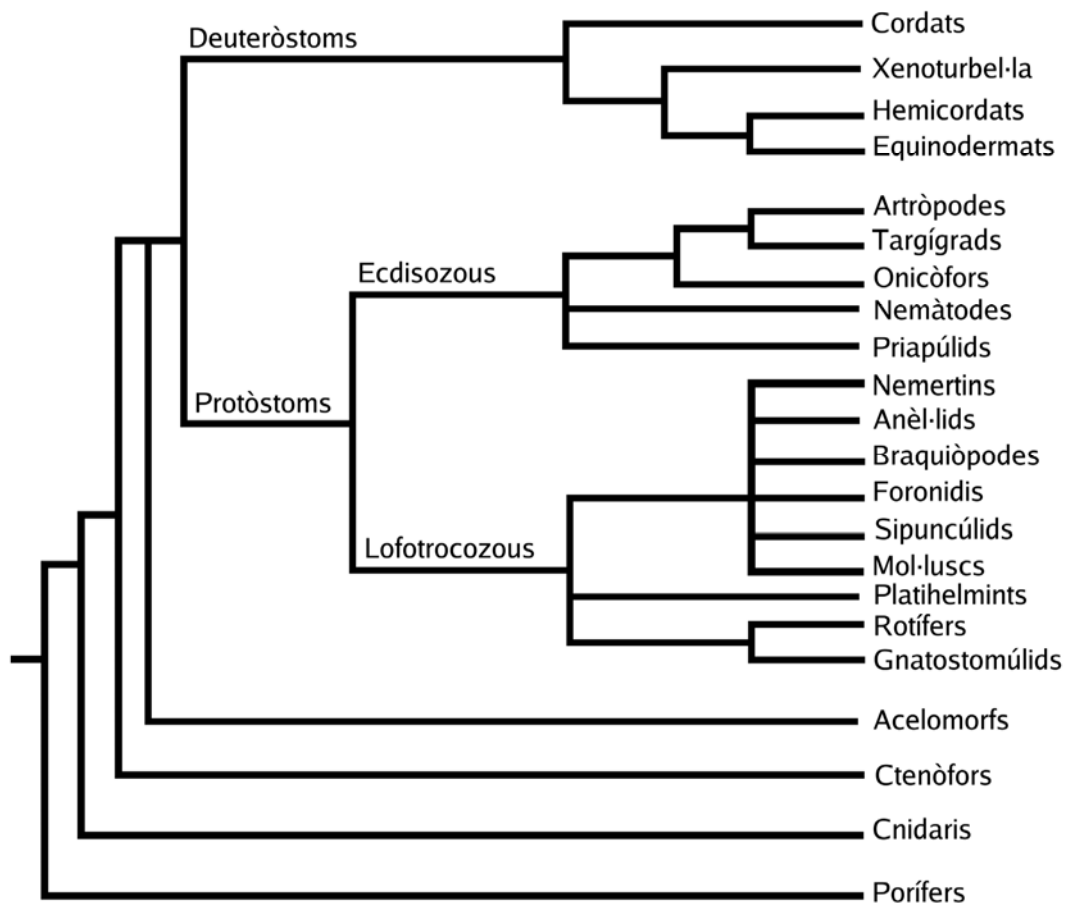


Figura 4: El regne animal s'ha dividit clàssicament en dos grans grups: els diblàstics, aquells organismes que es desenvolupen a partir de dues fulles blastodèrmiques, i els triblàstics, aquells que ho fan a partir de tres. Els triblàstics es classifiquen en protòstoms, organismes en què durant el desenvolupament embrionari el blastòpor origina la boca, i deuteròstoms, en els quals el blastòpor origina l'anús. Els protòstoms agrupen principalment els fílums dels artròpodes, targígrads, onicòfors, nemàtodes, priapulids, nemertins, anèl·lids, braquiòpodes, foronidis, sipuncúlids, mol·luscs, rotífers, platihelminths i gnatostomúlids, mentre que els deuteròstoms agrupen principalment els fílums dels equinodermes, hemicordats i cordats. Adaptat de Halanych 2004.

l'avantatge que permet la caracterització de TEs en el seu entorn genòmic tot i que no garanteix la troballa de cap TE. Altres metodologies més específiques es basen en la cerca per semblança a elements prèviament descrits. Així, es poden cercar elements mitjançant el crivellatge de genoteques amb sondes heteròlogues o dissenyant encebadors degenerats a partir de la comparació d'un domini conservat en diferents còpies. Aquesta darrera tècnica va permetre la caracterització de 2 retroelements a *Anemonia sulcata* (Greenwood 2005). Ara bé, totes aquestes metodologies s'han vist superades per la seqüenciació de genomes. En disposar de la seqüència de tot el genoma d'un organisme, o si més no de la regió eucromàtica, es poden identificar i quantificar tots els tipus coneguts d'elements presents en aquesta fracció del genoma.

2.2 TEs en el genoma dels protòstoms

Els protòstoms són un grup molt nombrós i divers d'organismes entre els quals es troben molts animals models tals com *Drosophila melanogaster*, *Anopheles gambiae*, *Apis mellifera*, *Caenorhabditis elegans*, *Actinia equina* i *Schmidtea mediterranea* entre altres. L'estudi d'aquests models ha permès caracteritzar elements transposables en molts d'ells però manquen estudis a nivell genòmic de la majoria d'organismes característics d'aquests fílums.

Els artròpodes representen un fílum de protòstoms que agrupa la majoria d'animals del planeta. Aquest grup heterogeni d'organismes s'ha modelitzat a partir del gènere *Drosophila* i altres insectes com *A. gambiae*. El genoma haploid de *D. melanogaster* té una mida de 0,18 picograms (pg) que corresponen a unes 180 Mb, 120 Mb de les quals d'eucromatina, i on s'han predit uns 14.000 gens. Amb un contingut GC del 41,1% (Adams 2000) és un genoma poc metilat, detectable únicament als estadis larvaris (Lyko 2000). El genoma d'*A. gambiae* té una mida de 278 Mb (0,27 pg), s'ha predit 13.600 gens i presenta un 35,2% en contingut GC (Holt 2002) amb uns nivells baixos de metilació (Regev 1998).

L'anàlisi del genoma de *D. melanogaster* mostra l'existència de més de 60 tipus d'elements transposables en baix nombre de còpies. Aquests elements representen un 3,6% del genoma eucromàtic, on cap dels tipus descrits supera les 150 còpies per genoma haploid i 2/3 parts són còpies truncades (Kaminker 2002). Els elements amb un major èxit colonitzador del genoma de *D. melanogaster* són els retrotransposons LTR ja que es troben 35 tipus que representen un 2,5% del genoma eucromàtic, un 0,8% correspon a 17 tipus de retrotransposons non-LTR, un 0,3% correspon a 14 tipus d'elements TIR -com els elements *P*- i un 0,04% correspon a un únic tipus d'elements FB (Taula 1). Els elements no autònoms no sembla que hagin prosperat a *D. melanogaster* ja que no s'ha trobat cap còpia d'elements tipus SINE ni MITE. Per altra banda, l'anàlisi del genoma d'*A. gambiae* revela que existeixen més de 40 tipus d'elements transposables. Els més abundants són els retroelements, especialment els retrotransposons LTR i els SINEs però també trobem transposons de DNA com els MITEs. Aquests elements ocupen el 16% del genoma eucromàtic però arriben al 60% de la fracció heterocromàtica (Holt 2002). De fet, en ambdós genomes els elements transposables es troben preferentment a regions heterocromàtiques, principalment a centròmers i telòmers, les quals tenen menor recombinació gènica (Eickbush 2002).

Aquests dos models d'artròpodes poden no ser representatius de tots els protòstoms, ni tan sols dels artròpodes. Ambdós són dípters amb genomes majors que els genomes més petits d'aquest fílum (0,08 pg a l'àcar *Tetranychus urticae*) però relativament petits si els comparem amb alguns crustacis (38 pg, en el gènere *Hymenodora sp*). Malauradament, no s'han fet anàlisis exhaustives d'altres espècies de protòstoms dels quals es disposa del seu genoma com *C.*

elegans o *Apis mellifera* que haguessin donat una visió més àmplia del tipus i freqüència dels TEs en aquest grup d'animals.

2.3 TEs en el genoma de deuteròstoms

2.3.1 TEs en el genoma de deuteròstoms no vertebrats

Els equinoderms s'han modelitzat a partir de l'eriçó de mar *Strongylocentrotus purpuratus*. El genoma d'aquest organisme s'està seqüenciant i en breu es podrà realitzar l'anàlisi global del seu genoma. Existeixen, però, estudis previs a diferents equinoderms on es descriu la presència d'elements de tipus FB (1000 còpies per genoma haploid de l'element Tsp, Cohen 1985), retrotransposons LTR (400 còpies de l'element SURL, Springer 1991), retrotransposons non-LTR (R2Mr, Kojima 2005) i SINEs (800 còpies de l'element SURF-1, Nisson 1988) però manca una visió exhaustiva dels TEs en aquests genomes.

Actualment no disposem de dades sobre elements transposables en els hemicordats, el segon fílum de deuteròstoms no cordats, ni tampoc hi ha cap projecte de seqüenciació genòmica finalitzat. Per tant no podem establir la fracció del genoma d'aquests organismes formada per TEs, ni valorar-ne el seu impacte en l'estructura i complexitat del genoma d'aquests organismes.

El fílum dels cordats es divideix en 3 subfílums, els urocordats o tunicats, els cefalocordats i els vertebrats. Les dades que corresponen a urocordats i cefalocordats són molt minses ja que no existeix cap revisió exhaustiva dels TEs en cap cordat no vertebrat. A l'inici d'aquesta Tesi únicament s'havien descrit elements aïllats tant de la classe I com de la classe II en l'urocordat *Ciona intestinalis* (Simmen 1999) (veure més endavant, apartat 4 d'aquesta introducció, per una descripció detallada d'aquests elements).

2.3.2 TEs en el genoma de vertebrats: àgnats, peixos i tetràpodes no mamífers

Els àgnats, coneguts també com a ciclostomats o vertebrats no mandibulats, són els vertebrats més basals i agrupen a les lamprees i les mixines o *hagfish*. No existeix cap revisió exhaustiva sobre aquests organismes però sí que existeixen treballs sobre elements concrets (veure més endavant, apartat 4 d'aquesta introducció, per una descripció detallada d'aquests elements).

En els grups dels vertebrats mandibulats o gnatostomats, excloent els mamífers, existeixen dades molt parcials sobre els elements transposables ja que únicament s'ha completat l'anàlisi genòmic del DNA repetitiu als tetraodontiformes *T. rubripes* i *T. nigroviridis* i a l'au *Gallus gallus*. Existeixen però projectes genòmics en curs com el de l'amfibi *Xenopus tropicalis* i diversos treballs puntuals que permeten establir un esquema general dels TEs d'aquests organismes.

Els genomes dels tetraodontiformes es consideren com a model dels

vertebrats ja que contenen un conjunt gènic similar però amb una mida molt menor. De fet, presenten els genomes vertebrats més petits que es coneixen (0,35pg, *Chelonodon fluviatilis*; 0,4 pg, *T. rubripes*). El genoma de *T. rubripes* ha estat seqüenciat recentment (Aparicio 2002), té una mida de 390 Mb (Brenner 1993) de les quals s'han seqüenciat 332 Mb, s'han predit de 20000 a 35000 gens (Jaillon 2004) i presenta un 45,2% en contingut GC (Jabbari 2004). El seu genoma allotja més de 20 tipus d'elements transposables que representen entre un 1,7% (Volff 2003) i un 2,7% (Aparicio 2002) del genoma eucromàtic. Ambdós valors són subestimacions ja que no consideren la fracció heterocromàtica però es troben clarament per sota dels valors observats a altres vertebrats. L'anàlisi de les dades genòmiques mostra que els elements majoritaris són els retrotransposons non-LTR ocupant un 1,3% del genoma i els retrotransposons LTR que representen el 0,3%. L'anàlisi del genoma de *T. nigroviridis* mostra un comportament molt similar en quan al tipus d'elements i les proporcions entre ells tot i que el nombre de còpies es menor; el 0,8% del genoma eucromàtic correspon a retrotransposons non-LTR i el 0,1% a retrotransposons LTR (Volff 2003) en un genoma també menor de 0,35 pg (Jaillon 2004). A part de les anàlisis derivades dels projectes de seqüenciació genòmica, s'ha estudiat els TEs en altres espècies de peixos com els retrotransposons non-LTR Rex descrits al gènere *Xiphophorus* però presents en moltes altres espècies en elevats nombres de còpies (revisat a Volff 2001). Tot i tractar-se d'anàlisis no genòmiques, sembla que els peixos no tetraodontiformes presenten més TEs que no pas els tetraodontiformes. Finalment, els peixos pulmonats representen la baula entre els vertebrats marins i els terrestres. Són els vertebrats amb els genomes més grans (132,8 pg a *Protopterus aethiopicus*) com a conseqüència de fenòmens de poliploidització. La mida enorme dels seus genomes suggereix la presència abundant d'elements transposables però només es coneix la presència d'un retrotransposó non-LTR a *Neoceratodus forsteri* que es troba present amb un moderadament elevat nombre de còpies (17500 còpies de l'element NfCR1, Sirijovski 2005).

Les dades sobre genomes d'amfibis són també escasses. Aquests organismes presenten alguns dels genomes més grans dels vertebrats degut a fenòmens de poliploidització. La mida enorme dels genomes d'aquests organismes permet dir *a priori* que la colonització de TEs serà important. Trobem evidències en el genoma de les salamandres del gènere *Hydromantes* on s'han caracteritzat retrotransposons LTR amb un elevat nombre de còpies (Marracci 1996). Al genoma dels gripaus del gènere *Xenopus* s'ha trobat la presència de diferents transposons de DNA que representen almenys un 1% del genoma haploid (Sinzelle 2005) i diversos elements de tipus MITE (Garrett 1989).

Les dades de rèptils són encara més escasses que en el cas dels amfibis ja que no hi ha cap organisme model d'aquest grup. Existeixen dades aïllades de diferents elements a diferents rèptils, els MITEs *Chompy* de cocodrils (Ray

2005), retrotransposons non-LTR com *Bov-b* a serps del gènere *Vipera* (Kordis 1998) o PsCR1 a la tortuga *Platemys spixii* (Kajikawa 1997). Malauradament, en tots els estudis s'analitza un element en concret i per tant no existeix la visió global del genoma tot i que com els elements descrits es troben presents amb un nombre de còpies moderadament alt i ja que els genomes d'aquests organismes no són petits, podem pensar que la fracció de DNA repetitiu d'aquests és significativa.

Les aus s'han modelitzat a partir del genoma de *G. gallus*. El genoma del pollastre és tres vegades més petit que l'humà (1,25 pg) probablement com a conseqüència de la pèrdua de DNA repetitiu ja que el nombre i la mida de gens és similar a la trobada en el nostre genoma (Hillier 2004). La fracció corresponent al DNA repetitiu representa un 9% del genoma i conté elements transposables de classe II (0,8% del genoma), retrotransposons LTR (1,3%) i retrotransposons non-LTR (6,4% amb CR1 com a únic tipus d'element). Cal destacar l'absència d'elements de tipus SINE tot i la presència d'un retrotransposó non-LTR que en altres organismes sí que presenta SINEs associats (Kajikawa 2002).

2.3.3 TEs en el genoma de vertebrats: mamífers

Existeixen moltes dades referents a genomes de mamífers tant a nivell puntual com a nivell genòmic. El genoma humà és el més ben caracteritzat a tots els nivells ($3,4 \cdot 10^9$ pb o 3,5 pg, (Lander 2001) i és molt similar al del ratolí ($2,8 \cdot 10^9$ pb o 2,9 pg, Waterston 2002), el de rata ($2,9 \cdot 10^9$ pb o 3 pg, Gibbs 2004) i el de gos ($3,1 \cdot 10^9$ pb o 3,2 pg, Lindblad-Toh 2005). Tots ells es caracteritzen per la massiva presència de DNA repetitiu en forma de repeticions disperses. Aquestes repeticions són de 4 tipus diferents; retrotransposons non-LTR, SINEs, retrotransposons LTR, i transposons de DNA que constitueixen una càrrega molt important per al genoma d'aquests organismes (44,83% a l'home, 38,55% al ratolí, 40,31% a la rata i 35,15% al gos). Els retrotransposons LTR i els transposons de DNA representen una part significativa d'aquests genomes, però els retrotransposons non-LTR i els SINEs són els principals elements transposables dels genomes dels mamífers. Els retrotransposons non-LTR caracteritzats són principalment del tipus L1, tot i trobar elements L2 i L3/CR1, amb múltiples llinatges i són la principal càrrega del genoma haploid, aproximadament un 20% amb milers i milers de còpies (868000 còpies que ocupen el 20,42% en l'home, 660000 còpies i el 19,2% en el ratolí, 657000 i el 23,11% en la rata i 1065000 i el 18,74% en el gos). La majoria d'aquestes còpies són truncades i no funcionals; a l'home s'estima que menys d'un centenar de còpies són actives (Kazazian 2004). Els SINEs han guanyat la partida als genomes dels mamífers ja que tot i ser elements petits no autònoms representen el 10% del genoma de mamífers amb més d'un milió de còpies per genoma haploid (1558000 còpies que ocupen el 13,14% en l'home, 1498000 i el 8,22% en el ratolí, 1360000 i el 7,05% en la rata i 1505000 i el 10,57% en el gos).

Els retrotransposons LTR de mamífers representen menys del 10% del genoma haploid amb milers de còpies (443000 còpies que ocupen un 8,29% en l'home, 631000 i el 9,87% en el ratolí, 556000 i el 9,04% en la rata i 304000 i el 3,68% en el gos). Els transposons de DNA no han tingut gaire èxit colonitzador en els genomes dels mamífers. Representen menys d'un 3% del genoma amb milers de còpies (294000 còpies que ocupen un 2,84% en l'home, 112000 i el 0,88% en el ratolí, 108000 i el 0,81% en la rata i 246000 i el 1,98% en el gos).

Introducció

organisme		mida (Mb/pg)	ORFs	seq. rep. (%)	LTR (%)	non-LTR (%)	SINE (%)	DNA (%)	Refs
<i>E. histolytica</i>	protozous	237/-	9938	6	0	6		0	Van Dellen 2002
<i>A. gambiae</i>	artròpode	278/0,27	13600	>16	-	-	-	-	Holt 2002
<i>D. melanogaster</i>	artròpode	137/0,12	14000	3,6	2,5	0,8	0	0,34	Adams 2000
<i>B. mori</i>	artròpode	530/0,53	18500	21,1	11,1	6,7	1,4	0	Mita 2004
<i>C. intestinalis</i>	urocordat	117/0,2	16000	30-35	-	-	-	-	Dehal 2002
<i>F. rubripes</i>	actinopterigi	400/0,4	20-35000	1,7-2,7	0,3	1,3	-	-	Aparicio 2002
<i>T. nigroviridis</i>	actinopterigi	385/0,38	28000	0,9	0,12	0,77	-	-	Volff 2003
<i>Gallus gallus</i>	Au	1200/1,25	20-23000	15	0,6	3,1	-	0,57	Hillier 2004
<i>M. musculus</i>	mamífer	2855/2,92	24174	38,5	9,9	19,2	8,2	0,9	Waterston 2002
<i>R. norvegicus</i>	mamífer	2915/2,98	21166	40,3	9	23,1	7	0,8	Gibbs 2004
<i>C. familiaris</i>	mamífer	3130/3,2	<22000	35,1	3,7	18,7	10,6	2	Lindblad-Toh 2005
<i>H. sapiens</i>	mamífer	3430/3,5	26966	44,4	8,1	20,4	13,1	2,8	Lander 2001

Taula 1: Comparació dels estudis genòmics realitzats en diferents organismes. La mida del genoma en pg es determina a partir de diferents aproximacions experimentals (www.genomesize.com). ORFs: nombre de gens predits, seq. rep.: fracció de DNA repetitu, LTR: fracció corresponent a retrotransposons LTR, non-LTR: fracció corresponent a retrotransposons non-LTR, SINE: fracció corresponent a SINEs i DNA: fracció corresponent a transposons de DNA. “-” indica que no s'ha determinat la mesura.

3 Contribució dels elements transposables en els genomes

3.1 Conseqüències de la presència de TEs

La càrrega addicional que suporten els genomes deguda a l'acumulació massiva de DNA repetitiu semblaria *a priori* deletèria, pròpia d'elements paràsits, però l'abundant presència d'aquests elements en els genomes eucariotes i l'evident domesticació que s'observa en alguns casos indiquen que la relació que s'estableix entre l'hoste -el *genoma*- i el paràsit -l'*element*- sembla tendir cap al mutualisme. Així els genomes "funcionen" més lentament però tenen més recursos a l'hora d'enfrontar-se a noves necessitats gèniques o genòmiques (Vinogradov 2003). Les mutacions són fonamentals per a assolir els canvis evolutius i això planteja la pregunta: les mutacions causades pels elements transposables són deletèries per a l'hoste? *A priori*, la presència de qualsevol seqüència en un genoma es justifica suposant que la seva existència augmenta la *fitness* de l'organisme i per tant la resposta a la pregunta no és ni senzilla ni òbvia.

Els TEs interaccionen amb el seu entorn genòmic i n'augmenten el potencial evolutiu mitjançant la generació de punts calents de recombinació, aportant el mecanisme de *exon/genomic shuffling* i una font de motius aptes per a ser utilitzats com a elements reguladors transcripcionals, senyals de poliadenilació i dominis proteics entre altres (Makalowski 2000) (Figura 5).

La presència de múltiples còpies del mateix element fa dels TEs substrats ideals per a la recombinació homòloga i no homòloga. A *S. cerevisiae* s'ha observat que en els punts de trencament produïts durant les delecions, duplicacions, inversions i translocacions s'hi troben elements Ty (Roeder 1983). Un altre exemple del paper que juguen els TEs en la recombinació és l'elevada densitat de seqüències Alu que es troben a les fronteres de regions duplicades del genoma humà (Bailey 2003). La recombinació també pot modificar els mateixos TEs ja que es pot produir entre diferents còpies d'un TE, promovent fenòmens de conversió gènica a partir dels quals es poden reparar elements truncats o mutats o fins i tot generar noves variants o el cas contrari, escapar o delectonar els elements (Roy-Engel 2002).

Els TEs posseeixen una sèrie d'elements reguladors transcripcionals que poden conferir nous patrons espacials o temporals d'expressió als gens hostes flanquejants. Alguns d'aquests canvis han estat adaptatius i s'han conservat a plantes i animals (revisat a McDonald 1995). L'evidència més clara que suporta la hipòtesi de la participació dels TEs en l'adquisició de seqüències reguladores en *cis* prové dels casos de diferències fixades evolutivament en el patró d'expressió de gens causades per la inserció de TEs. Un exemple clar és l'expressió dels gens similars a les beta-globines a cèl·lules eritroides humanes que es troben regulades pel LCR (*Locus Control Region*) (Long 1998). A l'extrem 5' d'aquest es van trobar 982 pb d'un LTR del retrovirus ERV-9 que contenia repeticions en tàndem a la regió activadora U3 que actuaven com a *enhancer*

específic de teixit, tant a humans com a goril·les. També queda clar el paper dels TEs en altres promotors, *enhancers* i *silencers* (Norris 1995, Vansant 1995 i Zhou 2002). Finalment, cal destacar que l'anàlisi de les regions promotores del genoma humà va mostrar que almenys el 24% dels promotors i un 2.5% dels elements reguladors en *cis* contenen seqüències derivades de transposons (Lander 2001 i van de Lagemaat 2003).

Els TEs han participat en l'evolució gènica no només aportant seqüències reguladores a gens ja existents sinó que també han participat creant noves funcions cel·lulars a partir de gens quimeres derivats de TEs. Aquest fenomen, conegut com a "domesticació molecular" (Miller 1992) és evident a: el manteniment dels telòmers de *Drosophila* pels retrotransposons Het-A i TART

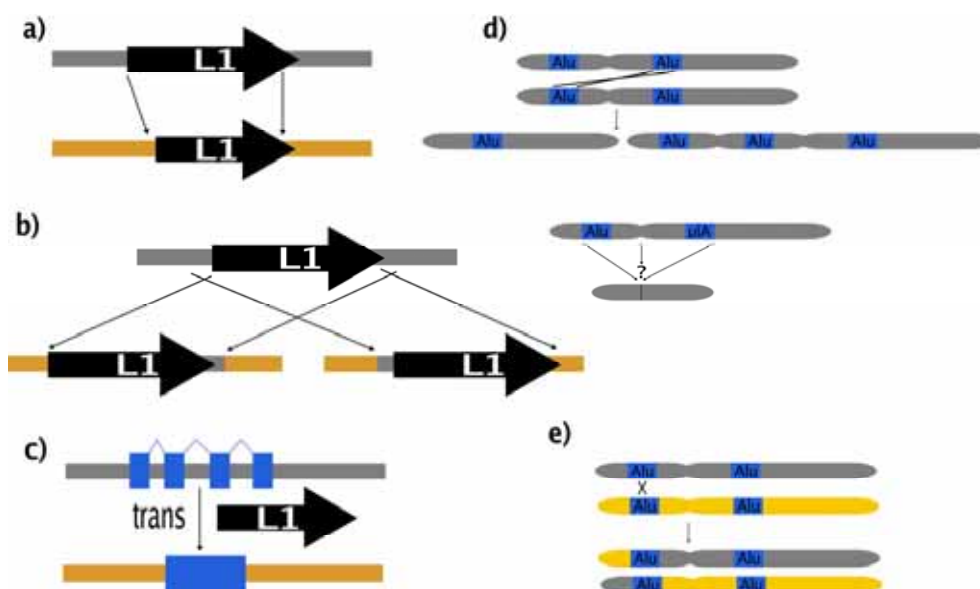


Figura 5: Representació d'alguns dels possibles impactes genòmics dels elements transposables. a) transposició d'un LINE generant una nova còpia truncada per 5' en una nova localització genòmica. Aquesta nova còpia no és funcional però permet els fenòmens de recombinació a part d'alterar l'entorn genòmic on s'insereix. b) transducció 5' i 3' o exon shuffling. c) l'acció en trans de l'activitat RT permet la generació i inserció de pseudogens processats. d) fenòmens de recombinació homòloga no balancejada. e) fenòmens de recombinació homòloga balancejada. Adaptat de Deininger 2003.

(Traverse 1988), el gen murí *Fv1* que és un gen *gag* sota el control cel·lular, associat a la immunitat contra el virus de la leucèmia murina (Best 1996), i els gens RAG1 i RAG2 que provenen d'un transposó, essencials en les recombinacions somàtiques V(D)J per a la generació d'anticossos i receptors de cèl·lules T del sistema immunitari humà (Agrawal 1998). Un segon mecanisme de generació de noves proteïnes és mitjançant empalmaments alternatius. Els elements Alu contenen en la seva seqüència consens 8 donadors d'empalmament i 3 acceptors d'empalmament putatius (Makalowski 1994 i Nekrutenko 2001). Per tant, la inserció de TEs pot ser una de les causes de l'elevada freqüència d'empalmaments alternatius en els gens humans (Brett 2000 i Krehling 2004). Finalment, els TEs participen en la generació de retropseudogens, ja que

aquests poden insertar-se sota el domini d'un promotor i ser així gens actius. Un exemple d'aquest fenomen el trobem en el gen *jgw* de *Drosophila yakuba* i *D. teissieri*, gen actiu de funció poc clara format per 4 exons, 3 de petits i un de gros. Aquest darrer corresponia al mRNA del gen *adh* retrotranscrit mentre que els 3 exons a 5' varen ser reclutats a partir del gen *ymp* amb posterioritat a la inserció del pseudogen (Long 1999).

La capacitat de mobilització que tenen els TEs els converteixen en potents agents mutàgens insercionals ja sigui disrupcionant exons o bé insertant-se en introns i provocant així empalmament incorrecte (Ostertag 2003). Al genoma humà s'han descrit molts casos on la inserció d'un element transposable s'associa a una malaltia com per exemple en el cas de l'hemofília per alteració dels factors VIII i IX o la distròfia muscular de Duchenne per la inserció de TEs (Li 2001) revisat a (Deininger 2003). A més, els senyals de poliadenilació són fàcilment generables a partir de mutacions puntuals de la cua poliA de determinats TEs. Així, la inserció de TEs a 3' d'un gen pot generar noves senyals de poliadenilació. Aquest és el cas del gen de la timidilat sintasa murina on la inserció d'un element L1 activa un senyal de poliadenilació críptica que afegeix una cua de poliA no habitual a nivell del codó d'atur del mRNA (Harendza 1990). Al ratolí també observem la formació de trànscrips anòmals per la inserció d'elements B2 en el gen *gamma*-fosforilasa quinasa (Makalowski 2000).

Finalment, no cal oblidar el fenomen d'*exon shuffling* produït per transducció a 5' o 3' d'un TE. En el cicle replicatiu de l'element L1 es demostra que, com a conseqüència de l'alteració del reconeixement probablement per la RNA polIII dels senyals que promouen tant la iniciació com la terminació de la transcripció, es poden incorporar seqüències que es troben a 3' o 5' d'un element L1 (Moran 1999 i Pickeral 2000). Malgrat que encara hi ha molts interrogants per resoldre, aquests estudis mostren com la co-mobilització és molt eficient ja que durant el procés de transposició es poden incorporar varies Kb de DNA forani a l'element. S'estima que l'1% del genoma humà podria haver estat barrejat per la transducció produïda pels elements L1 (Goodier 2000).

Per tant, tot i que els TEs són potencialment deleteris, la seva presència en situacions determinades i sota mecanismes de control que en minimitzin els efectes perniciosos pot ser beneficiosa al augmentar el potencial evolutiu del genoma.

3.2 Mecanismes de control

L'elevada variació en la càrrega de TEs observada en diferents organismes reflecteix la tolerància dels genomes vers la invasió dels TEs, així com la diferent capacitat de colonització dels elements i el grau de desenvolupament en la maquinària de defensa utilitzada per diferents organismes vers aquests elements. La naturalesa d'aquesta maquinària és font de controvèrsia. Inicialment es va proposar que a plantes i vertebrats els TEs podrien estar

transcripcionalment silenciats gràcies als mecanismes generals de metilació de les citosines en els genomes d'aquests organismes (Bennetzen 1994, Bird 1995 i Yoder 1997) però aquest mecanisme no explica les observacions fetes en genomes com el de *Ciona intestinalis*, on els TEs descrits no es troben metilats tot i presentar metilació gènica (Simmen 1999), o el de molts protozous que tot i tenir genomes enormes no presenten aquest mecanisme (Regev 1998). Recentment s'ha postulat que sistemes de regulació que participen en el control de l'expressió gènica en casos de redundància podrien també jugar un paper clau en el control d'aquests elements. Així, a eucariotes inferiors com *Neurospora crassa* i *Ascobolus immersus* s'ha descrit un sistema que reconeix aquestes seqüències duplicades i les metila. Com a conseqüència de la desaminació de les citosines metilades, la seqüència duplicada incorpora moltes mutacions (Rossignol 1994 i Selker 1997). Els canvis generats eviten la recombinació homòloga entre les seqüències duplicades alhora que suprimeixen la transcripció de la còpia mutada. A més, independentment de l'origen i de la seva evolució, a protistes, animals i plantes s'han descrit fenòmens de bloqueig a nivell transcripcional o postranscripcional específic de seqüències duplicades tot i que encara no està clar el seu funcionament (Wu-Scharf 2000, Bingham 1997 i Kooter 1999). En aquest mecanisme hi participen membres del sistema d'RNAi i de silenciament de la cromatina. Sembla doncs una hipòtesi plausible que la màquinaria que controla l'expressió gènica i evita les recombinacions il·legítimes de seqüències duplicades, ambdós processos cabdals en l'evolució de genomes complexos per endoduplicació, pugui alhora controlar la proliferació dels TEs en els genomes.

Finalment, hem de recordar que en l'evolució dels genomes de molts organismes s'han produït fenòmens de poliploidització que poden tamponar els efectes deleteris de les insercions, delecions, duplicacions, inversions i els rearranjaments genòmics induïts pels elements transposables i el seu cicle replicatiu. Per tant, els genomes duplicats podrien disposar dels recursos genètics necessaris per suportar els efectes derivats de la col·lonització dels TEs abans d'assolir els mecanismes de control necessaris per a la seva regulació.

4 Els cordats en l'escala evolutiva

Tot i que a grans trets els TEs es poden considerar com a paràsits genòmics que colonitzen el genoma i hi generen efectes neutres o deleteris, algunes de les mutacions i canvis genòmics que provoquen participen clarament en l'evolució dels hostes. En aquest sentit, aquests elements es podrien anomenar paràsits útils. És obvi que aquests elements no són l'únic factor que participa en l'evolució genòmica, però sembla que poden estar implicats en l'origen de fenòmens evolutivament importants com la reestructuració cromosòmica, creació de nous gens o noves funcions (Capy 2000). Per tant, els elements transposables poden considerar-se com a acceleradors evolutius, especialment quan els genomes es troben en una situació d'estrès poblacional o ambiental. Es conegut que una situació destacada per a l'aparició sobtada de noves estructures i funcions es va produir dins del fílum dels cordats amb l'aparició dels vertebrats. El fílum dels cordats es subdivideix en 3 subfílums, els urocordats, els cefalocordats i els vertebrats que comparteixen una sèrie de característiques, que els són exclusives i que permeten definir-los en el fílum: presència en almenys algun dels estadis del desenvolupament de notocorda, cordó nerviós dorsal tubular buit, fenedures faríngies i cua postanal. Els ascidis com *Ciona intestinalis*, del subfílum dels urocordats, presenten totes les característiques dels cordats en l'estadi larvari, però després de la metamorfosi únicament conserven les fenedures faríngies (Corbo 2001). Els amfioxos com *Branchiostoma floridae*, del subfílum dels cefalocordats, presenten totes les característiques del fílum abans i després de la metamorfosi, també presenten altres característiques presents en els vertebrats com la presència de ronyons i musculatura paraxial segmentada (revisat a (Holland 2004). Urocordats i cefalocordats presenten genomes relativament petits, entre 0,07 pg d'*Oikopleura dioica* i els 0,59 pg de *Branchiostoma lanceolatum*.

El subfílum dels vertebrats es subdivideix alhora en les superclasses dels àgnats o vertebrats no mandibulats i gnatostomats o vertebrats mandibulats. Els àgnats, també anomenats ciclòstoms, es troben representats per les lamprees i els *hagfish* o mixines, no tenen una mandíbula derivada dels arcs brànquials i presenten com a sinapomorfies l'absència del tercer canal semicircular de la oïda interna, notocorda llarga, dents calloses i absència d'apèndixs parells entre altres (revisat a Takezaki 2003). La mida del genoma d'aquests organismes és clarament superior als cordats no vertebrats, des de 1,42 pg per a la lamprea *Lampetra fluviatilis* a 4,59 pg per al hagfish *Myxine gamani*. La superclasse dels gnatostomats agrupa a la majoria dels vertebrats, des de peixos fins a mamífers i es caracteritza per presentar, a part dels trets dels cordats, sinapomorfies pròpies com la presència de mandíbula o una parella d'extremitats pèlviques.

Històricament, s'havia considerat que els cefalocordats eren el grup d'organismes vius més proper als vertebrats, formant així els *euchordata* (Figura

6a), i que els urocordats representaven el primer llinatge de cordats. Aquesta relació es fonamentava tant en similituds morfològiques com en l'aparent major complexitat de cefalocordats i vertebrats en relació a la major simplicitat dels urocordats (Schaeffer 1987). Ara bé els estudis filogenètics no eren concloents i també es va plantejar la existència del clade *olfactores* que agrupa vertebrats i urocordats. Recentment, gràcies a la publicació dels genomes de diversos organismes d'aquests llinatges, s'ha fet un estudi filogenètic de 146 gens nuclears que recolza el clade dels *olfactores* (Delsuc 2006). A més, aquesta anàlisi suggereix que els cordats podrien no ser un grup monofilètic ja que els cefalocordats s'agruparien amb els equinoderms (Figura 6b). Aquesta nova visió implica que l'ancestre de tots els cordats era més complex del que es creia i que els cefalocordats haurien retingut els caràcters ancestrals alhora que els urocordats els haurien perdut. Per tant, els cefalocordats serien més ancestrals que no pas evolucionats i els urocordats serien més derivats que no pas primitius.

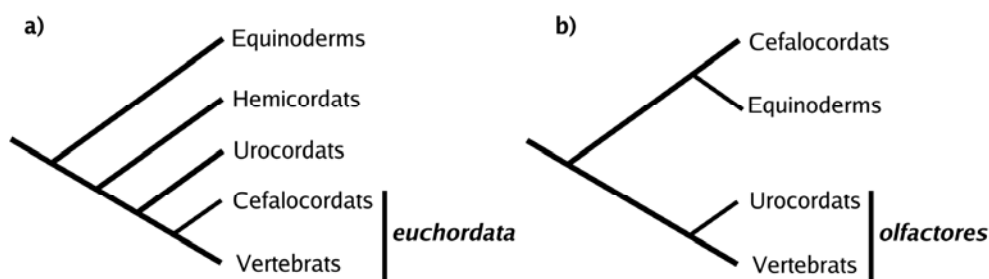


Figura 6: Relació entre els diferents llinatges dels cordats on es comparen les agrupacions proposades dels *euchordata* (a) vs els *olfactores* (b). Adaptat de Delsuc 2006.

El llinatge dels vertebrats es caracteritza per un increment en el nombre de gens i per innovacions importants en el desenvolupament. Al 1970, Ohno proposà que les duplicacions genòmiques podien ser la font de material necessària per incrementar la complexitat dels vertebrats (Ohno 1970) i suggerí que en el llinatge dels vertebrats s'haurien produït dues rondes de duplicacions del genoma, seguides per processos de pèrdues d'alguns gens i de diferenciació funcional d'alguns altres. Les dades moleculars mostren com moltes famílies gèniques tenen un únic representant als cordats no vertebrats però més d'un representant en diverses espècies de vertebrats, donant així suport a la hipòtesi d'Ohno (Holland 1994, Canestro 2000). Encara avui, tot i acceptar que aquestes duplicacions s'han produït, existeix molta controvèrsia sobre l'abast de les duplicacions i el moment d'aquestes (revisat a Furlong 2002). La hipòtesi més acceptada proposa una primera duplicació (total o parcial) a l'origen del llinatge dels vertebrats i una segona dins del grup dels vertebrats mandibulats després de la separació dels àgnats (Skrabanek 1998). També cal destacar que s'han produït altres fenòmens duplicatius dins d'alguns llinatges de vertebrats com la duplicació (total o parcial) en el llinatge dels actinopterigis o les dues

(totals o parcials) en el llinatge dels àgnats (Lundin 1993, Holland 1996). En tot aquest procés, els llinatges dels ascidis i amfioxos, independentment de la relació filogenètica dins els cordats, haurien divergit abans de les duplicacions gèniques produïdes durant l'evolució d'aquest fílum. Per tant, es considera que els estudis estructurals i funcionals en aquests organismes es veuen beneficiats per l'absència de redundància gènica dels vertebrats. Això converteix els ascidis i els amfioxos en models ideals per als estudis evolutius de genòmica comparada i per poder escatir l'impacte dels TEs en la transició des dels genomes pre-duplicatius dels invertebrats als dels vertebrats (Simmen 1998). Dins dels vertebrats, els *hagfishs*, probablement són els representants més primitius que existeixen avui en dia i per tant imprescindibles en aquests estudis comparatius. Amb un genoma altament duplicat, la seva anàlisi permet valorar l'evolució de les duplicacions genòmiques i els processos de colonització dels TEs en els genomes dels vertebrats més basals.

El coneixement que tenim dels elements transposables a la base dels cordats, urocordats, cefalocordats i àgnats és molt pobre. En l'inici d'aquest projecte, les úniques dades de què es disposava d'aquests organismes es limitaven als retrotransposons non-LTR *Cili-1* i *Cili-2*, el retrotransposó LTR *Cigr-1* i els SINEs *Cimi-1* i *Cics-1* a *Ciona intestinalis* (Simmen 1999 i Simmen 2000) i un transposó *Tes1* de tipus *Mariner* a l'àgnat *Eptatretus stouti* (Heierhorst 1992). Els elements descrits al urocordat *C. intestinalis* varen ser trobats en seqüenciar 3 còsmids d'aquest organisme i es caracteritzen per la presència en baixos nombres de còpies (50 còpies de l'element *Cili-1*, 50 de l'element *Cili-2*, 75 de l'element *Cigr-1*, 17000 de l'element *Cimi-1* i 40000 de l'element *Cics-1*) i per trobar-se en la fracció no metilada del genoma. L'element *Tes1* va ser descrit en un intró del gen de la vasotocina; tot i no saber el nombre de còpies d'aquest element, la presència de l'element *Tes1* en múltiples gens expressats a cervell d'*Eptatretus stouti* suggereix que pot ser abundant en el genoma d'aquest animal.

Durant la realització d'aquest projecte de Tesi s'han caracteritzat diferents elements en els genomes de l'ascidi *Ciona intestinalis*, el larvací *Oikopleura dioica*, i els amfioxos *Branchiostoma floridae* i *Branchiostoma lanceolatum* i la mixina *Eptatretus burgeri*. A més, el recent projecte genoma de *Ciona intestinalis* ha permès establir les característiques genòmiques d'aquest cordat invertebrat. La seqüenciació ha determinat la mida del genoma ($1.53\text{-}1.59\cdot 10^9\text{pb}$), la composició nucleotídica (65% AT), el contingut gènic (15852 prediccions gèniques) i la densitat gènica (1 gen/7.5 Kb). Aquestes característiques han permès definir l'organització del genoma de *Ciona* com a mig camí entre la de protòstoms i la de vertebrats. No obstant, en l'estudi inicial que es va fer del genoma de *C. intestinalis* (Dehal 2002) no es va analitzar en profunditat el DNA repetitiu, el qual es va estimar que representava el 30-35% del genoma. Recentment, la cerca d'*old* LINEs gràcies a la inserció d'aquests en dianes específiques ha permès la caracterització dels

elements L1Ci, R2Ci i YURECi a *Ciona intestinalis* (Kojima 2004) i R2Eb a *Eptatretus Burgeri* (Kojima 2005). La seqüenciació del reduït genoma del larvací *Oikopleura dioica* ha permès establir que aquest organisme presenta un únic tipus de retrotransposó non-LTR, anomenat *Odin*, diferent als 14 clades descrits prèviament i 4 retrotransposons LTR similars als elements de tipus *gypsy* però molt divergents, *Tor-1*, *Tor-2*, *Tor-3* i *Tor-4*, tots ells en baixos nombre de còpies (Volff 2004). L'esforç que ha fet la comunitat científica per a caracteritzar el genoma de *Branchiostoma floridae* ha permès la caracterització de TEs en el seu genoma. Un element no autònom de classe II anomenat ATE-1 va ser identificat a les espècies *B. floridae* i *B. lanceolatum*; la seva estructura suggereix que podria estar format a partir de la fusió de dos MITEs, i en ambdues espècies és present en baixos nombres de còpia (Canestro 2003). Recentment, s'ha determinat la presència dels elements *LanceleTn* de tipus MITE que haurien participat en la generació d'elements com l'ATE-1 (Osborne 2006). La caracterització d'aquests elements ha permès establir que es troben en un moderat nombre de còpies, de 2200 a 4800 en funció del tipus d'element. Finalment, a la regió reguladora del gen *FoxD* de l'amfiox s'ha trobat el retroelement no autònom *Amphi-Alu*, el primer SINE de tipus *Alu* descrit fora de primats (Holland 2006). Destaca el reduït nombre de còpies en que es troba, 50-100, en comparació als seus homòlegs de primats.

Tot i que el genoma de *Ciona intestinalis* i *Oikopleura dioica* ja s'han seqüenciat i que els cefalocordats i els àgnats tenen projectes genoma en curs que permetran la caracterització de la càrrega mòbil, a data d'avui no es té una visió global dels TEs, o almenys dels retrotransposons non-LTR, en aquests organismes. Per tant, la caracterització dels en aquests organismes és cabdal per a entendre els canvis evolutius que es van produir a nivell genòmic durant la transició des dels cordats més primitius fins als vertebrats actuals.