



Genètica de la Conservació,  
Institut de Recerca i Tecnologia Agroalimentàries



Dept. de Ciències Experimentals i de la Salut,  
Universitat Pompeu Fabra



Dept. de Genètica i de Microbiologia,  
Universitat Autònoma de Barcelona

***Origen, evolució i aplicacions en genètica  
de la conservació de DNA repetitiu en  
carnívors***

Francesc López Giráldez

*Barcelona, 19 de maig de 2006*

Memòria presentada per Francesc López Giráldez per optar al grau de Doctor per la Universitat Autònoma de Barcelona i amb el programa de doctorat en Ciències de la Salut i de la Vida (Universitat Pompeu Fabra, UPF).

Aquest treball ha estat realitzat en el Departament de Ciències Experimentals i de la Salut (UPF) i en el laboratori de Genètica de la Conservació (Institut de Recerca i Tecnologia Agroalimentàries, IRTA) sota la co-direcció del Dr. Xavier Domingo i Roura (IRTA), malauradament traspassat, i del Dr. Alfredo Ruíz Panadero (Universitat Autònoma de Barcelona, UAB).

Dr. Alfredo Ruíz Panadero  
Co-director de la Tesi

Francesc López Giráldez

*A en Xavier*

*A la meva àvia*

*Als meus pares i les meves germanes*

## AGRAÏMENTS

Els agraïments segurament és l'apartat més llegit de les tesis. Per tant, en aquestes línies intentaré no deixar-me ningú per no quedar malament. Així doncs, comencem pel principi de tot... Gràcies a l'espermatozoide, per ser tu i no un altre, el que fecundés l'òvul dins del ventre de la meva mare, gràcies als metges que van ajudar la meva mare a parir un 27 de març de fa ja uns anys, gràcies a la professora de la guarderia que em va començar a ensenyar el que era ser un ésser més o menys civilitzat, gràcies a aquells companys de l'hora del pati on jugàvem a fer curses amb pneumàtics... Encara estàs llegint? Veig que segueixes amb ganes de trobar-te entre aquestes línies.

Bé, en tot cas, no sé si hauríem d'anomenar aquest apartat "Agraïments" ja que si només fossin això no ser si valdria la pena escriure'ls. En aquestes línies hi ha una barreja d'agraïments, nostàlgia, patiment, desitjos,.... difícils d'expressar però que molts de vosaltres ja sabreu interpretar.

Segurament els hauria de començar amb el meu primer contacte amb el món de la recerca. Aquest es va produir a Girona al laboratori d'Ictiologia Genètica quan tot just acabava segon de carrera. Em va costar una mica atrevir-me a demanar si volien algun estudiant en pràctiques per la ben coneguda falta de carisma del Dr. José Luís. Doncs a ell li agraeixo que em permetés introduir-me en aquest món i als companys d'aquest laboratori, especialment en Martí Cortey i en Jordi Vinyas, els agraeixo tot el que van fer per ensenyar-me les tècniques de genètica molecular i alguna cosa més. Poc després vaig realitzar les Pràctiques en Empreses en el mateix laboratori molt ben dirigides també pel Dr. Jose Luís i una Beca de Col·laboració dirigida per la Dra. Maria Inés Roldán, però en aquesta última etapa no sóc jo qui hauria d'escriure agraïments.

Al Novembre del 2000 en Xavier em va proposar incorporar-me a la Unitat de Biologia Evolutiva de la Universitat Pompeu Fabra. A ell és a qui més haig d'estar agraït per haver-me permès dur a terme aquesta tesi

doctoral, que poc té a veure amb el projecte inicial, però ell sempre va confiar en mi i en les meves capacitats per tirar-la endavant. També li haig d'agrair les oportunitats de treballar en laboratoris punters en genètica de la conservació d'Estats Units i les experiències personals viscudes junts al llarg de la tesi (que en són moltes). Malauradament, en Xavier no ha pogut veure el resultat final de tota la tesi, però crec que n'estaria orgullós.

Durant gairebé 4 anys de tesi a la Pompeu Fabra la llista de persones que haurien d'aparèixer en aquest apartat és enorme i difícil d'oblidar. Aquí hauria de començar per en Guerau i en Marmota, que són els que ja estaven amb en Xavier i els que van introduir-me en el funcionament del laboratori de BioEvo i d'algunes tècniques "noves". En Guerau ens va deixar aviat, però conservo bons records dels moments viscuts junts. En Josep ha compartit moltes hores de feina al laboratori i, com bons becaris, aquestes també formen part de la nostra vida social. Durant aquest temps hem patit i gaudit moltes coses de les quals he après molt, com per exemple que difícilment les idees surten d'un sol individu. El grup de Genètica de la Conservació no va parar de créixer amb l'arribada d'altres becaris i postdocs als quals les meves paraules difícilment podran agrair tot el que hem compartit. La Lluïsa, amb qui, a part de feina, he compartit moltes xerrades molt freakies; l'Olga, sempre somrient i alegre, excepte algun cop que he portat la seva paciència al límit; la Montse, sempre tan eficient i amb qui hem seguit traient la tesi endavant; en Bruno, que m'ha ensenyat una petita part del seu art del *savoir faire* i la diplomàcia; la Roby, l'última a incorporar-se, però que ràpidament s'ha adaptat a "la vida" de Barcelona, i que sempre ha trobat algun moment per corregir-me l'anglès i compartir bons moments fora del laboratori. A més, la Yolanda, l'última incorporació un cop ja estàvem a l'IRTA, potser la que menys m'ha patit.

Al laboratori de BioEvo no estàvem sols. He conegut gent molt interessant que m'ha ofert la seva amistat i ajuda desinteressada. Des del principi he compartit bons moments, tant dins com fora del laboratori, amb la Marta, l'Aida, en David, la Mònica, l'Stephi, l'Óscar, en Clarimon, l'Anna P, la

Mònica, en Calafell, l'Arcadi, en Carles, l'Elena. Molts d'ells han seguit la mateixa aventura i amb molt d'ells també he compartit les soporíferes hores de les classes de doctorat. Altres m'han ajudat a millorar les tècniques apreses i me n'han ensenyat de noves amb molta paciència, dedicació i amabilitat, i m'han ofert una ajuda impagable en les anàlisis de dades. I els que no m'han ensenyat ciència m'han ensenyat altres coses de la vida. En els meus darrers temps a BioEvo s'han incorporat noves persones a la Unitat, la Lourdes, l'Anna R, en Tomàs, l'Anna G, la Michelle, l'Andrés, la Gemma, l'Anna "Petita", i la Elodie, entre d'altres, que també s'han convertit en uns bons companys. Sí, molta gent nova, que evidentment comparteixes diferents coses més o menys intenses. Amb molta gent de BioEvo només cal que ens fem una mirada perquè ens vinguin al cap algunes de les coses viscudes i no calen paraules per recordar-les. Aquí també hauria d'incloure, els visitants foranis del laboratori i laboratoris veïns, amb qui no només hem compartit la feina, com en Thomas, la Fede i en Wille entre d'altres.

Però la llista de persones no només es queda reduïda al laboratori de BioEvo. La meua etapa a la Pompeu Fabra m'ha ofert experiències i amistats impagables. Amistats com les d'en Magano, a qui tinc moltes coses a agrair: que perdés tant de temps amb mi ensenyant-me tots els secrets de l'anàlisi de fragments i que em salvés dels obsolets Hoffers, tot i ser del laboratori de Genètica, però, sobretot, que m'oferís una de les millors amistats. La cosa no només era de dos sinó que l'Alberto també formava part de les aventures i les escapades de cap de setmana. Del trio ja vam passar a un *gangbang* amb en Jesús-de-genètica. Amb tots ells he compartit moltes bones històries fora del laboratori. De la gent del laboratori de genètica (l'Anna, en Jaume, la Raquel i en Pérez-Jurado) també conservaré bons records per la seva ajuda i els moments compartits. No puc deixar-me la Maria, que a part de compartir moltes de les escapades amb els companys de la Pompeu, m'ha aguantat durant molt temps els canvis d'humor provocats pel doctorat i amb qui he viscut molt bon moments difícils d'oblidar.

A finals del 2004 ens vam mudar a l'IRTA de Cabrils. Allà no m'han vist gaire el pèl, però el temps suficient per conèixer l'última fase del becari: la del malhumorat, borde i gruñón. Tot i així, he fet bones migas amb la gent d'allà, especialment amb la Mireia i l'Iria, i fins i tot una nova companya de pis, la Kathy.

Durant el temps de doctorat també he pogut viatjar i enriquir-me, tant personalment com professionalment, tot i que el treball realitzat no està reflexat en aquesta tesi. Estant lluny de casa s'agraeix molt l'hospitalitat i que et facilitin les coses, per això, gràcies a la Dra. Jennifer Leonard per acollir-me a casa seva a Washington i per les seves ensenyances al laboratori, tot i que algun cop em compliqués l'existència; gràcies també al Dr. Robert Fleischer per acollir-me al seu laboratori. També gràcies al Dr. Bob Wayne per acollir-me dues vegades al seu laboratori a LA, temps del qual conservo molts bons records per tota la gent que vaig conèixer, especialment de la Isabel que em va ajudar moltíssim, de la Deb per acollir-me a casa seva i de l'Adán, en Borja, en Benajmín i en John pels bons moments. Encara que més curta, també vaig fer una estada a la terra del mormons, al laboratori del Dr. Keith Crandall, gràcies a ell i especialment a en Marcos, la Megan i la Katharina a qui agraixo la seva hospitalitat, amistat i la gran ajuda amb l'anàlisi de les meves dades.

I com no, vull fer una menció especial als amics de tota la vida, els del poble ("Cassà fa cagar") per tot el que hem compartit. Tot i que alguns no saben ben bé el que faig i el poc que saben ho utilitzen per enfotressen, si més no saben veure-li la part còmica (això sempre ho han tingut). Doncs la llista és llarga, però, gràcies: Parés, Bou, Enric, Sancho, Genís, Richal, Moski, Merca, Esteve, Pollastre, Miles, Anna, Anna,... i també gràcies a altres amics que he fet durant aquest temps a Barcelona i puc resumir en dues paraules: Son-iguales i Live-Band-Karaoke. També una doble menció especial per a aquells amics que a part de compartir moments d'amistat, també han estat companys de pis, com en Genís, l'Esteve i, especialment en Richal a Camèlies, que s'ha involucrat en gairebé totes les fases de la meva tesi. Gràcies Camèlies, t'estic agraït pels moments passats en aquest pis, on també he fet amistat amb en Miquel i en Julio amb qui,

juntament amb en Ricard i l'Esteve, hem passat moltes hores explicant els nostres problemes amb els doctorats, els abusos de poder i arreglant el món corrupte de la Ciència. Ai! Si algú ens escoltés.....

També vull recordar els que van viure amb mi els anys de la carrera, com la Marta de Cassà i especialment les meves "companyes de pis" a Girona, la Cris, la Yole, la M<sup>a</sup> Rosa i la Sònia, tot i que em feien fregar els plats molt sovint ;-p. I, com no, tots els col·legues que he fet allà com en David, amb qui encara segueixo compartint moltes coses i que així continuï.

La meva evolució com a becari és clara. Com tot bon becari, he patit la típica metamorfosi: es comença amb la pubertat del becari, sent un becari tímid, ingenu i optimista, i t'acabes convertint en el becari veterà, serio i malhumorat, que va fotent canya als becaris més jovenets. Doncs bé, aquestes línies són per a aquells que, a part d'ajudar-me moltíssim, han tingut una paciència infinita per aguantar tots aquests canvis. Aquí vull fer una menció especial del jurat a la gran paciència i les ganes d'ajudar-me que han tingut la Montse, l'Olga i l'Aïnhua en les últimes fases del doctorat perquè aquest manuscrit arribés on ha arribat. De veritat, moltes gràcies. També aquí incloc a l'Anna que ha passat a formar part de la meva vida en el moments més difícils i que tant m'ha ajudat. A més, és qui més directament ha patit les conseqüències del que costa finalitzar una tesi, des dels inicis de l'escriptura del manuscrit a la "burrocracia" de la fase final.

Per acabar, no puc deixar-me, és clar, una altra persona important que m'ha permès finalitzar la tesi, el Dr. Alfredo Ruíz, que no ha dubtat en cap moment a donar-me tot el seu suport per finalitzar el projecte que us presento a continuació.



*La realització d'aquesta tesi no hagués estat possible sense el finançament de les següents institucions – Ministerio de Educación y Cultura (Ref. PB98-1064); European Commission amb el projecte IMPRIMAT (Ref. INPRIMAT QLRI-CT-2002-01325) i els els projectes LIFE (Refs. LIFE 00/NAT/E 7229, LIFE 00/NAT/E 7331 i LIFE 00/NAT/E 7335) de conservació del visó europeu a Castilla y León, La Rioja i Álava; Departament d'Universitats, Recerca i Societat de la Informació, Generalitat de Catalunya (Ref. 2000SGR00093); la beca predoctoral que he gaudit durant aquests quatre anys concedida pel Departament d'Universitats, Recerca i Societat de la Informació, Generalitat de Catalunya (Ref. 2001FI-00625) – i tot un seguit de col·laboradors que ens han proporcionat mostres i que estan citats en els diferents capítols.*

# ÍNDEX

<b>PRESENTACIÓ</b> .....	<b>15</b>
<b>INTRODUCCIÓ</b> .....	<b>19</b>
1 EL DNA REPETITIU.....	20
1.1 <i>La paradoxa del valor C</i> .....	20
1.2 <i>Tipus de DNA repetitiu</i> .....	21
1.3 <i>DNA egoista</i> .....	26
2 ELS MICROSATÈL·LITS .....	29
2.1 <i>Descripció i principals característiques</i> .....	29
2.2 <i>Mecanismes mutacionals i origen</i> .....	31
2.3 <i>Models mutacionals</i> .....	35
2.4 <i>Mutacions complexes i homoplàsia</i> .....	37
2.5 <i>Aïllament i biaix de composició (ascertainment bias)</i> .....	40
3 ELS RETROPOSONS: ELS ELEMENTS SINES.....	43
3.1 <i>Descripció i classificació dels retroposons</i> .....	43
3.2 <i>Retrotransposició</i> .....	46
3.3 <i>Els SINES com a marcadors filogenètics</i> .....	48
3.4 <i>Aïllament de nous SINES</i> .....	51
4 ELS CARNÍVORS I L'ÚS DE TÈCNiques MOLECULARS PER A LA SEVA CONSERVACIÓ.....	53
4.1 <i>Els carnívors</i> .....	53
4.2 <i>Els mustèlids: el teixó comú i el visó europeu</i> .....	58
4.3 <i>Genètica de la conservació</i> .....	64
4.4 <i>Genètica forense i genotipar sense capturar</i> .....	68
<b>OBJECTIUS</b> .....	<b>71</b>
<b>RESULTATS</b> .....	<b>75</b>
CAPÍTOL I: HIGH INCIDENCE OF NON-SLIPPAGE MECHANISMS GENERATING VARIABILITY AND COMPLEXITY IN EURASIAN BADGER MICROSATELLITES .....	77
CAPÍTOL II: PHYLOGENETIC INFERENCE AND COMPARATIVE EVOLUTION OF A COMPLEX MICROSATELLITE AND ITS FLANKING REGIONS IN CARNIVORES .....	99
CAPÍTOL III: ANALYSES OF CARNIVORE MICROSATELLITES AND THEIR INTIMATE ASSOCIATION WITH TRNA- DERIVED SINES .....	125

CAPÍTOL IV:	
GENETIC DISTINCTION OF AMERICAN AND EUROPEAN MINK ( <i>MUSTELA VISON</i> AND <i>M. LUTREOLA</i> ) AND EUROPEAN POLECAT ( <i>M. PUTORIUS</i> ) HAIR SAMPLES BY DETECTION OF A SPECIES-SPECIFIC SINE AND A RFLP ASSAY .....	153
<b>DISCUSSIÓ .....</b>	<b>173</b>
1 ORIGEN I EVOLUCIÓ DELS MICROSATÈL·LITS .....	174
1.1 <i>Models mutacionals .....</i>	<i>174</i>
1.2 <i>Naixement i distribució dels microsatèl·lits en els genomes.....</i>	<i>178</i>
2 SELECCIÓ DEL MARCADOR MOLECULAR ADEQUAT.....	182
2.1 <i>Aplicacions de les regions repetitives en genètica de la conservació ..</i>	<i>186</i>
3 PERSPECTIVES DE FUTUR DELS MICROSATÈL·LITS .....	193
<b>CONCLUSIONS .....</b>	<b>195</b>
<b>BIBLIOGRAFIA .....</b>	<b>199</b>

# **PRESENTACIÓ**



El treball que presento a continuació pretén estudiar l'origen, evolució i formació dels microsatèl·lits i dels seus patrons complexes utilitzant com a model diferents espècies de carnívors. Els resultats del nostre grup i d'altres autors indiquen l'existència de microsatèl·lits que no s'adapten al patró d'evolució pel qual s'han desenvolupat la majoria d'índexs estadístics. Aquestes observacions i el fet que aquestes regions hipervariables són extremadament importants com a marcadors genètics, tant per a diagnòstics de malalties com per a estudis poblacionals i d'identificació individual, ens han portat a desenvolupar l'eix central de la meva tesi. Al mateix temps, hem utilitzat la informació obtinguda per resoldre qüestions taxonòmiques i de gestió d'espècies amenaçades d'aquest grup de mamífers, el qual presenta notables incongruències en les seves taxonomies i nombroses espècies en perill d'extinció.

El manuscrit d'aquesta tesi comença amb una *Introducció*, on presento breument els diferents tipus de DNA repetitiu del genoma dels mamífers, per després detallar en dos apartats diferents els microsatèl·lits i els retroposons, posant èmfasi en els elements SINE. Per finalitzar la introducció, parlaré de la biologia dels carnívors, especialment dels mustèlids, i de l'ús de tècniques moleculars en genètica de la conservació.

La tesi segueix amb els *Objectius*, on presento les principals fites que es pretenen assolir amb aquest treball, com són estudiar els patrons evolutius complexes dels microsatèl·lits, avançar en el coneixement dels processos mutacionals que originen els microsatèl·lits i explorar l'aplicació en genètica de la conservació dels resultats obtinguts en les anàlisis de les regions repetitives.

Els *Resultats* obtinguts s'exposen en quatre capítols. El primer capítol és un article, en revisió per la revista *Journal of Heredity*, on hem estudiat patrons d'evolució de microsatèl·lits en espècies de la família Mustelidae, ja que presenten abundants imperfeccions i patrons evolutius anormals que donen una variabilitat que no pot ser detectada a través de procediments estàndard de genotipatge. El segon capítol és un article, publicat a la revista *Genetical Research*, on hem aprofitat la conservació en carnívors de la regió flanquejant d'un microsatèl·lit compost complex per analitzar el seu origen i



diversificació. A més a més, hem utilitzat la regió flanquejant d'aquest microsatèl·lit per estudiar la informació filogenètica que aporta i compararla amb les controvèrsies actuals en filogènies de carnívors. El tercer capítol és un article, enviat a la revista *BMC Genomics*, on hem estudiat característiques de la co-evolució dels SINEs i els microsatèl·lits en carnívors i els mecanismes de gènesi de microsatèl·lits a partir dels SINEs. El darrer capítol és un article, publicat a la revista *Journal of Zoology*, on hem creat un marcador molecular utilitzant un SINE que ens permet diferenciar 3 espècies de mustèlids difícils de diferenciar sense capturar i amb prioritats de gestió molt diferents: el visó americà, que és una espècie invasora sobre la qual s'està aplicant un programa de control i d'erradicació; i dues espècies protegides, el turó i el visó europeu, el qual és una de les espècies de carnívors més amenaçades d'extinció a Europa.

Finalment, la *Discussió* és un assaig d'interpretació conjunta de tots els resultats obtinguts.

# **INTRODUCCIÓ**

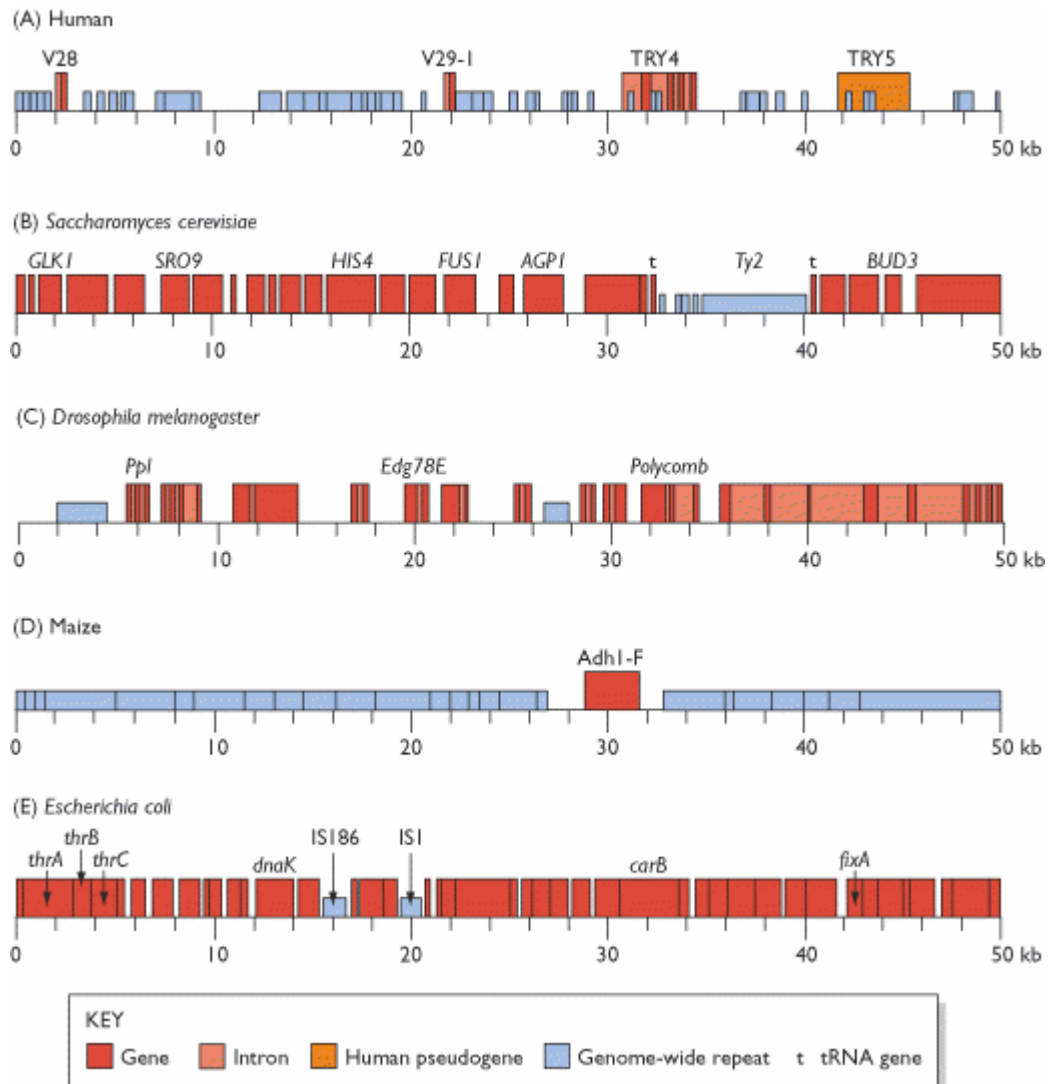
## 1 EL DNA REPETITIU

### 1.1 La paradoxa del valor C

La mida del genoma es defineix com a valor C (Swift 1950) o quantitat de DNA per genoma haploide. Se'l denomina C pel fet de ser (pràcticament) constant dins d'una mateixa espècie. Tanmateix, existeix una gran variació del valor C entre les espècies eucariotes i aquestes diferències no semblen tenir una relació ni amb la complexitat de l'organisme ni amb el nombre de gens que contenen. Per exemple, les amebes presenten genomes 200 vegades més grans que els humans, tot i que una ameba no és més complexa que un humà. Fins i tot, quan es comparen espècies amb complexitat similar trobem àmplies diferències en els seus valors C.

Aquesta falta de correspondència entre els valors C i la quantitat presumible d'informació genètica continguda dins dels genomes es coneix com a paradoxa del valor C (Rosbash et al. 1974; Zuckerkandl 1976). Com que no podem assumir que una espècie posseeixi menys quantitat de DNA de la necessària per complir les seves funcions vitals, hem de trobar una explicació a per què moltes espècies tenen tal quantitat d'excés. L'explicació a gran part d'aquesta paradoxa està en el coneixement que el genoma posseeix gran quantitat de DNA repetitiu, la majoria amb funció no coneguda, i que varia notablement entre espècies (Fig. 1). El DNA repetitiu pot arribar a representar fins al 90% del genoma de les plantes (Nagl et al. 1983) i en el genoma humà es parla que les seqüències repetitives es troben representades per valors superior al 50% (International Human Genome Sequencing Consortium 2001). D'aquesta manera, l'absència substancial de DNA repetitiu en el llinatge del gos i la menor mida alguns d'aquests elements s'utilitza per explicar la mida més petita del genoma del gos respecte al de l'home (Lindblad-Toh et al. 2005).

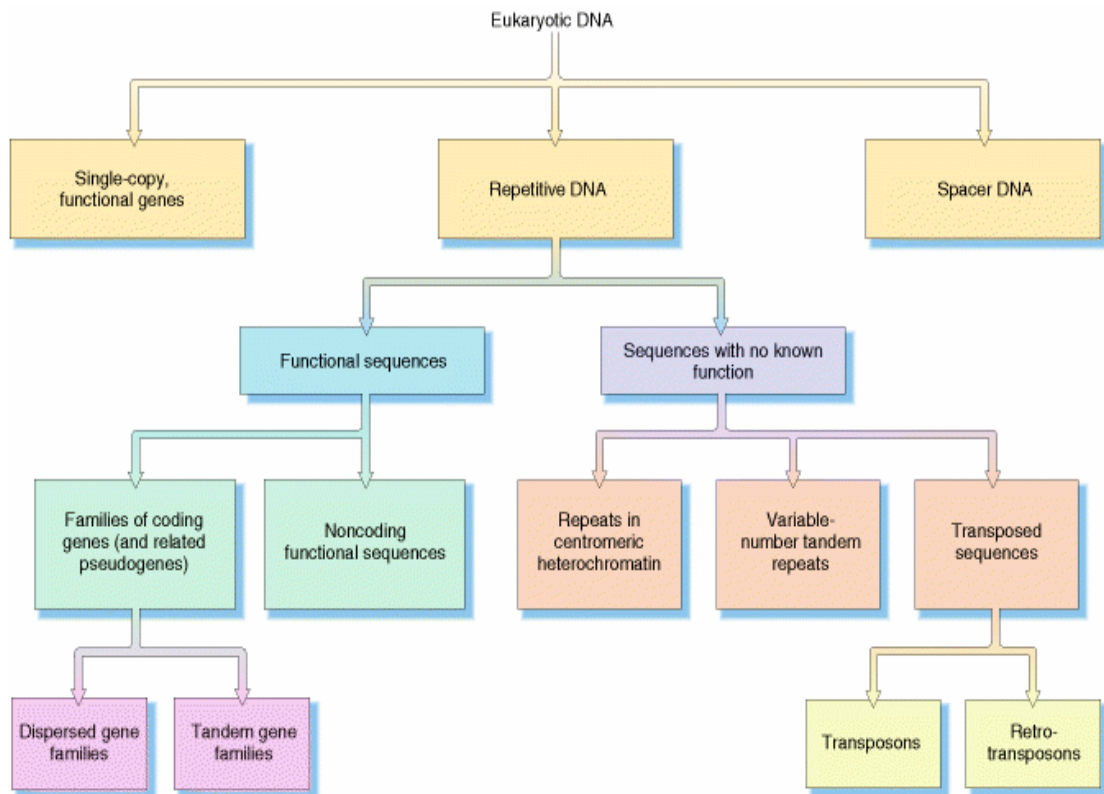




**Figura 1.1.** Comparació d'un fragment del genoma humà (A, receptor b de cel·lules T), del llevat (B, cromosoma III), de la mosca del vinagre (C), del blat de moro (D) i d'*Escherichia coli* (E). Figura extreta de Brown (2002).

## 1.2 Tipus de DNA repetitiu

Existeixen diferents tipus de DNA repetitiu i diversos sistemes de classificació. En la Figura 1.2 podem veure un sistema de classificació (Griffiths et al. 1999) on la primera divisió és entre seqüències de DNA repetides funcionals i seqüències repetides sense una funció coneguda. Utilitzant aquesta classificació, detallarem a continuació els diferents tipus de DNA repetitiu:



**Figura 1.2.** Classificació del DNA del genoma eucariota. Figura extreta de Griffiths et al. (1999).

➤ **Seqüències funcionals:**

Famílies de gens disperses: Molt gens que codifiquen per proteïnes pertanyen a famílies gèniques que es troben disperses per tot el genoma. Aquestes famílies poden ser d'unes poques còpies – p.e., la proteïna ovoalbumina està codificada per una família de 3 gens – o de moltes – p.e., els receptors olfactors són codificats per una família d'uns 900 gens. Alguns gens d'aquestes famílies han passat a ser no funcionals i han donat lloc a pseudogens no transcrits.

Famílies de gens en tàndem: La cèl·lula necessita grans quantitats de productes d'un determinat gen, com els que transcriuen per components transcripcionals, i aquests han evolucionat cap a famílies de gens repetits en tàndem – p.e., la regió de l'Organitzador Nuclear (NO), que és lloc de repeticions en tàndem d'RNA ribosomal.

Seqüències funcionals no codificants: Els telòmers (els extrems dels cromosomes) estan formats per tractes de repeticions en tàndem que no codifiquen ni per RNA ni per proteïnes, però, tot i així, tenen funcions relacionades amb el manteniment de l'estructura cromosòmica i el temps de vida de línies cel·lulars. L'enzim telomerasa és l'enzim encarregat d'afegir unitats de repeticions al final dels cromosomes - p.e., en humans aquestes unitats són TTAGGG.

- Seqüències sense una funció coneguda:

DNA altament repetit que flanqueja els centròmers: Aquestes seqüències de DNA consisteixen en repeticions en tàndem de seqüències curtes que arriben a formar tractes repetitius de centenars de kilobases. Podem trobar una o diverses unitats de repetició i generalment són de longitud inferior a 10 nucleòtids - p.e., en la mosca del vinagre la seqüència AATAACATAG es troba repetida en tàndem al voltant de tots els centròmers.

Repeticions en tàndem de número variable (VNTRs): Els VNTRs són repeticions de seqüències curtes de DNA organitzades en clústers que trobem dispersos per tot el genoma. El terme VNTR fa referència al fet que el nombre de repeticions d'aquestes seqüències pot ser variable entre individus, loci i al·lels del mateix locus. VNTR és una terminologia que va ser introduïda per Nakamura et al. (1987) i feia referència als loci minisatèl·lits. Des de llavors el terme s'ha utilitzat tant per a **minisatèl·lits** com per a **microsatèl·lits**, però, ja que són seqüències de DNA amb clares diferències, utilitzarem els termes específics.

Els minisatèl·lits són repeticions en tàndem on la unitat de repetició té un rang d'entre 14 i 100 parells de bases i que sovint té centenars de repeticions. El mecanisme de mutació predominant que produeix la seva variabilitat en el nombre de repeticions en aquestes seqüències sembla ser la conversió gènica i l'entrecreuament desigual. Els minisatèl·lits són utilitzats per



obtenir el que es coneix com a *DNA fingerprinting*, que ha estat àmpliament utilitzat en genètica forense.

Els microsatèl·lits són seqüències curtes d'entre un i sis parells de bases repetides en tàndem que arriben a formar estructures de fins a 100 unitats repetides. A diferència dels minisatèl·lits, el mecanisme mutacional predominant en els microsatèl·lits és el que es coneix com a *slippage* o *slip-strand mispairing*. En la secció 2 d'aquesta Introducció entrarem en més detall sobre aquest tipus de DNA repetitiu.

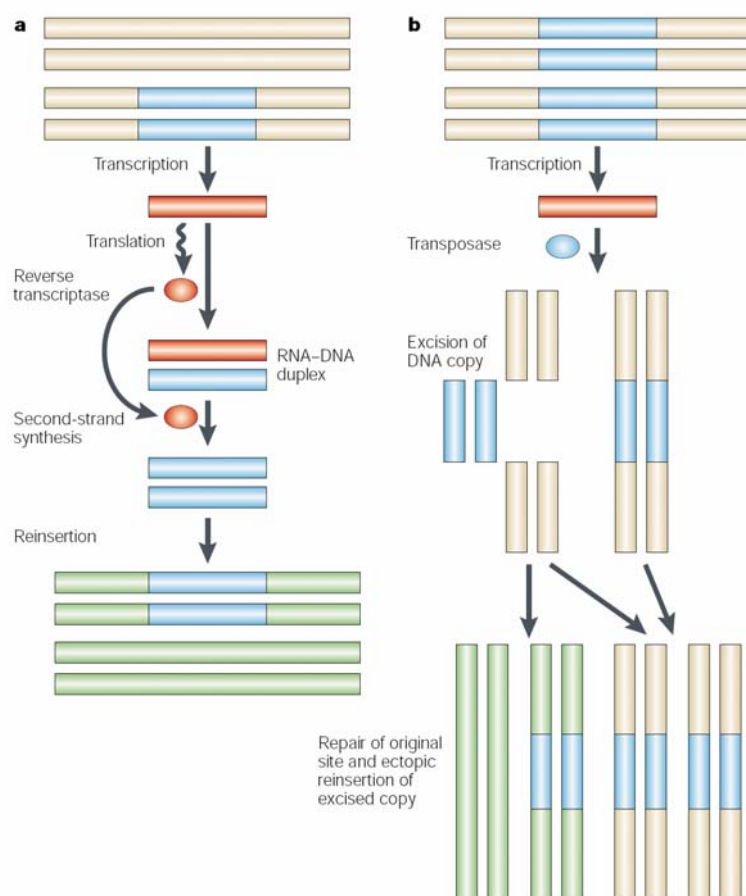
La diferència entre els termes de micro- o minisatèl·lits pot semblar no gaire òbvia però està motivada per la diferència entre els mecanismes mutacionals. La diferenciació entre microsatèl·lits i minisatèl·lits per unitats de repetició d'entre 7 i 14 parells de bases és, generalment, arbitrària. Es sol considerar que si l'estructura repetitiva forma llaços que no són reparats pel sistema de reparació del desaparellament del DNA se l'anomena minisatèl·lit.

Elements mòbils: Una gran porció dels genomes eucariotes està formada per seqüències que s'han propagat pel genoma fent còpies d'elles mateixes i que s'han pogut moure a nous llocs del genoma. Per exemple, els elements transponibles poden arribar a formar el 58% del genoma del blat (Messing et al. 2004), el 45% del genoma humà (International Human Genome Sequencing Consortium 2001), el 15% del genoma de la mosca del vinagre (Biemont & Cizeron 1999) i més del 34% del genoma del gos (*Canis familiaris*; Lindblad-Toh et al. 2005). També s'ha de tenir en compte que aquests valors són una infraestima ja que part del DNA restant pot derivar d'elements transponibles antics la seqüència dels quals ha divergit tant que actualment no els reconeixem com a tal.

Trobem dues grans classes d'elements mòbils:

Classe I: elements transponibles que transposen a través d'un intermediari d'RNA. Després que el DNA és transcrit, la còpia en RNA és convertida en DNA, freqüentment per activitat d'una

transcriptasa inversa que sol estar codificada pel propi element mòbil. Aquesta còpia de DNA es re-insertarà en el genoma en una posició diferent a l'original (Fig. 1.3a). Aquest procés es coneix com a retrotransposició i aquest tipus d'elements es coneixen com a **retroposons**. Dins d'aquesta categoria trobaríem els elements SINE (*Short Interspersed Elements*), els LINE (*Long Interspersed Elements*) i els retroposons amb LTRs (*Long Terminal Repeats*). Dins d'aquesta classe també és interessant distingir els elements pel seu grau d'autosuficiència en els mecanismes de replicació (Fig. 3.1), des dels autònoms – p.e., LINES –, que es creu que codifiquen essencialment per tota la maquinària que necessiten per a la seva mobilització, fins als no-autònoms – p.e., SINEs. Entrarem en més detall sobre aquests elements en la secció 3 d'aquesta Introducció.



**Figura 1.3.** Tipus d'elements transponibles. a) Classe I; b) Classe II. Figura extreta de Hurst & Werren (2001).



Classe II: elements que codifiquen per una transposasa que els escindeix del genoma. El punt d'escissió és reparat utilitzant la seqüència homòloga o la cromatina germana com a motlle i es crea un duplicat de l'element transponible. La còpia escindida s'insereix en un punt diferent del genoma. (Fig. 1.3b). Aquests elements es coneixen com a **transposons**.

### ***1.3 DNA egoista***

Una anàlisi comparativa entre els genomes de l'home, el ratolí i el gos (Lindblad-Toh et al. 2005) mostra que hi ha un 5% del genoma humà que és mantingut per selecció natural – cosa que suggereix que té alguna funció essencial. Només entre l'1-2% del genoma codifica per proteïnes, per tant hi hauria un conjunt d'elements funcionals no codificants comuns (3%). Aquests seqüències comunes podrien constituir, per exemple, elements reguladors i estructurals o gens d'RNA. Tanmateix, l'existència de DNA sense una funció coneguda presenta un dilema. Les idees prèvies sobre el poder de la selecció natural haurien predit que el DNA no funcional seria purgat per la selecció, ja que representaria una càrrega per a l'organisme només pel fet que sintetitzar-lo representa un cost energètic. Tanmateix, aquesta idea sembla inadequada, ja que aquest DNA podria tenir una funció de tipus estructural que permetria una divisió cel·lular o podria separar els elements funcionals per tal de ser regulats eficientment. Una altra hipòtesi seria que els elements repetitius, que constitueixen la major part del DNA sense funció, haurien trobat una manera d'evitar ser detectats per les forces de la selecció natural. Aquest tipus de DNA se l'ha anomenat DNA egoista, un tipus de DNA que existeix només pel propòsit d'existir i que no està exposat als rigors del fenotip. Tot i que els elements mòbils poden ocasionalment induir mutacions beneficioses, alguns autors (Charlesworth et al. 1994) creuen que l'explicació més parsimoniosa és que el DNA repetitiu es replica en el genoma com a DNA paràsit.

Tanmateix, el DNA repetitiu pot estar jugant un paper molt important en l'evolució dels genomes. Els elements transponibles són potencials agents mutagènics amb la capacitat de produir un ampli rang de canvis en els genomes dels hostes (Kidwell & Lisch 2000). Els elements mòbils són indubtablement responsables d'una porció significant de les diferències de mida entre genomes – per exemple doblant la mida del genoma del blat de moro en uns pocs milions d'anys (SanMiguel et al. 1998). A més de la seva habilitat d'incrementar la mida del genoma a partir de la transposició, els elements transponibles indueixen arranjaments cromosòmics tals com delecions, duplicacions, inversions i translocacions recíproques. També s'ha especulat que els elements mòbils podrien tenir un paper en l'aïllament reproductiu i en especiació (Ginzburg et al. 1984).

En la mosca del vinagre s'ha vist que els telòmers són construïts per elements transposables de la classe I (Danilevskaya et al. 1994). Ja que la replicació del DNA no és suficient per mantenir la llargada dels telòmers, els organismes tenen altres mecanismes per mantenir-la. A més a més, la telomerasa – l'enzim encarregat d'extendre els telòmers – té un component de transcriptasa inversa que és estructuralment similar a la transcriptasa inversa dels retroposons sense LTRs. Per tant, s'hipotetitzava que o bé aquests elements mòbils han evolucionat a partir de la telomerasa o bé a la inversa (Eickbush 1997).

En vertebrats, la inserció de transposons podria haver estat adquirida per tal de proporcionar un mecanisme per generar alts nivells de variació en immunoglobulines. Els vertebrats produeixen variació en les immunoglobulines mitjançant la recombinació somàtica. En aquest procés estan involucrats dos enzims, *RAG1* i *RAG2*, que presenten similituds en el mecanisme de funcionament amb una superfamília de transposons i poden afectar la transposició (Agrawal et al. 1998).

Els elements transponibles poden haver tingut un paper important en el disseny dels genomes per la seva capacitat de reestructurar cromosomes induint inversions i translocacions mitjançant la recombinació ectòpica entre seqüències d'elements transponibles homòlegs escampats pel genoma. En l'expansió dels genomes poden contribuir dos tipus de forces,



global o local. Els elements transponibles, juntament amb la poliplodització, són exemples de forces globals per l'expansió mentre que els microsatèl·lits, heterocromatina o altres elements repetits en tàndem representen forces locals de canvi. És àmpliament acceptat, doncs, que els transposons proporcionen material genètic i variació sobre els quals la selecció pot actuar i a partir d'aquí poden evolucionar noves funcions.



## 2 ELS MICROSATÈL·LITS

### 2.1 Descripció i principals característiques

Els microsatèl·lits són seqüències de DNA que consisteixen en la repetició en tàndem d'una unitat bàsica o motiu d'entre un i sis nucleòtids de longitud – p.e., CACACACACA. Per diferenciar-les del DNA de seqüència única, normalment es consideren microsatèl·lits aquelles seqüències repetitives amb un nombre mínim de repeticions – p.e.,  $(CA)_6$  –, tot i que alguns autors descriuen aquest límit amb un nombre mínim de parells de bases, i es consideren proto-microsatèl·lits per sota d'aquest número. En qualsevol cas, aquest límit és més una qüestió semàntica que científica. Una altra característica que contrasta amb el DNA de seqüència única és que la variació genètica a molts loci de microsatèl·lit es caracteritza per una elevada heterozigositat i per la presència de múltiples al·lels. Altres noms han estat utilitzats per descriure aquestes seqüències repetides en tàndem com *simple sequences* (Tautz 1989) o STRs (*short tandem repeats*; Edwards et al. 1991).

Els microsatèl·lits es troben distribuïts de forma més o menys ubíqua al llarg de tots els genomes analitzats fins ara, tot i que en més detall, els microsatèl·lits són rarament trobats en seqüències codificants. Aquest fet és particularment notable pels microsatèl·lits que no estan constituïts d'unitats repetitives de 3 nucleòtids i que, per tant, podrien produir una alteració en la pauta de lectura. Si afegim el fet que no es coneix una funció per aquest tipus de seqüència, la variació existent en els microsatèl·lits es consideraria selectivament neutra en la majoria dels casos.

En el primer informe de l'anàlisi de la seqüència del genoma humà (International Human Genome Sequencing Consortium 2001) es conclou que els microsatèl·lits representen el 3% del genoma i es parla d'un nombre superior al milió de loci de microsatèl·lits. També s'ha trobat que els microsatèl·lits amb unitats de dos nucleòtids són els més abundants,

tot i que aquestes dades poden variar segons com es defineix el que és un microsatèl·lit. Tanmateix, aquests resultats trobats en el genoma humà no tenen per què reflectir el que es produeix en els diferents genomes eucariotes. Al comparar el genoma humà amb el del ratolí (Mouse Genome Sequencing Consortium 2002), s'ha vist que en aquesta espècie existeixen entre 2 i 3 vegades més microsatèl·lits. També s'han trobat grans diferències comparant espècies més properes com l'humà i el ximpanzé (Webster et al. 2002) o entre diferents espècies del genera *Drosophila* (Pascual et al. 2000; Schlötterer et al. 2000).

Els microsatèl·lits es classifiquen normalment seguint dos criteris: segons el nombre de nucleòtids de la unitat bàsica de repetició i segons la composició nucleotídica de la seqüència repetitiva. Seguint el primer criteri, trobem els microsatèl·lits **mononucleòtids**, **dinucleòtids**, **trinucleòtids**, **tetranucleòtids**, **pentanucleòtids** i **hexanucleòtids**. Seguint el segon criteri trobem els microsatèl·lits **perfectes** - p.e., ACACACACACACACACAC - o **imperfectes** - p.e., ACACACTACACGACAC -, segons si estan formats per una única seqüència repetitiva sense cap interrupció en el patró de repetició o amb interrupcions per l'aparició d'algun altre nucleòtid en la seqüència repetitiva, respectivament. També podríem parlar de microsatèl·lits **compostos** - p.e., ACACACACACACAGAGAGAGAGAG - quan trobem dos microsatèl·lits diferents consecutius. No existeix cap consens per definir quin grau d'imperfeció és acceptable en una seqüència repetitiva per poder anomenar-la microsatèl·lit o quantes bases de separació calen per diferenciar un microsatèl·lit compost de dos microsatèl·lits. Per aquesta raó i d'altres que hem comentat anteriorment és important definir els criteris d'identificació d'un microsatèl·lit.

Els microsatèl·lits han estat estudiats per la seva relació amb malalties genètiques (Sutherland & Richards 1995), però també, gràcies a la seva àmplia distribució per tot el genoma, han estat extensament utilitzats per fer mapeig cromosòmic (Dietrich et al. 1994; Dib et al. 1996). Al mateix temps, els microsatèl·lits s'han imposat com els marcadors predilectes per realitzar identificacions individuals i proves de paternitat, ja que es troben entre les seqüències més variables del genoma. Posteriorment, han estat

àmpliament utilitzats en genètica de poblacions i altres àrees relacionades (Jarne & Lagoda 1996). A més a més, les anàlisis de microsatèl·lits basades en PCR han permès analitzar mostres amb quantitats mínimes de DNA – p.e., pèls (Taberlet et al. 1997; Frantz et al. 2004) – o DNA degradat – p.e., excrements (Wasser et al. 1997) i mostres de museu (Pertoldi et al. 2001).

## ***2.2 Mecanismes mutacionals i origen***

A diferència del DNA de seqüència única, el polimorfisme dels microsatèl·lits està causat principalment per la variació del nombre d'unitats repetides, més que per la variació de la seqüència primària, fet que genera un gran nombre d'al·lels. La taxa de mutació dels microsatèl·lits es situa al voltant de  $10^{-2}$  esdeveniments mutacionals per locus per replicació en *E. Coli* (Levinson & Gutman 1987b), entre  $10^{-4}$  i  $10^{-5}$  en llevats (Henderson & Petes 1992) i entre  $10^{-2}$  i  $10^{-4}$  en humans (Ellegren 2004). Aquestes taxes són clarament més elevades que la taxa de mutació de canvi puntual, que és de l'ordre de  $10^{-9}$  a  $10^{-10}$ .

S'han descrit dos models que explicarien aquesta elevada taxa de mutació. Una de les explicacions invoca la recombinació desigual, tot i que la seva influència a l'hora de generar variació és sovint considerada poc important (Eichler et al. 1995). L'altre és el que es coneix com a *slippage*, un error en la replicació del DNA que es considera el principal responsable de l'elevada taxa de mutació dels microsatèl·lits (Levinson & Gutman 1987a; Schlötterer & Tautz 1992). L'*slippage* té lloc durant la replicació del DNA quan la cadena naixent es dissocia de la cadena motlle. Quan es tracta de seqüències no repetitives aquest fet no suposa un problema perquè només hi ha una manera de realinear les dues cadenes, però si es tracta de seqüències microsatèl·lit les cadenes poden realinear-se fora de fase. Si la replicació continua després d'aquest error, la cadena naixent pot ser més llarga o més curta, segons si el mal realineament produeix un llaç a la cadena naixent o a la cadena motlle (Fig. 2.1).

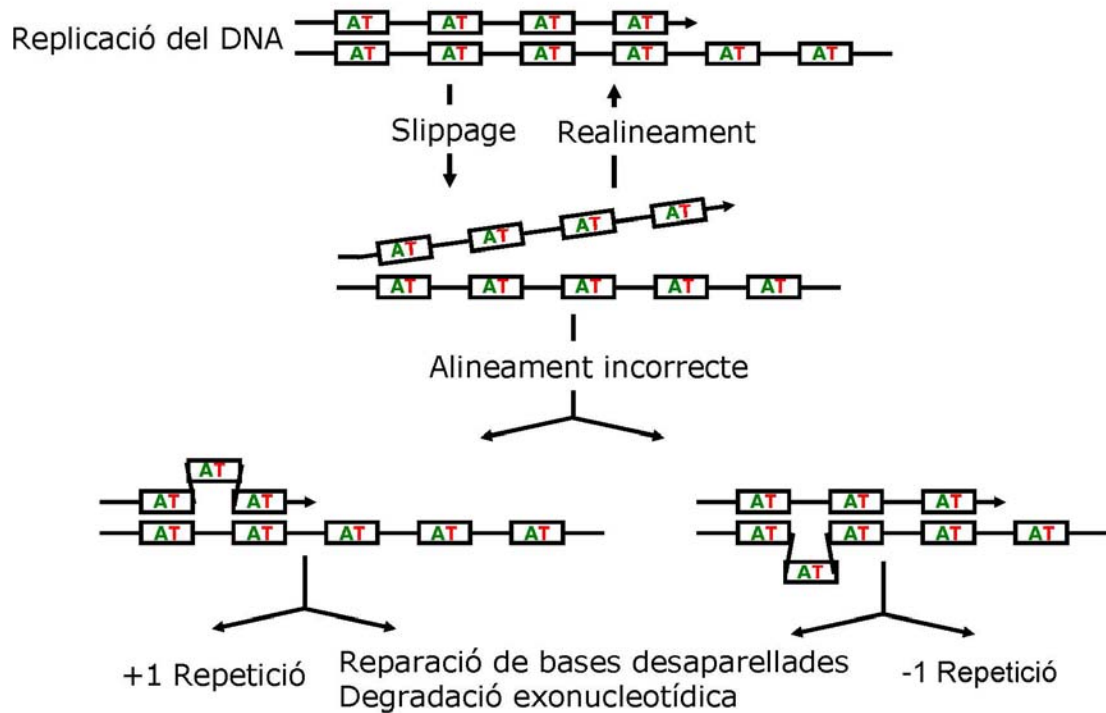


Figura 2.1. DNA *slippage*. Vegeu el text per més detalls.

Un fet que dóna suport a aquesta explicació és que s'ha vist que si el mecanisme cel·lular de reparació de bases desaparellades, que corregeix els llaços formats, no és funcional, la taxa de mutació dels microsatèl·lits augmenta entre 100 i 1000 vegades (Strand et al. 1993). En experiments *in vitro* que utilitzen enzims purificats d'eucariotes i procariotes també s'ha pogut veure que la DNA polimerasa és l'única activitat enzimàtica necessària per tal que es produeixi *slippage* (Schlötterer & Tautz 1992).

Tot i que el mecanisme d'*slippage*, i el paper que juga en la inestabilitat dels microsatèl·lits, està conservat, la taxa específica i els patrons varien entre espècies. Una possible explicació per a aquesta heterogeneïtat en les taxes de mutació és el fet que el mecanisme de reparació de bases desaparellades o la polimerasa presenten diferents taxes d'error entre les diferents espècies.

La taxa de mutació puntual, que pot variar entre organismes, també pot explicar l'heterogeneïtat detectada en la taxa de mutació dels microsatèl·lits, ja que les mutacions puntuals provoquen interrupcions en



el patró repetitiu del microsatèl·lit. Aquestes interrupcions semblen estabilitzar les repeticions, fent que els microsatèl·lits imperfectes siguin menys variables que els perfectes (Kruglyak et al. 1998). En un estudi amb loci ortòlegs de ratolí i rata es va trobar una correlació negativa entre la taxa de substitució a les regions que flanquegen els microsatèl·lits i la longitud del microsatèl·lit (Santibáñez-Koref et al. 2001).

Ara sabem que la taxa de mutació dels microsatèl·lits pot variar entre loci i, fins i tot, entre al·lels d'un mateix locus (Di Rienzo et al. 1998), ja que la naturalesa mateixa del microsatèl·lit hi juga un paper important. Existeixen diferents factors que poden afectar les diferències observades en la dinàmica evolutiva dels microsatèl·lits com són: el nombre de repeticions, la seqüència del motiu de repetició, la llargada de la unitat, la seqüència de la regió flanquejant, interrupcions en el tracte repetitiu i altres (vegeu Schlöterer 2000 per una revisió).

Un dels factors amb més pes és el nombre de repeticions, de manera que els microsatèl·lits més llargs acostumen a ser més polimòrfics. Aquest fet s'ha observat tant en dades obtingudes a partir de pedigrís (Weber & Wong 1993; Brinkmann et al. 1998) com a partir d'estudis poblacionals (Goldstein & Clark 1995). Intuïtivament, sembla comprensible que com més unitats de repetició existeixin més probabilitats hi ha que es produeixi *slippage*. Una taxa de mutació depenent de la llargada de microsatèl·lits explicaria part de l'heterogeneïtat observada a diferents nivells (p.e., la baixa taxa de mutació observada en la mosca del vinagre seria compatible amb el fet que presenten microsatèl·lits més curts que els vertebrats; Schug et al. 1997). Alguns estudis recents parlen d'una relació lineal (Sibly et al. 2003) mentre que d'altres parlen d'una relació exponencial (Leopoldino & Pena 2003) entre la mida del microsatèl·lit i la taxa de mutació.

Inicialment es va proposar que els tetranucleòtids mutaven més que els dinucleòtids (Weber & Wong 1993), però en estudis *in vitro* s'ha observat una correlació negativa entre la longitud del motiu de repetició i la taxa de mutació (Schlötterer & Tautz 1992), fet demostrat en treballs més recents (Lee et al. 1999).

També s'ha postulat que la seqüència flanquejant podria afectar la taxa d'*slippage*. Glenn et al. (1996) van detectar una correlació negativa entre el contingut de GC de les seqüències flanquejants i la taxa d'*slippage*. Altres estudis no troben aquesta variació (Balloux et al. 1998) i podria ser que aquesta correlació estigués causada per altres factors, com el nombre de repeticions, més que pel diferent contingut en GC.

Fins ara hem vist com muten els microsatèl·lits, però el seu procés de formació és menys conegut. S'assumeix que la formació d'un proto-microsatèl·lit és produïda a l'atzar. Aquests proto-microsatèl·lits poden aparèixer a partir de simples substitucions de bases, com una transició A-G en GCACGC a GCGCGC, però també una porció significant dels nous loci de dues repeticions apareixen a partir de mutacions d'inserció degudes a duplicacions de la seqüència adjacent (Zhu et al. 2000). Rose & Falush (1998) van determinar que era necessari un nombre mínim de repeticions per tal que l'*slippage* pogués estendre els proto-microsatèl·lits. Per una altra banda, Pupko & Graur (1999) van observar en llevats que aquest nombre mínim no és necessari i, en un altre estudi, Primmer & Ellegren (1998) van observar que l'*slippage* es pot produir en microsatèl·lits de tan sols dues repeticions. D'aquesta manera, els microsatèl·lits es podrien generar per mutacions puntuals produïdes a l'atzar seguides d'esdeveniments menys freqüents d'*slippage* que estendrien els proto-microsatèl·lits, fins que l'*slippage* fos el principal mecanisme mutacional.

Un mecanisme completament diferent de generació de microsatèl·lits s'ha trobat a partir de l'associació de microsatèl·lits amb elements retrotransponibles (Arcot et al. 1995; Nadir et al. 1996; Yandava et al. 1997). Els autors proposen que les cues de poli-A presents en els elements retrotransposats generarien la majoria de microsatèl·lits rics en A/T. Aquesta cua poli-A es produeix per la poliadenilació quan el retrotransposó es troba en forma d'RNA, de la mateixa manera que en els mRNAs. També és possible que la inserció d'un element transposable estigués afavorida per la preexistència d'un microsatèl·lit (Wilder & Hollocher 2001).

## 2.3 Models mutacionals

Els microsatèl·lits estan entre les seqüències més variables del genoma i, a més a més, són fàcils de genotipar, cosa que fa que siguin un dels marcadors genètics més potents i més utilitzats. En els últims anys, els microsatèl·lits han estat un dels marcadors predilectes en genètica de poblacions. L'estima de nombrosos paràmetres poblacionals (p.e., diferenciació genètica, nombre de migrants, etc.) a partir de dades empíriques (p.e., freqüències al·lèliques) depèn del model mutacional assumit pels marcadors utilitzats. La sensibilitat del model mutacional augmenta quan la taxa de mutació augmenta, d'aquesta manera la dependència del model mutacional és especialment forta en el cas dels microsatèl·lits.

Clàssicament, en els estudis amb microsatèl·lits, es consideren dos models teòrics oposats. Per un costat tenim l'SMM (*Stepwise Mutational Model*, Kimura & Ohta 1978), el qual proposa que les mutacions en els microsatèl·lits consisteixen en l'addició o pèrdua (amb igual probabilitat) d'una unitat repetitiva, conseqüentment els al·lèls poden mutar cap a al·lèls ja presents en la població. En contrast a aquest model trobem l'IAM (*Infinite Allele Model*, Kimura & Crow 1964), el qual proposa que cada mutació pot implicar qualsevol nombre de repeticions i generar un nou al·lèl no present en la població. De manera que els al·lèls sota l'IAM difereixen igualment en termes d'esdeveniments mutacionals, mentre que en l'SMM els al·lèls amb mides similars difereixen en menys esdeveniments que els que tenen més diferències.

Gràcies a l'elevada taxa de mutació dels microsatèl·lits ha estat possible observar mutacions *de novo* per un ampli rang de loci i espècies (Weber & Wong 1993; Shimoda et al. 2000; Gardner et al. 2000; Brohede et al. 2002). El patró general és coherent amb l'*slippage*, però, de fet, algunes vegades les mutacions impliquen més d'una unitat repetitiva, cosa que indica que l'SMM, estrictament parlant, no seria vàlid per a la majoria de casos. Una dada important d'aquestes observacions és que el procés mutacional sembla ser heterogeni pel que fa als loci, motius de repetició i organismes. També s'ha pogut observar que alguns loci són més

propensos que altres a grans canvis. Per tal d'incloure la possibilitat de canvis de més d'una unitat repetitiva, Di Rienzo et al. (1994) van proposar un altre model, el TPM (*Two-Phase Model*), derivat de l'SMM. Les mutacions en aquest model modificarien el nombre de repeticions en una unitat amb una probabilitat  $P$  i en més d'una unitat amb probabilitat  $1-P$ .

Sota aquests tres models bàsics d'evolució de microsatèl·lits, l'*slippage* seria un procés simètric, de manera que per mitjana el mateix nombre de repeticions serien afegides o eliminades. Aquest procés portaria inevitablement a la pèrdua del microsatèl·lit o a l'aparició de microsatèl·lits amb un nombre molt elevat de repeticions. Contràriament a les implicacions d'aquests models, trobem microsatèl·lits que estan conservats durant llargs períodes de l'escala evolutiva (300 milions d'anys (m.a.) en tortugues marines, FitzSimmons et al. 1995; 470 m.a. en peixos, Rico et al. 1996; 144 m.a. en vespes, Ezenwa et al. 1998) i, en el genoma, tampoc s'observen valors molt extrems en el nombre de repeticions dels microsatèl·lits. Aquesta última dada ha estat utilitzada com a evidència per hipotetitzar que la selecció podria actuar per mantenir la llargada dels microsatèl·lits dins d'un rang de mides (Bowcock et al. 1994; Garza et al. 1995). Una altra explicació està en dades observades en llevats (Wierdl et al. 1997), on s'observa com la mitjana de mutacions en els microsatèl·lits curts tendeix a allargar-los i en microsatèl·lits llargs, a escurçar-los.

Kruglyak et al. (1998) van proposar un altre model mutacional per explicar el fet que la distribució de la longitud dels microsatèl·lits està en equilibri. Aquest model estableix que l'equilibri és el resultat d'un balanç entre la longitud del tracte repetitiu i les mutacions puntuals. D'acord amb aquest model, dues forces mutacionals operarien sobre les seqüències dels microsatèl·lits: l'*slippage*, la taxa del qual augmentaria segons augmentés el nombre de repeticions; i les mutacions puntuals que trencarien el tracte repetitiu en peces més petites, fins i tot, fins a fer-lo desaparèixer. Tot i així, l'*slippage* pot eliminar les mutacions puntuals (Harr et al. 2000), de manera que aquestes no necessàriament arribarien a eliminar el microsatèl·lit, sinó que el portarien a un estat transitori. Per

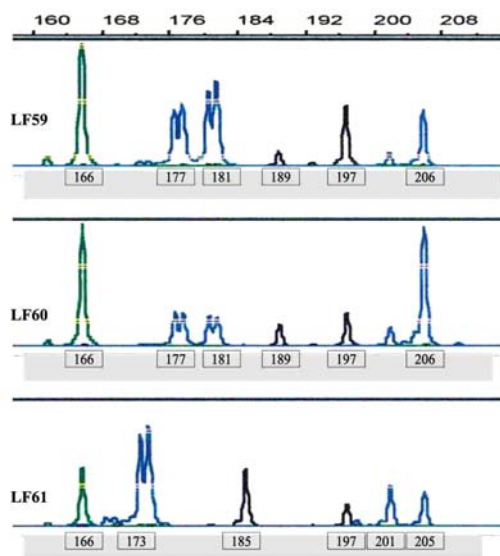


tant, aquest seria un altre factor que també hauria de ser inclòs en els models mutacionals.

L'evolució dels microsatèl·lits sembla ser més complexa del que es pensava inicialment i hi ha molts models que no poden explicar acuradament la distribució de les freqüències al·lèliques en les poblacions naturals. Els coneixements actuals sobre els processos mutacionals van augmentant, però encara queda camí per parametritzar adequadament els factors que es consideren rellevants per a l'evolució dels microsatèl·lits i d'aquí permetre'n una representació òptima en els models mutacionals.

## 2.4 Mutacions complexes i homoplàsia

*A priori*, les bases genètiques de la variabilitat dels microsatèl·lits són fàcilment apreciables. Uns encebadors específics són utilitzats per amplificar una regió genòmica que inclou la regió repetitiva, per després, mitjançant l'electroforesi dels productes de PCR obtinguts, revelar la variació present en el microsatèl·lit. Actualment, la PCR es realitza utilitzant uns encebadors marcats amb fluorocroms i a continuació s'analitzen els productes utilitzant un seqüenciador automàtic que ens permet determinar acuradament la longitud dels diferents al·lels (Fig. 2.1).



**Figura 2.2.** Resultat de la lectura, mitjançant un seqüenciador automàtic, de la mida dels fragments de DNA amplificats amb encebadors marcats amb fluorocroms. Corresponen a 3 mostres diferents (LF59, LF60 i LF61) genotipades per 4 microsatèl·lits (les dues primeres comparteixen el mateix perfil genotípic). Cada pic correspon a la senyal fluorescent de cada electromorf i l'alçada a la intensitat. L'escala en la part superior indica la longitud del fragment en parells de bases.

Òbviament, ens podem trobar amb microsatèl·lits que tinguin dos al·lels de la mateixa mida i, fins i tot, la mateixa seqüència, però que no siguin idèntics per descendència. Aquest fenomen que dos al·lels siguin idèntics en estat però no provinguin del mateix al·lel ancestral se'l coneix com a homoplàsia de grandària. Sovint, l'homoplàsia es considera com a "soroll" evolutiu, mentre que el contrari, l'homologia, es considera el "senyal" evolutiu. Alguns models mutacionals, com l'SMM, assumeixen cert nivell d'homoplàsia, ja que l'addició o eliminació d'unitats repetitives podria generar al·lels existents en la població. Per altra banda, l'IAM assumeix que les noves mutacions donen lloc a nous al·lels distingibles, és a dir, que els nous al·lels són sempre diferents als que ja existien en la població original, per la qual cosa en aquest cas l'homoplàsia no existeix o és menyspreable. Tanmateix, és d'esperar que l'homoplàsia sigui un fenomen freqüent en els microsatèl·lits ja que, com s'ha vist en dades poblacionals, la grandària dels al·lels es mou dins d'un rang, i, per tant, el nombre de possibles estats al·lèlics es redueix. A més a més, l'homoplàsia depèn de la taxa de mutació del locus i del temps de divergència entre individus. De manera intuïtiva, doncs, suposem que l'ocurrència d'homoplàsia augmenta amb la taxa de mutació i el temps de divergència. Així, l'homoplàsia és generalment major entre espècies que entre poblacions de la mateixa espècie i rara dins de poblacions.

Els models mutacionals teòrics que hem discutit en la secció anterior (excepte l'IAM) i les mesures de distàncies genètiques basades en aquests models més utilitzades (Goldstein et al. 1995; Slatkin 1995) assumeixen que les diferències al·lèliques són enterament degudes a canvis en la regió repetitiva del microsatèl·lit. Tanmateix, quan es mira la distribució per mides dels al·lels, s'observa que no sempre la diferència entre al·lels correspon a canvis d'unitats senceres de repetició. Per tant, presumiblement, aquesta distribució reflecteix processos mutacionals més complexes que els assumits pels models mutacionals. Aquestes mutacions introdueixen la possibilitat d'un nou nivell d'homoplàsia de grandària. La seqüenciació d'al·lels ha permès detectar esdeveniments d'insercions i delecions en la regió flanquejant dels microsatèl·lits al comparar diferents espècies (p.e., FitzSimmons et al. 1995; Angers & Bernatchez 1997), però

també, encara que més rarament, dins d'una mateixa espècie (p.e., Grimaldi & Crouau-Roy 1997; Anderson et al. 2000 ). Aquest tipus de mutacions poden portar a confusió a l'hora de determinar el nombre de repeticions dels microsatèl·lits si ens basem en la grandària dels al·lels. Fins i tot, aquest tipus de mutacions poden passar desapercebudes ja que els fragments delecionats o inserits poden mimetitzar una variació d'unitats senceres de repetició i, per tant, generar homoplàsia de grandària.

Com altres regions del genoma, els microsatèl·lits estan sotmesos a altres tipus de mutacions que no necessàriament són detectades per canvis en la mida dels al·lels. Tot i que la taxa de mutació puntual és més baixa que l'*slippage*, les interrupcions en el tracte repetitiu, a part d'introduir un altre nivell de polimorfisme, poden fer que els al·lels de la mateixa mida difereixin en el nombre i localització d'aquestes interrupcions. Les mutacions puntuals també poden generar diferències entre al·lels de la mateixa mida en la regió flanquejant del microsatèl·lit. Amb microsatèl·lits perfectes l'homoplàsia només es pot detectar a través de variacions en la regió flanquejant, ja que la taxa de mutació puntual en principi és molt més baixa que la taxa d'*slippage*.

Els microsatèl·lits compostos o altres regions repetitives dins del fragment amplificat també són una font de complexitat ja que probablement impedeixen un procés únic d'addició o eliminació de repeticions (una assumció dels models evolutius).

Com hem vist, hi ha una fracció de l'homoplàsia que pot ser accessible a nivell molecular ja que un electromorf pot amagar diferents seqüències. Tot i així, encara hi ha una altra fracció que es manté indetectable, per tant, tot i que a nivell de seqüència es pugui identificar homoplàsia, aquesta serà una subestima del total.

Molts loci de microsatèl·lits estan poc caracteritzats a nivell de seqüència i el fet de no tenir en compte l'homoplàsia porta a subestimar la diferència real entre poblacions i complica la interpretació dels patrons de variació de longitud dels al·lels. Aquest fet pot ser especialment problemàtic sota

l'SMM, ja que s'espera una quantitat substancial d'homoplàsia. A través de la seqüenciació es poden detectar els microsatèl·lits perfectes que tendeixen a desviar-se menys de l'SMM que els microsatèl·lits imperfectes o compostos.

## ***2.5 Aïllament i biaix de composició (ascertainment bias)***

Els microsatèl·lits es poden obtenir fent cerques en bases de dades genòmiques per aquells organismes en els quals s'estan aplicant projectes de seqüenciació o de mapeig cromosòmic. També poden ser amplificats en espècies relacionades pels loci homòlegs utilitzant els mateixos encebadors, tot i que el percentatge de loci amplificats disminueix a mesura que augmenta el temps de separació entre espècies. A més a més, els microsatèl·lits generalment es troben en regions no codificants on la taxa de mutació és més elevada que en les regions codificants i les probabilitats d'amplificar un microsatèl·lit amb encebadors dissenyats en una altra espècie són baixes. Generalment es considera que els microsatèl·lits estan poc conservats. Tanmateix, existeixen excepcions en les quals s'ha trobat que les seqüències flanquejants del microsatèl·lit es troben prou conservades com per obtenir el microsatèl·lit en espècies molt allunyades (300 m.a., FitzSimmons et al. 1995; 470 m.a. en peixos, Rico et al. 1996; 144 m.a. en vespes, Ezenwa et al. 1998).

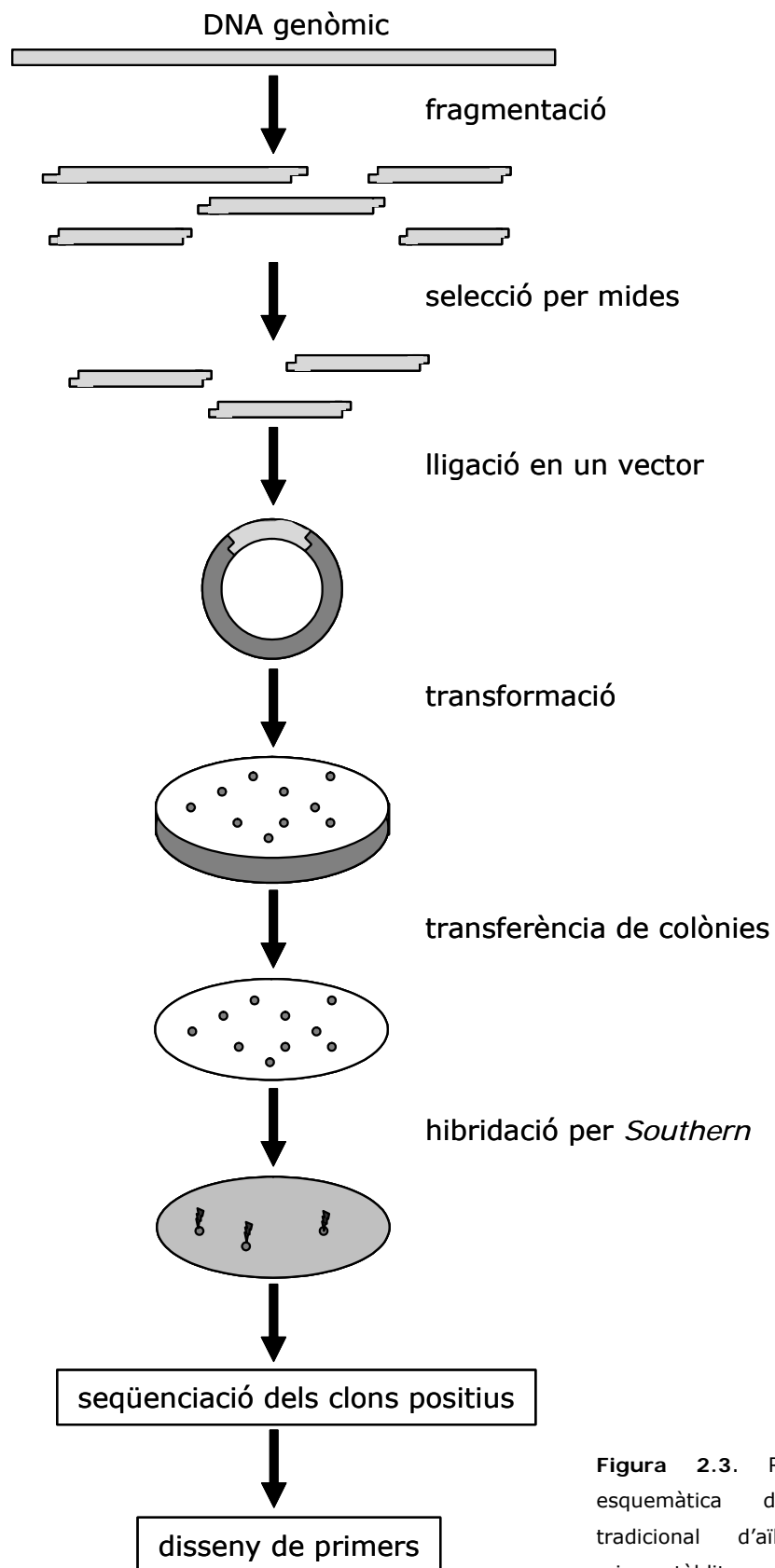
Tot i que el microsatèl·lit estigui conservat en altres espècies, no és garantia que el nivell de polimorfisme es mantingui. Varis estudis han demostrat que els microsatèl·lits aïllats en una espècie són consistentment més polimòrfics que en els ortòlegs d'una altra (Ellegren et al. 1995; Rubinsztein et al. 1995; Crawford et al. 1998). Aquest fet es podria explicar per causes poblacionals (Rubinsztein et al. 1995), com una mida poblacional més gran en l'espècie d'on s'ha aïllat el microsatèl·lit, però també s'ha proposat com a causa la selecció dels microsatèl·lits durant el clonatge (Ellegren et al. 1995). Els microsatèl·lits aïllats són seleccionats per presentar seqüències repetitives llargues i, per tant, s'espera que siguin la fracció més polimòrfica de tots els microsatèl·lits del genoma i

aquests probablement siguin més llargs que els seus ortòlegs en altres llinatges; aquest fenomen es coneix com a biaix de composició (*ascertainment bias*). És important recordar que moltes aplicacions dels microsatèl·lits depenen del fet que presentin un suficient nivell de polimorfisme. Per tant, en moltes espècies s'han d'aïllar *de novo*, una limitació significativa per la inversió inicial de recursos econòmics i d'experiència tècnica requerida pel clonatge i seqüenciació dels loci microsatèl·lit.

Tradicionalment, els microsatèl·lits s'han aïllat a partir de llibreries genòmiques parcials de l'espècie d'interès i explorant milers de clons a través d'hibridacions sobre colònies amb sondes que contenen seqüències repetitives (Rassmann et al. 1991). En aquest procés, el DNA genòmic de l'espècie d'interès és fragmentat per mitjà d'enzims de restricció, i després es fa una selecció de fragments per la seva mida, amb preferència pels de llargada curta (~200-1000bp). Els fragments seleccionats són lligats en vectors per transformar cèl·lules bacterianes que produiran milers de clons recombinants. La cerca dels clons que contenen els microsatèl·lits es realitza generalment mitjançant una hibridació per *Southern* utilitzant sondes que contenen seqüències repetitives. Els clons seleccionats són seqüenciats per tal de determinar la seqüència flanquejant al microsatèl·lit i poder dissenyar encebadors específics de locus (Fig. 2.2).

A partir d'aquest mètode s'obtenen entre el 12% i menys del 0.4% de clons amb microsatèl·lits (Zane et al. 2002), cosa que fa que aquest mètode sigui elevadament costós en termes de temps i diners, especialment quan l'espècie d'interès conté un baix nombre de microsatèl·lits. La majoria de notes tècniques publicades en els últims 15 anys fan referència a aquest mètode, tot i que més de la meitat dels últims articles publicats fan referència a mètodes que parteixen de llibreries enriquides amb microsatèl·lits (Zane et al. 2002). Existeixen diferents mètodes (p.e., Ender et al. 1996; Fisher et al. 1996; Paetkau 1999, Armour et al. 1994) que tracten de partir d'una llibreria rica en microsatèl·lits en comptes d'una llibreria de tot el genoma. Aquests últims mètodes són més ràpids, eficients i especialment útils quan l'espècie

d'interès conté un baix nombre de microsatèl·lits en el genoma (p.e., plantes i aus).



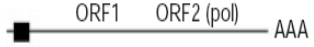

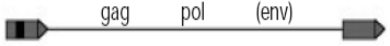
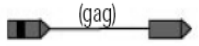
**Figura 2.3.** Representació esquemàtica del mètode tradicional d'aïllament de microsatèl·lits.

### 3 ELS RETROPOSONS: ELS ELEMENTS SINES

#### 3.1 Descripció i classificació dels retroposons

Com indicàvem en el primer apartat d'aquesta tesi, els retroposons són elements de DNA mòbils que transposen a través d'un intermediari d'RNA utilitzant una transcriptasa inversa (Fig. 1.3). Aquests elements estan presents en tots els organismes eucariotes estudiats fins al moment. En mamífers, la majoria d'aquests elements mòbils es caracteritzen per l'absència d'introns, una cua rica en As i repeticions directes que flanquegen l'element repetitiu produïdes en el moment de la inserció (Weiner 1986).

Aquests elements mòbils els podem dividir en dos grans categories: retroposons virals i no virals. També els podem classificar segons el grau d'autosuficiència en els mecanismes de replicació (Fig. 3.1), des dels autònoms (p.e., LINEs), que es creu que codifiquen essencialment per tota la maquinària que necessiten per la seva mobilització, fins als que no són capaços d'automobilitzar-se (p.e., SINEs).

			Length	Copy number	Fraction of genome
LINEs	Autonomous		6-8 kb	850,000	21%
SINEs	Non-autonomous		100-300 bp	1,500,000	13%
Retrovirus-like elements	Autonomous		6-11 kb 1.5-3 kb	450,000	8%
	Non-autonomous				

**Figura 3.1.** Classes de retroposons en el genoma humà. Figura modificada d'International Human Genome Sequencing Consortium (2001).

Retroposons virals o elements amb LTRs: Els retroposons virals, els quals discutirem breument, són abundants en llevats i en la mosca del vinagre, però en canvi s'estima que només formen el 8% del genoma humà (Fig. 3.1) i el 3% del genoma del gos (Kirkness et al. 2003; Lindblad-Toh et al. 2005). A més de les repeticions directes en els extrems 3' i 5', típiques en els llocs d'inserció de tots els elements mòbils, els retroposons virals es caracteritzen per la presència de repeticions d'entre 250 i 600 parells de bases conegudes com a *Long Terminal Repeats* (LTRs), que donen nom a aquest tipus de retroelements. En els elements autònoms, aquestes repeticions estan flanquejant una regió central que conté almenys 3 gens (*gag*, *pol*, i *env*; vegeu Fig. 3.1). La majoria (85%) dels elements "fòssils" derivats dels retroposons amb LTRs estan formats per un únic LTR sense la seqüència interna, perduda per recombinació homòloga entre els LTR flanquejants (International Human Genome Sequencing Consortium 2001). En mamífers, els elements amb LTRs són similars als retrovirus, excepte per l'absència d'un gen funcional per la càpsida (*env*; Ono et al. 1987). En mamífers estan classificats en 3 classes (I-III), cada una d'elles constituïdes per moltes famílies d'origens independents. Tot i la varietat de retroposons amb LTRs que existeixen, només els retrovirus endògens específics de vertebrats (ERVs) semblen estar actius en el genoma dels mamífers (Bénit et al. 1999).

Retroposons no-virals: Els retroposons no virals són els elements mòbils més comuns en mamífers; poden arribar a constituir més del 35% del genoma humà (Fig. 3.1) i el 29% del genoma del gos (Kirkness et al. 2003; Lindblad-Toh et al. 2005). Aquests elements no presenten LTRs i, en els genomes de mamífers, molts d'ells entren dins de dos classes de seqüències de DNA moderadament repetides: els LINEs (*long interspersed elements*) i els SINEs (*short interspersed elements*). En protozous, insectes i plantes s'han trobat seqüències repetides amb característiques similars als LINEs, però, per alguna raó desconeguda, aquests retroelements, igual que els SINEs, són particularment abundants en els genomes de mamífers.

Tot i que la majoria estan truncats per l'extrem 5', els **LINEs** són aproximadament de 6 kb de llarg i contenen dos marcs oberts de lectura



(ORF) que codifiquen per una endonucleasa i una transcriptasa inversa (Feng et al. 1996, Jurka 1997). A més a més, els LINEs posseeixen un promotor per l'RNA polimerasa II dins de la seva regió 5' no traduïda (UTR), i acaben amb una cua poli-A i un senyal de terminació dins de la seva regió 3' UTR (Moran et al. 1996). Els LINEs representen l'únic retroposó no viral actiu en el genoma dels primats, tot i que el seu origen s'estén des del principi de l'evolució eucariota (Eickbush 1992).

Els **SINES** són seqüències de DNA de menys de 600 bp i els trobem presents en centenars de milers de còpies disperses pel genoma, constitueixen aproximadament del 5 al 15% del genoma dels mamífers. Les famílies de SINE típicament deriven de tRNAs (Okada 1991), amb l'excepció de les famílies Alu d'humans i els elements B1 de rosegadors, que són homòlegs al 7SL RNA citoplasmàtic (Ullu & Tschudi 1984). El tRNA de la Lisina (tRNA-Lys) sembla ser la font més comú dels nous SINES en la natura (Okada 1991). Tot i la diversitat de famílies que podem trobar, els SINES presenten una sèrie d'estructures comunes. Els elements SINE contenen un promotor intragènic de l'RNA polimerasa III funcional que és capaç de portar a terme la transcripció de l'element complet mitjançant l'RNA polimerasa III (Jelinek 1982; Rogers 1985) i una cua de poli-A en l'extrem 3'. Tanmateix, com hem dit anteriorment, no contenen la capacitat de codificar les proteïnes implicades en el procés de retrotransposició, de manera que aquestes han de ser subministrades en *trans* per altres elements. Algunes vegades s'ha fet referència als SINES com a paràsits de paràsits per aquesta dependència d'altres elements per a la seva mobilitat (Shedlock & Okada 2000).

Existeix un debat obert sobre si els SINES són paràsits del genoma que treuen profit dels recursos cel·lulars o bé si són escampats pel seu potencial impacte positiu sobre l'evolució del genoma (Britten 1997; Hurst & Werren 2001). S'ha vist que els SINES poden utilitzar-se com a mutàgens inseribles, alterar l'expressió gènica o promoure la conversió gènica i la recombinació homòloga (Schmid 1998; Roy et al. 2000; Weiner 2002; Kazazian 2004). A més a més, la majoria de SINES tenen cues de poliadenines que poden ser origen de microsatèl·lits (Arcot et al. 1995; Nadir et al. 1996). S'ha suggerit que l'abundància de microsatèl·lits de

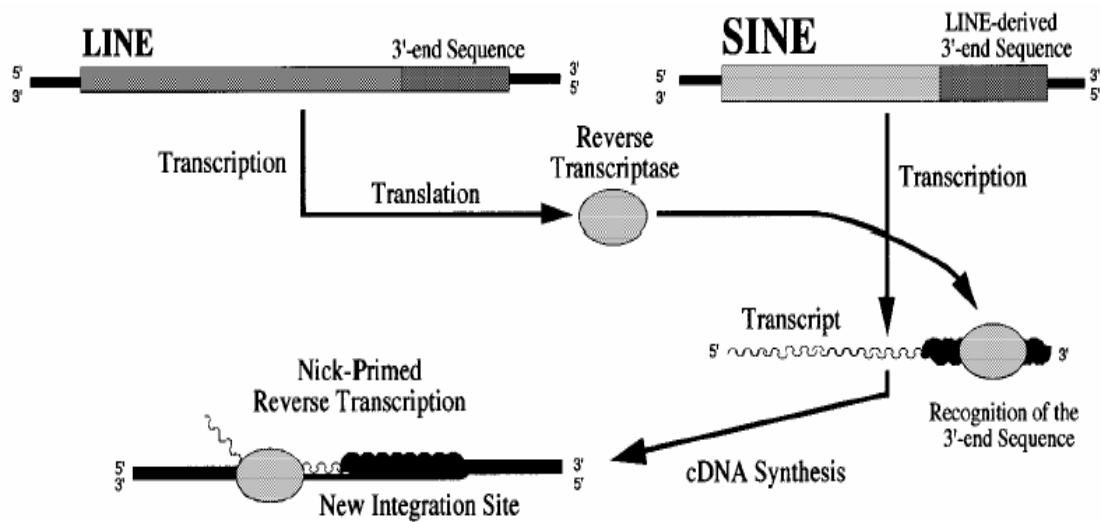
poli-A/T presents en el genoma humà és reflex de l'origen i de l'expansió a partir de retroposons (Beckmann & Weber 1992; Arcot et al. 1995; Nadir et al. 1996; Yandava et al. 1997). L'existència dels SINEs en el genoma, doncs, contribueix a augmentar la seva plasticitat donant-li forma de diferents maneres.

### ***3.2 Retrotransposició***

El procés de mobilitzar un retroposó s'anomena retrotransposició i això implica transcripció, processament de l'RNA, traducció, transcripció inversa i integració del nou producte de DNA dins d'una nova localització del genoma (Fig. 1.3a).

La retrotransposició dels LINEs és via un mecanisme conegut com a *Target Primed Reverse Transcription* (TPRT; Luan et al. 1993). El TPRT és un mecanisme a través del qual l'mRNA del LINE s'uneix a un 3'-OH lliure en l'osca creada per l'endonucleasa en els llocs 5'-TT/AAAA del genoma (Jurka 1997, Luan et al. 1993). La transcriptasa inversa transcriu l'mRNA del LINE a una còpia de cDNA que s'integra en el lloc diana (Luan & Eickbush 1995). Durant el procés d'integració es creen repeticions directes curtes que flanquegen el nou element inserit (Van Arsdell et al. 1981), aquestes són típiques de tots els retroposons.

Estudis recents han demostrat que els SINEs no poden mobilitzar-se sense proteïnes derivades dels LINEs (Dewannieux et al. 2003) i, com que posseeixen un extrem 3' similar a alguns dels LINEs trobats en la mateixa espècie, sembla que comparteixen el mateix mètode de mobilització (Fig. 3.2; Ohshima et al. 1996; Okada et al. 1997; Terai et al. 1998; Kajikawa & Okada 2002). Un cop inserits en el genoma, els SINEs generalment no són escindits i es mantenen degenerant-se al llarg del temps a causa de les mutacions puntuals (Deininger & Batzer 1993).



**Figura 3.2.** Model de l'amplificació dels SINEs per retrotransposició. L'extrem 3' de les seqüències dels SINEs deriven d'un LINE específic, el qual codifica per la transcriptasa inversa. La transcriptasa inversa reconeix sobre el transcrit del SINE la cua 3' compartida i permet la síntesi del cDNA i la inserció al nou lloc del genoma. Figura extreta de Shedlock et al. (2000).

Els llocs d'integració de la majoria de retroposons són rics en As, la qual cosa ha portat a la hipòtesi que la cua rica en Ts del DNA complementari del retroposó pot interaccionar de manera específica via aparellament de bases Watson-Crick amb els llocs d'inserció cromosòmica abans de la integració (Moos & Gallwitz 1983; Daniels & Deininger 1985; Nadir et al. 1996). Els retrotranscrits dels retroposons probablement experimenten una poliadenilació a l'extrem 3' similar a la dels mRNAs (Wahle 1995). En alguns casos la diana d'inserció utilitzada és altament específica; aquesta diana és reconeguda per una endonucleasa codificada per l'element. En altres casos, no és tant específica. Per mitjà d'hibridació *in situ* s'ha vist que les famílies de SINEs Alu, B1 i B2 s'acumulen predominantment en les bandes R que són riques en G+C, es repliquen abans i contenen la majoria dels gens *housekeeping* (Korenberg & Rykowski 1988; Boyle et al. 1990). Tanmateix s'ha vist que els Alus inserits més recentment es troben distribuïts a l'atzar per les bandes G i R en el cromosoma 19 (Arcot et al. 1998).

A partir d'evidències experimentals han sorgit dos models per explicar l'evolució dels SINEs: el model *master gene* (Shen et al. 1991) i el model *multiple source gene* (Matera et al. 1990). En el model *master gene* només un, o potser uns pocs, loci *masters* de SINEs molt actius són capaços d'amplificar i només donar lloc a còpies que no són actives. D'aquesta manera la taxa de mutació està molt lligada a l'activitat del gen *master* i augmentarà o disminuirà al llarg del temps en funció de les mutacions acumulades. Aquest model, doncs, prediu que el gen *master* té alguna funció per l'hoste o que està sota restriccions selectives per tal de mantenir la capacitat d'amplificació durant un període de temps evolutiu llarg. Oposadament a aquest model trobem el model de *multiple source gene* que proposa que les noves còpies generades també tenen el mateix potencial que la còpia *master* per propagar-se, a partir d'aquí les còpies poden ser múltiples fonts d'amplificació. Actualment, es creu que els SINEs s'amplifiquen d'acord amb aquest segon model i que es poden dividir en subfamílies, diferents fonts, que tenen la capacitat de propagar-se (Matera et al. 1990; Smit et al. 1996; Terai et al. 1998). La taxa d'amplificació dels gens font canviarà al llarg del temps evolutiu segons si les mutacions acumulades els desactiven més ràpidament o lentament (Schmid & Maraia 1992; Shedlock & Okada 2000).

### ***3.3 Els SINEs com a marcadors filogenètics***

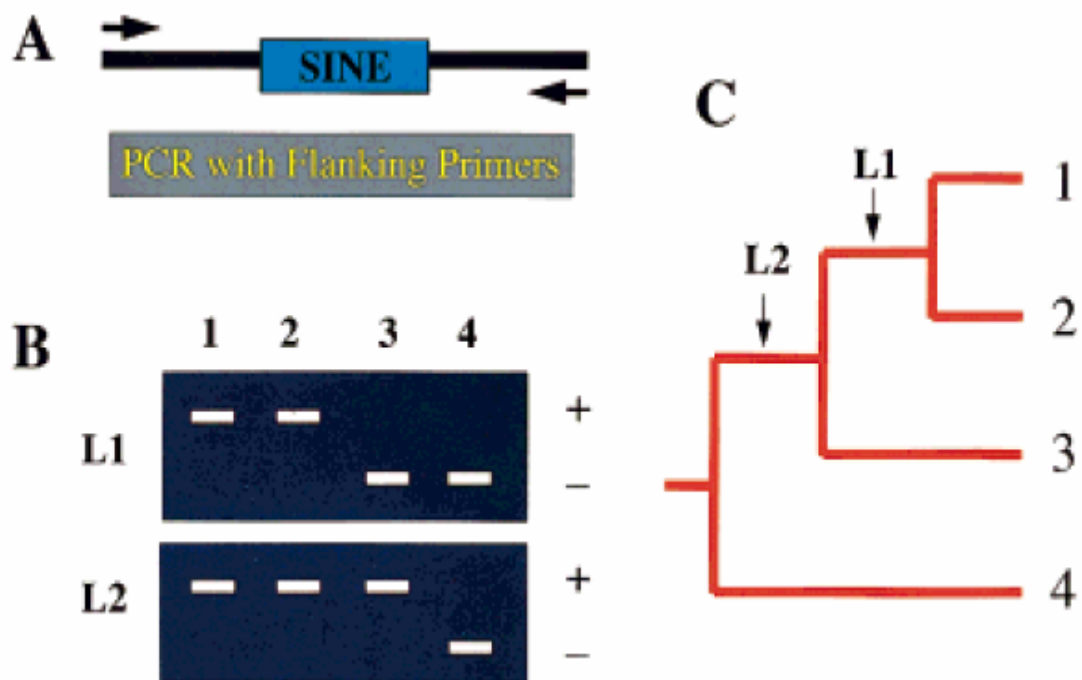
Els elements SINEs van ser en un principi detectats gràcies als estudis de renaturalització de DNA cromosòmic, que mostraven que la majoria d'organismes posseïen seqüències curtes, moderadament repetides i separades per llargs fragments de DNA de varies kb (Jelinek 1982). S'han identificat i caracteritzat SINEs en moltes espècies de mamífers (Smit & Riggs 1995). La majoria d'aquests es troben inserits en regions no funcionals i estan, en principi, lliures d'evolució paral·lela i mutacions reverses (Shedlock & Okada 2000). Aquestes característiques els fan ser una font excel·lent del registre de la història evolutiva i molt útils per inferències filogenètiques.

Els SINEs inserits en el mateix locus en diferent espècies probablement deriven a partir d'un element inserit en el genoma d'un antecessor comú, i d'aquesta manera ens indiquen els clades per a la reconstrucció filogenètica. La base d'aquestes inferències és que l'absència d'una inserció SINE en un locus determinat és l'estat ancestral. Aquest fet ens permet arrelar els arbres filogenètics sense fer referència a grups externs (*outgroups*), que són comunament utilitzats per inferir l'estat ancestral en els anàlisis amb substitucions nucleotídiques. Els experiments amb SINEs es basen en què els productes de PCR específics de locus que contenen una inserció SINE són més llargs que aquells productes que no tenen la inserció per una diferència de mida corresponent a la del SINE en qüestió (Fig. 3.3). Normalment, aquest patró és confirmat per seqüenciació dels productes de PCR amplificats. També ha de quedar clar que la no amplificació no és el mateix que l'absència del SINE.

És essencial que els SINEs estiguin fixats pel seu ús com a caràcters informatius pel diagnòstic d'ancestralitat comú entre espècies. Tanmateix, encara que no estiguin fixats, aquests també són idèntics per descendència. D'aquesta manera, els SINEs també són utilitzats per establir genealogies per sota del nivell d'espècie amb un mínim d'assumpcions, comparat amb altres marcadors estàndards com els microsatèl·lits, substitucions nucleotídiques, RFLPs (*restriction fragment length polymorphism*). Aquests altres marcadors poden ser idèntics per estat, resultat de mutacions independents a diferents temps que no són heretades a partir d'un antecessor comú (Batzer & Deininger 2001). Com a exemples més destacats, s'han utilitzat els SINEs per resoldre relacions internes entre ordres de mamífers i l'origen i la radiació de les balenes (Shimamura et al. 1997; Nikaido et al. 1999). Aquests retroposons també han servit per reconstruir la història demogràfica de les poblacions humanes (Stoneking et al. 1997; Watkins et al. 2003) i per inferir filogènies de primats (Schmitz et al. 2002; Salem et al. 2003).

La majoria de SINEs s'insereixen en regions no funcionals per tot el genoma sense implicar cap avantatge o desavantatge evident en l'hoste. Aquest fet fa que es considerin generalment marcadors genètics neutres. D'aquesta manera, la seva evolució pot ser modelada a partir de la teoria

neutra de la l'evolució i, per tant, el temps de fixació sota aquestes condicions generalment és més curt que el temps de separació entre dos llinatges. En aquests casos, la reconstrucció a partir de SINEs permet obtenir un arbre filogenètic robust. Si les espècies divergeixen en un procés ràpid abans que la fixació dels al·lels s'hagi produït – p.e., una radiació sobtada –, formes alternatives dels gens poden fixar-se en diferent llinatges de manera que no es reflecteix la veritable història de l'especiació i indica un polimorfisme ancestral.



**Figura 3.3.** Esquema de la detecció de SINEs per mitjà de la PCR. A: Encebadors, representats per fletxes, són dissenyats a la regió flanquejant del SINE. Els productes de PCR per diferents locus (L1 i L2) amplificats en diferents tàxons (1 al 4) són analitzats en un gel d'electroforesi per detectar presència (+) o absència (-) de la inserció SINE (B). C: Cladograma corresponent al patró de bandes de B. Cal notar que no és necessària la presència d'un outgroup per poder arrelar l'arbre ja que l'absència de SINE es considera l'estat ancestral. (Figura extreta de Shedlock & Okada 2000).

Com hem dit anteriorment, els SINEs pràcticament no presenten evolució paral·lela ni mutacions reverses. La probabilitat que dos insercions independents passin exactament en el mateix locus és essencialment zero

– p.e., 0.5% de 1336 Alu analitzats (Shedlock et al. 2004). Tot i que és estrany que un element SINE es deleccioni, quan ho fa queda una signatura molecular en la seqüència flanquejant del locus que pot ser detectada experimentalment a nivell nucleotídic (Edwards & Gibbs 1992). Tanmateix, tot i l'aparent idoneïtat de l'ús dels elements SINEs com a marcadors filogenètics, s'ha obert un debat sobre els límits del seu ús en relació a la seva evolució i a la freqüència dels patrons excepcionals – p.e., escissions precises o evolució paral·lela. Potser el major problema de l'ús dels SINE és que és necessari realitzar un esforç elevat per trobar suficients loci informatius per completar grans filogènies. A més a més, actualment encara és més barat i més fàcil comparar seqüències de DNA que aïllar i caracteritzar nous SINEs informatius. També, s'han trobat certes espècies on sembla haver-hi una falta de SINEs – p.e., la gallina i la *Drosophila* (Kazazian 2004). Un nombre baix de SINEs pot complicar l'ús pràctic d'aquesta eina en aquests organismes (Okada et al. 2004). Un altre problema dels SINEs és que, per causa de la degeneració dels SINEs per l'acumulació de mutacions puntuals, la seva detecció és difícil després de 150-200 m.a..

### ***3.4 Aïllament de nous SINEs***

Les caracteritzacions inicials de nous loci de SINEs estaven basades en comparacions de famílies ja estudiades i compilacions de seqüències de tRNAs en mamífers (Daniels & Deninger 1985; Sakamoto & Okada 1985). Tanmateix, aquesta informació era limitada i, amb l'objectiu d'ampliar el rang de famílies de SINEs, Endoh & Okada (1986) van desenvolupar un mètode més eficient per aïllar i caracteritzar nous SINEs a partir de diferents genomes: transcripció del DNA genòmic total *in vitro*. Aquesta tècnica permet discernir l'origen específic del tRNA de la nova família repetitiva. A més a més, ajuda en l'ús dels SINEs com a marcadors filogenètics, ja que amplia el nombre de tàxons a partir dels quals podem buscar nous loci útils en sistemàtica. Els SINEs són rarament transcrits *in vivo*, però, en canvi, aquests poden ser transcrits *in vitro* (Endoh & Okada 1986). Per a moltes espècies, la transcripció genòmica total dona com a



resultat un transcrit diferenciat que pot ser utilitzat com a sonda d'hibridació per aïllar fragments de DNA que contenen SINEs, clonats a partir d'una llibreria genòmica de l'espècie d'interès. Aquest assaig permet subclonar i seqüenciar, amb encebadors del vector, moltes noves famílies de SINEs i la seva regió flanquejant. L'estructura similar al tRNA i les regions promotores per l'RNA polimerasa III dels SINEs són essencials per distingir els nous SINEs d'altres elements repetitius del genoma (Shedlock & Okada 2000). Aquests nous elements són classificats en subfamílies segons mutacions diagnòstic en la seqüència de DNA. A més a més, la regió flanquejant adjacent als elements inserits pot ser utilitzada per dissenyar encebadors específics de locus per realitzar PCRs i utilitzar-los per a reconstruccions filogenètiques (vegeu apartat anterior).

Si ja tenim caracteritzada la família de SINEs en l'espècie o tàxon d'interès amb què poder dissenyar sondes marcades, podem construir una llibreria genòmica de l'espècie i detectar amb les sondes nous loci que continguin elements de la mateixa família. Actualment existeixen moltes eines de gran capacitat d'anàlisi que ens permeten realitzar projectes de seqüenciació de genomes complets i de les quals podem treure avantatge a l'hora de cercar nous SINEs. D'aquesta manera els SINEs poden ser detectats a partir de grans quantitats de seqüència primària utilitzant eines informàtiques – p.e., el programa RepeatMasker (Smit et al. no publicat) – que utilitzen bases de dades d'elements repetitius – p.e., Repbase Update (Jurka 2000) – per fer cerques per similitud de seqüències. Si assumim per evidències experimentals que una família típica de SINEs es troba present en més de  $10^4$  còpies en el genoma hoste, una aproximació a partir d'eines informàtiques hauria d'incloure sobre l'ordre de 0.6 Mb o més de seqüències genòmiques per capturar de manera fiable nous SINEs (Shedlock et al. 2004). La quantitat mínima de dades necessàries per aquesta aproximació perquè sigui efectiva varia segons l'estratègia de seqüenciació, l'estructura genòmica i el número de còpies dels SINEs.





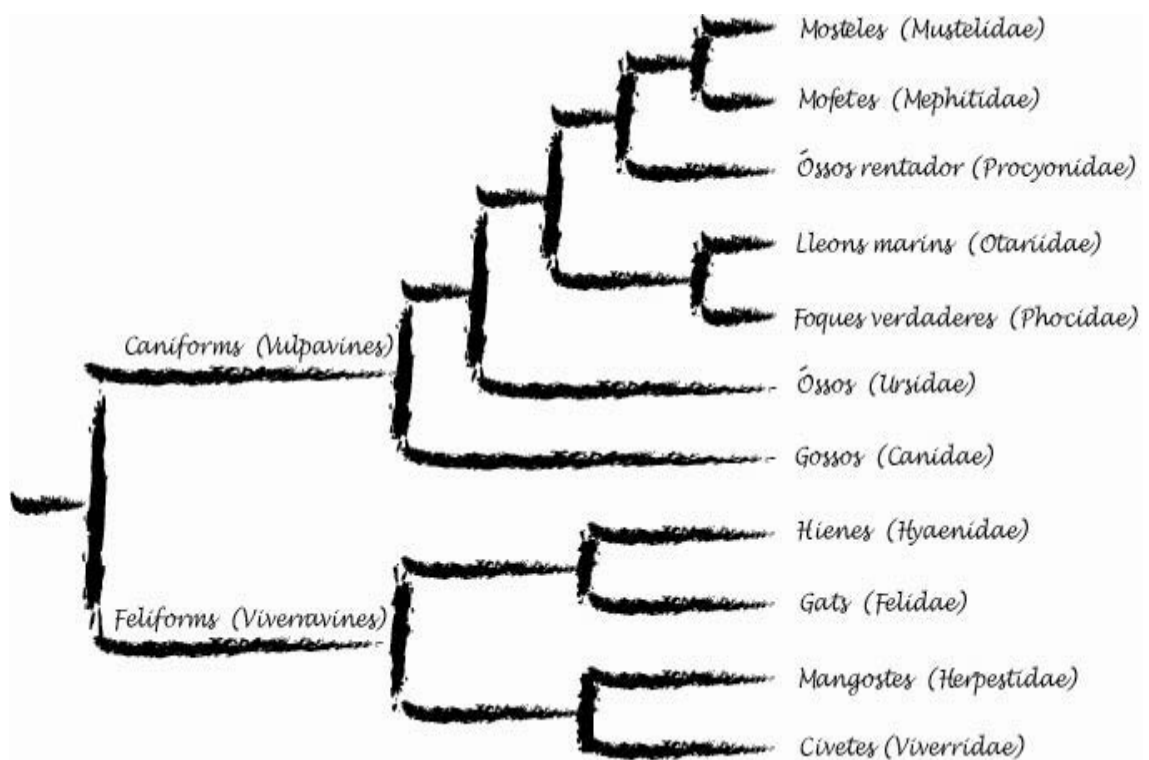
## 4 **ELS CARNÍVORS I L'ÚS DE TÈCNiques MOLECULARS PER A LA SEVA CONSERVACIÓ**

### **4.1 Els carnívors**

Els carnívors són un ordre de mamífers format per una col·lecció diversa de 271 espècies vives (segons Wozencraft 1993) que es troben en cada continent i ocupen hàbitats que van des dels boscos humits fins als deserts. La majoria dels carnívors són terrestres, però hi ha un grup molt especialitzat, els pinípeds (foques, lleons marins i morses), que fan vida als oceans. Uns pocs altres carnívors també viuen estretament lligats al medi aquàtic de manera més o menys especialitzada (p.e., la llúdriga marina, *Enhydra lutris*, també està especialitzada a la vida marina). A més a més, trobem marcades diferències en la riquesa d'espècies dins de cada família (p.e., hi ha 65 espècies de mustèlids i només 4 de hiènids). Els carnívors tenen una àmplia diversitat de mides que poden variar fins a 4 ordres de magnitud (més que qualsevol altre ordre de mamífers), però també són diversos en la seva fisiologia, estructura social, i, tot i el seu nom, en la seva alimentació. La seva dieta consisteix principalment en altres animals, però, dins d'aquest ordre, existeixen gran quantitat d'espècies que presenten una forma d'alimentació omnívora, frugívora o completament insectívora. Aquesta varietat de dietes ha donat al grup la capacitat d'adaptar-se a diferents tipus d'hàbitats i estils de vida. Una característica important del grup dels carnívors és el tipus de dentadura que presenten, denominada carnasial, que els ajuda a tallar i esquinçar amb més eficiència la carn de les preses capturades.

Des dels últims 50 milions d'anys aproximadament, els carnívors han estat separats en dos grans branques (Fig. 4.1), els Viverravines, que inclouen tots els parents del gat (*Felis domesticus*; feliforms); i els Vulpavines, que inclouen els parents del gos (caniforms). Avui en dia els feliforms estan representats per 4 famílies: Viverridae (civetes), Felidae (gats), Hyanidae (hienes) i Hesperistidae (mangostes). Hi ha més debat sobre quantes famílies formen els caniforms, generalment en són reconegudes 4:

Canidae (gossos), Ursidae (óssos), Mustelidae (mosteles) i Procionidae (ossos rentadors), tot i que alguns taxonomistes separen el panda vermell (*Ailurus fulgens*) en una família a part, Ailuridae. Encara hi ha altres relacions filogenètiques dins dels carnívors que es mantenen controvertides (p.e., la posició del panda vermell, els cànids, els fèlids i els hiènids) o que no han estat estudiades en gran detall (p.e., herpèstids, vivèrrids, i relacions intragenèriques dels procionids; vegeu capítol II de Resultats per a més detalls).



**Figura 4.1.** Divisió bàsica del les diferents famílies de carnívors segons Macdonald (2001).

La relacions entre els humans i els carnívors són molt variades. Per una banda trobem que el gos i el gat estan en tots els racons del planeta i, en canvi, certes espècies salvatges es troben greument amenaçades o en procés de desaparició. Es considera que el gos va ser el primer animal domesticat, fa més de 15000 anys, a partir del llop (Vila et al. 1997; Savolainen et al. 2002). En canvi, la domesticació del gat es creu que es va produir no més enllà de fa 4000 anys (Malek 1993). Actualment,



aquestes dues espècies amenacen els seus parents salvatges per mitjà de la hibridació i de la disseminació de malalties (Macdonald 2001). També trobem que espècies com el visó americà (*Mustela vison*), la foca (*Phoca vitulina*) i la guineu (*Vulpes vulpes*), entre d'altres, han estat àmpliament caçades i criades en granges pel seu ús en pelleteria. A més a més, la mala reputació d'alguns carnívors, generalment exagerada, com a depredadors de l'home o d'animals de granges ha justificat la seva caça indiscriminada. Mentre avui dia a Rússia encara es paguen recompenses pels llops (una femella amb cries es paga a 200 rubles), els conservacionistes, al mateix temps, estan lluitant per assegurar la supervivència de relictos poblacionals minúsculs – a Egipte al voltant de 30 individus i a Noruega menys de 10 (Macdonald 2001). Probablement el cas més conegut de carnívor amanegat d'extinció és el del linx ibèric (*Lynx pardina*), considerat el carnívor i el fèlid més amenaçat a Europa i al món, respectivament (Nowell & Jackson 1996). A més a més, existeixen altres espècies no tan populars però també greument amenaçades (vegeu taula 4.1). La caça i el trampeig de carnívors salvatges per obtenir-ne les pells i la destrucció dels hàbitats arreu del món estan conduint moltes espècies de carnívors a l'extinció i deixant en mans de l'home el seu destí.



**Taula 4.1.** Espècies de carnívors extintes en estat salvatge (EW), en perill crític (CR) o en perill (EN). La categoria d'amenaça està codificada d'acord amb la llista vermella d'espècies amenaçades publicada per la Unió Internacional per la Conservació de la Natura (UICN) al 2006.

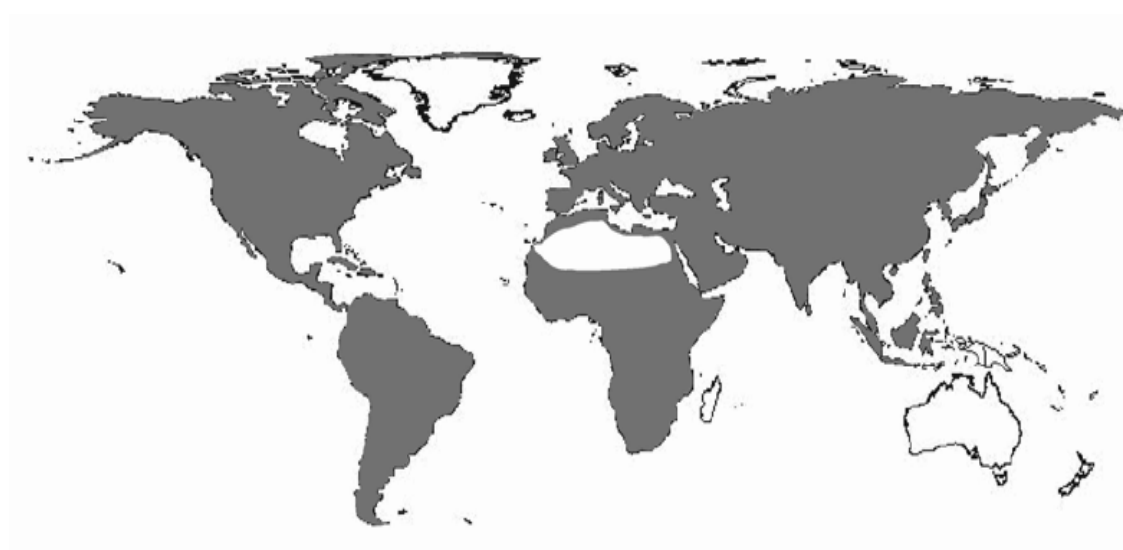
Família	Nom científic	Nom vulgar	Categoria	Distribució	
Mustelidae	<i>Enhydra lutris</i>	Liúdriga marina	EN A1ace	Costa pacífica nord-est i nord-oest	
	<i>Lontra felina</i>	Liúdriga de mar	EN A3ce	Costa pacífica d'Amèrica del Sud	
	<i>Lontra provocax</i>	Liúdriga de Xile	EN A3c	Argentina i Xile	
	<i>Mustela felipei</i>	Mustela de Colòmbia	EN B1+2ce	Colòmbia i Equador	
	<b><i>Mustela lutreola</i></b>	<b>Visó Europeu</b>	<b>EN A1ace</b>	<b>Vegeu text</b>	
	<i>Mustela lutreolina</i>	Mustela d'Indonèsia	EN B1+2c	Indonèsia	
	<i>Mustela nigripes</i>	Turó de peus negres	EW	Canadà i USA	
	<i>Pteronura brasiliensis</i>	Liúdriga gegant	EN A3ce	Amèrica del Sud	
	<i>Ailurus fulgens</i>	Panda vermell	EN C2a	Sud-est d'Àsia	
	Procyonidae	<i>Bassaricyon lasius</i>	Olingo costa-riqueny	EN D	Costa Rica
<i>Bassaricyon pauli</i>		Olingo fulvo	EN D	Panamà	
<i>Nasua nelsoni</i>		Coatí de l'Illa de Cozumel	EN D	Illa de Cozumel (Mèxic)	
<i>Procyon insularis</i>		Ós rentador de Tres Marias	EN D	Mèxic	
<i>Procyon maynardi</i>		Ós rentador de Bahames	EN C2a	Bahames	
<i>Procyon pygmaeus</i>		Ós rentador de l'Illa de Cozumel	EN C2a	Illa de Cozumel (Mèxic)	
<i>Procyon minor</i>		Ós rentador de Guadeloupe	EN C2b	Guadeloupe	
Otariidae		<i>Eumetopias jubatus</i>	Lleó marí d'Steller	EN A1b	Costa Pacífica nord-est i nord-oest
Phocidae		<i>Monachus monachus</i>	Foca monja del mediterrani	CR C2a	Mediterrani i Mar negre
		<i>Monachus schauinslandi</i>	Foca monja hawaiana	EN C2a	Hawaii



Ursidae	<i>Ailuropoda melanoleuca</i>	Ós panda	EN B1+2c, C2a	Xina	
Canidae	<i>Canis rufus</i>	Llop vermell	CR D	Amèrica del Nord	
	<i>Canis simensis</i>	Llop etíop	EN C2a(i); D	Etiòpia	
	<i>Cuon alpinus</i>	Gos salvatge asiàtic	EN C2a(ii)	Centre i Sud d'Àsia	
	<i>Lycaon pictus</i>	Licaó	EN C2a(i)	Sabana africana	
	<i>Pseudalopex fulvipes</i>	Guineu de Darwin	CR C2a(ii)	Xile (Los Lagos)	
	<i>Urocyon littoralis</i>	Guineu gris	CR A2be+3e	Califòrnia	
Felidae	<i>Catopuma badia</i>	Gat vermell de Borneo	EN C2a(ii)	Borneo	
	<i>Lynx pardinus</i>	Linx ibèric	CR C2a(i)	Sud de la Península Ibèrica	
	<i>Oreailurus jacobita</i>	Gat del Andes	EN C2a(i)	Andes	
	<i>Panthera tigris</i>	Tigre	EN C2a(i)	Índia, Manxúria, Xina i Indonèsia	
	<i>Uncia uncia</i>	Pantera de les neus	EN C2a(i)	Costa d'Ivori, Guinea i Libèria	
	Herpestidae	<i>Cryptoprocta ferax</i>	Gat fossa de Madagascar	EN C2a	Madagascar
		<i>Eupleres goudotii</i>	Fanaloca	EN C2a	Madagascar
		<i>Galidictis grandidieri</i>	Mangosta de ratlla gegant	EN B1+2b	Madagascar
		<i>Herpestes palustris</i>	Mangosta de Bangladesh	EN B1+2abcd	Est de l'Índia
		<i>Liberictis kuhni</i>	Mangosta de Libèria	EN B1+2c	Libèria
Viverridae	<i>Mungotictis decemlineata</i>	Mangosta de ratlla estreta	EN B1+2bc	Madagascar	
	<i>Cynogale bennettii</i>	Cibeta llúdriga	EN A1ce, C2a	Sud-est d'Àsia	
	<i>Genetta cristata</i>	Geneta crestada	EN B1+2c	Camerun i Nigèria	
	<i>Viverra civettina</i>	Cibeta de Malabar	CR C2a	Malabar (Índia)	

## 4.2 Els mustèlids: el teixó comú i el visó europeu

Els mustèlids (família Mustelidae) són un exemple d'èxit adaptatiu dins dels carnívors. Les aproximadament 57 espècies descrites dins d'aquesta família es distribueixen de forma natural arreu de cinc continents (Figura 4.2), excepte a les Índies occidentals, la major part de Groenlàndia i del Sàhara, Islàndia, Madagascar, Sulawesi i illes orientals, la major part de les Filipines, Nova Guinea, Austràlia, Nova Zelanda, l'Antàrtida i la majoria d'illes oceàniques menors (Nowak 1999; Macdonald 2001). Els mustèlids són carnívors de mida petita a mitjana – de 15 a 123 cm de longitud (sense considerar la cua) i de 30 a 45000 g de pes – i la majoria d'espècies presenten un cos allargat i prim amb extremitats curtes amb cinc dits a cada peu i ungles no-retràctils (Macdonald & Barrett 1993). Segons una de les classificacions més recents, recollida a Macdonald (2001), la família Mustelidae està dividida en cinc subfamílies: Mustelinae (p.e., la mostela), Lutrinae (p.e., la llúdriga comuna), Melinae (p.e., el teixó comú), Taxidiinae (p.e., el teixó americà) i Mellivorinae (p.e., el ratel).

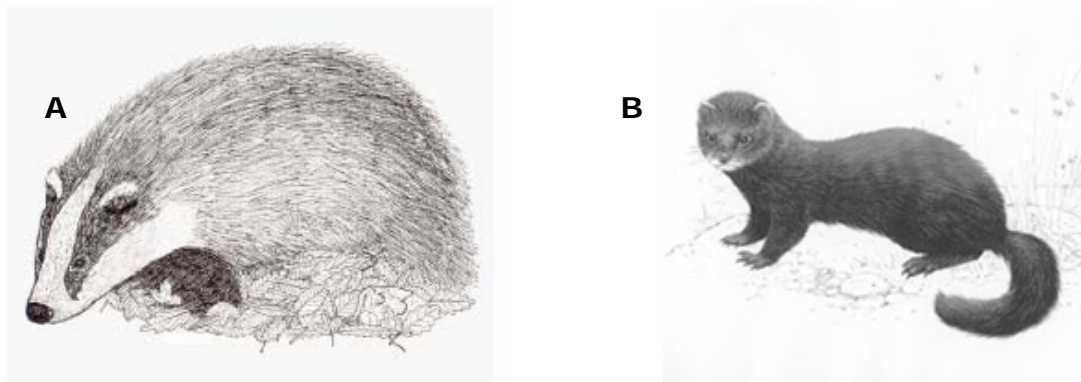


**Figura 4.2.** Distribució mundial, en gris, dels mustèlids. A més a més, la mostela comuna (*Mustela nivalis*), l'ermini (*Mustela erminea*) i el turó europeu (*Mustela putorius*) han estat introduïts per l'home a Nova Zelanda.

El registre fòssil de la família Mustelidae no és tan complet com el d'altres grups de mamífers ja que els mustèlids són animals de mida petita i ossos fràgils, la qual cosa dificulta la seva fossilització. Malgrat tot, el material descobert i estudiat ha permès fer un esborrany de la història evolutiva d'aquesta família, que s'està millorant notablement a partir de la informació proporcionada pels nombrosos estudis filogenètics realitzats en els darrers anys (vegeu capítol 2 dels Resultats per a més detalls).

La UICN ha classificat 22 espècies i 7 subespècies de mustèlids com a amenaçades. Dins de Mustelinae, el turó de peus negres, *Mustela nigripes*, natural d'Amèrica del Nord, està extingit de la natura i tan sols sobreviu en captivitat (Taula 4.1). A més a més, hi ha tres espècies més en perill d'extinció (la mostela d'Indonèsia, *Mustela lutrolina*; el visó europeu, *Mustela lutreola*; i la mostela colombiana, *Mustela felipei*) i tres més de classificades com a vulnerables (la mostela de dors llistat, *Mustela strigidorsa*; la marta de Nilgiri, *Martes gwatkinsi*; i el golut, *Gulo gulo*). Dins de Melinae hi ha dues espècies vulnerables (el teixó turó d'Everett, *Melogale everetti*; i el teixó pudent de Palawan, *Mydaus marchei*) i una en baix risc d'extinció (teixó turó oriental, *Melogale orientalis*). La subfamília Lutrinae és la que té més espècies amenaçades: quatre espècies de llúdrigues estan en perill d'extinció (llúdriga de mar, *Enhydra lutris*; llúdriga marina, *Lontra felina*; llúdriga de riu del sud, *Lontra provocax*; llúdriga gegant, *Pteronura brasiliensis*), tres són espècies vulnerables (llúdriga comuna, *Lutra lutra*; llúdriga de coll tacat, *Lutra maculicollis*; i llúdriga de pèl suau, *Lutrogale perspicillata*) i una està en baix risc d'extinció (llúdriga cendrosa, *Amblyonyx cinereus*). Moltes de les espècies restants d'aquesta família estan protegides o en perill d'extinció en determinats països dins de la seva àrea de distribució. En general, les principals amenaces per a aquestes espècies són la destrucció del seu hàbitat, la desaparició de les seves preses principals i la sobreexplotació (p.e., per proveir la indústria de la pelleteria). L'elevat nombre d'espècies amenaçades que presenten els mustèlids fa necessari tenir una filogènia i una taxonomia ben resoltes en aquesta família. Aquesta informació, juntament amb l'aportada per altres disciplines – com per exemple

l'ecologia –, pot ser molt útil i necessària per prioritzar i guiar plans de conservació dels mustèlids en el futur.



**Figura 4.3.** El teixó comú (A; *Meles meles*) i el visó europeu (B; *Mustela lutreola*).

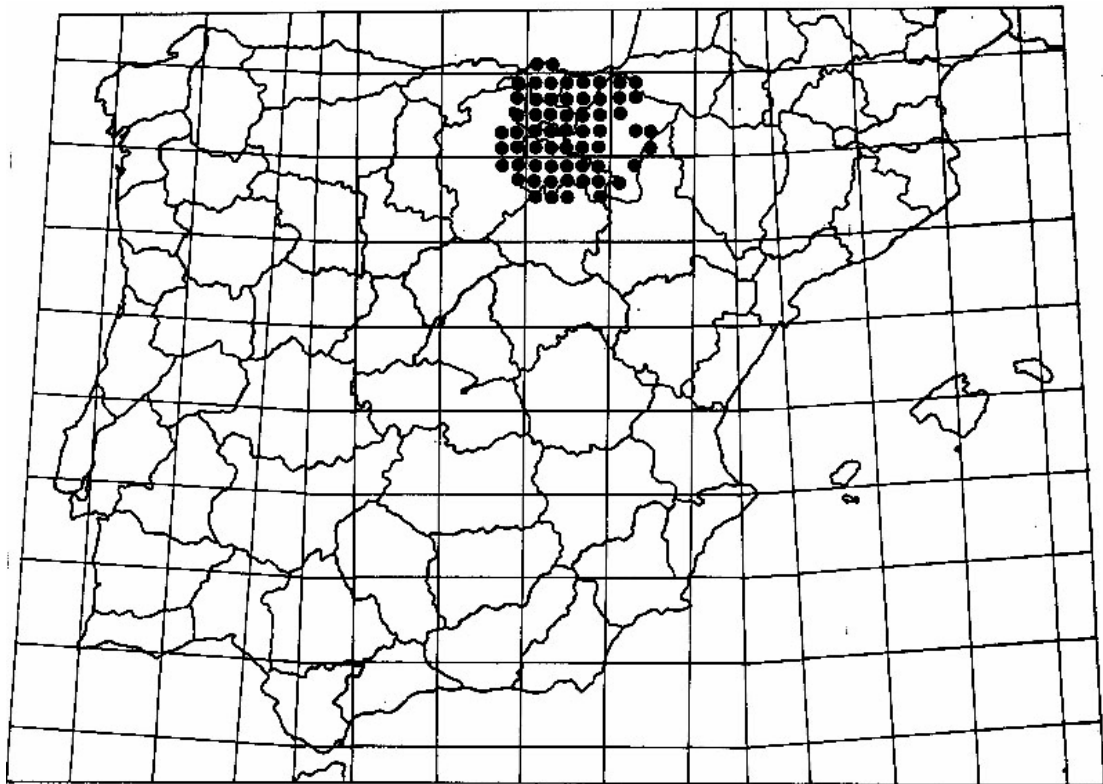
El teixó eurasiàtic o comú (Fig. 4.3; *Meles meles*) és un mustèlid de mida mitjana a gran, mesura de 67 a 81 cm de longitud (excepte la cua) i pesa de 10 a 12 kg (Macdonald 2001). És un animal robust i camacurt amb pelatge llarg i fort que presenta una coloració molt característica: el cap és de color blanc amb dues bandes negres laterals marcades des del morro fins a les orelles; el cos és grisós excepte el pit, el ventre i les extremitats, que són negres. El teixó comú és una de les espècies de mustèlids que té una àrea de distribució més extensa. A Europa, malgrat que el nombre d'individus de moltes poblacions roman estable o en augment, hi ha alguns països – com per exemple els Països Baixos, Albània, l'antiga Iugoslàvia i Grècia, concretament les illes de Creta i Rodes – on els teixons cada vegada són més rars o estan amenaçats d'extinció (Neal & Cheeseman 1996; Griffith & Thomas 1997). El teixó està inclòs a l'Apèndix III del Conveni de Berna i està protegit per lleis nacionals a: Portugal, Espanya, Itàlia, Irlanda, Gran Bretanya, Luxemburg, Bèlgica, Països Baixos, Albània, Grècia, Hongria i Estònia (Griffith i Thomas 1997). En àrees rurals, principalment de l'Europa de l'Est, els teixons són caçats per aprofitar-ne la carn, el greix (utilitzat pel tractament del dolor muscular), el pelatge i la pell. A partir dels seus pèls també s'elaboren brotxes d'afaitar de luxe (Domingo-Roura et al. 2006). Als països on no està



protegit, el teixó està considerat com una espècie de caça menor o fins i tot com un animal nociu (Neal & Cheeseman 1996). Malgrat tot, un dels majors problemes que té aquest animal a Europa és que ha estat identificat – a països com Irlanda i la Gran Bretanya – com el principal hoste en estat salvatge de *Mycobacterium bovis*, l'agent causant de la tuberculosi bovina (Gallagher & Clifton-Hadley 2000). Aquesta malaltia, que és endèmica d'Europa, causa importants pèrdues econòmiques en les explotacions de bestiar boví. Per controlar la malaltia, entre 1975 i 1982 una gran quantitat de teixons van ser exterminats a Anglaterra mitjançant cianur, però aquestes campanyes es van aturar perquè es consideraven inhumanes i no permetien una anàlisi *post-mortem* dels animals, ja que morien dins dels seus caus. La següent estratègia que ha seguit el Ministeri d'Agricultura, Pesca i Alimentació d'Anglaterra (MAFF) és capturar els animals mitjançant trampes i matar-los a trets (Neal & Cheeseman 1996). Aquest pla pel control de la tuberculosi bovina, que també suposa l'eliminació d'un gran nombre d'animals, en molts casos ha demostrat ser molt poc efectiu per controlar aquesta malaltia (Tuytens et al. 2000).

El visó europeu (Fig. 4.3; *Mustela lutreola*) és un mustèlid de mida mitjana amb un cos llarg, esvelt i arquejat, amb cames curtes i una cua curta i atapeïda, que mesura de 28 a 43 cm de longitud (excepte la cua) i pot arribar a pesar fins a 1 kg (Burten 2002). El pèl normalment és marró fosc tirant a negre amb una taca de pèl blanc distintiva al voltant dels llavis superior i inferior i ocasionalment també en el coll (Burten 2002). Aquesta última característica i la seva mida més petita són sovint utilitzades per distingir-lo del visó americà, però a vegades aquestes espècies poden ser tan semblants que només l'esquelet o les anàlisis genètiques poden garantir una correcta identificació (Macdonald 2001). El visó europeu és un dels carnívors més amenaçats d'extinció (Baillie & Groombridge 1996). A finals dels segle XX la seva distribució es va fragmentar i el declivi poblacional ha continuat des de llavors (Van Bree & Saint Girons 1966; Maran 1992; Tumanov 1992; Sidorovich 2000). Ara, l'espècie es troba en la part est de Bielorússia, en algunes zones de Rússia (Sidorovich 2000; Wolsan 1993) i en el delta del Danubi a Romania (Gotea & Kranz 1999). A l'oest, només hi ha dades de la seva presència a

7 departaments de la part més sud-oest de França (Maizeret et al. 2002) i en la part més alta de la vall del riu Ebre (Fig. 4.6; Ruiz-Olmo & Palazón 1991). A La Rioja, segons el cens de 1998, la població de visó europeu s'estima en 430 individus, el que representa més del 40% dels efectius presents a Espanya i el 20% dels de la Unió Europea (Boletín Oficial de La Rioja 2002). En canvi, la seva distribució històrica s'estenia des de Finlàndia fins a l'est dels Urals, nord d'Espanya i el Caucas. L'espècie va començar a extingir-se en algunes parts del centre d'Europa ja fa un centenar d'anys. Durant aquest últim segle, les poblacions s'han vist reduïdes en tota la seva distribució.



**Figura 4.4.** L'àrea de distribució del visó europeu a la península ibèrica (*Mustela lutreola*) està limitada a Navarra, La Rioja, País Basc, nord de Soria i nord-est de Burgos. Figura extreta de Palazón (2003).

**Taula 4.2.** Diferents factors utilitzats com a hipòtesi pel declivi del visó europeu.

Factor	Descripció
Pèrdua d'hàbitat	El drenatge de la terra, la canalització de rius i la deforestació són activitats humanes que des de finals dels segle XVIII han canviat l'ecosistema tan especialitzat del visó europeu (Danilov & Tumanov 1976).
Sobrecaça	Des dels primers moments el visó europeu ha patit una caça excessiva per la seva preuada pell. A principis del segle XX a l'antiga URSS es caçaven entre 40-60000 individus per any amb un rècord de 75000 (Novikov 1939).
Pol·lució	Tot i que no hi ha dades experimentals, es creu que la pol·lució ha estat una causa constant del declivi (Schropfer & Paliocha 1989)
Impacte climàtic	El període temperat dels anys 30 a Finlàndia podria haver estat desfavorable pel visó europeu (Voipio 1946)
Malaltia introduïda	El visó americà és portador i agent transmissor d'un Parvovirus que provoca la Malaltia Aleutiana del Visó (ADV), on una infecció crònica està associada a la disminució de la fertilitat i provoca aborts espontanis (Mañas et al. 2001).
Dependència d'una presa en declivi	Disminució de la presa principal del visó europeu, el cranc de riu, <i>Astacus astacus</i> (Henttonen 1992)
Relacions interespecífiques amb el visó americà	Gràcies a la seva major mida, el visó americà exclou l'europeu (Maran 1991). A més a més, el visó americà es reproduïx abans, de manera que pot deixar prenyades les femelles de l'europeu amb embrions que després són reabsorbits (Ternovskij 1977)
Relacions interespecífiques amb el turó europeu	S'han trobat híbrids salvatges de turó i visó europeu. A més a més, els canvis ambientals afavoreixen els hàbitats del turó (Schopfer & Paliocha 1989).
Predació i agressió intragremi	La desaparició de grans depredadors com el llop, fa que la guineu augmenti en nombre, dins la dieta de la qual entren mustèlids (Lindstrom et al. 1995).

S'han barrejat diverses hipòtesis per explicar la desaparició de l'espècie (Taula 4.2), però el que sembla clar és que cap dels factors hipotetitzats per separat podria explicar els esdeveniments de declivi en varis llocs en diferents moments (Maran & Henttonen 1995). Sembla que l'actual disminució del visó europeu a Estònia coincideix bé amb la presència del visó americà (Maran 1991), el qual va ser introduït a Europa a principis dels anys 20 i 30 per la seva cria en granges i, per fugides o alliberaments

intencionats, ha produït poblacions salvatges per tota Europa. Per altra banda, les disminucions inicials en el Centre d'Europa i més tard a Finlàndia van tenir lloc abans que aquesta espècie invasora s'escampés per Europa. Les disminucions inicials al centre d'Europa podrien haver estat causades per la destrucció dels ecosistemes del riu, especialment dels seus marges (Danilov & Tumanov 1976). Tot i així, a Finlàndia la major part de la desaparició es va produir bastant abans que els canvis ambientals dels ecosistemes dels rius. Tot i que les explicacions detallades sovint poden variar, hi ha un rerefons comú: el canvi mediambiental. Abans de l'arribada del visó americà, el canvi en l'hàbitat de preferència, petites rieres arenoses, o la desaparició de la font principal d'aliment podrien haver estat les principals causes del declivi. El visó europeu és una espècie més especialitzada que el visó americà i, amb la arribada d'aquest, l'europeu perd terreny, deixant de banda si hi ha canvi mediambiental o no. Actualment, el visó americà és la principal causa que porta a l'extinció, per diversos mecanismes (Taula 4.2), les poblacions remanents de visó europeu (Maran & Henttonen 1995, Maran et al. 1998, Sidorovich 2000).

### ***4.3 Genètica de la conservació***

Els científics que es dediquen a la conservació estudien els individus i poblacions que han estat afectades per pèrdua d'hàbitat, explotació, i/o canvis ambientals i la informació que extreuen és utilitzada per prendre decisions per assegurar la supervivència dels organismes en el futur. Si afegim les tècniques i coneixements genètics a la biologia de la conservació obtenim el que coneixem com a genètica de la conservació. Els estudis genètics proporcionen als científics que es dediquen a la conservació nous elements per entendre la diversitat entre els individus d'una població, sense els quals podríem conservar la població incorrecta o malgastar els recursos sobre una població que realment no està en perill. La genètica de la conservació, doncs, és una disciplina que aplica tècniques moleculars per tractar el manteniment, pèrdua i restauració de la diversitat. En aquesta disciplina es considera que una població o espècie

està en perill quan la diversitat genètica passa a ser baixa en molts gens dels individus que formen aquesta població o espècie. La importància de la genètica de la conservació queda clarament reflectida quan es veu el gran nombre d'espècies que estan en imminent perill d'extinció – p.e., la Llista Vermella d'espècies amenaçades de la UICN del 2006 presenta 1093 espècies de mamífers de les quals 79 són carnívors, aproximadament un 7.2% del total.

Les aplicacions en conservació fan referència a dos aspectes fonamentals: i) entendre els canvis genètics que afecten la supervivència de les espècies amenaçades i ii) proporcionar informació genètica utilitzable per a la millor gestió de les espècies en perill. Com que els qui es dediquen a la conservació d'espècies han de prendre decisions ràpides basades en les dades que tenen en el moment, els qui es dediquen a la genètica de la conservació busquen contínuament tècniques que detectin eficientment i ràpidament la variació genètica. En els últims anys s'han incorporat tècniques normalment utilitzades per a estudis sobre humans o espècies models – com per exemple la utilització d'SNPs (Zhang & Hewitt 2003; Aitken et al. 2004; Brumfield et al. 2003) i, fins i tot, s'ha contemplat la possibilitat d'utilitzar *microarrays* (Gibson 2002).

A continuació presento una llista de diferents exemples dels papers que pot jugar la genètica de la conservació:

- Resoldre l'estructura poblacional, la taxonomia i la identificació d'unitats de conservació: La identificació d'unitats de conservació és clarament un punt clau, tant en la genètica de la conservació com per altres del papers citats en aquesta llista. En permetre analitzar les relacions filogenètiques entre grups d'individus, els marcadors moleculars poden ser claus a l'hora de decidir si una població concreta mereix protecció. Els estudis filogenètics tenen aplicacions importants en la gestió i la conservació de les espècies. El reconeixement de categories sistemàtiques (subespècies, espècies, gèneres, famílies, etc.) aporta les bases legals per a la conservació (O'Brien 1994). L'espècie és el nivell taxonòmic bàsic en l'estudi de la diversitat biològica i de l'evolució i en la planificació de molts

programes de conservació, per tant, és molt important definir-ne els límits.

- Minimitzar la consanguinitat i la pèrdua d'heterozigositat: La reducció dràstica del nombre d'individus d'una població té conseqüències genètiques, que es sumen a les conseqüències demogràfiques i ecològiques. En teoria, els individus que pateixen alts nivells d'endogàmia, com aquells que es troben en poblacions molt petites i altament en perill, estan subjectes a la depressió per consanguinitat ja que l'homozigositat desemascararia mutacions deletèries. En petites poblacions la depressió per consanguinitat és molt comú pels efectes acumulatius de la deriva genètica (Hedrick & Kalinowski 2000).

- Documentar fenòmens d'hibridació i d'introgressió: Una de les amenaces per a espècies en perill és la pèrdua d'identitat genètica mitjançant hibridació extensiva amb grups relacionats.

- Aplicar la legislació mediambiental i utilització forense: Les anàlisis moleculars estan sent utilitzades com a eines per a la detecció de delictes contra el medi ambient, com és el tràfic d'espècies protegides o productes derivats. Un dels criteris per incloure espècies en el llistat d'espècies en perill de CITES (Convenció Internacional pel Comerç d'Espècies Amenaçades de Fauna i de Flora) és que el producte comercial derivat d'aquesta espècie ha de ser identificat (Vegeu l'apartat 4.4 per a més detalls).

- Definir llocs i genotips per a la reintroducció: La definició i estudi de diferents unitats de conservació és també important, de manera que la demografia i la genètica de la població utilitzada per a la reintroducció coincideixi amb aquelles unitats que estaven presents a l'àrea de reintroducció.

- Detectar i definir espècies invasores: Les espècies invasores estan provocant la pèrdua d'espècies i diversitat biològica en diversos ecosistemes de tot el món. Aquestes espècies poden envair l'ecosistema de diverses maneres, incloent la introducció voluntària sense la informació



adequada sobre l'espècie que s'està introduint ni de l'ecosistema receptor i la introducció accidental deguda al comerç, transport o turisme.

- Estimar la mida poblacional i el quocient sexual: La mida poblacional efectiva ( $N_e$ ) és el nombre d'individus que contribuiran amb els seus al·lels a la següent generació.  $N_e$  és gairebé sempre més petita que el nombre d'individus que formen la població. Aquests paràmetres ens ajudaran a entendre en quin grau de perill es troba la població. Per escapar de la retroalimentació entre el declivi genètic i demogràfic s'ha arribat a definir una grandària mínima perquè la població sigui viable (MVP; Shaffer 1981).

- Gestionar espècies en captivitat: La consanguinitat i altres processos demogràfics són amenaces que afecten les petites poblacions salvatges i també les poblacions en captivitat. S'ha vist que una de les causes de la reducció en l'eficàcia biològica de les poblacions en captivitat de les espècies amenaçades és la consanguinitat (Kalinowski et al. 2000).

Tanmateix, la genètica de la conservació es troba en un context on hi ha les dificultats de treballar a través de les barreres polítiques, els reptes econòmics i el fet que li cal afrontar la complexitat d'utilitzar la ciència per informar de les decisions de gestió. Malgrat el potencial que té la genètica de la conservació, hi ha experts que pensen que encara està en un estadi massa immadur per ser aplicada eficientment en el camp. De fet, encara no s'ha arribat a un consens sobre com incorporar la diversitat genètica en els plans de conservació (Moritz & Faith 1998). El paper de la genètica en la conservació de la biodiversitat ha estat sovint massa emfatitzat i s'han oblidat les limitacions teòriques i empíriques que presenta l'ús de la informació genètica per solucionar problemes ecològics i demogràfics (Domingo-Roura et al. 2001). En qualsevol cas, l'aproximació al problema de la conservació d'espècies o poblacions a partir d'anàlisis genètiques pot ser important per a l'èxit immediat de la població o espècie, però no podem oblidar que l'objectiu final de la biologia de la conservació hauria de ser la cura a llarg terme de les àrees on habiten els organismes en perill. En la gestió d'àrees la genètica de la conservació també pot jugar un paper molt important – p.e., definint àrees habitades per endemismes.

#### ***4.4 Genètica forense i genotipar sense capturar.***

Per definició, la genètica de la conservació ha de ser no invasora, per tant, el DNA de l'espècie d'estudi es recupera en el camp sense capturar, sense manipular o, fins i tot, sense veure l'animal. A més, hi ha algunes situacions en les quals capturar l'individu d'interès no és concebible: per exemple, en el cas de petites poblacions en perill o per estudis de comportament on la captura dels animals provocaria una distorsió en el sistema. En aquestes condicions, el mostreig no invasor es l'única manera d'aproximar el problema. Un altre factor a tenir en compte és que el mostreig no invasor pot augmentar el nombre d'animals que poden ser mostrejats en espècies reservades difícils d'observar – p.e., mustèlids –, i d'aquesta manera pot fer possible l'estima de paràmetres poblacionals importants.

L'ús de la PCR va obrir les portes a anàlisis sense la necessitat d'obtenir teixits d'animals vius (biòpsies) ja que permet obtenir una gran quantitat de DNA a partir de molt poques còpies inicials. No obstant, l'ús de mostres no invasives comporta una sèrie de limitacions ja que se n'obté una baixa quantitat de DNA i aquest és de baixa qualitat – p.e., DNA degradat, contaminació o presència d'inhibidors en la PCR. El DNA es comença a degradar a partir de la mort de l'individu a causa de la hidròlisi, l'oxidació, els efectes de les radiacions, la presència de microorganismes i les temperatures altes (Carter 2002). Darrerament, s'han dissenyat protocols per eliminar possibles inhibidors de PCR i evitar contaminacions, com són els mètodes d'extracció de DNA basats en sílica (Yang et al. 1998) o chelex (Singer-Sam et al. 1989). El perill de les contaminacions es pot minimitzar utilitzant material estèril o autoclavant els estris o solucions d'ús repetit, executant en habitacions separades els processos de pre i post-PCR, fent els experiments en laboratoris que no treballin habitualment amb els tàxons que volem diagnosticar i, evidentment, utilitzant controls negatius a les diferents fases del procés.

Actualment s'utilitza un ampli rang de fonts de DNA en genètica de la conservació, principalment femtes i pèls per la seva fàcil obtenció, però s'han utilitzat altres mostres més inusuals, com orina de llop a la neu

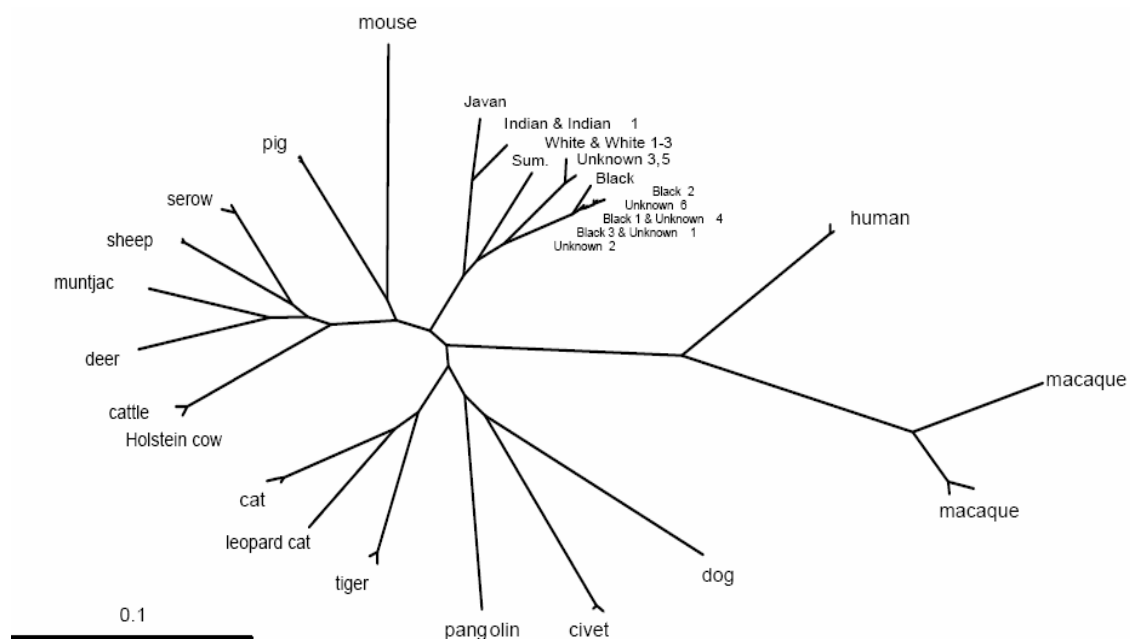


(Valiere & Taberlet 2000), cèl·lules bucals de ximpanzé obtingudes de restes de menjar mastegat (Sugiyama et al. 1993; Takenaka et al. 1993), mudes de pell de cetacis (Bricker et al 1996; Valsecchi et al. 1998) i d'ocells, material utilitzat per fer nius, cloves d'ou i orina (Morin et al. 1994; Pearce et al 1997; Nota & Takenaka 1999). Molts estudis utilitzen mostres de museus i mostres d'organismes extingits per detallar la genètica del passat d'espècies o poblacions en perill (Hofreiter et al. 2001). Aquests tipus d'estudis van des de mostres de museu de 100 anys d'antiguitat – p.e., varies espècies de carnívors (Hoelzel et al. 1993; Leonard et al 2000) – fins a la utilització de restes d'animals extingits fa milers d'anys – p.e., espècimens d'òssos (Paäbo 2000), peresós (Höss et al. 1996) i moas (Cooper et al. 1996). Aquestes tècniques, doncs, ens proporcionen una font de DNA que pot ser utilitzada no només per identificar l'espècie, sinó també individus i el sexe (Kohn & Wayne 1997; Taberlet et al. 1999). A més a més, proporcionen un gran potencial per fer estimacions d'abundància i determinar paràmetres de comportament, tals com àrea de distribució, hàbitat individual i preferències en dieta (Höss et al. 1992; Hofreiter et al. 2000; Symondson 2002), i algunes formes d'interacció social (Constable et al. 2001). També ens permeten substituir els mètodes clàssics d'obtenció de mostres que requereixen la captura de l'animal per fer estudis de parentesc, genètica de poblacions o anàlisis filogenètiques.

Tot i que hi ha una gran varietat de recursos per obtenir DNA de manera no invasora, encara és important aconseguir una col·lecció de teixits per obtenir DNA de qualitat. Actualment es fan esforços per emmagatzemar i arxivar mostres d'alta qualitat que provenen d'espècies en perill per preservar fonts de recursos genètics i, fins i tot, s'ha plantejat la possibilitat de clonar a partir de teixit d'un animal extingit (Lanza et al. 2000).

Un altre punt clau en conservació on la genètica forense pot intervenir és el seguiment del tràfic i comerç de productes derivats d'espècies en perill d'extinció, un problema que compromet el futur d'un gran nombre d'espècies. La majoria de vegades l'estat d'aquests productes està tan modificat que és pràcticament impossible saber a quina espècie pertanyen

a partir dels mètodes basats en el fenotip o el reconeixement visual dels productes. Per tant, la genètica forense pot ser una gran aliada a l'hora d'ajudar a complir les directrius de la CITES. La genètica forense ha estat decisiva en la detecció de comerç il·legal d'espècies en perill d'extinció. Per exemple, mitjançant seqüències de la regió control del DNA mitocondrial (mtDNA) s'ha demostrat el comerç il·legal de carn de balena amb gep, *Megaptera novaeangliae*, en mercats del Japó i de Corea del Sud (Baker & Palumbi 1994) i el comerç il·legal de brotxes d'afaitar fabricades amb pèls de teixó comú, *Meles meles*, a Europa (Domingo-Roura et al. 2006). Mitjançant seqüències del gen citocrom *b* s'ha demostrat la venda il·legal de carn de tigre, *Panthera tigris*, a la Xina (Wan & Fang 2003) i que sis escultures confiscades pel Departament d'Agricultura de Taiwan havien estat fabricades a partir de banyes de diferents espècies de rinoceront (Fig. 4.5; Hsieh et al. 2003).



**Fig 4.5.** Arbre filogenètic construït pel mètode de *neighbor-joining* a partir de seqüències parcials (402bp) del gen mitocondrial citocrom *b*. Les seqüències de diferents espècies de rinoceront extretes de Genbank estan indicades (Javan, Sumatran, Indian, black and white). Les mostres indicades com a desconegudes (unknown) serien les mostres de les 6 escultures que es trobarien agrupades i separades de la resta de mamífers a l'arbre. Figura extreta de Hsieh et al. (2003).

**OBJECTIUS**

Aquesta tesi tracta sobre l'estudi de la variació genètica en carnívors com a models per desxifrar característiques de l'evolució de les regions repetitives, per resoldre qüestions evolutives d'aquestes espècies i per les seves possibles aplicacions en genètica de la conservació. A continuació s'enumeren els principals objectius d'aquest treball:

1. *Estudiar els patrons evolutius complexos dels microsatèl·lits.*

Mitjançant la seqüenciació de les regions microsatèl·lits en teixons i altres mustèlids hem estudiat els patrons d'evolució d'aquestes regions repetitives dins d'aquesta família, ja que presenten abundants imperfeccions i patrons evolutius anormals. L'objectiu d'aquest apartat ha estat determinar l'adaptació d'aquests microsatèl·lits als models mutacionals que més sovint s'assumeixen en genètica de poblacions i determinar l'origen de la variabilitat que no pot ser detectada a través de procediments estàndards de genotipatge.

2. *Avançar en el coneixement dels processos mutacionals que originen els microsatèl·lits.*

En aquests treball, hem analitzat l'origen i diversificació d'un microsatèl·lit compost complexe treient avantatge de la conservació entre carnívors de la seva regió flanquejant. D'aquesta manera hem pogut observar els patrons d'evolució dels microsatèl·lits en una escala taxonòmica major. També hem abordat aquest objectiu analitzant el paper que poden tenir els retroposons en la generació i evolució dels microsatèl·lits de carnívors i com aquests poden afectar la distribució i el patró d'evolució de certs motius de repetició.



*3. Explorar l'aplicació en genètica de la conservació dels resultats obtinguts en les anàlisis de les regions repetitives.*

Mitjançant l'anàlisi de la informació de la variabilitat genètica continguda en les seqüències nuclears que flanquegen un microsatèl·lit – les quals presenten una baixa saturació de mutacions – hem intentat resoldre les relacions filogenètiques entre diferents espècies de carnívors. A més a més, hem creat un marcador molecular específic d'espècie barat i simple que pot ajudar en la gestió del visó europeu – un dels carnívors més amenaçats – a partir de mostres obtingudes de manera no invasiva.

# **RESULTATS**

# Capítol I

***High incidence of non-slippage  
mechanisms generating variability and  
complexity in Eurasian badger  
microsatellites***

Francesc López-Giráldez, Josep Marmi i Xavier Domingo-  
Roura

(en revisió per *Journal of Heredity*)



## HIGH INCIDENCE OF NON-SLIPPAGE MECHANISMS GENERATING VARIABILITY AND COMPLEXITY IN EURASIAN BADGER MICROSATELLITES

Francesc López-Giráldez<sup>1,2\*</sup>, Josep Marmi<sup>1,3</sup>, Xavier Domingo-Roura<sup>1,2</sup>

<sup>1</sup>*Genètica de la Conservació, Institut de Recerca i Tecnologia Agroalimentàries, Ctra. de Cabrils s/n, Cabrils (Barcelona), Spain.*

<sup>2</sup>*Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Dr. Aiguader 80, 08003 Barcelona, Spain.*

<sup>3</sup>*Departament de Ciència Animal i dels Aliments, Facultat de Veterinària, Universitat Autònoma de Barcelona, 08193 Cerdanyola del Vallès, Spain.*

\*Corresponding author: Genètica de la Conservació, Institut de Recerca i Tecnologia Agroalimentàries, Centre de Cabrils, Ctra. de Cabrils s/n, 08348 Cabrils (Barcelona), Spain. Tel: +34 937507511. Fax: +34 937533954. e-mail: [francesc.lopez@gmail.com](mailto:francesc.lopez@gmail.com)

---

**Abstract.** The use of microsatellites in population genetics is hindered by a lack of understanding of the pattern and origin of mutations, the need to develop more specific and better computational models, and a paucity of information about specific taxa and loci. We analyzed between four and ten allele sequences from ten different microsatellites in Eurasian badgers in order to determine the compliance of the sequences with step-wise mutation models and the origin of that variability which cannot be detected through standard genotyping procedures. All microsatellite loci exhibited imperfections and/or substitutions and indels in the flanking region, as well as additions or deletions of repeat units. Our dataset of sequences showed a higher number of imperfect repeats than other published badger and carnivore sequences. This could be attributed to the process of loci isolation, because when genetic variability is low, researchers may be more likely to use imperfect loci if these are variable in the population being studied. Locus *Me15* had two repetitive arrays: one was part of a polypyrimidine region of a CAN Short Interspersed Element (SINE) and the other was located in an A-rich region typical of these insertions. In spite of this complexity, heterozygosity was correlated with the maximum number of repeats. Thus, while new theoretical models are being evolved to cover complex patterns of microsatellite mutation, sequencing electromorphs is needed to identify microsatellites or portions of them whose evolution can be modeled under simple models.

---



## 1. Introduction

Microsatellites are abundant and widespread repetitive sequences in animal genomes (Kashi et al. 1990). On account of their ease to work, low cost, and high resolution, microsatellites are widely used as genetic markers and for individual identification and paternity testing. Their variability and evolutionary history also provide information about relationships between both populations and species. It is therefore tempting to use them in population genetics and for making phylogenetic inferences. However, patterns of mutation in microsatellites are often poorly understood and adequate analytical tools are rarely available (reviewed by Ellegren 2004; Goldstein & Pollock 1997; Jarne & Lagoda 1996; Schlötterer 2000).

The high mutation rates of microsatellites ( $10^{-2}$  to  $10^{-5}$  events per locus per generation) are mainly attributed to variations in the number of copies of the repeated motif, resulting from DNA (replication) slippage (Levinson & Gutman 1987; Schlötterer & Tautz 1992). Unequal crossing-over and recombination might also be involved, but their roles in the generation of mutation are less important (Eichler et al. 1995; Gordenin et al. 1997). A combination of these mechanisms together with base substitutions and indels results in three main sources of variability: the addition or deletion of repeat units, mutations interrupting the repeat, and substitutions and indels occurring in the microsatellite flanking regions (Makova et al. 2000).

The assumption that microsatellites evolve under the step-wise mutational model (SMM, Kimura & Ohta 1978) has been widely used since mutational mechanisms of slippage during replication are often consistent with this model (Schlötterer & Tautz 1992; Weber & Wong 1993) and the model is also attractive from a computational perspective. Microsatellite applications mainly involve comparing the mobility of different alleles through electrophoresis of PCR products. Nevertheless, size homoplasy, in which alleles have the same length but are not identical by descent, has been detected both within and between species. It is not therefore possible to determine allelic coalescence simply from information on allele



size (Garza & Freimer 1996; Grimaldi & Crouau-Roy 1997). Furthermore, the origin of simple mutational events cannot be assumed, and the role of point mutations or simple slippage in the origin of microsatellites cannot always be discerned (Gordon 1997).

The Mustelidae is a diverse and widespread carnivore family of increasing relevance in evolution and conservation genetics. Low levels of mitochondrial and nuclear variability, reduced repeat numbers, and a high number of imperfect repeats have been described in the Eurasian badger, *Meles meles* L. (Burke et al. 1996; Domingo-Roura et al. 2003). In this study, we sequenced four to ten alleles from each of ten Eurasian badger microsatellites in badgers and other mustelids. It has been established, based on the combination of morphological and molecular data, that these species shared a common ancestor around 20.8 million years ago (Bininda-Emonds et al. 1999). The sequences were obtained in order to detect microsatellites that follow SMM and to determine and inquire into the origin of variability that cannot be detected through standard genotyping procedures. Microsatellite variability and structure might be related to a wide range of processes, such as: i) genome-wide dynamics, including demographic history; ii) the technical particularities of each study, including library construction and the effort devoted to cloning and sequencing; and iii) the use of one or another definition of microsatellite categories. We standardized our microsatellite definition and explored whether the high incidence of imperfections and compound microsatellites were exclusive to our dataset or to Eurasian badgers by comparing our data with other Eurasian badger and carnivore sequences available from GenBank. We then assigned three possible sources of variability: the addition or deletion of repeat units (presumably following SMM); mutations interrupting the repeat; and substitutions and indels in the flanking regions of each of our microsatellites, in order to determine compliance with SMM. Finally, we correlated heterozygosity with the maximum number of perfect repeats in order to determine whether, in spite of this complexity, a portion of the microsatellite that can be modeled under SMM could be used for phylogenetic and population genetic inferences.

## 2. Materials and Methods

The present study was based on seven wild Eurasian badgers collected throughout their geographical distribution in Europe (Austria, Greece (Crete Island), Germany, Poland, Spain, Sweden, and the United Kingdom); two further Eurasian badgers from East Asia (one each from Japan and Mongolia); individuals from other mustelid species (a European river otter (*Lutra lutra* L.), a pine marten (*Martes martes* L.), a weasel (*Mustela nivalis* L.), an American mink (*Mustela vison* Schreber)); and a spotted skunk (*Spilogale putorius* L.). Samples were well-preserved, frozen or kept in ethanol and were obtained from museums and road kills.

Total DNA was extracted from either blood or muscular tissue according to standard phenol-chloroform protocols. DNA was amplified using the PCR primers *Me108-Me116* and *Me118*, which were designed from badger libraries constructed from the DNA of a British badger. Two non-enriched libraries were constructed following the conditions described in Domingo-Roura (2002) and Domingo-Roura et al. (2003). Briefly stated, genomic DNA extracted from muscle tissue was digested with *MboI* or *AluI*, and 100 to 500 bp fragments were recovered. These fragments were ligated into pUC18 vector and then transformed into XL-1 Blue cells (Stratagene, La Jolla, California). The resulting colonies were plated into nylon membranes, which were hybridized with (GC)<sub>10</sub>, (CA)<sub>15</sub>, (AT)<sub>10</sub>, (GAAA)<sub>10</sub>, or (GATA)<sub>10</sub> probes. An enriched DNA library was also constructed by Genetic Identification Services, as described in Kays et al. (2000).

Amplifications were performed with fluorescently labeled forward primers in a 15 µl reaction volume. Products were combined with size standard (ABI500 TAMRA, Applied Biosystems, Foster City, CA) and loaded into a 6% denaturing acrylamide gel in an ABI377 DNA sequencer. GeneScan v3.1 (Applied Biosystems) was used to determine fragment sizes.

Non-fluorescent versions of the same primers and the same protocol were used to sequence between three and nine electromorphs per marker; whenever possible, from homozygous individuals (Table 1). Twenty-five microlitres of PCR products were purified using GeneClean (BIO101, La



Jolla, CA). Neither of the badgers from Mongolia and Japan was homozygous for markers *Me/12*, *Me/15* and *Me/16*. In these cases to include geographically diverse samples, we sequenced alleles from heterozygous individuals after cloning. We also cloned both the otter and the German badger for marker *Me/14* because they were heterozygotes and their electromorph sizes were not congruent with the addition or deletion of whole repeat units. Overall, 26 electromorphs were sequenced from homozygous badgers and 10 from heterozygous specimens, including the clones sequenced in the original library. The amplification product from these markers was cloned and transformed into *E. coli* using the pMosBlue Blunt Ended Cloning Kit (Amersham Pharmacia Biotech, Piscataway, NJ). Plasmid DNA isolation was performed using Wizard<sup>®</sup> Plus SV Minipreps DNA Purification System (Promega, Madison, WI). Transgenic plasmids and simple PCR products were sequenced in both directions using Dye Terminator Cycle Sequencing Kit (Applied Biosystems) and an ABI377 DNA sequencer following the manufacturer's instructions. Resultant sequences were deposited in GenBank under accession numbers AJ309052-AJ309057, AJ309059, AJ309847-AJ309849, AJ489574, AJ489566, AJ489568, AJ489578, AJ489579, and AJ543956-AJ543992.

Alleles, including sequences from the original badger libraries, were visually aligned. Since automated scoring can produce slightly inaccurate estimates of allele size, fragment lengths obtained with Genescan v3.1 were adjusted with sequences. We defined a microsatellite as a sequence containing at least six consecutive repeats in at least one of the species sequenced. Microsatellite length, limits and classification as perfect, imperfect or compound were determined according to Weber (1990). Weber considered perfect repeat sequences as repeats without interruptions and without adjacent repeats of another sequence; imperfect repeat sequences as two or more runs of uninterrupted repeats separated by no more than three consecutive non-repeat bases and with terminal runs of at least three full repeats in length; and compound repeat sequences as runs of repeats separated by no more than three consecutive non-repeat bases from a run of  $\geq 5$  uninterrupted repeats of

other repeat motif. While Weber's work was based entirely on dinucleotide repeats, our study also considered tetranucleotide and mononucleotide repeat units (Table 1). The remaining sequence between the primers was considered the flanking region.

By sequencing different electromorphs for a given marker, we were able to determine which mutational records were present in the microsatellites: the addition and deletion of repeat units (presumably a result of slippage), the presence of interruptions in the repeat array, and changes in the flanking region including substitutions and indels (Table 2).

Fisher's exact tests were conducted to compare the number of single and compound repeats and the number of perfect and imperfect repeats obtained in this study with those of sequences available from GenBank of other Eurasian badgers (Carpenter et al. 2003) and of other carnivore species (*Mustela vison*, O'Connell et al. 1996; *Gulo gulo* L., Davis & Strobeck 1998; Walker et al. 2001; *Canis familiaris* L., Spriggs et al. 2003; *Ursus arctos* L., Taberlet et al. 1997; *Suricata suricatta* Schreber, Griffin et al. 2001; *Panthera tigris* L., Williamson et al. 2002; *Crocuta crocuta* Erxleben, Libants et al. 2000; Wilhelm et al. 2003). These microsatellites were isolated in a similar way to in our laboratory; from partial genomic libraries (selected for small insert size) of species of interest, and screening several thousands of clones through colony hybridization with repeat containing probes, most of which use (CA)<sub>n</sub> or (GT)<sub>n</sub>.

Finally, we explored whether – once alleles had been sequenced – the maximum length of perfect repeats correlated with variability. For this, we performed a Spearman's non-parametric coefficient for rank correlation of the number of alleles in British badgers from Wytham Woods, Oxfordshire (Domingo-Roura et al. 2003) and in the worldwide badger sample against the maximum number of consecutive perfect repeats.



**Table 1.** Species, microsatellite loci and electromorphs examined in this study. Sequenced electromorphs are in bold. NA = not amplified.

Microsatellite	Electromorphs						
	<i>Meles meles</i> (n = 9)	<i>Lutra lutra</i>	<i>Martes martes</i>	<i>Mustela nivalis</i>	<i>Mustela vison</i>	<i>Spilogale putorius</i>	
Me/08	<b>232</b>	<b>219</b>	<b>222</b>	<b>221</b>	437/438 <sup>a</sup>	<b>196</b>	
Me/09	<b>180, 187, 189, 191</b>	<b>158</b>	NA	NA	<b>158</b>	NA	
Me/10	<b>140, 142, 146, 148, 154, 156</b>	<b>132</b>	<b>161</b>	<b>152</b>	<b>143</b>	NA	
Me/11	<b>109, 117, 122, 124</b>	NA	84	84	84	NA	
Me/12	<b>119, 137, 139, 151, 153, 155, 157</b>	143	149/151	NA	NA	NA	
Me/13	<b>94, 96, 98, 100</b>	No cross-species amplification	No cross-species amplification				
Me/14	<b>170, 177, 178, 180, 182, 185, 186, 188, 190</b>	<b>183/189</b>	177	183/189	NA	NA	
Me/15	<b>228, 230, 234, 238, 242, 246, 254, 258, 270</b>	NA	<b>207</b>	199	NA	NA	
Me/16	290, <b>292, 293, 297, 316, 321, 325, 329, 333, 337, 345</b>	No cross-species amplification	No cross-species amplification				
Me/18	<b>339, 353, 356, 359, 367</b>	359/367	<b>352</b>	NA	NA	NA	

<sup>a</sup>Me/08 has a SINE insertion in this species (López-Giráldez et al., 2005)

### 3. Results and Discussion

#### (i) Information Extracted from Electromorphs

Cross-species amplification was possible for eight of the ten microsatellites, indicating the conservation of flanking regions over the approximately 20.8 million years that separate the badger from outgroup species (Bininda-Emonds et al. 1999; Table 1). The region flanking locus *Me/08* was able to amplify *Spilogale putorius* and species belonging to five different carnivore families (Domingo-Roura et al. 2005). This indicates the conservation of priming sites over approximately 35.5 million years. Amplification failed for the badgers from Crete and Austria using *Me/12* and *Me/18*, respectively. Since both markers could amplify across species, this probably indicated the presence of null alleles.

#### (ii) Sequencing to Explore Sources of Variation

One particular characteristic of our dataset was a high incidence of multiple repetitive arrays in a single clone, supporting a previously observed tendency for microsatellites to cluster (Bachtrog et al. 1999). *Me/09*, *Me/14* and *Me/18* each contained two separate repetitive arrays – one of which was a mononucleotide – between their priming sites (Figure 1), whereas *Me/08*, *Me/09*, *Me/10*, *Me/15* and *Me/18* each contained two adjacent arrays of different repetitive motif; in other words, they were compound microsatellites. All six microsatellite clones showed size variability in more than one repeat array across species, while four of them (*Me/09*, *Me/10*, *Me/14* and *Me/15*) also showed this variability within Eurasian badgers. The presence of more than one repeat array would probably preclude a unique stepwise mutation process which is an assumption of the SMM. Since *Me/08* was not a compound microsatellite in badgers, four out of the 14 repeat arrays in this species were compounds. The number of compound repeats was not significantly different from the number calculated from other Eurasian badger sequences available from GenBank (17 compound;  $n = 174$ , Fisher's exact test,  $P = 0.055$ ), and from other carnivore species (17 compound;  $n = 151$ , Fisher's exact test,  $P = 0.083$ ). The number of compound repeats was also not significantly



different between other Eurasian badger and carnivore sequences (Fisher's exact test,  $P = 0.378$ ). Note that the number of compound microsatellites reported in the original carnivore studies was lower than the number we used in these comparisons. This was due to differences between the criteria used to define compound microsatellites in these studies and our criteria, which were based on Weber (1990) and have already been detailed in the methods section. The evolution of compound microsatellites is still poorly understood, but the fact remains that clustering is an indicator that repeat arrays are unlikely to evolve independently.

When comparing sequences within the badger, all loci except *Me/08* and *Me/11* showed different perfect repeat copy numbers. In addition to this variability – that was presumably caused by slippage –, all of them except *Me/14* also showed interruptions in at least one repeat array. It has been suggested that point mutations break up perfect repeats, reducing the mutation rates of microsatellite loci and precluding infinite growth in repeat number (Kruglyak et al. 1998). This is in disagreement with another assumption of the SMM, which states that the number of repeats should vary at a fixed rate and does not consider an upper limit to their growth in copy number. None of the ten microsatellites analyzed showed a variable single repeat array without interruptions or indels in the flanking region (Table 2), demonstrating the complexity of our database. Ten repeats were perfect and eight were imperfect. The number of imperfect repeats was not significantly different from the number calculated for other badger sequences (47 imperfect;  $n = 192$ , Fisher's exact test,  $P = 0.064$ ), but it was significantly larger than the number obtained for other carnivore species (34 imperfect;  $n = 169$ , Fisher's exact test,  $P = 0.025$ ). The number of imperfect repeats was not significantly different between other badger and carnivore sequences (Fisher's exact test,  $P = 0.216$ ). It is not possible to differentiate between interruptions created by substitutions, those arising as a result of gains or losses of multiple repeat units – one of which is imperfect –, and those resulting from slippage involving partial repeat units (Palsbøll et al. 1999; Zhu et al. 2000).





**Table 2.** Molecular mechanisms generating allelic variation obtained from sequencing between 2 and 7 different electromorphs in 10 microsatellites from *Meles meles*. N1: total number of electromorphs sequenced; N2: number of sequences of the same size class obtained from different individuals; N3: number of true alleles detected by sequencing, including the microsatellite and flanking region. The mean number of interruptions in the repeat region (RR), the mean number of differences (transitions and transversions) between flanking region (FR) sequences calculated by MEGA v2.1, and the number of indels in the FR are shown. See Figure 1 for number of base pairs in the FR analyzed for each marker.

Locus	Repeat unit	Sequencing effort			True alleles		Rang of repeat numbers		Interruptions in RR	Differences in FR (bp)	Indels in FR (bp)
		N1	N2	N3	Min	Max					
Only one STR pattern											
<b>Me/11</b>	T <sub>n</sub>	4	2	3	11	27	2.25	0.67	0	0	
<b>Me/12</b>	(TG) <sub>n</sub>	6	1	6	11.5	30.5	1.00	0	0	0	
<b>Me/13</b>	(CA) <sub>n</sub>	6	3	4	12.5	16	1.17	0	0	0	
<b>Me/16</b>	(TTTC) <sub>n</sub>	4	1	3	18.25	30.75	3.75	2.00	1, 3	1, 3	
Two repeat motifs											
<b>Me/14</b> ( <i>Me/14.1</i> ) ( <i>Me/14.2</i> )	A <sub>n</sub> (CA) <sub>n</sub>	7	2	6	10 9	12 18	NO NO	0.40	0	0	
Compound											
<b>Me/08</b>	(CA) <sub>n</sub> A <sub>n</sub>	5	5	2	6/10	6/10	1.00	1.00	0	0	
<b>Me/10</b>	A <sub>n</sub> (CA) <sub>n</sub> T <sub>n</sub>	5	3	4	8/14/6	12/25/6	5.50	0	0	0	
Two repeat motifs and compounds											
<b>Me/09</b> ( <i>Me/9.1</i> ) ( <i>Me/9.2</i> )	(CT) <sub>n</sub> (CA) <sub>n</sub> (AATC) <sub>n</sub> A <sub>n</sub>	2	1	2	14/7/5 6	16/9/5 6	1.00 NO	1.00	1	1	
<b>Me/15</b> ( <i>Me/15.1</i> ) ( <i>Me/15.2</i> )	(TTTC) <sub>n</sub> T <sub>n</sub> (GA) <sub>n</sub>	5	1	5	4.5/7 18.5	14.5/7 21.5	NO 3.80	1.00	4, 12	4, 12	
<b>Me/18</b> ( <i>Me/18.1</i> ) ( <i>Me/18.2</i> )	(CTTC) <sub>n</sub> (CT) <sub>n</sub> A <sub>n</sub>	5	4	3	5/7 11	10/7 11	NO 1.00	1.18	0	0	







Thus, once microsatellite definition had been standardized, our dataset showed a high number of imperfect repeats. Since our badger sequences were isolated from four different genomic libraries in two different laboratories, it is unlikely that imperfections and complexities only result from a systematic bias in probing efficiency during library production (Domingo-Roura et al. 2003). Even so, our sequences were isolated from the British population inhabiting Wytham Woods, Oxfordshire, where variability is known to be low (Domingo-Roura et al. 2003), and where four different libraries were needed to achieve a moderate level of variability when trying to identify the individuals of this population. Thus, when variability is low and there is a need to obtain a large number of variable microsatellites, for instance for paternity testing, in addition to microsatellite clones with a large number of perfect repeats, researchers may tend to sequence, design primers, and include sequences from imperfect microsatellites in GenBank.

Even when electromorph sizes within badger and between species showed a pattern congruent with the addition or deletion of repeat units, the sequencing of alleles demonstrated the existence of other repeat arrays or indels that mimicked the addition or deletion of repeat units. Three loci (*Me/09*, *Me/15*, and *Me/16*) had one or more indels in the flanking regions (Table 2) in badgers, which again contradicted the SMM assumptions that only predict changes in size in the repeat array. This could also be a reason for misclassifying allele size relative to the number of repeat units and therefore also a possible form of homoplasy not assumed by SMM. In addition, all markers had either indels and/or base differences in their flanking regions across species and, therefore electromorph size is clearly inadequate for phylogenetic inference.

### *(iii) A CAN-SINE Insertion in the Dataset*

When exploring the sequences shown in Figure 1, the (complementary) sequence of *Me/15* was of particular interest as it contained an almost complete Short Interspersed Element (SINE). As with typical SINEs, the insertion contained internal promoters and terminators of RNA polymerase III, a polypyrimidine region and an A-rich 3'-tail (the T-rich tail in the

complementary sequence shown in the Figure). This element shared more than 80% sequence identity with other CAN-SINE sequences (data not shown), a family of short retroposons frequently repeated in carnivore genomes. One of the two repeat arrays present in this locus, the GA-based repeat (CT in the complementary chain), was part of the polypyrimidine region, which is variable both in sequence and length within the CAN SINE family (Vassetzky & Kramerov 2002). Accordingly, we found a large deletion in the badger from Crete and smaller deletions in martens, which both followed this GA-based repeat. The second repeat array in the *Me/15* locus had a (CTTT)<sub>n</sub> motif. Its complementary chain contained a (GAAA)<sub>n</sub> motif and was located in the A-rich tail of the SINE. We also found a CAN-SINE inserted in the flanking region of the *Me/08* locus in the American vison: that is useful for differentiating between this species from similar mustelid species (López-Giráldez et al. 2005). A-rich tails in Alu sequences have been shown as important sources for the genesis of primate microsatellites (Arcot et al. 1995; Nadir et al. 1996).

#### *(iv) Microsatellite Structure and Variation*

The alleles found in a badger population from Wytham Woods, Oxfordshire, United Kingdom (Domingo-Roura et al. 2003) and in our sample of nine badgers from nine different countries were available as variability data for each of the microsatellites that we studied. Only five markers (*Me/10*, *Me/12*, *Me/14*, *Me/15*, and *Me/16*) were variable in Wytham Woods, while the others were monomorphic. The number of alleles was positively correlated with the maximum number of consecutive perfect repeats ( $n = 10$ ,  $\rho = 0.751$ ,  $P = 0.012$ ), but not with microsatellite length, including imperfections ( $n = 10$ ,  $\rho = 0.208$ ,  $P = 0.564$ ). The comparison of allele numbers across Eurasia gave a better view of the relationship between structure and variation, since the number of alleles and heterozygosity values might differ considerably across populations for a single species. With the exception of monomorphic *Me/08*, between four and 11 electromorphs were found in the Eurasian sample. The number of electromorphs positively correlated with the maximum number of consecutive perfect repeats ( $n = 10$ ,



$\rho = 0.858$ ,  $P = 0.001$ ), and also, with a lower correlation coefficient, with microsatellite length, including imperfections ( $n = 10$ ,  $\rho = 0.640$ ,  $P = 0.046$ ). Thus, in spite of the complexity of the dataset, the number of perfect repeats was still a good predictor of variability and this parameter can be easily modeled under SMM.

It is becoming progressively more evident that the rate and patterns of microsatellite mutation depend on multiple factors including the sequence and chromosomal position of the loci involved, as well as the genotype, age, and sex of the individual hosting the microsatellite (reviewed in Ellegren 2004; and Schlötterer 2000). In addition, molecular mechanisms that affect microsatellites, such as the efficiency of slippage or recombination, are related to stress and environmental, demographic, and life history correlates (Li et al. 2002; Neff & Gross 2001). Theoretical models have been evolved to cover some of these complexities (Calabrese et al. 2001; Kruglyak et al. 1998; Nauta & Weissing 1996; Sainudiin et al. 2004), while new models, following distributions such as power-law and Bernouilli, have also been proposed (Li et al. 2002).

Mutational models and statistics applicable in population genetics need to be properly tested, but parameterizing complex models promises to be a difficult and long-term task. There is, therefore, a need to identify microsatellites whose evolution can be modeled under simpler models and/or to determine the longest array of perfect repeats, by sequencing different alleles of a single locus. At the same time, there is also a need to standardize as many variables related to microsatellite mutation as possible before statistics based on simple models, such as the SMM, can be applied.

## 5. References

- Arcot SS, Wang Z, Weber JL, Deininger PL, Batzer MA (1995) *Alu* repeats: a source for the genesis of primate microsatellites. *Genomics* 29: 136-144.
- Bachtrog D, Weiss S, Zangerl B, Brem G, Schlötterer C (1999) Distribution of dinucleotide microsatellites in the *Drosophila melanogaster* genome. *Mol Biol Evol* 16: 602-610.
- Bininda-Emonds ORP, Gittleman JL, Purvis A (1999) Building large trees by combining phylogenetic information: a complete phylogeny of the extant Carnivora (Mammalia). *Biol Rev* 74: 143-175.
- Burke T, Hanotte O, Van Pijlen I (1996) Minisatellite analysis in conservation genetics. In: *Molecular Genetic Approaches in Conservation* (Smith TB & Wayne RK, eds). Oxford University Press, New York.
- Calabrese PP, Durrett RT, Aquadro CF (2001) Dynamics of microsatellite divergence under stepwise mutation and proportional slippage/point mutation models. *Genetics* 159: 839-852.
- Carpenter PJ, Dawson DA, Greig C, Parham A, Cheeseman CL, Burke T (2003) Isolation of 39 polymorphic microsatellite loci and the development of a fluorescent marker set for the Eurasian badger (*Meles meles*) (Carnivora: Mustelidae). *Mol Ecol Notes* 3: 610-615.
- Davis CS, Strobeck C (1998) Isolation, variability, and cross-species amplification of polymorphic microsatellite loci in the family Mustelidae. *Mol Ecol* 7: 1176-1178.
- Domingo-Roura X (2002) Genetic distinction of marten species by the fixation of a microsatellite region. *J Mammal* 83: 907-912.
- Domingo-Roura X, Macdonald DW, Roy MS, Marmi J, Terradas J, Woodroffe RJ, Burke T, Wayne RK (2003) Confirmation of low genetic diversity and multiple breeding females in a social group of Eurasian badgers from microsatellite and field data. *Mol Ecol* 12: 533-539.



- Domingo-Roura X, López-Giráldez F, Saeki M, Marmi J (2005) Phylogenetic inference and comparative evolution of a complex microsatellite and its flanking regions in carnivores. *Genet Res* 85: 223-233.
- Eichler EE, Kunst CB, Lugenbeel KA, Ryder OA, Davison D, Warren ST, Nelson DL (1995) Evolution of the cryptic FMR1 CGG repeat. *Nat Genet* 11: 301-308.
- Ellegren H (2004) Microsatellites: simple sequences with complex evolution. *Nat Rev Genet* 5: 435-445.
- Estoup A, Jarne P, Cornuet JM (2002) Homoplasmy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Mol Ecol* 11: 1591-1604.
- Garza JC, Freimer NB (1996) Homoplasmy for size at microsatellite loci in humans and chimpanzees. *Genome Res* 6: 211-217.
- Goldstein DB, Pollock DD (1997) Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. *J Hered* 88: 335-342.
- Gordenin DA, Kunkel TA, Resnick MA (1997) Repeat expansion--all in a flap? *Nat Genet* 16: 116-118.
- Gordon AJ (1997) Microsatellite birth register. *J Mol Evol* 45: 337-338.
- Griffin AS, Nürnberger B, Pemberton JM (2001) A panel of microsatellites developed for meerkats (*Suricata suricatta*) by cross-species amplification and species-specific cloning. *Mol Ecol Notes* 1: 83-85.
- Grimaldi MC, Crouau-Roy B (1997) Microsatellite allelic homoplasmy due to variable flanking sequences. *J Mol Evol* 44: 336-340.
- Jarne P, Lagoda PJJ (1996) Microsatellites, from molecules to populations and back. *Trends Ecol Evol* 11: 424-429.



- Kashi Y, Tikochinsky Y, Genislav E, Iraqi F, Nave A, Beckmann JS, Gruenbaum Y, Soller M (1990) Large restriction fragments containing poly-GT are highly polymorphic in a variety of vertebrates. *Nucleic Acids Res* 21: 1111-1115.
- Kays RW, Gittleman JL, Wayne RK (2000) Microsatellite analysis of kinkajou social organization. *Mol Ecol* 9: 734-751.
- Kimura M, Ohta T (1978) Stepwise mutation model and distribution of allelic frequencies in a finite population. *Proc Natl Acad Sci USA* 75: 2868-2872.
- Kruglyak S, Durres RT, Schug MD, Aquadro CF (1998) Equilibrium distributions of microsatellite repeat length resulting from balance between slippage events and point mutations. *Proc Natl Acad Sci USA* 95: 10774-10778.
- Levinson G, Gutman GA (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol* 4: 203-221.
- Li Y-C, Korol AB, Fahima T, Beiles A, Nevo E (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Mol Ecol* 11: 2453-2465.
- Libants S, Olle E, Oswald K, Scribner KT (2000) Microsatellite loci in the spotted hyena *Crocuta crocuta*. *Mol Ecol* 9: 1443-1445.
- López-Giráldez F, Gómez-Moliner BJ, Marmi J, Domingo-Roura X (2005) Genetic distinction of American and European mink (*Mustela vison* and *M. lutreola*) and European polecat (*M. putorius*) hair samples by detection of a species-specific SINE and a RFLP assay. *J Zool, Lond* 265: 1-6.
- Makova, KD, Nekrutenko A, Baker RJ (2000) Evolution of microsatellite alleles in four species of mice (genus *Apodemus*). *J Mol Evol* 51: 166-172.



- Nadir E, Margalit H, Gallily T, Ben-Sasson SA (1996) Microsatellite spreading in the human genome: evolutionary mechanisms and structural implications. *Proc Natl Acad Sci USA* 93: 6470-6475.
- Nauta MJ, Weissing FJ (1996) Constraints in allele size at microsatellite loci: implications for genetic differentiation. *Genetics* 143: 1021-1032.
- Neff BD, Gross MR (2001) Microsatellite evolution in vertebrates: inference from AC dinucleotide repeats. *Evolution* 55: 1717-1733.
- O'Connell M, Wright JM, Farid A (1996) Development of PCR primers for nine polymorphic American mink, *Mustela vison*, microsatellite loci. *Mol Ecol* 5: 311-312.
- Palsbøll PJ, Bérubé M, Jørgensen H (1999) Multiple levels of single-strand slippage at cetacean tri- and tetranucleotide repeat microsatellite loci. *Genetics* 151: 285-296.
- Sainudiin R, Durrett RT, Aquadro CF, Nielsen R (2004) Microsatellite mutation models: insights from a comparison of humans and chimpanzees. *Genetics* 168: 383-395.
- Schlötterer C, Tautz D (1992) Slippage synthesis of simple sequence DNA. *Nucleic Acids Res* 20: 211-215.
- Schlötterer C (2000) Evolutionary dynamics of microsatellite DNA. *Chromosoma* 109: 365-371.
- Spriggs HF, Holmes NG, Breen MG, Deloukas PG, Langford CF, Ross MT, Carter NP, Davis ME, Knights CE, Smith AE, Farr CJ, McCarthy LC, Binns MM (2003) Construction and integration of radiation-hybrid and cytogenetic maps of dog Chromosome X. *Mamm Genome* 14: 214-221.
- Taberlet P, Camarra JJ, Griffin S, Uhres E, Hanotte O, Waits LP, Dubois-Paganon C, Burke T, Bouvet J (1997) Noninvasive genetic tracking of the endangered Pyrenean brown bear population. *Mol Ecol* 6: 869-876.
- Vassetzky NS, Kramerov DA (2002) CAN – a pan-carnivore SINE family. *Mamm Genome* 13: 50-57.

Walker CW, Vilà C, Landa A, Lindén M, Ellegren H (2001) Genetic variation and population structure in Scandinavian wolverine (*Gulo gulo*) populations. *Mol Ecol* 10: 53-63.

Weber JL, 1990. Informativeness of human (dC-dA)<sub>n</sub>·(dG-dT)<sub>n</sub> polymorphisms. *Genomics* 7: 524-530.

Weber JL, Wong C (1993) Mutation of human short tandem repeats. *Hum Mol Genet* 2: 1123-1128.

Wilhelm K, Dawson DA, Gentle LK, Horsfield GF, Schlötterer C, Greig C, East M, Hofer H, Tautz D, Burke T (2003) Characterization of spotted hyena, *Crocuta crocuta*, microsatellite loci. *Mol Ecol Notes* 3: 360-362.

Williamson JE, Huebinger RM, Sommer JA, Louis Jr EE, Barber RC (2002) Development and cross-species amplification of 18 microsatellite markers in the Sumatran tiger (*Panthera tigris sumatrae*). *Mol Ecol Notes* 2: 110-112.

Zhu Y, Strassmann JE, Queller DC (2000) Insertions, substitutions, and the origin of microsatellites. *Genet Res Camb* 76: 227-236.



**Acknowledgments.** We thank A. Arrizabalaga (Museu de Granollers), J. Brabec (University of Innsbruck), A. Bradshaw (University of Cardiff), P. Fakler (Universität Heidelberg), A. Laurie (Eastern Steppes Biodiversity Project), M. Miralles (Rectoria Vella), M. Nadolska (Agricultural University of Wrocław), M. Saeki (Wildlife Conservation Research Unit, University of Oxford), V. Sidorovich (National Academy of Sciences of Belarus), N. Yamaguchi (Wildlife Conservation Research Unit, University of Oxford), and R. Woodroffe (Wildlife Conservation Research Unit, University of Oxford) for providing samples used in the study. D. Macdonald (Wildlife Conservation Research Unit, University of Oxford) provided logistical support. We thank O. Andrés, A. Antonell, G. Fernández, L.F. Magano, A. Pérez-Lezaun, and M. Vallès for their help in the laboratory, and S. Baker for editorial assistance. This research has been co-funded by the European Commission under contract INPRIMAT (QLRI-CT-2002-01325) and the Comissionat per a Universitats i Recerca, Generalitat de Catalunya (ref. 2000SGR00093). During the development of the project, F. L-G. and J.M. received scholarships from the Departament d'Universitats, Recerca i Societat de la Informació, Generalitat de Catalunya (refs. 2001FI00625 and 2000FI00698, respectively).

## Capítol II

### ***Phylogenetic inference and comparative evolution of a complex microsatellite and its flanking regions in carnivores***

Xavier Domingo-Roura, Francesc López-Giráldez, Midori Saeki i Josep Marmi

(publicat en *Genetical Research* (2005) 85: 223–233)

## PHYLOGENETIC INFERENCE AND COMPARATIVE EVOLUTION OF A COMPLEX MICROSATELLITE AND ITS FLANKING REGIONS IN CARNIVORES

Xavier Domingo-Roura<sup>1,2</sup>, Francesc López-Giráldez<sup>1,2\*</sup>, Midori Saeki<sup>4</sup>, Josep Marmi<sup>2,4</sup>

<sup>1</sup>*Genètica de la Conservació, Institut de Recerca i Tecnologia Agroalimentàries, Centre de Cabrils, Ctra. de Cabrils s/n, 08348 Cabrils (Barcelona), Spain.*

<sup>2</sup>*Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Dr. Aiguader 80, 08003 Barcelona, Spain.*

<sup>3</sup>*Landscape and Ecology Division, Environment Department, National Institute for Land and Infrastructure Management, Ministry of Land, Infrastructure and Transport, Asahi 1, Tsukuba, Ibaraki 305-0804, Japan.*

<sup>4</sup>*Departament de Ciència Animal i dels Aliments, Facultat de Veterinària, Universitat Autònoma de Barcelona, 08193 Cerdanyola del Vallès, Spain.*

\*Corresponding author: Genètica de la Conservació, Institut de Recerca i Tecnologia Agroalimentàries, Centre de Cabrils, Ctra. de Cabrils s/n, 08348 Cabrils (Barcelona), Spain. Tel: +34 937507511. Fax: +34 937533954. e-mail: [francesc.lopez@gmail.com](mailto:francesc.lopez@gmail.com)

---

**Abstract.** We sequenced locus *Me/08*, with complex short repetitive motifs, in 24 carnivore species belonging to five different families in order to explore mutational changes in the region in the context of locus and species evolution. This non-coding locus includes up to four different parts or repetitive motifs showing size variability. The variability consists of repeat additions and deletions; substitutions, insertions and/or deletions creating interruptions in the repeat; and substitutions, insertions and deletions in the flanking regions. The locus has different repeat expansions in different carnivore subfamilies. We hypothesize that the complexity of this locus is due to a high mutation rate at an ancestral DNA sequence and, thus, prompts the emergence of repeats at mutational hotspots. High levels of homoplasy were evident, with nine electromorphs representing 28 haplotypes never shared across species. The variability in flanking regions was informative for phylogenetic inference and their evolutionary content. Tree topologies were congruent with relevant hypotheses on current conflicts in carnivore phylogenies, such as: (i) the monophyly of Lutrinae, (ii) the paraphyly of Mustelinae, (iii) the basal position of the Eurasian badger, *Meles meles*, in the Mustelidae, (iv) the classification of skunks as a separate family, Mephitidae, and (v) the placement of the red panda, *Ailurus fulgens*, as a monotypic family, Ailuridae, at a basal position in the Musteloidea.

---

## 1. Introduction

Repetitive regions are widespread in animal genomes and the determination and understanding of their variability is essential for their use in individual identification, population genetics or phylogenetics. Usually, repeats are considered microsatellites when they are composed of six or more consecutive repeat units of the same type, and short repeats or protomicrosatellites below this repeat number, although this might be more a semantic than scientific threshold.

Considerable effort has been devoted to understanding and modelling the mutational processes involved in microsatellite evolution (Bell & Jurka 1997; Kruglyak et al. 1998; Ellegren 2004; Sainudiin et al. 2004). However, little is known about the dynamics of shorter repetitive regions and their origin from standard DNA strings is not well understood. The information available on the evolution of groups of adjacent repeats is even scarcer.

DNA replication slippage (meaning that during DNA replication the nascent and template strands do not realign correctly) is the main force driving the dynamics of microsatellite evolution (Levinson & Gutman 1987). Recombination and unequal crossing-over might also act, although their influence in generating variation is often considered less important (Eichler et al. 1995). However, this might be different in short repeats. It has been proposed that point mutations can be important in generating variability at short repeat copy numbers (Stephan & Kim 1998) before slippage becomes the dominant mechanism (Schlötterer 2000). It has also been shown that, over long evolutionary periods, size expansion is possible from only two repeats (Primmer & Ellegren 1998). More general knowledge about repeat dynamics includes the fact that not only slippage but also deletions and insertions occur at higher frequencies at tandem repeats (Kroutil & Kunkel 1999), with short repeats having a 10- to 15-fold increased susceptibility to insertions and deletions compared with nonrepetitive sequences (Nishizawa & Nishizawa 2002).

The conservation of the flanking regions might permit the amplification of a target microsatellite in distantly related species. In this work we take advantage of the fact that among carnivores the regions flanking a complex repetitive locus are conserved to analyse their origin and diversification in order to advance in the understanding of the mutational processes leading to the origin of short and adjacent repetitive regions. We also check whether, despite the mutational complexity of this locus and the high levels of homoplasy observed, phylogenetic information can still be detected in both the repeat and flanking regions. As a part of this, we construct a phylogenetic tree based on these flanking regions and compare our results with other data from the literature that relate to relevant controversies in carnivore phylogeny (e.g., Bininda-Emonds et al. 1999; Koepfli & Wayne 2003). In particular we explore: (i) the monophyly of Lutrinae and Mustelinae; (ii) the position of the genus *Gulo* in relation to *Martes*; (iii) the basal position of the Eurasian badger, *Meles meles*, within the Mustelidae; (iv) the monophyly of Mustelidae, the placement of skunks either as a mustelid subfamily or as a separate family, and the possible sister relationship of mustelids and procyonids; and (v) the position of the red panda, *Ailurus fulgens*, in relation to other procyonids and carnivores.

## 2. Materials and methods

### (i) Amplification, sequencing and alignment

*Me/08* locus was isolated from a Eurasian badger (*Meles meles*) size-selected library (Domingo-Roura 2002). In the initial screening for variability, the marker showed a single electromorph of 232 base pairs (bp) in badgers from different origins but a shorter electromorph in other mustelid species. Polymerase chain reaction (PCR) amplification and sequencing was performed for all species shown in Table 1 as described in Domingo-Roura (2002). *Me/08* was highly conserved in the region of primers and we were able to amplify the locus in a wide range of carnivore species. Species nomenclature and classification follows Macdonald (2001) except for seals, which were classified according to Nowak (1991) because they were not classified by Macdonald at the subfamily level. We



sequenced a larger number of *Meles meles* and *Procyon lotor* individuals since these species had higher probabilities of showing polymorphism due to longer and presumably more variable repeat regions.

In addition to the species and individuals shown in Table 1, we tried unsuccessfully the amplification of the *Me/08* locus in the following species: *Vulpes vulpes*, *Canis familiaris*, *Melursus (Ursus) ursinus*, *Felis catus*, *Acinonyx jubatus*, *Genetta genetta*, *Suricata suricatta*, *Rhinolophus ferrumequinum*, *Myotis nattereu* and *Capra pyrenaica*. Unsuccessful amplifications included the design of degenerate primers, additional PCRs with a gradient of annealing temperatures between 47.9 and 56.1 °C using a Master Cycler Gradient Thermocycler (Eppendorf, Hamburg, Germany) and a concentration of 3.0 mM MgCl<sub>2</sub>.

We divided the sequences into repeats and flanking regions. The alignment of the flanking regions is unambiguous for all species (Fig. 1), whereas homology of repeat units within the repeat region cannot be assumed in our alignments. To facilitate reading and comparison, the repetitive regions were divided into four parts identified by a specific repeat motif and one of the parts was further divided into two subparts identified by a predominant repeat interruption (Fig. 2).

#### *(ii) Looking for evidences of coding*

In order to discard the possibility that the high degree of conservation in the priming region of *Me/08* could be due to the locus being part of transcribed DNA or gene, we looked for stop codons in our sequences considering the three possible reading frames and forward and reverse readings using the DnaSP package v3.51 (Rozas & Rozas 1999). We also investigated the presence of indels resulting from the addition or removal of three base pairs or a multiple of three base pairs in the flanking regions. Finally, using the USA National Center for Biotechnology Information (NCBI) interface (<http://www.ncbi.nlm.nih.gov/BLAST/>), we performed BLAST searches of the two flanking regions separately against the protein databases (blastx) and against the six-frame translations of a nucleotide sequence (tblastx). Both nr (all GenBank, EMBL, DDBJ and PDB



**Table 1.** Species included in the study and their geographic origin, the length (in nucleotides) of their Mel08 locus, and GenBank/EMBL accession numbers.

Species	Origin	Sample provider	<i>n</i>	Electromorph length	Repeat region length	Flanking regions length	GenBank/EMBL accession number
Mustelidae, Lutrinae							
<i>Enhydra lutris</i>	Alaska, USA	S. O'Brien, Laboratory of Genomic Diversity, National Cancer Institute	1	219	27	192	AJ489576
<i>Lutra lutra</i>	Wales, UK Verhnedvinsk, Belarus	A. Bradshaw, University of Cardiff V. Sidorovic, National Academy of Sciences of Belarus	1 1	219 219	27 27	192 192	AJ489574 AJ489574
<i>Amblyonyx chinereus</i>	Thailand	S. O'Brien, Laboratory of Genomic Diversity, National Cancer Institute	1	219	27	192	AJ489575
<i>Lontra canadensis</i>	Florida, USA	S. O'Brien, Laboratory of Genomic Diversity, National Cancer Institute	1	220	28	192	AJ489573
Mustelidae, Mustelinae							
<i>Mustela erminea</i>	Suffolk, UK Suffolk, UK	R. A. Macdonald, University of Bristol R. A. Macdonald, University of Bristol	1 1	221 222	29 30	192 192	AJ489565 AJ489564
<i>Mustela lutreola</i>	Gorodok, Belarus	V. Sidorovic, National Academy of Sciences of Belarus	1	220	28	192	AJ489560
<i>Mustela putorius</i>	Oxfordshire, UK Catalonia, Spain	A. Grogan, WildCRU, University of Oxford C. Rosell, Minuartia, Sant Celoni	1 1	220 220	28 28	192 192	AJ489562 AJ489562
<i>Mustela nigripes</i>	Captive	A. Kitchner, National Museums of Scotland	1	221	29	192	AJ489561
<i>Mustela nivalis</i>	Gorodoksky, Belarus	V. Sidorovic, National Academy of Sciences of Belarus	1	221	29	192	AJ489566
<i>Mustela altaica</i>	Mt Altai, Russia	E. I. Zholnerovskaya, Siberian Zoological Museum	1	222	30	192	AJ489563
<i>Martes martes</i>	Asturias, Spain Smolevichsky, Belarus Kihniö, Finland	S. Lavín, Universitat Autònoma de Barcelona V. Sidorovic, National Academy of Sciences of Belarus K. Kauhala, Finnish Game and Fisheries Research Institute	1 1 1	222 222 222	30 30 30	192 192 192	AJ489568 AJ489568 AJ489568
<i>Martes melampus</i>	Dijon, France Tokyo, Japan	X. Domingo-Roura, University Pompeu Fabra Y. Fukue, Tokyo University of Agriculture and Technology	1 1	222 222	30 30	192 192	AJ489568 AJ489567
<i>Martes foina</i>	Asturias, Spain	S. Lavín, Universitat Autònoma de Barcelona	1	222	30	192	AJ489569
<i>Gulo gulo</i>	Nesvizh, Belarus Northwest Territories, Canada	V. Sidorovic, National Academy of Sciences of Belarus M. A. Ramsay, University of Saskatchewan and C. M. Pond, Open University	1 1	222 222	30 30	192 192	AJ489570 AJ489571



Mustelidae, Melinae								
<i>Meles meles</i>								
	Austria	J. Brabec, University of Innsbruck	1	232	40	192	AJ309847	
	Tokyo, Japan	Y. Fukue, Tokyo University of Agriculture and Technology	2	232	40	192	AJ489572	
	Oxfordshire, UK	D. Macdonald and C. Newman, WildCRU, University of Oxford	1	232	40	192	AJ309847	
	Catalonia, Spain	M. Miralles, Rectoria Vella, Sant Celoni	1	232	40	192	AJ309847	
	Crete, Greece	D. Macdonald and R. Woodroffe, WildCRU, University of Oxford	1	232	40	192	AJ309847	
Mephitidae								
<i>Mephitis mephitis</i>	Maryland, USA	S. O'Brien, Laboratory of Genomic Diversity, National Cancer Institute	1	196	22	174	AJ489577	
<i>Spilogale putorius</i>	Oregon, USA	S. O'Brien, Laboratory of Genomic Diversity, National Cancer Institute	1	196	22	174	AJ489578-9	
Procyonidae, Procyoninae								
<i>Procyon lotor</i>	Michigan, USA	F. López-Giráldez, University Pompeu Fabra	1	220	31	189	AJ489580	
	Hokkaido, Japan	Y. Matoba and M. Asakawa, Rakuno Gakuen University	16	220	31	189	AJ489580	
	Hokkaido, Japan	Y. Matoba and M. Asakawa, Rakuno Gakuen University	2	220	31	189	AJ489581	
	London Zoo	D. Gottelli, Institute of Zoology, London	1	219	29	190	AJ489582	
Procyonidae, Ailurinae								
<i>Ailurus fulgens</i>	London Zoo and Barcelona Zoo	D. Gottelli, Institute of Zoology, London and J. Fernández, Parc Zoològic de Barcelona S.A.	2	207	21	186	AJ489583	
Otariidae, Arctocephalinae								
<i>Arctocephalus gazella</i>	South Georgia	J. P. Arnould, Macquarie University, and C. M. Pond, Open University	1	205	21	184	AJ489584	
Otariidae, Otariinae								
<i>Zalophus californianus</i>	London Zoo	D. Gottelli, Institute of Zoology, London	1	205	21	184	AJ489585	
Phocidae, Phocinae								
<i>Phoca vitulina</i>	London Zoo	D. Gottelli, Institute of Zoology, London	2	215	24	191	AJ489586	

sequences) and est (Expressed Sequence Tags from GenBank, EMBL and DDBJ sequences) databases were searched. For these comparisons we selected the first species of each genus as shown in the table and figures. To expand the possibility of detecting partial alignments in a wider region of our sequences, we increased the Expect value to 30 and we used Matrix BLOSUM45 with Gap Costs increased from default values to 13 for Existence and three for Extension. Searches were limited to eukaryotic organisms.

*(iii) Phylogenetic information in the flanking regions*

Homoplasy was evaluated by comparing the number of electromorphs against sequences representing true alleles. To understand the link between the evolution of species and the evolution of the repetitive locus, nucleotide substitutions in the flanking regions were used to estimate phylogenetic relationships among alleles. Indels in the flanking regions were recorded as a single transversion. The transversional model with rate heterogeneity (TVM+ $\Gamma$ ;  $\alpha=1.375$ ) was selected as the optimum substitution model for our data set using the Akaike Information Criterion (AIC) approach implemented in the program Modeltest v3.6 (Posada & Crandall 1998). We used this model and values of the parameters in a maximum-likelihood analysis to estimate phylogenetic relationships. Tree searches were performed using the genetic algorithm implemented in the Treefinder software, version of December 2004 (Jobb et al. 2004). The number of categories dividing the discrete approximation of the gamma distribution was 8. Confidence in the resulting relationships was assessed with 1000 bootstrap replicates. A pairwise matrix of substitutions and transversions per site was computed using the same substitution model with PAUP\* 4.0b10 (Swofford 2002) for the flanking regions. True seals (Phocidae) and eared seals (Otariidae) were considered outgroups.

*(iv) Mutational differences between subfamilies*

We also estimated how evolutionary changes in the flanking regions and repeat motifs are reflected in a time scale. We obtained the divergence time estimates between each pair of subfamilies from work based on

combining molecular and morphological data from different literature sources (Bininda-Emonds et al. 1999). After checking for normality in the distribution of variables by the Kolmogorov–Smirnov test, we calculated Spearman’s non-parametric coefficient for rank correlation ( $P$  values =  $P_S$ ) on the divergence time in million years (Mya) against pairwise differences between subfamilies in: (i) electromorph length; (ii) substitutions per site in flanking regions; (iii) transversions per site in flanking regions; (iv) differences in repeat length including imperfections in Part 1; (v) differences in longest perfect continuous nucleotide string in Part 1; (vi) differences in repeat length including imperfections in Part 3; and (vii) differences in longest perfect continuous nucleotide string in Part 3. These tests were performed using SPSS v9.0.1 (SPSS Inc.). Since it could be argued that values within variables are not independent,  $P$  values ( $P_M$ ) for the same correlations were double checked with Mantel’s test using the ARLEQUIN package (Schneider et al. 2000).

### 3. Results

#### *(i) The locus is not a coding region*

The sequences flanking locus *Me108* were conserved in a total of 24 different species belonging to five different carnivore families (Table 1), with an estimated time of divergence of 35.5 Mya (Bininda-Emonds et al. 1999). In spite of this high degree of conservation, the locus is not part of a coding region since we found stop codons in all species considering all reading frames in forward and reverse directions. In 98.5% of the sequences and reading frames, there was more than one stop codon. Deletion events were of 1, 2, 5, 7 and 16 nucleotides and never a multiple of 3. Furthermore, we could not find any pair of homologous sequences corresponding to the two flanking regions coding for a single protein or translated region through BLAST searches.

#### *(ii) Variability in flanking regions*

The nucleotide composition of the flanking regions was: T = 31.3%, C = 16%, A = 33.2% and G = 19.5%. Of 59 polymorphic sites in the flanking regions, 44 were parsimony informative. The nucleotide diversity

( $\pi$ ) was 0.103. The number of nucleotide differences among species ranged from 0 between closely related *Mustela putorius*, *M. nigripes* and *M. lutreola* to 34 between *Lontra canadensis* and *Zalophus californianus*. Flanking regions had a constant size within families with the exception of a nucleotide difference within the Procyonidae species (Table 1). The main differences in size were due to taxon-specific deletions. These deletions were often near the repeat motif (Fig. 1).

*(iii) Dissection of mutations in repetitive motifs by taxon*

Mutations accumulated through evolution at this locus have resulted in a complex repetitive sequence that can be separated in contiguous but clearly different repetitive patterns or parts. The most relevant characteristic of the locus is that all these parts are variable across species and that the expansion of these repeat motifs has been different in different subfamilies (Fig. 2). Part 1 is based on a mononucleotide T tract with G imperfections in several species and it is expanded in families Mephitidae, Procyonidae, Otariidae and Phocidae, whereas it is short in the Mustelidae family. This part is variable in sequence within *Procyon lotor*. *Meles meles* has an additional CCTG in Part 1 and a non-variable perfect (CA)<sub>6</sub> microsatellite in Part 2, which is absent in all other species. This repeat motif allowed the detection of the *Me108* locus during library hybridization (Domingo-Roura 2002). Part 3 is based on an imperfect C<sub>n</sub> tract and shows a high degree of complexity within the Mustelidae, whereas it is shorter and simpler in the remaining families. Part 4 is a mononucleotide G-based tract ranging from G<sub>2</sub> in the Mustelidae to G<sub>4</sub> in the Mephitidae. Several evolutionary mechanisms concentrated in a single DNA spot must have been involved in the generation of this complex repetitive region.

*(iv) Homoplasy*

In this study, we amplified a total of 55 individuals of 23 different carnivore species to detect a total of nine electromorphs (Table 1) and 28 haplotypes or true alleles. No pair of species shared any haplotype and homoplasy was detected at all taxonomic levels. Within species, there

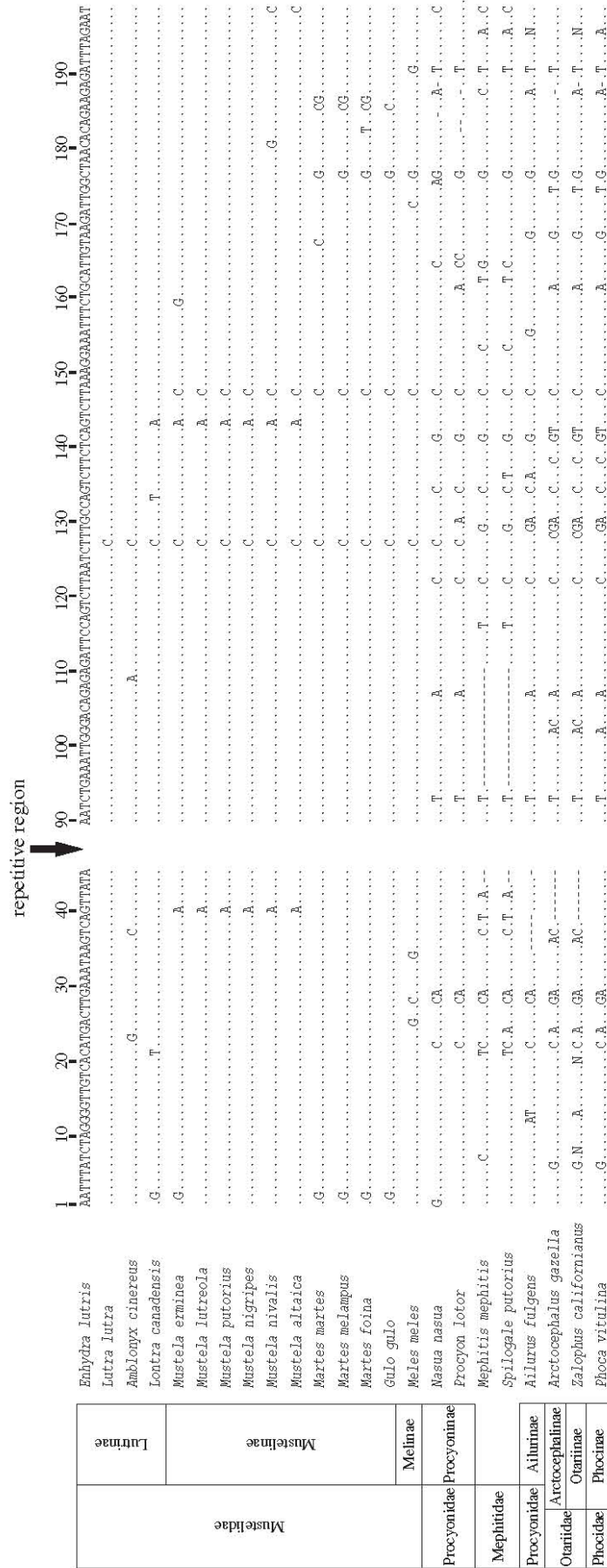


Fig. 1. Alignment of Me108 flanking region sequences. Dots (.) indicate nucleotide identity, hyphens (-) indicate nucleotide deletions and N indicate unresolved nucleotide positions. The location of repetitive sequences is also indicated.

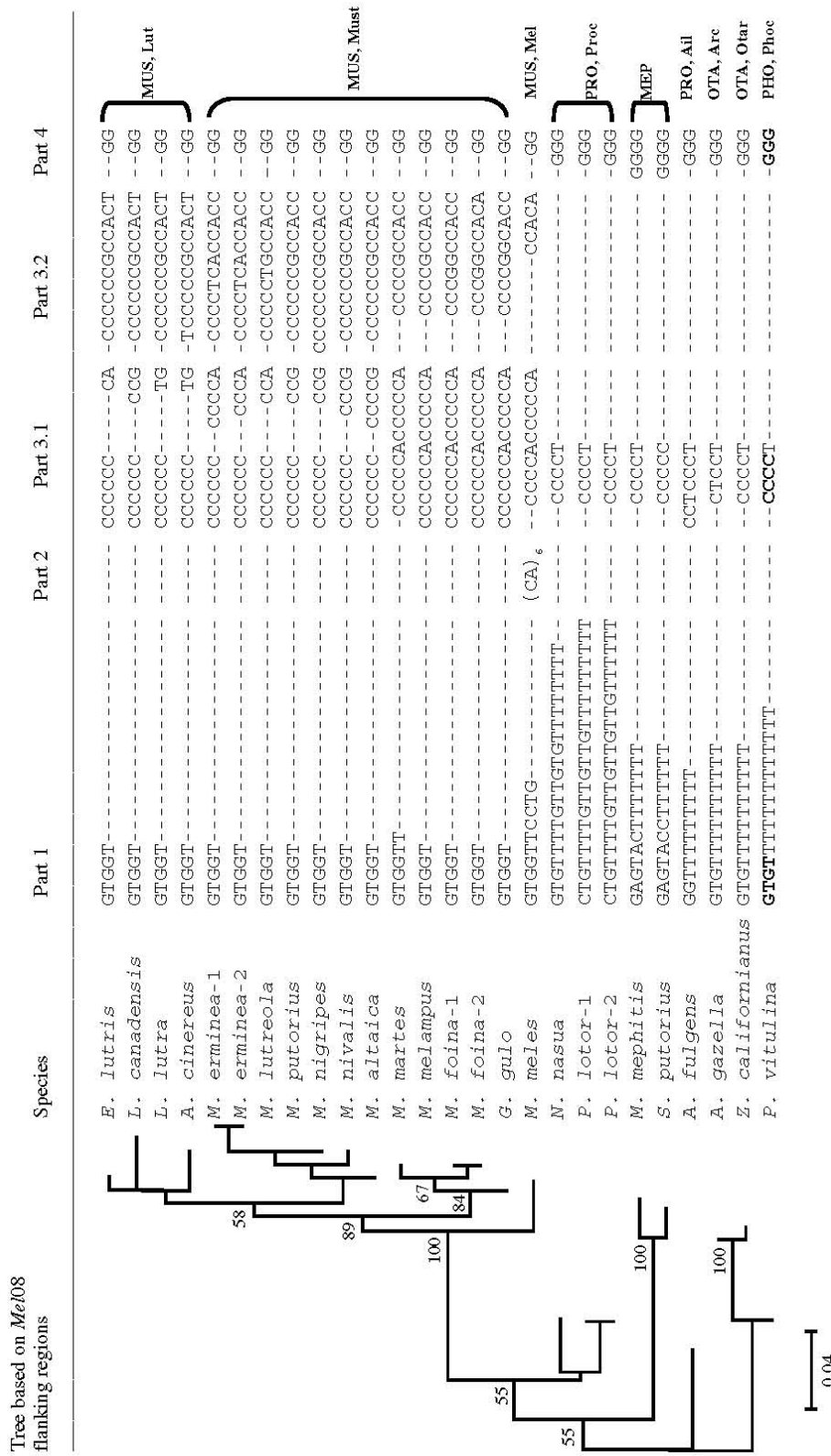


Fig. 2. Alignments of 26 allelic variants found in the repetitive region from different carnivore species that could be amplified at locus *Mei08*. The sequences have been divided into four parts according to the main repetitive pattern. Flanking regions are shown in Fig. 1. Dashes indicate the partial or complete absence of the different parts in given alleles. Bootstrap values higher than 50% of *Mei08* maximum likelihood tree are indicated. See Fig. 1 for complete generic names. Taxa abbreviations: MUS, Mustelidae; Lut, Lutrinae; Must, Mustelinae; Mel, Melinae; PRO, Procyonidae; Proc, Procyoninae; MEP, Mephitidae; Ail, Ailurinae; OTA, Otariidae; Arc, Arctocephalinae; Ota, Otariinae; PHO, Phocidae; Phoc, Phocinae.



were two different electromorphs within *Mustela erminea* and two different sequences of the same size in *Martes foina* and *Procyon lotor*. *Meles meles* and *Spilogale putorius* had withinspecies differences in the sequences flanking the repetitive region. The electromorph of 219 bp is shared among distant taxa, such as *Nasua nasua* and the otter (Lutrinae) subfamily. The electromorph of 220 bp is shared among three different genera and subfamilies. The low reliability of allele size for inferring true evolutionary relationships among alleles is demonstrated by the poor correlation between allele size or electromorph size differences and Mya of separation among subfamilies ( $\rho = 0.014$ , see below).

(v) *Phylogenetic information in the flanking regions*

The topology of the tree constructed from *Me/08* flanking regions is similar to published trees, such as the ones constructed from cytochrome *b*, although bootstrap values are moderate (Fig. 2) due to the short region analysed. True seals (Phocidae) and eared seals (Otariidae) were considered outgroups of a cluster formed by procyonids (Ailurinae and Procyoninae), skunks (Mephitidae) and mustelids (Melinae, Mustelinae and Lutrinae). The Lutrinae subfamily is monophyletic, whereas the Mustelinae subfamily is paraphyletic as Lutrinae was clustered with genus *Mustela* after the separation from genus *Martes*. *Gulo* clustered with *Martes* and *Meles* was at a basal position of the Mustelidae excluding skunks. *Procyon lotor* and *Nasua nasua* (Procyoninae) diverged from mustelids after the Mephitidae. *Ailurus fulgens* (Ailurinae) was not included in the cluster formed by the other procyonids (Procyoninae).

(vi) *Mutational differences between families*

The Kolmogorov–Smirnov test showed that divergence times (Mya) between subfamilies ( $Z = 1.794$ ,  $N = 36$ ,  $P = 0.003$ ) and differences in repeat length including imperfections in Part 3 ( $Z = 1.415$ ,  $N = 36$ ,  $P = 0.036$ ) significantly deviated from normality. Divergence time among subfamilies was significantly correlated with the following parameters: (i) substitutions in flanking regions ( $\rho = 0.442$ ,  $P_S = 0.007$ ,  $P_M = 0.019$ ); (ii) transversions in flanking regions ( $\rho = 0.411$ ,  $P_S = 0.013$ ,  $P_M = 0.017$ );

(iii) differences in the length of the largest perfect tract in Part 1 ( $\rho = 0.377$ ,  $P_S = 0.023$ ,  $P_M = 0.024$ ) and (iv) differences in electromorph length but only for Mantel's test ( $\rho = 0.014$ ,  $P_S = 0.936$ ,  $P_M = 0.020$ ). The variables that were not significantly correlated with time of divergence were differences in repeat length including imperfections in Part 1 ( $\rho = 0.030$ ,  $P_S = 0.860$ ,  $P_M = 0.256$ ), differences in repeat length including imperfections in Part 3 ( $\rho = 0.026$ ,  $P_S = 0.879$ ,  $P_M = 0.193$ ) and differences in the length of the longest perfect tract in Part 3 ( $\rho = 0.048$ ,  $P_S = 0.783$ ,  $P_M = 0.154$ ). The number of comparisons was 36 in all cases. All correlations were affected by the position of skunks as in the work of Bininda-Emonds et al. (1999) where they are considered to be mustelids, a hypothesis widely questioned in other works (e.g., Ledje & Arnason 1996; Marmi et al. 2004).

#### 4. Discussion

##### *(i) Where is variability located?*

Three main mechanisms have been proposed for generating new alleles in repeat regions: addition and deletion of repeats, substitutions and indels in flanking regions, and substitutions, insertions and deletions creating interruptions in the repeat (Kruglyak et al. 1998; Makova et al. 2000). The *Me/08* locus seems to have evolved through all three mechanisms. Clear evidence of nucleotides added by slippage was found in all parts and taxonomic levels. Slippage is evident in the C-based longest repeat of Part 3 in mustelids and in the T-based repeat of Part 1 in other species. Imperfections in the repeat are evident in Parts 1 and 3. These disruptions stabilize microsatellite loci by reducing the substrate for polymerase slippage and recombination. The high number of imperfections in Part 3 in mustelids and in Part 1 in other species supports the observation that single-base insertions occur at mononucleotide repeats with frequencies dramatically increasing with the increase in length of the repeat (Halangoda et al. 2001). However, several mutational mechanisms could result in the sequence observed: substitutions, insertions, single-strand slippage involving partial repeats, or the gain or loss of multiple repeats, one of which is imperfect. Imperfections, such as the ones found in Part 1

in *Nasua nasua* and *Procyon lotor* (where a G interruption is repeated several times), suggest that these interruptions are not independent events but are part of the addition and deletion mechanisms (Estoup et al. 1995). Some of the imperfections are species-specific. Examples include the T found at the beginning of Part 3.2 in *Amblonyx cinereus* and the T and A imperfections found in the middle of Part 3.2 in *Mustela erminea*. Other imperfections, such as the last T of Part 3.2 in the Lutrinae or the A in the middle of Part 3.1 in genus *Martes* and *Gulo gulo*, are found in related species, showing that these interruptions might have common ancestry, rather than being the result of convergence. The last nucleotide of Part 3.1 has mutated several times between A and G in different branches of the Mustelidae, suggesting a transitionprone position (Fig. 2). Substitutions and indels of different lengths in flanking regions are also common, especially when comparing across families (Fig. 1).

*(ii) The link between repeat generation and mutational hotspots*

The locus *Me108* shows high mutation rates in both the repetitive and flanking regions and still the conservation of priming sites across 35.5 Mya. More distant carnivore families could not be amplified. It should be noted that flanking regions as well as the repeat were already conserved in the badger, the species from which primers were designed. The level of homologous amplification is considerable in comparison with other published polymorphic microsatellite regions in mammals (e.g., 9 Mya in primates: Domingo-Roura et al. 1997; 40 Mya in primates: Clisson et al. 2000; 35–40 Mya in cetaceans: Schlötterer et al. 1991) but moderate in comparison with other microsatellites which have been amplified across longer evolutionary periods in other taxa (e.g., 300 Mya in turtles: FitzSimmons et al. 1995; 470 Mya in fishes: Rico et al. 1996; 144 Mya in wasps: Ezenwa et al. 1998). The conservation of flanking regions across different families is certainly unusual. We hypothesize that *Me108* represents a mutational hotspot with several different expansion and imperfection events accumulated from a single ancestral DNA sequence. We believe that the link between repeat generation and mutational hotspots exists because repetitive regions arise at

mutational/recombinational hotspots. Supporting this hypothesis it has been shown that hypervariability arises in the proximity of a meiotic hotspot where meiotic recombination is frequently observed (Jeffreys et al. 1998). Indirect evidence in support of the hypothesis of microsatellites arising at mutational hotspots would be that compound microsatellites are more common than expected (Estoup et al. 1995) and the tendency of microsatellites to group in clusters (Bachtrog et al. 1999). It is also known that substitution rates can be higher adjacent to microsatellites (Santibáñez-Koref et al. 2001), and that point mutation rates are higher within microsatellites than in the rest of the human genome (Sibly et al. 2003). Insertions resulting from mononucleotide slippage result in a direct repeat of the preceding bases (Cooper & Krawczak 1991). This seems to happen in the T repeated at the end of Part 1 in *Martes martes*, or in the G repeated in the middle of Part 3.2 in *Martes foina* and *Gulo gulo*.

Finally, mutation also depends on the characteristics of the locus, such as GC content. *Me/08*, with a 35.6% GC content in the flanking regions, supports a higher microsatellite density the further the GC content deviates from 50% (Dieringer & Schlötterer 2003).

### *(iii) Phylogenetic information in flanking regions*

Microsatellite flanking regions contain phylogenetic information (Zardoya et al. 1996; Makova et al. 2000). They can be highly mutable but are normally short and, thus, phylogenetic information is limited. In contrast, other nuclear regions are more conserved but larger fragments can be sequenced to compensate for this conservation (e.g., Amrine-Madsen et al. 2003; Koepfli & Wayne 2003; Sato et al. 2004). These regions can often be amplified with the same primers even across different families.

In spite of working with a short fragment, phylogenetic relationships inferred from *Me/08* flanking regions showed important congruencies with phylogenies based on fossil, morphological and sequence data. Our results give support to the monophyly of Lutrinae and the paraphyly of Mustelinae, a point often raised in other molecular-based work (based on five nuclear gene segments for a total of ~3000 bp: Koepfli & Wayne

2003; 1188 bp of the IRBP nuclear gene and the cytochrome *b*: Sato et al. 2003; cytochrome *b*: Marmi et al. 2004; two fragments of 1095 and 1188 bp of RAG1 and IRBP nuclear genes respectively: Sato et al. 2004), and work based on morphology (46 morphological characters: Bryant et al. 1993). *Gulo gulo* appears closely related to a monophyletic group of species of genus *Martes*, in agreement with other studies (Sato et al. 2003, 2004). In contrast, the paraphyly of *Martes* has also been claimed because of the sister association between *Gulo* and *Martes americana*, a species not analysed in our study (Koepfli & Wayne 2003). With a high level of confidence, *Meles meles* is placed at the base of the Mustelidae as suggested for badgers (from Taxidiinae and Melinae subfamilies) in other molecular studies (Koepfli & Wayne 2003; Sato et al. 2003, 2004; Marmi et al. 2004), but in disagreement with studies based on chromosome painting and G-banding (Nie et al. 2002) and morphology (Bryant et al. 1993).

There is also controversy regarding the phylogenetic relationships of skunks. Some reports based on morphological data suggest that skunks are closely related to taxa within the Mustelidae (Simpson 1945; Wolsan 1999). The phylogeny of the *Me108* flanking regions presented in this work supports the exclusion of skunks from the Mustelidae and their basal position in a group formed by Procyoninae and nonmephitine Mustelidae. Such a placement is supported by other works based on DNA sequences and morphology (cytochrome *b*: Ledge & Arnason 1996; Marmi et al. 2004; 851 bp of transthyretin intron I, cytochrome *b*, partial 12S rRNA, and morphological data: Flynn & Nedbal 1998; Sato et al. 2004).

A major issue relates to the position of the red panda in the tree. According to several morphological and molecular studies, the red panda is closely related to ursids or ursids and pinnipeds (based on morphological data: Wozencraft 1989; based on a combination of mitochondrial sequences and morphological data: Vrana et al. 1994), to the procyonids (based on the fossil record: Wang 1997), to the musteloids (mustelids plus procyonids, based on nuclear and mitochondrial genes: Flynn et al. 2000) or to the giant panda (*Ailuropoda melanoleuca*, based

on morphological data: Ginsburg 1982). However, other authors think that the red panda forms an unresolved monotypic lineage within the arctoid clade (Ledge & Arnason 1996). Although the position of the red panda has low bootstrap support values, the *Me/08* flanking region phylogeny places this species as a monotypic taxon at the base of the Musteloidea, which is consistent with other studies (Flynn & Nedbal 1998; Bininda-Emonds et al. 1999; Flynn et al. 2000; Marmi et al. 2004). Considering these results, we agree on placing the red panda within a monotypic family (Ailuridae) in the Musteloidea.

Our results demonstrate the potential for inferring phylogenies from sequences flanking microsatellites which are abundant in genetic databases, and usually discarded for individual identification because of their low variability. However, we detect homoplasy at all taxonomic levels and electromorph size is not an adequate parameter to explore evolutionary distance when complex loci are involved (Colson & Goldstein 1999). Despite homoplasy, differences in the length of the longest perfect tract were correlated with evolutionary distance for Part 1. However, imperfections in Part 1 and Part 3 prevent the correlation of the repeat length (including imperfections) with evolutionary distance. In conclusion, the sequencing of electromorphs provides not only valuable information contained in the flanking regions but also information on the number of perfect repeats, a parameter that can easily be modelled for population and phylogenetic inferences.

## 5. References

- Amrine-Madsen H, Koepfli KP, Wayne RK, Springer MS (2003) A new phylogenetic marker, apolipoprotein B, provides compelling evidence for eutherian relationships. *Mol Phylogenet Evol* 28: 225–240.
- Bachtrog D, Weiss S, Zangerl B, Brem G, Schlötterer C (1999) Distribution of dinucleotide microsatellites in the *Drosophila melanogaster* genome. *Mol Biol Evol* 16: 602–610.
- Bell GI, Jurka J (1997) The length distribution of perfect dimer repetitive DNA is consistent with its evolution by an unbiased single-step mutation process. *J Mol Evol* 44: 414–421.
- Bininda-Emonds OR, Gittleman JL, Purvis A (1999) Building large trees by combining phylogenetic information: a complete phylogeny of the extant Carnivora (Mammalia). *Biol Rev* 74: 143–175.
- Bryant HN, Russell AP, Fitch WD (1993) Phylogenetic relationships within the extant Mustelidae (Carnivora): appraisal of the cladistic status of the Simpsonian subfamilies. *Zool J Linn Soc-Lond* 108: 301–334.
- Clisson I, Lathuilliere M, Crouau-Roy B (2000) Conservation and evolution of microsatellite loci in primate taxa. *Am J Primatol* 50: 205–214.
- Colson I, Goldstein DB (1999) Evidence for complex mutations at microsatellite loci in *Drosophila*. *Genetics* 152: 617–627.
- Cooper DN, Krawczak M (1991) Mechanisms of insertional mutagenesis in human genes causing genetic diseases. *Hum Genet* 87: 409–415.
- Dieringer D, Schlötterer C (2003) Two distinct modes of microsatellite mutation processes: evidence from the complete genomic sequences of nine species. *Genome Res* 13: 2242–2251.
- Domingo-Roura X (2002) Genetic distinction of marten species by the fixation of a microsatellite region. *J Mammal* 83: 907–912.

- Domingo-Roura X, López-Giráldez T, Shinohara M, Takenaka O (1997) Hypervariable microsatellite loci in the Japanese macaque (*Macaca fuscata*) conserved in related species. *Am J Primatol* 43: 357–360.
- Eichler EE, Kunst CB, Lugenbeel KA, Ryder OA, Davison D, Warren ST, Nelson DL (1995) Evolution of the cryptic FMR1 CGG repeat. *Nat Genet* 11: 301–308.
- Ellegren H (2004) Microsatellites: simple sequences with complex evolution. *Nat Rev Genet* 5: 435–445.
- Estoup A, Tailliez C, Cornuet JM, Solignac M (1995) Size homoplasy and mutational processes of interrupted microsatellites in two bee species, *Apis mellifera* and *Bombus terrestris* (Apidae). *Mol Biol Evol* 12: 1074–1084.
- Ezenwa VO, Peters JM, Zhu Y, Arevalo E, Hastings MD, Seppa P, Pedersen JS, Zacchi F, Queller DC, Strassmann JE (1998) Ancient conservation of trinucleotide microsatellite loci in polistine wasps. *Mol Phylogenet Evol* 10: 168–177.
- FitzSimmons NN, Moritz C, Moore SS (1995) Conservation and dynamics of microsatellite loci over 300 million years of marine turtle evolution. *Mol Biol Evol* 12: 432–440.
- Flynn JJ, Nedbal MA (1998) Phylogeny of the Carnivora (Mammalia): congruence vs incompatibility among multiple data sets. *Mol Phylogenet Evol* 9: 414–426.
- Flynn JJ, Nedbal MA, Dragoo JW, Honeycutt RL (2000) Whence the red panda? *Mol Phylogenet Evol* 17: 190–199.
- Ginsburg L (1982) Sur la position systématique du petit panda, *Ailurus fulgens* (Carnivora, Mammalia). *Géobios, Mémoire Spécial* 6: 247–258.
- Halangoda A, Still JG, Hill KA, Sommer SS (2001) Spontaneous microdeletions and microinsertions in a transgenic mouse mutation



- deletion system: analysis of age, tissue, and sequence specificity. *Environ Mol Mutagen* 37: 311–323.
- Jeffreys AJ, Murray J, Neumann R (1998) Highresolution mapping of crossovers in human sperm defines a minisatellite-associated recombination hot spot. *Mol Cell* 2: 267–273.
- Jobb G, von Haeseler A, Strimmer K (2004) TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evol Biol* 4: 18.
- Koepfli KP, Wayne RK (2003) Type I STS markers are more informative than cytochrome *b* in phylogenetic reconstruction of the Mustelidae (Mammalia, Carnivora). *Syst Biol* 52: 571–593.
- Kroutil LC, Kunkel TA (1999) Deletion errors generated during replication of CAG repeats. *Nucleic Acids Res* 27: 3481–3486.
- Kruglyak S, Durrett RT, Schug MD, Aquadro CF (1998) Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. *Proc Natl Acad Sci USA* 95: 10774–10778.
- Ledge C, Arnason U (1996) Phylogenetic analyses of complete cytochrome *b* genes of the order Carnivora with particular emphasis on the Caniformia. *J Mol Evol* 42: 135–144.
- Levinson G, Gutman GA (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol* 4: 203–221.
- Macdonald DW (2001) *The New Encyclopedia of Mammals*. Oxford University Press, Oxford.
- Makova KD, Nekrutenko A, Baker RJ (2000). Evolution of microsatellite alleles in four species of mice (genus *Apodemus*). *J Mol Evol* 51: 166–172.
- Marmi J, López-Giráldez JF, Domingo-Roura X (2004) Phylogeny, evolutionary history and taxonomy of the Mustelidae based on

sequences of the *cyt b* gene and a complex repetitive-flanking region. *Zool Scr* 33: 481–499.

Nie W, Wang J, O'Brien PC, Fu B, Ying T, Ferguson-Smith MA, Yang F (2002) The genome phylogeny of domestic cat, red panda and five mustelid species revealed by comparative chromosome painting and G-banding. *Chromosome Res* 10: 209–222.

Nishizawa M, Nishizawa K (2002) A DNA sequence evolution analysis generalized by simulation and the Markov chain Monte Carlo method implicates strand slippage in a majority of insertions and deletions. *J Mol Evol* 55: 706–717.

Nowak RM (1991) *Walker's Mammals of the World*, 5th edn. Johns Hopkins University Press, Baltimore

Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics* 14: 817–818.

Primmer CR, Ellegren H (1998) Patterns of molecular evolution in avian microsatellites. *Mol Biol Evol* 15: 997–1008.

Rico C, Rico I, Hewitt G (1996) 470 million years of conservation of microsatellite loci among fish species. *P Roy Soc B-Biol Sci* 263: 549–557.

Rozas J, Rozas R (1999) DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* 15: 174–175.

Sainudiin R, Durrett RT, Aquadro CF, Nielsen R (2004) Microsatellite mutation models: insights from a comparison of human and chimpanzees. *Genetics* 168: 383–395.

Santibáñez-Koref MF, Gangeswaran R, Hancock JM (2001) A relationship between lengths of microsatellites and nearby substitution rates in mammalian genomes. *Mol Biol Evol* 18: 2119–2123.

- Sato JJ, Hosoda T, Wolsan M, Tsuchiya K, Yamamoto M, Suzuki H (2003) Phylogenetic relationships and divergence times among mustelids (Mammalia: Carnivora) based on nucleotide sequences of the nuclear interphotoreceptor retinoid binding protein and the mitochondrial cytochrome *b* genes. *Zool Sci* 20: 243–264.
- Sato JJ, Hosoda T, Wolsan M, Suzuki H (2004) Molecular phylogeny of arctoids (Mammalia: Carnivora) with emphasis on phylogenetic and taxonomic positions of the ferret-badgers and skunks. *Zool Sci* 21: 111–118.
- Schlötterer C (2000) Evolutionary dynamics of microsatellite DNA. *Chromosoma* 109: 365–371.
- Schlötterer C, Amos B, Tautz D (1991) Conservation of polymorphic simple sequence loci in cetacean species. *Nature* 354: 63–65.
- Schneider S, Roessli D, Excoffier L (2000) *Arlequin ver. 2.0: A Software for Population Genetics Data Analysis*. Genetics and Biometry Laboratory, University of Geneva, Geneva.
- Sibly RM, Meade A, Boxall N, Wilkinson MJ, Corne DW, Whittaker JC (2003) The structure of interrupted human AC microsatellites. *Mol Biol Evol* 20: 453–459.
- Simpson GG (1945) The principles of the classification and a classification of mammals. *B Am Mus Nat Hist* 85: 1–350.
- Stephan W, Kim Y (1998) Persistence of microsatellite arrays in finite populations. *Mol Biol Evol* 15: 1332–1336.
- Swofford DL (2002) *PAUP\*: Phylogenetic Analysis Using Parsimony (\*And Other Methods)*, version 4.0b8. Sinauer, Sunderland.
- Vrana PB, Milinkovitch MC, Powell JR, Wheeler WC (1994) Higher level relationships of the arctoid Carnivora based on sequence data and 'total evidence'. *Mol Phylogenet Evol* 3: 47–58.

Wang X (1997) New cranial material of *Symocyon* from China, and its implications for phylogenetic relationship to the red panda (*Ailurus*). *J Vertebr Palaeontol* 17: 184–198.

Wozencraft WC (1989) The phylogeny of the Recent Carnivora. In: *Carnivore Behavior, Ecology, and Evolution* (Gittleman JL ed). Comstock Publishing Associates, Ithaca.

Wolsan M (1999) Oldest mephitine cranium and its implications for the origin of skunks. *Acta Palaeontol Pol* 44: 223–230.

Zardoya R, Vollmer DM, Craddock C, Streelman JT, Karl S, Meyer A (1996) Evolutionary conservation of microsatellite flanking regions and their use in resolving the phylogeny of cichlid fishes (Pisces: Perciformes). *P Roy Soc B-Biol Sci* 263: 1589–1598.

**Acknowledgments.** We wish to thank researchers and institutions that provided samples listed in Table 1 as well as E. Garcia-Franquesa (Museu de Ciències Naturals de la Ciutadella, Barcelona) and S. Pérez (Taxidermia El Ciervo). We thank A. Pérez-Lezaun and M. Vallès for their help in the laboratory, F. Calafell, L. Pérez-Jurado and W. Matten (NCBI User Services) for their advice in data analyses, and M. Pulido for editorial assistance. D. Goldstein, R. Wayne and two anonymous reviewers provided useful comments to improve earlier versions of the manuscript. This project was financed by the Spanish Ministerio de Educación y Cultura (Ref. PB98-1064) and Generalitat de Catalunya (Ref. 2000SGR00093). F L-G and JM are supported by scholarships from the Departament d'Universitats, Recerca i Societat de la Informació, Generalitat de Catalunya (Refs. 2001FI-00625 and 2000FI-00698 respectively).

# Capítol III

## ***Analyses of carnivore microsatellites and their intimate association with tRNA- derived SINEs***

Francesc López-Giráldez, Olga Andrés, Xavier Domingo-  
Roura, Montserrat Bosch

(enviat a *BMC Genomics*)

## ANALYSES OF CARNIVORE MICROSATELLITES AND THEIR INTIMATE ASSOCIATION WITH tRNA-DERIVED SINES

Francesc López-Giráldez\*, Olga Andrés, Xavier Domingo-Roura\*,  
Montserrat Bosch

*Genètica de la Conservació, Institut de Recerca i Tecnologia Agroalimentàries, Centre de Cabrils, Ctra. de Cabrils s/n, 08348 Cabrils (Barcelona), Spain.*

\*Corresponding author: Genètica de la Conservació, Institut de Recerca i Tecnologia Agroalimentàries, Centre de Cabrils, Ctra. de Cabrils s/n, 08348 Cabrils (Barcelona), Spain. Tel: +34 937507511. Fax: +34 937533954. e-mail: [francesc.lopez@gmail.com](mailto:francesc.lopez@gmail.com)

\*This paper is dedicated to the memory of Dr. Xavier Domingo-Roura who died on 17th November 2005, aged 41 years, after a brave fight against cancer.

---

**Abstract.** The popularity of microsatellites has greatly increased in the last decade on account of their many applications. However, little is currently understood about the factors that influence their genesis and distribution among and within species genomes. In this work, we analyzed carnivore microsatellite clones from GenBank to study their association with interspersed repeats and elucidate the role of the latter in microsatellite genesis and distribution. Our results reveal a distinct juxtaposition of AG and A-rich repeats and tRNA<sup>Lys</sup>-derived SINEs suggesting their coevolution. Both microsatellites arose repeatedly in two regions of the element. These observations suggest two modes of microsatellite generation: the expansion and variation of pre-existing tandem repeats and the conversion of sequences with high cryptic simplicity into a repeat array; mechanisms which are not specific to tRNA<sup>Lys</sup>-derived SINEs. Microsatellite and interspersed repeat coevolution could also explain different distribution of repeat types among and within species genomes. Finally, microsatellites associated with tRNA<sup>Lys</sup>-derived SINEs showed the highest complexity and less potential instability. In light of these findings, care must be taken when using microsatellites associated with interspersed repeats as genetic markers.

---

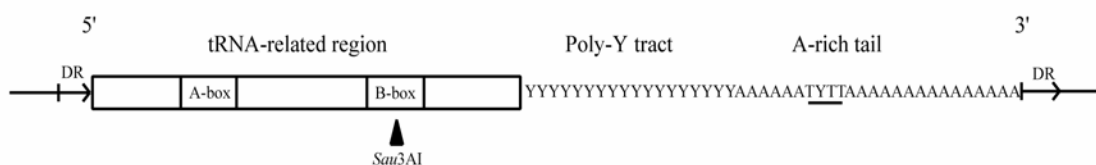
## 1. Introduction

Eukaryote genomes contain repetitive DNA sequences that can be classified into two groups: tandemly repeated sequences (e.g., micro- and minisatellites) and dispersed sequences (e.g., long interspersed elements – LINEs – and short interspersed elements – SINEs). Microsatellites (MSs) are tandem repeats of a DNA motif, one to six bases long, showing high levels of polymorphism based on changes in the repeat number. They are highly abundant and considered selectively neutral sequences and almost randomly distributed in the mammalian genome (Schlötterer 2000). In spite of the wide applications of MSs and their importance in genetic and evolutionary studies, the mechanisms of the genesis of these sequences are still not fully understood. It is thought that point mutation is the dominant source of generation of short repeat MSs before slippage becomes the dominant mechanism (Levinson and Gutman 1987; Messier et al. 1996; Rose and Falush 1998; Zhu et al. 2000). However, based on the close association described between retroposons and MSs in several mammalian species (in sheep; Buchanan et al. 1993; in pig, Alexander et al. 1995; in primates, Arcot et al. 1995; in humans, Nadir et al. 1996; in horse, Gallagher et al. 1999), a completely different mechanism for microsatellite generation has been proposed: it is thought that retrotranscripts undergo 3' polyadenylation – similar to mRNA polyadenylation (Whale 1995) – prior to their incorporation into the genome, and that the extension of this preexisting repeat can generate A-rich MSs (Arcot et al. 1995; Nadir et al. 1996). However, this mechanism would not explain the recent observations (in barley, Ramsay et al. 1999; in Dipterans, Wilder and Hollocher 2001) that MSs can also be associated with both 5' and internal regions of some retroposons. An explanation for the latter association is the mechanism described by Wilder and Hollocher (2001) which implies the conversion of an existing sequence with high cryptic simplicity located in Dipteran *mini-me* elements into tandemly repeated DNA. The importance of repeat elements as MS sources is still unclear. Whereas Nadir et al. (1996) proposed Alu elements as the preferential source for the origin of human MSs, Morgante



et al. (2002) showed a significant association between MSs and the single/low-copy fraction of the plant genome.

In mammals, most retroposons can be classified into three basic types: SINEs, LINEs, and LTR (Long Terminal Repeat) elements, with SINEs being the most abundant. SINEs are 80-400 bp long genomic repeats, apparently originating from tRNA (with the exception of human Alu and rodent B1 families which are derived from 7SL cytoplasmatic RNA; Okada et al. 1991). A typical SINE is flanked by short direct repeats and consists of three regions: a tRNA-related region, which contains an internal promoter for RNA polymerase III; a central family-specific or tRNA-unrelated region; and an A-rich tail (Fig. 1; for a review of SINEs, see Kramerov and Vassetzky 2005). In carnivores, the tRNA-derived SINEs (tRNA SINEs) are thought to derive from tRNA<sup>Lys</sup> (Vassetzky and Kramerov 2002). They are also characterized by a polypyrimidine region (poly-Y) in their central region, and a polyadenylation AATAAA signal and the RNA polymerase III TTTT or TCTT terminator in the A-rich tail (Pecon-Slattey et al. 2000; Vassetzky and Kramerov 2002). The tRNA-related region is more conserved than the unrelated part and the RNA polymerase III promoter is the most conserved, while poly-Y and A-rich tail are highly variable both in sequence and in length (Vassetzky and Kramerov 2002). At first, it was thought that tRNA SINE distribution was limited to doglike carnivores (Canoidea superfamily; e.g.: dogs, bears, raccoons, weasels, skunks, and seals). However, recent works (Pecon-Slattey et al. 2000; Vassetzky and Kramerov 2002) have detected the tRNA SINE in the genome of several species of catlike carnivores (Feloidea superfamily; e.g.: hyenas, cats, mongooses, and civets) but not beyond carnivores.



**Fig. 1.** Typical structure of a carnivore tRNA SINE with two promoter elements for RNA polymerase III (A-box and B-box), a polypyrimidine region and an A-rich tail with polymerase III termination signal (underlined). Direct repeats (DR) that result from the insertion process lie at both termini of the interspersed repeat. Restriction site for *Sau3AI* enzyme is indicated.

In this study we constructed a comprehensive database of carnivore MS clones from GenBank to explore the role of interspersed repeats in the generation of repeat arrays. We focused on tRNA SINEs which constitute the best characterized and most abundant interspersed repeat element in the carnivore genome. We observed two modes of MS genesis in two regions of tRNA SINEs in which MSs have repeatedly evolved. This observation led us to hypothesize that SINEs generating MSs could explain part of the different repeat array content and distribution among and within genomes. We also demonstrated that MSs associated with tRNA SINEs were more complex and had less potential instability than those not associated. The observations reported here therefore have practical implications since the use of MSs related with tRNA SINEs calls for special attention from the designing of experiments to the interpretation of results.

## 2. Material and Methods

### *(i) Construction of a sequence database of MS-containing clones*

We constructed a non-redundant database of clone sequences of carnivore MSs obtained from GenBank (Release 146.0). Firstly, we performed an *Entrez* (<http://www.ncbi.nlm.nih.gov/Entrez/>) MS search limited to Genomic DNA and to carnivores. We also arbitrarily limited our sequences to a minimum length of 200 bp to ensure the possibility of tRNA SINE detection and to obtain a reasonably sized dataset of identified MSs for analysis. We removed clone sequences which were not obtained following the traditional method of MS isolation – i.e. isolation from partial genomic libraries (selected for small insert size) or from microsatellites-enriched libraries of the species of interest, and screening several thousands of clones through colony hybridization with repeat-containing probes (Zane et al. 2002). To ensure this, we checked all MS publications and, for unpublished entries, we asked the authors to provide information concerning their MS isolation procedures. Secondly, we used the CLEANUP program (Grillo et al. 1996; <http://bighost.area.ba.cnr.it/BIG/CleanUP/>) to detect and eliminate duplicated sequences and to facilitate the

identification of different alleles of the same locus, as well as loci shared among species. Finally, we used the NCBI tool VecScreen (<http://www.ncbi.nlm.nih.gov/VecScreen/>) to eliminate possible vector contamination. Sequences which were reduced to below 200 bp after removing vector fragments were deleted from the database.

*(ii) Computer screening of the database for clones carrying interspersed repeats*

We used the RepeatMasker program (Smit et al. unpublished data; RepeatMasker open-3.0.8 at <http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker>) to detect and classify interspersed repeats in the clone database. To do this, RepeatMasker makes use of Repbase Update (Jurka 2000) which is a comprehensive database of repetitive element consensus sequences. We ran RepeatMasker at low speed using the interspersed repeat carnivore library as DNA source. In order to find possible repetitive DNA sequences not included in Repbase Update, the clones in which RepeatMasker did not find interspersed repeat sequences were analyzed by pairwise comparisons with the Basic Local Alignment Search Tool (BLAST; Altschul et al. 1997; <ftp://ftp.ncbi.nlm.nih.gov/blast/>).

To compare the repeat array content and evaluate the association between MSs and interspersed repeats, we further subdivided MS-containing clones into three different databases based on the RepeatMasker output: i) non-masked clones – i.e. not containing interspersed repeats; ii) clones masked as tRNA<sup>Lys</sup>-derived SINEs; and iii) the remaining masked clones– i.e. intimately associated with other interspersed repeats.

We roughly estimated the number of tRNA SINE copies for different carnivore species from the proportion of that element in bank sequences following the equation used by Bentolila et al. (1999):  $N = n \times 3 \times 10^9 / L$ ; where  $n$  is the number of tRNA SINEs found in a species with a total length ( $L$ ) represented in the database, and  $3 \times 10^9$  states for the haploid length of a mammalian genome.

*(iii) Computer identification and characterization of repeat arrays.*

To identify all repeat arrays in the MS-containing clones following a standard criterion, we used the modified version of the Sputnik program (Abajian, unpublished; <http://espressoftware.com/pages/sputnik.jsp>) used in Morgante et al. (2002) and accessible in <http://capb.dbi.udel.edu/main/tools.htm>. We looked for motifs of 1 to 5 bases repeated at least three times and with a total length of at least 12 bases. We allowed up to 10% variation between MS and a perfectly repeated motif of the same length (designated as imperfect and perfect MSs, respectively). We also considered a compound MS when two consecutive repeats detected by Sputnik were separated by no more than three consecutive non-repeat bases (Weber 1990). In all analyses, each repeat array of compound MSs was treated as an independent unit, unless comparison of compound MSs among databases was performed, in which case, the whole compound repeat was considered as a single array. Classification of MS sequences was carried out according to their repeat unit outputted by Sputnik, including all permutations on both strands (e.g., AAG represents the following: AAG, AGA, GAA, CTT, CTC, and TTC). Thus, the total number of theoretically possible repeat units is 151. Since two regions of tRNA SINEs have been associated with MSs (Vassetzky and Kramerov 2002), we classified the MSs associated with tRNA SINEs into three subtypes whenever possible: i) those which were positioned 3' or 5' from a transposable element; ii) those which had arisen at an internal sequence (poly-Y) and thus had transposable element sequences on both flanks; and iii) those which were part of the A-rich tail (Fig. 1).

*(iv) Statistical analyses.*

To evaluate the consistency with which repeat motifs were represented among databases – and to overcome the problem of low cell counts –, we used a Monte Carlo approximation ( $N = 100,000$ ) of Fisher's exact test. If differences were observed, we then used Fisher's exact tests to compare differences in specific motifs. Significance levels were adjusted using the standard Bonferroni method to take into account multiple tests on the same data set. To compare the length of the repeat array between

different databases, we performed a Mann-Whitney U-test, after first conducting a normality test. To investigate whether complex MS structures (e.g., compound and imperfect MSs) were associated with the presence of interspersed repeats, we compared the abundance of complex MSs found in the different databases using Fisher's exact test. All statistical analyses were performed with SPSS v11.0.1 (SPSS Inc.).

### 3. Results

#### *(i) Database description*

We obtained a non-redundant database of 1236 MS-containing clones for a total length of 515,359 bp representing 33 species from 11 of the 12 carnivore families (see Table 1). The Mephitidae family was not included. However, there was a clear overrepresentation of domestic cat (*Felis catus*), contributing 40.8% of the total number of clones and 33.6% of total length, and domestic dog (*Canis familiaris*), with 15.0% and 23.7%, respectively. The rest of the species contributed less than 8% of the two values, number of clones and sequence length.

#### *(ii) Identification and characterization of interspersed repeat elements*

RepeatMasker masked 687 (55.5%) clones; of these, 330 contained tRNA SINEs, 292 LINEs, 93 LTR elements, 75 MIRs (mammalian-wide interspersed repeat), and 48 DNA transposons. It was possible for the same clone to contain more than one kind of interspersed repeat. In the 330 tRNA SINE-containing clones, we found 362 tRNA SINEs of which 47 were full-size. Thirty-two clones therefore contained two SINEs, in 27 of which they were oriented in the same way. We also explored the non-masked clone sequences using pairwise comparisons with BLAST to search for potentially new repeat elements but no new elements were found.

We obtained a rough estimate of the number of tRNA SINE copies per haploid genome for those species with the highest number of clones (Table 1). We estimated  $2.1 \times 10^6$  tRNA SINEs in cat,  $9.1 \times 10^5$  in dog,  $2.3 \times 10^6$  in badger,  $2.0 \times 10^6$  in giant panda, and  $4.5 \times 10^6$  in spotted hyena.

However, these values may not be representative of their genomes because of different biases inherent in methods of microsatellite isolation (see Discussion section).

**Table 1.**  
**Distribution of MS-containing Clones by Species.**

Species	Family	Clones	bp	SINEs	Other repeats	Articles	Journals
<i>Felis catus</i>	Felidae	504(40.8%)	173240(33.6%)	125	130	3	2,8,11,17
<i>Canis familiaris</i>	Canidae	185(15.0%)	122264(23.7%)	37	67	19	2,6,10,12,14,17
<i>Meles meles</i>	Mustelidae	82(6.6%)	33008(6.4%)	25	29	3	2,15
<i>Ailuropoda melanoleuca</i>	Ursidae	80(6.5%)	40181(7.8%)	27	33	2	1,5,17
<i>Crocuta crocuta</i>	Hyaenidae	75(6.1%)	21993(4.3%)	33	10	3	3,15,16,17
<i>Mustela vison</i>	Mustelidae	46(3.7%)	21642(4.2%)	11	14	3	4,15,17
<i>Panthera leo</i>	Felidae	30(2.4%)	12733(2.5%)	15	2	1	16,17
<i>Lutra lutra</i>	Mustelidae	21(1.7%)	8652(1.7%)	3	9	3	7,15,16,17
<i>Odobenus rosmarus</i>	Odobenidae	14(1.1%)	9232(1.8%)	1	5	1	15
<i>Zalophus californianus</i>	Otariidae	14(1.1%)	8027(1.6%)	2	6	1	16,17
<i>Herpestes javanicus</i>	Herpestidae	13(1.1%)	3725(0.7%)	7	1	1	16
<i>Ursus arctos</i>	Ursidae	13(1.1%)	5201(1.0%)	0	6	1	15
<i>Panthera tigris</i>	Felidae	12(1.0%)	3114(0.6%)	6	2	1	16,17
<i>Mungos mungo</i>	Herpestidae	11(0.9%)	3791(0.7%)	1	1	1	16
<i>Martes americana</i>	Mustelidae	11(0.9%)	2979(0.6%)	6	2	1	15
<i>Halichoerus grypus</i>	Phocidae	11(0.9%)	2822(0.5%)	0	5	2	12,15
<i>Ursus americanus</i>	Ursidae	11(0.9%)	3607(0.7%)	0	6	2	15,17
<i>Leptonychotes weddellii</i>	Phocidae	10(0.8%)	3257(0.6%)	4	0	2	13,16
<i>Ailurus fulgens</i>	Procyonidae	10(0.8%)	3395(0.7%)	2	3	1	16
<i>Lontra canadensis</i>	Mustelidae	9(0.7%)	3825(0.7%)	4	1	1	16
<i>Suricata suricatta</i>	Herpestidae	9(0.7%)	4971(1.0%)	2	3	1	16
<i>Mustela erminea</i>	Mustelidae	8(0.6%)	2514(0.5%)	1	4	2	12,15
<i>Ursus thibetanus</i>	Ursidae	8(0.6%)	3460(0.7%)	2	5	1	15
<i>Phoca vitulina</i>	Phocidae	7(0.6%)	2099(0.4%)	1	4	3	2,15
<i>Fossa fossana</i>	Viverridae	6(0.5%)	1585(0.3%)	1	1	1	16
<i>Gulo gulo</i>	Mustelidae	6(0.5%)	1337(0.3%)	1	0	2	15
<i>Urocyon cinereoargenteus</i>	Canidae	6(0.5%)	3291(0.6%)	0	3	1	16
<i>Nopex lagopus</i>	Canidae	5(0.4%)	2039(0.4%)	0	3	1	9,14
<i>Cryptoprocta ferrox</i>	Viverridae	5(0.4%)	2388(0.5%)	4	0	1	15
<i>Hydrurga leptonyx</i>	Phocidae	4(0.3%)	1541(0.3%)	4	0	2	13,16
<i>Lobodon carcinophagus</i>	Phocidae	4(0.3%)	1335(0.3%)	1	1	2	13,16
<i>Lynx canadensis</i>	Felidae	3(0.2%)	1254(0.2%)	2	0	1	15
<i>Taxidea taxus</i>	Mustelidae	3(0.2%)	857(0.2%)	2	1	1	15
TOTALS		1236	515359	330	357	71	

SINEs: the number of clones associated with tRNA SINEs. Other repeats: the number of clones associated with other interspersed repeats. Articles: the number of publications in which these clones appeared. Journals: 1, Acta Genetica Sinica ( $N = 1$ ); 2, Animal Genetics ( $N = 19$ ); 3, Behavioural Ecology ( $N = 1$ ); 4, Canadian Journal of Animal Science ( $N = 1$ ); 5, Chinese Journal of Applied and Environmental Biology ( $N = 1$ ); 6, Chromosome Research ( $N = 1$ ); 7, Conservation Genetics ( $N = 1$ ); 8, Cytogenetics and Genome Research ( $N = 1$ ); 9, Electrophoresis ( $N = 1$ ); 10, Genetics ( $N = 1$ ); 11, Genomics ( $N = 1$ ); 12, Journal of Heredity ( $N = 3$ ); 13, Journal of Mammalogy ( $N = 1$ ); 14, Mammalian Genome ( $N = 2$ ); 15, Molecular Ecology ( $N = 16$ ); 16, Molecular Ecology Notes ( $N = 15$ ); and 17, Unpublished; where  $N$  shows the number of publications in each journal.

*(iii) Repeat motifs associated with interspersed repeat elements*

A Sputnik search in our database for MSs revealed 1695 repeat arrays in the 1236 clone sequences. Table 2 shows the most abundant (found >5 times) repeat motifs. We identified 58 out of the 151 possible motifs from monomer to pentamer. The 72.1% of the repeat arrays were dinucleotides and 16.4% were tetranucleotides, while mono-, tri- and pentanucleotides each accounted for less than 5% of the total. None of the representatives of the (CRG)<sub>n</sub> MS family, which are mainly located either within or very close to coding sequences (Stallings 1994), was found. All tri-, tetra- and pentanucleotides have an A in their repeat unit.

When Sputnik was applied to the masked clones, we found that 557 and 454 out of the 1695 repeat arrays were associated with tRNA SINEs and other repeats, respectively. Out of 58 repeat motifs, 10 were not represented in the masked sequences, but this could be due to their low abundance in the whole database (each was found only 1 to 3 times).

*(iv) Implications of tRNA SINEs for the genesis of MSs*

Overall, the abundance of motifs differs among databases ( $P < .0001$ ; Table 2). When pairwise comparisons were conducted, there were no significant differences in motif abundance between non-masked and other repeats clones ( $P = .0862$ ) but there were significant differences between tRNA SINE and the other two databases ( $P$  values  $< .0001$ ).

The tRNA SINE clones had a highly significant lack of AC repeats (22.5%, Fisher exact test,  $P < .0001$ ) and an abundance of AG (65.9%, Fisher exact test,  $P < .0001$ ) and AAAT (86.4%, Fisher exact test,  $P < .0001$ ) repeats compared with the combined values of the other two databases (Table 2). To know in detail how these differences were produced, we classified MSs in relation to the location in the tRNA SINE clones (Fig. 1 and Fig. 2; see also the Supplementary Material Table 1). Again, we found high significant differences in the overall motif distribution within the tRNA SINE clone sequence ( $P < .0001$ ) and in pairwise comparisons among the different regions: poly-Y, A-rich tail and other parts (all three comparisons yielded a  $P < .0001$ ). Poly-Y region had a highly significant abundance of

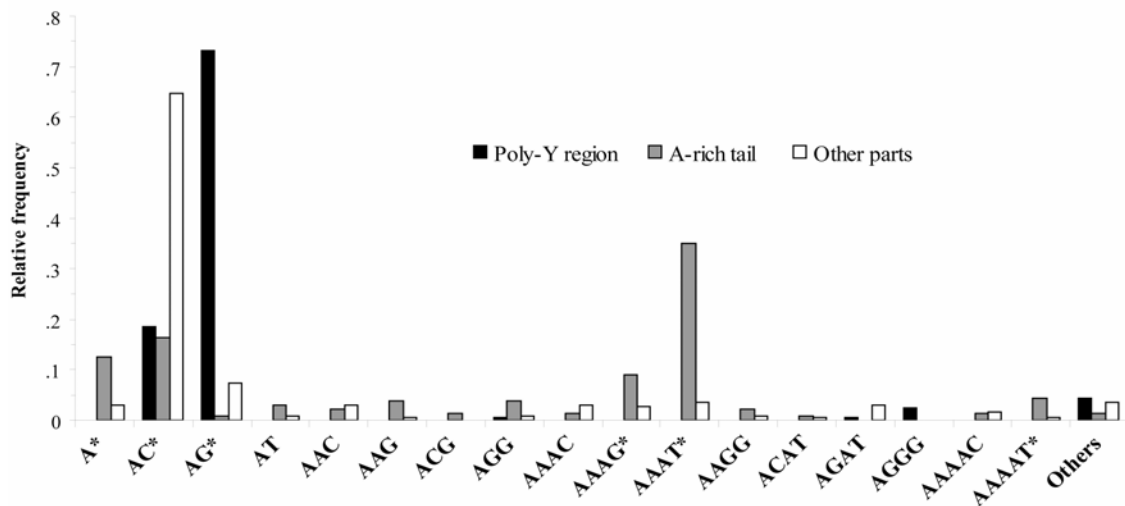
AG repeats (86.8%, Fisher exact test,  $P < .0001$ ) and the A-rich tail had a high number of A (70.8%, Fisher exact test,  $P < .0001$ ), AAAG (66.7%, Fisher exact test,  $P = .0002$ ), AAAT (85.5%, Fisher exact test,  $P < .0001$ ), and AAAAT (85.7%, Fisher exact test,  $P < .0014$ ) repeats. In addition, both regions showed a lack of AC repeats (15.1% and 11.1% respectively, Fisher exact test,  $P$ -values  $< .0001$ ). No MSs were found in any other parts of the SINE sequences.

**Table 2.**  
**Distribution of MSs in the Different Databases.**

Unit	Non-masked	tRNA SINEs	Other repeats	Total	$P$ -value <sup>a</sup>
A	14	24	11	49	0.0179
C	4	0	4	8	0.0590
AC	476	221*	285	982	$< .0001$
AG	45	141*	28	214	$< .0001$
AT	7	6	5	18	$> .9999$
CG	6	0	2	8	0.0590
AAC	5	11	11	27	0.4108
AAG	2	6	2	10	0.0896
ACC	4	0	5	9	0.0349
AGC	6	0	2	8	0.0590
AGG	4	8	0	12	0.0250
AAAC	6	9	2	17	0.1157
AAAG	19	19	16	54	0.7686
AAAT	3	57*	6	66	$< .0001$
AAGG	13	5	16	34	0.0256
ACAG	4	0	5	9	0.0349
ACAT	2	3	2	7	0.6901
ACGC	4	0	2	6	0.1863
AGAT	13	9	19	41	0.1772
AGGG	4	4	0	8	0.4509
ATCC	8	0	3	11	0.0201
AAAAC	6	6	1	13	0.3747
AAAAT	1	6	1	8	0.0180
ACACC	3	0	3	6	0.1863
Others	25	22	23	70	0.8968
Total	684	557	454	1695	

<sup>a</sup>Fisher's exact test for comparisons between specific tRNA SINE motifs and the combined values of the other two databases. Repeat motif frequencies which have a significant departure compared to Bonferroni-corrected alpha for 25 comparisons ( $P$ -value  $< .002$ ) are indicated with an asterisk (\*).





**Fig. 2.** Relative abundance of repeat motifs within tRNA SINE regions: poly-Y ( $N = 161$ ), A-rich tail ( $N = 134$ ), and other parts ( $N = 227$ ). Differences in specific motif abundance were tested using Fisher's exact tests comparing specific region/motif with the combined values of the other two regions. Repeat motif with frequencies which have a significant departure compared to Bonferroni-corrected alpha for 18 comparisons ( $P$ -value  $< .0028$ ) are indicated with an asterisk (\*). Thirty-five MSs were excluded because they were associated with SINEs which did not have a typical structure.

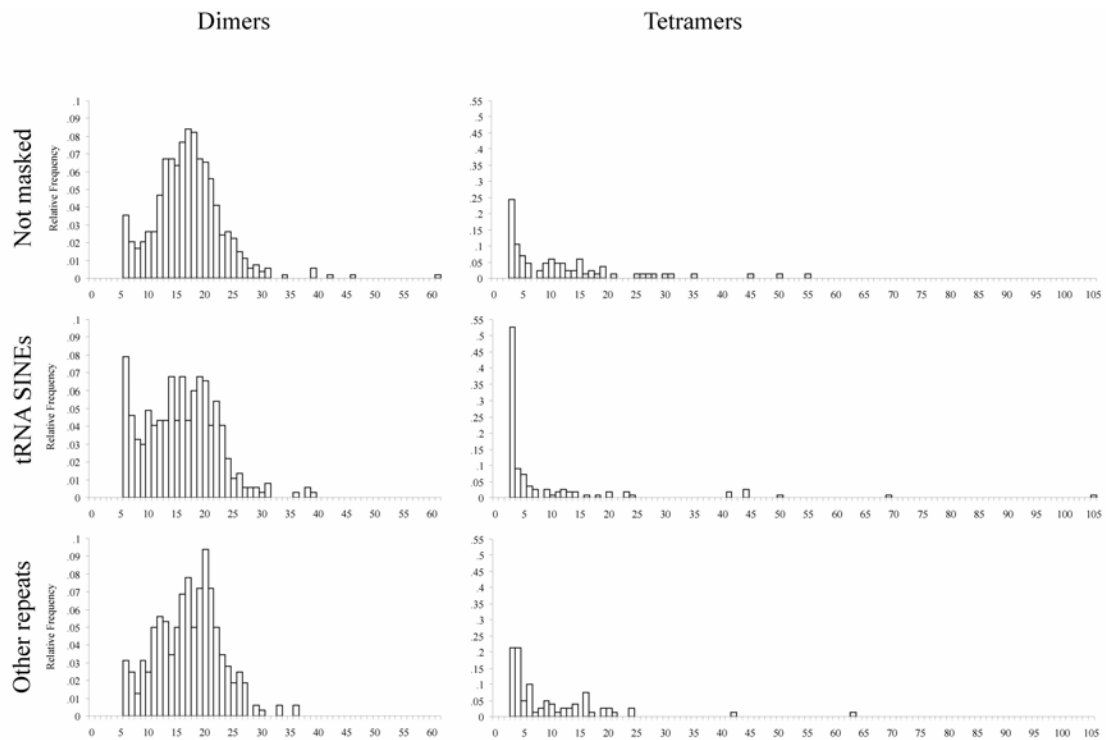
These results were based on the 330 tRNA SINE-containing clones but only 188 of these (15.1% of the total number of clones) had both poly-Y and A-rich tail (Fig. 1). In the other clones, these regions were either not clearly distinguishable ( $N = 25$ ) or there was only a fragment of the tRNA related region ( $N = 117$ ). Taking into account only those clones containing SINEs with both regions, we observed that 65.0% of the poly-Y and 53.2% of the A-rich tail generated MSs. In these clones, 314 MSs were found to be distributed as follows: 51.3% in poly-Y, 42.7% in A-rich tail, and 6.1% downstream from the SINE. In a detailed analysis of the SINE sequence (Fig. 2; Supplementary Material Table 1), the poly-Y tract produced 11 repeat motifs, which were mainly AG (73.3%) repeats. On the other hand, A-rich tail produced a higher motif diversity (17 repeat motifs), although these were limited to A-rich MSs. Thus, although only 188 clones had both regions, these differences in distribution of MSs among databases were mainly produced by poly-Y region which concentrated 55.1% of the AG and only 3.1% of AC repeats found in the whole database and by the A-rich tail which concentrated 71.2% of the AAAT and only 2.2% of AC repeats found in the whole database.

*(v) Implications of tRNA-SINEs for microsatellite instability*

Although we could not measure MS polymorphism, repeat array length correlates well with microsatellite instability. Mean array length can be used to summarize the location of the frequency distribution, although this does not capture the entire spectrum of variation in these non-normal distributions (Fig. 3; Kolmogorov-Smirnov test, Dimers  $N = 1222$ ,  $Z = 1.938$ ,  $P = .0011$ ; Tetramer  $N = 278$ ,  $Z = 4.736$ ,  $P < .0001$ ). On average, non-masked, tRNA SINE and other repeats clones carried, respectively, arrays with 34.109, 31.576 and 34.231 repeats for dimers, and 45.302, 36.893 and 37.200 repeats for tetramers. Over all three databases, dimers (Kruskal-Wallis rank sum test,  $df = 2$ ,  $\chi^2 = 11.381$ ,  $P = .0034$ ) and tetramers (Kruskal-Wallis rank sum test,  $df = 2$ ,  $\chi^2 = 18.993$ ,  $P < .0001$ ) differed in mean array sizes. Within the dimer class, tRNA SINE clones showed shorter arrays than non-masked (Mann-Whitney U test;  $N = 902$ ,  $Z = -2.781$ ,  $P = .0054$ ) and other repeats (Mann-Whitney U test;  $N = 688$ ,  $Z = -3.056$ ,  $P = .0022$ ) clones. Again, within the tetramer class, tRNA SINE clones showed shorter arrays than non-masked (Mann-Whitney U test;  $N = 198$ ,  $Z = -3.924$ ,  $P < .0001$ ) and other repeats (Mann-Whitney U test;  $N = 192$ ,  $Z = -3.383$ ,  $P = .0007$ ) clones.

*(vi) Implications of SINE elements for microsatellite complexity.*

To elucidate whether the contiguous existence of poly-Y and A-rich tail or similar structures in other interspersed repeats may be a source of compound MSs, we compared the presence of this kind of MSs among databases. Out of 550 non-masked clones, 66 contained compound arrays. The number of compounds was not significantly different from that calculated from the tRNA SINE (46 compounds in 330 clones, Fisher's exact test,  $P = .263$ ) and other repeats (50 compounds in 357 clones, Fisher's exact test,  $P = .249$ ) databases. Even when we calculated the number of compounds in the tRNA SINEs database taking into account only clones with both poly-Y tract and A-rich tail, there were no significant



**Fig. 3.** Frequency distribution of repeat array length (number of repeats) for dimer and tetramer motif classes.

differences between databases (data not shown). Moreover, we did not find any significant increase in the number of compound MSs in the SINE database when we changed the number of non-repeat nucleotides separating two adjacent arrays (data not shown). However, the presence of these two regions (poly-Y and A-rich tail) should imply a source of multiple arrays in the same clone. The number of clones with multiple arrays in tRNA SINE sequences was 146 which was significantly higher than the number calculated for non-masked sequences (128 multiples in 550 clones, Fisher's exact test,  $P < .0001$ ) and for other repeats (85 multiples in 357 clones, Fisher's exact test,  $P < .0001$ ).

On the other hand, of the 684 repeat arrays in non-masked clones, 181 were imperfect and this number was not significantly different from the number calculated for tRNA SINEs (156 imperfects, Fisher's exact test,  $P = .2928$ ) and for other repeats (115 imperfects, Fisher's exact test,  $P = .3613$ ).

#### 4. Discussion

Our database was generated from MS-containing clones of carnivore species which are generally used in population genetics and individual identification studies. These MSs had been isolated following the traditional method, which is known to produce a nonrepresentative sample of the genome. Two common procedures repeatedly used in traditional microsatellite isolation that may cause biases in our database are: i) the use of *Sau3AI* restriction enzyme to fragment genomic DNA, and ii) the use of AC probes to screen the libraries. However, and despite the biases inherent in isolation methods, the drawn conclusions are still pertinent or in some cases even reinforced (see below).

*Sau3AI* recognizes sites that are well-conserved in dispersed repeats (e.g., in porcine PRE-1, in rat L1 elements), and in the B-box of the tRNA polymerase III promoter of our tRNA SINEs. As a result, the SINEs found in our databases were often truncated (87.0%), conserving the 5' end in 127 cases and the 3' end, including the poly-Y and A-rich tail regions, in 188 cases. We found 47 full tRNA SINEs which could be used as phylogenetic markers, but only 6 of them were found in catlike species which probably indicates that the B-box is more conserved in these species. This fact has led some authors (e.g., Duffy et al. 1996; Bentolila et al. 1999) to state that the use of *Sau3AI* may result in bias towards the isolation of repeat-associated MSs. However, such a bias could not exist taking into account the high frequency of cutting owing to the restriction recognition site for *Sau3AI* (^GATC) is only four nucleotides in length.

The preferential use of (AC)<sub>n</sub> probes for library screening was reflected in the repeat motif content of our database where 58% were AC repeats, clearly overrepresented (Toth et al. 2000). Despite this, we found a significant lack of AC repeats in tRNA SINE clones: only 5.3% of the AC repeats in the whole database were located within the SINE sequence. Furthermore, 77.5% of AC repeats were not associated with these tRNA SINE clones. Similar results have been found in humans where more than 80% of AC repeats are not associated with Alu sequences (Arcot et al. 1995; Jurka and Pethiyagoda 1995; Nadir et al. 1996). Using AC probes

would therefore reduce the final number of SINEs captured in the MS isolation process.

Our tRNA SINE copy estimates per haploid genome in dog ( $9.1 \times 10^5$ ) were very similar to the values obtained from the dog genome sequence ( $1.06 \times 10^6$ ; Lindblad-Toh et al. 2005). It seems then reasonable to think that the values we obtained for the other carnivore species ( $2.1 \times 10^6$  tRNA SINEs in cat,  $2.3 \times 10^6$  in badger,  $2.0 \times 10^6$  in giant panda, and  $4.5 \times 10^6$  in spotted hyena) may also be realistic approximations. Even in the case where the above mentioned biases do apply – probably reducing tRNA SINE captured –, it is unlikely they would produce a difference of an order of magnitude with published estimates ( $10^5$  -  $3 \times 10^5$  in mustelids,  $2 \times 10^5$  in cats and true seals, and  $1.5 \times 10^5$  -  $4 \times 10^5$  in dogs and bears; Lavrent'eva et al. 1989; Bentolila et al. 1999; Vassetzky and Kramerov 2002; Pecon-Slattery et al. 2004).

We also observed that poly-Y (65.0%) and A-rich tail (53.2%) regularly gave rise to MSs. The different repeat motifs derived from these regions (mostly AG and AAAT, respectively) suggested two mechanisms for MS generation. The first mechanism was illustrated by the A-rich tail and it has already been well described in human Alu sequences (Arcot et al. 1995; Nadir et al. 1996). This mode of genesis implies the presence of a pre-existing MS and subsequent modifications by point mutation and slippage events. It has been suggested that the pre-existing MS could arise from the incorporation of the retrotranscript with an extended polyadenylated tail, a feature which may also serve to guide their retroposition in the genome (Nadir et al. 1996). Although the pre-existing microsatellite was an adenine tract, this evolved into more complex structures where we observe certain variability in the repeat motif, mostly centered in A rich MS (A 12.6% and  $A_{2-4}N$  57.8%).

The second mechanism is based on the fact that the poly-Y region has a nucleotide composition highly biased towards pyrimidines; it could therefore be defined as a site with high cryptic simplicity (Tautz et al. 1986). The functional significance of this structure currently remains unknown, but its presence in these elements makes the SINE an

important source of MS genesis. Generation of MSs at this site depends on base substitutions that create a tandemly duplicated motif, and on subsequent slippage mutations to increase the number of copies. Thus, a few C ↔ T transitions, the most frequent substitutions, are enough to transform cryptic simplicity sequences into tandemly repeated DNA. The initial bias in the base substitution was reflected in the repeat motifs generated, 73.4% of which were AGs. This tendency for invariability and the greater number of MSs produced by the poly-Y region suggest that slippage mutation is active during the early stages of MS genesis. This microsatellite-generating mutation process has also been shown in an internal region of the *mini-me* elements of Dipterans (Wilder and Hollocher 2001). The poly-Y region has also been found to be specific of other mammalian SINEs, such as rabbit C repeat (Cheng et al. 1984), rodent DIP (Serdobova and Kramerov 1998), bat VES (Borodulina and Kramerov 1999), and insectivore TAL, ERI-1 and ERI-2 (Borodulina and Kramerov 2001), showing that the action of the two mechanisms that generate MSs are not exclusive to the tRNA SINEs.

It is known that different microsatellite motifs, motif classes and even abundances are not equally represented in species belonging to different groups (Toth et al. 2000; Katti et al. 2001) or even within the genome of any one particular species (Schug et al. 1998; Kruglyak et al. 2000). These differences are still not well-understood and it has been hypothesized that they may be caused by species-specific differences in the DNA synthesis and repair machinery (Harr et al. 2002), selection (Nauta and Weissing 1996), or base composition (Dieringer and Schlötterer 2003). Although tRNA SINEs are not the only source of MSs in carnivore genomes, these elements could explain part of the differences in the distribution of MSs within a particular genome due to their high abundance and their preference for insertion at specific sites – such as around the R bands (Korenberg and Rykowski 1988; Chen and Manuelidis 1989) and clustering or insertion into other mobile genetic elements (Nadir et al. 1996; San Miguel et al. 1996; Bentolila et al. 1999). They could also explain differences among species or groups, since the interspersed repeat families generating MSs may be lineage specific. Along

these lines, our results would indicate that tRNA SINEs have a significant effect on the overall distribution of some repeat motifs in carnivores, especially AG and AAAT repeats.

The strong association between repeat elements and MSs has been largely used for different purposes, such as: i) to develop new, codominant multiplex marker technologies such as S-SAP (Waugh et al. 1997) or inter-AluPCR (Jarnik et al. 1996); ii) to build microsatellite-enriched libraries by amplifying *Sau3AI* inserts with a conserved SINE primer and a flanking vector primer (Band and Ron 1996); iii) to discover new SINEs, especially in species for which little information is available concerning their repeat element content (Buchanan et al. 1999); and iv) to discover new SINE loci which could be used to reconstruct phylogenies (Shedlock and Okada 2000). In this study we detected 47 completed SINEs whose flanking regions are targets for primer design and could be used as phylogenetic markers.

In spite of these applications, MSs associated with interspersed repeats not only distort estimates of the genomic distribution of MSs useful for genome mapping (Duffy et al. 1996), but also entail some methodological disadvantages. Firstly, genotyping with MSs associated with repeat elements is very hard. Placing one of the PCR primers within a highly repeated element might cause weak amplification, high background and difficulty in locus-specific amplification (Economou et al. 1990; Duffy et al. 1996; Gallagher et al. 1999). Moreover, if the primers were designed up- and downstream from the repeated element, the expected large size of the PCR product might cause problems in the resolution of the amplified products (Yandava et al. 1997). Our results also showed that potential instability in MSs associated with tRNA SINEs was lower than in non-associated MSs. Several studies (e.g., Weber and Wong 1993; Goldstein and Clark 1995; Wierdl et al. 1997) have shown that MS mutation rate increases with an increasing number of repeat units; this is considered the single most important factor affecting the mutation rate. The isolation of non-masked MS clones is therefore advisable on account of their high potential informative content. It has been argued that point mutations

break up perfect repeats and reduce the mutation rates of microsatellites (Kruglyak et al 1998). Since there are not significant differences in the number of imperfect repeat arrays among databases, the higher content of short repeat arrays in MSs associated with tRNA SINE cannot entirely be attributed to imperfections. Finally, most of the applications involving MSs as genetic markers are based on variations in the length of the PCR product, which is expected to vary according to single-step changes in the number of repeats. However, it has been shown that poly-Y is responsible for variation in length in the MS flanking region within species (López-Giráldez et al. unpublished data). This is probably due to the fact that cryptically simple sequences are susceptible to undergo slippage in a similar manner to MSs, but at lower rates (Hancock 1995). We also detected a larger number of clones with multiple MSs in the tRNA SINE database owing to the presence of poly-Y and A-rich tail. This may also explain the non-neutral observation of MS clustering (Bachtrog et al. 1999). Both cases mentioned preclude the basic assumptions of MS mutational models. As a result, the interpretation of the data obtained from MSs associated with tRNA SINEs may induce erroneous conclusions.

In this report we have shown how tRNA SINEs, the most abundant carnivore and a lineage-specific SINE, is clearly responsible for generating an important fraction of carnivore MSs. More specifically, we have demonstrated that not only the A-rich tails but also an internal region (poly-Y) of these elements regularly expand into lengthy MSs via two different mechanisms: the expansion of pre-existing tandem repeats and the conversion of sequences with high cryptic simplicity into tandemly repetitive DNA. The MS genesis in tRNA-SINEs is not only involved in complex patterns, such as multiple repeated arrays and length variation in the flanking regions, and is responsible for shorter repeat arrays, but may also explain differences in MS distribution among and within species genomes. The mechanism we have described in tRNA SINEs may also be generalized to other interspersed repeats. Based on the negative effect of the association between MSs and interspersed repeats, we recommend avoiding the use of these MSs as genetic markers. We suggest applying computer tools after initial sequencing in order to detect interspersed



repeats in MS-containing clones and taking special attention when designing isolation methods (e.g., not using *Sau3AI* and AG or AAAT probes).



## 5. References

- Alexander LJ, Rohrer CA, Stone RT, Beattie CW (1995) Porcine SINE-associated microsatellite markers: evidence for artiodactyls SINEs. *Mamm Genome* 6: 464-468.
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389-3402.
- Arcot SS, Wang Z, Weber JL, Deininger PL, Batzer MA (1995) Alu repeats: a source for the genesis of primate microsatellites. *Genomics* 29: 136-144.
- Bachtrog D, Weiss S, Zangerl B, Brem G, Schlötterer C (1999) Distribution of dinucleotide microsatellites in the *Drosophila melanogaster* genome. *Mol Biol Evol* 16: 602-610.
- Band M, Ron M (1996) Creation of a SINE enriched library for the isolation of polymorphic (AGC)<sub>n</sub> microsatellite markers in the bovine genome. *Anim Genet* 27: 243-248.
- Bentolila S, Bach JM, Kessler JL, Bordelais I, Cruaud C, Weissenbach J, Panthier JJ (1999) Analysis of major repetitive DNA sequences in the dog (*Canis familiaris*) genome. *Mamm Genome* 10: 699-705.
- Borodulina OR, Kramerov DA (1999) Wide distribution of short interspersed elements among eukaryotic genomes. *FEBS Lett* 457: 409-413.
- Borodulina OR, Kramerov DA (2001) Short interspersed elements (SINEs) from insectivores. Two classes of mammalian SINEs distinguished by A-rich tail structure. *Mamm Genome* 12: 779-786.
- Buchanan FC, Littlejohn RP, Galloway SM, Crawford AM (1993) Microsatellites and associated repetitive elements in the sheep genome. *Mamm Genome* 4: 258-264.

- Buchanan F, Crawford A, Strobeck C, Palsbøll P, Plante Y (1999) Evolutionary applications of MIRs and SINEs. *Anim Genet* 30: 47-50.
- Chen TL, Manuelidis L (1989) SINEs and LINEs cluster in distinct DNA fragments of Giemsa band size. *Chromosoma* 98: 309.
- Cheng JF, Printz R, Callaghan T, Shuey D, Hardison RC (1984) The rabbit C family of short, interspersed repeats. Nucleotide sequence determination and transcriptional analysis. *J Mol Biol* 176: 1-20.
- Dieringer D, Schlötterer C (2003) Two distinct modes of microsatellite mutation processes: evidence from the complete genomic sequences of nine species. *Genome Res* 13: 2242-2251.
- Duffy AJ, Coltman DW, Wright JM (1996) Microsatellites at a common site in the second ORF of L1 elements in mammalian genomes. *Mamm Genome* 7: 386-387.
- Economou EP, Bergen AW, Warren AC, Antonarakis SE (1990) The poly (A) tract of Alu repetitive elements is polymorphic in the human genome. *Proc Natl Acad Sci USA* 87: 2951-2954.
- Gallagher PC, Lear TL, Coogle LD, Bailey E (1999) Two SINE families associated with equine microsatellite loci. *Mamm Genome* 10: 140-144.
- Goldstein DB, Clark AG (1995) Microsatellite variation in North American populations of *Drosophila melanogaster*. *Nucleic Acids Res* 23: 3882-3886.
- Grillo G, Attimonelli M, Liuni S, Pesole G (1996) CLEANUP: a fast computer program for removing redundancies from nucleotide sequence databases. *Comput Appl Biosci* 12: 1-8.
- Harr B, Todorova J, Schlötterer C (2002) Mismatch repair driven mutational bias in *D. melanogaster*. *Mol Cell* 10: 199-205.
- Hancock JM (1995) The contribution of DNA slippage to eukaryotic nuclear 18S rRNA evolution. *J Mol Evol* 40: 629-639.

- Jarnik M, Tang J-Q, Korab-Lasowska M, Zietkiewicz E, Cardinal G, Gorska-Flipot I, Sinnott D, Labuda D (1996) Overall informativity, *OI*, in DNA polymorphisms revealed by inter-*Alu* PCR: detection of genomic rearrangements. *Genomics* 36: 388–398.
- Jurka J (2000) Repbase Update: a database and an electronic journal of repetitive elements. *Trends Genet* 16: 418-420.
- Jurka J, Pethiyagoda C (1995) Simple repetitive DNA sequences from primates: compilation and analysis. *J Mol Evol* 40: 120-126.
- Katti MV, Ranjekar PK, Gupta VS (2001) Differential distribution of simple sequence repeats in eukaryotic genome sequences. *Mol Biol Evol* 18: 1161–1167.
- Korenberg JR, Rykowski MC 1988. Human genome organization: *Alu*, lines, and the molecular structure of metaphase chromosome bands. *Cell* 53: 391.
- Kramerov DA, Vassetzky NS (2005) Short retroposons in Eukaryotic Genomes. *Int Rev Cytol* 247: 165-221.
- Kruglyak S, Durrett RT, Schug MD, Aquadro (1998) Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. *Proc Natl Acad Sci USA* 95: 10774-10778.
- Lavrent'eva MV, Rivkin MI, Shilov AG, Kobets ML, Rogozin IB, Serov OL (1989) B2-like repetitive sequence in the genome of the American mink. *Dokl Akad Nauk SSSR* 307: 226-228.
- Levinson G, Gutman G (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol* 4: 203–221.
- Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, Clamp M, Chang JL, Kulbokas EJ 3rd, Zody MC, et al. (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438: 803-819.

- Messier W, Li SH, Stewart CB (1996) The birth of microsatellites. *Nature* 381: 483.
- Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nat Genet* 30: 194-200.
- Nadir E, Margalit H, Gallily T, Ben-Sasson SA (1996) Microsatellite spreading in the human genome: evolutionary mechanisms and structural implications. *Proc Natl Acad Sci USA* 93: 6470-6475.
- Nauta, MJ, Weissing FJ (1996) Constraints on allele size at microsatellite loci: Implications for genetic differentiation. *Genetics* 143: 1021-1032.
- Okada N (1991) SINEs: short interspersed repeated elements of the eukaryotic genome. *Trends Ecol Evol* 6: 358-361.
- Pecon-Slattery J, Murphy WJ, O'Brien SJ (2000) Patterns of diversity among SINE elements isolated from three Y-chromosome genes in carnivores. *Mol Biol Evol* 17: 825-829.
- Pecon-Slattery J, Pearks Wilkerson AJ, Murphy WJ, O'Brien SJ (2004) Phylogenetic assessment of introns and SINEs within the Y chromosome using the cat family felidae as a species tree. *Mol Biol Evol* 21: 2299-2309.
- Ramsay L, Macaulay M, Cardle L, Morgante M, Degli Ivanissevich S, Maestri E, Powell W, Waugh R (1999) Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. *Plant J* 17: 415-425.
- Rose O, Falush D (1998) A threshold size for microsatellite expansion. *Mol Biol Evol* 15: 613-615.
- SanMiguel P, Tikhonov A, Jin YK, et al (11 co-authors) (1996) Nested retrotransposons in the intergenic regions of the maize genome. *Science* 274: 765-768.



- SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z, Bennetzen JL (1996) Nested retrotransposons in the intergenic regions of the maize genome. *Science* 274: 765-768.
- Schlötterer C (2000) Evolutionary dynamics of microsatellite DNA. *Chromosoma* 109: 365- 371.
- Schug MD, Wetterstrand KA, Gaudette MS, Lim RH, Hutter CM, Gaudette MS, Aquadro CF (1998) The distribution and frequency of microsatellite loci in *Drosophila melanogaster*. *Mol Ecol* 7: 57-70.
- Serdobova IM, Kramerov DA (1998) Short retroposons of the B2 superfamily: evolution and application for the study of rodent phylogeny. *J Mol Evol* 46: 202-214.
- Shedlock AM, Okada N (2000) SINE insertions: powerful tools for the molecular systematics. *BioEssays* 22: 148-160.
- Smit AFA, Hubley R, Green P. *RepeatMasker Open-3.0*. (1996-2004). <http://www.repeatmasker.org>.
- Stallings RL (1994) Distribution of trinucleotide microsatellites in different categories of mammalian genomic sequence: implications for human genetic diseases. *Genomics* 21: 116-121.
- Tautz D, Trick M, Dover G (1986) Cryptic simplicity in DNA is a major source of genetic variation. *Nature* 322: 652-656.
- Toth G, Gaspari Z, Jurka J (2000) Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res* 10: 967-981.
- Vassetzky NS, Kramerov DA (2002) CAN-a pan-carnivore SINE family. *Mamm Genome* 13: 50-57.
- Wahle E (1995) 3'-End cleavage and polyadenylation of messenger-RNA precursors. *Biochim Biophys Acta* 1261: 183-194.

Waugh R, McLean K, Flavell AJ, Pearce SR, Kumar A, Thomas BBT, Powell W (1997) Genetic distribution of Bare-1-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (SSAP). *Mol Gen Genet* 253: 687–694.

Weber JL (1990) Informativeness of human  $(dC-dA)_n \cdot (dG-dT)_n$  polymorphisms. *Genomics* 7: 524-530.

Weber JL, Wong C (1993) Mutation of human short tandem repeats. *Hum Mol Genet* 2: 1123–1128.

Wierdl M, Dominska M, Petes TD (1997) Microsatellite instability in yeast: dependence on the length of the microsatellite. *Genetics* 146: 769-779.

Wilder J, Hollocher H (2001) Mobile elements and the genesis of microsatellites in Dipterans. *Mol Biol Evol* 18: 384-392.

Yandava CN, Gastier JM, Pulido JC, Brody T, Sheffield V, Murray J, Buetow K, Duyk GM (1997) Characterization of Alu repeats that are associated with trinucleotide and tetranucleotide repeat microsatellites. *Genome Res* 7: 716-724.

Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. *Mol Ecol* 11: 1-16.

Zhu Y, Strassmann JE, Queller DC (2000) Insertions, substitutions, and the origin of 9 microsatellites. *Genet Res Camb* 76: 227-236.

**Acknowledgments.** We thank M. Hanafey and J. Abril for their help with modified Sputnik and F. Calafell, A. Ferrando, R. Lecis, R. Rodríguez and A. Ruíz for their useful comments on the manuscript. We also thank all researchers who shared information with us on their microsatellite isolation methods. This research was funded by the European Commission under contract (INPRIMAT QLRI-CT-2002-01325). During the development of the project, FL-G and OA received scholarships from the Departament d'Universitats, Recerca i Societat de la Informació, Generalitat de Catalunya (refs. 2001FI00625 and 2003FI00787, respectively).

**Supplementary Material Table 1.**  
**Distribution of MSs Associated with tRNA SINEs.**

Unit	Poly-Y region	A-rich tail	Other parts	Total	P-value <sup>a</sup>	P-value <sup>b</sup>	P-value <sup>c</sup>
A	0*	17*	7	24	0.0002	<.0001	0.2054
AC	30*	22*	147*	199	<.0001	<.0001	<.0001
AG	118*	1*	17*	136	<.0001	<.0001	<.0001
AT	0	4	2	6	0.1842	0.0403	0.7015
AAC	0	3	7	10	0.0358	0.7219	0.1112
AAG	0	5	1	6	0.1842	0.0050	0.2399
ACG	0	2	0	2	>.9999	0.0655	0.5075
AGG	1	5	2	8	0.4450	0.0295	0.4757
AAAC	0	2	7	9	0.0630	>.9999	0.0451
AAAG	0*	12*	6	18	0.0014	0.0002	0.4715
AAAT	0*	47*	8*	55	<.0001	<.0001	<.0001
AAGG	0	3	2	5	0.3302	0.1097	>.9999
ACAT	0	1	1	2	>.9999	0.4479	>.9999
AGAT	1	0	7	8	0.4450	0.1209	0.0240
AGGG	4	0	0	4	0.0088	0.5767	0.1363
AAAAC	0	2	4	6	0.1842	0.6495	0.4110
AAAAT	0	6*	1	7	0.1058	0.0014	0.1448
Others	7	2	8	17	0.4232	0.2611	0.8069
Total	161	134	227	522 <sup>d</sup>			

Fisher's exact test for comparisons between specific poly-Y<sup>a</sup>, A-rich tail<sup>b</sup> or other parts<sup>c</sup> motifs and the combined values of the other two regions. Repeat motif frequencies which have a significant departure compared to Bonferroni-corrected alpha for 18 comparisons ( $P$ -value < .0028) are indicated \*. <sup>d</sup>Thirty-five MSs were excluded because they were associated with SINEs which did not have a typical structure.



## **Capítol IV**

***Genetic distinction of American and European mink (Mustela vison and M. lutreola) and European polecat (M. putorius) hair samples by detection of a species-specific SINE and a RFLP assay***

Francesc López-Giráldez, Benjamín J. Gómez-Moliner,  
Josep Marmi i Xavier Domingo-Roura

(publicat en *Journal of Zoology*, Lond. (2005) 265: 405–410)

**GENETIC DISTINCTION OF AMERICAN AND EUROPEAN MINK (*MUSTELA VISON* AND *M. LUTREOLA*) AND EUROPEAN POLECAT (*M. PUTORIUS*) HAIR SAMPLES BY DETECTION OF A SPECIES-SPECIFIC SINE AND A RFLP ASSAY.**

Francesc López-Giráldez<sup>1,2\*</sup>, Benjamín J. Gómez-Moliner<sup>3</sup>, Josep Marmi<sup>2</sup>, Xavier Domingo-Roura<sup>1,2</sup>

<sup>1</sup>*Genètica de la Conservació, Institut de Recerca i Tecnologia Agroalimentàries, Centre de Cabriels, Carretera de Cabriels s/n, 08348 Cabriels, Spain.*

<sup>2</sup>*Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Dr. Aiguader 80, 08003 Barcelona, Spain.*

<sup>3</sup>*Departamento de Zoología y Biología Celular Animal, Universidad del País Vasco, Paseo de la Universidad 7, 01006 Vitoria, Spain.*

\*Corresponding author: Genètica de la Conservació, Institut de Recerca i Tecnologia Agroalimentàries, Centre de Cabriels, Ctra. de Cabriels s/n, 08348 Cabriels (Barcelona), Spain. Tel: +34 937507511. Fax: +34 937533954. e-mail: [francesc.lopez@gmail.com](mailto:francesc.lopez@gmail.com)

---

**Abstract.** The European mink *Mustela lutreola* is the focus of extensive conservation programmes whereas the American mink *M. vison* is exotic in Europe and a possible cause of European mink decline. Since both species, together with the European polecat *M. putorius* have similar hair and produce similar faeces, molecular techniques are essential for reliable species identification when disturbance is not appropriate. In this report a simple, cheap and reliable molecular method for the genetic distinction of mink is provided based on the amplification of the microsatellite locus *Me/08* that results in a longer PCR product (436 bp) in American mink than in other mustelid species, such as the European mink (221 bp) and the European polecat (221 bp). Sequencing these PCR products revealed, in the microsatellite-flanking region of American mink, a CAN-SINE insertion of 213 bp that was absent in 22 species of the Canoidea superfamily tested. When the *Me/08* PCR product is cut with restriction enzyme *AcI* and then run in an agarose gel, the restriction pattern observed is different in the three *Mustela* spp. The method was successfully tested with DNA amplified from a large sample of hair roots, thus making it suitable for non-invasive sampling.

---

## 1. Introduction

The genus *Mustela* includes 16 species distributed across Eurasia, America and North Africa (Macdonald 2001). This group of small and middle-sized carnivores includes weasels, polecats, and mink. In Europe, the European polecat *Mustela putorius*, the European mink *M. lutreola*, and the exotic American mink *M. vison* constitute the guild of semiaquatic mustelids, together with the European otter *Lutra lutra*. Molecular phylogenetic analyses have identified the European polecat as the most closely related congener of the European mink, their existence as separate species has been debated (Davison et al. 1999; Marmi et al. 2004), and the two species may hybridize in the wild (Tumanov & Abramov 2002). On the other hand, the American mink is the most basal member of the genus *Mustela* and is clearly separated from the rest of species of this genus (Bininda-Emonds et al. 1999; Marmi et al. 2004). The two mink species and the European polecat are sympatric in many parts of their respective ranges; mainly in forested wet areas with brooks, small rivers and lakes (Blanco 1998). The morphology of these three species is similar and, if the animal is not captured, they are often confused with each other when sighted in the field (Blanco 1998). These three species show very similar body size and weight for males and females, although both sexes of European mink are slightly smaller than American mink and polecats (Macdonald 2001). Mink have incomplete inter-digital membranes, which are not present in the European polecat. The latter species has white spots on the muzzle, above the eyes and at the edges of the ears, whereas a white patch on the upper lip, characteristic of the European mink, is also found in 10–20% of American mink (Macdonald 2001), making it difficult to distinguish the two mink species. Moreover, the three species are nocturnal, secretive and evasive and their faeces and hairs are difficult to distinguish in the field. Thus, genetic identification of the animals from their remains could become essential in the study and management of their populations.

Management priorities vary considerably between the three species. The European polecat is relatively widespread throughout Europe, although its

range is declining in several countries and it remains vulnerable to habitat loss and persecution (Birks & Kitchener 1999). The range of the European mink has shrunk to fragmented populations in Russia, Belarus, Romania, France and Spain; it is now considered to be one of Europe's most endangered carnivores and a priority for conservation throughout its distribution (Maran et al. 1998). The role of the American mink and various other factors in the decline of the European mink has been reviewed by Maran & Henttonen (1995) and by Macdonald et al. (2002). The American mink is an invasive alien species that became naturalized throughout Europe, and Central and East Asia, after it escaped from fur farms. It is therefore considered undesirable throughout Europe (Genovesi 2000), where eradication programmes have been implemented at both regional and national levels. Legal protection is strict for the European mink and somewhat less so for the polecat (Council of the European Communities 1992; International Union for Conservation of Nature and Natural Resources 2003).

While exploring the evolution of microsatellite loci among carnivore species, amplification of the locus *Me108*, originally isolated from the Eurasian badger *Meles meles* (Domingo-Roura 2002), was found to highlight an unexpectedly long polymerase chain reaction (PCR) product in American mink. Sequencing this PCR product revealed that the extra length could be attributed to the insertion of a short interspersed element (SINE) in the flanking region of *Me108* microsatellite. SINEs are mobile elements of < 500 base pairs (bp) that transpose through an RNA intermediate. Since their transposition is considered to be a unique event in evolution, SINEs are discrete characters whose presence/absence at orthologous loci in different species can be used to resolve phylogenetic relationships (Shimamura et al. 1997).

In this work the hypothesis was tested that a SINE has been inserted in the locus *Me108* after the separation of the American mink from other mustelids and, thus can be used as a species-specific marker for distinguishing exotic American mink from endangered European mink. Since European polecat remains are also similar to those of mink, our

study is complemented by identifying restriction sites in the *Me/08* PCR product that allow genetic differentiation of the European mink and the European polecat. The protocol was also tested with hair samples to assess its reliability with samples obtained non-invasively in the field, for instance using hair trapping, a technique successfully implemented for monitoring mustelids (e.g., Foran et al. 1997).

## 2. Materials and Methods

The present study analysed 2 different sets of samples (Table 1). The first consisted of good-quality DNA (undegraded and with high copy number) from 30 European mink, 39 American mink and 14 European polecats. This DNA was extracted using a standard phenol-chloroform protocol from muscle, liver or kidney tissue obtained from road kills or museums, and stored for several years either frozen (-20 °C), or in 96% ethanol. To test the feasibility of the same protocol using non-invasive samples that usually contain a reduced amount of degraded DNA, DNA extracted from samples that included at least 15 rooted hairs from a single individual was also analysed. This sample set consists of 75 European mink, 13 American mink and 7 European polecats (Table 1). Hair samples were obtained from dead individuals and from animals live-trapped during other field experiments a few weeks before the analyses. No animal was killed for the sake of this project. Individuals were assigned to species by clear morphological patterns (Youngman 1982; Blanco 1998) and according to the known distribution of each species. The same 5 European mink were included in both sets of samples to test for consistency of the technique.

DNA was amplified using the PCR primers for locus *Me/08* and following the procedure described by Domingo-Roura (2002). Amplifications were performed in a 15  $\mu$ l reaction volume; 2  $\mu$ l of PCR reaction product was then run in a 1.5% agarose gel to differentiate between animals with and without the SINE insertion. Gels were visualized using ethidium bromide (final concentration of 0.5  $\mu$ g/ml in the gel) and inspected on a transilluminator (Fig. 1a).



**Table 1.** Species included in the study, their geographic origin, the kind of tissue used, the number of samples digested with *Aci I*, and the University which processed the samples (UPF: Pompeu Fabra University or UPV: University of Pais Vasco).

Origin	Nº of animals	Tissue	<i>Aci I</i>	University
<b><i>Mustela lutreola</i> (n = 100)</b>				
Russia	3	Muscle	All	UPF
	9	Liver	2	UPV
Belarus	3	Muscle	All	UPF
France	6	Hair	All	UPV
Spain	3	Muscle	1	UPV
	3	Kidney	All	UPV
	4	Liver	3	UPV
	4	Muscle / Hair	All	UPV
	1	Ear / Hair	None	UPV
	64	Hair	32	UPV
<b><i>Mustela vison</i> (n = 52)</b>				
Russia	3	Muscle	None	UPF
Belarus	3	Muscle	1	UPF
Finland	10	Muscle	None	UPF
Holland	6	Muscle	1	UPF
Spain	4	Muscle	1	UPF
	13	Hair	6	UPV
United Kingdom	13	Muscle	None	UPF
<b><i>Mustela putorius</i> (n = 21)</b>				
Russia	2	Muscle	All	UPF
Belarus	1	Muscle	All	UPF
France	1	Muscle	All	UPF
	3	Hair	All	UPV
Spain	2	Muscle	All	UPF
	1	Liver	All	UPV
	4	Hair	All	UPV
United Kingdom	7	Muscle	All	UPF

To determine the sequence and size of the PCR product, a new amplification was performed in a 25  $\mu$ l reaction volume to sequence 5 American mink, 1 European mink and 2 European polecats. The PCR product was purified using Gene Clean (Qbiogene) and sequenced in both directions using a BigDye Terminator Cycle Sequencing Kit (Applied Biosystems) following the instructions of the manufacturer. Any excess of BigDye™ terminators was removed from the sequencing product using Centri-Sep (Princeton Separations Inc.). Sequences were run on an ABI3100 DNA sequencer (Applied Biosystems) and aligned by eye (Fig. 2).

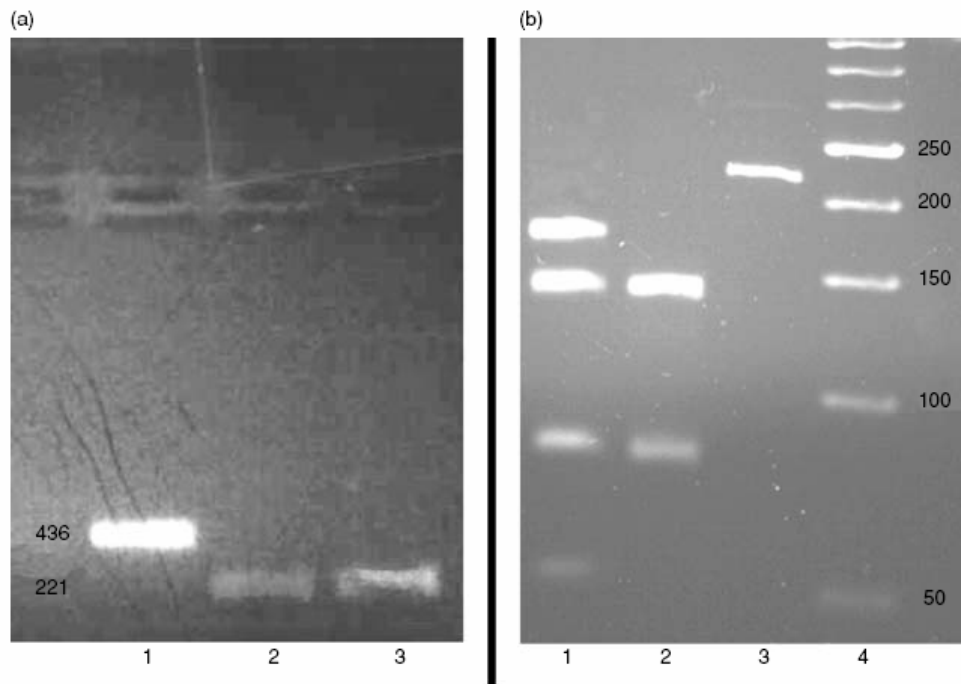
All sequences have been deposited in the GenBank/EMBL database under accession numbers AJ489560, AJ489562, and AY530949.

A single restriction enzyme *AcI* was identified as suitable for cutting at variable nucleotide positions when comparing European mink and European polecat sequences. In particular this enzyme recognized the presence of a CCGC sequence twice in the European polecat PCR product, but did not identify the same sequence in the European mink PCR product (Fig. 2). To confirm the suitability of *AcI* for differentiating the PCR products of these 2 species, 10  $\mu$ l of PCR product from 57 European mink, 21 European polecats, and 9 American mink (see Table 1) was mixed with 5 units of *AcI* restriction enzyme (New England BioLabs), 1.2  $\mu$ l of 10X buffer (100 mM NaCl, 50 mM Tris-HCl (pH 7.9), 10 mM MgCl<sub>2</sub>, and 1 mM DTT) provided by the manufacturer, and distilled water for a final volume of 12  $\mu$ l. The digestion mix was incubated for 4–12 h at 37 °C. Five  $\mu$ l of digested product were run in a 3% low-melting agarose gel. Gels were visualized using ethidium bromide (final concentration of 0.5  $\mu$ g/ml in the gel) and inspected on a transilluminator (Fig. 1b). Since the SINE and the restriction pattern were clearly species specific in all cases, no further analyses were performed.

### 3. Results

A total of 100 European mink, 52 American mink and 21 European polecats was analysed in this study. PCR products were obtained from all samples. A 1.5% agarose gel electrophoresis was sufficient to identify two distinctly different bands corresponding to the PCR products before digestion (Fig. 1a). The larger amplicon was present in all American mink and the smaller amplicon was present in all European mink and European polecats.

Sequencing revealed that the larger amplicon found exclusively in the American mink was 436 bp, whereas the smaller amplicon was 221 bp. The alignment of *Me/08* sequences for the three species is shown in Fig. 2. There were a total of eight variable positions in that portion of the locus that could be aligned in all three species. No intraspecific or interspecific



**Fig. 1.** *Me08* amplicon of an American mink *M. vison* (lane 1), a European polecat *M. putorius* (lane 2), and a European mink *M. lutreola* (lane 3) separated on a 1.5% agarose gel (a) and digested with the restriction enzyme *AcI* and separated on a 3% low-melting agarose gel (b). Sizes of PCR products (base pairs) and of the bands of a DNA weight marker (lane 4) are shown.

variation was detected in the microsatellite repeat number among European mink or European polecats. There were only two variable positions between the two sequences, and the *AcI* enzyme was able to recognize two restriction sites in all European polecats ( $n = 21$ ), producing three fragments of 7, 78 and 136 bp. The two larger fragments could be visualized in the 3% low-melting agarose gel, whereas a single band of 221 bp was visualized in all European mink samples tested ( $n = 57$ ). Digestion of the American mink amplicon ( $n = 9$ ) yielded four fragments of 48, 80, 139 and 169 bp (Fig. 1b). The alignment also revealed an insertion of 213 bp, exhibiting the typical structure of a CAN-SINE in the American mink sequence (Fig. 2). The insertion is limited by direct repeats that are duplications of the insertion site. When the sequence was tested for homology in GenBank/EMBL, the insertion showed an 89% sequence identity with a B-2-like repeat of the American mink (Lavrentieva et al. 1991), the first described member of the CAN-SINE family. The five





sequences obtained from the American mink samples were identical and the only differences that occurred between sequences were in the number of Cs repeated, from  $C_{10}$  to  $C_{13}$  in the longest perfect repeat. The 436 bp amplicon had a  $C_{10}$  repeat (Fig. 2) and will equate to 439 bp in animals with the  $C_{13}$  repeat. Finally, all five European mink whose DNA was extracted both from tissue and hair samples produced the same pattern.

The predicted *AcI* digestion pattern was also analysed for the *MeI08* sequences available from GenBank/EMBL for other mustelids either from genus *Mustela* or with similarities to our focal species. The number and sizes of the fragments should be: three fragments of 8, 79 and 135 bp for the black-footed ferret *M. nigripes* (AJ489561); three fragments of 7, 81 and 135 bp for the mountain weasel *M. altaica* (AJ489563); a single fragment of 223 bp for the ermine *M. erminea* (AJ489564); three fragments of 7, 80 and 135 bp for the least weasel *M. nivalis* (AJ489566); three fragments of 31, 88 and 104 bp for the pine marten *Martes martes* (AJ489568); two fragments of 31 and 192 bp for the beech marten *Martes foina* (AJ489570); and two fragments of 85 and 135 bp for the European otter (AJ489574). The interspecific variability in the repeat tract for these species is very low and intraspecific variability in repeat tract is expected to be almost null. Therefore, *AcI* could be expected to differentiate successfully between the pine marten and the beech marten, which are similar and sympatric species, and both of these from all other species tested. Whereas *MeI08* sequences for the European otter and the European polecat exhibit similar digestion patterns, the *AcI* enzyme clearly differentiates European otters from mink.

#### 4. Discussion

SINEs have been identified and characterized in many mammalian species and they constitute 5–10% of the mammalian genome (Rogozin et al. 2000). CAN-SINEs are present in all carnivore families (Vassetzky & Kramerov 2002) and it has been estimated that the mink genome has  $1-2 \times 10^5$  CAN-SINE copies (Lavrantieva et al. 1991). A-rich tails in Alu sequences (another family of SINEs found in primates) have also been

considered a source of microsatellites (Arcot et al. 1995). In common with other retroposons, SINEs have a polypyrimidine region and an A-rich tail which might be a target during the hybridization step in microsatellite library construction. However, *Me/08* has different repetitive motifs (Cs, Ts and CAs) in different carnivores and none of them is directly related to the polypyrimidine tract or the A-rich tail of a typical CAN-SINE. In addition, the CAN-SINE insertion was only found in the American mink and, thus, is not responsible for generating the repetitive motifs in locus *Me/08*.

The SINE insertion described here in locus *Me/08* was found in all 52 American mink and was absent in all 100 European mink, and all 21 European polecats analysed in this study. Animals from separate geographic origins were included among the samples of both mink species and European polecat to maximize the detection of possible polymorphisms. Thus, it seems that the SINE insertion is specific to American mink, being absent in other species of the Mustelidae family. This finding indicates that the transposition occurred as a unique event in evolution, and that it is essentially an irreversible character state, arising after the American mink separated from the rest of the mustelids. The evolution of cytochrome *b* in the Mustelidae family suggests that American mink separated from all other *Mustela* species *c.* 6.6–9.5 Ma (Marmi et al. 2004).

The protocol successfully resolved the species of origin in all of the samples in this study (including 95 hair roots). Therefore, it is a non-ambiguous, reliable and robust method for the genetic identification of American mink. For this species, it is not subject to possible failures of restriction enzymes, to possible sequence polymorphisms that would modify the results of RFLPs, or to mutations generated by the polymerase during PCR amplification. The failure of one allele to amplify (allelic dropout), often found when working with low-quantity DNA (Taberlet et al. 1996), would not affect the reliability of our procedure since the locus *Me/08* has been found to be monomorphic in both muscle and hair samples for each species. There could also be errors performed by the polymerase generating new false alleles that are one or a few repeats

longer or shorter than expected by slippage. It is thought that false alleles would not affect the performance of our protocol (if it existed it is unlikely it would be detected in the agarose gels) since the SINE would still be present and it is unlikely that slippage would result in mutations affecting the restriction sites. In addition, no evidence was found of any of false alleles generated by slippage in the individuals sequenced.

The technique is not only useful for detecting the presence of the exotic American mink (which is of high ecological concern), but it can also identify the presence of European mink or European polecats in a given area in combination with RFLP techniques. Both species are protected, and the European mink in particular requires strict protection and the designation of special areas to promote its conservation (Council of the European Communities 1992). However, since the separation of the European mink and the European polecat has been quite recent and hybrids do occur, the existence of shared ancestral polymorphisms cannot be completely discarded, specially when shared mitochondrial haplotypes and lack of reciprocal monophyly were found for these two species (Davison et al. 2000). The distinction of the European mink and the European polecat is based on two different nucleotide positions (and it is unlikely that both positions would show intraspecific polymorphisms or homoplasy at the same time to affect the protocol) and the molecular patterns described in this work have been found to be species-specific and fixed within species, behaving as good species diagnostic characters. In countries such as the U.K., where there are no European mink, the simple amplification of locus *Me108* will be useful for distinguishing between the American mink and the European polecat, a native species currently undergoing a population recovery after being near extinction (Langley & Yalden 1977; Birks & Kitchener 1999).

Since *Me108* is a nuclear locus, our protocol in combination with mitochondrial RFLP or sequencing could be used to detect possible hybrids between the species involved, in particular between the closely related European mink and European polecat, species known to hybridize

(Tumanov & Abramov 2002). However, the method will need to be tested in known hybrids before it can be used in the wild.

The advantage of our new method over that of amplifying mitochondrial DNA followed by either cutting with restriction enzymes or sequencing of the PCR product (Awise 1994; Hansen & Jacobsen 1999; Gómez-Moliner et al. 2004) is that our method requires only a single PCR amplification and an agarose electrophoresis gel to identify American mink samples. The new technique therefore provides a far cheaper and faster alternative. Use of a single enzyme, *AccI*, to distinguish between European mink and the European polecat also provides a cheaper alternative to the combination of restriction enzymes that is often currently applied (e.g., Paxinos et al. 1997; Gómez-Moliner et al. 2004). However, the fact that mitochondrial DNA sequences occur at a higher frequency per cell than do nuclear DNA sequences and the relative larger size of the amplification product will limit the applicability of this method to samples yielding DNA of low quality and low quantity. Nevertheless, amplification and further digestion of the nuclear *Me/08* locus can be applied to DNA from non-invasively collected hair. DNA has previously been successfully extracted and amplified from a single freshly plucked human hair (Higuchi et al. 1988). Also, endangered northern hairy-nosed wombats *Lasiiorhinus krefftii* have been individually identified using microsatellites extracted from single hairs obtained by sticky tape (Sloane et al. 2000). In our study at least 15 hair roots per individual were used, but the use of <10 hairs has been shown to result in allelic dropout problems (Goossens et al. 1998). The technique could be further tested using faeces or hair that has been shed naturally, although shed hair contains only a fraction of the DNA isolated from freshly plucked hair (Higuchi et al. 1988). Faecal material is the most abundant and generally available source of non-invasively collected biological material. However, faeces are more likely to contain substances that damage DNA, inhibit the activity of polymerases or introduce errors into the amplification procedure (Taberlet et al. 1999), although protocols to extract DNA from faeces are becoming increasingly efficient, enhancing the possibility to amplify nuclear loci.

Combining efficient hair-trapping techniques with our new molecular protocol provides a significantly more cost-effective alternative to other currently available techniques (including capturing animals) for identification of mink species and European polecats. Hair trapping, however, will need to be optimized for each species. Such a method could greatly enhance the efficiency of both American mink removal exercises (where these are deemed necessary) and the conservation efforts favouring endangered mustelid species.

## 5. References

- Arcot SS, Wang Z, Weber JL, Deininger PL, Batzer MA (1995) *Alu* repeats: a source for the genesis of primate microsatellites. *Genomics* 29: 136–144.
- Avise JC (1994) Speciation and hybridization. In *Molecular markers, natural history and evolution*. Chapman & Hall, New York.
- Bininda-Emonds ORP, Gittleman JL, Purvis A (1999) Building large trees by combining phylogenetic information: a complete phylogeny of the extant Carnivora (Mammalia). *Biol Rev* 74: 143–175.
- Birks JDS, Kitchener A (1999) *The distribution and status of the polecat Mustela putorius in Britain in the 1990s*. Vincent Wildlife Trust, London.
- Blanco JC (1998) *Guía de campo de mamíferos de España 1, Insectívoros, quirópteros, primates y carnívoros de la península Ibérica, Baleares y Canarias*. Planeta, Barcelona.
- Council of the European Communities (1992) Directive on the conservation of natural and semi-natural habitats and of wild fauna and flora (92/43/EEC) *Official J Counc Eur Commun* L206: 7–50.
- Davison A, Birks JDS, Griffiths HI, Kitchener AC, Biggins D, Butlin RK (1999) Hybridization and the phylogenetic relationship between polecats and domestic ferrets in Britain. *Biol Conserv* 87: 155–161.
- Davison A, Griffiths HI, Brookes RC, Maran T, Macdonald DW, Sidorovich VE, Kitchener AC, Irizar I, Villate I, González-Esteban J, et al. (2000) Mitochondrial DNA and palaeontological evidence for the origins of endangered European mink, *Mustela lutreola*. *Anim Conserv* 4: 345–355.
- Domingo-Roura X (2002) Genetic distinction of marten species by fixation of a microsatellite region. *J Mammal* 83: 907–912.
- Foran DR, Minta SC, Heinemeyer KS (1997) DNA-based analysis of hair to identify species and individuals for population research and monitoring. *Wildl Soc Bull* 25: 840–847.

- Genovesi P (2000) *Guidelines for eradication of terrestrial vertebrates: a European contribution to the invasive alien species issue*. Council of Europe, Strasbourg.
- Gómez-Moliner BJ, Cabria MT, Rubines J, Garin I, Madeira MJ, Elejalde A, Aihartza J, Fournier P, Palazón S (2004) PCR-RFLP identification of mustelid species: European mink (*Mustela lutreola*), American mink (*M. vison*) and polecat (*M. putorius*) by analysis of excremental DNA. *J Zool* 262: 311–316.
- Goossens B, Waits LP, Taberlet P (1998) Plucked hair samples as a source of DNA: reliability of dinucleotide microsatellite genotyping. *Mol Ecol* 7: 1237–41.
- Hansen MM, Jacobsen L (1999) Identification of mustelid species: otter (*Lutra lutra*), American mink (*Mustela vison*) and polecat (*Mustela putorius*), by analysis of DNA from faecal samples. *J Zool* 247: 177–181.
- Higuchi R, Von Beroldingen CH, Sensabaugh GF, Erlich HA (1988) DNA typing from single hairs. *Nature* 332: 543–546.
- International Union for Conservation of Nature and Natural Resources (2003) *2003 IUCN red list of threatened species* (<http://www.redlist.org>). IUCN, Gland.
- Langley PJW, Yalden DW (1977) The decline of the rarer carnivores in Great Britain during the nineteenth century. *Mamm Rev* 7: 95–116.
- Lavrentieva MV, Rivkin MI, Shilov AG, Kobetz ML, Rogozin IB, Serov OL (1991) B2-like repetitive sequence from the X chromosome of the American mink (*Mustela vison*). *Mamm Genome* 1: 165–170.
- Macdonald DW (2001) *The new encyclopaedia of mammals*. Oxford University Press, Oxford.
- Macdonald DW, Sidorovich VE, Maran T, Kruuk H (2002) *European mink, Mustela lutreola: analyses for conservation*. Wildlife Conservation Research Unit, Oxford.



- Maran T, Henttonen H (1995) Why is the European mink (*Mustela lutreola*) disappearing? –A review of the process and hypothesis. *Ann Zool Fenn* 32: 47–54.
- Maran T, Macdonald DW, Kruuk H, Sidorovich V, Rozhnov VV (1998) The continuing decline of the European mink, *Mustela lutreola*: evidence for the intraguild aggression hypothesis. *Symp Zool Soc Lond* 71: 297–323.
- Marmi J, López-Giráldez JF, Domingo-Roura X (2004) Phylogeny, evolutionary history and taxonomy of the Mustelidae based on sequences of the cyt *b* gene and a complex repetitive-flanking region. *Zool Scr* 33: 481– 499.
- Paxinos E, McIntosh C, Ralls K, Fleischer R (1997) A non-invasive method for distinguishing among canid species: amplification and enzyme restriction of DNA from dung. *Mol Ecol* 6: 483–486.
- Rogozin IB, Mayorov VI, Lavrentieva MV, Milanesi L, Adkinson LR (2000) Prediction and phylogenetic analysis of mammalian short interspersed elements (SINEs). *Brief Bioinform* 1: 260–274.
- Shimamura M, Yasue H, Ohshima K, Abe H, Kato H, Kishiro T, Goto M, Munechika I, Okada N (1997) Molecular evidence from retroposons that whales form a clade within eventooed ungulates. *Nature* 388: 666–670.
- Sloane MA, Sunnucks P, Alpers D, Beheregaray LB, Taylor AC (2000) Highly reliable genetic identification of individual northern hairy-nosed wombats from single remotely collected hairs: a feasible censusing method. *Mol Ecol* 9: 1233–1240.
- Taberlet P, Griffin S, Goossens B, Questiau S, Manceau V, Escaravage N, Waits LP, Bouvet J (1996) Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Res* 15: 3189–3194.
- Taberlet P, Waits LP, Luikart G (1999) Noninvasive genetic sampling: look before you leap. *Trends Ecol Evol* 14: 323–327.



Tumanov IL, Abramov AV (2002) A study of the hybrids between the European mink *Mustela lutreola* and the polecat *M. putorius*. *Small Carniv Conserv* 27: 29.

Vassetzky NS, Kramerov DA (2002) CAN – a pan-carnivore SINE family. *Mamm Genome* 13: 50–57.

Youngman PM (1982) Distribution and systematics of the European mink *Mustela lutreola* Linnaeus 1761. *Acta Zool Fenn* 166: 1–48.

**Acknowledgements.** We thank G. Barreto, A. Grogan, and N. Yamaguchi (WildCRU, University of Oxford), H. Jansman (Alterra), A. Kitchner (Nat. Museums of Scotland), M. Miralles (Rectoria Vella), S. Pérez (Taxidermia El Ciervo), C. Rosell (Minuartia), V. Sidorovic (National Academy of Science of Belarus); and the projects “Conservación del visón Europeo (*Mustela lutreola*) en La Rioja” LIFE 00/NAT/E 7331, “Conservación del visón Europeo (*Mustela lutreola*) en Álava” LIFE 00/NAT/E 7335, and “Conservación del visón Europeo (*Mustela lutreola*) en Castilla y León” LIFE 00/NAT/E 7229 for providing the samples used in this study. S. Baker (WildCRU, University of Oxford), provided editorial assistance. This work has been supported by grants from Fundació Territori i Paisatge, Caixa de Catalunya (Spain) and Comissionat per a Universitats i Recerca de la Generalitat de Catalunya (Spain) (ref. 2000SGR00093). During the development of the project, JFL-G and JM received scholarships from the Departament d’Universitats, Recerca i Societat de la Informació de la Generalitat de Catalunya (refs. 2001FI00625 and 2000FI00698, respectively).

# **DISCUSSIÓ**

A cada capítol de l'apartat de Resultats he presentat una discussió corresponent i específica del treball que s'hi presenta. En aquest apartat exposaré tot un seguit de reflexions generals sobre els darrers avanços que s'han produït paral·lelament a la realització d'aquesta tesi sobre l'evolució de microsatèl·lits i l'ús de marcadors moleculars en genètica de la conservació. A més a més, discutiré sobre l'aplicació en genètica de la conservació dels elements repetitius que hem vist en aquesta tesi comparant-los amb altres marcadors. Per finalitzar aquesta discussió, exposaré algunes reflexions sobre el futur dels microsatèl·lits, que han estat el fil conductor del meu treball.

## **1 ORIGEN I EVOLUCIÓ DELS MICROSATÈL·LITS**

Una de les motivacions per estudiar els patrons evolutius en microsatèl·lits de teixó aïllats en el nostre laboratori era la identificació i selecció dels millors loci microsatèl·lit per realitzar estudis poblacionals. Per això, a part de seleccionar els marcadors més variables, també volíem seleccionar marcadors que complissin les assumpcions dels models evolutius pels quals s'han desenvolupat la majoria d'índexs estadístics (p.e., *Stepwise Mutation Model*). Entendre el procés mutacional dels microsatèl·lits és essencial abans d'inferir les relacions existents entre la variació observada i les distàncies genètiques o l'estructura d'una població.

### **1.1 Models mutacionals**

Els dos models bàsics d'evolució de microsatèl·lits, l'SMM i l'IAM (vegeu la secció 2.3 de la Introducció), han estat provats amb dades empíriques i amb simulacions per ordinador. Tot i que l'IAM té una hipòtesi nul·la matemàticament convenient i que hi ha estudis que no poden refutar-lo com a model evolutiu (p.e., Anderson et al. 2000), els microsatèl·lits no evolucionen d'acord amb aquest model. Les dades obtingudes de mutacions directes i d'anàlisis estadístiques (Weber & Wong 1993; Kayser et al. 2000) indiquen que les mutacions en els



microsatèl·lits són del tipus SMM. Aquest model segueix una sèrie d'assumpcions: i) els guanys i pèrdues de repeticions es produeixen amb una taxa constant; ii) només hi ha un procés de guany o pèrdua per esdeveniment; i iii) no hi ha direccionalitat en les mutacions.

A mesura que s'anaven obtenint seqüències de microsatèl·lits i dades de la variabilitat en les poblacions es va veure que les desviacions d'aquest model eren comuns i van començar a aparèixer modificacions d'aquest model bàsic. Aquestes modificacions inclouen: i) la introducció d'un *two-phase stepwise mutation model* (TPM; Di Rienzo et al. 1994), ii) permetre variacions en la taxa de mutació entre loci i la introducció de límits en la grandària dels al·lels (Feldman et al. 1997), iii) processos mutacionals esbiaixats (Kimmel et al. 1996; Zhivotovsky et al. 1997), i iv) la limitació de grandària per deleció o mutació (Garza et al. 1995; Kruglyak et al. 1998, 2000). Normalment, els paràmetres d'aquests models matemàtics han estat testats contra mesures de variabilitat, tals com heterozigositat, variància en el nombre de repeticions i asimetria de la distribució de la longitud dels microsatèl·lits, que són observades dins de les poblacions.

Més recentment, els models han estat testats contra la distribució dels microsatèl·lits en *sets* de dades genòmiques (Renwick et al. 2001; Dieringer & Schlötterer 2003). Les dades obtingudes en el treball de Dieringer & Schlötterer (2003), que van analitzar seqüències de genomes complets de 9 espècies, indiquen que les modificacions del model bàsic no podrien explicar els patrons observats en la distribució dels microsatèl·lits. Els seus resultats suggereixen que els microsatèl·lits curts evolucionen per un procés similar a l'*slippage*, però aquest seria independent de la llargada del tracte repetitiu. Aquesta dada es pot justificar per les observacions que les insercions sovint copien la seqüència flanquejant, cosa que crea microsatèl·lits curts (Zhu et al. 2000; Nishizawa & Nishizawa 2002). D'aquesta manera, amb aquest model, Dieringer & Schlötterer (2003) van mostrar que amb la combinació de l'*slippage* que succeeix en repeticions curtes amb substitucions nucleotídiques i l'*slippage* dependent de llargada en repeticions llargues es podia obtenir una distribució de microsatèl·lits similar a les seves observacions.

Les dades utilitzades en aquests models han estat obtingudes directament de les seqüències, però hem de recordar que els al·lels són comunament identificats únicament per la llargada del producte de PCR, detectada via mobilitat electroforètica, i aquests fragments, a part de la regió repetitiva, també tenen una regió flanquejant. Quan realitzem aquest tipus de genotipatge assumim que la variació en la mida dels al·lels només està produïda per variacions en el nombre de repeticions. Tot i així, al mateix temps que els models d'evolució adquirien més paràmetres, l'anàlisi de seqüències d'al·lels de microsatèl·lit (p.e., Capítols I i II de Resultats, Primmer et al. 1997; Colson & Goldstein 1999; Makova et al. 2000) demostrava que les diferències en el nombre de repeticions no eren l'única forma de variació entre els al·lels d'un locus. Els al·lels també poden diferir per interrupcions, tals com mutacions puntuals en el tracte repetitiu per substitucions nucleotídiques i/o indels en la regió flanquejant. Juntament amb aquestes mutacions complexes, també ens podem trobar amb diverses regions repetitives variables en un mateix locus microsatèl·lit que poden crear confusions en el genotipatge (vegeu Capítol I de Resultats). No obstant, una de les dades que podem extreure de la seqüenciació dels al·lels microsatèl·lits analitzats en el Capítol I de Resultats és que, tot i la seva complexitat, la fracció perfecta de la regió repetitiva mostra una clara correlació amb la variabilitat i, per tant, podria ser modelada. A més a més, sense haver de considerar la llargada total de la regió repetitiva, amb només la fracció perfecta, continuaríem tenint un bon indicador de la variabilitat esperada i, per tant, un indicador dels loci microsatèl·lits més interessants.

Aquesta variació addicional present als loci microsatèl·lits pot provocar que productes de llinatges evolutius diferents siguin classificats com al·lels de la mateixa mida. Aquest fenomen és el que coneixem com a homoplàsia de grandària (vegeu la secció 2.4 de la Introducció). Aquests tipus de polimorfismes només poden ser detectats per seqüenciació i poden passar inadvertits en cas d'analitzar individus mitjançant amplificació per PCR i electroforesi. Els loci de microsatèl·lits compostos són especialment susceptibles a presentar homoplàsia de grandària, ja que al tenir dues regions amb potencial de variació, dos al·lels de la



mateixa mida poden ser resultat de la combinació de diferent nombre de repeticions de les dues regions. S'ha observat que l'homoplàsia de grandària és una característica comuna dels loci microsatèl·lit i ha estat detectada en individus de la mateixa població (p.e., Viard et al. 1998; Makova et al. 2000), de diferents poblacions de la mateixa espècie (p.e., Orti et al. 1997; Taylor et al. 1999) i entre espècies (p.e., Garza & Freimer 1996; Angers & Bernatchez 1997). En teoria, en els microsatèl·lits s'espera certa quantitat d'homoplàsia, ja que, com dèiem, els microsatèl·lits presenten un patró de mutació del tipus SMM, una elevada taxa de mutació i un rang limitat de mides. Aquestes característiques redueixen les possibilitats dels estats al·lèlics i afavoreixen, d'aquesta manera, certa quantitat d'homoplàsia (Nauta & Weissing 1996; Estoup & Cournuet 1999). Hem d'entendre, doncs, que l'homoplàsia que detectem seqüenciant és només una fracció de l'homoplàsia total, ja que, tot i tenir seqüències idèntiques, pot ser que dos al·lèls aparentment idèntics no ho siguin per descendència. L'homoplàsia pot ser una font d'error en estudis poblacionals o d'evolució perquè la taxa de mutació dels microsatèl·lits està relacionada de forma complexa amb la mida i la composició de l'al·lel. En estudis on s'ha detectat homoplàsia a partir d'anàlisis moleculars, s'ha vist que aquesta redueix el nombre d'al·lèls per població, el nombre d'individus heterozigots i la diversitat gènica (Estoup et al. 1995; Garza & Freimer 1996; Jarne & Lagoda 1996; Angers & Bernatchez 1997; Viard et al. 1998; Taylor et al. 1999; Angers et al. 2000). En alguns estudis, la detecció de l'homoplàsia de grandària ha resultat en la detecció d'una major diferenciació significativa entre poblacions quan, a més dels electromorfs, també es consideren les seqüències (Viard et al. 1998; Angers et al. 2000). Tot i que hi ha varis estudis que s'han realitzat per detectar l'homoplàsia a nivell molecular, la relació amb l'homoplàsia total és difícil de definir i només s'ha descrit per simulacions per ordinador (Estoup et al. 2002). Amb aquests nous mètodes analítics es conclouïa que l'homoplàsia no era un problema significatiu per a moltes anàlisis genètiques si s'utilitzava un model evolutiu apropiat i amb el suficient nombre de loci variables. En aquest treball, es predeïa que l'homoplàsia de grandària era problemàtica només en situacions que involucraven mides poblacionals grans, taxes de mutacions elevades i fortes



constriccions en la mida dels al·lels. Per tant, el guany associat amb la detecció d'homoplàsia en estudis de genètica de poblacions sembla ser mínim en la majoria de casos i no justificaria els esforços de seqüenciació, *a posteriori*, per eliminar-la. No obstant, els resultats presentats en aquest treball (vegeu el Capítol I i III de Resultats) suggereixen que és convenient seleccionar, *a priori*, els marcadors que fem servir i obtenir un nombre elevat de microsatèl·lits variables per tal de disminuir els efectes de l'homoplàsia (Estoup et al. 2002).

## ***1.2 Naixement i distribució dels microsatèl·lits en els genomes***

Com hem vist en els capítols I i III de Resultats, encara hi ha un factor més que complica l'evolució dels microsatèl·lits, els elements transponibles. Els llocs de naixement dels microsatèl·lits poden sorgir per atzar (vegeu secció 2.2 de la Introducció), però aquests llocs també poden ser proporcionats, amb més freqüència del que ens imaginem, per certs elements transponibles que funcionarien com a generadors de microsatèl·lits. Al mateix temps, sembla que certs microsatèl·lits faciliten la transposició actuant com a llocs d'inserció (Ramsay et al. 1999) o com a navegadors de la retrotransposició (Nadir et al. 1996). Ja fa més d'una dècada que se sap que hi ha certa relació entre l'evolució dels retroposons i els microsatèl·lits (Buchanan et al. 1993; Alexander et al. 1995; Arcot et al. 1995; Nadir et al. 1996). Tanmateix, existeixen molt pocs estudis que analitzen el nivell d'influència dels elements transponibles com a precursors dels microsatèl·lits i encara es tenen menys en compte quan s'intenta modelar la seva evolució. Per exemple, Nadir et al. (1996) van analitzar 2.84 Mb del genoma humà i van trobar que els microsatèl·lits rics en A (>70%) estaven associats amb els SINEs Alu. D'una manera similar, Arcot et al. (1995) també van detectar aquesta associació amb seqüències humanes i de ximpanzés. Ambdós grups proposen que les cues de poli-A presents en els elements retrotransposats generarien la majoria de microsatèl·lits rics en A/T. Els autors hipotetitzen que aquesta cua poli-A es produeix per la poliadenilació quan el retrotransposó es troba en forma



d'RNA, de la mateixa manera que en els mRNAs. Recentment, també s'ha vist que la cua de poli-A està en relació amb l'activitat dels elements repetitius dispersos i són els inserits més recentment els que presenten cues més llargues (Odom et al. 2004). En estudis més recents, s'ha demostrat que aquesta relació no estava limitada a microsatèl·lits rics en adenines ni al seu extrem 3' terminal. Ramsay et al. (1999) també van trobar, en plantes, una associació amb l'extrem 5' i regions internes d'alguns retroposons. Wilder & Hollocher (2001) també van trobar una associació entre microsatèl·lits i regions internes dels retroposons *mini-me* dels dípters.

Tot i que no hi ha estudis que ho demostrin, per la gran abundància dels retroposons en els genomes, la coevolució entre aquests elements i els microsatèl·lits podrien explicar diferents patrons evolutius en els microsatèl·lits. A continuació detallaré un seguit d'observacions que s'han realitzat sobre la distribució dels microsatèl·lits en els genomes i com aquests fenòmens poden ser explicats, almenys en part, pels elements transponibles.

De les anàlisis dels genomes complets realitzats per Dieringer & Schlötterer (2003), s'observa que els microsatèl·lits llargs estan sobrerepresentats. Curiosament, la longitud a partir de la qual s'observa aquesta major sobrerepresentació varia entre tipus de repeticions i entre espècies. Tot i que el model d'*slippage* assumeix que els microsatèl·lits llargs són generats a partir de mutacions a l'atzar que allarguen o escurcen els microsatèl·lits – i que ocasionalment resulten en un nombre llarg de repeticions –, altres escenaris són possibles, sota els quals les repeticions llargues són generades de cop pels elements transponibles. De forma similar, també s'ha descrit la generació *de novo* de microsatèl·lits GT durant la reparació no homòloga dels trencaments de la doble cadena (Liang et al. 1998).

Generalment, es creu que la densitat dels microsatèl·lits tendeix a estar positivament correlacionada amb la mida del genomes (Primmer et al. 1997; Hancock 1999; Toth et al. 2000). Tanmateix, també trobem excepcions a aquesta correlació en espècies properes. Tot i que el genoma

humà és més gran que el del ratolí, aquesta espècie presenta de 2 a 3 vegades més microsatèl·lits (*Mouse Genome Sequencing Consortium* 2002). El fet que la distribució dels microsatèl·lits sigui diferent en espècies relacionades pot ser un fenomen generalitzat. De fet, s'han trobat diferències entre espècies estretament relacionades com entre ximpanzés i humans (Webster et al. 2002) i dins del gènere *Drosophila* (Pascual et al. 2000; Schlötterer & Harr 2000). També s'ha vist que els loci de microsatèl·lits són menys comuns en ocells que en altres animals (Primmer et al. 1997). Primmer et al. (1997) especulen que la menor freqüència de microsatèl·lits en el genoma dels ocells és el resultat de la reducció en la mida del genoma imposada pel vol, el qual a la vegada resulta en una menor quantitat de DNA no codificant. Els rat-penats tenen un genoma més petit i que conté menys microsatèl·lits que els genomes d'altres mamífers, i aparentment, doncs, donen suport a aquesta hipòtesi (Van Den Bussche et al. 1995). Tanmateix, com comentàvem, no sempre trobem aquesta associació i podem trobar genomes més petits que el dels ocells que contenen més microsatèl·lits (Neff & Gross 2001). Com alternativa a l'explicació anterior estaria el fet que en els ocells també trobem una falta d'elements repetitius que actuarien com a precursors dels microsatèl·lits, especialment SINEs i LINEs. En plantes, però, s'ha observat que la freqüència dels microsatèl·lits està negativament correlacionada amb la mida del genoma (Morgante et al. 2002). Això s'ha atribuït al fet que els microsatèl·lits estan sotarepresentats en les parts repetitives del genoma de la planta involucrades en l'expansió del genoma, tal com els elements amb LTRs (Morgante et al. 2002). Per tant, sembla que el mecanisme preferent de generació de microsatèl·lits en plantes no és mitjançant els elements transponibles.

L'abundància i distribució dels microsatèl·lits no només no està representada de manera uniforme dins del genoma d'una espècie (Schug et al. 1998; Kruglyak et al. 2000), sinó que també s'han observat diferències en motius i classes d'unitats de repetició en espècies que pertanyen a diferents grups (Katti et al. 2001; Toth et al. 2000; Dieringer & Schlötterer 2003). Aquestes diferències fan pensar que les teories sobre l'*slippage* són insuficients per explicar la distribució característica dels



microsatèl·lits. El contrast en la distribució de motius de repetició entre genomes sembla indicar que hi ha una variació en el mecanisme de mutació/reparació de motius específics o de síntesi (Harr et al. 2002). Alternativament podria haver-hi variació en les constriccions selectives que estan associades amb diferents motius de repetició (Nauta & Weissing 1996) o en la composició nucleotídica dels genomes (Dieringer & Schlötterer 2003). Segons el model més recent de Dieringer & Schlötterer (2003), la composició de bases té una gran influència en la densitat dels microsatèl·lits. La composició de bases difereix entre espècies i el número esperat de microsatèl·lits depèn de la composició de bases. Per exemple, en els seus resultats, a un contingut balancejat de GC, mono- i dinucleòtids tenen menor densitat. El valor esperat de microsatèl·lits augmenta amb valors més esbiaixats del contingut en GC. Aquest efecte és torna més pronunciat pels microsatèl·lits llargs i menys pronunciat pels més curts. D'aquests treballs es pot deduir que l'associació entre microsatèl·lits i elements repetitius podria afectar la distribució dels microsatèl·lits dins i entre genomes, tot i que no ha estat analitzat. Ja que diferents genomes tenen diferents famílies d'elements transponibles, podrien tenir diferents precursors de microsatèl·lits. Les dades que presentàvem en el Capítol III de Resultats, doncs, podrien explicar la distribució diferencial dins d'espècie i entre diferents tàxons de microsatèl·lits com AG, AC i AAAT.

Tot i que encara està per demostrar clarament com els elements transponibles poden donar forma a la distribució genòmica dels microsatèl·lits, aquests elements estan afegint més complexitat tant al modelatge com a les aplicacions pràctiques dels microsatèl·lits (vegeu Capítol III de Resultats). Aquest és un motiu més per fer una selecció prèvia dels microsatèl·lits, seleccionant els menys complexes, abans de realitzar els estudis amb aquest marcador molecular.

## **2 SELECCIÓ DEL MARCADOR MOLECULAR ADEQUAT**

El procés de selecció d'un marcador molecular és un dels processos més crucials en genètica de la conservació, ja que la selecció d'un marcador inapropiat pot resultar en unes actuacions incorrectes en projectes de conservació. En els primers estudis moleculars, la selecció de marcadors moleculars estava molt influenciada per la disponibilitat de metodologies i material de laboratori. Avui dia, però, la situació ha canviat substancialment, ja que per la gran majoria d'organismes model existeix un repertori molt complet de marcadors disponibles. No obstant, pels organismes no model encara s'han de desenvolupar molts d'ells. Tanmateix, el desenvolupament i l'anàlisi poden ser contractats a una empresa externa al laboratori, si aquest últim no disposa de l'equipament necessari. D'aquesta manera, la decisió de quin marcador s'ha d'utilitzar es pot fer en base als objectius del treball. Evidentment, el gran inconvenient serà el cost de les anàlisis, sobretot tenint en compte que per estudis de conservació no es destinen tants recursos en comparació amb altres camps de recerca.

La selecció d'un marcador depèn, com ja hem dit, de les aplicacions a les quals està destinat. Generalment un marcador molecular ideal en genètica evolutiva ha de tenir almenys 4 propietats: i) ha de ser abundant i àmpliament distribuït pel genoma per tal d'evitar biaixos associats amb l'anàlisi d'un o uns pocs loci; ii) els models evolutius que se li apliquen han de ser ben coneguts per tal de facilitar la seva anàlisi i interpretació; iii) les tècniques moleculars han de permetre l'obtenció de dades a partir de molts loci en gran nombre de mostres; i, a més a més, iv) els resultats han de ser comparables amb resultats obtinguts en un altre laboratori utilitzant altres mètodes o tecnologies (Sunnucks 2000).

Els marcadors moleculars es poden classificar en dos grans tipus: marcadors nuclears i marcadors mitocondrials. El mtDNA es caracteritza per ser haploide, d'herència materna i no recombinant (amb excepcions puntuals) i per tenir una taxa de mutació entre 5 i 10 vegades superior als



exons del DNA nuclear, probablement deguda a una menor eficiència en els mecanismes de reparació del mtDNA (Weaver 1992); dins d'aquest, els gens mitocondrials estan més conservats mentre que el D-loop (o regió control) és hipervariable. Per altra banda, el DNA nuclear és d'herència biparental i presenta també una variació en les taxes de mutació, des dels exons fins als microsatèl·lits, els quals serien els marcadors més variables. D'aquesta manera, depenent de les característiques del marcador que escollim, observarem diferents signatures evolutives.

Els microsatèl·lits reuneixen, *a priori*, un seguit de característiques que els proposen com a marcador molecular ideal: presenten un elevat grau de polimorfisme, la seva herència és mendeliana simple, són codominants i són fàcils de mesurar i analitzar; a més a més, la tècnica de laboratori emprada és automatitzable i reproduïble. Per altra banda, s'assumeix que aquests es troben distribuïts de forma uniforme per tot el genoma, generalment en regions no codificants (Goldstein & Schlötterer 1999). Així doncs, els marcadors microsatèl·lits són una de les millors opcions per inferir patrons evolutius recents. Els microsatèl·lits han estat els marcadors predilectes en les àrees de genètica de poblacions, identificació individual i mapeig cromosòmic. Malgrat els avantatges aparents dels microsatèl·lits com a marcadors moleculars, existeixen també altres factors a considerar que poden disminuir el seu poder i/o la seva sensibilitat com a marcadors, o que poden ser una font d'errors que disminueix la seva utilitat en els estudis genètics. Com hem vist al llarg d'aquesta tesi, aquests factors són, bàsicament, els patrons de mutacions i el seu modelat i els problemes derivats d'aquests, com l'homoplàsia.

Els microsatèl·lits tampoc estan exempts de problemes tècnics que poden portar a errors en el genotipatge, tals com artefactes en la PCR (*stutter bands*), al·lells que no podem amplificar per PCR (al·lells nuls), l'amplificació d'un sol al·lel per atzar a causa de la poca quantitat de DNA o la generació d'un al·lel fals durant la PCR (per a més detalls, veure Taberlet et al. 1996; Dakin & Avise 2004).



Molt més recentment, s'han començat a desenvolupar tècniques que utilitzen els SNPs (Zhang & Hewitt 2003; Aitken et al. 2004; Brumfield et al. 2003). Els SNPs tenen moltes de les característiques d'un marcador ideal (Vignal et al. 2002; Brumfield et al. 2003; Morin et al. 2004) i podrien desbancar els microsatèl·lits com a marcador predilecte, ja que tenen àmbits d'aplicació similars. Aquest tipus de mutacions es troben típicament cada 300-1000 parells de bases en la majoria de genomes (Lindblad-Toh et al. 2005; Sachidanandam et al. 2001) i presenten una taxa de mutació relativament baixa ( $\sim 10^{-8}$ ), similar entre els diferents SNPs i en acord amb l'*Infinite Site Model*. El principal avantatge dels SNPs és que en podem genotipar milers en un simple assaig utilitzant tècniques com els *microarrays* (Chakraborty et al. 1999). La majoria d'aquests polimorfismes presenten dos al·lels i per aquest motiu també se'ls anomena marcadors bial·lèlics, tot i que algunes vegades els podem trobar en llocs hipermutable on aquesta assumptió és violada (Meyer et al. 1999). La informació continguda en un SNP és, doncs, limitada, especialment si un dels dos al·lels es troba en baixa freqüència. Per tant, amb la presència únicament de dos al·lels, la màxima heterozigositat esperada per cada SNP és de només un 50%, de manera que són menys informatius que les regions microsatèl·lit, les quals generalment presenten múltiples al·lels i valors d'heterozigositat que poden ser molt més elevats.

De forma general, per a la reconstrucció de mapes genètics a partir d'SNPs es requereixen fins a 3 vegades més marcadors que amb els microsatèl·lits (Kwok et al. 1996). Per a l'anàlisi de parentiu i d'identificació individual s'estima que el nombre d'SNPs necessaris per obtenir el mateix poder estadístic que amb 10-15 microsatèl·lits és del voltant de 30 a 50, depenent de les freqüències al·lèliques (Chakraborty et al. 1999; Krawczak 1999; Fries & Durstewitz 2001; Glaubitz et al. 2003). Probablement aquest número també seria adequat per a estudis poblacionals, tot i que un nombre major d'SNPs és més desitjable (Edwards & Beerli 2000; Pluzhnikov & Donnelly 1996; Luikart & England 1999; Kuhner et al. 2000; Nielsen 2000; Wakeley et al. 2001).



Tot i els possibles avantatges dels SNPs i l'increment del seu ús en humans i en estudis amb organismes models, encara són poc utilitzats en organismes no models (Aitken et al. 2004). Aquest fet es deu bàsicament a les limitacions tècniques per localitzar SNPs en genomes relativament poc coneguts. Com que aquest tipus de mutacions es produeixen a l'atzar, els SNPs no solen estar conservats entre espècies i, per tant, s'han d'identificar de nou per cada nova espècie. Existeixen diferents mètodes per a la identificació d'SNPs que varien en eficiència i cost. Morin et al. (2004) van comparar alguns d'aquests mètodes aplicats a estudis d'ecologia molecular. El descobriment d'SNPs requereix que aquests siguin seleccionats independentment de la freqüència al·lèlica, ja que el biaix en l'aïllament dels SNPs pot afectar significativament la inferència de paràmetres demogràfics històrics (Wakeley et al. 2001). Per tant, els mètodes basats en SNPs depenen enormement d'una correcció apropiada del biaix d'aïllament (Wakeley et al. 2001; Nielsen & Signorovitch 2003; Kuhner et al. 2000). Si l'objectiu és aïllar 50 SNPs que segreguen independentment, Aitken et al. (2004) estimen que per a un mamífer (no model) es necessita treballar amb 306 loci nuclears (introns). Un dels obstacles per desenvolupar marcadors SNPs *de novo* en una espècie no model ha estat la percepció que el descobriment d'SNPs, el desenvolupament dels assajos i el genotipatge serien considerablement més cars que en els microsatèl·lits. Aitken et al. (2004) estimen que caldrien 559 hores i 16000\$ per fer una exploració en 150 loci nuclears i 156 hores i 13080\$ per aïllar i optimitzar 30 loci microsatèl·lit *de novo*. També hem de tenir en compte que molts microsatèl·lits poden ser transferits entre espècies relacionades (Morin et al. 1998).

Per avaluar la utilitat dels SNPs en organismes no models és indispensable considerar la inversió necessària per desenvolupar el marcador i també el cost del genotipatge amb SNPs. Per tant, una qüestió essencial és valorar si el cost i l'efectivitat del genotipatge a gran escala amb SNPs compensen l'elevat cost del seu aïllament. Molts estudis en organismes no models, i en particular en estudis de genètica de la conservació, impliquen una mida mostral petita o moderada i, per tant, els microsatèl·lits són, per ara, els marcadors més cost-efectius.



## ***2.1 Aplicacions de les regions repetitives en genètica de la conservació***

Tal i com he mostrat en diferents apartats d'aquesta tesi (vegeu capítols II i IV), a més d'estudiar l'evolució dels microsatèl·lits, hem aplicat els marcadors estudiats en àrees on habitualment no són utilitzats.

### **2.1.1 Les regions flanquejants dels microsatèl·lits com a marcadors filogenètics**

Els caràcters morfològics van ser els primers a ser utilitzats per classificar els éssers vius i per a molts fòssils és l'única informació que disposem, però aquests caràcters presenten certs conflictes, especialment per l'evolució morfològica convergent dels organismes. Altres inconvenients que presenten els caràcters morfològics són que molts estan influenciats en major o menor grau per l'ambient; a vegades s'utilitzen caràcters difícilment quantificables que poden tenir un cert grau de subjectivitat i d'ambigüitat; i sovint no evolucionen d'una manera regular (Graur & Li 2000; Brown 2002). Un obstacle molt difícil de superar per la sistemàtica basada en morfologia és la complicació que representa realitzar comparacions entre tàxons que van divergir molt enrera en el temps – fins i tot dins dels vertebrats – a causa de les poques homologies que presenten (Crozier 1997).

Posteriorment es van incorporar les dades moleculars per fer reconstruccions filogenètiques. Aquestes estan basades generalment en la informació de les seqüències de DNA a través de les genealogies de gens, normalment gens mitocondrials. La seqüenciació de gens mitocondrials com a marcadors filogenètics té certs inconvenients com són l'anàlisi d'un únic locus, el patró d'herència (Hare 2001; Harpending et al. 1998) i problemes tècnics, com les insercions nuclears del DNA mitocondrial (Numts, Bensasson et al. 2001). Els Numts o pseudogens mitocondrials poden ser amplificats i seqüenciats per error, cosa que pot portar-nos a la reconstrucció de filogènies errònies. Diferents estratègies han estat revisades per Bensasson (2001) per identificar i treballar amb Numts. Existeixen alguns estudis que realitzen filogènies amb loci nuclears (p.e.,



Lindblad-Toh et al. 2005; Regier et al. 2005) i amb la combinació d'ambdós (p.e., Koepfli & Wayne 2003; Sato et al. 2003), però l'escassa disponibilitat de seqüències nuclears filogenèticament informatives per certes espècies limita molt la seva aplicació.

Tot i així, hi ha diferents exemples d'anàlisis combinades de dades morfològiques i moleculars en els quals s'han obtingut arbres més robustos que els obtinguts amb cada base de dades per separat, la qual cosa suggereix que les dades morfològiques sovint presenten un patró semblant al de les molècules i que no es poden descartar per fer sistemàtica (Freudenstein et al. 2003).

Els mustèlids són un exemple il·lustratiu del potencial que té la sistemàtica molecular per reconstruir filogènies i resoldre taxonomies quan la morfologia perd el seu poder de resolució, o per detectar errors deguts a la mala interpretació dels caràcters morfològics. Com s'ha pogut observar (vegeu capítol IV de Resultats), l'estudi del parentesc evolutiu dels diferents membres de la família Mustelidae mitjançant la morfologia presenta múltiples ambigüitats i és força confús. En canvi els marcadors moleculars tenen un poder resolutiu sensiblement superior. En aquesta tesi i en diferents treballs que he citat en el capítol IV de Resultats s'han descrit tres regions mitocondrials i set loci nuclears que han demostrat ser informatius per resoldre les relacions filogenètiques dels mustèlids. Tot i així, en les reconstruccions filogenètiques a partir de dades moleculars, la posició d'alguns tàxons a l'arbre filogenètic varia entre un treball i un altre, en funció del marcador utilitzat i del model evolutiu aplicat. Per exemple, les reconstruccions filogenètiques basades només en seqüències del citocrom *b* sol o en combinació amb un marcador nuclear no donen suport a la monofília de Lutrinae (Koepfli & Wayne 1998; Marmi et al. 2004). Aquest resultat és, en part, degut al fet que el citocrom *b* permet obtenir branques robustes en nodes recents, però no en els nodes més profunds dels arbres; mentre que amb gens nuclears succeeix el contrari (Koepfli & Wayne 2003; Sato et al. 2003).

Tot i la seva versatilitat, els microsatèl·lits no són considerats com a eines útils per als estudis de filogènies interespecífiques a causa de l'homoplàsia

que existeix a nivell interespecífic, tal com vèiem en els capítols I i II de Resultats. A meitats de la dècada dels 90, alguns treballs van demostrar que alguns loci microsatèl·lit homòlegs podien persistir durant uns 300 m.a., en el cas de les tortugues (FitzSimmons et al. 1995), o fins i tot 470 m.a., en el cas dels peixos (Rico et al. 1996), la qual cosa suggeria que les seves regions flanquejants estaven bastant conservades. L'aplicació de les regions flanquejants de microsatèl·lits en estudis filogenètics va ser explorada per primera vegada per estudiar les relacions filogenètiques dels principals llinatges de peixos de la família Cichlidae, juntament amb altres famílies dins del subordre Labroidei (Zardoya et al. 1996). Els resultats obtinguts van ser consistents amb altres treballs realitzats amb dades morfològiques i altres marcadors genètics mitocondrials i nuclears. Encara que des d'aleshores aquests marcadors genètics han estat poc utilitzats en estudis filogenètics, també s'han obtingut bons resultats en taurons (Martin et al. 2002) i rosegadors (Garza & Desmarais 2000; Makova et al. 2000). Tots aquests treballs han reafirmat la idea que les regions nuclears que flanquegen seqüències repetitives tenen un gran potencial que es pot explotar a l'hora de realitzar estudis filogenètics. En el capítol II dels Resultats, hem utilitzat la variació existent en la regió flanquejant del microsatèl·lit *Me108* per estudiar aquest potencial com a font d'informació filogenètica en diferents espècies de carnívors, especialment en mustèlids, que encara presenten relacions filogenètiques amb clares controvèrsies. Malgrat que hem obtingut seqüències curtes, de tan sols 154 parells de bases, aquest marcador ha demostrat ser filogenèticament força informatiu. La conservació dels microsatèl·lits en les diferents espècies d'interès converteix les seves regions flanquejants en possibles marcadors per a estudis genètics. A més a més, la combinació de varies d'aquestes regions pot aportar molta informació. Aquesta estratègia pot ser especialment beneficiosa per a l'estudi d'espècies relacionades de les quals no es disposa d'un nombre suficient de seqüències nuclears però sí de microsatèl·lits.



### 2.2.2 Creació d'un marcador específic d'espècie amb aplicacions forenses

Obtenir directament dades útils sobre espècies amenaçades o elusives pot ser logísticament difícil. El desenvolupament recent de tècniques moleculars aplicades a mostres obtingudes de forma no invasiva permet obtenir molta informació d'aquestes espècies com: la identificació de l'espècie, el sexe o l'individu; estimes de l'abundància, l'àrea de distribució o l'hàbitat individual; preferències en la dieta i, fins i tot, algunes formes d'interacció social (vegeu capítol 4.4 de la Introducció per a més detalls). Tanmateix, certes precaucions i estudis pilots són necessaris a causa de les limitacions inherents en aquest tipus de material biològic (vegeu secció 4.4 de la Introducció).

L'habilitat d'identificar de forma inequívoca l'espècie a la que pertany una mostra recollida de manera no invasiva utilitzant mètodes basats en PCR és, sense cap mena de dubte, una contribució important de la genètica a molts aspectes de la zoologia i de la biologia de la conservació. Tanmateix, per tal que aquesta tècnica sigui útil, ha de ser capaç de detectar consistentment caràcters fiables pel diagnòstic d'espècies i ha de ser tan simple i econòmica com sigui possible. La naturalesa del diagnòstic es basa en la fixació d'un patró dins d'una espècie i la falta de coincidència d'aquest patró en individus de diferents espècies. Per tal que la identificació sigui fiable, cal que aquests patrons es defineixin analitzant una mostra d'individus representativa – però inevitablement limitada – i amb un origen geogràfic tan ampli com sigui possible, per tal de tenir en compte qualsevol polimorfisme.

L'anàlisi d'RFLPs del DNA mitocondrial ha estat l'eina més estesa per identificar discontinuïtats genètiques a nivells interespecífics (revisat en Avise 1994). A més de disposar d'una base de dades amb més seqüències mitocondrials, el principal avantatge de treballar amb aquests tipus de marcadors és que el mtDNA té un gran nombre de còpies per cada cèl·lula. Un mitocondri conté entre 2 i 10 còpies del mtDNA, i poden haver-hi 1000 mitocondris per cèl·lula somàtica (Budowle et al. 2003).

Aquest fet és important quan, per exemple, es treballa amb mostres no invasives o degradades on el nombre de cèl·lules és baix.

Com dèiem prèviament, com a conseqüència de la seva elevada taxa de mutació recurrent, els microsatèl·lits no són utilitzats com a marcadors per comparacions interespecífiques. Tot i així, Domingo-Roura (2002) va dur a terme un treball on utilitzava la mida dels al·lels d'un microsatèl·lit per diferenciar entre dues espècies de mustèlids d'interès en conservació, la marta i la fagina. Per una altra banda, els elements repetitius poden ser utilitzats per diferenciar espècies. En el capítol IV de Resultats, hem presentat una eina molecular nova i simple amb aplicacions forenses: l'ús de la inserció d'un element SINE com a marcador específic d'espècie per diferenciar el visó americà del visó europeu i del turó. A més a més, també hem utilitzat RFLP per diferenciar aquestes dues últimes espècies que no quedarien diferenciades amb la primera tècnica. Aquest SINE es troba inserit en la regió flanquejant del ja abans mencionat microsatèl·lit *Me/08*. Les insercions de SINEs presenten propietats que els fan idonis per a la diferenciació d'espècies (vegeu secció 3.3 de la Introducció) com és, per exemple, la irreversibilitat de les insercions. Com hem vist en el capítol II de Resultats, l'única espècie dels carnívors analitzats que presentava el SINE inserit en el marcador *Me/08* era el visó americà. A més a més, aquest no té cap altra espècie filogenèticament propera no analitzada amb la qual hagués pogut compartir aquesta inserció. Aquest marcador va ser testat en una àmplia i diversa mostra d'individus de 7 països per les 3 espècies – més diversa que en altres treballs publicats de característiques semblants (p.e., Gómez-Moliner et al. 2004). Aquest és un mètode que pot ser aplicat per confirmar la presència d'alguna d'aquestes espècies en una àrea determinada a partir de restes biològiques. De manera que aquesta eina pot tenir implicacions directes en conservació per la importància d'aquestes espècies i per resoldre qüestions d'hibridació entre visó europeu i turó.

L'ús d'un marcador nuclear per diferenciar el visó europeu del turó presenta certes avantatges respecte a la utilització de marcadors mitocondrials. Aquest fet es deu a l'observació d'haplotips mitocondrials



compartits i una falta de monofília recíproca en aquestes dues espècies, cosa que indica un ancestre comú recent o bé fenòmens d'hibridació (Davison et al. 2000). Tot i que s'espera que aquests dos patrons siguin més probables per marcadors nuclears, a causa d'una població efectiva major en el genoma nuclear i, per tant, una taxa de fixació menor (Brown 1979), no hem trobat cap polimorfisme intraespecífic per les posicions variables en les mostres analitzades – una mostra d'individus més gran i diversa que en el treball de Davison et al. (2000). Per aquest fet, l'ús de marcadors mitocondrials per diferenciar aquestes espècies estaria limitat.

El mètode presentat dependrà del mètode de recol·lecció de pèl fresc arrencat mitjançant el mostreig amb trampes de pèl. El trampeig de pèls no és un mètode molt estès i en molts casos probablement necessita una optimització per cada espècie, segons les característiques de cada espècie i els seus hàbits. Tanmateix, les tècniques de trampeig de pèl ja han estat provades amb èxit en diferents espècies de mustèlids, fins i tot en poblacions petites i poc distribuïdes (Foran et al. 1997; Mowat & Paetkau 2002; Frantz et al. 2004). En un estudi recent, s'han obtingut dades de les grandàries poblacionals de grups de teixons a partir de DNA extret de pèl recollit de manera no invasiva i els seus resultats han demostrat ser més efectius que el mètode amb femtes (Frantz et al. 2004).

A diferència dels altres treballs presentats en aquesta tesi (vegeu capítols del I al III dels Resultats), aquest és un treball majoritàriament metodològic i els seu impacte dependrà de la utilitat pràctica de la tècnica i dels seus avantatges o complementarietats sobre altres tècniques amb la mateixa finalitat. Al *Centro de Recuperación de Fauna de Mártioda* (Àlava), on porten a terme un control de la població de visó americà i estudis de conservació del visó europeu, utilitzen la tècnica d'identificació que he presentat amb els animals que els arriben i, a més a més, estan provant la seva efectivitat amb femtes.

La qüestió de la identificació molecular d'espècies i la taxonomia basada en dades moleculars és un tema que ha reaparegut fortament en els últims anys per l'aparició del que es coneix com *DNA Barcoding*, proposat per Hebert et al. (2003). El *DNA Barcoding* proposa l'ús d'un marcador

genètic curt del DNA mitocondrial per identificar de manera ràpida i senzilla a quina espècie pertany una mostra desconeguda. Una regió de 684 parells de bases del gen citocrom oxidasa I (COI) del DNA mitocondrial ha estat seleccionada com a codi de barres potencial. Donada la llarga història de l'ús dels marcadors moleculars per a aquest propòsit, no hi ha res fonamentalment nou en el concepte *DNA Barcoding* a excepció de potenciar el seu ús i utilitzar un únic marcador estandarditzat per identificar espècies. EL *DNA Barcoding* no ha passat gens desapercbut pels científics, especialment pels sistemàtics, que mostren des d'una afiliació entusiasta (Janzen 2004) fins a una oposició total (p.e., vegeu el volum de febrer de 2003 de la revista *Trends in Ecology and Evolution*). Les principals crítiques a aquest mètode són, per exemple, la simplificació de la ciència de la taxonomia o que el COI no proporciona informació fiable a nivell d'espècies que han divergit recentment. Un altre factor a tenir en compte és que el DNA mitocondrial és heretat per via materna i els casos d'hibridació portaran a resultats erronis. Pel que fa al problema concret de la diferenciació dels tres mustèlids del capítol IV de Resultats, el *DNA barcoding* podria donar problemes donada la divergència recent del visó europeu i el turó. Per un altre banda, els més moderats consideren que el *DNA Barcoding* s'ha d'utilitzar conjuntament amb eines de taxonomia tradicional i altres tècniques de sistemàtica molecular de manera que les causes dels problemes poden ser identificades i els errors detectats.

Qualsevol font d'informació que identifiqui caràcters que es mostrin com a únics d'una espècie hauria de ser considerada com un codi de barres potencial. Tècniques moleculars com els RFLPs, AFLPs, RAPDs, microsatèl·lits, l'ordre dels gens, absència/presència de gens i SNPs serien mètodes apropiats per al descobriment d'identificadors únics per un codi de barres de DNA. A més a més, les tècniques d'anàlisi semblen avançar més ràpid que els nostres coneixements sobre el genoma de les espècies no models, per tant, en un futur, probablement siguem capaços d'obtenir informació de tot un genoma sense tenir-ne molta informació prèvia.



La identificació molecular d'espècies és un camp on el genotipatge de posicions variables entre espècies al llarg del genoma pot jugar un paper important. Els *microarrays* estan sent utilitzats pel diagnòstic de diferents espècies de bacteris (p.e., Kingsley et al. 2002; Rudi et al. 2003) o patògens vírics (p.e., Wang et al. 2002; Boonham et al. 2003) en anàlisis clíniques, i han mostrat que el potencial del seu ús és enorme. Es podria fer un xip de mamífers utilitzant posicions variables que inclogués múltiples tests redundants per identificar tots els mamífers d'Europa. Tanmateix, econòmicament només valdria la pena el desenvolupament d'un únic xip, produït en grans quantitats, que servís per diferents propòsits i respongués a moltes qüestions diverses d'ecologia (Rubenstein 2003). Encara estem lluny d'arribar a situacions com aquesta i, actualment, el que cal és desenvolupar marcadors simples i barats que puguin ajudar a resoldre qüestions crítiques en conservació, com per exemple la greu amenaça sobre el visó europeu i la invasió del visó americà.

### **3 PERSPECTIVES DE FUTUR DELS MICROSATÈL·LITS**

Com hem vist al llarg d'aquesta tesi, el procés evolutiu dels microsatèl·lits està lluny de ser simple. Una de les conseqüències de la complexitat de l'evolució dels microsatèl·lits és la necessitat de prendre precaucions quan s'utilitzen com a marcadors, especialment en genètica de poblacions. Molts d'aquests problemes podrien ser resolts si es destinés un esforç de seqüenciació per seleccionar els microsatèl·lits més apropiats per a cada estudi. Al mateix temps, el desenvolupament de nous algoritmes que descriuïn l'evolució dels microsatèl·lits ha d'incloure tantes formes d'heterogeneïtat mutacional com sigui possible. Tot i els inconvenients, els microsatèl·lits continuaran sent eines interessants per fer mapeig, identificacions individuals i per estudiar certs processos demogràfics. No obstant, a causa de les cada cop més evidents limitacions que presenten en genètica de poblacions, els SNPs es plantegen com una alternativa



interessant. Més recentment, s'ha vist que els microsatèl·lits poden tenir usos en estudis de desequilibri de lligament, en els quals es busquen associacions entre marcadors i loci de caràcters fenotípics en mostres poblacionals (Ohashi & Tokunaga 2003), i en *hitchhiking*, en els que es fan cerques per tot el genoma de regions que mostren signes de selecció (Schlötterer 2003).

A més a més, també trobem noves aplicacions pels microsatèl·lits. Donada la seva elevada taxa de mutació, els microsatèl·lits ofereixen medis per estudiar la manera com els factors mediambientals afecten la taxa de mutació genòmica. S'han observat elevades taxes de mutació en línies germinals en animals i plantes que han estat exposades a radiacions ionitzants (Ellegren et al. 1997; Kovalchuk et al. 2003), i observacions similars s'han fet per minisatèl·lits en humans (Dubrova et al. 1996). Estimar les taxes mutacionals dels microsatèl·lits en mostres que han estat exposades a diferents formes de radiacions o compostos tòxics podria ajudar a fer estimes de riscos per aquestes agressions. Per tant, per totes les noves aplicacions que van sorgint i per la falta de millors alternatives per algunes aplicacions, la recerca bàsica en evolució de microsatèl·lits continua sent de gran importància.



## **CONCLUSIONS**

Les conclusions que es poden extreure del present treball són:

1. La comparació intraespecífica de seqüències al·lèliques dels 10 microsatèl·lits analitzats de teixó comú mostra que cap d'ells té una única regió repetitiva variable sense interrupcions o indels en la regió flanquejant. Aquest fet limita notablement la seva aplicabilitat com a marcadors genètics basada en la grandària dels al·lels.
2. En l'anàlisi a nivell de seqüència tant dels 10 microsatèl·lits com en les comparacions del marcador *Me/08* en 24 espècies de carnívors observem homoplàsia de grandària a nivell intraespecífic, per la presència en tots aquests marcadors d'indels i/o diferències nucleotídiques a la regió flanquejant. Aquest fet indica que la mida dels electromorfs no és adequada per fer inferències filogenètiques.
3. El marcador microsatèl·lit *Me/15* presenta dues regions repetitives generades per un SINE, particularitat que difícilment permet un únic procés d'*slippage*. A més a més, la regió de polipirimidines del SINE també és responsable de generar variació intraespecífica de grandària en les regions flanquejants.
4. Tot i la complexitat detectada en els 10 microsatèl·lits de teixó comú, els valors d'heterozigositat es correlacionen positivament amb el nombre de repeticions perfectes de les diferents regions repetitives. Per tant, a pesar de la complexitat de la base de dades, el nombre de repeticions perfectes és un bon predictor de la variabilitat, i aquest paràmetre, obtingut per seqüenciació, pot ser modelat per models evolutius simples.
5. A pesar de ser una regió conservada al llarg de 35.5 milions d'anys – fenomen poc comú en microsatèl·lits –, el marcador *Me/08* presenta una variabilitat elevada dins i fora de la regió repetitiva. A més de la presència d'indels de varis parells de bases i diferències nucleotídiques en la regió flanquejant, aquest microsatèl·lit mostra una estructura



complexa amb quatre regions repetitives on totes mostren signes de processos d'*slippage*. Aquestes dades suggereixen que el microsatèl·lit s'ha generat en un punt calent de mutació.

6. La variabilitat present en la regió flanquejant del marcador microsatèl·lit *Me108* és prou informativa per fer inferències filogenètiques. La reconstrucció filogenètica realitzada a partir d'aquestes seqüències dona suport a: i) la monofília de la subfamília Lutrinae, ii) la parafília de la subfamília Mustelinae, iii) la posició basal del teixó comú respecte a la família Mustelidae, iv) la classificació de les mofetes com una família independent i v) la classificació del panda vermell com a una família monotípica en posició basal respecte a la superfamília Musteloidea.
7. Els tRNA<sup>Lys</sup> SINEs de carnívors generen freqüentment microsatèl·lits en la regió interna de polipirimidines i en la seva cua de poliadenines. Aquestes dues regions són responsables de la majoria de microsatèl·lits AG i rics en As aïllats en carnívors. Donat que el tRNA<sup>Lys</sup> SINE és l'element més abundant en el genoma de carnívors, molt probablement afecta la distribució dins i entre espècies d'aquest tipus de repeticions.
8. Existeix una elevada freqüència de clons de microsatèl·lits aïllats en carnívors associats amb tRNA<sup>Lys</sup> SINEs. L'ús d'aquests marcadors no és recomanable ja que, a part dels problemes tècnics que impliquen, els clons que presenten aquesta associació també porten amb més freqüència regions repetitives múltiples i tenen un menor potencial d'expandibilitat.
9. Tant la seqüenciació del genoma del gos com les nostres estimes del nombre de tRNA<sup>Lys</sup> SINEs en diferents espècies de carnívors indiquen que les estimes prèvies infravaloraven els valors reals i que, probablement, en moltes espècies de carnívors aquest nombre de tRNA<sup>Lys</sup> SINE estigui per sobre de 10<sup>6</sup> còpies per genoma haploide.



10. La inserció SINE en el microsatèl·lit *Me/08* pot ser utilitzada per diferenciar el visó americà del visó europeu i del turó a partir d'una simple PCR. A més a més, amb aquest mateix producte de PCR i mitjançant el patró de bandes d'un RFLP amb l'enzim *AcI*, podem diferenciar el visó europeu del turó. Aquest mètode, doncs, pot ser utilitzat d'una manera molt econòmica i simple per monitoritzar aquestes espècies.
  
11. El mètode de diferenciació del visó americà, del visó europeu i del turó pot ser utilitzat a partir de pèl dels animals recollit de manera no invasiva. També permetrà, probablement, estudiar fenòmens d'hibridació entre el visó europeu i el turó comú. Aquest mètode de diferenciació té especial interès en conservació pel greu perill en què es troba el visó europeu i els diferents problemes que causa el visó americà sobre aquesta espècie i altres mustèlids.

# **BIBLIOGRAFIA**

**BIBLIOGRAFIA**

- Agrawal A, Eastman QM, Schatz DG (1998) Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature* 394: 744-751.
- Aitken N, Smith S, Schwarz C, Morin PA (2004) Single nucleotide polymorphism (SNP) discovery in mammals: a targeted-gene approach. *Mol Ecol* 13: 1423-1431.
- Alexander LJ, Rohrer CA, Stone RT, Beattie CW (1995) Porcine SINE-associated microsatellite markers: evidence for artiodactyls SINEs. *Mamm Genome* 6: 464-468.
- Anderson TJ, Su XZ, Roddam A, Day KP (2000) Complex mutations in a high proportion of microsatellite loci from the protozoan parasite *Plasmodium falciparum*. *Mol Ecol* 9: 1599-1608.
- Angers B, Bernatchez L (1997) Complex evolution of a salmonid microsatellite locus and its consequences in inferring allelic divergence from size information. *Mol Biol Evol* 14: 230-238.
- Angers B, Estoup A, Jarne P (2000) Microsatellites size homoplasy, SSCP, and population structure: a case study in the freshwater snail *Bulinus truncatus*. *Mol Biol Evol* 17: 1926-1932.
- Armour JA, Neumann R, Gobert S, Jeffreys AJ (1994) Isolation of human simple repeat loci by hybridization selection. *Hum Mol Genet* 3: 599-565.
- Arcot SS, Wang Z, Weber JL, Deininger PL, Batzer MA (1995) Alu repeats: a source for the genesis of primate microsatellites. *Genomics* 29: 136-144.
- Arcot SS, Adamson AW, Risch GW, LaFleur J, Robichaux MB, Lamerdin JE, Carrano AV, Batzer MA (1998) High resolution cartography of recently integrated human chromosome 19-specific Alu fossils. *J Mol Biol* 281: 843-856.



- Avise JC (1994) *Molecular Markers: Natural History and Evolution*. Chapman and Hall, New York.
- Baker CS, Palumbi SR (1994) Which whales are hunted? A molecular genetic approach to monitoring whaling. *Science* 265: 1538-1539.
- Baillie J, Groombridge B (1996) *1996 IUCN Red List of Threatened Animals*. IUCN, Gland, Switzerland and Conservation International, Washington DC.
- Balloux F, Ecoffey E, Fumagalli L, Goudet J, Wytttenback A, Hausser J (1998) Microsatellite conservation, polymorphism, and GC content in shrews of the genus *Sorex* (Insectivora, Mammalia). *Mol Biol Evol* 15: 473-475.
- Beckmann JS, Weber JL (1992) Survey of human and rat microsatellites. *Genomics* 12: 627-631.
- Bénit L, Lallemand JB, Casella JF, Philippe H, Heidmann T (1999) ERV-L elements: a family of endogenous retrovirus-like elements active throughout the evolution of mammals. *J Virol* 73: 3301-3308.
- Bensasson D, Zhang D-X, Hartl DL, Hewitt GM (2001) Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends Ecol Evol* 16: 314-321.
- Biemont C, Cizeron G (1999) Distribution of transposable elements in *Drosophila* species. *Genetica* 105: 43-62.
- Boonham N, Walsh K, Smith P, Madagan K, Graham I, Barker I (2003) Detection of potato viruses using microarray technology: towards a generic method for plant viral disease diagnosis. *J Virol Methods* 108: 181-187.
- Boyle AL, Ballard SG, Ward DC (1990) Differential distribution of long and short interspersed element sequences in the mouse genome: chromosome karyotyping by fluorescence in situ hybridization. *Proc Natl Acad Sci USA* 87: 7757-7761.





- Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994) High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* 368: 455-457.
- Bricker J, Bushar LM, Reinert HK, Gelbert L (1996) Purification of high quality DNA from shed skins. *Herpetol Rev* 27: 133-134.
- Brinkmann B, Klintschar M, Neuhuber F, Huhne J, Rolf B (1998) Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *Am J Hum Genet* 62: 1408-1415.
- Brohede J, Primmer CR, Moller A, Ellegren H (2002) Heterogeneity in the rate and pattern of germline mutation at individual microsatellite loci. *Nucleic Acids Res* 30: 1997-2003.
- Brown WM, George M, Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. *Proc Natl Acad Sci USA* 76: 1967-1971.
- Brown TA (2002) *Genomes*. 2nd ed Oxford, UK: BIOS Scientific Publishers, Ltd.
- Brumfield R, Nickerson D, Beerli P, Edwards SV (2003) The utility of single nucleotide polymorphism in inferences of population history. *Trends Ecol Evol* 18: 249-256.
- Buchanan FC, Littlejohn RP, Galloway SM, Crawford AM (1993) Microsatellites and associated repetitive elements in the sheep genome. *Mamm Genome* 4: 258-264.
- Budowle B, Allard MW, Wilson MR, Chakraborty R (2003) Forensics and mitochondrial DNA: applications, debates, and foundations. *Annu Rev Genom Hum G* 4: 119-141.
- Burten J (2002) *Mammals of Britain and Europe*. Kingfisher Publications, London.
- Carter J (2002) DNA preservation in fluid preserved collections. *SPNHC Newsletter* 16: 14-15.



- Chakraborty R, Stivers DN, Su B, Zhong Y, Budowle B (1999) The utility of short tandem repeat loci beyond human identification: implications for development of new DNA typing systems. *Electrophoresis* 20: 1682–1696.
- Charlesworth B, Sniegowski P, Stephan W (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. A review of the population genetics of transposable and other repetitive elements, detailing why transposons should be regarded as parasites. *Nature* 371: 215–220.
- Colson I, Goldstein DB (1999) Evidence for complex mutations at microsatellite loci in *Drosophila*. *Genetics* 152: 617–27.
- Constable JL, Ashley MV, Goodall J, Pusey AE (2001) Noninvasive paternity assignment in Gombe chimpanzees. *Mol Ecol* 10: 1279–1300.
- Cooper A, Rhymer J, James HF, Olson SL, McIntosh CE, Sorenson MD, Fleischer RC (1996) Ancient DNA and island endemics. *Nature* 381: 484.
- Crawford AM, Kappes SM, Paterson KA, deGotari MJ, Dodds KG, Freking BA, Stone RT, Beattie CW (1998) Microsatellite evolution: testing the ascertainment bias hypothesis. *J Mol Evol* 46: 256–260.
- Crozier RH (1997) Preserving the information content of species: genetic diversity, phylogeny, and conservation worth. *Annu Rev Ecol S* 28: 243–268.
- Dakin EE, Avise JC (2004) Microsatellite null alleles in parentage analysis. *Heredity* 93: 504–509.
- Daniels GR, Deininger PL (1985) Repeat sequence families derived from mammalian tRNA genes. *Nature* 317: 819.
- Danilevskaya O, Slot F, Pavlova M, Pardue ML (1994) Structure of the *Drosophila* HeT-A transposon: a retrotransposon-like element forming telomeres. *Chromosoma* 103: 215–224.



- Danilov PI, Tumanov IL (1976) The ecology of the European and American mink in the Northwest of the USSR. In: *Ecology of birds and mammals in the Northwest of the USSR*. Akad Nauk Karelski filial, Inst Biol: 118-143
- Davison A, Griffiths HI, Brookes RC, Maran T, Macdonald DW, Sidorovich VE, Kitchener AC, Irizar I, Villate I, González-Esteban J, et al. (2000) Mitochondrial DNA and palaeontological evidence for the origins of endangered European mink, *Mustela lutreola*. *Anim Conserv* 4: 345–355.
- Dewannieux M, Esnault C, Heidmann T (2003). LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet* 35: 41–48.
- Deininger PL, Batzer MA (1993) Evolution of retroposons. *Evol Biol* 27: 157–196.
- Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E, et al. (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* 380: 152-154.
- Dieringer D, Schlötterer C (2003) Two distinct modes of microsatellite mutation processes: evidence from the complete genomic sequences of nine species. *Genome Res* 13: 2242–2251.
- Dietrich WF, Miller JC, Steen RG, Merchant M, Damron D, Nahf R, Gross A, Joyce DC, Wessel M, Dredge RD, et al. (1994) A genetic map of the mouse with 4,006 simple sequence length polymorphisms. *Nat Genet* 7: 220-245.
- Di Rienzo A, Peterson AC, Garza JC, Valdes Slatkin M, Freimer NB (1994) Mutational processes of simple-sequence repeat loci in human populations. *Proc Natl Acad Sci USA* 91: 3166–3170.
- Di Rienzo A, Donnelly P, Toomajian C, Sisk B, Hill A, Petzl-Erler M, Haines G, Barch DH (1998) Heterogeneity of microsatellite mutations within and between loci and implications for human demographic histories. *Genetics* 148: 1269–1284.



- Domingo-Roura X, Marmi J, López-Giráldez JF, Garcia-Franquesa E (2001) New molecular challenges in animal conservation. *Anim Biodiv Conserv* 24: 19-29.
- Domingo-Roura X (2002) Genetic distinction of marten species by the fixation of a microsatellite region. *J Mammal* 83: 907-912.
- Domingo-Roura X, Marmi J, Ferrando A, López-Giráldez F, Macdonald DW, Jansman H (2006) Badger hair in shaving brushes comes from protected Eurasian badgers. *Biol Conserv* 128: 425-430.
- Dubrova YE, Nesterov VN, Krouchinsky NG, Ostapenko VA, Neumann R, Neil DL, Jeffreys AJ (1996) Human minisatellite mutation rate after the Chernobyl accident. *Nature* 380: 683-686.
- Edwards A, Civitello A, Hammond HA, Caskey CT (1991) DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Am J Hum Genet* 49: 746-756.
- Edwards MC, Gibbs RA (1992) A human dimorphism resulting from loss of an Alu. *Genomics* 14: 590-597.
- Edwards SV, Beerli P (2000) Gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution* 54: 1839-1854.
- Eichler EE, Kunst CB, Lugenbeel KA, Ryder OA, Davison D, Warren ST, Nelson DL (1995) Evolution of the cryptic FMR1 CGG repeat. *Nat Genet* 11: 301-308.
- Eickbush TH (1992) Transposing without ends: the non-LTR retrotransposable elements. *New Biol* 4: 430-440.
- Eickbush TH. (1997) Telomerase and retrotransposons: which came first? *Science* 275: 911-912.
- Ellegren H, Primmer CR, Sheldon BC (1995) Microsatellite 'evolution': directionality or bias? *Nat Genet* 11: 360-362.



- Ellegren H, Lindgren G, Primmer CR, Moller AP (1997) Fitness loss and germline mutations in barn swallows breeding in Chernobyl. *Nature* 389: 593–596.
- Ellegren H (2000) Microsatellite mutations in the germline: implications for evolutionary inference. *Trends Genet* 16: 551–558.
- Ellegren H (2004). Microsatellites: simple sequences with complex evolution. *Nat Rev Genet* 5: 435–445.
- Ender A, Schwenk K, Stadler T, Streit B, Schierwater B (1996) RAPD identification of microsatellites in *Daphnia*. *Mol Ecol* 5: 437–441.
- Endoh H, Okada N (1986) Total DNA transcription in vitro: a procedure to detect highly repetitive and transcribable sequences with tRNA-like structures. *Proc Natl Acad Sci USA* 83: 251–255.
- Estoup A, Tailliez C, Cornuet J-M, Solignac M (1995) Size homoplasy and mutational processes of interrupted microsatellites in two bee species, *Apis mellifera* and *Bombus terrestris* (Apidae). *Mol Biol Evol* 12: 1074–1084.
- Estoup A, Cornuet J-M (1999) Microsatellite evolution: inferences from population data. In: *Microsatellites: evolution and applications*, Goldstein DB, Schlötterer C eds. Oxford University Press, Oxford: 49–65.
- Estoup A, Jarne P, Cornuet JM (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Mol Ecol* 11: 1591–1604.
- Ezenwa VO, Peters JM, Zhu Y, Arevalo E, Hastings MD, Seppa P, Pedersen JS, Zacchi F, Queller DC, Strassmann JE (1998) Ancient conservation of trinucleotide microsatellite loci in Polistine wasps. *Mol Phylogenet Evol* 10: 168–177.
- Feldman MW, Bergman A, Pollock DD, Goldstein DB (1997) Microsatellite genetic distances with range constraints: analytic description and problems of estimation. *Genetics* 145: 207–216.



- Feng Q, Moran JV, Kazazian HH Jr, Boeke JD (1996) Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* 87: 905-916.
- Fisher PJ, Gardner RC, Richardson TE (1996) Single locus microsatellites isolated using 5' anchored PCR. *Nucleic Acids Res* 24: 4369-4371.
- FitzSimmons NN, Moritz C, Moore SS (1995) Conservation and dynamics of microsatellite loci over 300 million years of marine turtle evolution. *Mol Biol Evol* 12: 432-440.
- Foran DR, Minta SC, Heinemeyer KS (1997) DNA-based analysis of hair to identify species and individuals for population research and monitoring. *Wildl Soc Bull* 25: 840-847.
- Frantz AC, Schaul M, Pope LC, Fack F, Schley L, Muller CP, Roper TJ (2004) Estimating population size by genotyping remotely plucked hair: the Eurasian badger. *J Appl Ecol* 41: 985-995
- Freudenstein JV, Pickett KM, Simmons MP, Wenzel JW (2003) From basepairs to birdsongs: phylogenetic data in the age of genomics. *Cladistics* 19: 333-347.
- Fries R, Durstewitz G (2001) Digital DNA signatures for animal tagging. *Nat Biotechnol* 19: 508
- Gallagher J, Clifton-Hadley RS (2000) Tuberculosis in badgers; a review of disease and its significance for other animals. *Res Vet Sci* 69: 203-217.
- Gardner MG, Bull CM, Cooper SJ, Duffield GA (2000) Microsatellite mutations in litters of the Australian lizard *Egernia stokesii*. *J Evol Biol* 13: 551-560.
- Garza JC, Slatkin M, Freimer NB (1995) Microsatellite allele frequencies in humans and chimpanzees, with implications for constraints on allele size. *Mol Biol Evol* 12: 594-603.
- Garza JC, Freimer NB (1996) Homoplasmy for size at microsatellite loci in humans and chimpanzees. *Genome Res* 6: 211-217.



- Garza JC, Desmarais E (2000) Derivation of a simple microsatellite locus from a compound ancestor in the genus *Mus*. *Mamm Genome* 11: 1117-1122.
- Gibson G (2002) Microarrays in ecology and evolution: a preview. *Mol Ecol* 11: 17-24.
- Ginzburg LR, Bingham PM, Yoo S (1984) On the theory of speciation induced by transposable elements. *Genetics* 107: 331-341.
- Glaubitz JC, Rhodes OE, Dewoody JA (2003) Prospects for inferring pairwise relationships with single nucleotide polymorphisms. *Mol Ecol* 12: 1039-1047.
- Glenn TC, Stephan W, Dessauer HC, Braun MJ (1996) Allelic diversity in alligator microsatellite loci is negatively correlated with GC content of flanking sequences and evolutionary conservation of PCR amplifiability. *Mol Biol Evol* 13: 1151-1154
- Gómez-Moliner BJ, Cabria MT, Rubines J, Garin I, Madeira MJ, Elejalde A, Aihartza J, Fournier P, Palazón YS (2004) PCR-RFLP identification of mustelid species: European mink (*Mustela lutreola*), American mink (*M. vison*) and polecat (*M. putorius*) by analysis of excremental DNA. *J Zool* 262: 311-316.
- Goldstein DB, Clark AG (1995) Microsatellite variation in North American populations of *Drosophila melanogaster*. *Nucleic Acids Res* 23: 3882-3886.
- Goldstein DB, Linares AR, Cavalli-Sforza LL, Feldman MW (1995) An evaluation of genetic distances for use with microsatellite loci. *Genetics* 139: 463-471.
- Goldstein D, Schlötterer C (1999) *Microsatellites: evolution and applications*. Oxford University Press, New York.
- Gotea V, Kranz A (1999). The European mink (*Mustela lutreola*) in the Danube Delta. *Small Carniv Conserv* 21: 23-25.



- Graur D, Li WH (2000) *Fundamentals of Molecular Evolution* (2nd edition). Sinauer Associates Inc, Massachusetts.
- Griffiths AJF, Gelbart WM, Miller JH, Lewontin RC (1999) *Modern Genetic Analysis*. WH Freeman & Co., New York.
- Grimaldi MC, Crouau-Roy B (1997) Microsatellite allelic homoplasy due to variable flanking sequences. *J Mol Evol* 44: 336-340
- Hancock JM (1999) Microsatellites and other simple sequences: genomic context and mutational mechanisms. In: *Microsatellites: Evolution and Applications*, Goldstein DB, Schlötterer C eds, Oxford University Press, Oxford.
- Hare MP (2001) Prospects for nuclear gene phylogeography. *Trends Ecol Evol* 16: 700-706.
- Harpending HC, Batzer MA, Gurven M, Jorde LB, Rogers AR, Sherry ST (1998) Genetic traces of ancient demography. *Proc Natl Acad Sci USA* 95: 1961-1967
- Harr B, Zangerl B, Schlötterer C (2000) Removal of microsatellite interruptions by DNA replication slippage: phylogenetic evidence from *Drosophila*. *Mol Biol Evol* 17: 1001-1009.
- Harr B, Todorova J, Schlötterer C (2002) Mismatch repair driven mutational bias in *D. melanogaster*. *Mol Cell* 10: 199-205.
- Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identification through DNA barcodes. *Proc R Soc Lond Ser B* 270: 313-321.
- Hedrick PW, Kalinowski ST (2000) Inbreeding depression in conservation biology. *Annu Rev Ecol S* 31: 139-162.
- Henderson ST, Petes TD (1992) Instability of simple sequence DNA in *Saccharomyces cerevisiae*. *Mol Cell Biol* 12: 2749-2757.





- Henttonen H (1992) Vesikko (*Mustela lutreola*). In: *Maailman uhanalaiset eläimet*, Elo U ed: 46-48.
- Hofreiter M, Poinar HN, Spaulding WG, Bauer K, Martin PS, Possnert G, Paabo S (2000) A molecular analysis of ground sloth diet through the last glaciation. *Mol Ecol* 9: 1975-1984.
- Hofreiter M, Serre D, Poinar HN, Kuch M, Paabo S (2001) Ancient DNA. *Nat Rev Genet* 2: 353-359.
- Hoelzel AR, Halley J, O'Brien SJ, Campagna C, Arnborn T, Le Boeuf B, Ralls K, Dover GA (1993) Elephant seal genetic variation and the use of simulation models to investigate historical population bottlenecks. *J Hered* 84: 443-449.
- Höss M, Kohn M, Pääbo S, Knauer F, Schroder W (1992) Excrement analysis by PCR. *Nature* 359: 199.
- Höss M, Dilling A, Currant A, Paabo S (1996) Molecular phylogeny of the extinct ground sloth *Myiodon darwini*. *Proc Natl Acad Sci USA* 93:181-185.
- Hsieh H-M, Huang L-H, Tsai L-C, Kuo Y-C, Meng H-H, Linacre A, Lee JC-I (2003) Species identification of rhinoceros horns using the cytochrome *b* gene. *Forensic Sci Int* 136: 1-11.
- Hurst GD, Werren JH (2001) The role of selfish genetic elements in eukaryotic evolution. *Nat Rev Genet* 2: 597-606.
- International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409: 860-921.
- Janzen DH (2004) Now is the time. *Philos Trans R Soc Lond B Biol Sci* 359: 731-732.
- Jarne P, Lagoda JL (1996) Microsatellites, from molecules to populations and back. *Trends Ecol Evol* 11: 424-429.



- Jelinek WR (1982) Repetitive sequences in eukaryotic DNA and their expression. *Annu Rev Biochem* 51: 813-844.
- Jurka J (1997) Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. *Proc Natl Acad Sci USA* 94: 1872-1877.
- Jurka J (2000) RepBase update: a database and an electronic journal of repetitive elements. *Trends Genet* 9: 418- 420.
- Kalinowski ST, Hedrick PW, Miller PS (2000) Inbreeding depression in the Speke's gazella captive breeding program. *Conserv Biol* 14: 1375-1384.
- Katti MV, Ranjekar PK, Gupta VS (2001) Differential distribution of simple sequence repeats in eukaryotic genome sequences. *Mol Biol Evol* 18: 1161-1167.
- Kajikawa M, Okada N (2002) LINEs mobilize SINEs in the eel through a shared 3' sequence. *Cell* 111: 433-444.
- Kayser M, Roewer L, Hedman M, Henke L, Henke J, Brauer S, Kruger C, Krawczak M, Nagy M, Dobosz T, et al. (2000) Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. *Am J Hum Genet* 66: 1580-1588
- Kazazian HH Jr (2004) Mobile elements: drivers of genome evolution. *Science* 303: 1626-1632.
- Kidwell MG, Lisch DR (2000) Transposable elements and host genome evolution. *Trends Ecol Evol* 15: 95-99.
- Kimmel M, Chakraborty R, Stivers DN, Deka R (1996) Dynamics of repeat polymorphisms under a forwardbackward mutation model: within- and between-population variability at microsatellite loci. *Genetics* 143: 549-555.
- Kimura M, Crow JF (1964) The number of alleles that can maintained in a finite population. *Genetics* 49: 725-738.

- Kimura M, Ohta T (1978) Stepwise mutation model and distribution of allelic frequencies in a finite population. *Proc Natl Acad Sci USA* 75: 2868-2872.
- Kingsley MT, Straub TM, Call DR, Daly DS, Wunschel SC, Chandler DP (2002) Fingerprinting closely related *Xanthomonas* pathovars with random nonamer oligonucleotide microarrays. *Appl Environ Microbiol* 68: 6361-6370.
- Kirkness EF, Bafna V, Halpern AL, Levy S, Remington K, Rusch DB, Delcher AL, Pop M, Wang W, Fraser CM, Venter JC (2003) The dog genome: survey sequencing and comparative analysis. *Science* 301: 1898-1903.
- Koepfli KP, Wayne RK (1998) Phylogenetic relationships of otters (Carnivora: Mustelidae) based on mitochondrial cytochrome *b* sequences. *J Zool Lond* 246: 401-416.
- Koepfli KP, Wayne RK (2003) Type I STS markers are more informative than cytochrome *b* in phylogenetic reconstruction of the Mustelidae (Mammalia: Carnivora). *Syst Biol* 52: 571-593.
- Kohn MH, Wayne RK (1997) Facts from feces revisited. *Trends Ecol Evol* 12: 223-227.
- Korenberg JR, Rykowski MC (1988) Human genome organization: Alu, lines, and the molecular structure of metaphase chromosome bands. *Cell* 53: 391-400.
- Kovalchuk O, Kovalchuk I, Arkhipov A, Hohn B & Dubrova YE (2003) Extremely complex pattern of microsatellite mutation in the germline of wheat exposed to the post-Chernobyl radioactive contamination. *Mutat Res* 525: 93-101.
- Krawczak M (1999) Informativity assessment for biallelic single nucleotide polymorphisms. *Electrophoresis* 20: 1676-1681.
- Kruglyak S, Durrett RT, Schug MD, Aquadro CF (1998) Equilibrium distributions of microsatellite repeat length resulting from a balance



between slippage events and point mutations. *Proc Natl Acad Sci USA* 95: 10774-10778.

Kruglyak S, Durrett RT, Schug MD, Aquadro CF (2000) Distribution and abundance of microsatellite in the yeast genome can be explained by a balance between slippage events and point mutations. *Mol Evol Biol* 17: 1210-1219.

Kuhner MK, Beerli P, Yamato J, Felsenstein J (2000) Usefulness of single nucleotide polymorphism data for estimating population parameters. *Genetics* 156: 439-447.

Kwok P, Deng Q, Zakeri H, Nickerson DA (1996) Increasing the information content of STS based genome maps: Identifying polymorphism in mapped STSs. *Genomics* 23: 138-144.

Lanza RL, Cibelli JB, Diaz F, Moraes CT, Farin PW, Farin EF, Hammer CJ, West MD, Damiani P (2000) Cloning of an endangered species (*Bos guarus*) using interspecies nuclear transfer. *Cloning* 2: 79-90.

Lee JS, Hanford MG, Genova JL, Farber RA (1999) Relative stabilities of dinucleotide and tetranucleotide repeats in cultured mammalian cells. *Hum Mol Genet* 8: 2567-2572.

Leonard JA, Wayne RK, Cooper A (2000) Population genetics of ice age brown bears. *Proc Natl Acad Sci USA* 97: 1651-1654.

Leopoldino AM, Pena SD (2003) The mutational spectrum of human autosomal tetranucleotide microsatellites. *Hum Mutat* 21: 71-79.

Levinson G, Gutman GA (1987a) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol* 4: 203-221.

Levinson G, Gutman GA (1987b) High frequencies of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in *Escherichia coli* K-12. *Nucleic Acids Res* 15: 5323-5338.



- Liang F, Han M, Romanienko PJ, Jasin M (1998) Homology-directed repair is a major double-strand break repair pathway in mammalian cells. *Proc Natl Acad Sci* 95: 5172–5177.
- Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, Clamp M, Chang JL, Kulbokas EJ 3rd, Zody MC, et al. (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438: 803-819.
- Lindstrom ER, Brainerd SM, Helldin JO, Overskaug K (1995) Pine marten – red fox interactions: a case of intraguild predation? *Ann Zool Fennici* 32: 123-130.
- Luan DD, Korman MH, Jakubczak JL, Eickbush TH (1993) Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: A mechanism for non-LTR retrotransposition. *Cell* 72: 595–605.
- Luan DD, Eickbush TH (1995) RNA template requirements for target DNA-primed reverse transcription by the R2 retrotransposable element. *Mol Cell Biol* 15: 3882–3891.
- Luikart G, England PR (1999) Statistical analysis of microsatellite DNA data. *Trends Ecol Evol* 14: 253–256.
- Macdonald DW, Barrett P (1993) *Mammals of Britain and Europe*. Harper Collins, London.
- Macdonald DW (2001) *The New Encyclopedia of Mammals*. Oxford University Press, Oxford.
- Maizeret C, Migot P, Rosoux R, Chusseau J-P, Gatelier T, Maurin H, Fournier-Chambrillon C (2002) The distribution of European mink (*Mustela lutreola*) in France: Towards a short term extinction? *Mammalia* 66: 525-532.



- Makova KD, Nekrutenko A, Baker RJ (2000) Evolution of microsatellite alleles in four species of mice (genus *Apodemus*). *J Mol Evol* 51: 166-172.
- Malek J (1993) The cat in ancient Egypt. British Museum Press, London.
- Mañas S, Cena JC, Ruiz-Olmo J, Palazón S, Domingo M, Wolfenbarger JB, Bloom ME (2001) Aleutian mink disease parvovirus in wild riparian carnivores in Spain. *J Wild Dis* 37: 138-144.
- Marmi J, López-Giráldez F, Domingo-Roura X (2004) Phylogeny, evolutionary history and taxonomy of Mustelidae based on cyt *b* gene and a complex repetitive flanking region sequences. *Zool Scr* 33: 481-499
- Martin AP, Pardini AT, Noble LR, Jones CS (2002) Conservation of a dinucleotide simple sequence repeat locus in sharks. *Mol Phylogenet Evol* 23: 205-213.
- Maran T (1991) Distribution of the European mink, *Mustela lutreola*, in Estonia: A historical review. *Folia Theriol Estonica* I: 1-17.
- Maran T (1992) The European mink, *Mustela lutreola*, in protected areas in the former Soviet Union. *Small Carniv Conserv* 7: 10-12.
- Maran T, Henttonen H (1995) Why is the European mink (*Mustela lutreola*) disappearing? – A review of the process and hypothesis. *Ann Zool Fenn* 32: 47-54.
- Maran T, Macdonald DW, Kruuk H, Sidorovich V, Rozhnov VV (1998) The continuing decline of the European mink, *Mustela lutreola*: evidence for the intraguild aggression hypothesis. *Symp Zool Soc Lond* 71: 297-323.
- Matera AG, Hellman U, Schmid CW (1990) A transpositionally and transcriptionally competent Alu subfamily. *Mol Cell Biol* 10: 5424-5432.
- Messing J, Bharti AK, Karlowski WM, Gundlach H, Kim HR, Yu Y, Wei F, Fuks G, Soderlund CA, Mayer KF, Wing RA (2004) Sequence composition



- and genome organization of maize. *Proc Natl Acad Sci USA* 101: 14349-14354
- Meyer S, Weiss G, von Haeseler A (1999) Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. *Genetics* 152: 1103-1110.
- Moos M, Gallwitz D (1983) Structure of two human beta-actin-related processed genes one of which is located next to a simple repetitive sequence. *EMBO J* 2: 757-761.
- Moran JV, Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD, Kazazian Jr HH (1996) High frequency retrotransposition in cultured mammalian cells. *Cell* 87: 917-927.
- Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nature Genet* 30: 194-200.
- Morin PA, Messier J, Woodruff DS (1994) DNA extraction, amplification, and direct sequencing from hornbill feathers. *J Sci Soc Thailand* 20: 31-41.
- Morin PA, Mahboubi P, Wedel S, Rogers J (1998) Rapid screening and comparison of human microsatellite markers in baboons: allele size is conserved, but allele number is not. *Genomics* 53: 12-20.
- Morin PA, Luikart G, Wayne RK, SNP Workshop group (2004) Applications of SNPs in ecology, evolution, and conservation. *Trends Ecol Evol* 19: 208-216.
- Moritz C, Faith DP (1998) Comparative phylogeography and the identification of genetically divergent areas for conservation. *Mol Ecol* 7: 419-429.
- Mouse Genome Sequencing Consortium (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* 420: 520-562



- Mowat G, Paetkau D (2002) Estimating marten *Martes americana* population size using hair capture and genetic tagging. *Wild Biol* 8: 201-209
- Nadir E, Margalit H, Gallily T, Ben-Sasson SA (1996) Microsatellite spreading in the human genome: evolutionary mechanisms and structural implications. *Proc Natl Acad Sci USA* 93: 6470-6475.
- Nagl W, Jeanjour M, Kling H, Kühner S, Michels I, Müller T, Stein B (1983) Genome and chromatin organisation in higher plants. *Biol Zentralbl* 102: 129-148.
- Nakamura Y, Leppert M, O'Connell P, Wolff R, Holm T, Culver M, Martin C, Fujimoto E, Hoff M, Kumlin E, et al. (1987) Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* 235: 1616-1622.
- Nauta MJ, Weissing FJ (1996) Constraints on allele size at microsatellite loci: Implications for genetic differentiation. *Genetics* 143: 1021-1032.
- Neff BD, Gross MR (2001) Microsatellite evolution in vertebrates: inference from AC dinucleotide repeats. *Evolution* 55: 1717-1733.
- Neal E, Cheeseman C (1996) *Badgers*. T & AD Poyser Ltd, London.
- Nielsen R (2000) Estimation of population parameters and recombination rates from single nucleotide polymorphisms. *Genetics* 154: 931-942.
- Nielsen R, Signorovitch J (2003) Correcting for ascertainment biases when analyzing SNP data: applications to the estimation of linkage disequilibrium. *Theor Popul Biol* 63: 245-255.
- Nikaido M, Rooney AP, Okada N (1999) Phylogenetic relationships among cetartiodactyls based on evidence from insertions of SINEs and LINEs: hippopotamuses are the closest extant relatives of whales. *Proc Natl Acad Sci USA* 96: 10261-10266.
- Nishizawa N, Nishizawa K (2002) A DNA sequence evolution analysis generalized by simulation and the Markov Chain Monte Carlo method





- implicates strand slippage in a majority of insertions and deletions. *J Mol Evol* 55: 706–717.
- Nota Y, Takenaka O (1999) DNA extraction from urine and sex identification of birds. *Mol Ecol* 8: 1235–1238.
- Novikov GA (1939) *The European mink*. Izd. Leningradskogo Gos. Univ., Leningrad.
- Nowak RM (1999) *Walker's mammals of the world, 6th ed.* Baltimore: Johns Hopkins University Press.
- Nowell K, Jackson P (1996) *Wild cats: status survey and conservation action plan*. IUCN publications. Cambridge: Burlington Press.
- Odom GL, Robichaux JL, Deininger PL (2004) Predicting mammalia SINE subfamily activity from A-tail length. *Mol Biol Evol* 21: 2140–2148.
- Ohashi J, Tokunaga K (2003) Power of genome-wide linkage disequilibrium testing by using microsatellite markers. *J Hum Genet* 48: 487–491.
- Okada N (1991) SINEs: short interspersed repeated elements of the eukaryotic genome. *Trends Ecol Evol* 6: 358–361.
- Okada N, Hamada M, Ogiwara I, Ohshima K (1997) SINEs and LINEs share common 3' sequences: a review. *Gene* 205: 229–243.
- Okada N, Shedlock AM, Nikaido M (2004) Retroposon mapping in molecular systematics. *Methods Mol Biol* 260: 189–226.
- Ono M, Kawakami M, Takezawa T (1987) A novel human nonviral retroposon derived from an endogenous retrovirus. *Nucleic Acids Res* 15: 8725–8737.
- Orti G, Pearse DE, Avise JC (1997) Phylogenetic assessment of length variation at a microsatellite locus. *Proc Natl Acad Sci USA* 94: 10745–10749.



- Ohshima K, Hamada M, Terai Y, Okada N (1996) The 30 ends of tRNA-derived short interspersed repetitive elements are derived from the 30 ends of long interspersed repetitive elements. *Mol Cell Biol* 16: 3756–3764.
- Paetkau D (1999) Microsatellites obtained using strand extension: an enrichment protocol. *Biotechniques* 26 690: 694-697.
- Palazón S (2003) Visión europeo - *Mustela lutreola*. In: *Enciclopedia Virtual de los Vertebrados Españoles*. Carrascal LM, Salvador A eds. MNCN, Madrid. <http://www.vertebradosibericos.org/>
- Pascual M, Schug MD, Aquadro CF (2000) High density of long dinucleotide microsatellites in *Drosophila subobscura*. *Mol Biol Evol* 17: 1259–1267.
- Pearce JM, Fields RL, Scribner KT (1997) Nest material as a source of genetic data for avian ecological studies. *J Field Ornithol* 68: 471-481.
- Pertoldi C, Hansen MM, Loeschcke V, Madsen AB, Jacobsen L, Baagoe H (2001) Genetic consequences of population decline in the European otter (*Lutra lutra*): an assessment of microsatellite DNA variation in Danish otters from 1883 to 1993. *Proc R Soc Lond B* 268: 1775-1781.
- Pluzhnikov A, Donnelly P (1996) Optimal sequencing strategies for surveying molecular genetic diversity. *Genetics* 144: 1247–1262.
- Primmer CR, Raudsepp T, Chowdhary BP, Møller AP, Ellegren H (1997) Low frequency of microsatellites in the avian genome. *Genome Res* 7: 471–482.
- Primmer CR, Ellegren H (1998) Patterns of molecular evolution in avian microsatellites. *Mol Biol Evol* 15: 997-1008.
- Pupko T, Graur D (1999) Evolution of microsatellites in the yeast *Saccharomyces cerevisiae*: role of length and number of repeated units. *J Mol Evol* 48: 313-316.
- Ramsay L, Macaulay M, Cardle L, Morgante M, Degli Ivanissevich S, Maestri E, Powell W, Waugh R (1999) Intimate association of microsatellite



- repeats with retrotransposons and other dispersed repetitive elements in barley. *Plant J* 17: 415–425.
- Rassmann K, Schlötterer C, Tautz D (1991) Isolation of simple-sequence loci for use in polymerase chain reaction-based DNA fingerprinting. *Electrophoresis* 12: 113–118.
- Regier JC, Wilson HM, Shultz JW (2005) Phylogenetic analysis of Myriapoda using three nuclear protein-coding genes. *Mol Phylogenet Evol* 34: 147–158.
- Renwick A, Davison L, Spratt H, King JP, Kimmel M (2001) DNA dinucleotide evolution in humans: fitting theory to facts. *Genetics* 159: 737–747.
- Rico C, Rico I, Hewitt G (1996) 470 million years of conservation of microsatellite loci among fish species. *Proc R Soc Lond B* 263: 549–557.
- Rogers JH (1985) The origin and evolution of retroposons. *Int Rev Cytol* 93: 187–279.
- Rosbash M, Ford PJ, Bishop JO (1974) Analysis of the C-value paradox by molecular hybridization. *Proc Natl Acad Sci USA* 71: 3746–3750.
- Rose O, Falush D (1998) A threshold size for microsatellite expansion. *Mol Biol Evol* 15: 613–615.
- Roy AM, Carroll ML, Nguyen SV, Salem A-H, Oldridge M, Wilkie AOM, Batzer MA, Deininger PL (2000) Potential gene conversion and source genes for recently integrated Alu elements. *Genome Res* 10: 1485–1495.
- Rubinsztein DC, Amos W, Leggo J, Goodburn S, Jain S, Li SH, Margolis RL, Ross CA, Ferguson-Smith MA (1995) Microsatellite evolution-evidence for directionality and variation in rate between species. *Nat Genet* 10: 337–343.



- Rudi K, Katla T, Naterstad K (2003) Multi locus fingerprinting of *Listeria monocytogenes* by sequence-specific labeling of DNA probes combined with array hybridization. *FEMS Microbiol Lett* 220: 9-14.
- Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, et al. (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409: 928-933.
- Sakamoto K, Okada N (1985) Rodent type 2 Alu family, rat identifier sequence, rabbit C family, and bovine or goat 73 bp repeat may have evolved from tRNA genes. *J Mol Evol* 22: 134-140.
- Salem A-H, Ray DA, Xing J, Callinan PA, Myers JS, Hedges DJ, Garber RK, Witherspoon DJ, Jorde LB, Batzer MA (2003) Alu elements and hominid phylogenetics. *Proc Natl Acad Sci USA* 100: 12787-12791.
- SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL (1998) The paleontology of intergene retrotransposons of maize. *Nat Genet* 20: 43-45.
- Santibáñez-Koref MF, Gangeswaran R, Hancock JM (2001) A relationship between lengths of microsatellites and nearby substitution rates in mammalian genomes. *Mol Biol Evol* 18: 2119-2123.
- Sato JJ, Hosoda T, Wolsan M, Tsuchiya K, Yamamoto Y, Suzuki H (2003) Phylogenetic relationships and divergence times among mustelids (Mammalia: Carnivora) based on nucleotide sequences of the nuclear interphotoreceptor retinoid binding protein and mitochondrial cytochrome *b* genes. *Zool Sci* 20: 243-264.
- Savolainen P, Zhang Y, Luo J, Lundeberg J, Leitner T (2002) Genetic evidence for an East Asian origin of domestic dogs. *Science* 298: 1610-1613.
- Schmid CW, Maraia R (1992) Transcriptional and transpositional selection of active SINE sequences. *Curr Opin Genet Develop* 2: 874-882.



- Schmid CW (1998) Does SINE evolution preclude Alu function? *Nucleic Acids Res* 26: 4541–4550.
- Schlötterer C, Tautz D (1992) Slippage synthesis of simple sequence DNA. *Nucleic Acids Res* 20: 211–215.
- Schlötterer C, Harr B (2000) *Drosophila virilis* has long and highly polymorphic microsatellites. *Mol Biol Evol* 17: 1641–1646.
- Schlötterer C (2003) Hitchhiking mapping — functional genomics from the population genetics perspective. *Trends Genet* 19: 32–38.
- Schmitz J, Ohme M, Suryobroto B, Zischler H (2002) The Colugo (*Cynocephalus variegatus* Dermoptera): the primates gliding sister? *Mol Biol Evol* 19: 2308–2312.
- Schropfer R, Paliocha E (1989) Zur historischen und rezenten Bestanderung der Nerze *Mustela lutreola* (L. 1761) und *Mustela vison* Schreber 1777 in Europaeine Hypothesendiskussion. *Wiss Beitr Univ Halle* 37: 303–321.
- Schug MD, Mackay TF, Aquadro CF (1997) Low mutation rates of microsatellite loci in *Drosophila melanogaster*. *Nature Genet* 15: 99–102.
- Schug MD, Wetterstrand KA, Gaudette MS, Lim RH, Hutter CM, Aquadro CF (1998) The distribution and frequency of microsatellite loci in *Drosophila melanogaster*. *Mol Ecol* 7: 57–70.
- Shedlock AM, Okada N (2000) SINE insertions: powerful tools for molecular systematics. *Bioessays* 22: 148–160.
- Shedlock AM, Milinkovitch MC, Okada N (2000) SINE Evolution, Missing Data, and the Origin of Whales. *Syst Biol* 49: 808–817.
- Shedlock AM, Takahashi K, Okada N (2004) SINEs of speciation: tracking lineages with retroposons. *Trends Ecol Evol* 19: 546–553.
- Shen MR, Batzer MA, Deininger PL (1991) Evolution of the master Alu gene(s). *J Mol Evol* 33: 311–320.



- Shimamura M, Yasue H, Ohshima K, Abe H, Kato H, Kishiro T, Goto M, Munechika I, Okada N (1997) Molecular evidence from retroposons that whales form a clade within even-toed ungulates. *Nature* 388: 666–670.
- Shimoda N, Knapik EW, Ziniti J, Sim C, Yamada E, Kaplan S, Jackson D, de Sauvage F, Jacob H, Fishman MC (2000) Zebrafish genetic map with 2000 microsatellite markers. *Genomics* 58: 219-232.
- Sibly RM, Meade A, Boxall N, Wilkinson MJ, Corne DW, Whittaker JC (2003) The structure of interrupted human AC microsatellites. *Mol Biol Evol* 20: 453-459.
- Sidorovich V (2000). The on-going decline of riparian mustelids (European mink, *Mustela lutreola*, polecat, *Mustela putorius*, and stoat, *Mustela erminea*) in eastern Europe: a review of the results to date and an hypotheses. In *Mustelids in a modern world. Management and conservation aspects of small carnivore: human interactions*, Griffiths HI Ed: 295–317.
- Singer-Sam JR, Tanguay C, Rigs AD (1989) Use of Chelex to improve PCR signal from a small number of cells. *Amplifications* 3:11.
- Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139: 457–462.
- Smit AFA, Riggs AD (1995) MIRs are classic, tRNA-derived SINEs that amplified before the mammalian radiation. *Nucleic Acids Res* 23: 98–102.
- Smit AF (1996) The origin of interspersed repeats in the human genome. *Curr Opin Genet Dev* 6: 743-748.
- Stoneking M, Fontius JJ, Clifford SL, Soodyall H, Arcot SS, Saha N, Jenkins T, Tahir MA, Deininger PL, Batzer MA (1997) Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa. *Genome Res* 7: 1061–1071



- Strand M, Prolla TA, Liskay RM, Petes TD (1993) Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* 365: 274-276.
- Sugiyama Y, Kawamoto S, Takenaka O, Kumizaka K, Norikatsu W (1993) Paternity discrimination and inter-group relationship of chimpanzees at Bossou. *Primates* 34: 545-552.
- Sunnucks P (2000) Efficient genetic markers for population biology. *Trends Ecol Evol* 15: 199-206.
- Sutherland GR, Richards RI (1995) Simple tandem DNA repeats and human genetic disease. *Proc Natl Acad Sci USA* 92: 3636-3641.
- Swift H (1950) The constancy of deoxyribose nucleic acid in plant nuclei. *Proc Natl Acad Sci USA* 36: 643-654.
- Symondson WOC (2002) Molecular identification of prey in predator diets. *Mol Ecol* 11: 627-641.
- Taberlet P, Griffin S, Goossens B, Questiau S, Manceau V, Escaravage N, Waits LP, Bouvet J (1996) Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Res* 24: 3189-3194.
- Taberlet P, Camarra JJ, Griffin S, Uhrès E, Hanotte O, Waits LP, Dubois-Paganon C, Burke T, Bouvet J (1997) Noninvasive genetic tracking of the endangered Pyrenean brown bear population. *Mol Ecol* 6: 869-876.
- Taberlet P, Waits LP, Luikart G (1999) Noninvasive genetic sampling: look before you leap. *Trends Ecol Evol* 14: 323-327.
- Takenaka O, Takasaki H, Kawamoto S, Arakawa M, Takenaka A (1993) Polymorphic microsatellite DNA amplification customized for chimpanzee paternity testing. *Primates* 34: 27-35.
- Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res* 17: 6463-6471.



- Taylor JS, Sanny JSP, Breden F (1999) Microsatellite allele size homoplasy in the guppy (*Poecilia reticulata*). *J Mol Evol* 48: 245–247.
- Terai Y, Takahashi K, Okada N (1998) SINE cousins: the 3' end tails of the two oldest and distantly related families of SINEs are descended from the 39 ends of LINEs with the same genealogical origin. *Mol Biol Evol* 15: 1460–1471.
- Ternovskij DV (1977) *Biology of mustelids (Mustelidae)*. Nauka Publ, Novosibirsk.
- Toth G, Gaspari Z, Jurka J (2000) Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res* 10: 967–981.
- Tumanov IL (1992) The number of the European mink (*Mustela lutreola* L.) in the eastern area and its relation to the American mink. In: *Semiaquatische Säugetiere*, Wiss Beitr Univ Halle: 329–335.
- Tuytens FAM, Delahay RJ, Macdonald DW, Cheeseman CL, Long B, Donnelly CA (2000) Spatial perturbation caused by a badger (*Meles meles*) culling operation: implications for the function of territoriality and the control of bovine tuberculosis (*Mycobacterium bovis*). *J Anim Ecol* 69: 815–828.
- Ullu E, Tschudi C (1984) Alu sequences are processed 7SL RNA genes. *Nature* 312: 171–172.
- Valiere N, Taberlet P (2000) Urine collected in the field as a source of DNA for species and individual identification. *Mol Ecol* 9: 2149–2154.
- Valsecchi E, Glockner-Ferrari D, Ferrari M (1998) Molecular analysis of the efficiency of sloughed skin sampling in whale population genetics. *Mol Ecol* 7: 1419–1422.
- Van Arsdell SW, Denison RA, Bernstein LB, Weiner AM, Manser T, Gesteland RF (1981) Direct repeats flank three small nuclear RNA pseudogenes in the human genome. *Cell* 26: 11–17.





- Van Den Bussche RA, Longmire JL, Baker RJ (1995) How bats achieve a small C-value: frequency of repetitive DNA in *Macrotus*. *Genome* 6: 521–525.
- Viard F, Franck P, Dubois M-P, Estoup A, Jarne P (1998) Variation of microsatellite size homoplasy across electromorphs, loci and populations in three invertebrate species. *J Mol Evol* 47: 42–51.
- Vignal A, Milan D, SanCristobal M, Eggen A (2002) A review on SNP and other types of molecular markers and their use in animal genetics. *Genet Sel Evol* 34: 275–305.
- Vila C, Savolainen P, Maldonado JE, Amorim IR, Rice JE, Honeycutt RL, Crandall KA, Lundeberg J, Wayne RK (1997) Multiple and ancient origins of the domestic dog. *Science* 276: 1687–1689.
- Voipio P (1946) Kuolevatko hilleri ja vesikko sukupuuttoon maassamme. *Metsästys ja Kalastus* 35: 40-43.
- Wakeley J, Nielsen R, Liu-Cordero SN, Ardlie K (2001) The discovery of single-nucleotide polymorphisms — and inferences about human demographic history. *Am J Hum Genet* 69: 1332–1347.
- Wan Q-H, Fang S-G (2003) Application of species-specific polymerase chain reaction in the forensic identification of tiger species. *Forensic Sci Int* 131: 75-78.
- Wang D, Coscoy L, Zylberberg M, Avila PC, Boushey HA, Ganem D, DeRisi JL (2002) Microarray-based detection and genotyping of viral pathogens. *Proc Natl Acad Sci USA* 99: 15687-15692.
- Wasser SK, Houston CS, Koehler GM, Cadd GG, SR Fain (1997). Techniques for application of faecal DNA methods to field studies of Ursids. *Mol Ecol* 6: 1091-1097.
- Watkins WS, Rogers AR, Ostler CT, Wooding S, Bamshad MJ, Brassington AM, Carroll ML, Nguyen SV, Walker JA, Prasad BV, et al. (2003) Genetic



- variation among world populations: inferences from 100 Alu insertion polymorphisms. *Genome Res* 13: 1607–1618
- Weaver RF, Hedrick PW (1992) *Genetics*. 2nd ed. Wm. C Brown, Dubuque, Iowa.
- Weber JL, Wong C (1993) Mutation of human short tandem repeats. *Hum Mol Genet* 2: 1123-1128.
- Webster MT, Smith NG, Ellegren H (2002) Microsatellite evolution inferred from human–chimpanzee genomic sequence alignments. *Proc Natl Acad Sci USA* 99: 8748–8753.
- Weiner AM, Deininger PL, Efstratiadis A (1986) Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. *Annu Rev Biochem* 55: 631-661.
- Weiner AM (2002) SINEs and LINEs: the art of biting the hand that feeds you. *Curr Opin Cell Biol* 14: 343–350.
- Wahle E (1995) 3'-End cleavage and polyadenylation of messenger-RNA precursors. *Biochim Biophys Acta* 1261: 183-194.
- Wierdl M, Dominska M, Petes TD (1997) Microsatellite instability in yeast: dependence on the length of the microsatellite. *Genetics* 146: 769-779.
- Wilder J, Hollocher H (2001) Mobile elements and the genesis of microsatellites in dipterans. *Mol Biol Evol* 18: 384-392.
- Wolsan M (1993) Phylogeny and classification of early European Mustelida (Mammalia: Carnivora). *Acta Theriol* 38: 345-384.
- Yandava CN, Gastier JM, Pulido JC, Brody T, Sheffield V, Murray J, Buetow K, Duyk GM (1997) Characterization of Alu repeats that are associated with trinucleotide and tetranucleotide repeat microsatellites. *Genome Res* 7: 716-724.



- Yang DY, Eng B, Wayne JS, Dudar C, Saunders SR (1998) Technical note: improved DNA extraction from ancient bones using silica-based spin columns. *Am J Phys Anthropol* 105: 539-543.
- Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. *Mol Ecol* 11: 1-16.
- Zardoya R, Vollmer DM, Craddock C, Streelman JT, Karl S, Meyer A (1996) Evolutionary conservation of microsatellite flanking regions and their use in resolving the phylogeny of cichlid fishes (Pisces: Perciformes). *Proc R Soc Lond Ser B* 263: 1589-1598.
- Zhang DX, Hewitt GM (2003) Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Mol Ecol* 12: 563-584.
- Zhivotovsky LA, Feldman LA, Grishchkin SA (1997) Biased mutation and microsatellite variation. *Mol Biol Evol* 14: 926-933.
- Zhu Y, Strassmann JE, Queller DC (2000) Insertions, substitutions, and the origin of microsatellites. *Genet Res* 76: 227-236.
- Zuckerklund E (1976) Gene control in eukaryotes and the c-value paradox "excess" DNA as an impediment to transcription of coding sequences. *J Mol Evol* 9: 73-104.