Genet. Res., Camb. (2005), **85**, *pp.* 223–233. *With 2 figures.* © 2005 Cambridge University Press
doi:10.1017/S0016672305007512 Printed in the United Kingdom

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

XAVIER DOMINGO-ROURA[1,2], FRANCESC LÓPEZ-GIRÁLDEZ[1,2]*, MIDORI SAEKI[3,4] AND JOSEP MARMI[2,5]

[1] *Genètica de la Conservació, Institut de Recerca i Tecnologia Agroalimentàries, Centre de Cabrils, Ctra. de Cabrils s/n, 08348 Cabrils (Barcelona), Spain*
[2] *Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Dr. Aiguader 80, 08003 Barcelona, Spain*
[3] *Wildlife Conservation Research Unit, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK*
[4] *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*
[5] *Departament de Ciència Animal i dels Aliments, Facultat de Veterinària, Universitat Autònoma de Barcelona, 08193 Cerdanyola del Vallès, Spain*

(*Received 12 July 2004 and in revised form 8 March 2005*)

Summary

We sequenced locus *Mel*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 proto-microsatellites 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

* 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, xavier.domingo@irta.es

https://doi.org/10.1017/S0016672305007512 Published online by Cambridge University Press

https://doi.org/10.1017/S0016672305007512 Published online by Cambridge University Press

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	*n*	Electromorph length	Repeat region length	Flanking regions length	GenBank/EMBL accession number
Mustelidae, Lutrinae							
Enhydra lutris	Alaska, USA	S. O'Brien, Laboratory of Genomic Diversity, National Cancer Institute	1	219	27	192	AJ489576
Lutra lutra	Wales, UK	A. Bradshaw, University of Cardiff	1	219	27	192	AJ489574
	Verhnedvinsk, Belarus	V. Sidorovic, National Academy of Sciences of Belarus	1	219	27	192	AJ489574
Amblonyx cinereus	Thailand	S. O'Brien, Laboratory of Genomic Diversity, National Cancer Institute	1	219	27	192	AJ489575
Lontra canadensis	Florida, USA	S. O'Brien, Laboratory of Genomic Diversity, National Cancer Institute	1	220	28	192	AJ489573
Mustelidae, Mustelinae							
Mustela erminea	Suffolk, UK	R. A. Macdonald, University of Bristol	1	221	29	192	AJ489565
	Suffolk, UK	R. A. Macdonald, University of Bristol	1	222	30	192	AJ489564
Mustela lutreola	Gorodok, Belarus	V. Sidorovic, National Academy of Sciences of Belarus	1	220	28	192	AJ489560
Mustela putorius	Oxfordshire, UK	A. Grogan, WildCRU, University of Oxford	1	220	28	192	AJ489562
	Catalonia, Spain	C. Rosell, Minuartia, Sant Celoni	1	220	28	192	AJ489562
Mustela nigripes	Captive	A. Kitchner, National Museums of Scotland	1	221	29	192	AJ489561
Mustela nivalis	Gorodoksky, Belarus	V. Sidorovic, National Academy of Sciences of Belarus	1	221	29	192	AJ489566
Mustela altaica	Mt Altai, Russia	E. I. Zholnerovskaya, Siberian Zoological Museum	1	222	30	192	AJ489563
Martes martes	Asturias, Spain	S. Lavín, Universitat Autónoma de Barcelona	1	222	30	192	AJ489568
	Smolevichsky, Belarus	V. Sidorovic, National Academy of Sciences of Belarus	1	222	30	192	AJ489568
	Kihniö, Finland	K. Kauhala, Finnish Game and Fisheries Research Institute	1	222	30	192	AJ489568
	Dijon, France	X. Domingo-Roura, University Pompeu Fabra	1	222	30	192	AJ489568
Martes melampus	Tokyo, Japan	Y. Fukue, Tokyo University of Agriculture and Technology	1	222	30	192	AJ489567
Martes foina	Asturias, Spain	S. Lavín, Universitat Autónoma de Barcelona	1	222	30	192	AJ489569
	Nesvizh, Belarus	V. Sidorovic, National Academy of Sciences of Belarus	1	222	30	192	AJ489570
Gulo gulo	Northwest Territories, Canada	M. A. Ramsay, University of Saskatchewan and C. M. Pond, Open University	1	222	30	192	AJ489571
Mustelidae, Melinae							
Meles meles	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							
Mephitis mephitis	Maryland, USA	S. O'Brien, Laboratory of Genomic Diversity, National Cancer Institute	1	196	22	174	AJ489577
Spilogale putorius	Oregon, USA	S. O'Brien, Laboratory of Genomic Diversity, National Cancer Institute	1	196	22	174	AJ489578–9
Procyonidae, Procyoninae							
Procyon lotor	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
Nasua nasua	London Zoo	D. Gottelli, Institute of Zoology, London	1	219	29	190	AJ489582
Procyonidae, Ailurinae							
Ailurus fulgens	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							
Arctocephalus gazella	South Georgia	J. P. Arnould, Macquarie University, and C. M. Pond, Open University	1	205	21	184	AJ489584
Otariidae, Otariinae							
Zalophus californianus	London Zoo	D. Gottelli, Institute of Zoology, London	1	205	21	184	AJ489585
Phocidae, Phocinae							
Phoca vitulina	London Zoo	D. Gottelli, Institute of Zoology, London	2	215	24	191	AJ489586

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 non-repetitive 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*

*Mel*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).

https://doi.org/10.1017/S0016672305007512 Published online by Cambridge University Press

*Mel*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 *Mel*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 $MgCl_2$.

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 *Mel*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 v 3.51 (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 (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 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 + Γ; $\alpha = 1{\cdot}375$) 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)

https://doi.org/10.1017/S0016672305007512 Published online by Cambridge University Press

https://doi.org/10.1017/S0016672305007512 Published online by Cambridge University Press

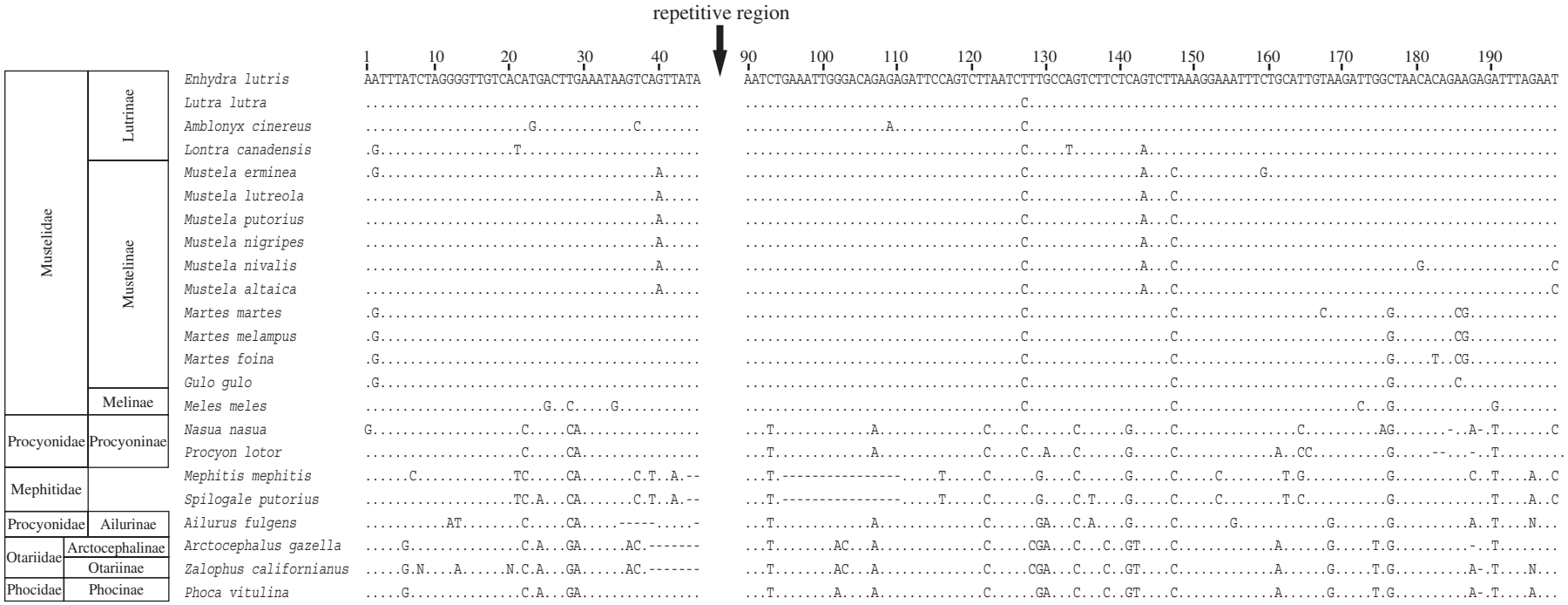

Fig. 1. Alignment of *Mel*08 flanking region sequences. Dots (.) indicate nucleotide identity, hyphens (-) indicate nucleotide deletions and N indicate unresolved nucleotide positions. The location of repetitive sequences is also indicated.

https://doi.org/10.1017/S0016672305007512 Published online by Cambridge University Press

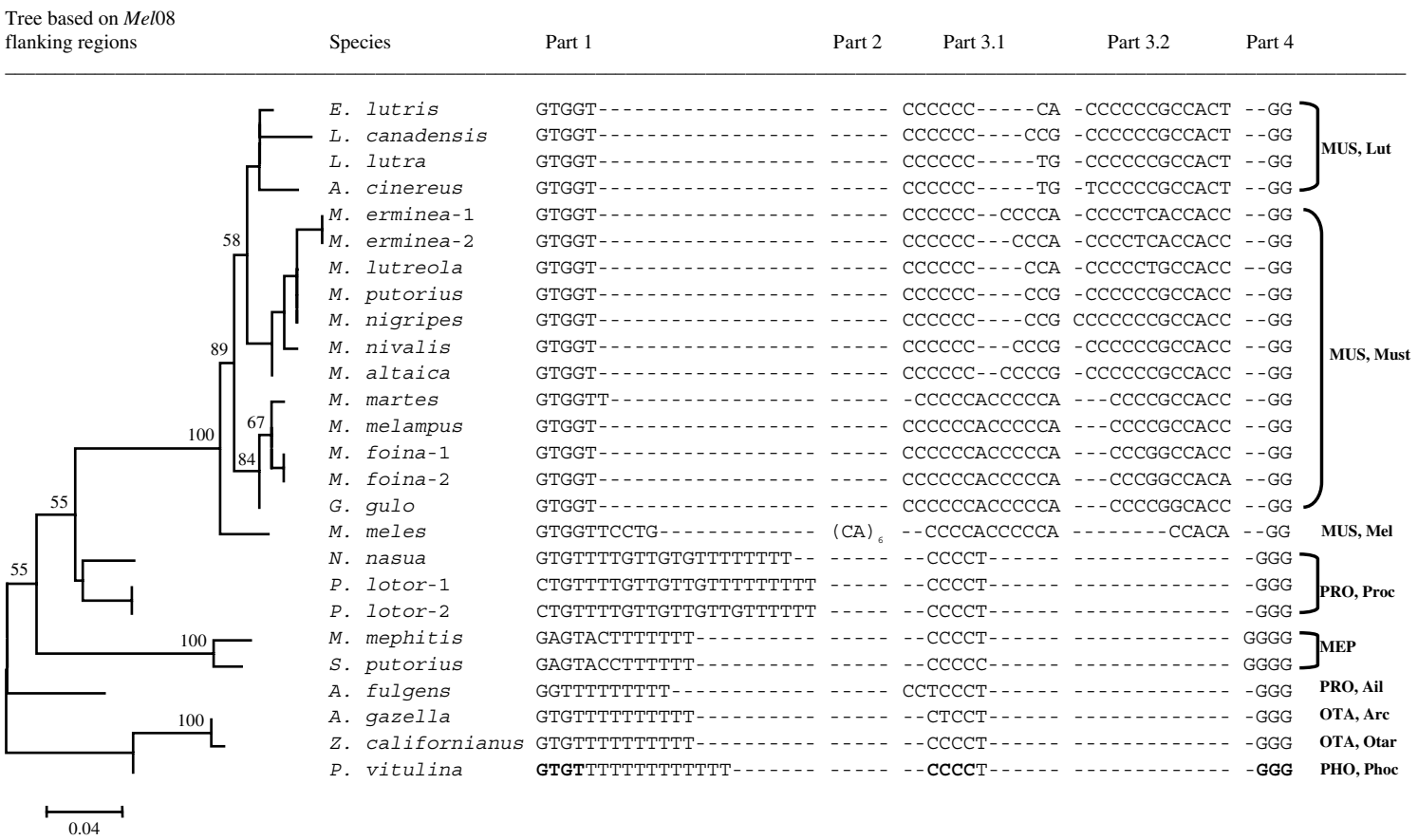

Fig. 2. Alignments of 26 allelic variants found in the repetitive region from different carnivore species that could be amplified at locus *Mel*08. The sequences have been divided into four parts according to the main repetitive pattern. Flanking regions are shown in Fig. 1. Dashes indicate the partial or complete absence of the different parts in given alleles. Bootstrap values higher than 50% of *Mel*08 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.

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 *Mel*08 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 (π) 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 $(CA)_6$ microsatellite in Part 2, which is absent in all other species. This repeat motif allowed the detection of the *Mel*08 locus during library hybridization (Domingo-Roura, 2002). Part 3 is based on an imperfect C_n 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_2 in the Mustelidae to G_4 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 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 within-species 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 ($\rho = 0{\cdot}014$, see below).

(v) *Phylogenetic information in the flanking regions*

The topology of the tree constructed from *Mel*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)

https://doi.org/10.1017/S0016672305007512 Published online by Cambridge University Press

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 ($Z=1{\cdot}794$, $N=36$, $P=0{\cdot}003$) and differences in repeat length including imperfections in Part 3 ($Z=1{\cdot}415$, $N=36$, $P=0{\cdot}036$) significantly deviated from normality. Divergence time among subfamilies was significantly correlated with the following parameters: (i) substitutions in flanking regions ($\rho=0{\cdot}442$, $P_S=0{\cdot}007$, $P_M=0{\cdot}019$); (ii) transversions in flanking regions $\rho=0{\cdot}411$, $P_S=0{\cdot}013$, $P_M=0{\cdot}017$); (iii) differences in the length of the largest perfect tract in Part 1 ($\rho=0{\cdot}377$, $P_S=0{\cdot}023$, $P_M=0{\cdot}024$) and (iv) differences in electromorph length but only for Mantel's test ($\rho=0{\cdot}014$, $P_S=0{\cdot}936$, $P_M=0{\cdot}020$). The variables that were not significantly correlated with time of divergence were differences in repeat length including imperfections in Part 1 ($\rho=0{\cdot}030$, $P_S=0{\cdot}860$, $P_M=0{\cdot}256$), differences in repeat length including imperfections in Part 3 ($\rho=0{\cdot}026$, $P_S=0{\cdot}879$, $P_M=0{\cdot}193$) and differences in the length of the longest perfect tract in Part 3 ($\rho=0{\cdot}048$, $P_S=0{\cdot}783$, $P_M=0{\cdot}154$). 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 *Mel*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 transition-prone 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 *Mel*08 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 *Mel*08 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

https://doi.org/10.1017/S0016672305007512 Published online by Cambridge University Press

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. *Mel*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 *Mel*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 *Mel*08 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 non-mephitine Mustelidae. Such a placement is supported by other works based on DNA sequences and morphology (cytochrome *b*: Ledge & Arnason, 1996; 851 bp of transthyretin intron I, cytochrome *b*, partial 12S rRNA, and morphological data: Flynn & Nedbal, 1998; Marmi *et al*., 2004; 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 *Mel*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

https://doi.org/10.1017/S0016672305007512 Published online by Cambridge University Press

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.

Supplementary material

Sequences have been deposited in the GenBank/EMBL database with accession numbers AJ309847, AJ489325, AJ489560–AJ489586.

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. 2000SGR 00093). F. L.-G. and J. M. 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).

References

Amrine-Madsen, H., Koepfli, K. P., Wayne, R. K. & Springer, M. S. (2003). A new phylogenetic marker, apolipoprotein B, provides compelling evidence for eutherian relationships. *Molecular Phylogenetics and Evolution* **28**, 225–240.

Bachtrog, D., Weiss, S., Zangerl, B., Brem, G. & Schlötterer, C. (1999). Distribution of dinucleotide microsatellites in the *Drosophila melanogaster* genome. *Molecular Biology and Evolution* **16**, 602–610.

Bell, G. I. & Jurka, J. (1997). The length distribution of perfect dimer repetitive DNA is consistent with its evolution by an unbiased single-step mutation process. *Journal of Molecular Evolution* **44**, 414–421.

Bininda-Emonds, O. R., Gittleman, J. L. & Purvis, A. (1999). Building large trees by combining phylogenetic information: a complete phylogeny of the extant Carnivora (Mammalia). *Biological Reviews of the Cambridge Philosophical Society* **74**, 143–175.

Bryant, H. N., Russell, A. P. & Fitch, W. D. (1993). Phylogenetic relationships within the extant Mustelidae (Carnivora): appraisal of the cladistic status of the Simpsonian subfamilies. *Zoological Journal of the Linnean Society* **108**, 301–334.

Clisson, I., Lathuilliere, M. & Crouau-Roy, B. (2000). Conservation and evolution of microsatellite loci in primate taxa. *American Journal of Primatology* **50**, 205–214.

Colson, I. & Goldstein, D. B. (1999). Evidence for complex mutations at microsatellite loci in *Drosophila*. *Genetics* **152**, 617–627.

Cooper, D. N. & Krawczak, M. (1991). Mechanisms of insertional mutagenesis in human genes causing genetic diseases. *Human Genetics* **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 Research* **13**, 2242–2251.

Domingo-Roura, X. (2002). Genetic distinction of marten species by the fixation of a microsatellite region. *Journal of Mammalogy* **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. *American Journal of Primatology* **43**, 357–360.

Eichler, E. E., Kunst, C. B., Lugenbeel, K. A., Ryder, O. A., Davison, D., Warren, S. T. & Nelson, D. L. (1995). Evolution of the cryptic FMR1 CGG repeat. *Nature Genetics* **11**, 301–308.

Ellegren, H. (2004). Microsatellites: simple sequences with complex evolution. *Nature Reviews in Genetics* **5**, 435–445.

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). *Molecular Biology and Evolution* **12**, 1074–1084.

Ezenwa, V. O., Peters, J. M., Zhu, Y., Arevalo, E., Hastings, M. D., Seppa, P., Pedersen, J. S., Zacchi, F., Queller, D. C. & Strassmann, J. E. (1998). Ancient conservation of trinucleotide microsatellite loci in polistine wasps. *Molecular Phylogenetics and Evolution* **10**, 168–177.

FitzSimmons, N. N., Moritz, C. & Moore, S. S. (1995). Conservation and dynamics of microsatellite loci over 300 million years of marine turtle evolution. *Molecular Biology and Evolution* **12**, 432–440.

Flynn, J. J. & Nedbal, M. A. (1998). Phylogeny of the Carnivora (Mammalia): congruence vs incompatibility among multiple data sets. *Molecular Phylogenetics and Evolution* **9**, 414–426.

Flynn, J. J., Nedbal, M. A., Dragoo, J. W. & Honeycutt, R. L. (2000). Whence the red panda? *Molecular Phylogenetics and Evolution* **17**, 190–199.

Ginsburg, L. (1982). Sur la position systemátique du petit panda, *Ailurus fulgens* (Carnivora, Mammalia). *Géobios, Mémoire Spécial* **6**, 247–258.

Halangoda, A., Still, J. G., Hill, K. A. & Sommer, S. S. (2001). Spontaneous microdeletions and microinsertions in a transgenic mouse mutation deletion system: analysis of age, tissue, and sequence specificity. *Environmental and Molecular Mutagenesis* **37**, 311–323.

Jeffreys, A. J., Murray, J. & Neumann, R. (1998). High-resolution mapping of crossovers in human sperm defines a minisatellite-associated recombination hot spot. *Molecular Cell* **2**, 267–273.

Jobb, G., von Haeseler, A. & Strimmer, K. (2004). TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evolutionary Biology* **4**, 18.

Koepfli, K. P. & Wayne, R. K. (2003). Type I STS markers are more informative than cytochrome *b* in phylogenetic reconstruction of the Mustelidae (Mammalia, Carnivora). *Systematic Biology* **52**, 571–593.

Kroutil, L. C. & Kunkel, T. A. (1999). Deletion errors generated during replication of CAG repeats. *Nucleic Acids Research* **27**, 3481–3486.

Kruglyak, S., Durrett, R. T., Schug, M. D. & Aquadro, C. F. (1998). Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. *Proceedings of the National Academy of Sciences of the USA* **95**, 10774–10778.

https://doi.org/10.1017/S0016672305007512 Published online by Cambridge University Press

Ledge, C. & Arnason, U. (1996). Phylogenetic analyses of complete cytochrome *b* genes of the order Carnivora with particular emphasis on the Caniformia. *Journal of Molecular Evolution* **42**, 135–144.

Levinson, G. & Gutman, G. A. (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology and Evolution* **4**, 203–221.

Macdonald, D. W. (2001). *The New Encyclopedia of Mammals*. Oxford: Oxford University Press.

Makova, K. D., Nekrutenko, A. & Baker, R. J. (2000). Evolution of microsatellite alleles in four species of mice (genus *Apodemus*). *Journal of Molecular Evolution* **51**, 166–172.

Marmi, J., López-Giráldez, J. F. & 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. *Zoologica Scripta* **33**, 481–499.

Nie, W., Wang, J., O'Brien, P. C., Fu, B., Ying, T., Ferguson-Smith, M. A. & Yang, F. (2002). The genome phylogeny of domestic cat, red panda and five mustelid species revealed by comparative chromosome painting and G-banding. *Chromosome Research* **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. *Journal of Molecular Evolution* **55**, 706–717.

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

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

Primmer, C. R. & Ellegren, H. (1998). Patterns of molecular evolution in avian microsatellites. *Molecular Biology and Evolution* **15**, 997–1008.

Rico, C., Rico, I. & Hewitt, G. (1996). 470 million years of conservation of microsatellite loci among fish species. *Proceedings of the Royal Society of London, Series B* **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, R. T., Aquadro, C. F. & Nielsen, R. (2004). Microsatellite mutation models: insights from a comparison of human and chimpanzees. *Genetics* **168**, 383–395.

Santibáñez-Koref, M. F., Gangeswaran, R. & Hancock, J. M. (2001). A relationship between lengths of microsatellites and nearby substitution rates in mammalian genomes. *Molecular Biology and Evolution* **18**, 2119–2123.

Sato, J. J., 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. *Zoological Science* **20**, 243–264.

Sato, J. J., 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. *Zoological Science* **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*. Geneva, Switzerland: Genetics and Biometry Laboratory, University of Geneva.

Sibly, R. M., Meade, A., Boxall, N., Wilkinson, M. J., Corne, D. W. & Whittaker, J. C. (2003). The structure of interrupted human AC microsatellites. *Molecular Biology and Evolution* **20**, 453–459.

Simpson, G. G. (1945). The principles of the classification and a classification of mammals. *Bulletin of the American Museum of Natural History* **85**, 1–350.

Stephan, W. & Kim, Y. (1998). Persistence of microsatellite arrays in finite populations. *Molecular Biology and Evolution* **15**, 1332–1336.

Swofford, D. L. (2002). *PAUP*: Phylogenetic Analysis Using Parsimony* (**And Other Methods*), *version 4.0b8*. Sunderland, MA: Sinauer.

Vrana, P. B., Milinkovitch, M. C., Powell, J. R. & Wheeler, W. C. (1994). Higher level relationships of the arctoid Carnivora based on sequence data and 'total evidence'. *Molecular Phylogenetics and Evolution* **3**, 47–58.

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

Wozencraft, W. C. (1989). The phylogeny of the Recent Carnivora. In *Carnivore Behavior, Ecology, and Evolution* (ed. J. L. Gittleman), pp. 495–535. Ithaca, NJ: Comstock Publishing Associates.

Wolsan, M. (1999). Oldest mephitine cranium and its implications for the origin of skunks. *Acta Palaeontologica Polonica* **44**, 223–230.

Zardoya, R., Vollmer, D. M., Craddock, C., Streelman, J. T., Karl, S. & Meyer, A. (1996). Evolutionary conservation of microsatellite flanking regions and their use in resolving the phylogeny of cichlid fishes (Pisces: Perciformes). *Proceedings of the Royal Society of London, Series B* **263**, 1589–1598.

https://doi.org/10.1017/S0016672305007512 Published online by Cambridge University Press