Faecal DNA template as non-invasive tools in order to distinguish the endangered Pyrenean desman (Galemys pyrenaicus, Eulipotyphla, Talpidae) from Mediterranean water shrews (Neomys anomalus, Soricomorpha, Soricidae)

José Luis Fernández-García*, María del Prado Vivas-Cedillo
Genetic and Animal Breeding, Universidad de Extremadura, 10071 Cáceres, Spain

Keywords: cytochrome b gene droppings Galemys pyrenaicus micro-mammals mitochondrial DNA molecular identification non-invasive tools PCR-RFLP design

Abstract
The molecular genetic tools using non-invasive samples such as faeces have an important role for the monitoring of wildlife and identification of elusive and rare species of semi-aquatic vertebrates. However, under aquatic environmental conditions the faeces undergo physical changes that harm the direct identification by visual inspection and leads to the confusion of the deposits from different cohabiting species. Consequently, the need of a refinement of molecular procedures is suggested. In this study this task was focused on the study of the reliability of the PCR-RFLP method to easily distinguish the faeces of the desman (Galemys pyrenaicus) from ecologically related species, including Neomys anomalus, but using an in silico-RFLP strategy. In this work, the faecal samples were collected from 2013 to 2014 from the LIC “Sierra de Gredos y Valle del Jerte” in the Mountain Central System of Spain, where the southernmost European population of the Pyrenean desman still survives. One primer pair was designed to amplify a very short portion of the mitochondrial cytochrome b gene containing targets for the restriction enzymes Hinf I, Mbo I and Hpy CH4 V, because the combination provides species-specific identifications. Pyrenean desman or Mediterranean water shrews were undoubtedly identified in the collected faeces. Moreover, the degree of homoplasy was studied under different scenarios of the three enzymes combination. Finally, the in silico-RFLP profiles given by the three enzymes combination have contributed with valuable information for a successful discrimination of faeces of the Pyrenean desman from those derived of N. fodiens, N. anomalus, Nectogale elegans and Cinclus cinclus.

Introduction
The desmans are part of the tribe Desmanini — within the subfamily Desmaninae Mivart, 1871 family Talpidae, order Eulipotyphla — with two genera and only one extant species each: (1) Galemys pyrenaicus E. Geoffroy, 1811; representing the Iberian or Pyrenean desman, and (2) Desmana moschata Linnaeus, 1758; the Russian desman. Recent molecular data support monophyly for these two desmanines which situate both as the last representatives of a relict line of great evolutionary and ecological interest (Igea et al., 2013). The Pyrenean desman (G. pyrenaicus) is a small semi-aquatic mammal endemic of the Iberian Peninsula. The species distribution roughly ranges from the Central Mountain System (Spain) and Sierra Da Estrela (Portugal) to the northern Iberian including the French side of the Pyrenees. The Pyrenean desmans live in typical mountain habitats but where the streams flow with clean cold well-oxygenated water and rich on benthic invertebrates (Nores et al., 2007; Igea et al., 2013). This species is characterized by large webbed hind feet, a double layered fur, a long tail and a mobile prehensile snout, which make it a specialist in finding and feeding on larvae of benthonic macroinvertebrates (Palmeirim and Hoffman, 1983; Richard, 1986). They live in pairs with a lifespan of about 3.5 years. The body size is between 11 and 16 cm in length; the tail measures between 12 and 16 cm and weighs between 35-80 g. It is considered by the IUCN (Fernandes et al., 2008) as a vulnerable relic species and, recently, experiencing elevated historical extinction rates (Igea et al., 2013). The genetic variability found at the cytochrome b (Cyt b) gene have permitted to regroup the Pyrenean desman in four phylogenetic lineages, with strong parapatric distribution and only one population declared to be admixed at contact zone (Igea et al., 2013). This strong signature of isolation among lineages has proved to be useful for unambiguously ascribing them to geographic localities (Igea et al., 2013).

Conservation of this endangered species is critically dependent on the knowledge of their populations (Frankham, 2003). Moreover, anthropogenic activities modify their natural habitat resulting in fragmentation, habitat loss or varying the distribution range of the species, which plays a role in shaping the structure of wildlife populations too (Gillet et al., 2014). All this information is essential to try to find molecular methods dedicated to the detection of this elusive species and to develop management strategies (Némoz and Bertrand, 2009) based on collected samples in as many localities as possible, rather than discrete populations (Igea et al., 2013).

Previous researches (Nores et al., 1998; González-Esteban et al., 2003) suggested that Pyrenean desman can be detected by faecal sampling or live-trapping. However, this small semi-aquatic vertebrate species is highly vulnerable to human management during live-trapping method (Gillet et al., 2014), which is supposed to be a potential risk for its survival. The development of non-invasive tech-
The tools for monitoring endangered Pyrenean desmans include techniques like collecting faecal deposits (Igea et al., 2013), made sampling less harmful for monitoring rare or elusive wild species (Lerone et al., 2014). However, these techniques may pose different issues such as a mistakenly assigned Pyrenean desman sample belonging to other semi-aquatic insectivorous species which shares habitats and diet. In the case of the desman a differential assignment should be performed, respect to mainly, three extant Neomys species (Family Soricidae): *N. fodiens* (Pennant, 1771), *N. anomalus* (Cabrera, 1907) and *N. teres* (Miller, 1908) (Corbet, 1978; Spitzenberger, 1990; Hutterer, 1993), since they have overlapped geographic distribution. The Eurasian water shrew (*N. fodiens*) is a species of palaearctic distribution which can be found across Europe from the north of the Iberian Peninsula to lake Baikal (including the British Isles to Russia, Norwest Mongolia, China and Nord Korea) but it is absent in the middle and southern Spain. The Mediterranean water shrew (*N. anomalus*, Cabrera 1907) is a palaearctic species that occupies most of the woodlands from southern and central Europe and extends eastwards (Black Sea, Ukraine and Asia Minor) except for the absence in much of the south and west of France, which has caused the isolation of the Iberian and the French Pyrenees populations with respect to the rest of the continental species. It is widely distributed in the northern half of the Iberian Peninsula and also in the eastern half of Andalusia and Extremadura (Spain). Based on somatic dimensions and genetic data, the two recognized subspecies differ significantly from the central European morphotype *N. a. milleri* minor than the nominal subspecies *N. a. anomalus* which is being considered the autochthonous morphotype of the Iberian peninsula (Ventura, 2002 and references therein). However, the geographic range of the transcaucasian water shrew (*N. teres*) is comparatively small, being found in Armenia, Azerbaijan, Georgia, Iran and Turkey. Other less related Soricomorpha species like *Nectogale elegans* (Milne-Edwards 1870) has been reported in China, North Sikkim (India) and Eastern Nepal (Moltur et al., 2005) being only interesting as a molecular control. The White-throated dipper (*Cinclus cinclus*, Linnaeus 1758) is found throughout the Palaearctic region associated to fast flowing streams, so contributing to confound the desmans’ faeces.

The PCR-Restriction Fragment Length Polymorphism (PCR-RFLP) approach (where a PCR product is digested by restriction endonucleases digestions targeting short DNA motifs to generate a specific genetic profile) has been widely used to survey from faeces (Mukherjee et al., 2010) but also it has been proved to be useful to distinguish up to a maximum of 7 sympatric species (Cossíos and Anger, 2006; Bidlack et al., 2007). Hence, when a vast diversity of species occur in sympatry, a set of protocols like PCR-RFLP for assigning scats accurately to species is required (Mukherjee et al., 2010), mainly if PCR product sizes do not differ among species (Montes et al., in press). Recently, in order to discard the dropings of sympatric species, Gillet et al. (2014) have been implementing a PCR-RFLP technique when some confusion can occur with not-fresh, water-soaked faeces or other reasons. However, Gillet et al. (2014) did not include the *N. anomalus* among them. Accordingly, a most complete identification is therefore of great importance to detect the presence of the desman. This is the reason why the aim of this study was (1) to study the reliability of the PCR-RFLP method to avoid and control the wrong assignments of species in small semi-aquatic and ecologically related species, including specially *N. a. anomalus*, (2) to implement this procedure within the context of non-invasive methods able to easily distinguish desmans’ faeces and (3) to prevent mistake of assignations of species attributable to inter-specific homoplasy at some sites by using in silico-RFLP on disposable Gen-Bank sequences.

**Material and methods**

**Sampling**

All the samples from faeces and tissues were collected with the authorizations licensed by the Extremadura Government (see acknowledgement). The tissue samples of desman were collected from fortuitously dead specimens (personal communication of the senders). The faeces of supposedly desman (n=97) were collected by professionally qualified personnel from three geographical protected areas belonging to LIC (“Lugares de Importancia Comunitaria”: officially declared as sites of community importance) of the Red Natura2000 “Sierra de
Gredos y Valle del Jerte**: Ambroz Valley district, Jerte Valley district and La Vera district from north-western to north-eastern, respectively, of the Spanish Central Mountain System but inside the Extremadura autonomy (south-western Spain) (Fig. 1). Here the two semi-aquatic *N. anomalus* and *G. pyrenaicus* overlap. The field collection was done in the summer from 2013 to 2015 and the faeces preserved using a 1.8 ml sterile cryo-tube with screw cap (certified premium Line, LABBOX Germany) filled with 70% ethanol and stored at least to -20°C until the next step.

Tissue samples of the Pyrenean desman (n=6) and Mediterranean water shrew (n=1), were used here as PCR controls. Except water shrew sample, the rest of control samples were selected from populations genetically unrelated from those in the LIC sites (Igea et al., 2013) for best monitoring possible cross contaminations after sequencing positive scats. Tissues were preserved in ethanol 70% like described for faeces.

DNA extraction

Genomic DNA from faecal samples were extracted using the Real Pure Spin Food Stool (Real), following the manufacturer protocols (DUR-VIZT S.L. Spain). The genomic DNA from tissues were extracted using a salting out procedures (Fernández-García, 2012). All DNA extractions were made in a dedicated room keeping clean the working area and under a Bunsen burner during the entire procedure but a UV-sterilised cabinet where no tissue samples had been previously managed were used for faeces. DNA concentration was measured using a Nano-Drop ND-1000 spectrophotometer (NanoDrop Technologies from Thermo Fisher Scientific). In our faecal extraction the DNA concentration was 7-25 ng/µl or even less, but it was not possible to discard the presence of contaminant DNA as those from food of the desman or other sources. Then, the DNA from tissues should be diluted to around 10-20 ng/µl to accommodate to the most common DNA concentration from faeces.

**Primer design**

Due to the low quality of DNA extracted from faeces (Paetkau, 2003; Lerone et al., 2014) more efficient primer pairs should be designed under the following criteria: (1) to amplify a small size portion of the Cyb gene, since DNA extracted from faeces should be frequently degraded, (2) flanking species-specific Restriction Endonuclease (RE) targets for the species of interest, mainly those with overlapping geographic distribution as occurred for the desman and two shrews (*N. fodiens* and *N. anomalus*) in Spain, (3) that the number of RE cutting site inside the amplicon should have from zero to a maximum of three target, respectively. The shorter portion flanked by the designed primer pair, that is, 202 bp the amplicon should have from zero to a maximum of three target, namely those that can be used for PCR and the sequences of all these species (see Acc. number in Tab. 1). The RESTRICTIONMAPPER 3.0 software (http://www.restrictionmapper. org/) was used to predict suitable RE with species-specific target within the trimmed sequences. A panel of three RE (*Mbo I*, *Hin f I* and *Hpy CH4 V*) were selected on the basis of inter-species different cutting sites found and according to the criteria of primer selection (see primer design; Material & Methods). The checking of cleavage targets across all the downloaded sequences was conducted using MEGA 4.0 software (Tamura et al., 2007).

**Partial cytchrome b amplifications using conventional PCR and PCR-RFLP**

All faeces samples were subjected to conventional PCR amplification with *CybNeA*-marker (this study) for after positive *G. pyrenaicus* extract being studied with *Cyb1*-marker, *Cyb2*-marker, *Cyb3*-marker according to Igea et al. (2013) (data not showed). The primer of the *CybNeA*-marker described here amplified a fragment of 202bp that expanded from the nucleotide 368 to 569 of the Cyb gene. This 202 bp fragment overlapped the last and the first portion amplified with *Cyb1*-marker and *Cyb2*-marker, respectively (Igea et al., 2013). Positive reactions to *CybNeA*-marker were subjected to PCR-RFLP analysis as described below. Positive samples to desman were further amplified with *Cyb1* to *Cyb3*-marker and positive PCR were sequenced to identify desman haplotypes (unpublished data). PCR reactions were prepared in a final volume of 25 µl containing 4-5 µl DNA, 1 µM of each primer (forward primer NeAF and reverse primer NeAR; *CybNeA*-marker), 0.25 µM dNTPs, 1.5 mM MgCl2, 1X buffer reaction and 0.5 U of Biotaq DNA polymerase (Ecogen®). Amplifications were performed with a 2720 thermal cycler (Life Technologies®) under the next protocol: one denature step at 94°C for 5 min followed by 40 cycles (denaturation at 94°C for 30 seconds, annealing at 54°C for 60 seconds, extension at 72°C for 30 seconds) and final extension step at 72°C for 7 minutes. The protocol used when the DNA template consisted of DNA from faecal samples was as follow: the first and penultimate tubes of the racks were reserved as negative controls but the final one was loaded with the positive control using *N. anomalus* template. PCR products were electrophoresed on 1.6% low melting agarose gels (ethidium bromide stained, 0.5 µg/ml) for 1 h at 100 V, but using the first slot for the molecular weight marker (100bp Ladder; NZlitech®).

Finally, the surplus PCR reaction from positive tissue and faecal samples (including positive controls) were divided into aliquots of 5 µl to be digested with each one of the selected restriction enzymes. The positive enzymatic digestions were performed in 10 µl volume (5 µl of the PCR product) according to manufacturer recommendations. Digestions were performed for 4 hours at 37°C in a thermal cycler. 10 µl of the digestion were electrophoresed on 2% (Low Melting) agarose gel (ethidium bromide stained, 0.5 µg/ml) for 1 hour at 100 V and, again, the first slot loading with molecular weight marker (100 bp Ladder; NZlitech®).

**Species identification by Restriction Enzyme using in silico-RFLP**

The BLAST tool (Zhang et al., 2000) at NCBI (www.ncbi.nlm.nih.gov) was used to download all the disposable sequences deposited in Gen-Bank of the mitochondrial Cyb (Tab. 1) from the following species: *G. pyrenaicus*, *D. moschata*, *N. anomalus*, *N. fodiens*, *N. teres*, *N. elegans* and *C. cinctus* but for these ones only in silico analysis was performed. The primer mispriming by species can be seen in Tab. S1.
Table 1: Species, subspecies, size as sequence numbers and geographic locations of the GenBank sequences used for the in silico-RFLP study. Below this first information, the in silico-RFLP haplotypes and its corresponding fragment size (FS) for each restriction enzyme. Acc. N. = Accession Numbers. FS = Fragment size. * Asterisk for Cinclus cinclus because analyzed sequences was treated as a whole which ranged from Northern Africa to Eastern Europe.

<table>
<thead>
<tr>
<th>Species</th>
<th>G. prenasus</th>
<th>D. moschata</th>
<th>N. anomalus</th>
<th>N. foliaceus</th>
<th>N. teres</th>
<th>N. elegans</th>
<th>Cinclus cinclus*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subspecies</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>(n=135)</td>
<td>(n=2)</td>
<td>(n=13)</td>
<td>(n=5)</td>
<td>(n=8)</td>
<td>(n=21)</td>
<td></td>
</tr>
<tr>
<td><strong>Acc. N. by location</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unspecified</td>
<td>AR000466</td>
<td>AR000465</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Combined RFLP-Haplotype</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hpy CH4 V (FS)</td>
<td>C:142,60</td>
<td>C:142,60</td>
<td>C:142,60</td>
<td>C:142,60</td>
<td>C:142,60</td>
<td>C:142,60</td>
<td></td>
</tr>
</tbody>
</table>

Caucasus
-iskey
Central Asia
-iskey
Eastern Asia
-iskey
Central Europe
-iskey
Eastern Europe
-iskey
Northern Europe
-iskey
Southern Europe
-iskey
Combined RFLP-Haplotype

FS = Fragment size, * Asterisk for Cinclus cinclus because analyzed sequences was treated as a whole which ranged from Northern Africa to Eastern Europe.
by species and by district within the LIC can be seen in Fig. S2 as deduced from our result. With the support of in silico-RFLP analysis (Tab. 1), all field samples carried the most common RFLP-haplotype from desman (A1B1C; Fig. 2). Once the positive faeces were assigned as desman, a next stage was performed. These samples were used to be amplified by PCR using the three Cyb primers pairs described in Igea et al. (2013) and those successful amplifications subjected to sequencing methods so as to be assessed again (Fernández-García et al., 2015). The 17% of faeces assigned as shrews showed exclusively the A2B3U-haplotype (Tab. 1) which may be highly frequent in the N. a. anomalus (Fig. 2) from the Iberian Peninsula as suggested by the in silico-RFLP assay. Consequently, these samples were rejected for further PCR assays.

The in silico-RFLP analysis of the available GenBank sequences belonging to the desman and the nominal subspecies of the water shrew were always consistently concordant with the tissue controls. According to all these results, the digestion with three restriction enzymes leads to an unambiguous identification of both species, desman vs Mediterranean water shrew. In spite of this fact, it was demonstrated that genetic variability of the Cyb across different species can cause some confusing species assignments (Hansen and Jacobsen, 1999; Gomez-Moliner et al., 2004; Gillet et al., 2014). For example, this some confusing species assignments (Hansen and Jacobsen, 1999; M. M. Hansen M.M., Jacobsen L., 1999. Identification of mustelid species: otter (Lutra lutra), man and the rest of species tested, both in silico-RFLP analysis and empirically. As stated by Galtier et al. (2006), the mtDNA is greatly characterized by a high level of homoplasy, i.e., phylogenetic/genealogic conflict between sites, which should be solved using exhaustive data set whatever the origin of the homoplasy: the existence of mutation hot spots or ancestral polymorphism. However, the combination of the three restriction patterns was found a valid diagnostic method as deduced from all known sequences of the Pyrenean desman, that is, in the 45 haplotypes described by Igea et al. (2013) and Cabria et al. (2006). Moreover, in silico-RFLP should be performed in other species as N. fodiens vs. N. anomalus, N. elegans vs C. cinscus (Tab. 1) in order to study and compare to others conflicting genetic profiles. But the last species should be considered important because it is also a frequent inhabitant of Iberian rivers.

Finally, in the present study our conclusion is that the accuracy of the PCR-RFLP methods with the aim to identify ecologically cohabiting species depends critically on the abundance and quality of the molecular data sets tested. Hence, apart from genetic information (Fagundes, 2005), well identified faeces provide a wealth of information to conduct ecological studies (Costa et al., 2016) focusing on habitat use (Rivero et al., 2005), diet (Stewart et al., 2003; Najera-Hillman and Mandujano, 2013), density (Pertiaga and Leynaud, 2009) or endocrino- logical research (Pereira et al., 2006).

References


Tools for monitoring endangered Pyrenean desmans


Supplemental information

Additional Supplemental Information may be found in the online version of this article:

Table S1 Primers used to amplify the cytochrome b fragment, and mismatch distribution of primer bias in different sequences of each species or subspecies within the homologous selected primer sites.

Figure S2 Number of PCR positive samples distributed by species.