U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

Development of a new methodology for species and individual identification in canids

Diana Carina da Costa Lobo

Dissertação de Mestrado apresentada à
Faculdade de Ciências da Universidade do Porto em
Biodiversidade, Genética e Evolução

2013

Development of a new methodology for species and individual identification in canids

Diana Carina da Costa Lobo

Biodiversidade, Genética e Evolução
Centro de Investigação em Biodiversidade e Recursos Genéticos
2013

Orientador
Raquel Godinho, phD, CIBIO
Alejandro Rodríguez, phD, EBD/CSIC

Coorientador
Francisco Álvares, phD, CIBIO
José López-Bao, phD, SLU

Agradecimentos

A tese aqui apresentada foi fruto de um trabalho coletivo fundamental que envolveu o desempenho e empenhamento de muitas pessoas para que se pudesse concretizar, razão pela qual não poderia deixar de agradecer-vos uma última vez.

Naturalmente que o meu primeiro agradecimento vai para toda a minha família, aqueles que sempre estiveram do meu lado. Um obrigada infinito aos meus pais por todo o apoio durante todos estes anos, especialmente pelo enorme esforço financeiro e por nunca terem duvidado deste momento!

Obrigada à minha irmã Cris, que tão gentilmente me emprestou o seu computador nos meus momentos de desespero quando a lentidão do meu quase me enlouquecia.

Aos meus orientadores, porque desde o primeiro dia me acolheram como um membro da equipa, sempre me apoiaram e partilharam muito do vosso enorme conhecimento e paixão. Muito obrigada Francisco pelas incontáveis viagens ao canil e todo o apoio no “campo” e por teres sempre uma palavra carinhosa! Raquel não poderia agradecer o suficiente a tua constante presença, o apoio e conselhos que me deste ao longo deste ano! Alejandro e José-Vicente senti-vos sempre por perto, o vosso constante esforço em estar sempre disponíveis para mim foi excepcional e o conhecimento que me passaram foi tremendo, obrigada!

Sem a vossa ajuda e o vosso ânimo na reta final não teria sido possível!

Aos meus colegas de mestrado, a todos eles, que fizeram com que esta fosse uma experiência inesquecível, sem vocês tudo teria sido diferente! Sem as nossas clássicas “Noites no Porto” o mundo teria outra cor! Obrigada por toda a ajuda e partilha de conhecimento, principalmente ao André pelos “elevados” conhecimentos em estatística e ao Hugo (nosso companheiro de velhas andanças), por estar sempre um passo à frente.

A toda a equipa do CTM que me recebeu com imensa generosidade e me ensinaram da forma mais carinhosa e profissional que poderia esperar! O trabalho que vocês desenvolvem todos os dias é fenomenal! Uma palavra de agradecimento em particular à Diana, minha mentora desde o início que me ensinou e ajudou em todos os momentos do desenvolvimento laboratorial desde trabalho, a ela devo muito do

êxito que conseguimos. À Susana, que no meio de tanto trabalho conseguia sempre receber-me com todo o carinho; foste fundamental. Por fim, à Filipa, companheira de trabalho sem a qual todos aqueles dias não teriam sido tão divertidos, és uma pessoa muito especial!

A todos os funcionários do canil "A Cerca", foram fabulosos na ajuda que me deram durante os 2 meses que ai estive, o vosso trabalho é de louvar! Aos funcionários do canil do ICBAS/UP que me prestaram todo o auxílio e foram muito acolhedores comigo.

À Corticeira Amorim por ter cedido gratuitamente cortiças para este trabalho.

À Comissão de Ética do ICBAS/UP por ter analisado este estudo e ter liberado uma licença para a manipulação dos animais durante o plano de trabalhos.

E para terminar, ao Ricardo, meu companheiro de todas as horas! Pelos inúmeros momentos de apoio, carinho e amor, sem ti ao meu lado nada teria sido possível! Obrigada por seres a minha pessoa!

Por tudo isto e muito mais, um eterno obrigada a todos!

Resumo

A necessidade de métodos de amostragem que permitam realizar estimativas populacionais precisas de canídeos selvagens é cada vez maior, no entanto, a monitorização destes predadores pode ser bastante difícil. A genética não-invasiva tem assumido um papel cada vez mais importante na resolução destes problemas. Contudo, apesar da vasta gama de aplicações, o ADN obtido em amostras não-invasivas normalmente apresenta qualidade e quantidade bastante reduzidas, implicando a realização de múltiplas réplicas no processo de amplificação, para evitar erros de genotipagem. Neste trabalho descrevemos o desenvolvimento de um novo procedimento de amostragem não-invasiva, cujo objetivo é obter ADN de um substrato (cortiça, esferovite e madeira) revestido por um isco (paté de cão e sardinha) atrativo aos canídeos, que deverá manter-se por tempo suficiente na boca dos animais de forma a absorver grandes quantidades de saliva. Estes substratos e iscos foram testados em cães de canil para avaliar o efeito do substrato, isco e da interação entre ambos no(a) (i) quantidade de saliva observada; (ii) concentração de ADN obtida e (iii) sucesso de identificação específica e individual através da análise genética utilizando, respetivamente, um fragmento de 435-pb de ADN mitocondrial e um conjunto de dez microssatélites autossómicos desenvolvidos para cão. Os nossos resultados demonstraram que a quantidade de saliva observada foi apenas afetada pelo tipo de substrato, sendo a madeira o substrato onde os valores registados foram mais elevados. Segundo, foram observados efeitos significativos dos tipos de isco e de determinadas combinações entre o substrato e o isco na concentração de ADN, onde se verificou que a combinação madeira/paté providencia maior quantidade de ADN e que o isco sardinha apresenta os valores mais baixos, possivelmente indicando que a composição das sardinhas em lata degrada as moléculas de ADN. Ainda assim, a concentração média de ADN obtida (11.9 ± 24.2 ng/µl) foi incrivelmente elevada para um método não-invasivo. Terceiro, os diferentes tipos de substrato e isco não influenciaram significativamente o sucesso de identificação específica e individual (85 e 84%, respetivamente). Os erros de genotipagem (“ADO” = 1.6%; “FA”= 1.3%) foram praticamente residuais em todos os tipos de substrato e isco. Adicionalmente, foi possível obter uma média de duas réplicas necessárias para identificar os indivíduos, um número bastante inferior aos que são apresentados para a maioria de amostras não-invasivas.

Desta forma conclui-se que este método é muito eficaz no âmbito da genética não-invasiva em canídeos, permitindo obter elevadas concentrações de ADN e elevadas taxas de sucesso na identificação específica e individual, com registo residual de erros de genotipagem, e reduzir a necessidade de múltiplas amplificações que, consequentemente, diminui os custos de genotipagem, possibilitando em larga medida a monitorização e conservação de canídeos selvagens.

PALAVRAS-CHAVE: amostragem não-invasiva, canídeos, identificação específica, identificação individual, isco, genética não-invasiva, substrato.

Abstract

Despite the need for reliable methods providing accurate population estimates to manage and conserve wild canids, monitoring these predators could be difficult and poses many practical problems. Non-invasive genetics has been an important tool for researchers to overcome these limitations. Beyond the set of applications and potentialities, DNA obtained from non-invasive sampling methods is usually of low quality and quantity and requires multiple genotyping replicates to avoid genotyping errors. Here we described a new non-invasive sampling procedure for obtaining DNA from a substrate (cork, styrofoam and wood) covered with a bait (dog's pâté and sardine) attractive to canids, which should remain in animal's mouth enough time to absorb saliva. We tested these substrates and baits in captive domestic dogs in order to evaluated the effect of substrate, bait and their interaction on (i) the amount of saliva, (ii) the concentration of extracted DNA and (iii) the success of species and individual identification from genetic analysis using a 425-bp fragment of mitochondrial DNA and a set of 10 dog-specific autosomal microsatellites, respectively. First, our results showed that the amount of saliva observed was only affected by substrate types, being wood the substrate with highest values. Second, significant effects of bait types and certain substrate/bait combinations on DNA concentration were observed. The wood/pâté combination provided higher quantities of DNA while sardine had the lowest values, suggesting that canned sardines may have certain substances in their composition that degrade DNA molecules. Nevertheless, the average of DNA concentration obtained (11.9 ± 24.2 ng/µl) was remarkably high for a non-invasive method. Third, we found that different types of substrate and bait have no significant effect on species and individual identification success (85 and 84%, respectively). Genotyping error rates were residual in all types of substrate and bait (ADO = 1.6%; FA = 1.3%). Additionally, this method allowed to achieve an average of two replicates required to identify individuals, a lower number than those generally needed for non-invasive samples.

This method proved to be efficient in non-invasive genetics of canids, allowing to achieve remarkably higher amounts of DNA and higher success on species and individual identification, with residual genotyping errors, and to reduce the need of multiple PCR amplifications and thus to reduce costs of management and conservation of wild canids.

KEYWORDS: bait, canids, individual identification, non-invasive genetics, non-invasive sampling, species identification, substrate.

List of contents

List of tables

List of figures

List of abbreviations

ADO	Allelic dropout
ADN	Ácido desoxirribonucleico
AIC	Akaike Information Criterion
ANOVA	Analysis of Variance
bp	Base Pair
DA	Dog's attributes
DB	Dog's behaviour
DNA	Deoxyribonucleic acid
e.g.	For example

et al	And others
Exp	Exposure time
FA	False alleles
Fw	Forward
GLM	Generalized Linear Models
GLMM	Generalized Linear Mixed Models
ID	Identification
Ind	Individual
LT	Time licking
MT	Time in mouth
mtDNA	Mitochondrial DNA
ns	Not significant
pb	Pares de bases
PBS	Phospate buffered saline
PCR	Polymerase Chain Reaction
PI	Probability of Identity
Rep	Replicates
rpm	Revolutions per minute
Rv	Reverse
Sard	Sardine
STRs	Short Tandem Repeats
Styr	Styrofoam
Sub	Substrate
TT	Handling time
UV	Ultraviolet
w	Weight
wi	Akaike weights

Chapter 1
Introduction

Mammalian carnivores face a myriad of challenges in our overcrowded world, as habitat loss and fragmentation continue to increase (Gese, 2004). Under this scenario, many carnivore populations have experienced dramatic range contractions and are in urgent need of protection (Laliberte and Ripple, 2004), which pressure conservation biologists, more than ever, to produce relevant and sound data pertaining to carnivore distribution, habitat use and other biological and ecological traits (Gese, 2004).

Canids are the main group of carnivores for which worldwide conservation or management problems have been identified (Gese, 2001; Gese, 2004; Ginsberg and Macdonald, 1990; Wilson and Delahay, 2001), which increases the need for knowledge related with their status and ecological requirements. Several threats such as habitat loss, human persecution, prey depletion, disease and increased competition with other carnivores due to reduced space and habitat, have led several canid species to face extinction or to suffer a strong reduction of their original range (Barea-Azcón *et al*., 2007; Gese, 2001). These constraints to their survival turn imperative to obtain reliable methods to estimate species distribution and abundance, in order to provide scientific-based data to support management actions.

1.1. Sampling methods for wild canids

Determining the occurrence and estimating population abundance of species is fundamental for their conservation, research and management. However, this could be difficult and poses many practical problems on a large spatial scale and in long-term monitoring because many wild canids are secretive, nocturnal, wide ranging in densely vegetated habitats or remote areas, or occur at extremely low densities (Caughley and Sinclar, 1994; Gese, 2004; Sadlier *et al.,* 2004; Silveira *et al*., 2003). Few methods are suitable for monitoring elusive and low-density species (Mills *et al.,* 2000) in spite of the amount of available monitoring methods (Williams *et al.,* 2002).

Animal abundance may be monitored indirectly by counting presence signs, or by direct methods of detecting the animals themselves (Gese, 2004). Direct counts may use either dead animals (e.g. harvest reports, mortality records) or alive animals

(e.g. trapping, sightings) (Burnham *et al.,* 1980; Gese, 2004). The most common sampling methods for direct counts are road mortality records (Clark and Andrews, 1982), remote camera traps (Harrison *et al.,* 2002), capture-mark-recapture (Caughley, 1977) and radio-telemetry (Fuller, 1989). For wild canids the most classical method to detect their presence and estimate relative abundance is the use of indirect methods such as signs surveys (scats, footprints, dens, vocalizations) (Barea-Azcón *et al*., 2007; Bider, 1968; Lyra-Jorge *et al*., 2008), because of the great difficulty to directly contact or capture such animals. Sign surveys can be passive (e.g. the observer simply records the number of signs found along a transect) or may use attractants that elicit the approximation of the target species (Gese, 2004; Lyra-Jorge *et al.,* 2008). One of the most common sign surveys used for indexing canid abundance is scent-stations surveys (Linhart and Knowlton, 1975; Roughton and Sweeny, 1982; Schauster *et al.,* 2002). A scent-station is a smoothed natural surface, which includes an attractant to lure wild canids (Linhart and Knowlton, 1975). Scent-stations are inexpensive and can be successfully deployed in great quantities over large areas by researchers with very little training, also allowing the simultaneous monitoring of a diverse array of species (Ray and Zielinski, 2008). Usually, the frequency of visits is used as an index of abundance. However, visits to multiple stations by a single animal may result in overestimation of canid abundance (Gese, 2004). Nevertheless, field-based species identification may be ambiguous or unfeasible so additional efforts and highly skilled and experienced trackers are needed to validate the identification of species or individuals. This weakness related to species identification combined with the limited availability of appropriate tracking means or conditions, the ephemeral and weather-dependent character of tracks and the inconsistent survey designs and quality control procedures, have resulted in a growing criticism of track surveys and the need to improve survey efforts to meet more rigorous standards, as non-invasive genetic sampling (Gese, 2004). The possibility to extract DNA from samples collected non-invasively during track surveys (e.g. scats, hairs, urine, saliva, menstrual blood) provides an important opportunity to confirm species identification (specially for rare species) and distinguish individuals, an important measured to determine wild canids relative abundance together with the traditional track surveys methods (Heinemeyer *et al.*, 2008).

1.2. Non-invasive genetic sampling

Non-invasive sampling allows studying animals without disturbing them (Beja-Pereira *et al.,* 2009), making genetic sampling much easier. Non-invasive genetic studies have used DNA as a diagnostic tool generally to achieve information about species with ecological or behavioural requirements that make them difficult-to-study (Schwartz and Monfort, 2008). Popular uses of non-invasive DNA include species identification (Farrel *et al.,* 2000; Ruell and Crooks, 2006; Sundqvist *et al.,* 2008; Valière *et al.,* 2003; Valière and Taberlet, 2000), individual identification (Clevenger and Sawaya, 2010; Ruell and Crooks, 2006; Schwartz *et al.,* 2007; Sloane *et al.,* 2000; Taberlet and Luikart, 1999; Valière and Taberlet, 2000) and sex identification (Pilgrimm *et al.,* 2005; Taberlet *et al.,* 1997), which can be used for a panoply of studies covering molecular ecology (Bellemain *et al.,* 2005; Frantz *et al*., 2004; Hung *et al*., 2004; Kohn *et al.,* 1999; Palomares *et al.,* 2002; Zielinski *et al.,* 2006), interspecific hybridization (Adams *et al*., 2003; Schwartz *et al*., 2004), paternity and relatedness (Banks *et al.,* 2002; Schwartz *et al.,* 2004), phylogeography and population genetics (Iyengar *et al.,* 2005; Triant *et al.,* 2004), among many others. Beyond the set of applications and potentialities, DNA obtained from non-invasive sampling methods is usually of low quality and quantity (Gagneux *et al.,* 1997; Taberlet *et al.,* 1999; Taberlet and Luikart, 1999) which entails multiple genotyping replicates to avoid errors as allelic dropout (amplification failure of an allele in a heterozygote) or the amplification of false alleles (Gagneux *et al.,* 1997; Gossens *et al.,* 1998; Taberlet *et al.,* 1996). Both types of error can have dramatic consequences on inferences (genetic, ecological, behavioural, among others) made over the data (e.g. Bonin *et al.,* 2004; Broquet and Petit, 2004; Gagneux *et al.,*1997; Miller *et al*., 2002; Mills *et al.,* 2000; Roon *et al.,* 2005; Schwartz and Monfort, 2008).

DNA analysis has been attempted on several biological materials ranging from pelts kept in Museums to regurgitates found on trails, but hair (Morin and Woodruff, 1992; Taylor *et al*., 1997) and scats (Prugh *et al*., 2005; Taberlet *et al.,* 1997) are the most common source of non-invasive genetic material collected for wildlife research and monitoring. Scats are widely used when studying mammals because they are easy to find in the wild and provide additional information (e.g. diet, stress hormones, parasites; Goymann, 2005; Kohn and Wayne, 1997; Schwartz and Monfort, 2008). Scats contain many sloughed epithelial cell and most carnivore's fecal samples are large enough to allow multiple attempts at DNA recovery (Schwartz and Monfort, 2008). The greatest constraint to scat analysis is the presence of inhibitors that

restrict DNA amplification (Schwartz and Monfort, 2008). Also, amounts and quality of fecal DNA are known to vary by species, temperature, age and season at time of collection, preservation method, diet, storage time and extraction protocol (Maudet *et al.,* 2004; Murphy *et al.,* 2002; Nsubuga *et al.,* 2004; Piggott and Taylor, 2003; Santini *et al.,* 2007). Moreover, scats may be difficult to find for some species and their number may be too small to genotype a significant fraction of the population (Schwartz and Monfort, 2008). Hair collected in hair traps provides DNA of better quality than scats mostly due to the lower amount of inhibitors, but avoiding cross-contamination might be difficult because multiple individuals can visit the same hair trap (Kendall and McKelvey, 2008). Urine is also used for non-invasive studies in mammals (Hausknecht *et al*., 2006), however, has a lower amplification success rate when compared with other non-invasive samples (Hedmark *et al.,* 2004).

Saliva is also known as a good source of non-invasive DNA (Chiappin *et al.,* 2007; Ng *et al.,* 2006; Quinque *et al.,* 2006) often used in wildlife forensics for predator identification (Blejwas *et al.,* 2006; Sundqvist *et al*., 2008; Williams *et al.,* 2003). New saliva swab sampling techniques could improve non-invasive sampling studies and genotyping, involving material such as eggshells or any surface that animals come into contact, rub against, lick or bite (e.g. rocks, sticks) (Beja-Pereira *et al.,* 2009; Pang and Cheung, 2007). However, further evaluation of this and other sampling methods are needed to improve non-invasive studies based on saliva and increase amplification success rates.

1.3. Molecular markers

Genetic markers are simply heritable characters with multiple states at each character, reflecting differences in DNA sequences (Allison, 2007; Avise, 2004; Schlotterer, 2004). Deciding whether to examine sections of the nuclear or mitochondrial genome depends largely on the goals of the study. Concerning the use of non-invasive samples in molecular ecology studies, two goals are generally in mind: identify species, for which the mitochondrial genome is typically used (Farrel *et al.,* 2000; Ruell and Crooks, 2006; Schwartz and Monfort, 2008; Sundqvist *et al.,* 2008; Valière *et al.,* 2003) and/or to identify individuals for which nuclear DNA is often preferred (Avise *et al.,* 1987; Clevenger and Sawaya, 2010; Schlotterer, 2004; Schwartz *et al.,* 2007; Taberlet and Luikart, 1999). Amplifying and analyzing mitochondrial DNA (mtDNA) is generally easier than nuclear DNA, due to several unique properties of this molecule, including high copy number (highly useful on

degraded samples) and single maternal inheritance which prevents recombination, simplifying interpretation of results (Allison, 2007; Avise, 2004). One of the most variable regions of mammals mtDNA, is the control region, because of its non-coding function, which have made it the region of choice for studies of population history in canids (e.g. dogs and wolves) (Pilot *et al.,* 2006; Vilà *et al.,* 1997, among others). Despite the advantages of high copy number and lack of recombination, mtDNA mutation rates generally do not allow individual identification (Allison, 2007; Avise, 2004). To complete that task, it is necessary to apply highly variable molecular markers, as is the case of microsatellites (or STRs – Short Tandem Repeats) (Allison, 2007; Avise, 2004).

Microsatellites are tandemly repeated sequences, with short repeat motifs generally not extending overall more than 100 base pairs (Schlotterer, 2004). Microsatellites properties like being codominant markers, have high mutation rates and be easy to score together with the ability to be inexpensively searched in the genome of a particular species (Schlotterer, 2004; Schwartz and Monfort, 2008), have made them a popular tool for accessing individual and sex identification, relatedness estimates, hybridization patterns, pedigree reconstruction, estimates of census and effective population size, and the level of genetic polymorphism within or between populations (Pilgrim *et al.,* 2005; Queller *et al.,* 1993; Taberlet *et al.,* 1999). Currently, these markers are one of the most common genetic tools for the use in non-invasive samples, related both with the properties described and with their small size, highly convenient for low quality DNA.

1.4. Objectives

The main goal of this study is to develop a new non-invasive sampling procedure for wild canids by obtaining DNA from saliva.

We explore the possibility of obtaining DNA from a surface (substrate) covered with a bait attractive to canids, which should remain into the animal mouth long enough to absorb saliva, but simultaneously would provoke repulsion and be expelled, increasing the probability of recovering the substrate.

Three substrates (cork, styrofoam and wood) were selected to test for differences in attributes such as hardness (related with the probability of detection and expulsion) and porosity (storage capacity of saliva) and two baits (dog's pâté and sardine) were used to test dog's preferences and the presence of substances that could degrade DNA, difficult its extraction, or inhibit PCR. We tested these

substrates and baits in captive domestic dogs in order to evaluate the effect of substrate, bait and the interaction between both on: (i) the amount of saliva obtained; (ii) the concentration of extracted DNA and (iii) the success for species and individual identification through genetic analysis.

The principal outcome of this work, after selected the most profitable substrate and bait, was to provide a new and improved non-invasive tool for sampling wild canids, allowing, species and individual identification with a high success rate.

Chapter 2
Material and methods

2.1. Substrates and baits selected and sampling protocol

The three substrates selected, cork (6 x 3.5 cm), styrofoam (4 x 5.5 cm) and wood (3.5 x 5.5 cm) (Figure 1) were covered with the two baits, dog's pâté and sardine. Previously to the experiment, all substrates were exposed to ultraviolet (UV) light for a period of 30 minutes, to sterilize and eliminate all traces of foreign DNA.

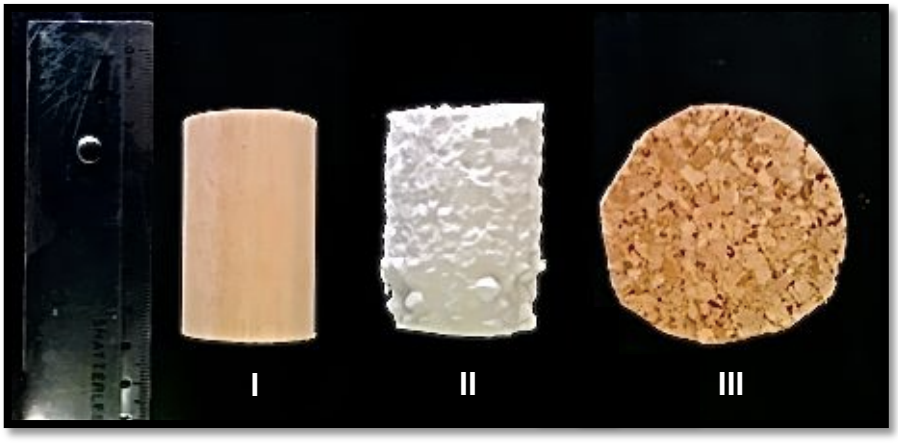

Fig.1 – Selected substrates tested in this study: I – wood; II – styrofoam and III – cork.

A total of 170 handled substrates (wood, n = 58; styrofoam, n = 54 and cork, n = 58) were obtained from 29 dogs kept in kennels. Were performed five experimental sets within each, six different dogs were sampled (except in set 3, in which only 5

dogs were sampled), for the six possible combinations of substrate/bait. Each experimental set was performed during nine days, with 24 and 48 hours intervals after days 2 and 5, respectively, to avoid habituation of dogs to bait and procedure. In each of the days, the six dogs were maintained in individual cages during the experiment at which the substrate/bait combination was introduced. The combinations of substrate/bait assign to each dog were randomly selected in order to avoid bias for statistical analysis. Dog behaviour while handling substrates was recorded by video camera to measure (i) the handling time, (ii) the substrate/bait in mouth time and (iii) the licking time. The substrates were collected after dogs leaving them for periods longer than 1 minute and do not return to handling the substrates, except when it was not possible (e.g. dogs swallow or destroy substrates, n = 4). The quantity of saliva present in each substrate was evaluated, considering a seven level scale from 1 to 4 categories with the respective intermediate values (1.5, 2.5 and 3.5), ranging from category 1 (no saliva) to category 4 (substrate surface completely saturated with saliva).

Substrates were preserved in ethanol 96% until DNA extraction.

Additionally, dogs were weighed and hair samples were taken, as a reliable source of DNA, for individual identification.

2.2. Laboratory procedures

From the total of 170 handled substrates were used 134 substrates (wood, n = 46; styrofoam, n = 42 and cork, n = 46) from 24 dogs to perform the genetic analysis. DNA was extracted using QIAamp DNA Micro Kit. Before DNA extraction, substrates were incubated at 37°C for two days to remove the remaining ethanol. Dry substrates were washed with 15ml of PBS solution to release the cells from the substrate surface and the resulting solution was centrifuged at 4000 rpm for 25 minutes. After discard supernatant, all proceedings were made according QIAamp DNA Micro Kit manufacturer's instructions (DNA was eluted with 80µl of Elution Buffer). A maximum of fourteen samples and one negative control (in order to detect exogenous DNA contamination) were handled in each extraction round (see complete DNA extraction protocol in figure A1, in appendix A). DNA from hair samples (n = 24) was also extracted using QIAamp DNA Micro Kit according to manufacturer's instructions (DNA was eluted with 50µl of Elution Buffer) along with a negative control. Two elutions were performed per extraction for both hair and saliva samples. Extraction rounds were carried under meticulous conditions to avoid

contamination, which included special clothing, mascara and gloves. DNA quality from all samples was assessed by electrophoresis at 0.8% agarose gel.

DNA concentration in saliva samples (both elutions) was measure by fluorescence using Wallac Victor fluorometer (PerkinElmer Life Sciences) set at 480nm for excitation and 520nm for emission according Quant-it PicoGreen dsDNA Assay Kit manufacturer's instructions. Each sample was quantified three times and values were averaged to accept the final concentration of each sample.

In order to perform species identification (despite the fact that we already know the species, we want to perform the species identification to evaluated the percentage of success of our method in this type of analysis), a 425-bp fragment of the mitochondrial control region was amplified, using universal primers Thr-L 15926 and DL-H 16340, as described by Vilà *et al.,* (1999). Mitochondrial amplification was prepared in an 11µl final volume reaction containing 2µl of DNA (first elution), using the Multiplex PCR Kit (QIAGEN) following PCR conditions given in the manufacturer's instructions and with the annealing temperature set at 52°C (PCR mix and conditions are shown in table A1 and A2, respectively, in appendix A). Successful amplifications were purified using enzymes exonuclease I and Shrimp alkaline phosphatase and sequenced with BigDye chemistry (Applied Biosystems). Electropherograms were verified using SEQSCAPE 2.5 (Applied Biosystems).

Individual multilocus genotypes were determined using a set of ten dog autosomal microsatellites (see loci description in table A3, in appendix A). All loci were amplified in a single multiplex reaction using the Multiplex PCR Kit (QIAGEN) following PCR conditions indicated in the manufacturer's instructions, prepared in an 10µl final volume reaction containing 2µl of DNA (first elution). Thermocycling used a touchdown profile with the annealing temperature decreasing from 60°C to 57°C in seven cycles, followed by twenty-two cycles with constant annealing temperature set to 57°C and more eight cycles with constant annealing temperature set to 53°C (PCR mix and conditions are shown in tables A1 and A2, respectively, in appendix A). Three replicates were performed (multi-tubes approach, Navidi *et al.,* 1992; Taberlet *et al.,* 1996) for each amplification. DNA from hair samples was genotyped as described above for individual genotyping, but without replicates. PCR products were separated by size on an ABI3100xl genetic analyzer using the 350ROX size standard. Alleles were determined using GENEMAPPER 4.0 (Applied Biosystems) and checked manually.

All amplifications were performed in Bio-Rad Thermal cyclers, always using negative controls to monitor possible contaminants. PCR products were visualized by electrophoresis at 2% agarose gel.

2.3. Data analysis

Amount of saliva

To evaluate if the substrate type and the remaining sampling factors (Table 1) could interfere or induce in the amount of saliva we specified a set of eighteen candidate models that could potentially predict the amount of collected saliva, therefore restricting the model selection process to a few meaningful combinations of predictors. The variable saliva was transformed in proportion (dividing each one of the seven levels by seven) and after applied a square root arcsine transformation to normalize data.

Table 1 - Fixed and random effects used to model relative amount of saliva.

Variable	Code	Definition	Type of factor
Substrate	sub		
Cork	cork	Type of substrate tested	Fixed
Styrofoam	styr	Type of substrate tested	Fixed
Wood	wood	Type of substrate tested	Fixed
Bait	bait		
Dog's pâté	pâté	Type of bait tested	Fixed
Sardine	sard	Type of bait tested	Fixed
Dog's attributes	DA		
Sex	sex	Dog's gender: female/male	Fixed
Weight	w	Dog's weight (kg)	Fixed
Dog's Behaviour	DB		
Time in mouth	MT	Time that dog's spent with substrate in mouth (sec)	Fixed
Time licking	LT	Time that dog's spent licking substrate (sec)	Fixed
Handling time	TT	Total time that dog's spent handling substrate (sec)	Fixed
Exposure time	Exp	Dog's exposure time to the experience (h): Day 1 represents 0 exposure; day 2 represents 24h exposure and so one.	Fixed
Individual	Ind	Dogs used in the study	Random
Replicate	Rep	Number of replicates/experimental sets	Random
Day	Day	Number of days in each experimental set	Random

We considered the substrate and bait as the factors of interest and also considered confounding fixed factors, as i) the dog attributes (sex and weight), to evaluate if differences between genders could affect the production of saliva, as also if bigger dogs would produce more saliva; ii) dog behaviour (time that dogs spent with substrate in mouth, licking and the total handling time), since longer handlings might induce salivation as also the type of manipulation and iii) the dog's exposure time to the experience, to control the effect of a possible habituation from dogs to the experience in the amount of saliva produced. The identity of the individual, the replicate and the day of the experimental set were included as random factors. Interactions between substrate with bait and dog's behaviour with substrate/bait combinations were also tested in models, in order to evaluate if a specific combination of substrate/bait could induce more saliva or influence the dog's handling time. We fitted an intercept-only equation in order to test improvement over the null model of no effect. Fitted models were compared using corrected Akaike Information Criterion (AIC_c) and models weights (Burnham and Anderson, 2004). We rank models by their AIC_c values and determined the model averaged parameter estimates (Burnham and Anderson, 2004). The relative variable importance of predictor variable *x* (w_x) is determined as the sum of the w_i across all models where *x* occurs. Larger w_x values indicate a higher relative importance of variable *x* compared to other variables (Burnham and Anderson, 2004). These parameters were calculated in R 2.15.2 (R Development Core Team, 2011) using the package qpcR (Spiess, 2013).

We select models that present Δi ($AIC_i - AIC_{min}$) ≤ 2, representing the models that have substantial support (evidence) (Burnham and Anderson, 2004). Select model was fitted using Generalized Linear Mixed Models (GLMM) with a Gaussian error distribution and an identity link in R using the package lme4 (Bates, 2013), and *p*-values were estimated using the package languageR (Baayen, 2011). Multiple comparisons of means using Tukey Contrast test were used, after selected the significant variables among models, to detect significant relationships between the subgroups of the significant factors.

DNA concentration

To evaluate if DNA concentration was correlated with the amount of saliva observed in substrates we perform a Pearson correlation test between DNA concentration and the amount of saliva.

To assess the effect of substrate/bait combinations on DNA concentration, we specified a set of four candidate models, considering the substrate and bait as the

factors of interest and also the interaction between them. DNA concentration variable was normalized using logarithm transformation and outliers were removed (after performing a Grubbs test, with 95% confidence level). We fitted an intercept-only equation and fitted models were compared using AIC_c and models weights as described above for the amount of saliva.

Selected model was fitted using Generalized Linear Models (GLM) with a Gaussian error distribution and an identity link in R using the package lme4 (Bates, 2013). After selected the significant factors among models, multiple comparisons of means using Tukey Contrast test were used, to detect significant relationships between the subgroups of the significant factors.

Species and Individual identification

To evaluate the success of species identification using DNA obtained in substrates, we set a list of eight candidate models, considering the effect of DNA (to evaluate if DNA concentration influenced the success of species identification), substrate and bait (to assess if substrate and bait compositions could interfere with amplification in PCR). We attributed 1 when species identification was successful or 0 when identification failed. We fitted an intercept-only equation and model selection was made by AIC_c and models weights as described before. Selected models were fitted using GLM with a Binomial error distribution and a logit link in R using the package lme4 (Bates, 2013).

The success of individual identification was assessed through the amplification of ten loci. For each of the three replicates in each sample, we compared the genotype observed at each locus with the true genotype known by the analysis of the corresponding hair sample and attributed 1 when genotype was corrected or 0 if genotype was not correct or no amplification was observed. Then, we calculated the percentage of individual identification success using the ratio between the number of loci with correct genotype and the total number of loci. We analyzed the percentage of success in the three replicates from each sample and assumed the higher value achieved as the final percentage of success for that sample. We also determined the rate of genotyping errors (allelic dropout and false alleles) between the ten loci for substrate and bait, using GIMLET 1.3.3 (Valière, 2002b). We performed ANOVA in R, to test significant differences between the results of ADO (allelic dropout) and FA (false alleles) from the different types of substrate and bait.

The success rate of individual identification was evaluated through eight models with an intercept-only equation, considering substrate, bait and DNA concentration as principal factors and selects the best models by AIC_c criteria and models weights.

Selected models were fitted using GLM with a Binomial error distribution and a logit link in R using the package lme4 (Bates, 2013).

Additionally, we assessed the number of genotyping replicates required for accurate individual identification for each substrate and bait, since a low number of replicates mean lower lab costs. We compared the three replicates from each sample with the corresponding hair sample and evaluated the necessary number of replicates (1,2,3 or 3+) to achieve a complete genotype for our set of ten loci and for eight loci which were enough for individual identification; PI values (probability of identity) varying between 0.269 to 0.033 (determined using GIMLET). For both variables, we considered, independently, the effect of substrate, bait and DNA concentration on eight models with an intercept-only equation, selected by AIC_c criteria and models weights. Selected models were fitted using GLM with a Poisson error distribution and a log link in R using the package lme4 (Bates, 2013)

Finally, we used the three genotyping replicates to calculate the success rate of individual identification without making a direct comparison to the control hair sample. This is particular relevant because we ultimately intend to develop a useful method to be applied in wildlife studies, for which control DNA samples are generally not available. For that we determine the consensus genotype of each sample. We first estimate ADO and FA rates using a maximum likelihood approach implemented in PEDANT software (Johnson and Haydon, 2007). Based on ADO and FA results, we estimated the minimum number of repetitions that confirms an allele per locus (consensus threshold) using a simulation based method (1000 simulations) implemented in GEMINI 1.3.0 (Valière, 2002a). Then, using the consensus threshold we estimate the consensus genotype for each sample using GIMLET. The consensus genotype of each sample was compared to the respective reference genotype and the individual identification success was determined. Mean values of success for individual identification for each substrate and bait were after compared between consensus and hair samples.

Chapter 3
Results

3.1. Amount of saliva

The best approximating model ($\Delta AIC_c \leq 2$) explaining the observed amount of saliva after dog handling, only included the substrate; showing the highest weight ($w_i > 0.9$; see Table B1, in appendix B). The substrate significantly affect the amount of saliva ($p < 0.001$), being the major differences recorded between the styrofoam and cork ($p < 0.001$) and styrofoam and wood ($p < 0.001$), while between wood and cork the difference was marginally significant ($p = 0.05$; see Tables B2, B3 and B4, in appendix B). Wood substrate has a higher amount of saliva, while styrofoam presents the lowest values (Figure 2).

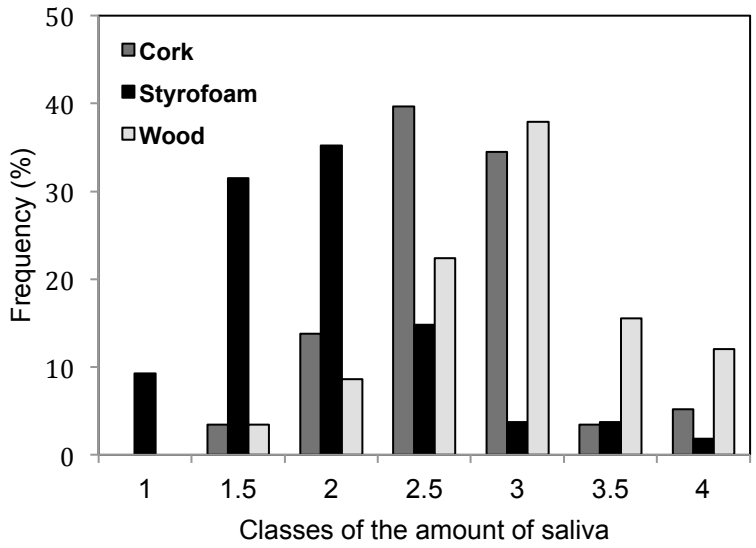

Fig.2 – Frequency of substrates (cork, styrofoam and wood) observed for the seven classes defined for the amount of saliva (1 to 4).

3.2. DNA concentration

No correlation was found between the DNA concentration and the quantity of saliva ($r < 0.2$, $p = 0.99$).

The best model ($\Delta AIC_c \leq 2$) explaining the DNA concentration included substrate, bait and interaction between the bait and substrate ($w_i > 0.8$; see Table B5, in appendix B).

We found that bait type has a significant effect on the DNA concentration ($p < 0.001$), the dog's pâté allowing to obtain more DNA than the sardine (Table 2). Combinations of substrate/bait also appeared to significantly affect DNA concentration ($p < 0.05$), with major differences between wood/sardine and cork/pâté ($p < 0.05$), styrofoam/sardine and styrofoam/pâté ($p < 0.05$), wood/sardine and cork/sardine ($p < 0.05$), and, the most significant between the wood/sardine and wood/pâté ($p < 0.005$), and wood/sardine and styrofoam/pâté ($p < 0.001$). These differences are clearly observed on figure 3, in which, the combination wood/pâté allowed to obtain more DNA, while wood/sardine has the lowest values of DNA concentration. Relatively to substrate, cork presented the highest values for recovering DNA, while wood registered the lowest values (Table 2). However, the effect of substrate on DNA concentration was not statistically significant ($p > 0.05$; see Tables B6, B7 and B8, in appendix B).

Although the concentration of recovered DNA varied among substrates and baits, the average of DNA concentrations obtained was considerably high (mean $_{[DNA]}$ = 11.9 ± 24.9 ng/µl).

Table 2 - DNA concentration (ng/µl) obtained in the different substrates (cork, styrofoam and wood) and baits (dog's pâté and sardine).

	Number of samples (*n*)	**DNA concentration (mean±SD)**
Substrate		
Cork	46	15.6 ± 36.3
Styrofoam	42	13.7 ± 20.5
Wood	46	6.4 ± 8.1
Bait		
Dog's pâté	69	13.3 ± 17.0
Sardine	65	10.4 ± 31.3

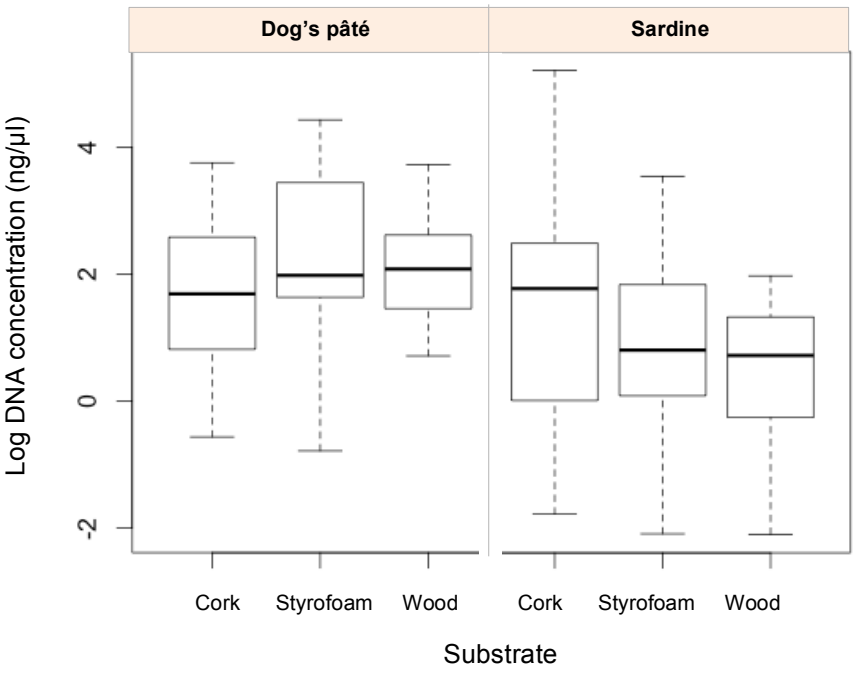

Fig.3 – DNA concentration (ng/µl) (log-transformed) obtained in the different combinations of substrates (cork, styrofoam and wood) with baits (dog's pâté and sardine). Central bars correspond to median values, box corresponds to the superior and inferior quantile containing 50% of data and dashed lines correspond to minimum and maximum values.

3.3. Species Identification

For species identification two models showed $\Delta AIC_c \leq 2$. Those included DNA concentration ($w_i > 0.4$) and, DNA concentration and bait ($w_i > 0.2$; see Table B9, in appendix B). Differences between pairs of substrates do not explained or consistently affected the probability of species identification ($\Delta AIC_c > 2$; see Table B9, in appendix B), although, cork and styrofoam presented higher success rates than wood (Table 3).

The concentration of DNA did not significantly affect the probability of species identification ($p > 0.05$) since even with low DNA concentrations it was possibly to perform a correct identification of the species. In relation to bait type, sardine shows higher values of success rate in identifying species than dog's pâté. However, only slightly differences were observed which did not have significant effect on the probability of species identification ($p > 0.05$; see Table B10 and B11, in appendix B).

3.4. Individual Identification

Concerning individual identification, we identified three models with $\Delta AIC_c \leq 2$, included as predictors DNA concentration ($w_i > 0.3$), DNA concentration and bait ($w_i > 0.2$) and, DNA concentration and substrate ($w_i > 0.1$; see Table B12, in appendix B).

The concentration of DNA appeared to significantly affect the probability of correct individual identification ($p < 0.001$), in which high success rates are positively correlated with greater amounts of DNA, as expected. Substrate and bait types did not affect significantly individual identification ($p > 0.05$; see Table B13 and B14, in appendix B). Although, styrofoam presented the higher success concerning this measure, while cork presented the lowest. Sardine and dog's pâté showed very similar effects on individual identification (Table 3).

Allelic dropout and false alleles rates were practically residual. Relatively to substrate, cork presented the lowest values for both type of genotyping error rates, while styrofoam and wood register the highest values for ADO and FA, respectively (Table 3). However, these differences did not have statistical significance ($p > 0.05$; see Table 15, in appendix B). Differences between baits showed a marginally significance on ADO rates ($p = 0.07$), in which sardine has the highest values (Table 3), and, no significant effects on FA rates ($p > 0.05$; see Table 15 in appendix B), with very similar results for the both type of baits (Table 3).

Table 3 - Frequency of success (%) observed for species identification, individual identification and the percentage of genotyping errors, such as allelic dropout (ADO) and false alleles (FA), using the different types of substrate (cork, styrofoam and wood) and bait (dog's pâté and sardine).

	Number of samples (*n*)	**Frequency of success (%)**		**Genotyping errors (%)**	
		Species ID	Individual ID	ADO	FA
Substrate					
Cork	46	87	78	1.37	0.90
Styrofoam	42	88	91	1.93	1.36
Wood	46	80	82	1.66	1.53
Bait					
Dog's pâté	69	81	83	0.75	1.39
Sardine	65	89	84	2.27	1.12

The number of replicates necessary to achieve individual identification using ten [a] or eight [b] loci gave similar responses, both including as predictors on the best approximating models ($\Delta AIC_c \leq 2$), the effect of DNA concentration ([a] $w_i > 0.5$ and [b] $w_i > 0.4$) and, DNA concentration and bait ([a] and [b] $w_i > 0.2$; see Table B16, in appendix B). Different substrates did not produce significant effects for these variables ($\Delta AIC_c > 2$; see Table B16, in appendix B). In general, all substrates presented an average of two replicates required for accurate individual identification.

As expected, the quantity of DNA appeared to significantly affect the number of replicates necessary to obtain correct individual identifications ($p < 0.05$), for which higher concentrations of DNA are negatively correlated with the number of required replicates. Bait did not affect significantly the number of required replicates ($p > 0.05$; see Table B17 and B18, in appendix B). Sardine and dog's pâté also presented an average of two replicates for individual identification both for 8 or 10 loci.

The comparisons between individual identification success using consensus genotypes (obtained after three replicates) or the best amplification replicate, showed exactly identical results for styrofoam (91%) and cork (78%) and very similar results for wood (83% with consensus versus 82% with the best replicate). The same was observed in the case of baits, for which the success rate was the same either using consensus (dog's pâté: 83% and sardine: 84%) or the best replicate.

Chapter 4
Discussion

In the last few years, many efforts were made to develop DNA-based methodologies to identify species and individuals from non-invasive samples with increasing accuracy. In this study, we developed and evaluated an innovative procedure for collecting saliva as non-invasive source of DNA using a new "swab" technique, based on substrates that canids could lick and bite in order to leave saliva for genetic identification.

The amount of saliva observed after dogs handling showed marked differences between the three types of substrates that were tested, being wood the substrate in which the higher amounts of saliva were quantified, opposite to styrofoam that

presented the lowest values. These results seems to be related to the storage capacity from substrates, since styrofoam and cork are more porous than wood, saliva is quickly absorbed difficult its visualization in substrate surface. It could be expected that the time while dogs spent handling substrates, together with the dog's behaviour related to manipulation (e.g. time inside the mouth or licking), could have influence on the quantity of saliva produced. However, the absence of statistically significant effects of dog's handling time on the amount of saliva observed in substrates did not support that idea. Furthermore, these results are also consistent with the absence of a clear effect of bait type on the amount of saliva, in contrary to the expectation that dogs could have some trend preference for a particular bait. A vast majority of studies using saliva from dogs have been documented in the literature for measuring cortisol, as a predictor of the stress response (Beerda *et al*., 1996; Bennett and Hayssen, 2010; Dreschel and Granger, 2009). Frequently, for collecting saliva in dogs are used swab cottons, although, the amount of saliva obtained can be very low (Beerda *et al.,* 1999; Bergeron *et al.,* 2002; Dreschel and Granger, 2009). To overcome this problem, the utilization of salivary stimulants, such as citric acids, is normally implemented to obtain higher quantities of saliva, as described by Beerda *et al.* (1999), Bennet and Hayssen (2010), Bergeron *et al.* (2002), Coppola *et al.* (2006) and Dreschel and Granger (2009). Increasing the amount of saliva by stimulants has revealed not be necessary in this study, since, the bait appealing seems to be sufficient to ensure high amounts of saliva. Weerth *et al*. (2007) also tested a new device in humans, for collecting saliva for cortisol determination that was normally sampled by swabbing the mouth with a cotton roll. The new device, an eye sponge, also presented many advantages over the traditional methodology, principally by providing high quantities of collected saliva.

The amount of saliva observed was not correlated with the concentration of DNA gathered from the substrates, which is an unexpected result. A possible explanation for it could be that porous substrates were impregnated with saliva that would not be perceivable by eye. This possibility would be consistent with the non-significant results observed for the effect of substrate on DNA concentration, reinforcing the hypothesis that all substrates could absorb similar amounts of saliva and hence similar amounts of DNA. Despite this, some combinations of substrate/bait and bait type showed clear effects on DNA concentration, and sardine seems to decrease the overall amount of DNA regardless of the substrate as well as presents higher variation within each substrate. Evaluation of the results for the use of dog's pâté showed mean values of DNA concentration significantly higher than for sardine and

more discrete differences between substrates as well as less variation within each substrate, reveling more consistent results. These results suggest that i) dogs actually revealed a bait preference (dog's pâté), although we could not perceive it from the amount of saliva observed, or ii) there is an effect of bait composition on the final quality/quantity of DNA. In this case, we could suspect that canned sardines used as bait could have substances in their composition, such as oils, that degraded DNA molecules resulting in a significant lower amount of DNA recovered. Indeed, according to Rogers and Bendich (1994), DNA molecules could suffer irreversible oxidation by phenolic components, which will make it inaccessible for restriction enzymes during DNA extraction. Furthermore, the oils used on the fish processing industries exhibit high concentrations of polyphenols (Mesquita, 2006), which could have interfere with DNA from saliva. The average of DNA concentration obtained, 11.9 ± 24.2 ng/µl, was remarkably high for a non-invasive method for which the average DNA concentration achieved normally ranges from 1 to 5 ng/µl. For instance, i) Morin *et al.* (2001) using feces and hairs from wild chimpanzees to evaluated the DNA concentration, achieved an average of 0.2 and 0.02 ng/µl, respectively, and, for comparison also tested human hairs, obtaining 0.3 ng/µl; ii) when study faecal samples from wild mountain gorillas, Nsubuga *et al*. (2004) got a maximum for DNA concentration of 1.3 ng/µl and, (iii) Ball *et al.* (2007) used faecal samples from swift fox and had an average of 7.5 ng/µl of DNA concentration.

Species identification success was not significantly affected by DNA concentration, contrary to individual identification success that correlated positively with DNA concentration. This result could be expected because species identification was performed using mtDNA which is a high copy number molecule (Allison, 2007; Kovach *et al.,* 2003; Williams and Johnston, 2004). The remarkable amounts of DNA recovered, even for less favorable substrate/bait conditions, allowed a high success on species identification, independently of the PCR inhibitors that non-invasive DNA is prone to have (Beja-Pereira *et al.,* 2009; Gagneux *et al.,* 1997; Taberlet *et al.,* 1999; Taberlet and Luikart, 1999). Additionally, different types of substrate and bait have no significant effects on species identification success, reinforcing the efficiency of our method regardless the type of substrate/bait combination used.

Individual identification success was positively correlated with the concentration of DNA recovered from the substrates, showing that DNA presented a good quality and, in general, the selected substrates and baits did not prevent amplification success that was mainly limited by DNA concentration. Additionally, it is well known

that non-invasive samples are highly susceptible to the quantity and quality of DNA concerning individual identification (Beja-Pereira *et al.,* 2009; Gagneux *et al.,* 1997; Goossens *et al.,* 1998; Smith *et al.,* 2000; Taberlet *et al.,* 1999), but our method enable individual identification with high accuracy in all types of substrate/bait combinations, since the differences observed between substrates and baits have no significant effects on individual identification.

Success rates obtained, both from species identification (mean $_{\text{species ID}}$ = 85%) and individual identification (mean $_{\text{individual ID}}$ = 84%) were remarkably high for non-invasive samples, while genotyping error rates (mean $_{\text{ADO}}$ = 1.6% and mean $_{\text{FA}}$ = 1.3%) were residual in all types of substrate and bait, which highlight the high accuracy of our method. In fact, both theoretical and empirical studies (Krenke *et al.,* 2002; Leclair *et al.,* 2003; Morin *et al.,* 2001; Taberlet and Luikart, 1999) estimated approximately DNA concentrations of 10 ng/µl for avoiding genotyping errors (allelic dropout), which is in accordance with our results. The slight differences observed between substrates do not produced significant effects on genotyping error rates, while differences between baits have marginally significant effects on allelic dropout, being sardine the bait in which the higher percentage was quantified (2.27%), opposite to dog's pâté, that presented the lowest values (0.75%). This result could be explained again by the idea that canned sardines could have substances in their composition that affect the final quantity/quality of DNA, increasing the allelic dropout rate. This possibility would be consistent with the significant results observed for the effect of bait on DNA concentration. Although, because the quantity of DNA obtained was generally high even with the less favorable bait (sardine) this negative result would not be perceivable on individual identification success. Additionally, achieving high individual success does not mean that allelic dropout was not detected in samples. Indeed, it is possibly achieve the complete genotype among the different replicates and, at the same time have genotyping errors among the replicates, or, in opposite, achieve low individual identification success, due to non-amplification, and, the allele dropout be residual. This could explain the different results observed for individual identification success and allelic dropout rate, in substrates.

A wide range of values had been documented in the literature for species and individual identification success, and for genotyping errors. For example, i) using feces from coyotes, Kohn *et al.* (1999) achieved a success of 79% for species identification and 96.6% for individual identification; ii) when study wolf scat samples, Lucchini *et al* (2002) got a success for species identification of 84% while obtaining 53% for individual identification success and a percentage of allelic dropout of 18%;

iii) using hairs to perform species identification in wolves and domestic dogs, Valière *et al*. (2003) achieved a success of 55%, however, when using feces, the success rate increased to 94.2%; iv) Valière and Taberlet (2000) used urine collected in the field from wolf and had success rates of 83.3% and 63.9%, respectively for species identification and individual identification, while reporting an allele dropout of 39.4% and v) Sundqvist *et al*. (2008) used saliva collected around bite wounds on prey to species identification (wolf versus dog) and got a success of 13-54% with an allelic dropout rate of 27-69%.

Regarding the number of replicates necessary to achieve individual identification, the differences between baits and substrates showed no influence on this. As expected, the number of required replicates presented a negative correlation with DNA concentration. In the multi-tubes approach each sample run multiple times for each locus to ensure genotype consistency, which can exhaust DNA extracted and be expensive (Miller *et al.,* 2002; Morin *et al.,* 2001; Valière *et al*., 2003; Vigilant, 2002). For those reasons, major efforts were made to reduce the required number of replicates and at the same time do not compromise genotyping accuracy, however, the quality of majority non-invasive samples makes that task extremely difficult. With our method we achieved an average of two replicates required to identify individuals that corresponds to 100% of amplification success. This number is much smaller than the number of replicates defended by some authors for mainly non-invasive samples (6-10 similar genotypes to define an individual as homozygous or heterozygous with 100% accuracy; Miller *et* al., 2002; Navidi *et al.,* 1992; Taberlet *et al.,* 1996). For example, i) Bellemain and Taberlet (2004) using faecal DNA from brown bears repeated each amplification eight times to perform individual identification; ii) using feces and hairs samples from bobcat, mountain lion, coyote and gray fox, Ruell and Crooks (2007), used three replicates in multi-tubes approach for individual identification and iii) Santini *et al.,* (2007) repeated each amplification four times, using excremental DNA samples from wolf, also to performed individual identification.

The use of hair from captive dogs allowed us to get a trusted source of DNA for each saliva sample, in order to further compare and certify the obtained genotype, which is often impossible in wildlife studies using non-invasive samples. The success of individual identification evaluated through consensus genotypes showed exactly the same results as for the control samples, which definitely encourage the use of this new non-invasive methodology in wild canid studies.

Implications for genetic sampling and future research

Non-invasive genetic approaches have demonstrated an increasing value in many fields, such as molecular ecology and conservation genetics. However, the efficiency of non-invasive studies is still affected by the low quality/quantity of DNA in samples, which may lead to genotyping errors, as allelic dropout and false alleles. New and improved methods need to developed, increasing the efficiency and implementation of non-invasive sampling across different taxa.

In this study, we developed a new non-invasive methodology for species and individual identification in canids that showed remarkable results regardless the type of substrate/bait combinations used. However, despite all the substrate/bait combinations showed high efficiency in all the variables in study, our results and the feasibility of this methodology in the field may suggest that wood/pâté is the more profitable combination for future application in wild. Concerning the choice of bait, substances present in canned sardine may degrade DNA and might have significantly decreased DNA concentration and induced to a higher allelic dropout rate, making dog's pâté a more appealing bait. Although differences between substrates does not have significant effects on the majority of variables studied, based on field collection, wood revealed to be the most robust and thus advantageous substrate since cork and styrofoam have broken a few times during the experience due to dog's handling. Since substrate collection might occur several days after the handling in the wild, broken substrate may disperse throughout the sampling area or be swallowed by the animals.

The main overcome of our method is the remarkably DNA concentration obtained from substrates and the higher success in species and individual identification, with residual genotyping errors. Furthermore, laboratory costs and time are substantially reduced since the high quantity and quality of DNA obtained using this method allow reliable individual genotypes to be obtained from only two PCRs as opposed to multiple independent microsatellite amplifications that are typically required for non-invasive samples with low target DNA. Reduced costs also allow more samples to be analyzed, which could be important in lower budget projects. Additionally, the preparation and transportation of this "saliva-traps" are simpler and inexpensive, and they can be successfully deployed in great quantities over large areas. This "saliva-traps" can be used along with alternative survey methods in wild, as camera trapping, which could be especially useful in studies that aim to correlate genetic parameters with phenotypic features. By using "saliva-traps", cross-contamination is

very unlikely to happen, since during the handling bait is completely consumed, making substrates unattractive to other animals.

In future studies, this methodology should be implemented and evaluated in wild canids, and possibly can be adapted to other taxa (e.g. small mammals). This method has both field and laboratory benefits that could greatly facilitate management and conservation of wild canids.

References

Adams JR, Kelly BT and Waits LP (2003). Using fecal DNA sampling and GIS to monitor hybridization between red wolves (*Canis rufus*) and coyotes *(Canis latrans). Molecular Ecology,* 12: 2175-2186.

Allison LA (2007). Genome Analysis: DNA typing, genomics, and beyond. In: *Fundamental Molecular Biology,* pp. 581-617. Blackwell Publishing, USA.

Avise JC (2004). Molecular Markers, Natural History and Evolution. Sinauer Associates Inc., USA, 2th edition.

Avise JC, Arnold J, Ball RM Jr *et al.* (1987). Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics,* 18: 489-522.

Baayen RH (2011). R package languageR. Available: http://cran.r-project.org/web/packages/languageR/index.htlm. Accessed in 12 of August 2013.

Ball MC, Pither R and Manseau M (2007). Characterization of target nuclear DNA from faeces reduces technical issues associated with the assumptions of low-quality and quantity template. *Conservation genetics,* 8: 577-586.

Banks SC, Skerratt LF and Taylor AC (2002). Female dispersal and relatedness structure in common wombats (*Vombatus ursinus*). *Journal of Zoology,* 256: 389-399.

Barea-Azcón JM, Virgós E, Ballesteros-Duperón E *et al.* (2007). Surveying carnivores at large scales: a comparison of four broad-applied methods. *Biodiversity Conservation,* 16: 1213-1230.

Bates D, Bolker B, Maechler M *et al*. (2013). R package lme4: Linear mixed-effects models using Eigen and S4. Available: http://cran.r-project.org/web/packages/lme4/index.html. Accessed in 12 of August 2013.

Beerda B, Schilder MBH, Bernardina W *et al.* (1999). Chronic stress in dogs subjected to social and spatial restriction: Hormonal and immunological responses. *Physiology and Behavior,* 66: 243-254.

Beerda B, Schilder MBH, Jenseen *et al.* (1996). The use of saliva cortisol and catecholamine measurements for a noninvasive assessment of stress responses in dogs. *Hormones and Behavior,* 30: 272-279.

Beja-Pereira A, Oliveira R, Alves PC *et al.* (2009). Advancing ecological understandings through technological transformations in noninvasive genetics. *Molecular Ecology Resources,* 9: 1279-1301.

Bellemain E and Taberlet P (2004). Improved noninvasive genotyping method: application to brown bear (*Ursus arctos*) faeces. *Molecular Ecology Notes,* 4: 519-522.

Bellemain E, Swenson JE, Tallmon D *et al.* (2005). Estimating population size of elusive animals with DNA from hunter-collected faeces: four methods for brown bear. *Conservation Biology,* 19: 150-161.

Bennett A and Hayssen V (2010). Measuring cortisol in hair and saliva from dogs: coat color and pigment differences. *Domestic Animal Endocrinology*, 39: 171-180.

Bergeron R, Scott SL, Émond J *et al*. (2002). Physiology and behavior of dogs during air transport. *Canadian Journal of Veterinary Research,* 66: 211:216.

Bider JR (1968). Animal activity in uncontrolled terrestrial communities as determined by a sand transect technique. *Ecological Monographs,* 38: 269-308.

Blejwas KM, Williams CL, Shin GT *et al.* (2006). Salivary DNA evidence convicts breeding male coyotes of killing sheep. *Journal of Wildlife Management*, 70: 1087-1093.

Bonin A, Bellemain E, Eidensen PB *et al.* (2004). How to track and assess genotyping errors in population genetic studies. *Molecular Ecology,* 13: 3261-3273.

Brenn M, Jouquand S, Renier C *et al.* (2001). Chromosome specific single-locus FISH probes allow anchorage of an 1800-marker integrated radiation-hybrid/linkage map of the domestic dog genome to all chromosomes. *Genome Research,* 11: 1784-1795.

Broquet T and Petit E (2004). Quantifying genotyping errors in non-invasive population genetics. *Molecular Ecology,* 13: 3601-3608.

Burnham KP and Anderson DR (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods and Research,* 33: 261-304.

Burnham KP, Anderson DR and Laake JK (1980). Estimating of Density from Line Transect Sampling of Biological Populations. *Wildlife Monographs, 72*: 1-202.

Caughley G (1977). Analysis of vertebrate populations. *John Wiley and Sons, Ltd*., New York, USA

Caughley G and Sinclair AR (1994). Wildlife ecology and Management. *Blackwell Science Publications,* Oxford.

Chiappin S, Antonelli G, Gatti R *et al.* (2007). Saliva specimen: A new laboratory tool for diagnostic and basic investigation. *Clinica Chimica Acta,* 383: 30-40.

Clark WR and Andrews RD (1982). Review of population indices applied in furbearer management. In: *Midwest furbearer management,* (eds. G.C. Sanderson), pp. 11-22. Proceedingd of the 43rd Midwest Fish and Wildlife Conference, Wichita, KS, USA.

Clevenger AP and Sawaya MA (2010) Piloting a Non-Invasive Genetic Sampling Method for Evaluating Population-Level Benefits of Wildlife Crossing Structures. *Ecology and Society,* 15: 1-22.

Coppola CL, Grandin T and Enns RM (2006). Human interaction and cortisol: Can human contact reduce stress for shelter dogs? *Physiology and Behavior,* 87: 537-541.

Dreschel NA and Granger DA (2009). Methods of collection for salivary cortisol measurement in dogs. *Hormones and Behavior,* 55: 163-168.

Farrell RK, Leader RW and Johnston SD (2000). Dietary separation of sympatric carniores identified by molecular analysis of scats. *Molecular Ecology*, 9: 1583-1590.

Francisco LV, Langston AA, Mellersh CS *et al.* (1996). A class of highly polymorphic tetra-nucleotide repeats for the canine genetic mapping. *Mammal Genome,* 7: 329-362.

Frantz AC, Schaul M, Pope LC *et al.* (2004). Estimating population size by genotyping remotely plucked hair: the Eurasian badger. *Journal of Applied Ecology,* 41: 985-995.

Fredholm M and Wintero AK (1995). Variation of short tandem repeats within and between species belonging to the *Canidae* family. *Mammal Genome,* 6:11-18.

Fuller TK (1989). Population dynamics of wolves in north-central Minnesota. *Wildlife Monographs*, 105: 1-41.

Gagneux P, Boesch C and Woodruff DS (1997). Microsatellite scoring errors associated with noninvasive genotyping based on nuclear DNA amplified from shed hair. *Molecular Ecology,* 6: 39-48.

Gese EM (2001). Monitoring of terrestrial carnivore populations. In: *Carnivore conservation,* (eds. J.L. Gittleman, S.M. Funk, D.W. Macdonald and R.K. Wayne), pp. 372-396. Cambridge University Press, Cambridge, UK.

Gese EM (2004). Survey and Census Techniques for Canids. In: *Canids: foxes, wolves, jackals, and*

dogs (eds. C. Sillero-Zubiri, M. Hoffman and D.W. Macdonald), pp. 273-279. Wildlife Damage Management Publications, Switzerland and Cambridge, United Kingdom.

Ginsberg JR and Macdonald DW (1990). Foxes, wolves, jackals and dogs: an action plan for the conservation of canids. IUCN, Gland, Switzerland.

Gossens B, Waits LP and Taberlet P (1998). Plucked hair samples as a source of DNA: Reliability of dinucleotide microsatellite genotyping. *Molecular Ecology*, 7: 1237-1241.

Goymann W (2005). Noninvasive monitoring of hormones in bird droppings: physiological validation, sampling, extraction, sex differences, and the influence of diet on hormone metabolite levels. *Annals of the New York Academy of Sciences*, 1046: 35–53.

Harrison RL, Barr DJ and Dragoo JW (2002). A comparison of population survey techniques for swift foxes (*Vulpes velox)* in New Mexico. *American Midland Naturalist,* 148: 320-337.

Hausknecht R, Gula R, Pirga C *et al.* (2006). Urine – a source for noninvasive genetic monitoring in wildlife. *Molecular Ecology Notes,* 10: 1-5.

Hedmark E, Flagstad O, Segerstrom P *et al.* (2004). DNA-based individual and sex identification from wolverine (*Gulo gulo*) faeces and urine. *Conservation Genetics,* 5: 405-410.

Heinemeyer KS, Ulizio TJ and Harrison RL (2008). Natural Sign: Tracks and Scats. In: *Noninvasive Survey Methods for Carnivores*, (eds. R.A. Long, P. MacKay, W.J. Zielinski and J.C. Ray), pp. 45-75. Island Press, Washington, USA.

Hung CM, Li SH and Lee LL (2004). Faecal DNA typing to determine the abundance and spatial organization of otters (*Lutra lutra)* along two strea systems in Kinmen. *Animal Conservation,* 7: 301-311.

Iyengar A, Babu VN, Hedges S *et al.* (2005). Phylogeography, genetic structure and diversity in the dhole (*Cuon alpinus*). *Molecular Ecology,* 14: 2281-2297.

Kendall KC and McKelvey KS (2008). Hair collection. In: *Noninvasive Survey Methods for north American Carnivores* (eds. R.A. Long, P. MacKay, W.J. Zielinski and J.C. Ray), pp. 135-176. Island Press, Washington, USA.

Kohn MH and Wayne RK (1997). Facts from faeces revisited. *Trends in Ecology and Evolution*, 6: 223–227.

Kohn MH, Yorks EC, Kamradt DA *et al.* (1999). Estimating population size by genotyping faeces. *Proceedings of the Royal Society of London,* 266: 657-663.

Kovach AI, Litvaitis MK and Litvaitis JA (2003). Evaluation of fecal mtDNA analysis as a method to determine the geographic distribution of a rare lagomorph. *Wildlife Society Bulletin,* 31: 1061-1065.

Krenke BE, Tereba A, Anderson SJ *et al.* (2002). Validation of a 16-locus fluorescent multiplex system. *Journal of Forensic Science,* 47: 773-785.

Laliberte AS and Ripple WJ (2004). Range contractions of North American carnivores and ungulates. *Biological science*, 54: 123-138.

Leclair B, Sgueglia JB, Wojtowicz PC *et al.* (2003). STR DNA typing: increased sensitivity and efficient sample consumption using reduced PCR reaction volumes. *Journal of Forensic Science,* 48: 1001-1013.

Linhart SB and Knowlton FF (1975). Determining the relative abundance of coyotes by scent station lines. *Wildlife Society Bulletin,* 3: 119-124.

Luchini V, Fabbri E, Marucco F *et al.* (2002). Noninvasive molecular tracking of colonizing wolf (*Canis lupus)* packs in the western Italian Alps. *Molecular Ecology,* 11: 857-868.

Lyra-Jorge MC, Ciocheti G, Pivello VR *et al.* (2008). Comparing methods for sampling large- and medium-sized mammals: camera traps and track plots. *European Journal of Wildlife Research*, 54: 739-744.

Maudet C, Luikart G, Dubray D *et al.* (2004). Low genotyping error rates in wild ungulate faeces sampled in winter. *Molecular Ecology Notes,* 4: 772-775.

Mesquita DPB (2006). Monitorização e Análise de Imagem de um Reactor Descontínuo Sequencial de Lamas Activadas. Tese de Mestrado em Biotecnologia/Engenharia de Bioprocessos. Escola de Engenharia/Universidade do Minho. 122pp.

Miller C, Joyce P and Waits LP (2002). Assessing allelic dropout and genotype reliability using maximum likelihood. *Genetics,* 160: 357-366.

Mills LS, Citta JJ, Lair KP *et al.* (2000). Estimating animal abundance using noninvasive DNA sampling: promise and pitfalls. *Ecological Applications*, 10: 283-294.

Moore M, Sarah B and Benjamin S (2010). Thirty-one short red fox (*Vulpes vulpes*) microssatelite markers. *Molecular Ecology Resources,* 1-10.

Morin PA and Woodruff DS (1992). Paternity exclusion using multiple hypervariable microsatellite loci amplified from nuclear DNA of hair cells. In: *Paternity in primates: genetic tests and theories,* (eds. R.D. Martin, A.F. Dixon and E.J. Wickings), pp.63-81. Karger, Basel, Switzerland.

Morin PA, Chambers KE, Boesh C *et al* (2001). Quantitative polymerase chain reaction analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (*Pan troglodytes verus*). *Molecular Ecology,* 10: 1835-1844.

Murphy MA, Waits LP, Kendall KC *et al.* (2002). An evaluation of long-term preservation methods for brown bear (*Ursus arctos)* faecal samples. *Conservation Genetics,* 3: 347-359.

Navidi W, Arnheim N, Wareman MS (1992). A multiple-tubes approach for accurate genotyping of very small DNA samples by using PCR: statistical considerations. *American Journal of Human Genetics,* 50: 347-359.

Neff MW, Broman KW, Mellersh CS *et al.* (1999). A second-generation genetic linkage map of the domestic dog, *Canis familiaris. Genetics,* 151: 803-820.

Ng DPK, Koh D and Chia ACK (2006). Saliva as a viavle alternative source of human genomic DNA in genetic epidemiology. *Clinica Chimica Acta,* doi:10.1016/j.cca.2005.11.024.

Nsubuga AM, Robbins MM, Roeder AD *et al.* (2004). Factors affecting the amount of genomic DNA extracted from ape faeces and the identification of an improved sample storage method. *Molecular Ecology,* 13: 2089-2094.

Oberbauer AM, Grossman DI, Iron DN *et al.* (2003).The genetics of Epilepsy in the Belgian Tervuren and Sheepdog. *Journal of Heredity,* 94: 57-63.

Ostrander EA, Mapa FA, Yee M *et al.* (1995). One hundred and one simple sequence repeat-based markers for the canine genome. *Mammal Genome,* 6: 192-195.

Palomares F, Godoy JA, Piriz A *et al.* (2002). Faecal genetic analysis to determine the presence and distribution of elusive carnivores: design and feasibility for the Iberian lynx. *Molecular Ecology,* 11: 2171-2182.

Pang BCM and Cheung BKK (2007). Double swab technique for collection touched evidence. *Legal Medicine,* 9: 181-184.

Piggot MP and Taylor AC (2003). Remote collection of animal DNA and its applications in conservation management and understanding the population biology of are and cryptic species. *Wildlife Research,* 30: 1-13.

Pilgrim KL, McKelvey KS, Riddle AE *et al.* (2005). Felid sex identification based on genetic samples. *Molecular Ecology Notes,* 5: 60-61.

Pilot M, Jedrzejewski W, Branicki *et al.* (2006). Ecological factors influence population genetic structure of European grey wolves. *Molecular Ecology*, 15: 4533-4553.

Prugh LR, Ritland CE, Arthur SM *et al.* (2005). Monitoring coyote population dynamics by genotyping faeces. *Molecular Ecology,* 14: 1585-1596.

Queller DC, Strassmann JE and Hughes CR (1993). Microsatellites and kinship. *Trends in Ecology and Evolution*, 8: 285-288.

R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN: 3-900051-07-0, URL: http://www.R-project.org/.

Ray CJ and Zielinski WJ (2008). Track Stations. In: *Noninvasive Survey Methods for Carnivores*, (eds. R.A. Long, P. MacKay, W.J. Zielinski and J.C. Ray), pp. 75-109. Island Press, Washington, USA.

Rogers SO and Bendich AJ (1994). Extraction of total cellular DNA from plants, algae and fungi. In: *Plant molecular biology manual,* (eds. S.B. Gelvin, R.A. Schilperoort), pp. 129. Kluwer Academic Publishers, Dordrecht, Netherlands.

Roon DA, Thomas ME, Kendall K *et al.* (2005). Evaluating mixed samples as a source of error in noninvasive genetic studies using microsatellites. *Molecular Ecology,* 14: 195-201.

Roughton RD and Sweeny MW (1982). Refinements in scent-station methodology for assessing trends in carnivore populations. *Journal of Wildlife Management,* 46: 217-229.

Ruell EW and Crooks KR (2006). Evaluation of Noninvasive Genetic Sampling Methods for Felid and Canid Populations. *Journal of Wildlife Management,* 71: 1690-1694.

Sadlier LMJ, Webbon CC, Baker PJ *et al.* (2004). Methods of monitoring red foxes Vulpes *vulpes* and badgers *Meles meles:* are field signs the answer? *Mammal Review*, 34: 75-98.

Santini A, Lucchini V, Fabbri E *et al.* (2007). Ageing and environmental factors affect PCR success in wolf (*Canis lupus)* excremental DNA samples. *Molecular Ecology Notes*, 10: 1-7.

Schauster ER, Gese EM and Kitchen AM (2002). An evaluation of survey methods for monitoring swift fox abundance. *Wildlife Society Bulletin*, *30:* 464-477.

Schlotterer C (2004). The Evolution of molecular markers – just a matter of fashion? *Nature Reviews Genetics,* 5: 63-69.

Schwartz MK and Monfort SL (2008). Genetic and endocrine tools for carnivore surveys. In: *Noninvasive Survey Methods for Carnivores* (eds. R.A. Long, P. MacKay, W.J. Zielinski and J.C. Ray), pp. 238-250. Island Press, Washington, USA.

Schwartz MK, Luikart G and Waples RS (2007). Genetic monitoring as a promising tool for conservation and management. *Trends in Ecology and Evolution*, 22: 25-33.

Schwartz MK, Pilgrim KL, McKelvey KS *et al.* (2004). Hybridization between Canada lynx and bobcats: genetic results and management implications. *Conservation Genetics,* 5: 349-355.

Silveira L, Jácomo AT and Diniz-Filho JA (2003). Camera trap, line transect census and track surveys: a comparative evaluation. *Biological Conservation,* 114: 351-355.

Slone MA, Sunnucks P, Alpers D *et al.* (2000). Highly reliable genetic identification of individual northern hairy-nosed wombats from single remotely collected hairs: a feasible censusing method. *Molecular Ecology,* 9: 1233-1240.

Smith KL, Alberts SC, Bayes MK *et al.* (2000). Cross-species amplification, non-invasive genotyping,

and non-Mendelian inheritance of human STRs in Savannah baboons. *American Journal of Primatology,* 51: 219-227.

Spiess AN (2013). R package: qpcR: Modelling and analysis of real-time PCR data. Available: http://cran.r-project.org/web/packages/qpcR/index.html. Accessed in 12 of August 2013.

Sundqvist AK, Ellegren H and Vilá C (2008). Wolf or dog? Genetic identification of predators from saliva collected around bite wounds on prey. *Conservation Genetics,* 9: 1275-1279.

Taberlet P and Luikart G (1999). Non-invasive genetic sampling and individual identification. *Biological Journal of the Linnean Society,* 68: 41-55.

Taberlet P, Camarra JJ, Griffin S *et al.* (1997). Noninvasive genetic tracking of the endangered Pyrenean brown bear population. Imolecular Ecology, *6*: 869-876.

Taberlet P, Griffin S, Goosens B *et al.* (1996). Reliable genotyping of the sample with very low DNA quantities using PCR. *Nucleic Acids Research,* 24: 3189-3194.

Taberlet P, Waits LP and Luikart G (1999). Noninvasive genetic sampling: look before you leap. *Trends in Ecology and Evolution,* 14: 323-327.

Taylor AC, Horsup A, Johnson CN *et al.* (1997). Relatedness structure detected by microsatellite analysis and attempted pedigree reconstruction in an endangered marsupial, the northern hairy-nosed wombat, *Lasiorhinus kreffiti. Molecular Ecology,* 6: 9-19.

Triant DA, Pace RM and Stine M (*2004*). Abundance, genetic diversity and conservation of Louisiana black bears (*Ursus americanus luteolus*) as detected through noninvasive sampling. *Conservation Genetics,* 5: 647-659.

Valière N (2002a). Gemini: software for testing the effects of genotyping errors and multitubes approach for individual identification. *Molecular Ecology Notes,* 9: 2150-2152.

Valière N (2002b). Gimlet: a computer program for analyzing genetic individual identification data. *Molecular Ecology Notes*, 2: 377-379.

Valière N and Taberlet P (2000). Urine collected in the field as a source of DNA for species and individual identification. *Molecular Ecology,* 9: 2149-2154.

Valière N, Fumagalli L, Gielly C *et al.* (2003). Long-distance wolf recolonization of France and Switzerland inferred from non-invasive genetic samples over a period of 10 year. *Animal Conservation,* 6: 83-92.

Vigilant L (2002). Technical challenges in the microsatellite genotyping of a wild chimpanzee population. *Evolutionary Anthropology Supplement,* 1: 162-165.

Vilà C, Amorim IR, Leonard JA *et al.* (1999). Mitochondrial DNA phylogeography and population history of the grey wolf *Canis lupis. Molecular Ecology,* 8: 2089-2103.

Vilà C, Savolainen P, Maldomado JE *et al.* (1997). Multiple and Ancient Origins of the Domestic Dog. *Science,* 276: 1687-1689.

Weber JL and May PE (1989). Abundant class of human DNA polymorphism which can be typed using the polymerase chain reaction. *The American Journal of Human Genetics*, 44: 388-396.

Weerth C, Jansen J, Vos MH *et al.* (2007). A new device for collecting saliva for cortisol determination. *Psychoneuroendocrinology, Psychoneuroendocrinology,* 32: 1144-1148.

Williams BK, Nichols JD and Conroy MJ (2002). Analysis and management of animal populations: modeling, estimation and decision making. *Academic Press,* San Diego, USA: 1-1040.

Williams CL and Johnston JJ (2004). Using Genetic Analyses to Identify Predators. *Sheep & Goat Research Journal,* 19: 85-88.

Williams CL, Blejwas K, Johnston JJ *et al.* (2003). A coyote in sheep´s clothing: predator identification

from saliva. *Wildlife Society Bulletin,* 31: 926-932.

Wilson GJ and Delahay RJ (2001). A review of methods to estimated the abundance of terrestrial carnivores using field and observations. *Wildlife Research*, 28: 151-164.

Zielinski WJ, Schlexer FV, Pilgrim KL *et al.* (2006). The efficacy of wire and glue hair snares in identifying mesocarnivores. *Wildlife Society Bulletin,* 34: 1152-1161.

Appendix

Appendix A

This appendix is related to materials and methods.

Fig. A1 - DNA extraction protocol for saliva samples (adapted from QIAamp DNA Micro Kit).

1. Remove ethanol from samples storage recipient and put in incubator at 37°C for 2 days.

2. Wash samples with 15 ml of PBS and agitate for 30 seconds.

3. Transfer PBS solution to a 15 ml tube.

4. Centrifuge the tubes at 4000 rpm for 25 minutes.

5. Discard supernatant fluid.

6. Immediately add 300 µl Buffer ATL (Micro Kit Quiagen) and 20 µl proteinase K and mix by vortexing until pellet dissolve.

7. Place tubes on incubator at 56°C for 2/3 hours.

8. Briefly centrifuge the tubes at 4000rpm

9. Transfer solution obtained in step 8 to a clean 1.5 ml tube.

10. Add 300µl Buffer AL and mix by pulse-vortexing for 15 seconds.

11. Incubate for 10/ 15 minutes at 70°C.

12. Briefly centrifuge the 1.5 ml tube to remove drops from inside the lid.

13. Carefully transfer the entire lysate from step 12 to the QIAamp MinElute column (in a 2 ml collection tube) without wetting the rim. Centrifuge at 8000 rpm for 1 minute. Place the QIAamp MinElute column in a clean 2 ml collection tube, and discard the collection tube containing the flow-through.

14. Carefully add 500 µl Buffer AW1 without wetting the rim. Centrifuge at 8000 rpm for 1 minute. Place the column in a clean 2 ml collection tube and discard the collection tube containing the flow-through.

15. Add 500 µl Buffer AW2 and repeat proceedings described in step 14.

16. Centrifuge at full speed (14 000 rpm) for 3 minutes to dry the membrane completely.

17. Place the column in a clean 1.5 ml microcentrifuge tube and discard the collection tube containing the flow-through. Add 80µl Buffer AE to the center of the membrane and incubate at room temperature for 5 minutes.

18. Centrifuge at full speed for 1 minute.

19. Keep the first elution tubes and place the column in a clean 1.5 ml microcentrifuge and repeat proceedings described in step 17 and 18.

20. Keep the second elution tubes.

Table A1 - List of reagents and respective volume of PCR reaction for the ten loci multiplex and 425-bb fragment of mitochondrial control region.

	Reagents	Volume (µl)
Multiplex (10 loci)	Water	2
	Mastermix (Multiplex PCR Kit)	5
	Mix Primers	1
	DNA	2
Mitochondrial fragment	Water	3.2
	Mastermix (Multiplex PCR Kit)	5
	Primer DL-H 16340 (Fw)	0.4
	Primer Thr-L 15926 (Rv)	0.4
	DNA	2

Table A2 – PCR protocols for the ten loci multiplex and 425-bp fragment of mitochondrial control region.

	Temperature (°C)	Time	#cycles
Multiplex (10 loci)	95	15'	7X (-0.5°C/ciclo)
	95	30"	
	60	45"	
	72	45"	
	95	30"	22x
	57	45"	
	72	45"	
	95	30"	8x
	53	45"	
	72	45"	
	60	30'	
	10	∞	
Mitochondrial fragment	95	15'	1
	95	30"	40
	52	30"	
	72	45"	
	60	30'	1
	10	∞	

Table A3 – Description of the ten dog autosomal microsatellite used, indicating the repeat type, primer concentration (pM/µl), fluorescent dye and reference.

Locus	Repeat Type	Primer concentration (pM/µl)	Dye	Reference
C08.410	di	0.20	VIC	Moore *et al*. (2010)
C08.618	di	0.16	VIC	Moore *et al*. (2010)
C09.474	di	0.28	PET	Neff *et al*. (1999)
C22.763	di	0.36	PET	Oberbauer *et al*. (2003)
CPH5	di	0.16	FAM	Fredholm & Wintero (1995)
CPH2	di	0.20	NED	Fredholm & Wintero, (1995)
CPH9	di	0.16	NED	Fredholm & Wintero (1995)
CXX.459	di	0.16	VIC	Ostrander *et al*. (1995)
FH2161	tetra	0.28	NED	Francisco *et al*. (1996)
REN64E19	di	0.16	FAM	Breen *et al*. (2001)

Appendix B

This appendix is related to results.

Table B1 – Summary of predictive models for the amount of saliva and model selection estimators; AIC_Cvalues; $\Delta i = (AIC_c)i - (AIC_c)_{min}$; Akaike weights. Top models ($\Delta AIC_c \leq 2$) are shown in bold.

Model	AIC	Δi	Wi
Substrate	**-40.289**	**0**	**0.975**
Substrate+Bait	-32.900	7.389	0.024
Subs+bait+DA	-23.442	16.847	0
Substrate+bait+ subs*isco	-21.580	18.709	0
Subs+bait+Exp	-18.679	21.610	0
Subs+bait+DA+Exp	-9.212	31.077	0
Subs+bait+DB	-7.775	32.514	0
Subs+bait+DB+DA	2.885	43.174	0
Subs+bait+DB+Exp	4.168	44.457	0
Subs+bait+DB+DA+Exp	14.585	54.874	0
Null model (intercept only)	21.126	61.415	0
Subs+bait+DB+DA+Exp + subs*isco	25.956	66.245	0
Bait	27.906	68.195	0
Dog's attributes	30.257	70.546	0
Dog's behaviour	33.484	73.773	0
Exposure time	35.157	75.446	0
Subs+bait+DB+subs*isco*MT	62.043	102.332	0
Subs+bait+DB+subs*isco*LT	65.834	106.123	0
Subs+bait+DB+subs*isco*TT	69.079	109.368	0

Table B2 – Results of GLMM analysis to test the effect of substrate on the amount of saliva. Significant variables are shown in bold. (*P<0.05; **P <0.01; ***P<0.001; · p=0.05; ns not significant).

Effects				
	Estimate	Std. Error	t value	Pr(>|t|)
(*Intercept)*	0.927	0.042	21.859	**0.000** ***
subStyr	-0.221	0.033	-6.681	**0.000** ***
subWood	0.101	0.032	3.118	**0.002** **

Table B3 – Results of ANOVA, calculated with "F test" (95% confidence level), to test the effect of substrate on the amount of saliva. Significant variables are shown in bold. (*P<0.05; **P <0.01; ***P<0.001; · p=0.05; ns not significant).

	Df	Deviance Resid.	Df	Resid. Dev	F	Pr(>F)
NULL			169	11.340		
substrate	2	3.093	167	8.247	31.315	**0.000** ***

Table B4 – Results of Tukey Contrasts test (95% confidence level), to test significant differences among the levels of substrate. Significant variables are shown in bold. (*P<0.05; **P <0.01; ***P<0.001; · p=0.05; ns not significant).

Linear Hypotheses:				
	Estimate	Std. Error	z value	Pr(>|z|)
Styr – Cork == 0	-0.221	0.033	-6.681	**< 0.001** ***
Wood – Cork == 0	0.101	0.032	3.118	**0.005** **
Wood – Styr == 0	0.323	0.033	9.735	**< 0.001** ***

Table B5 – Summary of predictive models for DNA concentration and model selection estimators; AIC_Cvalues; $\Delta i = (AIC_c)i - (AIC_c)_{min}$; Akaike weights. Top models ($\Delta AIC_c \leq 2$) are shown in bold.

Model	AICc	Δ*i*	*Wi*
Subtrate+bait+subs*bait	**468.958**	**0**	**0.813**
Substrate+bait	472.945	3.987	0.108
Bait	473.706	4.748	0.076
Substrate	486.701	17.743	0
Null model (intercept only)	487.013	18.055	0

Table B6 – Results of GLM analysis to test the effect of substrate, bait and the interaction between them, on DNA concentration. Significant variables are shown in bold. (*P<0.05; **P <0.01; ***P<0.001; · p=0.05; ns not significant).

Coefficients:				
	Estimate	Std.Error	z value	Pr(>|z|)
(Intercept)	1.656	0.296	5.597	**0.000 *** **
subStyr	0.633	0.414	1.528	0.129
subWood	0.163	0.414	0.394	0.694
baitSard	-0.058	0.414	-0.140	0.889
subStyr*baitSard	-1.364	0.597	-2.286	**0.024 ***
subWood*baitSard	-1.484	0.585	-2.535	**0.013 ***

Table B7 – Results of ANOVA, calculated with "F test" (95% confidence level), to test the effect of substrate, bait and the interaction between them, on DNA concentration. Significant variables are shown in bold. (*P<0.05; **P <0.01; ***P<0.001; · p=0.05; ns not significant).

	Df	Deviance Resid.	Df	Resid. Dev	F	Pr(>F)
NULL			131	300.10		
substrate	2	9.645	129	290.45	2.504	0.086
bait	1	32.679	128	257.77	16.970	**0.000 *** **
substrate:bait	2	15.136	126	242.64	3.923	**0.022***

Table B8 – Results of Tukey Contrasts test (95% confidence level), to test significant differences among the levels of bait and between the substrate/bait combinations. Significant variables are shown in bold. (*P<0.05; **P <0.01; ***P<0.001; · p=0.05; ns not significant).

Bait				
	diff	lwr	upr	p adj
Sardine-pate	-0.994	-1.473	-0.516	**0.000 *** **
Substrate*Bait				
Styr*pate – Cork*pate	0.632	-0.565	1.830	0.647
Wood*pate – Cork*pate	0.163	-1.035	1.361	0.999
Cork*sard – Cork*pate	-0.058	-1.256	1.140	0.100
Styr*sard – Cork*pate	-0.790	-2.048	0.468	0.458
Wood*sard – Cork*pate	-1.379	-2.589	-0.168	**0.016 ***
Wood*pate – Styr*pate	-0.469	-1.654	0.715	0.861
Cork*sard – Styr*pate	-0.690	-1.875	0.494	0.543
Styr*sard – Styr*pate	-1.422	-2.667	-0.177	**0.015 ***
Wood*sard – Styr*pate	-2.011	-3.209	-0.813	**0.000 *** **
Cork*sard – Wood*pate	-0.221	-1.405	0.963	0.994
Styr*sard – Wood*pate	-0.953	-2.198	0.292	0.238
Wood*sard – Wood*pate	-1.542	-2.739	-0.344	**0.004 ** **
Styr*Sard – Cork*sard	-0.732	-1.977	0.513	0.533
Wood*Sard – Cork*sard	-1.321	-2.518	-0.123	**0.022 ***
Wood*Sard – Styr*Sard	-0.589	-1.847	0.669	0.754

Table B9 – Summary of predictive models for species identification success and model selection estimators; AIC_Cvalues; $\Delta i = (AIC_c)i - (AIC_c)_{min}$; Akaike weights. Top models ($\Delta AIC_c \leq 2$) are shown in bold.

Model	AIC	Δ*i*	*Wi*
[DNA]	**107.584**	**0**	**0.454**
Bait+[DNA]	**108.463**	**0.879**	**0.293**
Substrate+[DNA]	110.027	2.443	0.134
Substrate+Bait+[DNA]	111.112	3.528	0.078
Substrate+Bait+[DNA]+subs*isco	114.528	6.944	0.014
Null model (intercept only)	114.939	7.355	0.011
Bait	115.195	7.611	0.010
Substrate	117.763	10.179	0.003
Substrate+Bait	117.920	10.336	0.003

Table B10 – Results of GLM analysis to test the effect of (i) DNA concentration and (ii) DNA concentration and bait, on species identification success. Significant variables are shown in bold. (*P<0.05; **P <0.01; ***P<0.001; ˙p=0.05; ˙ p=0.05; ns not significant).

Model 1: DNA concentration				
Coefficients:				
	Estimate	Std.Error	z value	Pr(>|z|)
(Intercept)	2.196	0.405	5.425	**0.000 *** **
[DNA]	-0.217	0.176	-1.230	0.219

Model 2: DNA concentration and bait				
Coefficients:				
	Estimate	Std.Error	z value	Pr(>|z|)
(Intercept)	1.862	0.505	3.685	**0.0002 *** **
baitSard	0.581	0.558	1.040	0.298
[DNA]	-0.160	0.188	-0.854	0.393

Table B11 – Results of ANOVA, calculated with “Chisq test” (95% confidence level), to test the effect of DNA concentration and bait, on species identification success. Significant variables are shown in bold. (*P<0.05; **P <0.01; ***P<0.001; ˙ p=0.05; ns not significant).

	Df	Deviance Resid.	Df	Resid. Dev	Pr(>Chi)
NULL			131	105.15	
[DNA]	1	1.569	130	103.58	0.210
isco	1	0.915	127	101.11	0.339

Table B12 – Summary of predictive models for Individual identification success and model selection estimators; AIC_Cvalues; $\Delta i = (AIC_c)i - (AIC_c)_{min}$; Akaike weights. Top models ($\Delta AIC_c \leq 2$) are shown in bold.

Model	AICc	Δ*i*	*Wi*
Individual identification			
[DNA]	**105.176**	**0**	**0.106**
Bait+[DNA]	**106.168**	**0.992**	**0.247**
Substrate+[DNA]	**106.975**	**1.799**	**0.165**
Substrate+bait+[DNA]	108.139	2.963	0.092
Substrate+bait+[DNA]+subs:bait	108.209	3.033	0.089
Null model (intercept only)	124.916	19.740	0
Bait	126.963	21.787	0
Substrate	127.209	22.033	0
Substrate+bait	129.3138	24.138	0

Table B13 – Results of GLM analysis to test the effect of (i) DNA concentration, (ii) DNA concentration and bait and (iii) DNA concentration and substrate, on individual identification success. Significant variables are shown in bold. (*P<0.05; **P <0.01; ***P<0.001; · p=0.05; ns not significant).

Model 1: DNA concentration

Coefficients:

	Estimate	Std.Error	z value	Pr(>|z|)
(Intercept)	1.082	0.279	2.886	**0.000** ***
[DNA]	0.561	0.172	3.258	**0.001** **

Model 2: DNA concentration and bait

Coefficients:

	Estimate	**Std.Error**	**z value**	**Pr(>|z|)**
(Intercept)	0.672	0.434	1.550	0.121
[DNA]	**0.647**	**0.191**	**3.395**	**0.000** ***
baitSard	0.657	0.556	1.181	0.238

Model 3: DNA concentration and substrate

Coefficients:

	Estimate	**Std.Error**	**z value**	**Pr(>|z|)**
(Intercept)	0.620	0.431	1.440	0.150
[DNA]	**0.580**	**0.178**	**3.257**	**0.001** **
subStyr	1.030	0.683	1.507	0.132
subWood	0.505	0.584	0.865	0.387

Table B14 – Results of ANOVA, calculated with "Chisq test" (95% confidence level), to test the effect of DNA concentration, substrate and bait, on individual identification success. Significant variables are shown in bold. (*P<0.05; **P <0.01; ***P<0.001; · p=0.05; ns not significant).

	Df	Deviance Resid.	Df	Resid. Dev	Pr(>Chi)
NULL			131	78.406	
[DNA]	1	11.895	130	66.511	**0.000** ***
sub	2	2.476	128	64.035	0.290
isco	1	1.306	127	62.729	0.253

Table B15 – Results of ANOVA (95% confidence level), to test significant differences between the results of ADO and FA from the different types of substrate and bait. Significant variables are shown in bold. (*P<0.05; **P <0.01; ***P<0.001; · p=0.05; ns not significant).

[i] Effect of bait in ADO and FA rates

ADO rate

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
bait	1	0.001	0.001	3.657	0.071 .
Residuals	18	0.005	0.000		

FA rate

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
bait	1	0.000	0.000	0.0831	0.776
Residuals	18	0.008	0.000		

[ii] Effect of substrate in ADO and FA rates

ADO rate

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Sub	2	0.000	0.000	0.170	0.845
Residuals	27	0.012	0.000		

FA rate

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Sub	2	0.000	0.000	0.100	0.905
Residuals	27	0.029	0.001		

Table B16 - Summary of predictive models for the number of replicates necessary to achieve individual identification using (i) eight or (ii) ten loci and model selection estimators; AIC_Cvalues; $\Delta i = (AIC_c)i - (AIC_c)_{min}$; Akaike weights. Top models ($\Delta AIC_c \leq 2$) are shown in bold.

Model	AICc	Δ*i*	*Wi*
[i] using eight loci			
[DNA]	**395.684**	**0**	**0.504**
Bait+[DNA]	**397.146**	**1.462**	**0.243**
Substrate+[DNA]	398.303	2.619	0.136
Substrate+bait+[DNA]	399.757	4.073	0.066
Substrate+bait+[DNA]+ subs*bait	400.260	4.576	0
Null model (intercept only)	414.148	18.464	0
Substrate	415.837	20.153	0
Bait	415.994	20.310	0
Substrate+bait	417.730	22.046	0
[ii] using ten loci			
DNA]	**436.646**	**0**	**0.464**
Bait+[DNA]	**437.767**	**1.121**	**0.265**
Substrate+[DNA]	438.778	2.132	0.160
Substrate+bait+[DNA]	439.917	3.271	0.090
Substrate+bait+[DNA]+subs:bait	442.918	6.272	0.020
Null model (intercept only)	450.881	14.235	0
Substrate	451.586	14.940	0
Bait	452.880	16.234	0
Substrate+bait	453.584	16.938	0

Table B17 – Results of GLM analysis to test the effect of (i) DNA concentration and (ii) DNA concentration and bait, on the number of replicates necessary to achieve individual identification using [i] eight and [ii] ten loci. Significant variables are shown in bold. (*P<0.05; **P <0.01; ***P<0.001; · p=0.05; ns not significant).

[i] using eight loci

Model 1: DNA concentration

Coefficients:

	Estimate	Std.Error	z value	Pr(>|z|)
(Intercept)	0.743	0.080	9.207	**0.0000** ***
[DNA]	-0.151	0.043	-3.559	**0.0003** ***

Model 2: DNA concentration and bait

Coefficients:

	Estimate	Std.Error	z value	Pr(>|z|)
(Intercept)	0.809	0.120	6.738	**0.0000** ***
[DNA]	-0.163	0.046	-3.581	**0.0003** ***
baitSard	-0.104	0.142	-0.733	0.463

[ii] using ten loci

Model 1: DNA concentration

Coefficients:

	Estimate	Std.Error	z value	Pr(>|z|)
(Intercept)	0.904	0.075	12.016	**0.000** ***
[DNA]	-0.104	0.039	-2.706	**0.007** **

Model 2: DNA concentration and bait

Coefficients:

	Estimate	Std.Error	z value	Pr(>|z|)
(Intercept)	0.980	0.109	8.968	**0.000** ***
[DNA]	-0.118	0.041	-2.857	**0.004** **
baitSard	-0.119	0.127	-0.937	0.349

Table B16 – Results of ANOVA, calculated with "Chisq test" (95% confidence level), to test the effect of DNA concentration and bait, on the number of replicates necessary to achieve individual identification using [i] eight and [ii] ten loci. Significant variables are shown in bold. (*P<0.05; **P <0.01; ***P<0.001; · p=0.05; ns not significant).

[i] using eight loci

	Df	Deviance Resid.	Df	Resid. Dev	Pr(>Chi)
NULL			131	97.392	
[DNA]	1	12.421	130	84.971	**0.000** ***
isco	1	0.538	129	84.433	0.463

[ii] using ten loci

	Df	Deviance Resid.	Df	Resid. Dev	Pr(>Chi)
NULL			131	107.929	
[DNA]	1	7.223	130	100.705	**0.007** **
isco	1	0.879	129	99.826	0.348