International Congress Series 1288 (2006) 103–105

www.ics-elsevier.com

Analysis of inter-specific mitochondrial DNA diversity for accurate species identification

Filipe Pereira [a,b,*], José Meirinhos [a], António Amorim [a,b], Luísa Pereira [a]

[a] *Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP), Portugal*
[b] *Faculdade de Ciências, Universidade do Porto, Portugal*

Abstract. We describe a new strategy for species identification in class *Mammalia* based on the use of *s*hort *m*tDNA *i*nformative *r*egions (SMIRs). SMIRs were operatively defined as having high levels of diversity flanked by conservative regions, so that suitable primers for a broad species range could be designed. Seventeen such regions were selected from the 12srRNA, 16srRNA, Cytb, ND1, ND2 and ND5 genes using a sliding window scan of diversity patterns throughout different mtDNA regions. All SMIRs presented high levels of diversity and are flanked by conservative region for which primers, suitable for a broad range of species amplification, were designed. These primers will be particularly useful in the development of a multiplex-PCR of short amplicons (most SMIRs with less than 200 bp) in different mtDNA regions for post-sequencing analysis, avoiding some drawbacks of previous works (only one fragment analysed, usually with more than 350 bp). The discriminatory efficiency of all SMIRs was demonstrated with different "in silico" tests, using available databases, showing that all SMIRs have enough informative resolution for species identification, even for closely related species from the same genus.

Keywords: Species identification; mtDNA; *Mammalia*

1. Introduction

Elucidation of several forensic casework studies relies on the precise identification of the species of origin for a variety of biological materials. With the advent of DNA based techniques, this correct identification has become of primary importance in different fields, such as in criminal investigations, food industry, protection of endangered species, etc. [1,2]. Nevertheless, the correct assignment of the biological samples sent to forensic

* Corresponding author. IPATIMUP, R. Dr. Roberto Frias s/n, 4200-465 Porto, Portugal. Tel.: +351 22 5570700; fax: +351 22 5570799.
E-mail address: fpereira@ipatimup.pt (F. Pereira).

0531-5131/
doi:10.1016/j.ics.2005.09.125

laboratories has frequently proven to be a difficult task due to the high level of degradation and low quality DNA present in many samples as in the case of ancient materials (bone or teeth remains), stomach contents, hair, processed food (dairy products, roasted meat), etc. In this work we attempt to develop a strategy to produce more sensitive and reliable results for species identification based on the determination of short mtDNA informative regions flanked by conservative regions ideal for species-wide amplification.

2. Method

We constructed a large database for the two rRNA and the 13 protein coding mtDNA genes using all available reference sequences for the class *Mammalia* (123 records). After alignment of all sequences, diversity patterns across these genes were assessed using a sliding window method to calculate the nucleotide diversity (π) in short windows of 100 bp overlapped by 1 bp (*DNAsp* ver.4.10.2; [3]). Primers for the conserved regions surrounding SMIRs were designed and annealing temperatures (Table 1) calculated with Primer3 software [4].

Table 1
Description of the 17 SMIRs

Location	Oligonucleotides primers[a] (5′–3′)	*T* (°C)	Size (bp)	No. of haplotypes[b]
12s rRNA	803-CCC CAC GGG ATA CAG CAG	61	58–64	117
	899-CGG TGG CTG GCA CGA AAT T	60		
	881-AAT TTC GTG CCA GCC ACC G	60	165–194	123
	1092-CAT AGT GGG GTA TCT AAT CCC	60		
	1072-GGG ATT AGA TAC CCC ACT ATG	60	65–70	119
	1180-ACC GCC AAG TCC TTT GAG TTT	60		
	1072-GGG ATT AGA TAC CCC ACT ATG	60	166–179	121
	1279-GGT TTG CTG AAG ATG GCG G	60		
	1350-GGG AAG AAA TGG GCT ACA TTT TC	61	103–129	122
	1495-GGG TGA CGG GCG GTG TGT	63		
	1478-ACA CAC CGC CCG TCA CCC	63	42–68	121
	1570-CCT TGT TAC GAC TTG TCT CCT C	62		
16s rRNA	2173-GGC CTA AAA GCA GCC ATC AA	58	287–356	123
	2507-TTT TTG GTA AAC AGG CGG GG	58		
	2488-CCC CGC CTG TTT ACC AAA AAC	61	204–217	121
	2735-CTC CAT AGG GTC TTC TCG TC	60		
	2716-GAC GAG AAG ACC CTA TGG AG	60	67–90	123
	2827-TCC GAG GTC ACC CCA ACC	61		
	2716-GAC GAG AAG ACC CTA TGG AG	60	182–232	123
	2948-GAT TGC GCT GTT ATC CCT AGG	61		
	2810-GGT TGG GGT GAC CTC GGA	61	91–124	123
	2948-GAT TGC GCT GTT ATC CCT AGG	61		
Cyt b	15,149-TGA GGA CAA ATA TCA TTC TGA GG	59	175	123
	15,367-GTT GTT GGA TCC TGT TTC GTG	60		
	15,347-CAC GAA ACA GGA TCC AAC AAC	60	186	123
	15,576-GCA AAT AGG AAG TAT CAT TCT GG	59		
	15,554-CCA GAA TGA TAC TTC CTA TTT GC	59	176	122
	15,773-CTG GTT GGC CTC CAA TTC ATG	61		
ND1	3709-GCA GTA GCC CAA ACA ATC TC	58	122	123
	3874-GGG TTG AGA TAA ATC ATA TTA TGG	58		
ND2	4860-CTA ACA TGA CAA AAA CTA GCA CC	59	133–138	122
	5038-ATT CAT CCT ATG TGG GCG ATT G	60		
ND5	13,396-CAA GAC ATC CGA AAA ATA GGA GG	61	124	123
	13,561-GGG CTC AGG CGT TGG TAT A	60		

[a] Numbered according to the human reference sequence NC_001807.
[b] For 123 reference sequences of the class *Mammalia*.

To test the information content in the SMIRs, all of them from four selected species (*Homo sapiens*, *Balaenoptera musculus*, *Canis familiaris* and *Ovis aries*) were submitted as a query to the *BLASTn* sequence similarity search in the NCBI databases [5]. We also test our SMIRs databases performing Neighbor-joining (NJ) analysis with the *DNADIST* and *NEIGHBOR* ver3.5c programs [6].

3. Results

Inspection of the sliding window nucleotide diversity throughout the mtDNA genes allowed us to identify 17 SMIRs in six mtDNA genes, all of them with high haplotype diversity (Table 1). For the highly conservative regions surrounding SMIRs we designed 17 pairs of primers with annealing temperatures between 58 and 63 °C (Table 1).

In order to verify if these short fragments have enough informative resolution for species identification, the 17 SMIRs from four species were submitted to a BLAST search. For *H. sapiens, O. aries* and *C. familiaris* total homology was observed only intraspecifically, with only one exception in sheep (*O. dalli*) and seven in dog (*C. lupus*). For the *B. musculus* the most similar sequences from closest species always presented at least 2% of differences. When different haplotypes were available in GenBank, for the mentioned species, we constructed NJ trees adding to the 123 sequences dataset the most divergent ones. For the SMIRs trees calculated, the sequence used as test always cluster together with the reference for the same species. Finally, the discriminatory efficiency of the SMIRs was also demonstrated for two pairs of closely related species (*Bos grunniens*/*Bos taurus* and *Ursus americanus*/*Ursus arctos*) that presented differences for all the SMIRs analyzed.

4. Discussion

Information available in any of the 17 SMIRs databases proved to be informative enough for species identification in class *Mammalia*, even for species from the same genus (as shown for *Bos* and *Ursus*). The same result was found even taking into account intra-specific diversity—"blind test" with the most divergent haplotypes found in GenBank. This information will be particularly useful in the development of a multiplex-PCR of short amplicons in different mtDNA regions for post-sequencing analysis (all the pairs of primers with similar annealing temperatures). The main advantages of this system are (i) the use of short amplicons (most SMIRs with less than 200 bp) very useful for old and/or degraded samples—most of the previous studies rely on the amplification of fragments with more than 350 bp [1,2]; (ii) and the use of more than one amplicon that will increase the discriminatory efficiency avoiding the complete absence of results due to the non-amplification of one region.

Acknowledgements

This work was partially supported by a research grant to FP (SFRH/BD/19585/2004) from Fundação para a Ciência e a Tecnologia and IPATIMUP is supported by POCTI, Quadro Comunitário de Apoio III.

References

[1] C. Bellis, et al., A molecular genetic approach for forensic animal species identification, Forensic Sci. Int. 134 (2003) 99–108.
[2] W. Parson, et al., Species identification by means of the cytochrome *b* gene, Int. J. Leg. Med. 144 (2000) 23–28.
[3] J. Rozas, et al., DnaSP, DNA polymorphism analyses by the coalescent and other methods, Bioinformatics 19 (2003) 2496–2497.
[4] http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi.
[5] http://www.ncbi.nlm.nih.gov/BLAST.
[6] (c) Copyright 1986–1993 by Joseph Felsenstein and by the University of Washington.