Nuclear copies of mitochondrial genes: another problem for ancient DNA

Robert-Jan den Tex · Jesus E. Maldonado ·
Richard Thorington · Jennifer A. Leonard

Abstract The application of ancient DNA techniques is subject to many problems caused primarily by low quality and by low quantity of DNA. For these reasons most studies employing ancient DNA rely on the characterization of mitochondrial DNA, which is present in many more copies per cell than nuclear DNA and hence more copies are likely to survive. We used universal and taxon specific mitochondrial primers to amplify DNA from museum specimens, and found many instances where the amplification of nuclear copies of the mitochondrial gene (numts) instead of the targeted mitochondrial fragment had occurred. Furthermore, the likelihood of amplifying numts increased dramatically when universal primers were utilized. Here we suggest that ancient DNA practitioners must consider the possibility that numts can be amplified at higher rates than previously thought. This is another complication for ancient DNA studies, but it also suggests that more extensive inclusion of nuclear markers in ancient DNA studies should be feasible.

R.-J. den Tex · J. A. Leonard (✉)
Department of Evolutionary Biology, Uppsala University,
Norbyvägen 18D, 752 36 Uppsala, Sweden
e-mail: JLeonard@ebd.csic.es

J. E. Maldonado · R. Thorington
Department of Vertebrate Zoology, National Museum of Natural
History Smithsonian Institution, MRC 108, Washington, DC
20013-7012, USA

J. E. Maldonado · J. A. Leonard
Center for Conservation and Evolutionary Genetics, Smithsonian
Conservation Biology Institute, National Zoological Park,
Washington, DC 20008, USA

J. A. Leonard
Estación Biológica de Doñana-CSIC,
Avd. Americo Vespucio s/n, 41092 Sevilla, Spain

Introduction

Nuclear copies of mitochondrial genes (numts) have been reported in many species (Sorenson and Quinn 1998; Bensasson et al. 2000; Williams and Knowlton 2001; Tourmen et al. 2002; Pereira and Baker 2004; Richly and Leister 2004; Kim et al. 2006; Behura 2007), including rodents (Mirol et al. 2000; Triant and DeWoody 2007a, b; Gonzalez-Ittig and Gardenal 2008). These numts may be long (i.e. Kim et al. 2006) and may be numerous (Richly and Leister 2004; Triant and DeWoody 2007b). Amplification of numts can be a problem in phylogenetic and systematic studies which are based on mitochondrial genes when numts are amplified in some taxa and mitochondrial copies in others, because the sequences being compared are not homologous and therefore are unlikely to reveal the evolutionary relationships of the taxa in question (Arctander 1995; Zhang and Hewitt 1996; Triant and DeWoody 2007a).

There are many more copies of the mitochondrial genome than the nuclear genome per cell, and after death DNA degrades with time. For these reasons, most historic and ancient DNA studies have focused on mitochondrial DNA through the amplification of multiple fragments (Wayne et al. 1999; Hofreiter et al. 2001). Numts are not generally considered to be an issue in studies based on historic or ancient DNA, as it is considered less likely for nuclear DNA to survive in a sufficient quantity to amplify.

Here we describe the frequent amplification of numts from historic squirrel material, despite evidence of DNA degradation characteristic of ancient DNA. These results

highlight the need to be aware of the possibility of numts even in degraded or ancient material, but also suggest that the inclusion of nuclear markers in some studies based on degraded and/or historic material may be more feasible than previously envisioned. Although the most common techniques for avoiding numts are not appropriate for use with historic or ancient DNA, using taxon specific primers instead of universal primers substantially reduced the risk of amplifying numts.

Materials and methods

Samples

The mitochondrial cytochrome b (cob) gene was targeted in different species of southeast Asian tree squirrels (sub-family Callosciurini). Squirrel samples included 44 museum specimens ranging in date from 1902 to 1994 (Table 1) as well as 18 fresh muscle tissue samples (den Tex et al. 2010).

Molecular methods

DNA from the fresh tissue samples was extracted by phenol–chloroform extraction and alcohol precipitated (Sambrook et al. 1989). DNA was extracted from historic material with phenol–chloroform and purified and concentrated in YM-30 Centricons (Millipore) as in Leonard et al. (2005) or by DNeasy animal tissue and blood extraction kit (Qiagen, Hilden, Germany) following the animal tissue protocol as in Den Tex et al. (2010). To monitor for contamination, negative extractions were included in every batch of extractions.

The whole cob gene region was amplified from muscle tissue samples in a single amplification in order to minimize the risk of amplifying nuclear copies, and used as a reference to design more squirrel specific primers for use in the museum material (den Tex et al. 2010). Historic squirrel specimens were amplified using 52 different primer sets, some of which were mammal "universal" (n = 23), and some of which were designed for these squirrels, and will be referred to here as "specific" primer sets (n = 29). When possible, the universal primers were designed to exclude human and cow, two ubiquitous contaminants in commercial reagents (Leonard et al. 2007). The fragments amplified from historic material ranged in size from 141 to 586 base pairs (bp) (including primers). Adjacent fragments always overlapped in order to yield comparative sequence.

included 1*9* Gold Buffer (Applied Biosystems, Foster

Table 1 Museum numbers, species and year of collection of the historic museum specimens included in this study

Museum number	Species	Year	Fragments	Numt
USNM 481196	Callosciurus prevostii	1969	18	Yes
USNM 488368	Callosciurus prevostii	1971	19	Yes
USNM 311468	Lariscus insignis	1958	20	Yes
USNM 488597	Lariscus insignis	1971	17	Yes
NHM 1919.11.5.47	Sundasciurus altitudinis[a]	1919	14	No
USNM 271038	Sundasciurus altitudinis[a]	1939	10	No
USNM 271039	Sundasciurus altitudinis[a]	1939	16	Yes
USNM 488400	Sundasciurus brookei[a]	1971	15	Yes
USNM 90570	Sundasciurus brookei[a]	1971	13	Yes
USNM 121628	Sundasciurus fraterculus[a]	1902	7	No
USNM 252339	Sundasciurus fraterculus[a]	1925	3	No
USNM 141032	Sundasciurus hippurus[a]	1904	13	No
USNM 142273	Sundasciurus hippurus[a]	1905	13	Yes
USNM 300957	Sundasciurus hippurus[a]	1953	12	No
USNM 300958	Sundasciurus hippurus[a]	1953	10	No
USNM 311463	Sundasciurus hippurus[a]	1958	14	No
USNM 488406	Sundasciurus hippurus[a]	1968	18	Yes
USNM 488407	Sundasciurus hippurus[a]	1968	15	Yes
USNM 477851	Sundasciurus hoogstraali[a]	1962	21	Yes
USNM 477855	Sundasciurus hoogstraali[a]	1962	13	Yes
USNM 292576	Sundasciurus jentinki[a]	1951	16	Yes
USNM 292578	Sundasciurus jentinki[a]	1951	7	Yes
USNM 477865	Sundasciurus juvencus[a]	1962	12	Yes
USNM 477866	Sundasciurus juvencus[a]	1962	12	Yes
USNM 488447	Sundasciurus lowii[a] [b]		19	Yes
USNM 488451	Sundasciurus lowii[a]	1971	14	Yes
USNM 462199	Sundasciurus mindanensis[a]	1975	18	Yes
	mindanensis[a]			

Table 1 continued

Museum number	Species	Year	Fragments	Numt
USNM 477946	Sundasciurus moellendorffi[a]	1962	19	Yes
USNM 477947	Sundasciurus moellendorffi[a]	1962	11	No
RMNH 38320	Sundasciurus philippinensis[a]	1994	6	No
USNM 239214	Sundasciurus philippinensis[a]	1907	5	No
USNM 477985	Sundasciurus rabori[a]	1962	13	Yes
USNM 477986	Sundasciurus rabori[a]	1962	9	Yes
USNM 458736	Sundasciurus samarensis[a]	1987	14	Yes
RMNH 38321	Sundasciurus steerii[a]	1994	5	No
USNM 477948	Sundasciurus steerii[a]	1962	19	No
USNM 477964	Sundasciurus steerii[a]	1961	15	No
USNM 311333	Sundasciurus tahan[a]	1958	15	Yes
USNM 311334	Sundasciurus tahan[a]	1958	12	Yes
USNM 87085	Sundasciurus tahan[a]	1970	16	No
USNM 198746	Sundasciurus tenuis[a]	1914	2	No
USNM 488458	Sundasciurus tenuis[a]	1969	19	Yes
USNM 488459	Sundasciurus tenuis[a]	1969	13	Yes

Museum abbreviations are: NMH, Natural History Museum, London; RMNH, National Museum of Natural History 'Naturalis', Leiden; USNM, United States National Museum of Natural History, Smithsonian Institution. The number of successfully sequenced different PCR products is indicated in the column 'Fragments'. If any numt was identified for an individual, there is a yes in the column 'Numt', and if no numt was ever identified, there is a no

[a] From den Tex et al. 2010

[b] Date unknown

City, CA, USA), 2.5 mM MgCl$_2$, 0.8 mM dNTPs (0.2 mM each), 1 lM of each primer, 1.25 U AmpliTaq Gold DNA polymerase (Applied Biosystems) and between 10 and 50 ng of DNA. The PCR program started with an initial denaturation step of 95°C for 10 min followed by 36 cycles of 95°C for 30 s, annealing of 50–60°C for 30 s and extension of 72°C for 45 s; with a final extension of 72°C for 10 min. In all cases negative controls were included to identify possible contamination.

The PCR products were checked on a 2% agarose gel stained with ethidium bromide. Successful amplifications were purified in 25 ll reactions containing 21 ll PCR product, 16.8 U of Exonuclease I (New England Biolabs, Ipswich, MA, USA) and 1.68 U of Shrimp Alkaline Phosphatase (USB Corporation, Staufen, Germany) incubated at 37°C for 15 min followed by 80°C for 15 min. Both strands of each PCR product were sequenced with BigDye (Applied Biosystems) according to manufacturer's recommendations with the same primers as used for amplification. Sequencing

reactions were separated and analysed on an automated ABI 3730xl DNA Analyzer (Applied Biosystems).

All sequences were checked and corrected by eye in Sequencher v. 4.8 (Gene Codes). Usable sequences were subject to a BLAST search (blast.ncbi.nlm.nih.gov/Blast) in order to make sure the sequences were rodent in origin, as opposed to any possible human contaminant. If chimeric sequences generated through jumping PCR had been created, they would also be identified at this stage. No human or other contamination or chimeric sequences were identified.

Statistical analyses

Statistical comparisons were carried out to determine if differences in rates of amplification of nuclear versus mitochondrial copies could be attributed to differences in the size of fragments amplified by specific and universal primer sets. The difference between the variance in the two distributions of amplicon sizes was tested with an F-test. Differences between the means were then tested using a parametric t-test. In order to determine if the use of universal or specific primer sets was more likely to yield numts, a contingency table was constructed and the significance was calculated with a Pearson's v^2 test. All statistical tests were carried out using SPSS v18 (IBM).

Numt identification

In order to determine if each amplified and sequenced fragment was nuclear or mitochondrial in origin, the following steps were followed. First, electropherograms were inspected for double signal and clean sequences were compared to public data in GenBank through a BLAST search. If a double signal was identified (two clear peaks at one or at each of a few base pairs), it was considered to be a mix of mitochondrial and nuclear sequences. Heteroplasmy could also cause double signal, but that is rare and many of these double signals were independently identified as nuclear copies using the other criteria. Next, clean sequences were translated. If a frame shift mutation, or a mutation causing an inappropriate stop codon was identified, it was considered a nuclear copy. Last, each fragment was compared to other amplifications of the same and overlapping fragments from the same individual. Differences at the same base pair in different amplifications (both with the same and with different primer sets) were used to indicate that one of the fragments was of nuclear DNA origin. In order to determine which of the fragments originated from nuclear DNA, multiple fragments from the same individual were compared to each other and to amplifications from other individuals of the same and related species.

Results

A total of 97 out of 584 fragments amplified from 44 historic squirrel samples with primers targeting mitochondrial fragments. were identified as nuclear in origin. The range in size of the fragments that were determined to be numts and those that were mitochondrial in origin overlapped completely (Fig. 1).

Variance in amplicon size from universal and specific primer sets was not significantly different ($F_{22,28} = 1.374$, $P = 0.247$). In addition, amplicon size derived from universal and specific primer sets did not differ significantly ($t = 1.828$, $df = 50$, $P = 0.074$). Out of the 23 universal primer sets, 15 amplified at least one nuclear copy (65%). However, out of the 29 specific primer sets, only five amplified at least one nuclear copy (17%; Table 2). Universal primer sets amplified nuclear fragments at a significantly higher rate than specific primer sets ($v^2 = 12.474$, $df = 1$, $P = 0.001$).

Eighty-six of the nuclear copies were identified by checking for sequences with double signal in the

Table 2 Table comparing the characteristics of the universal and specific primer pairs used on the tree squirrels in this study

Primer set	Number	Numts	mtDNA	Length (in bp)[a]
Universal	23	15	8	345 (±132)
Specific	29	5	24	284 (±107)

Number is the number of primer sets in each category, numts is the number of primer sets that ever amplified a numt, mtDNA is the number of primer sets that only amplified mitochondrial DNA, and length is the average length of the fragment amplified with the primer sets in the different categories, with the standard deviation following in parentheses

[a] The fragment lengths are not statistically different, see "Results"

electropherograms, two due to problems in translation and 54 from comparison to other amplifications of the same or overlapping fragments. The numbers add to more than 97 (the total number of numts identified) because many fragments were identified as nuclear in origin by more than one method. At least one numt was amplified in 15 out of the 18 species (83%; Table 1).

Discussion

The existence of nuclear copies of mitochondrial genes or gene fragments has been recognized for many years now (Zhang and Hewitt 1996). These copies, if not properly identified, may lead to confusion in phylogeographic and systematic studies due to the comparison of non-homologous loci (mitochondrial markers in some individuals and nuclear copies others). The problems associated with the inadvertent amplification of numts in systematics and evolutionary biology have been summarized clearly in various reviews (i.e. Zhang and Hewitt 1996; Sorenson and Quinn 1998; Triant and DeWoody 2007a).

Mitochondrial DNA has been the overwhelmingly most popular marker in historic and ancient DNA studies to date. There is a logic behind this- DNA degrades quickly after an organism dies, and there are many more copies of the mitochondrial genome per cell than the nuclear genome. Therefore, it is logical to presume that there is a higher chance that more copies of a particular fragment of the mitochondria than of a nuclear fragment of the same length will survive. Ancient DNA practitioners have extended this line of reasoning to conclude that since nuclear DNA is difficult to extract from suboptimal ancient material, numts are not a problem in ancient DNA studies. Here we show that numts are frequently amplified in historic squirrel specimens. Numts were amplified in almost all specimens for which mtDNA was also amplified. This problem appears not to be limited to historic specimens. Numts have also been inadvertently amplified in much older Pleistocene material (Orlando et al. 2003; Kolokotronis et al. 2007).
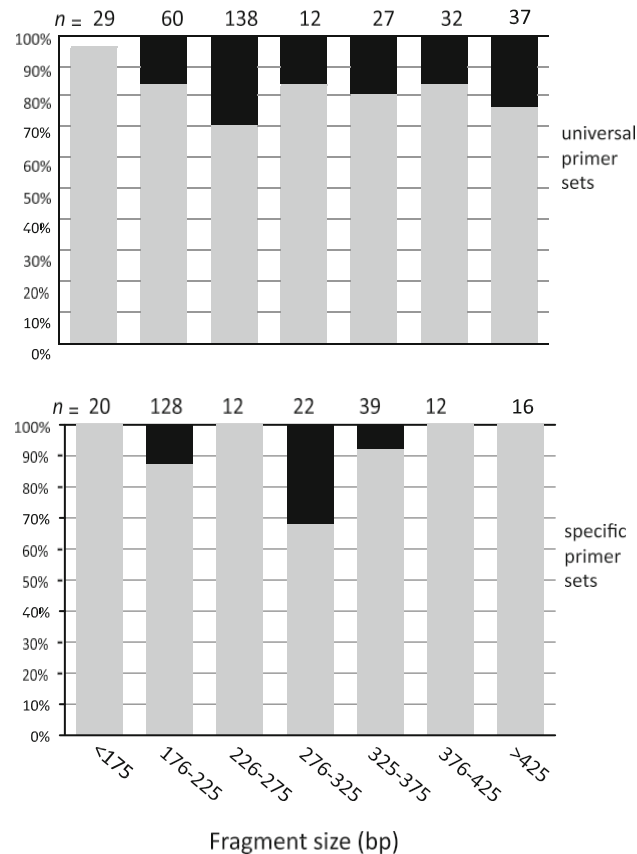

Fig. 1 Graphs showing proportion of mitochondrial (in light gray) and nuclear (in black) copies amplified with universal (top) and specific (bottom) primer sets across the range of amplicon sizes. Absolute number of fragments in size category is marked above each column. Range of fragment size in number of base pairs (bp) is labeled below the columns and is the same for the top and bottom graphs

Together these data suggest that this issue may be more widespread in ancient DNA studies than previously anticipated. Although there are many problems associated with the application of ancient DNA techniques, the accidental amplification of numts has not been on that list (Wayne et al. 1999; Hofreiter et al. 2001). The data presented here suggests that even when working with historic or ancient specimens, it is important to evaluate results critically and consider the possibility that nuclear copies could be amplified.

Once numts have been identified in a project, additional steps need to be taken to ensure that numts and mitochondrial markers are properly categorized. Several methods have been proposed to avoid the amplification of numts or to be able to detect numts once amplified and sequenced (Triant and DeWoody 2007a). These include the isolation of entire mitochondrial genomes, selecting tissues particularly rich in mitochondrial DNA (i.e. muscle), amplification of very long mitochondrial fragments, and amplification from expressed genes (through conversion of mRNA into cDNA). However, these methods are not appropriate for use in historic or ancient material. Of the methods described by Triant and DeWoody (2007a), only two seem reasonable to apply to historic and ancient material: phylogenetic analysis and comparative sequence analysis.

Phylogenetic analysis in order to identify numts involves constructing trees of sequences with known and unknown origin. Numts may have much shorter branch lengths because of the slower mutation rate in nuclear DNA. They may also fall in an 'incorrect' and often basal position in the tree. These results may also be expected from ancient DNA sequences that stopped evolving long before the sequences derived from recent material. Unfortunately, this makes phylogenetic reconstruction a less reliable indicator when applied to data from ancient materials.

The final method suggested by Triant and DeWoody (2007a) is comparative sequence analysis. Comparative sequence analysis depends on the different selective and mutational environments in the nucleus and mitochondrion. Mitochondrial genes that are copied into the nuclear genome are thought not to be functional, so there is no selection against mutations that would cause frame shifts, changes in secondary structure or truncated genes. Selection against changes at second codon positions, which is observed in the mitochondrion, should also be absent in numts. Identification of the pattern of higher mutation rates at third codon positions, followed by first and then second position, can be identified from ancient material. This pattern is also different from the random pattern of apparent mutations that can be a result of degradation of DNA in ancient material (Hofreiter et al. 2001).

Although most of the methods used for identifying and avoiding numts in recent material are not applicable in historic or ancient material, some tools are left. We found a large and significant difference in probability of amplifying numts with universal versus specific primer sets. Many studies use universal primers for a number of good reasons, but we found here that they are much more likely to amplify numts. The use of primers designed specifically from known mtDNA sequences of the target or closely related species significantly reduced the probability of amplifying numts. Since most ancient DNA projects rely on the amplification and sequencing of many overlapping fragments, the careful design of these primers can both help identify and avoid numts. Numts can be avoided by using more specific primers, which preferentially amplify mitochondrial copies of the targeted gene. Good primer design can be used to identify numts by making sure there is sufficient overlap between adjacent fragments. These overlapping fragments can then be used to compare sequences of the same fragment amplified with different primers, greatly increasing the chance of identifying numts if they are present in a dataset.

The data presented here highlight yet another problem that needs to be carefully considered by people working with ancient DNA. However, and perhaps more importantly, these data also highlight the vast potential to incorporate nuclear markers into ancient DNA studies based on material from a wide variety of sources from historic to truly ancient. The incorporation of nuclear markers has become routine and expected in systematic studies based on DNA from optimal samples. The data presented here suggest that many types of ancient material can be incorporated into these studies. Also, more nuclear markers are being characterized in a wide variety of species for a wide variety of reasons. This provides the very important background information necessary to develop appropriate primer sets for ancient DNA, and provides data sets with which to compare results. Single nucleotide polymorphisms (SNPs) are a particularly exciting type of marker for the field of ancient DNA because such small fragments are targeted. In combination with next generation sequencing technology and newly developed PCR techniques where a large number of loci are amplified in a first round, and then individual markers amplified out of the pool in a second PCR round, potentially large numbers of these nuclear markers could be reasonably typed from many ancient DNA sources. Perhaps ancient DNA has a greater potential for looking at genome wide genetic variation than we previously thought.

References

Arctander P (1995) Comparison of a mitochondrial gene and a corresponding nuclear pseudogene. Proc R Soc Lond B 262: 13–19

Behura SK (2007) Analysis of nuclear copies of mitochondrial sequences in honeybee (Apis mellifera) genome. Mol Biol Evol 24:1492–1505

Bensasson D, Zhang D-X, Hewitt GM (2000) Frequent assimilation of mitochondrial DNA by grasshopper nuclear genomes. Mol Biol Evol 17:406–415

den Tex R-J, Thorington R, Maldonado JE, Leonard JA (2010) Speciation dynamics in the SE Asian tropics: putting a time perspective on the phylogeny and biogeography of Sundaland tree squirrels, Sundasciurus. Mol Phylogen Evol 55:711–720

Gonzalez-Ittig RE, Gardenal CN (2008) Co-amplification of mitochondrial pseudogenes in Calomys musculinus (Rodentia, Cricetidae): a source of error in phylogeographic studies. Genome 51:73–78

Hofreiter M, Serre D, Poinar HN, Kuch M, Pääbo S (2001) Ancient DNA. Nat Rev Genet 2:353–359

Kim J-H, Antunes A, Lou S-J, Menninger J, Nash WG, O'Brien SJ, Johnson WE (2006) Evolutionary analysis of a large mtDNA translocation (numt) into the nuclear genome of the Panthera genus species. Gene 366:292–302

Kolokotronis S-O, MacPhee RDE, Greenwood A (2007) Detection of mitochondrial insertions in the nucleus (NuMts) of Pleistocene and modern muskoxen. BMC Evol Biol 7:67

Leonard JA, Vilà C, Wayne RK (2005) Legacy lost: genetic variability and population size of extirpated US gray wolves (Canis lupus). Mol Ecol 14:9–17

Leonard JA, Shanks O, Hofreiter M, Kreuz E, Hodges L, Ream W, Wayne RK, Fleischer RC (2007) Animal DNA in PCR reagents plagues ancient DNA research. J Archaeo Sci 34:1361–1366

Mirol PM, Mascheretti S, Searle JB (2000) Multiple nuclear pseudogenes of mitochondrial cytochrome b in Ctenomys (Caviomorpha, Rodentia) with either great similarity to or high divergence from the true mitochondrial sequence. Heredity 84: 538–547

Orlando L, Leonard JA, Laudet V, Guerin C, Hänni C (2003) Ancient DNA analysis reveals wooly rhino evolutionary relationships. Mol Phylo Evol 28:76–90

Pereira SL, Baker AJ (2004) Low number of mitochondrial pseudogenes in the chicken (Gallus gallus) nuclear genome: implications for molecular inference of population history and phylogenetics. BMC Evol Biol 4:17. doi:10.1186/1471-2148-4-17

Richly E, Leister D (2004) NUMTs in sequenced eukaryotic genomes. Mol Biol Evol 21:1081–1084

Sambrook E, Fritsch F, Maniatis T (1989) Molecular cloning. Cold Spring Harbor Press, Cold Spring Harbor

Sorenson MD, Quinn TW (1998) Numts: a challenge for avian systematics and population biology. Auk 115:214–221

Tourmen Y, Baris O, Dessen P, Jacques C, Malthièry Y, Reynier P (2002) Structure and chromosomal distribution of human mitochondrial pseudogenes. Genomics 80:71–77

Triant DA, DeWoody JA (2007a) The occurrence, detection, and avoidance of mitochondrial DNA translocations in mammalian systematics and phylogeography. J Mamm 88:908–920

Triant DA, DeWoody JA (2007b) Extensive mitochondrial DNA transfer in a rapidly evolving rodent has been mediated by independent insertion events and by duplications. Gene 401: 61–70

Wayne RK, Leonard JA, Cooper A (1999) Full of sound and fury: the recent history of ancient DNA. Annu Rev Ecol Syst 30:457–477

Williams ST, Knowlton N (2001) Mitochondrial pseudogenes are pervasive and often insidious in the snapping shrimp genus Alpheus. Mol Biol Evol 18:1484–1493

Zhang D-X, Hewitt GM (1996) Nuclear integrations: challenges for mitochondrial DNA markers. Trends Evol Ecol 11:247–251