Indian Journal of Animal Sciences **84** (12): 1283–1288, December 2014/Article
https://doi.org/10.56093/ijans.v84i12.45383

Species specific mitochondrial *Cytochrome c oxidase* gene sequence of Manipuri pony

KSH. MIRANDA DEVI [1], MONIKA AHANTHEM[2] and SANKAR KUMAR GHOSH[3]

Assam University, Silchar, Asom 788 011 India

Received: 2 August 2012; Accepted: 24 July 2014

ABSTRACT

The Manipuri pony, a unique indigenous horse breed of India, is known for its fastness, intelligence, surefooted moves and high endurance. The use of DNA barcodes, short DNA sequences from a standardized region of the mitochondrial (mt) genome, has recently been proposed as a tool to facilitate species identification. However, for this emblematic species, there is lacking in the development of DNA barcode which will remain as the molecular tag in the future. A specific molecular identification tag of Manipuri pony was developed under the Accession no. JN228963, and analysis within this family found that the individuals of a single species grouped closely together. Using a set of primer (forward-5´CCAACCACAAAGACATTGGCAC 3´ and reverse- 5´ CTTCTGGGTGGCAA AGAATCA 3´), PCR amplification based on the total genomic DNA extracted from hair samples of Manipuri pony gave an amplification product of 669bp which lies within the barcode region of *COI* gene of the mitochondrial genome. The partial sequence of *COI* gene, which is the DNA barcode of Manipuri pony will remain as the molecular identification mark for this species in the future. Additionally, it will also enhance the conservation of genetic resources of Manipuri pony. *COI* sequence divergence for conspecific individuals of Equidae family was 0.46%, whereas those for congeneric species averaged 6.75% (3.3% to 9.5%). The present finding reaffirmed a very close genetic similarity among the Equidae species. The results showed that analysis based on mt *COI* gene can be useful for explaining the phylogenetic relationships in the family Equidae.

Key words: Biodiversity, *Cytochrome c oxidase*, DNA barcode, Equidae, Manipuri pony

India has a rich biodiversity, which also includes the equine, particularly the horse population in addition to indigenous donkey and wild ass. Six distinctive breeds of horses—Marwari, Kathiawari, Manipuri, Zanskari, Bhutia and Spiti—have been identified in India (National Commission on Agriculture 1976). They are described into 2 main groups; one group represented by the ponies of the Himalayan region, namely Zanskari, Spiti, Bhutia and Manipuri, and the other group represented by breeds adapted to hot arid regions like Marwari and Kathiawari (Behl *et al*. 2007). The Manipuri pony was most likely brought to India by invading tartar tribes. In the past, these ponies were utilized in warfare, long-distance travel, polo, traditional functions, etc., but now it is restricted to polo games and very occasionally in customary functions. According to the Indian Council of Agricultural research (ICAR, Govt. of India) Manipuri pony is "one of the purest breeds of equines of India." Its demand in various areas of the country is increasing being one of the most suitable pack ponies in the hilly terrains. However, the population is decreasing and as per the livestock census (2007) the population of the Manipuri pony has declined to 1,218 against the 1,893 ponies as per the Livestock Census (2003).

Studies were conducted to compare different horse breeds or assessed the genetic structure of single breeds based on microsatellites (Aberle *et al*. 2004, Kakoi *et al*. 2007, Mittmann *et al*. 2010). Analyses of DNA from horses of Chinese tombs from the third century BC (Keyser-Tracqui *et al*. 2005) and from the Bronze Age (Lei *et al*. 2009) showed that the high mtDNA diversity is of ancient origin. A total of over 100 distinct equine mtDNA haplotypes are described in multiple studies focusing on the domestication of horses in general, or on the origin of specific breeds based on D-loop (Lopes *et al*. 2005, Lira *et al*. 2010, Achilli *et al*. 2012). However, there is a great dearth of research based information on this animal in the scientific literature. So, to initiate the research on this animal, the development of DNA barcode as molecular tag is the primary need.

DNA barcoding is a species specific molecular identification technique, and Hebert *et al*. (2003a, b) have shown that the analysis of short, standardized genomic regions (DNA barcodes) can discriminate morphologically recognized animal species. In particular, they suggested that the mitochondrial gene *cytochrome c oxidase* subunit 1 (*COI*) can serve as a uniform target gene for a bio-identification system. It also has gained great attention as a

Present address: [1,2]Research Scholar, [3]Professor (drsankarghosh@gmail.com), Department of Biotechnology.

universal means for the identification of organisms, assigning many specimens to species and in discovering cryptic species (Hebert *et al*. 2003b, Smith *et al*. 2006, Gomez *et al*. 2007). There is an urgent need to characterize this valuable horse to set priorities for its conservation. Characterization at the morphological and genetic levels is the first step towards formulating breeding policies and prioritizing the species for conservation in an effective and meaningful way. Maintaining the viable population of this species is a crucial factor in biodiversity conservation, and this requires the development of species specific molecular tag as DNA barcode. This is the first study that has been carried out to tackle the DNA barcoding of Manipuri pony and also to enhance the conservation of natural resources. Further, the genetic closeness between the different species of Equidae was also detected.

MATERIALS AND METHOD

Sample: Hair samples of Manipuri pony were collected from Imphal West district of Manipur, India. The samples were collected from the mane, withers and croup regions of the animal and while collecting, the hairs were twitched and not combed out so that the hair roots could be intact. They were transferred to sterile sample collection vials containing absolute ethanol sufficient enough to keep the samples immersed under it. The samples were then brought to the laboratory. The subject in our study is Manipuri Pony, which is a semi- domesticated animal in Manipur. While collecting the samples (i.e. only hair), the consent of the owner has been taken.

DNA extraction

Hair roots (50–100) were cut aseptically and suspended in 500µl TES buffer (10 mM Tris.HCl, pH 8.0; 10 mM EDTA, pH 8; 50 mM NaCl) in sterile 1.5 ml micro centrifuge tubes. The hair samples were washed 2–3 times with sterile deionised water by vortex at a maximum speed. After removing the solution, samples were resuspended in a solution containing 500µl TES buffer, 50 µl 10% SDS and 35 µl proteinase K (200µg/µl) and were incubated at 56ºC for 2 h for lysis/digest to occur. After this phenol-chloroform-isoamyl alcohol (25:24:1) extraction method was done twice followed by extractions with chloroform-isoamyl alcohol (24:1) to get the genomic DNA.

PCR amplification and sequencing: The primer pair having the sequence, forward-5´CCAACCACAAAGACAT TGGCAC 3´ and reverse- 5´ CTTCTGGGTGGCCAAAG AATCA 3´ which is available in our laboratory can amplify the barcode region of Manipuri pony. Targeted *COI* sequence was amplified from 10 to 50 ng of genomic DNA in a 25 µl final reaction volume. The reaction mixture contained 0.2mM of each dNTP, 10 pmol of each primer, 0.5 units of high fidelity Taq polymerase enzyme, 1X PCR buffer and 1.5 mM $MgCl_2$. PCR reactions were conducted on a Veriti 96-well thermal cycler with the following conditions: an initial denaturation step of 94ºC (2 min) followed by 30 cycles of 94ºC (1 min), 49ºC (1 min) and 72ºC (1 min) followed by 72ºC for 10 min.

Table 1. Species under Equidae with their GenBank Accession No. from NCBI used in present study for sequence analysis with the partial *COI* gene sequence of Manipuri pony

Accession no.	Species	Authors
EF597513.1	*Equus caballus* (Naqu)	Xu *et al*. (2007)
EF597514.1	*Equus caballus* (Deqin)	Xu *et al*. (2007)
JN228963	*Equus caballus* (Manipuri)	Present study
AY584828	*Equus caballus* (Jeju)	Unpublished
NC_001640	*Equus caballus*	Xu and Arnason (1994)
JN398402.1	*Equus przewalski*	Achilli *et al*. (2012)
JN398403.1	*Equus przewalski*	Achilli *et al*. (2012)
HQ439484.1	*Equus przewalski*	Unpublished
X97337	*Equus asinus*	Xu *et al*. (1996)
NC_001788	*Equus asinus*	Xu *et al*. (1996)
AP012271	*Equus asinus somalicus*	Goto *et al*. (2011)
EF568640	*Equus grevyi*	Unpublished
NC_016061	*Equus hemionus*	Luo *et al*. (2011)
HM118851	*Equus hemionus*	Luo *et al*. (2011)
NC_001779	*Rhinoceros unicornis*	Xu *et al*. (1996)

The size and quality of amplified PCR products were examined on 1.5% agarose gel stained with ethidium bromide (10 mg/ml) and then purified with a commercial kit following manufacturer's instructions and was bidirectionally sequenced in an automated ABI 13500 DNA sequencer.

Data analysis: The sequence electropherogram data for the DNA barcode sequence was edited manually and validated by comparing both strands using Chromas 2.33 software (www.technelysium.com.au). The default parameter was performed to determine the identity and closest known relatives of the partial *COI* gene sequence using BLAST program and GenBank nucleotide database. Open reading frame (ORF) sequence of amino acid is also determined by using the NCBI Sequence analysis, ORF finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). Partial mt *COI* gene sequences of various species under Equidae which are submitted in BOLD (Table 1) were downloaded and were used for sequence analysis using MEGA version 4 (Tamura *et al*. 2007). Sequence divergence was calculated using the Kimura 2 parameter (K2P) model (Kimura 1980) in MEGA.

RESULTS AND DISCUSSION

The mitochondrial *COI* gene of Manipuri pony, which was successfully amplified in the present study was 669 bp long. The sequence corresponded to the position 5401 to 6069 in the whole mt-genome. The ORF of this sequence revealed the absence of stop codon. The nucleotide sequence of partial *COI* gene region was deposited in the NCBI GenBank nucleotide database as the barcode sequence of Manipuri pony using the BankIt web based submission tool (http://www.ncbi.nlm.nih.gov) under the Accession number JN228963. A similarity and homology search using NCBI-BLAST and GenBank database showed closest similarity

Table 2. Nucleotide composition of barcode region in Equidae family

Equus species	A (%)	T (%)	G (%)	C (%)	AT (%)	GC (%)
Equus caballus	25.2	28	17.8	29	53.2	46.8
Equus asinus	24.9	28.5	18	28.7	53.4	46.7
Equus grevyi	25.1	27.7	17.7	29.5	52.8	47.2
Equus hemionus	25.8	29.5	17.2	27.6	55.3	44.8
Average	25.3	28.4	17.6	28.7	53.6	46.4

with the partial *cytochrome c oxidase* gene of "Unspecified Iranian horse" with only one mismatch. For sequence analysis, the aligned sequences were trimmed off from both 5′ and 3′ ends to make dataset of equal aligned length of 617 bp. Analysis within the horse (*Equus caballus*) species found 12 varying sites in which 2 sites specific for Manipuri pony were shared with *Equus caballus* (Deqin). The *COI* gene sequence analysis among the Equidae revealed 81 variable sites, and 61 were found to be parsimony informative. Out of these variable sites, 23 were the species specific sites for *Equus caballus.* The species which are having only one haplotype are represented by that haplotype only and those having more than one are represented by different haplotypes (Fig.1). Within this family, *Equus caballus* and *Equus asinus* nucleotide composition is same as that of the family average, though certain species within

	2	3	6	7	7	7	8	1	1	1	**1**	1	1	1	1	1	1	1	1	1	1	2	2	**2**	2
	2	3	6	7	7	7	8	0	1	1	**2**	3	3	4	5	5	5	6	7	8	8	8	9	0	0	**1**	1
	7	6	0	3	5	8	7	2	1	7	**1**	2	8	1	0	3	6	8	1	3	7	9	2	1	4	**2**	3
E c1	T	C	C	C	A	A	T	C	C	T	**T**	T	T	G	C	T	A	C	A	C	C	G	A	A	T	**C**	T
E c2	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
E c3	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
E c4	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
E c5	.	.	.	.	G	.	.	.	.	.	**C**	.	.	.	.	.	.	.	.	.	.	.	.	.	.	**G**	.
E p	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
E h	C	.	.	.	.	.	C	T	T	C	.	.	C	A	.	.	G	T	G	.	T	A	.	G	.	.	.
E g	C	T	.	.	.	.	C	T	.	C	.	.	C	A	T	.	G	T	G	.	T	A	G	.	.	.	.
E a	C	.	T	.	G	G	.	T	T	.	.	C	C	A	.	C	G	T	G	T	T	A	.	.	C	.	C
E a s	C	.	T	.	.	G	.	T	.	.	.	C	C	A	.	C	G	T	G	T	T	A	.	.	C	.	C
	2	2	2	2	2	2	2	**3**	3	3	3	3	3	3	3	3	3	3	3	3	3	4	4	4	4	4	4
	1	4	4	7	7	7	9	**0**	2	2	3	4	4	5	5	7	7	8	8	9	9	0	0	2	2	4	5
	6	4	9	3	6	9	7	**0**	1	7	9	2	8	4	7	2	5	1	4	6	9	2	5	3	6	1	0
E c1	C	T	T	C	T	C	T	**C**	T	T	T	G	T	A	C	C	T	T	C	G	G	C	G	C	T	T	A
E c2	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.
E c3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.
E c4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	G
E c5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
E p	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.
E h	.	.	.	T	C	.	C	A	C	C	.	A	C	G	.	T	.	.	.	T	A	T	A	T	C	C	.
E g	.	C	C	T	.	T	C	T	C	C	.	A	C	G	.	.	.	.	.	T	A	T	A	.	C	C	.
E a	T	.	.	T	.	.	C	T	.	C	.	A	C	G	T	.	C	C	T	T	A	T	A	T	C	C	.
E a s	T	.	.	T	.	.	C	T	.	C	C	A	C	G	T	.	C	.	T	T	A	T	A	T	C	C	.
	4	4	4	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	6	6	6	6
	5	5	6	6	6	7	8	9	0	2	2	2	3	3	4	4	5	6	7	7	8	8	9	0	0	1	1
	6	9	2	5	8	7	0	5	1	2	5	8	1	4	6	9	5	1	7	9	5	8	7	0	6	2	5
E c1	A	C	A	C	A	C	C	T	T	C	A	C	C	G	A	C	C	T	C	G	T	T	C	C	A	G	T
E c2	G	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.
E c3	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.
E c4	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.
E c5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
E p	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.
E h	.	.	.	.	.	T	.	.	.	T	.	.	.	A	.	T	.	.	.	A	C	C	.	T	.	A	C
E g	.	.	.	T	.	.	T	.	.	T	G	.	T	A	G	T	.	C	.	A	C	C	.	T	G	A	C
E a	.	.	G	.	G	.	T	C	C	.	.	T	T	A	.	T	T	.	T	A	C	C	T	T	G	.	.
E a s	.	.	G	.	G	T	T	C	C	.	.	T	T	A	.	T	T	.	T	A	C	C	.	T	G	.	.

Fig. 1. Variable sites among the species of Equidae. The shaded sites are the species specific sites of *Equus caballus*. The two variation sites which correspond to the amino acid changes are given in bold. Ec1-Ec5 represents the *Equus caballus* Accession No. NC_001640, EF597513, JN228963, EF597514, AY584828; Ep, *Equus przewalski*; Eh, *Equus hemionus*;Eg,*Equus grevyi*; Ea, *Equus asinus*; Ea s, *Equus asinus somalicus*. Numbering is in respect to the nucleotide position from 5428-6044 of the mitochondrial genome of horse.

Table 3. Mean sequence divergence (K2P) within (bold number on diagonal), among (below diagonal) and standard error (above diagonal) of the Equidae species

	Equus caballus	*Equus hemionus*	*Equus grevyi*	*Equus asinus*
Equus caballus	**0.007**	0.011	0.012	0.013
Equus hemionus	0.070	**0.000**	0.007	0.009
Equus grevyi	0.082	0.033	**n/c***	0.010
Equus asinus	0.095	0.059	0.066	**0.007**

this family have wide variation. *Equus grevyi* have a slightly less AT% while the *Equus hemionus* are towards the higher side and vice versa with the GC content (Table 2). The average nucleotide composition across all the species was T=28.4%, A=25.3%, C=28.7% and G=17.6%.

The translation of nucleotide sequences revealed identical sequences of amino acids for all the *Equus* species of this family except the *Equus caballus* (Jeju) having the substitution of 2 amino acids. These variation sites are at the position 41 (phenylalanine to leucine) and 71 (alanine to glycine) of the analyzed amino acid sequence, and these were caused by a transition from T to C at 121 position and a transversion from C to G at 212 position of the nucleotide sequence (Fig. 1).

The pair-wise genetic distance among the 4 species of *Equus* ranges from 3.3% to 9.5% with an average of 6.75% (Table 3). Conspecific individuals of this family showed the sequence divergence in the range of 0.00% to 0.7% with a mean of 0.46%. The phylogenetic relationship between the Equidae species taking *Rhinoceros unicornis* as outgroup are represented in Fig. 2. The NJ tree showed that all intraspecific haplotypes grouped together forming

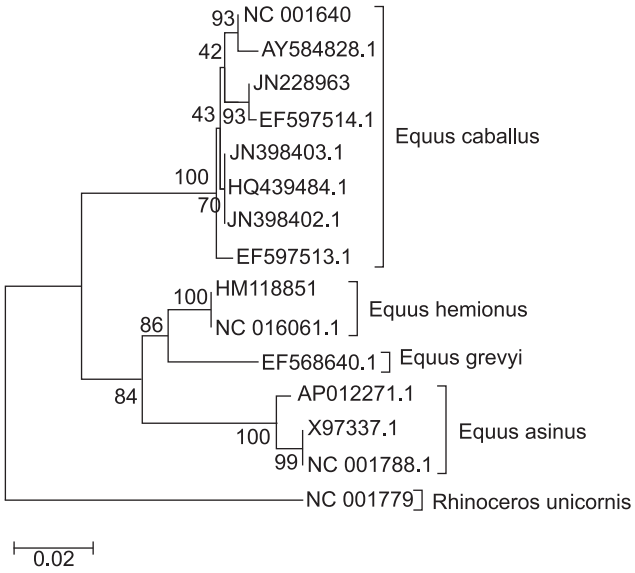

Fig. 2. Neighbour-Joining (NJ) analysis of Equidae family using Kimura 2-parameter (K2P).The analysis is based on 617bp aligned nucleotide position. *Rhinoceros unicornis* was taken as outgroup. JN228963 is the Accession no. of Manipuri pony barcode sequence under study.

cohesive and strongly supported groups. Bootstrap values higher than 40% are indicated.

Manipuri pony barcode sequence based on *COI* gene, demonstrating simplicity and unambiguity was developed. Absence of stop codon in the amplified sequence revealed that nuclear DNA sequences originating from mtDNA sequences (numts) were not sequences since vertebrate nuclear DNA sequences originating from mtDNA sequences are smaller than 600 bp (Ward *et al.* 2005). But sometimes numts had been detected in a small percentage (Clare *et al.* 2011). So, while doing barcoding one should check for the presence of pseudogenes and DNA extraction should be performed in mitochondrion rich tissues.

It was observed that within this family, 75 nucleotide changes occurred at the third codon position, 5 changes at the first, and only 1 at the second codon position. This reflected the fact that synonymous mutation occurred at the third codon position, with a few at the first and least at the second codon position as also observed in most of the species (Ward *et al.* 2005). In this family, only 2 non-synonymous mutations between amino acids were found. This is in consistent with the finding that within the coding region, when the highest proportion of varied sites was found in protein-coding genes, there was an excess of synonymous mutations (Achilli *et al*. 2012). These changes in amino acid were also found when compared with other horse breed sequences (Xu *et al.* 2007). It is possible that this mutation (phenylalanine-leucine; alanine-glycine) will escape the effects of selection because of their similar chemical properties. It is more likely that most of the observed amino acid substitutions have low-impact changes that escape purifying selection, particularly because linkage between genetic loci reduces the effectiveness of purifying selection (Paland and Lynch 2006).

The mean/average genetic distance within the species of Equidae was found much smaller than the mean divergence between congeners. The average intraspecific distance in this family corresponded to the lepidoptera, 0.46% (Hajibabaei *et al.* 2006), but it was lower than the intraspecific barcode variations in other mammals. For example, the intraspecific variation averaged 1.1% in primates (Lorenzen *et al.* 2008), 0.60% in Neotropical bats (Clare *et al.* 2011), and 1% in small mammal communities (Borisenko *et al.* 2008).

From the study of phylogenetic analysis part we can concluded that individuals of a single species grouped closely together, regardless of from where they were collected and the neighbor-joining tree showed shallow intraspecific and deep interspecific divergences. The study of phylogenetic relationship within this family recognized 2 distinct clades as 2 subtree. One clad is formed by the caballines (true horse: *Equus caballus* and *Equus przewalski*) and the other is formed by the non caballines: hemionids (*Equus hemionus*), African wild ass (*Equus asinus*) and zebras (*Equus grevyi*) (Oakenfull and Clegg 1998, Lorenzen *et al.* 2008). The second sub tree has 2 clads one representing the clustering of *Equus grevyi* and *Equus*

hemionus and the other clad with the clustering of conspecific individuals of *Equus asinus* with 89% and 100% bootstrap support respectively. The same topology is also coming from the maximum likelihood method with the K2+I as the best model. The clustering of *Equus hemionus* with *E. grevyi* is concordant with the result of Orlando *et al.* (2009) where Grevyi's zebra nest within paraphyletic hemionids which is in striking contrast with traditionally accepted taxonomy based on coat patterns, behavior, morphology, geographical separation (Oakenfull and Clegg 1998). Further studies with more sequence or more individuals for each case should be performed to clarify the issue.

The present study was to highlight the divergence ability of *COI* sequences in distinguishing closely related species and the intraspecific distances are lower than the interspecific distance which was proven beyond doubt. It could also be applied when traditional methods are inefficacious, as in the identification of eggs and larval forms, and in the analysis of stomach contents or excreta to determine food webs (Vernooy *et al.* 2010). Besides its usefulness in taxonomy, the barcode methodology is expected to be of great utility in conservation biology, for example, when performing biodiversity surveys. Depositing this barcode sequence in public database will make this species identification technique widely accessible. The assembly of a DNA barcode library for mammals will not only aid species recognition, but will also lead to the development of an automated identification system, which would be particularly valuable for law enforcement and allow conservation officials to identify poachers and smugglers. It further reaffirmed the validity of DNA barcoding in species identification. Geographical differentiation does not pose a significant problem in species identification as individuals of a single species cluster together irrespective of their region.

ACKNOWLEDGEMENT

We are grateful to the local Manipuri pony owners and Dr Ranadhir Thiyam of Central Agricultural University (CAU), Imphal, Manipur for helping us in hair sample collection. We would also like to extend our heartiest gratitude towards Department of Biotechnology, Government of India for providing the basic infra-structural facilities and chemicals for the research work.

REFERENCES

Aberle K S, Hamann H, Drogemuller C and Distl O. 2004. Genetic diversity in German draught horse breeds compared with a group of primitive, riding and wild horses by means of microsatellite DNA markers. *Animal Genetics* **35**: 270–77.

Achilli A, Olivieri A, Soares P, Lancioni H, Hooshiar K B, Perego U A, Nergadze S G, Carossa V, Santagostino M, Capomaccio S, Felicetti M, Al-Achkar W, Penedo M C, Verini-Supplizi A, Houshmand M, Woodward S R, Semino O, Silvestrelli M, Giulotto E, Pereira L, Bandelt H J and Torroni A. 2012. Mitochondrial genomes from modern horses reveal the major haplogroups that underwent domestication. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 2449–54.

Behl R, Behl J, Gupta N and Gupta S C. 2007. Genetic relationships of five Indian horse breeds using microsatellite markers. *Animal* **1**: 483–88.

Borisenko A V, Lim B K, Ivanova N V, Hanner R H and Hebert P D. 2008. DNA barcoding in surveys of small mammal communities: a field study in Suriname. *Molecular Ecology Resources* **8**: 471–79.

Clare E L, Lim B K, Fenton M B and Hebert P D. 2011. Neotropical bats: estimating species diversity with DNA barcodes. *PLoS One* **6**: e22648.

Gomez A, Wright P J, Lunt D H, Cancino J M, Carvalho G R and Hughes R N. 2007. Mating trials validate the use of DNA barcoding to reveal cryptic speciation of a marine bryozoan taxon. *Proceedings of the Royal Society - Biological Sciences* **274**: 199–207.

Hajibabaei M, Janzen D H, Burns J M, Hallwachs W and Hebert P D. 2006. DNA barcodes distinguish species of tropical Lepidoptera. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 968–71.

Hebert P D, Ratnasingham S and deWaard J R. 2003a. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society - Biological Sciences* **270** Suppl 1: S96–99.

Hebert P D, Cywinska A, Ball S L and deWaard J R. 2003b. Biological identifications through DNA barcodes. *Proceedings of the Royal society - Biological Sciences* **270**: 313–21.

Kakoi H, Tozaki T and Gawahara H. 2007. Molecular analysis using mitochondrial DNA and microsatellites to infer the formation process of Japanese native horse populations. *Biochemical Genetics* **45**: 375–95.

Keyser-Tracqui C, Blandin-Frappin P, Francfort H P, Ricaut F X, Lepetz S, Crubezy E, Samashev Z and Ludes B. 2005. Mitochondrial DNA analysis of horses recovered from a frozen tomb (Berel site, Kazakhstan, 3rd Century BC). *Animal Genetics* **36**: 203–09.

Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16**: 111–20.

Lei C Z, Su R, Bower M A, Edwards C J, Wang X B, Weining S, Liu L, Xie W M, Li F, Liu R Y, Zhang Y S, Zhang C M and Chen H. 2009. Multiple maternal origins of native modern and ancient horse populations in China. *Animal Genetics* **40**: 933–44.

Lira J, Linderholm A, Olaria C, Brandstrom Durling M, Gilbert M T, Ellegren H, Willerslev E, Liden K, Arsuaga J L and Gotherstrom A. 2010. Ancient DNA reveals traces of Iberian Neolithic and Bronze Age lineages in modern Iberian horses. *Molecular Ecology* **19**: 64–78.

Lopes M S, Mendonca D, Cymbron T, Valera M, da Costa-Ferreira J and Machado Ada C. 2005. The Lusitano horse maternal lineage based on mitochondrial D-loop sequence variation. *Animal Genetics* **36**: 196–202.

Lorenzen E D, Arctander P and Siegismund H R. 2008. High variation and very low differentiation in wide ranging plains zebra (Equus quagga): insights from mtDNA and microsatellites. *Molecular Ecology* **17**: 2812–24.

Mittmann E H, Lampe V, Momke S, Zeitz A and Distl O. 2010. Characterization of a minimal microsatellite set for whole genome scans informative in warmblood and coldblood horse

breeds. *Journal of Heredity* **101**: 246–50.

Oakenfull E A and Clegg J B. 1998. Phylogenetic relationships within the genus Equus and the evolution of alpha and theta globin genes. *Journal of Molecular Evolution* **47**: 772–83.

Orlando L, Metcalf J L, Alberdi M T, Telles-Antunes M, Bonjean D, Otte M, Martin F, Eisenmann V, Mashkour M, Morello F, Prado J L, Salas-Gismondi R, Shockey B J, Wrinn P J, Vasil'ev S K, Ovodov N D, Cherry M I, Hopwood B, Male D, Austin J J, Hanni C and Cooper A. 2009. Revising the recent evolutionary history of equids using ancient DNA. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 21754–59.

Paland S and Lynch M. 2006. Transitions to asexuality result in excess amino acid substitutions. *Science* **311**: 990–92.

Smith M A, Woodley N E, Janzen D H, Hallwachs W and Hebert P D. 2006. DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera: Tachinidae). *Proceedings of the National Academy of Sciences of the United States of America* **103**: 3657–62.

Tamura K, Dudley J, Nei M and Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24**: 1596–99.

Vernooy R, Haribabu E, Muller M R, Vogel J H, Hebert P D, Schindel D E, Shimura J and Singer G A. 2010. Barcoding life to conserve biological diversity: beyond the taxonomic imperative. *PLoS Biology* **8**: e1000417.

Ward R D, Zemlak T S, Innes B H, Last P R and Hebert P D. 2005. DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society B - Biological Sciences* **360**: 1847–57.

Xu S, Luosang J, Hua S, He J, Ciren A, Wang W, Tong X, Liang Y, Wang J and Zheng X. 2007. High altitude adaptation and phylogenetic analysis of Tibetan horse based on the mitochondrial genome. *Journal of Genetics and Genomics* **34**: 720–29.