Indian Journal of Animal Sciences **83** (4): 428–430, April 2013/Short Communication

DNA barcoding of indigenous pig in northeast India

MONIKA AHANTHEM[1], K S H MIRANDA DEVI[2] and SANKAR KUMAR GHOSH[3]

Assam University, Silchar, Asom 788 011 India

Received: 29 September 2012; Accepted: 4 November 2012

Key words: Indigenous pigs, Northeast India, DNA barcoding, Species identification

The North-Eastern region (NE) of India accounted over one-quarter of the pigs in India, i.e. 3.5 million pigs (Anjani *et al.* 2007). Several studies of genetic relationship between different pig breeds based on both nuclear and mitochondrial DNA have been well documented. However, the latter offers several advantages over the former for having rapid evolution, limited exposure to recombination, lack of introns, and high copy number (Ji *et al.* 2011). Accordingly, there has been a considerable focus on the use of the mitochondrial genome as a barcode locus in animals. More recently, the use of partial cytochrome C oxidase I (COI) has been proposed to be a species level marker due to its high interspecific variation, low intraspecific variation, universality of priming site for large taxonomic group (van Velzen *et al.* 2012) and high availability of COI sequences for many species that are publicly accessible from the Barcode of Life Database (BOLD) (Ghosh 2012). The aim of this study was to develop the DNA barcode data of the indigenous pig of NE India using designed primer pair as well as to elucidate the degree of sequence divergence and the phylogenetic relationship among different Suidae species around the world.

The DNA was extracted using standardized phenol-chloroform method (Sambrook and Russell 2001) from tissue and blood samples and PCR amplification was performed using the designed primer pair, i.e. EcoFOR (5' CCTCTGTCCTTAGATTTACAGTC 3') and EcoREV (5' CGTGTATCGACGTCT ATTCC 3'). The reaction mixture contained 50 ng of template DNA, 10 µl of 10×PCR buffer, 10 mM Tris-HCl (pH 8.3) and 50 mM KCl, 0.01% NP–40, 1.5 mM $MgCl_2$, 200 mM of each dNTP, 1 unit Taq polymerase, 5 µl of each primer (20 picomoles) and the remaining ddH_2O to make final volume of 50 µl. Thirty-cycles of amplification were programmed as initial denaturation at 94°C for 2 min, strand denaturation at 94°C for 45 s, annealing at 47°C for 45 s, extension at 72°C for 45 s and a final extension at 72°C for 10 min. After running on 1% agarose gel, DNA fragment was excised and purified with the QIAquick Gel Extraction kit and sequenced bidirectionally using ABI 3500 from Bose Institute, Kolkata. The forward and reverse sequences were read using Seq Scanner and 100% homologous segment of 652 bp was retrieved. The dataset was complemented by 20 additional COI sequences from GenBank database for wide coverage of intraspecific and interspecific diversity (Table 1). The study was limited to only three species as the data of other Suidae species were not available in the database. The comparisons of nucleotide and amino acid sequences were handled with BLAST and FASTA programs. Preliminary multiple alignments were obtained by using the CLUSTALW program (Thompson *et al.* 1994). Nucleotide base and skew were calculated with MEGA 5 (Tamura *et al*. 2011) and Acua (Vetrivel *et al.* 2007) respectively. The phylogenetic tree based on the Neighbor-Joining K2P model was constructed from the developed sequence as well as additional sequences obtained from GenBank using MEGA 5 with 1 000 bootstraps replicates.

Table 1. Partial *COI* sequences used in this study collected from NCBI GenBank

Species name	Scientific name	Accession number
Pig	*Sus scrofa*	AF486864.1, AF486873.1, AY574047.1, AF486859.1, JN245997.1, AF486865.1, EF545586.1, EF545585.1,DQ207755.1,FJ237000.1, FJ237001.1, FJ237002.1, FJ236999.1, FJ237003.1
Common warthog	*Phacochoerus africanus*	NC_008830.1, DQ409327.1
Red river hog	*Potamochoerus porcus*	GQ144632.1, GQ144635.1, GQ144636.1 GQ144638.1, JN632688.1

Present address: [1,2]Research Scholar (lukhoibi@yahoo.co.in, minni.ksh@gmail.com), [3]Professor (drsnkarghosh@gmail.com), Department of Biotechnology, Assam University (Central University), Silchar.

```
FJ237001.1|EWB          CGCTATGTCCG
FJ237002.1|EWB          ...........
FJ237000.1|EWB          ...........
FJ236999.1|EWB          ...........
FJ237003.1|EWB          ..........A
EF545585.1|CWB          TAAA..ACT..
DQ207755.1|KWB          ...AGCACT..
AF486873.1|CIB          ...AGCACT..
AF486864.1|CIB          ...AGCACT..
JN245997.1|NEI          ...AGCACT..
AY574047.1|KWB          ...AGCACT..
AF486859.1|CIB          ...AGCACT..
AF486865.1|CIB          ...AGCACT..
EF545586.1|CWB          ...A.CACTT.
```

*EWB-European wild boar, CWB-Chinese wild boar, KWB-Korean wild boar, CIB-Chinese indigenous breed, NEI-North East India indigenous breed

Fig. 1. Polymorphic sites defining the 14 partial COI regions uncovered in the species pig (*Sus scrofa*). Gray shaded portion indicate synonymous sites found in a particular sequence and yellow shaded portion indicate unique synonymous sites that categorize European and Asian pig (*Sus scrofa*). Polymorphic sites positions are indicated above the synonymous sites with respect to COI gene.

Table 2. Mean base composition of partial *COI* sequences representing 3 species of Suidae family

Species name	No. of samples	T%	C%	A%	G%	AT%	GC%
Sus scrofa	14	29.2	26.7	26.9	17.2	56.1	43.9
Potamochoerus porcus	5	30.9	24.8	26.4	18	57.3	42.7
Phacochoerus africanus	2	29.7	26.1	28.2	15.9	58	42

The sequence homology search showed a high degree of homology (99–100%) to corresponding sequences in GenBank and BOLD indicating that amplified partial COI gene sequence could be used for species identification. The sequence was then deposited in NCBI GenBank database with accession number JN245997.1. The developed sequence and the additional COI sequences downloaded from the GenBank were aligned to yield an equal length of 609 bp with no gaps and no indels. The aligned 609 bp portion of pig sequences possessed 11 variable sites, 598 conserved sites and 6 parsimony informative sites. All the variable sites found in the pig sequences were synonymous. Among these sites, four unique sites were observed that categorize European and Asian pig (Fig. 1). The ranges of nucleotide composition in the 609 bp partial COI region of three Suidae species representing three genera were: A (26.4–28.2), T (29.2–30.9), C (24.8–26.7) and G (15.9–18) (Table 2). This is in agreement to the fact that the frequency of nucleotide G is lowest among the four nucleotide bases in most of the mammalian species (Lin *et al.* 1999). In general, the percentage of average AT content is more compared to GC content while the average GC content of pig was higher than the other two species (Table 3). Such differences were mainly to the first and third codon positions since the second codon position was almost conserved. These differences between the codon positions reflected the divergence could be due to selective constraint (Min and Hickey 2007). The AT skew extended into the negative range but there was no positive value (average –0.025 to –0.073) indicating the occurrence of more (T) s than (A) s (Table 3). This is because the strand asymmetry might have counteracted by strong selection for codons containing T at the second codon position which encode hydrophobic amino acids. The GC skew was strongly negative (average –0.168 to –0.242) indicating heavy bias toward (C) s and against (G) s (Table 3). This is because the nucleotide C is found more frequently on the coding strand while the complementary G is found on the template strand in the mammalian mitochondrial genome (Min and Hickey 2007).

The intraspecific and interspecific genetic distances of three Suidae species representing three genera (Sus, Potamochoerus and Phacochoerus) are shown in Table 4. The mean intraspecific K2P distance was 0.32% (ranged from 0 to 0.62%). Such a low level of intraspecific variation was reported in previous studies (Waugh 2007). The mean interspecific K2P distance was 12.5% (ranged from 9.9% to 14%) which was 39 fold higher than the average intraspecific K2P distance. The least divergence (9.9%) was observed between red river hog and warthog while the highest divergence (14%) was between pig and red river hog. This study further verified that the significance of DNA barcoding based species level differentiation is due to lesser genetic variation within a species than between species (Dasmahapatra and Mallet 2006). A 609 bp COI portion of the studied species was used for the construction of a

Table 3. Average mean composition of GC and AT content of the first, second and third codon positions including average mean of AT and GC skew

Species name	No. of sample	AT1	AT2	AT3	GC1	GC2	GC3	AT skew	GC skew
Sus scrofa	14	45	57	67	55.5	42.9	33.2	–0.041	–0.216
Potamochoerus porcus	5	48	57	67	52.1	42.9	33.2	–0.073	–0.168
Phacochoerus africanus	2	46	57	70	53.7	42.9	29.6	–0.025	–0.242

Table 4. Genetic divergence of partial COI sequences representing 3 species of Suidae family

Pairwise K2P	Range	Mean	Average S.E
Intra-species	0–0.62	0.32	0.12
Inter-species	9.9–14	12.5	1.6

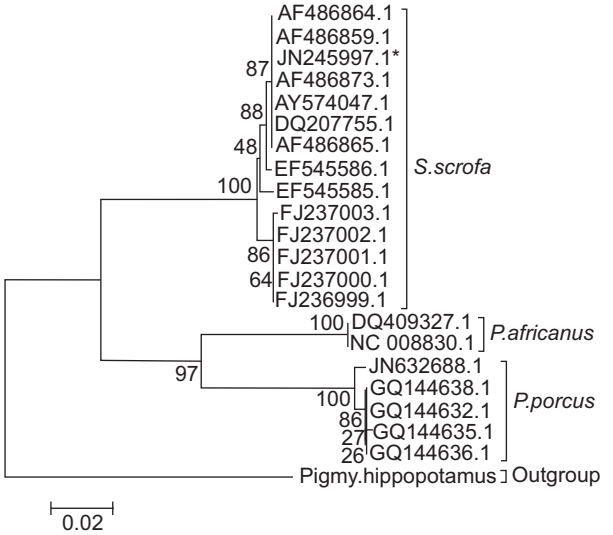

Fig. 2. Neighbour-joining tree of 21 partial COI sequences for 3 species belonging to three genera *Sus, Potamochoerus* and *Phacochoerus*. Bootstrap support values are indicated on the branches. Pigmy hippopotamus (*Hexaproton liberiensis*) is taken as an out-group.

Neighbor-Joining tree (Fig. 2). Each species clustered as a separate assemblage (no individuals were misplaced) with 100% bootstrap support. One distinct clade with high bootstrap values of 97% was observed showing close relationship between red river hog and warthog and the same was also observed by Wu (2006). Both the European and Asian pigs formed 2 clusters with bootstrap support of 86% and 46% respectively that was also in parity with earlier studies (Kim *et al.* 2002). High bootstrap support for each species node as observed in this study as well as in previous studies suggested that Neighbour-Joining analysis of partial COI sequences would be widely effective (Kerr *et al.* 2007). Although this study used only 609 bp of the COI gene, it could still resolve the terminal nodes of the tree quite efficiently. This explains the preeminence of DNA barcoding based on partial COI gene as a successful marker for species level identification of animals (Hajibabaei *et al.* 2006).

SUMMARY

The indigenous pig has been the basis used for pig production for a long period of time in the Northeast India. Here, we developed species-specific COI sequence as DNA barcode sequences from the northeast local pig and submitted to NCBI (Acc. No. JN245997.1). The data set was complemented by 20 additional COI sequences from GeneBank for wide coverage of intra specific and inter specific diversity, and the K2P distance was 0.32% and 12.5% respectively. It reaffirmed that DNA barcode of partial COI gene can be used to accurately discriminate species within the family Suidae as well as other animals. Eco-geographical variations play no rule in barcode mediated identification of *Suidae* species as the barcode sequences are species specific.

REFERENCES

Anjani K, Steven S, Elumalai K and Dhiraj K. 2007. Livestock Sector in North-Eastern Region of India: An appraisal of performance. *Agricultural Economics Research Review* **20**: 255–72.

Dasmahapatra K K and Mallet J. 2006. Taxonomy: DNA barcodes: recent successes and future prospects. *Heredity* **97**(4): 254–5.

Ghosh S K. 2012. *Text book on DNA Barcoding*. Books Space, Kolkata

Hajibabaei M, Janzen D H, Burns J M, Hallwachs W and Hebert P D. 2006. DNA barcodes distinguish species of tropical Lepidoptera. *Proceedings of National Academy of Science* **103**(4): 968–71.

Ji Y Q, Wu D D, Wu G S, Wang G D and Zhang Y P. 2011. Multi-locus analysis reveals a different pattern of genetic diversity for mitochondrial and nuclear DNA between wild and domestic pigs in East Asia. *PLoS One* **6**(10): e26416

Kerr K C, Stoeckle M Y, Dove C J, Weigt L A, Francis C M and Hebert P D. 2007. Comprehensive DNA barcode coverage of North American birds. *Molecular Ecology Notes* **7**(4): 535–43.

Kim K I, Lee J H, Li K, Zhang Y P, Lee S S, Gongora J and Moran C. 2002. Phylogenetic relationships of Asian and European pig breeds determined by mitochondrial DNA D-loop sequence polymorphism. *Animal Genetics* **33**(1): 19–25.

Lin C S, Sun Y L, Liu C Y, Yang P C, Chang L C, Cheng I C, Mao S J and Huang M C. 1999. Complete nucleotide sequence of pig (Sus scrofa) mitochondrial genome and dating evolutionary divergence within Artiodactyla. *Gene* **236**(1): 107–14.

Min X J and Hickey D A. 2007. DNA barcodes provide a quick preview of mitochondrial genome composition. *PloS one* **2**(3): e325.

Sambrook J and Russell D W. 2001. *Molecular Cloning: A Laboratory Manual*. New York,: Cold Spring Harbor Laboratory Press.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M and Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28**(10): 2731–9.

Thompson J D, Higgins D G and Gibson T J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**(22): 4673–80.

van Velzen R, Weitschek E, Felici G and Bakker F T. 2012. DNA barcoding of recently diverged species: relative performance of matching methods. *PLoS One.* **7**(1): e30490

Vetrivel U, Arunkumar V and Dorairaj S. 2007. ACUA: a software tool for automated codon usage analysis. *Bioinformation* **2**(2): 62–3.

Wu G S, Pang P J and Zhang Y P. 2006. Molecular phylogeny and phylogeography of Suidae. *Zoological Research* **27**: 197–201.