

Complete mitochondrial genome sequence and phylogenetic status of Halang pig (*Sus scrofa*)

ABSTRACT

The complete mitochondrial circular genome sequence of indigenous Halang pig was first determined with 16,722 base pairs (bp) in length (GenBank accession number: KY800118). The nucleotide composition had the order A>C>T>G feature, namely of these were calculated to be 34.67%, 26.20%, 25.78% and 13.32% respectively. In this genome's structure, the gene organization was typical of other pigs (*Sus scrofa*) that contain 13 protein-coding genes, 22 tRNA genes, two rRNA genes and one D-loop region with 1,285 bp in size, all of which are arranged similar in other vertebrates. The derived data of this Vietnamese pig's mitochondrial genome (mtDNA) and others from Asian domestic pigs and wild boars were used to phylogeny reconstruction by Bayesian inference and Maximum likelihood methods. These results indicate that the closest evolution relationship between Halang pig and Lantang pig from South China and some domestic and wild pigs from other nearby geographic regions.

Keywords: *mitochondrial genome, phylogeny, Halang pig, Sus scrofa.*

1. INTRODUCTION

The **Halang pigs** are a long-standing Vietnamese native breed in Cao Bang, a northern border province. This pig breed has been recognized and preserved as a local gene source since 2007. The Halang pigs have thin skin, short snout, a saddle cavity on the body, high reproductive performance and high fat rates and much gluttonous. Historically, there have been many studies about pig breeds between northern mountainous Vietnam and China, especially in their genetic relationships [1, 2]. In recent years, well-known domestic pigs such as Meishan, Jinhua and Mong Cai have been established and used as a genetic source to develop pig breeds [1, 3]. In addition, Vietnam and southern China is thought to be one of the points of origin of the earliest domestic pigs [4].

As known, the origin of domestic animals can be studied by analyzing mtDNA. The gene organization of animal mtDNA is simple and conserved, and does not seem to undergo genetic recombination. Moreover, mtDNA is usually maternally monoclonal. Once a mtDNA type is formed in a female, all the descendants of that female carry it, and therefore, the inheritance pattern is clonal through the maternal lineage [1]. The polymorphism of the D-loop sequence in the mtDNA reveals an insight into maternal genetic lineages among species. This is based on high deoxynucleotide substitution rates and rare recombination [5, 6]. The repeat regions have the high variable about respective number so it is the reason why respective region have been often removed in genetic analysis on D-loop [7].

Here, the first complete mitochondrial genome and general structure of the Halang pig breeds is presented by polymerase chain reaction (PCR)- based Sanger sequencing. From these results, **there** were an extra recognition in the regarding genetic relationships of the Halang pigs with other pig breeds. The phylogenetic relationship between the Halang pig and 17 indigenous and wild Asian pig breeds in nearby geographic regions **was** determined by analysis of the polymorphism D-loop region and complete coding region on the whole **mt DNA** sequences. **Analyzing the phylogenetic trees showed that the Halang pigs have a close relationship to originated pigs from South China.** The hypothesis of migration and formation processes of the Halang pigs has been addressed and discussed. This study was undertaken to assist in the future genetic conservation and recovery of this breed.

2. MATERIAL AND METHODS

2.1. **Sampling**

The genealogy information of Halang pig was investigated and provided by National Institute of Animal Sciences (Hanoi, Vietnam) and the local livestock conservation center (Cao Bang, Vietnam). Individuals of Halang pig population were randomly selected for sampling. Genomic DNA was extracted from Halang pig's blood samples by standard phenol-chloroform method described by [8].

2.2. DNA amplification and sequencing

Entire sequences of the mtDNA were amplified by PCR in thermal cycler using the following 30 primers as Table 1. The PCR amplification reaction in 25 μ L total volume consisted of 12.5 μ L GoTaq® Green Master Mix (Promega, Wisconsin, USA), 1.0 μ L DNA template, 0.5 μ L of each primer (10 ppm), and 10.5 μ L deionizer water. The amplify reaction profiles included an initial denaturation at 94°C for 5 min, followed by 25 cycles, each consisting of 30-45 sec denaturation at 94°C, 30 sec primer annealing in range 53 - 55°C (depend on composition of primers), 30 sec extension at 72°C, and then a final 8 min extension at 72°C. The PCR products were evaluated by electrophoresis through 2.0% (wt/vol) agarose gel which was stained with 0.2 μ g/ ml ethidium bromide solution and visualized under UV light. Post amplified DNA was purified utilizing silica-based membrane technology in the form of a spin column by GeneJET™ PCR Purification Kit (Thermo Fisher Scientific, Henderson Road, Singapore).

Sequencing of PCR products was carried out according to Sanger's method [9]. Sequencing reaction volumes of 10 μ l were performed in 96 well plates on the automated ABI 3500 Genetic Analyzer (Applied Biosystems) at Institute of Genome Research (Hanoi, Vietnam).

Table 1. Thirty primer pairs used for PCR

No	Primer's sequence (5'-3')		Annealing T°C
	Forward	Reverse	
D-loop	AGGAGACTAACTCCGCCAT	GCGGATACTTGCATGTGT	54°C
1	ACTAAGTCAATGCCTATTCTG	CAAATGTATGAAACCTCAG	54°C
2	CTACACAATAACCTCCCATA	TGGCACGAGATTTACCAACT	54°C
3	GCTCATAACGCCTTGCTC	ATTCTTTCATCTTTCCCTT	54°C
4	CACAACCATGCAAGAAGAGACA	ACAACCAGCTATCACCAGGC	54°C
5	CCGTAAGGGAAAGATGAAAG	TATGGTTATTTTGA CTGGT	54°C
6	CCGTGCAAAGGTAGCATA	CCAACATCGAGGTCGTAA	55°C
7	TGGGGTGACCTCGGAGTAC	AATATGGCGAAAGGTCCGG	54°C
8	CGAGCAGTAGCCCAAACA	GGTCGTATCGGAATCGTG	55°C
9	GTATCAGGCTTTAACGTAGA	TGGTAATACTGCTGTCATTC	55°C
10	CACAGAAGCAGCCACAAA	ATGGGATAGGGATAAAGT	55°C
11	ACATAGGATGAATGACAGC	TGGTGGAAGTAGTCAGAAAC	55°C
12	GCACTGCCTTGAGCCTAC	GTGTTCAAGTTGCGGTCT	55°C
13	CCCATTATGATTGGGGGTTT	TGCTGTGTATGCGTCAGGAT	55°C
14	CACTTTGTAATCATATTCGTAG	TAGTTGGAAAGGGTAAGC	53°C
15	TTCATCTCACTAACAGCAG	TTGAGTTCGGTTGATTCTG	55°C
16	GCTTCATGCCATTGTAC	TTATAGCGGAATCCTGTG	55°C
17	GCAAGCCCAGAATCAACCG	CGAGGAGGATTGAGGTGTT	55°C
18	ATACCACATAGTAAACCCAA	CCTGTAGCCACAAAGAAA	55°C
19	CTAAACACCTCAATCCTCC	TTGGACGTAATCGGTACCG	55°C
20	CCTTG CAGGGTTACTTAT	TTCGGGTTGTGGTTTCTT	53°C
21	CGGTACCGATTACGTCAA	CCGATTAGATTGATGGATG	55°C

22	ACCAGCTCTATCTGCTTA	GAGGCTTTGATGTTGTTA	55°C
23	ATGATGACTAATAGCAAGCC	GGGATGTAGTCCGAATTG	55°C
24	CATCGGAGACATTGGATT	AGTTGGCTTGAAGTTGAG	55°C
25	CCTACTCCTAGCTGCAGCAG	ATTATGGAGATTACTCGTGG	55°C
26	TCCGCATCATCATTACTA	TTTATGGTGGACTTGGGT	55°C
27	TAATTACCACGAGTAATCTC	TTCTACGAGGTCTGTTCCG	55°C
28	GGAGCATCCATATTCTTT	GGTGTAGTTGTCTGGGTCT	53°C
29	TCGTAGAATGAATCTGAGG	GGTGATACGCATGTTGACTG	55°C

2.3. Data analyses

All used sequences were partitioned into the 4 geographic regions as Mekong Region, Yellow River Valley, South China, and Yangtze River Region that referred in previous study [10]. Entire sequences of the control region and coding region were assembled by overlapping forward and reverse fragment with EditSeq software (DNASTAR Inc., Madison, WI, USA; Hein and Støvlbæk, 1996) and DNADragon v1.6.0 software (SequentiX, Germany). The tandem repeat motifs 5'CGTGCGTACA3' and 5'ACACAAACC3' of the D-loop sequence were removed from the analysis and multiple sequence alignment was performed [11, 12]. Annotation was done using Dogma and MITOS [13, 14]. All annotations were manually verified by BLAST analysis against GenBank [15, 16]. The data of complete mt genome also deposited in NCBI GenBank with accession number KY800118.

2.4. Phylogenetic analysis

To ensure the confidence in origin and evolution analysis, both complete coding and D-loop region sequences were used to construct the two separate phylogenetic trees for Halang pigs and other Asian wild boar and domestic pig breeds. Multiple alignments of D loop region of mtDNA sequence were performed using Cluster W algorithm [17] of MEGA 7 version 6.0.6 [18]. Phylogenetic and molecular evolutionary analyses were conducted using MEGA version 7. The results were converted into a readable FASTA format for finding the best fitted model of each sequence group. The model with lowest Bayesian information criterion is HKY+G. The pair wise distance calculations were estimated by using maximum composite likelihood in MEGA 7. Molecular phylogenetic analysis was performed by the discrete data based on Bayesian Inference method in the Hasegawa-Kishino-Yano model. In the second program, in BEAST v1.8.3 software [19] with Markov Chain Monte Carlo (MCMC) of 10000000 was used to approximate the posterior probabilities of trees (Huelsenbeck and Ronquist, 2001). After that, the best supporting phylogenetic tree was found by Tree Annotator v.1.8.4. Finally, the Figure Tree v1.4.2 software was used to read exporting format file for the phylogenetic tree construction. The tree is rooted using a homologous sequence of Malaysia wild boar (*Sus barbatus*). Bootstrap confidence levels of phylogenetic trees were estimated by 1,000 bootstrap replicates, re-sampling all characters of the control region in each replicate [20]. Genetic distances between Halang pig and other pig breeds were estimated by the Kimura 2-parameter distance matrix using MEGA software version 7 [18].

3. RESULTS AND DISCUSSION

The genome of Halang pig breed is 16,722 bp in its total length that contains 13 protein-coding genes, 22 tRNA genes, two rRNA genes and one control region (D-loop region) between tRNA-*Phe* and tRNA-*Pro* with 1,285 bp in size (Figure 1 and Table 2).



Table 2. Sequence component and location of genes in the mitochondrial genome of the Halang pig

Gene	Codon		Strand	Position		Size (bp)	
	Start	Stop		Anti-codon	Start		Stop
D-loop			H	1	1285	1285	
tRNA <i>Phe</i>			GAA	H	1286	1355	70
12S <i>rRNA</i>			H	1356	2318	963	
tRNA <i>Val</i>			TAC	H	2318	2385	68
16S <i>rRNA</i>			H	2384	3955	1572	
tRNA <i>Leu2</i>			TAA	H	3956	4030	75
ND1	ATG	TAG	H	4033	4989	957	
tRNA <i>Ile</i>			GAT	H	4988	5056	69
tRNA <i>Gln</i>			TTG	L	5054	5126	73
tRNA <i>Met</i>			CAT	H	5128	5197	70
ND2	ATA	TAG	H	5198	6241	1044	
tRNA <i>Trp</i>			TCA	H	6240	6307	68
tRNA <i>Ala</i>			TGC	L	6314	6381	68
tRNA <i>Asn</i>			GTT	L	6383	6457	75

tRNA Cys			GCA	L	6490	6555	66
tRNA Tyr			GTA	L	6556	6620	65
COX1	ATG	TAA		H	6622	8166	1545
tRNA Ser2			TGA	L	8170	8238	69
tRNA Asp			GTC	H	8246	8313	68
COX2	ATG	T--		H	8314	9001	688
tRNA Lys			TTT	H	9002	9068	67
ATPase8	ATG	TAA		H	9070	9273	204
ATPase6	ATG	TAA		H	9231	9911	681
COX3	ATG	T--		H	9911	10694	784
tRNA Gly			TCC	H	10695	10763	69
ND3	ATA	T--		H	10764	11109	346
tRNA Arg			TCG	H	11111	11179	69
ND4l	GTG	TAA		H	11180	11476	297
ND4	ATG	T--		H	11470	12847	1378
tRNA His			GTG	H	12848	12916	69
tRNA Ser1			GCT	H	12917	12975	59
tRNA Leu1			TAG	H	12976	13045	70
ND5	ATA	TAA		H	13046	14866	1821
ND6	ATG	TAA		L	14853	15380	528
tRNA Glu			TTC	L	15378	15446	69
Cytb	ATG	AGA		H	15451	16590	1140
tRNA Thr			TGT	H	16591	16658	68
tRNA Pro			TGG	L	16658	16722	65

Abbreviations: bp: base pairs; rRNA: ribosomal RNA; 16S rRNA: large rRNA subunit; 12S rRNA: small rRNA subunit; tRNA: transfer RNA and italic words are replaced by one amino acid code; ND1-6 and ND4l: genes encoding nicotinamide dinucleotide dehydrogenase subunits 1 to 6 and 4l; ATPase6 and 8: genes encoding adenosine triphosphatase subunits 6 and 8; COX1 to 3: genes encoding cytochrome c oxidase subunits I to III; Cytb: gene encoding cytochrome b. T-- indicates the incomplete termination codon.

The overall base composition is A (34.67%), C (26.20%), T (25.78%) and G (13.32%), in the order A>C>T>G. The location sizes and anti-codon of genes in mitochondrial genome sequence were showed on the Table 2. In Halang pig's whole mt **gemone**, all of genes are arranged similar as in other *Sus scrofa* and vertebrates. The tandem repeat motif (5'- CGTGCGTACA- 3') has 24 repeat sequences, that is much higher than other pig breeds and also to be a unusual character in mitochondrial **gemone**. The more repeat motif numbers are the more chances for hairpin structure leading to errors in copying process. It is thought to be some effect to phylogenetic relationships and evolution of breeds [22].

On both control region and complete coding region trees, Halang pig was reconstructed to be the breed which was most related to the Lantang pig. It is demonstrated by that they were each other's sister taxon and had a shortest genetic distance (0.001) (Figure 2 and 3). **There are two major claded to be showed as two main geographic regions that are Mekong region and Chinese regions.** Obviously, Halang pig was clearly clustered into the Asian clade whereas had significant differences with other pig groups such as Mekong region and some Chinese regions. Although two wild boar (WB) breeds of WB-Jiangxi and WB-Fujian were included to a sub-clade, they were closed with Halang and Langtan pig breeds.

154

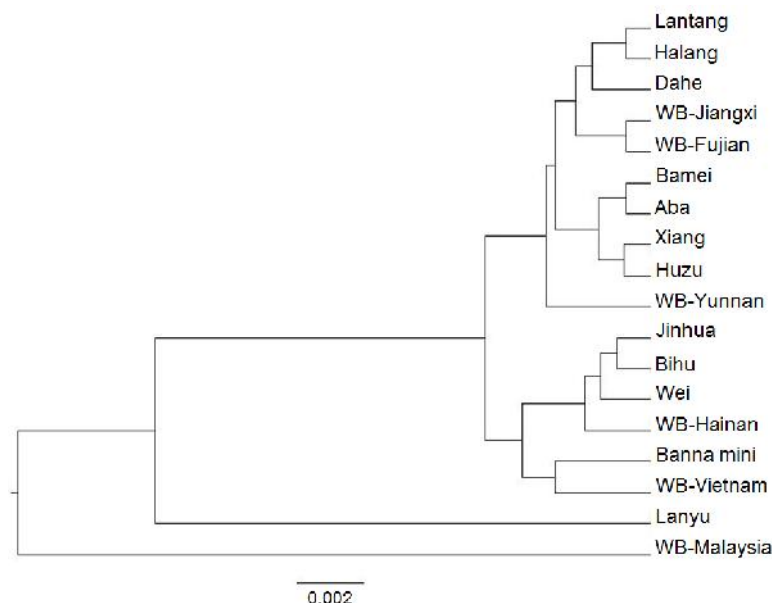


Fig. 2. The phylogenetic relationship was analyzed using discrete data based method (Bayesian Inference) by BEAST v1.8.3 software [19]. The phylogenetic tree was reconstructed by comparison of control region sequences of mitogenome of Halang pig and 17 pig breeds by Tree Annotator software v.1.8.4.

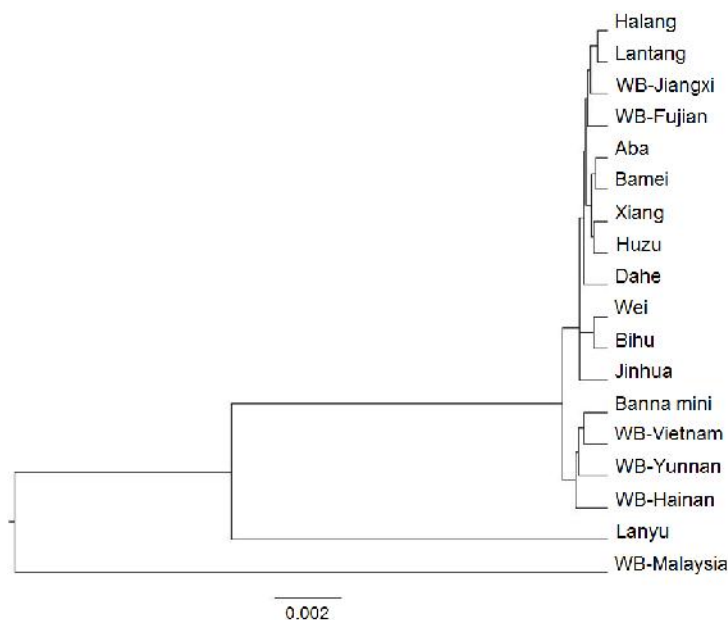


Fig. 3. The phylogenetic relationship was analyzed using discrete data based method (Bayesian Inference) by BEAST v1.8.3 software [19]. The phylogenetic tree was reconstructed by comparison of complete coding region sequences of mitogenome of Halang pig and 17 pig breeds by Tree Annotator software v.1.8.4.

Examination of the phylogenetic trees produced from the D-loop and complete coding region of mtDNA sequence revealed some geographic relationships among the breeds. The trees confirmed again the monophyletic position of Asian domestic pigs and wild boars. It presented two divergent clusters, including the Malaysian wild boar (*Sus barbatus*) that fell outside the two major-clades consisting pigs of Mekong regions and Chinese regions [12]. Similarly with the previous studies, the

Lanyu breed found more far away from other Chinese pigs [10, 23]. Wild boar Malaysian pig was of independent branches, in which the genetic distance appeared 0.036 with both of the other domestic and wild boar pigs. However, these pig breeds probably still belong to the Asian type.

In both phylogenetic trees of control region and complete coding region, Halang pig falls in the subgroup with South Chinese pigs. They would have a close relationships and which probably were formed from closely related maternal ancestors, but it remains to be investigated to make more clearly. This offers a historical trade hypothesis between Cao Bang and the provinces of Southern China. Beside the commodity products, domestic animals such as Halang pigs are also traded, and underwent thousands of years they gradually became indigenous animals of Vietnam. It is quite possible because Cao Bang has a geographical location that is a bordering province with Southern China. Previous study was also indicated that pigs might have been domesticated independently from subspecies of the European and Asian wild boar populations [24]. From the phylogenetic trees revealed that wild board subclades (e.g., WB Jiangxi, WB Fujian, WB Vietnam, WB Yunnan, and WB Hainan) were mainly distributed in South Asia, South East Asia and Chinese regions.

4. CONCLUSION

In conclusion, our results suggest that there were close phylogenetic relationships of Halang pig with other Asian pig breed, especially with Lantang pig breed from South China region. Together with published report, there is a hypothesis was the Halang pig's origin belongs to South Chinese region. However, the other evolutionary evidences should be further studied and the origin of Halang pig needs to make more clarify by other evolutionary research tools. Besides, the sequencing and phylogenetic analysis of the whole mt genome of Halang pig would be useful for genetic study such as disease-resistance varieties and further evolutionary researches.

REFERENCES

1. Lan H, Shi L. The origin and genetic differentiation of native breeds of pigs in southwest China: an approach from mitochondrial DNA polymorphism. *Biochem Genet.* 1993;31(1-2):51-60.
2. Tran TNT, Nia P, Chena J, Leb TT, Stevec K, Han J, Wanga H, Zhao S. The complete mitochondrial genome of Mong Cai pig (*Sus scrofa*) in Vietnam. *Mitochondrial DNA Part B* 2016;1(1):226-227.
3. Watanabe T, Hayashi Y, Kimura J, Yasuda Y, Saitou N, Tomita T, Ogasawara N. Pig mitochondrial DNA: polymorphism, restriction map orientation, and sequence data. *Biochem Genet.* 1986;24(5-6):385-96.
4. Piper PJ. The Origins and Arrival of the Earliest Domestic Animals in Mainland and Island Southeast Asia: A Developing Story of Complexity, in *New Perspectives in Southeast Asian and Pacific Prehistory*, Piper PJ, Matsumura H, Bulbeck D, Editors. 2017, ANU Press, The Australian National University: Canberra, Australia. 251-273.
5. Cummins J. Mitochondrial DNA and the Y chromosome: parallels and paradoxes. *Reproduction, Fertility and Development.* 2002;13(8):533-542.
6. Gongora J, Fleming P, Spencer PB, Mason R, Garkavenko O, Meyer J-N, Droegemueller C, Lee JH, Moran C. Phylogenetic relationships of Australian and New Zealand feral pigs assessed by mitochondrial control region sequence and nuclear GPIP genotype. *Molecular Phylogenetics and Evolution.* 2004;33(2):339-348.
7. Imes DL, Wictum EJ, Allard MW, Sacks BN. Identification of single nucleotide polymorphisms within the mtDNA genome of the domestic dog to discriminate individuals with common HVI haplotypes. *Forensic Science International: Genetics.* 2012;6(5):630-639.
8. Sambrook J. PM, David Russell. *Molecular Cloning: A Laboratory Manual.* New York, USA: Cold Spring Harbor Laboratory Press. 2000.
9. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A.* 1977;74(12):5463-7.
10. Yu G, Xiang H, Wang J, Zhao X. The phylogenetic status of typical Chinese native pigs: analyzed by Asian and European pig mitochondrial genome sequences. *J Anim Sci Biotechnol.* 2013;4(1):9.
11. Ghivizzani SC, Mackay SL, Madsen CS, Laipis PJ, Hauswirth WW. Transcribed heteroplasmic repeated sequences in the porcine mitochondrial DNA D-loop region. *Journal of molecular evolution.* 1993;37(1):36-47.

- 234 12. Wu C, Jiang Y, Chu H, Li S-H, Wang Y, Li Y, Chang Y, Ju Y. The type I Lanyu pig has a
235 maternal genetic lineage distinct from Asian and European pigs. *Animal genetics*.
236 2007;38(5):499-505.
- 237 13. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30(14):2068-
238 9.
- 239 14. Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, Pütz J, Middendorf M,
240 Stadler PF. MITOS: Improved de novo metazoan mitochondrial genome annotation.
241 *Molecular phylogenetics and evolution*. 2013;69(2):313-319.
- 242 15. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW.
243 GenBank. *Nucleic acids research*. 2013;41(D1):D36-D42.
- 244 16. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped
245 BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic*
246 *acids research*. 1997;25(17):3389-3402.
- 247 17. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput.
248 *Nucleic Acids Res*. 2004;32(5):1792-7.
- 249 18. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: molecular evolutionary genetics analysis
250 (MEGA) software version 4.0. *Molecular biology and evolution*. 2007;24(8):1596-1599.
- 251 19. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the
252 BEAST 1.7. *Mol Biol Evol*. 2012;29(8):1969-73.
- 253 20. Felsenstein J. Confidence Limits on Phylogenies: An Approach Using the Bootstrap.
254 *Evolution*. 1985;39(4):783-791.
- 255 21. Ran ML, Liu Z, Yang AQ, Li Z, Chen B. The complete sequence of the mitochondrial genome
256 of Lantang pig (*Sus scrofa*). *Mitochondrial DNA A DNA Mapp Seq Anal*. 2016;27(2):1376-7.
- 257 22. MacKav DG. The theoretical epistemology: A new perspective on some long-standing
258 methodological issues in psychology, in *A handbook for data analysis in the behavioral*
259 *sciences: Methodological issues*, Lewis GKC, Editor. 1993, University of California: Los
260 Angeles, USA. 229-255.
- 261 23. Chen C-H, Huang H-L, Yang H-Y, Lai S-H, Yen N-T, Wu M-C, Huang M-C. Mitochondrial
262 genome of taiwan pig (*Sus scrofa*). *African Journal of Biotechnology*. 2011;10(13):2556-2561.
- 263 24. Okumura N, Kurosawa Y, Kobayashi E, Watanobe T, Ishiguro N, Yasue H, Mitsuhashi T.
264 Genetic relationship amongst the major non-coding regions of mitochondrial DNAs in wild
265 boars and several breeds of domesticated pigs. *Animal genetics*. 2001;32(3):139-147.