

# The Robust Phylogeny of Korean Wild Boar (*Sus scrofa coreanus*) Using Partial D-Loop Sequence of mtDNA

In-Cheol Cho<sup>1</sup>, Sang-Hyun Han<sup>1</sup>, Meiying Fang<sup>2</sup>, Sung-Soo Lee<sup>1</sup>, Moon-Suck Ko<sup>1</sup>, Hang Lee<sup>3</sup>, Hyun-Tae Lim<sup>4</sup>, Chae-Kyoung Yoo<sup>4</sup>, Jun-Heon Lee<sup>5</sup>, and Jin-Tae Jeon<sup>4,\*</sup>

In order to elucidate the precise phylogenetic relationships of Korean wild boar (*Sus scrofa coreanus*), a partial mtDNA D-loop region (1,274 bp, NC\_000845 nucleotide positions 16576–1236) was sequenced among 56 Korean wild boars. In total, 25 haplotypes were identified and classified into four distinct subgroups (K1 to K4) based on Bayesian phylogenetic analysis using Markov chain Monte Carlo methods. An extended analysis, adding 139 wild boars sampled worldwide, confirmed that Korean wild boars clearly belong to the Asian wild boar cluster. Unexpectedly, the Myanmar/Thai wild boar population was detected on the same branch as Korean wild boar subgroups K3 and K4. A parsimonious median-joining network analysis including all Asian wild boar haplotypes again revealed four maternal lineages of Korean wild boars, which corresponded to the four Korean wild boar subgroups identified previously. In an additional analysis, we supplemented the Asian wild boar network with 34 Korean and Chinese domestic pig haplotypes. We found only one haplotype, C31, that was shared by Chinese wild, Chinese domestic and Korean domestic pigs. In contrast to our expectation that Korean wild boars contributed to the gene pool of Korean native pigs, these data clearly suggest that Korean native pigs would be introduced from China after domestication from Chinese wild boars.

## INTRODUCTION

Wild boars (*Sus scrofa*) are distributed throughout Asia, Europe and Northwest Africa, and at least 16 subspecies are currently recognised (Epstein, 1984; Herre and Rohrs, 1977; Ruvinsky and Rothschild, 1998). The phylogeography of wild boars provides insights into expansion of human populations (Larson et al., 2007), and analysis of the domestication of wild boars suggests that this process led to one of the most important socio-

economic transitions in human history (Larson et al., 2005).

Two wild boar subspecies are found in Japan: the Japanese wild boar (*Sus scrofa leucomystax*) inhabits the main Japanese islands of Honshu, Shikoku and Kyushu, and the Ryukyu wild boar (*Sus scrofa riukiuanus*) is found on the Ryukyu Islands. Previous phylogenetic studies using the mtDNA control region have indicated that Korean wild boars are indigenous to Asia but differ from the Japanese subspecies (Hongo et al., 2002; Naya et al., 2003). Based on the fossil record of Japanese land mammals, Kawamura (2007) estimated that a land bridge formed between the Korean Peninsula and Kyushu Island between 300,000 and 500,000 years before present (YBP). They proposed that continental wild boars migrated to Japan at that time and that the progenitor of Japanese wild boars was a Eurasian subspecies.

The wild boar was the progenitor of the domestic pig (Giuffra et al., 2000; Kijas and Andersson, 2001; Larson et al., 2005). Domestication of the pig is likely to have occurred first in the Near East ~9,000 YBP and may have occurred repeatedly from local populations of wild (Bökonyi, 1974). Analysis of mtDNA *cytB* sequences indicates that the time since divergence of European and Asian wild boar populations is of the order of 500,000 YBP, before domestication ~9,000 YBP (Giuffra et al., 2000). Moreover, previous studies suggest that Asian domestic pigs are very closely related to Chinese wild boar with respect to genetic distance and genealogical classification (Fang et al., 2005; Kim et al., 2002; Li et al., 2004; Yang et al., 2003).

The Korean wild boar (*Sus scrofa coreanus*) inhabits the Korean mainland except for the largest island, Jeju. Archaeological evidence from terrestrial macromammalian fauna suggests that Korean wild boars appeared on the peninsula in the mid-Pleistocene (780,000 to 130,000 YBP) (Norton, 2000). However, there haven't been any independent studies using sufficient Korean wild boar samples for the analyses of genetic structure or history of this population. In the phylogeographic analyses using a large set of world-wide wild boar samples, a

<sup>1</sup>Jeju Sub-Station, National Institute of Animal Science, Rural Development Administration, Jeju 690-150, Korea, <sup>2</sup>College of Animal Science and Technology, China Agricultural University, Beijing 100094, China, <sup>3</sup>Conservation Genome Resource Bank for Korean Wildlife, Brain Korea 21 Program for Veterinary Science and College of Veterinary Medicine, Seoul National University, Seoul 151-742, Korea, <sup>4</sup>Division of Applied Life Science (Brain Korea 21 Program), Graduate School of Gyeongsang National University, Jinju 660-701, Korea, <sup>5</sup>Division of Animal Science and Resources, College of Agriculture and Life Sciences, Chungnam National University, Daejeon 305-764, Korea

\*Correspondence: jtjeon@gnu.ac.kr

few Korean wild boar samples have been ever involved (Larson et al., 2005), but the samples were dealt as a minor portion in East Asia. On the other hand, Chinese and Japanese wild boars have been being considered and dealt as major sub-populations in East Asia due to the accumulated genetic information through extensive phylogenetic analysis.

Therefore, in this study, samples of 56 Korean wild boars from the Southern Korean Peninsula which is the largest number ever used for the phylogenetic study of Korean wild boar were collected to investigate their genetic diversity and estimate their evolutionary relationships based on partial mtDNA D-loop sequences. Several phylogeny analysis programs and tools (refer the "Materials and Methods") were applied to deduce valuable genetic information focusing on Korean wild boar. Additionally, evidence that the Korean wild boar contributed to pig domestication in Asia was also investigated.

## MATERIALS AND METHODS

### Animals and DNA extraction

Tissue samples were kindly provided by the Conservation Genome Resource Bank for Korean Wildlife and National Veterinary Research & Quarantine Service. In total, 56 Korean wild boars (the appearance is presented in Supplementary Fig. 3A) were collected from 18 locations, which could be grouped into 4 geographical areas (designated as Region 1 to Region 4 in Fig. 4) in the southern Korean Peninsula. Blood samples of 1 Korean native pigs (the feature is shown in Supplemental Fig. 3B) from Jeju Island were collected from the abattoir. Total DNA was isolated using the sucrose-proteinase K method (Birren et al., 1997).

### PCR, cloning and sequencing of the mtDNA D-loop region

Using the highly conserved tRNA-Pro and tRNA-Phe sequences within the mtDNA (GenBank accession number NC\_000845, nucleotide positions 16576-1236), we designed a primer set to PCR-amplify 1,274 bp of the porcine mtDNA control region. The primers were designated pDF (5'-AGCACCCA-AAGCTGAAATTC-3') and pDR (5'-AGCTGTGAGGCTCATCT-AGG-3'). The 25  $\mu$ l reactions contained 10 mM of each primer, 200  $\mu$ M dNTPs, 1  $\times$  reaction buffer, 1.5 U *Taq* DNA polymerase (Promega, USA), and ~50 ng DNA. Then 5  $\mu$ l of each PCR product was separated by electrophoresis on a 1% agarose gel, stained with ethidium bromide and visualised under UV illumination to determine amplification success. The remainder of the reaction was purified using the QIAquick PCR purification kit (Qiagen, Germany).

The purified PCR products were cloned using the TOPO TA Cloning Kit (Invitrogen, USA), according to the manufacturer's instructions. DNA sequencing of cloned DNA fragments was carried out using the DYEnamic ET Dye terminator Cycle Sequencing kit (Amersham Pharmacia, Sweden). The boundaries of the D-loop were determined by comparison with the pig mtDNA genome sequence (NC000845; Lin et al., 1999). The Sequencher 4.6 software package (Gene Codes Corporation, USA) was used to produce a consensus sequence of each DNA sample by correcting sequencing errors and to identify possible polymorphisms among consensus sequences.

### Sequence alignment and phylogenetic analyses

We produced sequences from 56 Korean wild boars and 10 Korean native pigs in this study and collected an additional 137 sequences that represent worldwide populations of wild boar from the GenBank database (sequence information is available in Supplementary Table 1). A multiple alignment of the D-loop

sequences was performed using CLUSTAL W (Thompson et al., 1994), implemented in DAMBE 4.5.50 (Xia and Xie, 2001). Data format conversion between Fasta and Nexus and between Nexus and Arlequin/Roehl was done using DAMBE and DnaSP 4.50.3 (Rozas et al., 2003), respectively. Indices of total sequence variation, nucleotide diversity ( $\pi$ ), number of nucleotide differences ( $K$ ), number of nucleotide substitutions per site ( $D_{XY}$ ), and haplotype structure were estimated using DnaSP. The aligned sequences were used to construct the following 5 data sets: 1) 1,072 bp including indels for the phylogenetic analysis of 56 Korean wild boars (see Fig. 2A); 2) 620 bp including indels for the phylogeographic analysis of 193 wild populations from the islands of Southeast Asia (ISEA), India, East Asia and Europe (see Fig. 2B); 3) 602 bp excluding indels for the network analysis of Asian wild boars (see Fig. 3A); 4) 602 bp excluding indels for the domestication analysis in East Asia, containing 109 Asian wild boars and 60 Asian domestic pigs (see Fig. 3B and Supplementary Fig. S2) and 5) 612 bp for checking feral signatures in Korean wild boars using 193 worldwide wild boars and 51 domestic pigs (see Supplementary Fig. 4).

Phylogenetic reconstruction was performed using multiple steps. The appropriate analysis model for the data sets was determined using hierarchical likelihood-ratio tests performed by MrModeltest 2.3 (<http://darwin.uvigo.es/software/modeltest.html>), as implemented in PAUP\* 4.0, Windows Beta v10 (Swofford, 2003). The Hasegawa-Kishino-Yano (HKY) (Hasegawa et al., 1985) evolutionary model with invariable sites (HKY + I;  $nst = 2$  and rates = propinv) was identified as the best model for data set 1, and the General Time Reversible (GTR) (Tavare, 1986) model with gamma-distributed rate variation across sites and a proportion of invariable sites (GTR + I + G;  $nst = 6$  and rates = invgamma) was identified as the best model for data set 2 and 5. A Bayesian-derived consensus tree was constructed using MR BAYES 3.1 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003). The software default priors, such as the number of iterations and the sampling frequency, were assumed after a pilot run with 1 million iterations. The main analysis was run until the average standard deviation of split frequencies fell below 0.01. Of the sampled parameter values, 25% were summarised to output a cladogram with the posterior probabilities for each split and a phylogram with mean branch lengths. The consensus tree was edited using TreeView 1.6.6 (<http://taxonomy.zoology.gla.ac.uk/rod/treeview.html>) and illustrated using FigTree version 1.0 (<http://tree.bio.ed.ac.uk/software/figtree/>). The relationships between haplotypes of Asian wild boars and Asian domestic pigs were estimated by a parsimonious median-joining method (Bandelt et al., 1999) and visualised using Network 4.5.1.0 and Network Publisher 1.1.0.6 (<http://www.fluxus-engineering.com>). Nucleotide weighting ( $\omega$ ) was adjusted to reflect the difference in mutational frequency between transversions ( $\omega = 20$ ) and transitions ( $\omega = 10$ ), where the rare-common event received the higher value.

### Analysis of population expansion

The observed number of pair-wise differences was estimated in order to search for signs of population expansion using Fu's  $F_s$  statistic (Fu, 1997) by Arlequin 3.01 (Excoffier et al., 2005). Observed mismatch distributions within haplogroups K2, K3 and K4 were fitted against model parameters, assuming a sudden expansion model (Rogers, 1995). Then, a P-value test, as described by Schneider and Excoffier (1999), was performed. For those sequence sets that fit the spatial expansion model (Excoffier, 2004), the time in years since commencement of the expansion ( $t$ ) was estimated from  $t = \pi/2\mu$  using a nonlinear-stepwise least squares approach implemented in Arlequin 3.01

**Table 1.** Summary of phylogenetic variation and relationships among subgroups of Korean wild boars used in this study

	Subgroup				Total
	K1	K2	K3	K4	
Sequence	13	12	14	17	56
Variable site <sup>a</sup>	13	19	6	7	47
Haplotype	5	9	5	6	25
Nucleotide diversity <sup>b</sup> ( $\pi$ )	3.51	3.03	1.61	1.68	
Nucleotide difference <sup>c</sup> ( $K$ )	2.13	1.83	0.98	1.02	
$D_{xy}$ <sup>d</sup>					
K1		(0.19)	(0.42)	(0.38)	
K2	1.67		(0.22)	(0.32)	
K3	1.68	1.05		(0.15)	
K4	1.47	1.55	0.67		

<sup>a</sup>Variable sites are single nucleotide polymorphism.

<sup>b</sup>Nucleotide diversity ( $\times 10^{-3}$ ) was calculated excluding indels.

<sup>c</sup>Average number of nucleotide differences between haplotypes within a haplogroup

<sup>d</sup>Nucleotide substitution per site (%) between haplogroups using the Jukes and Cantor correction (Nei, 1987). Numbers in parentheses over the diagonal are standard deviation.

(Excoffier et al., 2005). A confidence interval ( $\alpha < 0.05$ ) was calculated using 10,000 parametric bootstrap replicates. Tau ( $\tau$ ) is the empirical peak of the mismatch distribution and  $r = m_t\mu$ , where  $m_t$  is the sequence length (612 bp) and  $\mu$  is the substitution rate. A general substitution rate for mammalian mtDNA,  $\mu_{\text{general}} = 1 \times 10^{-8}$  substitutions/site/year inferred from a sequence divergence of ~2% per million years for mtDNA which has been found among mammalian species (Brown et al., 1979), was used to estimate the time in years since commencement of the expansion.

### Analysis of regional differentiation

A four by four (haplogroup  $\times$  region) contingency table was constructed for a chi-square test of independence. The probability level ( $\alpha$ ) for the null hypothesis that haplogroups are evenly distributed in each geographical area (designated as 1 to 4 in Fig. 4B) was estimated. A global test of population differentiation (Goudet et al., 1996) and an analysis of molecular variance (AMOVA) (Weir, 1996) based on haplotype frequencies found in each area were performed using the genetic structure setting of Arlequin 3.01. To estimate the exact P-value for pair-wise differences representing independence between Korean wild boar populations classified by area, a Markov chain of length 100,000 steps was applied.

## RESULTS

### Sequence polymorphism of the mtDNA D-loop region

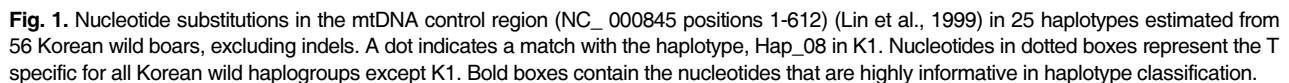
Within the 612 bp mtDNA D-loop region, 45 variable positions were found, including 10 indel events and 35 single nucleotide polymorphisms (SNPs). There were 15 parsimony informative sites among the SNPs. A total of 25 haplotypes, classified into four haplogroups, were detected using the SNPs (Fig. 1). The average number of nucleotide differences between haplotypes ( $K$ ) was 6.38, and the haplotype diversity ( $H_d$ ) was  $0.91 \pm 0.02$ . The Korean wild boar population showed a similar level of heterogeneity to the Japanese wild boar population used in this study ( $K = 5.01$  and  $H_d = 0.97 \pm 0.02$  from 18 haplotypes from 24 individuals). While haplogroup K1 showed the greatest nucleotide diversity ( $\pi \times 10^3 = 3.51$ ) and divergence ( $K = 2.13$ ), K3 showed the least phylogenetic variation ( $\pi \times 10^3 = 1.61$  and  $K = 0.98$ ) (see Table 1). Based on comparison with sequences from

wild boars worldwide, 90% (18 out of 20) of Korean wild boar haplotypes, except for K1, had a specific T at the 280<sup>th</sup> nucleotide position (Fig. 1). However, 1.2% (1 out of 83 haplotypes) of the worldwide wild boar haplotypes also possessed a T at this position.

### Phylogenetic relationships

In order to investigate the genetic diversity of Korean wild boars, a phylogenetic tree was constructed using a Bayesian Markov chain Monte Carlo method. Korean wild boars were split into four subgroups, K1 to K4 (Fig. 2A), representing 23.2, 21.4, 25.0 and 30.4% of the tree components, respectively. The posterior probabilities for the nodes were between 100% and 83%. For the phylogeographic analysis, Korean wild boar sequences were added to an alignment of mtDNA D-loop sequences from 137 worldwide wild boars. In order to make the sequence alignment of consistent length, the Korean wild boar sequences were trimmed to 602 bp. In data set 2, 108 haplotypes and 105 nucleotide polymorphisms were identified. The worldwide wild boars formed four main geographical clusters: Europe, India, the Islands of Southeast Asia (ISEA) and Asia (Fig. 2B). There were two exceptions (about 1%): an Iranian and an Indonesian wild boar (AY884638 and AY884688 respectively) were located within the Asian cluster (Supplementary Fig. 1). Korean wild boars were clearly located within the Asian cluster, again divided into four subgroups. The posterior probability of nodes for dividing the Korean wild boar subgroups were 80 to 57%. These relatively low probabilities were due to trimming of sequence data up from 1,072 bp (data set 1) to 620 bp (data set 2). Interestingly, an unexpected relationship among the Myanmarese/Thai group and K3 and K4 was revealed.

A maximum-parsimony network was produced using the Asian wild boar haplotypes (Fig. 3A). A radiation from K2 to K3, followed by K4, is clearly identified. K1 is located in a region in which the Chinese, Russian, Taiwanese and Japanese wild populations show a complex pattern of reticulation. K1 shows a relationship with K2 via two Taiwanese and one Japanese wild haplotypes. Together with the fact that K1 possesses a C at the 280<sup>th</sup> nucleotide position (Fig. 1), these features suggest that K2's progenitors were introduced independently from those of K1 and contributed to forming the major subspecies that inhabit the Korean peninsula. Therefore, K2 could be the most ancient

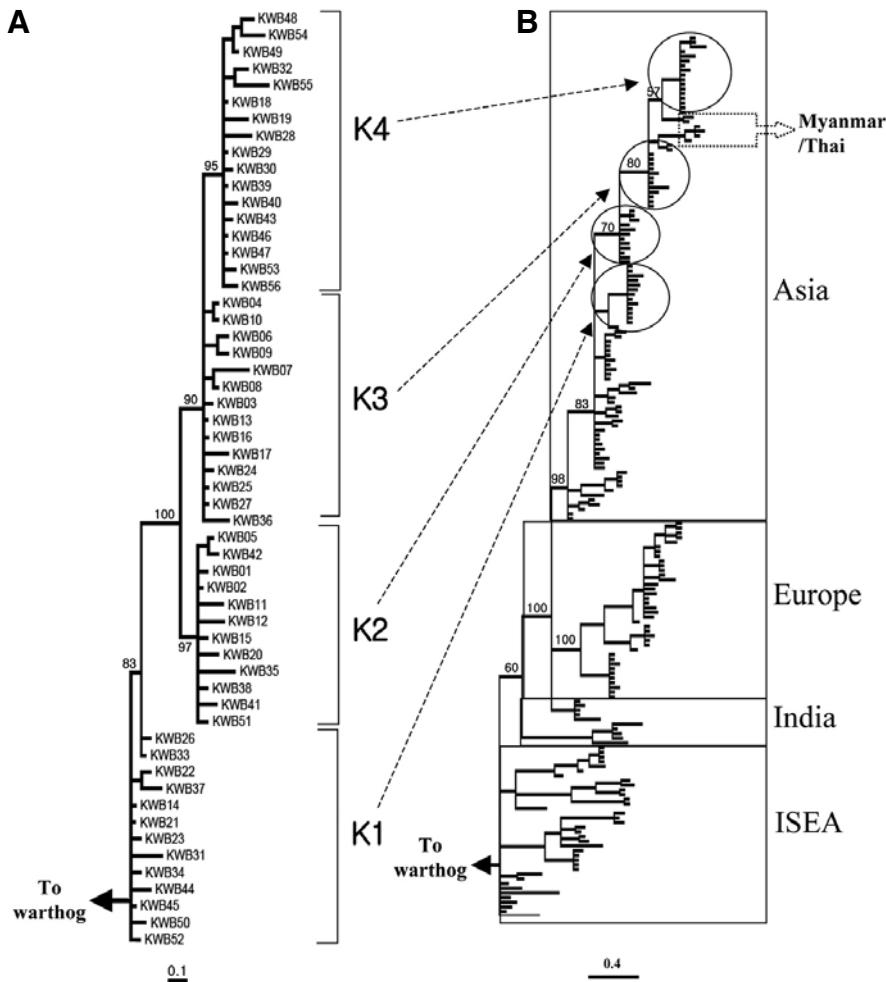


years (K2 to K3 = 527,000 years and K3 to K4 = 336,000 years) when  $\mu_{general}$  was applied. If Norton's *TODs* (780,000 and 130,000 years) become fixed factors, K2 haplotypes were the first to immigrate into Korea, and K4 haplotypes were the most recently established, then two  $\mu_{Kus}$ ,  $1.11 \times 10^{-8}$  and  $6.64 \times 10^{-8}$  substitutions/site/year corresponding to the lower and upper bounds, could be simply calculated by a reverse operation.

### Rejection of the possible contribution of Korean wild boars to pig domestication

In total, 34 haplotypes from Korean and Chinese domestic pigs were added to the Asian wild boar network in order to determine the domestication centre in East Asia and the core haplotypes shared by wild boars and domestic pigs. Two haplotypes C31 and C33, are shared by both wild and domestic pigs (Fig. 3B). Particularly in C31, two Korean native pigs show 100% sequence identity with Chinese domestic pigs (Meishan, Jiaangquhai, 2 Tongcheng, Wananhua and Yushanhei) and a Chinese wild boar. In contrast to our expectation that Korean wild boars contributed to the gene pool of Korean native pigs these data clearly suggest that Korean native pigs would be introduced from China after domestication from Chinese wild boars. Consequently, the possibility of another domestication centre in Korea was rejected. Although a clear star-like burst of the domestic pig from haplotype C31 is evident, the overall diffuse nature of the network (Supplementary Fig. 2) suggests domestication from more than one wild boar lineage in East Asia, as hypothesised by Larson et al. (2005). An extended phylogenetic tree containing 193 worldwide wild boars, including 51 European and Asian domestic pigs, indicates that there

Novel substitution rates that are specific for the Korean wild boar population ( $\mu_{KUS}$ ) were inferred using the upper and lower times of divergence ( $TOD_{KUS}$ ) bounds, 780,000 and 130,000 YBP, respectively, which were estimated from an archaeological investigation of wild boars' appearance in Korea (Norton, 2000). The estimate of  $TOD_{KUS}$  from K2 to K4 was 863,000



**Fig. 2.** Consensus phylogenetic trees reconstructed using the statistics for the taxon bipartitions and clade credibility (posterior probability) values. (A) A consensus tree for Korean wild boars using 1,072 bp of mtDNA sequence (NC\_000845 positions 1-1,072). The Bayesian Markov chain Monte Carlo analysis was run with the Hasegawa-Kishino-Yano (HKY) (Hasegawa et al., 1985) evolutionary model with invariable sites (HKY + I;  $\text{nst} = 2$  and rates = propinv) implemented in MrBays 3.1. The average standard deviation of split frequencies reached 0.008161 after 2 million iterations. The tree components were calculated from the sampled parameters after burn-in using the first 25% of the samples. (B) A consensus tree using 612 bp of D-loop sequences (NC\_000845 positions 1-612) from worldwide wild boars, including Korean wild boars. The General Time Reversible (GTR) (Tavare, 1986) model with gamma-distributed rate variation across sites and a proportion of invariable sites (GTR + I + G;  $\text{nst} = 6$  and rates = invgamma) was applied. The average standard deviation of split frequencies reached 0.006923 after 20 million iterations. Warthog (AB046876) was used as an outgroup. The log likelihood of the cold chain increased from -9,856.003 (initial) to -3,221.777 (final). We inferred a split from ISEA to India followed by Europe and East Asia, then dispersal from India to West Europe through Armenia/Iran. Korean wild boars belong to the Asian cluster, and Myanmar/Thai wild boars are located with K3 and K4 on the same branch. Refer to Supplementary Fig. 3 and Supplementary Table 1 to view fully labelled taxa and decode the shortened representation of each taxon.

ter, and Myanmar/Thai wild boars are located with K3 and K4 on the same branch. Refer to Supplementary Fig. 3 and Supplementary Table 1 to view fully labelled taxa and decode the shortened representation of each taxon.

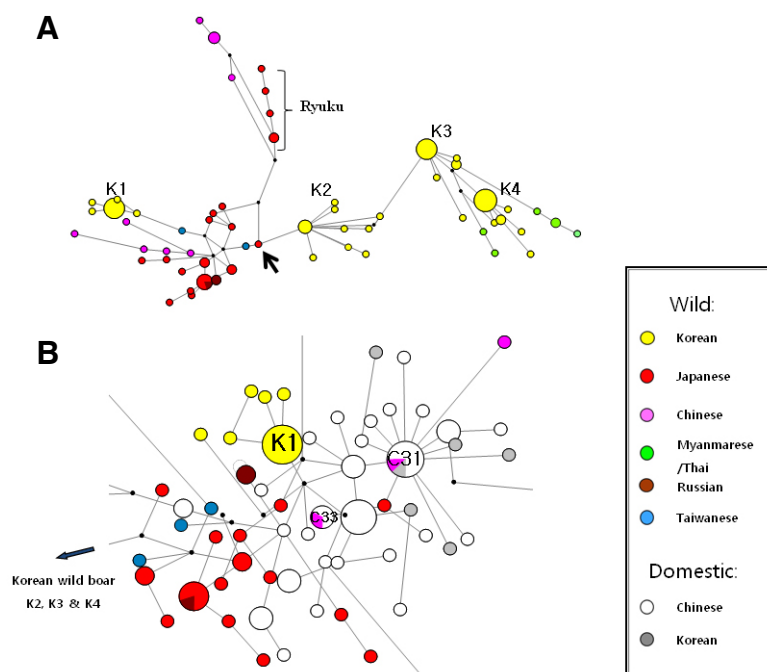
is no feral signal in the Korean wild boar population (Supplemental Fig. 4).

## DISCUSSION

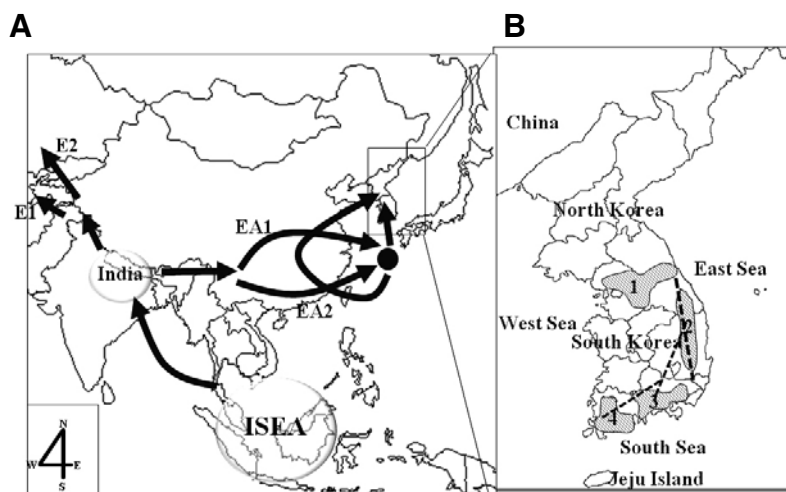
This is the first study to sample a relatively large number of Korean wild boars, facilitating the discovery of four major maternal lineages. K2 was identified as the haplogroup initially present within South Korea, which gave rise to K3 followed by K4. As shown in Fig. 1, there are eight nucleotide positions that are critical to classify the haplotypes into the four subgroups: 1) two nucleotides at the 179<sup>th</sup> and 448<sup>th</sup> positions, showing complete linkage disequilibrium that discriminate between K1 and the remaining haplogroups; 2) a C fixed in K2 at the 212<sup>th</sup> nucleotide position; 3) a C specific for K4 at the 455<sup>th</sup> nucleotide positions; 3) two Cs specifically fixed in both K2 and K3 at the 265<sup>th</sup> and 304<sup>th</sup> nucleotide positions, respectively; 4) a C and a G specific for both K3 and K4 at the 321<sup>st</sup> and 439<sup>th</sup> nucleotide positions, respectively. In the AMOVA analysis for the Korean wild boars classified by haplogroup, 80.47% of the variance is due to variation among haplogroups, and only 19.53% is due to variation of haplotypes within haplogroups.

The Korean wild boar samples used in this study were collected in four regions (Fig. 4B). The wild boars in Region 1 were

from Yeoncheon in Gyeonggi-do and Cheolwon in Gangwon-do, close to the border between North and South Korea where the demilitarised zone is located and civilians are not permitted free access. The natural habitats of Region 1 are well preserved; therefore, formation of larger wild populations in this region than in other regions may be possible. Region 2 (in Gyeongsangbuk-do) is the Taebaek Mountains area (the dotted line on Region 2) (Fig. 4B), where wild boars rove without human interference. Regions 3 (in Gyeongsangnam-do) and 4 (in Jeollanam-do) are areas of the Soback and Charyoung Mountains, derived from the Taebak Mountains. In the AMOVA analysis using wild boars classified by region, 25.16% of the total variance is due to variation among regions, and 78.84% is due to variation of haplotypes within regions. The chi-square test of independence using a four regions  $\times$  four haplogroups contingency table rejects the null hypothesis ( $P < 0.001$ ), meaning that some haplotypes are limited to certain regions. In fact, K1 is completely restricted to Region 1, and K4 is limited to Regions 1 and 2. In a differentiation test for pairs of regions, the three pairs Regions 1 and 2, Regions 2 and 3, and Regions 3 and 4 show no significance; however, other region pairs are significantly different ( $P < 0.05$ ). From these results, we conclude that 1) Region 1 is the largest pool of Korean wild haplotypes; 2) the Korean wild haplotypes, except for those in K1,



**Fig. 3.** Parsimonious median-joining networks depicting the relationships between the Asian wild and domestic pigs that inhabit East Asia. (A) A network illustrating the relationships of Asian wild haplotypes. A Chinese/Ryukyu lineage derived from ISEA wild boars is clearly identified. The grouping of Korean wild boars into K1 to K4 (see Fig. 1) is verified. The relationship between K3 and the Myanmar/Thai population is also confirmed. The Japanese wild haplotype (AB015085) designated by an arrow is unique evidence that the Korean wild boars' progenitors were introduced from the Okinawa Islands to Korea through Japan. (B) A median-joining network depicting the relationships between 45 Chinese and Korean domestic pigs and the Asian wild boars used in Fig. 3A, located mainly in East Asia. The haplotype C31 is shared by the Chinese wild, Chinese domestic and Korean domestic pigs, and C33 is found in the Chinese wild and Chinese domestic pigs. There is no signature suggesting that Korean wild boars contributed to the domestication of Chinese or Korean domestic pigs.



**Fig. 4.** Hypothesis for the formation and migration of the extant Korean peninsula wild boar populations. (A) The main radiation of wild boars is illustrated on the basis of the results shown in Figs. 2B and 3A. The initial dispersal from ISEA to India, followed by Europe and East Asia, is marked. A split from India to Italy (E1) and to western Europe through Armenia/Iran (E2), is confirmed. A black spot indicates the point at which the Chinese/Ryukyu lineage converged, which is related to Korean wild boars through a few Japanese and Taiwanese wild boars. A novel westward dispersal route of Korean wild boars, back to Indochina, is suggested in this study. (B) The four regions where the Korean wild boar samples were collected are marked on the map. The dotted line indicates the mountains in each region (marked with slash lines). Region 1 is the area of Yeoncheon in Gyeonggi-do and

Cheolwon in Gangwon-do, close to the border between North and South Korea. Region 2, in Gyeongsangbuk-do, is the Taebaek Mountains. Region 3, in Gyeongsangnam-do, and Region 4, in Jeollanam-do, are areas of the Soback and Charyoung Mountains, respectively.

have dispersed into the neighbouring regions through the mountains; and 3) each region, however, is still somewhat isolated. Differentiation signals were not identified between pairs of regions that are one step apart, for example, Regions 1 and 3 ( $P = 0.0161 \pm 0.0023$ ) and Regions 2 and 4 ( $P = 0.01826 \pm 0.0016$ ), etc. By the samples from South Korea used in this study, the explanation for the presence of K1 haplotypes only in Region 1 is limited. If samples from North Korea were added, the dispersal of K1 haplotypes to and from the northern area of the Korean Peninsula could be explored.

The phylogenetic relationships between Korean wild boars and other populations worldwide are clearly delineated in this study. The 218<sup>th</sup> nucleotide position is highly specific for Korean wild boars, except for K1 (Fig. 1). This polymorphism supports the idea that the Korean wild haplogroups did not originate from

a single progenitor; that is, K1 had a separate origin from K2. The dispersal route of wild haplotypes is illustrated (Fig. 4A) on the basis of the results shown in Figs. 2B and 3A. The initial dispersal from ISEA to India, followed by Europe and East Asia, is clearly shown. A split from India to Italy (E1) and to western Europe through Armenia/Iran (E2), is also confirmed. The consistency of our data with the widely accepted phylogeographic analysis of Larson et al. (2005) reinforces our conclusions regarding Korean wild boars. Focusing on the Asian cluster (Fig. 3A), a Chinese/Ryukyu (EA1/EA2 in Fig. 4A) lineage that radiated from ISEA is clearly identified. It is related to Korean wild boars through one Japanese (AB015085) and two Taiwanese wild boars (AY884708 and AY884706). With only these features in this study, the Okinawa Island area may have been a bridge of radiation of the Chinese/Ryukyu lineage into Korea and

Japan. However, this suggestion is not supported by any phylogenetic studies using Japanese wild boars. Watanobe et al. (2004) indicated that Japanese wild boars are more closely related to the Northeast wild boars than Ryukyu wild boar and a group genetically similar to contemporary Japanese wild boars was found mainly on Honshu, Hokkaido and Izu islands. Apparently, the K1 to K4 haplogroups are different from Japanese wild boar, as suggested by Hongo et al. (2002) and Naya et al. (2003). However, perhaps some of the Korean wild boars' ancestors migrated from Japan through land bridges formed between the Korean peninsula and Kyushu island (500,000 and 300,000 YBP) (Kawamura, 2007). This inference is supported by unique evidence from a Japanese wild haplotype (AB015085) found on Honshu island (Okumura et al., 1996), which is directly linked with K2 (Fig. 3A). The Japanese mainland was connected to the Asian continent via a land bridge three times during the Quaternary, as evidenced by the migration of large land animals from the continent (Dobson and Yoshinari, 1998). The ancestors of Japanese wild boars may have spread from Southeast to Northeast Asia during the Middle to Late Pleistocene (780,000-10,000 years ago) (Watanobe et al., 2003).

We found a novel phylogenetic relationship between Korean and the Myanmar/Thai wild boar. During the glacial epoch in the Pleistocene of East Asia, sea level dropped by 100-140 meters, and most continental shelves were exposed. Land bridges were formed in the sea around Korea and southern China (Choi, 2003). A westward dispersal to Indochina, on the basis of the relationship between K3 and Myanmar/Thai wild boar would be suggested. However, as there is no clear evidence for the migration, the coincidence of the nucleotide mutations in D-loop region would be more reasonable explanation for the relationship.

In conclusion, we propose a robust phylogeny and phylogeography of Korean wild boars using a relatively large number of samples. Our data enable us to reject the contribution of Korean wild boars to pig domestication in East Asia. Further studies can solve unanswered questions, such as the lack of prominent dispersals from haplogroup K1.

*Note: Supplementary information is available on the Molecules and Cells website (www.molcells.org).*

## ACKNOWLEDGMENTS

This work was supported by grant no. 20080401034053 from the BioGreen 21 Program, Rural Development Administration, Korea. C.K. Yoo was supported by a scholarship and H. Lee was partially supported by a grant from Brain Korea 21 Program, the Ministry of Education, Science and Technology, Korea. The authors thank Soo-Ho Kim and the Korea Expressway Corporation for the kind donation of wild boar specimens to Conservation Genome Resource Bank for Korean Wildlife (CGRB).

## REFERENCES

Bandelt, H.J., Forster, P., and Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16, 37-48.

Birren, B., Green, E.D., Klapholz, S., Myers, R.M., and Roskams, J. (1997). *Genome Analysis: A Laboratory Manual* (USA: Cold Spring Harbor Laboratory Press).

Bökönyi, S. (1974). History of domestic mammals in central and Eastern Europe. *Academiai Kiado.*

Brown, W.M., George, Jr. M., and Wilson, A.C. (1979). Rapid evolution of animal mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* 76, 1967-1971.

Choi, D.K. (2003). *Understanding the Earth*. Seoul National University press.

Dobson, M., and Yoshinari, K. (1998). Origin of the Japanese land mammal fauna: Allocation of extant species to historically-based categories. *Daiyonki Kenkyu.* 37, 385-395.

Epstein, H. (1984). Pig. In *handbook of evolution of domesticated animals*, I.L. Mason, eds. (London : Longman), pp. 145-162.

Excoffier, L. (2004). Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. *Mol. Ecol.* 13, 853-864.

Excoffier, L., Laval, G., and Schneider, S. (2005). Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evol. Bioinform. Online* 1, 47-50.

Fang, M., Hu, X., Jiang, T., Braunschweig, M., Hu, L., Du, Z., Feng, J., Zhang, Q., Wu, C., and Li, N. (2005). The phylogeny of Chinese indigenous pig breeds inferred from microsatellite markers. *Anim. Genet.* 36, 7-13.

Fu, Y.X. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147, 915-925.

Giuffra, E., Kijas, J.M.H., Amarger, V., Carlborg, O., Jeon, J.T., and Andersson, L. (2000). The origin of the domestic pig: independent domestication and subsequent introgression. *Genetics* 154, 1785-1791.

Goudet, J., Raymond, M., Meeüs, T., and Rousset, F. (1996). Testing differentiation in diploid populations. *Genetics* 144, 1933-1940.

Hasegawa, M., Kishino, H., and Yano, T. (1985). Dating of the humanape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* 22, 160-174.

Herre, W., and Rohrs, M. (1977). Zoological considerations on the origins of farming and domestication. In *Handbook of Origins of Agriculture*, C.A. Reed, eds. (Mouton : The Hague), pp. 245-279.

Hongo, H., Ishiguro, N., Watanobe, T., Shigehara, N., Anezaki, T., Long, V.T., Bihn, D.V., Tien, N.T., and Nam, N.H. (2002). Variation in mitochondrial DNA of Vietnamese pigs: relationships with Asian Domestic pigs and Ryukyu wild boars. *Zool. Sci.* 19, 1329-1335.

Huelsenbeck, J.P., and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* 17, 754-755.

Kawamura, Y. (2007). Last glacial and Holocene land mammals of the Japanese islands : their fauna, extinction and immigration. *The Quaternary Research* 46, 171-177.

Kijas, J.M.H., and Andersson, L. (2001). A phylogenetic study of the domestic pig estimated from the near-complete mtDNA genome. *J. Mol. Evol.* 52, 302-308.

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. *Anim. Genet.* 33, 19-25.

Larson, G., Dobney, K., Albarella, U., Fang, M., Matisoo-Smith, E., Robins, J., Lowden, S., Finlayson, H., Brand, T., Willerslev, E., et al. (2005). Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science* 307, 1618-1621.

Larson, G., Albarella, U., Dobney, K., Rowley-Conwy, P., Schibler, J., Tresset, A., Vigne, J.D., Edwards, C.J., Schlumbaum, A., Dinu, A., et al. (2007). Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *Proc. Natl. Acad. Sci. USA* 104, 15276-15281.

Li, S.J., Yang, S.H., Zhao, S.H., Fan, B., Yu, M., Wang, H.S., Li, M.H., Liu, B., Xiong, T.A., and Li, K. (2004). Genetic diversity analyses of 10 indigenous Chinese pig populations based on 20 microsatellites. *J. Anim. Sci.* 82, 368-374.

Lin, C.S., Sun, Y.L., and Liu, C.Y. (1999). Complete nucleotide sequence of pig (*Sus scrofa*) mitochondrial genome and dating evolutionary divergence within Artiodactyla. *Gene* 236, 107-114.

Naya, Y., Horiuchi, M., Ishiguro, N., and Shinagawa, M. (2003). Bacteriological and genetic assessment of game meat from Japanese wild boars. *J. Agric. Food. Chem.* 51, 345-349.

Nei, M. (1987). *Molecular Evolutionary Genetics*. Columbia University Press.

Norton, C.J. (2000). The current state of Korean paleoanthropology. *J. Hum. Evol.* 38, 803-825.

Okumura, N., Ishiguro, N., Nakano, M., Hirai, K., Matsui, A., and Sahara, M. (1996). Geographic population structure and sequence divergence in the mitochondrial DNA control region of the Japanese wild boar (*Sus scrofa leucomystax*), with reference to those of domestic pigs. *Biochem. Genet.* 34, 179-189.

- Rogers, A. (1995). Genetic evidence for a Pleistocene population explosion. *Evolution* 49, 608-615.
- Ronquist, F., and Huelsenbeck, J.P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572-1574.
- Rozas, J., Sánchez-DelBarrio, J.C., Messeguer, X., and Rozas, R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19, 2496-2497.
- Ruvinsky, A., and Rothschild, M.F. (1998). Systematics and evolution of the pig. In *Handbook of The Genetics of the Pig*, M.F. Rothschild, and A. Ruvinsky, eds. (Oxfordshire, CAB International), pp. 1-16.
- Schneider, S., and Excoffier, L. (1999). Estimation of demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: Application to human mitochondrial DNA. *Genetics* 152, 1079-1089.
- Swofford, D.L. (2003). PAUP\*. Phylogenetic analysis using parsimony (\*and other methods), Version 4. Sinauer Associates.
- Tavare, S. (1986). Some probabilistic and statistical problems on the analysis of DNA sequences. *Lectures. Lect. Math. Life. Sci.* 17, 57-86.
- Thompson, J.D., Higgins, D.G., and Gibson, T.I. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673-4680.
- Watanobe, T., Ishiguro, N., and Nakano, M. (2003). Phylogeography and population structure of the Japanese wild boar *Sus scrofa leucomystax*: mitochondrial DNA variation. *Zoolog. Sci.* 20, 1477-1489.
- Watanobe, T., Ishiguro, N., Nakano, M., Matsui, A., Hongo, H., Yamazaki, K., and Takahashi, O. (2004) Prehistoric Sado Island populations of *Sus scrofa* distinguished from contemporary Japanese wild boar by ancient mitochondrial DNA. *Zoolog. Sci.* 21, 219-228.
- Weir, B.S. (1996). *Genetic data analysis II: methods for discrete population genetic data*. Sinauer Associates.
- Xia, X., and Xie, Z. (2001). DAMBE: software package for data analysis in molecular biology and evolution. *J. Hered.* 92, 371-373.
- Yang, J., Wang, J., Kijas, J., Liu, B., Han, H., Yu, M., Yang, H., Zhao, S., and Li, K. (2003). Genetic diversity present within the near-complete mtDNA genome of 17 breeds of indigenous Chinese pigs. *J. Hered.* 94, 381-385.