

# Multi-Locus Analysis Reveals A Different Pattern of Genetic Diversity for Mitochondrial and Nuclear DNA between Wild and Domestic Pigs in East Asia

Yin-Qiu Ji<sup>1,2</sup>, Dong-Dong Wu<sup>1</sup>, Gui-Sheng Wu<sup>1,3</sup>, Guo-Dong Wang<sup>1</sup>, Ya-Ping Zhang<sup>1,3\*</sup>

**1** State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China, **2** The Graduate School of the Chinese Academy of Sciences, Beijing, China, **3** Laboratory for Conservation and Utilization of Bio-resource, Yunnan University, Kunming, China

## Abstract

**Background:** A major reduction of genetic diversity in mtDNA occurred during the domestication of East Asian pigs. However, the extent to which genetic diversity has been lost in the nuclear genome is uncertain. To reveal levels and patterns of nucleotide diversity and to elucidate the genetic relationships and demographic history of domestic pigs and their ancestors, wild boars, we investigated 14 nuclear markers (including 8 functional genes, 2 pseudogenes and 4 intergenic regions) from 11 different chromosomes in East Asia-wide samples and pooled them with previously obtained mtDNA data for a combined analysis.

**Principal Findings:** The results indicated that domestic pigs and wild boars possess comparable levels of nucleotide diversity across the nuclear genome, which is inconsistent with patterns that have been found in mitochondrial genome.

**Conclusions:** This incongruence between the mtDNA and nuclear genomes is suggestive of a large-scale backcross between male wild boars and female domestic pigs in East Asia. Our data reveal the impacts of founder effects and backcross on the pig genome and help us better understand the complex demographic histories of East Asian pigs, which will be useful for future work on artificial selection.

**Citation:** Ji Y-Q, Wu D-D, Wu G-S, Wang G-D, 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. doi:10.1371/journal.pone.0026416

**Editor:** Arthur J. Lustig, Tulane University Health Sciences Center, United States of America

**Received:** December 15, 2010; **Accepted:** September 26, 2011; **Published:** October 31, 2011

**Copyright:** © 2011 Ji et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by grants from the National Basic Research Program of China (973 Program, 2007CB815704), National program for transgenic research from Ministry of Agriculture of China, National Natural Science Foundation of China (U0836602, 31061160189), Chinese Academy of Sciences, and Bureau of Science and Technology of Yunnan Province. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: zhangyp@mail.kiz.ac.cn

## Introduction

Domestication has been generally seen as a process that starts with a small number of wild individuals, which through succeeding generations of breeding, are integrated into human societies and thereby lose almost all opportunities to mate with their wild relatives [1,2,3,4,5]. In view of this, the general expectation is that founder effects should cause a general loss of genetic diversity in domesticated species relative to their wild progenitors due to a domestication bottleneck, and this has been observed in crop plants [6,7,8]. However, the impacts of domestication on the genetic diversity of domesticated animals, livestock, is largely unknown since most of the putative wild ancestors of livestock are extinct or are being threatened by extinction and are therefore themselves genetically depauperate [9]. One of the few exceptions is the wild boar, the ancestor of domestic pigs. Both wild boars and domestic pigs are widely distributed in the Old World [10], which provides an excellent opportunity to study the impacts of domestication on livestock genetic diversity.

To date, many previous studies have focused on the origin and the distribution histories of domestic pigs [10,11,12,13,14,15,16,17], and future work will focus more and more on identifying

economically important genes under artificial selection. However, existing evidence has implicated artificial selection as another major cause of founder effects, in addition to domestication bottlenecks [7,18,19,20,21,22]. The difference between these two factors is that domestication bottlenecks reduce genetic diversity across the entire genome, while artificial selection is expected to reduce diversity only at selected loci, plus linked regions. This latter pattern has been observed, for example, in maize [7,21]. Therefore, testing for any domestication bottleneck events in the pig genome will aid ongoing efforts to identify economically important loci that have been subject to artificial selection. If there have been genome-wide domestication bottlenecks, merely showing that any particular locus has low diversity will not in itself be sufficient to conclude that selection has acted on that locus.

Some studies have compared genetic diversity between wild boars and domestic pigs using mtDNA and nuclear DNA markers [14,16,23,24,25,26,27], and these studies have found no evidence for loss of genetic diversity, with the exceptions of Wu et al. (2007) and Scandura et al. (2008). These two studies showed that domestic pigs have lower mtDNA diversity than do wild boars in East Asia (Wu et al. 2007) and in Europe (Scandura et al. 2008). On the other hand, data from autosomal microsatellites,

mitochondrial and Y-chromosome polymorphisms in Ramirez et al. (2009) and three nuclear functional genes (*FABP4*, *FABP5* and *IGF2*) in the studies of Ojeda et al. [23,24,25] showed that genetic diversity is similar in domestic pigs and wild boars. This difference among the studies might have resulted from different sampling ranges: Wu's dataset was only from Asia, Scandura's was from Europe, Ramirez et al.'s and Ojeda et al.'s were from a much larger range, a sample of pigs and wild boars with a worldwide distribution.

Several investigations on the origin of pigs [10,11,12,28] have shown that pigs have been independently domesticated multiple times in multiple sites, resulting in clear phylogeographic structures in both wild boars and domestic pigs, with multiple, highly differentiated mtDNA gene pools, which, if pooled for analysis, would result in an overestimation of mtDNA diversity and the loss of any signal of founder effects. For this reason, Ramirez et al. (2009) might have overestimated mtDNA diversity of pigs and showed a different result from Wu et al. (2007) and Scandura et al. (2008).

The disagreement between the reduced mtDNA diversity found in Wu et al. (2007) and Scandura et al. (2008) and failure to observe reduced diversity in nuclear DNA [16,23,25] might also have resulted from differences in size of sampling ranges. The exception is Ojeda et al. (2008) [24], which compared the genetic diversities of *IGF2* across continents and showed no genetic loss.

It is interesting to find such an unexpected difference between mitochondrial and nuclear DNA, which may suggest different evolutionary histories between male and female lineages during pig domestication. But more data are needed because *IGF2* is a possible selection target locus and cannot represent the whole nuclear genome. Here, we provide a study of the levels and patterns of genetic variation in domestic pigs, from 41 breeds, and wild boars, from 14 localities, in East Asia (Figure 1) that is based on multiple markers. We characterized 14 nuclear markers from 11 different chromosomes, including 8 functional genes, 2 pseudogenes, and 4 intergenic sequences, and pooled them with

previously published mtDNA data [14] to compare genetic diversities of domestic pigs and wild boars, to detect any founder effect of pig domestication and to infer any demographic events that have had important impacts on the pig genome.

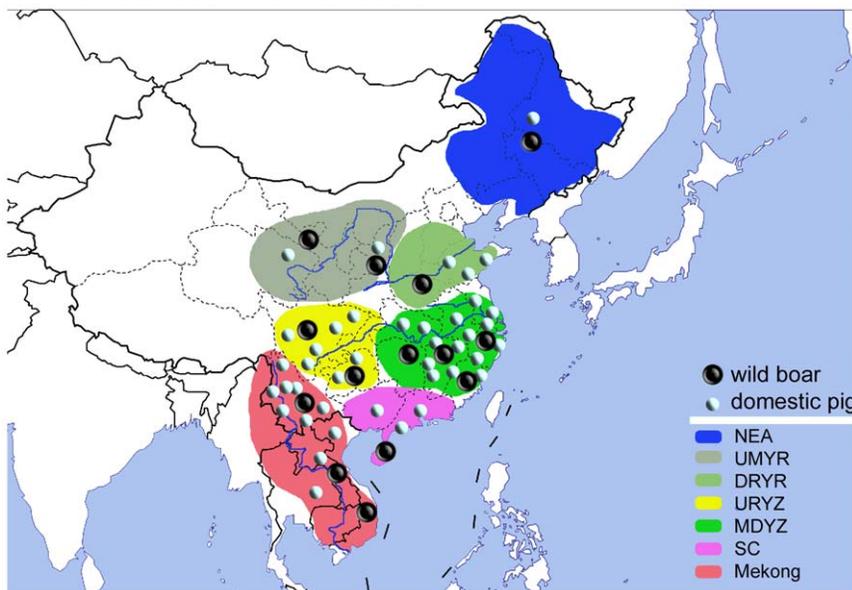
## Materials and Methods

### Samples and markers studied

In this study, our samples were mostly taken from individuals that were collected from the East Asian mainland and had been used in our previous analysis of mtDNA [14]. Wild boars were from 14 geographic sites and domestic pigs were from 41 local breeds (Figure 1).

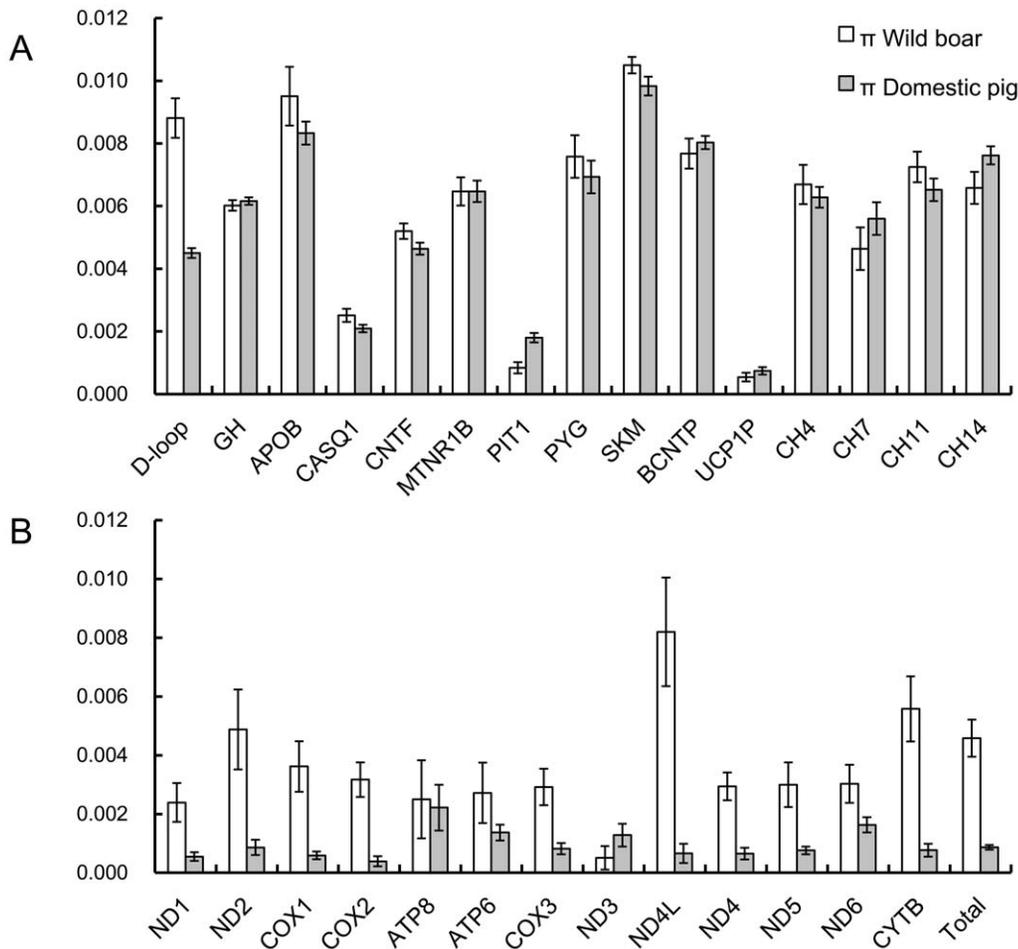
Because the genetic diversity of a subdivided population will be overestimated, we first sampled at least five individuals in most geographic populations or breeds for the gene *GH*, a nuclear marker with a similar sample size to that of the D-loop region of mtDNA. The data from both *GH* and *D-loop* showed that wild boars and domestic pigs in East Asia can each be treated as a single population without clear subdivided structures (Figure S1). Therefore, the following analyses were mainly based on the whole East Asian population of wild boars or domestic pigs, and breed information was not used. For other nuclear markers, only 1 to 3 individuals were sampled in most geographic populations or breeds, since this sample size is enough to provide a  $\pi$  value with a relatively low standard deviation (less than 10%) when genetic diversity of the whole population of wild boars or domestic pigs was estimated (Figure 2A). We endeavored to collect samples from unrelated individuals, using information provided by owners and local farmers. Dataset S1 presents detailed information on the samples.

The markers used in this study represent 14 nuclear regions on 11 different chromosomes in the pig. The reference sequences of these nuclear markers were obtained from GenBank [accession numbers: *GH* - M17704, *APOB* - M22646, *CASQ1* - AJ488283, *CNTF* - U57644, *MTNR1B* - AJ276454, *PIT1* - U00793, *PYG* -



**Figure 1. Geographical and group distribution of the East Asian wild boars and domestic pigs sampled.** All the samples are sorted into seven groups: NEA = Northeast China, UMYR = the upper and middle catchment of the Yellow River, DRYR = the downstream catchment of the Yellow River, URYZ = the upper catchment of the Yangtze River, MDYZ = the middle and downstream catchment of the Yangtze River, SC = South China, and Mekong = the Mekong River catchment.

doi:10.1371/journal.pone.0026416.g001



**Figure 2. Comparison of  $\pi$  between wild boars and domestic pigs.** These comparisons were based on the samples from all East Asia. Error bars indicate one standard deviation of the measurement. (A) Comparison of  $\pi$  in *D-loop* and 14 nuclear markers. (B) Comparison of  $\pi$  in 13 mitochondrial genes and the whole mitochondrial genome. doi:10.1371/journal.pone.0026416.g002

AJ507153, SKM – U23954, BCNTP – AB213482, UCP1P – DQ372918, CH4 – NW\_001886157, CH7 – NW\_001886412, CH11 – NW\_001885204, CH14 – NW\_001885377]. In total, we sequenced 9,859 bp, comprised of 1,872 bp of coding sequence and 7,987 bp of noncoding sequence. Detailed information on the location, functional association [29,30,31,32,33,34,35,36,37], length of each marker, and primer sequences used is in Table 1. To ensure that intergenic sequences were not influenced by neighboring functional regions, we chose markers for which there were no genes for at least 10 kb to either side [27]. Genomic DNA was extracted from whole blood, tissue, or hair by standard phenol/chloroform methods [38].

### PCR and sequencing

PCR was performed in a 50  $\mu$ l reaction mixture containing 50–100 ng genomic DNA, 5  $\mu$ l 10 $\times$  buffer, 8  $\mu$ l 2.5 mM dNTPs, 2  $\mu$ l of 10  $\mu$ M/L each primer, and 2.5 unit of Taq DNA polymerase (TaKaRa Biotech, Dalian, China). Thermocycling was 95°C for 3 min, 35 cycles of 94°C for 30 s, optimal Ta for 30 s (Table 1), and 72°C for 1 to 2 min, and a final extension of 72°C for 5 min. PCR products were purified on spin columns (Watson Biotechnologies, Shanghai) and sequenced on an ABI 3730 automated sequencer (Applied Biosystems), using BigDye v3.1 terminator chemistry (Applied Biosystems). To reduce

potential errors from PCR, we independently amplified and sequenced each individual twice.

Direct sequencing of PCR products revealed that many individuals were heterozygous. To determine haplotypes unambiguously, we used two different methods. For the *GH* gene, PCR products of heterozygous samples were cloned into the PMD 18-T Vector (TaKaRa Biotech, Dalian, China) and transformed into *Escherichia coli* JM109 cells (TaKaRa Biotech, Dalian, China). Plasmids were extracted from positive clones and sequenced using the above primers. Typically, 2 to 6 clones were sequenced for each individual to determine the exact point mutations. For all other markers, haplotypes were identified with the software package PHASE v2.1.1 [39] using default options. Only those haplotypes with high probability ( $P > 0.8$ ) were retained for data analyses. All haplotypes have been submitted to GenBank [accession numbers: EU684330–EU684446, GU328959–GU329035 and GU348654–GU348822]. All genotype information for each sample at each marker is presented in Dataset S1. The numbers of sequences obtained for each marker in wild boars and domestic pigs are listed in Table 2.

### Data analysis

The mtDNA data of Wu et al. (2007) were reanalyzed here because the emphasis before was placed on the origin of

**Table 1.** Summary of the nuclear markers surveyed and the primers used in the study.

Marker	Chromosome Location	Functional Association	Alignment Length (bp)			Primer (5' to 3')	Ta
			Total	Coding	Noncoding		
<i>GH</i>	12	Growth hormone	1,971	648	1323	pGH1F: GTCGACGGGAACAGGAT GAGTGGGAGGAGGTT pGH1R: AAGC TTGCCGGGTCAACCA TCATTCAGTGTCT pGH1-2F: CCGA AGATGCTATCAGG TGAGTGTA pGH1-2R: TTGGAG AAGGACAAAGAGGGAAGA pGH1-3F: TGTTTGGCAC CTCAGACCGC pGH1-3R: GG GTCAACCATTCAGTGTCTA	60°C
<i>APOB</i>	3	Apolipoprotein B	645	173	472	APOBF: GCTTGCCG AGTCCTAAC APOBR: GGCC AGTCTGACTATTCTAGTT	52°C
<i>CASQ1</i>	4	Calsequestrin 1	583	131	452	CASQ1F: CAAAGCCCA GAGATGTTAAGA CASQ1R: GCCAGATCGGGTTAGAAT	55°C
<i>CNTF</i>	2	Ciliary neurotrophic factor	759	114	645	CNTFF: ATGCCAG TGGATTAGTC CNTFR: TGAC AGGCCTTAGTAAAGA	50°C
<i>MTNR1B</i>	9	Melatonin receptor 1B	568	320	248	MTNR1BF: CGCAGGAAGG TCAAGTCGGACAA MTNR1BR: GGCGAACGAGGTGAGCGAGAG	60°C
<i>PIT1</i>	13	POU-domain protein	650	211	439	PIT1F: CAATACAACAT AAACATGAGTAGCGT PIT1R: AGGCTCTGCAAAA GTTACTGATAAGATT	51°C
<i>PYG</i>	2	Glycogen phosphorylase	578	101	477	PYGF: AGCCAGAAA TGCGGACA PYGR: GGTAG CCACTCTGCGGTGAT	57°C
<i>SKM</i>	1	Skeletal muscle calpain	784	174	610	SKMF: CCCAACACACC AAATAACTAGACA SKMR: GCTTGCCAGAAGTGAATGCTA	56°C
<i>BCNTP</i>	1	Bcmt pseudogene for bucentaur	669	0	669	BCNTPF: GGGATGCT ATGGAGGAAT BCNTPR: ATGGGAAACAGTCGATGA	51°C
<i>UCP1P</i>	8	Uncoupling protein 1 pseudogene	635	0	635	UCP1PF: TTCAGATCC AAGGCAAAT UCP1PR: GTCCACCAAGTTCAGTCAAT	52°C
<i>CH4</i>	4	Intergenic sequence	350	0	350	CH4F: TCCGAGCCGTATCTGCAAC CH4R: CAGGGGTTTTCAGGGTTTTATG	55°C
<i>CH7</i>	7	Intergenic sequence	641	0	641	CH7F: GGGAGGGC AGTCACGAAGTC CH7R: TGTA GAGCAGAAATACACGCAACA	55°C
<i>CH11</i>	11	Intergenic sequence	506	0	506	CH11F: TCATGACG GAGTGTACGA CH11R: GCAGCCAATCCAATTAC	51°C
<i>CH14</i>	14	Intergenic sequence	520	0	520	CH14F: GGGCTCTG GGACAGTCTCGT CH14R: TGTC CAGAAGAGGCAAACCCATAA	56°C
<b>Total</b>			<b>9,859</b>	<b>1,872</b>	<b>7,987</b>		

doi:10.1371/journal.pone.0026416.t001

domestication. Sequence data from the 14 nuclear markers were assembled and edited using seqman (DNASTAR, Madison, WI), and were aligned using a combination of methods implemented in ClustalW 1.81 [40] and BioEdit version 5.0.9 [41], with further manual refinements. The alignments are available from the authors upon request.

For the analyses of population structure, we placed our samples into the following groups (Figure 1) according to principles

previously described in [14] (geographic fauna and possible pig domestication sites): (1) Region NEA, Northeast China, including Jilin, Liaoning, Heilongjiang, northeast Inner Mongolia; (2) Region UMYR, the upper and middle catchment of the Yellow River, including Gansu, east Qinghai, northwest Sichuan, south Inner Mongolia, Ningxia, Shaanxi, Shanxi, and west Henan; (3) Region DRYR, the downstream catchment of the Yellow River, including east Henan, Hebei, and Shandong; (4) Region URYZ,

**Table 2.** Summary statistics for the *D-loop* region and all nuclear markers.

Marker	Wild boars									Domestic pigs									$F_{ST}$	$\Phi_{CT}$
	N	S	$\pi$	$\Theta$ (MPE)	$\Theta$ (95%)	D	$D^*$	$F^*$	N	S	$\pi$	$\Theta$ (MPE)	$\Theta$ (95%)	D	$D^*$	$F^*$				
GH	226	78	6.02	16.7	9.6–25.6	–0.36	0.44	0.09	416	65	6.16	6.8	4.6–10.6	0.6	0.23	0.49	0.88	–0.03		
APOB	54	25	9.51	10.9	6.5–17.3	0.38	1.46*	1.28	96	21	8.33	10.8	6.6–17.3	0.92	0.48	0.77	2.82	0.90		
CASQ1	64	9	2.51	3.9	1.8–8.6	–0.62	–0.05	–0.28	108	7	2.09	3.7	1.6–7.0	–0.2	0.3	0.16	6.49	5.74*		
CNTF	62	12	5.2	4.8	1.9–10.8	1.55	0.91	1.33	106	14	4.64	6.5	4.0–10.9	0.86	0.97	1.11	5.11	3.82		
MTNR1B	54	11	6.47	5.3	2.4–10.1	1.47	1.44	1.71 *	96	12	6.47	8.1	4.6–13.8	1.5	1.47	1.76*	1.40	–1.14		
PIT1	58	4	0.84	1.4	0.3–3.4	–0.81	–0.14	–0.4	112	7	1.8	3	1.2–5.7	–0.27	1.18	0.82	4.85	1.24		
PYG	68	19	7.58	12.5	6.5–18.6	0.32	1.27	1.1	102	18	6.93	5.2	3.0–9.2	0.44	1.2	1.1	0.41	–1.20		
SKM	62	20	10.5	4.3	2.6–7.8	2.82**	0.92	1.90*	108	20	9.83	3.6	2.3–6.6	2.89**	0.38	1.59	–0.56	–1.34		
BCNTP	58	14	7.68	3.5	1.6–6.0	2.07*	1.55*	2.04**	112	12	8.03	2.7	1.2–5.3	3.58***	1.46	2.65**	2.79	1.44		
UCP1P	56	4	0.54	1.7	0.4–3.9	–1.34	–0.13	–0.59	94	4	0.74	1.7	0.6–3.7	–0.81	–0.25	–0.5	1.52	–0.98		
CH4	56	11	6.69	9.9	3.7–19.1	–0.06	–0.33	–0.29	202	12	6.28	9.6	3.7–17.4	0.19	1.43	1.16	6.45	5.07		
CH7	56	15	4.64	6.5	3.2–12.1	–0.27	0.63	0.38	104	19	5.6	9.5	5.0–14.9	–0.06	0.78	0.56	0.52	–3.13		
CH11	60	21	7.25	10.7	4.6–21.1	–0.58	–0.16	–0.37	204	17	6.52	10.9	5.1–17.3	0.37	–0.06	0.13	2.30	1.12		
CH14	54	12	6.58	6.4	4.0–11.5	0.87	1.48	1.51	96	15	7.62	10.7	6.9–19.9	0.99	0.51	0.82	1.00	–0.62		
<i>D-loop</i>	106	34	8.81	26.2	21.9–41.4	–0.99	–1.69	–1.67	197	21	4.5	14.4	11.6–19.9	–0.52	0.79	0.33	6.90	6.19***		

N, number of sequences; S, number of SNPs (excluding insertion-deletions);  $\pi$ , nucleotide diversity estimated in DnaSP;  $\theta$ , most probable estimate in LAMARC;  $\theta$ (95%), 95% credible intervals; all  $\pi$  and  $\theta$  are  $\times 10^3$ ; D, Tajima's D-statistic;  $D^*$ , Fu and Li's D-statistic;  $F^*$ , Fu and Li's F-statistic;  $F_{ST}$  and  $\Phi_{CT}$ , genetic differentiation between wild boars and domestic pigs; all  $F_{ST}$  and  $\Phi_{CT}$  are  $\times 10^2$ ;

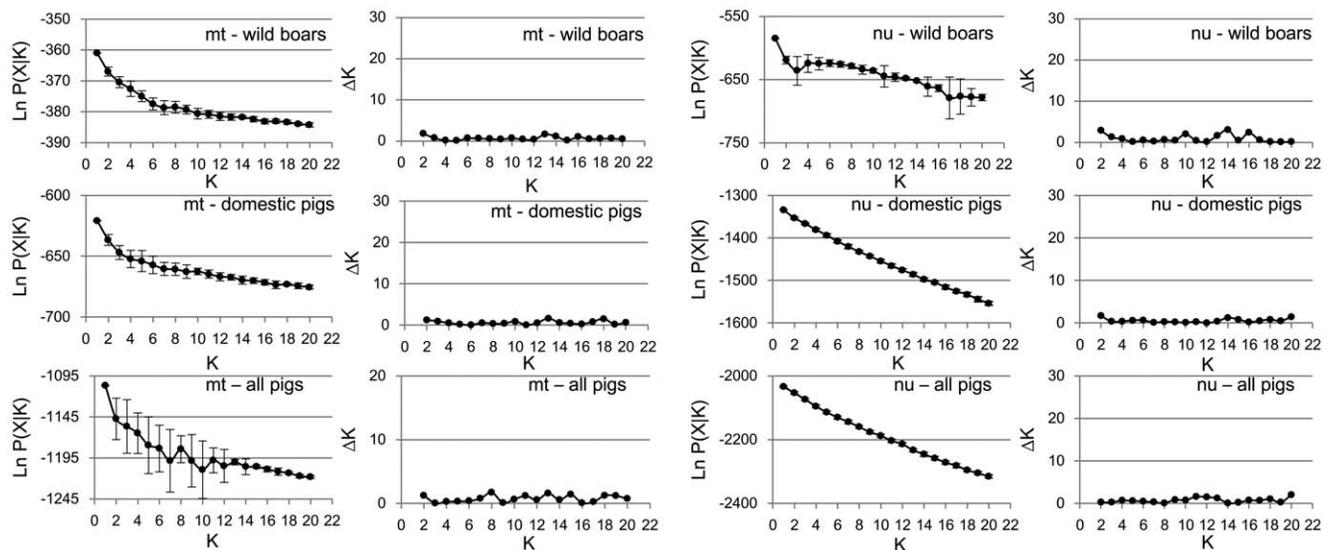
\*,  $P < 0.05$ ;  
 \*\*,  $P < 0.01$ ;  
 \*\*\*,  $P < 0.001$ .

doi:10.1371/journal.pone.0026416.t002

the upstream catchment of the Yangtze River, including Sichuan, Chongqing, Guizhou, northeast Yunnan, west Hubei, and northwest Hunan; (5) Region MDYZ, the middle and downstream catchment of the Yangtze River, including east Hubei, northeast Hunan, Anhui, Jiangxi, Fujian, Zhejiang, Jiangsu, and Shanghai; (6) Region SC, South China, including Guangdong, south and

southeast Guangxi, south Hunan and Hainan; (7) Region Mekong, including northwest, south and southeast Yunnan, Laos, Vietnam and Thailand.

We used Bayesian clustering within STRUCTURE version 2.3.3 [42] to deduce if subpopulations exist in East Asian samples. The *D-loop* region was used to detect mtDNA subdivision and the



**Figure 3. Bayesian clustering results.** Plots of  $\ln P(X|K)$  ( $\pm$ SD) and  $\Delta K$  versus the number of *K* at *D-loop* (mt) or in multi locus (*CH4*, *CH7*, *CH11*, and *CH14*) nuclear DNA (nu) for wild boars, domestic pigs and both combined.

doi:10.1371/journal.pone.0026416.g003

**Table 3.** Some statistics of all the 13 protein-coding mitochondrial genes and the whole mitochondrial genome.

Marker	Wild boars					Domestic pigs				
	$\pi$	$\theta$	$N_A$	$N_S$	$N_A/N_S$	$\pi$	$\theta$	$N_A$	$N_S$	$N_A/N_S$
ND1	2.39	3.21	0	9	0.00	0.55	1.58	1	5	<b>0.20</b>
ND2	4.88	5.89	6	12	0.50	0.86	1.93	5	3	<b>1.67</b>
COX1	3.62	4.64	1	20	0.05	0.59	1.8	4	7	<b>0.57</b>
COX2	3.17	5.95	3	9	0.33	0.39	1.47	1	3	0.33
ATP8	2.5	3.35	2	0	NA	2.22	6.19	4	1	4.00
ATP6	2.72	4.01	3	5	0.60	1.37	2.59	6	1	<b>6.00</b>
COX3	2.92	3.92	4	5	0.80	0.82	1.93	1	5	0.20
ND3	0.51	0.96	1	0	NA	1.28	3.54	3	2	1.50
ND4L	8.2	9.2	2	6	0.33	0.66	1.7	0	2	0.00
ND4	2.94	3.65	12	4	3.00	0.65	1.8	8	2	<b>4.00</b>
ND5	3	4.5	9	15	0.60	0.76	1.8	3	10	0.30
ND6	3.03	4.53	6	1	6.00	1.63	1.91	3	1	3.00
CYTb	5.58	7.79	5	21	0.24	0.77	2.07	5	4	<b>1.25</b>
Total	4.58	4.58	53	105	0.50	0.87	1.98	43	45	<b>0.96</b>

$\pi$  and  $\theta$ , nucleotide diversity estimated in DnaSP and are  $\times 10^3$ ;  $N_A$ , the number of nonsynonymous mutations;  $N_S$ , the number of synonymous mutations; NA, not available.

The  $N_A/N_S$  that were larger in domestic pigs were highlighted in bold.

doi:10.1371/journal.pone.0026416.t003

four intergenic regions combined were used to detect nuclear DNA subdivision. The analyses were performed by using the data of boar-only, pig-only and all samples combined. Structure clusters based on a user-supplied number of populations  $K$  and are given  $\text{LnP}(X/K)$ , the log-likelihood of the posterior probability for each  $K$ , which is used to discern the optimal number of population clusters. We ran 1,000,000 steps after a burn-in of 100,000 under the admixture model without population data. We ran 20 separate iterations for each value of  $K$  from 1 to 20 and used the average value of  $\text{LnP}(X|K)$  to select  $K$  in addition to  $\Delta K$  [43].

In addition, a median-joining network [44] was drawn for each nuclear marker using the program Network 4.5 to visualize phylogeographic structure within East Asia and the relationship between wild boars and domestic pigs (Figure S1).

To apportion the variation between groups (defined above), within and between populations (breeds in domestic pigs and geographic subpopulations in wild boars), and to estimate genetic differentiation ( $\Phi_{CT}$ ) between wild boars and domestic pigs, analyses of molecular variance (AMOVA) [45] were carried out in ARLEQUIN version 3.01 [46] (Dataset S2).

For 14 nuclear markers and the mtDNA D-loop region, we calculated the number of segregating sites, the number of haplotypes, and nucleotide diversity,  $\pi$  (the average number of nucleotide substitutions per site between two sequences) [47]. The level of LD (linkage disequilibrium) was estimated by determining the significance of the associations between all the possible informative nucleotide position pairs within markers with Fisher's exact test after Bonferroni correction. In each marker, the tests of Tajima's  $D$  [48] and  $D^*$  and  $F^*$  of Fu and Li [49] were performed in wild boars and domestic pigs separately. We also calculated the estimator of genetic differentiation ( $F_{ST}$ ) at each marker between wild and domestic pigs. For 8 functional genes,  $N_A$  (the number of nonsynonymous mutations),  $N_S$  (the number of synonymous

mutations) and the ratio of  $N_A/N_S$  were estimated. All the above analyses were carried out in DnaSP 5.00 [50].

To investigate the genetic pattern of the 13 mitochondrial coding genes, we chose all the East Asian samples that had a complete mitochondrial genome sequence in Wu et al. (2007), including 11 wild boars and 30 domestic pigs. Detailed information on samples can be found in Wu et al. (2007). For these 13 mitochondrial coding genes and the whole mitochondrial genome, we calculated  $\pi$ ,  $\theta$  (the proportion of segregating sites) [51],  $N_A$  (the number of nonsynonymous mutations),  $N_S$  (the number of synonymous mutations) and the ratio of  $N_A/N_S$ . All these analyses were carried out in DnaSP 5.00.

We used Modeltest 3.7 [52] and PAUP 4b10 [53] to identify the best-fit model parameters, which were used to simulate a model in LAMARC 2.1.3 [54], using Bayesian MCMC analyses with 10,000 recorded genealogies sampled every 20 steps with a burn-in of 1,000 genealogies, to estimate the neutral parameters  $\theta$  ( $4N_e\mu$  for autosomal markers and  $2N_e\mu$  for mtDNA,  $N_e$ = effective population size of females and  $\mu$ = mutation rate), the migration rates  $M$  ( $m/\mu$ ) from wild boars into domestic pigs, and the recombination rates  $r$  ( $\rho/\mu$ ). Analyses were repeated three times to verify that parameters estimated converged within and among runs. The migration rate  $M$  ( $m/\mu$ ) was multiplied by  $\theta$  ( $4N_e\mu/2N_e\mu$ ) of domestic pigs to calculate  $M_N$  ( $4N_e m/2N_e\mu$ , independence from mutation rate), the average number of effective migrants per generation.

We calculated allele frequency of each SNP that we found in the 14 nuclear markers and used a chi-square test to examine the significance of the allele frequency changes in all the SNP sites between wild boars and domestic pigs. In addition, we calculated and compared LD ratios (the proportion of significant pairwise comparisons to total possible pairwise comparisons) between wild boars and domestic pigs.

## Results and Discussion

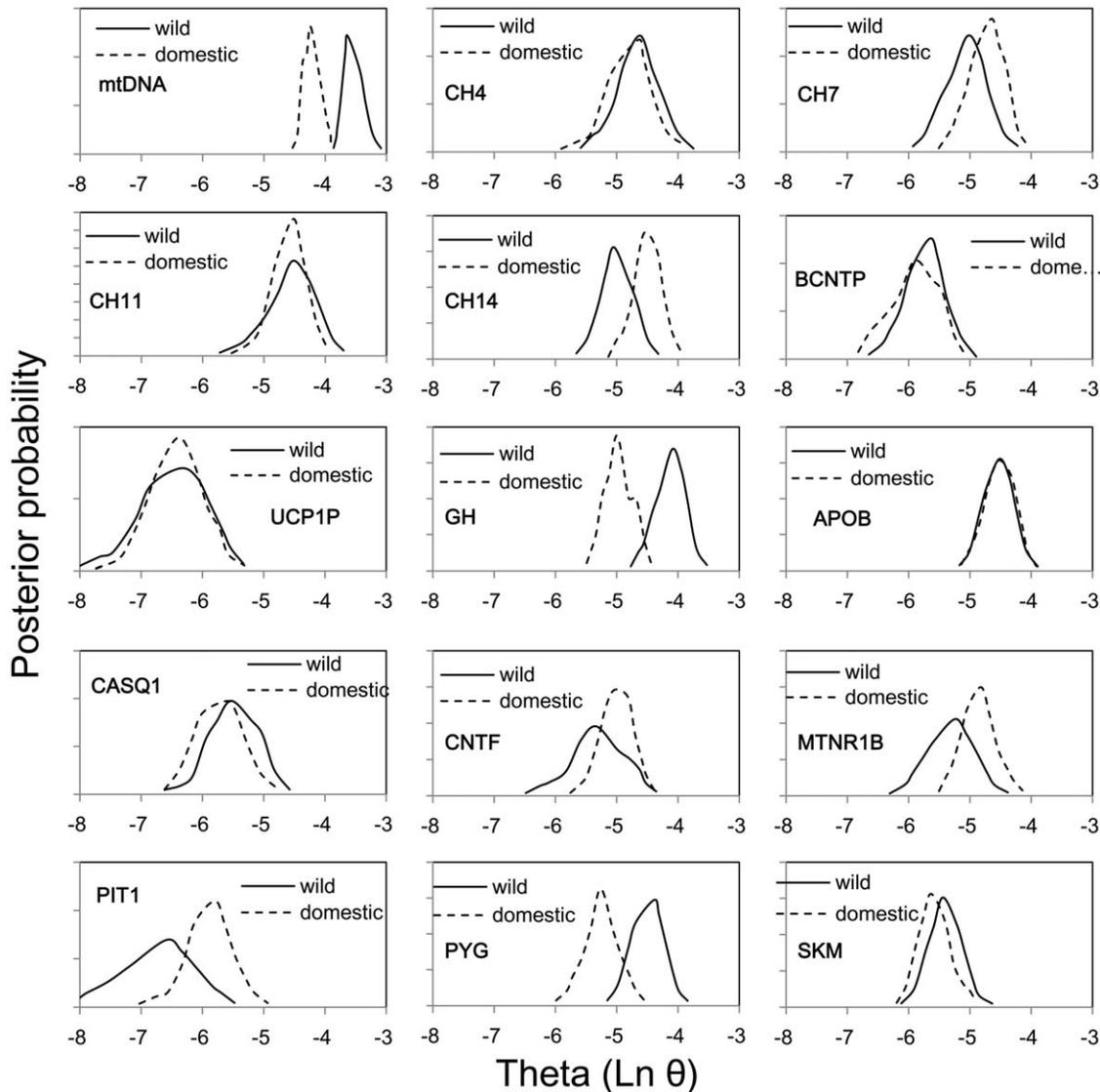
### Population structure

Before estimating and comparing genetic diversity between wild boars and domestic pigs, we first investigated population structure within East Asia. If the population is subdivided, all diversity analyses should be performed within each subpopulation.

To start, we used the clustering method in STRUCTURE [42] to deduce the optimal number of subpopulations ( $K$ ) in East Asian samples. We conducted the analyses for wild boars and domestic pigs separately in addition to all samples combined since the hierarchical structure of wild boars or domestic pigs might not be detected if they both were combined into a single analysis. Because selection might confound the detection of substructure, only *D-loop* and the four nuclear intergenic markers that are considered to be neutral were used. The results are shown in Figure 3. For all the structure analyses, the  $K$  with the highest  $\text{LnP}(X|K)$  and lowest standard deviation was 1. Although the method of  $\Delta K$  [43] did not give the corresponding value of  $K = 1$ , all the plots of  $\Delta K$  versus  $K$  show no peak from  $K = 2$  to 20, further suggesting that 1 is the correct  $K$  number. The results of structure analyses not only indicated no substructure in either wild boars or domestic pigs but also showed a very high level of admixture between boar and pig.

Furthermore, the networks of all these markers do not show a clear phylogeographic structure in either wild boars or domestic pigs of East Asia (Figure S1), which is unlike the strong structure of pig mitochondrial sequences globally [10], probably resulting from high gene flow among different populations in the East Asian mainland.

To further investigate population structure in East Asia, we performed AMOVA analyses on all markers in wild boars and



**Figure 4. Comparison of  $\theta$  in *D-loop* and 14 nuclear markers between wild boars and domestic pigs.** The values of  $\theta$  ( $4N_e\mu$  for autosomal markers and  $2N_e\mu$  for mtDNA,  $N_e$ =effective population size of females and  $\mu$ =mutation rate) were calculated by using Lamarc analysis. doi:10.1371/journal.pone.0026416.g004

domestic pigs separately. The results are in Dataset S2 and show that for all the markers analyzed, most genetic variation derives from the difference within populations in both wild boars and domestic pigs. In wild boars, only three genes (*APOB*, *CNTF* and *PIT1*) show significant genetic differentiation among groups ( $P < 0.05$ ), whereas in domestic pigs, only *D-loop* does so, which might be the result of subdivision since the *D-loop* region is usually considered a neutral marker. Overall, most AMOVA results show low genetic differentiation among East Asian groups, which is consistent with the results of STRUCTURE and the network analysis. Consequently, the following analyses were based on the whole sampled population of East Asian wild boars or domestic pigs.

#### Comparison of genetic diversity between wild boars and domestic pigs

The loss of genetic diversity due to founder effects is considered to be a general consequence of domestication, so the most straightforward strategy for testing for the effect of domestication is

to compare nucleotide diversity between domestic species and their wild progenitors.

For *D-loop* and each nuclear marker, estimated nucleotide diversity ( $\pi$ ) is listed in Table 2. For the 13 mtDNA coding genes and the whole mitochondrial genome,  $\pi$  is listed in Table 3. A summary comparison of  $\pi$  values between wild boars and domestic pigs is shown in Figure 2. *D-loop* and most mtDNA coding genes (*ND1*, *ND2*, *COX1*, *COX2*, *COX3*, *ND4L*, *ND4*, *ND5*, *ND6*, and *CYTb*) confirm that domestic pigs have a clearly lower level of genetic diversity than do wild boars, but nuclear DNA data reveal that domestic pigs and wild boars exhibit comparable levels of diversity in all the markers except *PIT1*, which shows higher diversity in domestic pigs than in wild boars.

Given that the demographic histories of wild and domestic pigs are quite complex, we also used Bayesian MCMC analysis, which has the advantage of explicitly handling uncertainty in parameter estimates, implemented in LAMARC, to estimate genetic diversities  $\theta$  at *D-loop* and at each nuclear marker. The  $\theta$  values are listed in

**Table 4.** Linkage disequilibrium.

Marker	Wild boars			Domestic pigs		
	P <sub>s</sub> <sup>a</sup>	P <sub>T</sub> <sup>b</sup>	ratio <sup>c</sup>	P <sub>s</sub> <sup>a</sup>	P <sub>T</sub> <sup>b</sup>	ratio <sup>c</sup>
GH	220	2926	7.52%	439	2080	<b>21.11%</b>
APOB	47	300	15.67%	49	210	<b>23.33%</b>
CASQ1	1	36	2.78%	1	21	<b>4.76%</b>
CNTF	15	66	22.73%	14	91	15.38%
MTNR1B	11	55	20%	16	66	<b>24.24%</b>
PIT1	1	6	16.67%	2	21	9.52%
PYG	31	171	18.13%	37	153	<b>24.18%</b>
SKM	119	190	62.63%	119	190	62.63%
BCNTP	51	91	56.04%	66	66	<b>100%</b>
UCP1P	0	6	0%	0	6	0%
CH4	3	55	5.45%	16	66	<b>24.24%</b>
CH7	21	105	20%	38	171	<b>22.22%</b>
CH11	8	210	3.81%	42	136	<b>10.88%</b>
CH14	12	66	18.18%	22	105	<b>20.95%</b>
D-loop	49	528	9.28%	9	210	4.29%

<sup>a</sup>P<sub>s</sub>, number of significant pairwise comparisons by Fisher's exact test after Bonferroni correction.

<sup>b</sup>P<sub>T</sub>, number of all possible pairwise comparisons.

<sup>c</sup>ratio = P<sub>s</sub>/P<sub>T</sub>.

The ratios that were greater in domestic pigs were highlighted in bold.

doi:10.1371/journal.pone.0026416.t004

Table 2, and comparisons between wild and domestic pigs are shown in Figure 4. LAMARC analyses show that the  $\theta$  of *D-loop* in wild boars is higher than that in domestic pigs without overlap in the 95% confidence interval (CI), but overlapping CIs are observed in all the nuclear markers. Based on the most probable estimates (MPEs) of  $\theta$ , domestic pigs have retained 55% ( $\theta = 0.0144$ ) of the mtDNA diversity present in wild boars ( $\theta = 0.0262$ ) but have a similar or slightly higher diversity than wild boars for most nuclear markers, which is consistent with the data from  $\pi$  (Figure 2A).

Overall, the mtDNA data confirm previous results [14] in providing strong evidence for the loss of mtDNA genetic diversity in domestic pigs, which could result from the founder effects of domestication bottlenecks and/or from artificial selection. However, in the nuclear genome, regardless of whether we analyzed intergenic markers, pseudogenes, or functional genes, all but one (*PIT1*) reveal that domestic pigs and wild boars have indistinguishable levels of nuclear genetic diversity.

#### Founder effects of domestication in nuclear DNA

Even though reduced genetic diversity was not detected in nuclear DNA, data from low-frequency alleles nonetheless supports the persistence of at least some founder effects. Our analyses found that the 131 SNP alleles (excluding indels) with frequencies <10% were detected across all the nuclear markers in wild boars, but only 83 were present in domestic pigs (Dataset S3), a 36.6% loss of low-frequency alleles in domestic pigs. Meanwhile, Tajima's D appears to be higher in domestic pigs than in wild boars for most markers (Table 2), which is also consistent with the loss of low-frequency alleles. Out of all 293 SNPs (excluding indels), 102 (34.8%) had significantly different frequency distributions across pig and boar (Dataset S3). In addition, 10 of 14 nuclear markers in domestic pigs were detected to have a higher proportion of pairs with significant LD than in wild boars (Table 4),

suggesting a slight trend toward increasing LD in nuclear DNA during domestication or breeding, which is consistent with founder effects.

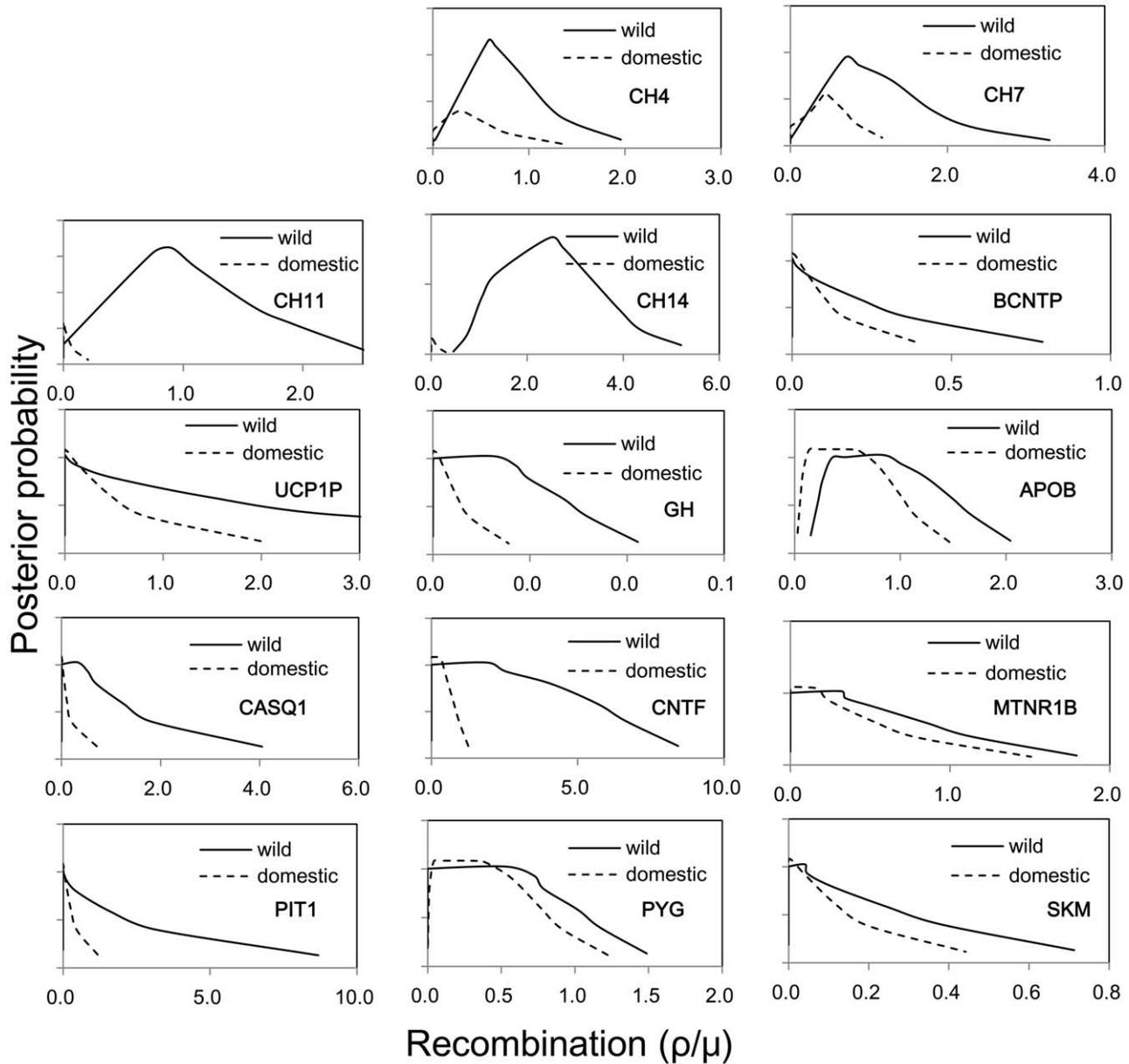
#### What caused the incongruence between mtDNA and nuclear DNA?

If only the nuclear data are considered, the observation of no loss in overall genetic diversity in domestic pigs relative to wild boars could be explained by post-domestication bottleneck events in wild boars, such as widespread hunting. It is plausible that such bottlenecks have occurred in East Asian populations of wild boars, but such bottlenecks cannot explain the inconsistency between mtDNA and nuclear DNA. A bottleneck event will affect mitochondrial and nuclear genomes simultaneously and should in fact affect the mitochondrial genome more strongly because of its smaller effective population size. Thus, what has caused the incongruence between mtDNA and nuclear DNA?

There are three major factors influencing genetic diversity: recombination, selection and demographic events. Firstly, to detect if recombination has increased nuclear genetic diversity in domestic pigs, we calculated the recombination rates ( $r$ ) of each nuclear marker by using Bayesian MCMC analysis implemented in LAMARC and compared  $r$  between wild boars and domestic pigs. We did not calculate the  $r$  of *D-loop* because mtDNA is almost entirely inherited from one parent only. The results showed that 95% CIs of recombination rates overlap between wild and domestic pigs for all comparisons (Figure 5), indicating no elevated recombination rate in the nuclear genome of domestic pigs relative to that of wild boars. This suggests that recombination is not a likely explanation for the similar genetic diversity between pigs and boars.

Selection could also potentially explain the incongruence between mitochondrial and nuclear DNA in two ways. Since mtDNA is inherited without recombination, selection at protein-coding mtDNA loci may reduce genetic diversity across the entire mtDNA genome. However, the samples sequenced for the whole mitochondrial genome were chosen based on the data of the *D-loop* region to avoid the same haplotype between individuals, which might affect the results of statistical methods that capture the information about the frequency spectrum of alleles. Because of this, we only performed the tests of Tajima's  $D$  [48] and  $D^*$  and  $F^*$  of Fu and Li [49] for *D-loop*, and we calculated and compared the ratios of  $N_A/N_S$  at each mitochondrial gene and the whole mitochondrial genome. The results of tests (Table 2) showed no significant departures from neutral expectation in either wild or domestic pigs. From this, we could not determine if mtDNA was under diversity-reducing selection because signatures of selection are always confounded by the effects of historical demographic factors. Domestication is a complicated process including many demographic factors, which could blur the signal of selection. However, comparisons of the  $N_A/N_S$  ratio, which is robust to demographic influences, showed that domestic pigs have a higher ratio than wild boars in the total genome and in over half of comparable genes (6/11 = 54.55%) (Table 3), reflecting a general relaxation of selection on the mitochondrial genome of domestic pigs, similar to that observed in dog domestication [55]. By itself, reduced selective pressure can result in higher genetic diversity. In addition, the low value of LD in domestic pigs (Table 4) excludes the hypothesis of a selective sweep in the *D-loop* region. Thus, on balance, the available data do not suggest that selection has led to reduced mtDNA diversity in domestic pigs, and if anything, reduced selective pressure has resulted in the opposite pattern.

Another hypothesis is that artificial selection has increased the genetic diversity of domestic pigs at the nuclear markers used in



**Figure 5. LAMARC analysis showing the comparison of recombination estimates between wild and domestic pigs.** Recombination rates were calculated and compared at all the 14 nuclear markers. doi:10.1371/journal.pone.0026416.g005

this study. First, we performed the tests of Tajima's  $D$  [48] and Fu and Li's  $D^*$  and  $F^*$  [49] for each nuclear marker, and we found that most markers do not show significant departures from neutral expectations in either wild boars or domestic pigs (Table 2). The four exceptions are three functional genes (*APOB*, *MTNR1B* and *SKM*) and one pseudogene (*BCNTP*). Two of these loci, *SKM* and *BCNTP*, are located on the same chromosome (Table 1) and present similar statistics (Table 2). To elucidate the relationship between *SKM* and *BCNTP*, we conducted a NCBI BLAST search in the pig genome database and found that both markers are located in the same genomic contig [GenBank: NW\_001885768], indicating that these two markers are closely linked and can be influenced together. Although these four markers were found to have significant departures from neutrality, their diversity was not necessarily influenced by artificial selection. At *APOB*, only wild

boars were found to have significant departure, and at the other three loci, both domestic pigs and wild boars show similar departures from neutrality. We then calculated and compared the ratios of  $N_A/N_S$  at the 8 functional nuclear genes. The results are different from mtDNA in that domestic pigs had a lower level of  $N_A/N_S$  than did wild boars in the only 2 genes that showed difference in ratio between wild and domestic pigs (Table 5). This suggests that there has been no relaxation of selection in nuclear DNA in domestic pigs and that there might even be a history of selection at some nuclear loci, which should decrease domestic pig genetic diversity. In addition, most nuclear markers, including most functional genes, pseudogenes, and all intergenic regions from 11 different chromosomes exhibit the same incongruence with the mitochondrial DNA, strongly suggesting that the entire nuclear genome in domestic pigs exhibits high genetic diversity,

**Table 5.** Comparison of ratio  $N_A/N_S$  between wild and domestic pigs at nuclear functional genes.

Marker	Wild boars			Domestic pigs		
	$N_A$	$N_S$	$N_A/N_S$	$N_A$	$N_S$	$N_A/N_S$
<i>GH</i>	9	11	0.82	6	11	0.55
<i>APOB</i>	6	3	2	5	3	1.67
<i>CASQ1</i>	0	0	NA	0	0	NA
<i>CNTF</i>	0	1	0	0	1	0
<i>MTNR1B</i>	0	3	0	0	3	0
<i>PIT1</i>	0	0	NA	1	0	NA
<i>PYG</i>	2	0	NA	2	0	NA
<i>SKM</i>	0	0	NA	0	0	NA
<i>Total</i>	17	18	0.94	14	18	0.78

NA, not available.

doi:10.1371/journal.pone.0026416.t005

which is unlikely to be explained by selection. Furthermore, all the intergenic region markers do not exhibit high levels of LD, suggesting that they are not influenced by selection acting on their neighboring functional regions and are therefore neutral. Thus, we reject the hypothesis of selection increasing genetic diversity in domestic pigs.

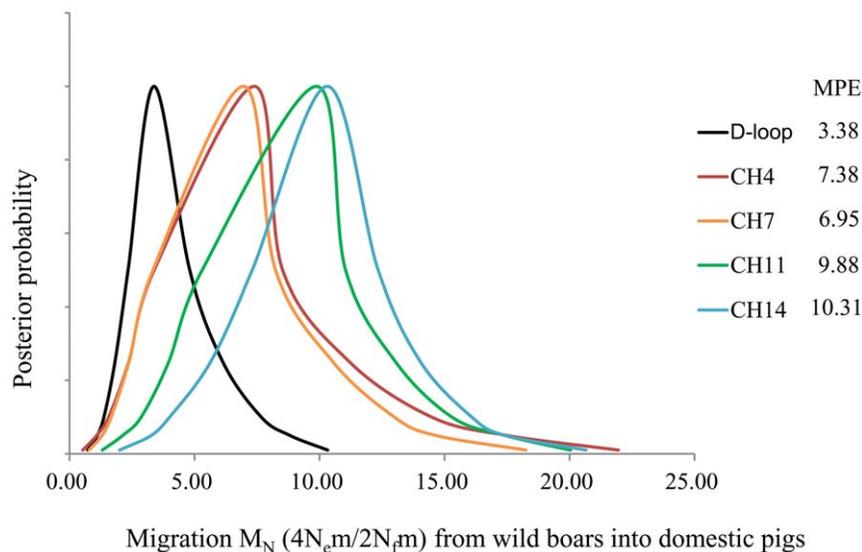
We therefore lean toward accepting the last factor, demographic events, to explain the incongruence between mtDNA and nuclear DNA. Both population subdivision and genetic introgression can, under some conditions, increase total population diversity. If population subdivision is the cause, we would expect DNA regions with higher mutation rates to have a higher level of polymorphisms, due to genetic drift [56]. Given that mtDNA usually evolves faster than nuclear DNA [57,58] and that the AMOVA analyses showed that mtDNA had a higher degree of subdivision than nuclear DNA in domestic pigs, the estimated mitochondrial

diversity of domestic pigs should be higher than that of the nuclear genome. However, the opposite is observed in our data. Therefore, we infer that genetic introgression is the main cause of incongruence between mtDNA and nuclear DNA. Inputs of genetic diversity could either originate from other centers of domestication or from wild boars. We lean toward wild boars because there is no evidence of introgression in East Asian local breeds from other centers based on the microsatellite data [16,59].

To further evaluate the hypothesis of genetic introgression from wild boars to domestic pigs, we estimated migration for the mtDNA D-loop region and for four nuclear intergenic regions that are considered to be neutral and therefore free of effects of selection. The results showed that the confidence intervals of  $M_N$  overlapped for all these regions, but the MPEs of nuclear regions were at least 2 times that of mtDNA (Figure 6), indicating that the number of effective male migrants was similar or higher than that of female migrants per generation. And if independent from effective population size ( $m = M_N/4N_e$ ), the true migration rate of nuclear DNA is probably higher and the male migrants greater because there has been a sex bias ( $N_f > N_m$ ) in domestic pigs in East Asia for thousands of years [48] and the effective population size  $N_e$  should be smaller than  $2N_f$  [49]. Consequently, it is highly possible for males to have contributed more to gene flow from wild into domestic pigs.

### The backcross hypothesis

Backcross is defined here as post-domestication introgression with sex bias. Vila et al. [60] previously proposed that a high level of diversity in nuclear DNA (but not in mtDNA) resulted from mating between wild progenitor males and domesticated females. However, Vila et al.'s analysis was based on MHC, the high diversity of which is maintained by balancing selection. Such markers might therefore overestimate the founder number of domestic pigs because in their simulations, the maximum number of founder populations tested was six, which might underestimate the real number of origins of pig domestication [10,15,17]. Our study provides more and stronger evidence in favor of Vila et al.'s

**Figure 6.** LAMARC analysis showing the migration estimates from wild boars into domestic pigs. The migration estimates were only calculated at 4 intergenic regions and the mtDNA D-loop region.

doi:10.1371/journal.pone.0026416.g006

proposal of backcross: 1) higher diversity in nuclear DNA relative to mtDNA; 2) shared haplotypes, showing that domestic pig samples from Northeast Asia (NEA), the Yellow River catchment (UMYR and DRYR) and the upstream catchment of the Yangtze River (URYZ) share some haplotypes with the local wild boars in nuclear DNA (Figure S1), but none of them share even the same haplogroup in mtDNA [14]; 3) analyses of migration rates, showing that the male lineage highly possibly has a greater gene flow than female lineage from wild into domestic pigs, which is inconsistent with the observation that the domesticated pig populations typically retain more females than males for production [61,62]. Furthermore, the hypothesis of backcross is consistent with the observation that pig husbandry in China historically allowed pigs to range freely, and this custom persists in some regions today.

## Conclusions

Although backcrossing has been demonstrated for some domestic species [63,64], the extent of its impacts to the gene pool of modern livestock was uncertain. The main objective in this study is to investigate its impacts on the domestic pig genome in addition to supporting its existence in pigs. Our data are consistent with the hypothesis that backcross events have increased the diversity of nuclear DNA in East Asian domestic pigs, resulting in a different genetic pattern between male and female lineages. Therefore, any founder effects of a domestication bottleneck have been weakened, resulting in no observed loss of overall nuclear genetic diversity. In addition, the high diversity with a very weak population structure that is observed in nuclear DNA of domestic pigs is expected when most polymorphisms are from wild boars through backcross and not from artificial selection or subdivision in breeding. In conclusion, our data help us better understand the complex demographic histories of East Asian pigs, and bring a practical benefit for the future work on artificial selection. High nuclear genetic diversity increases statistical power in the search for genes that have been subject to artificial selection during domestication, since there is less need to consider the alternative hypothesis of founder effects as a cause of lower diversity in candidate loci for artificial selection.

## Supporting Information

**Figure S1** The network of 14 nuclear markers and D-loop in domestic pigs and wild boars of East Asia. These samples are from Northeast Asia (NEA), the upper and middle catchment of the Yellow River (UMYR), the downstream catchment of the Yellow River (DRYR), the upstream catchment of the Yangtze River

(URYZ), the middle and downstream catchment of the Yangtze River (MDYZ), the Mekong catchment and South China (SC). Each haplotype is represented by a circle, with the area of the circle proportional to its frequency. Samples from different regions were indicated by different colours. The length of each branch was proportional to the number of mutations on the respective branch. The network of *D-loop* was cited from Wu et al. (2007).

(PDF)

**Dataset S1** Sample information. List of the detail information of samples used in this study. The samples of domestic pigs were named by their breeds, following the number code. For each individual, haplotype/genotype is given at each marker. Furthermore, in the mtDNA D-loop region, the clade in Larson et al. (2010) that each individual belongs to is also given.

(XLS)

**Dataset S2** AMOVA results. This file listed the AMOVA results of wild boars and domestic pigs. All samples are classified into 7 geographic groups, which have been defined in materials and methods. Populations are divided based on sampling plot or breed. The P-values of  $\Phi_{CT}$  lower than 0.05 are emphasized by yellow background.

(XLS)

**Dataset S3** The allele frequency of every SNP site in wild boar and domestic pig. This file listed and compared the allele frequency of every SNP site between wild boar and domestic pig. The alleles which have a frequency lower than 0.05 in wild boars were highlighted in yellow. The low-frequency alleles which were not present in domestic pig were highlighted in red.  $\chi^2$  test was performed to examine the significance of allele frequency changes in all SNPs, and the changes that were significant were highlighted in purple.

(XLS)

## Acknowledgements

We thank Drs. Shan-Yuan Chen, Yong-Gang Yao, Peng Shi and Li Yu for helpful suggestions and comments. We thank Dr. Douglas W. Yu for help in writing. We thank Yong-Yi Shen for his help in experiment. We thank Felicia Yap Chai Lee for her help in sample collection. We also thank anonymous reviewers for their helpful comments.

## Author Contributions

Conceived and designed the experiments: Y-QJ Y-PZ. Performed the experiments: Y-QJ G-SW G-DW. Analyzed the data: Y-QJ D-DW Y-PZ. Contributed reagents/materials/analysis tools: Y-QJ D-DW Y-PZ. Wrote the paper: Y-QJ Y-PZ.

## References

- Zeuner FE (1963) A History of Domesticated Animals. New York: Harper & Row.
- Darwin C (1968) Variation of animals and plants under domestication. London: John Murray.
- Clutton-Brock J (1989) The Walking Larder: Patterns of Domestication, Pastoralism, and Predation London: Unwin Hyman.
- Budiandy S (1992) The Covenant of the Wild: Why Animals Chose Domestication. New Haven: Yale University Press.
- Darwin C (1868) Variation of animals and plants under domestication. London: John Murray.
- Eyre-Walker A, Gaut RL, Hilton H, Feldman DL, Gaut BS (1998) Investigation of the bottleneck leading to the domestication of maize. Proc Natl Acad Sci U S A 95: 4441–4446.
- Tenaillon MI, U'Ren J, Tenaillon O, Gaut BS (2004) Selection versus demography: a multilocus investigation of the domestication process in maize. Mol Biol Evol 21: 1214–1225.
- Tanksley SD, McCouch SR (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. Science 277: 1063–1066.
- Scherf BD (1995) World Watch List for Domestic Animal Diversity, 2nd Edition. Rome: FAO.
- Larson G, Dobney K, Albarella U, Fang M, Matisoo-Smith E, et al. (2005) Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. Science 307: 1618–1621.
- Giuffra E, Kijas JM, Amarger V, Carlborg O, Jeon JT, et al. (2000) The origin of the domestic pig: independent domestication and subsequent introgression. Genetics 154: 1785–1791.
- Kijas JM, Andersson L (2001) A phylogenetic study of the origin of the domestic pig estimated from the near-complete mtDNA genome. J Mol Evol 52: 302–308.
- Lum JK, McIntyre JK, Greger DL, Huffman KW, Vilar MG (2006) Recent Southeast Asian domestication and Lapita dispersal of sacred male pseudohermaphroditic “tuskers” and hairless pigs of Vanuatu. Proc Natl Acad Sci U S A 103: 17190–17195.
- Wu GS, Yao YG, Qu KX, Ding ZL, Li H, et al. (2007) Population phylogenomic analysis of mitochondrial DNA in wild boars and domestic pigs revealed multiple domestication events in East Asia. Genome Biol 8: R245.

15. Larson G, Albarella U, Dobney K, Rowley-Conwy P, Schibler J, et al. (2007) Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *Proc Natl Acad Sci U S A* 104: 15276–15281.
16. Ramirez O, Ojeda A, Tomas A, Gallardo D, Huang LS, et al. (2009) Integrating Y-chromosome, mitochondrial, and autosomal data to analyze the origin of pig breeds. *Mol Biol Evol* 26: 2061–2072.
17. Larson G, Liu R, Zhao X, Yuan J, Fuller D, et al. (2010) Patterns of East Asian pig domestication, migration, and turnover revealed by modern and ancient DNA. *Proc Natl Acad Sci U S A* 107: 7686–7691.
18. Fang M, Larson G, Ribeiro HS, Li N, Andersson L (2009) Contrasting mode of evolution at a coat color locus in wild and domestic pigs. *PLoS Genet* 5: e1000341.
19. Li J, Yang H, Li JR, Li HP, Ning T, et al. (2010) Artificial selection of the melanocortin receptor 1 gene in Chinese domestic pigs during domestication. *Heredity*.
20. Sutter NB, Bustamante CD, Chase K, Gray MM, Zhao K, et al. (2007) A single IGF1 allele is a major determinant of small size in dogs. *Science* 316: 112–115.
21. Clark RM, Linton E, Messing J, Doebley JF (2004) Pattern of diversity in the genomic region near the maize domestication gene *tb1*. *Proc Natl Acad Sci U S A* 101: 700–707.
22. Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, et al. (2005) The effects of artificial selection on the maize genome. *Science* 308: 1310–1314.
23. Ojeda A, Rozas J, Folch JM, Perez-Enciso M (2006) Unexpected high polymorphism at *FABP4* gene unveils a complex history for pig populations. *Genetics* 174: 2119–2127.
24. Ojeda A, Wang L, Ren J, Angiolillo A, Cho IC, et al. (2008) Selection in the making: A Worldwide Survey of Haplotypic Diversity around a Causative Mutation in Porcine *IGF2*. *Genetics* 178: 1639–1652.
25. Ojeda A, Estelle J, Folch JM, Perez-Enciso M (2008) Nucleotide variability and linkage disequilibrium patterns at the porcine *FABP5* gene. *Anim Genet* 39: 468–473.
26. Scandura M, Iacolina L, Crestanello B, Pecchioli E, Di Benedetto MF, et al. (2008) Ancient vs. recent processes as factors shaping the genetic variation of the European wild boar: are the effects of the last glaciation still detectable? *Mol Ecol* 17: 1745–1762.
27. Amaral AJ, Megens HJ, Crooijmans RP, Heuven HC, Groenen MA (2008) Linkage disequilibrium decay and haplotype block structure in the pig. *Genetics* 179: 569–579.
28. Fang M, Andersson L (2006) Mitochondrial diversity in European and Chinese pigs is consistent with population expansions that occurred prior to domestication. *Proc Biol Sci* 273: 1803–1810.
29. Vize PD, Wells JR (1987) Isolation and characterization of the porcine growth hormone gene. *Gene* 55: 339–344.
30. Maeda N, Ebert DL, Doers TM, Newman M, Hasler-Rapacz J, et al. (1988) Molecular genetics of the apolipoprotein B gene in pigs in relation to atherosclerosis. *Gene* 70: 213–229.
31. Yu TP, Schmitz CB, Rothschild MF, Tuggle CK (1994) Expression pattern, genomic cloning and RFLP analyses of the swine *PIT-1* gene. *Anim Genet* 25: 229–233.
32. Briley GP, Riggs PK, Womack JE, Hancock DL, Bidwell CA (1996) Chromosomal localization of the porcine skeletal muscle calpain gene. *Mamm Genome* 7: 226–228.
33. Weaks RL, Ramsoondar JJ, Gallagher DS, Jr., Noguez C, Piedrahita JA (1997) Isolation, characterization and chromosomal localization of the porcine ciliary neurotrophic factor (*CNTF*) gene. *Anim Genet* 28: 354–357.
34. Skinner TM, Lopez-Corrales NL, Anderson SI, Loudon AL, Haley CS, et al. (2000) Genetic and physical mapping of the porcine melatonin receptor 1B gene (*MTNR1B*) to chromosome 9. *Cytogenet Cell Genet* 90: 53–55.
35. Knoll A, Stratil A, Reiner G, Peelman IJ, Van Poucke M, et al. (2002) Linkage and radiation hybrid mapping of the porcine caseinogenase 1 (*CASQ1*) gene to chromosome 4q. *Anim Genet* 33: 390–392.
36. te Pas MF, Leenhouders JI, Knol EF, Booi M, Priem J, et al. (2003) Marker polymorphisms in the porcine genes for muscle glycogen synthase (*GYS1*) and muscle glycogen phosphorylase (*PYGM*). *Anim Genet* 34: 157–158.
37. Berg F, Gustafson U, Andersson L (2006) The uncoupling protein 1 gene (*UCP1*) is disrupted in the pig lineage: a genetic explanation for poor thermoregulation in piglets. *PLoS Genet* 2: e129.
38. Blin N, Stafford DW (1976) A general method for isolation of high molecular weight DNA from eukaryotes. *Nucleic Acids Res* 3: 2303–2308.
39. Li N, Stephens M (2003) Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics* 165: 2213–2233.
40. Thompson JD, Higgins DG, Gibson TJ (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.
41. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* 41: 95–98.
42. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
43. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14: 2611–2620.
44. Bandelt HJ, Forster P, Rohlf A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16: 37–48.
45. Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479–491.
46. Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online* 1: 47–50.
47. Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105: 437–460.
48. Tajima F (1989) The effect of change in population size on DNA polymorphism. *Genetics* 123: 597–601.
49. Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics* 133: 693–709.
50. Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19: 2496–2497.
51. Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* 7: 256–276.
52. Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14: 817–818.
53. Swofford DL (2003) PAUP\*: Phylogenetic Analysis Using Parsimony (\* and other methods). Version 4.0b10. SunderlandMA: Sinauer Associates.
54. Kuhner MK (2006) LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics* 22: 768–770.
55. Bjornerfeldt S, Webster MT, Vila C (2006) Relaxation of selective constraint on dog mitochondrial DNA following domestication. *Genome Res* 16: 990–994.
56. Ingvarsson PK (2004) Population subdivision and the Hudson-Kreitman-Aguade test: testing for deviations from the neutral model in organelle genomes. *Genet Res* 83: 31–39.
57. Pesole G, Gissi C, De Chirico A, Saccone C (1999) Nucleotide substitution rate of mammalian mitochondrial genomes. *J Mol Evol* 48: 427–434.
58. Kumar S, Subramanian S (2002) Mutation rates in mammalian genomes. *Proc Natl Acad Sci U S A* 99: 803–808.
59. Megens HJ, Crooijmans RP, San Cristobal M, Hui X, Li N, et al. (2008) Biodiversity of pig breeds from China and Europe estimated from pooled DNA samples: differences in microsatellite variation between two areas of domestication. *Genet Sel Evol* 40: 103–128.
60. Vila C, Seddon J, Ellegren H (2005) Genes of domestic mammals augmented by backcrossing with wild ancestors. *Trends Genet* 21: 214–218.
61. Luo YB, Tao Y, Zhu JY (2009) Preliminary Observation of Pig-bones Burials in the Qinglongquan Site. *Jiangnan archaeology* 3: 58–65.
62. Yuan J (2010) Zooarchaeological study on the domestic animals in ancient china. *Quaternary sciences* 30: 298–306.
63. Vonholdt BM, Pollinger JP, Lohmueller KE, Han E, Parker HG, et al. (2010) Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature* 464: 898–902.
64. Beja-Pereira A, Caramelli D, Lalueza-Fox C, Vernesi C, Ferrand N, et al. (2006) The origin of European cattle: evidence from modern and ancient DNA. *Proc Natl Acad Sci U S A* 103: 8113–8118.