

ORIGINAL ARTICLE

Origins of domestic dog in Southern East Asia is supported by analysis of Y-chromosome DNA

Z-L Ding^{1,2,7}, M Oskarsson^{3,7}, A Ardalan³, H Angleby³, L-G Dahlgren³, C Tepeli⁴, E Kirkness⁵, P Savolainen³ and Y-P Zhang^{1,6}

Global mitochondrial DNA (mtDNA) data indicates that the dog originates from domestication of wolf in Asia South of Yangtze River (ASY), with minor genetic contributions from dog–wolf hybridisation elsewhere. Archaeological data and autosomal single nucleotide polymorphism data have instead suggested that dogs originate from Europe and/or South West Asia but, because these datasets lack data from ASY, evidence pointing to ASY may have been overlooked. Analyses of additional markers for global datasets, including ASY, are therefore necessary to test if mtDNA phylogeography reflects the actual dog history and not merely stochastic events or selection. Here, we analyse 14 437 bp of Y-chromosome DNA sequence in 151 dogs sampled worldwide. We found 28 haplotypes distributed in five haplogroups. Two haplogroups were universally shared and included three haplotypes carried by 46% of all dogs, but two other haplogroups were primarily restricted to East Asia. Highest genetic diversity and virtually complete phylogenetic coverage was found within ASY. The 151 dogs were estimated to originate from 13–24 wolf founders, but there was no indication of post-domestication dog–wolf hybridisations. Thus, Y-chromosome and mtDNA data give strikingly similar pictures of dog phylogeography, most importantly that roughly 50% of the gene pools are shared universally but only ASY has nearly the full range of genetic diversity, such that the gene pools in all other regions may derive from ASY. This corroborates that ASY was the principal, and possibly sole region of wolf domestication, that a large number of wolves were domesticated, and that subsequent dog–wolf hybridisation contributed modestly to the dog gene pool.

Heredity (2012) **108**, 507–514; doi:10.1038/hdy.2011.114; published online 23 November 2011

Keywords: dog; *canis familiaris*; domestication; Y-chromosome DNA; genetic diversity; phylogeography

INTRODUCTION

Based on genetic, morphological and behavioural data (Wayne, 1993; Clutton-Brock, 1995) it is clear that the domestic dog originates from the wolf. However, there is yet no consensus concerning in which geographical region the domestication of wolf occurred. Studies of mitochondrial DNA (mtDNA) from dogs worldwide have strongly indicated the southern part of East Asia (dubbed Asia South of Yangtze River, ASY) (Savolainen *et al.*, 2002; Pang *et al.*, 2009; Klütsch and Savolainen, 2011). Archaeological data has instead indicated an origin from Europe or Southwest (SW) Asia or from multiple regions (Clutton-Brock, 1995), and a recent study of autosomal single nucleotide polymorphism (SNP) data suggested SW Asia as the major source of genetic diversity for dogs (Vonholdt *et al.*, 2010). However, both the archaeological- and the autosomal-SNP datasets suffer from geographical bias, in that they almost totally lack data from ASY (Klütsch and Savolainen, 2011).

Thus, the only comprehensive study of dog origins so far, that includes data from virtually throughout the world, is based on analysis of mtDNA (Pang *et al.*, 2009). It distinctly indicates that the domestic dog originated in ASY, with only limited contribution of wolf from

other regions, probably from later crossbreeding between dog and wolf. The study showed that dogs worldwide share a common gene pool of 3 principal haplogroups containing totally 10 subhaplogroups. Most dogs carry haplotypes shared by virtually every population, with for example, 95% of dogs in Europe and SW Asia carrying a haplotype identical to or differing by a single mutation from haplotypes carried by dogs in East Asia. Thus, there is considerable genetic homogeneity across the world, which is likely the result of all three principal haplogroups originating from a single domestication event. However, the genetic homogeneity is coupled with a gradient of diversity, from distinctly highest values in ASY, through much lower values in North China (N China) and SW Asia, to a minimum in Europe (Pang *et al.*, 2009). It has been claimed that African dogs have as high mtDNA diversity as East Asian dogs (Boyko *et al.*, 2009) but this has been shown to be incorrect (Pang *et al.*, 2009); diversity is distinctly higher in ASY than in all other regions. The difference in diversity reflects that only ASY harbours the full range of variety of the universal gene pool (all 10 subhaplogroups) whereas 5 subhaplogroups were found in, for example, N China and SW Asia and only 4 in Europe. Thus, all haplotypes can be traced to a possible origin in ASY whereas dogs in

¹Laboratory for Conservation and Utilization of Bio-resource and Key Laboratory for Microbial Resources of the Ministry of Education, Yunnan University, Kunming, China;
²Kunming Large-scale Instrument Regional Center of Biodiversity, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China; ³Department of Gene Technology, KTH-Royal Institute of Technology, Science for Life Laboratory, Solna, Sweden; ⁴University of Selcuk, Faculty of Veterinary Medicine, Department of Animal Science, Konya, Turkey; ⁵The J Craig Venter Institute, Rockville, MD, USA and ⁶Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China

⁷These authors contributed equally to this work

Correspondence: Dr P Savolainen, Department of Gene Technology, KTH-Royal Institute of Technology, Science for Life Laboratory, Solna 17121, Sweden.

E-mail: savo@kth.se or Professor Y-P Zhang, Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China.
 E-mail: zhangyp@mail.kiz.ac.cn

Received 18 May 2011; revised 28 September 2011; accepted 14 October 2011; published online 23 November 2011

SW Asia and Europe carry only ~50% of the full range of genetic diversity, from which the full gene pool cannot have derived. In addition, 2.6% of the dogs globally carried haplotypes belonging to four regionally occurring haplogroups, indicating that dogs have hybridised with wolves at a few occasions through history. The mtDNA data also indicates that the dogs originated <16 300 years ago from at least 51 female wolves (Pang *et al.*, 2009).

This picture challenges what has so far been the dominating theories about dogs origins which, based on the archaeological record, have instead suggested Europe or SW Asia as the probable region of dog origins, since the oldest reasonably firm evidence (Raisor, 2005; Wang and Tedford, 2008) of domestic dog is from 11 500 years ago in SW Asia (Dayan, 1994) and 10 000 years ago in Europe (Chaix, 2000), while the oldest well documented evidence of dog in ASY is from only 6500 years ago (Underhill, 1997). However, the archaeological record suffers from geographical bias. Archaeological excavations in general, and detailed analysis of animal materials in particular, have been carried out very unevenly, with the majority of work performed in Europe and the Middle East (Underhill, 1997; Raisor, 2005). Therefore, the archaeological record does not give a comprehensive picture of the geographical distribution of dogs through time and evidence for early domestic dog in ASY may remain undetected. A similar bias against ASY affects the dataset of a recent study of genome-wide SNP variation among domestic dogs and wolves (Vonholdt *et al.*, 2010). The study showed domestic dogs to share more unique multilocus haplotypes with wolves from the Middle East than with wolves from Northern China, Europe and America. However, this kind of analysis relies heavily on adequate sampling of the wolf populations. If the wolf population from which dogs were actually domesticated is not sampled it cannot be identified, and instead another wolf population, showing the greatest haplotype sharing with dogs, would be erroneously identified as the origin of dogs. In this case, no wolves from ASY were included in the SNP study implying that if ASY is actually the region of dog origins this would have gone undetected by these analyses. Furthermore, the SNPs used for the analyses were almost exclusively identified from just two European dogs, and are thus affected by strong ascertainment bias distorting comparison of genetic diversity worldwide (see for example, Morin *et al.*, 2004 and Schuster *et al.*, 2010).

So far, the mtDNA data is the only dataset based on a global sample of dogs or wolves, and it distinctly indicates ASY as the geographical origin of dogs. However, mtDNA is a single genetic marker inherited only through the female line. This leaves open a possibility that the phylogeographical pattern observed (most importantly, the universal sharing of haplotypes but full coverage of diversity only in ASY) may reflect stochastic events, selection or sex bias rather than dog population history. Consequently, global datasets need to be analysed using additional independent markers inherited also through the male line, to confirm or refute the mtDNA-based results. Here, we analyse Y-chromosome DNA sequence in a worldwide sample of dogs.

Two dog genome sequences have been published, from a male standard Poodle (sequenced with 1.5×sequence coverage; Kirkness *et al.*, 2003) and a female Boxer (7.5×sequence coverage; Lindblad-Toh *et al.*, 2005), but the Y-chromosome sequence is largely unidentified since the male sequence, because of the low sequence coverage, is fragmentary. However, 24 000 bp of dog Y-chromosome DNA sequence has been identified (Natanaelsson *et al.*, 2006). From these data, 14 437 bp were also sequenced for 10 dogs from across the world, yielding 14 substitutions that defined nine different haplotypes. The minimum number of wolf founder haplotypes was estimated at five (Natanaelsson *et al.*, 2006).

Here, we analyse these 14 437 bp of Y-chromosome DNA in the first comprehensive study of Y-chromosome diversity among dogs worldwide, to produce the second global dataset for studies of dog origins. Hereby, we obtain genetic data for a second independently inherited marker to evaluate the scenario for the origins of domestic dogs suggested by mtDNA data (Savolainen *et al.*, 2002; Pang *et al.*, 2009). It is unlikely that independently inherited markers would be affected by selection in the same way, or would by chance obtain the same phylogeographical pattern. Therefore, if the same global phylogeographical pattern would be found for Y-chromosome DNA as for mtDNA, this would corroborate the scenario for dog origins indicated by the mtDNA data. We also obtain information not accessible from the maternally inherited mtDNA, most importantly about the number of male founders from wolf and the extent of crossbreeding between female dog and male wolf.

MATERIALS AND METHODS

Analysed DNA sequence

The dog Y-chromosome sequence is largely unidentified since the male genome sequence (Kirkness *et al.*, 2003) has only 1.5×sequence coverage and therefore is fragmentary. However, 24 000 bp of dog Y-chromosome DNA sequence has been identified through Blast search of the male genome shotgun sequences against the female genome and human Y-chromosome sequences, and testing for male specificity through PCR screening (Natanaelsson *et al.*, 2006). In the present study, we analysed 14 437 bp of this Y-chromosome sequence. As it consists of shotgun sequences, the analysed sequence is distributed in 18 different amplification fragments (see Supplementary Dataset 1, 2).

Samples

The Y-chromosome DNA sequence was studied for totally 165 canids: 151 dogs (10 of which from (Natanaelsson *et al.*, 2006)), 12 wolves and 2 coyotes. The dog samples were collected to represent different parts of the world (Table 1; see also below and in Supplementary Dataset 3 for complete list of samples). Care was taken to obtain comprehensive and representative sampling by collecting across geographical regions and normally only a single sample from each location, and for breed dogs mostly a single dog per breed. Samples were assumed to represent geographical regions based on either (i) being from a region (mostly rural) with small influx of foreign dogs or (ii) belonging to a

Table 1 Genetic diversity in geographical regions

Region ^a	N	HTs ^b	HD (s.e.) ^c	Res 23 ^d	Res 14 ^d	Res 9 ^d
ASY	23	13	0.901 (0.051)	13	9.30	6.69
SW Asia	25	10	0.863 (0.044)	9.58	7.35	5.67
Europe	32	7	0.734 (0.064)	6.50	5.31	4.33
Southw ASY	16	11	0.950 (0.036)	—	10.1	7.35
Fertile Cr	14	8	0.923 (0.044)	—	8	6.46
Siberia	9	7	0.917 (0.092)	—	—	7
N China	12	5	0.848 (0.059)	—	—	4.69
C China	14	7	0.824 (0.078)	—	7	5.22
S China	16	9	0.858 (0.077)	—	8.24	6.04
SW Asia East	11	5	0.709 (0.136)	—	—	4.43
C Europe	10	4	0.644 (0.149)	—	—	3.80
S Europe	10	4	0.644 (0.149)	—	—	3.80
Africa	19	7	0.848 (0.046)	—	6.19	5.14
America	9	3	0.639 (0.128)	—	—	3

^aASY, Asia South of Yangtze River (South China, Southeast Asia); Fertile Cr, Fertile Crescent region (West Iran, Israel, East Turkey); N, north; C, central; S, south; Southw ASY, Southwest ASY (Yunnan, Guangxi, Southeast Asia); N/C/S China, China North of Yellow/between Yellow and Yangtze/South of Yangtze Rivers; SW Asia East, eastern part of SW Asia (East Iran, Afghanistan).

^bHTs=Number of haplotypes.

^cHD=Haplotype Diversity (with s.e.).

^dRes=Number of haplotypes obtained from resampling of size 23/14/9 (1000 replications) to adjust for differences in sample size.

breed with known historical geographical origin. Dogs not belonging to a specific breed had mostly specialized morphology and were kept by their owners for a specific use; all dogs sampled had an owner, that is, none were stray. For the European dog population, with its special history of intense breeding often involving severe, breed specific, genetic bottlenecks (Clutton-Brock, 1995), sampling bias was avoided by sampling a single individual per breed, among different morphological types and across most parts of Europe.

The geographical distribution of the dog samples (see also Table 1 and Supplementary Dataset 3): Africa—Northern Africa: Niger (Azawakh, $n=1$), Mali (Azawakh, $n=2$), Misc (Sloughi, $n=2$); Southern Africa—South Africa ($n=4$), Benin (Basenji, $n=1$), Cameroon ($n=1$), Kenya ($n=2$), DR Congo (Basenji, $n=3$), Sudan (Basenji, $n=1$), Lesotho ($n=2$); America—North America (Xoloitzcuintle, $n=2$, Chihuahua, $n=1$, Alaskan Malamute, $n=2$, Inuit Sled Dog, $n=1$, Greenland Dog, $n=1$), South America (Perro Sin Pelo Del Peru, $n=2$); East Asia—China (Guangdong, $n=1$, Guangxi, $n=5$, Guizhou, $n=2$, Hainan, $n=2$, Hebei (Pug, $n=1$), Chinese Crested Dog, $n=1$, Chow chow, $n=1$, Shih tzu, $n=1$), Heilongjiang, $n=2$, Hunan, $n=2$, Jiangxi, $n=2$, Liaoning, $n=4$, Qinghai, $n=2$, Shaanxi, $n=2$, Shanxi, $n=2$, Sichuan, $n=6$, Tibet (Tibetan Terrier, $n=1$, Tibetan Spaniel, $n=1$, Tibetan Mastiff, $n=1$, Lhasa Apso, $n=1$), Yunnan, $n=4$), Japan (Akita, $n=1$, Iwate matagi, $n=1$, Kisyu, $n=1$, Shiba, $n=1$, Hokkaido, $n=1$), Siberia (Aboriginal Sled Dog, $n=2$, East Siberian Laika, $n=2$, West Siberian Laika, $n=1$, Siberian Husky, $n=1$, Samoyed, $n=1$, Chukotka Sled Dog, $n=2$); Southeast Asia—Cambodia ($n=1$), Thailand (Thai Ridgeback, $n=2$, Misc., $n=2$), Vietnam (Phuquoc Dog, $n=1$, Misc., $n=1$), Misc: (Sredneasiatskaja Ovtjarka, $n=1$); Europe—Britain (Yorkshireterrier, $n=1$, Golden Retriever, $n=1$, English Springer Spaniel, $n=1$, Border Collie, $n=1$, Shetland Sheep Dog, $n=1$, Cavalier King Charles Spaniel, $n=1$); Central Europe (German Shepherd, $n=1$, Bouvier des Flandres, $n=1$, Chart Polski, $n=1$, Dachshund, $n=1$, Puli, $n=1$, Leonberger, $n=1$, Polski Owczarek Nizinny, $n=1$, Pumi, $n=1$, Groenendael, $n=1$, Slovenský Kopov, $n=1$); South Europe (Poodle, $n=1$, Lagotto Romagnolo, $n=1$, Maremma-abruzzese, $n=1$, Piccolo Levriero Italiano, $n=1$, Volpino Italiano, $n=1$, Galgo Espanol, $n=1$, Pyrenean Mastiff, $n=1$, Portuguese Water Spaniel, $n=1$, Cane Corso, $n=1$, Bracco Italiano, $n=1$); Scandinavia (Swedish Elkhound, $n=1$, Finnish Lapphund, $n=1$, Norwegian Lundehund, $n=1$, Norwegian Elk Hound, $n=1$, Karelo-Finnish Laika, $n=1$); Misc. (Russo-European Laika, $n=1$); SW Asia: Afghanistan (Afghan Hound, $n=1$), Iran ($n=10$; from 10 different locations across Iran), Israel (Canaan Dog, $n=4$), Turkey ($n=7$; from 6 different locations across Turkey), Misc., $n=3$.

Definition of geographical regions and abbreviations used in the study (see also Supplementary Dataset 3); Central Europe: Germany, Belgium, Poland, Hungary and Slovakia; South Europe: France, Italy, Spain and Portugal; Fertile Crescent region (Fertile Cr): the Fertile Crescent with adjacent regions (South Eastern Turkey, Israel and Western Iran) where domestication of farm animals took place and/or the earliest Neolithic occurred (for example, Catal Hoyuc); SW Asia East: Central and Eastern Iran and Afghanistan; N China: China North of the Yellow River; Central China: China between the Yellow River and Yangtze River; South China: China South of Yangtze River; ASY: South China and Southeast Asia; Southw ASY (Southwestern ASY): Yunnan, Guangxi and Southeast Asia.

The geographical distribution of wolf samples: Europe: Scandinavia ($n=1$); America: North America ($n=1$); East Asia: China: Heilongjiang ($n=6$), Qinghai ($n=1$), Sichuan ($n=1$), Shanxi ($n=1$), Yunnan ($n=1$). The two coyotes were sampled in Sonoma county and Butte county, CA, USA.

DNA extraction, amplification and sequencing

Samples, buccal epithelial cells on Whatman FTA cards or heparin treated blood, were extracted according to (Natanaelsson *et al.*, 2006). PCR amplification was performed in two steps using nested outer- and inner-primer pairs (See Supplementary dataset 4) for increased target specificity, with PCR conditions described in (Natanaelsson *et al.*, 2006). DNA sequencing was performed with primers (See Supplementary dataset 4) giving forward and reverse sequence reads for all nucleotide positions, using ABI Big Dye Terminator chemistry and analysis on an ABI 3700 DNA sequencer as described (Natanaelsson *et al.*, 2006). The DNA sequences were edited, assembled into contigs and aligned using Sequencher 4.1 (Gene codes corporation, Ann Arbor, MI, USA).

Calculation of phylogenetic relations and genetic diversity

A most parsimonious phylogenetic tree was created by calculation of pairwise genetic distances among haplotypes, in number of substitutions, using Arlequin ver 3.1 (Excoffier *et al.*, 2005), and the tree was drawn by hand. For comparing the number of haplotypes among populations with different sample size, the sample sizes were normalised by resampling populations (without replacement, 1000 replications) for an equal number of samples, using an in-house developed programme. Haplotype diversity (HD) was calculated in Arlequin ver 3.1, and s.e. estimated by bootstrap resampling (1000 replications).

RESULTS

Y-chromosome sequence variation among domestic dogs

We here present the first study of Y-chromosome DNA sequence diversity among dogs worldwide, hereby, obtaining genetic data for a second independently inherited marker to evaluate the scenario for the origins of domestic dogs. We analysed 151 dogs sampled from throughout the world and, for a reference, also 12 wolves and 2 coyotes, for 14 437 bp of Y-chromosome DNA sequence (see Table 1, Materials and methods, Supplementary Dataset 1, 2 and 3). In total, there were 49 nucleotide positions with binary substitutions (1 substitution/295 bp) and 14 with indels, and among dogs 30 substitutions (1 substitution/481 bp) and 11 indels. The 49 substitutions define 32 haplotypes: 28 found among dogs (1 shared with wolf), 2 wolf specific and 2 coyote specific. The genetic relations between haplotypes were reconstructed in a most parsimonious phylogenetic tree (Figure 1a), without homoplasy in any nucleotide position.

Single basecalls are expected at every nucleotide position of the haploid Y-chromosome. However, three positions situated in one amplification fragment (fragment G; see Materials and methods and Supplementary Dataset 2) had double basecalls resembling diploid variation (~50% of each nucleotide) for some individuals. The haplotypes with double basecalls (haplotypes with names including ‘*’) group in three parts (separate for each position) of the phylogeny (Figure 1a); the double basecalls are obviously caused by a duplication of the DNA segment and subsequent substitution in one of the copies at three different positions.

In accordance with dogs originating from wolf (Wayne, 1993; Clutton-Brock, 1995), the wolf haplotypes differed by 0–4 substitutions and the coyote haplotypes by at least 15 substitutions from the closest-dog haplotype. Among the 12 wolf samples there were three haplotypes: H23*, which was shared between dog and one Chinese wolf, H26 (found in one American wolf), which was separated from two dog haplotypes by one substitution and H27 (found in nine Chinese and one Scandinavian wolf), which differed by four substitutions from the closest-dog haplotype (Figure 1a).

Five major dog haplogroups but at least 13 male founders

The dog haplotypes clustered in five major groups (dubbed haplogroups HG1, HG3, HG6, HG9 and HG23 after their respective central haplotypes) consisting of one or two frequently occurring central haplotypes surrounded by less frequent haplotypes (Figure 1a). This pattern is suggestive of an origin of dogs from five wolf founders, carrying five haplotypes from which all other haplotypes subsequently derived through substitutions within the dog population. However, calculations based on the number of substitutions expected to have occurred among the 151 dog lineages since the time of the dog origins indicate a larger number of founders from wolf. We estimated the substitution rate from the mean number of substitutions between dog/wolf and coyote (18.1 substitutions (17.5–18.6, 95% confidence limits) or 1.25×10^{-3} substitutions per site (1.22×10^{-3} – 1.29×10^{-3})), and the time since the split between the lineages leading to wolf and coyote. There is no exact archaeological calibration point for this separation;

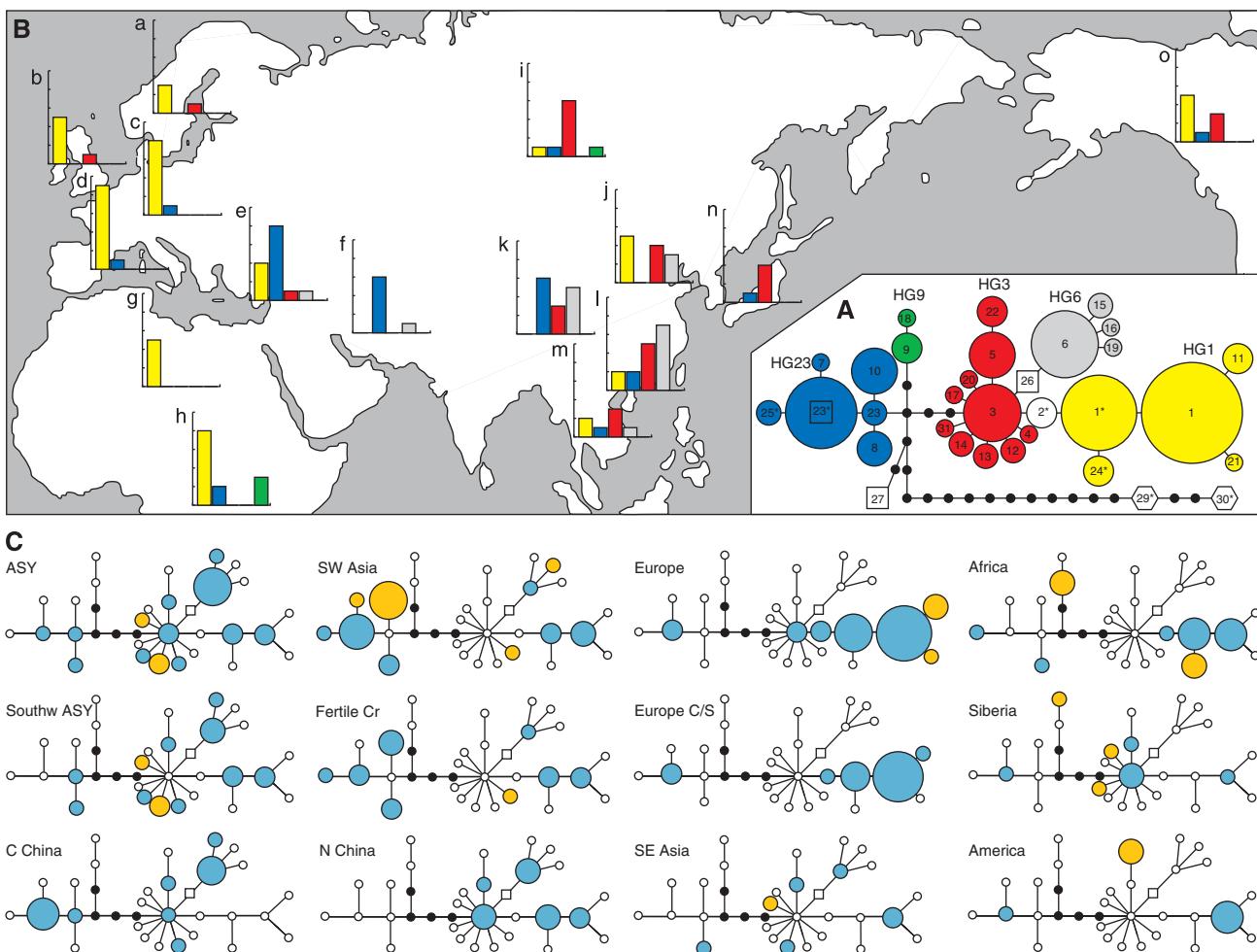


Figure 1 Phylogenetic and geographical distribution of haplotypes. **(A)** Most parsimonious phylogenetic tree. Haplotypes (symbolized by circles for dog, squares for wolf and hexagons for coyote; black dots are hypothetical intermediates) are separated by one substitutional step. The area of the circles is proportional to the frequency of the haplotype among dogs. Haplogroups (see text) are indicated by colour; haplotype 2* cannot be assigned to HG1 or HG3 and therefore white. **(B)** Geographical distribution of haplogroups. Graphs show number of individuals carrying each haplogroup, colours referring to haplogroups according to (a). Populations: a, Scandinavia; b, Britain; c, Central Europe; d, South Europe; e, Fertile Cr; f, SW Asia East; g, Northern Africa; h, Southern Africa; i, Siberia; j, North China; k, Central China; l, South China; m, Southeast Asia (l and m jointly forming ASY); n, Japan; o, America. For definitions of geographical regions, see Note to Table 1, and Materials and methods. **(C)** Trees (see a) showing representation (blue, shared with other regions; yellow, unique to the region; white, not present) and frequency (proportional to area) of haplotypes among dogs in geographical regions. Europe C/S, Central and South Europe; SE Asia, Southeast Asia.

1.8–2.5 (Nowak, 2003) or around 3.2 million years ago (Tedford *et al.*, 2009) has been suggested but it may have occurred 1.5–4.5 million years ago (Nowak, 2003; Tedford *et al.*, 2009). The substitution rate was thus calculated as a broad range at 1.39×10^{-10} – 4.18×10^{-10} substitutions per site per year (1.35×10^{-10} – 4.31×10^{-10} , 95% confidence limits) or one substitution per 165 746–497 238 years (160 782–513 078 years). This is less than half the rate estimated for the human Y-chromosome (Xue *et al.*, 2009), which is possibly related to the method used for identifying the dog Y-chromosome sequences (see Materials and methods), which involved selection for similarity to human Y-chromosome sequences, possibly enriching conserved regions of the dog Y-chromosome.

Assuming that dogs originated 11 500–16 000 years ago, according to the archaeological record (Dayan, 1994; Chaix, 2000; Raisor, 2005; Wang and Tedford, 2008; Napierala and Uerpmann, 2010), mtDNA data (Pang *et al.*, 2009) and autosomal SNP data (Skoglund *et al.*, 2010), at average 0.023–0.097 substitutions (0.022–0.100, 95% confidence limits) would have occurred per dog lineage since the time of

origin from wolf. Further, assuming conservatively that all 151 dogs represent independent lineages leading back to the dog origins, 3.5–14.6 substitutions (3.4–15.0, 95% confidence limits) would have occurred among the 151 dog lineages since domestication. Given a total of 28 haplotypes among the domestic dogs, this implies that 13.4–24.5 (13.0–24.6) of the 28 haplotypes, rounded down to 13–24, are intact from the wolf founders. Thus, our data indicate that the Y-chromosome gene pool of the relatively limited number of dog samples in this study originates from at least 13–24 different wolf Y-chromosome haplotypes. The formation of the dog haplotypes in five star-like clusters must therefore partly stem from the relations between haplotypes in the founder wolf population(s). Notably, an origin of dogs from numerous male wolves is in line with both mtDNA data indicating that dogs originated from a minimum of 51 female wolf lineages (Pang *et al.*, 2009) and MHC data (from the low diversity European dog population) indicating an origin from at least 21 wolves (Vilà *et al.*, 2005). Therefore, multiple genetic datasets indicate that dogs originate from a large number of domesticated wolves.

Two of four principal haplogroups are shared universally but two are almost exclusive to East Asia

The dog Y-chromosome gene pool was to a large degree shared among the populations across the world (Figures 1b and c). Two of the five haplogroups (HG1 and HG23) were virtually universally represented and carried by 62% of all dogs in the study. The three central haplotypes within these haplogroups, H1, H1* and H23*, were carried by almost half (46%) of the dogs and shared by dogs in Europe, SW Asia and China, by 75%, 44% and 32% of the individuals, respectively.

However, there were also distinct differences in the geographical representation and distribution of haplogroups and haplotypes. The other three haplogroups were also distributed across relatively large distances but not universally spread. HG3 was found in East Asia (including Siberia) and America, and at lower frequency in SW Asia, Scandinavia and Britain, but not in samples from the European continent and Africa. HG6 was found in East Asia and at low frequency in SW Asia, but was absent elsewhere. Finally, HG9 was found in only totally four individuals, but as far apart as East Siberia (one individual) and Central Africa (three individuals).

As the sample sizes were relatively limited, haplogroups with low frequency, for example, HG9 may have remained undetected in some populations. However, the general pattern was that the four main haplogroups were relatively equally represented in the eastern part of the world, whereas west of the Himalayas and the Urals haplogroups HG1 and HG23 were represented by 89% of the individuals, and HG6 and HG3 rare or absent. Thus, HG1 and HG23 were universally represented, whereas HG3 and HG6 had restricted distributions. Only in East Asia and SW Asia all four major haplogroups were represented.

Highest genetic diversity in Southwestern ASY

Accordingly, except for the practically universal representation of haplotypes H1, H1* and H23*, the representation and frequencies of haplotypes differed considerably among regions, as demonstrated in the phylogenetic trees (Figure 1c). In some regions, for example, Europe, frequency was very high for a few haplotypes, mainly H1, H1* and H23*, and other parts of the phylogeny was empty. Other regions, for example, ASY, had a larger number of haplotypes at more even frequencies and representation across the phylogeny. These differences in genetic coverage are reflected in difference in genetic diversity measured as the number of haplotypes per sampled individual and HD (Table 1). In many cases the samples were too small to yield significant differences, but the general trend was that the highest values for genetic diversity among all regions were found within ASY, there were medium values in other parts of East Asia, and in SW Asia and Africa, and low in Europe and America.

Comparing the three major regions suggested as potential origins for dogs, ASY had the highest diversity with 13 haplotypes among 23 samples, and a HD of 0.901, to compare with SW Asia and Europe, which had 9.58 and 6.50 haplotypes at resampling of 23 samples, and a HD of 0.863 and 0.734, respectively (Table 1). Importantly, except for haplogroup HG9, practically the full diversity for dog Y-chromosome DNA was covered in ASY, such that all haplotypes in other regions were maximally one step from haplotypes in ASY (Figure 1c). The highest diversity worldwide was found within the Southwestern part of ASY (Southw ASY; Southeast Asia and the adjacent Chinese provinces Yunnan and Guangxi) with 11 haplotypes among 16 individuals, 10.10 haplotypes at resampling size 14, and a HD of 0.950. In contrast, at the other end of Eurasia, Europe had 7 haplotypes among 32 samples and 5.31 haplotypes at resampling size 14, almost half compared with Southw ASY. The remarkably low diversity for Europe is related to high frequency of haplotypes H1

(carried by 47% of the individuals) and H1* (22%) and that the other parts of the phylogeny are largely empty. This pattern was shared across Europe, by the north and south parts of the continent as well as Britain, and must therefore stem from the first origin of the European-dog population and not from later intense breeding, as it is unlikely that all haplogroups but HG1 would have been lost independently in several different lineages leading to today's breeds. SW Asia had 10 haplotypes among 25 samples and 7.35 haplotypes at resampling size 14, and had much higher frequency of haplogroup HG23 (68%) than other regions, whereas only one and two samples carried HG3 and HG6, respectively. Within SW Asia, the Fertile Crescent region (Fertile Cr; West Iran, Israel and East Turkey) had a higher diversity, with HD higher, but the number of haplotypes lower than ASY. Also here the frequency of HG23 was high (57%), but all four main haplogroups were represented. Among other regions, Siberia had especially high diversity, with marginally lower values than Southw ASY for number of haplotypes and haplotype diversity. Central and N China and Africa had medium diversity values and the small sample of American dogs had three haplotypes among nine samples.

Thus, diversity differences were generally small across the Old World, but Southw ASY had the highest diversity of all regions. The large difference between the opposite sides of the Eurasian continent is striking, and further highlighted by comparing the samples from Europe, having seven haplotypes among 32 samples, and Southeast Asia with six haplotypes (distributed among all four major haplogroups) among only 7 samples (Figure 1c).

With this study, two independently inherited markers have shown genetic diversity among dogs worldwide to be highest within ASY. It is also notable that, in similarity to the mtDNA data, ASY had the most comprehensive coverage of the phylogenetic diversity of all regions. The haplotypes were distributed across the four major haplogroups such that all haplotypes in other regions were at most one substitution from a haplotype found in ASY (Figure 1c). Therefore, except HG9, all haplotypes across the world were identical to or differed by a single substitution from a haplotype found in ASY, and may potentially have derived from haplotypes present in ASY.

A possible single origin of all haplogroups in ASY, but not in SW Asia or Europe

The haplogroups were geographically distributed in a distinct pattern (Figure 1b). HG1 had a frequency close to 100% in Europe and Africa, and HG23 a high frequency in SW Asia and Central China, but both haplogroups were also represented at lower and relatively even frequency virtually worldwide. In contrast, HG3 and HG6 were almost exclusively restricted to East Asia, at moderate frequency. This pattern may be explained by an origin of all four haplogroups from a single (not necessarily homogenous) founder population somewhere in East Asia, for example ASY, and genetic bottlenecks reducing diversity in other populations. However, separate origins of the haplogroups in different regions followed by non-symmetrical migrations between populations are also possible.

The high frequency (81%) and large number of haplotypes (four) of HG1 in Europe could possibly be explained by an origin of HG1 in Europe, after which only two of four haplotypes derived from the wolf founders would have spread to other regions. However, because of the high frequency in Europe of this haplogroup, a larger number of derived haplotypes are expected than in other regions. Among the 26 European lineages carrying HG1, 0.60–2.51 substitutions (0.57–2.60, 95% confidence limits) would be expected to have occurred during the 11 500–16 000 years since the origins of dogs. This indicates that only the universal haplotypes, H1 and H1* were inherited from wolf and

the others derived from mutations within the European dog population. Therefore, HG1 being virtually universally represented, its geographical origins cannot be definitely identified based on this dataset. Similarly, SW Asia had a high frequency (68%) and the largest number of haplotypes (five) of HG23. In this case, 0.39–1.64 (0.37–1.70) substitutions would be expected among the 17 lineages in SW Asia, weakly indicating that HG23 may have originated in SW Asia. It is notable that the Fertile Cr had four haplotypes among eight individuals carrying HG23. However, HG23 was represented almost universally and also ASY had a high diversity for HG23, with three haplotypes among three samples. For HG3, ASY had six haplotypes among eight lineages. Only 0.18–0.77 (0.18–0.80) substitutions would be expected since the origins of dogs, leaving the majority of haplotypes identical to the haplotypes carried by wolves; the star-like formation of HG3 was obviously inherited from the founder wolf population. The large number of HG3 haplotypes in ASY indicates an origin of this haplogroup in ASY or adjacent regions, but a relatively high diversity (four haplotypes among six individuals) in Siberia is also notable. Finally, HG6 being found almost exclusively in East Asia most probably originated somewhere in this region.

Consequently, it is not possible to definitely point out from where each haplogroup originated. However, it can with greater certainty be concluded from where the haplogroups did not originate. Thus, it seems very unlikely that haplogroups HG3, HG6 and HG23 would have originated in Europe or Africa, or haplogroups HG3 and HG6 in SW Asia. Therefore, three out of four of the dogs Y-chromosome gene pool clearly originates from outside Europe as only HG1 may have originated there. Importantly, the extremely low diversity in Europe cannot be linked to the intense breeding of European dogs in historic times (see Discussion). It also seems clear that a maximum of roughly 50% of the gene pool (HG23 and HG1) may have originated in SW Asia. In contrast, the full dog Y-chromosome gene pool may have originated somewhere in East Asia, including ASY. ASY is especially likely considering that, uniquely, all haplotypes of the four major haplogroups differed by at most one substitution from haplotypes in ASY.

To conclude, the Y-chromosomal DNA data indicates that if the domestic dog originated from a single geographical region this could have happened in ASY but not in SW Asia or Europe. If the dog originated from several regions, at most 50% of the gene pool may have originated in SW Asia or Europe. Thus the Y-chromosome data indicates that wolves in ASY were the major source of genetic diversity for dogs.

DISCUSSION

With this study, analyses of two independently inherited DNA markers, the only two studies based on global samples of dogs performed so far, give strikingly similar pictures of dog phylogeography. Thus, both the present study of Y-chromosomal DNA and earlier studies of mtDNA (Pang *et al.*, 2009) show that ~50% of dog genetic diversity is shared in a universal gene pool, but whereas most regions harbour only these 50%, ASY has virtually the full range of genetic diversity from which the complete gene pools in other regions may derive. It is unlikely that the two datasets would by chance have obtained the same phylogeographical pattern or that selection would have affected both markers similarly. Therefore, these results offer strong evidence that domestication of wolf occurred primarily and possibly exclusively in ASY, with only small genetic contributions from wolf in other regions, through dog–wolf hybridisation.

This is in conflict with conclusions normally drawn from analyses of the archaeological record (Clutton-Brock, 1995) and in a recent study

of autosomal SNPs (Vonholdt *et al.*, 2010), suggesting SW Asia and/or Europe as the principal regions of origin. However, both the archaeological record and the SNP study suffer from geographical bias in a lack of data from ASY (Klütsch and Savolainen, 2011). Therefore, there is a clear possibility that these datasets failing to identify ASY's central role in dog origins may reflect the lack of sampling specifically from this region. Arguably, the Y-chromosome DNA and mtDNA datasets represent only two genetic markers, and the Y-chromosome data includes relatively small samples. Therefore, analyses of further markers are desirable; when based on comprehensive sampling, large-scale studies of genome wide polymorphisms, for example, autosomal SNPs will help to reveal dog history in unprecedented detail. However, in the light of the mtDNA and Y-chromosomal data, comprehensive sampling from ASY is necessary for any study aimed at unravelling the origins and earliest history of dogs. It is especially notable that, for both Y-chromosome and mtDNA data, diversity is much lower in N China and Central China than in ASY, and instead more similar to that of other regions, for example, SW Asia. Therefore, samples from China or East Asia in general cannot compensate for lack of samples from ASY.

The exact geographical origin of each Y-chromosome haplogroup cannot be determined based on the present dataset. However, it seems clear that at most 50% of the genetic diversity can have originated from SW Asia or from Europe, and it is possible, especially considering that all haplotypes of the four principal haplogroups differ by at most a single substitution from a haplotype found in ASY, that 100% of the Y-chromosome gene pool originated in ASY in a single domestication event. The strongest indication against this is the high frequency and relatively high diversity of HG23 in SW Asia. In Fertile Cr >50% of the samples had HG23 and four of the six haplotypes were represented, suggesting the possibility of a separate origin of HG23, through independent domestication or crossbreeding of dog and wolf. However, in the case of independent domestication a high frequency would be anticipated also in the neighbouring regions, but instead the frequency of HG23 was exceptionally low in, for example, Europe (6%). Considering the large impact of the spread of farming and the related farm animals from the SW Asia to Europe (Bellwood, 2005), it would be anticipated that European dogs, if originating from SW Asia, would have a high frequency of the SW Asian haplotypes. An alternative possibility is that HG23 originated from crossbreeding of dog with wolf in SW Asia. The mtDNA data gives a clear indication of crossbreeding in SW Asia, haplogroup d2 being found only in SW Asia and the Mediterranean at a frequency of ~2% (Pang *et al.*, 2009; Klütsch *et al.*, 2010). However, in crossbreeding of wolf into an already established dog population the novel haplotypes would be expected to remain at low frequency, like the mtDNA haplogroup d2, and not above 50% as HG23. The geographical origin of HG23 is therefore unclear, but an origin in ASY, where three different HG23 haplotypes were found among only three dogs, cannot be excluded.

There was not a single example of regionally restricted Y-chromosome haplogroups and therefore no clear sign that crossbreeding between male wolf and female domestic dog have contributed extensively to the evolution of the domestic dog. However, haplotypes deriving from crossbreeding would normally have limited geographical spread unless a superior phenotype would have evolved (Pang *et al.*, 2009; Klütsch *et al.*, 2010), and may have gone undetected in this study. So far, the only clear genetic evidence of wolf–dog crossbreeding is the regionally restricted mtDNA haplogroups d1 (restricted to Scandinavia), d2 (restricted to the Middle East and the Mediterranean), and F (found only in a few extant Japanese dogs and samples from extinct Japanese wolf) (Ishiguro *et al.*, 2009; Pang *et al.*, 2009; Klütsch *et al.*, 2010).

Care was taken to obtain extensive and representative samples from each geographical region, by collecting across the regions and normally a single sample from each location. It is therefore noticeable that several extensive regions had one haplotype at very high frequency (See Supplementary Dataset 3), a pattern not seen for mtDNA (Pang *et al.*, 2009). For example, 6 of 10 samples from across Iran carried haplotype H23*, all 4 samples from (different parts of) the Japanese main island Honshu, carried H5 and 2 out of 2 samples from each of the South Chinese provinces Guizhou and Hunan carried H6. At analysis of Y-chromosome microsatellites according to (Bannasch *et al.*, 2005) all samples had different haplotypes (See Supplementary Text), showing that the sharing of SNP-based haplotypes is not the result of events in modern time. Therefore, the dominance of a single Y-chromosome haplotype across large regions possibly reflects involvement of relatively few males in some migrations and population founder events.

Considering the intense breeding of European dogs during the last few 100 years, giving severe breed-specific bottlenecks (Clutton-Brock, 1995), special care was taken to avoid sampling bias by sampling a single individual per breed, from different morphological types and from across Europe. The extremely low diversity, 81% of European dogs carrying HG1, must therefore stem from before breeding started, as it is unlikely that all haplogroups but HG1 would have been lost independently in several different lineages leading to today's breeds. For mtDNA the picture is even clearer, with the European population lacking 6 of the 10 principal haplogroups, 5 of which are missing also in SW Asia (Pang *et al.*, 2009), showing that the loss of diversity occurred before the European and SW Asian populations were originally formed. Therefore, the low genetic diversity of the European population, and its separate grouping in analysis of autosomal SNPs (Vonholdt *et al.*, 2010), seem to reflect the geographical position at the far end of the Eurasian continent compared with ASY, rather than recent intense breeding.

The Y-chromosome data, as well as mtDNA (Pang *et al.*, 2009) and autosomal MHC data (Vilà *et al.*, 2005), indicates that a large number of wolves were founders for the domestic dog population. Considering the relatively small sample of dogs in this study and that some domesticated wolves probably carried identical HTs, a minimum of 13 Y-chromosome haplotypes and 51 mtDNA haplotypes (Pang *et al.*, 2009) deriving from the wolf founders indicates that the origin of dogs involved taming of several hundred wolves and was a major event in the related human culture.

The phylogeographical data is not detailed enough to indicate exactly where this domestication may have taken place, since several South Chinese provinces and also, for example, Burma have not been analysed for either Y-chromosome DNA or mtDNA. The possibility that dogs originated in connection with the transition from hunter gathering to farming of rice (Bellwood, 2005) has been suggested, based on mtDNA indicating dogs to have originated approximately at this time (Pang *et al.*, 2009). This would place the origin of dogs in Northern/Central ASY where the earliest evidence of rice cultivation has been found (Underhill, 1997; Bellwood, 2005). However, the highest Y-chromosomal diversity was found in Southw ASY, which was also the only region harbouring the full set of the principal mtDNA haplogroups (Pang *et al.*, 2009). The southern range of wolves would define the southern limit for possible domestication of wolf, but the historical range of wolf in the region is not known. Thus, although the principal region of dog origins has probably been identified, many details remain to be studied. However, analyses based on denser sampling and application of the new generation of powerful DNA sequence analysis has the potential of producing a very

detailed phylogeographic map of the region, promising a detailed picture of the first steps in dog origins.

CONCLUSION

With this study of Y-chromosome diversity among dogs worldwide, we present a second global dataset, in addition to mtDNA, for studies of dog origins. These two independently inherited genetic markers give strikingly similar pictures of dog phylogeography. Most importantly, both markers show that ~50% of dog genetic diversity is shared in a universal gene pool, but whereas most regions harbour only these 50%, ASY has virtually the full range of genetic diversity from which the complete gene pools in other regions may have derived.

This offers strong evidence that domestication of wolf occurred primarily and possibly exclusively, in ASY. Both markers also indicate that a large number of wolves, probably several hundred, were domesticated, which suggests that taming of wolf was an important cultural trait in the related human populations. Subsequent hybridisation between dog and wolf seems to have occurred only rarely.

Studies of the archaeological record and autosomal SNP data have not indicated ASY to be the region of dog origins but, because of an almost complete lack of samples from ASY in these studies, evidence indicating ASY may have been overlooked. In the light of the Y-chromosomal and mtDNA data it is clear that comprehensive sampling from across the world, and especially ASY, is necessary for studies of early dog history.

Based on this knowledge, analyses of haplotypic and autosomal genome-wide markers on geographically dense sample collections and systematic archaeological investigations of canid material in neglected regions, can now be initiated. Hereby, elucidation of details, such as the more exact location(s) of dog origins in ASY, the possibility that independent domestication of wolf also occurred in regions other than ASY and the extent of crossbreeding of dog and wolf through history seems within reach.

DATA ARCHIVING

All new haplotypes identified have been deposited in GenBank with accession numbers HQ389365–HQ389435.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by grants from the Swedish Research Council and the Carl Trygger Foundation. Peter Savolainen is a Royal Swedish Academy of Sciences Research Fellow supported by a grant from the Knut and Alice Wallenberg Foundation.

- Bannasch DL, Bannasch MJ, Ryun JR, Famula TR, Pedersen NC (2005). Y chromosome haplotype analysis in purebred dogs. *Mamm genome* **16**: 273–280.
Bellwood P (2005). *First Farmers: The origins of Agricultural Societies*. Blackwell Publishing: UK.
Boyko AR, Boyko RH, Boyko CM, Parker HG, Castelhano M, Corey L *et al.* (2009). Complex population structure in African village dogs and its implications for inferring dog domestication history. *Proc Natl Acad Sci USA* **106**: 13903–13908.
Chaix L (2000). A preboreal dog from the northern Alps (Savoie, France). In: Crockford SJ (eds). *Dogs Through Time: An Archaeological Perspective*. British archaeological reports: Oxford. pp 49–59.
Clutton-Brock J (1995). Origins of the dog: domestication and early history. In: Serpell J (eds). *The Domestic Dog, its Evolution, Behavior and Interactions With People*. Cambridge University Press: Cambridge. pp 7–20.
Dayan T (1994). Early domesticated dogs of the near East. *J Archaeol Sci* **21**: 633–640.

- Excoffier L, Laval G, Schneider S (2005). Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evol Bioinform Online* **1**: 7–50.
- Ishiguro N, Inoshima Y, Shigebara N (2009). Mitochondrial DNA analysis of the Japanese wolf (*Canis Lupus Hodophilax* Temminck, 1839) and comparison with representative wolf and domestic dog haplotypes. *Zoolog Sci* **26**: 765–770.
- Kirkness EF, Bafna V, Halpern AL, Levy S, Remington K, Rusch DB et al. (2003). The dog genome: survey sequencing and comparative analysis. *Science* **301**: 1898–1903.
- Klütsch CFC, Savolainen P (2011). Geographical origin of the domestic dog. In *Encyclopedia of Life Sciences*. John Wiley & Sons, Ltd: Chichester.
- Klütsch CFC, Seppälä EH, Fall T, Uhlén M, Hedhammar Å, Lohi H et al. (2010). Regional occurrence, high frequency but low diversity of mitochondrial DNA haplogroup d1 suggests a recent dog-wolf hybridization in Scandinavia. *Anim Genet* **42**: 100–103.
- Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M et al. (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**: 803–819.
- Morin PA, Luikart G, Wayne RK (2004). SNPs in ecology, evolution and conservation. *Trends Ecol Evol* **19**: 208–216.
- Napierala H, Uerpmann H-P (2010). A 'new' palaeolithic dog from Central Europe. *Int J Osteoarchaeol* (doi:10.1002/oa.1182).
- Natanaelsson C, Oskarsson MC, Angleby H, Lundeberg J, Kirkness E, Savolainen P (2006). Dog Y chromosomal DNA sequence: identification, sequencing and SNP discovery. *BMC Genet* **7**: 45.
- Nowak RM (2003). Wolf evolution and taxonomy. In: Mech LD, Boitani L (eds). *Wolves: Behavior, Ecology and Conservation*. University of Chicago Press: Illinois. pp 239–258.
- Pang J-F, Klütsch C, Zou X-J, Zhang A-B, Luo L-Y, Angleby H et al. (2009). mtDNA data indicates a single origin for dogs south of Yangtze River, less than 16 300 years ago, from numerous wolves. *Mol Biol Evol* **26**: 2849–2864.
- Raison MJ (2005). *Determining the Antiquity of Dog Origins*. British archaeological reports: Oxford.
- Savolainen P, Zhang Y-P, Luo J, Lundeberg J, Leitner T (2002). Genetic evidence for an East Asian origin of dogs. *Science* **298**: 1610–1613.
- Schuster SC, Miller W, Ratan A, Tomsho LP, Giardine B, Kasson LR et al. (2010). Complete Khoisan and Bantu genomes from Southern Africa. *Nature* **463**: 943–947.
- Skoglund P, Götherström A, Jakobsson M (2010). Estimation of population divergence times from non-overlapping genomic sequences: examples from dogs and wolves. *Mol Biol Evol* **28**: 1505–1517.
- Tedford RH, Wang X, Taylor BE (2009). Phylogenetic systematics of the North American fossil caninae (carnivore: canidae). *Bull Am Mus Nat Hist* **325**: 1–218.
- Underhill AP (1997). Current Issues in Chinese Neolithic archaeology. *J World Prehistory* **11**: 103–160.
- Vilà C, Seddon JM, Ellegren H (2005). Genes of domestic mammals augmented by backcrossing with wild ancestors. *Trends Genet* **21**: 214–218.
- Vonholdt BM, Pollinger JP, Lohmueller KE, Han E, Parker HG, Quignon P et al. (2010). Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature* **464**: 898–903.
- Wang X, Tedford RH (2008). *Dogs, their Fossil Relatives and Evolutionary History*. Columbia University Press: New York.
- Wayne RK (1993). Molecular evolution of the dog family. *Trends Genet* **9**: 218–224.
- Xue Y, Wang Q, Long Q, Ng BL, Swerdlow H, Burton J et al. (2009). Human Y chromosome base-substitution mutation rate measured by direct sequencing in a deep-rooting pedigree. *Curr Biol* **19**: 1453–1457.



This work is licensed under the Creative Commons Attribution-NonCommercial-No Derivative Works 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>

Supplementary Information accompanies the paper on Heredity website (<http://www.nature.com/hdy>)