International Institute for
Applied Systems Analysis
Schlossplatz 1
A-2361 Laxenburg, Austria

Tel: +43 2236 807 342
Fax: +43 2236 71313
E-mail: publications@iiasa.ac.at
Web: www.iiasa.ac.at

Interim Report **IR-09-078**

mtDNA data indicate a single origin for dogs south of Yangtze River, less than 16,300 years ago, from numerous wolves

Jun-Feng Pang
Cornelya Kluetsch
Xiao-Ju Zou
Ai-bing Zhang
Li-Yang Luo
Helen Angleby
Arman Ardalan
Camilla Ekström
Anna Sköllermo
Joakim Lundeberg
Shuichi Matsumura
Thomas Leitner
Ya-Ping Zhang
Peter Savolainen

Approved by

Ulf Dieckmann
Leader, Evolution and Ecology Program

June 2010

Contents

mtDNA Data Indicates a Single Origin for Dogs South of Yangtze River, less than 16,300 Years Ago, from Numerous Wolves

Research Article

Jun-Feng Pang[1,2 #], Cornelya Kluetsch[3 #], Xiao-Ju Zou[2 #], Ai-bing Zhang[3 #], Li-Yang Luo[1,4], Helen Angleby[3], Arman Ardalan[3,5,6], Camilla Ekström[3], Anna Sköllermo[3], Joakim Lundeberg[3], Shuichi Matsumura[7,8], Thomas Leitner[9], Ya-Ping Zhang[1,2 *], Peter Savolainen[3 *]

1 State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China

2 Laboratory for Conservation and Utilization of Bio-resource, Yunnan University, Kunming 650091, China

3 Department of Gene Technology, School of Biotechnology, KTH-Royal Institute of Technology, Roslagstullsbacken 21, 10691 Stockholm, Sweden

4 College of Wildlife Resource, Northeast Forestry University, No. 26 Hexing Road, Harbin 150040, China

5 National Institute of Genetic Engineering and Biotechnology (NIGEB), 14965/161 Tehran, Iran

6 School of Agronomy and Animal Science, University of Tehran, 4111 Karaj, Iran

7 Evolution and Ecology Program, International Institute for Applied Systems Analysis, Schlossplatz 1, 2361 Laxenburg, Austria

8 Department of Fish Biology and Ecology, Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Muggelseedamm 310, 12587 Berlin, Germany

9 Theoretical Biology and Biophysics, MS K710, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

 # These authors contributed equally to this work.

*** Corresponding authors:**

Peter Savolainen. Department of Gene Technology, School of Biotechnology, KTH-Royal Institute of Technology, Roslagstullsbacken 21, 10691 Stockholm, Sweden. Tel: +46-8-55378335. Fax: +46-8-55378481. E-mail: savo@biotech.kth.se

Ya-Ping Zhang. State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China. Tel: +86-871-5032804. Fax: +86-871-5032804. E-mail: zhangyp@mail.kiz.ac.cn

Key words: dog, Canis familiaris, domestication, mitochondrial DNA

Abstract

There is no generally accepted picture of where, when, and how the domestic dog originated. Previous studies of mitochondrial DNA (mtDNA) have failed to establish the time and precise place of origin because of lack of phylogenetic resolution in the so far studied control region (CR), and inadequate sampling. We therefore analysed entire mitochondrial genomes for 169 dogs to obtain maximal phylogenetic resolution, and the CR for 1,543 dogs across the Old World for a comprehensive picture of geographical diversity. Hereby, a detailed picture of the origins of the dog can for the first time be suggested. We obtained evidence that the dog has a single origin in time and space, and an estimation of the time of origin, number of founders and approximate region, which also gives potential clues about the human culture involved. The analyses showed that dogs universally share a common homogenous gene pool containing ten major haplogroups. However, the full range of genetic diversity, all ten haplogroups, was found only in south-eastern Asia south of Yangtze River, and diversity decreases following a gradient across Eurasia down to only four haplogroups in Europe. The mean sequence distance to ancestral haplotypes indicates an origin 5,400-16,300 years ago from at least 51 female wolf founders. These results indicate that the dog originated in southern China less than 16,300 years ago, from several hundred wolves. The place and time coincide approximately with the origin of rice agriculture, suggesting an origin among sedentary hunter-gatherers or early farmers, and the numerous founders indicate that wolf taming was an important culture trait.

1. Introduction

The dog was probably the first domesticated animal and the only one accompanying humans to every continent in ancient times (Clutton-Brock 1995), and has therefore a central position in human history. However, despite extensive archaeological and genetic research, there is not full agreement on where and when dogs originated. Consequently, the related human culture and the mechanisms for domestication of wolf are unknown. It is clear that dogs originated from wolves (Clutton-Brock 1995; Lindblad-Toh et al. 2005), which historically were distributed throughout most of Eurasia and North America (Nowak 2003). Archaeological evidence indicates that domestic dogs existed by 11,500 years ago (ya) (Davis and Valla 1978; Dayan 1994; Raisor 2005) (earlier dates in Europe have been reported, but the evidence does not seem conclusive (Wang and Tedford 2008), see Supplementary Material for details). However, it has failed to tell where and at how many different places the dog originated, because of the difficulty in discriminating between small wolves and domestic dogs, and the large difference in amount of archaeological work and systematic surveys of animal materials in different parts of the world (see Supplementary Material for details). The earliest reasonably firm archaeological evidence for dog (Raisor 2005) is now at 11,500 ya in SW Asia (Davis and Valla 1978), 10,000 ya in Europe (Chaix 2000), 8,100 ya in America (Morey and Wiant 1992), and 7,100 ya in China (Underhill 1997). However, a morphological feature of the jaw, the "turned-back apex of the coronoid process of the ascending ramus", is found in dogs and in Chinese wolves, but is absent in wolves from other regions (Olsen and Olsen 1977).

Also genetic studies have failed to give definite answers about the time and place of origin. In an initial major study of dog and wolf mitochondrial DNA (mtDNA) data (Vilà et al. 1997), it was suggested that the dog may have originated >100,000 ya, much earlier than indicated by archaeological evidence, and at several different times and places, based on the large age of the main phylogenetic group of dog mtDNA sequences (see Supplementary Material for details about this hypothesis). However, in a later study it was suggested that a single much more recent origin is more probable; the universal sharing of mtDNA haplotypes but highest diversity among East Asian dogs indicated an origin in Asia east of the Urals, possibly ~15,000 ya (Savolainen et al. 2002). However, neither time nor place could be definitely established. Studies of ancient samples have shown that ancient and modern dogs share identical mtDNA haplotypes, and that the American dogs originated from the Old World dog population (Leonard et al. 2002; Malmström et al. 2008).). Notably, in a recent

study of African village dogs (Boyko et al. 2009) it was claimed that the reported high diversity for mtDNA in East Asia compared to other parts of the world (Savolainen et al. 2002), was the result of sampling bias. However, in the present study (see "Results") we show this assertion to be incorrect.

Since neither the time nor exact geographical location for the dog origins was established in Savolainen et al. (2002), the possibility of a very ancient and/or multiple origin, as suggested in Vila et al. (1997), is still mostly maintained in the literature (Lindblad-Toh et al. 2005; Morey 2006; Ostrander and Wayne 2005; Vilà, Seddon and Ellegren 2005; Zeder et al. 2006; Gray et al. 2009, Boyko et al. 2009). Thus, there is no generally accepted picture of the time and place of origin, and the only geographical indication so far (Savolainen et al. 2002) is for a vast territory covering the Asian continent East of the Urals and the Himalayas. Consequently, the human culture that performed the wolf taming, and the mechanisms by which the domestication took place, remain unknown. Knowledge of the exact time and place of origin is necessary for identifying the related human culture. The timing is crucial also for understanding the mechanisms of wolf domestication, since the time of origin reflects the number of founder animals. An origin ~15,000 ya (Savolainen et al. 2002) requires a large number of founders (wolves) to explain today's mtDNA diversity, which would indicate that the taming of wolves was an important custom of the related human culture. An ancient origin (>100,000 ya) (Vilà et al. 1997) could, on the other hand, have involved a single female wolf at a singular chance event.

The earlier studies of dog mtDNA (Vilà et al. 1997; Savolainen et al. 2002) have failed to determine the time of origin and number of founders for dogs because the studied region, 582 bp or less of the control region (CR), does not give the necessary phylogenetic resolution. Since the mutation rate is at most 1 substitution per 40,000 year for this region, the sequences in today's dogs would be largely identical to those of the wolf founders, in the case the dog originated 11,500 ya as indicated by archaeological data. Therefore, if some founder haplotypes (from wolf) differed by just one or two substitutions, they and their respective derived haplotypes would not have resolved into separate, identifiable, haplogroups by today, leaving it impossible to determine the number of founders (see Supplementary Material for a detailed discussion). Accordingly, estimates of the time of domestication based on mtDNA CR data have, for a number of domestic animals, consistently given dates that are much earlier than indicated by the archaeological evidence (Ho and Larson 2006), and this can possibly be attributed to underestimation of the number of founders. Therefore, analysis of a much larger part of the mtDNA genome is necessary for obtaining the resolution needed for

identifying the number of founders and time of origin for dogs. Furthermore, incomplete sampling of dogs has so far hindered an exact determination of the region of origin of dogs. For an effective intraspecific study, knowledge of the full extent of the global genetic diversity, as well as a detailed knowledge of key regional populations, is necessary.

Thus, in order to obtain conclusive information about the time and place of origin of dogs, and the number of founders, improvement of both phylogenetic resolution and phylogeographical fine mapping is necessary. Therefore, we performed the most comprehensive study of dog mtDNA so far, concerning both geographical representation and phylogenetic detail. We first made an almost exhaustive survey of the pattern of geographical diversity for dog mtDNA, by analysis of 582 bp of the CR for 1,543 domestic dogs across the Old World. Since an earlier study indicated an origin for dogs somewhere in eastern Asia (Savolainen et al. 2002) this area was especially densely sampled, revealing considerable differences within eastern Asia. Based on this survey we could select a representative subsample of 169 dogs across mtDNA diversity for analysis of almost the entire mtDNA genome (16,195 bp). This gave a detailed and highly resolved picture of the phylogenetic structure for dog mtDNA, which revealed distinct phylogenetic subgroups and thereby allowed an estimation of the time of origin and number of founders. Through these analyses, we obtain evidence that the domestic dog had a single origin in time and space, and an indication of the approximate place and date as well as the number of founders from the wolf. This also gives some potential clues to how, why and by whom the wolf was domesticated.

2. Materials and Methods

2.1 Samples

1,576 dogs and 40 wolves were studied for 582 bp of the CR, and 169 dogs and 8 wolves for 16,195 bp of the mtDNA genome (excluding repetitive and difficult-to-align regions). A list of all samples, with information about haplotype, geographical origin, and breed/type (if applicable), is given in editable format in Supplementary Datasets S1 and S2. The DNA sequence was generated in this study for 907 of the dogs and 1 wolf for the 582 bp region (yielding 103 haplotypes deposited at GenBank under accession numbers EU816456-EU816558), and for 155 dogs and 2 wolves for the 16,195 bp region (yielding 152 haplotypes

deposited at GenBank under accession numbers EU789638-EU789789). For the remaining samples the DNA sequence was retrieved from GenBank.

The geographical distribution of dog samples, for the analysis of the 582 bp region, is given in table 1, table S1 in Supplementary Material, and in some more detail as follows; Europe: north (n=91) and south (n=59) European continent, Britain (n=108), Scandinavia (n=58), Misc. (n=20); Southwest (SW) Asia: Israel (n=17), Iran (n=51), Turkey (n=25), Misc. (n=40); Africa: North Africa (n=14), central Africa (n=17), southern Africa (n=22), Misc. (n=4); Southeast Asia: Thailand (n=41), Vietnam (n=11), Cambodia (n=7); North China: Heilongjian (n=52), Liaoning (n=6), Hebei (n=17), Shanxi (n=23); Central China: Shaanxi (n=91), Shichuan (n=48); South China: Guangdong (n=14), Guangxi (n=35), Hunan (n=54), Guizhou (n=57), Jiangxi (n=46), Yunnan (n=75); China also Tibet (with Quinghai and Nepal, n=37) and Hainansanya (n=31). Samples were assumed to represent geographical regions based on that they either (i) were from a region (mostly rural villages) with small influx of foreign dogs or (ii) belonged to a breed with known historical geographic origin. Dogs not belonging to a specific breed had mostly specialized morphology and were generally kept by their owners for a specific (often multipurpose) use: as watchdog, as pet, or for herding, hunting or for consumption of the meat. Only a few dogs from Spain, India and Southeast Asia were stray dogs. Importantly, breed dogs outside Europe (e.g. in Siberia, and the Israeli Canaan dog) are mostly not intensely bred, but have been formed by a broad sampling of indigenous dogs. The dog breeds were mostly represented by at most 5 individuals each to avoid sampling bias, the greatest exception being Korean dogs which were almost all of either of two breeds: Jindo (n=53) and Pungsan (n=40). Dogs from Europe (totally 117 dog breeds), Japan, Korea and Siberia belonged mostly to specific breeds, dogs from SW Asia were both of a breed (mainly from Israel and Turkey, and Sighthounds from various regions) and of no breed (Iran), while dogs from India, Africa, China and Southeast Asia were mostly not of a specialized breed. The geographical origin of wolf samples analysed for the 582 bp region is shown in the legend to fig. 1a.

For analysis of mtDNA genomes, 169 dog samples were analysed. The samples represent five of the six main phylogenetic groups for the CR sequence: dog clades A (n=112), B (n=29), C (n=22), D (n=5) and E (n=1). The samples were chosen to cover the mtDNA diversity for dog clades A, B and C according to the CR-based MS networks (fig. 1c, fig. S1 in Supplementary Material, Supplementary Dataset S2), so that nearly every CR haplotype was represented either itself, or by a neighbour haplotype differing by a single substitution. Some haplotypes, primarily UTs (Universal Types: the 14 CR haplotypes that

were represented in both Europe, SW Asia and East Asia, see Results), were represented by several samples. For those parts of clades A, B and C shared between the West (Europe, SW Asia, India and Africa) and Eastern Asia (fig. 1b), 63 samples were from the West, and 58 from Eastern Asia, while 41 additional samples from Eastern Asia were from parts of clades A and B unique, or almost unique, to Eastern Asia. Information about the 169 dog samples (haplotype for the 582 bp region, geographical origin, breed/type [if applicable], and GenBank accession number) is given in editable format in Supplementary Dataset S2. The origin of wolf samples is shown in fig. 2a.

2.2 DNA Sequence Analysis

For samples analysed for the DNA sequence in this study, DNA was extracted from blood or hair (Savolainen et al. 2002), or buccal swabs (Natanaelsson et al. 2006), and analysed by PCR and DNA sequencing (determined in both forward and reverse direction for all nucleotide positions; Angleby and Savolainen 2005) as described. The aligned DNA sequences for the CR and the mtDNA genome sequence data, and PCR primers for the mtDNA genome, are given in editable format in Supplementary Dataset S3.

2.3 Phylogenetic Analyses

The phylogenetic tree for the whole mtDNA genome (WG) data was inferred by an optimized and parallelized ML code based on fastDNAml and DNArates (Olsen et al. 1994), in iterative steps to refine the tree and substitution model parameters, as described (Korber et al. 2000). The phylogenetic tree for the 582 bp region of the CR was inferred using a neighbor-joining method (BioNJ) with a HKY+I+G model (I=0.7799, shape=0.5921) and maximum likelihood (ML) evaluated ingroup midpoint rooting as described (Savolainen et al. 2002; Savolainen et al. 2004). The inference robustness of the clades for the CR tree was evaluated by non-parametric bootstrapping (1000 replicates) using BioNJ with a HKY+I+G model, and for the WG tree by bootstraping as well as Bayesian analysis. Minimum spanning networks (MS networks) were inferred using Arlequin (Excoffier, Laval and Schneider 2005), and drawn manually.

2.4 Calculation of substitution rate and time estimates

The average number of substitutions accumulated since the separation of the dog/wolf and coyote lineages for the WG data was estimated by using branch lengths in the ML tree without a clock constriction as well as Bayesian Markov Chain Monte Carlo (BMCMC) searches explicitly assuming various clock and population models. For the non-clock constrained estimate the average tree height for the dog/wolf clade was calculated after a least-squares optimization of the rooting point (TreeRate. beta0.9 ed. http://www.hiv.lanl.gov/content/sequence/TREERATE/combinedBranchlength.html; Athreya G, Kothari M, Maljkovic Berry I and Leitner T, unpublished data). The BMCMC estimates were performed using BEAST (Drummond and Rambaut 2007) with a constant clock (Poisson distributed) with three different population growth models (constant size, exponential growth and a skyline coalescent growth model), and with a relaxed exponential clock and a relaxed lognormal clock with the skyline coalescent growth model. All runs were 10.000.000 or 50.000.000 MCMC steps with sampling every 1000 steps, and a 10% burn-in. We used tree height estimates from each clock model only if the estimated sample size reached above 100. The lower and upper 95% highest probability distribution cut-offs were estimated to examine the credibility interval for the tree heights of each clock and coalescent model combination. Subsequently, the substitution rate for the WG data was derived by dividing the tree height with the separation time of dog/wolf and coyote according to the fossil record (1.5-4.5 million years ago, see Results). The mean distance to the most recent common ancestor (MRCA) for clades and sub-clades was estimated through the BMCMC searches. The time of geographical expansion of clades A, B and C was calculated using the statistic ρ (the mean number of substitutions for a set of sequences to their common ancestral haplotype [MRCA]) (Forster 1996), by calculating the mean value of ρ (summing several sub-subclades with different MRCAs identified in the ML tree) for each clade. The standard error of mean for ρ was calculated by resampling, with the same size as the original number of individuals, in 1000 replicates. For the control region (582 bp), the average number of substitutions accumulated since the separation of the dog/wolf and coyote lineages was calculated from the average genetic distance between dog/wolf and coyote in the phylogenetic tree (fig. 1a), and estimated at 0.057 substitutions $site^{-1}$ (range 0.050-0.065, considering variation within the dog/wolf clade and the coyote clade). Calibrated with the time for the separation between wolf and coyote (1.5-4.5 million years ago, see results), this gives a rate of $6.4x10^{-6} – 2.5x10^{-5}$ substitutions $site^{-1}$ $year^{-1}$, or 1 substitution per 40,000-155,000 years.

3. Results

3.1 Dog mtDNA haplotypes are distributed across the wolf diversity, in six phylogenetic clades

We analysed 582 bp of the CR for 1,543 dogs from across the Old World, 33 dogs from Arctic America, and 40 Eurasian wolves (table 1, table S1 in Supplementary Material). Phylogenetic analysis (fig. 1a) grouped all dog sequences into the previously described (Vilà et al. 1997; Savolainen et al. 2002) six distinct haplogroups, clades A-F. In similarity to earlier studies of the CR, poor support (bootstrap values <50%) was found for clades A and B (fig. S2 in Supplementary Material). However, the analysis of entire mtDNA genomes (see below) recreated all clades with high support by neighbor-joining (NJ) based bootstrap values (≥92%) and Bayesian values (100%), confirming the CR based topography. The six dog clades were dispersed across most of the distribution of haplotypes for Eurasian wolf (fig. 1a). Furthermore, the two major dog haplogroups, clades A and B, contained wolf haplotypes: clade A had three haplotypes from North Chinese and Mongolian wolves differing by one or two substitutions from the closest dog haplotypes; clade B had four wolf haplotypes, two (from North Chinese and Romanian wolf) that were identical to dog haplotypes, and two (from Afghani and Yugoslavian wolf) that differed by two substitutions from the closest dog haplotypes. However, the available sample of extant wolf is restricted, and extinction of wolf in large parts of the world, e.g. most parts of Europe and large parts of southern Asia, precludes its use for completely recreating the diversity for wolf at the time of domestication. Therefore, instead of drawing conclusions from comparison of dog and wolf populations, we perform an intraspecific study of the distribution of genetic diversity for extant dog across the world, in order to trace it back to the place and time for the domestication of wolf.

3.2 Dogs across the Old World share a common homogenous mtDNA genepool, but diversity follows a gradient from high values in southeastern Asia to low in Europe

Among the dogs, the gene pool of the Old World is remarkably homogenous in several ways. Clades A, B and C were represented in every population, by totally 97.4% of all dogs

and in most regions by 100 % of the dogs (table 1, table S1 in Supplementary Material, fig. 1b) (clades D, E and F were rare and geographically restricted, and possibly derive from post-domestication wolf-dog hybridisation, they are therefore not further considered in this study: see Supplementary Material for a detailed discussion). The proportion of individuals having clades A, B and C was also very similar among the geographical regions across the Old World (fig. 3a, table 1). A distinct example of this is Britain and Japan, two island groups situated outside the opposite parts of the immense Eurasian continent (Britain: 75.9%, 20.4% and 3.7%; Japan: 64.4%, 20.3% and 15.3%, of A, B and C). Finally, the majority of individuals had one of a relatively small number of haplotypes that were shared by virtually every population of the Old World, as demonstrated in the minimum-spanning networks (MS networks) (fig. 1b). Fourteen of these haplotypes (9 in clade A, 2 in clade B and 3 in clade C) were represented in Europe, Southwest Asia as well as East Asia (China, Southeast Asia and Japan), and we termed these "universally" occurring haplotypes Universal Types (UTs). The UTs were universally very frequent (fig. 1b, table 1), especially in the western part of the Old World (west of the Urals and the Himalayas, which we term "the West") where 81.5% of dogs in Europe and SW Asia had a UT (table 1, fig. 3a) and consequently the same haplotype as a dog in East Asia. However, the frequency of UTs was considerably lower in East Asia (54.2%), and in the extreme south-eastern part of Asia (China south of Yangtze River and Southeast Asia, a region which we term "Asia south of Yangtze River" [ASY]) it reached a minimum of 40.8% for the whole of ASY, with a range between 18.4 and 54.7% among its provinces (fig. 3 and table S1 in Supplementary Material).

It is even more striking that, except for the UTs, almost all other haplotypes in the West differ by a single mutation from a UT (fig. 1b). Thus, 98.7 % of dogs in Europe and 94.6 % in SW Asia had a haplotype which is either a UT or derives from a UT by a single mutation (which we collectively term UT-derived: UTd) (table 1, fig. 3a). This implies that almost all dogs in the West had a haplotype that can be traced to a haplotype found also in East Asia. In contrast, 69.2 % of dogs in East Asia and only 53.0% in ASY had a UTd. Thus, in addition to the great genetic homogeneity among regions, there is also a distinct difference in diversity across the Old World For virtually every measure of diversity, e.g. the frequency of UTs and UTds, and the number of haplotypes (also when adjusted for different sample size) and unique haplotypes, there is a distinct maximum in ASY (table 1, fig 3a and 3b, table S1 in Supplementary Material). The genetic diversity follows a gradient from the maximum levels in ASY, decreasing through East Asia and further across Eurasia to low values in Europe at the other end of the continent. The difference in genetic variation among regions is directly

visible as a difference in coverage of the MS networks, most distinctly for clade A (fig. 1b). Western populations lack several parts of clade A, and it is largely the same parts missing for all these populations. Eastern populations have a more complete coverage, but the only region with almost complete representation across clade A is ASY.

Thus, populations in the West have a haplotype pool consisting almost exclusively of the 14 UTs and surrounding haplotypes. East Asian populations, and in particular ASY, have a large proportion of dogs with unique haplotypes on a large distance from the UTs, and consequently from haplotypes found in the West. It is also noteworthy that all the western populations lack almost exactly the same parts of the MS networks for clades A and B; the representation is largely identical for Europe and SW Asia. Therefore, the remarkably low genetic diversity for European dogs does not seem to be caused by bottlenecks in connection with the relatively recent development of the European dog breeds (Clutton-Brock 1995), but must stem from a time before the formation of the European and SW Asian dog populations.

Importantly, we also compared the data to the recent study by Boyko et al. (2009), in which it was claimed that African village dogs have at least as high diversity as East Asian dogs, measured as the number of haplotypes (ignoring indels) per sampled individual. They therefore question the conclusions by Savolainen et al. (2002) about an East Asian origin of dogs, arguing that the results (high diversity in East Asia) were biased from sampling village dogs in East Asia and primarily (inbred) breed dogs elsewhere. However, the African village dog sample (Boyko et al. 2009) had 41 haplotypes among 318 dogs, to compare with, e.g., 71 haplotypes among 281 dogs in the sample from South China in the present study (table 1; see also table S1 in Supplementary Material for subregions in southern East Asia, all but one considerably more diverse than the African subpopulations in Boyko et al.(2009)). Thus, a direct comparison shows that the smaller South Chinese sample has 73% more haplotypes than the African one; the assertion by Boyko et al. (2009) is the result of not adequately compensating for differences in sample size between the relatively small East Asian samples in Savolainen et al. (2002) and the larger African samples. The African sample has also all the other characteristics of the "western" dog populations: The haplotypes fall in the same parts of the MS networks as for other western populations, leaving large parts unique to East Asia (data not shown); and values are high for UT (66.7%) and UTd (90.9%), and number of unique haplotypes low (12) (compare with e.g. South China: UT (42.0%), UTd (53.4%), and number of unique haplotypes (40; i.e. only one less than the total number of haplotypes in the African sample!)). To conclude, the sample of African village dogs in Boyko et al. (2009), like all "western" samples, has considerably lower genetic variation than the populations in

ASY. There are, therefore, no indications that the results in Savolainen et al. (2002) or the present study are influenced by a "village dog sample bias".

3.3 Analysis of complete mtDNA genomes reveals 10 subclades in clades A, B and C, with geographical representation following the East-to-West gradient

There is clear difference in coverage of clade A among geographical regions, especially between ASY and the rest of the world (fig. 1b). This indicates that clade A, rather than being a single dense clade, may consist of several different phylogenetic subgroups with different geographical spread, groups that cannot be resolved based on the CR data. To study this geographical pattern in detail, and to obtain sufficient resolution for dating the dog origins and estimating the number of founders, we analysed almost the entire mtDNA genomes for 169 dogs and 8 wolves (16,195 bp analysed, repetitive and difficult-to-align regions were excluded). The samples were chosen to cover most of the mtDNA diversity for dog clades A, B and C according to the CR-based MS networks (fig. 1c), for the West (Europe, SW Asia, India and Africa) as well as for East Asia (Supplementary Dataset S2, fig. S1 in Supplementary Material). Phylogenetic analysis of the mtDNA genomes improved the resolution considerably, compared to analysis of the CR (fig. 2a). The two major phylogenetic clades, A and B, which were weakly supported in the CR based tree, obtained Bayesian support values of 100% in the genome based tree (fig. S3 and S4 in Supplementary Material).

More importantly, the analysis also revealed a distinct substructure within clades A, B and C. Thus, the seemingly dense clades A, B and C are composed of a substructure of subclades (fig. 2a and b). Clade A had six major subclades, and B and C two each, giving a total of 10 subclades (or haplogroups), with high bootstrap and Bayesian support values (fig. 2a, fig. S3 and S4 in Supplementary Material), and separated by relatively large genetic distances (fig. 2b). For the CR part of the genome sequences, the 10 subclades group almost perfectly in separate parts of the MS networks (fig. 1c). Importantly, 5 of the 6 subclades of clade A, corresponding to those parts of the CR-based MS network which are empty for populations in the western populations, were found only in East Asia (fig. 2b). Accordingly, when all 1,576 CR sequences are assorted into the 10 subclades based on diagnostic mutations (see Supplementary Material for details), the geographical distribution of the subclades follow a distinct gradient; the complete set of 10 subclades is found only in ASY, while 7 are represented in Central China and Japan, 5 in North China, India and SW Asia, and

only 4 in Europe (table 2, fig. 3a and 3b, table S2 in Supplementary Material). Only 1 of the 6 subclades of clade A is represented in Europe and SW Asia, and the missing 5 subclades correspond to the empty parts of the CR-based MS networks (fig. 1b and 1c). To conclude, the full extent of diversity for clades A, B and C, all the 10 major phylogenetic groups, is represented in the region comprising China south of Yangtze River and Southeast Asia, ASY. Outside this region only part of the total diversity is found, but it can be traced to a subset of the gene pool in ASY, basically the 14 universally occurring haplotypes, the UTs, which are distributed in 4 of the 10 subclades. Thus, the facts that nearly 100% of dogs in Europe and SW Asia have CR-based haplotypes closely related to the 14 UTs while Eastern populations have a large number of unique and distinct haplotypes, and that parts of the CR-based MS networks are empty for the western populations, can be attributed to the almost complete absence of 6 out of the 10 major phylogenetic groups in the western part of the Old World. Within ASY, there was no single subregion having all 10 subclades, but in relatively small samples from Yunnan (n=75), Southeast Asia (n=59) and Guizhou (n=57), 9, 9 and 8 subclades, respectively, were found (fig. 3b, table S2 in Supplementary Material). The smallest region containing all ten haplogroups comprises Yunnan and Southeast Asia, in the southwest of ASY. The simplest explanation for the observed geographical distribution of the 10 subclades of clades A, B and C is that they had a single origin within or close to ASY, and that only a subset of the original gene pool spread to the rest of the world.

3.4 Similar proportions of clades A, B and C across geographical regions indicates a simultaneous origin for the three clades

A strong indication that clades A, B and C actually originated at a single place and time is given by the very similar proportion of the three clades, among populations in different parts of the Old World (fig. 3a, table 1). The simplest explanation for this pattern is that all populations worldwide originate from a single population having approximately these proportions of clades A, B and C. If the clades had originated in different regions, from separate independent domestications of wolf, a majority of each clade in their respective region of origin would be anticipated. Had they originated at different times (the younger clades then from wolf-dog hybridisations in already established dog populations), the younger clades would have had difficulty spreading to already populated regions (the latter is possibly the case for clades D, E and F, see Supplementary Material). Only very thorough mixing across the entire Eurasian continent could have counteracted these patterns. Thus, had the

three clades originated in different regions and/or at different times, a large proportion of regionally unique and distinct haplotypes, or at least distinctly varying proportions of the clades, would be expected. In order to test whether multiple origins in time or space are compatible with the observed homogenous distribution of clades A, B and C, we carried out population genetic simulations based on a simple stepping-stone model (shown in Supplementary Material). The proportions of the three clades, obtained under different scenarios for dog origins (single or multiple origins in time and/or space), were followed for five regions across Eurasia (Britain, Continental Europe, Southwest Asia, China and Japan) and compared to the observed values for these regions. This analysis showed that multiple origins in time are virtually impossible, demanding extreme migration rates for sufficiently mixing all populations, in order to end up with today's even proportions of the clades. This implies that it is highly unlikely that clade B or C would have originated after clade A, through hybridisation between dog and wolf. Multiple origins in space, if occurring almost simultaneously, are also unlikely unless the migration rate was very high (e.g. 30% of dogs in each population migrating to the neighbouring population, in every generation since the time of origin, if an origin 20,000 ya and a four year generation time (Fuller, Mech and Cochrane 2003) is assumed). The only scenario not rejected at moderate migration rates is a single origin in time and space for the three clades. Therefore, the most probable reason for the similar proportions of clades A, B and C is that all three clades originate from domestication of wolf at a single time and place.

3.5 Robust estimate of dog and wolf evolutionary rate

For estimation of the age of phylogenetic groups, we calculated the substitution rate for the mtDNA genome data. The rate of substitutions over time was estimated using branch lengths in the ML tree without a clock constriction as well as Bayesian Markov Chain Monte Carlo (BMCMC) searches explicitly assuming various clock and population models. The height estimated from the ML tree was 0.031 substitutions site^{-1}, and the BMCMC estimates (skyline growth model) were only slightly lower at 0.0296 [0.0269-0.0325, 95% limits of the highest probability distribution (HPD)] substitutions site^{-1} for the Bayesian constant clock and 0.0288 [0.0205-0.0371] substitutions site^{-1} for the Bayesian relaxed lognormal clock. The Bayesian relaxed exponential clock could not be estimated reliably; although we ran 50.000.000 states with sampling every 1000, the estimated sample size (ESS) for the tree

height only reached 18 (while posterior and likelihood ESS were high; 488 and 4064, respectively). A Bayes factor analysis showed that the log normal relaxed clock was better than the constant clock (log_{10} Bayes factor difference of the harmonic means of the tree likelihoods were 5.7, where values above 5 indicate "substantial" support) (Suchard, Weiss and Sinsheimer 2001; Drummond and Rambaut 2007; Jeffreys 1998). Furthermore, the distance estimates for all clades and sub-clades to their most recent common ancestors (MRCAs) using all clocks agreed very well with that of the non-constricted tree estimates (R>0.92, table S3 in Supplementary Material) without any significant differences (P>0.16, t test). Thus, we had a robust estimate of the average substitution rate across all clocks tested with a confidence range that may follow the relaxed lognormal clock. We therefore estimate the substitution rate at 0.0288 [0.0205-0.0371, 95% HPD] substitutions $site^{-1}$ $time^{-1}$, where *time* is the number of years since the split of dog/wolves and coyotes.

Since the fossil record is incomplete, there is no exact calibration point for the separation time of the dog/wolf and coyote lineages. The lineages were well separated by ~1,5 million ya, but it is not clear when the actual split occurred; a separation 1,8-2,5 million ya seems most probable, but it possibly occurred up to 4.5 million ya (Nowak 2003). We therefore used the time range 1.5-4.5 million ya as calibration value, giving the time-calibrated substitution rate as a range: $6.4x10^{-9}$ – $1.92x10^{-8}$ substitutions $site^{-1}$ $year^{-1}$, [with 95% HPD $4.56x10^{-9}$ – $2.47x10^{-8}$ substitutions $site^{-1}$ $year^{-1}$] or 1 substitution per 3,200 – 9,600 years [2,500 – 13,500 years].

3.6 Time to MRCAs for samples across Eurasia indicate a simultaneous spread of clades A, B and C across the Old World 5,400-16,300 ya

If clades A, B and C originated from wolf haplotypes approximately simultaneously, it should be possible to identify phylogenetic subgroups, each originating from a single wolf haplotype (the founder haplotype for the dog haplotypes), of similar age for all three clades. For the ten major subclades, the distance to the MRCAs varies considerably, between $1.96x10^{-4}$ ($1.29x10^{-4}$ - $2.78x10^{-4}$, 95% HPD) and $9.45x10^{-4}$ ($7.82x10^{-4}$ - $1.13x10^{-3}$) substitutions per site (fig. 2a and b, table S3 in Supplementary Material), corresponding to a time to the MRCAs of 10,200-30,600 years (6,700-43,400 years, 95% HPD) for the youngest and 49,200-147,000 (40,700-176,600) years for the oldest subclade. However, most of the subclades contain a substructure of younger sub-subclades, possibly originating from separate wolf haplotypes (fig. 2a and b). It is not possible to identify at which level of these subclades

and sub-subclades the wolf haplotypes (dog founder haplotypes) are situated, and thereby date the origin of the dogs. However, there are several sub-subclades, formed primarily by samples having the universally occurring haplotypes (the UTs) for the CR sequence, which are represented by samples from across the Old World (fig. 2b). By dating the time to the MRCAs for these universal sub-subclades, the time since the spread out of the region of origin for these lineages can be estimated. Thus, in the MS networks (fig. 2b) we identify twelve sub-subclades in clades A, B and C, each with a common still existing MRCA (fig. 2a), for samples from Eastern Asia as well as from Europe and/or SW Asia and Africa (MRCA A19_3 [with derived haplotypes for Europe/SW Asia/ Africa: A21, A19_2]; A18_1 [A36, A18_4, A18_5, A18_6, A165, A18_8, A27]; A11_1 [A11_6, A26_1]; A11_2 [A11_3]; A11_4 [A11_5]; A2_1 [A2_2, A1]; A16_1 [A33_1, A33_2]; A17_1 [A17_2, A30]; B1_1 [B1_12, B1_11, B12, B1_3, B1_4, B1_7]; B6_1 [B8, B6_2]; C3_1 [C3_5, C3_7, C3_8]; C1_1 [C15]). For the European, SW Asian and African samples, the mean distance in substitutional steps to the universal MRCAs, the statistic ρ (Forster 1996), are similar for clades A, B and C: 1.78 (s.e.m. = 0.068), 1.64 (0,034) and 1.5 (0,40) substitutions, respectively. Assuming an origin in East Asia for these subclades, this indicates that clades A, B and C spread globally simultaneously, in agreement with the simulation analysis. The mean distance to the MRCAs for all three clades is 1.70 (s.e.m. = 0.035) substitutions, corresponding to a time to the MRCAs (and thus spread of dogs from East Asia) 5,400-16,300 ya [4,100-24,000 ya, 95% HPD]. This agrees with the archaeological evidence for dogs in Europe by at least 10,000 ya and in SW Asia probably by 11,500 ya, if the dogs spread across the Eurasian continent within a few thousand years. To conclude, it is not possible to directly date the origins of dogs from this data set, but the date of spread from the centre of origin across Eurasia can be estimated. Assuming, further, that the dog spread shortly after the domestication of wolves, the genetic data indicates that dogs originated approximately 5,400-16,300 ya or shortly earlier. Considering also the archaeological evidence, an origin for dogs 11,500-16,300 ya would be indicated. Importantly, for most of the individuals having a UT (a haplotype identical in the CR for samples from across Eurasia), the samples from the West and from East Asia do not have an identical haplotype when the whole mtDNA genome is considered (fig.2b). Since one substitution corresponds to 3,200 – 9,600 years for the mtDNA genome, this implies that the sharing of haplogroups and UTs across Eurasia is not caused by recent (during the last few thousand years) mixing of the populations, but stem from ancient events.

3.7 An origin of dogs ≤16,300 ya entails a minimum of 51 female wolf founders

As noted above, it is not possible to identify exactly which haplotypes were carried by the domesticated wolves, and thereby count the number of founders for the dog. However, we may estimate the minimum number of lineages, for the whole genome data, that existed at the time of spread of dogs across the Old World (fig. 2b). The probability is low that more than 6 substitutions have occurred in any of the 169 dog lineages since the time of the global spread of dogs, 1.70 substitutions back in time ($p<0.002$ per lineage, Poisson distribution). Therefore, haplotypes separated by more than 12 substitutions (having a MRCA more than 6 substitutions back in time) should originate from different lineages existing at the time of spread. Counting the minimum number of groups of haplotypes with a maximum distance of 12 substitutions between haplotypes within the group, we identify 51 lineages (20 lineages in subclade a1; 12 in a2; 3 in a3; 1 in a4; 5 in a5; 1 in a6; 2 in b1; 4 in b2; 2 in c1; 1 in c2) leading to today's dog haplotypes (fig. 2b). Thus, assuming that the 10 subclades of clades A, B and C formed almost simultaneously and shortly before spreading to the rest of the world, today's mtDNA gene pool must have originated from a minimum of 51 wolf haplotypes. This number probably represents a conservative estimate since: (i) it is possible that some haplotypes differing by 12 substitutions or less originate from different founder lineages, (ii) there may be lineages present among extant dogs not detected in our data set, and (iii) some lineages may have become extinct since the time of domestication. Furthermore, some of the domesticated wolves may have had identical haplotypes. It is therefore reasonable to assume that several hundred female wolves were domesticated. It is possible that some of the 51 lineages, the ones that are unique to East Asia (basically haplotypes of subclades a2-a6 and b2) originate from hybridisation between dog and wolf, in ASY, at some time after the spread of dogs across the Old World, rather than from the initial domestication. However, outside ASY there are indications for hybridisation between female wolf and male dog in only three cases, two (in Scandinavia and SW Asia) forming clade D, and one (in Japan) forming clade F (see Supplementary Material for a detailed discussion). Therefore, hybridisation between female wolf and male dog leading to offspring passed on in the dog population seems rare.

4. Discussion

4.1 Summary of the mtDNA data

This study shows that there is a common gene pool for mtDNA among dogs across the Old World; in most populations 100% of the dogs have a haplotype belonging to the main phylogenetic groups, clade A, B and C, the proportions of the three clades is very similar among populations, and a majority of individuals have a haplotype (for the CR) which is shared all across the Old World. However, the full extent of diversity, all ten major subclades within clades A, B and C, is found only within a region in south-eastern Asia south of the Yangtze River, which we call Asia South of Yangtze, ASY. From this region and across the Old World the diversity decreases by a gradient, to a minimum in Europe, for the representation of subclades as well as other measures of diversity; the number of haplotypes and unique haplotypes are considerably higher in ASY than in any other region. Importantly, the gene pools in all other populations across the Old World consist of a subset of haplotypes that are identical to, or very similar to, the haplotypes found in ASY. Furthermore, simulations show that the even proportions of clades A, B and C across Eurasia strongly indicates that the three clades originated almost simultaneously, and probably also in a single place. In accordance with this, dating of the genetic distance to universal MRCAs indicates that the three clades spread across the Old World at the same time, approximately 5,400-16,300 ya. Finally, at this time there existed at least 51 different mtDNA lineages leading to today's dogs.

Thus, with this data we establish the earlier findings (Savolainen et al 2002) that mtDNA diversity is much higher in East Asia than in other parts of the world, rejecting the assertion (Boyko et al. 2009) that these results were derived from sampling bias. More importantly, with the increased phylogeographical resolution we can identify a much more precisely defined region within East Asia, ASY, harbouring considerably higher diversity than all other regions. We also obtain for the first time reasonable estimates of the time of origin and number of founders from the wolf. We can therefore definitely dismiss the assertion (Vila et al. 1997) that mtDNA data indicates an origin of dogs much earlier than the 10.000-15,000 years ago indicated by the archaeological record (see Supplementary Material for a discussion).

4.2 Validity of the mtDNA data

In this study we try to reconstruct the population genetic history of the mtDNA gene pool among dogs, and by that means draw conclusions about the origin of the dog. It is therefore important that the phylogeographical patterns we study stem mainly from ancient population events, and have not been disturbed by more recent bottlenecks and migrations. For example, the European dog population has a very special history compared to other regions, with development of a large number of specialized dog breeds during the last few hundred years which normally involved severe genetic bottlenecks (Clutton-Brock 1995). Therefore, one might suspect that the low genetic diversity for the European dogs, for example the presence of only 4 out of the 10 major phylogenetic subclades (fig. 3a, table 2), might have been caused by these recent bottlenecks, rather than by ancient events. However, it is very unlikely that exactly the same 6 subclades would have been independently lost in all the separate bottlenecks leading to the 117 different European dog breeds in this study. More importantly, the gene pool for the European dogs is almost identical to that of the other regions in the West, e.g. SW Asia, in that they all lack the same subclades (fig. 1b, table 2). Therefore, the 6 mtDNA subclades must have been missing already before the European dog population, and the other western populations, was originally formed. Obviously, while the formation of the European dog breeds must have constituted a severe bottleneck for each breed, it does not seem to have severely reduced the diversity for the population of European dogs as a whole. Thus, the remarkable fact that the European dog population, which harbours the largest morphological variation with several hundred dog breeds, has the lowest genetic diversity across Eurasia (according to the proportions of UTs and UTds it has approximately 50% of the original genetic variation among dogs) reflects its position as the most peripheral of the Eurasian populations, and not modern breeding history. Another concern might be that some of the 10 major subclades, rather than originating from ASY, may have originated in Europe and then spread to East Asia, for example in connection with the European colonization of Asian countries in the last few hundred years, thereby giving the full representation of subclades only in ASY (for example if subclade a1 originated in Europe and a2-a6 in ASY, and a1 spread from Europe to East Asia giving representation of all subclades of clade A only in ASY). However, almost no samples from Europe and East Asia, respectively, had identical mtDNA genomes (fig. 3b); obviously the European and East Asian haplotypes separated several thousand years ago. Finally, we have rejected the assertion (Boyko et al. 2009) that the high genetic diversity for East Asian dogs is an artifact caused by

sampling of village dogs in East Asia and breed dogs elsewhere. Importantly, in the present study the majority of dogs in most regions, except Europe, are non-breed dogs, and as discussed above, even for the European population the breeding of dogs cannot be accountable for the low mtDNA diversity. We therefore conclude that the phylogeographical patterns for mtDNA, most prominently the universal sharing of CR haplotypes and the gradient of diversity across the Old World, have not been severely influenced by modern population events or by sampling bias but seem to reflect ancient events in dog history.

Furthermore, it has been noted, for a number of domestic animals including the dog, that estimates of the time of domestication that are based on mtDNA data have given dates that are much earlier than indicated by the archaeological evidence (Ho and Larson 2006), casting doubt upon the use of mtDNA data for dating domestication events. However, these studies were all based on analysis of only the CR, which, because of lack in resolution, does not seem suitable for such calculations (see Supplementary Material for a detailed discussion). In the present study, we show that the number of founders for the dog is considerably underestimated when the CR is used for identifying phylogenetic groups with separate origin from wolf haplotypes, but that analysis of mtDNA genomes gives the necessary resolution and results in a dating of dog origins in good agreement with the archaeological data.

4.3 Conclusions drawn from the mtDNA diversity pattern

The observed phylogeographical pattern for dog mtDNA (most importantly the presence of all major phylogenetic groups only in ASY, and the gradual decrease in number of groups with the distance from ASY) is a strong indication that the domestic dog had a single origin in space and time from ASY. The pattern is similar to that observed for human mtDNA (Ingman 2000), which gives a strong argument for the “Out of Africa” scenario for human origins. Similarly, the simplest explanation for the emergence of the dog mtDNA diversity pattern is a single origin for dogs in ASY and a gradual loss of diversity as the dog spread around the world. Other scenarios cannot be ruled out, but demand more complex explanations.

There are a number of conclusions that can be drawn from this data set, but with different degrees of certainty. To begin with, there are a few conclusions that seem solid: (i) Given the large extent of diversity in the small sample of wolves in this study (fig. 1a), it is unlikely that universally identical or almost identical haplotypes among dogs would derive from several independent domestications of wolf in different parts of the world. Therefore,

the universal sharing of clades A, B and C, and of the 14 UTs therein, strongly indicates that dogs worldwide have a common origin from the same wolves, either from a single geographical region for all three clades, or from a different place for each clade followed by very effective mixing of haplotypes. It can therefore be concluded that dogs across the world share a single gene pool originating from the same wolves, and in this sense there is clearly a single genetic origin for all dogs. (ii) The largest number, and largest variety, of mtDNA lineages exist in ASY. Furthermore, practically all haplotypes outside ASY may potentially derive from a subset of the gene pool in ASY, but not vice versa. We can therefore conclude that the largest part of the genetic diversity for mtDNA among dogs, and possibly all the diversity (except clades D, E and F), originated in ASY. Thus, a single origin for all dogs in ASY is possible, but a single origin outside ASY seems impossible. (iii) The similar proportions of the frequencies for clades A, B and C across regions strongly indicates that the three clades originated from wolf approximately simultaneously. Simulations show that it is virtually impossible for haplotypes from later domestications or wolf-dog crossbreedings to spread effectively into already populated regions. Accordingly, the mean distance to universal MRCAs for samples in western Eurasia are similar for clades A, B and C, indicating a simultaneous origin of the three clades. Thus, it can with great certainty be concluded that all domestic dogs originate from a single gene pool, which derives from a number of wolves which were domesticated at approximately the same time, and that the main contribution, and possibly the total contribution, of genetic diversity from wolf into the original dog gene pool comes from ASY.

There are also a number of inferences which are not equally certain but seem the most plausible: (i) The similar proportions of clades A, B and C across Eurasia (fig. 3a), and simulations about their distribution under different origin scenarios and migration rates, show that unless there has been an extreme rate of migration across the continent during thousands of years they must have had a common geographical origin. Furthermore, the full genetic diversity for clades A, B and C exists in ASY, and only in ASY, and the diversity decreases following a gradient across the Eurasian continent. All this indicates that there was a single origin in time and space for clades A, B and C, in ASY. (ii) The time to universal MRCAs for samples in western Eurasia are similar for clades A, B and C, approximately 5,400-16,300 years, indicating a simultaneous spread of the three clades, in accordance with the simulations. (iii) At the time of this spread, there existed at least 51 mtDNA lineages leading to today's dog mtDNA haplotypes. This indicates that, unless large-scale crossbreeding with

female wolves have later occurred in ASY, at least 51 female wolves with different mtDNA haplotypes, and therefore probably several hundred wolves in total, were domesticated.

The large number of founders agrees with studies of the dog MHC, identifying 41 founder alleles and therefore a minimum of 21, but according to simulations more probably several hundred, wolf founders (Vilà, Seddon and Ellegren 2005). Importantly, the dogs were almost entirely of European breeds which, judging from the mtDNA data in the present study, probably harbour only a subset of the global allele population. The many founders were hypothesized to come largely from hybridisation between dogs and wolves, but based on the present mtDNA data they more probably originate from the original domestication of wolves.

A large number of founders is also suggested by demographic modelling of autosomal SNP data in wolf and dog (Gray et al. 2009). The simulations suggest that the domestication of wolves resulted in a modest population contraction and a small loss of nucleotide diversity, indicating that dogs originated from a large number of wolves.

Considering all this, we conclude that the mtDNA data in this study strongly indicates a single origin in time and space common to all domestic dogs, in southern East Asia approximately 5,400-16,300 ya (11,500-16,300 ya taking archaeological evidence into account), and that several hundred wolves were domesticated in this process. Alternative explanations for the dog origins cannot be excluded but demand far more complicated scenarios (see Supplementary Material for a discussion).

The exact part of ASY, or its surroundings, where the dog may have originated is not clear. Several south and central Chinese provinces are not represented in this study, and for the regions represented sample sizes are only around 50. The full representation of diversity for clades A, B and C, all 10 major subclades, was found in a region encompassing Yunnan and Southeast Asia, but diversity was high also in other parts of ASY. For example, Guizhou has a higher number of haplotypes than Yunnan, when normalised for sample size, and only one subclades less (fig. 3b, table S1 in Supplementary Material). It is possible that the domestication of wolves was a widespread practice that took place in several parts of ASY. The most dramatic decrease of genetic diversity across the Old World is between ASY and North China. For example, the number of subclusters decreases from 10 to 5 and the proportion of individuals having a UTd increases from 53.4% to 89.8% (fig. 3a, table 1). In fact, the North Chinese diversity is more similar to that of the West; the similarity in phylogenetic coverage between North China and the western regions is visible in fig. 1b, and values for pairwise difference, Φ_{ST}, for North China are lower compared to Europe, SW Asia and Japan than to ASY (data not shown). A scenario where subclades a2-a6 and b2 mainly

originated in the south and the universal subclades a1, b1,c1, c2 more to the north of ASY, from where the universal spread would then have taken place, would explain this pattern.

The mtDNA data presented here strongly indicates that the domestic dog has a single origin from southern East Asia, but further genetic studies are necessary to corroborate this. Independent markers, inherited also through the male lineages, should be investigated to see whether the phylogeographical patterns, e.g. the worldwide sharing of haplotypes and largest diversity in southern East Asia, are consistent across markers. They may also show if the extent of crossbreeding between female dog and male wolf has been as rare as that between male dog and female wolf (only three or four cases through time, as indicated by the region specific clades D, E and F, (see Supplementary Material for details)). More complex population genetic simulations than the relatively simple ones performed in this study may be valuable for understanding the mechanisms of genetic spread. Finally, analysis of mtDNA from archaeological samples may be of large value to investigate the earliest steps in the dog origins, if performed at a large enough scale to allow population genetic inferences. For example, the number of independent domestications may be definitely established when we know whether the earliest dog populations in different parts of Eurasia all shared the same mix of clades A, B and C as today's dogs, or if the different clades initially occurred regionally.

4.4 The genetic evidence in the light of other data

How does the mtDNA based scenario for the dog origins that we suggest fit with the other sources of information about dog origins, chiefly archaeological and osteological data?

First of all, since the dog originated from wolf, presence of wolf would be necessary in the presumed region of origin. The wolf is now extinct south of the Yangtze River, but it was present in practically every Chinese province until the 1950s (Gao 1997; Gao 2006). There is no record of wolf in Southeast Asia (Higham 1996), but there is no clear information about how far south in East Asia wolves may once have occurred. Thus, wolf has been present south of Yangtze River, and wolf domestication within ASY was therefore possible. It is notable that, since wolf is now exterminated south of the Yangtze river, it would not have been possible to identify the region of origin for the dogs based on a genetic comparison of extant dog and wolf populations. Therefore, intraspecific studies of dog, such as this, remains the only possibility for studying dog origins based on extant populations.

One osteological detail gives some support for an origin of dogs in China. Olsen and Olsen (1977) observed that a morphologic feature of the jaw, the "turned-back apex of the coronoid process of the ascending ramus", which is diagnostic for dogs is also found in Chinese wolves, but is absent in wolves from other regions. Based on this, and the small size of Chinese wolves, they drew the conclusion that the Chinese wolf is a likely ancestor of Chinese and American dogs.

Compared to the archaeological evidence, indicating the presence of dogs in SW Asia by 11,500 ya, the time of origin suggested by the mtDNA data is in good agreement if the dogs spread across Eurasia within a few thousand years, which seems possible. There is for example clear evidence for the spread of agriculture over large distances in Europe at a rate in excess of 5 km/year (Price 2000). At this rate, dogs would have spread from Yunnan to Israel (<10,000 km) in less than 2,000 years. The dog did not necessarily spread in connection with human migrations; a spread through trade to regions previously without dogs could possibly have been very fast, if dogs were a desirable resource. It can here be noted that there are few, if any, signs in the archaeological record of contact between East Asia and Europe or SW Asia at this time. However, it is noticeable that dogs seem to have been uniquely mobile; they are the only domestic animal accompanying humans to every continent in ancient times, and therefore seem to be a much more mobile "sign of contact" than most material artefacts. For example, the Australian dingo originated from East Asian dogs approximately 5,000 ya (Corbett 1995; Savolainen et al. 2004) but there is no other sign of contact between China and Australian aboriginals in ancient times.

While there is reasonable agreement about the time of origin between the mtDNA and archaeological data, there is some disagreement concerning the place. The earliest archaeological evidence giving a strong indication for the presence of domestic dog is from at least 11,500 ya in SW Asia (from the Natufian culture in today's Israel (Davis and Valla 1978; Dayan 1994; Tchernov and Valla 1997); earlier dates in Europe have been reported, but the evidence does not seem conclusive (Wang and Tedford 2008), see Supplementary Material)), but only by at least 7,100 ya in North China, 6,500 ya in South China (Underhill 1997), and 4,000 ya in Southeast Asia (Higham 1996). However, there are large difficulties with the interpretation of the archaeological record of early dogs (see Supplementary Material for a detailed discussion). Discrimination between wolf and dog is hard even for experts, and osteological traits used to distinguish domesticated dogs from wolves are not totally exclusive but occasionally found also among wolves (Musil 2000). More importantly, archaeological work, and systematic surveys of animal remains in particular, have been executed to very

varying extent in different parts of the world. For example, the Natufian dogs do not have the full repertoire of distinguishing morphological details, and the conclusion that they are domestic dogs relies on extensive analyses (Dayan 1994). In contrast, archaeological evidence for dogs in North China by 9,700-10,800 ya has been reported (Underhill 1997; Jin and Xu 1992), but no morphological details are described and the evidence can therefore not be evaluated. Therefore, it is possible that early evidence for dog in East Asia has so far been overlooked because of the lack of systematic search for the subtle morphological differences between early domestic dogs and wolves. It should be noted also, that this is a study of the mtDNA diversity of the extant dog population, and therefore concerns only the domestication of wolves that were actually ancestral to today's dog population. Therefore, if there were any taming or domestication, through history, of wolves or other canids that were not ancestors of today's dogs, this would go unrecorded in this study.

To conclude, an origin of the dog from Chinese wolf is supported by osteological data, and the time of origin agrees with archaeological data assuming that the spread across Eurasia occurred within a few thousand years, which seems plausible. The earliest strong archaeological evidence for domestic dog is from SW Asia, but this may be an effect of bias in the amount of archaeological research. Importantly, nothing contradicts that the SW Asian dogs were introduced from outside. The indications from the mtDNA data, that southern East Asia was the main centre of origin for the domestic dog, urges for intensified archaeological studies of canid remains in this region.

4.5 Indications of how, why and by whom the wolf was domesticated

This study gives for the first time some potential clues about the human culture and the mechanisms that may have been behind the origins of dogs, based on the indication of time and place, and that several hundred wolves were domesticated.

It is noticeable that the approximate time and place coincide with the transition from hunting and gathering to farming in connection with the emergence of rice agriculture. China had two centres of plant domestication and early agriculture, of millet by the Yellow River and of rice in the Yangtze River area, both at least 8,500 ya (Underhill 1997; Bellwood 2005). Rice domestication probably started to develop first, at least 9,000 ya but possibly as early as 11,500 ya. There is also evidence for pottery by 14,000-11,200 ya in southern China (Underhill 1997; Bellwood 2005). It is possible that the dog originated in this cultural context of increasingly sedentary hunter-gatherers or early rice farmers. The large number, probably

hundreds, of domesticated wolves indicates that the dog originated in an ordered and widely distributed culture involving large numbers of humans. It seems probable that some degree of sedentary life would have facilitated the wolf taming and domestication process. The earliest evidence for rice agriculture is found in Hunan and Jiangxi (Underhill 1997; Bellwood 2005), which lacked two (the two least frequent) of the ten major mtDNA subclades. The full representation, all ten subclades, was found 1,000 km to the southwest within Yunnan and Southeast Asia. Importantly, this study has a relatively small number of samples, around 50, from each province in ASY, and several south and central Chinese provinces are not represented. Analysis of more samples is therefore necessary to get the full picture of the mtDNA diversity of the region, to study more precisely the region for the dog origins and the possible connection with the origins of agriculture.

The large number of domesticated wolves, and subsequently kept female lineages, shows that the domestication of wolves was not a singular chance event involving a small number of wolves but probably a widespread and important custom of the human culture involved. On the other hand, the mtDNA data also indicates that domestication of wolf took place only once in history. The taming and domestication of wolf, to begin with probably not intentionally done, was presumably not straightforward and possibly difficult and dangerous. However, once it had started it was performed at a large scale and was possibly simple when the appropriate methods were employed.

Finally, it is worth noting that, in contrast to most other parts of the world, dogs have been used as food on a large scale in southern East Asia, from ancient times until today (Higham, Kijngam and Manly 1980; Simoons 1991; Ren 1995). It may therefore be speculated that the wolf was domesticated for its use as a source of food rather than for hunting, guarding or companionship as mostly suggested, perhaps under influence of a European non-dog eating perspective. In nature, wolves (in contrast to the omnivorous dogs) are practically strict carnivores (Thorne 1995), and feeding meat to a meat animal may seem an illogical expense. However, wolves are able to obtain all necessary nutrients from vegetable material (Thorne 1995) and Italian wolves, whose habitats have been severely encroached by human settlement, are estimated to obtain 60-70% of their food from garbage dumps, including a large proportion of vegetable substances, e.g. spaghetti (Boitani 1982). Possibly, the transition in behaviour from carnivore to omnivore was an early step in the domestication process, perhaps in an initial “self-domestication” process (Crockford 2000) in which wolves approached human camp sites in search for food left-overs.

These theories are so far only loosely founded; the approximate coincidence in time and space of the origins of dogs and of rice agriculture may be the result of mere chance. However, itshould be possible to test these hypotheses through detailed studies of dog in southern East Asia, by genetic studies of extant dogs as well as ancient canid samples, and by thorough archaeological studies of canid remains. Therefore, a precise picture of how the domestic dog originated now seems feasible.

Supplementary Material

Supplementary Material (containing Supplementary text, and tables S1-S6 and figures S1-S6), and Supplementary Datasets 1, 2 and 3, is available at *Molecular Biology and Evolution* online.

Acknowledgements

This work was supported by grants from the National Basic Research Program of China (973 Program, 2007CB815700), Chinese Academy of Sciences (KSCX2-YW-N-018), Bureau of Science and Technology of Yunnan Province, National Natural Science Foundation of China (30621092), the Swedish Research Council, OE and Edla Johanssons Scientific Foundation, the Carl Trygger Foundation, the Wenner-Gren Foundations, and the Swedish Kennel Club. Peter Savolainen is a Royal Swedish Academy of Sciences Research Fellow supported by a grant from the Knut and Alice Wallenberg Foundation.

References

Angleby H, Savolainen P. 2005. A study of the forensic usefulness of the mitochondrial DNA variation among and within populations, breeds and types of domestic dogs. Forensic Sci Int. 154**:**99-110.

Bellwood P. 2005. First Farmers: The origins of Agricultural Societies. UK: Blackwell Publishing.

Boitani L. 1982. Wolf management in intensively used areas of Italy. In: Parrington FH, Paquet PC, editors. Wolves of the world. p. 158-172.

Chaix L. 2000. A preboreal dog from the northern Alps (Savoie, France). In: Crockford SJ, editor. Dogs through time: an archaeological perspective. Oxford: British archaeological reports. p. 49-59.

Clutton-Brock J. 1995. Origins of the dog: domestication and early history. In: Serpell J, editor. The domestic dog, its evolution, behavior and interactions with people. Cambridge: Cambridge University Press. p. 7-20.

Corbett, L. 1995. The Dingo in Australia and Asia. Sydney: University of New South Wales Press.

Crockford SJ. 2000. Dog evolution: a role for thyroid hormone physiology in domestication changes. In: Crockford SJ, editor. Dogs through time: an archaeological perspective. Oxford: British archaeological reports. p. 11-20.

Davis SJM, Valla FR. 1978. Evidence for domestication of the dog 12,000 years ago in the Natufian of Israel. Nature. 276:608-610.

Dayan T. 1994. Early domesticated dogs of the Near East. J Arch Sci. 21:633-640.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. 7:214.

Excoffier L, Laval G, Schneider S. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evol Bioinformatics Online. 1:7-50.

Forster P, Harding R, Torroni A, Bandelt HJ. 1996. Origin and evolution of Native American mtDNA variation: a reappraisal. Am J Hum Genet. 59:935-945.

Fuller KF, Mech LD, Cochrane JF. 2003. Wolf population dynamics. In: Mech LD, Boitani L, editors. Wolves: behavior, ecology, and conservation. Illinois: University of Chicago Press. p. 161-191.

Gao, ZS. 1997. Geographic distribution and current population status of wolves worldwide. Chin Wildlife. 18:27-28. (In Chinese)

Gao Z-X. 2006. Review of the Research on Wolf in China. Chin J Zool. 41:134-136. (in Chinese)

Gray MM, Granka JM, Bustamante CD, Sutter NB, Boyko AR, Zhu L, Ostrander E, Wayne R. 2009. Linkage Disequilibrium and Demographic History of Wild and Domestic Canids. Genetics. forthcoming.

Higham CFW. 1996. A review of Archaeology in Mainland Southeast Asia. J Arch Res. 4:3-49.

Higham CFW, Kijngam A, Manly BFJ. 1980. An analysis of prehistoric canid remains from Thailand. J Arch Sci. 7:149-165.

Ho SY, Larson G. 2006. Molecular clocks: when times are a-changin'. Trends Genet. 22:79-83.

Ingman M, Kaessmann H, Pääbo S, Gyllensten U. 2000. Mitochondrial genome variation and the origin of modern humans. Nature. 408:708-713.

Jeffreys H. 1998. Theory of probability (3 edn.). USA: Oxford University Press.

Jin J, Xu H. 1992. Opinions about the early Neolithic site of Nanzhuangtou at Xushui. Kaogu. 11:1018-1022. (In Chinese)

Korber B, Muldoon M, Theiler J, Gao F, Gupta R Lapedes A, Hahn BH, Wolinsky S, Bhattacharya T. 2000. Timing the ancestor of the HIV-1 pandemic strains. Science. 288:1789-1796.

Leonard JA, Wayne RK, Wheeler J, Valadez R, Guillén S, Vilà C. 2002. Ancient DNA evidence for Old World origin of New World dogs. Science. 298:1613-1616.

Lindblad-Toh K, Wade CM, Mikkelsen TS, et al. (236 co-authors) 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature. 438:803-819.

Malmström H, Vilà C, Gilbert MT, Storå J, Willerslev E, Holmlund G, Götherström A. 2008. Barking up the wrong tree: modern northern European dogs fail to explain their origin. BMC Evol Biol. 8:71.

Morey DF. 2006. Burying key evidence: the social bond between dogs and people. J Archaeol Sci. 33:158-175.

Morey DF, Wiant MD. 1992. Early Holocene domestic dog burials from the North American Midwest. Curr Anthrop. 33:224-229.

Musil R. 2000. Evidence for the domestication of wolves in central European Magdalenian sites. In: Crockford SJ, editor. Dogs through time: an archaeological perspective. Oxford: British archaeological reports. p. 21-28.

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, editors. Wolves: Behavior, ecology, and conservation. Illinois: University of Chicago Press. p. 239-258.

Olsen GJ, Matsuda H, Hagstrom R, Overbeek R. 1994. fastDNAml: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. Comput Appl Biosci. 10:41-48.

Olsen SJ, Olsen JW. 1977. The Chinese wolf, ancestor of New World dogs. Nature. 197:533-535.

Ostrander EA, Wayne RK. 2005. The Canine genome. Genome Res. 15:1706-1716.

Price TD. 2000. Europe's first farmers. Cambridge: Cambridge University Press.

Raisor MJ. 2005. Determining the antiquity of dog origins. Oxford: British archaeological reports.

Ren S. 1995. Important results regarding Neolithic cultures in China earlier than 5000 B.C. Kaogu. 1:37-49. (In Chinese)

Savolainen P, Leitner T, Wilton AN, Matisoo-Smith E, Lundeberg J. 2004. A detailed picture of the origin of the Australian dingo, obtained from the study of mitochondrial DNA. Proc Natl Acad Sci USA. 101:12387-12390.

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.

Simoons FJ. 1991. Dog flesh. In: Food in China, a cultural and historical inquiry. Boston: CRC Press. p. 200-252.

Suchard MA, Weiss RE, Sinsheimer JS. 2001. Bayesian selection of continuous-time Markov chain evolutionary models. Mol Biol Evol. 18**:**1001-1013.

Tchernov E, Valla FF. 1997. Two new dogs, and other Natufian dogs, from the Southern Levant. J Arch Sci. 24:65-95.

Thorne C. 1995. Feeding behaviour of domestic dogs and the role of experience. In: Serpell J, editor. The domestic dog, its evolution, behavior and interactions with people. Cambridge: Cambridge University Press. p. 103-114.

Underhill AP. 1997. Current Issues in Chinese Neolithic Archaeology. J World Prehistory. 11:103-160.

Vilà C, Savolainen P, Maldonado JE, Amorim IR, Rice JE, Honeycutt RL, Crandall KA, Lundeberg J, Wayne RK. 1997. Multiple and ancient origins of the domestic dog. Science. 276:1687–1689.

Vilà C, Seddon JM, Ellegren H. 2005. Genes of domestic mammals augmented by backcrossing with wild ancestors. Trends Genet. 21:214-218.

Wang X, Tedford RH. 2008. Dogs, their fossil relatives and evolutionary history. New York: Columbia University Press.

Zeder MA, Emshwiller E, Smith BD, Bradley DG. 2006. Documenting domestication: the intersection of genetics and archaeology. Trends Genet. 22:139-155.

Table 1. Genetic diversity for CR data across the Old World

Region[a]	N[b] ABC(DEF)	nA(%)[c]	nB(%)[c]	nC(%)[c]	HT[d]	HTun[e]	HTres[f]	PropUT[g]	PropUTd
East Asia	730 (5)	549(75.2)	124(17.0)	57(7.8)	121	84	28.5 ± 3.0	54.2	69.2
West	558 (27)	384(68.8)	123(22.0)	51(9.1)	68	41	23.0 ± 2.5	76.2	95.5
Europe	313 (23)	217(69.3)	68(21.7)	28(8.9)	39	19	19.5 ± 2.0	81.5	98.7
SW Asia	130 (3)	72(55.4)	45(34.6)	13(10.0)	30	8	19.5 ± 2.1	81.5	94.6
Africa	56 (1)	48(85.7)	6(10.7)	2(3.6)	22	6	22.0 ± 0.0	57.1	91.1
India	59 (0)	47(79.7)	4(6.8)	8(13.6)	22	6	21.5 ± 0.6	54.2	81.4
Siberia	60 (2)	39(65.0)	13(21.7)	8(13.3)	20	7	19.7 ± 0.5	46.7	75.0
Japan	118 (3)	76(64.4)	24(20.3)	18(15.3)	25	6	19.7 ± 1.7	58.5	84.7
Korea	90 (7)	80(88.9)	6(6.7)	4(4.4)	24	5	19.1 ± 1.6	63.3	77.8
N China	98 (0)	65(66.3)	25(25.5)	8(8.2)	24	5	18.8 ± 1.6	79.6	89.8
C China	141 (0)	109(77.3)	21(14.9)	11(7.8)	27	8	17.4 ± 1.9	70.9	85.8
S China	281 (0)	223(79.4)	44(15.7)	14(5.0)	71	40	27.7 ± 2.8	42.0	53.4
SE Asia	57 (2)	50(87.7)	3(5.3)	4(7.0)	30	11	29.7 ± 0.5	35.1	50.9
ASY	338 (2)	273(80.8)	47(13.9)	18(5.3)	87	53	30.1 ± 3.0	40.8	53.0

[a]East Asia – China (N/C/S China, Tibet, Quinghai), Southeast (SE) Asia, Japan; West - Europe, Southwest (SW) Asia, Africa, India; N/C/S China – China north of Yellow River/between Yellow and Yangtze/south of Yangtze River; ASY – Asia South of Yangtze River (S China, SE Asia)

[b]N – number of individuals having clade A, B or C, and D, E or F (within parentheses)

[c]n – number of individuals having clade A, B and C (within parentheses percentage based on only clades A, B and C)

[d]HT – number of haplotypes

[e]HTun – number of unique haplotypes

[f]HTres – number of haplotypes (with s.e.m.) obtained from resampling of size 56 (500 replications) to adjust for different sample size

[g]PropUT/UTd – proportion of individuals carrying a UT and UTd.

Table 2. Geographical representation of the subclades of clades A, B and C[a]

Region[b]	a1	a2	a3	a4	a5	a6	b1	b2	c1	c2
East Asia	322	166	17	6	26	3	111	14	31	26
West	377	2	-	-	-	-	123	4	25	26
Europe	215	-	-	-	-	-	68	-	14	14
SW Asia	71	-	-	-	-	-	42	4	9	4
Africa	46	1	-	-	-	-	6	-	-	2
India	45	1	-	-	-	-	4	-	2	6
Siberia	36	3	-	-	-	-	13	-	8	-
Japan	61	13	-	2	-	-	20	4	4	14
Korea	62	11	-	1	-	-	5	1	2	2
N China	55	10	-	-	-	-	25	-	7	1
C China	89	18	-	1	1	-	21	-	10	1
S China	86	99	14	-	18	3	35	9	8	6
SE Asia	20	13	3	1	7	-	3	1	2	2
ASY	106	112	17	1	25	3	38	10	10	8

[a]Distribution of CR sequences (number of individuals) to the 10 subclades. Eight haplotypes, for totally 22 individuals, could not be assigned to a specific subclade, see Supplementary Material and table S2.

[b]Definition of geographical regions, see Note to table 1.

Fig. 1. Genetic relationships between the mtDNA CR sequences (582 bp).

a) Maximum likelihood evaluated neighbor-joining tree for the dog (unlabeled) and wolf (filled square) haplotypes, rooted by coyote (Coy) sequences (branch length reduced by 55%). The 6 main phylogenetic clades (A-F) of dog haplotypes are indicated. The geographical origin of wolf samples are, for the wolf haplotypes (filled squares) from top to bottom: Mongolia (1 individual); China (1); Mongolia (1); Afghanistan (5); Romania (1); China (1); Yugoslavia (4); Canada (1); China (1); China (1); Spain (1); Spain (1); Sweden (1); Saudi Arabia (1); China (4); China (1); China (1) and Mongolia (1); China (2); China (1); Mongolia (2); Yugoslavia (3); Russia (1); Saudi Arabia (1); Mongolia (2).

b) Minimum spanning networks showing the relationships between the haplotypes in dog clades A, B and C, and the representation of the haplotypes in different geographical regions. The haplotypes are symbolized by circles (coloured or white), and are separated by one substitutional step (ignoring indels); black dots are hypothetical intermediates. The 14 "universally occurring" haplotypes (UTs) are indicated with bold lining. The representation of haplotypes in the geographical regions is shown by the colour and size of the circles. A coloured circle indicates the presence of the haplotype (blue for haplotypes shared with other regions and orange for haplotypes unique to the region); white denotes non-represented haplotypes. The size of the circle is proportional to the frequency of the haplotype in the region. To simplify the figure, the circle size is reduced for the most frequent haplotypes for East Asia (eight most frequent haplotypes reduced by 30%) and West (most frequent haplotype reduced by 50%, next seven most frequent haplotypes by 30%). For definition of geographical regions, see Note to table 1 (The region denoted "Sib, Mon, Kor, Jap" represents Siberia, Mongolia, Korea and Japan).

c) The CR haplotypes chosen for analysis of the mtDNA genome, indicated by colour. The different colours refer to the subclades (six for clade A, two each for B and C) subsequently identified (see fig. 2a). Clade A: subclade a1 orange, a2 green, a3 red (upper part of network), a4 red (lower part of network), a5 blue, a6 yellow. Clade B: subclade b1 orange, b2 green. Clade C: subclade c1 green, c2 orange.

Fig. 2. Genetic relationships between the mtDNA genome sequences (16,195 bp).

a) Maximum likelihood tree for dog and wolf sequences, rooted by coyote sequences (branch length reduced by 87%). The sample identity for the dogs is given as the name of the CR haplotype (purple indicates that the CR haplotype is a UT ["universally occurring haplotype"]), and an individual number for samples with identical CR haplotype. The geographical origin of the dog samples is given in fig. 2b. "W" indicates wolf (geographical origin from top to bottom: China, Saudi Arabia, Saudi Arabia, Sweden, Canada, Spain, China, Russia) and "Coy" coyote. Dog clades A, B and C (marked by *) were supported by bootstrap values ≥92% and Bayesian values =100%, and their subclades a1-a6, b1-b2 and c1-c2, by bootstrap values ≥87% and Bayesian values ≥99%.

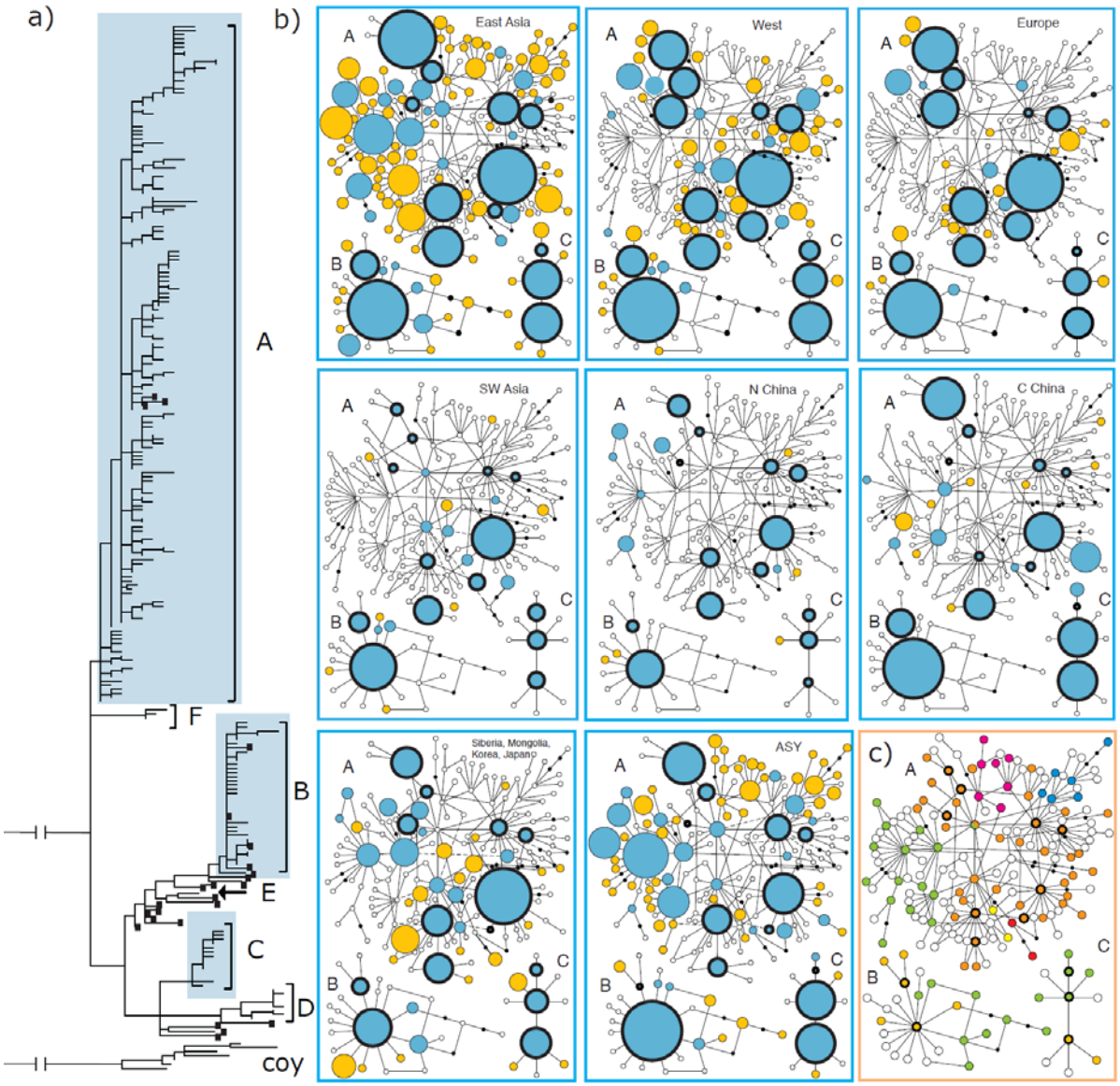
b) Minimum spanning networks showing the relationships between the haplotypes in dog clades A, B and C. The haplotypes are represented by ellipses (UTs indicated with bold lining) and are separated by the number of substitutions indicated next to the connecting lines, no number indicating a single substitution. The sample identity is the same as in fig. 2a (identical samples are given within the same ellipse). Colours indicate geographical origin: Red, Eastern Asia (China, Southeast Asia, Japan, Korea, Mongolia, Siberia); Blue, Europe; Green, Southwest Asia; Purple, Africa; Yellow, India.

Fig. 3. Genetic diversity for CR data, among regions across the Old World.

a) Genetic diversity across the Old World. Pie diagrams show the proportion of individuals having clades A (blue), B (red) and C (yellow). Boxes show (i) UT: the proportion of individuals having one of the 14 "universally occurring" haplotypes, (ii) UTd: the proportion having a UT-derived haplotype, i.e. a haplotype which is either a UT or differs by a single substitution from a UT, and (iii) SC: the representation of the 6 non-universal (of the totally 10) subclades of clades A, B and C. (Eur N/S cont - Europe North/South continent; N/C China - North/Central China [north of Yellow River/between Yellow and Yangtze River]; ASY- Asia South of Yangtze River).

b) Genetic diversity in south eastern Asia. Boxes show UT, UTd and SC as defined in fig. 3a. Sampled regions are indicated by colour. LHS - Liaoning, Hebei and Shanxi; GG - Guanxi and Guangdong. SE Asia – Vietnam, Thailand and Cambodia. Tibet includes samples from Qinghai and Nepal.

Savolainen_fig1



Savolainen_fig2