Complex population structure in African village dogs and its implications for inferring dog domestication history

Adam R. Boyko[a,1], Ryan H. Boyko[b], Corin M. Boyko[b], Heidi G. Parker[c], Marta Castelhano[d], Liz Corey[d], Jeremiah D. Degenhardt[a], Adam Auton[a], Marius Hedimbi[e], Robert Kityo[f], Elaine A. Ostrander[c], Jeffrey Schoenebeck[c], Rory J. Todhunter[d], Paul Jones[g], and Carlos D. Bustamante[a]

[a]Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY 14853; [b]Department of Anthropology and Graduate Group in Ecology, University of California, Davis, CA 95616; [c]National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892; [d]Department of Clinical Sciences and the Medical Genetic Archive, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853; [e]Department of Biological Sciences, University of Namibia, Windhoek, Namibia; [f]Department of Zoology, Makerere University, Kampala, Uganda; and [g]The Walthan Centre for Pet Nutrition, Waltham on the Wolds, Leicestershire LE14 4RT, United Kingdom

High genetic diversity of East Asian village dogs has recently been used to argue for an East Asian origin of the domestic dog. However, global village dog genetic diversity and the extent to which semiferal village dogs represent distinct, indigenous populations instead of admixtures of various dog breeds has not been quantified. Understanding these issues is critical to properly reconstructing the timing, number, and locations of dog domestication. To address these questions, we sampled 318 village dogs from 7 regions in Egypt, Uganda, and Namibia, measuring genetic diversity >680 bp of the mitochondrial D-loop, 300 SNPs, and 89 microsatellite markers. We also analyzed breed dogs, including putatively African breeds (Afghan hounds, Basenjis, Pharaoh hounds, Rhodesian ridgebacks, and Salukis), Puerto Rican street dogs, and mixed breed dogs from the United States. Village dogs from most African regions appear genetically distinct from non-native breed and mixed-breed dogs, although some individuals cluster genetically with Puerto Rican dogs or United States breed mixes instead of with neighboring village dogs. Thus, African village dogs are a mosaic of indigenous dogs descended from early migrants to Africa, and non-native, breed-admixed individuals. Among putatively African breeds, Pharaoh hounds, and Rhodesian ridgebacks clustered with non-native rather than indigenous African dogs, suggesting they have predominantly non-African origins. Surprisingly, we find similar mtDNA haplotype diversity in African and East Asian village dogs, potentially calling into question the hypothesis of an East Asian origin for dog domestication.

Canis familiaris | microsatellites | principal component analysis | single nucleotide polymorphisms

In many respects, dogs have a unique relationship to humans. They were the first domesticated species, serve as valuable companions and service animals, and have been bred to exhibit more phenotypic diversity than any other mammal (1–3). Dogs were probably domesticated from Eurasian wolves at least 15,000–40,000 years ago (4–6), although the process by which domestication took place, including the specific selected traits and the manner in which selection was performed, is very poorly understood (7, 8).

After domestication somewhere in Eurasia, dogs quickly spread throughout the continent and into Africa, Oceania and the Americas (9). These early dogs, like modern day "village dogs" (7), almost certainly lived as human commensals that were not subject to the same degree of intense artificial selection and closed breeding practices that characterize modern dog breeds. Like ancient human populations, these ancient dog populations developed genetic signatures characteristic of their geographic locale. These signatures would persist in both modern day village dog populations that descend from these ancient populations and in dog breeds that were founded from them. We refer to such dogs as "indigenous" in the sense that they carry characteristic genetic signatures appropriate for their geographic region.

Today, semiferal village dogs are nearly ubiquitous around human settlements in much of the world, and such animals comprise a large proportion of the global dog population (7). However, the popularity of modern breeds has led to the widespread transport of mostly European-derived breed dogs into many areas containing village dogs, so it is likely that many modern village dogs are not derived solely from indigenous ancestors. We refer to village dogs that descend from these foreign dogs as "non-native" and expect that genetic markers can differentiate these village dogs from indigenous dogs. We believe most of these dogs will be complex mixtures of several non-native breeds and/or mixtures of both non-native breeds and indigenous village dogs ("intermediate" ancestry).

The distinction between indigenous and non-native dogs is important because indigenous, but not non-native, village dogs are likely to contain genetic variants that are not found in any of today's >400 recognized dog breeds. Furthermore, they are expected to be more informative regarding dog population history and are likely to be more adapted to local environmental conditions and more genetically related to the first prebreed domestic dogs than breed or breed-admixed individuals. To our knowledge, the degree to which village dogs consist of indigenous versus non-native individuals has not been quantified.

In one of the most comprehensive surveys of village and breed dogs to date, Savolainen et al. (6) examined mtDNA diversity in a global panel of 654 dogs. Their results confirmed previous mtDNA evidence of dog domestication from Eurasian wolves (5), showed that East Asian dogs had the highest mtDNA diversity of any region, suggesting an East Asian origin of domestication. However, subsequent work by Pires et al. (10) has shown that mtDNA does not show significant population structure in village dogs. Because Savolainen et al. included many East Asian village dogs but few village dogs from other regions, their conclusion of high levels of
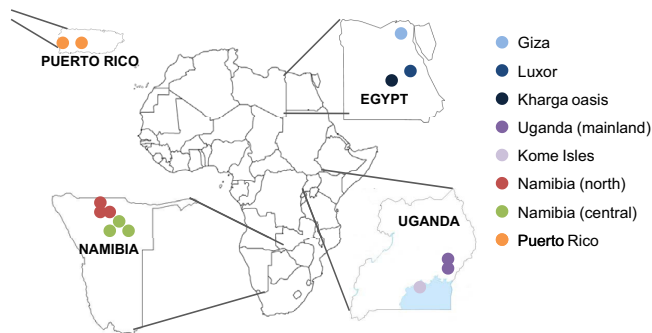
GENETICS

Fig. 1. Map of village dog sampling locations. Colors denote each distinct region and dots show approximate range of sampling within each region. See Table S1 for full description.

Fig. 2. STRUCTURE analysis across 389 SNP and microsatellite loci in African village and American mixed breed dogs.

East Asian diversity is likely a consequence of high levels of mitochondrial diversity in village dogs and not necessarily an indication of East Asian domestication.

Other genetic markers have been shown to exhibit significant population structure in village dogs. Microsatellites and MHC types both separate Bali street dogs from New Guinea singing dogs, dingoes, and breed dogs (11, 12). Both studies demonstrated high diversity in the Bali dogs, consistent either with an indigenous, prebreed ancestry or with a complex admixture history from a large number of breeds. Therefore, given a large enough sample of village and breed dogs, microsatellite and single nucleotide polymorphism (SNP) markers seem well suited to studying population structure and the possibility breed admixture in village dogs.

In this study, we analyzed mtDNA, microsatellite, and SNP markers in 318 African village dogs to characterize population structure and genetic diversity. In addition, we analyzed 16 Puerto Rico street dogs, 102 known mixed-breed dogs from the United States, and several hundred dogs from 126 breeds, including 129 dogs from five African and Middle Eastern breeds, to determine the degree of non-native admixture in African village dogs. Our sampling effort concentrated on seven regions from three geographically separated African countries (Fig. 1): Egypt: We sampled three distinct locales: a Giza animal shelter, a Luxor animal shelter and surrounds, and a rural desert oasis (Kharga). Although the geographic distance between Giza and Luxor is greater than that between Kharga and Luxor, we hypothesized that the desert would be a strong barrier to gene flow, making the latter populations more genetically distinct.

Uganda: We sampled >100 dogs from a cluster of villages east of Kampala and 30 dogs from three neighboring isles of the Kome Island group in Lake Victoria. Despite the islands being close to each other and the mainland (<20 km), we expected the lake might act as a dispersal barrier.

Namibia: We sampled from over a dozen villages and urban areas in the northern and central parts of the country. No natural dispersal barriers existed between sampling locations, although a cordon fence is maintained to keep livestock diseases out of the southern part of the country. Dogs are permitted to be taken across the cordon and likely have little difficulty getting through the fence themselves, but the cordon is significant in that it demarcates the extent of European colonization influence in the country [with southern and central Namibia colonization history being roughly similar to that of South Africa while northern Namibia resembles the rest of sub-Saharan Africa (13)]. We sampled dogs within 100 km of both sides of the cordon, including populations within 10–20 km of the barrier.

For comparison, we also sampled from two shelters in Puerto Rico, known mixed-breed dogs (see *Methods*) from the United States, and dogs from 126 breeds, including five African and
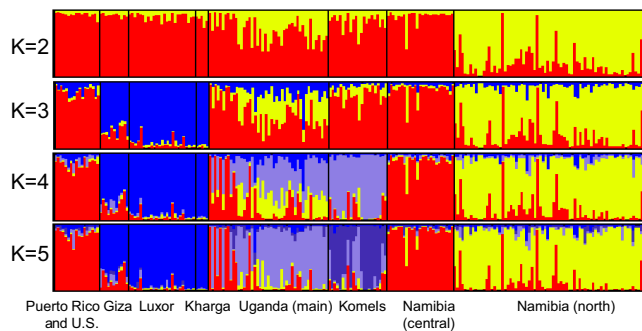
near-African breeds (putative origin in parentheses): Afghan hounds (Sinai, Egypt), Basenjis (Congo), Pharaoh hounds (near Mediterranean), Rhodesian ridgebacks (Zimbabwe), and Salukis (Iraq).

Results

Inference of Population Structure and Degree of Breed Admixture in African Village Dogs. A subset of 223 unrelated African village dogs from seven African locales were typed on a panel of 89 microsatellite markers or 300 SNP markers (206 village dogs, 15 Puerto Rican dogs, and two United States mixed-breed dogs were typed on both panels). Using the Bayesian clustering program STRUCTURE (14), we found that Puerto Rican street dogs clustered with the mixed-breed dogs from the United States, indicating these dogs are all breed admixtures. STRUCTURE analysis at K = 5 consistently showed the same five groupings: Egyptian dogs, Ugandan mainland dogs, Kome Island dogs, Northern Namibian dogs, and admixed (including all Puerto Rican and U.S. dogs, nearly all Central Namibian dogs, and a few other African village dogs; Fig. 2 and Fig. S1). At K = 4, STRUCTURE clustered Ugandan dogs together (mainland and Kome Islands), and at K >5, STRUCTURE subdivided Ugandan dogs further, although these clusters were inconsistent (Fig. S2).

We quantified admixture in each village dog as the mean proportion of the genome assigned to the American (United States + Puerto Rico) cluster by STRUCTURE across 10 runs at K = 5 (admixture estimates using K = 4 or K = 6 mean proportions were nearly identical; R^2 = 0.984 and 0.992, respectively). In total, 84% of African village dogs outside of central Namibia showed little or no evidence of non-native admixture (estimated admixture proportion <25% in 152 of 181 dogs), whereas all central Namibian dogs had >25% admixture, and most had >60% (24 of 25; Table 1). Principal component analysis showed a clear separation of Egyptian from sub-Saharan populations in PC1 and separation between

Table 1. Number of indigenous (<25% inferred admixture), uncertain (25%–60% inferred admixture) and breed admixed (>60% inferred admixture) village dogs by region from the 223 unrelated genotyped dogs

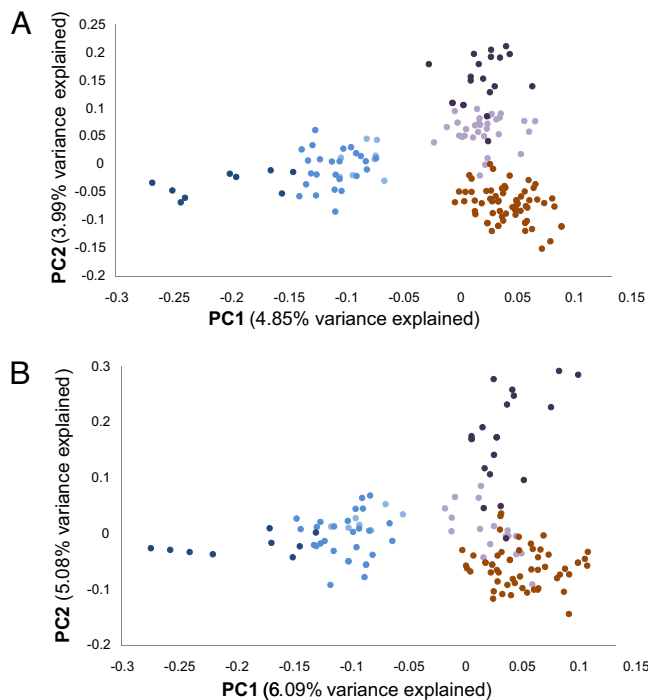| country	region	indigenous	uncertain	admixed
Egypt	Giza	7	4	0
Egypt	Luxor	25	0	0
Egypt	Kharga	5	0	0
Uganda	mainland	34	4	7
Uganda	isles	19	3	0
Namibia	central	0	1	24
Namibia	north	62	7	4
Puerto Rico		0	0	15

Boyko et al.

Fig. 3. Principal component analysis of indigenous African village dogs. (*A*) PCA with the 89 microsatellite loci (*n* = 152). (*B*) PCA with the 300 SNP loci (*n* = 126).

Ugandan and Namibian populations in PC2 for indigenous African village dogs for both SNP and microsatellite markers (Fig. 3). When admixed African and American dogs were included, PCA, like STRUCTURE, always clustered them together, and the interpretation of the principal components became more complicated (Fig. S2).

To clarify the relationship between the Puerto Rican and African dogs that clustered with the two known mixed-breed dogs genotyped on the full 389 marker panel, we ran STRUCTURE on the 300-SNP dataset with an additional 100 known breed-admixed dogs from the United States that were genotyped on this SNP panel (Fig. S3). The groupings of African dogs and the inference of non-native admixed individuals are highly consistent with the earlier analyses until K = 5, when STRUCTURE starts to detect groupings within the admixed individuals. The substructure found within admixed individuals may be a consequence of different ancestral breeds in different individuals; STRUCTURE analysis of the village dogs and dogs from 126 breeds shows that the putatively indigenous village dogs cluster with ancient breeds (specifically Basenjis) while the putatively non-native dogs cluster with modern breed groups in various proportions (Fig. S4).

F_{ST} calculations confirm that central Namibian dogs show virtually no genetic differentiation from American dogs (pairwise F_{ST} based on SNP markers = 0.011; microsatellite F_{ST} = 0.0025). The

pairwise F_{ST} between Egyptian dogs from Giza and Luxor was also low (SNP F_{ST} = 0.0024; microsatellite F_{ST} = 0.0057), whereas other village dog populations had pairwise F_{ST} values of 0.025–0.133 (Table 2). Dogs from Kharga were the most distinct (F_{ST} of 0.0735–13.3) whereas dogs from mainland Uganda and northern Namibia ($\approx$2,900 km apart) show only moderate differentiation (F_{ST} = 0.0237–0.0254). Heterozygosity was high across all genetic marker types in all village dog populations except those of the Kharga oasis and the Kome islands and low in all of the breed dogs (Table 3).

Origin of Putatively African Breeds. We included individuals from five breeds with presumed African or Middle Eastern ancestry in our principal component analyses to see whether this approach could detect which sampled village dog populations are closest to the founding population for each breed. For the SNP loci, PC1 and PC2 differentiated three breed groups—Basenjis, Salukis/Afghan hounds, and Rhodesian ridgebacks/Pharaoh hounds—while village dogs were clustered closer to the origin (Fig. 4). Notably, the village dog cluster still exhibited geographical structuring with Egyptian village dogs lying closest to the Saluki/Afghan hound cluster, indigenous Namibian and Ugandan dogs lying closest to the Basenji cluster, and breed-admixed Namibian and American dogs lying closest to the Rhodesian ridgeback/Pharaoh hound cluster. PCA of the microsatellite loci revealed the same clustering affinities (Egyptian village dogs nearest to Salukis/Afghan hounds, etc.) as the SNP PCA although the breed clusters were less well defined (Fig. S5).

Analysis of Mitochondrial Diversity. We sequenced 680 bp of the mitochondrial D-loop, including the 582-bp region described in ref. 6. We found 47 haplotypes in the African dogs as well as 9 haplotypes in the Puerto Rican dogs, two of which were also found in the sampled United States mixed breed dogs (see Table S1 and Table S3). All haplotypes were in the A (33 African haplotypes), B (6 African haplotypes), or C (8 African haplotypes) clades (Fig. S6), the clades that are believed to contain >95% of domestic dogs (6). Over the region sequenced in (6) and ignoring indels, we found 18 African haplotypes that were not described by (6); 14 in A clade [one of which was found in Africa by (10)], one in B clade, and three in C clade. The Puerto Rican and United States mixed-breed dogs had 8 A clade and one B clade haplotypes (only one haplotype, a Puerto Rican A clade haplotype, ws not previously described in ref. 6).

Surprisingly, local mtDNA diversity did not differ systematically between African regions and similarly sized regions in East Asia, the purported origin of domestic dogs. Across the 582-bp region analyzed in refs. 6 and 10, and this study, the number of haplotypes observed in a region closely matches the neutral expectation (Fig. 5). Differences in regional haplotype diversity appear to be driven by sampling artifacts rather than by distance from an hypothetical domestication origin, with the highly sampled and fractionated subpopulations of Japan exhibiting the most diversity, and nearby Sichuan (China) probably exhibiting the least (Fig. 5). Neither Africa nor East Asia appears to contain private haplogroups

Table 2. Pairwise F_{ST} in village dogs between regions based on 300 SNPs

	Giza	Kharga	Luxor	NA_cent	NA_north	UG_isles	UG_main	America
Giza	—							
Kharga	8.98%	—						
Luxor	0.62%	8.31%	—					
NA_cent	3.48%	12.52%	5.92%	—				
NA_north	3.87%	10.91%	4.75%	4.75%	—			
UG_isles	4.90%	13.27%	6.28%	4.70%	4.93%	—		
UG_main	3.38%	11.78%	4.95%	4.44%	2.54%	3.75%	—	
America	2.70%	12.86%	5.15%	1.14%	4.97%	5.00%	4.51%	—

Boyko et al.

GENETICS

Table 3. Gene diversity (expected heterozygosity) at 89 microsatellite markers, 300 SNP markers, and the mitochondrial D-loop in African village dogs and five breeds

	microsatellites		SNPs		mtDNA
	all dogs	indigenous	all dogs	indigenous	all dogs
Egypt (Giza)	0.677 (11)	0.684 (8)	0.438 (11)	0.438 (8)	0.890 (11)
Egypt (Luxor)	0.666 (25)	0.664 (24)	0.419 (25)	0.417 (24)	0.936 (26)
Egypt (Kharga)	0.553 (5)	0.553 (5)	0.360 (5)	0.360 (5)	0.427 (5)
Uganda (mainland)	0.669 (43)	0.660 (30)	0.432 (19)	0.424 (16)	0.901 (118)
Uganda (isles)	0.633 (20)	0.630 (16)	0.435 (20)	0.429 (16)	0.858 (30)
Namibia (north)	0.638 (71)	0.631 (60)	0.429 (61)	0.415 (50)	0.929 (91)
Namibia (central)	0.637 (25)	0	0.466 (18)	0	0.916 (28)
America	0.648 (17)	0	0.459 (18)	0	0.909 (17)
Afghan Hound	0.333 (5)		0.317 (18)		
Basenji	0.356 (5)		0.184 (19)		
Pharaoh Hound	0.217 (4)		0.283 (16)		
Rhodesian Ridgeback	0.353 (5)		0.368 (28)		
Saluki	0.330 (5)		0.355 (24)		

Sample sizes are given in parentheses.

(haplotypes that are highly differentiated from those found on other continents; Fig. S6).

Discussion

This study analyzed a large number of genetic markers to characterize the level of non-native admixture in a geographically widespread set of semiferal village dog populations. African village dogs exhibit complex population structure because of the effects of geography, gene flow barriers, and the presence of non-indigenous dogs in some populations. Notably, the vast majority of the African village dogs could be classified as indigenous (<25% non-African ancestry) or non-native (>60% non-African ancestry), with only 7% showing intermediate levels of African ancestry (Table 1). Classification of individuals as indigenous versus non-native was consistent between runs, and remained consistent even when the number of mixed-breed dogs included in the analysis was substantially increased (Fig. S3).

With two exceptions, African village dogs did not exhibit a region-specific level of non-African admixture, but rather contained dogs with completely indigenous ancestry (or nearly so) that were often intermingling with a few highly admixed individuals. The lack of consistent levels of admixture within regions suggests that non-indigenous dog genes are quickly removed from village dog populations, or that admixture with non-indigenous dogs is a very

recent phenomenon in these areas. The two exceptions were central Namibia, where every dog had significant levels of non-indigenous admixture (see below), and Giza, where all dogs showed some, usually low, level of admixture. This background level of admixture in Giza could reflect older mixing with breed dogs around this ancient city, or it could simply reflect the relative proximity of Giza to Eurasia, the ancestral home of most modern breed dogs. STRUCTURE analyses including dogs from 126 breeds suggest it is the latter—Egyptian dogs cluster partially with ancient (mostly Asian) breeds and the sub-Saharan (Basenji + village dog) cluster and do not appear to cluster significantly with any of the (mostly European) modern breed groups (Fig. S4).

Dispersal barriers significantly affected population structure. The 230 km of desert separating the Kharga oasis from Luxor led to much stronger population differentiation (F_{ST} = 0.084) than the 500 km Nile corridor between Luxor and Giza (F_{ST} = 0.0024). Likewise, the Kome islands which lie 10–20 km from the mainland in Lake Victoria were much more differentiated from mainland Uganda than were northern Namibian populations 2,900 km away (F_{ST} = 0.051 vs. F_{ST} = 0.033). Most surprising, the 20–100 km distance between northern and central Namibian populations that coincided with that country's Red Line veterinary cordon fence
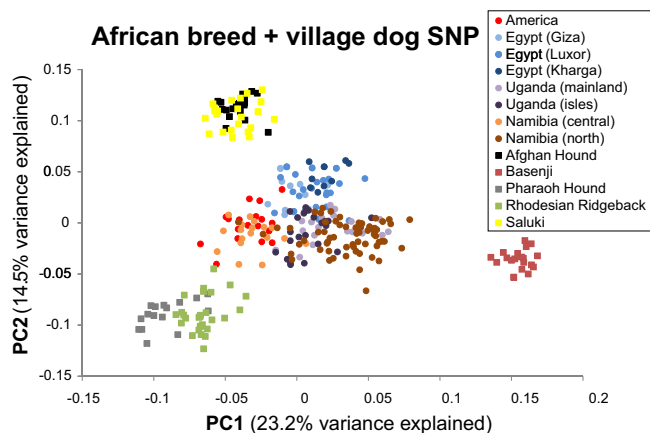
Fig. 4. Principal component analysis of village dogs and dogs from 5 putatively African and Middle Eastern breeds across 300 SNP markers in 186 village dog and 105 breed dogs.
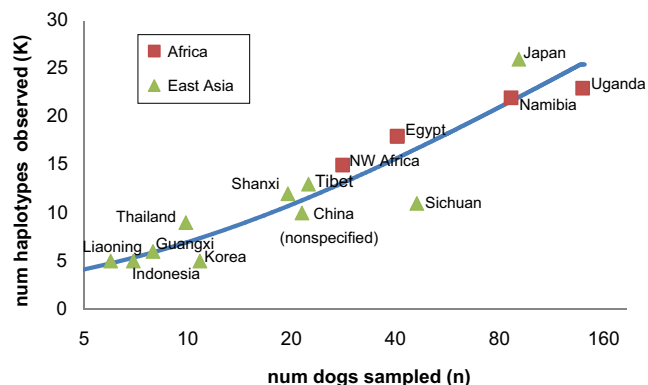
Fig. 5. Number of haplotypes (excluding indels) versus number of dogs sampled within Africa and East Asian geographic regions. Note log scale of x axis. East Asian samples from (6); African samples from this study or by (10). See Table S4 for a list of the areas used to construct this figure. The blue line depicts the expected number of haplotypes from Ewens's sampling formula (29), which assumes an infinite alleles model; $E(K) = \sum_{j=0}^{n-1} \theta/(j+\theta)$. Using Levenberg-Marquardt nonlinear regression, we estimate θ to be 8.654 (95% C.I. = [7.41, 9.89]).

Boyko et al.

represented a stark population boundary—dogs north of the cordon averaged 87% indigenous African ancestry while those south of the cordon were only 9% African. The cordon has separated the indigenous human populations (to the north) from white settlement areas (to the south) for the last 100 years and is currently used to restrict livestock (but not humans or dogs) from crossing southward (13). During this time, indigenous dogs have apparently been extirpated from central Namibia, and the selective pressures on dogs in each region must be strong and disparate enough to maintain a sharp genetic boundary along this porous chain-link fence. That Puerto Rico also seems to contain few, if any, indigenous dogs highlights the degree to which colonization history affects dog populations.

STRUCTURE and principal component analysis revealed strikingly similar patterns of genetic variation—indigenous African dogs clearly clustered by country and away from non-indigenous dogs in each analysis (Figs. 2–4). PCA showed slight differences between the SNP and microsatellite results: SNP but not microsatellite markers led to PC1 separating out dogs based on admixture (Fig. S2), although PCA with only indigenous African dogs resulted in the same axes of variation in both sets (Fig. 3). Breeds were clustered more cleanly with the SNP dataset than the microsatellite dataset, although this result could be an effect of the larger number of breed dogs that were typed on the SNP panel rather than a consequence of using SNPs versus microsatellites per se (Fig. 4 and Fig. S5). Nevertheless, both marker sets clustered Salukis and Afghan hounds nearest to Egyptian village dogs and Basenjis nearest to indigenous Ugandan and Namibian dogs, as expected by each breed's history. In contrast, Rhodesian ridgebacks and Pharaoh hounds clustered nearest to admixed dogs, suggesting these breeds have been recreated from admixture with non-African dogs. These results are consistent with the STRUCTURE results from (15, 16), showing that Salukis, Afghan hounds, and Basenjis cluster with ancient, non-European breeds, while Pharaoh hounds and Rhodesian ridgebacks do not. Although this coarse sampling (3 countries) is suitable for detecting truly indigenous versus reconstituted ancestry in putatively African breeds, analysis including village dogs from more regions will be necessary to better localize the ancestral origins of these breeds.

Village dog populations had higher levels of diversity than purebred dogs across all markers (see (17) for purebred mtDNA diversity estimates), although for SNP markers, non-native/admixed dogs had even higher diversity estimates. The high heterozygosity found in breed-admixed dogs is likely because of SNP ascertainment; by preferentially genotyping SNPs that are highly polymorphic in breed dogs, inferences based on SNP diversity in village dogs may be biased. Microsatellite ascertainment bias is less likely to have this effect since even microsatellites that are highly polymorphic in breeds can exhibit new alleles when genotyped in other populations. This suggests that careful control of ascertainment, or a denser SNP marker set that enables haplotype-based inference, is desirable for SNP markers. However, the high degree of concordance of SNP and microsatellite markers in both PCA and STRUCTURE analyses shows that these methods are robust to these effects.

African village dogs exhibited a similar level of mitochondrial D-loop diversity to that of the dogs sampled by (6) in East Asia, the putative site of dog domestication. Although we do not suggest that Africa is actually the site of dog domestication, we do believe that an East Asian origin of dogs should be further scrutinized, especially as Africa also has numerous private haplotypes and East Asia has no private haplogroups, with the possible exception of clade E, which is poorly represented numerically (1 haplotype, 3 individuals) and is rather similar to clade C. The data appear consistent with a rapid spread of dogs after original domestication and high effective population sizes and gene flow between continents, as there is no clear signal of decreasing haplotype diversity away from any origin.

Interestingly, Ugandan and northern Namibian populations that appear relatively undifferentiated using nuclear markers also have large overlap in their mitochondrial sequences. Thus, long-distance gene flow may be occurring, leading to a lower total number of haplotypes in these areas, whereas areas in Egypt with less chance for gene flow between them may harbor more diversity in the aggregate. This underscores the need to design a sampling and interpretation scheme to compare populations as opposed to coarse geographic areas. These areas could have features such as islands and deserts that may increase the number of haplotypes found only because one is sampling multiple populations.

Besides the discovery of 18 haplotypes, we have also expanded the geographic range of some previously reported dog mtDNA haplotypes. For example, we found haplotype A29, the predominant mtDNA haplotype of Australian dingoes, in a Puerto Rican dog even though this haplotype has never been reported in a dog outside of East Asia or the American Arctic (18). Either Puerto Rican dogs descend from some non-European (probably Asian) dogs that still carry this haplotype, or this is an indigenous New World haplotype that has persisted in Puerto Rico despite widespread historical European admixture.

Our results clearly demonstrate the need for further research with indigenous village dogs. Indigenous dog populations can be largely eliminated, as in Puerto Rico and central Namibia, by European colonization, and it is unclear the degree to which other populations will be able to maintain their genetic identity and persist in the face of modernity. The dog, although certainly a species uniquely suited as a model organism for genomics, can also serve as an invaluable organism for comparative studies of evolution and adaptation. Like other domesticated animals (e.g., cats, horses, and pigeons), dogs consist of breeds intensely selected for specific traits and feral populations that have been left to adapt to local conditions with "random" breeding. Dense genotyping and resequencing in these species should reveal genes underlying domestication in random-bred populations, instead of just those that have been under strong artificial selection in breed animals, and whether the relaxation of selective constraint observed in these species (19) is a product of recent breeding practices or domestication per se. Resequencing in indigenous village dogs will also be necessary to obtain markers free of ascertainment bias to estimate the amount of genetic variation in dogs that is absent in existing modern breeds, and the degree to which present-day indigenous village dogs represent populations that have been randomly breeding since dog domestication versus remnants of ancient, indigenous breeds.

Mitochondrial sequencing alone does not seem well-suited to determining the timing and location of domestication. Dog mitochondrial haplogroups seem more or less cosmopolitan, and inferences based on mtDNA diversity statistics can be easily skewed by sampling effort and misled by the inability to distinguish indigenous from non-native dogs. In the absence of finding multiple highly diverged and highly localized mitochondrial haplogroups, genome-wide autosomal markers will be needed to unravel the story of the first domesticated species.

Materials and Methods

Sampling Protocol. Dogs were sampled from animal shelters or were brought to the researchers for sampling by owners and villagers. In accordance with Cornell IACUC protocol 2007–0076, 3–5 mL of blood drawn from the cephalic or lateral saphenous vein into K2-EDTA blood collection tubes. At the field site, the blood cells were lysed with an ammonium chloride solution and spun at 1,100 × g with a portable centrifuge. After discarding the supernatant, cell pellets were resuspended in an EDTA-Tris-SDS solution for transport to the DNA Bank at Cornell Baker Institute for Animal Health. DNA was isolated from the lysate using ammonium acetate and alcohol and was suspended in Tris-EDTA buffer. Concentrations were determined by A260 on a NanoDrop ND1000 spectrophotometer. Stock DNA was stored in −20 °C freezers by the Cornell Medical Genetics Archive. Dilutions were made from a 200 μg/mL working stock as needed for sequencing and genotyping. A similar protocol was followed for the 102 United

States dogs, except that we also verified that they were mixtures of several different breeds by using the Wisdom MX breed test (Mars Inc.).

Microsatellite Genotyping. Two hundred twenty-seven village dogs were typed on a 96-microsatellite panel described in (15, 16). Microsatellites were amplified individually in the presence of a fluorescently labeled universal primer and were combined post-PCR into sets of 1 to 4 markers for capillary electrophoresis on an ABI3730xl (ABI). Standard PCR conditions have been described in ref. 15 while adjustments made to individual markers are listed in Table S5. Each 96-well plate of samples included a previously genotyped control sample for size verification and binned using GeneMapper 4.0. All genotype calls were checked manually and markers were scanned individually for the appearance of new alleles outside the existing bins. After genotyping, 7 markers were excluded on the basis of high missing rates (>20%) or heterozygote deficits ($P < 0.01$) in a majority of the 8 regional populations because this suggests the presence of null alleles at these loci. These data were combined with dogs from 126 breeds previously genotyped for breed structure studies (15, 16).

SNP Genotyping. One hundred sixty-eight village dogs, 102 mixed-breed dogs, and dogs from 126 breeds were genotyped using the sequenom iPLEX platform on a 321-SNP panel described in ref. 20. For each sample, 2 μL of dog genomic DNA was aliquoted into 13 separate microtiter wells for PCR amplification. Each genomic aliquot was amplified in a total volume of 10 μL >45 cycles with up to 28 primer pairs. Each reaction was treated with shrimp alkaline phosphatase for 40 min before heat inactivation. Primer extension reactions were carried out in a standard thermocycler according to the sequenom iPLEX gold protocol. Each reaction was desalted before spotting and shooting a SpectroChip on the Compact MassARRAY system (Sequenom). Results were interpreted automatically using cluster plots with the Histogram tabular view active in SpectroTyper-TyperAnalyzer (Sequenom). SNP genotypes were loaded into PLINK version 1.0.4 (21) and 15 SNPs with high missingness (>20%) and 1 SNP with an extreme heterozygote deficiency ($P < 10^{-7}$ below Hardy-Weinberg equilibrium) were removed from further analysis.

Mitochondrial Sequencing. A 680-bp fragment of the mitochondrial D-loop was amplified in two overlapping reactions. Region-1 was amplified using forward primer H15422: 5'-CTCTTGCTCCACCATCAGC-3', and reverse primer L15781: 5'-GTAAGAACCAGATGCCAGG-3'. Region-2 was amplified using forward primer H15693 5'-AATAAGGGCTTAATCACCATGC-3' and reverse primer L16106: 5'-AAACTATATGTCCTGAAACC-3' (primer names correspond to 3' most position of primer, relative to the published dog mitochondrial genome as in (6)). PCR was carried out under the following protocol using 10 ng genomic DNA: Denaturation: 94 °C (40 s); annealing: 54 °C (1 min); amplification: 72 °C (1 min) for 35 total cycles followed by a 5 min final annealing step at 72 °C. Sequencing reactions

were carried out on an ABI 3730 sequencer using BigDye Terminator chemistry using the Region-1 reverse primer and Region-2 forward primer. Any reads with ambiguous bases were rerun in the opposite direction. Sequences were edited, assembled, and aligned with Sequencher 4.8 (Gene Codes Corporation) and submitted to GenBank with Sequin (http://www.ncbi.nlm.nih.gov/Sequin/).

Statistical Analyses. We used two approaches—principal component analysis with EIGENSOFT v2.0 (22) and clustering analysis with STRUCTURE v2.2 (14)—to classify individuals as indigenous or non-native and to describe the genetic structure of indigenous African village dogs and their relationship to dogs from putatively African breeds. We relied primarily on STRUCTURE to determine the proportion of non-African admixture present in each village dog because structure allows for probabilistic assignment of individuals to classes and explicit modeling of admixture (22). In contrast, PCA makes no assumptions regarding discrete versus clinal population structure and is well suited for describing the principal axes of genetic variation between populations. In practice, STRUCTURE and PCA usually reveal very similar patterns of genetic variation (22).

Before running these clustering methods, we removed markers in high LD with other markers [r^2>0.5, see (23)] using Arlequin v3.11 (24) and removed 9 village dogs that showed high relatedness to another dog in the genotyping panel ($\pi_{hat} > 0.3$). All STRUCTURE runs were done using the admixture model with correlated allele frequencies, no prior population information, and default parameter settings with a burnin period of 100,000 iterations followed by 500,000 MCMC repetitions, with 10 runs per K, and averaged using CLUMPP v1.1.2 (25). In contrast, PCA was carried out separately for the SNP and microsatellite markers. Microsatellite loci with $n > 2$ alleles were recorded as n-1 biallelic loci before running PCA in Eigensoft.

Expected heterozygoisty (h) was calculated in Arlequin after removing 10 dogs that appeared to be r approximately 0.5 related. F_{ST} based on SNP loci was computed with a custom C++ implementation of Eq. **6** from (26); microsatellite F_{ST} was computed using Arlequin. Unless otherwise noted, statistical tests were performed in R v2.6.2 (27). STRUCTURE results were plotted using Distruct v1.1 (28).

1. Wayne R (2001) Consequences of domestication: Morphological diversity of the dog. In *The Genetics of the Dog*, eds Ruvinsky A, Sampson J (CABI Publishing, Oxon, UK), pp 43–60.
2. Clutton-Brock J (1995) Origins of the dog: Domestication and early history. In *The Domestic Dog, Its Evolution, Behavior and Interactions with People*, ed Serpell J CUP, Cambridge), pp 7–20.
3. Vilà C, Maldonado J, Wayne R (1999) Phylogenetic relationships, evolution, and genetic diversity of the domestic dog. *J Hered* 90:71–77.
4. Germonpré M, et al. (2009) Fossil dogs and wolves from Palaeolithic sites in Belgium, the Ukraine and Russia: Osteometry, ancient DNA and stable isotopes. *J Arch Sci* 36:473–490.
5. Vilà C, et al. (1997) Multiple and ancient origins of the domestic dog. *Science* 276:1687–1689.
6. Savolainen P, Zhang Y, Luo J, Lundeberg J, Leitner T (2002) Genetic evidence for an East Asian origin of domestic dogs. *Science* 298:1610–1613.
7. Coppinger R, Coppinger L (2001) in *Dogs: A Startling New Understanding of Canine Origin, Behavior and Evolution* (Scribner, New York).
8. Dobney K, Larson G (2006) Genetics and animal domestication: New windows on an elusive process. *J Zool* 269:261–271.
9. Miklosi A (2008) in *Dog Behaviour, Evolution, and Cognition* (Oxford Univ Press, Oxford), p 304.
10. Pires A, et al. (2006) Mitochondrial DNA sequence variation in Portuguese native breed dogs: diversity and phylogenetic affinities. *J Hered* 97:318–330.
11. Irion D, Schaffer A, Grant S, Wilton A, Pedersen N (2005) Genetic variation analysis of the Bali street dog using microsatellites. *BMC Genet* 6:6.
12. Runstadler J, Angles J, Pedersen N (2006) Dog leucocyte antigen class II diversity and relationships among indigenous dogs of the island nations of Indonesia (Bali). *Australia and New Guinea Tissue Antigens* 68:418–426.
13. (2008) Police Zone. Encyclopædia Britannica. Online Ed.
14. Pritchard J, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–949.
15. Parker HG, et al. (2004) Genetic structure of the purebred domestic dog. *Science* 304:1160–1164.
16. Parker H, et al. (2007) Breed relationships facilitate fine-mapping studies: A 7.8-kb deletion cosegregates with Collie eye anomaly across multiple dog breeds. *Genome Res* 2007:1562–1571.
17. Gundry R, et al. (2007) Mitochondrial DNA analysis of the domestic dog: Control region variation within and among breeds. *J Forensic Sci* 52:562–572.
18. Savolainen P, Leitner T, Wilton A, 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.
19. Björnerfeldt S, Webster M, Vilà C (2006) Relaxation of selective constraint on dog mitochondrial DNA following domestication. *Genome Res* 16:990–994.
20. Jones P, et al. (2008) Single-nucleotide-polymorphism-based association mapping of dog stereotypes. *Genetics* 179:1033–1044.
21. Purcell S, et al. (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559–575.
22. Patterson N, Price A, Reich D (2006) Population structure and eigenanalysis. *PLoS Genet* 2:e190.
23. Kaeuffer R, Réale D, Coltman D, Pontier D (2007) Detecting population structure using STRUCTURE software: Effect of background linkage disequilibrium. *J Hered* 99:374–380.
24. Excoffier L, Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evol Bioinform Online* 1:47–50.
25. Jakobsson M, Rosenberg N (2007) CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23:1801–1806.
26. Weir B, Cockerman C (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358–1370.
27. R Development Core Team (2008) in *R: A language and environment for statistical computing* (R Foundation for Statistical Computing, Vienna, Austria).
28. Rosenberg N (2004) DISTRUCT: A program for the graphical display of population structure. *Mol Ecol Notes* 4:137–138.
29. Ewens W (1972) The sampling theory of selectively neutral alleles. *Theor Pop Biol* 3:87–112.

Boyko et al.

Supporting Information

Boyko et al. 10.1073/pnas.0902129106
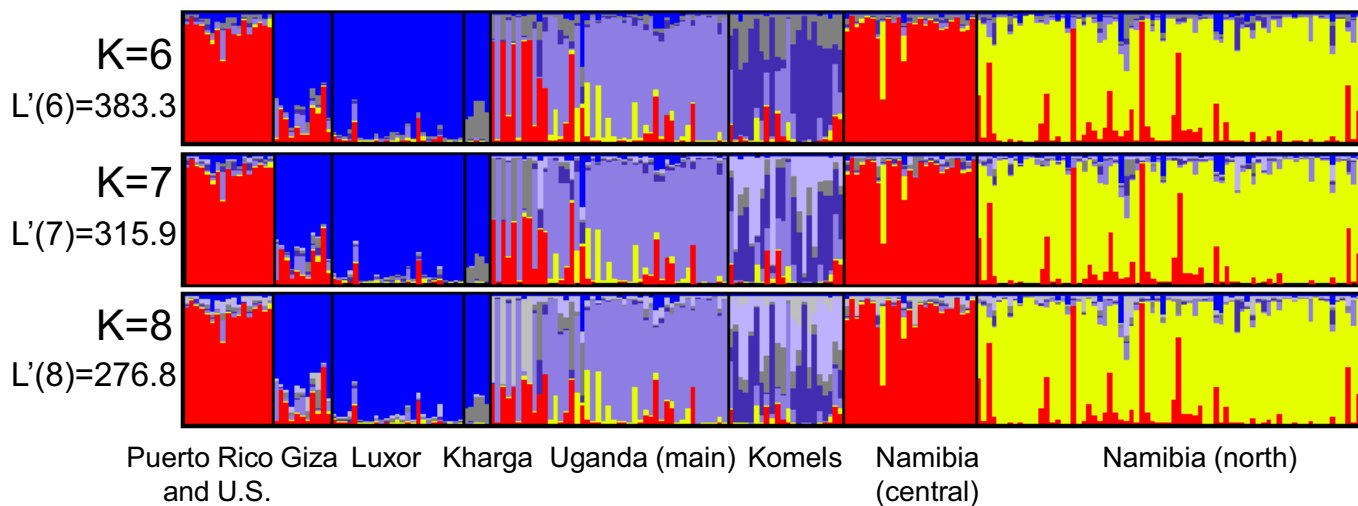
Fig. S1.   STRUCTURE results as in Fig. 2 for higher K. L'(K) defined as L(K)-L(K-1) where L(K) is the average of the mean ln likelihood across all 10 runs for a given K. L'(K) = 1499.4, 1214.3, and 646.7 for K = 3, 4, and 5, respectively.
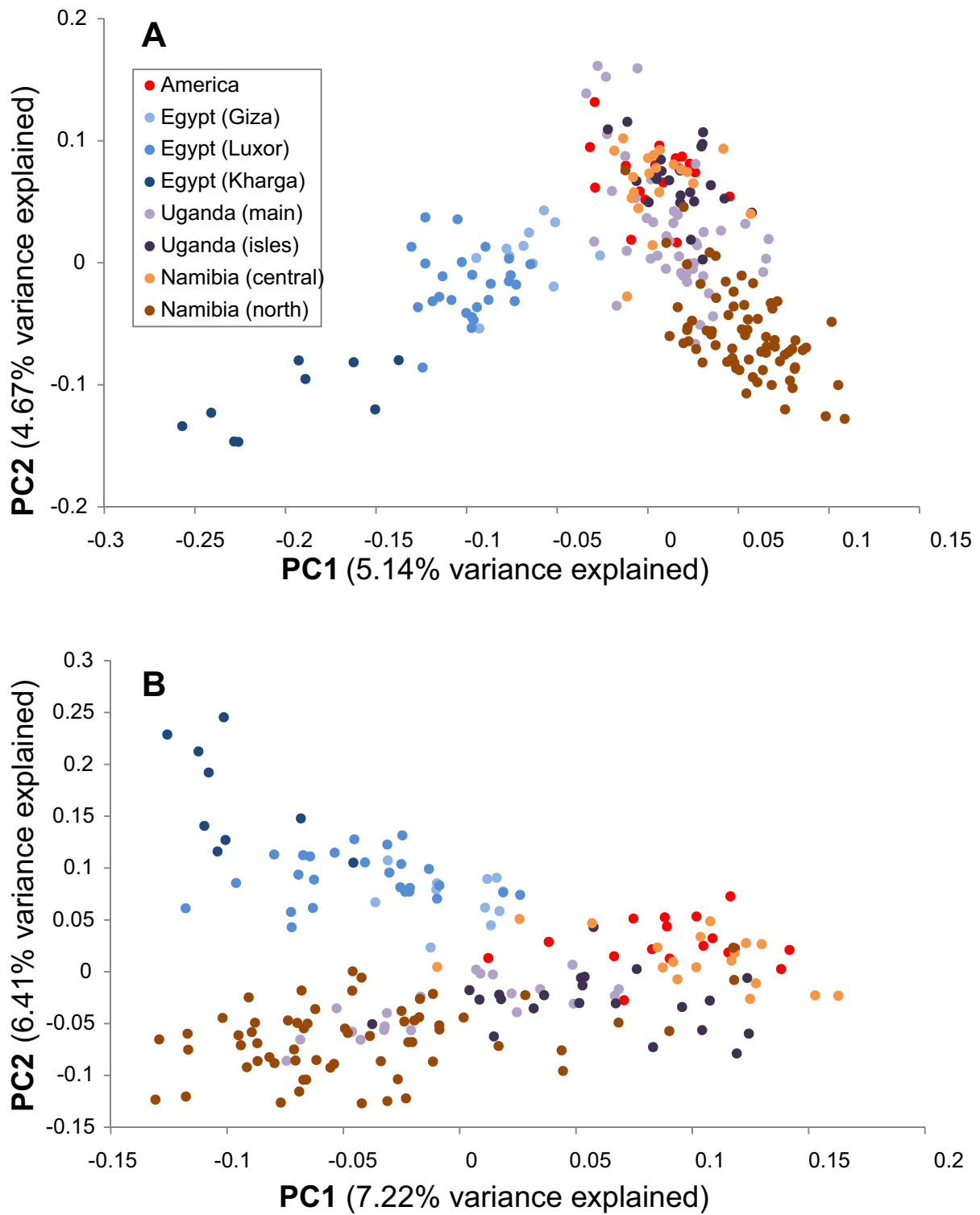
Fig. S2. Principal component analysis of African village dogs (all including admixed individuals) and American dogs. (*A*) PCA with the 89 microsatellite loci (*n* = 227). (*B*) PCA with the 300 SNP loci (*n* = 186).
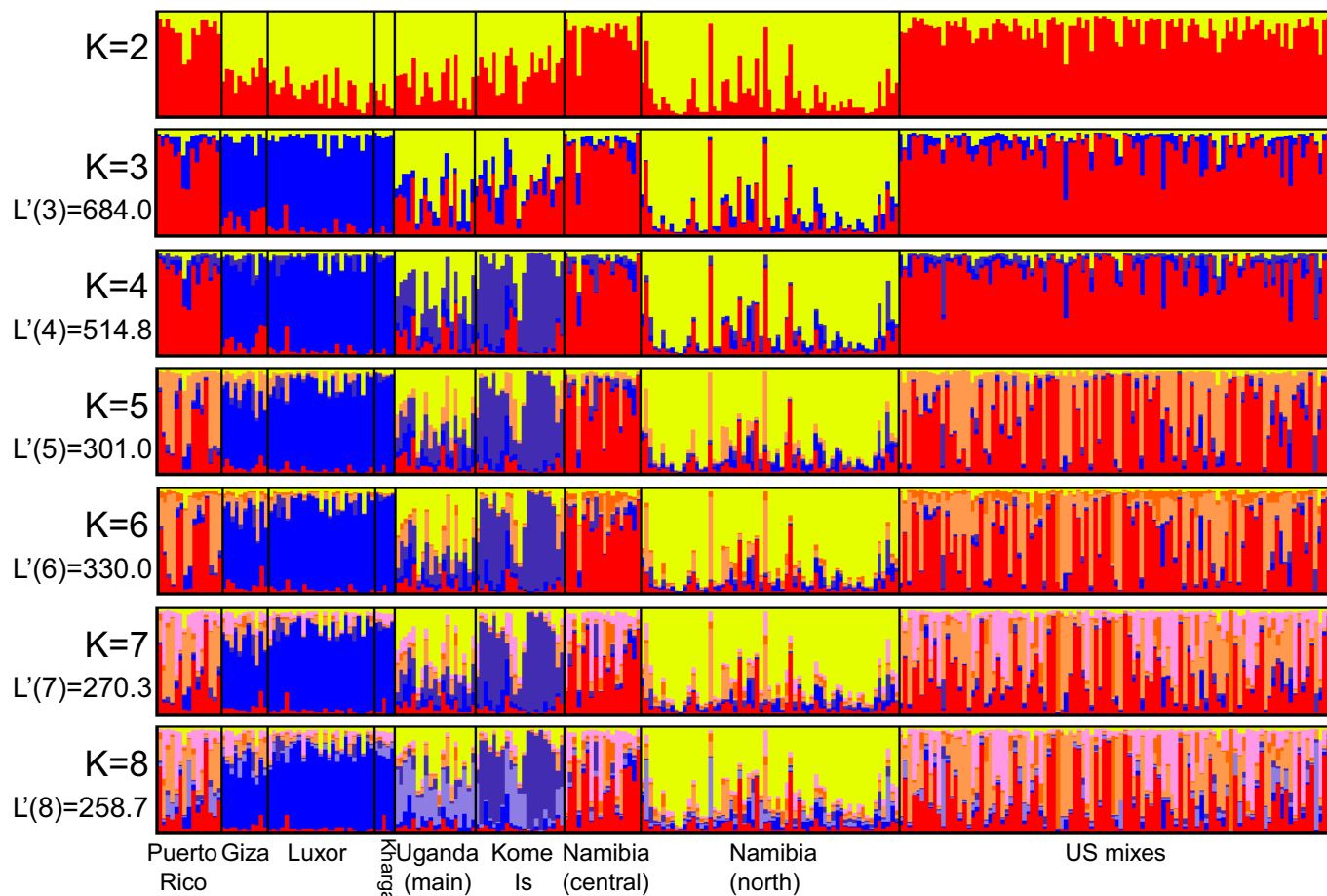
Fig. S3. STRUCTURE analysis of 184 village dogs and 102 complex mixed breed dogs from the United States at 300 SNP loci.
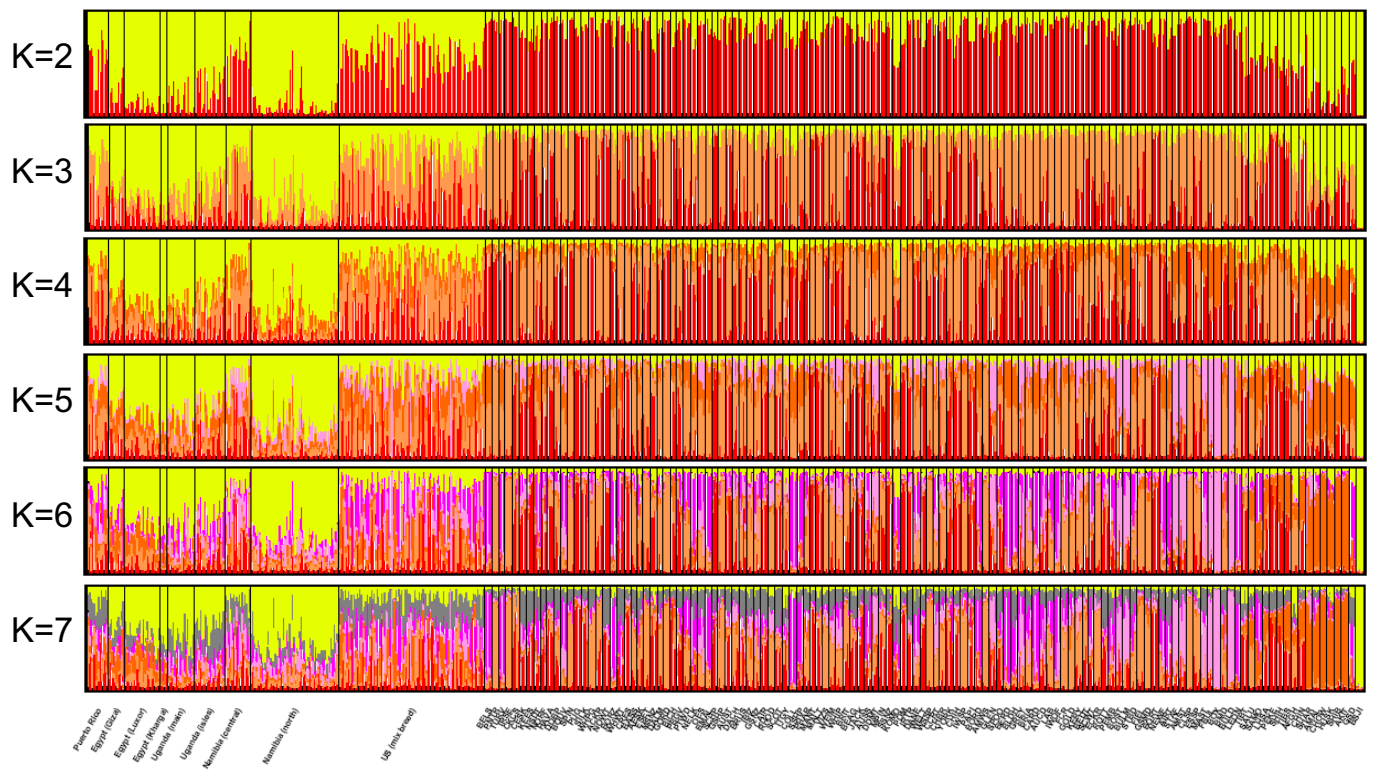
Fig. S4. STRUCTURE analysis of village dogs and dogs from 126 dog breeds using 300 SNP loci. An African cluster is apparent at K = 2 while subsequent clusters detect structure between breed dog groups. L'(K) declines sharply beyond K = 6 (L'(K) = 2280.5, 1899.8, 1638.5, 1609.2, and 1122.0 for K = 3–7, respectively).

African village + breed dog microsatellite PCA
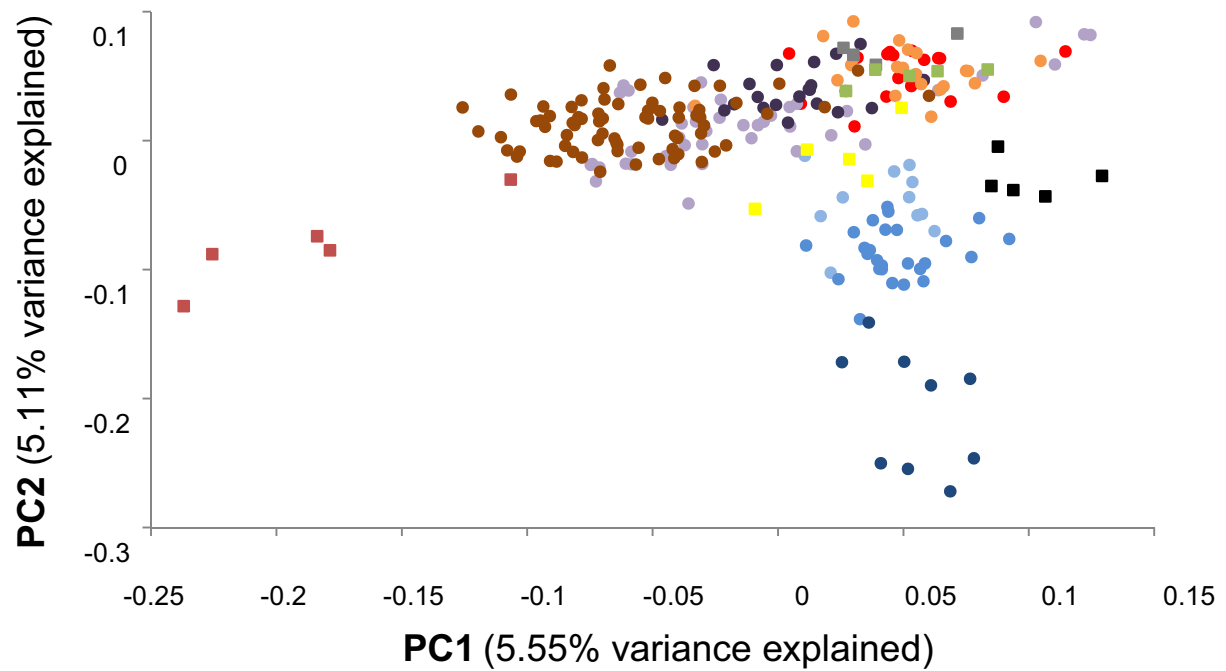
Fig. S5. Principal component analysis of village dogs and dogs from five putatively African and Middle Eastern breeds across 89 microsatellite markers in 227 village dogs and 24 breed dogs.
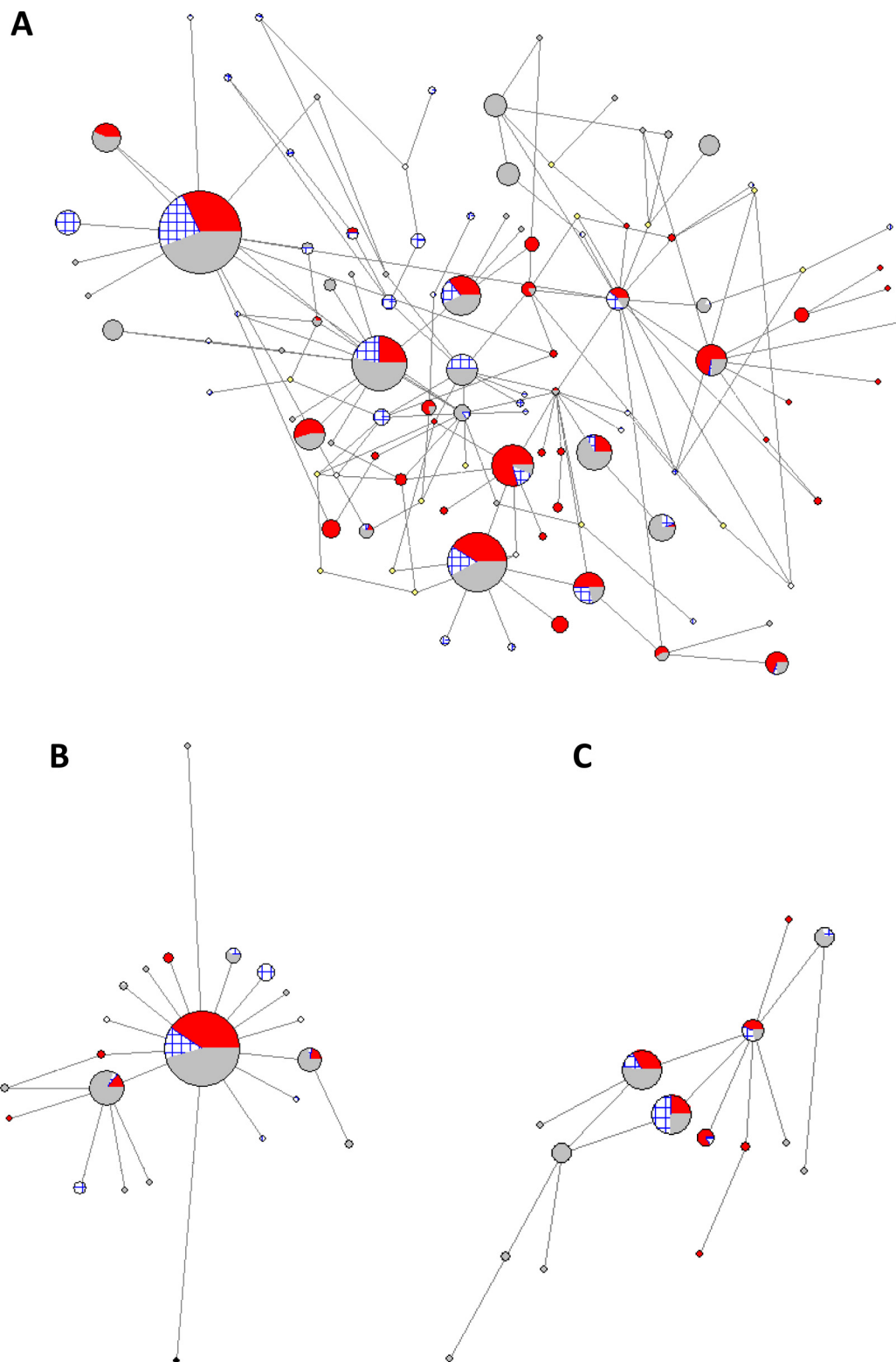
Fig. S6. Haplotype network graphs of 582-bp region excluding indels for mtDNA clades *A–C*. Data from (6, 10) and this study. Red = Africa, blue hatching = East Asia; gray = other. Area of node proportional to number of dogs with that haplotype.

Table S1. Summary of sampling locations

Country	Region	Location name	N	N (nonrel)	N (auto)	Mal	Fem	I/?/M	Location	Elev	Date	mtDNA haplotypes
Egypt	Giza	Animal Friends shelter	11	11	11	6	5	7/4/0	30° 02.724′ N, 31° 14.190′ E	80′	08/15/07	vilA1 (2), vilA2 (3), vilA5 (2), vilC3a, vilC2, vilA11a, vilA12
Egypt	Luxor	ACE (Animal Care in Egypt)	10	9	9	2	8	9/0/0	25° 41.406′ N, 32° 39.139′ E	240′	08/17/07	vilA3 (3), vilC1a (2), vilA4, vilA2, vilC2, vilA6
Egypt	Kharga	Bagdad	7	4	7	3	4	4/0/0	24° 44.819′ N, 30° 35.524′ E	82′	08/18/07	vilB1a (3), vilB2
Egypt	Kharga	Jeddah	2	1	2	1	1	1/0/0	24° 52.949′ N, 30° 33.912′ E	138′	08/18/07	vilB1a
Egypt	Luxor	Gouahera	6	6	5	2	4	5/0/0	25° 41.406′ N, 32° 39.139′ E	240′	08/19/07	vilA1, vilA3, vilA7 (2), vilA8 (2)
Egypt	Luxor	Armanti	1	1	1	0	1	1/0/0	25° 37.028′ N, 32° 32.565′ E	261′	08/19/07	vilC1a [or vilC1b]
Egypt	Luxor	La Galta	1	1	1	1	0	1/0/0	25° 40.974′ N, 32° 36.876′ E	266′	08/19/07	vilA6
Egypt	Luxor	Ezba	2	2	2	2	0	2/0/0	25° 41.641′ N, 32° 37.311′ E	276′	08/19/07	vilB1b, vilC1b
Egypt	Luxor	Comb	2	2	3	0	2	3/0/0	25° 42.163′ N, 32° 36.417′ E	251′	08/19/07	vilB1b, vilC1b
Egypt	Luxor	Bairat	2	2	2	0	2	2/0/0	25° 42.718′ N, 32° 37.428′ E	254′	08/19/07	vilA9, vilA10
Egypt	Luxor	Jazeera	4	3	4	4	0	3/0/0	25° 42.421′ N, 32° 37.782′ E	258′	08/19/07	vilA3, vilA7, vilC4
Uganda	mainland	Napoli	10	10	3	7	3	1/0/2	1° 02.99′ N, 34° 10.892′ E	3780′	07/17/07	vilA1 (3), vilA2, vilA11a (2), vilA11b (2), vilA13, vilB1b
Uganda	mainland	Walchwaba	10	10	5	4	6	2/0/3	1° 02.805′ N, 34° 10.957′ E	3780′	07/17/07	vilA6, vilA11a (2), vilA11b (2), vilA13 (5)
Uganda	mainland	Namabasa	9	9	3	4	5	1/1/1	1° 06.39′ N, 34° 09.726′ E	3653′	07/17/07	vilA11a (4), vilA15, vilB1b (3), vilC3b
Uganda	mainland	Mooni	20	20	7	9	11	5/1/1	1° 03.532′ N, 34° 12.078′ E	3703′	07/18/07	vilA6 (2), vilA11a (3), vilA13 (4), vilA32 (5), vilB1b (5), vilC3b
Uganda	mainland	Busoba	10	10	5	6	4	5/0/0	0° 58.133′ N, 34° 10.237′ E	3926′	07/18/07	vilA2, vilA6 (4), vilA13, vilA25, vilA33 (2), vilB4
Uganda	mainland	Buchenda	1	1	1	1	0	1/0/0	0° 57.514′ N, 34° 18.194′ E	3925′	07/18/07	vilA16
Uganda	mainland	Bunanimi	6	6	3	2	4	3/0/0	0° 57.596′ N, 34° 11.597′ E	4026′	07/18/07	vilA2, vilA11a (3), vilA16, vilA27
Uganda	mainland	Nugisu	7	7	3	5	2	3/0/0	1° 05.376′ N, 34° 10.766′ E	3665′	07/18/07	vilA1, vilA2, vilA14, vilA16, vilA25 (2), vilB1b
Uganda	mainland	Navigyo	4	4	1	3	1	1/0/0	1° 05.095′ N, 34° 10.977′ E	3631′	07/18/07	vilA5 (3), vilC2
Uganda	mainland	Butandiga	16	15	4	9	7	3/1/0	1° 04.334′ N, 34° 12.093′ E	3814′	07/19/07	vilA2, vilA6, vilA7, vilA13, vilA16, vilA23, vilA32, vilB1b (7), vilB4
Uganda	mainland	Kongoidi	7	7	3	4	3	3/0/0	1° 11.915′ N, 34° 07.152′ E	3777′	07/19/07	vilA2, vilA6 (4), vilB1b (2)
Uganda	mainland	Komosingo	20	19	8	10	10	6/1/0	1° 12.584′ N, 34° 07.950′ E	3804′	07/19/07	vilA2, vilA6 (8), vilA11a (2), vilA15, vilA24, vilA29, vilB1b (5)
Uganda	isles	Mwoma (Ddamba Is.)	8	7	6	6	2	5/0/0	0° 2.414′ S, 32° 46.557′ E	1135′	08/10/07	vilA23 (2), vilA30 (4), vilB1b
Uganda	isles	Busiro (Koome Is.)	3	2	2	3	0	2/0/0	0° 1.952′ S, 32° 44.843′ E	1135′	08/10/07	vilA30, vilB1b
Uganda	isles	Tabaliro (Nsadzi Is.)	5	4	5	3	2	3/2/0	0° 5.203′ S, 32° 35.750′ E	1135′	08/10/07	vilA5 (3), vilA6
Uganda	isles	Kisigala (Koome Is.)	7	6	5	5	2	5/0/0	0° 5.603′ S, 32° 41.094′ E	1135′	08/10/07	vilA31 (4), vilB1b (2)
Uganda	isles	Zingoola (Koome Is.)	11	11	5	9	2	3/2/0	0° 3.319′ S, 32° 42.941′ E	1135′	08/11/07	vilA2, vilA11a, vilA23, vilA31 (3), vilB1b (5)
Namibia	central	Otavi	6	6	5	3	3	0/0/5	19° 39.192′ S, 17° 27.937′ E	5035′	08/23/07	vilA2, vilA13 (3), vilC3a (2)

Country	Region	Location name	N	N (nonrel)	N (auto)	Mal	Fem	I/?/M	Location	Elev	Date	mtDNA haplotypes
Namibia	north	Oshivelo	4	4	4	2	2	2/0/2	18° 36.957′ S, 17° 09.995′ E	3627′	08/23/07	vilA13 (2), vilB1b, vilC5
Namibia	north	Oshikango	6	6	6	3	3	6/0/0	18° 36.957′ S, 17° 09.995′ E	3626′	08/24/07	vilA11a, vilA13, vilA15 (3), vilB1b
Namibia	north	Onhuno	6	6	4	4	2	3/1/0	17° 30.905′ S, 15° 54.126′ E	3629′	08/24/07	vilA2, vilA13 (2), vilA17, vilB1b (2)
Namibia	north	Ongha	6	6	5	6	0	4/0/1	17° 38.623′ S, 15° 55.651′ E	3615′	08/24/07	vilA7, vilA11a, vilA16, vilA18, vilB1b, vilB1a
Namibia	north	Indiangungu	6	5	4	6	0	4/0/0	17° 47.694′ S, 16° 00.280′ E	3650′	08/24/07	vilA7, vilA11a, vilA18, vilA19, vilB3a
Namibia	north	Ondandwa	6	6	6	2	4	5/1/0	17° 54.945′ S, 15° 58.605′ E	3595′	08/24/07	vilA11a, vilA13, vilA15, vilA16, vilA20, vilB3b
Namibia	north	Olund	6	6	4	3	3	2/1/1	17° 55.261′ S, 16° 0.123′ E	3573′	08/25/07	vilA5, vilA6, vilA14 (2), vilA19 (2)
Namibia	north	Omaarara	7	7	4	5	2	4/0/0	17° 51.249′ S, 15° 52.894′ E	3606′	08/25/07	vilA5 (3), vilA11a, vilA13, vilA19 (2)
Namibia	north	Okanbjengedhi	11	11	7	6	5	6/0/1	17° 47.145′ S, 15° 43.760′ E	3597′	08/25/07	vilA11a (3), vilA13 (2), vilA15, vilA20, vilA21, vilB1b (2), vilC5
Namibia	north	Oshikuku	6	6	4	4	2	3/1/0	17° 39.419′ S, 15° 29.081′ E	3606′	08/25/07	vilA11a, vilA20, vilA22 (2), vilB1b (2)
Namibia	north	Omavela	5	5	5	3	2	5/0/0	17° 42.038′ S, 15° 43.372′ E	3588′	08/25/07	vilA11a, vilA13, vilA23, vilB1b, vilB3a
Namibia	north	Endola	7	7	6	5	2	6/0/0	17° 35.841′ S, 15° 43.313′ E	3630′	08/26/07	vilA5, vilA11a, vilA13 (3), vilA17, vilA20
Namibia	north	Onyvulae	9	9	8	9	0	8/0/0	18° 04.315′ S, 16° 31.653′ E	3656′	08/26/07	vilA7, vilA11a (3), vilA13 (2), vilA22, vilA24, vilA34
Namibia	north	Onyati	3	3	2	1	2	2/0/0	18° 13.108′ S, 16° 24.927′ E	3609′	08/26/07	vilA2 (3)
Namibia	north	Cham-Cham	5*	4	5	2	2	1/3/0	18° 28.495′ S, 16° 57.160′ E	3603′	08/26/07	vilA13 (2), vilA16, vilC6
Namibia	central	Tsumeb	11	11	9	7	4	0/1/8	19° 14.942′ S, 17° 42.197′ E	4223′	08/26/07	vilA13 (2), vilA14 (3), vilB1b (2), vilB1a, vilC2 (3)
Namibia	central	Grootfontaine	12	11	11	6	6	0/0/11	19° 35.234′ S, 18° 06.170′ E	4693′	08/27/07	vilA7, vilA11a (4), vilA13, vilA16 (4), vilC3a
America	Indiana	Tippecanoe Cty Humane Society	1	1	1	0	1	0/0/1	40° 24.089′ N, 86° 53.753′ W	604′	08/23/07	vilA27
America	Virginia	Norfolk SPCA	1	1	1	1	0	0/0/1	36° 51.190′ N, 76° 14.878′ W	23′	07/05/07	vilB1b
America	Puerto Rico	Albergue de Mayaguez	5	5	5	4	1	0/0/5	18° 12.760′ N, 67° 7.708′ W	948′	06/29/07	vilA13 (2), vilA27, vilA28, vilB1b
America	Puerto Rico	Albergue de Ponce	11	10	11	4	7	0/0/10	18° 00.356′ N, 66° 38.911′ W	94′	06/30/07	vilA2 (2), vilA11a (3), vilA16, vilA25, vilA26, vilB1b (2)

I/?/M denotes number of indigenous (<25% admixed), uncertain (25–60% admixed), and admixed (>60% admixed) dogs (unrelated genotyped dogs only) N (nonrel) denotes number of dogs sequenced across D-loop, excluding relatives; N (auto) denotes number of dogs with SNP and/or microsatellite genotyping
*Sex was not recorded for one dog from Cham-Cham

Table S2. Pairwise F$_{ST}$ in village dogs between regions based on 89 microsatellite markers

	Giza	Kharga	Luxor	NA_cent	NA_north	UG_isles	UG_main	America
Giza	—							
Kharga	7.79%	—						
Luxor	0.57%	7.35%	—					
NA_cent	2.22%	12.40%	4.42%	—				
NA_north	2.85%	11.01%	3.90%	3.71%	—			
UG_isles	4.35%	13.24%	5.62%	4.97%	5.18%	—		
UG_main	1.62%	10.33%	3.22%	2.50%	2.37%	2.47%	—	
America	2.38%	12.73%	4.13%	0.25%	3.55%	4.33%	2.22%	—

Table S3. Multiple sequence alignment of the 60 mtDNA haplotypes found in the 337 village dogs

vilA1	TG-TCCTCAACTTTCTGTTTAAATGTCCTACTCATTCACAGCAC
vilA2	..-. . . . . .A.C.T. T
vilA3	..-. . . . . C. . . . .
vilA4	..-. . . . . C. . . C. . . . .
vilA5	..-. . . . . .A.C.T. . . . . .C. . . . .T
vilA6	..-. . . . . C. . . . . T.T
vilA7	..-. . . . . CA.C.T. . . . . T
vilA8	..-. . . . . .A.C. . . . . A. . . . .
vilA9	..-. . . . . C. . . . C. . . . .
vilA10	..-. . . . . C. . . . . .G. . . . .
vilA11a	..-. . . . . .A.C. . . . .
vilA11b	..C-. . . . . .A.C. . . . .
vilA12	..-. . . . . G. . . . . .
vilA13	..-. . . . . C. . . . C. . . . . T.T
vilA14	..-. . . . . C. . . C. . . . . C. . . . .
vilA15	..-. . . . . .A.C.T. . . . .
vilA16	..-. . . . . C. . . . C. T
vilA17	..-. . . . . .A.C. . . . . T.T
vilA18	..-. . . . . .A.C. . . G. . . . .T
vilA19	..-. . . . . .A.C. . . G. . . . .
vilA20	..-. . . . . .A.C. . . . . T.T. . . . . T
vilA21	..-. . . . . C. . . . . AC. . . . . T. . . . . . T
vilA22	..-. . . C. . . . . .T. . . C. . . . . T
vilA23	..-. . . . . C. . . . .
vilA24	..-. . . . . T. . . C. T
vilA25	..-.T. . . . . .C. . . . . T
vilA26	..-. . . . . C. . . . . A. . . . . T
vilA27	..-.T. . . . . .C. . . . . .T. . . . T
vilA28	..-. . . . . T.A.C. . . . . T.T. . . . . T
vilA29	..-. . . . . C. . . . . A. . . . . T..
vilA30	..-. . . . . C. . . . . A. . . . .
vilA31	..-. . . . . C. . . CC. . . . . T.T
vilA32	..-. . . . . .A.C.T. . . . . T. . . . . . C. . . . . ..T
vilA33	..-. . . . . C. . . . .G. . . . . .T.T
vilA34	..-. . . . . G. . . . . .A.C. . . . . T
vilB1a	.A-. . . . T..T.C. . . A. . . GG..A. . . . . . C. . . C..T.-TGT
vilB1b	.A-. . . . T..T.C. . . A. . . GG..A. . . . . . C. . . C..T.-TGT
vilB2	.A-. . . . T.TT.C. . . A. . . GG..A. . . . . . C. . . C..T.-TGT
vilB3a	.A-. . . . T. . . . C. . . A. . . GG..A. . . . . . C. . . C..T-.TGT
vilB3b	.A-. . . . T. . . . C. . . A. . . GG..A. . . . . . C. . . C..T..TGT
vilB4	.A-. . . . T..T.C. . . A. . . G. . . A. . . . . . C. . . C..T..TGT
vilC1a	.A-..T.TG..C. . . . A.C.G..C. . . . . . .C. . . . . . T.-TGT
vilC1b	CA-..T.TG..C. . . . A.C.G..C. . . . . . .C. . . . . . T.-TGT
vilC2	.A-..T.T. . . C. . . . A.C.G..C..T. . . . C. . . . . . T.-TGT
vilC3a	.A-..T.T. . . C. . . . A.C.G..C. . . . . . .C. . . . . . T.-TGT
vilC3b	.A-..T.T. . . C. . . . A.C.G..C. . . . . . .C. . . . . . T..TGT
vilC4	.A-..T.T. . . . . A.C.G..C. . . . . . .C. . . . . . T.-TGT
vilC5	.A-..T.T. . . C..T.A.C.G..C. . . . . . .C. . . . . T.-TGT
vilC6	.A-..T.T. . . C..T.A.C.G..C. . . . C..C. . . . . . T.-TGT

Table S4. Haplotype diversity in various regions in Africa (Pires et al., 2006 and this study) and East Asia (Savolainen et al., 2002).

Region	# Samples	# Haplotypes	Area 10^3 km^2
Egypt	42	18	1002
NW Africa	29	15	163
Uganda	148	23	241
Namibia (north)	91	22	824
Japan	96	26	378
Sichuan	48	11	485
Tibet	23	13	1230
China (Non-specified)	22	10	9598
Shanxi	20	12	157
Korea	11	5	220
Thailand	10	9	513
Guangxi	8	6	237
Indonesia	7	5	1905
Liaoning	6	5	146
Mongolia	2	2	1564
Cambodia	2	2	181
Anhui	2	1	139
New Guinea	1	1	463
Philippines	1	1	300
Vietnam	1	1	332

Table S5. Microsatellite markers used in this study (see Parker et al., 2004, 2007)

M#	Marker name	Chrom number	Dye label*	Annealing temp	MgCl$_2$ conc	Electrophoresis conditions†	Notes
1	REN285G14	1	PET	55	2	2ul,multiplexed	
2	C01.673	1	VIC	58	1.5	2 ul, multiplexed	
3	REN112I02	1	PET	58	1.5	2 ul, multiplexed	
4	REN172C02	1	6-FAM	55	2	2 ul, multiplexed	
5	FH2793	1	6-FAM	58	1.5	2 ul, multiplexed	
6	REN143K19	1	6-FAM	55	2	2 ul, multiplexed	
7	FH2890	2	VIC	55	2	2 ul, multiplexed	
8	C02.466	2	NED	58	1.5	2 ul, multiplexed	
9	C02.894	2	PET	55	1.5	2 ul, multiplexed	
11	FH2895	3	PET	58	1.5	2 ul, multiplexed	removed (missingness)
12	REN157C08	3	VIC	55	2	2 ul, multiplexed	
13	C03.445	3	NED	55	1.5	2 ul, multiplexed	
14	FH2732	4	6-FAM	58	1.5	2 ul, multiplexed	
15	FH2776	4	VIC	58	1.5	2 ul, multiplexed	
16	REN160J02	4	PET	58	1.5	2 ul, multiplexed	
17	REN262N08	4	6-FAM	55	2	5–10 ul	
18	REN92G21	5	6-FAM	55	1.5	2 ul, multiplexed	removed (het deficit)
19	REN285I23	5	PET	55	2	2 ul, multiplexed	
20	C05.414	5	PET	55	1.5	2 ul, multiplexed	
21	FH2752	6	6-FAM	58	1.5	2 ul, multiplexed	
22	REN210I14	6	NED	55	2	2 ul, multiplexed	
23	REN37H09	6	PET	58	1.5	2 ul, multiplexed	
24	REN97M11	7	6-FAM	55	2	2 ul, multiplexed	
25	REN286L19	7	VIC	58	1.5	2 ul, multiplexed	
26	FH2860	7	6-FAM	55	2	2 ul, multiplexed	
27	REN204K13	8	NED	55	2	2 ul, multiplexed	
28	C08.373	8	VIC	58	1.5	2 ul,multiplexed	
29	C08.618	8	NED	55	2	2 ul, multiplexed	
30	C09.173	9	NED	58	1.5	2 ul, multiplexed	
31	C09.474	9	6-FAM	55	2	2 ul, multiplexed	
32	FH2885	9	VIC	55	2	2 ul, multiplexed	
33	C10.781	10	6-FAM	55	2	2 ul, multiplexed	
34	REN73F08	10	VIC	55	2	2 ul, multiplexed	
35	REN154G10	10	NED	55	2	2 ul, multiplexed	
36	REN164B05	11	PET	55	2	2 ul, multiplexed	
37	FH2874	11	VIC	55	2	2 ul, multiplexed	
38	C11.873	11	VIC	58	1.5	2 ul, multiplexed	
40	REN213F01	12	VIC	55	2	2 ul, multiplexed	
41	REN208M20	12	6-FAM	58	1.5	2 ul, multiplexed	
42	REN94K11	12	VIC	55	2	2 ul, multiplexed	
44	REN286P03	13	PET	55	1.5	2 ul, multiplexed	
45	C13.758	13	6-FAM	55	2	2 ul, multiplexed	
46	C14.866	14	VIC	55	2	2 ul, multiplexed	
47	FH3072	14	NED	55	2	2 ul,multiplexed	
48	FH3802	15	PET	55	2	2 ul, multiplexed	
49	REN06C11	15	PET	55	1.5	2 ul, multiplexed	
50	REN144M10	15	6-FAM	58	1.5	2 ul, multiplexed	
51	REN85N14	16	6-FAM	58	1.5	2 ul, multiplexed	
52	FH3096	16	NED	55	2	2 ul, multiplexed	
53	C17.402	17	NED	58	1.5	2 ul, multiplexed	
54	REN50B03	17	PET	58	1.5	2 ul, multiplexed	
55	REN112G10	17	VIC	55	2	2 ul, multiplexed	
56	REN186N13	18	VIC	55	1.5	2 ul, multiplexed	
57	FH2795	18	NED	55	1.5	2 ul, multiplexed	
58	C18.460	18	PET	58	1.5	2 ul, multiplexed	
59	FH2783	19	NED	55	2	2 ul, multiplexed	
60	REN91I14	19	VIC	55	1.5	2 ul, multiplexed	
61	REN274F18	19	NED	58	1.5	2 ul, multiplexed	
62	FH2887	20	6-FAM	55	2	2 ul, multiplexed	
63	FH3109	20	PET	55	1.5	2 ul, multiplexed	
64	REN293N22	20	VIC	55	1.5	2 ul, multiplexed	
65	FH2914	21	VIC	55	2	2 ul, multiplexed	
66	FH3069	21	NED	55	2	2 ul, multiplexed	
67	REN49F22	22	PET	55	2	5–10 ul	removed (missingness)

M#	Marker name	Chrom number	Dye label*	Annealing temp	MgCl$_2$ conc	Electrophoresis conditions†	Notes
68	REN107H05	22	NED	55	2	2 ul, multiplexed	removed (het deficit)
69	REN78I16	22	PET	55	1.5	2 ul, multiplexed	
70	FH3078	23	VIC	55	2	2 ul, multiplexed	
71	C23.277	23	NED	55	2	2 ul, multiplexed	
72	REN181K04	23	NED	58	1.5	2 ul, multiplexed	removed (het deficit)
73	REN106I06	24	PET	55	2	2 ul, multiplexed	
74	FH3083	24	PET	55	2	2 ul, multiplexed	removed (het deficit)
75	REN54E19	25	6-FAM	55	2	2 ul, multiplexed	
77	REN87O21	26	VIC	55	2	2 ul, multiplexed	
78	C26.733	26	NED	55	2	2 ul, multiplexed	
79	C27.442	27	VIC	55	2	2 ul, multiplexed	
80	C27.436	27	NED	55	2	2 ul, multiplexed	removed (het deficit)
81	REN72K15	27	PET	55	2	2 ul, multiplexed	
82	FH2759	28	PET	55	2	2 ul, multiplexed	
83	FH2785	28	6-FAM	55	2	2 ul, multiplexed	
84	REN239K24	29	VIC	55	2	2 ul, multiplexed	
85	FH3082	29	NED	55	2	2 ul, multiplexed	
86	REN51C16	30	VIC	55	2	2 ul, multiplexed	
87	FH3053	30	NED	55	2	2 ul, multiplexed	
88	REN43H24	31	NED	55	2	2 ul, multiplexed	
89	FH2712	31	PET	55	2	2 ul, multiplexed	
90	FH2875	32	6-FAM	55	2	5–10 ul	
91	FH2790	33	VIC	55	2	2 ul, multiplexed	
92	REN291M20	33	NED	58	1.5	2 ul, multiplexed	
93	REN160M18	34	PET	55	2	2 ul, multiplexed	
94	FH3060	34	6-FAM	55	2	2 ul, multiplexed	
95	REN314H10	34	VIC	55	2	2 ul, multiplexed	
96	REN01G01	35	NED	55	2	2 ul, multiplexed	
97	REN112C08	35	PET	55	2	2 ul, multiplexed	
98	REN106I07	36	PET	55	2	5–10 ul	
99	FH2708	37	VIC	55	2	2 ul, multiplexed	
100	REN86G15	38	NED	55	2	2 ul, multiplexed	

*Fluorescent dye labeled primers are available through Applied Biosystems (ABI).
†Multiplexed markers were combined into groups of 2 to 4 different dyes colors post PCR.