

Genetic structure in village dogs reveals a Central Asian domestication origin

Laura M. Shannon^a, Ryan H. Boyko^b, Marta Castelhan^c, Elizabeth Corey^c, Jessica J. Hayward^a, Corin McLean^d, Michelle E. White^a, Mounir Abi Said^e, Baddley A. Anita^f, Nono Ikombe Bondjengo^g, Jorge Calero^h, Ana Galovⁱ, Marius Hedimbi^j, Bulu Imam^k, Rajashree Khalap^l, Douglas Lally^m, Andrew Mastaⁿ, Kyle C. Oliveira^a, Lucía Pérez^o, Julia Randall^p, Nguyen Minh Tam^q, Francisco J. Trujillo-Cornejo^r, Carlos Valeriano^h, Nathan B. Sutter^r, Rory J. Todhunter^c, Carlos D. Bustamante^s, and Adam R. Boyko^{a,1}

^aDepartment of Biomedical Sciences, Cornell University, Ithaca, NY 14853; ^bDepartment of Epidemiology of Microbial Diseases, Yale School of Public Health, Yale University, New Haven, CT 06510; ^cDepartment of Clinical Sciences, Cornell University, Ithaca, NY 14853; ^dBiogen Idec, Cambridge, MA 02142; ^eBiology Department, American University of Beirut, Beirut, Lebanon; ^fHoniara Veterinary Clinic and Surgery, Honiara, Solomon Islands; ^gDépartement de l'environnement, Faculté des Sciences, Université de Mbandaka, Mbandaka, Democratic Republic of Congo; ^hAcadémico de Arqueología, Universidad Nacional de San Antonio Abad del Cusco, Cusco, Peru; ⁱDepartment of Animal Physiology, University of Zagreb, Zagreb 10000, Croatia; ^jMicrobiology, University of Namibia, Windhoek, Namibia; ^kSanskriti Centre, Hazaribagh, Jharkhand, India 825 301; ^lThe INDog Project, Maharashtra, India; ^mThe Mongolian Bankhar Project, Ulaanbaatar, Mongolia; ⁿSchool of Medicine and Health Sciences, University of Papua New Guinea, Boroko, Port Moresby, National Capital District, 111, Papua New Guinea; ^oInstituto de Biología, Universidad Nacional Autónoma de México, Mexico City, Federal District, Mexico; ^pUniversity of Massachusetts Medical School, Worcester, MA 01655; ^qVietnam National Museum of Nature, Vietnam Academy of Science and Technology, Hanoi, Vietnam; ^rDepartment of Biology, La Sierra University, Riverside, CA 92505; and ^sDepartment of Genetics, Stanford University School of Medicine, Stanford, CA 94305

Edited by David M. Hillis, The University of Texas at Austin, Austin, TX, and approved September 11, 2015 (received for review August 19, 2015)

Dogs were the first domesticated species, originating at least 15,000 y ago from Eurasian gray wolves. Dogs today consist primarily of two specialized groups—a diverse set of nearly 400 pure breeds and a far more populous group of free-ranging animals adapted to a human commensal lifestyle (village dogs). Village dogs are more genetically diverse and geographically widespread than purebred dogs making them vital for unraveling dog population history. Using a semicustom 185,805-marker genotyping array, we conducted a large-scale survey of autosomal, mitochondrial, and Y chromosome diversity in 4,676 purebred dogs from 161 breeds and 549 village dogs from 38 countries. Geographic structure shows both isolation and gene flow have shaped genetic diversity in village dog populations. Some populations (notably those in the Neotropics and the South Pacific) are almost completely derived from European stock, whereas others are clearly admixed between indigenous and European dogs. Importantly, many populations—including those of Vietnam, India, and Egypt—show minimal evidence of European admixture. These populations exhibit a clear gradient of short-range linkage disequilibrium consistent with a Central Asian domestication origin.

admixture | domestication | linkage disequilibrium | introgression | haplotype diversity

The domestic dog, *Canis lupus familiaris*, is found living with and around humans throughout the globe. Selective breeding of dogs has been practiced for thousands of years, but the majority of modern breeds are less than 200 y old and of European ancestry (1, 2). Most dogs in the world are not purebred or even mixed-breed dogs, but rather belong to free-breeding human-commensal populations (“village dogs”) (1, 3, 4). The history and lineage of most modern breeds is well established (5, 6), but the genetic relationships among village dog populations and between village dogs and breeds is less understood.

Global surveys of mitochondrial and Y chromosome diversity in dogs have concluded that domestication occurred in southern China less than 16,500 yBP (7–10). In contrast, the earliest archeological evidence for dog-like canids occurs in Europe and Siberia, and Mt haplotypes found in ancient and modern gray wolves appear to be consistent with an origin of dogs from European wolves (11). These conflicting observations could be due to demographic processes after domestication (bottlenecks, migration, and admixture), altering patterns of genetic diversity or simply a consequence of a sparse archeological record in East Asia

during this period. Archeologists and geneticists agree that dogs evolved from Eurasian gray wolves at least 15,000 yBP (2), but precise determination of the domestication origin(s) is elusive.

Whereas the Y and Mt chromosomes are just two inherited loci, autosomal markers offer a vastly richer picture of the patterning of genetic variation genome-wide and better resolution for demographic inference. Efforts to identify the basis of phenotypic diversity and genetic diseases in dogs have yielded large genomic datasets of purebred dogs readily available for demographic inference (6, 12, 13). Genomic comparisons of purebred dogs and wolves show Middle Eastern wolves have more haplotype sharing with dogs than other wolf populations (6), but this is likely due to dog-wolf introgression in the Middle East (14) rather than an indication of Middle Eastern origins.

Inference of early population history using purebred dogs is hampered by the confounding effects of artificial selection and bottlenecks and by the relative dearth of breeds without European

Significance

Dogs were the first domesticated species, but the precise timing and location of domestication are hotly debated. Using genomic data from 5,392 dogs, including a global set of 549 village dogs, we find strong evidence that dogs were domesticated in Central Asia, perhaps near present-day Nepal and Mongolia. Dogs in nearby regions (e.g., East Asia, India, and Southwest Asia) contain high levels of genetic diversity due to their proximity to Central Asia and large population sizes. Indigenous dog populations in the Neotropics and South Pacific have been largely replaced by European dogs, whereas those in Africa show varying degrees of European vs. indigenous African ancestry.

Author contributions: R.H.B., J.C., A.G., D.L., N.B.S., R.J.T., C.D.B., and A.R.B. designed research; L.M.S., R.H.B., M.C., E.C., J.J.H., C.M., M.E.W., M.A.S., B.A.A., N.I.B., J.C., A.G., M.H., B.I., R.K., D.L., A.M., K.C.O., L.P., J.R., N.M.T., F.J.T.-C., C.V., N.B.S., R.J.T., and A.R.B. performed research; L.M.S. and A.R.B. analyzed data; and L.M.S. and A.R.B. wrote the paper.

Conflict of interest statement: A.R.B. and R.H.B. are cofounders and officers of Embark Veterinary, Inc., a canine genetics testing company.

This article is a PNAS Direct Submission.

Data deposition: Genotype and geographical data have been deposited in Dryad, datadryad.org (doi:10.5061/dryad.v9t5h).

¹To whom correspondence should be addressed. Email: boyko@cornell.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1516215112/-DCSupplemental.

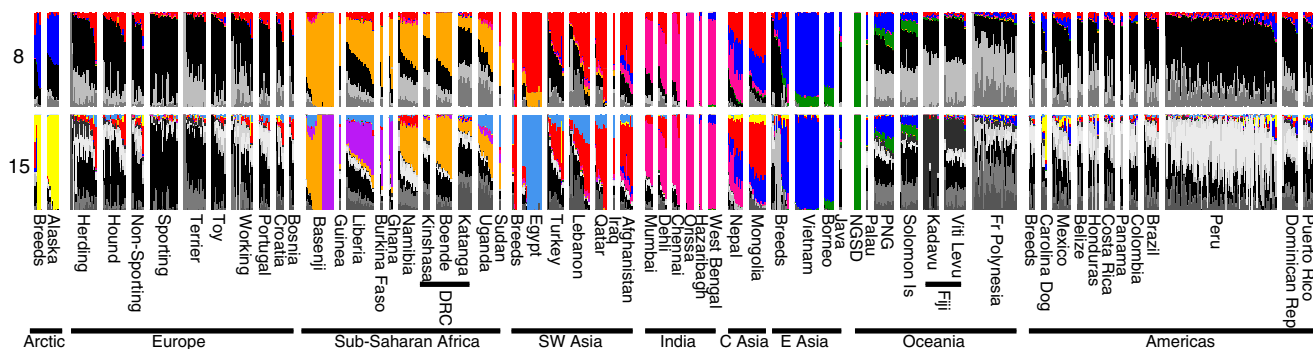


Fig. 2. ADMIXTURE analysis at $K = 8$ and 15 for unrelated village and breed dogs. European ancestry components are in grayscale ancestry. Vertical lines are individual dogs with one dog per breed (except for Basenjis, NGSDs, and Carolina Dogs).

admixture is strongest in Katanga and Kinshasa (86–88% and 74–75%, respectively). African dogs exhibit high Y and Mt haplotype diversity (Table 1), and haplotypes are geographically structured, including a high frequency Y haplogroup found only in Africa and Mongolia, and a cluster of closely related uniquely African Mt haplotypes (Fig. 3), demonstrating significant diversification of dogs within Africa (SI Appendix, Table S9).

In Eurasia, regional ancestry components predominate, with appreciable European ancestry occurring mainly in regions proximate to Europe (e.g., Turkey and Lebanon), consistent with gene flow between neighboring populations (SI Appendix, Fig. S17 and Table S10). Fine-scale population structure is also evident; for example, in Namibia one sample came from north of the Red Line veterinary cordon and clearly clusters with European rather than Namibian dogs, consistent with a previous study (4). In Lebanon and Egypt, dogs near Beirut and Cairo have more European ancestry than dogs from elsewhere. Similar patterns in Papua New Guinea (Port Moresby), Nepal (Kathmandu), and India (Mumbai) suggest that foreign dogs are more likely to be brought to (or survive in) urban areas than in more remote regions. Although some regions (e.g., Peru) were extensively sampled in remote areas, in other regions, sampling was largely limited to urban areas (e.g., Java and Turkey) or certain isolated populations (e.g., Roatán Island, Honduras), so caution is warranted in interpreting results from these areas.

Global Patterns of LD Indicate a Central Asian Origin for Dogs. The average rate of decay of LD between nearby autosomal markers is strongly influenced by population history. Four specific patterns are (i) populations with large N_e show faster decay rates (30), (ii) populations with recent admixture show greater LD than unadmixed populations, particularly at large inter-SNP distances (31), (iii) bottlenecked populations show elevated LD (32), and (iv) expanding populations have a greater slope of LD decay versus stable or contracting populations as long-range LD is mainly determined by recent recombination events but short-range LD is affected by more ancient events (33).

As expected, village dog populations show elevated short-range LD compared with gray wolves, reflecting a bottleneck during domestication (Fig. 5 and SI Appendix). The rate of LD decay is greater in dogs than wolves, reflecting larger N_e post-domestication for dogs vs. wolves and leading to similarly low LD in dogs and wolves at large inter-SNP distances (> 0.05 cM). Across village dog populations, LD is lowest in Asia. Specifically, LD is lowest in Afghanistan and Central Asia at short inter-SNP distances (< 0.0005 cM) and lowest in Vietnam at intermediate distances (0.01–0.05 cM), with rates increasing in other populations depending on their isolation and distance from Asia. These patterns of LD decay strongly suggest a Central Asian origin for domestic dogs with a subsequent population expansion

(larger contemporary N_e) in East Asia and elsewhere. These patterns are consistent if physical, rather than genetic, inter-SNP distance is measured, or if different subsets of dogs are used for each population (SI Appendix, Fig. S18).

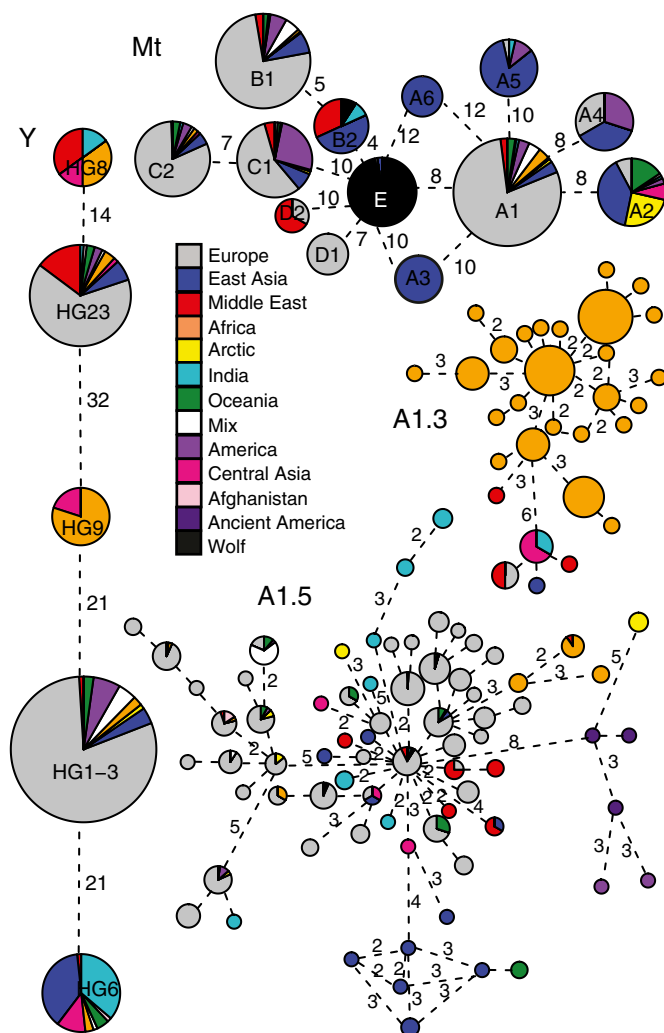


Fig. 3. Minimum spanning networks of Y and Mt haplogroups. Haplotypes within A1.3 and A1.5 Mt haplogroups also shown. Within each network, circle size is proportional to haplotype/group frequency and line length is determined by number of mutations separating haplotypes (indicated by numerals when greater than one mutation).

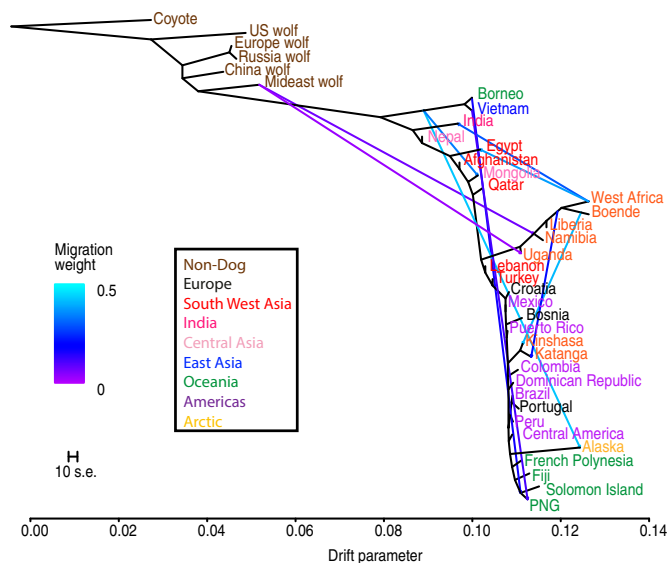


Fig. 4. Maximum likelihood tree of village dogs and select wolves with 10 migration edges. Edge color shows the proportion of source population ancestry found in the sink population. Populations are colored by region (outgroups are brown).

Discussion

This study represents the largest-ever survey of worldwide canine genetic diversity using nuclear, Y, and Mt markers. We confirm high diversity and low LD in village compared with purebred dogs (4, 12, 19) and show how village dog populations improve inference of dog evolutionary history. This increased geographic and genetic resolution reveals the effects of bottlenecks and admixture in extant populations, as well as evidence for an origin of dogs in Central Asia.

Like previous studies, we find high levels of Mt and Y haplotype diversity in East Asia (8–10, 29, 34), but we also find high levels of Mt and Y diversity in India and Southwest Asia, respectively (Table 1). Whereas previous studies have used the high levels of uni-parentally inherited haplotype diversity as evidence for an East Asian, specifically Southern Chinese, origin for dogs, genome-wide LD patterns among populations suggests a different process. Namely, domestication occurred in Central Asia where early dogs carrying nearly the full complement of Mt and Y haplotypes spread to nearby Asian regions, including Afghanistan, India, and Vietnam. The substantial N_e in these regions, particularly East Asia, allowed these haplogroups to survive and diversify to a greater extent than in Central Asia. Higher N_e in East vs. Central Asia is supported both by census estimates (35) and by the more negative slope of the LD decay curve in East vs. Central Asia (Fig. 5), because recent population history has a greater impact on long-range vs. short-range LD (30, 33).

Gray wolves were clearly present in Central Asia during the Mesolithic, and both wolves and human hunter-gatherers were exploiting large mammals during this time (36). Increasing human population density, blade and hunting technology, and/or climate change during the Late Paleolithic in Central Asia (28) may have altered prey densities and made scavenging crucial to the survival of some wolf populations. Adaptations to scavenging such as tameness, small body size, and a decreased age of reproduction would reduce hunting efficiency further, eventually leading to obligate scavenging (37). Whether these earliest dogs were simply human-commensal scavengers or they played some role as companions or hunters that hastened their spread is uncertain, but clearly adaptation to conditions outside this initial

domestication origin [e.g., efficient starch digestion (38) and aseasonal breeding (39, 40)] has also been important in dog evolution.

Although SNP array data are poorly suited for estimating the timing of ancient population events, it does shed light on the conflicting estimates of dog origins in previous genetic studies. Because there is incomplete lineage sorting between dogs and wolves, estimates based on Mt or Y haplotype diversity are sensitive to assumptions regarding the number of founder haplotypes in early dogs (8, 41). Nuclear datasets offer better resolution for parameterizing demographic models, but two such studies have yielded widely varying results [14 vs. 32 kya (14, 42)]. Our LD data support a relatively strong domestication bottleneck in dogs followed by substantial population expansion, particularly in East Asia. An ancient origin with a weak domestication bottleneck and small current N_e in Asian village dogs is also consistent with the allele frequency data in ref. 42, but a more recent, stronger domestication bottleneck, and large current N_e could be consistent with both allele frequency data and LD decay rates, bringing the inferred timing of dog origins more in line with archeological estimates.

Central Asia has been considered a likely domestication origin for dogs by some archeologists (43), but it has been poorly represented in previous genetic studies of dog origins. The pattern of reduced short-range LD in populations near Central Asia is most parsimoniously explained by an origin of dogs somewhere in this region, but we cannot rule out the possibility that dogs were domesticated elsewhere and subsequently, either through migration or a separate domestication event, arrived and diversified in Central Asia. For example, European dog populations have undergone extensive turnover over the last 15,000 y (44), erasing the genomic signatures of early European population history. Although it is difficult to explain the clear gradient of short-range LD out of Central Asia if dogs were domesticated from a far-flung region, studies of extant dogs cannot exclude the possibility of earlier domestication events that subsequently died out, or were overwhelmed by more modern populations.

Further analysis of diverse dogs throughout Central Asia and surrounding regions is crucial for precisely resolving the origin and history of early dogs. Refining the timing of dog domestication

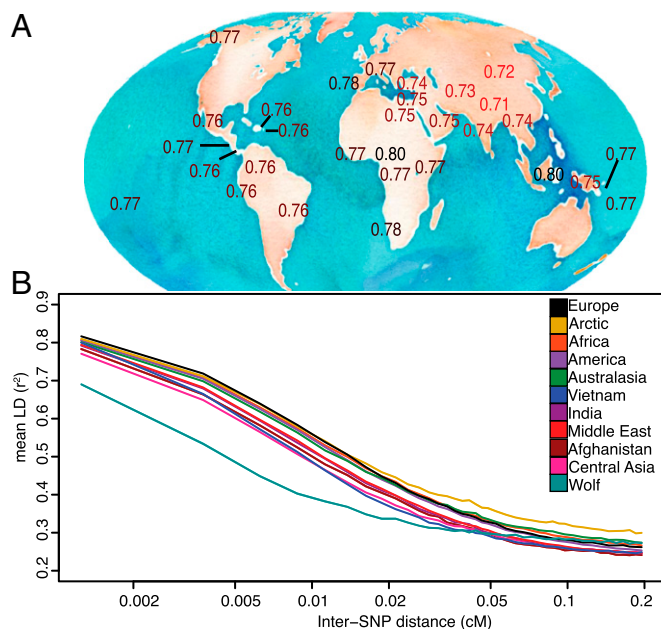


Fig. 5. LD decay for village dog populations worldwide. (A) Mean LD for village dog populations at short inter-SNP distances (<0.005 cM). (B) LD decay curves for select regions calculated from $N = 6$ dogs per region (100 replicates).

could yield substantial insights into the process by which dogs became domesticated and the wolf and human population(s) involved. Ancient DNA analysis will surely contribute to our understanding of early dog populations, but where ancient specimens are unavailable, village dogs are often the best proxy we have to ancient populations. Many indigenous populations have already succumbed to swamping gene flow from foreign dogs, so further work characterizing remaining indigenous populations genetically, morphologically, and behaviorally, is vital for building an improved understanding of dog evolutionary history.

Methods

Sample Collection. The majority of samples used in this study come from blood stored in the Cornell Veterinary Biobank collected in accordance with Cornell animal care protocols 2005-0151 and 2011-0061. These samples include 4,676 purebred dogs from 161 breeds, 167 mixed breed dogs, and 549 village dogs from 38 different countries (*SI Appendix, Table S11*). Blood was stored in EDTA, and DNA was extracted by salt precipitation.

Genotyping. Samples were genotyped on a semicustom Illumina SNP array containing 173,662 SNPs from the CanineHD array (13) and 12,143 markers identified using whole genome sequencing (45). A total of 166,171 markers remained after filtering markers with >5% missing data, discordant genotypes between technical replicates, or extreme divergence from Hardy-Weinberg expectations (HWEs; observed heterozygosity vs. HWE ratio <0.25 or >1.0). Genotype and geographical data have been deposited in Dryad (datadryad.org, doi:10.5061/dryad.v9t5h).

PCA. PCA of unrelated village dogs was run using the smartpca program distributed in the Eigenstrat v5.0.1 software package (46). Village dogs were used to define the PCA space, and Basenjis, Carolina Dogs, New Guinea Singing Dogs (NGSDs), and one dog each from other breeds were projected onto it.

Haplotype Analysis. The array included 582 Mt markers, of which 367 were polymorphic and passed quality control filtering. Additionally, seven markers that introduced multiple cycles in the haplotype network suggesting genotyping error were removed. We added 431 additional dogs with published complete Mt sequences (8, 11, 23–27) based on their genotypes at the marker positions. Haplotypes were named according to published convention (8), with some published haplotypes mapping to multiple haplotypes in this study due to the markers we used outside the control region. These haplotypes were split and are indicated with a letter. Conversely, some of the sequenced haplotypes are identical across all 360 marker positions on the array and are included as a single combined haplotype (e.g., C1_2).

The array included 336 Y chromosome markers, of which 207 were polymorphic and passed quality control filtering. One of these was removed because it introduced several cycles in the haplotype network. Haplotypes were named to correspond with Ding et al. (10), subdividing haplotypes to account for our enriched marker set.

Haplotype networks were constructed in R v3.1.0 using the Ape and Pegas packages (47–49). The distance matrix was calculated based on the count of differences, and then a minimum spanning forest was calculated. Networks were visualized in R using the igraph package (50). We defined haplogroups as groups of haplotypes at least 2 SDs further apart than the average distance between haplotypes, as measured by number of differences at the array marker positions.

Regional haplotype diversity (H) was computed with regions defined by geography and PCA. To control for sample size differences, we subsampled

($N = 20$) dogs 100 times within regions and counted the number of observed haplotypes.

LD Decay. LD is a reflection of N_e , with LD at proximate SNPs reflecting historic N_e and LD at distant SNPs reflecting N_e in more recent times (30, 33). To ensure estimates of LD were not biased by particular individuals or by the choice of sample size, we used the PLINK 1.0.7 (51) `--genome` command to remove related (π outliers) and the `--het` command to remove inbred ($F > 0.25$) individuals. We then performed two parallel analyses, one retaining 6 individuals per population and one retaining 20, randomly selecting the individuals 100 times to compute means and SEs. LD was calculated using the command `--maf 0.3 --r2 --ld-window 999 --ld-window-r2 0 --ld-window-kb 200`, and averaging within bins based on inter-SNP distance was performed using a C script (12).

Admixture. Ancestry of individual dogs was determined using ADMIXTURE software (52). For a global view of village dog ancestry, we included all unrelated individuals from NGSDs, Basenjis, Carolina dogs, and village dogs, and a single individual from select dog breeds. For each K , 10 replicates were run using a different random seed. The replicate with the lowest cross-validation score for each K is reported.

To estimate historical relationships between populations we used TreeMix (53). We built admixture trees for dog breeds with gray wolves as the root, and for village dogs with coyotes as the root. For the village dog tree, we combined our data with previously published wolf and coyote Affymetrix v2 data (13). Only SNPs genotyped on both arrays and passing quality control were included (36,358 total). Trees were calculated using a range of numbers of migration events (m); we report the trees where further migration edges do not appreciably improve the fit. With the same populations, we calculated pairwise F_{st} values using a custom C script.

We formally tested for admixture (indigenous vs. European) for African and Pacific Island dogs by computing f_3 statistics using the Admixtools package (54) for each population with Europe as one source population and Basenji, Vietnam, or Borneo as the other. Populations were the same as for the village dog TreeMix analysis, and wolves were the outgroup for testing the bounds of the admixture percentage.

For American dogs, a suitable unadmixed source population was not available, so we used a PCA-based approach (55) to identify the extent of European ancestry in individual breed and village dogs. Village dogs from Europe, Alaska, and Vietnam were used to define the PCA space, and then dogs from the Americas were projected onto it. The same approach was used to investigate indigenous versus European ancestry proportions from East Asian breeds using village dogs from Europe, Borneo, Vietnam, Mongolia, and Vietnam to define the principal components.

ACKNOWLEDGMENTS. We thank the countless dog owners and enthusiasts who facilitated sample collecting, including Carol Beuchat, Laura Colín, Jon Curby, Gautam Das, Ricardo de Matos, Baird Fleming, George Hicks, Gary S. Johnson, Warren Johnson, Janice Koler-Matznick, Leonard Kuwale, Kris Kvam, Judith Liggio, Stephanie Little Wolf, Gaby Matshimba, Mark Neff, Casey Quimby, Sue Ann Sandusky, Myrna Shibolet, Jo Thompson, Steve Wooten, Asociación de Amigos por los Animales de Sosúa-Judy's Pet Lodge, Animal Care in Egypt, Animals Fiji, Beirut for the Ethical Treatment of Animals, Liberia Animal Welfare and Conservation Society, Mongolian Bankhar Dog Project, Gump South Pacific Research Station, and the Qatar Animal Welfare Society. We thank Joy Li (Cornell Veterinary Biobank) and the Cornell University Genomics Core Facility for technical help. We thank A. G. Clark, S. Gravel, S. M. Myles, and R. K. Wayne for critical input. Funding for this project came from National Science Foundation Grant 0516310, National Geographic Society Expedition Council Grants EC0492-11 and 1P-14, Zoetis Animal Health, Cornell University Center for Advanced Technology, Cornell University, and dozens of PetriDish donors, including Sandra Coliver, Buck Farmer, Richard Gardner, Kathryn Sikkink, and Elaine and Chris McLeod.

- Parker HG, et al. (2004) Genetic structure of the purebred domestic dog. *Science* 304(5674):1160–1164.
- Larson G, et al. (2012) Rethinking dog domestication by integrating genetics, archeology, and biogeography. *Proc Natl Acad Sci USA* 109(23):8878–8883.
- Coppinger R, Coppinger L (2001) *Dogs: A Startling New Understanding of Canine Origin, Behavior & Evolution* (Simon and Schuster, New York).
- Boyko AR, et al. (2009) Complex population structure in African village dogs and its implications for inferring dog domestication history. *Proc Natl Acad Sci USA* 106(33):13903–13908.
- Clubb AK (2006) *The Complete Dog Book* (Ballantine Books, New York), 20th Ed.
- Vonholdt BM, et al. (2010) Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature* 464(7290):898–902.
- Savolainen P, Zhang YP, Luo J, Lundeberg J, Leitner T (2002) Genetic evidence for an East Asian origin of domestic dogs. *Science* 298(5598):1610–1613.
- Pang JF, et al. (2009) mtDNA data indicate a single origin for dogs south of Yangtze River, less than 16,300 years ago, from numerous wolves. *Mol Biol Evol* 26(12):2849–2864.
- Brown SK, et al. (2011) Phylogenetic distinctiveness of Middle Eastern and Southeast Asian village dog Y chromosomes illuminates dog origins. *PLoS One* 6(12):e28496.
- Ding ZL, et al. (2012) Origins of domestic dog in southern East Asia is supported by analysis of Y-chromosome DNA. *Heredity (Edinb)* 108(5):507–514.
- Thalmann O, et al. (2013) Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs. *Science* 342(6160):871–874.
- Boyko AR, et al. (2010) A simple genetic architecture underlies morphological variation in dogs. *PLoS Biol* 8(8):e1000451.
- Vaysse A, et al.; LUPA Consortium (2011) Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping. *PLoS Genet* 7(10):e1002316.

14. Freedman AH, et al. (2014) Genome sequencing highlights the dynamic early history of dogs. *PLoS Genet* 10(1):e1004016.
15. Reich DE, et al. (2001) Linkage disequilibrium in the human genome. *Nature* 411(6834):199–204.
16. Lindblad-Toh K, et al. (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438(7069):803–819.
17. Gray MM, et al. (2009) Linkage disequilibrium and demographic history of wild and domestic canids. *Genetics* 181(4):1493–1505.
18. Axelsson E, Webster MT, Ratnakumar A, Ponting CP, Lindblad-Toh K; LUPA Consortium (2012) Death of PRDM9 coincides with stabilization of the recombination landscape in the dog genome. *Genome Res* 22(1):51–63.
19. Irion DN, Schaffer AL, Grant S, Wilton AN, Pedersen NC (2005) Genetic variation analysis of the Bali street dog using microsatellites. *BMC Genet* 6:6.
20. Pedersen N, Liu H, Theilen G, Sacks B (2013) The effects of dog breed development on genetic diversity and the relative influences of performance and conformation breeding. *J Anim Breed Genet* 130(3):236–248.
21. Boyko AR (2011) The domestic dog: Man's best friend in the genomic era. *Genome Biol* 12(2):216.
22. Witt KE, et al. (2015) DNA analysis of ancient dogs of the Americas: Identifying possible founding haplotypes and reconstructing population histories. *J Hum Evol* 79:105–118.
23. Kim KS, Lee SE, Jeong HW, Ha JH (1998) The complete nucleotide sequence of the domestic dog (*Canis familiaris*) mitochondrial genome. *Mol Phylogenet Evol* 10(2):210–220.
24. Björnerfeldt S, Webster MT, Vilà C (2006) Relaxation of selective constraint on dog mitochondrial DNA following domestication. *Genome Res* 16(8):990–994.
25. Baranowska I, et al. (2009) Sensory ataxic neuropathy in golden retriever dogs is caused by a deletion in the mitochondrial tRNATyr gene. *PLoS Genet* 5(5):e1000499.
26. Webb KM, Allard MW (2009) Mitochondrial genome DNA analysis of the domestic dog: Identifying informative SNPs outside of the control region. *J Forensic Sci* 54(2):275–288.
27. Matsumura S, Inoshima Y, Ishiguro N (2014) Reconstructing the colonization history of lost wolf lineages by the analysis of the mitochondrial genome. *Mol Phylogenet Evol* 80:105–112.
28. Unesco (1992) *History of Civilizations of Central Asia* (UNESCO Publishing, Paris), Vol 1.
29. Sacks BN, et al. (2013) Y chromosome analysis of dingoes and southeast asian village dogs suggests a neolithic continental expansion from Southeast Asia followed by multiple Austronesian dispersals. *Mol Biol Evol* 30(5):1103–1118.
30. Hill WG (1981) Estimation of effective population size from data on linkage disequilibrium. *Genet Res* 38(03):209–216.
31. Pritchard JK, Rosenberg NA (1999) Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Genet* 65(1):220–228.
32. Slatkin M (2008) Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. *Nat Rev Genet* 9(6):477–485.
33. Hayes BJ, Visscher PM, McPartlan HC, Goddard ME (2003) Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Res* 13(4):635–643.
34. Ardalan A, et al. (2011) Comprehensive study of mtDNA among Southwest Asian dogs contradicts independent domestication of wolf, but implies dog-wolf hybridization. *Ecol Evol* 1(3):373–385.
35. Dong W, et al. (2014) The dog-human-wildlife interface: Assessing the scope of the problem. *Free-Ranging Dogs and Wildlife Conservation*, ed Gompper ME (Oxford Univ Press, Oxford, UK), pp 9–45.
36. Dong W, et al. (2014) Late Pleistocene mammalian fauna from Wulanmulan Paleolithic Site, Nei Mongol, China. *Quat Int* 347(9):139–147.
37. Coppinger R, Feinstein M (2015) *How Dogs Work* (Univ of Chicago Press, Chicago).
38. Axelsson E, et al. (2013) The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* 495(7441):360–364.
39. Driscoll CA, Macdonald DW, O'Brien SJ (2009) From wild animals to domestic pets, an evolutionary view of domestication. *Proc Natl Acad Sci USA* 106(Suppl 1):9971–9978.
40. Lord K, Feinstein M, Smith B, Coppinger R (2013) Variation in reproductive traits of members of the genus *Canis* with special attention to the domestic dog (*Canis familiaris*). *Behav Processes* 92:131–142.
41. Vilà C, Seddon J, Ellegren H (2005) Genes of domestic mammals augmented by backcrossing with wild ancestors. *Trends Genet* 21(4):214–218.
42. Wang GD, et al. (2013) The genomics of selection in dogs and the parallel evolution between dogs and humans. *Nat Commun* 4:1860.
43. Derr M (2013) *How the Dog Became the Dog: From Wolves to Our Best Friends* (Overlook Books, New York).
44. Malmström H, et al. (2008) Barking up the wrong tree: Modern northern European dogs fail to explain their origin. *BMC Evol Biol* 8:71.
45. Auton A, et al. (2013) Genetic recombination is targeted towards gene promoter regions in dogs. *PLoS Genet* 9(12):e1003984.
46. Price AL, et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38(8):904–909.
47. Paradis E, Claude J, Strimmer K (2004) APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* 20(2):289–290.
48. Paradis E (2010) pegas: An R package for population genetics with an integrated-modular approach. *Bioinformatics* 26(3):419–420.
49. R Core Team (2014) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna).
50. Csardi G, Nepusz T (2006) The igraph software package for complex network research, InterJournal, Complex Systems 1695. Available at igraph.org.
51. Purcell S, et al. (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81(3):559–575.
52. Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19(9):1655–1664.
53. Pickrell JK, Pritchard JK (2012) Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet* 8(11):e1002967.
54. Patterson N, et al. (2012) Ancient admixture in human history. *Genetics* 192(3):1065–1093.
55. Bryc K, et al. (2010) Genome-wide patterns of population structure and admixture in West Africans and African Americans. *Proc Natl Acad Sci USA* 107(2):786–791.

Patterns of gene flow in Eurasia do not explain LD decay in Central Asia

Gene flow clearly occurs in dogs, with population structure exhibiting isolation by distance everywhere, and admixture with European dogs in particular regions. Using f3 admixture statistics, we find that indeed all Old World populations geographically intermediate between two other populations appear to be genetically intermediate as well, consistent with this model of bidirectional gene flow (Table S1).

As Central Asia is a population intermediate between regions previously implicated in genetic theories of dog origins (East Asia and Middle East / Europe), we used ms simulations¹ (Table S2) to examine the effect of gene flow on central and peripheral populations to ensure that the intermediate position of Central Asia within Eurasia was not driving a spurious signature of a domestication origin in the LD decay data. We explored two alternate models (Fig S1). In the first, dogs arise in a single population; this population splits as dogs move westward (serial founder model). The non-founder populations each experience a bottleneck as they split off. Gene flow then occurs between the centrally located population (Population 2) and both the founder population (Population 1) and the most diverged population (Population 3). This is the pattern we would expect in Eurasia if dogs were from South East Asia and then had spread west (or were from Europe or the Middle East, and spread east). In the second model, dogs originate in a central population, two populations branch off from the central populations and after going through a bottleneck both the Eastern and Western population experience gene flow with the central population. This is the model we would expect if dogs originated in Central Asia and spread in both directions.

Under various levels of migration, in the first model Population 2 never exhibited lower LD than the founder population (Fig S2). As migration rates increased populations became more similar in LD curve. For model 2 the founding population was also the population with lowest LD. This is consistent with the interpretation that dogs arose in the Central Asia and spread outward, but not consistent with dogs arising in either East Asia or Europe / Middle East.

More complicated patterns of admixture may influence patterns of LD in more complex ways. Perhaps the most likely way for LD to be reduced in a region (apart from being a domestication origin) would be substantial ancient admixture from an archaic population. While there is evidence of localized admixture with wolves in our dataset (Fig S3) and previously published datasets [14], we did not see any D tests where Central Asian dogs were outliers, indicating outsized gene flow from another population. East Asian dogs showed clear evidence of admixture with Chinese wolves, and all other dog populations showed evidence of admixture with Middle Eastern wolves.

We note that Freedman et al 2014 demonstrated that admixture with Middle Eastern wolves was bidirectional, although TreeMix only identified admixture from Middle Eastern wolves into dogs in our dataset (Fig 4). We believe the lack of a reciprocal admixture edge biases the TreeMix topology, artificially putting Middle Eastern wolves as the nearest neighbor to dogs. Other TreeMix models run with different numbers of admixture edges are inconsistent about the grouping of Middle Eastern wolves with dogs. In some models, Chinese wolves are closest (but no admixture edge is seen between Chinese wolves and dogs despite highly significant D-tests supporting such admixture) and in others wolves are monophyletic. Thus, while we believe admixture edges shown in Figure 4 reflect real genetic patterns, some admixture events are not robustly detected by TreeMix, leading to inconsistent topologies at the most basal branches of dogs and wolves.

Table S1: Significant F3 tests. We tested all combinations of village dog populations from: Afghanistan, Africa, Europe, India, Middle East, Mongolia, Nepal, and Vietnam and we report the significant results.

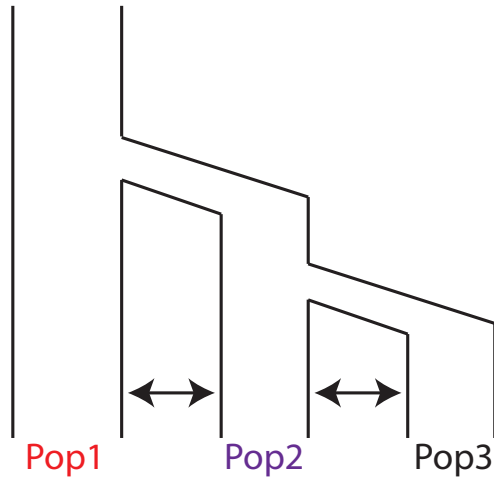
Source1	Source2	Target	f3	Z	Lower	Upper	p-value
Europe	India	Middle East	-0.00919	-21.63	0.608	0.818	4.69E-104
Europe	Vietnam	Mongolia	-0.0160	-21.062	0.489	0.818	8.88E-99
Middle East	Vietnam	Mongolia	-0.0130	-20.119	0.623	0.772	2.52E-90
Vietnam	India	Nepal	-0.0124	-16.499	0.257	0.418	1.87E-61
Vietnam	Africa	Nepal	-0.0133	-16.192	0.303	0.58	2.87E-59
Middle East	Vietnam	Nepal	-0.0115	-15.497	0.531	0.767	1.82E-54
Vietnam	Africa	Mongolia	-0.0111	-13.385	0.299	0.553	3.70E-41
Europe	India	Afghanistan	-0.00849	-13.07	0.399	0.665	2.44E-39
Vietnam	Afghanistan	Mongolia	-0.00780	-11.134	0.191	0.273	4.29E-29
Europe	Vietnam	Nepal	-0.0102	-10.896	0.374	0.818	6.02E-28
Vietnam	Afghanistan	Nepal	-0.00787	-10.392	0.196	0.362	1.35E-25
Europe	Vietnam	Afghanistan	-0.00607	-7.276	0.614	0.872	1.72E-13
Middle East	Vietnam	Afghanistan	-0.00441	-6.266	0.812	0.955	1.85E-10
Europe	Nepal	Afghanistan	-0.00376	-6.046	0.383	0.711	7.42E-10
Europe	Nepal	Middle East	-0.00264	-6.035	0.598	0.856	7.95E-10
India	Mongolia	Nepal	-0.00310	-5.439	0.254	0.59	2.68E-08
Europe	Africa	Middle East	-0.00135	-4.247	0.26	0.558	1.08E-05
Africa	Mongolia	Middle East	-0.00145	-4.081	0.542	0.072	2.24E-05
Vietnam	Africa	Afghanistan	-0.00322	-3.665	0.133	0.403	0.000124
Europe	Vietnam	Middle East	-0.00142	-2.231	0.749	0.927	0.0128

¹Hudson, R. R.(2002) Generating samples under a Wright-Fisher neutral model. Bioinformatics 18:337-8

Table S2: Code for reported ms simulations

Migration Rate	Model 1	Model 2
2.5×10^{-6}	ms 36 100 -t 80 -r 80 100000 -l 3 12 12 12 -n 1 1 -n 2 1 -n 3 1 -m 2 1 0.2 -m 2 3 0.2 -m 1 2 0.2 -m 3 2 0.2 -eM 0.01 0 -en 0.025 3 0.01 -ej 0.0251 3 2 -en 0.0495 2 0.01 -ej 0.05 2 1	ms 36 100 -t 80 -r 80 100000 -l 3 12 12 12 -n 1 1 -n 2 1 -n 3 1 -m 2 1 0.2 -m 1 3 0.2 -m 1 2 0.2 -m 3 1 0.2 -eM 0.04 0 -en 0.049 3 0.01 -en 0.0495 2 0.01 -ej 0.0497 3 1 -ej 0.05 2 1
1.25×10^{-5}	ms 36 100 -t 80 -r 80 100000 -l 3 12 12 12 -n 1 1 -n 2 1 -n 3 1 -m 2 1 1 -m 2 3 1 -m 1 2 1 -m 3 2 1 -eM 0.01 0 -en 0.025 3 0.01 -ej 0.0251 3 2 -en 0.0495 2 0.01 -ej 0.05 2 1	ms 36 100 -t 80 -r 80 100000 -l 3 12 12 12 -n 1 1 -n 2 1 -n 3 1 -m 2 1 1 -m 1 3 1 -m 1 2 1 -m 3 1 1 -eM 0.04 0 -en 0.049 3 0.01 -en 0.0495 2 0.01 -ej 0.0497 3 1 -ej 0.05 2 1
2.5×10^{-5}	ms 36 100 -t 80 -r 80 100000 -l 3 12 12 12 -n 1 1 -n 2 1 -n 3 1 -m 2 1 2 -m 2 3 2 -m 1 2 2 -m 3 2 2 -eM 0.01 0 -en 0.025 3 0.01 -ej 0.0251 3 2 -en 0.0495 2 0.01 -ej 0.05 2 1	ms 36 100 -t 80 -r 80 100000 -l 3 12 12 12 -n 1 1 -n 2 1 -n 3 1 -m 2 1 2 -m 1 3 2 -m 1 2 2 -m 3 1 2 -eM 0.04 0 -en 0.049 3 0.01 -en 0.0495 2 0.01 -ej 0.0497 3 1 -ej 0.05 2 1
1×10^{-4}	ms 36 100 -t 80 -r 80 100000 -l 3 12 12 12 -n 1 1 -n 2 1 -n 3 1 -m 2 1 8 -m 2 3 8 -m 1 2 8 -m 3 2 8 -eM 0.01 0 -en 0.025 3 0.01 -ej 0.0251 3 2 -en 0.0495 2 0.01 -ej 0.05 2 1	ms 36 100 -t 80 -r 80 100000 -l 3 12 12 12 -n 1 1 -n 2 1 -n 3 1 -m 2 1 8 -m 1 3 8 -m 1 2 8 -m 3 1 8 -eM 0.04 0 -en 0.049 3 0.01 -en 0.0495 2 0.01 -ej 0.0497 3 1 -ej 0.05 2 1
0.001	ms 36 100 -t 80 -r 80 100000 -l 3 12 12 12 -n 1 1 -n 2 1 -n 3 1 -m 2 1 80 -m 2 3 80 -m 1 2 80 -m 3 2 80 -eM 0.01 0 -en 0.025 3 0.01 -ej 0.0251 3 2 -en 0.0495 2 0.01 -ej 0.05 2 1	ms 36 100 -t 80 -r 80 100000 -l 3 12 12 12 -n 1 1 -n 2 1 -n 3 1 -m 2 1 80 -m 1 3 80 -m 1 2 80 -m 3 1 80 -eM 0.04 0 -en 0.049 3 0.01 -en 0.0495 2 0.01 -ej 0.0497 3 1 -ej 0.05 2 1
0.05	ms 36 100 -t 80 -r 80 100000 -l 3 12 12 12 -n 1 1 -n 2 1 -n 3 1 -m 2 1 4000 -m 2 3 4000 -m 1 2 4000 -m 3 2 4000 -eM 0.01 0 -en 0.025 3 0.01 -ej 0.0251 3 2 -en 0.0495 2 0.01 -ej 0.05 2 1	ms 36 100 -t 80 -r 80 100000 -l 3 12 12 12 -n 1 1 -n 2 1 -n 3 1 -m 2 1 4000 -m 1 3 4000 -m 1 2 4000 -m 3 1 4000 -eM 0.04 0 -en 0.049 3 0.01 -en 0.0495 2 0.01 -ej 0.0497 3 1 -ej 0.05 2 1

Model 1- East Asian Origin



Model 2- Central Asian Origin

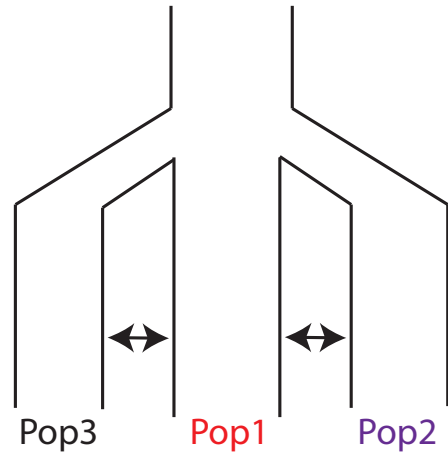


Fig. S1: Models for MS simulations.

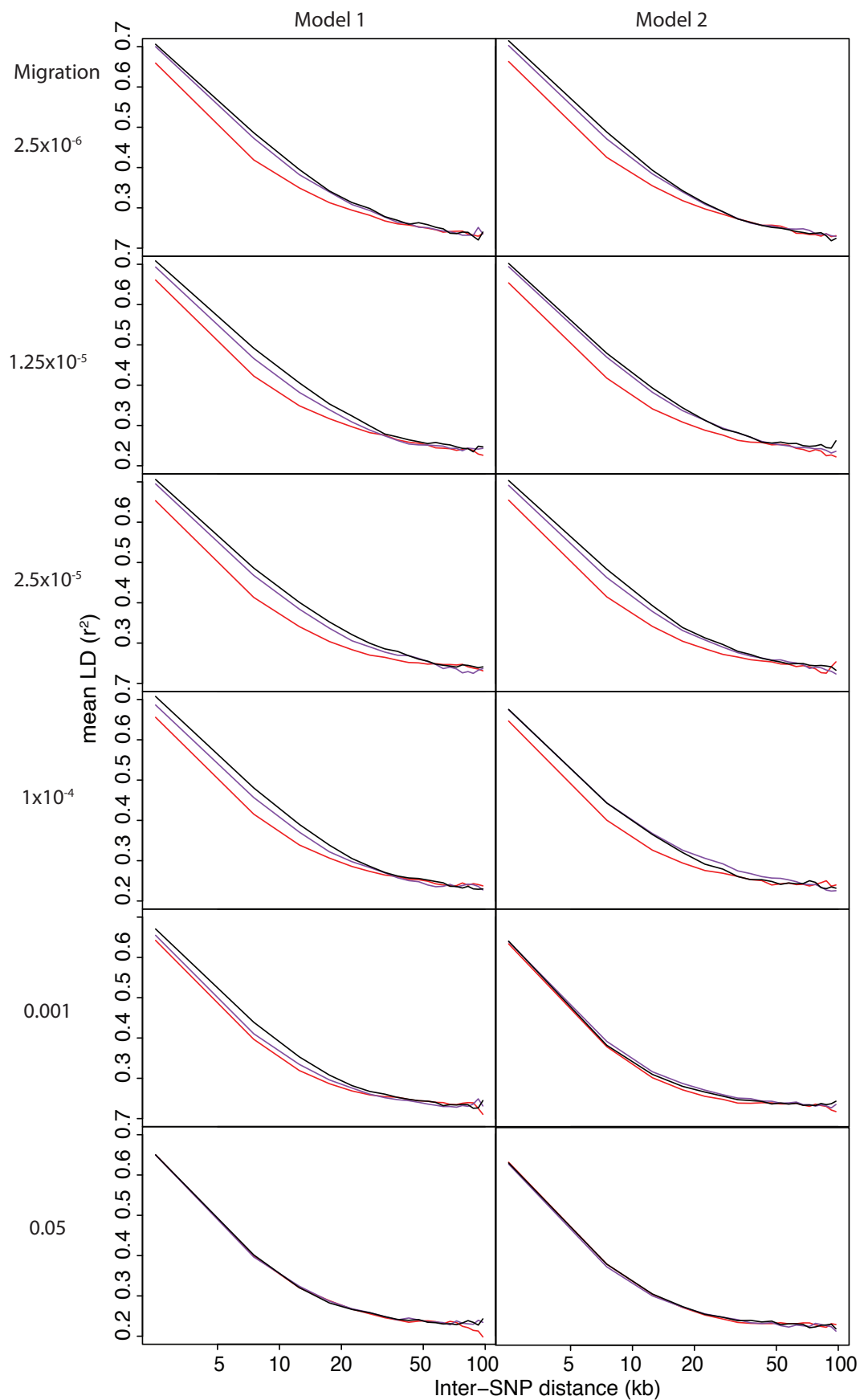


Fig. S2: LD curves for two contrasting models of dog history featuring various migration rates (percentage of sink population made up of source population per generation). In all cases the red population is the founding population.

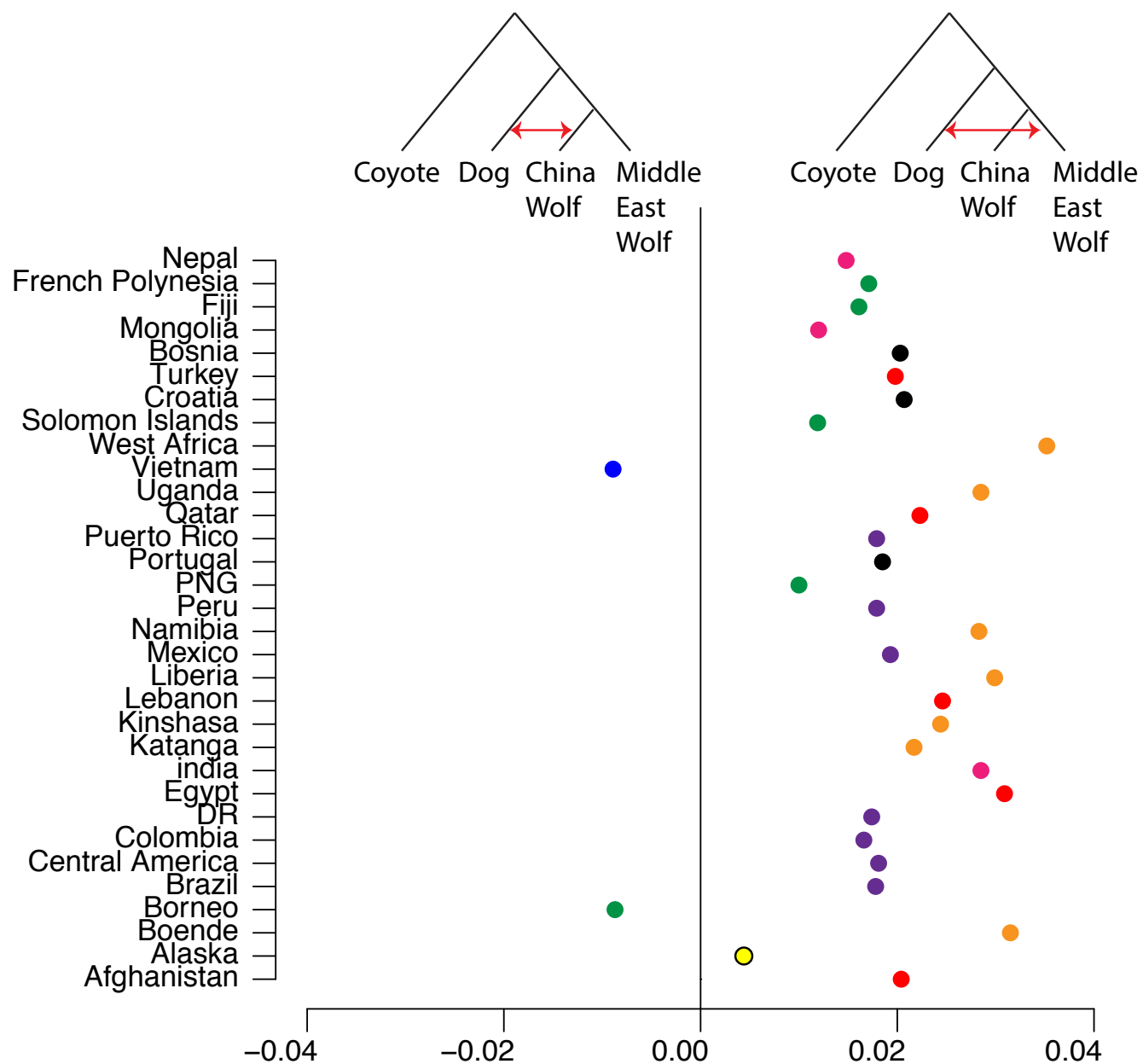


Fig. S3: D tests for all village dog populations in relation to Middle Eastern and Chinese wolves. Points are colored by region.

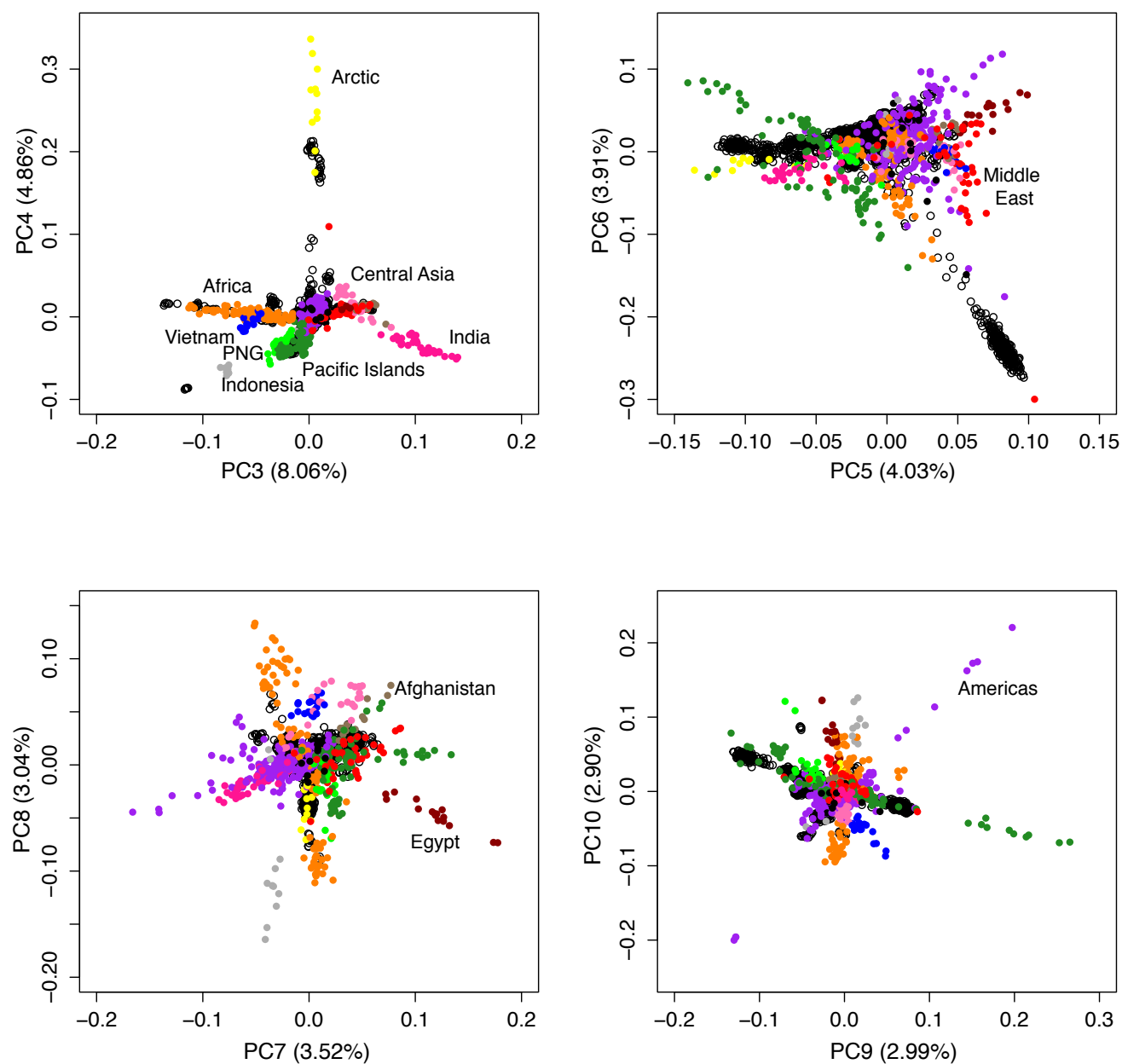
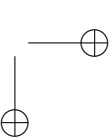
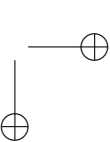
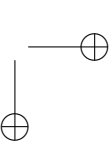


Fig. S4: Principal component analysis for village dogs (PC3–PC10). Village dogs are represented by colored dots while breed dogs, which have been projected onto the PCA space, are represented by open black circles.



Footline Author





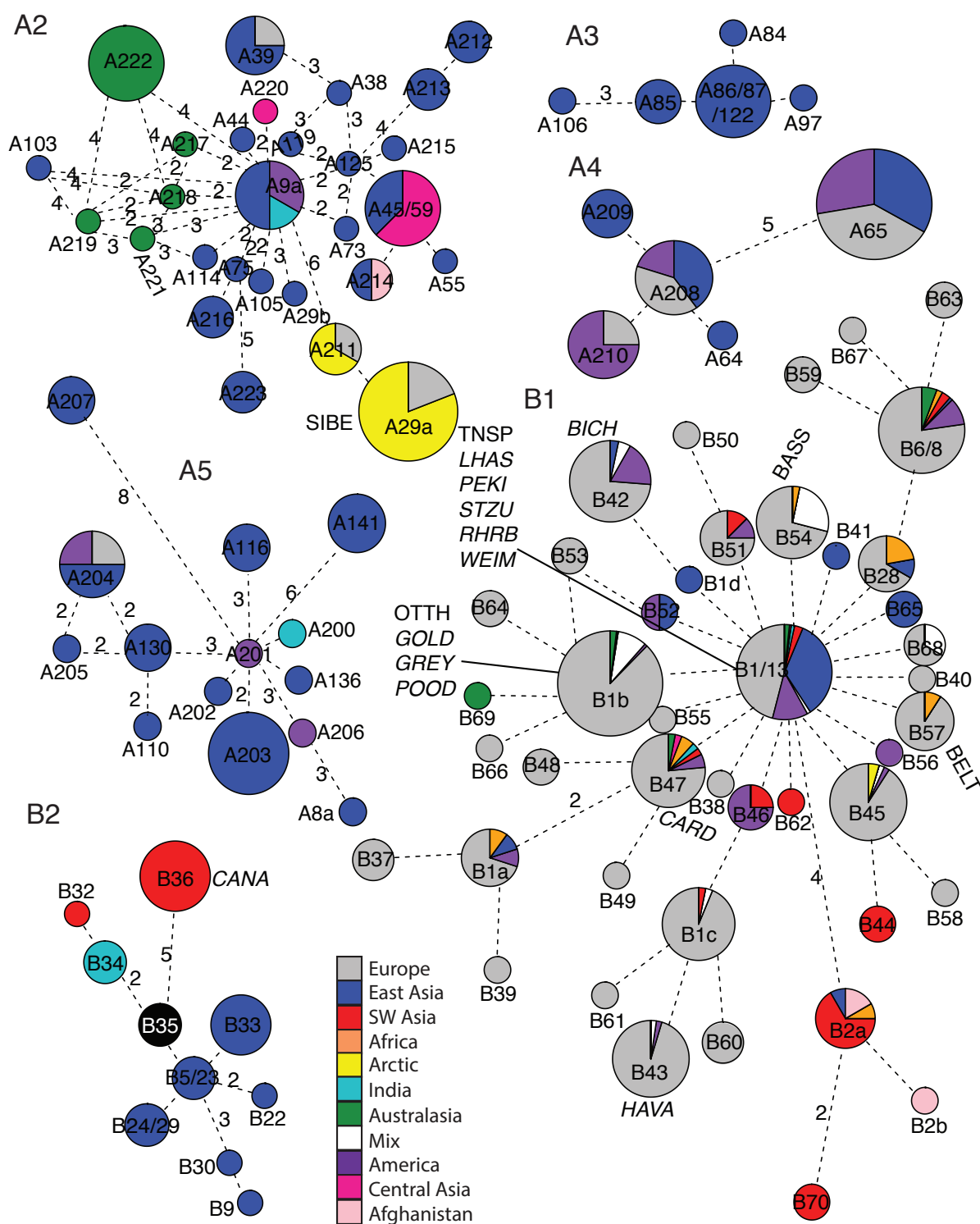


Fig. S8: Minimum spanning Mt haplotype networks for A and B haplogroups. Circle size is proportional to haplotype frequency and line length is determined by number of mutations separating haplotypes (indicated by numerals when > 1 mutation). Previously reported haplotypes have maintained their published names. Novel haplotypes, or published haplotypes which did not follow the naming convention from Pang et al [8], have names larger than A200 or B32. Some haplotypes are particularly common in individual breeds. Breed abbreviation is printed next to a haplotype if we had at least 5 individuals from that breed and more than 80% (or 50% in italics) of those individuals carried the haplotype. Ancient American haplotypes are from Thalmann et al. [11].

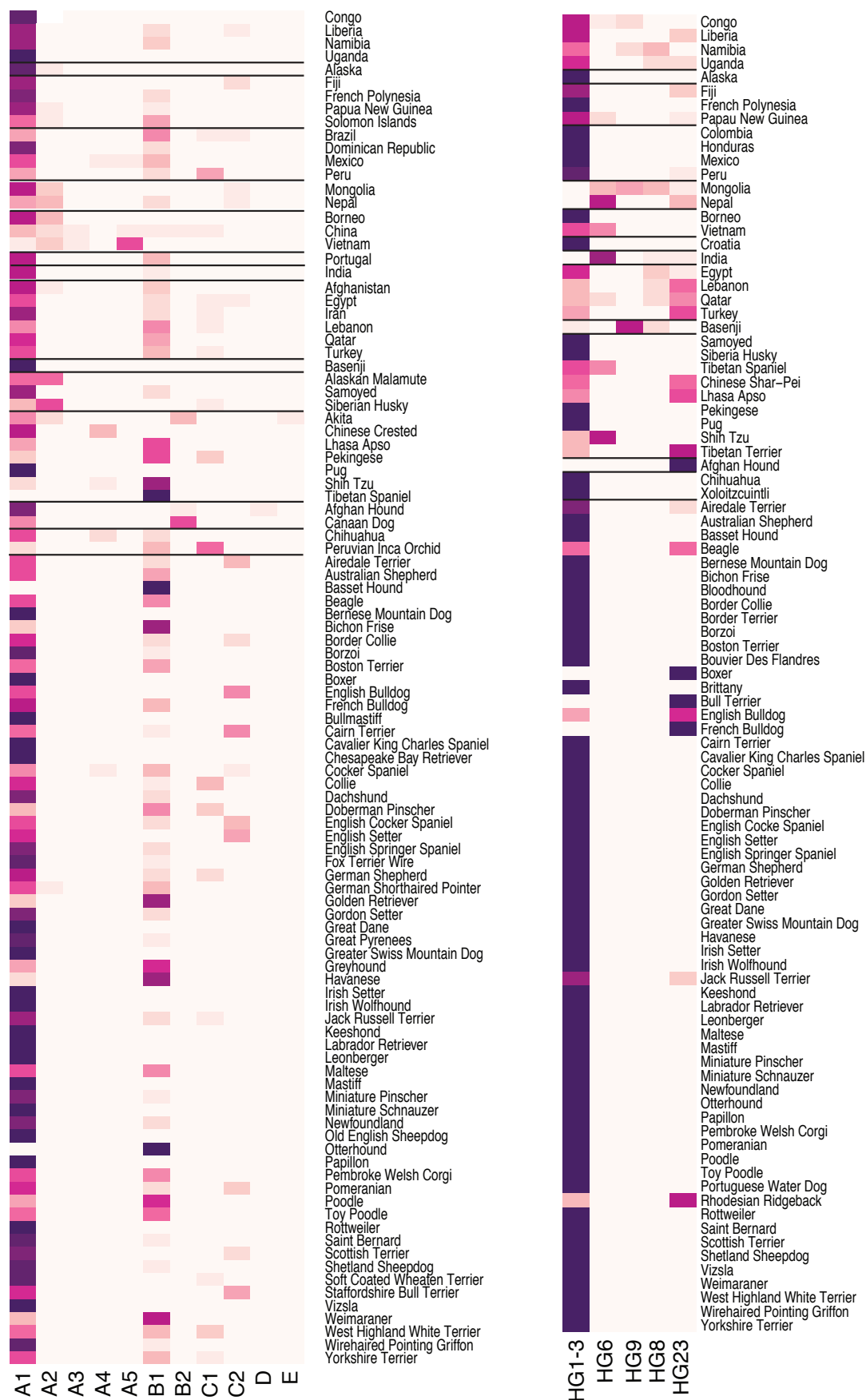


Fig. S9: Heat maps for haplogroups found within each population containing at least 10 dogs (at least 6 males). All Mt haplogroups (left) and Y haplogroups (right) are listed. Dark purple indicates that a group of dogs is fixed for a haplogroup while white indicates that the haplogroup is not found within that group of dogs.

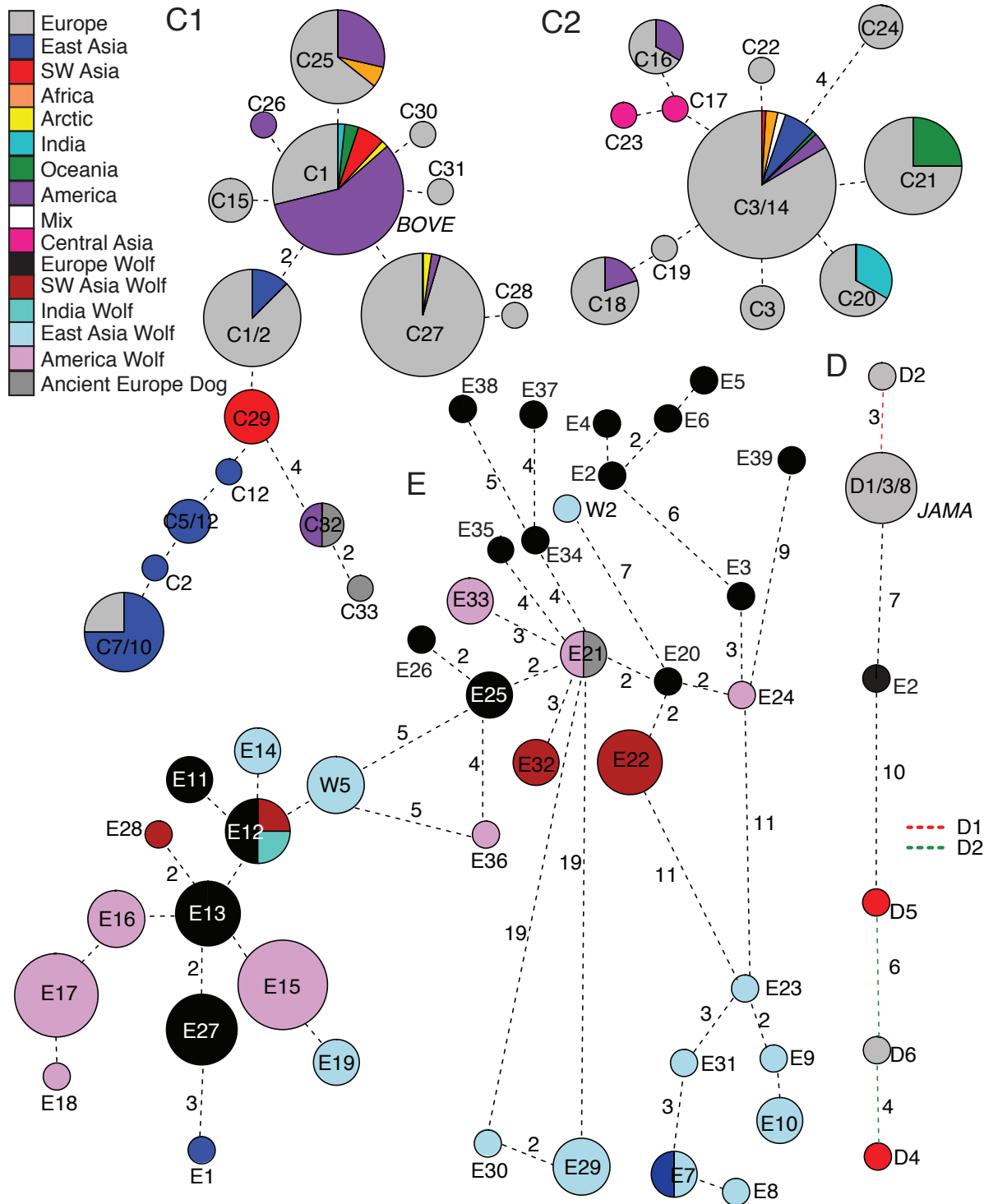


Fig. S10: Minimum spanning haplotype networks for C, D, and E haplogroups. Circle size is proportional to haplotype frequency and line length is determined by number of mutations separating haplotypes (indicated by numerals when > 1 mutation). Previously reported haplotypes have maintained their published names. Novel haplotypes, or published haplotypes which did not follow the naming convention from Pang et al. [8], have names larger than C15, or E1 (all D haplotypes have been previously reported). Some haplotypes are particularly common in individual breeds. Breed abbreviation is printed next to a haplotype if we had at least 5 individuals from that breed and more than 80% (or 50% in *italics*) of those individuals carried the haplotype. Ancient European wolf haplotypes are from Thalmann et al. [11].

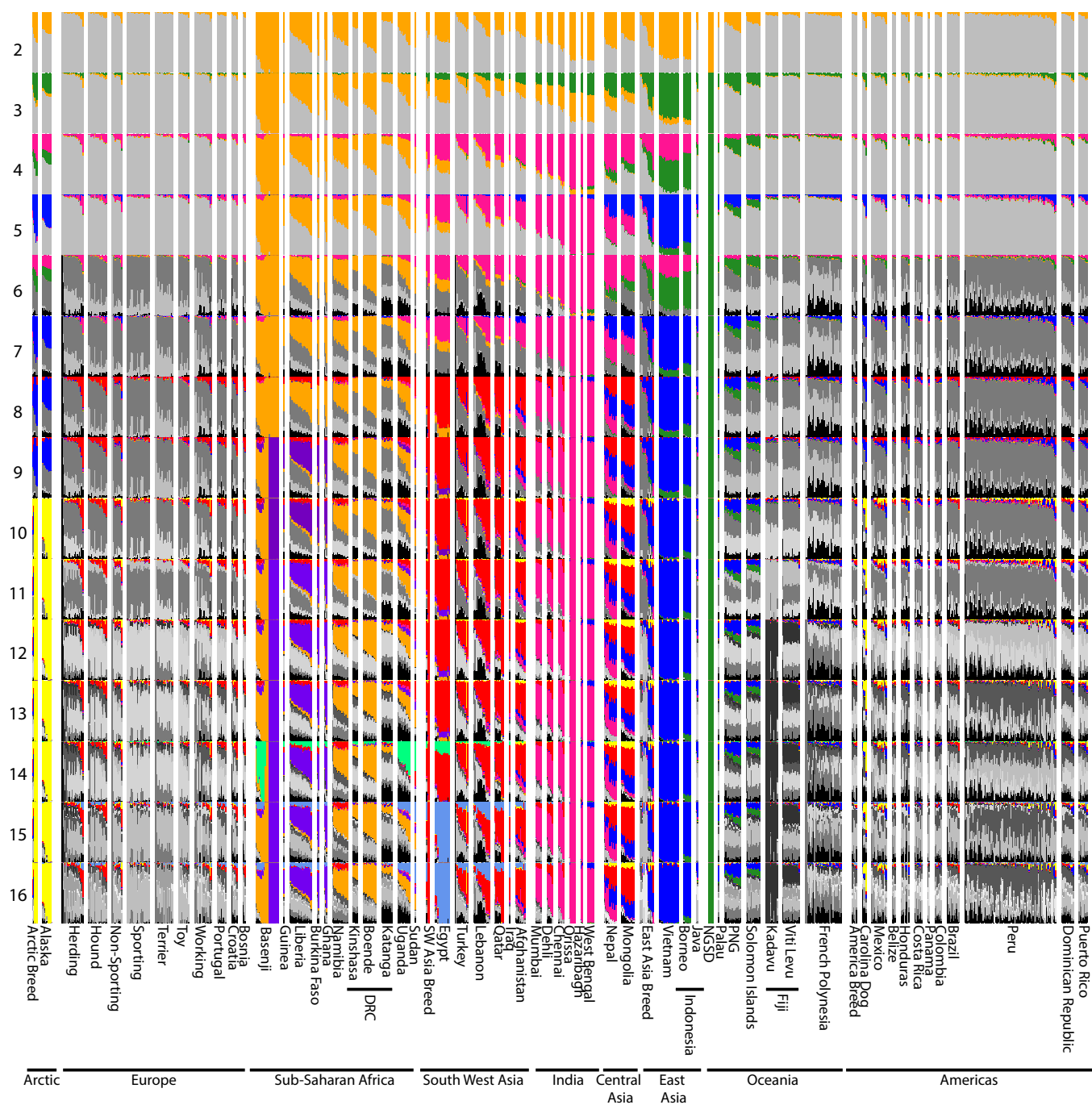


Fig. S11: ADMIXTURE analysis of village and purebred dogs at all K between 2 and 16. European components are in gray scale while regional components are in color. Each vertical line represents an individual, breeds are represented by a single individual with the exception of Basenjis, New Guinea Singing Dogs, and Carolina Dogs for which all unrelated individuals are included.

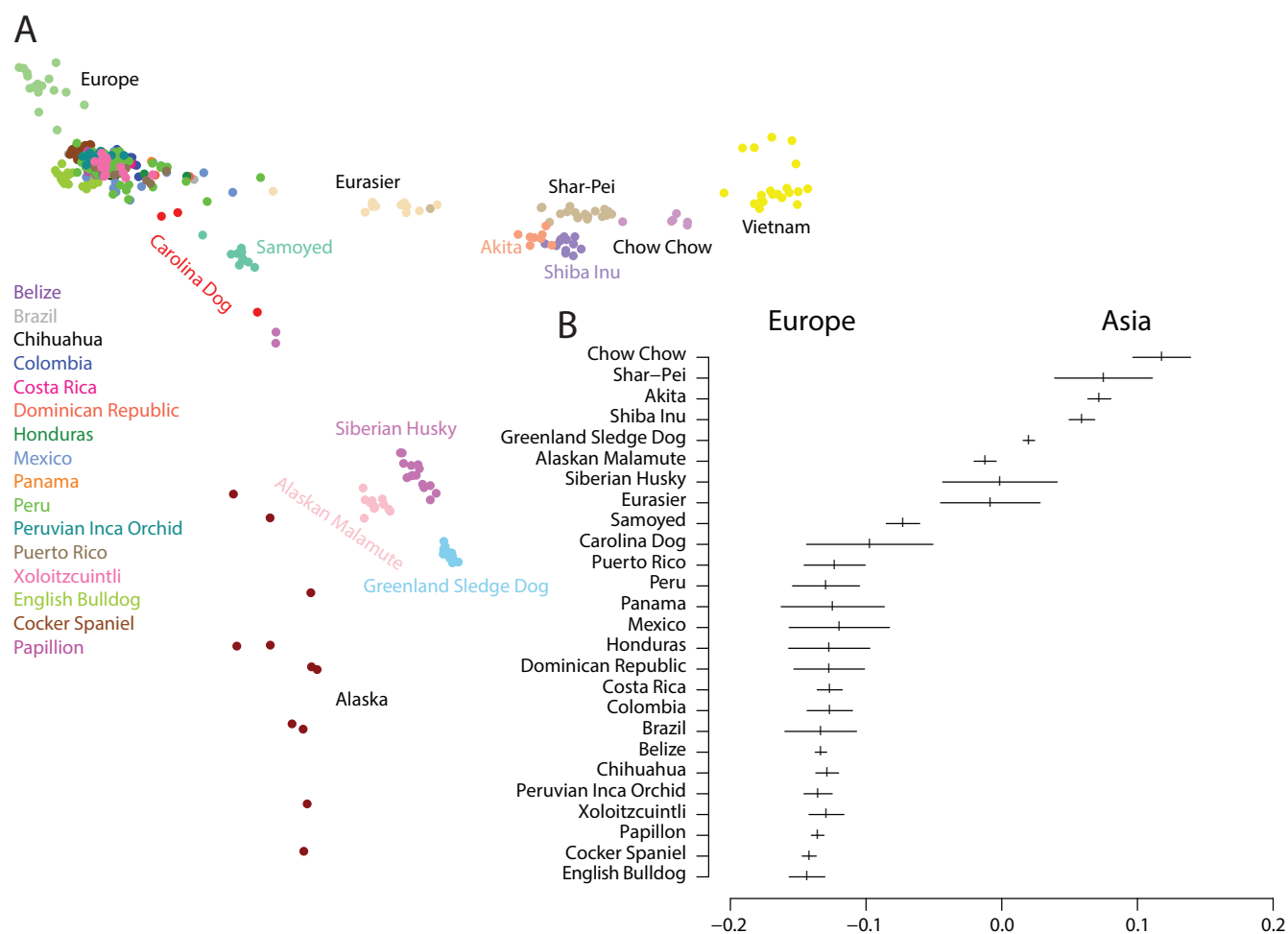


Fig. S12: A. Principal component analysis of European, Vietnamese, and Alaskan village dogs with American dogs (and representative European and Asian breed dogs) projected onto the PCA space. B. Mean value for PC1 of breed and village populations. Error bars indicate two standard deviations from the mean.

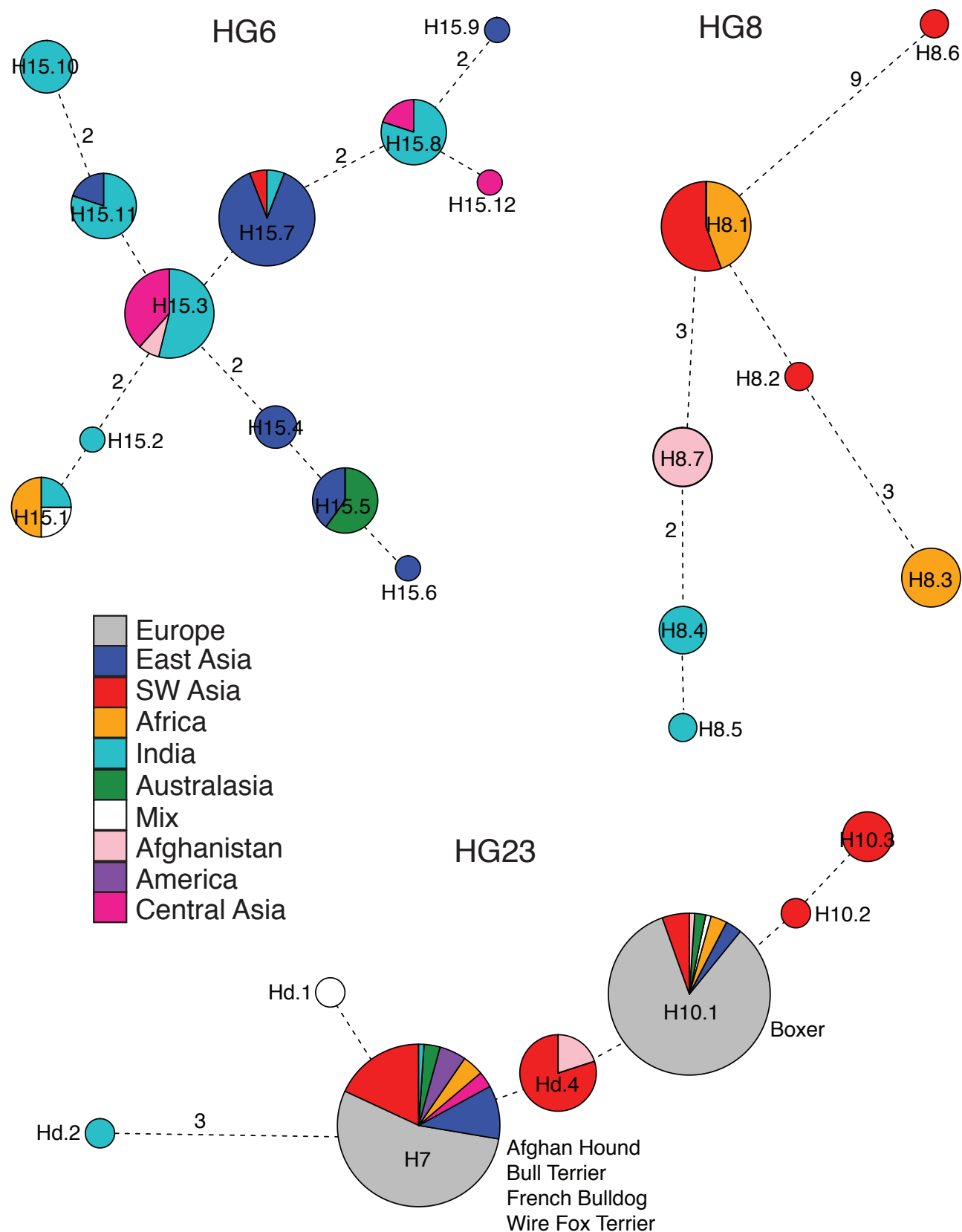


Fig. S13: Minimum spanning networks for the smaller Y haplotypes. Circle size is proportional to haplotype frequency and line length is determined by number of mutations separating haplotypes (indicated by numerals when > 1 mutation). Previously reported haplotypes have maintained their published names. Some haplotypes are particularly common in individual breeds. Breed name is printed next to a haplotype if we had at least 5 individuals from that breed and more than 90% of those individuals carried the haplotype.

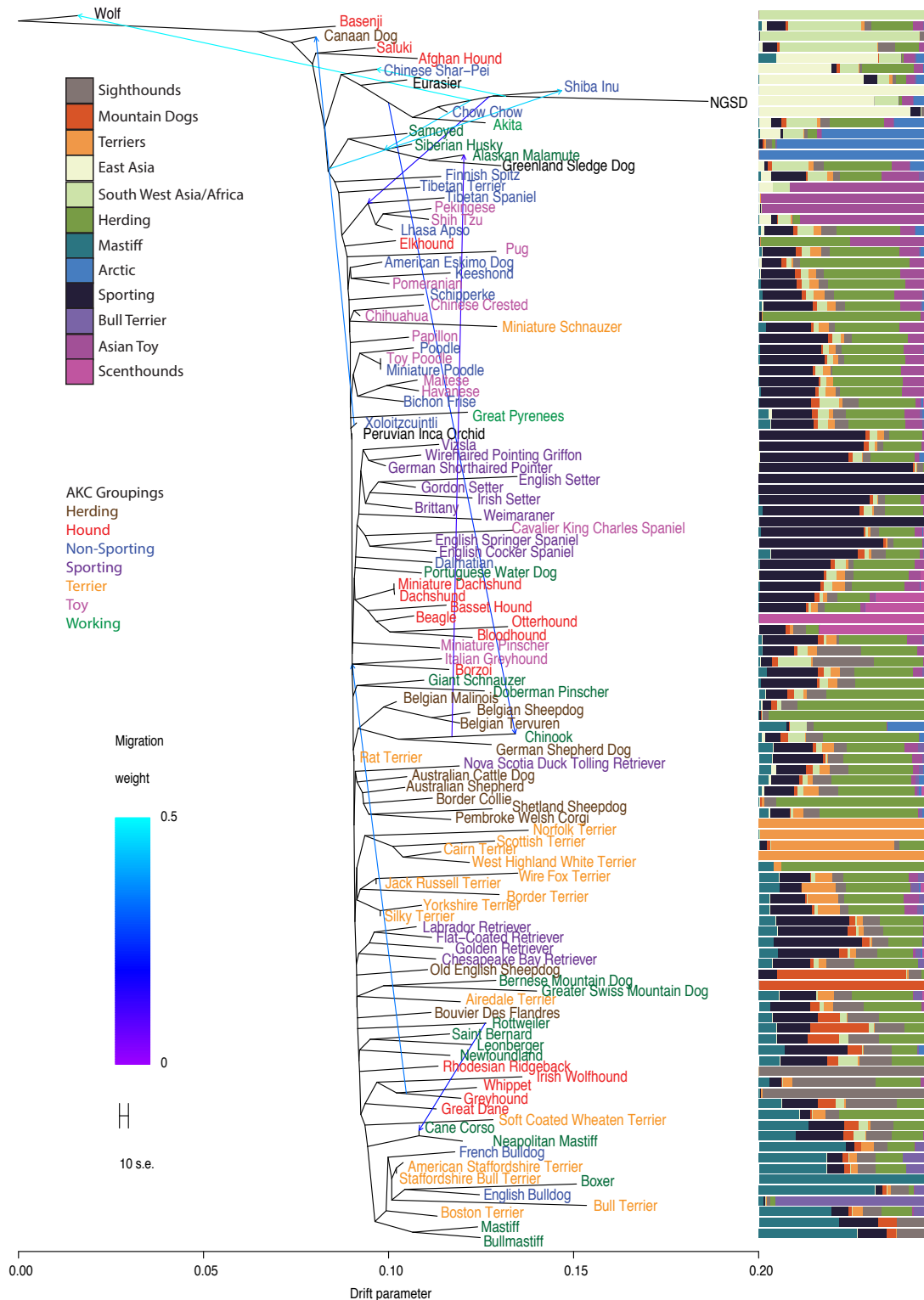


Fig. S14: Breeds form genetic clusters primarily according to geography in TreeMix (left) and ADMIXTURE (right). In both analyses, four basal clusters correspond to the Middle East, Africa, East Asia, and the Arctic with the fifth (European) cluster containing distinct subgroups corresponding to Sight Hounds, Scent Hounds, Sporting Dogs, Retrievers, Terriers, Mastiffs, and Herding Dogs, although many AKC groups (e.g. Toy, Working and Non-sporting) are spread throughout the tree. A few non-European breeds (Rhodesian Ridgeback, Chihuahua, Xoloitzcuintli and Peruvian Inca Orchid) group with European dogs, having been reconstituted over time from European stock. Four Asian breeds—Tibetan Spaniel, Lhasa Apso, Pekingese and Shih Tzu—also lie within the European cluster but retain between 18% and 44% Asian ancestry depending on breed (Fig. S15), consistent with their origins in East Asia followed by introgression from European dogs during modern times. Other admixture edges are consistent with known origins of modern breeds—Cane Corso as a mix of Rottweiler and Mastiff, Chinooks as a mix of German Shepherds and other dogs. The remaining admixture edges suggest gene flow between more ancient populations.

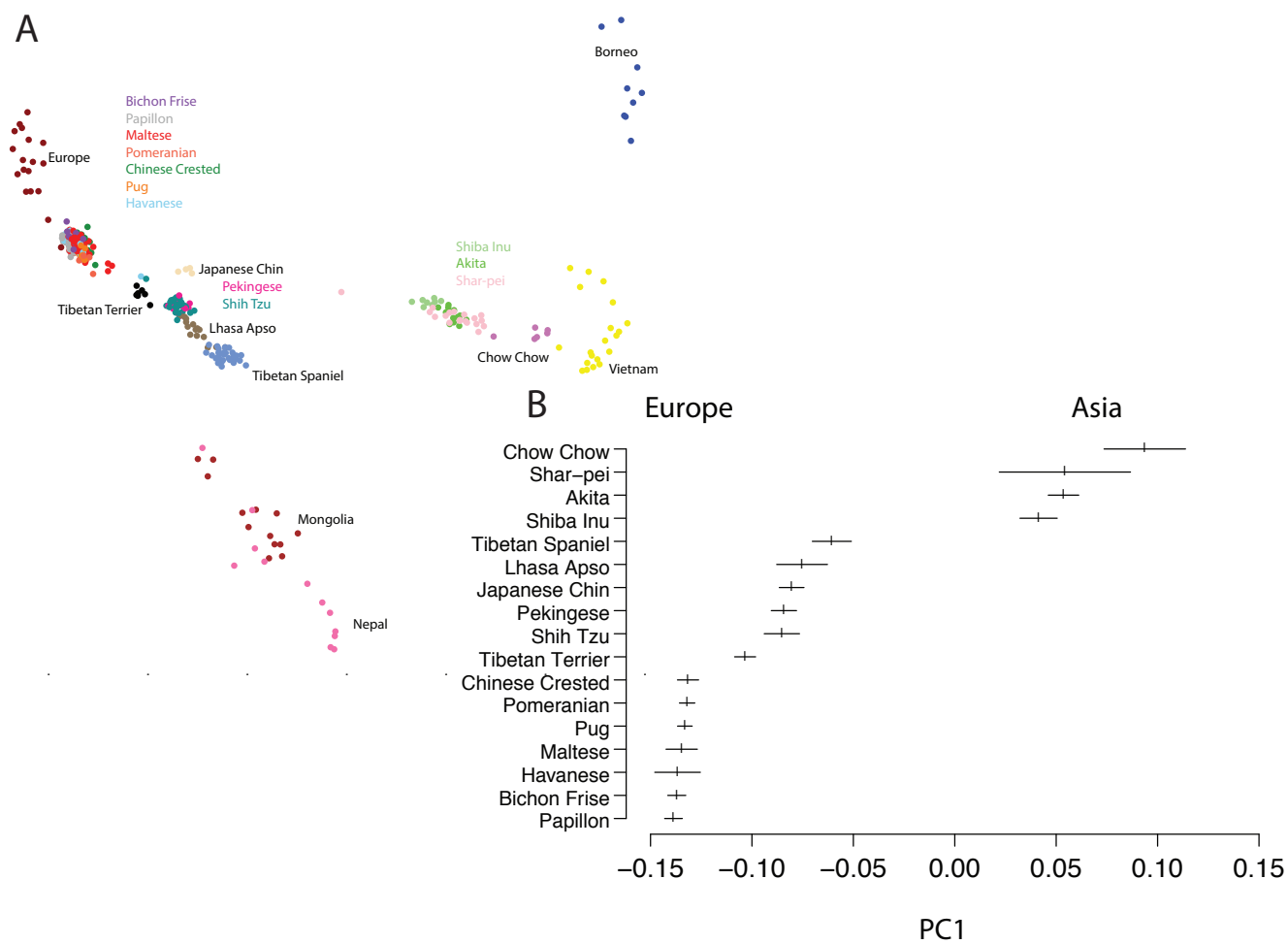


Fig. S15: A. Principal component analysis of European, Vietnamese, Bornese, Mongolian, and Nepalese village dogs with toy dogs (and representative Asian breed dogs) projected onto the PCA space. B. Mean value for PC1 of breed and village populations. Error bars indicate two standard deviations from the mean.

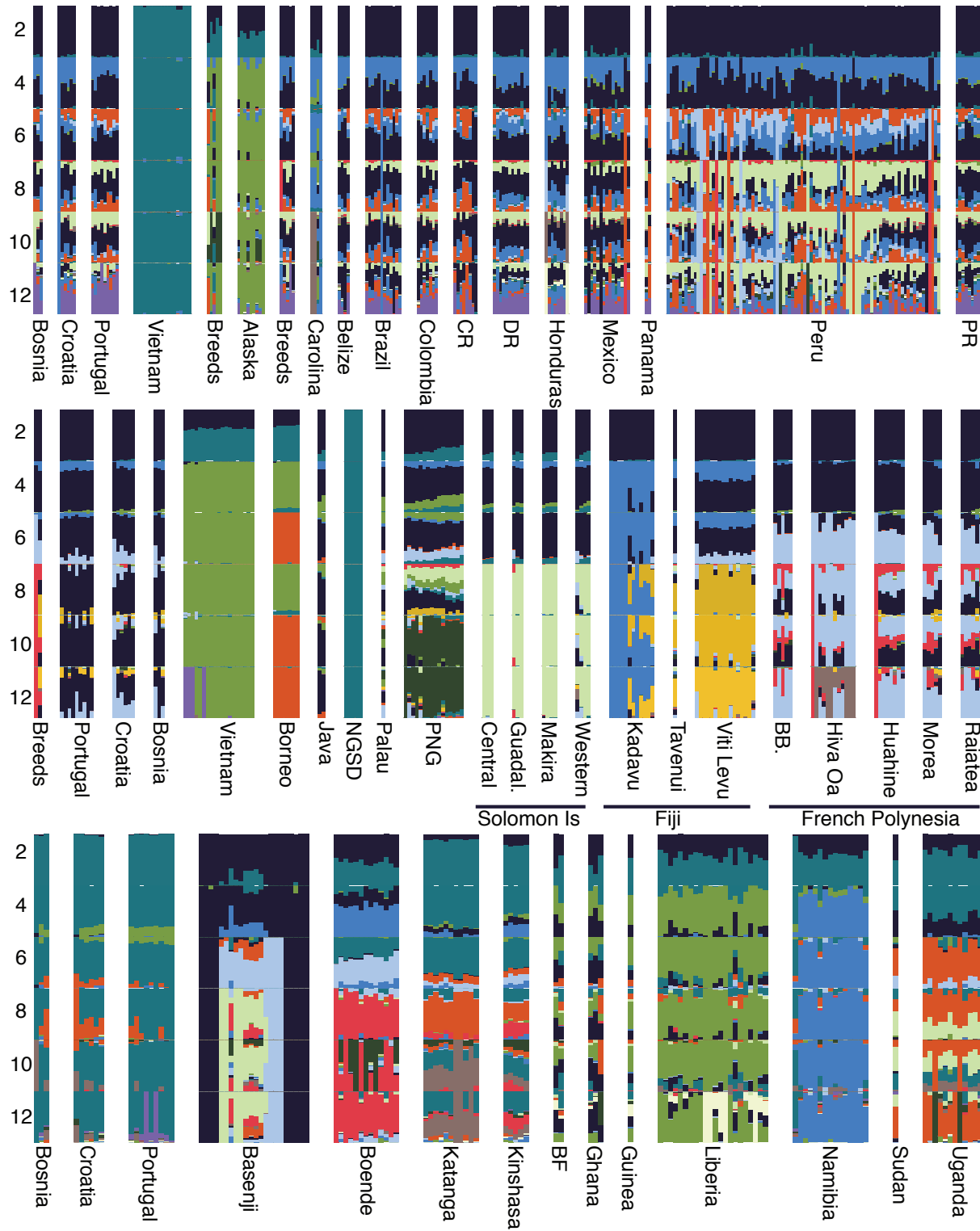


Fig. S16: Regional ADMIXTURE analyses. Top: New World Dogs, including Carolina dogs, American and Arctic breeds (Chihuahua, Peruvian Inca Orchid, Xoloitzcuintli, Greenland Sledge Dog, Alaskan Malamute, Siberian Husky, Eurasier, and Samoyed) with one individual per breed, and European and East Asian village dogs as outgroups. PR is Puerto Rico, DR is Dominican Republic, and CR is Costa Rica. Middle: Village dogs from Oceania and Island Southeast Asia, with select East Asian and European dogs. BB stands for Bora Bora. The breeds referenced here are Staffordshire Bull Terrier and Mastiff. Bottom: Village dogs from Africa with European village dogs as an out group. BF stands for Burkina Faso.

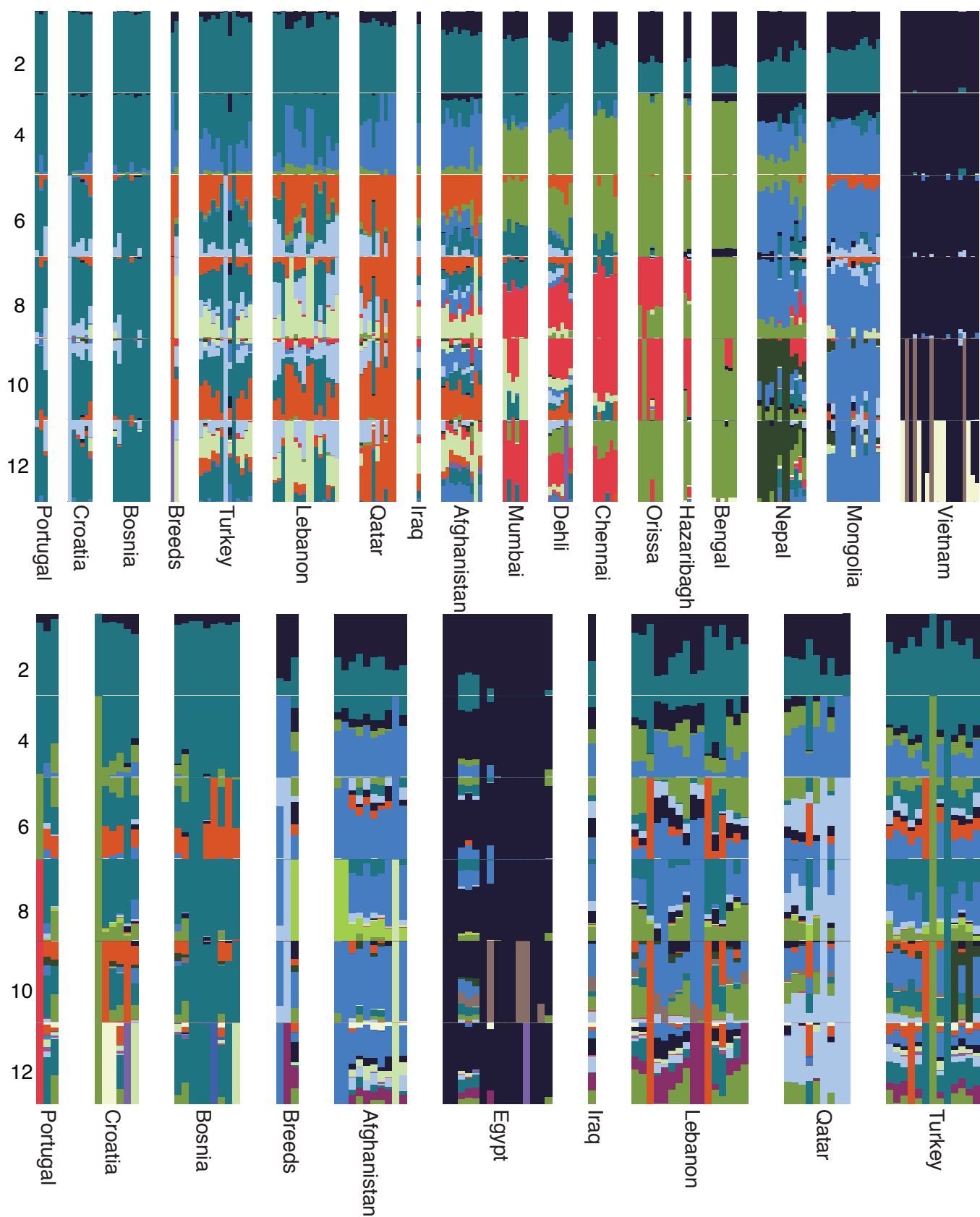


Fig. S17: Regional ADMIXTURE analyses, with two subsets of Eurasian village dogs—spanning the entire continent (top) and focused on Europe and South West Asia(bottom). In both cases an Afghan Hound and a Saluki are included as breeds.).

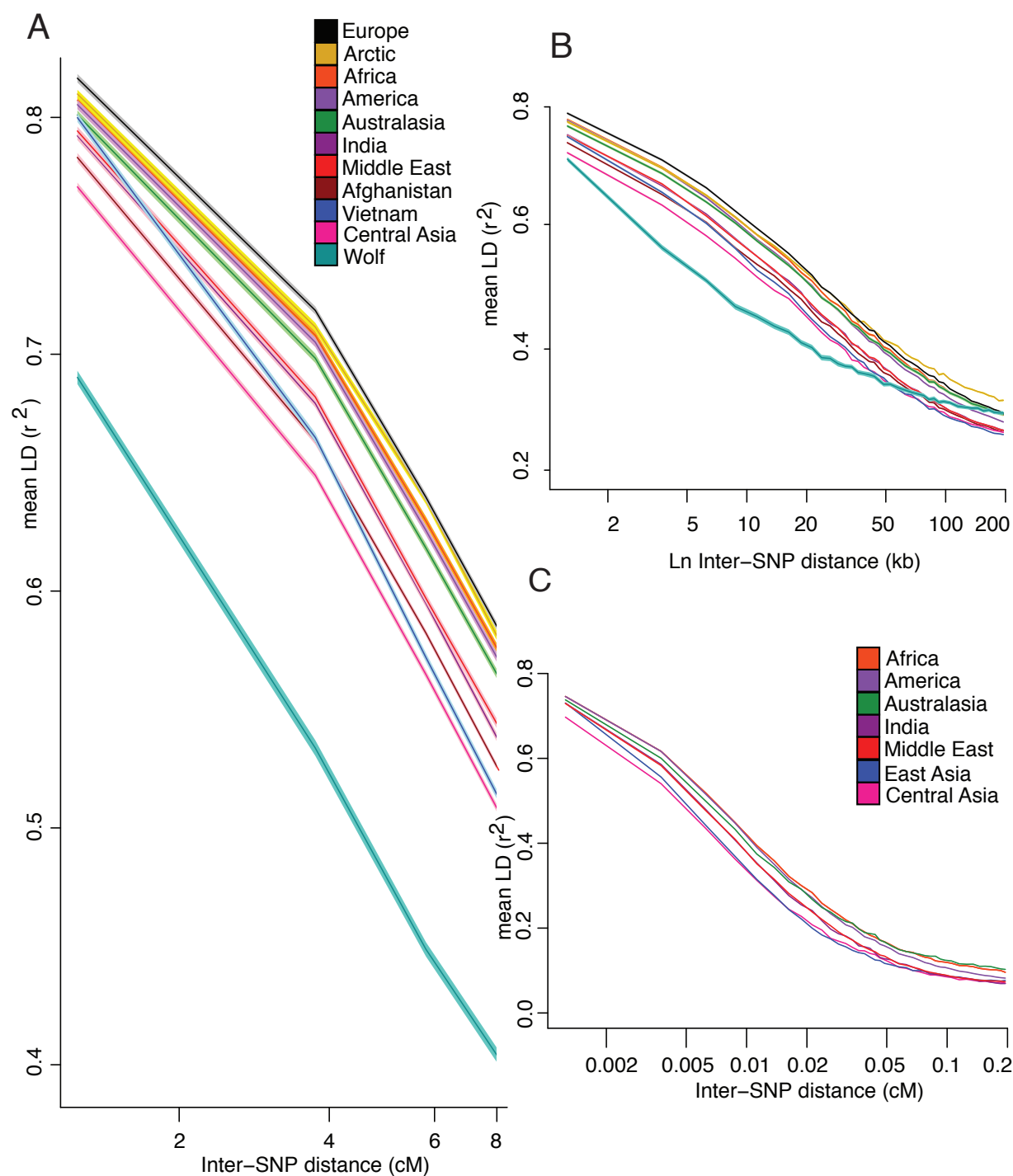


Fig. S18: LD decay for village dog populations worldwide. (A) LD at small inter-SNP distances. Dark lines represent means and lighter shading represent 1 s.e. interval calculated from 100 replicates. For each replicate, $N = 6$ individuals were selected randomly. This is a subsection of Fig 5 enlarged to make the 1 s.e. intervals visible. (B) LD curves for the same populations calculated in the same way as in A but with inter-SNP distance measured in kb. (C) LD curves calculated in the same way as A but for $N = 20$, a smaller number of populations were considered based on which populations contained more than 20 individuals. Although the mean LD is lower, the relative ordering of populations stays the same.

Table S3: Shared vs. regional haplotypes in 651 village dogs

region	population	Mt		Y	
		both	local	both	local
Africa	Burkina Faso	0	2	0	1
	Congo	18	15	16	6
	Ghana	4	1	1	1
	Guinea	1	0	1	0
	Liberia	10	11	15	1
	Namibia	13	2	6	5
	Nigeria	0	1	-	-
	Sudan	0	1	-	-
	Uganda	7	5	9	2
America	Belize	4	0	3	0
	Brazil	11	2	4	0
	Colombia	9	0	6	0
	Costa Rica	7	1	2	0
	Dominican Republic	10	2	2	0
	Honduras	8	0	6	0
	Mexico	15	0	10	0
	Panama	2	0	-	-
	Peru	83	8	61	0
Arctic	Puerto Rico	7	2	5	0
	Alaska	7	4	4	2
Oceania	Fiji	29	0	14	0
	French Polynesia	35	1	15	0
	Palau	1	0	-	-
	Papua New Guinea	16	4	9	6
	Solomon Island	14	1	3	0
Central Asia	Mongolia	5	9	1	10
	Nepal	6	6	2	4
East Asia	Borneo	2	7	0	6
	China	29	51	-	-
	Java	2	0	2	0
	Thailand	4	3	-	-
	Vietnam	0	20	6	10
Europe	Bosnia	3	0	3	0
	Croatia	6	0	6	0
	Portugal	6	3	5	0
India	India	15	20	2	25
South West Asia	Afghanistan	8	5	1	2
	Egypt	7	8	2	6
	Iran	12	1	-	-
	Iraq	1	0	1	0
	Lebanon	9	7	9	6
	Qatar	5	4	2	5
	Turkey	11	2	9	1

Table S4: Mt haplotype diversity in novel and combined Mt data sets

Population	Semi-Custom CanineHD Array				Combined			
	N	H	HT(HG)	HT(SD)	N	H	HT(HG)	HT(SD)
Africa Breed	30	0.93	15(1)	11.06(3.36)	33	0.93	15(1)	11.92(1.10)
Africa Village	91	0.96	41(4)	14.05(3.81)	91	0.96	41(4)	14.91(1.66)
America Breed	46	0.92	20(6)	12.06(1.53)	48	0.93	21(6)	12.38(1.67)
America Village	171	0.94	49(7)	13.45(1.92)	171	0.94	49(7)	13.82(1.66)
Arctic Breed	39	0.84	11(4)	8.77(1.07)	42	0.83	12(4)	8.92(1.24)
Arctic Village	11	0.82	5(2)	-	11	0.82	5(2)	-
Oceania Breed	5	0	1(1)	-	6	0.33	2(1)	-
Oceania Village	101	0.93	29(5)	12.44(1.64)	105	0.93	29(5)	12.48(1.79)
Central Asia Village	25	0.95	16(5)	13.83(0.95)	26	0.95	17(5)	13.92(1.04)
East Asia Breed	151	0.85	32(10)	10.68(1.76)	166	0.86	38(10)	11.67(2.05)
East Asia Village	31	0.96	19(5)	14.36(1.42)	118	0.99	74(10)	17.95(1.11)
Europe Breed	3414	0.95	151(7)	13.99(1.82)	3534	0.95	164(8)	14.20(1.92)
Europe Village	18	0.96	15(3)	-	18	0.96	15(3)	-
India Village	31	0.98	23(7)	16.41(1.12)	35	0.97	24(7)	15.44(1.44)
South West Asia Breed	31	0.85	10(4)	8.60(0.88)	37	0.90	16(4)	11.00(1.36)
South West Asia Village	67	0.97	34(6)	15.26(1.47)	80	0.96	36(6)	14.74(1.67)

Sample size (N) and haplotype heterozygosity (H) by region. HT(HG) is the number of unique haplotypes (haplogroups) and HT(SD) is the average number (standard deviation) of haplotypes from a sub-sample of 20 dogs.

Table S5: Y haplotype diversity in breed dogs (top) and village dogs (bottom)

Population	N	H	HT(HG)	HT(SD)
Africa	10	0.51	3(3)	-
America	26	0.67	7(1)	5.91(0.78)
Arctic	15	0.79	5(1)	-
East Asia	83	0.88	12(3)	9.17(1.35)
Europe	1708	0.91	34(3)	10.65(1.64)
South West Asia	14	0.59	5(3)	-
Africa	64	0.93	16(5)	11.58(1.49)
America	99	0.86	15(2)	8.09(1.33)
Arctic	6	0.87	4(1)	-
Oceania	47	0.94	19(3)	12.36(1.34)
Central Asia	17	0.84	6(4)	-
East Asia	24	0.92	13(2)	11.68(0.85)
Europe	14	0.93	9(1)	-
India	27	0.89	12(4)	10.19(1.01)
South West Asia	46	0.94	20(4)	12.41(1.38)

Sample size (N) and haplotype heterozygosity (H) by region. HT(HG) is the number of unique haplotypes (haplogroups) and HT(SD) is the average number (standard deviation) of haplotypes from a sub-sample of 20 dogs.

Table S6: Ancestry percentages for African and island populations estimated by F3 tests

Source population 1	Source population 2	Target population	F3	Z score	Ancestry proportion from source 1
Europe	Vietnam	French Polynesia	0.01	14.82	0.958-0.991
Europe	Borneo	French Polynesia	0.01	13.02	0.958-0.981
Europe	Vietnam	Fiji	0.01	15.34	0.952-0.966
Europe	Borneo	Fiji	0.01	12.57	0.951-0.959
Europe	Vietnam	Solomon Islands	-0.00	-2.28	0.785-0.855
Europe	Borneo	Solomon Islands	-0.01	-6.48	0.783-0.863
Europe	Vietnam	PNG	-0.02	-32.26	0.680-0.715
Europe	Borneo	PNG	-0.02	-30.68	0.678-0.784
Europe	Basenji	Katanga	-0.01	-5.02	0.884-0.856
Europe	Basenji	Kinshasa	-0.03	-27.66	0.749-0.730
Europe	Basenji	Uganda	-0.03	-40.38	0.580-0.638
Europe	Basenji	Namibia	-0.03	-32.83	0.525-0.609
Europe	Basenji	Liberia	-0.03	-39.95	0.522-0.599
Europe	Basenji	Boende	-0.031	-33.72	0.394-0.355

Table S7: F_{st} between American populations

	Alaska	ArcB	Bos	Cro	EAB	EurB	Por	Vie	Mon	AmB	Car	Mex	Hon	CR	Col	Bra	Peru	DR
Arctic breed	0.02																	
Bosnia	0.08	0.06																
Croatia	0.07	0.06	0.00															
East Asia breed	0.04	0.03	0.02	0.02														
Europe breed	0.05	0.03	0.00	0.00	0.03													
Portugal	0.07	0.07	0.01	0.00	0.03	0.00												
Vietnam	0.11	0.09	0.15	0.15	0.06	0.13	0.15											
Mongolia	0.05	0.04	0.05	0.04	0.02	0.04	0.05	0.08										
America breed	0.09	0.06	0.00	0.01	0.02	0.00	0.02	0.14	0.06									
Carolina Dog	0.06	0.05	0.02	0.02	0.02	0.00	0.03	0.13	0.04	0.04								
Mexico	0.06	0.06	0.01	0.00	0.02	0.00	0.00	0.13	0.04	0.01	0.02							
Honduras	0.07	0.07	0.02	0.01	0.03	0.00	0.01	0.15	0.05	0.02	0.02	0.01						
Costa Rica	0.07	0.06	0.01	0.00	0.02	0.00	0.01	0.14	0.04	0.01	0.02	0.00	0.01					
Colombia	0.07	0.07	0.01	0.01	0.02	0.00	0.01	0.14	0.04	0.02	0.02	0.00	0.01	0.01				
Brazil	0.07	0.07	0.01	0.00	0.03	0.00	0.00	0.14	0.04	0.02	0.02	0.00	0.01	0.00	0.01			
Peru	0.06	0.04	0.00	0.00	0.03	0.00	0.01	0.13	0.04	0.00	0.01	0.00	0.01	0.01	0.00	0.00		
Dominican Rep	0.06	0.07	0.01	0.01	0.03	0.00	0.01	0.14	0.04	0.02	0.02	0.00	0.01	0.01	0.01	0.00	0.00	
Puerto Rico	0.06	0.06	0.01	0.01	0.02	0.00	0.00	0.13	0.04	0.02	0.02	0.00	0.01	0.00	0.00	0.00	0.00	0.00

Table S8: F_{st} between Pacific Island populations

	Kadavu	Europe	Huahine	Moorea	Raiatea	PNG	NGSD	Borneo	Vietnam	Hiva Oa	Solomon Islands	Bora Bora
Europe	0.04											
Huahine	0.04	0.01										
Moorea	0.04	0.01	0.00									
Raiatea	0.04	0.01	0.00	0.00								
PNG	0.04	0.03	0.02	0.02	0.02							
NGSD	0.39	0.14	0.41	0.46	0.44	0.29						
Boreno	0.17	0.09	0.16	0.17	0.17	0.10	0.32					
Vietnam	0.16	0.14	0.15	0.15	0.15	0.08	0.26	0.05				
Hiva Oa	0.04	0.02	0.01	0.01	0.01	0.03	0.39	0.17	0.15			
Solomon Islands	0.04	0.03	0.03	0.02	0.03	0.01	0.32	0.12	0.11	0.032		
Bora Bora	0.04	0.01	0.00	0.00	0.00	0.02	0.46	0.17	0.15	0.01	0.03	
Viti Levu	0.02	0.02	0.02	0.01	0.02	0.02	0.37	0.16	0.14	0.02	0.03	0.02

Table S9: F_{st} between African populations

	Basenji	Boende	Bosnia	Croatia	Egypt	Europe Breeds	Ghana	Katanga	Kinshasa	Liberia	Namibia	Portugal
Boende	0.06											
Bosnia	0.21	0.09										
Croatia	0.21	0.08	0.00									
Egypt	0.13	0.06	0.06	0.05								
Europe Breeds	0.17	0.07	0.00	0.00	0.05							
Ghana	0.06	0.04	0.08	0.08	0.06	0.02						
Katanga	0.17	0.06	0.02	0.01	0.05	0.01	0.07					
Kinshasa	0.14	0.04	0.02	0.01	0.05	0.00	0.05	0.01				
Liberia	0.08	0.03	0.05	0.05	0.04	0.05	0.02	0.04	0.03			
Namibia	0.09	0.03	0.06	0.05	0.05	0.04	0.04	0.04	0.02	0.03		
Portugal	0.20	0.08	0.01	0.00	0.05	0.00	0.08	0.02	0.02	0.05	0.05	
Uganda	0.10	0.04	0.04	0.04	0.04	0.03	0.04	0.03	0.02	0.02	0.03	0.04

Table S10: F_{st} between Asian populations

	Vie ^a	Bas ^b	Ben ^c	Bos ^d	Che ^e	Cro ^f	Dehli	EAB ^g	Egypt	EB ^h	Leb ⁱ	Mon ^j	Mum ^k	Nepal	Ori ^l	Por ^m	Qatar	SWAB ⁿ	Tur ^o
Bas	0.25																		
Ben	0.13	0.17																	
Bos	0.15	0.21	0.13																
Che	0.14	0.16	0.03	0.08															
Cro	0.15	0.21	0.12	0.00	0.07														
Dehli	0.12	0.16	0.03	0.05	0.01	0.04													
EAB	0.06	0.13	0.07	0.02	0.05	0.02	0.03												
Egypt	0.14	0.13	0.07	0.06	0.05	0.05	0.03	0.05											
EB	0.13	0.17	0.08	0.00	0.05	0.00	0.03	0.03	0.05										
Leb	0.13	0.16	0.08	0.02	0.05	0.01	0.03	0.03	0.03	0.01									
Mon	0.08	0.18	0.07	0.05	0.05	0.04	0.03	0.02	0.05	0.04	0.03								
Mum	0.13	0.16	0.04	0.05	0.01	0.05	0.00	0.03	0.04	0.03	0.03	0.04							
Nepal	0.07	0.16	0.04	0.06	0.04	0.05	0.02	0.02	0.04	0.05	0.04	0.02	0.03						
Ori	0.15	0.17	0.01	0.13	0.02	0.12	0.03	0.08	0.07	0.07	0.09	0.08	0.03	0.05					
Por	0.15	0.20	0.12	0.01	0.08	0.00	0.05	0.03	0.05	0.00	0.02	0.05	0.05	0.06	0.12				
Qatar	0.13	0.16	0.08	0.03	0.05	0.03	0.03	0.03	0.03	0.02	0.01	0.03	0.04	0.04	0.08	0.03			
SWAB	0.14	0.18	0.09	0.02	0.05	0.04	0.03	0.03	0.04	0.01	0.02	0.04	0.04	0.04	0.09	0.04	0.01		
Tur	0.13	0.17	0.09	0.01	0.06	0.00	0.03	0.02	0.04	0.01	0.00	0.03	0.03	0.04	0.10	0.01	0.01	0.02	
Afg	0.11	0.16	0.06	0.03	0.04	0.03	0.02	0.02	0.03	0.02	0.01	0.02	0.02	0.02	0.07	0.03	0.01	0.02	0.01
^a Vietnam																			
^b Basenji																			
^c Bengal																			
^d Bosnia																			
^e Chennai																			
^f Croatia																			
^g East Asia Breed																			
^h Europe Breed																			
ⁱ Lebanon																			
^j Mongolia																			
^k Mumbai																			
^l Orissa																			
^m Portugal																			
ⁿ South West Asian Breed																			
^o Turkey																			

Table S11: Sampling locations for village dogs

region	country	location	N_{mal}	N_{fem}	notes
Africa	Guinea	Kankan	1	0	US import
Africa	South Sudan	Sudan	0	1	US import
Africa	Nigeria	unknown	1	0	sampled in Monrovia
Africa	Namibia	northern	10	4	see Boyko et al 2009
Africa	Namibia	central	1	0	see Boyko et al 2009
Africa	Uganda	mainland	11	1	see Boyko et al 2009
Africa	Liberia	Lofa	13	6	various districts
Africa	Liberia	Monrovia	3	1	
Africa	DR Congo	Boende	8	7	
Africa	DR Congo	Kinshasa	5	1	
Africa	DR Congo	Katanga	9	3	
Africa	Ghana	Cape Coast	2	3	
Africa	Burkina Faso	Bobo-Dioulasso	1	1	
America	Peru	Arequipa	11	5	various communities
America	Peru	Cusco	18	7	various communities
America	Peru	Ica	7	1	Nazca
America	Peru	Loreto	13	12	various communities
America	Peru	Puno	12	5	various communities
America	Panama	various	0	2	US imports
America	Mexico	Mexico City	5	1	
America	Mexico	Morelia	5	4	
America	Brazil	various	4	9	
America	Dominican Republic	Sosua	2	10	
America	Honduras	Roatán	6	2	
America	Colombia	Colombia	6	3	
America	United States	Puerto Rico	5	4	see Boyko et al 2009
America	Belize	Belize	3	1	
America	Costa Rica	Costa Rica	2	6	
Arctic	United States	Alaska	6	5	coastal and interior Alaskan Village Huskies
Central Asia	Mongolia	various	11	2	Bankhar
Central Asia	Nepal	various	6	6	
East Asia	Indonesia	Borneo	6	3	
East Asia	Vietnam	Lào Cai	4	0	
East Asia	Vietnam	Hà Giang	4	1	
East Asia	Vietnam	Cao Bang	4	2	
East Asia	Vietnam	Lang Son	4	1	
East Asia	Indonesia	Jakarta	2	0	
Europe	Portugal	various	5	4	
Europe	Croatia	various	6	0	
Europe	Bosnia	various	3	0	
India	India	Orissa	6	0	
India	India	Chennai	4	2	
India	India	Dehli	6	0	
India	India	Hazaribagh	1	1	
India	India	Mumbai	6	0	
India	India	West Bengal	5	1	
Australasia	Papua New Guinea	Port Moresby	8	1	
Australasia	Papua New Guinea	Eastern Highlands	7	4	
Australasia	Palau	Palau	0	1	US import
Australasia	French Polynesia	Society Islands	10	14	Bora Bora, Huahine, Moorea, Raiatea
Australasia	French Polynesia	Marquesas	5	7	Hiva Oa
Australasia	Fiji	Viti Levu	5	11	various villages
Australasia	Fiji	Kadavu	8	4	
Australasia	Fiji	Tavenui	1	0	sampled in Kadavu
Australasia	Solomon Islands	Guadalcanal	1	2	
Australasia	Solomon Islands	Makira	1	3	
Australasia	Solomon Islands	Central	1	3	
Australasia	Solomon Islands	Western	0	4	
Middle East	Qatar	various	7	2	
Middle East	Egypt	Luxor	8	4	see Boyko et al 2009
Middle East	Egypt	Giza	1	2	see Boyko et al 2009
Middle East	Egypt	Kharga	1	0	see Boyko et al 2009
Middle East	Lebanon	Beruit	8	1	
Middle East	Lebanon	Bekaa	7	0	
Middle East	Iraq	various	1	0	US import
Middle East	Afghanistan	various	3	10	US imports
Middle East	Turkey	Istanbul	8	2	
Middle East	Turkey	Giresun	2	1	