

# Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication

Bridgett M. vonHoldt<sup>1</sup>, John P. Pollinger<sup>1</sup>, Kirk E. Lohmueller<sup>2</sup>, Eunjung Han<sup>3</sup>, Heidi G. Parker<sup>4</sup>, Pascale Quignon<sup>4</sup>, Jeremiah D. Degenhardt<sup>2</sup>, Adam R. Boyko<sup>2</sup>, Dent A. Earl<sup>5</sup>, Adam Auton<sup>2</sup>, Andy Reynolds<sup>2</sup>, Kasia Bryc<sup>2</sup>, Abra Brisbin<sup>2</sup>, James C. Knowles<sup>1</sup>, Dana S. Mosher<sup>4</sup>, Tyrone C. Spady<sup>4</sup>, Abdel Elkahloun<sup>4</sup>, Eli Geffen<sup>6</sup>, Malgorzata Pilot<sup>7</sup>, Włodzimierz Jedrzejewski<sup>8</sup>, Claudia Greco<sup>9</sup>, Ettore Randi<sup>9</sup>, Danika Bannasch<sup>10</sup>, Alan Wilton<sup>11</sup>, Jeremy Shearman<sup>11</sup>, Marco Musiani<sup>12</sup>, Michelle Cargill<sup>13</sup>, Paul G. Jones<sup>14</sup>, Zuwei Qian<sup>15</sup>, Wei Huang<sup>15</sup>, Zhao-Li Ding<sup>16</sup>, Ya-ping Zhang<sup>17</sup>, Carlos D. Bustamante<sup>2</sup>, Elaine A. Ostrander<sup>4</sup>, John Novembre<sup>1,18</sup> & Robert K. Wayne<sup>1</sup>

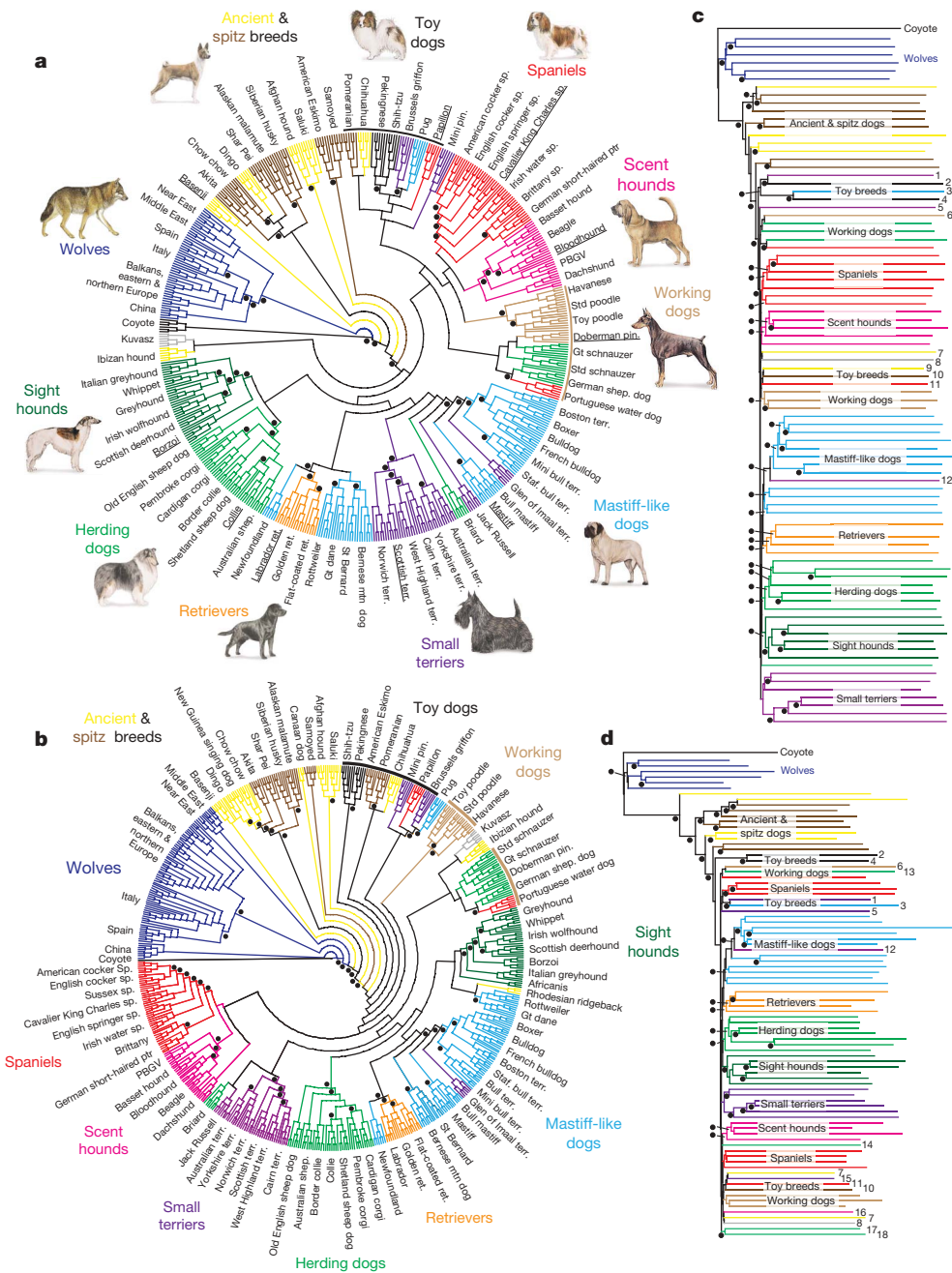
Advances in genome technology have facilitated a new understanding of the historical and genetic processes crucial to rapid phenotypic evolution under domestication<sup>1,2</sup>. To understand the process of dog diversification better, we conducted an extensive genome-wide survey of more than 48,000 single nucleotide polymorphisms in dogs and their wild progenitor, the grey wolf. Here we show that dog breeds share a higher proportion of multi-locus haplotypes unique to grey wolves from the Middle East, indicating that they are a dominant source of genetic diversity for dogs rather than wolves from east Asia, as suggested by mitochondrial DNA sequence data<sup>3</sup>. Furthermore, we find a surprising correspondence between genetic and phenotypic/functional breed groupings but there are exceptions that suggest phenotypic diversification depended in part on the repeated crossing of individuals with novel phenotypes. Our results show that Middle Eastern wolves were a critical source of genome diversity, although interbreeding with local wolf populations clearly occurred elsewhere in the early history of specific lineages. More recently, the evolution of modern dog breeds seems to have been an iterative process that drew on a limited genetic toolkit to create remarkable phenotypic diversity.

The dog is a striking example of variation under domestication, yet the evolutionary processes underlying the genesis of this diversity are poorly understood. To understand the geographic and evolutionary context for phenotypic diversification better, we analysed more than 48,000 single nucleotide polymorphisms (SNPs) typed in a panel of 912 dogs from 85 breeds as well as an extensive sample of 225 grey wolves (the ancestor of the domestic dog<sup>4,5</sup>) from 11 globally distributed populations (Supplementary Table 1). We constructed neighbour-joining trees using individuals and populations as units of analysis based on both individual SNP and haplotype similarity (Fig. 1 and Supplementary Note A). All trees identify dogs as a distinct cluster. Moreover, using as few as 20 diagnostic SNPs, all dog and wolf samples can be correctly assigned to species of origin with high confidence (see Supplementary Table 2, Supplementary Figs 1–4 and Supplementary Note B). Applying the Bayesian clustering

method implemented in STRUCTURE<sup>6</sup>, we found strong evidence for admixture with wolves only in a minority of breeds (Supplementary Fig. 5). Neighbour-joining trees reveal that most of these breeds (basenji, Afghan hound, Samoyed, saluki, Canaan dog, New Guinea singing dog, dingo, chow chow, Chinese Shar Pei, Akita, Alaskan malamute, Siberian husky and American Eskimo dog) are highly divergent from other dog breeds (Fig. 1 and Supplementary Figs 6–11). These highly divergent breeds have been identified previously and termed ‘ancient’ breeds (as opposed to ‘modern’)<sup>4</sup> because, consistent with their high levels of divergence, historical information suggests that most have ancient origins (>500 years ago)<sup>7–9</sup>. The limitation of evidence for admixture to only a few breeds is striking given that backcrossing between dogs and wolves is known to occur<sup>10</sup> and dogs and wolves coexist widely. Given that modern breeds are the products of controlled breeding practices of the Victorian era (circa 1830–1900)<sup>4,7–9</sup>, the lack of detectable admixture with wolves is consistent with the strict breeding regimes recently implemented by humans.

To identify the primary source of genetic diversity for domestic dogs, we used three approaches. First, we assessed whether a single wolf population clustered with dogs in neighbour-joining trees based on allele sharing of SNPs and sharing of 5- and 10-SNP haplotypes for individuals and breed/population groupings (Fig. 1, Supplementary Figs 6–11 and Supplementary Note A). Only in individual SNP and 5-SNP haplotype trees were specific populations of Middle and Near Eastern grey wolves found to be most similar to domestic dogs (Fig. 1b and Supplementary Fig. 10). In all other trees, wolves form a single genetic group and are not informative with regard to the wolf population that is most similar to dogs. We further tested the approach of a previous mitochondrial DNA (mtDNA) sequence study that suggested dogs have an origin in east Asia because diversity was highest in east Asian dog breeds<sup>3,11</sup>. We find that genetic diversity of dogs does not vary with geography in a consistent pattern. Specifically, breeds of east Asian origin do not have the highest level of nuclear variability, even when the SNP discovery scheme differed

<sup>1</sup>Department of Ecology and Evolutionary Biology, 621 Charles E. Young Drive South, University of California, Los Angeles, California 90095, USA. <sup>2</sup>Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, New York 14853-2601, USA. <sup>3</sup>Department of Biostatistics, University of California, Los Angeles, California 14853, USA. <sup>4</sup>Cancer Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. <sup>5</sup>Department of Biomolecular Engineering, University of California, Santa Cruz, California 95064, USA. <sup>6</sup>Department of Zoology, Tel Aviv University, Tel Aviv 69978, Israel. <sup>7</sup>Museum and Institute of Zoology, Polish Academy of Sciences, Wilcza 64, 00-679 Warszawa, Poland. <sup>8</sup>Mammal Research Institute, Polish Academy of Sciences, 17-230 Białowieża, Poland. <sup>9</sup>Instituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA), 40064 Ozzano Emilia (B), Italy. <sup>10</sup>Department of Population Health and Reproduction, School of Veterinary Medicine, University of California, Davis, California 95616, USA. <sup>11</sup>School of Biotechnology and Biomolecular Sciences and Clive and Vera Ramaciotti Center for Gene Function Analysis, University of New South Wales, Sydney NSW 2052, Australia. <sup>12</sup>Faculty of Environmental Design, University of Calgary, 2500 University Drive NW, Calgary, Alberta T2N 1N4, Canada. <sup>13</sup>Affymetrix Corporation, 3420 Central Expressway, Santa Clara, California 95051, USA. <sup>14</sup>The WALTHAM Centre for Pet Nutrition, Waltham on the Worlds, Leicestershire LE14 4RT, UK. <sup>15</sup>Affymetrix Asia Pacific, Scientific Affairs and Collaborations, 1233 Lujiazui Ring Road, AZIA Center, Suite 1508, Shanghai 200120, China. <sup>16</sup>Laboratory for Conservation and Utilization of Bio-resources, Yunnan University, Kunming 650091, China. <sup>17</sup>State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China. <sup>18</sup>Interdepartmental Program in Bioinformatics, 621 Charles E. Young Drive South, University of California, Los Angeles, California 90095, USA.

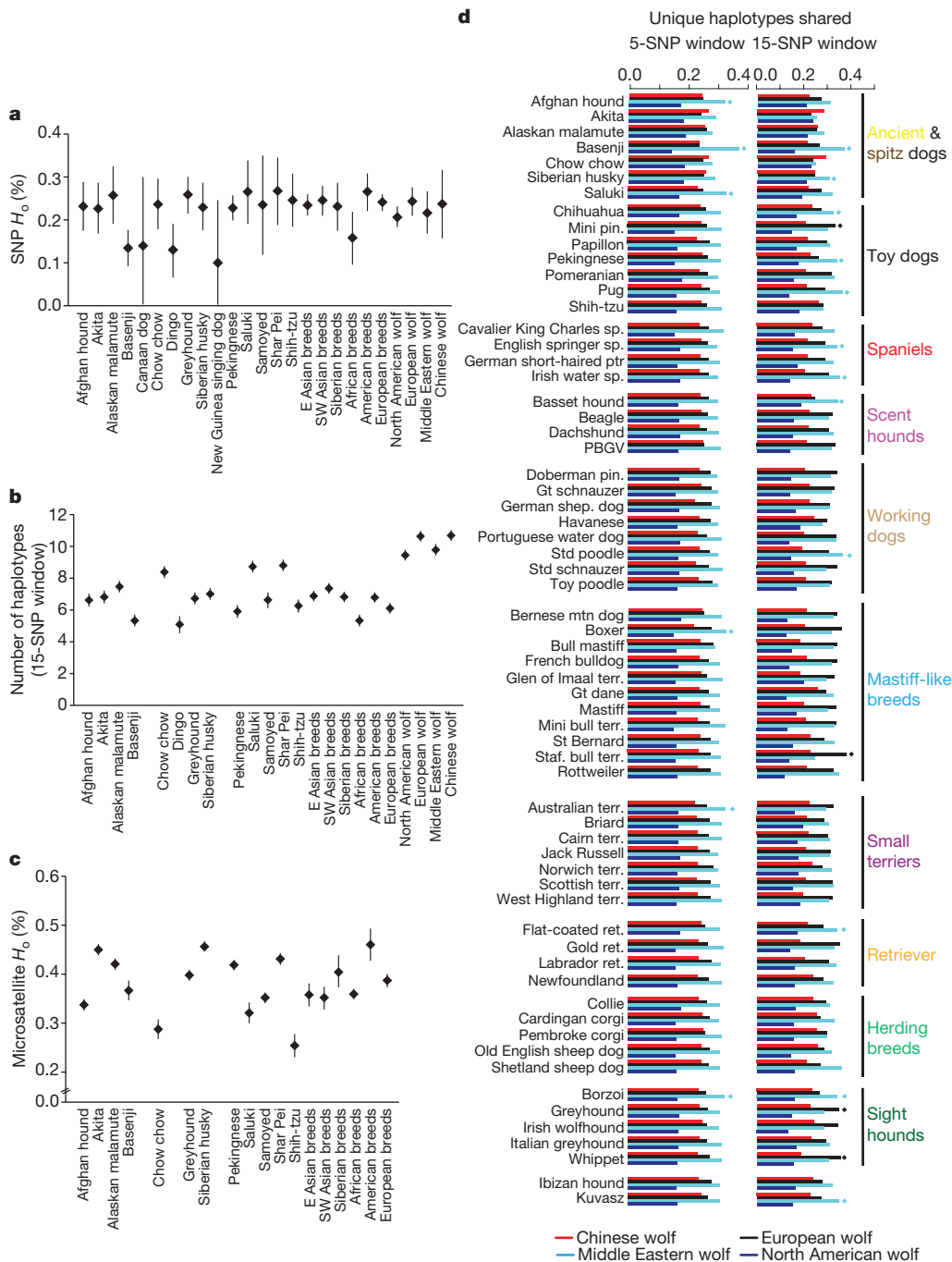


**Figure 1 | Neighbour-joining trees of domestic dogs and grey wolves.** Branch colour indicates the phenotypic/functional designation used by dog breeders<sup>8,9</sup>. A dot indicates  $\geq 95\%$  bootstrap support from 1,000 replicates. **a**, Haplotype-sharing cladogram for 10-SNP windows ( $n = 6$  for each breed and wolf population). **b**, Allele-sharing cladogram of individuals based on individual SNP loci. **c**, Haplotype-sharing phylogram based on 10-SNP windows of breeds and wolf populations. **d**, Allele-sharing phylogram of individual SNPs for breeds and wolf populations. For **c** and **d**, we note breeds where genetic assignments conflict with phenotypic/functional designations as follows: 1, Brussels griffon; 2, Pekingese; 3, pug; 4, Shih-tzu; 5, miniature pinscher; 6, Doberman pinscher; 7, Kuvasz; 8, Ibizian hound; 9, chihuahua; 10, Pomeranian; 11, papillon; 12, Glen of Imaal; 13, German shepherd; 14, Briard; 15, Jack Russell; 16, dachshund; 17, great schnauzer; and 18, standard schnauzer. Gt, great; mtn, mountain; PBGV, petit basset griffon vendéen; pin., pinscher; ptr, pointer; ret., retriever; shep., shepherd; sp., spaniel; Staf., Staffordshire; std, standard; terr., terrier. Canine images not drawn to scale. Wolf image adapted from ref. 31; dog images from the American Kennel Club (<http://www.akc.org>).

or haplotype measures of diversity were used to minimize ascertainment bias<sup>12</sup> (Fig. 2a, b and Supplementary Figs 12a and 13). Furthermore, we confirmed an absence of geographic patterns in nuclear variation through a reanalysis of previously published microsatellite data<sup>4,7</sup> (Fig. 2c and Supplementary Fig. 12b). For example, the two ancient breeds with highest SNP haplotype diversity, saluki and Chinese Shar Pei, originated in widely different areas (the Middle East and China, respectively<sup>8,9</sup>; Fig. 2b). However, ancient and island breeds are exceptions in consistently having lower diversity (basenji, Canaan dog, dingo, New Guinea singing dog; Fig. 2a–c). Thus, in contrast to previous mtDNA sequence results, current levels of autosomal diversity do not support an east Asian origin (or any other location). Indeed, if demographic history has varied substantially in dog breeds across geographic regions after domestication, current levels of genetic diversity may not directly reflect the oldest, ancestral population as it does in other species such as humans<sup>12–14</sup>. In addition, we note that recently, the use of genetic diversity to infer centres of domestication has been questioned by studies of semi-feral village dogs from Africa and Puerto Rico that found levels of mtDNA diversity

as high or higher than those in east Asia<sup>11,15</sup>. High diversity in African dog populations reflects the added contribution of ancient indigenous dogs to the gene pool, which elsewhere is often dominated by modern breeds<sup>15</sup>.

Consequently, as a third approach to determine the primary centre of dog domestication, we considered haplotype sharing of modern and ancient dog breeds with specific wolf populations (see Supplementary Note A). Haplotype diversity patterns have been shown to be less sensitive to ascertainment biases<sup>12</sup>, and the sharing of SNPs that are otherwise private to specific wolf populations provides a unique signal to support ancestry or admixture. We analysed haplotype sharing between 64 well-sampled ( $n \geq 9$ ) dog breeds and wolf populations from Europe, the Middle East and China for 500-kilobase (kb) haplotype windows containing 5 and 15 SNPs drawn at random (Fig. 2d and Supplementary Table 3, Supplementary Note A). The Middle East and China have been implicated as centres for dog origination based on the archaeological record or mtDNA diversity<sup>3,11,16–18</sup>. We also assessed haplotype sharing between dog breeds and North American wolves as a negative



**Figure 2 | Genome-wide analysis of SNP variation in domestic dogs and grey wolves.** **a**, Average observed heterozygosity ( $H_o$ ). **b**, Average number of haplotypes per breed or group for phased SNP loci (15-SNP windows). **c**, Average observed heterozygosity of microsatellite data<sup>4,7</sup>. **d**, Fraction of unique haplotypes shared between 64 dog breeds and wolf populations for 5- (left) and 15- (right) SNP windows. Diamonds indicate significant sharing ( $P < 0.05$ ) using permutation test 2 (Supplementary Note A). Six (a–c) or nine (d) individuals represent each breed and wolf population. Error bars indicate s.e.m. E, east; SW, southwest.

control because dogs did not originate there<sup>19</sup>. Across all breeds, and for both window sizes, levels of sharing between dogs and North American wolves are substantially lower than the analogous comparison with Old World wolves, as expected (Fig. 2d). For 5-SNP haplotype windows, we found that haplotype sharing was uniformly higher between modern dog breeds and Middle Eastern wolves than between other wolf populations (Fig. 2d, left). For 15-SNP windows (Fig. 2d, right), the majority of breeds show the most sharing with Middle Eastern wolves, including some dog breeds of diverse geographic origins (for example, basenji, chihuahua, basset hound and borzoi). Notably, significant sharing with European wolves is found in miniature pinschers, Staffordshire bull terriers, greyhounds and whippets. The increased haplotype sharing between some European breeds and European wolves in the 15-SNP analysis may not be revealed in the 5-SNP windows because the European and Middle Eastern wolf haplotypes are less readily distinguished when based on fewer SNPs. Finally, only two east Asian breeds (Akita and chow

chow) had higher sharing with Chinese wolves, although the results were not significant. In an analysis with fewer chromosomes per breed ( $n \geq 6$ ), four east Asian breeds—the Akita, Chinese Shar Pei, chow chow and dingo—showed most sharing with Chinese wolves (the latter two breeds showing significantly more sharing than expected), corroborating STRUCTURE clustering results (Supplementary Figs 5 and 14).

Notably, in both 5-SNP and 15-SNP window analyses, the basenji, a breed of Middle Eastern origin, had a greater proportion of shared haplotypes with Middle Eastern wolves than any other domestic dog (Fig. 2d and Supplementary Table 3). This result suggests that basenjis had a larger effective population size early in domestication or that they have more recently backcrossed with wolves. Overall, these data implicate the Middle East as a primary source of genetic variation in the dog, with potential secondary sources of variation from Europe and east Asia. In contrast to the mtDNA results, east Asian wolves are a predominant source



of haplotype diversity for only a few east Asian dog breeds that have a long history in that region.

Neighbour-joining trees based on SNP data provide an explicit framework for investigating hypotheses of breed history and the genesis of phenotypic diversity. Consistent with previous microsatellite results<sup>4,7</sup>, topological analyses often define three well-supported groups of highly divergent, ancient breeds: an Asian group (dingo, New Guinea singing dog, chow chow, Akita and Chinese Shar Pei), a Middle Eastern group (Afghan hound and saluki) and a northern group (Alaskan malamute and Siberian husky) as being distinct from modern domestic dogs (Fig. 1a, b and Supplementary Figs 6–11). In addition, we find that the basenji often appears as the most divergent breed in allele- and haplotype-sharing trees (Fig. 1a, b and Supplementary Figs 6–11). This finding and high haplotype sharing, as well as a long recorded history<sup>8,9</sup>, suggest that this breed is one of the most ancient extant dog breeds.

The radiation of modern dog breeds has been difficult to resolve because most have originated recently and lack deep, detailed histories<sup>8,9</sup>. Consequently, the evolutionary process underlying the genesis of phenotypic/functional groupings is obscure. Specifically, many breeds have been documented as originating through crosses of genealogically or geographically distant stocks<sup>9</sup> and thus, parallel evolution and genetic heterogeneity within phenotypic/functional breed groupings is expected. Nonetheless, we discern distinct genetic clusters within modern dogs that largely correspond to those based on phenotype or function, including spaniels, scent hounds, mastiff-like breeds, small terriers, retrievers, herding dogs and sight hounds (see Fig. 1). Most genetic groups have short internodes and often low bootstrap support, reflecting the rapid formation of modern breeds in the Victorian era<sup>8,9</sup>. Notably, toy and working dogs have a more varied relationship to genetic groupings, which is consistent with their known histories involving crosses between breeds from divergent genetic lineages (Supplementary Table 4). The heterogeneous composition of toy breeds may specifically indicate their frequent origin as a cross between a larger dog from a distinct breed grouping and a toy or dwarfed breed (Supplementary Table 4). Finally, within each breed, there is a remarkable concordance with known origin as all dogs are correctly assigned to the breed or population from which they were sampled, with one exception (bull terrier and miniature bull terrier; Fig. 1a, b). The contribution of these groupings to genetic variation was assessed by an analysis of molecular variance (AMOVA; Supplementary Table 5) which showed that 65% of the variation is due to variation within dog breeds, and 31% is due to variation within breed groups, similar to that reported for microsatellite data<sup>4,7</sup>. However, our analysis also showed that 3.8% of the variation is between phenotypic/functional breed groups ( $P < 0.001$ ). Consequently, although most variation is within breeds, phenotypic/functional breed groups represent a relatively small but significant component of variation.

The process of domestication involves strong selection of specific phenotypes; therefore, a signal of this selection should be evident in the genome<sup>20</sup>. Given the genome-wide coverage of our panel of SNPs, we searched for genomic regions that might contain adaptive substitutions due to positive selection during the initial phase of dog domestication (rather than breed formation, see Supplementary Note C). For each SNP, we calculated the fixation index ( $F_{ST}$ ) and cross-population extended haplotype homozygosity (XP-EHH) values between non-admixed wolves and modern dogs and considered SNPs with extreme values as candidates for recent positive selection<sup>20</sup>. These statistics measure population differentiation and relative levels of genetic diversity, both of which are robust indicators of positive selection for recently domesticated species<sup>21</sup>. We found that SNPs within the top 5% of  $F_{ST}$  values and SNPs within the highest 1% of XP-EHH values are each significantly enriched for SNPs in genic regions ( $P = 0.04$  for  $F_{ST}$ ,  $P = 0.02$  for XP-EHH, one-sided exact conditional test, controlling for the ascertainment panel; Supplementary Fig. 15). This result is consistent with a history

of adaptive divergence in genic regions. To identify specific regions that are candidates for recent adaptive evolution, we normalized  $F_{ST}$  and XP-EHH values within ascertainment categories, and targeted regions that have several SNPs with extreme values for both statistics (Supplementary Table 6, see Supplementary Note C for results). Notably, two of our top three signals are near genes that have been implicated in memory formation and/or behavioural sensitization in mouse or human studies (ryanodine receptor 3 (OMIM accession 180903; ref. 22), adenylate cyclase 8 (OMIM accession 103070; Supplementary Note C)). Furthermore, we observed a single SNP with a high  $F_{ST}$  value located near the *WBSCR17* gene responsible for Williams–Beuren syndrome in humans (OMIM accession 194050; Supplementary Fig. 16), which is characterized by social traits such as exceptional gregariousness. These outlier SNPs provide specific candidate regions for fine-scale mapping of genes that are important in the early domestication of dogs.

Our results show domestic dogs have genetic structure on three fundamental levels resulting from distinct evolutionary processes. First, within dog breeds, nearly all dogs are assigned to a breed of origin. This result is supported by previous microsatellite research<sup>4,7</sup> and reflects the limited number of founders, inbreeding and small effective population size characteristic of many breeds<sup>23,24</sup>. Second, breed groupings are evident at a finer scale than previously described, and mirror breed classification based on form and function. We propose that this result reflects the tendency of dog breeders to develop new breeds by crossing individuals within specific functional and phenotypic groups to enhance abilities such as retrieving and herding, or further develop specific morphological traits<sup>4</sup>. However, heterogeneity within toy breeds and other breed groupings suggests the importance of discrete phenotypic mutations in the evolution of phenotypic diversity in the dog. Recent genetic studies have established that variation in coat colour<sup>25</sup> and texture<sup>26</sup>, body size<sup>2</sup>, relative leg length<sup>27</sup> and body proportions (A.R.B. *et al.*, manuscript submitted) in different dog breeds are due to variation in shared genes of large phenotypic effect. For example, at least 19 distinct dog breeds with foreshortened limbs all uniquely share the same retrotransposed version of *Fgf4* that is strongly implicated as the genetic basis for this phenotype<sup>27</sup>. Once such discrete mutations are fixed in a breed they can readily be crossed into unrelated lineages and thus enhance the process of phenotypic diversification. This process has perhaps produced more phenotypic diversity in dogs than other domesticated species because they are selected for many functions of value to humans (for example, defence, herding, retrieving, hunting, speed and companionship) as well as for novelty, which culminated in the ‘fancy breeds’ of the Victorian era<sup>8,9</sup>. Last, we identify divergent lineages of dogs distinct from those breeds that radiated during the nineteenth century and that probably derive from ancient geographically indigenous breeds. This finding mirrors recent genetic discoveries in sheep<sup>28</sup> and cattle<sup>29</sup> and suggests that some canine lineages may have persisted from antiquity or have more recently admixed with wolves. The latter seems unlikely given that some of these breeds have known ancient histories, exist in areas where wolves are absent, and are phenotypically highly derived<sup>8,9</sup>. For example, the chow chow originated more than 2,000 years ago<sup>8,9</sup>. Similarly, the dingo and New Guinea singing dog were probably established over 4,000 years ago and exist in areas without wolves. However, given their close proximity to extensive wolf populations, divergent northern breeds such as the Alaskan malamute, Siberian husky and American Eskimo dog may be better candidates for recent admixture.

Our haplotype sharing analysis evaluates the contribution of specific wolf populations to the genome of dogs, and reveals significant Middle Eastern and, for certain breeds, European ancestry. This result is consistent with the archaeological record that identified the earliest dog remains in the Middle East (12,000 years ago)<sup>16</sup>, Belgium (31,000 years ago)<sup>17</sup>, and the Bryansk region in western Russia (15,000 years ago)<sup>17</sup>, as well as the finding of high mtDNA diversity in ancient Italian dogs<sup>18</sup>. However, some ancient east Asian

breeds show affinity with Chinese wolves, which suggests that they were derived from Chinese wolves or admixed with them after domestication<sup>10</sup>. The domestic dog seems comparable to other domestic species in containing several sources of variation from wild relatives. This dynamic process enriched the dog genome through interbreeding with wolves early in the domestication process. Similarly, mutations that have occurred since domestication, such as the mutation responsible for black coat colour, have been transferred to grey wolves<sup>30</sup>. Our genome-wide SNP analysis provides a new evolutionary framework for understanding the rapid phenotypic diversification unique to the domestic dog.

## METHODS SUMMARY

**SNP genotyping.** Genomic DNA was isolated from blood samples of domestic dogs (*Canis familiaris*,  $n = 912$ ) and from tissue and blood samples of grey wolves (*C. lupus*,  $n = 225$ ) and coyotes (*C. latrans*,  $n = 60$ ; see Supplementary Methods and Supplementary Table 1). The samples were genotyped and quality control filters were applied (see A.R.B. *et al.*, manuscript submitted) to obtain high-quality genotypes from 48,036 autosomal SNP loci.

**Cluster analysis.** To visualize genetic relationships suggested by our SNP data we used principal component analysis (PCA) ( $n_{\text{dog\_breed}} = 2$ ) and STRUCTURE<sup>6</sup> ( $n_{\text{dog\_breed}} = 1$ ). For tree reconstruction, we analysed two data sets. First, for individual-based allele-sharing distance analyses, we used 574 individuals ( $n_{\text{dogs}} = 490$ ;  $n_{\text{Old\_World\_wolves}} = 84$ ). This data set consisted of 75 dog breeds where six individuals were genotyped from each breed and an additional five dog breeds where five or fewer individuals were genotyped. The second data set was created for the population-level and haplotype-sharing distance-based analyses and used a subset of 530 individuals to provide comparable sample sizes from 79 dog breeds ( $n_{\text{per\_breed}} = 6$ ) or wolf populations from China ( $n = 6$ ), Middle East ( $n = 7$ ), central Asia ( $n = 6$ ) and Europe ( $n = 31$ ). Coyotes from the western United States ( $n = 6$ ) were used for rooting.

**SNP haplotype analysis.** From phased genotypes, we divided the genome into 500-kb windows to identify haplotypes and estimated haplotype diversity. The level of haplotype sharing was assessed between a dog breed ( $n_{\text{individuals}} \geq 9$  per breed,  $n_{\text{breeds}} = 64$ ) and each wolf population (China, Europe, Middle East and North America).

**Detecting selection.** Population differentiation and extended haplotype homozygosity test statistics were calculated between modern dog breeds and grey wolves. We identified outlier SNP loci based on normalized scores and ranking in the 95th and 99th percentile.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 1 September 2009; accepted 19 January 2010.

Published online 17 March 2010.

- Liti, G. *et al.* Population genomics of domestic and wild yeasts. *Nature* **458**, 337–341 (2009).
- Sutter, N. *et al.* A single *IGF1* allele is a major determinant of small size in dogs. *Science* **316**, 112–115 (2007).
- Savolainen, P., Zhang, Y. P., Luo, J., Lundeberg, J. & Leitner, T. Genetic evidence for an East Asian origin of domestic dogs. *Science* **298**, 1610–1613 (2002).
- Parker, H. *et al.* Genetic structure of the purebred domestic dog. *Science* **304**, 1160–1164 (2004).
- Vilà, C., Maldonado, J. & Wayne, R. Phylogenetic relationships, evolution, and genetic diversity of the domestic dog. *J. Hered.* **90**, 71–77 (1999).
- Pritchard, J., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
- Parker, H. *et al.* Breed relationships facilitate fine-mapping studies: a 7.8-kb deletion cosegregates with Collie eye anomaly across multiple dog breeds. *Genome Res.* **17**, 1562–1571 (2007).
- Wilcox, B. & Walkowicz, C. *The Atlas of Dog Breeds of the World* 5th edn (TFH Publications, 1995).
- American Kennel Club. *The Complete Dog Book* 20th edn (Ballantine Books, 2006).
- Vilà, C., Seddon, J. & Ellegren, H. Genes of domestic mammals augmented by backcrossing with wild ancestors. *Trends Genet.* **21**, 214–218 (2005).
- Pang, J. *et al.* 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**, 2849–2864 (2009).
- Conrad, D. *et al.* A worldwide survey of haplotypes variation and linkage disequilibrium in the human genome. *Nature Genet.* **38**, 1251–1260 (2006).

- Jakobsson, M. *et al.* Genotype, haplotype and copy-number variation in worldwide human populations. *Nature* **451**, 998–1003 (2008).
- Li, J. *et al.* Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **319**, 1100–1104 (2008).
- Boyko, A. *et al.* Complex population structure in African village dogs and its implication for inferring dog domestication history. *Proc. Natl Acad. Sci. USA* **106**, 13903–13908 (2009).
- Dayan, T. Early domesticated dogs of the Near East. *J. Archaeol. Sci.* **21**, 633–640 (1999).
- Germonpre, M. *et al.* Fossil dogs and wolves from Palaeolithic sites in Belgium, the Ukraine and Russia: osteometry, ancient DNA and stable isotopes. *J. Archaeol. Sci.* **36**, 473–490 (2009).
- Verginelli, F. *et al.* Mitochondrial DNA from prehistoric canids highlights relationships between dogs and South-East European wolves. *Mol. Biol. Evol.* **22**, 2541–2551 (2005).
- Leonard, J. *et al.* Ancient DNA evidence for Old World origin of New World dogs. *Science* **298**, 1613–1616 (2002).
- Sabeti, P. *et al.* Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913–918 (2007).
- Innan, H. & Kim, Y. Pattern of polymorphism after strong artificial selection in a domestication event. *Proc. Natl Acad. Sci. USA* **101**, 10667–10672 (2004).
- Balschun, D. *et al.* Deletion of the ryanodine receptor type 3 (RyR3) impairs forms of synaptic plasticity and spatial learning. *EMBO J.* **18**, 5264–5273 (1999).
- Cruz, F., Vilà, C. & Webster, M. The legacy of domestication: accumulation of deleterious mutations in the dog genome. *Mol. Biol. Evol.* **25**, 2331–2336 (2008).
- Gray, M. *et al.* Linkage disequilibrium and demographic history of wild and domestic canids. *Genetics* **181**, 1493–1505 (2009).
- Candille, S. *et al.* A  $\beta$ -defensin mutation causes black coat color in domestic dogs. *Science* **318**, 1418–1423 (2007).
- Cadiou, E. *et al.* Coat variation in the domestic dog is governed by variants in three genes. *Science* **326**, 150–153 (2009).
- Parker, H. *et al.* An expressed *Fgf4* retrogene is associated with breed-defining chondrodysplasia in domestic dogs. *Science* **325**, 995–998 (2009).
- Pedrosa, S. *et al.* Evidence of three maternal lineages in near eastern sheep supporting multiple domestication events. *Proc. R. Soc. Lond. B* **272**, 2211–2217 (2005).
- Elsik, C. *et al.* The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science* **324**, 528–532 (2009).
- Anderson, T. *et al.* Molecular and evolutionary history of melanism in North American gray wolves. *Science* **323**, 1339–1343 (2009).
- Macdonald, D. W. & Barrett, P. *Mammals of Britain and Europe* (Collins, 1993).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** Grants from NSF and NIH (R.K.W.; C.D.B. and J.N.), the Polish Ministry of Science and Higher Education (M.P. and W.J.), European Nature Heritage Fund EURONATUR (W.J.), National Basic Research Program of China (Y.-p.Z.), and Chinese Academy of Sciences (Y.-p.Z.) supported this research. J.N. was supported by the Searle Scholars Program. B.M.vH. was supported by a NIH Training Grant in Genomic Analysis and Interpretation. K.E.L. was supported by a NSF Graduate Research Fellowship. E.A.O., D.S.M., T.C.S., A.E. and H.G.P. are supported by the intramural program of the National Human Genome Research Institute. M.P. was supported by the Foundation for Polish Science. Wolf samples from central and eastern Europe and Turkey were collected as a result of a continuing project on genetic differentiation in Eurasian wolves. We thank the project participants (B. Jedrzejewska, V. E. Sidorovich, M. Shkvyrya, I. Dikiy, E. Tsingarskaya and S. Nowak) for their permission to use 72 samples for this study. We acknowledge R. Hefner and the Zoological collection at Tel Aviv University for Israeli wolf samples. We thank the American Kennel Club (AKC) for the dog images reproduced in Fig. 1. We also gratefully acknowledge the dog owners who generously provided samples, the AKC Canine Health Foundation, and Affymetrix Corporation. We thank B. Van Valkenburgh, K.-P. Koepfli, D. Stahler and D. Smith for reviewing the manuscript.

**Author Contributions** Samples were contributed by E.G., M.P., W.J., C.G., E.R., D.B., A.W., J.S., M.M., E.A.O. and R.K.W. The experiment was designed and carried out with the help of B.M.vH., J.P.P., H.G.P., P.Q., D.S.M., T.C.S., A.E., A.W., J.S., M.C., P.G.J., Z.Q., W.H., Z.-L.D., Y.-p.Z., C.D.B., E.A.O. and R.K.W. The genotyping program was written by A.R.B., A.A., A.R., K.B., A.B. and C.D.B. and further programming was completed by K.E.L., J.D.D., D.A.E., E.H. and J.N. The analyses were conducted by B.M.vH., J.P.P., K.E.L., E.H., H.G.P., J.D.D., A.R.B., D.A.E., A.A., A.R., J.C.K. and J.N. The manuscript was written by B.M.vH., K.E.L., C.D.B., E.A.O., J.N. and R.K.W.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.K.W. ([rwayne@ucla.edu](mailto:rwayne@ucla.edu)).

## METHODS

**The CanMap sample collection.** Our study is part of the CanMap project (A.R.B. *et al.*, manuscript submitted), which isolated genomic DNA from blood samples collected from domestic dogs (*C. familiaris*,  $n = 912$ ) and from tissue and blood samples from grey wolves (*C. lupus*,  $n = 225$ ), coyotes (*C. latrans*,  $n = 60$ ), putative dog–wolf hybrids ( $n = 17$ ), red wolves (*C. rufus*,  $n = 12$ ), Mexican wolves (*C. l. baileyi*,  $n = 10$ ), Ethiopian wolves (*C. simensis*,  $n = 4$ ), black-backed jackals (*C. mesomelas*,  $n = 6$ ), golden jackals (*C. aureus*,  $n = 2$ ), and a side-striped jackal (*C. adustus*,  $n = 1$ ; Supplementary Table 1). Domestic dog samples were obtained through American Kennel Club (AKC) sanctioned dog shows, speciality events, breed clubs, and veterinary clinics. Three-to-twelve dogs from each breed from each of 81 AKC-recognized breeds and four semi-domestic lineages (Africanis, Canaan dog, dingo, and New Guinea singing dog) were used in the analysis. Specifically, the semi-domestic dingo and New Guinea singing dog are ancient breeds that were probably established more than 4,000 years ago and have existed in isolation from wolves<sup>32</sup>.

The Affymetrix Canine Mapping Array version 2 contains SNPs that were ascertained by aligning sequence reads to the boxer genome assembly (CamFam2). A large number of SNPs were discovered as heterozygous sites in the boxer genome (here denoted as boxer  $\times$  boxer SNPs), and further SNPs were found by aligning sequence reads from other breeds or wild canids to the boxer genome. These extra SNPs can be categorized by the sequence aligned to the boxer as follows: (1) the standard poodle (CanFam1); (2) one of nine dog breeds; (3) one of four wolf populations (Alaskan, Chinese, Indian or Spanish wolves); and (4) coyote (see Supplementary Table 7).

**Breed groupings.** Several analyses are based on specific dog breed groupings for comparison purposes (Supplementary Table 1). We define ancient breeds as those that are divergent genetically (Fig. 1), corroborated by previous genetic studies<sup>33</sup> and, in most cases, are known to have originated in ancient cultures more than 500 years ago<sup>8,9</sup>. Furthermore, we used previously defined geographical breed groupings<sup>3</sup> (Africa, America, east Asia, Europe, Siberia, southeast Asia and southwest Asia breed groups) and functional and phenotypic breed groupings in common usage by dog breeders<sup>8,9,33</sup> (ancient, spitz, toy, spaniels, scent hounds, working dogs, mastiff-like breeds, small terriers, retrievers, herding and sight hounds).

**Identification of recently related individuals in the sample.** We used PLINK<sup>34</sup> to obtain pairwise estimates of identity by state (IBS). From the Yellowstone National Park wolves in the data set ( $n = 19$ ), known pedigree relationships were used to calibrate IBS scores<sup>35</sup>. A minimum score of IBS  $> 0.8$  indicated a relatedness status of half-siblings, and values below this level were used to identify a set of unrelated wild canids.

**Single-SNP measures of genetic diversity.** Single-marker descriptive statistics (observed/expected heterozygosity and polymorphism) were estimated using PLINK<sup>34</sup> for the complete SNP data set (Fig. 2). Observed heterozygosity was also estimated using only SNPs ascertained from the grey wolf and boxer sequence comparisons (Supplementary Fig. 12). We used microsatellite genotype data from a previous study<sup>4</sup> for an independent comparison of observed heterozygosity from loci with different mutational properties and ascertainment schemes (Fig. 2 and Supplementary Fig. 13).

**STRUCTURE analysis.** We used the Bayesian inference program STRUCTURE<sup>36</sup> to assess genetic partitions and admixture for the 43,953-SNP data set (linkage disequilibrium (LD) pruned,  $r^2 < 0.5$ ). We used 5,000 burn-in iterations and 15,000 Markov chain Monte Carlo (MCMC) iterations in STRUCTURE, with three repetitions of these parameter settings. The alpha and likelihood statistics were verified to reach convergence before 5,000 burn-in and 15,000 MCMC iterations were completed during each repetition for each number of assumed populations analysed. We analysed domestic dogs and Old World wolves to resolve the ancestry of domestic dogs; hence, we included only one dog per breed for the analysis ( $n = 85$ ; Supplementary Fig. 5). We also included only Old World wolf populations because they may be closely related to the direct ancestors of domestic dogs and we used unrelated individuals from IBS estimates. We sampled China ( $n = 9$ ), central Asia ( $n = 3$ ), the Middle East ( $n = 7$ ) and Europe ( $n = 43$ ). We excluded wolves from highly inbred populations (Italy, Spain and Sweden)<sup>24</sup> to avoid their early partitioning in the analysis. No dog–wolf hybrids were found in the full sample of modern breeds ( $n = 801$ ) as determined with the program smartpca in the Eigensoft package<sup>37</sup>.

From the dog–wolf PCA (see Supplementary Note B) we identified 20 SNPs with the highest loadings on PC1 as input for an additional STRUCTURE analysis to determine the posterior probability of assignment for dogs and wolves to their corresponding species (Supplementary Table 2). Results were plotted using the circular visualization program CIRCOS (<http://mkweb.bcgsc.ca/circos/>; Supplementary Fig. 5).

After the initial partitioning of modern domestic dogs from wild canids for  $K = 2$  (Supplementary Fig. 5), ancient breeds are separated from other canids when a third population is assumed ( $K = 3$ ). Our results show uniquely that Canaan dog, dingo, New Guinea singing dog, and Alaskan Eskimo dog are members of this cluster of ancient breeds, and confirm previous results showing basenji, Afghan hound, samoyed, Saluki, Canaan dog, chow chow, Chinese Shar Pei, Akita, Alaska malamute, and Siberian husky belong in that group<sup>3,4,8,9</sup>.

**Analysis of molecular variance.** The IBS matrix was put into ARLEQUIN v3 to analyse molecular variance with 10,098 permutations for significance testing<sup>38</sup>. We defined three different analysis groups (see Fig. 1 and Supplementary Table 5): (1) breed groups in Fig. 1; (2) geographical dog breed groups (Supplementary Table 1); and (3) wolves and dogs as separate groups.

32. Savolainen, P., Leitner, T., Wilton, A., Matisoo-Smith, E. & Lundeberg, J. 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 (2004).
33. Sablin, M. & Khlopachev, G. The earliest Ice Age dogs: evidence from Eliseevichi I. *Curr. Anthropol.* **43**, 795–799 (2002).
34. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
35. vonHoldt, B. *et al.* The genealogy and genetic viability of reintroduced Yellowstone grey wolves. *Mol. Ecol.* **17**, 252–274 (2008).
36. Pritchard, J., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
37. Price, A. *et al.* Principle components analysis corrects for stratification in genome-wide association studies. *Nature Genet.* **38**, 904–909 (2006).
38. Schneider, S., Roessli, D. & Excoffier, L. Arlequin: a software for population genetics data analysis v.2.000 (Genetics and Biometry Lab, Department of Anthropology, University of Geneva, 2000).