Historic dog Furs Unravel the Origin and Artificial Selection of Modern Nordic Lapphund and Elkhound dog Breeds

Shi-Zhi Wang,[1,2,†] Yu Yan,[3,†] Malin Widlund,[4] Chen-Chang Qian,[1,5] Liang-Liang Zhang,[6] Shao-Jie Zhang,[1,5] Zi-Mai Li,[7] Peng Cao,[8] Qing-Yan Dai,[8] Xiao-Tian Feng,[8] Feng Liu,[8] Lu Wang,[3] Chao Gao,[7] Qiao-Mei Fu ,[8] Marjo K Hytönen,[9,10] Hannes Lohi,[9,10] Peter Savolainen,[4,*] and Guo-Dong Wang [1,5,*]

[1]Key Laboratory of Genetic Evolution & Animal Models, Yunnan Key Laboratory of Molecular Biology of Domestic Animals, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China

[2]Yunnan Animal Science and Veterinary Institute, Kunming, China

[3]State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan, Yunnan University, Kunming, China

[4]Department of Gene Technology, KTH Royal Institute of Technology, Science for Life Laboratory, Stockholm, Sweden

[5]Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming, China

[6]Xingzhi College, Zhejiang Normal University, Jinhua, China

[7]State Key Laboratory of Microbial Technology, Shandong University, Qingdao, China

[8]Key Laboratory of Vertebrate Evolution and Human Origins, Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, Beijing, China

[9]Department of Medical and Clinical Genetics, Department of Veterinary Biosciences, University of Helsinki, Helsinki, Finland

[10]Folkhälsan Research Center, Helsinki, Finland

[†]These authors contributed equally to this work.

*Corresponding authors: E-mails: wanggd@mail.kiz.ac.cn; peter.savolainen@scilifelab.se.

Associate editor: Evelyne Heyer

Abstract

The origins and extreme morphological evolution of the modern dog breeds are poorly studied because the founder populations are extinct. Here, we analyse eight 100 to 200 years old dog fur samples obtained from traditional North Swedish clothing, to explore the origin and artificial selection of the modern Nordic Lapphund and Elkhound dog breeds. Population genomic analysis confirmed the Lapphund and Elkhound breeds to originate from the local dog population, and showed a distinct decrease in genetic diversity in agreement with intense breeding. We identified eleven genes under positive selection during the breed development. In particular, the *MSRB3* gene, associated with breed-related ear morphology, was selected in all Lapphund and Elkhound breeds, and functional assays showed that a SNP mutation in the 3′UTR region suppresses its expression through miRNA regulation. Our findings demonstrate analysis of near-modern dog artifacts as an effective tool for interpreting the origin and artificial selection of the modern dog breeds.

Key words: historic dog fur, lapphund, elkhound, genetic origin, artificial selection, population genomics.

Introduction

There are today more than 450 dog breeds with a plethora of morphologies, which makes the dog the most phenotypically diverse mammal (Plassais et al. 2019). Most dog breeds evolved during the last 200 years, through intense selective breeding resulting in an unparalleled burst of rapid evolution (Akey et al. 2010). However, while the ancient origins and evolution of the domestic dog is increasingly explored based on studies of both modern and ancient samples (Vilà et al. 1997, Savolainen et al. 2002, Boyko et al. 2009, Pang et al. 2009, Freedman et al. 2014, Shannon et al. 2015, Frantz et al. 2016, Wang et al. 2016, Bergström et al. 2020, Sinding et al. 2020, Feuerborn et al. 2021, Bergström et al. 2022), the origins of the modern dog breeds, and the rapid evolution of their diverse morphologies and correlated inbreeding related disorders, remain to be thoroughly studied. The reason for this is that the founder populations of the pedigree dog breeds, i.e. the local populations of indigenous non-breed dogs, are mostly extinct, implying that there is no reference population to base studies of geographical origin and genetic development upon.

Received: January 31, 2024. **Revised:** May 28, 2024. **Accepted:** May 31, 2024

Animal skins from numerous species were extensively collected by museums during the last few hundred years. These samples have provided a source for historic DNA (hDNA) which has been used to uncover the origin and evolution of several extinct populations and species (Hung et al. 2014; van der Valk et al. 2019; Wang et al. 2019; Raxworthy and Smith 2021), but the indigenous non-breed population of European dogs is almost absent in the museum collections. However, in the 19th century, clothes made of furs from local dogs were commonly manufactured in northern Sweden, among both the Sámi reindeer herding people and the Swedish-speaking rural inhabitants, offering a unique source of DNA from a historic non-breed dog population. The dog furs were obtained from the indigenous non-breed dogs of the region, typically used for reindeer herding or hunting, from which the modern Swedish Lapphund (SL) and Swedish Elkhound (SE) breeds were developed starting 130 years ago (Svanberg and Lindin 1997). This offers the possibility to directly compare the modern Lapphund and Elkhound breeds with their source population, which enables the first detailed study of the evolution of a dog breed from its source population of non-breed dogs. In this study, we extracted hDNA from eight dog fur samples obtained from traditional North Swedish clothing manufactured 100 to 200 years ago. Based on this unique source of DNA material, we investigated the genetic origin and diversity of the modern Lapphund and Elkhound breeds, and we identified genes positively selected during the breed development and inspected these for their possible relevance for morphology and disease, achieving a unique view into the genomic mechanisms behind the origin and evolution of modern dog breeds.

Results

Sample Collection and Sequencing

Historic DNA was extracted from eight pieces of dog fur obtained from North Swedish traditional clothing and a sleeping bag. The genomes were sequenced to a coverage of 8.56 to 12.97. The nucleotide misconception of C to T substitution at the 1st bp of the 5′-end of sequencing reads ranged from 0.007 to 0.013 and of G to A substitution at the 1st bp of the 3′-end of sequencing reads ranged from 0.007 to 0.016 among the historic fur samples (Table 1). This indicates that the degree of DNA damage after the repair by uracil-DNA-glycosylase (UDG) from *E. coli* and endonuclease (*Endo* VIII) during the process of library construction is suitable for genomic analysis using standard methods. The dog fur objects were manufactured approximately 100 to 200 years ago, before the pedigree breeding of today's Swedish Lapphund and Swedish Elkhound breeds was initiated, allowing a direct comparison of the source population of indigenous non-breed dogs with the modern breeds. We compared the genome sequence data from the eight museum samples to publicly available data from modern samples of Swedish Lapphund

Table 1 Background information about the eight historic dog fur samples sequenced in this study

Sample	Object	Museum	Date of fabrication	Sequencing reads	Average length (bp)	Pairwise mapped	Frequency of deaminated cytosine residues at the 1st bp of sequencing reads		Genome coverage
							C -> T on 5′-end	G -> A on 3′-end	
AJ2810	Coat	Ajtte museum (Jokkmokk)	Early 20th century	568,532,880	68	97.97%	0.007	0.007	9.92
SM11850	Glove	Skellefteå museum	na	599,967,803	74	96.19%	0.011	0.011	10.50
SM16271	Glove	Skellefteå museum	1920s	711,167,861	58	97.37%	0.010	0.011	10.73
Ly1	Sleeping bag	Lycksele museum	na	492,112,479	46	97.49%	0.009	0.010	8.82
Ly2	Sleeping bag	Lycksele museum	na	614,294,546	44	93.27%	0.012	0.016	9.17
Ly3	Sleeping bag	Lycksele museum	na	488,627,552	48	98.62%	0.010	0.015	8.56
JM1711	Pelt	Ajtte museum (Jokkmokk)	Early 20th century	659,323,743	82	98.34%	0.007	0.007	12.97
SMOKAND1	na	Skellefteå museum	na	715,057,313	76	98.09%	0.013	0.013	12.89

(SL) and Swedish Elkhound (SE) and to samples of the closely related Lapphund breeds Finnish Lapphund (FL) and Lapponian Herder (LH) and the Elkhound breed Norwegian Elkhound (NE), four to six dogs from each breed, from the study Wang et al. (2016) and Meadows et al. (2023). These samples were analyzed in the context of data from 30 dogs of other breeds from across Europe, Asia, and America, from 10 Southeast Asian indigenous non-breed dogs, and from 15 European wolves and one dhole (for more details see supplementary table S1, Supplementary Material online). Before formal bioinformatic analysis, genomic principal component analysis and calculation of pairwise genomic relationships were performed to inspect that the samples were obtained from different individuals without close relationship. The genomic principal component analysis of the eight dog fur samples showed no duplicated genotypes (supplementary fig. S1, Supplementary Material online), even though samples Ly1-Ly3 were extracted from the same object (a sleeping bag), and pairwise genomic relationships calculated using GCTA (Yang et al. 2011) gave similarities far below 0.125 (supplementary table S2, Supplementary Material online), showing the samples not to be obtained from closely related individuals.

Genetic Origin of dog Breeds

Phylogenetic analysis based on autosomal genomes was performed to investigate whether the modern Lapphund and Elkhound breeds were developed from the historic population of indigenous dogs in northern Scandinavia and Finland (Fig. 1a for more details, see supplementary fig. S2, Supplementary Material online). The phylogenetic results showed that the modern Lapphund and Elkhound breeds and the historic fur samples group together, separated from all other dog samples. The fur samples form a separate clade basal to the modern breeds, which in turn separate into one clade with the three Lapphund breeds (SL, FL, and LH) and another clade with the two Elkhound breeds (SE and NE). This indicates that the modern Lapphund and Elkhound breeds originate from the historic population of indigenous non-breed dogs represented by the fur samples. Furthermore, directly basal to the clade of the fur samples and the Lapphund and Elkhound breeds is the clade of European breeds (e.g. Border Collie, French Bulldog, and German Spitz), while arctic breeds from Siberia (East Siberian Laika and Siberian Husky) and North America (Alaskan Malamute and Greenland Dog) are more distant. This indicates a genetic ancestry for the modern Lapphund and Elkhound breeds primarily from European dogs rather than from the morphologically more similar Siberian dogs.

To test the potential genetic origin of each of the Lapphund and Elkhound samples, outgroup-f3 statistics for the Lapphund and Elkhound samples compared to all other dog samples and the fur samples were performed using dhole as outgroup (Fig. 1b; for more details, see supplementary table S3, Supplementary Material online). This analysis showed that outgroup-f3 values were highest compared to four of the eight fur samples (Ly1, Ly2, Ly3, and SM16271), followed by the other four fur samples and a Spanish Greyhound and a French Bulldog, and then decreasingly by other European breeds, breeds from the Middle East and Siberia, and finally Southeast Asian indigenous non-breed dogs, as illustrated in Fig. 1b for one SL (sample SL_01). Thus, in agreement with the phylogenetic analysis, the outgroup-f3 statistics indicated that the modern Lapphund and Elkhound breeds originate from the historic population of indigenous non-breed dogs represented by the fur samples and that modern Lapphund and Elkhound breeds have a genetic origin primarily from European rather than Siberian dogs.

We also performed a phylogenetic analysis based on mitochondrial genomes, analysing the fur samples in the context of 169 dogs and eight wolves from across the Old World, from Pang et al. (2009) (Fig. 2; for more details,

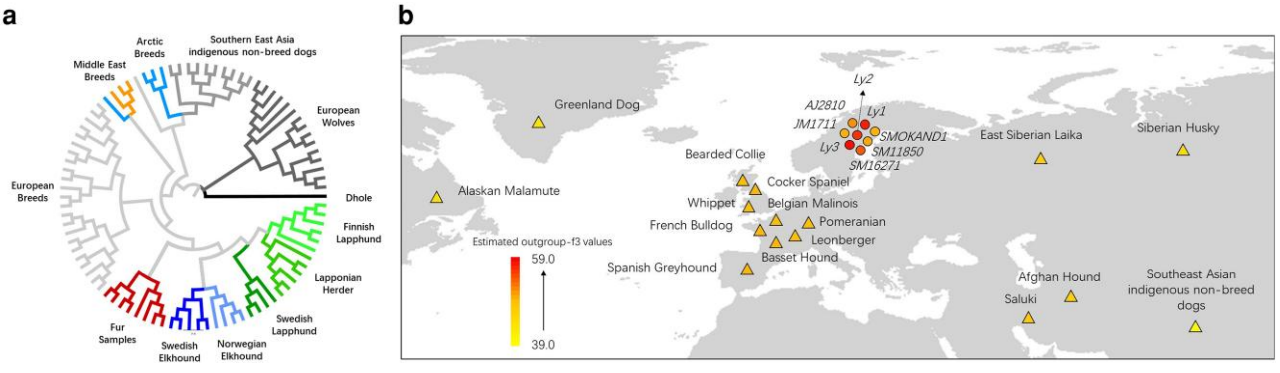

Fig. 1. Origin of Lapphund and Elkhound breeds. (a) Phylogenetic analysis of historic dog fur samples, Lapphund and Elkhound breed dogs, breed dogs from Europe, Asia and America, Southeast Asian indigenous non-breed dogs, European wolves and Dhole (for further details about samples, see supplementary table S1, Supplementary Material). (b) Genetic relationship of the Lapphund and Elkhound breed dogs, compared to the eight historic dog fur samples (AJ2810, SM11850, SM16271, Ly1, Ly2, Ly3, JM1711, SMOKAND1) and all other dogs, using dhole as outgroup. Outgroup-f3 values are shown for one Swedish Lapphund (SL_01). For complete results, see supplementary table S2, Supplementary Material online.

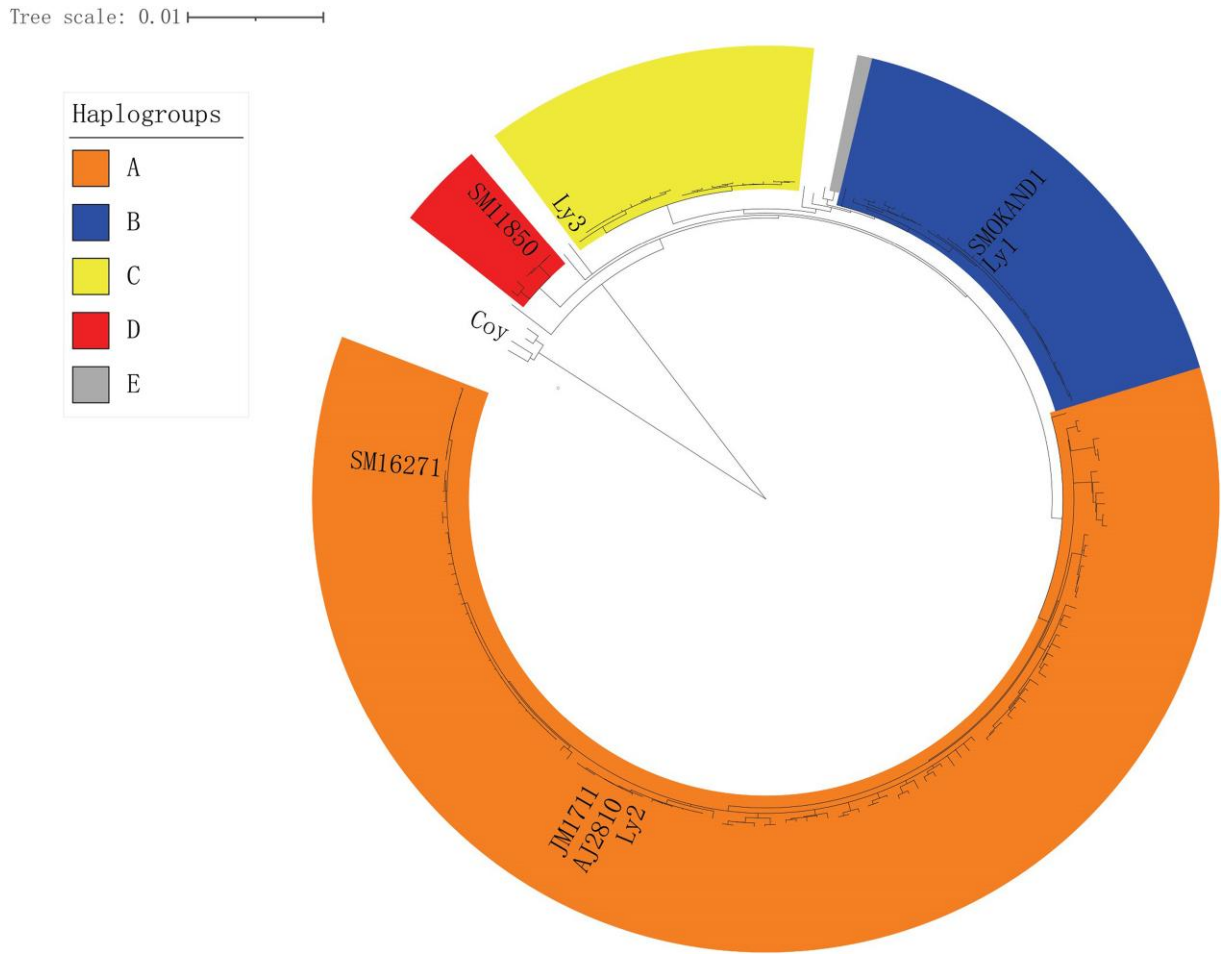

Fig. 2. Phylogenetic analysis of mitochondrial genomes. A neighbour-joining tree based on the mitochondrial genomes from the eight historic dog fur samples and from 169 domestic dogs and eight wolves from across the Old World, with four coyotes as the outgroup. The eight fur samples are indicated by their respective sample name. The five major dog haplogroups A–E are indicated by colours. Branches without colour are wolf and coyote, the four coyotes indicated by Coy. A maximum likelihood tree with sample names for all samples and bootstrap values is shown in supplementary fig. S3, Supplementary Material online.

see supplementary fig. S3, Supplementary Material online). This showed one of the eight fur samples (SM11850) to have a haplotype belonging to the d1 sub-haplogroup, which is carried by more than 50% of the individuals among the modern Lapphund and Elkhound breeds (Klütsch et al. 2011). Since this haplogroup has been found almost exclusively among these Scandinavian and Finnish dog breeds, it is assumed to originate from a local wolf-dog crossbreeding a few hundred or thousand years ago. The other seven fur samples had haplotypes typical for European breed dogs, six of which are typical also for modern Lapphund and Elkhound.

Gene Flow With European Wolves

A *D*-statistic test was performed to compare the fur samples and the Lapphund and Elkhound breeds with three European wolf populations (wolves from Portugal, European Russia, and from the modern Swedish population which most probably originates from recent migration from Finland or Russia) (Fig. 3; for more details see supplementary table S4, Supplementary Material online). Compared to wolves from European Russia, half of both the fur samples (Ly1, Ly2, Ly3, and SM16271) and the Lapphund and Elkhound samples (FL (5 out of 6 samples), SL (3/4), LH (2/6), NE (2/5), and SE (1/4)) had a *Z*-score below −3.00, indicative of gene flow. Similar *z*-score values were obtained compared to Portuguese wolves, while compared to the Swedish wolves, only one fur sample (Ly3) had a *Z*-score below −3.00. Notably, the four fur samples indicated to have had gene flow with wolves were the same four samples indicated to be most closely related to the Lapphund and Elkhound samples according to the outgroup-f3 statistics. To further investigate gene flow with the wolf population we performed TreeMix analysis based on two scenarios, grouping the eight historic fur samples either as a single group, or as two groups of four fur samples each, with and without indication of gene flow with European wolves according to the *D*-statistic (Fig. 3). However, neither of the scenarios indicate gene flow between the fur sample populations and European wolves, based on the optimal number of migrations, while the second scenario indicates gene flow between wolves

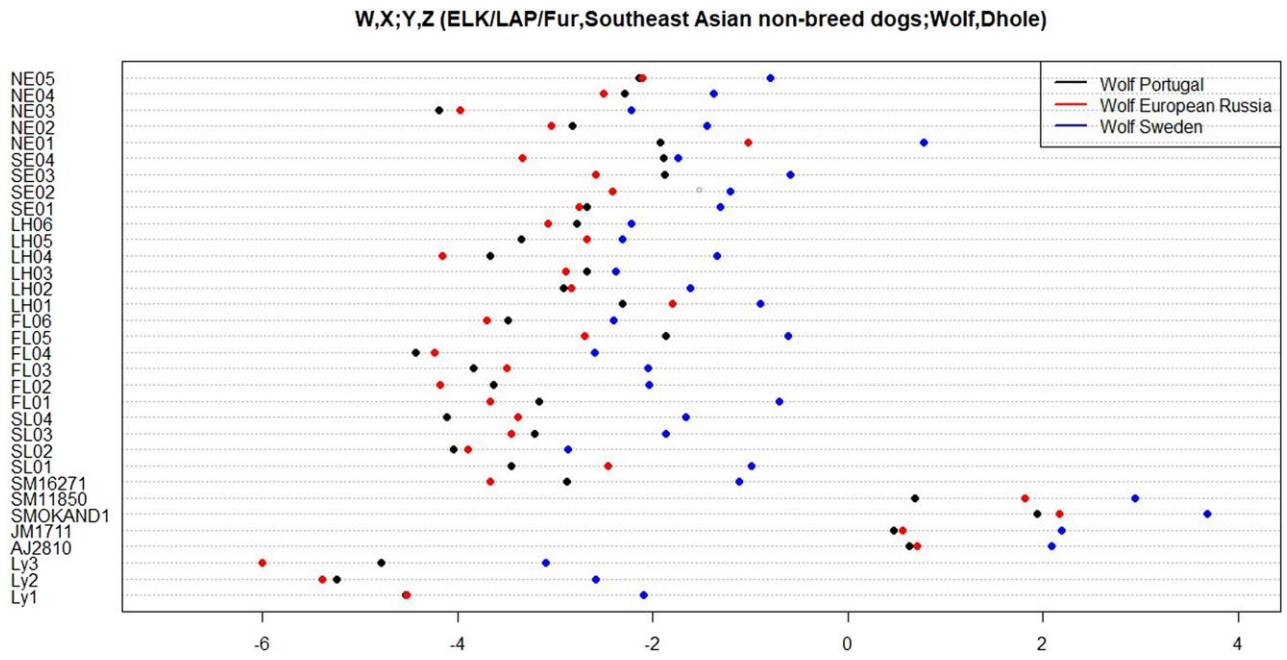

Fig. 3. Gene flow with the wolf. *Z* scores estimated by a *D*-statistic test for the Lapphund and Elkhound breed dogs (NE, Norwegian Elkhound; SE, Swedish Elkhound; LH, Lapponian Herder; FL, Finnish Lapphund and SL, Swedish Lapphund) and the eight historic dog fur samples (AJ2810, SM11850, SM16271, Ly1, Ly2, Ly3, JM1711, SMOKAND1), compared to the three European wolf populations.

and dog breeds from the Middle East (supplementary fig. S4, Supplementary Material online). Furthermore, the proportion of genomic components in the modern Lapphund and Elkhound samples coming from wolves but not present in the historic dog fur samples was estimated by the qpAdm method, showing that the proportion of genomic components from the Swedish wolves in each Lapphund and Elkhound sample ranged from 0.00 to 0.12 (supplementary table S5, Supplementary Material online), and that Lapphund samples generally had a larger wolf component than the Elkhound samples.

Decreased Genetic Diversity

To investigate how the formation and further selective breeding of the modern breeds from the population of indigenous non-breed dogs have influenced genetic diversity, the inbreeding coefficient based on runs of homozygosity (F_{ROH}) and the linkage disequilibrium decay (LD decay) were compared between the five modern Lapphund and Elkhound breeds, the historic fur samples and the Southeast Asian indigenous non-breed dogs. The average of F_{ROH} in modern Lapphund and Elkhound breeds ranged from 0.09 to 0.25, which was considerably higher than for the fur samples (0.02) and the Southeast Asian indigenous non-breed dogs (0.01), and the rates of LD decay for each modern Lapphund and Elkhound breed were all lower than the rates among the fur samples and Southeast Asian indigenous non-breed dogs (supplementary fig. S5, Supplementary Material online). Thus, both analyses indicated a fast decrease in the genetic diversity during the process of breed development of the modern Lapphund and Elkhound breeds (Lindblad-Toh et al. 2005).

Genes Under Positive Selection

To identify genes under artificial selection during the development of the Lapphund and Elkhound breeds from the local non-breed population, genome-wide scans for selection signatures were performed using the XP-EHH and Fst methods. The group of fur samples was regarded as a baseline for the Lapphund breed group and the Elkhound breed group, respectively, to calculate XP-EHH and Fst values based on 50 kb sliding windows. Overlaps of the Top 1% windows between the two methods were considered genomic regions under strong positive selection (for more details, see supplementary table S6, Supplementary Material online). Based on this, eleven genes with nonsynonymous SNP mutations were identified (Table 2), two of which, *MSRB3* and *MGAM*, have been reported to be associated with phenotype in dogs. *MSRB3* is associated with the phenotype of drop/prick ear in dogs (Vaysse et al. 2011; Plassais et al. 2019) as well as with ear shape in sheep (Paris et al. 2020) and ear size in pigs (Che et al. 2018). *MSRB3* gene was under positive selection in both breed groups, in line with the breed standard for the prick ear phenotype for all modern Lapphund and Elkhound breeds giving a strong selection during breed development. The *MGAM* gene has been reported to be associated with improved starch digestion during domestication from wolf to dog (Axelsson et al. 2013; Wang et al. 2013; Zhou et al. 2022) and adaptation to the arctic environment in dogs (Sinding et al. 2020), and it is related to Glycogen Storage Disease II (GSDII) in humans (Zhang et al. 2006), which is also a breed-specific disorder in modern Lapphund breeds (Seppälä et al. 2013). In line with this, we found that *MGAM* was under positive

Table 2 Nonsynonymous SNP mutations in protein-coding regions and corresponding genes in genomic regions under strong positive selection were identified in the modern Lapphund and Elkhound breed groups through comparison with the historic fur samples. (Breed group: LAP, Lapphund; ELK, Elkhound)

Breed group	Chromosome	Position of SNP	50 K bp-window based values		Gene
			Fst	XP-EHH	
LAP	1	72,686,026	0.419833	0.707071	*FBP1*
	9	44,061,937	0.369166	0.868852	*EFCAB5*
	9	44,062,086			
	9	44,062,146			
	9	44,098,687	0.369166	0.868852	*ENSCAFG00805006392*
	10	8,394,629	0.619465	0.59375	*MSRB3*
	10	8,394,668			
	10	8,394,697			
	17	54,064,250	0.346586	0.599078	*SLC22A15*
	19	1,664,103	0.345019	0.944954	*INPP4B*
	25	31,936,485	0.339326	0.733333	*ENSCAFG00805023321*
	25	31,936,495			
	25	31,936,562			
	25	31,936,573			
	25	31,936,877			
	25	31,937,113			
	25	31,937,704			
	25	31,937,808			
ELK	10	8,394,629	0.419657	0.693878	*MSRB3*
	10	8,394,668			
	10	8,394,697			
	16	6,857,878	0.408825	0.760563	*MGAM*
	16	6,886,346	0.408825	0.760563	*MGAM2*
	16	6,887,268			
	16	6,887,275			
	16	11,977,025	0.518138	0.84466	*MTPN*
	17	38,492,856	0.424649	0.939024	*THNSL2*
	17	38,494,462			

selection in the Elkhound breed group but not in the Lapphund breed group.

Functional Assays of the *MSRB3* Gene

The *MSRB3* gene was identified to have three SNP variants, two of which were nonsynonymous variants located in the protein-coding region (chr10:8394629, A > G and chr10:8394668, G > A) and one located in either the protein-coding or the 3′ UTR region (chr10:8394697, T > C) based on different MSRB3 transcripts. To further verify the potential molecular function of these three SNP variants, an enzyme activity experiment and a dual luciferase assay was performed. These experiments showed that the two nonsynonymous SNP variants in the protein-coding region did not affect the enzyme activity (supplementary fig. S6, Supplementary Material online), while the third SNP variant (located in the 3′ UTR or protein coding region) suppressed the expression level of *MSRB3* gene through miRNA regulation of cfa-miR-127 (Fig. 4 and supplementary fig. S7, Supplementary Material online).

Discussion

The development of the unparalleled morphological diversity among the modern dog breeds during a period of less than 200 years represents a unique event of rapid evolution. However, the genomics behind this important historic stage is poorly studied, because the source populations for the dog breeds are unknown or extinct which leaves the genomic status at the starting point unknown. An alternative is to analyze historic samples of the non-breed population from the time of the start of breeding, but cultural objects made from dog tissue are scarce in Europe. However, in the harsh climate of northern Scandinavia, furs made from dogs were extensively used for clothing up to 100 years ago. To study the genomics behind the origin of the Lapphund and Elkhound dog breeds, we obtained eight museum samples from dog fur objects originating from the putative time and place of origin for these breeds.

Phylogenetic and outgroup-f3 statistics confirmed that today's Lapphund and Elkhound breeds are derived from the local non-breed population represented by the eight fur samples (Fig. 1), allowing further analysis of the genomic evolution of the breeds. The analyses also show that the modern Lapphund and Elkhound breeds have a genetic ancestry primarily from European dogs rather than from the morphologically more similar Siberian dogs. On the other hand, TreeMix analysis indicated gene flow from the root of the historic dog fur population to the Arctic dog breeds group (one Alaskan Malamute, one Greenland dog, and one Siberian Husky) (supplementary fig. S4, Supplementary Material online), suggesting ancient

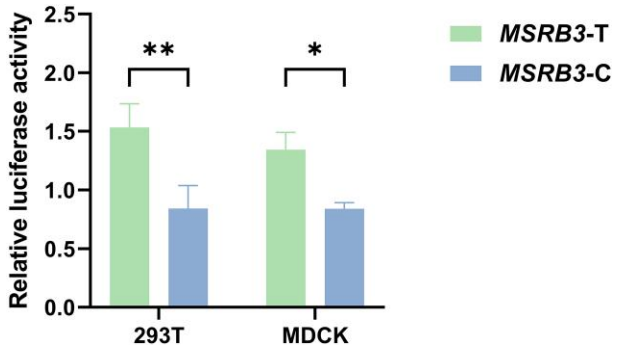

Fig. 4. Functional assay of *MSRB3* gene (Chr10:8394697, T > C). The expression of MSRB3 luciferase activity was regulated by cfa-miR-127, showing by the dual-luciferase reporter experiments enhancer assay using psiCHECK-2-promoter vectors with one human and one canine cell line: HEK293 (human embryonic kidney) and MDCK (Madin-Darby Canine kidney). The experiments were replicated at least three times, and the results were analyzed using a two-tails Student's *t*-test (*represents $P < 0.05$; **represents $P < 0.01$) and visualized by GraphPad software.

admixture among Arctic populations. To which extent the similarity in morphology to Siberian dogs of the modern Lapphund and Elkhound breeds stems from parallel evolution of the Arctic-related traits or from introgression of Siberian dogs into a European-derived population and subsequent breeding on these traits remains to be investigated.

D-statistic analysis indicated gene flow with European wolves for the fur samples as well as the Lapphund and Elkhound breeds, but for only around half of these individuals (Fig. 3). Notably, the fur samples with and without gene flow with wolves were divided into two subclades in the phylogenetic analysis, and the outgroup-f3 statistics showed the four samples with gene flow to be most closely related to the Lapphund and Elkhound samples. To further investigate the subdivision of the fur samples and their gene flow with the wolf population we performed TreeMix analysis, based on two scenarios, one with all eight historic fur samples as a single group and another separating the fur samples with and without indication of gene flow with European wolves (supplementary fig. S4, Supplementary Material online). However, neither of these scenarios indicates gene flow between the fur sample populations and European wolves at the optimal number of migrations, and higher number of migrations gave topologies inconsistent with the phylogenetic analysis. Therefore, while gene flow with wolves was indicated for half of both the fur samples and the Lapphund and Elkhound breed dogs, indicating substructures in these populations, the exact relations between these could not be determined.

The intense selective breeding of today's dog breeds has led not only to desired traits but also to loss of genetic variation and related inbreeding health problems due to random enrichment of disease-causing mutations (Leroy 2011). As expected, the inbreeding coefficient was considerably higher for the modern Lapphund and Elkhound breed dogs than for the fur samples (supplementary fig. S5, Supplementary Material online), indicating loss of diversity during the development of the breeds. Accordingly, the modern Lapphund and Elkhound breeds have a high incidence of several disorders. Glycogen Storage Disease II (GSDII) is a disorder affecting modern Lapphund breeds but not found in Elkhounds. In line with this, genome-wide scans for selection signatures indicated that the gene *MGAM*, which is related to GSDII in humans (Zhang et al. 2006), was under positive selection in the Elkhound breed group but not in the Lapphund breed group. The intense breeding for phenotype is reflected in the positive selection on the gene *MSRB3*, which is associated with the phenotype of drop/prick ear in dogs (Vaysse et al. 2011; Plassais et al. 2019) and with ear morphology in both sheep and pigs (Che et al. 2018; Paris et al. 2020). The *MSRB3* gene was under positive selection in both the Lapphund and the Elkhound breed group, which is in accordance with the breed standard for the prick ear phenotype for all modern Lapphund and Elkhound breeds putting a strong selection on this trait during breed development. There are also morphological traits with potentially strong selection for specific Lapphund and Elkhound breeds, e.g. the coat colours of SL (solid black), SE (grey) and NE (grey or solid black) described in the breed standards. However, because the sample size of each of the modern breeds (four to six samples per breed) was not sufficient for detecting selection signatures, the analysis was performed based on breed groups, concealing signals for breed specific traits. A trait common to all the Lapphund and Elkhound breed standards, as well as to all polar spitz dogs, is the double coat, but since this trait was shared with the historic fur samples, which formed the baseline for the genome-wide scans for selection signatures, potential selection signatures would not be detected.

In summary, studies of the genetic origin and development of the modern European dog breeds from their non-breed founder populations have been hampered by the lack of surviving founder populations, and by the complicated history of European breeds hindering conclusions based on breed-to-breed comparisons. With this study we have shown that 100 to 200 years old museum samples could be used to represent the local non-breed dog population from which the Lapphund and Elkhound breeds derived, allowing genomic analyses of the origin and artificial selection of these modern dog breeds.

Materials and Methods

Samples

We analyzed the genome sequence of two datasets: (i) modern Nordic spitz breed dogs (Lapphund and Elkhound breeds) and (ii) 100 to 200 years old dog fur samples from northern Sweden. We studied five modern Lapphund and Elkhound breeds: Swedish Lapphund (SL), Finnish Lapphund (FL), Lapponian Herder (LH), Swedish Elkhound (SE), and Norwegian Elkhound (NE), four to six dogs from each breed. These breeds are genetically related

Nordic spitz breeds (Klütsch et al. 2010), which originate from the Sápmi area (the region traditionally inhabited by the Sámi people) shared by northern Scandinavia, Finland and the Russian Kola peninsula. The three Lapphund breeds (SL, Fl, LH) originate from dogs used by the Sámi primarily for reindeer herding, and the two Elkhound breeds (SE, NE) originate from dogs used by the Swedish speaking and Norwegian speaking rural populations mainly for hunting.

We extracted and analyzed DNA from eight dog fur samples collected from the Swedish part of the Sápmi area. In northern Sweden, clothes made of fur from dogs were commonly manufactured among both the Sámi reindeer herding people and the Swedish rural inhabitants during the 19th century up until the start of World War I in 1914, which effectively stopped the large-scale manufacturing of dog furs. The dog furs in this trade were obtained from the local populations of indigenous dogs used by the Sámi and Swedish rural populations, and the modern Swedish Lapphund and Swedish Elkhound breeds are believed to have been developed from these populations (Svanberg and Lindin 1997). The eight dog fur samples analyzed in this study were kindly provided by Ájtte Museum in Jokkmokk, Skogs-och samemuseet in Lycksele, and Skellefteå Museum in Skellefteå. The eight samples (Table 1) were obtained from the following objects: a Sámi woman's traditional fur coat (sample AJ2810), two dog fur gloves of different origin (samples SM11850 and SM16271), a sleeping bag made from dog and reindeer fur (samples Ly1, Ly2, and Ly3; three dog furs, from three different individuals), a single piece of Lapphund pelt (sample JM1711), and remains of an unknown object (sample SMOKÄND1; most probably a fur coat). Three objects have museum records dating their fabrication to the early 20th century (AJ2810, JM1711) and 1920s (SM16271), but the pelts may have been prepared earlier. The other five samples have no exact dating since the donors, because of the substantial age of the objects, did not have a precise record of their origin, but based on the tradition of fur preparation and the design of the clothes their fabrication can be approximately dated to 100 to 200 years ago.

Structured breeding of the Swedish Elkhound breed started in the 1930s, based on indigenous dogs from the Swedish-speaking rural population. Structured breeding of Swedish Lapphund was started in the early 1890s based on indigenous dogs from the Sámi people, forming a subpopulation of registered pedigree dogs with a low unidirectional gene flow from the Sámi-related population to the pedigree breed population, but not vice versa. This first pedigree population of Swedish Lapphund nearly went extinct, but in the 1950s less than five of these dogs were bred with additional dogs from the Sámi people, forming the basis for today's registered Swedish Lapphund breed. Therefore, the two breeds Swedish Lapphund and Swedish Elkhound can be assumed to have originated from the same indigenous dog population as the eight fur samples, a few decades up 110 years after the fur samples were manufactured. The three other Nordic breeds investigated, Finnish Lapphund, Lapponian Herder and Norwegian Elkhound, have similar histories as Swedish Lapphund and Swedish Elkhound, and the five breeds can therefore be assumed to belong to the same general population of indigenous dogs in northern Scandinavia and Finland.

DNA Extraction and Whole Genome Sequencing

DNA was extracted separately from each of the eight fur samples. The furs were first shaved with a sterilized razor blade to remove the hairs. For each sample, 25 mg of skin was cut into small pieces of size <1 mm^3, using sterilized scissors between each sample. The pieces were transferred into a PCR-clean 2.0 mL DNA LoBind tube (Eppendorf, cat. No. 0030108078) containing 70% ethanol (Sigma Aldrich, cat. No. E7023). The mixture was vortexed at maximum speed for one minute and then spun at 13,200 rpm in a table centrifuge for 1 min, and the supernatant was removed. The vortex step was repeated three times, and the tube was then left open for five minutes at 40 °C for complete ethanol evaporation. After that, the fur sample in each tube was used to prepare 50 uL of DNA extract per sample, using the DNA extraction method described in Dabney et al. (2013). Preparation of samples was performed in a clean room at the Laboratory of Molecular Paleontology at the Institute of Vertebrate Paleontology and Paleoanthropology (IVPP), Chinese Academy of Sciences, Beijing, China. All tubes and other experiment materials were UV irradiated for 40 min, and reagents were UV irradiated for 20 min. All laboratory procedures were conducted using contamination controls, such as the use of full body coverings, bleach decontamination, and UV irradiation of tools and work area before and between use. The DNA extracts were processed to produce double-stranded genomic libraries, which were prepared and treated with uracil-DNA-glycosylase (UDG) from *E. coli* and endonuclease (*Endo* VIII) to remove deaminated cytosine residues. The libraries were amplified using Pfu Turbo Cx Hotstart DNA Polymerase DNA with 35 cycles of amplification. Finally, the libraries were sequenced using the Illumina HiSeq X Ten platform.

Read Alignment and Variant Calling

Raw sequence reads were mapped to the dog reference genome of UU_Cfam_GSD_1.0 (Wang et al. 2021) using bwa-mem2 (Vasimuddin et al. 2019). GATK (DePresto et al. 2011) was adopted for marking duplicates and sorting in one step with option MarkDuplicatesSpark, for base calibration with option BQSR to produce final BAM files and for creation of a per-sample GVCF file for subsequent cohort short variant identification with option HaplotypeCaller. Final BAM files for each of the eight fur samples were separately analyzed for DNA damage using mapDamage with default settings (Ginolhac et al. 2011). GATK with option GatherVcfs was adopted to combine the GVCF files of 15 wolves and 73 dogs into one VCF

file, which was then used as a training set for adding the GVCF file of a dhole (as an outgroup). Finally, a VCF file was generated, composed of one dhole, 15 wolves and 73 dogs, and SNPs were called from autosomal chromosomes, including 26,744,144 SNPs with option SelectVariants (filtration threshold QD < 2.0 || FS > 60.0 || MQ < 40.0 || QUAL < 50.0 || SOR > 3.0 || MQRankSum < −12.5) for later genomic analysis. For more information about samples used in this study, see supplementary table S1, Supplementary Material online. Before formal bioinformatic analysis, genomic principal component analysis and calculation of pairwise genomic relationships were performed to inspect that the samples were obtained from different individuals without close relationship. Genomic principal component analysis (PCA) was performed by PLINK v 1.9 (Purcell et al. 2007) using option–pca, after dataset pruning (option–indep 50 5 2) for the eight fur samples. The first three genomic principal components were visualized by a R package "rgl" into 3D view in supplementary fig. S1, Supplementary Material online. A genomic relationship matrix for the eight dog furs samples was calculated using GCTA (Yang et al. 2011) using option–make-grm (supplementary table S2, Supplementary Material online).

Phylogenetic Analysis

Evolutionary relationships among all genome sequences (one dhole, 15 wolves, and 73 dogs) were analyzed (parameters: -a 38 -A -t 8 -c 1 -L 0.1 -b -o Dhole_01) using genome-wide SNPs by SNPhylo (Lee et al. 2014). The topology of evolutionary relationships was built based on the maximum likelihood algorithm and visualized by FigTree (Rambaut 2010), as shown in supplementary fig. S2, Supplementary Material online. To facilitate the interpretation of the phylogenetic tree, individual samples were highlighted with different colours to indicate geographical location or breed group (Fig. 1a).

Outgroup-f3 Test

Using ADMIXTOOLS (Patterson et al. 2012) and regarding dhole as an outgroup, outgroup-f3 values were estimated for each Lapphund and Elkhound breed sample compared to the fur samples and all other dog samples. The estimated outgroup-f3 values were marked on a map (Fig. 1b) showing the geographical location of breed development or sample collection of each dog and indicating f3 values by colour (yellow to red). For the sake of limited space in the figure, only nine breeds (out of 22) from Europe were marked on the map while the full results are found in supplementary table S3, Supplementary Material online.

Phylogenetic Analysis of Mitochondrial Genomes

The mitochondrial genomes from the eight dog furs were analyzed with the mitochondrial genomes from 169 domestic dogs and eight wolves from across the Old World and four coyotes, obtained from Pang et al. (2009). The DNA sequences were aligned, and a neighbor-joining tree was constructed, using Mega X (Kumar et al. 2018). The tree was visualized and graphically edited using iTOL (Letunic and Bork 2021), presented in Fig. 2 (for more details, see supplementary fig. S3, Supplementary Material online). A maximum likelihood analysis was constructed using IQ-TREE (Trifinopoulos et al. 2016) with 1,000 ultrafast bootstrap replicates.

D-Statistic Test

D-statistic test was performed using ADMIXTOOLS. For the order (W, X; Y, Z), each fur sample and each sample of the Lapphund and Elkhound breeds was independently set as W; the group of ten Southeast Asian indigenous non-breed dogs was set as X; three populations of European wolves (Portugal, European Russia and Sweden) were independently set as Y, and the dhole was regarded as an outgroup, set as Z. The estimated values of the *D*-statistic are presented in Fig. 3 (for more details, see supplementary table S4, Supplementary Material online).

TreeMix Analysis

TreeMix (Pickrell and Pritchard 2012) was adopted to test the potential direction of gene flow among breed groups and populations. We defined groups as: WF_RE (seven wolves from European Russia), WF_SWE (six wolves from Sweden), WF_PT (two wolves from Portugal), SEA (ten Southeast Asian indigenous non-breed dogs from South China and Vietnam), Arctic (one Alaskan Malamute, one Greenland dog, and one Siberian Husky), ME (one Afghan Hound, one Saluki, and one Sloughi), Fur (the eight historic dog fur samples), Fur1 (the four historic dog fur samples having gene flow with European wolves according to the D-statistic test: samples Ly1, Ly2, Ly3, and SM16271), Fur 2 (the four historic dog fur samples without gene flow with European wolves according to the D-statistic test: samples AJ2810, JM1711, SMOKAND1, and SM11850), LAP (six FL, six LH, and four SL) and ELK (four SE and five NE). Two scenarios were tested: one (scenario "Fur") with all the eight historic fur samples grouped together and another (scenario "Fur1 + Fur2") with the historic dog fur samples divided into the two separate groups with and without gene flow with European wolves according to the *D*-statistic test (Fig. 3). For each scenario, a phylogenetic tree was first generated based on the maximum likelihood method, the optimal number of migrations between groups was then explored using the R package OptM (Fitak 2021), and the final TreeMix graphs were generated (supplementary fig. S5, Supplementary Material online).

qpAdm Analysis

qpAdm analysis was performed using ADMIXTOOLS to estimate the proportion of genomic components of wolves in each Lapphund and Elkhound sample not present in the historic dog fur samples. The four fur samples without gene flow with European wolves according to the *D*-statistic test (AJ2810, JM1711, SMOKAND1, and

SM11850) and Swedish wolves were regarded as the source populations, and four groups of dogs were regarded as reference populations: Southeast Asian indigenous non-breed dogs (ten dogs from South China and Vietnam), Arctic breeds (one Alaskan Malamute, one Greenland dog, and one Siberian Husky), breeds from the Middle East (one Afghan Hound, one Saluki, and one Sloughi), and European breed dogs (in total 22 different breed dogs). The results are presented in supplementary table S5, Supplementary Material online.

Genetic Diversity

The genetic diversity of the five modern Lapphund and Elkhound breeds was compared to the fur samples and the Southeast Asian indigenous non-breed dogs, using the inbreeding coefficient based on runs of homozygosity (F_{ROH}) and linkage disequilibrium decay (LD decay). However, because the rate of LD decay is affected by the sample size of each breed/population, four samples were randomly taken from each breed/population to both the F_{ROH} and the LD decay analysis. Runs of homozygosity were scanned by PLINK v 1.9 (Purcell et al. 2007) using option–homozyg with default settings, after dataset pruning (option–indep 50 5 2) of all samples together. F_{ROH} was calculated for each sample as the total length of runs of homozygosity divided by the total length of the dog autosome (2,228,550,668 bp). Finally, the results of average F_{ROH} for each breed/population were presented in a boxplot using R software (supplementary fig. S5A, Supplementary Material online). PopLDdecay (Zhang et al. 2019) software was used for the LD decay analysis, analysing each breed/population individually and then visualizing them together in supplementary fig. S5B, Supplementary Material online.

Selection Signature Study

The three Lapphund breeds (FL, LH, and SL) and two Elkhound breeds (SE and NE) were compared, as two independent cohorts, to the eight fur samples for detecting potential selection signatures during their breed development stages. VCFtools (Danecek et al. 2011) and selscan (Szpiech and Hernandez 2014) were adopted to perform Fst (Fixation Index) and XP-EHH (Cross-population Extended Haplotype Homozygosity) using 50 kb sliding windows (no overlap between windows). After acquiring the Top 1% windows with the highest estimated values from the two methods, Ensembl Variant Effect Predictor annotated SNP loci from overlapping windows (McLaren et al. 2016), as listed in supplementary table S6, Supplementary Material online. Particularly, loci containing mutations with nonsynonymous SNPs only in protein-coding regions were listed in Table 2.

Enzymatic Assays of *MSRB3*

For the two non-synonymous SNP mutations (Chr10:8394629, A > G and Chr10:8394668, G > A) in the exon region of the *MSRB3* gene, the genes encoding wild-type MSRB3 and its mutation MSRB3$^{T144A/G155S}$ were synthesized by Tongyong Biosystem Co., Ltd (China) and cloned into bacterial vector pET28a to construct pET28a-MSRB3 and pET28a-MSRB3$^{T144A/G155S}$, respectively. The *E. coli* BL21(DE3) strains harbouring pET28a-MSRB3 and pET28a-MSRB3$^{T144A/G155S}$ were grown to an OD600 of 0.8 in LB medium at 37 °C and then induced for 16 h with 1 mM IPTG at 25 °C. The cells were harvested, washed, and resuspended with buffer A (20 mM sodium phosphate and 500 mM sodium chloride, pH 7.4) and lysed by sonication on ice. The cell lysate was centrifuged at 12,000 × g for 20 min to remove cell debris. The resultant supernatant was loaded onto a HisTrap HP column and the target proteins were eluted with buffer B (20 mM sodium phosphate, 500 mM sodium chloride, and 500 mM imidazole, pH 7.4). The purified pET28a-MSRB3 and pET28a-MSRB3$^{T144A/G155S}$ were analyzed by 13% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), and protein concentrations were determined using the Bradford protein assay kit (Sangon, China). The activity of MSRB3 and MSRB3$^{T144A/G155S}$ was detected as follows: A reaction mixture (200 µL) containing 50 mM PBS (pH 7.4), 250 µM DABS-L-MetSO, 15 mM DTT, and 0.05 mg/mL MSRB3 and MSRB3$^{T144A/G155S}$ was incubated at 37 °C for 5 min. The reaction was stopped by adding 200 µL of 1 M sodium acetate (pH 6.0). Then, 100 µL of acetonitrile and 1 mL of n-hexane were added, and the concentration of produced DABS-L-Met in the reaction mixture was assayed by HPLC. One unit (U) of MSRB3 activity was defined as the amount of enzyme that catalyzed the production of 1 µmol of DABS-L-Met per min. The results are presented in supplementary fig. S6, Supplementary Material online.

Dual-Luciferase Reporter Assays of MSRB3

For the dual-luciferase reporter assays of a non-synonymous *MSRB3* gene SNP mutation (Chr10:8394697, T > C), which was predicted to be located in the 3′ UTR region based on transcript MSRB3-201 (ENSCAFT 00000044828.2), we first performed miRNA target prediction for this SNP using six databases: miRWaLK, Starbase, RNA22, miRDB, miRanda, and MicroT-CDS. Referring to the sequence complementarity between miRNA and mRNA, the SNP-associated mmu-miR-7021-3p was predicted as a post-transcriptional regulatory factor of *MSRB3* based on the RNA22 and miRanda databases. The sequence of mmu-miR-7021-3p is highly conserved and consistent with the sequence of cfa-miR-127 (supplementary fig. S7A, Supplementary Material online). The 3′-UTR of *MSRB3* gene, which contained the putative target site of cfa-miR-127, is also conserved (supplementary fig. S7B, Supplementary Material online), and cfa-miR-127 could therefore be used in subsequent dual-luciferase reporter assays (supplementary fig. S7C, Supplementary Material online). The 3′-UTR of *MSRB3*, which contained the putative target site of cfa-miR-127 was synthesized and ligated into a psiCHECK-2 construct (Kunming Animal Branch, Chinese Academy of Sciences) to construct wild-type and mutant-type MSRB3 dual-luciferase

reporter vectors (supplementary fig. S7D, Supplementary Material online). The day before transfection, human embryonic kidney 293T (HEK 293T) cell line and canine kidney (MDCK) cell line (released from the cell platform of Kunming Animal Branch, Chinese Academy of Sciences) were inoculated into the culture medium. Both cell lines were negative for *Mycoplasma*. Cells were cultured at 37 °C in a humid environment with 5% CO2 in a DMEM high-sugar medium with 10% fetal bovine serum. After this, two types of MSRB3 dual-luciferase reporter vectors (wild and mutant, each 200 ng per well) and cfa-miR-127 mimics (20 μM, 0.1 μL per well), were co-transfected into the cells. After 24 h of transfection, the cells were collected and the luciferase activity was estimated using the Dual-Luciferase Reporter Assay System (Promega), normalized to renilla reniformis luciferase. The statistical significance of the data was analyzed using a two-tailed *t*-test. The results are presented in Fig. 4.

Supplementary Material

Supplemental information can be found on the documents below: supplementary figs. and supplementary tables, Supplementary Material online. Supplementary material is available at *Molecular Biology and Evolution* online.

Acknowledgments

We thank Ájtte museum, Skogs- och samemuseet and Skellefteå museum for providing the dog fur samples, and Prof. Greger Larson and Prof. Laurent Frantz for helpful comments and suggestions. This work was supported by NSFC (32170436, 32360122), National Key R&D Program of China (2019YFA0707101), Agria och SKK Forskningsfond för sällskapsdjur (project N2021-0019), STI2030-Major Projects (2021ZD0203900), Spring City Plan: the High-level Talent Promotion and Training Project of Kunming (2022SCP001) and the project GREKF21-10 from State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences. Guo-Dong Wang was supported by Yunnan Fundamental Research Projects (202201AV070011). Hannes Lohi was supported by Jane and Aatos Erkko Foundation and HiLIFE. This work was also supported for sample storage by the Animal Branch of the Germplasm Bank of Wild Species, Chinese Academy of Sciences.

Data Availability

New data generated from this study can be downloaded from the SRA database of NCBI (project number: PRJNA889984).

References

Akey JM, Ruhe AL, Akey DT, Wong AK, Connelly CF, Madeoy J, Nicholas TJ, Neff MW. Tracking footprints of artificial selection in the dog genome. *Proc Natl Acad Sci U S A*. 2010:**107**(3):1160–1165. https://doi.org/10.1073/pnas.0909918107.

Axelsson E, Ratnakumar A, Arendt ML, Maqbool K, Webster MT, Perloski M, Liberg O, Arnemo JM, Hedhammar A, Lindblad-Toh K. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature*. 2013:**495**(7441):360–364. https://doi.org/10.1038/nature11837.

Bergström A, Frantz L, Schmidt R, Ersmark E, Lebrasseur O, Girdland-Flink L, Lin AT, Storå J, Sjögren KG, Anthony D, *et al.* Origins and genetic legacy of prehistoric dogs. *Science*. 2020:**370**(6516):557–564.

Bergström A, Stanton DWG, Taron UH, Frantz L, Sinding MS, Ersmark E, Pfrengle S, Cassatt-Johnstone M, Lebrasseur O, Girdland-Flink L, *et al.* Grey wolf genomic history reveals a dual ancestry of dogs. *Nature*. 2022:**607**(7918):313–320. https://doi.org/10.1038/s41586-022-04824-9.

Boyko AR, Boyko RH, Boyko CM, Parker HG, Castelhano M, Corey L, Degenhardt JD, Auton A, Hedimbi M, Kityo R, *et al.* Complex population structure in African village dogs and its implications for inferring dog domestication history. *Proc Natl Acad Sci U S A*. 2009:**106**(33):13903–13908. https://doi.org/10.1073/pnas.0902129106.

Che C, Liu C, Xiong X, Fang S, Yang H, Zhang Z, Ren J, Guo Y, Huang L. Copy number variation in the MSRB3 gene enlarges porcine ear size through a mechanism involving miR-584-5p. *Genet Sel Evol*. 2018:**50**(1):72. https://doi.org/10.1186/s12711-018-0442-6.

Dabney J, Knapp M, Glocke I, Gansauge MT, Weihmann A, Nickel B, Valdiosera C, García N, Pääbo S, Arsuaga JL, *et al.* Complete mitochondrial genome sequence of a middle pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc Natl Acad Sci U S A*. 2013:**110**(39):15758–15763. https://doi.org/10.1073/pnas.1314445110.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, *et al.* The variant call format and VCFtools. *Bioinformatics*. 2011:**27**(15):2156–2158. https://doi.org/10.1093/bioinformatics/btr330.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011:**43**(5):491–498. https://doi.org/10.1038/ng.806.

Feuerborn TR, Carmagnini A, Losey RJ, Nomokonova T, Askeyev A, Askeyev I, Askeyev O, Antipina EE, Appelt M, Bachura OP, *et al.* Modern Siberian dog ancestry was shaped by several thousand years of Eurasian-wide trade and human dispersal. *Proc Natl Acad Sci U S A*. 2021:**118**(39):e2100338118.

Fitak RR. Optm: estimating the optimal number of migration edges on population trees using Treemix. *Biol Methods Protoc*. 2021:**6**(1):bpab017. https://doi.org/10.1093/biomethods/bpab017.

Frantz LA, Mullin VE, Pionnier-Capitan M, Lebrasseur O, Ollivier M, Linderholm PA, Mattiangeli A, Teasdale V, Dimopoulos MD, A E, *et al.* Genomic and archaeological evidence suggest a dual origin of domestic dogs. *Science*. 2016:**352**(6290):1228–1231. https://doi.org/10.1126/science.aaf3161.

Freedman AH, Gronau I, Schweizer RM, Ortega-Del VD, Han E, Silva PM, Galaverni M, Fan Z, Marx P, Lorente-Galdos B, *et al.* Genome sequencing highlights the dynamic early history of dogs. *PLoS Genet*. 2014:**10**(1):e1004016. https://doi.org/10.1371/journal.pgen.1004016.

Ginolhac A, Rasmussen M, Gilbert MT, Willerslev E, Orlando L. mapDamage: testing for damage patterns in ancient DNA sequences. *Bioinformatics*. 2011:**27**(15):2153–2155. https://doi.org/10.1093/bioinformatics/btr347.

Hung CM, Shaner PJ, Zink RM, Liu WC, Chu TC, Huang WS, Li SH. Drastic population fluctuations explain the rapid extinction of the passenger pigeon. *Proc Natl Acad Sci U S A*. 2014:**111**(29):10636–10641. https://doi.org/10.1073/pnas.1401526111.

Klütsch CF, Seppälä EH, Fall T, Uhlén M, Hedhammar A, Lohi H, Savolainen P. Regional occurrence, high frequency but low diversity of mitochondrial DNA haplogroup d1 suggests a recent dog-wolf hybridization in Scandinavia. *Anim Genet*. 2011:**42**(1):100–103. https://doi.org/10.1111/j.1365-2052.2010.02069.x.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA x: molecular evolutionary genetics analysis across computing platforms.

Mol Biol Evol. 2018:**35**(6):1547–1549. https://doi.org/10.1093/molbev/msy096.

Lindblad-Toh K, Wade CM, Mikkelse TS, Karlsson EK, Jaffe DB, Kamal M, Clamp M, Chang JL, Kulbokas EJ, Zody MC, *et al*. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*. 2005:**438**(7069):803–819. https://doi.org/10.1038/nature04338.

Lee TH, Guo H, Wang X, Kim C, Paterson AH. SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genom*. 2014:**15**(1):162. https://doi.org/10.1186/1471-2164-15-162.

Leroy G. Genetic diversity, inbreeding and breeding practices in dogs: results from pedigree analyses. *Vet J*. 2011:**189**(2):177–182. https://doi.org/10.1016/j.tvjl.2011.06.016.

Letunic I, Bork P. Interactive Tree of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res*. 2021:**49**(W1):W293–W296. https://doi.org/10.1093/nar/gkab301.

McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F. The ensembl variant effect predictor. *Genome Biol*. 2016:**17**(1):122. https://doi.org/10.1186/s13059-016-0974-4.

Meadows JRS, Kidd JM, Wang GD, Parker HG, Schall PZ, Bianchi M, Christmas MJ, Bougiouri K, Buckley RM, Hitte C, *et al*. Genome sequencing of 2000 canids by the Dog10 K consortium advances the understanding of demography, genome function and architecture. *Genome Biol*. 2023:**24**(1):187. https://doi.org/10.1186/s13059-023-03023-7.

Pang JF, Kluetsch C, Zou XJ, Zhang AB, Luo LY, Angleby H, Ardalan A, Ekström C, Sköllermo A, Lundeberg 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*. 2009:**26**(12):2849–2864. https://doi.org/10.1093/molbev/msp195.

Paris JM, Letko A, Häfliger IM, Ammann P, Drögemüller C. Ear type in sheep is associated with the MSRB3 locus. *Anim Genet*. 2020:**51**(6):968–972. https://doi.org/10.1111/age.12994.

Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. Ancient admixture in human history. *Genet*. 2012:**192**(3):1065–1093. https://doi.org/10.1534/genetics.112.145037.

Pickrell JK, Pritchard JK. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet*. 2012:**8**(11):e1002967. https://doi.org/10.1371/journal.pgen.1002967.

Plassais J, Kim J, Davis BW, Karyadi DM, Hogan AN, Harris AC, Decker B, Parker HG, Ostrander EA. Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. *Nat Commun*. 2019:**10**(1):1489. https://doi.org/10.1038/s41467-019-09373-w.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, *et al*. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007:**81**(3):559–575. https://doi.org/10.1086/519795.

Rambaut A. *Figtree v1.3.1*. Edinburgh: Institute of Evolutionary Biology, University of Edinburgh; 2010. http://tree.bio.ed.ac.uk/software/figtree/

Raxworthy CJ, Smith BT. Mining museums for historical DNA: advances and challenges in museomics. *Trends Ecol Evol*. 2021:**36**(11):1049–1060. https://doi.org/10.1016/j.tree.2021.07.009.

Savolainen P, Zhang YP, Luo J, Lundeberg J, Leitner T. Genetic evidence for an east Asian origin of domestic dogs. *Science*. 2002:**298**(5598):1610–1613. https://doi.org/10.1126/science.1073906.

Seppälä EH, Reuser AJ, Lohi H. A nonsense mutation in the acid α-glucosidase gene causes pompe disease in Finnish and Swedish Lapphunds. *PloS one*. 2013:**8**(2):e56825. https://doi.org/10.1371/journal.pone.0056825.

Shannon LM, Boyko RH, Castelhano M, Hayward CE, McLean JJ, White C, Abi ME, Anita SM, Bondjengo BA, I N, *et al*. Genetic structure in village dogs reveals a central Asian domestication origin. *Proc Natl Acad Sci U S A*. 2015:**112**(44):13639–13644. https://doi.org/10.1073/pnas.1516215112.

Sinding MS, Gopalakrishnan S, Ramos-Madrigal J, de Manuel M, Pitulko VV, Kuderna L, Feuerborn TR, Frantz LAF, Vieira FG, Niemann J, *et al*. Arctic-adapted dogs emerged at the Pleistocene-Holocene transition. *Science*. 2020:**368**(6498):1495–1499. https://doi.org/10.1126/science.aaz8599.

Svanberg I, Lindin L. Den jämtländska spetsen-en pälshund? Hundskinn och pälshundsuppfödning i norra Sverige. *Svenska Landsmål och Svenskt Folkliv*. 1997:**332**:133–158.

Szpiech ZA, Hernandez RD. Selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. *Mol Biol Evol*. 2014:**31**(10):2824–2827. https://doi.org/10.1093/molbev/msu211.

Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res*. 2016:**44W1**(W1):W232–W235. https://doi.org/10.1093/nar/gkw256.

Vasimuddin M, Misra S, Li H, Aluru S. Efficient architecture-aware acceleration of BWA-MEM for multicore systems. 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS); Rio de Janeiro, Brazil; 2019. p. 314–324. https://doi.org/10.1109/IPDPS.2019.00041.

Vaysse A, Ratnakumar A, Derrien T, Axelsson E, Rosengren PG, Sigurdsson S, Fall T, Seppälä EH, Hansen MS, Lawley CT, *et al*. Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping. *PLoS Genet*. 2011:**7**(10):e1002316. https://doi.org/10.1371/journal.pgen.1002316.

van der Valk T, Díez-Del-Molino D, Marques-Bonet T, Guschanski K, Dalén L. Historical genomes reveal the genomic consequences of recent population decline in eastern gorillas. *Curr Biol*. 2019:**29**(1):165–170.e6. https://doi.org/10.1016/j.cub.2018.11.055.

Vilà C, Savolainen P, Maldonado JE, Amorim IR, Rice JE, Honeycutt RL, Crandall KA, Lundeberg J, Wayne RK. Multiple and ancient origins of the domestic dog. *Science*. 1997:**276**(5319):1687–1689. https://doi.org/10.1126/science.276.5319.1687.

Wang GD, Zhai WW, Yang HC, Fan RX, Cao X, Zhong L, Wang L, Liu F, Wu H, Cheng LG, *et al*. The genomics of selection in dogs and the parallel evolution between dogs and humans. *Nat Commun*. 2013:**4**(1):1860. https://doi.org/10.1038/ncomms2814.

Wang GD, Zhai WW, Yang HC, Wang L, Zhong L, Liu YH, Fan RX, Yin TT, Zhu CL, Poyarkov AD, *et al*. Out of southern east Asia: the natural history of domestic dogs across the world. *Cell Res*. 2016:**26**(1):21–33. https://doi.org/10.1038/cr.2015.147.

Wang GD, Zhang M, Wang X, Yang MA, Cao P, Liu F, Lu H, Feng X, Skoglund P, Wang L, *et al*. Genomic approaches reveal an endemic subpopulation of gray wolves in southern China. *iScience*. 2019:**20**:110–118. https://doi.org/10.1016/j.isci.2019.09.008.

Wang C, Wallerman O, Arendt ML, Sundström E, Karlsson Å, Nordin J, Mäkeläinen S, Pielberg GR, Hanson J, Ohlsson Å, *et al*. A novel canine reference genome resolves genomic architecture and uncovers transcript complexity. *Commun Biol*. 2021:**4**(1):185. https://doi.org/10.1038/s42003-021-01698-x.

Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*. 2011:**88**(1):76–82. https://doi.org/10.1016/j.ajhg.2010.11.011.

Zhang H, Kallwass H, Young SP, Carr C, Dai J, Kishnani PS, Millington DS, Keutzer J, Chen YT, Bali D. Comparison of maltose and acarbose as inhibitors of maltase-glucoamylase activity in assaying acid alpha-glucosidase activity in dried blood spots for the diagnosis of infantile Pompe disease. *Genet Med*. 2006:**8**(5):302–306. https://doi.org/10.1097/01.gim.0000217781.66786.9b.

Zhang C, Dong SS, Xu JY, He WM, Yang TL. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics*. 2019:**35**(10):1786–1788. https://doi.org/10.1093/bioinformatics/bty875.

Zhou QJ, Liu X, Zhang L, Wang R, Yin T, Li X, Li G, He Y, Ding Z, Ma P, *et al*. A single-nucleus transcriptomic atlas of the dog hippocampus reveals the potential relationship between specific cell types and domestication. *Natl Sci Rev*. 2022:**9**(11):nwac147. https://doi.org/10.1093/nsr/nwac147.