

RESEARCH ARTICLE

Open Access

# Genetic diversity and population structure of the Sapsaree, a native Korean dog breed



Chandima Gajaweera<sup>1,5†</sup>, Ji Min Kang<sup>1†</sup>, Doo Ho Lee<sup>1</sup>, Soo Hyun Lee<sup>1</sup>, Yeong Kuk Kim<sup>1</sup>, Hasini I. Wijayananda<sup>1</sup>, Jong Joo Kim<sup>3</sup>, Ji Hong Ha<sup>4</sup>, Bong Hwan Choi<sup>2\*</sup> and Seung Hwan Lee<sup>1\*</sup> 

## Abstract

**Background:** The Sapsaree is a breed of dog (*Canis familiaris*) native to Korea, which became perilously close to extinction in the mid-1980s. However, with systematic genetic conservation and restoration efforts, this breed was rescued from extinction and population sizes have been gradually increasing over the past few decades. The aim of this study was to ascertain novel information about the genetic diversity, population structure, and demographic history of the Sapsaree breed using genome-wide single nucleotide polymorphism data. We characterized the genetic profile of the Sapsaree breed by comparison with seven foreign dog breeds with similar morphologies to estimate genetic differentiation within and among these breeds.

**Results:** The results suggest that Sapsarees have higher genetic variance compared with the other breeds analyzed. The majority of the Sapsarees in this study share a discrete genetic pattern, although some individuals were slightly different, possibly as a consequence of the recent restoration process. Concordant results from analyses of linkage disequilibrium, effective population size, genetic diversity, and population structural analyses illustrate a relationship among the Sapsaree and the Tibetan breeds Tibetan terrier and Lhasa Apso, and a small genetic introgression from European breeds. The effective population size of the Sapsaree has contracted dramatically over the past generations, and is currently insufficient to maintain long-term viability of the breed's genetic diversity.

**Conclusions:** This study provides novel insights regarding the genetic diversity and population structure of the native Korean dog breed Sapsaree. Our results suggest the importance of a strategic and systematic approach to ensure the genetic diversity and the authenticity of the Sapsaree breed.

**Keywords:** Sapsaree, Genetic diversity, Population structure

## Background

The domestic dog (*Canis familiaris*) is the most phenotypically diverse mammalian species, and one of the first animals to be domesticated by humans [1–3]. While dogs are the closest animal companion of humans, they are still used for specialized tasks including herding, hunting, retrieving, pulling sleds, and even for military tasks [4–6]. The gray wolf (*Canis lupus*) is the common ancestor of domesticated dogs, which have since been differentiated through artificial selection of the hugely

diverse features of modern breeds [7, 8]. It has been hypothesized that the domestication of dogs began nearly 33,000 years ago in South East Asia. Ancestral canines accompanied humans in a migration to Africa and the Middle East around 15,000 years ago, and then to Europe around 10,000 years ago [6, 9–11].

Although evidence suggests dogs have been present on the Korean peninsula for a long period of time, the specifics of canine domestication are not well understood. Some have hypothesized that current dog breeds on the Korean peninsula were gradually introduced with the influx of humans. Today, there are more than 150 dog breeds on the Korean peninsula, and over 400 recognized dog breeds worldwide [12, 13]. Among the native Korean dog breeds, the Jindo, Sapsaree, and Donggyeong are protected as a designated 'natural monument' by the Korean government (Cultural Heritage Administration

\* Correspondence: [bhchoi@korea.kr](mailto:bhchoi@korea.kr); [slee46@cnu.ac.kr](mailto:slee46@cnu.ac.kr)

<sup>†</sup>Chandima Gajaweera and Ji Min Kang contributed equally to this work.

<sup>2</sup>Animal Genomics & Bioinformatics Division, National Institute of Animal Science, RDA, Wanju 55365, Republic of Korea

<sup>1</sup>Division of Animal & Dairy Science, Chungnam National University, Daejeon 34134, Republic of Korea

Full list of author information is available at the end of the article



of Korea, #54, 368, and 540 respectively) [12, 14, 15]. The Poongsan breed was also designated as a natural monument during the Japanese colonial period (number 128), but the designation was removed by the Korean government in 1962 [12, 16].

The Sapsaree is a shaggy-haired and droopy-eared dog breed believed to reflect the character of the Korean people. They have a medium body size (54–62 cm in height) and two distinguishable coat colors: the ‘Chung,’ or blue Sapsaree, and the ‘Hwang,’ or yellow Sapsaree [12, 16, 17].

Historical evidence suggests that Sapsarees were used as military dogs by nobles of the Silla dynasty. Following the collapse of the unified Silla, Sapsarees were featured in the classical literary works of the Joseon dynasty and have since gained popularity throughout the Korean peninsula. Their disposition is friendly and gentle, and their loyalty has long been recognized [16, 18, 19].

The population size of Sapsaree was substantially decreased and became perilously close to extinction during the Japanese colonial period (1910–1945) and the Korean War (1950–1953). In 1969, a Sapsaree revival was initiated by Kyungpook National University, however the restoration process and systematic genetic conservation began by 1985 at the Sapsaree Breeding Research Institute in Gyeongsan, South Korea. In 1992, the Sapsaree was registered as a national treasure of Korea and their breeding and sale were strictly regulated to protect the purity of the breed [17–23]. Current total Sapsaree population is approximately 4000 including the 500 dogs maintained at the Sapsaree Breeding Research Institute [19]. The existing Sapsaree population size is relatively small, and it will therefore be necessary to expand the population size to maintain the sustainability of the breed.

Understanding the genetic diversity of domesticated species is important to establish effective conservation decisions and management strategies [24, 25]. Advances in genome technology and the availability of high density genome-wide single nucleotide polymorphism (SNP) data have facilitated the characterization of genetic diversity and breed composition [26, 27]. Linkage disequilibrium (LD), effective population size ( $N_e$ ), and heterozygosity are parameters widely used to understand the genetic diversity of populations [24]. The evolutionary history of a population is estimated through LD, by estimating the non-random association between two genetic markers that results from various evolutionary and demographic processes [28, 29]. Another important parameter for estimating the demographic history of a population is  $N_e$ , which estimates the rate of genetic drift, inbreeding, and the effects of evolutionary forces such as mutation, selection, and migration [30, 31]. Heterozygosity is also a widely used parameter to measure genetic variation within a population [23, 32]. Information regarding genetic diversity, LD,  $N_e$ , and heterozygosity would therefore be useful for

establishing a breeding program that avoids inbreeding while maintaining the breed purity of Sapsarees. However, there are a limited number of scientific studies on the genetic diversity of Sapsaree populations [20, 21, 23, 33]. In this study, we used high-density SNP data to estimate the genetic diversity of the Sapsaree. We characterized the genetic profile of the Sapsaree by comparison with seven foreign dog breeds with similar morphology and estimated the genetic differentiation within and among these breeds.

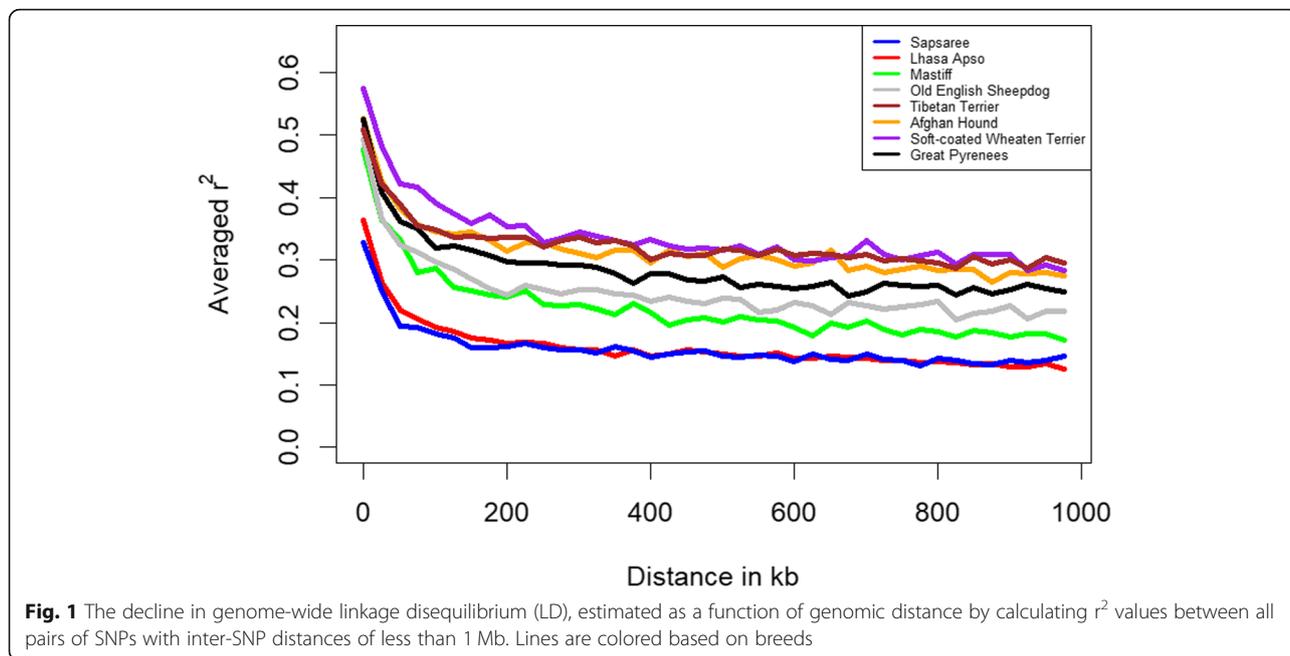
## Results

As LD is expected to decay with recombination and increase the physical distance between markers [48], Fig. 1 shows different estimates of genome-wide LD for each of the eight populations, and declines in LD with increasing genomic distance across and within breeds. However, the rates of decay were different among breeds. Large differences were observed between Sapsaree, Lhasa Apso, and the other breeds. LD dropped off rapidly over a short distance in all breeds. Sapsaree and Lhasa Apso showed the lowest average LD across the genome. The breeds with the highest average LD were the Soft-coated Wheaten Terrier at the short marker distance but, the Tibetan Terrier at the long-distance marker. However, the LD values of Tibetan Terrier and Soft-coated Wheaten Terrier were not significantly different toward the long-distance.

The estimated effective population size ( $N_e$ ) at  $t$  generations ago is shown in Fig. 2. The results suggest that  $N_e$  was lower in the recent past compared with the ancient past (Fig. 2). Based on the genomic data 11 generations ago, the highest  $N_e$  was for Sapsaree which approximately 54 individuals, followed by Lhasa Apso (51 individuals) and the lowest  $N_e$  was approximately 17 individuals for the Tibetan Terrier (Fig. 2). In the more distant past of 1400 generations ago, the  $N_e$  was highest for Sapsaree approximately 2098 then 1966 for Lhasa Apso, and lowest for Soft-coated Wheaten terrier (approximately 764).

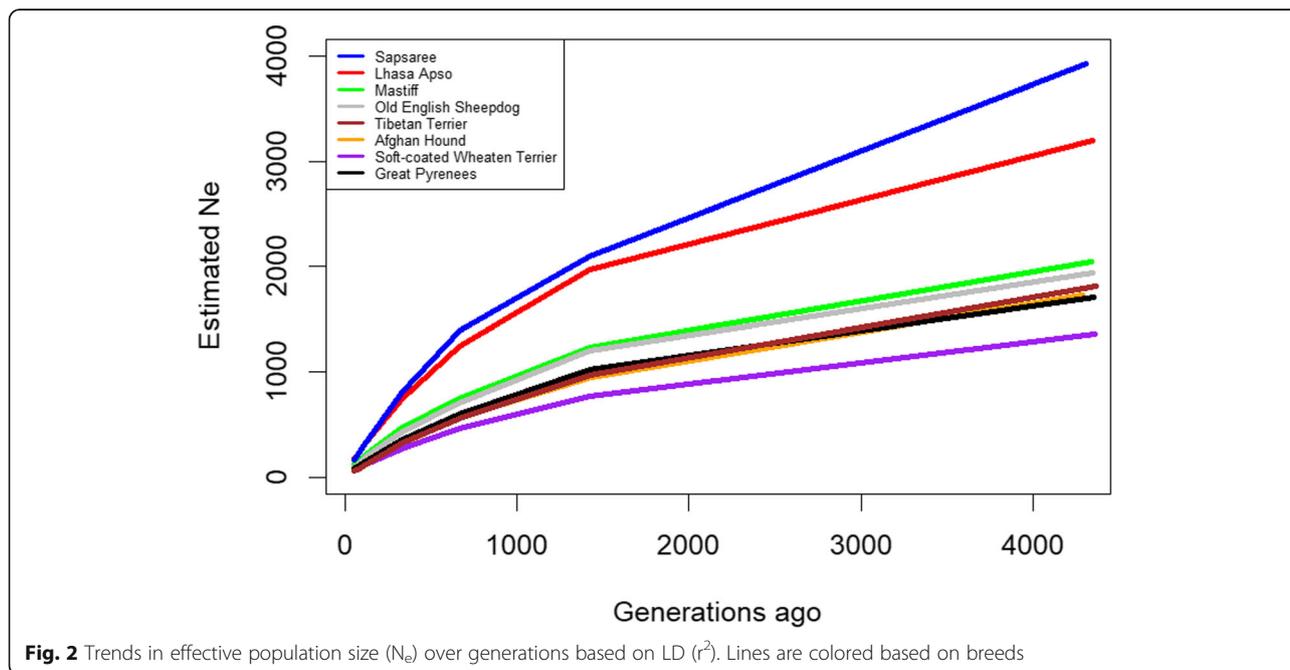
Heterozygosity was highest in the Sapsaree (0.342), followed by the Lhasa Apso (0.309) and Tibetan Terrier (0.273). The Old English Sheepdog (0.179) and Great Pyrenees (0.232) showed the lowest heterozygosity in the present generation (Fig. 3). Results suggest that heterozygosity will decline drastically in the future and is predicted to reduce by half within 25 generations. The estimated heterozygosity after 50 generations was also highest in the Sapsaree (0.118), with the Tibetan Terrier (0.003), Soft-coated Wheaten terrier (0.012), and Old English Sheepdog (0.000) showing the lowest values.

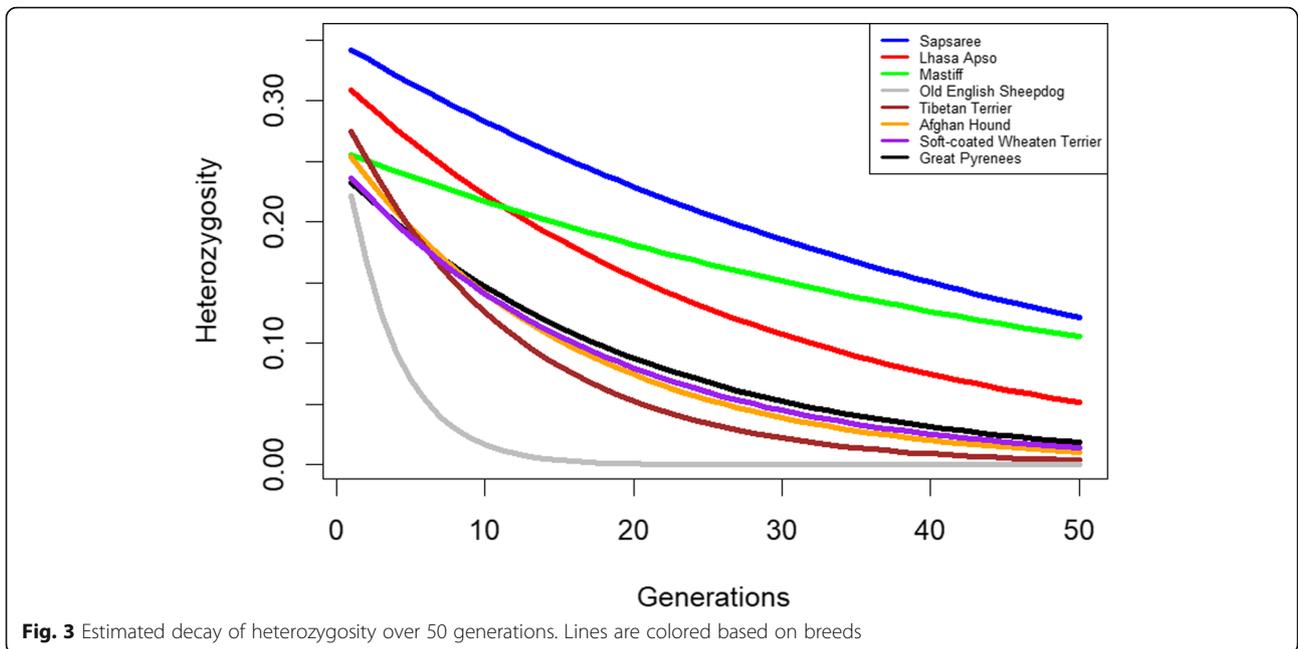
Ancestry-based models of admixture analysis were used to show the genetic structure and admixture proportion of the canine ancestors (Fig. 4 and Additional file 3: Figure S3). Additional file 1: Figure S1 shows that the lowest CV



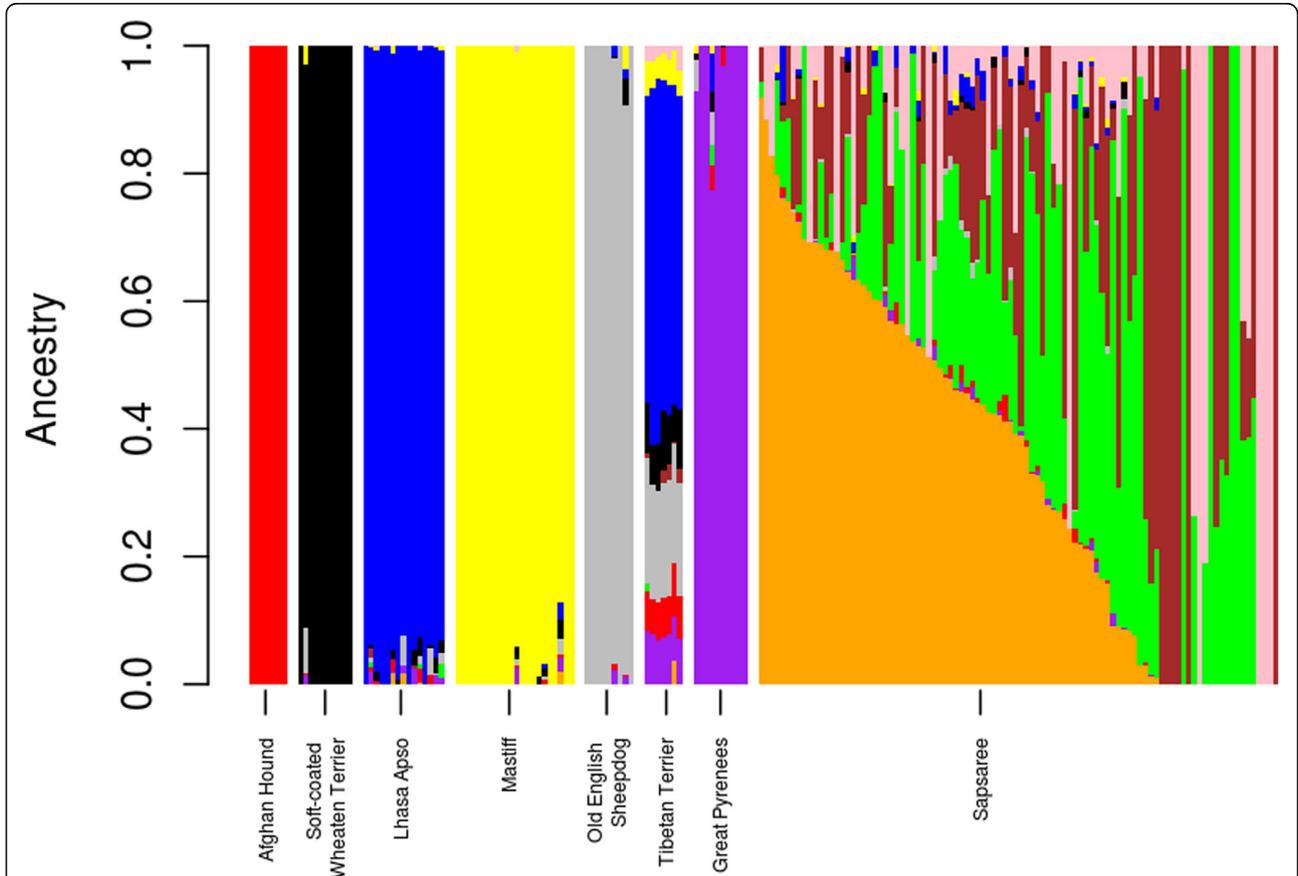
error (0.583) was obtained at  $K = 10$ . The relationship of ancestry for Sapsaree and other breeds was visualized using  $K = 10$ , where  $K$  is the number of ancestors. Admixture models illustrated the greater degree of diversity and admixture in Sapsaree than the other breeds. Moreover, the admixture analysis was done with several other related dog breeds based on the genetic distance (Additional file 4: Figure S4) also revealed a greater genetic heterogeneity within the Sapsaree breed. Afghan Hound, Lhasa Apso,

Great Pyrenees, Old English Sheepdog, Soft-coated Wheaten terrier, and Mastiff seem to have little or no admixture from other breeds, indicating that they have less remaining from other interacted ancestral breeds. Sapsaree indicated low levels of admixture with the Lhasa Apso and Tibetan terriers. Moreover, Sapsaree showed a small level of introgression with one of the oldest European breed Mastiffs ancestry, Great Pyrenees and the Old English Sheepdogs. However, admixture analysis indicated that





**Fig. 3** Estimated decay of heterozygosity over 50 generations. Lines are colored based on breeds



**Fig. 4** Population structure plots using K= 10 ancestry models. Each colored vertical line represents proportions of ancestral populations for each individual. K inferred the number of estimated ancestors and which differentiated by colors. Optimum K value was determined by Admixture's cross-validation (CV) procedure. (Additional file 1: Figure S1)

major ancestries of Sapsaree were not shared with the other breeds used in this study.

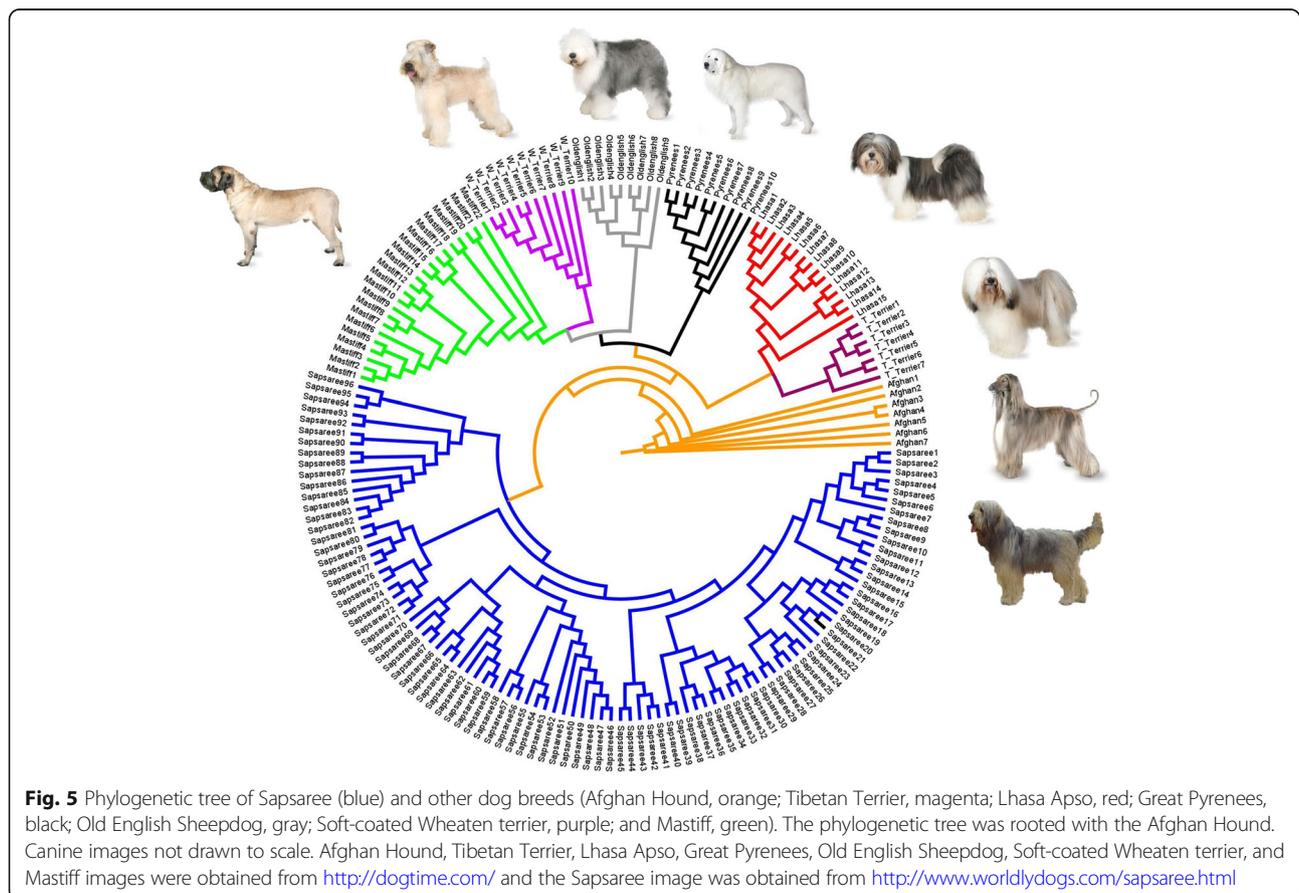
The phylogenetic tree clearly indicates a monophyletic clade of Sapsaree that is diverge from the other breeds, which supports the admixture analysis results (Fig. 5). The European breeds (Mastiff, Old English Sheepdog, Soft-coated Wheaten terrier, and Great Pyrenees) were grouped together in a single clade, and the Tibetan breeds (Tibetan Terrier and Lhasa Apso) comprise an adjacent monophyletic clade. The Afghan Hound was used as a root to construct the phylogenetic tree because it is an ancient breed, and more closer to a “real dog” than other domesticated breeds [7, 26, 49–51]. Our phylogenetic tree also indicates that the Afghan Hound is highly diverged from the other breeds.

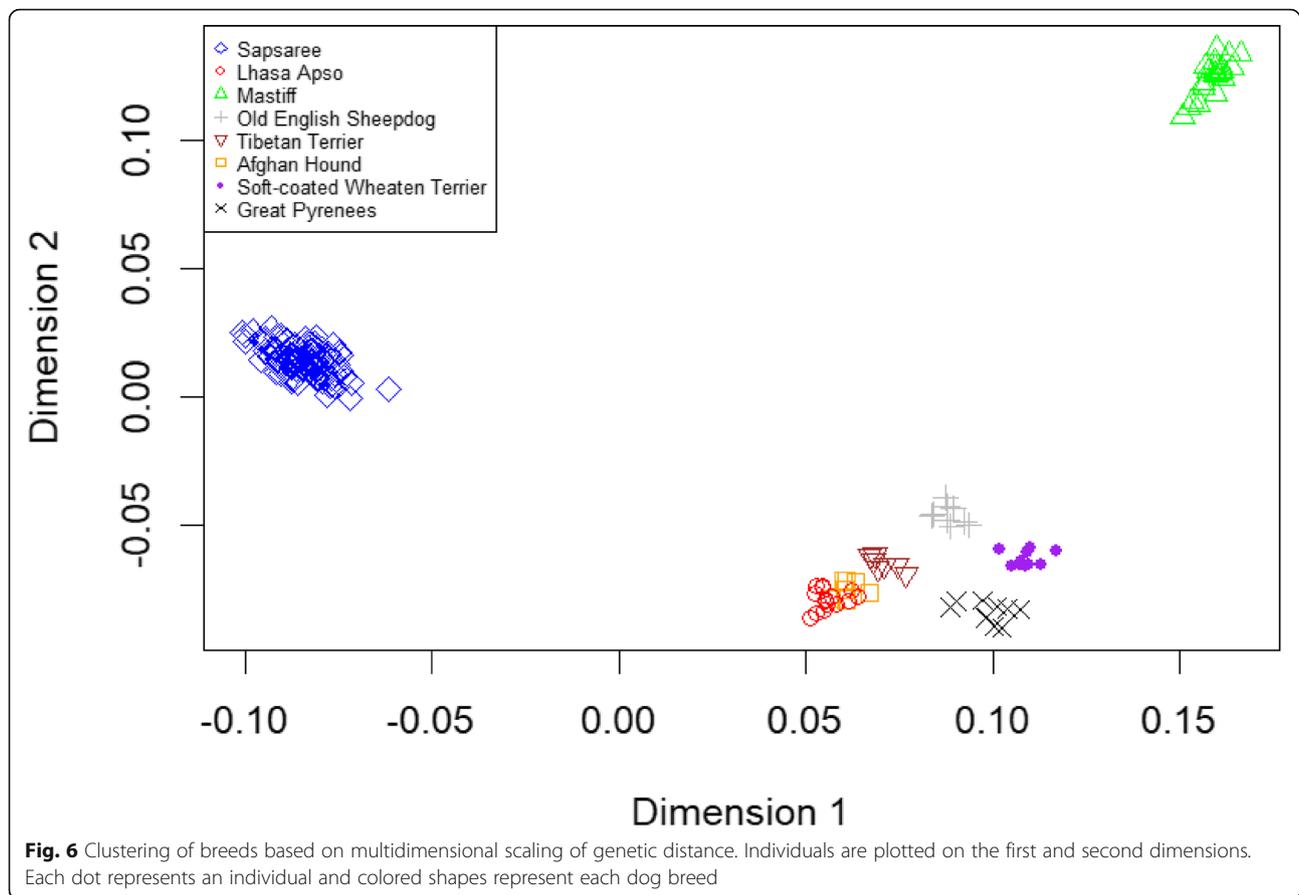
MDS analysis was used to visualize the quantitative estimates of genetic distance among the breeds (Fig. 6). Consistent with the admixture results, MDS also revealed that Sapsaree was clustered farthest from the other breeds, which supports assemblages into a single clade on the phylogenetic tree. However, Sapsaree clusters with the Mastiff, Old English Sheepdog, and Tibetan terrier when dimension 3 was plotted against dimension 4 (Additional file 2: Figure S2).

## Discussion

In this study, genome-wide SNP data was used to characterize the genetic diversity, population structure, and demographic history of an aboriginal Korean dog breed, the Sapsaree. The non-random association of genes at different loci is assessed as LD, which gives insight to the structure of present populations and evolutionary demographic events [52, 53]. Similar LD and  $N_e$  patterns in the Lhasa Apso and Sapsaree reflect their historical similarities [54]. Alam et al. [20] indicated that five generations ago, LD and  $N_e$  were approximately 0.2 and 64–75, respectively, which differs from our results. This variation may be due to discrepancies between samples and different algorithms used [6]. Ascertainment bias may have also caused the systematic deviation of population genetic structure from its theoretical expectations [55, 56].  $N_e$  has long been recognized as a useful criterion for evaluating conservation status and threats to the genetic health of a population [57]. Meuwissen. [58] suggested that a threshold level of 50 or 100 for  $N_e$  would be necessary to maintain viable genetic diversity. Our results also emphasize that care should be taken to maintain the reasonable genetic diversity of the Sapsaree breed.

Ancient events, as well as the recent breeding program, can lead to dramatic changes in the genetic diversity





among the individual dogs [6, 59–63]. Our analyses suggest that the Sapsaree has higher variance and discrete genetic compared to the other breeds studied here, consistent with the results of other studies [21, 23, 33]. Previous studies have also provided evidence that genetic diversity is high in dogs native to Korea [14, 21 9, 55] or East Asia [6, 64].

Heterozygosity is considered a useful parameter in estimating a population's genetic diversity [32, 52, 65], and the Sapsaree has shown greater heterozygosity compared with foreign breeds [21, 23, 33]. One study indicated that the observed and expected mean heterozygosities in the Sapsaree were 0.460 and 0.543, respectively [23]. A recent study by Choi et al. [55] has suggested high heterozygosity (0.4) in Korean dogs (Poongsan, Donggyengi and Jindo). However, compared with the previous studies, there was low heterozygosity in the Sapsarees in this study. We were also determined that the Tibetan Terrier exhibits greater heterozygosity than the Mastiff [66] and alignment with the present results Mortlock et al. [67] showed multiple-locus heterozygosity of Mastiff was 0.206.

Population bottlenecks can dramatically reduce the genetic diversity of populations [68–71], and Sapsarees have

experienced severe population bottlenecks during the Japanese colonial rule and the Korean War and subsequent economic crisis [18, 20, 23]. Interestingly, Sapsarees have still been able to maintain more genetic variation than other breeds.

Reductions in genetic variability or heterozygosity primarily depend on bottleneck size, rate of population growth, and mutation rate [72–74]. Although declines in genetic variability are expected following a bottleneck, variation may accumulate through mutations as the population size increases. Correspondingly, Kekkonen et al. [65] reported fairly high genetic diversity of white-tailed deer (*Odocoileus virginianus*) in Finland, even though the population was founded by four individuals in 1934 and remained isolated from other deer populations. In contrast, the German Leonberger breed had similar experience as which was nearly wiped out during World War I by violence and starvation. Their genetic variation drastically declined but was re-established in 1992 using five females and two males. However, their genetic variation was still low compared with other breeds [51, 69, 75].

Admixture, MDS, and phylogenetic analyses showed the unique diversity of the Sapsaree breed. Other studies

have also found that native Korean dogs have substantially different genetic patterns than other foreign dog breeds [33, 55]. Furthermore, admixture analysis (Fig. 4 and, Additional file 3: Figure. S3 and Additional file 4: Figure S4) and structure analysis (Additional file 5: Figure S5) revealed a greater genetic heterogeneity within the Sapsaree compared to the other breeds. The consequences of the restoration process might be a reason for the increased genetic diversity of the breed. In 1986, the Sapsaree population was restored using eight individuals collected based on their similar characteristics with the original breed such as color and body shape. A system of non-restricted selection was then established to increase the population size [18, 76]. In alignment with the present results, several previous studies showed a greater genetic diversity of Sapsaree compared with foreign dog breeds [21, 23, 33]. Moreover, a small fraction of Sapsaree deviated from major genetic patterns, also possibly as a consequence of recent restoration processes. Founder animals were collected based on phenotypic characteristics, which might lead some dogs having distance genetic pattern from majority of the Sapsaree population.

Correspondingly, Han et al. [22] showed that Sapsarees have greater genetic diversity based on several morphological traits such as tongue spots, dewclaws, tail-set, and coat, nose, and eye color. The Coat color of the Sapsaree also revealed the heterogeneous nature of the breed, indicating two distinct group of blue and yellow including several subdivisions such as blue black, grey black, deep yellow, yellow and light yellow [77]. On the other hand, some studies have also shown discrete phenotypic diversity such as Kim et al. [19] revealed that they can be divided into two groups based on gene expression patterns for physiological activities. Accordingly, the results suggest that systematic approach is needed to select the individuals for breeding to established the breed while ensuring the authenticity.

There was evidence of introgression into Sapsaree in the admixture analysis, which might have occurred prior to the restoration process when the population levels were low. Introgression from non-tested breeds could also have contributed to the high levels of genetic diversity noted in the Sapsaree. The admixture and MDS analyses provide compelling evidence that the ancestor of the Sapsaree is related to Tibetan long-haired breeds. The Tibetan Terrier and Lhasa Apso are native to Tibet, where they lived in nobles palaces and Buddhist monasteries as watch dogs, companions, and 'good luck charms'. There are definitive evidences that which used as a special gift, tokens of esteem and good fortune when spreading the Buddhism [78–82]. Buddhism was introduced to Korea in fourth century CE [83, 84], and the introgression of Tibetan dog breeds might be an outcome of that relationship.

Additionally, our results suggested the admixture of European dog breeds, which were introduced to the Korean peninsula as a result of cultural exchange. Christianity invaded Korea from Europe during the eighteenth century [81, 85], and some European dog breeds accompanied those missions. Afterwards, numerous European delegations and military correspondence with Korea occurred during World War I and the Korean War [86–88]. Furthermore, the Silk Road was a historical network of international trade routes from ancient China to Europe, stretching from Korea and Japan to the Mediterranean Sea. In addition to silk as the major commodity, companion animals were also exchanged on this route [89–91]. Comas et al. [92] suggested that genetic diversity was also traded along the Silk Road between Europe and eastern Asia. Consistent with our phylogenetic results, vonHoldt et al. [26] illustrated that European dog breeds, such as the Mastiff and Old English Sheepdog, are phylogenetically clustered, while Choi et al. [55] showed that the Tibetan Terrier and Lhasa Apso grouped into a single clade. Although, Jeong et al. [23] suggested a great genetic distance between Sapsaree and the European breeds, their structure analysis showed a low level of genetic sharing among them, which support the current findings.

## Conclusions

Our results provide novel information regarding the genetic diversity and population structure of the native Korean dog, Sapsaree. Consistent with previous studies, our results also revealed higher genetic diversity in Sapsarees compared with other breeds. The majority of the breed showed a discrete genetic pattern, while a small fraction was genetically divergent and might be a consequence of recent restoration process. The  $N_e$  of the breed has declined drastically and is currently insufficient to maintain long-term viability of genetic diversity. Therefore, we suggest a strategic and systematic approach to ensure the purity and genetic diversity of the Sapsaree breed, a Korean natural treasure. Admixture analysis revealed a complex pattern of Sapsaree, where major ancestries were not shared with the other breeds analyzed in this study. LD,  $N_e$ , genetic diversity, and population structural analyses indicate a relationship between Sapsaree and the long-haired breeds Tibetan Terrier and Lhasa Apso. Introgression from European breeds was also revealed.

## Methods

### Animals, genotyping, and quality control

All research methods were approved by the Institutional Animal Care and Use Committee of the Rural Development Administration in South Korea. To investigate the genetic origin of the Sapsaree breed, we selected seven foreign dog breeds analyzed in a previous study Shannon

et al. [34] based on their phenotypes, such as long haired and body conformation [35]. The Sapsaree ( $n = 96$ ), Lhasa Apso ( $n = 15$ ), Great Pyrenees ( $n = 10$ ), Tibetan Terrier ( $n = 7$ ), Afghan Hound ( $n = 7$ ), Old English Sheepdog ( $n = 9$ ), Soft-coated Wheaten Terrier ( $n = 10$ ), and Mastiff ( $n = 22$ ) dog breeds were categorized as ancient or modern breeds according to Vonholdt et al. [26] and Parker et al. [13]. Based on a memorandum of understanding between the research team and the Sapsaree conservation center, blood samples were collected by veterinarians in an ethical manner according to the animal health and welfare guidelines (Approval numbers: 2016–177).

Samples were genotyped using Illumina CanineSNP20 BeadChip. Other breeds were genotyped by [34] using the Illumina CanineHD array and merged into our dataset. The CanineSNP20 BeadChip is Illumina's first non-human standard genotyping panel contains more than 22,000 evenly spaced and validated SNP probes derived from the CanFam2.0 assembly. The CanineHD Genotyping BeadChip contains more than 170,000 markers placed also on the CanFam2.0 reference sequence. This presents an average of greater than 70 markers per megabase (Mb), providing ample SNP density for robust within-breed association and copy number variation (CNV) studies ([www.illumina.com](http://www.illumina.com)). The quality of SNP data was maintained with the use of PLINK 1.9 [36] to filter SNPs with low call rates (< 90%) or missing genotypes (> 10%). To reduce bias, the number of minor allele frequencies was limited to 1%, and deviations from Hardy-Weinberg equilibrium ( $P > 0.001$ ) were also excluded [37]. Non-autosomal SNPs were also removed from analyses.

#### Linkage disequilibrium, effective population size, and heterozygosity

The extent of LD between markers was measured using the squared correlation coefficient of allele frequencies at pairs of loci ( $r^2$ ) with inter-SNP distance within 1 Mb, both within a given breed and across all breeds [38]. Pairwise LD between adjacent SNPs was calculated for each chromosome using the default PLINK V1.9 approach [39]. Effective population size ( $N_e$ ) was estimated based on the LD value ( $r^2$ ) using the SNeP V1.1 tool [29, 40–42]. Heterozygosity over the next 50 generations was estimated as described by [43]. Statistical software package R [44] was used to produce graphical representations. Wright–Fisher model was used to calculate the forward derivation of heterozygosity, assuming that  $N$  diploid parents produce a large number of gametes, these gametes randomly unite to produce a large number of zygotes, and from these zygotes,  $N$  progeny are randomly chosen to form the next generation [43].

#### Genetic diversity and population structure

Population structure and genetic diversity were studied using multi-dimensional scaling (MDS) analysis, ancestor's admixture prediction, and phylogenetic comparisons. To create a matrix representation of interbreed relationships, MDS algorithms of pairwise genetic distances were implemented in PLINK [39] and depicted as coordinates in R. Population substructures and the extent of mixture between ancestral populations of Sapsaree and unrelated individuals of other studied breeds were evaluated through the model-based clustering algorithm using ADMIXTURE v.1.23 [45]. To reduce prediction error, admixture's cross-validation (CV) procedure was used to determine the optimal K-value by minimizing CV error. These results were graphed using R. A phylogenetic tree was developed using the SNPhylo software package and illustrated using FigTree software v. 1.4.2 to infer the evolutionary relationships among breeds [46, 47].

#### Additional files

**Additional file 1: Figure S1.** Cross-validation plot of admixture analysis. The x-axis represents the number of clusters (K) in the model and the y-axis represents cross-validation error values. (DOCX 33 kb)

**Additional file 2: Figure S2.** Clustering of animals from Sapsaree and other selected breeds based on multidimensional scaling of genetic distance. Individuals are plotted for the third and fourth dimension. (DOCX 22 kb)

**Additional file 3: Figure S3.** Population structure plots using K = 8 and K = 11 ancestry models. Each colored vertical line represents proportions of ancestral populations for each individual. K inferred the number of estimated ancestors and which differentiated by colors. (DOCX 214 kb)

**Additional file 4: Figure S4.** Ancestry model for Sapsaree including related dog breeds based on the genetic distance. Each colored vertical line represents proportions of ancestral populations for each individual. K inferred the number of estimated ancestors and which differentiated by colors. Optimum K value (K = 16) was determined by Admixture's cross-validation (CV) procedure. (DOCX 64 kb)

**Additional file 5: Figure S5.** The population structure bar plots generated by STRUCTURE software at K = 8. (DOCX 73 kb)

**Additional file 6: Figure S6.** Heat map of relatedness between the individuals of Sapsaree and other studied breeds. (DOCX 95 kb)

**Additional file 7: File S7.** Genotype information of Sapsaree. (BED 524 kb)

**Additional file 8: File S8.** Genotype information of Sapsaree. (BIM 697 kb)

**Additional file 9: File S9.** Genotype information of Sapsaree. (FAM 2 kb)

#### Abbreviations

Afghan: Afghan Hound; CV: Cross-validation; H: Heterozygosity; LD: Linkage Disequilibrium; Lhasa: Lhasa Apso; MDS: Multi-dimensional scaling;  $N_e$ : Effective Population size; Oldenglish: Old English Sheepdog; Pyrenees: Great Pyrenees; T\_Terrier: Tibetan Terrier; W\_Terrier: Soft-coated Wheaten Terrier

#### Acknowledgments

We gratefully acknowledge Shannon et al., 2015 for the genotype data of foreign dog breeds.

#### Authors' contributions

CJG and JMK interpreted the results, wrote the manuscript & editing, Data analysis & Visualization were performed by JMK, DHL, CJG, YKK, SHL2, HIW, and JJK and JHH Contributed to the study design and helped to draft the manuscript, BHC performed data collection, data generation, resources, review and editing, Role of

SHL1 were conceptualization, investigation, methodology, project administration, validation, results interpretation, review and editing. All authors have read and approved the final manuscript.

#### Funding

This study was supported by awards from the AGENDA project (grant no. PJ011950022017) of the National Institute of Animal Science at the Rural Development Administration, Republic of Korea. The funding body had no role in the design of the study, collection, analysis, and interpretation of data, or in writing the manuscript.

#### Availability of data and materials

All data generated or analyzed during this study are included in this published article [and its supplementary information files]. Sapsaree genotype data have been uploaded as Additional files 6, 7, 8 and 9. Genotype data of the other breeds are available from Shannon et al., 2015 (<https://doi.org/10.5061/dryad.v9t5h>).

#### Ethics approval and consent to participate

Advance approval was acquired from the Institutional Animal Care and Use Committee of the National Institute of Animal Science, of the Rural Development Administration, of South Korea (Approval numbers: 2016–177).

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Division of Animal & Dairy Science, Chungnam National University, Daejeon 34134, Republic of Korea. <sup>2</sup>Animal Genomics & Bioinformatics Division, National Institute of Animal Science, RDA, Wanju 55365, Republic of Korea. <sup>3</sup>School of Biotechnology, Yeungnam University, Gyeongsan 712-749, Republic of Korea. <sup>4</sup>School of Life Science, Kyungpook National University, Daegu 41940, Republic of Korea. <sup>5</sup>Department of Animal Science, Faculty of Agriculture, University of Ruhuna, Matara, Sri Lanka.

Received: 17 August 2018 Accepted: 20 June 2019

Published online: 05 August 2019

#### References

- Serpell J, Clutton-Broke J, Coppinger R, Schneider R, Willis MB, Benjamin L. In: Serpell J, editor. *The domestic dog: its evolution behavior and interactions with people*. Cambridge: Cambridge University press; 1995. p. 7–20.
- Shearman JR, Wilton AN. Origins of the domestic dog and the rich potential for gene mapping. *Genet Res Int*. 2011;2011. <https://doi.org/10.4061/2011/579308>.
- Vila C, Jennifer AL. Canid phylogeny and origin of the domestic dog. In: Ostrander EA, Ruvinsky A. *The genetics of the dog*. Oxfordshire: CAB; 2012. p. 1–9.
- Spady TC, Ostrander EA. Canine behavioral genetics: pointing out the phenotypes and herding up the genes. *Am J Hum Genet*. 2008;82:10–8. <https://doi.org/10.1016/j.ajhg.2007.12.001>.
- Ostermeier M. History of guide dog use by veterans. *Mil Med*. 2010;175(8):587–93.
- Wang GD, Zhai W, Yang HC, Wang L, Zhong L, Liu YH, et al. Out of southern East Asia: the natural history of domestic dogs across the world. *Nat Publ Gr*. 2015;26:21–3321. <https://doi.org/10.1038/cr.2015.147>.
- Tanabe Y. Phylogenetic studies of dogs with emphasis on Japanese and Asian breeds. *Proc Jpn Acad, Ser B*. 2006;82 [https://www.jstage.jst.go.jp/article/pjab/82/10/82\\_10\\_375/\\_pdf](https://www.jstage.jst.go.jp/article/pjab/82/10/82_10_375/_pdf). Accessed 13 Dec 2017.
- Fan Z, Silva P, Gronau I, Wang S, Armero AS, Schweizer RM, et al. Worldwide patterns of genomic variation and admixture in gray wolves. *Genome Res*. 2016;26:163–73. <https://doi.org/10.1101/gr.197517.115>.
- Savolainen P, Zhang Y, Luo J, Lundeberg J, Leitner T. Genetic evidence for an east Asian origin of domestic dogs. *Science*. 2002;298:1610–3. <https://doi.org/10.1126/science.1073906>.
- Ding ZL, Oskarsson M, Ardalán A, Angleby H, Dahlgren LG, Tepeli C, et al. Origins of domestic dog in southern East Asia is supported by analysis of Y-chromosome DNA. *Heredity (Edinb)*. 2012;108:507–14. <https://doi.org/10.1038/hdy.2011.114>.
- Wang J, Santiago E, Caballero A. Prediction and estimation of effective population size. *Heredity (Edinb)*. 2016;117:193–206. <https://doi.org/10.1038/hdy.2016.43>.
- Cho GJ. Microsatellite polymorphism and Genetic relationship in dog breeds in Korea. *Asian-Australasian J Anim Sci*. 2005;18:1071–4. <https://doi.org/10.5713/ajas.2005.1071>.
- Parker HG, Dreger DL, Rimbault M, Davis BW, Mullen AB, Carpintero-Ramirez G, Ostrander EA. Genomic analyses reveal the influence of geographic origin, migration, and hybridization on modern dog breed development. *Cell Rep*. 2017;19:697–708. <https://doi.org/10.1016/j.celrep.2017.03.079>.
- Kang BT, Kim KS, Min MS, Chae YJ, Kang JW, Yoon J, et al. Microsatellite loci analysis for the genetic variability and the parentage test of five dog breeds in South Korea. *Genes Genet Syst*. 2009;84:245–51 <http://www.ncbi.nlm.nih.gov/pubmed/19745574>. Accessed 18 Dec 2017.
- Yoo D, Kim K, Kim H, Cho S, Kim JN, Lim D, et al. The Genetic origin of short tail in endangered Korean dog, DongGyeonggi. *Sci Rep*. 2017;7:10048. <https://doi.org/10.1038/s41598-017-10106-6>.
- Lee HE, Choi BH, Lee DH, Kwon YJ, Eo J, Choi Y, et al. Polymorphism analysis of tyrosine hydroxylase gene variable number of tandem repeats in various Korean dogs. *Genes Genomics*. 2015;37:257–61. <https://doi.org/10.1007/s13258-014-0243-4>.
- Skabelund AH. *Empire of dogs : canines, Japan, and the making of the modern imperial world*, vol. 196. Ithaca: Cornell University Press; 2011.
- Ha JH, Alam M, Lee DH, Kim JJ. Whole genome association study to detect single nucleotide polymorphisms for behavior in Sapsaree dog (*Canis familiaris*). *Asian-Australasian J Anim Sci*. 2015;28:936–42. <https://doi.org/10.5713/ajas.14.0941>.
- Kim JE, Choe J, Lee JH, Kim WB, Cho W, Ha JH, et al. Whole-transcriptome analyses of the Sapsaree, a Korean natural monument, before and after exercise-induced stress. *J Anim Sci Technol*. 2016;58:17. <https://doi.org/10.1186/s40781-016-0097-1>.
- Alam M, Han KI, Lee DH, Ha JH, Kim JJ. Estimation of effective population size in the Sapsaree: a Korean native dog (*Canis familiaris*). *Asian-Australasian J Anim Sci*. 2012;25:1063–72. <https://doi.org/10.5713/ajas.2012.12048>.
- Kim KS, Tanabe Y, Park CK, Ha JH. Genetic variability in east Asian dogs using microsatellite loci analysis. *J Hered*. 2001;92:398–403. <https://doi.org/10.1093/jhered/92.5.398>.
- Han KI, Alam M, Lee YM, Lee DH, Ha JH, Kim JJ. A study on morphology and behavior of the Sapsaree : a Korean native dog (*Canis familiaris*). *J Anim Sci Technol*. 2010;52:481–90. <https://doi.org/10.5187/JAST.2010.52.6.481>.
- Jeong H, Choi B-H, Eo J, Kwon YJ, Lee HE, Choi Y, et al. Statistical analysis and genetic diversity of three dog breeds using simple sequence repeats. *Genes Genomics*. 2014;36:883–9. <https://doi.org/10.1007/s13258-014-0237-2>.
- Al-Mamun HA, A Clark S, Kwan P, Gondro C. Genome-wide linkage disequilibrium and genetic diversity in five populations of Australian domestic sheep. *Genet Sel Evol* 2015;47:90. doi:<https://doi.org/10.1186/s12711-015-0169-6>.
- Wulfsch C, Caragiulo A, Dias-Freedman I, Quigley H, Rabinowitz S, Amato G. Genetic diversity and population structure of Mesoamerican jaguars (*Panthera onca*): implications for conservation and management. *PLoS One*. 2016;11:e0162377. <https://doi.org/10.1371/journal.pone.0162377>.
- vonHoldt BM, Pollinger JP, Lohmueller KE, Han E, Parker HG, Quignon P, et al. Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature*. 2010;464:898–902. <https://doi.org/10.1038/nature08837>.
- Decker JE, McKay SD, Rolf MM, Kim J, Molina Alcalá A, Sonstegard TS, et al. Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. *PLoS Genet*. 2014;10:e1004254. <https://doi.org/10.1371/journal.pgen.1004254>.
- Falconer DS, Mackay TF. *Introduction to quantitative genetics* (4th edn). Pearson United Kingdom; 1996. p. 48–81.
- Hayes BJ, Visscher PM, McPartlan HC, Goddard ME. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Res*. 2003;13:635–43. <https://doi.org/10.1101/gr.387103>.
- Shin DH, Cho KH, Park KD, Lee HJ, Kim H. Accurate estimation of effective population size in the Korean dairy cattle based on linkage disequilibrium corrected by genomic relationship matrix. *Asian-Australasian J Anim Sci*. 2013;26:1672–9. <https://doi.org/10.5713/ajas.2013.13320>.
- Frankham R. Conservation genetics. *Annu Rev Genet*. 1995;29:305–27. <https://doi.org/10.1146/annurev.ge.29.120195.001513>.
- Toro MA, Caballero A. Characterization and conservation of genetic diversity in subdivided populations. *Philos Trans R Soc B Biol Sci*. 2005;360:1367–78. <https://doi.org/10.1098/rstb.2005.1680>.

33. Lee EW, Choi SK, Cho GJ. Molecular Genetic diversity of the Gyeongju Donggyeong dog in Korea. *J Vet Med Sci.* 2014;76:14–189. <https://doi.org/10.1292/jvms.14-0189>.
34. Shannon LM, Boyko RH, Castelano M, Corey E, Hayward JJ, McLean C, et al. Genetic structure in village dogs reveals a central Asian domestication origin. *Proc Natl Acad Sci U S A.* 2015;112(44):13639. <https://doi.org/10.1073/pnas.1516215112>.
35. Young A, Bannasch D. Morphological variation in the dog. In: Ostrander EA, Giger U, Lindblad-Toh K. The dog and its genome. Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 2006. p. 47–63.
36. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience.* 2015;4:7. <https://doi.org/10.1186/s13742-015-0047-8>.
37. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. *Nat Protoc.* 2010;5:1564–73. <https://doi.org/10.1038/nprot.2010.116>.
38. Hill WG, Robertson A. Linkage disequilibrium in finite populations. *Theor Appl Genet.* 1968;38(6):226–31. <https://doi.org/10.1007/bf01245622>.
39. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PI, Daly MJ, Sham PC. Plink: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81:559–75. <https://doi.org/10.1086/519795>.
40. Sved JA. Linkage disequilibrium and homozygosity of chromosome segments. *Theor Popul Biol.* 1971;14:125–41. [https://doi.org/10.1016/0040-5809\(71\)90011-6](https://doi.org/10.1016/0040-5809(71)90011-6).
41. Sargolzaei M, Schenkel FS, Jansen GB, Schaeffer LR. Extent of linkage disequilibrium in Holstein cattle in North America. *J Dairy Sci.* 2008;91:2106–17.
42. Barbato M, Orozco-terWengel P, Tapio M, Bruford MW. SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Front Genet.* 2015;6:109.
43. Genetic Drift HP. Effective population size, chapter 4. In: Genetics of populations, 4th edition, Jones and Bartlett publishers; 2011. p. 196–8.
44. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2016; ISBN 3–900051–07-0, URL <http://www.R-project.org/>.
45. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 2009;19(9):1655–64. <https://doi.org/10.1101/gr.094052.109>.
46. Lee TH, Guo H, Wang X, Kim C, Paterson AH. SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genomics.* 2014;15:162. <https://doi.org/10.1186/1471-2164-15-162>.
47. Rambaut, A. FigTree v1.4.2 molecular evolution, phylogenetics and epidemiology, Institute of Evolutionary Biology, University of Edinburgh. 2014; (<http://tree.bio.ed.ac.uk/software/figtree/>).
48. Brito LF, Jafarikia M, Grossi DA, Kijas JW, Porto-Neto LR, Ventura RV, et al. Characterization of linkage disequilibrium, consistency of gametic phase and admixture in Australian and Canadian goats. *BMC Genet.* 2015;16:67. <https://doi.org/10.1186/s12863-015-0220-1>.
49. Bettany S, Daly R. Figuring companion-species consumption: a multi-site ethnography of the post-canine afghan hound. *J Bus Res.* 2008;61:408–18. <https://doi.org/10.1016/J.JBUSRES.2006.08.010>.
50. Niblock M. The afghan hound : a definitive study. Arco Pub; 1980. [https://bepl.ent.sirsi.net/client/en\\_US/default/search/detailnonmodal/ent:\\$002f5002f5D\\_ILS5002f0\\$002f5D\\_ILS:185175/ada](https://bepl.ent.sirsi.net/client/en_US/default/search/detailnonmodal/ent:$002f5002f5D_ILS5002f0$002f5D_ILS:185175/ada). Accessed 5 Dec 2017.
51. Dobson JM. Breed-predispositions to cancer in pedigree dogs. *ISRN Vet Sci.* 2013;2013:941275. <https://doi.org/10.1155/2013/941275>.
52. Sharma A, Lee SH, Lim D, Chai HH, Choi BH, Cho Y. A genome-wide assessment of genetic diversity and population structure of Korean native cattle breeds. *BMC Genet.* 2016;17:139. <https://doi.org/10.1186/s12863-016-0444-8>.
53. Sutter NB, Eberle MA, Parker HG, Pullar BJ, Kirkness EF, Kruglyak L, et al. Extensive and breed-specific linkage disequilibrium in *Canis familiaris*. *Genome Res.* 2004;14:2388–96. <https://doi.org/10.1101/gr.3147604>.
54. Hayes BJ. Ben Hayes Course Notes, Toulouse. 2011. [http://snp.toulouse.inra.fr/~alegarra/ben\\_hayes\\_course/toulouse\\_course\\_notes.pdf](http://snp.toulouse.inra.fr/~alegarra/ben_hayes_course/toulouse_course_notes.pdf). Accessed 4 Jan 2018.
55. Choi BH, Wijayananda HI, Lee SH, Lee DH, Kim JS, Oh S II, et al. genome-wide analysis of the diversity and ancestry of Korean dogs. *PLoS One.* 2017; 12:e0188676. <https://doi.org/10.1371/journal.pone.0188676>.
56. Lachance J, Tishkoff SA. SNP ascertainment bias in population genetic analyses: why it is important, and how to correct it. *BioEssays.* 2013;35:780–6. <https://doi.org/10.1002/bies.201300014>.
57. Hare MP, Nunney L, Schwartz MK, Ruzzante DE, Burford M. Understanding and estimating effective population size for practical application in marine species management. *Conserv Biol.* 2011;25(3):438–49. <https://doi.org/10.1111/j.1523-1739.2010.01637.x>.
58. Meuwissen T. Genetic management of small populations: a review. *Acta Agric Scand Sect A - Anim Sci.* 2009;59:71–9. <https://doi.org/10.1080/09064700903118148>.
59. Hague MTJ, Routman EJ. Does population size affect genetic diversity? A test with sympatric lizard species. *Heredity (Edinb).* 2016;116:92–8. <https://doi.org/10.1038/hdy.2015.76>.
60. Jansson M, Laikre L. Pedigree data indicate rapid inbreeding and loss of genetic diversity within populations of native, traditional dog breeds of conservation concern. *PLoS One.* 2018;13:e0202849. <https://doi.org/10.1371/journal.pone.0202849>.
61. Vila C, Maldonado J, Wayne R. Phylogenetic relationships, evolution, and genetic diversity of the domestic dog. *J Hered.* 1999;90:71–7. <https://doi.org/10.1093/jhered/90.1.71>.
62. Kimura M. The neutral theory of molecular evolution: Cambridge University Press; 1983.
63. Oldenbroek K, Van der Waaij L. Textbook animal breeding and genetics for BSc students. Chapter 6.2.2: loss of genetic diversity: selection: Centre for Genetic Resources. The Netherlands and Animal Breeding and Genomics Centre; 2015.
64. Olsen SJ, Olsen JW, Luo J, Lundeberg J, Leitner T. The Chinese wolf, ancestor of New World dogs. *Science (80- ).* 1977;197:533–5. <https://doi.org/10.1126/science.197.4303.533>.
65. DeGiorgio M, Jankovic I, Rosenberg NA. Unbiased estimation of gene diversity in samples containing related individuals: exact variance and arbitrary ploidy. *Genetics.* 2010;186:1367–87. <https://doi.org/10.1534/genetics.110.121756>.
66. Parker GH, Sutter NB, Ostrander EA. Understanding genetic relationship among purebred dogs: the phyDO project, chapter 9. 2006. In: Ostrander, E. a., Giger, U., and Lindblad-Toh, K. the dog and its genome: Cold Spring Harbor Laboratory Press; 2006. p. 141–55.
67. Mortlock SA, Khatkhat MS, Williamson P. Comparative analysis of genome diversity in bullmastiff dogs. *PLoS One.* 2016;11:e0147941. <https://doi.org/10.1371/journal.pone.0147941>.
68. Marsden CD, Ortega-Del Vecchyo D, O'Brien DP, Taylor JF, Ramirez O, Vilà C, et al. Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *Proc Natl Acad Sci U S A.* 2016;113:152–7. <https://doi.org/10.1073/pnas.1512501113>.
69. Wayne RK, Ostrander EA. Lessons learned from the dog genome. *Trends Genet.* 2007;23:557–67. <https://doi.org/10.1016/J.TIG.2007.08.013>.
70. Kekkonen J, Wikström M, Brommer JE. Heterozygosity in an isolated population of a large mammal founded by four individuals is predicted by an individual-based Genetic model. *PLoS One.* 2012;7:e43482. <https://doi.org/10.1371/journal.pone.0043482>.
71. Ostrander E. A genetics and the shape of dogs. *Amer Sci.* 2007;95:406–13. [http://www.lualdalmatians.com/ostrander\\_article.pdf](http://www.lualdalmatians.com/ostrander_article.pdf). Accessed 12 Dec 2017.
72. Nei M, Maruyama T, Chakraborty R. The bottleneck effect and Genetic variability in populations. *Evolution (N Y).* 1975;29(1). <https://doi.org/10.2307/2407137>.
73. Chakraborty R, Nei M. Bottleneck effects on average heterozygosity and Genetic distance with the stepwise mutation model. *Evolution (N Y).* 1977; 31:347. <https://doi.org/10.2307/2407757>.
74. Cabe PR. The effects of founding bottlenecks on genetic variation in the European starling (*Sturnus vulgaris*) in North America. *Heredity (Edinb).* 1998;80:519–25. <https://doi.org/10.1046/j.1365-2540.1998.00296.x>.
75. Clark LA, Starr-Moss A. Genetics and genomics of the domestic dog. Chapter. 12 in; Khatib H. molecular and quantitative animal genetics. Hoboken: Wiley; 2015. p. 121–9.
76. The Korean Sapsaree Foundation. Sapsaree and the Korean Nation; Restoration of the Sapsaree <http://koreansapsaree.weebly.com/history.html> and <http://www.sapsaree.org/html/>. Accessed 8 Feb 2019.
77. Ha JH, Chung WB, Lee SL, Tak RB, Kim JB. Studies on the Characteristics of Coat Color and Genetic Inter-Relationship among Korean Native Dog, Sapsaree. *Korean J Genetics.* 1991;13(4):247–54. <http://www.papersearch.net/thesis/article.asp?key=166664>.
78. Debroy B. Sarama and her children : the dog in Indian myth. New Delhi: Penguin Books; 2008. p. 1–16.
79. Cunliffe J. Lhasa Apso : a comprehensive guide to owning and caring for your dog. Freehold: Kennel Club Books; 2012. p. 9–25.
80. Sefton F, Schneider E. Know your Lhasa apso. Pet Library; 2011. p. 5–8.

81. Clark RD. Medical, Genetic & Behavioral Risk Factors of Tibetan Terriers. Bloomington: Xlibris; 2014. p. 1–2.
82. Hafer T, Hafer J. 101 Amazing Things About Dog Lovers Broadstreet Publishing Group, LLC, vol. 5; 2017.
83. Grayson JH. Early Buddhism and Christianity in Korea : a study in the emplantation of religion. Leiden: E.J. Brill; 1985. p. 1–20.
84. Lancaster LR, Yu CS. Introduction of Buddhism to Korea : new cultural patterns. Fremont: Asian Humanities Press; 1989. p. 1–4.
85. Buswell RE, Lee TS. Christianity in Korea. Honolulu: University of Hawai'i; 2006. p. 7–21.
86. Schratz PR. Submarine commander : a story of world war II and Korea. Lexington: University Press of Kentucky; 1988. p. 52–100.
87. Edwards PM. United Nations participants in the Korean war : the contributions of 45 member countries. Jefferson: McFarland & Company, Inc. Publishers; 2013. p. 19–29.
88. Millett AR. The Korean war. Korea Institute of Military History. Lincoln: University of Nebraska Press; 2000.
89. Liu X. The silk road in world history: Oxford University Press, Inc; 2010. p. 42–62. [http://14.139.206.50:8080/jspui/bitstream/1/2626/1/Liu - the Silk Road in World History.pdf](http://14.139.206.50:8080/jspui/bitstream/1/2626/1/Liu%20-%20the%20Silk%20Road%20in%20World%20History.pdf). Accessed 29 Dec 2017
90. Wood F. The silk road : two thousand years in the heart of Asia. Berkeley: University of California Press; 2002. p. 8–28.
91. Whitfield S, Sims-Williams U, Library B. The silk road : trade, travel, war and faith, vol. 24. Chicago: Serindia Publications; 2004.
92. Comas D, Calafell F, Mateu E, Pérez-Lezaun A, Bosch E, Martínez-Arias R, et al. Trading genes along the silk road: mtDNA sequences and the origin of central Asian populations. *Am J Hum Genet.* 1998;63:1824–38. <https://doi.org/10.1086/302133>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

