Characterization of the genetic profile of five Danish dog breeds[1]

C. Pertoldi,*†‡[2] T. N. Kristensen,*†§ V. Loeschcke,* P. Berg,§ A. Praebel,§ A. V. Stronen,# H. F. Proschowsky,‖ and M. Fredholm¶

*Department of Bioscience, Aarhus University, Ny Munkegade 116, DK-8000 Aarhus C, Denmark; †Department of Biotechnology, Chemistry and Environmental Engineering - Section of Biology and Environmental Science, Aalborg University, Aalborg, Sohngårdsholmsvej 57, DK-9000, Denmark; ‡Aalborg Zoo, Aalborg, Denmark, §NordGen– Nordic Genetic Resource Center, P.O. Box 115, NO-1431 Ås, Norway; #Mammal Research Institute, Polish Academy of Sciences, 17-230 Białowieża, Poland; ‖The Danish Kennel Club, Parkvej1, DK-2680 Solrød Strand, Denmark; and ¶Department of Veterinary Clinical and Animal Sciences, University of Copenhagen, Grønnegårdsvej 3, DK-1870 Frederiksberg C, Denmark

ABSTRACT: This investigation presents results from a genetic characterization of 5 Danish dog breeds genotyped on the CanineHD BeadChip microarray with 170,000 SNP. The breeds investigated were 1) Danish Spitz (DS; *n* = 8), 2) Danish-Swedish Farm Dog (DSF; *n* = 18), 3) Broholmer (BR; *n* = 22), 4) Old Danish Pointing Dog (ODP; *n* = 24), and 5) Greenland Dog (GD; *n* = 23). The aims of the investigation were to characterize the genetic profile of the abovementioned dog breeds by quantifying the genetic differentiation among them and the degree of genetic homogeneity within breeds. The genetic profile was determined by means of principal component analysis (PCA) and through a Bayesian clustering method. Both the PCA and the Bayesian clustering method revealed a clear genetic separation of the 5 breeds. The level of genetic variation within the breeds varied. The expected heterozygosity (H_E) as well as the degree of polymorphism (P%) ranked the dog breeds in the order DS > DSF > BR > ODP > GD. Interestingly, the breed with a tenfold higher census population size compared to the other breeds, the Greenland Dog, had the lowest within-breed genetic variation, emphasizing that census size is a poor predictor of genetic variation. The observed differences in variation among and within dog breeds may be related to factors such as genetic drift, founder effects, genetic admixture, and population bottlenecks. We further examined whether the observed genetic patterns in the 5 dog breeds can be used to design breeding strategies for the preservation of the genetic pool of these dog breeds.

Key words: CanineHD BeadChip, inbreeding, , PLINK, single nucleotide polymorphisms, STRUCTURE

 J. Anim. Sci. 2013.91:5122–5127
doi:10.2527/jas2013-6617

INTRODUCTION

Many domestic breeds have effective population sizes (N_E) ranging from less than 100 to a few hundred individuals, which suggests that genetic drift is likely to diminish the variation within breeds (e.g., Leroy, 2011). Among the domestic breeds, several dog breeds are also considered to have a small effective population size (Rooney, 2009). Distinct dog breeds have been observed since antiquity and separation of dog populations into closed breeds during the 19th century, together with selection for specific physical attributes, have led to an increase in differentiation among breeds (Clutton-Brock, 1999). In some breeds crossbreeding or temporary open studbooks should be considered due to small N_E.

More than 300 dog breeds are recognized by the International Dog Society (FCI; www.fci.be). Each breed is under the responsibility of a specific country. Two breed registries are associated with Denmark via Kennel Clubs and similar breed organizations although not fully recognized by the FCI: the Danish Spitz (**DS**) and the Danish-Swedish Farm Dog (**DSF**). A total of 130 dogs were chosen as original founders but the studbook remained open. Furthermore, 3 dog breeds, which are recognized by the FCI, are considered to "belong to"

[1]CP was supported by a Marie Curie Transfer of Knowledge Fellowship (project BIORESC in the 6th FP, contract no. MTKD-CT-2005-029957). We thank the Danish Natural Science Research Council (grant numbers: 11-103926, 09-065999, 95095995), the Carlsberg Foundation (grant number 2011-01-0059) and the Danish Finance Act 2012 § 24.21.02.35 for financial support to the project.

[2]Corresponding author: Cino Pertoldi, biocp@nf.au.dk

Received April 20, 2013.
Accepted August 5, 2013.

Denmark: the Broholmer breed (**BR**), the Old Danish Pointing Dog (**ODP**), and the Greenland Dog (**GD**).

The aims of this investigation were to characterize the genetic profile of the abovementioned dog breeds by quantifying the genetic differentiation among and within these breeds. We genotyped dogs from the 5 breeds using the CanineHD BeadChip, which allows genotyping of up to 170,000 single nucleotide polymorphism (SNP) markers (Lequarré et al., 2011). We quantified variation within and among the 5 breeds and examined genetic patterns for each breed to determine whether the results can be applied for designing breeding strategies aimed at preserving the genetic variation of these dog breeds.

MATERIALS AND METHODS

Animal Care and Use Committee approval was not obtained for this study because no animals were used.

Extraction of DNA and Genotyping

We collected EDTA-stabilized blood samples from the 5 breeds ($n = 95$ dogs). The samples have been sampled from privately owned dogs during the period from 2003 to 2012. With respect to DS, DSF, BR, and ODP only individuals unrelated at the parental level were included in this study. The GD was sampled in Greenland where studbooks are not maintained. Although the sampling was made to avoid close relationship between dogs, the GD dogs might be more closely related than the other dogs included in this study. Samples were genotyped for 172,155 loci using the CanineHD BeadChip microarray from Illumina (Illumina, Inc., San Diego, CA). Samples included 1) DS ($n = 8$), 2) DSF ($n = 18$), 3) BR ($n = 22$), 4) ODP ($n = 24$), and 5) GD ($n = 23$).

We used GenomeStudio and accompanying guidelines from Illumina (www.illumina.com) to identify individuals suitable for analyses of the genetic profile. Single nucleotide polymorphism calling rates ranged from 99.03 to 99.80% and the average call rate was 99.69%.

We performed additional quality control in GenomeStudio, which resulted in a set of 169,106 loci with an average call rate of 99.92%. This data set was filtered in PLINK (Purcell et al., 2007) to retain loci with a minor allele frequency (**MAF**) of 0.01 (MAF = 0.01) and a maximum per-SNP missing rate of 0.02 (genotypes [geno] = 0.02).

Following data evaluation in GenomeStudio, including removal of SNP on the X and Y chromosomes, we estimated genetic variation, including observed heterozygosity and expected heterozygosity ($\mathbf{H_E}$) and percent polymorphic loci degree of polymorphism (**P%**) in PLINK. To test if the H_E values within breeds were significantly different from each other we performed a 1-way ANOVA followed by pairwise Tukey's tests.

The number of SNP retained for calculations after the pruning process which consisted in (1) selecting only SNP with MAF > 0.01 (2) geno > 0.02, (3) max individual missing rate (mind) > 0.4, which corresponds to the removal of samples with sample call rate < 60% (4) removal of SNP with pairwise genotypic associations (r2) > 0.8 within a window of 50 SNP: Plink command: "indep-pairwise 50 5 0.8" per breed were 69,775 for DS, 107,717 for DSF, 74,191 for BR, 75,388 for oDH, and 88,404 for GD. The different number of SNP retained in the different breeds reflect different levels of linkage disequilibrium among loci and different amounts of minor alleles.

Statistical Analyses of Genetic Profiles

We evaluated the population genetic profiles using a Bayesian inference model in the program STRUCTURE 2.3.3 (Pritchard et al., 2000). We used 10,000 burn-in runs followed by 10,000 Markov Chain Monte Carlo repetitions and evaluated 4 possible population clusters ($K = 3$–6). Each parameter setting was repeated 3 times. We used the admixture model and correlated allele frequencies option. We used STRUCTURE HARVESTER v.06.92 (Earl and VonHoldt, 2012) and CLUMPP v1.1.2 (Jakobsson and Rosenberg, 2007) to summarize the output, which included estimates for delta K (Evanno et al., 2005), and plotted individual assignments with Distruct v1.1 (Rosenberg, 2004). The STRUCTURE approach has become a standard method of evaluating the number of genetic clusters in a data set while assuming equilibrium genetic conditions (Hardy-Weinberg and linkage equilibrium). These conditions may nonetheless not be fulfilled in all populations. Therefore, we also evaluated the data with principal component analyses (**PCA**) methods that are without such equilibrium assumptions using the *adegenet* package (Jombart, 2008) in R 2.14.2 (R Development Core Team, 2012).

The identity-by-state (alleles that are the same, irrespective of whether they are inherited from a recent ancestor) between pairs of dogs within the same breed was calculated for every possible pair of dogs and the mean, median, minimum, and maximum identity-by-descent (**IBD**; alleles that are descended from a common ancestor in a base population) were estimated. The estimates of pairwise IBD were used to find pairs of individuals who look too similar to each other (more than what would have been expected by chance in a random sample).

RESULTS

Genetic Variation

The H_E values differed significantly among breeds (1-way ANOVA) and all pairwise comparisons of H_E were also highly significant ($P < 0.0001$). The dog breeds were ranked relative to genetic variation expressed by both H_E and P%: DS > DSF > BR > ODP > GD (Table 1).

Table 1. Degree of polymorphism. The mean and median expected heterozygosity (H_E) estimated for the 5 dog breeds investigated

	DS[1]	DSF[2]	BR[3]	ODP[4]	GD[5]
P% SNP[6], %	43.60	36.50	35.90	31.90	26.90
Mean H_E[7]	0.36 ± 0.13	0.35 ± 0.08	0.33 ± 0.10	0.3 ± 0.09	0.27 ± 0.06
Median H_E[8]	0.4	0.4	0.37	0.33	0.26

[1]DS = Danish Spitz.
[2]DSF = Danish-Swedish Farm Dog.
[3]BR = Broholmer.
[4]ODP = Old Danish Pointing Dog.
[5]GD = Greenland Dog.
[6]Proportion of polymorphic SNP.
[7]Mean (±SD) of the H_E estimated for the dog breeds.
[8]Median of the H_E estimated for the dog breeds.

Population Genetic Profile

The STRUCTURE results supported the presence of 5 genetic clusters ($K = 5$; Fig. 1). With K = 3 DS, DSF and GD clustered together indicating a higher genetic similarity between these 3 breeds compared to the other 2 (BR and ODP). Among the Broholmer dogs the genetic profile of 1 dog deviated strongly. This was in agreement with the PCA results (see below). From the studbook of this dog it could be seen that both the father and the dog itself were accepted for breeding based on phenotypic resemblance. For $K = 6$ the DSF breed was split into 2 different clusters (Fig. 1).

The variation explained by the first, second, and third eigenvalue obtained from PCA were 18.6, 16.1, and 8.2%, respectively (Fig. 2), considering the fact that in total the PCA revealed highly differentiated dog breeds (Fig. 3a is the first and second axes, Fig. 3b is the second and third axes, and Fig. 3c is first and third axes). The Broholmer showed the presence of an outlier, which was consistent with the results from STRUCTURE.

The IBD between pairs of dogs within breeds, which was calculated for every possible pair of dogs, revealed large differences between mean and median IBD, with values ranging from zero (completely unrelated individuals) to 0.60 (highly related individuals; Table 2).

DISCUSSION

Genetic Variation

The lowest level of H_E and average IBD was observed for GD suggesting an ancient bottleneck where genetic variation was lost. This finding could be explained by the fact that this breed is considered one of the oldest in the world. Moreover, the breed has been kept isolated from other breeds for more than 1,000 yr because of a ban on importing other dog breeds to Greenland (the area north of 66° N latitude on the west coast and the entire east coast down to Cape Farewell east of 44° E longitude).

The low H_E observed in the ODP breed can be attributed to a strong founder effect, as the extant breed derives from only 20 individuals in the 1940s followed by genetic drift, which presumably has depleted the amount of genetic variation. The origin of the ODP can be traced back to about 1710 when gypsy and farm dogs were crossed for 8 generations with selection to fix the piedbald brown and white pattern. These dogs are the founders of the present day population of Old Danish Pointing dogs. We also attribute the relative low H_E observed for the BR breed to the same causes (founder effect and genetic drift). The BR breed, which was established from a cross between English Mastiff and local dogs in Germany in the 18th century, was believed to be extinct for more than 50 yr, partly as a consequence of strife during the Second World War. Nevertheless, the breed was successfully reconstructed based on a few individuals with a typical Broholmer phenotype and by using dogs of the Spanish and English Mastiff breeds in the 1970s. Founder effects and genetic drift have clearly limited the gene pool of the extant breed. Crossbred dogs of suitable phenotype were advertised for but only a handful of founders were identified and used to reconstruct the breed. The Broholmer was approved as an official Danish dog breed by the FCI in 1998.

The fact that DSF and DS showed a relatively high H_E compared to the other dog breeds can probably be explained by the methods used for the reconstruction of these breeds. Although the DS went through a strong population bottleneck recently, it was subsequently crossed with the Samoyed, which has boosted its genetic variability. The relatively high H_E observed in the DSF, which is an old native breed that historically lived on farms in the eastern part of Denmark and the southernmost part of Sweden, serving as a farm dog and hunting dog, can be attributed to a considerably higher number of founders of this breed. This breed was in fact also nearly extinct but reconstructed by crossing DSF dogs with other dogs that showed phenotypic resemblance with the DSF. In fact, the DSF breeding club still keeps an open studbook allowing

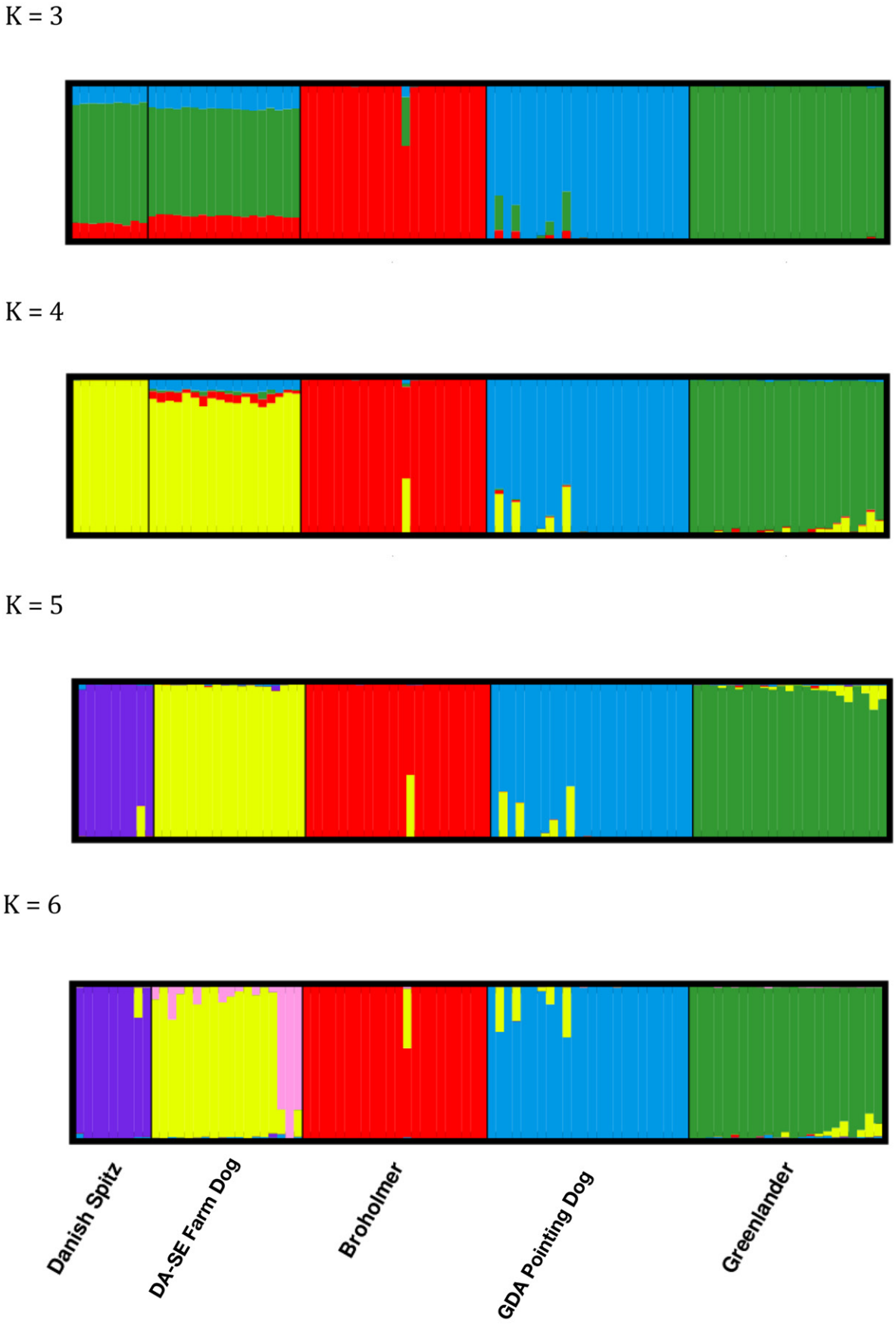

Figure 1. Estimated population structure ($K = 3$, 4, 5, and 6) derived using the program STRUCTURE 2.3.3 (Pritchard et al., 2000) for 5 dog breeds: Danish Spitz, Danish-Swedish (DA-SE) Farm Dog, Broholmer, Old Danish (GDA) Pointing Dog, and Greenland Dog (Greenlander). Each individual is represented by a thin vertical line, which is partitioned into K colored segments that represent the individual's estimated membership proportion in each of the clusters. See online version for figure in color.

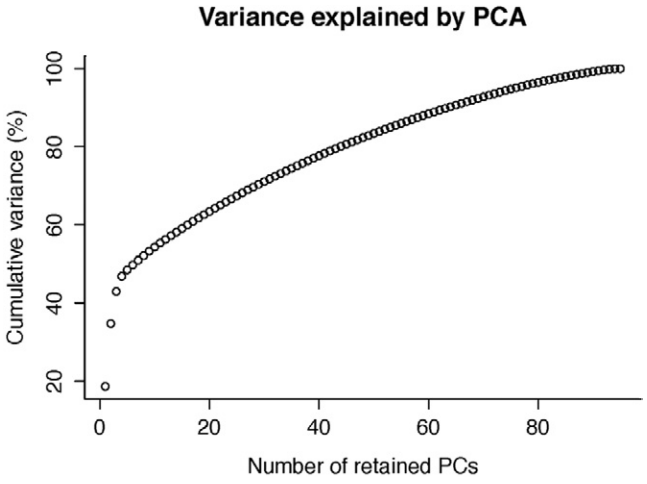

Figure 2. Principal component analysis (PCA) showing the amount of variation explained by each component. The first 3 eigenvalues explained 42.9% of the total variation: 18.6% for the first, 16.1% for the second, and 8.2% for the third. PC = principal component.

dogs with DSF resemblance to enter the breeding program after an evaluation by an authorized judge. Such crosses have clearly augmented the gene pool of the breed. The polymorphism within each breed seems to be correlated with the H_E within breeds as expected in a genetically depauperate population (low N_E).

Population Genetic Profile

The PCA plots reflect a clear separation between the breeds. The PCA and STRUCTURE results were consistent and both revealed the presence of a BR dog outlier, indicating that the genetic profile of this dog differs markedly from the other genotyped members of the breed. From all the 3 plots (Fig. 3a, 3b, and 3c) we see that ODP are more loosely clustered than the other breeds. The fact that STRUCTURE indicates $K = 5$ as the most likely number of clusters shows that the 5 dog breeds make up distinct units with uniform genetic profiles, with exception of certain outliers. The PCA plots, however, may not reflect the real genetic distance between the different breeds. For example, we would have expected a much higher proximity between the GD and the DS as both breeds belong to the Spitz breeds, a group of dog characterized by their prick ears, curly tails, and thick coats. However, these results could partly be a result of genetic drift. The influence of genetic drift is expected to be high, especially in the breeds with small effective population size. For $K = 3$, STRUCTURE nonetheless suggested genetic similarity between the DS and GD consistent with expectations based on their shared ancestry.

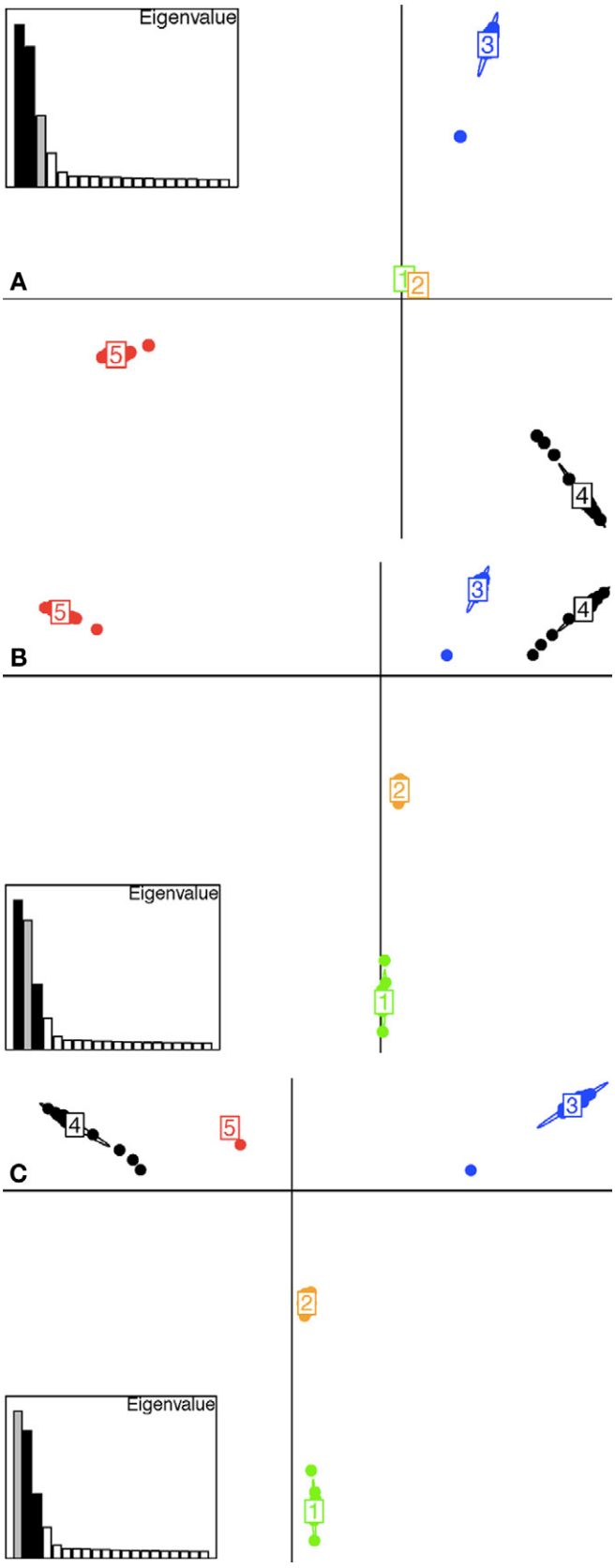

Figure 3. Principal component analysis of the 5 dog breeds: Danish Spitz, Danish-Swedish Farm Dog, Broholmer, Old Danish Pointing Dog, and Greenland Dog. Genetic differentiation is represented by distance and color: 3a shows first and second axes, 3b shows second and third axes and 3c shows first and third axes. The variance explained by the first, the second and the third axes are 18.6%, 16.1% and 8.2% respectively. See online version for figure in color.

Table 2. The identity-by-descent (IBD) estimates between every possible pair of dogs within the same breed

	DS[1]	DSF[2]	BR[3]	ODP[4]	GD[5]
NC[6]	35	153	231	276	253
Minimum[7]	0.08	0	0	0	0
Maximum[8]	0.52	0.56	0.55	0.6	0.53
Mean[9]	0.24	0.03	0.12	0.1	0.02
SE[10]	0.02	0	0	0	0
Median[11]	0.21	0	0.1	0.1	0

[1]DS = Danish Spitz.
[2]DSF = Danish-Swedish Farm Dog.
[3]BR = Broholmer.
[4]ODP = Old Danish Pointing Dog.
[5]GD = Greenland Dog.
[6]NC = number of comparisons
[7]Minimum identity-by-descent observed.
[8]Maximum identity-by-descent observed.
[9]Mean identity-by-descent.
[10]Standard error of the mean identity-by-descent.
[11]Median of the identity-by-descent.

Perspectives

The conservation of genetic resources for domestic breeds is becoming an important issue in conservation genetics, which needs urgent actions. A major challenge for many dog breeds is to reduce the rate of inbreeding and the frequency of deleterious dominant and recessive alleles, thereby reducing the incidence of hereditary diseases within the breeds. Many dog breeds have health problems, often caused by high rates of inbreeding, genetic drift, and breeding for characters that are problematic from an animal welfare point of view (Collins et al., 2011). Common problems in some dog breeds are undesirable temperament, impairment of eyesight and weakened immune system, high frequency of dysplasia, etc. Pedigree information can be used to monitor and control inbreeding in a population, but molecular data can be used more efficiently to do so. First, data from studies such as this can provide guidelines useful for breeding decisions and for evaluating if it would be relevant to open the studbook to allow dogs resembling the breed standards from other breeds to be included. Our results show the value of an open studbook when crossbreeding is the preferred strategy. Second, the heterogeneity found when estimating the IBD between pairs of dogs ranged from 0 to 0.60 within the same breed. This suggests that an appropriate breeding strategy based on IBD could be developed for all the 5 breeds investigated. Information from the SNP chip could be used as an accurate tool for guiding which individuals should mate, to optimize the optimal contribution of animals to the next generation. This can be in the form of a specific list with suggested matings or guidelines on the number of matings that given dogs should be engaged in during a given number of generations. Developing a SNP chip with a subset of SNP that are polymorphic across the 5 breeds investigated here could be useful for this purpose and make these goals more realistic from an economic point of view.

LITERATURE CITED

Clutton-Brock, J. 1999. A natural history of domesticated animals. Cambridge Univ. Press, Cambridge, UK.

Collins, L. M., L. Asher, J. Summers, and P. McGreevy. 2011. Getting priorities straight: Risk assessment and decision-making in the improvement of inherited disorders in pedigree dogs. Vet. J. 189:147–154.

Earl, D. A., and B. M. VonHoldt. 2012. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. 4:359–361.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Mol. Ecol. 14:2611–2620.

Jakobsson, M., and N. A. Rosenberg. 2007. CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801–1806.

Jombart, T. 2008. Adegenet: A R package for the multivariate analysis of genetic markers. Biochem. Genet. 10:149–163.

Lequarré, A. S., L. Andersson, C. André, M. Fredholm, C. Hitte, T. Leeb, H. Lohi, K. Lindblad-Toh, and M. Georges. 2011. LUPA: A European initiative taking advantage of the canine genome architecture for unravelling complex disorders in both human and dogs. Vet. J. 189:155–159.

Leroy, G. 2011. Genetic diversity, inbreeding and breeding practices in dogs: Results from pedigree analyses. Vet. J. 189:177–182.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. de Bakker, M. J. Daly, and P. C. Sham. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81:559–575.

R Development Core Team. 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rooney, N. J. 2009. The welfare of pedigree dogs: Cause for concern. J. Vet. Behav. 4:180–186.

Rosenberg, N. A. 2004. DISTRUCT: A program for the graphical display of population structure. Mol. Ecol. Notes 4:137–138.