**Characterization of the genetic profile of five Danish dog breeds**

Pertoldi, Cino; Kristensen, T. N.; Loeschcke, Volker; Berg, P.; Praebel, A.; Stronen, A. V.; Proschowsky, H. F.; Fredholm, Merete

**Characterization of the genetic profile of five Danish dog breeds**

Cino Pertoldi, Torsten Nygaard Kristensen, Volker Loeschcke, Peer Berg, Anne Praebel, Astrid Vik Stronen, Helle Friis Proschowsky and Merete Fredholm

*J ANIM SCI* published online August 29, 2013

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://www.journalofanimalscience.org/content/early/2013/08/28/jas.2013-6617

American Society of Animal Science

www.asas.org

# Characterization of the genetic profile of five Danish dog breeds

Cino Pertoldi*†‡[1] , Torsten Nygaard Kristensen*§, Volker Loeschcke*, Peer Berg§, Anne Praebel§, Astrid Vik Stronen#, Helle Friis Proschowsky¶, and Merete Fredholm||

**Department of Bioscience, Aarhus University, Ny Munkegade 116, DK-8000 Aarhus C, Denmark

†Department 18/Section of Environmental Engineering, Aalborg University, Aalborg, 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

||Department of Veterinary Clinical and Animal Sciences, University of Copenhagen, Grønnegårdsvej 3, DK-1870 Frederiksberg C, Denmark

[1]Corresponding author: biocp@nf.au.dk

1

**Abstract**

This investigation presents results from a genetic characterization of five Danish dog breeds genotyped on the CanineHD BeadChip microarray with 170K SNPs. The breeds investigated were 1) Danish Spitz (DS, n = 8), 2) Danish-Swedish farm Dog (DSF n = 18), 3) Broholmer (B, 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 five 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 > B > 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 five dog breeds can be used to design breeding strategies for the preservation of the genetic pool of these dog breeds.

## Introduction

Many domestic breeds have effective population sizes ($N_E$) ranging from less than one hundred 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) (http://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: 1) The Danish Spitz (DS) and 2) The Danish-Swedish farm Dog (DSF). A total of 130 dogs were chosen as original founders but the studbook remained open. Furthermore, three dog breeds, which are recognized by the FCI are considered to 'belong to' Denmark: 3) The Broholmer breed (B), 4) The Old Danish Pointing Dog (ODP) and 5) 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 five breeds using the CanineHD

3

BeadChip, which allows genotyping of up to 170K single nucleotide polymorphism (SNP) markers (Lequarré et al., 2011). We quantified variation within and among the five 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

*DNA extraction and genotyping*

We collected EDTA-stabilized blood samples from the five 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, B, 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, California, USA). Samples included: 1) Danish Spitz (DS, n = 8), 2) Danish-Swedish farm dog (DSF, n = 18), 3) Broholmer (B, n = 22), 4) Old Danish Pointing Dog (ODP, n = 24), 5) Greenland dog (GD, n = 23).

We used GenomeStudio and accompanying guidelines from Illumina (2010) to identify individuals suitable for analyses of the genetic profile. SNP calling rates ranged from 99.03% to 99.80% and the average call rate was 99.69%. Further quality control steps resulted in removing uncalled loci; sorting normalized R value for AA,

4

AB and BB clusters and remove all loci with value < 0.12 - sort Orig Score and remove all loci with value < 0.4 - sort Cluster Sep and remove all loci with value < 0.25 - remove all loci with 100% no call - remove all loci with AB Freq > 0.8 - remove all loci with 10% GC < 0.4).

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 of 0.01 (maf 0.01), and a maximum per-SNP missing rate of 0.02 (geno 0.02).

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

The number of SNPs retained for calculations (after pruning; maf > 0.01, geno > 0.02, max individual missing rate (mind) > 0.4 --indep-pairwise 50 5 0.8), (gplink code: -----maf 0.01 --geno 0.02 --mind 0.4 --indep-pairwise 50 5 0.8 --gplink) per breed were: DS: 69,775, DSG: 107,717, B: 74,191, GDH: 75,388, GD: 88,404. The different number of SNPs 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 MCMC repetitions and evaluated four possible population clusters (K = 3 – 6). Each parameter setting was repeated three times. We used the

5

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. Thus, 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 (IBS) (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 was utilized to find pairs of individuals who look too similar to each other (more than what would have been expect by chance in a random sample).

**Results**

*Genetic variation*

6

The $H_E$ values differed significantly among breeds (one-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 > B > ODP > GD (Table 1).

*Population genetic profile*

The STRUCTURE results supported the presence of five genetic clusters (K = 5) (Supplementary Material and Figure1). K=3 clustered DS, DSF and GD together indicating a higher genetic similarity between these three breeds compared to the other two (B and ODP). Among the Broholmer dogs the genetic profile of one 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 two different clusters (Figure 1).

The variation explained by the first, second and third eigenvalue obtained from PCA were 18.6 %, 16.1 % and 8.2 %, respectively (Figure 2), considering the fact that in total the PCA revealed highly differentiated dog breeds (Figure 3a = 1st, 2nd), (Figure 3b = 2nd and 3rd), (Figure 3c = 1st and 3rd). 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).

7

**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 (Brown et al., 2013). Moreover, the breed has been kept isolated from other breeds for more than 1000 years because of a ban to import 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 1940's (http://en.wikipedia.org/wiki/Old_Danish_Pointer), 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 eight 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 B breed to the same causes (founder effect and genetic drift). The B 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 years, 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 1970's. Founder effects and genetic drift have clearly

8

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 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 B dog outlier, indicating that the genetic profile of this dog differs markedly from the other

9

genotyped members of the breed. From all the three plots (Figure 3a, b, c) 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 five dog breeds comprise 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.

*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

10

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 are the preferred strategy. Secondly, 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 five breeds investigated. Information from the SNP chip could be used as an accurate tool for guiding which individuals should mate, in order 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 SNPs that are polymorphic across the five breeds investigated here could be useful for this purpose and make these goals more realistic from an economic point of view.

**Acknowledgements**

## References

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

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. The Veterinary Journal 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. Res. 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. Bioch. Genet. 10: 149-163.

12

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. The Veterinary Journal 189: 155-159.

Leroy, G. 2011. Genetic diversity, inbreeding and breeding practices in dogs: Results from pedigree analyses. The Veterinary Journal 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, *et al.* 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am. J. Hum. Genet., 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.

14

Table 1. Number of polymorphic loci expressed in percentage (P%). The mean and median expected heterozygosity ($H_E$) estimated for the five dog breeds investigated.

| | DS[4] | DSF[5] | B[6] | ODP[7] | GD[8] |
|---|---|---|---|---|---|
| [1]P% SNPs | 43.60% | 36.50% | 35.90% | 31.90% | 26.90% |
| [2]Mean $H_E$ | $0.36 \pm 0.13$ | $0.35 \pm 0.08$ | $0.33 \pm 0.10$ | $0.3 \pm 0.09$ | $0.27 \pm 0.06$ |
| [3]Median $H_E$ | 0.4 | 0.4 | 0.37 | 0.33 | 0.26 |

[1] Proportion of polymorphic SNPs expressed in percentage (%).
[2] Mean ($\pm$ SD) of the expected heterozygosity ($H_E$) estimated for the dog breeds.
[3] Median of the expected heterozygosity ($H_E$) estimated for the dog breeds.
[4] Danish Spitz.
[5] Danish-Swedish Farm Dog.
[6] Broholmer.
[7] Old Danish Pointing Dog.
[8] Greenland Dog.

15

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

| | DS[7] | DSF[8] | B[9] | ODP[10] | GD[11] |
|---|---|---|---|---|---|
| [1]NC | 35 | 153 | 231 | 276 | 253 |
| [2]Min | 0.08 | 0 | 0 | 0 | 0 |
| [3]Max | 0.52 | 0.56 | 0.55 | 0.6 | 0.53 |
| [4]Mean | 0.24 | 0.03 | 0.12 | 0.1 | 0.02 |
| [5]Std. error | 0.02 | 0 | 0 | 0 | 0 |
| [6]Median | 0.21 | 0 | 0.1 | 0.1 | 0 |

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

**Figure legends**

Figure 1. Estimated population structure (K = 3, 4, 5 and 6) derived using the program STRUCTURE 2.3.3 (Pritchard et al. 2000) for five dog breeds. 1) Danish Spitz, 2) Danish-Swedish farm dog, 3) Broholmer, 4) Old Danish Pointing Dog, 5) Greenland dog. 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.

Figure 2. Principal Component Analysis showing the amount of variation explained by each component. The first 3 eigenvalues explained 42.9% of the total variation; 1st - 18.6 %, 2nd - 16.1 % and 3rd - 8.2 %.

Figure 3. Principal component analysis of the five dog breeds:. 1) Danish Spitz, 2) Danish-Swedish farm dog, 3) Broholmer, 4) Old Danish Pointing Dog, 5) Greenland dog. Genetic differentiation is represented by distance and colour (3a = 1st, 2nd axes), (3b = 2nd and 3rd axes), (3c = 2nd and 3rd axes). 1st - 18.6 %, 2nd - 16.1 % and 3rd - 8.2 %.
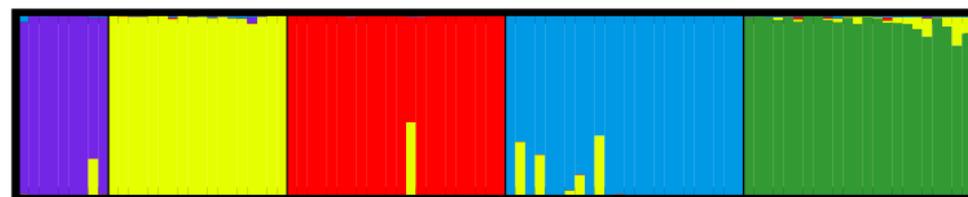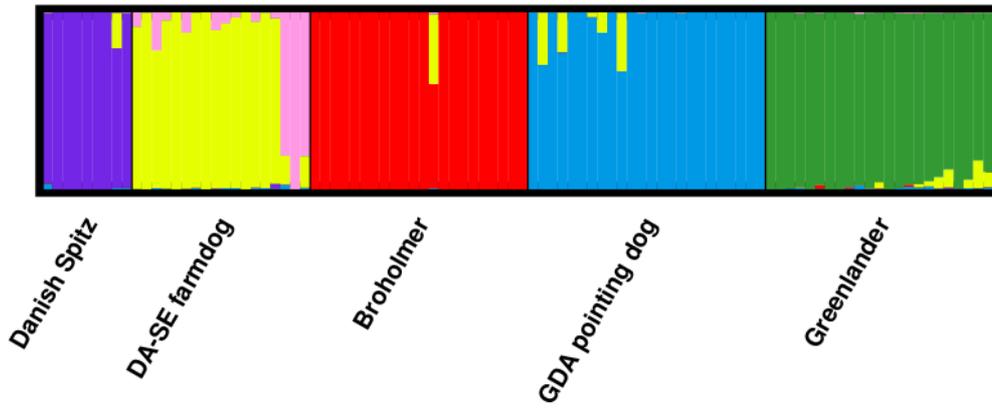
18

Figure 1

K = 3
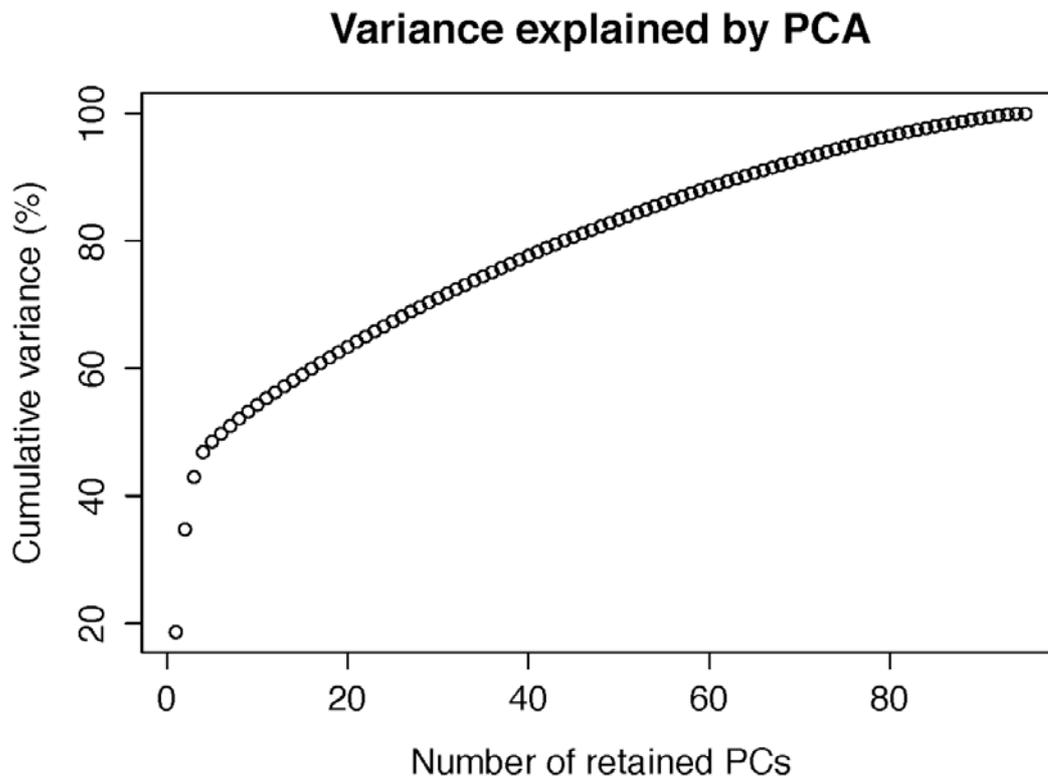


K = 4



K = 5
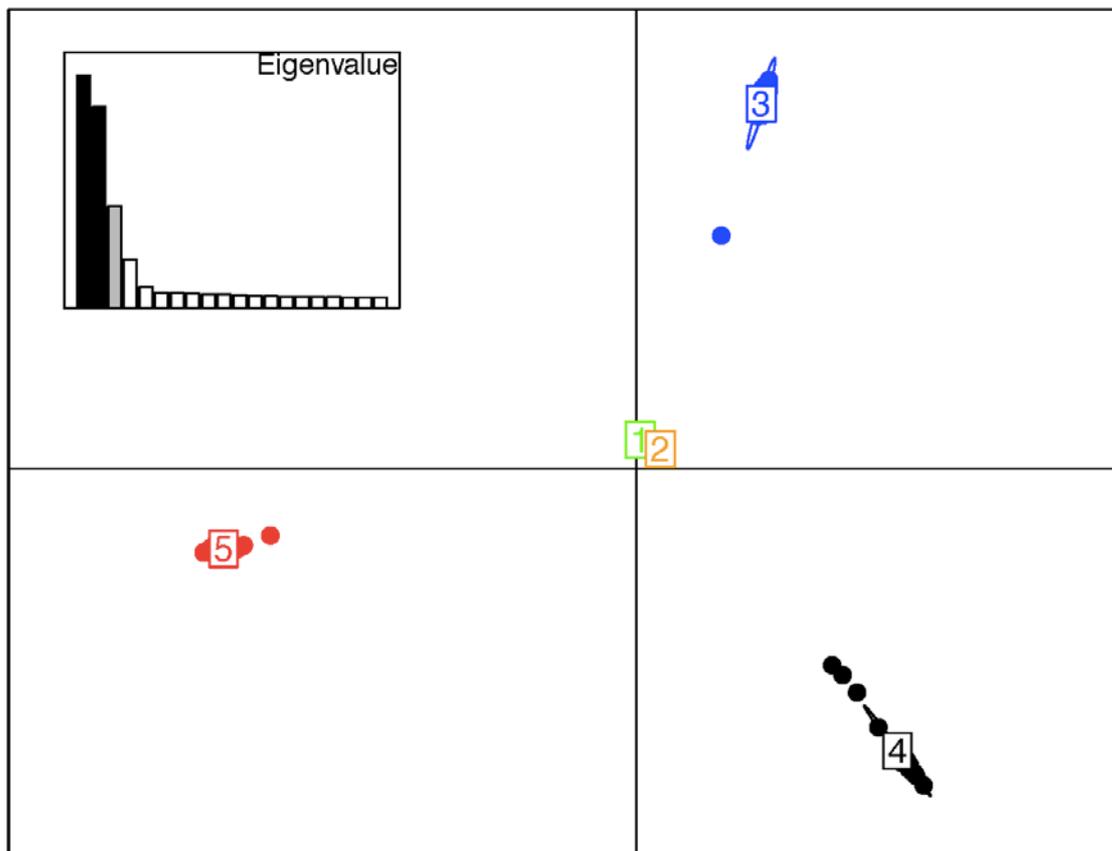
K = 6

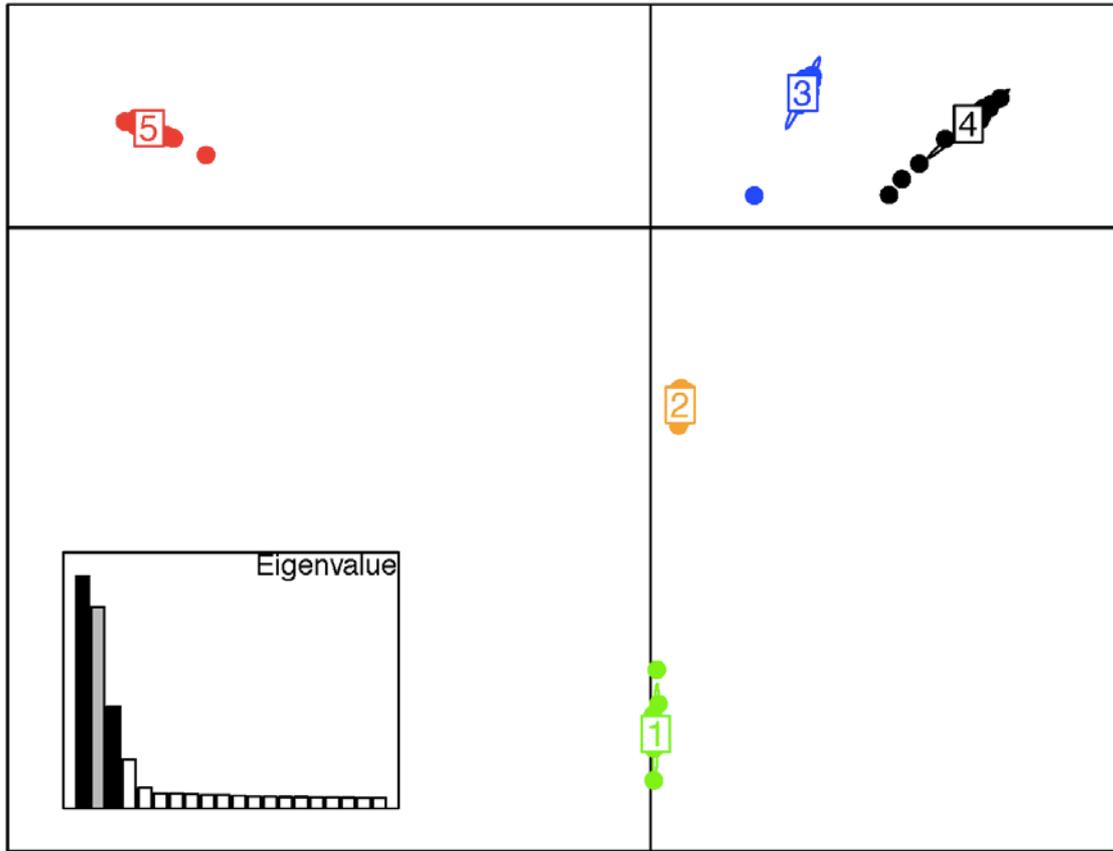Figure 2

# Variance explained by PCA
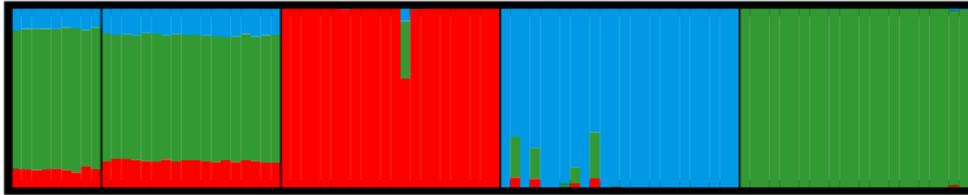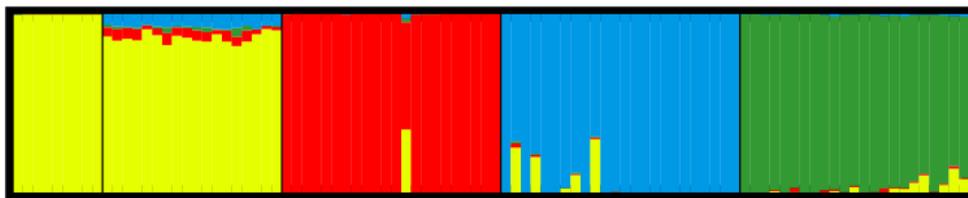
Figure 3a

Figure 3b
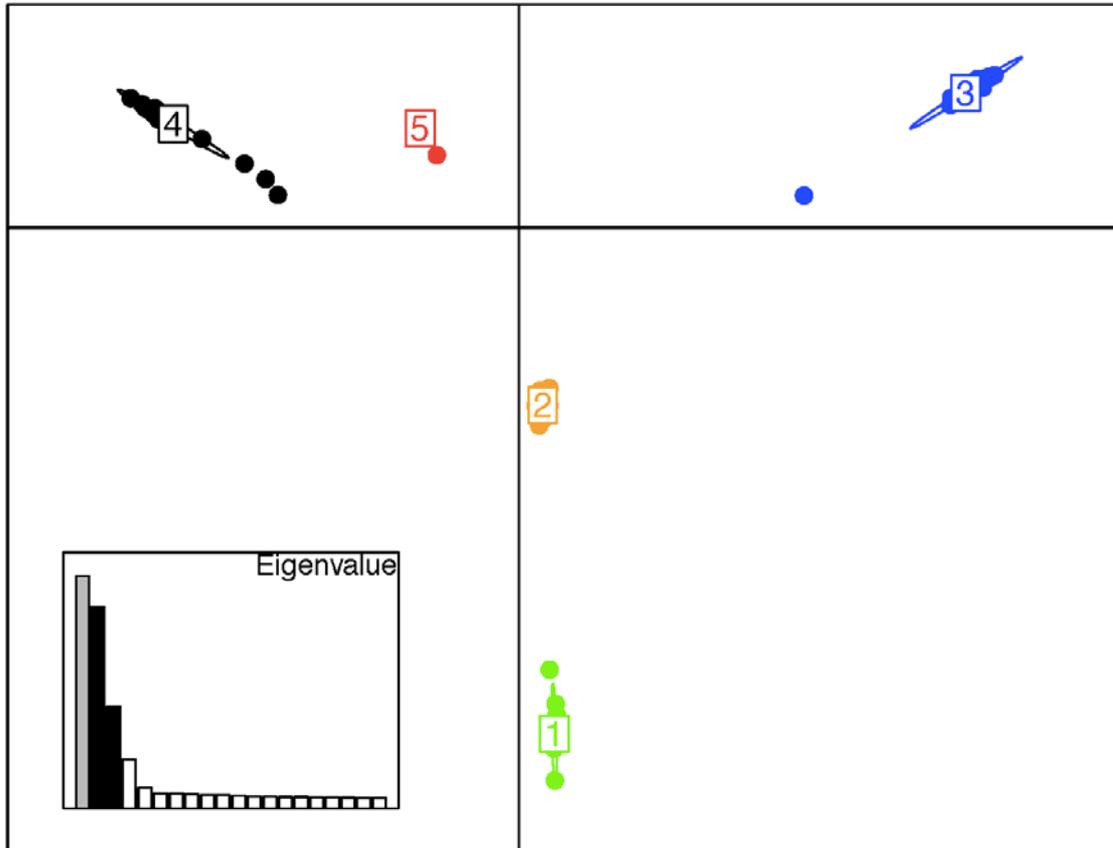
Figure 3c

Calculations for the number of genetic clusters indicated by STRUCTURE. The most probable number of subpopulations in bold.

| K[1] | L(K)[2] | SD[3] | ΔK[4] |
|------|---------|-------|-------|
| 3 | -1553124.3 | 346.3 | - |
| 4 | -1482858.4 | 7173.7 | 3.8 |
| **5** | **-1439816.4** | **909** | **38.9** |
| 6 | -1432109.1 | 239.2 | - |

[1] Number of clusters.
[2] Mean value of the posterior probability of K calculated for K values 3-6 (L(K)) –
[3] Standard deviation.
[4] Second order rate of change of the log probability of data between successive K values (ΔK) –.