ANIMAL GENETICS Immunogenetics, Molecular Genetics and Functional Genomics

doi:10.1111/j.1365-2052.2009.01849.x

Molecular structure in peripheral dog breeds: Portuguese native breeds as a case study

A. E. Pires*,†, I. R. Amorim‡,§, C. Ginja¶,,††, M. Gomes‡‡, I. Godinho*, F. Simões†, M. Oom*, F. Petrucci-Fonseca*, J. Matos† and M. W. Bruford§§**

*Departamento de Biologia Animal/Centro de Biologia Ambiental, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal. †Instituto Nacional de Engenharia e Tecnologia e Inovação, Grupo de Biologia Molecular, 1649-038 Lisboa, Portugal. ‡Departamento de Ciências Agrárias – CITAA, Universidade dos Açores, Terra-Chã, 9700-851 Angra do Heroísmo, Terceira, Açores, Portugal. §Grupo Lobo, Faculdade de Ciências de Lisboa, Bloco C2, 5° piso, Departamento de Biologia Animal, 1749-016 Lisboa, Portugal. ¶Estação Zootécnica Nacional, Departamento de Genética e Melhoramento Animal, Fonte Boa, 2005-048 Vale de Santarém, Portugal. **Instituto Superior de Agronomia, Tapada da Ajuda, 1349-017 Lisboa, Portugal. ††Veterinary Genetics Laboratory, University of California, 980 Old Davis Road, Davis, CA 95616, USA. ‡‡Parque Natural Montesinho, 5391-901 Bragança, Portugal. §§Cardiff University, School of Biosciences, Biomedical Building, Museum Avenue, Cardiff CF10 3US, UK

Summary

Genetic variability in purebred dogs is known to be highly structured, with differences among breeds accounting for ~30% of the genetic variation. However, analysis of the genetic structure in non-cosmopolitan breeds and local populations is still limited. Nine Portuguese native dog breeds, and other peripheral dog populations (five) with regional affinities, were characterized using 16 microsatellites and 225 amplified fragment length polymorphism (AFLP) markers, and the pattern of genetic differentiation was investigated. Although the level of breed differentiation detected is below that of other dog breeds, there is in most cases a correlation between breed affiliation and molecular structure. AFLP markers and Bayesian clustering methods allowed an average of 73.1% of individuals to be correctly assigned to source populations, providing robust genotypic assessment of breed affiliation. A geographical genetic structure was also detected, which suggests a limited influence of African dogs on the Iberian breeds. The sampling effect on the estimation of population structure was evaluated and there was a 2.2% decrease in genetic differentiation among breeds when working animals were included. Genetic diversity of stray dogs was also assessed and there is no evidence that they pose a threat to the preservation of the gene pool of native dog breeds.

Keywords amplified fragment length polymorphism, microsatellites, native dog breeds, population genetic structure, stray dogs.

Introduction

The presence of the domestic dog in South Western Europe is known to be very ancient. The oldest bones from Portugal date to 6010–5850 cal BC (2δ) (Cardoso 2002) and references to dog breeds date back to the 16th century (Frutuoso 1977). Native breeds were developed to perform tasks predominantly associated with a rural context, such as livestock guarding and herding, hunting and fishing. Currently, there are 10 established dog breeds registered in the Portuguese kennel club, of which eight are internationally recognized (http://www.fci.be) and three represent important reservoirs for domestic dog mitochondrial diversity (Pires 2006). Based on the current number of potential breeding females and following the Food and Agriculture Organization (FAO) classification, some of the Portuguese native dog breeds are Endangered (<1000 breeding females). Although FAO categories suggest a risk of extinction because of demographic stochastic changes, they may not be accurate indicators of the genetic diversity of the breeds. These breeds are currently not managed as closed breeding populations, because new individuals can still be recruited, mainly from rural sources (working dogs). Pedigreed and working dogs are distinct groups within the native breeds, which are managed

Address for correspondence

A. E. Pires, Instituto Nacional de Engenharia e Tecnologia e Inovação, Grupo de Biologia Molecular, 1649-038 Lisboa, Portugal.
E-mail: elisabete.pires@ineti.pt

Accepted for publication 27 November 2008

differently: the former being raised for their appearance, whereas the latter are bred to perform a specific task.

The study of peripheral breeds, including their molecular composition and genetic structure, is of interest. Additional genotypes and novel evolutionary groups may be revealed (Tapio *et al.* 2005; Beja-Pereira *et al.* 2006), and it may also prove critical for the conservation of animal genetic resources at a local scale (Björnerfeldt 2007), because of its potential impact on economic resource allocation. Microsatellites and amplified fragment length polymorphisms (AFLPs) are among the most informative markers used in population genetic studies. There are several microsatellite markers described for dogs (Sargan *et al.* 2007), and these have been highly useful for addressing dog breed genetic differences. However, because of issues, such as size homoplasy (Estoup *et al.* 2002; O'Reilly *et al.* 2004) and due to sampling bias (Hedrick 1999), a reliable signal of population differentiation may sometimes be difficult to assess with highly variable microsatellite loci.

Although the use of AFLPs on dog population genetic studies has been limited (Kim *et al.* 2001a; Pires 2006), they may in fact offer a higher statistically discriminatory power for population analysis, particularly in cases of weak differentiation (Campbell *et al.* 2003).

Here, we use for the first time a combination of AFLPs and microsatellites to compare and contrast patterns of genetic variation in Portuguese native dog breeds and other marginal breeds with regional affinities. We investigate correlation between breed affiliation and molecular structure, examine phylogeographic structure and historical events, compare breed and stray dogs in the same area, and assess whether breeds represent a distinct geographical distribution of alleles.

Materials and methods

Sampling and DNA extraction

Samples from 12 native dog breeds (from Portugal, Spain and North Africa) and from two stray dog populations were collected at several locations. Breed information is shown in Table S1. Animals were selected based on breed standards (registry books), and whenever information was available, related individuals back to three generations were excluded. Working animals for all native breeds, although frequently of unknown ancestry, but fulfilling morphological and behavioural profiles, were also sampled. Stray dogs in Portugal whose phenotypes could not be assigned to any breed were also sampled at several shelters in the Azores archipelago, Estrela Mountain and Alentejo regions. Sampling from North Africa dogs included the two dog breeds registered in Morocco – Aidi and Sloughi, and stray dogs from Tunisia, where there are no formally established breeds.

Blood samples (1–2 ml) were collected into vacutainers with EDTA (10% w/v) and kept frozen until processed. Pulled hairs were kept dry at room temperature. Tissue samples (ear punches, 20–50 mg) were preserved in a dimethyl sulphoxide salt solution buffer, at −20 °C.

Genomic DNA was extracted from whole blood and tissue using a standard proteinase K/Phenol–Chloroform protocol (Sambrook *et al.* 1989), the Nucleospin Blood QuickPure kit (Macherey-Nagel) or a high salt method (Montgomery & Sise 1990). DNA was extracted from hair roots in a 20% Chelex solution (Walsh *et al.* 1991).

Microsatellite genotyping

A total of 16 microsatellites, which include nine dinucleotides (*AHT121*, *AHTh171*, *AHTk253*, *C22.279*, *INRA21*, *CXX.109*, *CXX.173*, *CXX.225* and *C09.250*) and seven tetranucleotides (*FH2001*, *FH2054*, *FH2247*, *FH2010*, *FH2159*, *FH2611* and *PEZ08*) (Ostrander *et al.* 1993; Francisco *et al.* 1996; Mariat *et al.* 1996; Mellersh *et al.* 1997; Neff *et al.* 1999), were analysed through multiplex PCR amplifications using fluorescent-labelled primers. Sequences are available on GenBank, except for *INRA21*, available at http://www.isag.org.uk/ISAG/all/2005ISAGPanelDOG.pdf, and *PEZ08* (Neff *et al.* 1999). The QIAGEN Multiplex PCR Kit was used. Reactions were carried out in 10 μl using 2 μl genomic DNA (∼25–50 ng) following the manufacturer's instructions. Gel electrophoresis was performed on ABI Genetic Analysers (310 or 3730) or on a 4200 Li-Cor sequencer and alleles scored with the recommended software. The size standards used were Genescan 350 ROX and STR Marker LI-COR (4000–44B) for the ABI and LI-COR instruments respectively. Negative and positive reaction controls were always included. In detail, 4% of samples, heterozygote genotypes only and spanning the allele size range for each locus, were used as positive controls across instruments. Some allele sizes obtained from the ABI 3730 instrument were adjusted to be consistent with those obtained from the Licor or the ABI 310 instruments. Therefore, adjustments for 12 loci varied between 1 and 3 bp as follows: ABI 3730/Licor – *AHT121* (−2 bp), *C22.279* (−2), *FH2001* (−2), *FH2247* (+1), *FH2611* (−2), *INRA21* (−2), *PEZ08* (−3); ABI 3730/ABI 310 – *CXX.109* (+2), *CXX.173* (−1), *CXX.225* (+1), *C09.250* (+2) and *FH2010* (+2).

Scoring of AFLP markers

AFLP marker profiles were generated following Ajmone-Marsan *et al.* (1997). Selective amplification was performed with five EcoRI/TaqI primer combinations according to polymorphism, consistency and number of bands. The five combinations used in the second selective amplification (selective nucleotides: AAC/ACT, AAC/CCA, AAG/CCA, ACT/CAC, ACT/CCA) generated markers from 100 to 800 bp in size. In each case, the EcoRI selective primer was labelled with FAM or TET dyes (Qiagen) and fragments

electrophoresed on an ABI 377 automated sequencer. Products were sized using the GS TAMRA 2500 and local Southern interpretation. GENOGRAPHER 1.4 (Montana State University, 1988) was used for visualization and scoring of bands ('thumbnail' option). Band presence, above a threshold of 100 fluorescent units, or absence, was recorded for each sample in a binary character matrix for statistical analysis.

Microsatellite Statistical Analysis GeneAlex 6 (Peakall & Smouse 2006) was used to estimate observed and unbiased within-population expected heterozygosity, mean number of alleles per locus, to assess private alleles, to test for deviations from Hardy–Weinberg equilibrium using Fisher's test, and to calculate pairwise populations F_{ST} values. Genetic differentiation among all populations was estimated using θ (Weir & Cockerham 1984) in GENETIX 4.03 (Belkhir *et al.* 1996–2004). θ is the Weir and Cockerham's measure of Wright's F_{ST}. Confidence interval for θ-values (obtained by bootstrapping loci 15 000 times) were calculated with GDA 1.0 (http://lewis.eeb.uconn.edu/lewishome/software.html).

In previous studies, dog breeds were sampled using mainly pedigreed dogs. This may introduce an underestimation of within-breed gene diversity and an overestimation of genetic differentiation among breeds. Therefore, to test the effect of sampling on the estimation of population structure, we analysed two datasets: one with pedigreed dogs only, and another set with both pedigreed and working animals.

ARLEQUIN 2.0 (Schneider *et al.* 2000) was used for analysis of molecular variance (Excoffier *et al.* 1992). Three groups were defined – Portugal, Spain and North African dogs. Significance was evaluated through Monte Carlo simulation with 10 000 replicates. ARLEQUIN was also used to calculate pairwise populations ϕ_{ST} values. Population structure was further investigated using STRUCTURE 2.2 (Pritchard *et al.* 2000). Ten independent runs [K = 1–16; 50 000 Markov Chain Monte Carlo (MCMC) iterations, burn-in = 50 000] were carried out to estimate the most likely number of partitions, independent of breed affiliation. To assure a more accurate estimation of K, we followed the Evanno *et al.* (2005) approach, which was carried out when all populations were sampled, and for subsets of data, in order to assess nested structure or sub-structuring until no evidence for further sub-structuring was found. The modal value of the distribution of ΔK was used as an indicator of the signal strength for the genetic structure detected by the software (Evanno *et al.* 2005). The program was run using the admixture model and considering correlated allele frequencies (Falush *et al.* 2003).

Assignment tests, i.e. allocation of individuals to populations, were performed with GENECLASS 2.0 (Piry *et al.* 2004) using the frequency-based algorithm of Paetkau *et al.* (1995) and a simulation approach (10,000 genotypes) as proposed by Paetkau *et al.* (2004). Animals were considered to be correctly assigned when their genotype probability was higher than the P-value threshold (0.05) in their source population and lower in all others (Manel *et al.* 2002).

AFLP statistical analysis

Outlier loci, which must be removed from the dataset prior to the estimation of demographic parameters and statistics (Allendorf & Luikart 2007), were identified using the software Dfdist (Beaumont & Balding 2004). A null distribution was generated based on 50 000 simulated loci and aberrant loci were identified based on two rounds of simulations (confidence level = 95%).

Analyses of genetic diversity and population differentiation were thus performed on a subset of neutral AFLPs. Gene diversity was estimated by the Bayesian method implemented in Hickory 1.0 (burn-in = 50 000; sample = 250 000; thinning factor = 50) (Holsinger *et al.* 2002). The same software was used to estimate θ^B (F_{ST} Bayesian analogue), using the f free model (posterior distributions: α = 0.98 and β = 0.99 for f; α = 211.28 and β = 369.63 for θ). Hierarchical structuring of genetic variation based on Euclidean distances (AMOVA) between AFLPs multilocus and populations pairwise F_{ST} distances was evaluated using ARLEQUIN 2.0 (Schneider *et al.* 2000) and Monte Carlo simulation with 10 000 replicates. Alternative regional groupings were also tested.

Population structure based on AFLPs was further investigated using STRUCTURE (Pritchard *et al.* 2000). Input files were prepared with AFLPDAT (Ehrich 2006) and AFLP scores were re-coded as in Evanno *et al.* (2005) (absent bands coded 0/0; present bands coded 1/−9). The remaining analysis was essentially performed as described for microsatellites.

For AFLP data, individual assignment tests were implemented in STRUCTURE (Pritchard & Wen 2004) using source population prior information (K = 13), burn-in = 20 000 and 50 000 MCMC iterations, and assuming admixture and correlated allele frequencies between populations (Falush *et al.* 2003). For each individual, we estimated the proportion of the genotype (q) in its source population and the probability of ancestry in other populations in the present, first or second previous generations. The percentage of individuals correctly assigned was calculated for $q > 0.95$ and 0.999.

The relationship between current population size and genetic diversity of each breed and stray dogs' populations was investigated by the Pearson's correlation.

Results

Molecular markers

No microsatellite loci significantly deviated from Hardy–Weinberg expectations and therefore all markers were included in the subsequent analyses. A total of 227 alleles

were found for the 16 loci across the studied populations; allele number per locus ranged from six (*CXX.109* and *CXX.225*) to 33 (*FH2159*), with an average of 14.19 ± 7.66 alleles/locus. Thirteen breed-specific alleles (with frequencies >5%) were detected at 12 loci (not shown). Except for the Castro Laboreiro Watchdog and the Portuguese Pointer, all breeds showed private alleles.

For the AFLPs, 25 of 225 markers (11.1%) exhibited F_{ST} outside the 95% confidence limits of neutral expectation after two checking rounds.

Genetic variation

Microsatellite allelic richness ranged from 4.7 in the Portuguese Pointer to 9.4 in Portuguese stray dogs, with an average of 6.8 ± 1.56 alleles/locus/population. The average expected heterozygosity corrected for sample size (H_E n.b) over all loci ranged from 0.63 in the Portuguese Pointer to 0.81 in Tunisian dogs, while observed heterozygosity (H_O) varied from 0.60 (Portuguese Pointer) to 0.89 (Tunisian dogs) (Table S2). The least diverse breeds were the Portuguese Pointer and the Portuguese Sheepdog, followed by the Castro Laboreiro Watchdog. For AFLPs, genetic diversities across populations (Hs) ranged from 0.09 (Azores Cattle and Portuguese stray dogs) to 0.15 (Estrela Mountain Dog and Sloughi) with an average of 0.13 ± 0.006 (Table S2). Estrela Mountain Dog and Sloughi were the most diverse breeds, followed by the Alentejo Shepherd Dog and Portuguese Warren Hound.

Genetic differentiation

The average microsatellite differentiation among all breeds was 0.057, which is significantly different from zero (0.0454–0.0733). For AFLPs, pairwise ϕ_{ST} values ranged from 0.07 (Portuguese Warren Hound and Azores Cattle Dog) to 0.60 (Portuguese Sheepdog and Portuguese Pointer). A concise table for both pairwise populations F_{ST} and ϕ_{ST} values is presented (Table S3).

The estimated degree of genetic differentiation (θ) among only the Portuguese native dog breeds differs according to whether the pedigreed or the pedigreed plus working dogs dataset was used, the first being 0.092 (0.073–0.113) whilst the latter was 0.070 (0.055–0.088). A 2.2% decrease in genetic differentiation (mean θ) among breeds was detected when working animals were included.

For microsatellites, subdivision among dog breeds was detected with AMOVA (ϕ_{ST} = 4%, $P < 0.001$) (Table S4). Approximately 92% of the variation can be explained by individual differences, and no geographic structure was detected (ϕ_{CT} = −0.006, NS). The AMOVA analyses for AFLPs also revealed that genetic variance among populations is significant, with a global ϕ_{ST} of 0.32 ($P < 0.001$) and approximately 36% of the genetic variation observed among individuals; differentiation among breeds within regions explained 23.20% of the total variance ($P < 0.001$) and 8.44% of the variance could be attributed to geographic structure ($P < 0.05$). For the AFLP dataset, the mean value of the Bayesian F_{ST}, analogue of θ^B, was 0.35.

For the structure analysis of the microsatellite dataset, the modal value for the distribution of ΔK (175.45) was at $K = 2$. The first partition segregated all the Portuguese livestock guarding dog breeds, Castro Laboreiro Watchdog, Estrela Mountain Dog, Alentejo Shepherd Dog and Transmontano Mastiff, from the other breeds (Fig. 1a). Further genetic structure within the first subgroup was detected, corresponding to the Castro Laboreiro Watchdog and Transmontano Mastiff breeds (19.74), whereas the Alentejo Shepherd Dog and Estrela Mountain Dog clustered together. The Transmontano Mastiff showed within-breed sub-structuring related to the time of sample collection (before vs. after official breed establishment). In the second group, the Portuguese Pointer, Water Dog, Sheepdog and Azores Cattle Dog clustered separately (13.51) from the Spanish Mastiff, Aidi, Sloughi, Portuguese Warren Hound, Portuguese stray dogs and Tunisian dogs, whereas the latter six populations were not differentiated from each other. Further sub-structuring (show vs. working dogs) was detected within the Azores Cattle Dog.

For the AFLP markers, a modal value for the distribution of ΔK was also found at $K = 2$ (165.70). However, with these markers, the partition segregated Castro Laboreiro Watchdog, Portuguese Sheepdog, Portuguese Water Dog, Aidi and Sloughi from all other breeds (Fig. 1b). Further analyses including only the breeds Castro Laboreiro Watchdog, Portuguese Sheepdog, Portuguese Water Dog, Aidi and Sloughi revealed two groups, with Castro Laboreiro Watchdog separated from all other breeds (59.11), and at $K = 5$ all breeds were differentiated (61.00). For the remaining dog populations, the modal ΔK value was again obtained at $K = 2$ (21.48), with Estrela Mountain Dog and Alentejo Shepherd Dog clustering together, independent of other populations. For the Spanish Mastiff, Portuguese Pointer, Azores Cattle Dog, Portuguese Warren Hound, Portuguese stray dogs and Tunisian dogs, further structure was detected, with the first three differentiated and the latter three populations remaining undifferentiated (11.67) (Fig. 1b).

Breed assignment

The overall percentage of individuals correctly assigned to their source population based on microsatellite loci was only 13% (P-value = 0.05), with no individuals being classified within the Spanish Mastiff, Aidi, Portuguese Pointer, Sloughi, Portuguese Sheepdog and Tunisia populations (Table S5). The Transmontano Mastiff shows the highest percentage of correctly assigned individuals (67%). In general, misclassified individuals were assigned to the most heterogeneous group: Portuguese Stray dogs. The

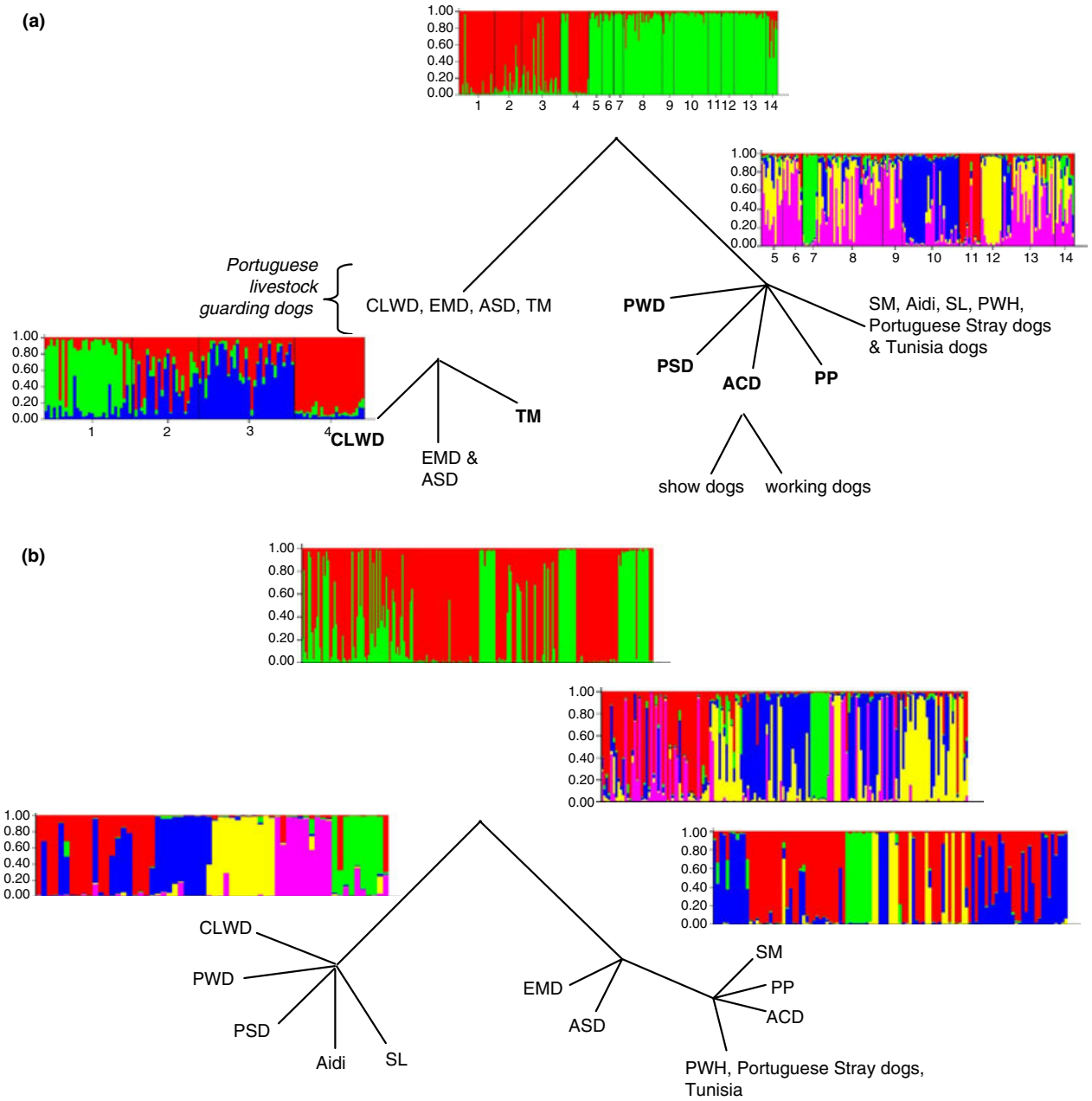

Figure 1 Population partitioning suggested by STRUCTURE based on (a) microsatellites and (b) AFLP markers. Junctions show where the data were split into the *K* populations and re-run on the sub-dataset until the *K*-value of all resultant clusters was 1. For breed acronyms, see Table S1.

percentage of individuals with maximum genotype probabilities in the source population was 53 overall and varied between 9% for the Spanish Mastiff and 75% for the Transmontano Mastiff.

In stark contrast to the microsatellite analysis, the AFLP results show that the average *Q*-value (proportion of membership of each pre-defined population in each of the 13 clusters) for these dog populations is high (0.988 ± 0.112), and varied between 0.933 (Sloughi) and 1 (Spanish Mastiff, Portuguese Sheepdog, Portuguese Pointer, Portuguese Waterdog and Aidi) (Table S6). Depending on the threshold q value (estimated proportion of each individual genotype in each population or cluster) defined, the average percentages of individuals correctly assigned are 93.9 and 73.1% for $q > 0.95$ and 0.999 respectively. For the Spanish Mastiff, Portuguese Sheepdog, Portuguese Pointer, Portuguese Waterdog, Aidi breeds and Tunisia dogs, all individuals sampled were classified within their source population with high q (>0.999). For the Estrela Mountain Dog, Alentejo Shepherd Dog, Azores Cattle Dog, Portuguese Warren Hound, Sloughi and Portuguese stray dogs, the percentages of individuals correctly assigned

with $q > 0.999$ ranged between 46.4 and 70%, and individuals with possible admixed ancestry were detected.

Gene diversity and effective population size

The microsatellite data show no correlation between gene diversity and the effective number of breeding females (N_eF) (Pearson correlation coefficient, $r = 0.11$; $P = 0.79$) (Fig. 2a). In contrast, AFLP diversity increased significantly with the increase in the effective number of breeding females in each population ($r = 0.69$, $P = 0.056$) (Fig. 2b).

Discussion

Genetic diversity

Not surprisingly, expected heterozygosity was much higher with microsatellites compared with AFLPs. The low frequencies typically associated with even the commonest microsatellite alleles lead to higher estimates of expected panmitic heterozygosity.

The highest expected microsatellite heterozygosities detected in this study (0.63–0.81) are higher than those reported in other studies (0.56–0.72, Koskinen & Bredbacka 2000; 0; 0.45–0.75, Altet *et al.* 2001; 0; 0.31–0.72; Kim *et al.* 2001b; 0; 0.39–0.71, Irion *et al.* 2003; 0; 0.56–0.72, Koskinen 2003; 0; 0.62–0.68, Parra *et al.* 2008). Heterozygosity values estimated for AFLP are, to our knowledge, the first to be reported for domestic dog breeds, and thus comparisons with other breeds cannot be made. Although values of genetic diversity based on microsatellites and AFLPs cannot be compared directly, breeds did not rank in the same order when comparing variation estimates using these markers. While microsatellites reveal the signature at very recent or ongoing demographic processes, AFLPs, because of their lower evolutionary rate and polymorphism, may differ in their sensitivity to population bottlenecks and demographic recovery, thus retaining the signal of past genetic structure more effectively.

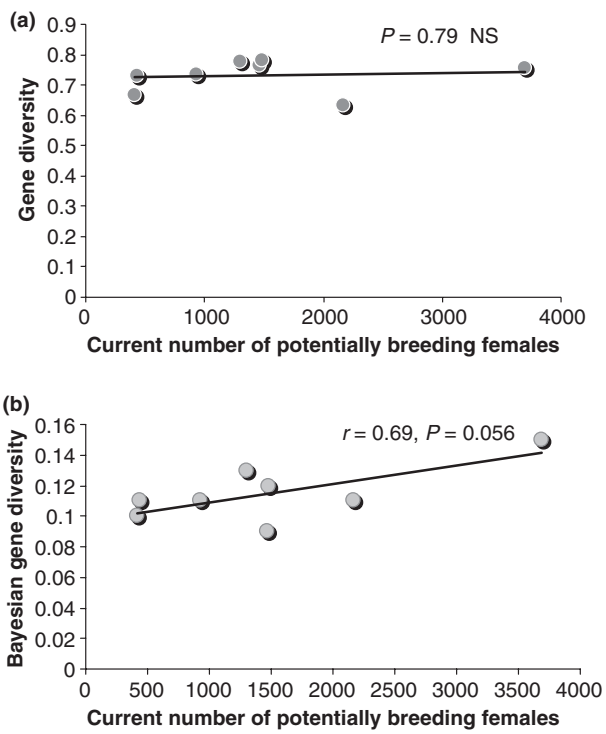

Figure 2 Pearson's correlation of gene diversity and current female effective population size. (a) Microsatellite and (b) AFLP data.

Fine-scale population genetic patterning

The Bayesian analysis of AFLP data produced, in general, results similar to those obtained using microsatellites, whilst still allowing higher resolution among breeds, such as Spanish Mastiff, Aidi and Sloughi. There is no evidence of genetic differentiation between Alentejo Shepherd Dog and Estrela Mountain Dog. Historical evidence supports their close genetic relationship, because the Estrela Mountain Dog is the ancestor of the Alentejo Shepherd Dog (Alpoim 1999) and these breeds maintained contact because of transhumance with possible interbreeding. Transhumance, the migration of livestock, shepherds and dogs twice a year observed in Mediterranean areas from plains to mountains, was important up to the 19th century and may have contributed to admixture among Iberian livestock guarding dog breeds.

The distinctiveness of Castro Laboreiro Watchdog as indicated by both Bayesian analyses is very well supported by mitochondrial DNA data (Pires *et al.* 2006). Several individuals of this breed showed exactly the same mtDNA haplotype, which is unique in the context of these native breeds (see van Asch *et al.* 2005). However, this breed did not show any private microsatellite alleles. The Transmontano Mastiff is also genetically very distinctive, although its morphotype resembles that of the Alentejo Shepherd Dog and thus it has been considered an ecotype of that breed. The Transmontano Mastiff registry was only established in 2004 based on 170 founders, 93 males and 77 females, and the sub-structuring revealed by the Bayesian method within this breed corresponds to samples collected before and after breed registration. The Transmontano Mastiff is nowadays genetically cohesive and if individuals sampled before official breed establishment were among breed founders, they are no longer represented in the current population. In turn, the within-breed genetic structure revealed by microsatellite for the Azores Cattle Dog is most likely because of the fact that breeding between show and working dogs has not been favoured.

Pairwise θ-values were significantly correlated between markers (not shown); however, θ-values obtained with microsatellites were lower than for AFLPs. Genetic differentiation values were also marker dependent: as high as

5.7% with microsatellites, and 35% with AFLPs. Both datasets concur that there is genetic differentiation among populations. The high evolutionary rate and hence polymorphism of microsatellites may contribute to a homogenizing effect and lower F_{ST} values (Balloux & Lugon-Moulin 2002). In contrast, AFLP markers can provide upwardly biased estimates of differentiation because of their dominant inheritance pattern (Gaudeul *et al.* 2004). However, the different number of loci used in this study prevents straightforward comparisons. Nonetheless, it seems clear that the degree of genetic differentiation among Iberian and North African peripheral dog breeds is below that observed in many other dog populations.

The transversal within-breed sampling strategy used in this study is radically different from the sampling designed by other authors, who report very high levels of genetic differentiation among dog breeds (Koskinen 2003; Parker *et al.* 2004). The degree of breed differentiation decreases by 2.2% when working dogs were included in the dataset. Most of the working dogs sampled were not registered in the Portuguese kennel Club, however, their morphological and behaviour characteristics were carefully evaluated and the animals selected correspond entirely to breed designations (breed standards). Therefore, we consider that working dogs represent additional genetic variability that should be taken into account when characterizing native dogs. Thus, the approximately 30% differentiation determined by Parker *et al.* (2004) could be regarded as the maximum value found among dog breeds. For Portuguese native breeds, the lower genetic differentiation can be explained by the fact that these breeds are not closed breeding populations. Occasional recruitment of unregistered (non-pedigreed) animals can lead to high levels of genetic diversity, higher breed heterogeneity and thus a lower differentiation. Portuguese native dog breed standards date mostly from the first half of the 20th century, and the short period of time since breed divergence could also account for a lower differentiation.

Our results with Portuguese native breeds show how important it is to perform wide sampling within a breed, because working animals that underwent historic selection for a specific task may carry a suite of different genotypes.

AFLPs allowed higher resolution of geographical genetic structure and detected significant genetic differentiation among the geographic regions of Portugal, Spain and North Africa (ϕ_{CT} = 8.44%, $P < 0.05$). Historically, the Iberian Peninsula was in close connection with North Africa mainly because of the Islamic (Arab and Berber) occupation (Ribeiro & Saraiva 2004), which explains the North African mtDNA influence in Iberian people that is not detected elsewhere in Europe (Pereira *et al.* 2000). During the Islamic occupation, animals were probably also introduced from Africa into Iberia. Cymbron *et al.* (1999) and Beja-Pereira *et al.* (2002) detected admixture among bovines based on mtDNA and casein haplotypes respectively and the gene pool of Iberian sheep was also improved during the Muslim period (Pereira *et al.* 2006; Davis 2008). However, this has not been detected for the Portuguese native domestic dogs based on analysis of their mtDNA (Pires *et al.* 2006), and this is confirmed here. Religious factors may account for such lack of 'African print' in Iberian dog breeds, because dogs are considered impure by Muslims (Coppinger & Coppinger 2002) and are not generally part of households (Gallant 2002).

Breed assignment

The low breed assignment rates achieved with microsatellites may result from a combination of factors such as reduced level of breed differentiation (θ = 5.7%), relatively small number of sampled individuals for some of the populations and number of loci in the analysis (<25). The influence of these factors in assignment tests has been shown for closely related populations in other species (Maudet 2001; Paetkau *et al.* 2004).

The very high percentages of breed affiliation obtained with AFLP profiles and Bayesian clustering methods was surprising, reinforcing the suitability of this marker for population differentiation and breed assignment studies (Campbell *et al.* 2003; Negrini *et al.* 2007). The high number of AFLP loci analysed allowed, in general, the correct allocation of individuals from relatively recent and somehow related dog breeds. Thus, DNA-based dog breed identification at the individual level, in cases of weak genetic differentiation, is possible using AFLP profiles. Nevertheless, because AFLP scoring needs a considerable amount of good quality DNA, it would be difficult to implement individual breed assignment in forensic cases.

Our combined data show that Castro Laboreiro Watchdog, Transmontano Mastiff, Portuguese Pointer, Portuguese Sheepdog, Portuguese Water Dog and Azores Cattle Dog are breeds with more specific genotypic compositions corresponding to a distinct geographical distribution of alleles.

Effective population size

The conservative Pearson correlation coefficient between the current female effective population size and within-breed genetic diversity was significant only for AFLP markers, where almost 70% of the genetic variation is explained by differences in breeding female population size among breeds. Other breed-specific factors, such as disease outbreaks, inbreeding and population isolation may explain the remaining variation (30%). This correlation is mainly affected by the presence of the Estrela Mountain dog. This is one of the breeds with a large number of potentially breeding females and shows the highest value for AFLP gene diversity. Our results suggest that the correlation between effective population size and genetic diversity might be more obvious when genome-wide molecular markers

(e.g. AFLPs) are used and when populations have a large breeding female effective population size. This problem requires further investigation, but in the case of Portuguese dog breeds, all possible breeds were sampled with one exception, the Terceira Cattle Dog (recently established).

This research is one of the first cases where stray dogs are compared with pedigree dogs, which allows for an improved description of the current dog diversity. Portuguese stray dogs showed low values of genetic distances with the established native breeds, which may be because of the fact that the former are a genetically heterogeneous group comprising different lineages, and are not under strong artificial selection. Portuguese native dog breeds, namely the Estrela Mountain Dog, Alentejo Shepherd Dog and the Azores Cattle Dog, are genetically differentiated from stray dogs that co-occur in the same area. Thus, there is no strong indication that stray dogs in Portugal are a concern for the preservation of the gene pools of native dog breeds. However, stray dogs should be strictly controlled, particularly in rural areas where feral dog predation on livestock is mistakenly attributed to wolves, leading to human–wolf conflicts and difficulties in the conservation of the Iberian Wolf (Petrucci-Fonseca *et al.* 2000; Sundqvist *et al.* 2008).

Acknowledgements

We thank C. Cruz, S. Ribeiro, dog owners, breed clubs and the Portuguese Kennel Club who contributed with samples and information; D. Mendonça and F. Sobral for genotyping technical assistance; M. Fernandes and C. João for critically reviewing early versions of the manuscript, and T. Bray for suggestions with STRUCTURE. Research supported by Fundação para a Ciência e a Tecnologia (PRAXIS XXI/BD/21677/99), Programa AGRO (project 31106) and Grupo Lobo, Portugal.

References

Ajmone-Marsan P., Valentini A., Cassandro M., Vecchiotti-Antaldi G., Bertoni G. & Kuiper M. (1997) AFLP markers for DNA fingerprinting in cattle. *Animal Genetics* **28,** 418–26.

Allendorf F.W. & Luikart G. (2007) *Conservation and the Genetics of Populations*. Blackwell Publishing, Oxford, UK.

Alpoim J.A. (1999) *O Rafeiro do Alentejo. Monografia de uma Raça [The Alentejo Shepherd dog. Monograph of a Breed]*. Câmara Municipal de Monforte, Monforte, Portugal.

Altet L., Francino O. & Sanchez A. (2001) Microsatellite polymorphism in closely related dogs. *Journal of Heredity* **92,** 276–9.

van Asch B., Pereira L., Pereira F., Santa-Rita P., Lima M. & Amorim A. (2005) MtDNA diversity among four Portuguese autochthonous dog breeds: a fine-scale characterisation. *BMC Genetics* **6,** 37.

Balloux F. & Lugon-Moulin N. (2002) The estimation of population differentiation with microsatellite markers. *Molecular Ecology* **11,** 155–65.

Beaumont M.A. & Balding D.J. (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology* **13,** 969–80.

Beja-Pereira A., Erhardt G., Matos C., Gama L. & Ferrand N. (2002) Evidence for a geographical cline of casein haplotypes in Portuguese cattle breeds. *Animal Genetics* **33,** 295–300.

Beja-Pereira A., Caramelli D., Lalueza-Fox C. *et al.* (2006) The origin of European cattle: evidence from modern and ancient DNA. *Proceedings of the National Academy of Sciences of the United States of America* **103,** 8113–8.

Belkhir K., Borsa P., Chikhi L., Raufaste N. & Bonhomme F. (1996–2004) GENETIX *logiciel sous Windows TM pour la génétique des populations*. Laboratoire Génome, Populations, Interactions, CNRS, Université de Montpellier II, Montpellier, France.

Björnerfeldt S. (2007) *Consequences of the Domestication of Man's Best Friend the Dog*. PhD Thesis, Uppsala University, Sweden.

Campbell D., Duchesne P. & Bernatchez L. (2003) AFLP utility for population assignment studies: analytical investigation and empirical comparison with microsatellites. *Molecular Ecology* **12,** 1979–91.

Cardoso J.L. (2002) *Pré-História de Portugal*. Editorial Verbo, Lisboa.

Coppinger R. & Coppinger L. (2002) *Dogs: A New Understanding of Canine Origin, Behavior, and Evolution*. University of Chicago Press, Chicago, IL.

Cymbron T., Loftus R.T., Malheiro M.I. & Bradley D.G. (1999) Mitochondrial sequence variation suggests an African influence in Portuguese cattle. *Proceedings Biological Sciences* **266,** 597–603.

Davis J.M. (2008) Zooarcheological evidence for Moslem and Christian improvements of sheep and cattle in Portugal. *Journal of Archaeological Science* **35,** 991–1010.

Ehrich D. (2006) AFLPDAT: a collection of R function for convenient handling of AFLP data. *Molecular Ecology Notes* **6,** 603–4.

Estoup A., Jarne P. & Cornuet J.-M. (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology* **11,** 1591–1604.

Evanno G., Regnaut S. & Goudet J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14,** 2611–20.

Excoffier L., Smouse P.E. & Quattro J.M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131,** 479–91.

Falush D., Stephens M. & Pritchard J.K. (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164,** 1567–87.

Francisco L.V., Langston A.A., Mellersh C.S., Neal C.L. & Ostrander E.A. (1996) A class of highly polymorphic tetranucleotide repeats for canine genetic mapping. *Mammalian Genome* **7,** 359–62.

Frutuoso G. (1977) *Saudades da Terra (written in 1580–1591)*. Instituto Cultural de Ponta Delgada, Ponta Delgada.

Gallant J. (2002) *The Story of the African Dog*. University of Natal Press, Natal, South Africa.

Gaudeul M., Till-Bottraud I., Barjon F. & Manel S. (2004) Genetic diversity and differentiation in *Eryngium alpinum* L. (Apiaceae): comparison of AFLP and microsatellite markers. *Heredity* **92,** 508–18.

Hedrick P.W. (1999) Perspective: Highly variable loci and their interpretation in evolution and conservation. *Evolution* **53,** 313–8.

Holsinger K.E., Lewis P.O. & Dey D.K. (2002) A Bayesian approach to inferring population structure from dominant markers. *Molecular Ecology* **11,** 1157–64.

Irion D.N., Schaffer A.L., Famula T.R., Eggleston M.L., Hughes S.S. & Pedersen N.C. (2003) Analysis of genetic variation in 28 dog breed populations with 100 microsatellite markers. *Journal of Heredity* **94,** 81–7.

Kim K.S., Jeong H.W., Park C.K. & Ha J.H. (2001a) Suitability of AFLP markers for the study of genetic relationships among Korean native dogs. *Genes & Genetic Systems* **76,** 243–50.

Kim K.S., Tanabe Y., Park C.K. & Ha J.H. (2001b) Genetic variability in East Asian dogs using microsatellite loci analysis. *Journal of Heredity* **92,** 398–403.

Koskinen M.T. (2003) Individual assignment using microsatellite DNA reveals unambiguous breed identification in the domestic dog. *Animal Genetics* **34,** 297–301.

Koskinen M.T. & Bredbacka P. (2000) Assessment of the population structure of five Finnish dog breeds with microsatellites. *Animal Genetics* **31,** 310–7.

Manel S., Berthier P. & Luikart G. (2002) Detecting wildlife poaching: identifying the origin of individuals with Bayesian Assignment tests and multilocus genotypes. *Conservation Biology* **16,** 650–9.

Mariat D., Kessler J.L., Vaiman D. & Panthier J.J. (1996) Polymorphism characterization of five canine microsatellites. *Animal Genetics* **27,** 434–5.

Maudet C. (2001) *Diversité et caractérisation génétique des races bovines et caprines originaires de la région Rhône-Alpes*. Docteur. Université Joseph Fourier, Grenoble, France.

Mellersh C.S., Langston A.A., Acland G.M., Fleming M.A., Ray K., Wiegand N.A., Francisco L.V., Gibbs M., Aguirre G.D. & Ostrander E.A. (1997) A linkage map of the canine genome. *Genomics* **46,** 326–36.

Montgomery G.W. & Sise J.A. (1990) Extraction of DNA from sheep white blood cells. *New Zealand Journal of Agricultural Research* **33,** 437.

Neff M.W., Broman K.W., Mellersh C.S., Ray K., Acland G.M., Aguirre G.D., Ziegle J.S., Ostrander E.A. & Rine J. (1999) A second-generation genetic linkage map of the domestic dog, *Canis familiaris*. *Genetics* **151,** 803–20.

Negrini R., Milanesi E., Colli L., Pellecchia M., Nicoloso L., Crepaldi P., Lenstra J.A. & Ajmone-Marsan P. (2007) Breed assignment of Italian cattle using biallelic AFLP markers. *Animal Genetics* **38,** 147–53.

O'Reilly P.T., Canino M.F., Bailey K.M. & Bentze P. (2004) Inverse relationship between F_{ST} and microsatellite polymorphism in the marine fish, walleye pollock (*Theragra chalcogramma*): implications for resolving weak population structure. *Molecular Ecology* **13,** 1799–1814.

Ostrander E.A., Sprague G.F. Jr & Rine J. (1993) Identification and characterization of dinucleotide repeat (CA)n markers for genetic mapping in dog. *Genomics* **16,** 207–13.

Paetkau D., Calvert W., Stirling I. & Strobeck C. (1995) Microsatellite analysis of population structure in Canadian polar bears. *Molecular Ecology* **4,** 347–54.

Paetkau D., Slade R., Burden M. & Estoup A. (2004) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology* **13,** 55–65.

Parker H.G., Kim L.V., Sutter N.B., Carlson S., Lorentzen T.D., Malek T.B., Johnson G.S., DeFrance H.B., Ostrander E.A. & Kruglyak L. (2004) Genetic structure of the purebred domestic dog. *Science* **304,** 1160–4.

Parra D., Mendez S., Canon J. & Dunner S. (2008) Genetic differentiation in pointing dog breeds inferred from microsatellites and mitochondrial DNA sequence. *Animal Genetics* **39,** 1–7.

Peakall R. & Smouse P.E. (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6,** 288–95.

Pereira L., Prata M.J. & Amorim A. (2000) Diversity of mtDNA lineages in Portugal: not a genetic edge of European variation. *Annals of Human Genetics* **64,** 491–506.

Pereira F., Davis S.J., Pereira L., McEvoy B., Bradley D.G. & Amorim A. (2006) Genetic signatures of a Mediterranean influence in Iberian Peninsula sheep husbandry. *Molecular Biology and Evolution* **23,** 1420–6.

Petrucci-Fonseca F., Pires A.E., Ribeiro S. *et al.* (2000) Cães de gado e a conservação do lobo em Portugal [Livestock guarding dogs and wolf conservation in Portugal]. *Galemys* **12,** 135–48.

Pires A.E. (2006) *Phylogeny, Population Structure and Genetic Diversity of Dog Breeds in the Iberian Peninsula and North Africa*. PhD Thesis, University of Lisbon, Portugal.

Pires A.E., Ouragh L., Kalboussi M., Matos J., Petrucci-Fonseca F. & Bruford M.W. (2006) Mitochondrial DNA sequence variation in Portuguese native dog breeds: diversity and phylogenetic affinities. *Journal of Heredity* **97,** 318–30.

Piry S., Alapetite A., Cornuet J.M., Paetkau D., Baudouin L. & Estoup A. (2004) GENECLASS2: a software for genetic assignment and first-generation migrant detection. *Journal Heredity* **95,** 536–9.

Pritchard J.K. & Wen W. (2004) *Documentation for the STRUCTURE software Version 2*. http://pritch.bsd.uchicago.edu/structure.html.

Pritchard J.K., Stephens M. & Donnelly P. (2000) Inference of population structure using multilocus genotype data. *Genetics* **155,** 945–59.

Ribeiro A. & Saraiva J.H. (2004) *História de Portugal. A Formação do Território – da Lusitânia ao Alargamento do País. [History of Portugal Territory Formation – from Lusitania to the Expansion of the Country]*. Quidnovi®, Matosinhos, Portugal.

Sambrook E., Fritsch F. & Maniatis T. (1989) *Molecular Cloning*. Cold Spring Harbor Press, New York, NY.

Sargan D.R., Aguirre-Hernandez J., Galibert F. & Ostrander E.A. (2007) An extended microsatellite set for linkage mapping in the domestic dog. *Journal of Heredity* **98,** 221–31.

Schneider S., Roessli D. & Excoffier L. (2000) *ARLEQUIN VER. 2.000: A Software for Population Genetics Data Analysis*. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Sundqvist A.-K., Ellegren H. & Vilà C. (2008) Wolf or dog? Genetic identification of predators from saliva collected around bite wounds on prey. *Conservation Genetics* **9,** 1275–9.

Tapio M., Tapio I., Grislis Z., Holm L.E., Jeppsson S., Kantanen J., Miceikiene I., Olsaker I., Viinalass H. & Eythorsdottir E. (2005) Native breeds demonstrate high contributions to the molecular

variation in northern European sheep. *Molecular Ecolology* **14,** 3951–63.
Walsh P.S., Metzger D.A. & Higuchi R. (1991) Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *BioTechniques* **10,** 506–13.
Weir B.S. & Cockerham C.C. (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution* **38,** 1358–70.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Information regarding the studied dog breeds.
Table S2 Genetic diversity of dog breeds/populations.
Table S3 Pairwise F_{ST} values for both markers.
Table S4 Analysis of molecular variance (AMOVA).
Table S5 GENECLASS assignment results.
Table S6 Percentage of animals correctly assigned in each population.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors.