doi:10.1111/j.1365-2052.2008.01842.x

Genetic diversity of dog breeds: within-breed diversity comparing genealogical and molecular data

G. Leroy*,†,‡, E. Verrier*,†, J. C. Meriaux§ and X. Rognon*,†

*AgroParisTech, UMR1236 Génétique et Diversité Animales, 16 rue Claude Bernard, F-75321 Paris, France. †INRA, UMR1236 Génétique et Diversité Animales, Domaine de Vilvert, F-78352 Jouy-en-Josas, France. ‡Société Centrale Canine, 155 av. Jean Jaurès, F-93535 Aubervilliers, France. §LABOGENA, Domaine de Vilvert, F-78352 Jouy-en-Josas, France

Summary

The genetic diversity of 61 dog breeds raised in France was investigated. Genealogical analyses were performed on the pedigree file of the French kennel club. A total of 1514 dogs were also genotyped using 21 microsatellite markers. For animals born from 2001 to 2005, the average coefficient of inbreeding ranged from 0.2% to 8.8% and the effective number of ancestors ranged from 9 to 209, according to the breed. The mean value of heterozygosity was 0.62 over all breeds (range 0.37–0.77). At the breed level, few correlations were found between genealogical and molecular parameters. Kinship coefficients and individual similarity estimators were, however, significantly correlated, with the best mean correlation being found for the Lynch & Ritland estimator ($r = 0.43$). According to both approaches, it was concluded that special efforts should be made to maintain diversity for three breeds, namely the Berger des Pyrénées, Braque Saint-Germain and Bull Terrier.

Keywords breed, diversity, dog, heterozygosity, inbreeding, microsatellite.

Introduction

As more and more inherited diseases have been identified in purebred dogs during recent years (Patterson 1993), management of genetic variation has become a major concern for people involved in dog breeding. Kennel clubs are therefore more and more interested in parameters that evaluate the genetic variability in order to make decisions about selection, particularly because inbreeding is sometimes considered as a selection tool by breeders (Leroy *et al.* 2007).

To evaluate the genetic diversity among domestic breeds, one may use genealogical data. Such an approach has been extensively used on dog breeds (e.g. Mäki *et al.* 2001; Leroy *et al.* 2006; Głażewska 2007). The most commonly used tool to evaluate diversity in a population is the coefficient of inbreeding, defined as the probability that two alleles at a given locus are identical by descent (Malécot 1948). Also, it has often been suggested that attempts be made to maximize diversity by minimizing kinship among individuals (Caballero & Toro 2000), kinship being considered as the inbreeding coefficient for the hypothetical offspring of two individuals. Effective population size and ancestral approaches constitute two other useful tools to evaluate the diversity through genealogical data. Such approaches are, however, limited by incomplete knowledge of the genealogies, depending on populations, even if some indicators are less influenced by this factor than others (Boichard *et al.* 1997). The use of genetic markers constitutes a more recent approach to evaluate the within-breed diversity. It has been used in several studies (Irion *et al.* 2003; Parker *et al.* 2004; Parra *et al.* 2008). The two main indicators of molecular diversity are the rate of heterozygosity (H) and the allelic richness (A_r). Only a few studies have, however, combined the two different tools, pedigrees and molecular markers, at the level of domestic breeds. In the domestic dog, the study by Lüpke & Distl (2005) on the Hannoverian Hound breed is the only one to have used both genealogical data and molecular markers.

The aim of this study was to evaluate the within-breed genetic diversity, using both genealogical and molecular data, on a large set of dog breeds, being representative of French dog breeds and covering a range of management conditions. In addition, correlations between some of the indicators computed using the different approaches were analysed at the level of breeds and individuals.

Address for correspondence

G. Leroy, AgroParisTech, UMR1236 Génétique et Diversité Animales, 16 rue Claude Bernard, F-75321 Paris, France.
E-mail: gregoire.leroy@agroparistech.fr

Accepted for publication 7 November 2008

Materials and methods

Breeds studied

Among the 300 breeds raised in France, a set of 61 breeds was sampled in order to represent the 10 groups of the Fédération Cynologique Internationale (FCI) nomenclature (abbreviations for each breed are shown in Table 1). According to the French Kennel Club data, 33 of these 61 breeds showed the largest number of dogs registered in France. Among the other breeds, there were some local (CUR) or rare (BAR, BSG, CUR) French breeds. Because the French Kennel Club (SCC) allows registration of Jack Russell Terrier/Parson Russell Terrier litters into both breeds (according to the puppy sizes), the two breeds were considered here as a single one (RUT).

Pedigree analysis

The national pedigree file managed by the SCC was used, including all dogs born from registered litters. The analyses were performed using the PEDIG software (Boichard 2002). For more details on the methods used, see for instance Boichard *et al.* (1997) and Leroy *et al.* (2006). For each breed, the cohort (or reference population) was defined as all animals born from 2001 to 2005. The number of animals in this reference population and the number of breeders having registered litters during this period are given in Table 1. The following parameters were computed for the reference population. The number of equivalent complete generations traced (EqG) was used as an indicator of pedigree completeness. The generation length (T) was computed in the four pathways as the average age of parents at the birth of their offspring; here, only the 'useful' offspring were considered, i.e. offspring, which became parents themselves. The effective number of founders (f_e) and the effective number of ancestors (f_a) are defined as the reciprocal of the probability that two genes drawn at random in the reference population come from the same founder or from the same ancestor respectively. Founder animals were defined as animals with no known parents. Ancestors, founders or not, are defined as the animals having the most expected genetic contribution to the reference population and were detected by the iterative method proposed by Boichard *et al.* (1997). By nature, the effective number of ancestors is lower than the effective number of founders, the difference being due to bottlenecks that have occurred from the base population to the reference population. Average coefficient of inbreeding (F) and average coefficient of kinship (Φ) were also computed.

For each breed, the annual rate of inbreeding was estimated by linear regression over time, and the realized effective size (N_{er}) was computed from the rate of inbreeding per generation (ΔF), using the following formula: $N_{er} = 1/2\Delta F$. Because some breeds were seldom present in France until the 1990s, ΔF and N_{er} were computed only for breeds having an EqG higher than 4, over the period 1980–2005.

Animals sampled and choice of markers

The biological samples used had two origins: mainly, buccal swabs made during dog shows or dog trials and, to a lesser extent, samples sent by some laboratories or breeders. For each breed, the animals sampled were chosen to fulfil specific conditions. First, the presence of close relatives (e.g. full sibs) was avoided. Second, we checked the absence of significant difference ($P < 0.05$) between the sample and the whole reference population for the values of some simple parameters of the pedigree: pedigree completeness level (EqG), the proportions of null coefficients of inbreeding (F) and of null coefficients of kinship (Φ), and the average F and Φ. In the CUR breed, because of a lack of pedigree data, the animals were chosen from different litters. A total of 1514 animals were sampled, with a mean of 25 dogs per breed.

The 21 autosomal microsatellite markers of the ISAG panel (http://www.isag.org.uk/ISAG/all/2005ISAGPanelDOG.pdf) were chosen to perform the molecular analysis. For the entire sample, the amplification and the genotyping were performed by the same laboratory, namely Labogena, using a capillary sequencer (ABI PRISM 3100 Genetic Analyzer; Applied Biosystems).

Analysis of the marker polymorphism

Observed (H_o), or non-biased heterozygosity (H_e), and Wright's F_{IS} coefficient were estimated using GENETIX software (Belkhir *et al.* 1996). Allelic richness A_r was estimated using FSTAT (Goudet 2001) by the rarefaction method (El Mousadik & Petit 1996). The Nei's genetic differentiation parameter, G_{ST}, was computed, as well as allelic richness differentiation ρ_{ST}, using the following formula (El Mousadik & Petit 1996): $\rho_{ST} = 1 - A_{rm}/A_{rT}$, where A_{rm} is the average A_r computed over all breeds, and A_{rT} is the allelic richness over the whole sampling, according to the rarefaction method. For each breed/locus, Hardy–Weinberg equilibrium (HWE) was checked, with sequential Bonferroni corrections, using the GENEPOP software (Raymond & Rousset 1995). Global tests across loci and samples were performed using the Fisher method. A simple general linear model was used to test FCI breed group effect on genealogical and molecular indicators.

Comparison of genealogical and molecular results

Because of the influence of pedigree knowledge on genealogical results, only the 24 breeds showing an EqG higher than 6 were considered for the comparison. Canonical correlation analysis was performed at the breed level using the SAS software (SAS Institute Inc. 2004). The following factors were included in the analysis: reference population

Table 1 Characteristics of the data files of the 61 dog breeds, with the reference population being considered as the animals born from 2001 to 2005.

Code used in this paper	Full name	FCI group	No. dogs in pedigree file	Reference population size	No. breeders	EqG	*T*	No. dogs sampled
ARI	Ariégeois	6	9108	2806	235	3.9	4.3	22
ASD	Australian Shepherd	1	10 424	6154	225	4.0	3.4	22
AST	American Staffordshire Terrier	3	41 900	24 803	1654	5.6	3.3	29
BAF	Basset fauve de Bretagne	6	25 750	5182	467	7.0	4.4	30
BAR	Barbet	8	875	112	9	3.1	4.2	20
BEA	Beagle	6	54 420	11 690	814	7.9	4.5	20
BECo	Bearded Collie	1	10 978	2649	85	5.0	5.2	30
BEN	Beauceron	1	98 503	19 072	844	9.1	4.5	30
BLD	Bulldog	2	12 324	4347	350	4.0	3.2	30
BLF	Bouledogue français	9	31 347	14 718	974	7.6	3.0	30
BMD	Bernese Mountain Dog	2	33 666	12 325	551	5.0	3.7	22
BOCo	Border Collie	1	18 995	6415	589	2.9	4.6	20
BOX	German Boxer	2	88 357	12 732	829	5.3	3.7	20
BRP	Berger des Pyrénées	1	12 738	3742	179	6.7	4.7	28
BRZ	Borzoi	10	11 970	1324	74	4.9	4.2	25
BSD	Belgian Shepherd Dog Malinois	1	64 780	18 276	985	6.7	5.2	29
BSG	Braque Saint-Germain	7	2167	348	31	8.0	5.0	20
BUT	Bull Terrier	3	9327	3378	258	4.0	3.1	23
CAT	Cairn Terrier	3	33 875	7869	292	6.7	4.2	20
CKC	Cavalier King Charles Spaniel	9	68 797	27 392	1150	6.6	3.7	30
COT	Coton de Tuléar	9	30 556	8755	252	6.2	4.4	25
CSP	English Cocker Spaniel	8	122 845	23 464	906	6.7	4.1	30
CUR	Cursinu	5	198	42	5	0.3	3.7	22
CWD	Czeslovakian Wolfdog	1	856	244	21	2.2	3.0	22
DAL	Dalmatian	6	17 778	4473	236	5.5	4.2	20
DBM	Dobermann	2	61 738	10 402	535	4.4	4.1	30
DOA	Dogo Argentino	2	14 229	7541	488	4.1	3.3	21
DOB	Dogue de Bordeaux	2	12 212	4055	256	6.9	3.3	30
EPB	Epagneul Breton	7	156 492	27 703	1714	9.2	4.7	30
ESE	English Setter	7	138 934	26 575	2340	6.4	5.1	20
GBG	Griffon bleu de Gascogne	6	11 713	4234	362	3.7	4.1	27
GPD	German Short-haired Pointing Dog	7	70 337	8392	623	7.5	4.8	30
GRD	Great Dane	2	52 678	8799	366	7.1	3.4	30
GRT	Golden Retriever	8	79 473	36 258	1315	4.8	4.2	20
GSD	German Shepherd Dog	1	353 117	56 583	1864	5.1	4.1	30
ICD	Italian Corso Dog	2	12 890	8672	452	3.5	4.0	30
ISE	Irish Red Setter	7	33 066	3143	188	6.2	5.3	30
KCS	King Charles Spaniel	9	4199	1320	101	5.7	3.7	20
KOR	Griffon d'arrêt à poil dur Korthals	7	40 720	6711	502	8.5	5.1	27
LEO	Leonberger	2	24 217	7546	306	6.5	4.2	20
LRT	Labrador Retriever	8	177 421	41 670	1640	5.5	4.4	22
MOP	Chien de Montagne des Pyrénées	2	18 498	1941	85	6.1	4.5	29
NFL	Newfoundland	2	35 336	5800	344	4.6	3.9	30
PNT	English Pointer	7	59 460	7881	765	5.8	4.8	20
POO	Poodle	9	89 826	8808	324	6.0	4.7	24
RGCo	Collie Rough	1	81 873	6529	251	5.7	3.8	27
ROT	Rottweiler	2	76 511	28 544	1849	4.8	3.9	20
RUT	Parson/Jack Russell Terrier	3	23 138	13 668	716	3.2	3.4	30
RWD	Romagna Water Dog	8	713	229	21	2.5	3.0	23
SDH	Smooth-haired Dachshund	4	36 441	3098	138	5.8	4.6	23
SHI	Shih Tzu	9	47 162	10 907	558	5.9	4.1	27
SHP	Shar Pei	2	20 829	5801	343	4.7	3.7	20

Table 1 Continued.

Code used in this paper	Full name	FCI group	No. dogs in pedigree file	Reference population size	No. breeders	EqG	*T*	No. dogs sampled
SHU	Siberian Husky	5	53 946	4221	228	4.8	5.0	25
SPI	German Spitz	5	19 867	2853	188	6.7	4.2	20
SSP	English Springer Spaniel	8	31 116	8457	620	6.3	4.4	23
SWD	Saarloos Wolfdog	1	651	201	22	3.1	5.1	20
WDH	Wire-haired Dachshund	4	78 119	14 360	1183	5.9	4.7	21
WEI	Short-haired Weimaraner	7	25 179	6736	513	8.4	4.5	24
WHI	Whippet	10	32 255	4506	315	5.4	4.9	20
WHT	West Highland White Terrier	3	55 887	14 100	669	5.8	3.9	24
YOT	Yorkshire Terrier	3	142 382	24 907	1121	6.6	4.3	28

EqG, number of equivalent generations; *T*, average generation length in years.
FCI groups: 1 – Sheepdogs and cattle dogs (except Swiss cattle dogs); 2 – Pinscher and Schnauzer, Molossoid breeds, Swiss mountain and cattle dogs and other breeds; 3 – Terriers; 4 – Dachshunds; 5 – Spitz and primitive types; 6 – Scent hounds and related breeds; 7 – Pointing dogs; 8 – Retrievers, Flushing dogs, Water dogs; 9 – Companion and toy dogs; 10 – Sighthounds.

size, inbreeding F and kinship Φ of the reference population, F/Φ ratio, N_{er}, f_e and f_a as demographic/genealogical parameters, H_e, A_r and F_{IS} as molecular parameters.

For each of the 24 breeds, Pearson correlations were computed between individual genealogical kinship and different estimators of molecular resemblance. IDENTIX was used (Belkhir *et al.* 2002) to compute various estimators: R_{QG} from Queller & Goodnight (1989), R_M from Mathieu *et al.* (1990) and R_{LR} from Lynch & Ritland (1999). MOLKIN (Gutiérrez *et al.* 2005) estimated molecular similarity (S_{EM}) (Eding & Meuwissen 2001) and shared allele distance (DAS) (Chakraborty & Jin 1993). The D_a distance (Nei *et al.* 1983) was also computed between animals using the POPULATION software (Olivier Langella; http://www.pge.cnrs-gif.fr/bioinfo/populations/). Differences between each pair of correlations were tested across breeds using Wilcoxon matched pair tests.

Results

Demographical and genealogical results

The breeds studied showed variable demographical situations, the number of dogs registered between 2001 and 2005 ranging from 42 (CUR) to 56 583 (GSD) (Table 1). Six breeds (BAR, BSG, CUR, CWD, RWD, SWD) had <500 dogs registered during this period. The number of breeders having registered litters during the period ranged from 5 (CUR) to 2340 (ESE). According to the breed, the mean number of dogs registered by breeders ranged from 8 (CUR) to 35 (COT). The mean generation length (T) was on average 4.2 years for all breeds, with a range from 3.0 (BLF) to 5.3 (ISE). T values were significantly different across breeds ($P < 0.0001$) and across FCI groups ($P = 0.001$), with pointing dog breeds (group 7) having higher T-values (4.9) than groups 2, 3 and 9 (3.8, 3.7 and 3.9 respectively). Depending on the breed, the pedigree knowledge varied from almost non-existent (CUR breed with an EqG of 0.3 generations), to largely complete for breeds such as BEN and EPB, with values of EqG equal to 9.1 and 9.2 respectively. Twenty-four breeds had an EqG value higher than 6. FCI breed groups also had a significant effect on EqG ($P < 0.04$), with pointing dog breeds having deeper pedigree knowledge than other FCI groups (7.5 vs. 5.5 on average). When comparing reference population size and number of breeders, the number of dogs produced by breeders during the 2001–2005 period ranged from 8 (CUR) to 35 (COT) (18 on average), showing that most of them were occasional and/or hobby breeders. FCI groups affected neither this ratio nor any other genealogical or molecular parameters.

Results on the genealogical estimators of diversity are shown in Table 2 for all breeds. Regarding F and Φ, the BRP breed showed the highest values (respectively 8.8% and 5.5%), while no inbreeding coefficient could be computed in the CUR breed because of lack of pedigree knowledge. The POO breed showed the lowest mean kinship (0.4%). The realized effective size N_{er} was not computed for BLF and GSD breeds, although their EqG was higher than 4, because the rate of inbreeding ΔF computed over the period 1980–2005 was found to be null and negative for these two breeds due to a large number of imports. Among the other breeds, N_{er} ranged from 46 (ISE) to 2136 (WHT), with the mean value being 226. According to ancestral approaches, the POO breed showed the highest diversity with respective f_e and f_a of 656 and 209, and the BAR breed had the lowest diversity ($f_e = 10$ and $f_a = 9$). Across breeds, the ratio f_a/f_e ranged from 0.2 (BSG) to 0.95 (BOCo).

Heterozygosities and Hardy–Weinberg equilibrium

Amongst the 21 markers, 240 alleles were identified, with the number of alleles per marker ranging from 7 to 18 (Table S1). Results showed a G_{ST} of 0.24 and a ρ_{ST} of 0.44.

Table 2 Genealogical and molecular estimators of the genetic diversity of the 61 breeds studied [N_{er} was not computed for breeds with EqG lower than 4 and for the two breeds with null or negative ΔF (GSD and BLF)].

Breed code	Genealogical estimators					Molecular estimators			
	F (%)	Φ (%)	N_{er}	f_e	f_a	H_e	H_o	F_{IS} (%)	A_r
ARI	3.2	1.6	_	58	48	0.70	0.67	4.2	5.5
ASD	1.1	1.2	_	167	55	0.66	0.65	1.4	5.2
AST	4.6	2.3	60	88	29	0.56	0.54	4.0	4.1
BAF	4.1	2.2	88	85	41	0.70	0.68	2.7	5.4
BAR	6.4	5.5	_	10	9	0.70	0.72	−2.8	5.2
BEA	4.6	1.9	76	163	44	0.69	0.66	4.1	5.1
BECo	4.4	2.2	60	111	37	0.55	0.57	−3.3	3.6
BEN	6.1	3.7	56	75	34	0.65	0.62	4.5	4.9
BLD	1.3	0.7	1216	276	84	0.53	0.52	1.7	3.5
BLF	3.2	2.4	_	105	41	0.60	0.59	2.3	4.1
BMD	1.8	1.2	167	154	58	0.53	0.49	8.0	3.8
BOCo	0.8	0.7	_	95	90	0.66	0.60	8.2	5.2
BOX	2.4	1.2	231	81	68	0.47	0.46	2.4	3.2
BRP	8.8	5.7	30	33	15	0.67	0.67	−0.9	4.9
BRZ	2.7	1.1	147	213	80	0.62	0.58	5.3	4.2
BSD	4.3	2.1	161	106	44	0.72	0.69	4.4	5.6
BSG	7.5	8.8	29	64	13	0.59	0.59	−0.4	4.1
BUT	1.1	1.3	295	133	53	0.37	0.41	−10.1	2.3
CAT	3.1	1.7	145	182	49	0.61	0.58	5.0	4.4
CKC	3.3	1.4	150	200	61	0.47	0.45	4.1	3.1
COT	6.2	3.9	56	27	25	0.73	0.70	3.6	6.0
CSP	2.3	0.6	199	512	140	0.65	0.56	13.5	4.4
CUR	_	_	_	_	_	0.77	0.77	0.4	6.9
CWD	0.2	1.2	_	61	22	0.61	0.60	1.3	4.4
DAL	2.4	1.9	120	136	43	0.58	0.59	−0.8	3.9
DBM	2.3	1.0	187	98	71	0.38	0.40	−5.4	2.7
DOA	1.5	1.0	310	128	61	0.64	0.66	−3.3	4.4
DOB	3.9	3.3	993	58	34	0.52	0.50	2.3	3.4
EPB	5.2	3.3	78	69	31	0.68	0.71	−3.2	5.0
ESE	2.1	1.3	195	186	58	0.66	0.64	2.7	5.1
GBG	2.0	1.4	_	67	54	0.74	0.70	6.4	6.1
GPD	3.5	2.5	81	150	40	0.70	0.69	1.8	5.4
GRD	4.4	1.1	375	235	94	0.67	0.64	4.6	4.8
GRT	1.3	0.6	219	243	106	0.58	0.58	0.2	3.9
GSD	1.8	0.7	_	152	129	0.55	0.52	4.3	3.7
ICD	1.4	0.9	_	140	61	0.71	0.68	3.5	5.3
ISE	5.8	2.0	46	183	43	0.70	0.65	6.9	5.3
KCS	2.8	2.8	218	105	27	0.44	0.46	−5.3	2.9
KOR	5.6	4.3	49	68	29	0.69	0.71	−2.7	5.0
LEO	2.8	2.3	722	97	50	0.59	0.61	−4.1	3.9
LRT	2.2	0.7	122	345	97	0.60	0.58	2.1	4.4
MOP	5.8	3.1	112	67	35	0.60	0.60	0.2	4.5
NFL	2.3	1.0	158	198	64	0.63	0.66	−5.1	4.7
PNT	2.4	1.5	161	99	50	0.62	0.57	8.0	4.8
POO	4.7	0.4	54	656	209	0.72	0.60	17.3	5.6
RGCo	3.9	1.2	95	165	63	0.45	0.44	2.8	3.0
ROT	1.7	0.6	274	189	124	0.55	0.54	0.5	4.0
RUT	2.2	1.2	_	63	56	0.77	0.75	2.6	6.3
RWD	0.3	2.1	_	52	22	0.65	0.67	−2.1	4.9
SDH	5.0	1.1	82	241	78	0.63	0.57	9.2	4.7
SHI	2.8	0.8	176	310	91	0.63	0.62	0.7	4.4
SHP	2.9	1.2	89	148	63	0.72	0.67	7.2	5.8
SHU	2.6	0.8	90	197	94	0.64	0.60	6.4	4.4

Table 2 Continued.

Breed code	Genealogical estimators					Molecular estimators			
	F (%)	Φ (%)	N_{er}	f_e	f_a	H_e	H_o	F_{IS} (%)	A_r
SPI	6.6	2.1	52	207	48	0.71	0.63	11.7	5.6
SSP	2.9	1.8	128	126	46	0.71	0.64	10.3	5.3
SWD	6.4	3.6	_	25	14	0.52	0.53	−2.5	3.6
WDH	3.0	0.8	109	261	99	0.69	0.66	5.3	5.3
WEI	5.6	4.6	50	90	23	0.64	0.66	−3.3	4.3
WHI	3.5	1.1	87	226	66	0.64	0.56	12.5	4.5
WHT	2.2	0.9	2136	317	93	0.46	0.47	−0.5	3.0
YOT	3.4	0.9	117	145	86	0.70	0.71	−2.4	4.9

F, mean inbreeding coefficient; Φ, mean kinship coefficient; N_{er}, realized effective size; f_e, effective number of founders; f_a, effective number of ancestors; H_e, non-biased heterozygosity; H_o, observed heterozygosity; A_r, allelic richness.

According to the breed, H_e values ranged from 0.37 (BUT) to 0.77 (CUR and RUT) with a mean value around 0.62 (Table 2). A_r values rose from 2.3 (BUT) to 6.9 (CUR) with a mean value around 4.56. F_{IS} ranged from −0.1 (BUT) to 0.17 (POO). Out of the 1281 HWE tests performed, 14 showed a significant deficit after sequential Bonferonni corrections, concerning different loci and breeds. Using global tests, seven breeds (SDH, GBG, PNT, ISE, CSP, POO, WHI) and four loci (*AHTh171*, *FH2848*, *INU005*, *REN247M23*) were found to have a deficit. No test was found to be significant for heterozygote excess.

Canonical correlation analysis

Only the first canonical correlation of genealogical and molecular data was significant ($r = 0.80$, $P = 0.03$). The first genealogical canonical axis was primarily associated with f_e ($r = 0.910$), while the first molecular axis was associated primarily with F_{IS} ($r = 0.88$). As shown in Table 3, H_e and A_r were strongly correlated as expected ($r = 0.94$, $P < 0.0001$), as well as f_e and f_a ($r = 0.90$, $P < 0.0001$) and Φ and *F* ($r = 0.75$, $P < 0.0001$). Φ and f_a were negatively correlated ($r = 0.71$, $P < 0.0001$), while F_{IS} was positively correlated with f_e ($r = 0.66$, $P = 0.0004$) and F/Φ ratio ($r = 0.50$, $P = 0.01$), and negatively with Φ ($r = -0.46$, $P = 0.03$).

Correlations between genealogical and molecular estimators of individual similarities

Values of mean correlations over the 24 breeds ranged from 0.33 (S_{EM} and $1 - D_a$) to 0.43 (R_{LR}) (Table 4). Correlations dropped from 0.71 (R_{LR} estimator on the BSG breed) to 0.05 (S_{EM} on MOP breed) depending on the breed and the estimator used (Table S2). Except for this last case, all correlations with genealogical kinship were found to be significant ($P < 0.05$) for the 24 breeds and the six coefficients tested. Correlations were found to be significantly

Table 3 Correlation among eight genealogical/molecular parameters and their canonical variable, based on 24 breeds.

	Correlations between parameters and their canonical variable	Reference population size	Φ	*F*	F/Φ	N_{er}	f_e	f_a	H_e	A_r	F_{IS}
Reference population size	−0.57	1	0.06	−0.07	−0.36	−0.04	−0.33	−0.24	−0.46	−0.26	−0.26
Φ	−0.49	0.06	1	0.75	−0.63	−0.07	−0.60	−0.71	−0.12	−0.11	−0.46
F	−0.24	−0.07	0.75	1	−0.19	−0.22	−0.49	−0.61	0.18	0.18	−0.21
F/Φ	0.74	−0.36	−0.63	−0.19	1	−0.17	0.71	0.80	0.25	0.12	0.50
N_{er}	0.05	−0.04	−0.07	−0.22	−0.17	1	−0.06	−0.03	−0.21	−0.26	−0.04
f_e	0.91	−0.33	−0.60	−0.49	0.71	−0.06	1	0.90	0.01	−0.12	0.66
f_a	0.74	−0.24	−0.71	−0.61	0.80	−0.03	0.90	1	0.01	−0.11	0.49
H_e	0.20	−0.46	−0.12	0.18	0.25	−0.21	0.01	0.01	1	0.94	0.20
A_r	−0.00	−0.26	−0.11	0.18	0.12	−0.26	−0.12	−0.11	0.94	1	0.16
F_{IS}	0.88	−0.26	−0.46	−0.21	0.50	−0.04	0.66	0.49	0.20	0.16	1

Φ, mean kinship coefficient; *F*, mean inbreeding coefficient; N_{er}, realized effective size; f_e, effective number of founders; f_a, effective number of ancestors; H_e, non-biased heterozygosity; A_r, allelic richness.

Table 4 Average values across breeds of correlations between genealogical kinship and different estimators of marker similarity or distance, computed on pairs of individuals of 24 breeds.

	S_{EM}	1 − DAS	R_{QG}	R_M	R_{LR}	1 − D_a
Mean correlation	0.33	0.35	0.37	0.35	0.43	0.33

Estimators of molecular resemblance and distances: S_{EM} (Eding & Meuwissen 2001); DAS (Chakraborty & Jin 1993); R_{QG} (Queller & Goodnight 1989); R_M (Mathieu *et al.* 1990); R_{LR} (Lynch & Ritland (1999); D_a (Nei *et al*. 1983).

higher ($P < 0.05$) across breeds when using R_{LR} and R_{QG} than when using other estimators.

Discussion

The aim of this study was to assess within-breed genetic diversity on a large set of breeds showing various demographic conditions and breeding programmes. Some differences found in genealogical indicators of variability could be related to historical origins, specific breeding practices and physiological parameters of the breeds. Among the five breeds with the deepest pedigree knowledge, four (BEN, BSG, EPB, KOR) were of French origin and four (BSG, EPB, KOR, WEI) were pointing dog breeds, which can be related, in both cases, to a low number of imported dogs. In pointing dog breeds, the larger proportions of occasional breeders (Leroy *et al.* 2007), which rarely use imported dogs, may explain this fact. Moreover, pedigrees are generally better known in French breeds than in foreign breeds, as a large number of the worldwide population is raised in France. The pedigree knowledge was, however, low in the case of recently recognized or partially reconstructed French breeds (CUR and BAR) or French breeds where a large number of dogs are usually registered without pedigree (ARI and GBG). Concerning mean generation lengths, T was quite low for breeds from some FCI groups: group 2, which includes large-sized breeds with small longevity (Jones *et al.* 2008) that end their reproductive career earlier (Leroy *et al.* 2007), and groups 3 and 9, involving small-sized breeds, which are allowed by French kennel club rules to begin their reproductive career earlier. In contrast, in pointing dog breeds (group 7), a large number of bitches are used for hunting and could begin their reproductive career later. Such practices could explain the higher generation length found for those breeds.

The results may be compared with those found in other studies. For instance, Mäki *et al.* (2001) found genealogical results of the same order for GSD, GRT and LRT Finnish populations: for a number of equivalent generations of 4.3 years for the three breeds (against 5.1, 4.8 and 5.5 in our study respectively), the mean F was slightly higher in the Finnish study with 2.3%, 3.0% and 2.3% against 1.8%, 1.3% and 2.2% in France respectively. This can be easily explained by differences in population sizes and situations. In a study on the Polish Hound, Głażewska (2007) found a very high mean inbreeding coefficient (37%), but the population combined a very small size with a complete pedigree knowledge. Only small differences were found with the results obtained on the nine French breeds studied by Leroy *et al.* (2006), and most of them can be explained by differences between the cohorts and the whole data set. Since 2006, new information has been added in SCC pedigree files. The only striking difference was found for BSG, where f_e was 21 in the previous study and 64 in the present study, whereas f_a was similar (13) in both studies. Therefore, using the cohort of Leroy *et al.* (2006) and the current genealogical data set, f_e was also found to be 64, while EqG was higher than in the Leroy *et al.* (2006) study (7.1 against 5.9). This difference in the number of known generations for the same reference population modified the founder population, which could probably explain the increase in f_e.

The comparison between f_e and the effective number of ancestors f_a can reveal bottlenecks (Boichard *et al.* 1997). Consequently, we can suppose that BSG breed has gone through an early bottleneck that was not detected in the first study. Regarding molecular data, the results on heterozygosity values were quite similar with those of a US-based study on 28 breeds and 100 microsatellites (Irion *et al.* 2003) that showed large differences according to the breed: in both studies, BUT had the lowest heterozygosity in all breeds (0.39 and 0.37 in our study), while the RUT breed showed the highest heterozygosity (0.76 and 0.77 respectively). When compared with another previous study that used 21 microsatellite markers (Parra *et al.* 2008), heterozygosity values were similar for the four common breeds (ESE, EPB, GPD, PNT), ranging between 0.60 and 0.70 in both studies.

The results of this study clearly illustrate the differences in information that can be gleaned about genetic diversity using genealogical and molecular data. For molecular indicators, if heterozygosity values and allelic richness are highly correlated over populations, breeds are found to be more differentiated in allelic richness than in heterozygosity. Such a result is in agreement with Foulley & Ollivier's (2006) on pig breeds. Two breeds (CUR and RUT) showed the highest H_e (0.77), but the level of A_r was much higher in the CUR breed (6.9) than in the RUT breed (6.3). This result may be explained by the properties of allelic richness, which are more sensitive to bottlenecks than heterozygosity (Foulley & Ollivier 2006). The CUR breed has been recognized only recently (2003) and has had no real breeding programmes until now. Therefore, in contrast to other breeds, this breed has probably not suffered from bottlenecks, which could explain its higher A_r value.

Differences found between genealogical and molecular estimators can be explained by the differing characteristics of the two approaches. Indeed, genealogical analysis provides a comprehensive view of the evolution of genetic variability from the base population, particularly if the

pedigree data are complete and reliable. Molecular data are obtained on a limited number of markers and thus there may be a sampling effect. It is generally assumed that breeds with small population sizes, high inbreeding and low genealogical diversity parameters (N_{er}, f_a…) should have low heterozygosity values and allelic richness, but it is much more difficult to infer genealogical parameters from molecular data (Toro *et al.* 2008). Analysing six horse breeds with genealogical and allozyme/blood groups data, Moureaux *et al.* (1996) found no simple relationship between marker and genealogical data at the breed level, while in the study of Toro *et al.* (2002) on two pig breeds, the population with the highest genealogical kinship had also the lowest molecular diversity. In our study, correlations between both kinds of estimators were generally not significant. The two parameters primarily used to compute the canonical variable (f_e and F_{IS}) were positively correlated, although no correlation should have been expected. The results of canonical correlation analysis should therefore be taken with caution. Among genealogical and demographic indicators, a positive correlation across breeds between realized effective size (N_{er}) and reference population size was expected. However, several factors influencing N_{er}, such as variations in number of unregistered/imported dogs over time (which influences the estimation of inbreeding), existence of subpopulations, or inbreeding practices, could explain why we did not observe such a correlation. Furthermore, a negative correlation could be expected between the current average values of inbreeding (F) and expected heterozygosity (H_e) across breed, which was not observed. If each population had the same initial values of F and H_e, then the current values of F and H_e should therefore be negatively correlated. Thus, different factors related either to initial conditions or to the breed management may influence the value of the correlation between F and H_e. In particular, the breeds with the highest current average inbreeding could be, by chance, the breeds with the highest initial heterozygosity, so that their current value of H_e remains higher than those of breeds with lower inbreeding. Unfortunately, because of lack of biological samples from the ancient generations, such a hypothesis cannot be checked.

One may also suggest that heterozygote animals were favoured by selection (Falconer & Mackay 1996). In some breeds like the BAR breed, despite a high inbreeding coefficient underestimated by a low pedigree knowledge, heterozygosity was found to be very high (0.70). This is probably linked to the influence of some recent crosses (Leroy *et al.* 2006). Thus, genealogical estimators may show the effect of recent events, whilst molecular data report on the cumulated effects of past selection, migration and drift that have taken place in the population. Discrepancies between genealogical parameters and molecular estimators of genetic diversity may justify a closer look at a breed's history and management. Furthermore, such discrepancies may be also due to pedigree errors, as suggested by Slate *et al.* (2004).

The positive correlation found between F/Φ ratio and F_{IS} was expected. Björnerfeldt *et al.* (2008) showed that the existence of more or less closed varieties in the POO breed was related to high F_{IS}-values and a significant heterozygote deficit (Walhund effect). When they exist, such varieties represent genetic isolates that may be quite inbred. It is therefore not surprising that in our data set, the POO and CSP breeds, which are known to have closed varieties (based on size and/or colours), had both high F/Φ ratio and F_{IS}-value. Furthermore, we cannot exclude the possibility that in some breeds, inbreeding practices could lead to heterozygote deficit, negative F_{IS}-values and high F/Φ ratio (Moureaux *et al.* 1996).

While our results showed some inconsistency at the breed level, links between genealogical and molecular parameters were shown at the individual level by the correlations between genealogical and molecular kinships. The estimators R_{QG} (Queller & Goodnight 1989) and R_{LR} (Lynch & Ritland 1999) exhibited significantly higher correlations than other estimators, which can be explained by the fact that both estimators take into account the allele frequencies at the breed level. By using two pig breeds and 10 different molecular estimators, Toro *et al.* (2002) found that within-breed correlations ranged from 0.19 to 0.69, which is of the same order of magnitude as our results. Moreover, for the sampling of the 24 breeds, mean Φ was quite low (2.79 on average): as correlations depend on relatedness composition (Csillery *et al.* 2006), correlations could probably be higher when variances of Φ increased (Slate *et al.* 2004). Within our breeds, large variances of Φ were linked with large mean Φ (Table S2). This explains why the two breeds (BRP and BSG) with the highest mean Φ (respectively 5.7 and 8.8) always had correlations between Φ and molecular coefficients larger than 0.55.

Fernandez *et al.* (2005) wrote that marker information (except if the number of markers is very high) is not a particularly useful tool to establish conservation strategies that should be based on minimization of global coancestries. Instead, these researchers suggested the favouring of pedigree information where possible. Bijma (2000) discovered that, on the basis of pedigree data, acceptable values of ΔF per generation should be lower than a value between 0.5 and 1% in order to limit the extent of inbreeding depression. This means that the realized effective size N_{er} should not be lower than 50. As N_{er} and f_a values were found to be particularly low in the BRP and BSG breeds, breeding decisions should take into account the genetic diversity, even if heterozygosity values are not particularly low in these breeds.

In this study, 10 breeds had a number of equivalent generations lower than 4, which is due to two main reasons: either a short history of the French recognized population, or a large number of imported dogs. In such cases,

molecular approaches provide useful information for the evaluation of genetic variability. In the BUT breed, the extremely low heterozygosity rate (0.37), may be because of a bottleneck occurring in the 20th century (Irion *et al.* 2003). This should be taken into account by the breeders: contributions of reproducing animals should be optimized (Toro *et al.* 2008) by maximizing their coancestry, and more simply by maximizing their number.

To conclude, as genotyping with a very large number of markers is becoming less and less expensive, the molecular approach is expected to become an increasingly useful tool to manage within-population diversity, with better correlations between genealogical and molecular estimators of diversity being obtained.

Acknowledgements

The authors thank the Société Centrale Canine and Association Nationale de la Recherche Technique for both financial and logistic support, Prof. Stéphane Robin for useful discussions about statistical analyses, breeders and CERCA for samples provided, Wendy Brand-Williams for linguistic revision and Michèle Tixier-Boichard for useful comments.

References

Belkhir K., Borsa P., Chikhi L., Raufaste N. & Bonhomme F. (1996) GENETIX, *logiciel sous WindowsTM pour la génétique des populations*. Laboratoire Génome, Populations, Interactions CNRS UMR 5000, Université de Montpellier II, Montpellier.

Belkhir K., Castric V. & Bonhomme F. (2002) IDENTIX, a software to test for relatedness in a population using permutation methods. *Molecular Ecology Notes* **2,** 611–4.

Bijma P. (2000) Long-term genetic contributions: prediction of rates of inbreeding and genetic gain in selected populations. PhD Thesis, Department of Animal Sciences, Wageningen University, Wageningen, p. 225.

Björnerfeldt S., Hailer F., Nord M. & Vila C. (2008) Assortative mating and fragmentation within dog breeds. *BMC Evolutionary Biology* **8(28)**. Doi:10.1186/1471-2148-8-28.

Boichard D. (2002) PEDIG: a Fortran package for pedigree analysis suited for large populations. In: *7th World Congress of Genetics Applied to Livestock Production* (Ed. by WCGALP Organizing Committee), CD-ROM no. 28-13. Montpellier, France.

Boichard D., Maignel L. & Verrier E. (1997) Value of using probabilities of gene origin to measure genetic variability in a population. *Genetics Selection Evolution* **29,** 5–23.

Caballero A. & Toro M.A. (2000) Interrelations between effective population size and other pedigree tools for the management of conserved populations. *Genetical Research* **75,** 331–43.

Chakraborty R. & Jin L. (1993) Determination of relatedness between individuals by DNA fingerprinting. *Human Biology* **65,** 875–95.

Csillery K., Johnson T., Beraldi D., Clutton-Brock T., Coltman D., Hansson B., Spong G. & Pemberton J.M. (2006) Performance of marker-based relatedness estimators in natural populations of outbred vertebrates. *Genetics* **173,** 2091–101.

Eding H. & Meuwissen T. (2001) Marker-based estimates of between and within population kinships for the conservation of genetic diversity. *Journal of Animal Breeding and Genetics* **118,** 141–59.

El Mousadik A. & Petit R.J. (1996) High level of genetic differentiation for allelic richness among populations argan tree [*Argania spinosa* (L. Skeels)] endemic to Morocco. *Theoretical and Applied Genetics* **92,** 832–6.

Falconer D.S. & Mackay T.F. (1996) *Introduction to Quantitative Genetics*, 4th edn. Longman, Harlow, Essex.

Fernandez J., Villanueva B., Pong-Wong R. & Toro M.A. (2005) Efficiency of the use of pedigree and molecular marker information in conservation programs. *Genetics* **170,** 1313–21.

Foulley J.L. & Ollivier L. (2006) Estimating allelic richness and its diversity. *Livestock Science* **101,** 150–8.

Głażewska I. (2007) Genetic diversity in Polish hounds estimated by pedigree analysis. *Livestock Science* **113,** 296–301.

Goudet J. (2001) FSTAT, a Program to Estimate and Test Gene Diversities and Fixation Indices (Version 2.9.3). Available at: http://www2.unil.ch/popgen/softwares/fstat.htm (last accessed on 23 December 2008).

Gutiérrez J.P., Royo L.J., Alvarez I. & Goyache F. (2005) MOLKIN v2.0: a computer program for genetic analysis of populations using molecular coancestry information. *Journal of Heredity* **96,** 718–21.

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.

Jones P., Chase K., Martin A., Ostrander E.A. & Lark K.G. (2008) SNP based association mapping of dog stereotypes. *Genetics* **179,** 1033–44.

Leroy G., Rognon X., Varlet A., Joffrin C. & Verrier E. (2006) Genetic variability in French dog breeds assessed by pedigree data. *Journal of Animal Breeding and Genetics* **123,** 1–9.

Leroy G., Verrier E., Wisner-Bourgeois C. & Rognon X. (2007) Breeding goals and breeding practices of French dog breeders: results from a large survey. *Revue de Médecine Vétérinaire* **158,** 496–503.

Lüpke L. & Distl O. (2005) Microsatellite marker analysis of the genetic variability in Hanoverian Hounds. *Journal of Animal Breeding and Genetics* **122,** 131–9.

Lynch M. & Ritland K. (1999) Estimation of pairwise relatedness with molecular markers. *Genetics* **152,** 1753–66.

Mäki K., Groen A.F., Liinamo A.E. & Ojala M. (2001) Population structure, inbreeding trend and their association with hip and elbow dysplasia in dogs. *Animal Science* **73,** 217–28.

Malécot G. (1948) *Les mathématiques de l'hérédité*. Masson, Paris.

Mathieu E., Autem M., Roux M. & Bonhomme F. (1990) Epreuves de validation dans l'analyse de structures génétiques multivariées: comment tester l'équilibre panmictique? *Revue de Statistique Appliquée* **38,** 47–66.

Moureaux S., Verrier E., Ricard A. & Mériaux J.C. (1996) Genetic variability within French race and riding horse breeds from genealogical data and blood marker polymorphism. *Genetics Selection Evolution* **28,** 83–102.

Nei M., Tajima F. & Tateno Y. (1983) Accuracy of estimated phylogenetic trees from molecular data. *Journal of Molecular Evolution* **19,** 153–70.

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

Patterson D.F. (1993) Understanding and controlling inherited diseases in dog and cat. *Tidjschrift voor Diergeneeskunde* **118,** 23–7.

Queller D.C. & Goodnight K.F. (1989) Estimating relatedness using molecular markers. *Evolution* **43,** 258–74.

Raymond M. & Rousset F. (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86,** 248–9.

SAS Institute Inc. (2004) *SAS 2004: Version 9.1.2*. Cary, NC, USA.

Slate J., David P., Dodds K.G., Veenvliet B.A., Glass B.C., Broad T.E. & McEwan J.C. (2004) Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: theoretical expectations and empirical data. *Heredity* **93,** 255–65.

Toro M.A., Barragán C., Óvilo C., Rodriganez J., Rodriguez C. & Silio L. (2002) Estimation of coancestry in Iberian pigs using molecular markers. *Conservation Genetics* **3,** 309–20.

Toro M.A., Fernandez J. & Caballero A. (2008) Molecular characterization of breeds and its use in conservation. *Livestock Science* in press. Doi:10.1016/j.livsci.2008.07.003.

Supporting information

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

Table S1 Size range, number of alleles, polymorphism information content (PIC) and observed heterozygosities for the 21 loci studied.

Table S2 Correlation between genealogical kinship and different estimators of molecular similarity or distance, computed on pairs of individuals of 24 breeds.

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