npg

ORIGINAL ARTICLE

Toward a genome-wide approach for detecting hybrids: informative SNPs to detect introgression between domestic cats and European wildcats (*Felis silvestris*)

R Oliveira[1,2], E Randi[3,4], F Mattucci[3], JD Kurushima[5], LA Lyons[5,6] and PC Alves[1,2,7]

Endemic gene pools have been severely endangered by human-mediated hybridization, which is posing new challenges in the conservation of several vertebrate species. The endangered European wildcat is an example of this problem, as several natural populations are suffering introgression of genes from the domestic cat. The implementation of molecular methods for detecting hybridization is crucial for supporting appropriate conservation programs on the wildcat. In this study, genetic variation at 158 single-nucleotide polymorphisms (SNPs) was analyzed in 139 domestic cats, 130 putative European wildcats and 5 captive-bred hybrids ($N = 274$). These SNPs were variable both in wild ($H_E = 0.107$) and domestic cats ($H_E = 0.340$). Although we did not find any SNP that was private in any population, 22 SNPs were monomorphic in wildcats and pairwise F_{CT} values revealed marked differences between domestic and wildcats, with the most divergent 35 loci providing an average $F_{CT} > 0.74$. The power of all the loci to accurately identify admixture events and discriminate the different hybrid categories was evaluated. Results from simulated and real genotypes show that the 158 SNPs provide successful estimates of admixture, with 100% hybrid individuals (two to three generations in the past) being correctly identified in STRUCTURE and over 92% using the NEWHYBRIDS' algorithm. None of the unclassified cats were wrongly allocated to another hybrid class. Thirty-five SNPs, showing the highest F_{CT} values, provided the most parsimonious panel for robust inferences of parental and first generations of admixed ancestries. This approach may be used to further reconstruct the evolution of wildcat populations and, hopefully, to develop sound conservation guidelines for its legal protection in Europe.
Heredity (2015) **115**, 195–205; doi:10.1038/hdy.2015.25; published online 24 June 2015

INTRODUCTION

Hybridization creates exceptional challenges in conservation biology. It is known that hybridization can have an important role in the evolution of many organisms; however, it also can produce irreversible damages, namely the extinction of species (Allendorf *et al.*, 2001). Usually these two contrasting effects are associated with human indirect or direct interference on natural processes. One of the main consequences of hybridization is the introgression of genes from one species to another, which can result in the extinction of native gene pools. The amount of introgressive hybridization can be exceptionally increased when hybridization is mediated by humans, namely when contacts between wild and domesticated counterparts are promoted because of habitat loss of the wild species.

The current situation of European wildcat (*Felis silvestris silvestris*) is a remarkable example of the consequences of anthropogenic hybridization. The survival and conservation of indigenous populations of the European wildcat might be locally threatened by introgressive hybridization with feral domestic cats (*Felis silvestris catus*). Over the last decade, the genotyping of highly polymorphic molecular markers

(specifically microsatellites, short tandem repeats) and partial mitochondrial DNA sequences, combined with new Bayesian statistical tools, have radically improved the knowledge on wildcat population genetics and admixture with the domestic cat (for example, Beaumont *et al.*, 2001; Randi *et al.*, 2001; Pierpaoli *et al.*, 2003; Kitchener *et al.*, 2005; Lecis *et al.*, 2006; Oliveira *et al.*, 2008a, b; O'Brien *et al.*, 2009; Hertwig *et al.*, 2009; Eckert *et al.*, 2010; Mattucci *et al.*, 2013). Wildcats have been domesticated from African wildcat (*F.s. libyca*) ancestors ~ 10 600 years ago (Vigne *et al.*, 2012), and since then wild and domesticated forms have remained fully interfertile (Robinson, 1977; Ragni and Possenti, 1996). Hybridization between wildcat subspecies is thought to have initiated when feral domestic cats started their expansion across the range of wildcats (Driscoll *et al.*, 2009), thus occurring perhaps for several thousands of years in some regions. In some areas, where *taxa* boundaries are probably maintained and wildcat populations less fragmented, introgression may be minimal (for example, Pierpaoli *et al.*, 2003; Kitchener *et al.*, 2005; Lecis *et al.*, 2006). However, in particular historical or ecological conditions, widespread admixture might produce hybrid swarms, likely leading

[1]CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, InBIO—Laboratório Associado, Vairão, Portugal; [2]Departamento de Biologia, Faculdade de Ciências da Universidade do Porto, Porto, Portugal; [3]Laboratorio di Genetica, Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA), Bologna, Italy; [4]Department 18/Section of Environmental Engineering, Aalborg University, Aalborg, Denmark; [5]Department of Population Health and Reproduction, School of Veterinary Medicine, University of California—Davis, Davis, CA, USA; [6]Department of Veterinary Medicine and Surgery, College of Veterinary Medicine, University of Missouri—Columbia, Columbia, MI, USA and [7]Wildlife Biology Program, University of Montana, Missoula, MT, USA
Correspondence: Professor PC Alves, CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, Campus Agrário de Vairão, Vairão 4485-661, Portugal.
E-mail: pcalves@fc.up.pt
Received 12 January 2014; revised 13 December 2014; accepted 3 February 2015; published online 24 June 2015

to the genetic extinction of the wildcat parental populations (Allendorf et al., 2001; Beaumont et al., 2001; Brumfield, 2010; Fitzpatrick et al., 2010). European wildcats have apparently experienced both extremes. Wildcats in Scotland and Hungary show widespread hybridization and deep genetic introgression with domestic cats (Beaumont et al., 2001; Lecis et al., 2006), whereas only sporadic hybridization or no detectable introgression have been observed in Italy, Iberia and northeast France (Pierpaoli et al., 2003; Lecis et al., 2006; Oliveira et al., 2008a, b; O'Brien et al., 2009).

Although wildcat and domestic cat hybridization has been addressed in several studies, detecting hybrids and introgressed individuals, and understanding the causes limiting or favoring introgression, are still complex and controversial issues affecting wildcat research and conservation. The recent and intricate domestication of cats may implicate an overall shallow differentiation between the domestic and the wildcat counterpart subspecies, and thus the detection of hybridization between these forms is expected to be demanding. Combinations of markers, such as microsatellites and mitochondrial DNA (mtDNA; Driscoll et al., 2011), have improved hybrid detection; however, resolution remains limited, namely when involves several backcrosses. The development of a larger suite of molecular tools, applicable in invasive and noninvasive samples, is essential to increase the power of admixture analysis, which is mandatory for the adequate conservation planning of European wildcat populations.

High-throughput technologies improved genomic resources, such as single-nucleotide polymorphism (SNP) arrays and sequence assemblies, and have enabled the genome-wide genotyping of several species, namely the domesticated and their wild relatives (for example, wolf, vonHoldt et al., 2010; bison, Pertoldi et al., 2010; and bighorn sheep, Poissant et al., 2010). The European wildcat is an example of such a 'genome-enabled' taxon (Kohn et al., 2006), benefitting from the cross-species applicability of domestic cat data. Specifically, the recent sequencing of the domestic cat genome (Pontius et al., 2007; Mullikin et al., 2010; Montague et al., 2014; Tamazian et al., 2014), which has included SNP discovery in the African wildcat subspecies (Felis silvestris cafra), provides useful reference data for the discovery of new nuclear markers for assessing the introgression of domestic cat genes in the wild counterparts. Nussberger et al. (2013) recently described a set of 48 nuclear SNPs for identifying European wildcats, domestic cats and their admixed progeny. However, this work used a limited number of SNPs and reference samples solely from Switzerland.

Here we examined the power of anonymous SNPs in the domestic cat, to estimate the depth of introgression in conspecific wildcats sampled from several European populations. We identified the minimum number of highly divergent SNPs needed for accurate admixture analyses, hybrid identification and individual assignment to the wild or domestic parental populations. We expect that the proposed panel of SNPs, in combination with the existing ones in the recent literature, will provide an easier, standardized, cheaper and more accurate methodology to assess hybridization between domestic and the European wildcat. Accurate estimates of introgression and level of hybridization are crucial for prioritizing conservation efforts for European wildcat populations.

MATERIALS AND METHODS
European wildcats and domestic cats
Morphologically identified wildcats (Schauenberg, 1969, 1970; Ragni and Possenti, 1996) were selected from ISPRA (Istituto Superiore per la Protezione e la Ricerca Ambientale) and CIBIO/UP (Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto) tissue bank

Table 1 Information and number (*N*) on the cats used for SNP analyses of introgression

Location	N	No. of FSI	Pop
Putative European wildcats (FSI)	130	82	
Belgium	4	4	4
Bosnia	1	0	7
Bulgaria	5	5	9
Germany	10	9	5
Hungary	11	0	8
Italy	41	34	6
Luxembourg	2	2	4
Portugal	19	11	1
Scotland	16	0	3
Slovenia	7	7	7
Spain	14	10	2
Random-bred cats (FCA)	139		
Germany	29		
Italy	29		
Turkey	51		
Cyprus	30		
Known hybrids (HYB)—Italy	5		10
All cats	274		

Abbreviation: SNP, single-nucleotide polymorphism.
Putative European wildcats (FSI), random-bred domestic cats (FCA) and known wild×domestic cat hybrids (HYB). *Pop* is the population designation used in the analyses (codes used in Figure 1). The number of wildcats (No. FSI) indicates the number of cats considered to be best representatives of non-introgressed European wildcats in our sampling.
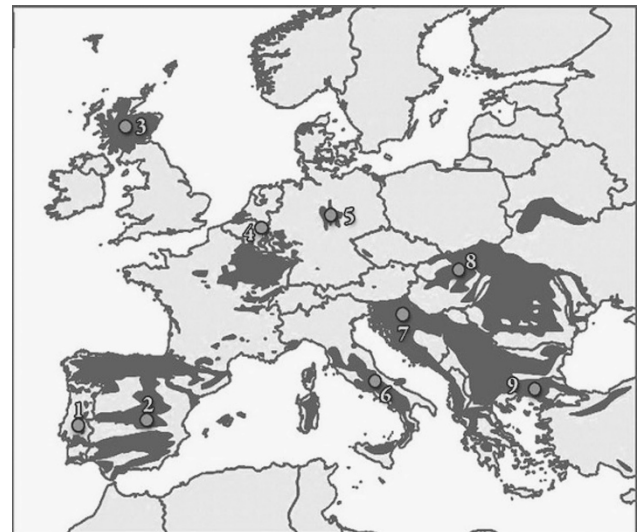

Figure 1 Sampling locations of putative European wildcats. Shaded areas correspond to the approximate current distribution of *F. silvestris* in Europe (adapted from Grabe and Worel, 2001).

collections (F. s. silvestris, *n* = 130) taking in consideration the natural distribution of European wildcat and the fragmentation of its populations in Europe (Pierpaoli et al., 2003). Sampling was performed across diverse European geographic localities by randomly selecting a few available samples from each location (Table 1; Figure 1). Five known wildcats by domestic cat hybrids obtained in captivity were included in the analyses (Pierpaoli et al., 2003). Domestic cats (F. s. catus, *n* = 139) living in regions sympatric to the wildcats but in urban areas were also genotyped (Table 1; Lipinski et al., 2008; Kurushima et al., 2012). DNA from tissue samples was extracted as described by

Pierpaoli et al. (2003). Buccal swabs from domestic cats from Cyprus were obtained from the Cyprus Malcolm Cat Sanctuary and were prepared as previously described (Kurushima et al., 2012). DNA from putative European wildcats and captive bred hybrids was whole-genome-amplified according to the manufacturer's recommendations using the REPLI-g Midi Kit (Qiagen Inc, Hilden, Germany).

SNP genotyping

A total of 158 SNPs from 18 cat autosomes ($n = 154$) and X chromosome ($n = 4$) were selected and used to genotype all cats. Most of the SNPs (134) were randomly selected from Kurushima et al. (2012), including 48 in intragenic regions, and were opportunistically applied to this study. Two phenotypic SNPs, associated with cat coat colors, TYR (Siamese—points) and TYRP1 (brown), were also examined (Lyons et al., 2005a, b). Additional 22 SNPs in intragenic regions were genotyped, including the following: (i) nine that revealed at least one polymorphic position between European wildcat and domestic cats (Johnson et al., 2006); (ii) one for which high variability was known among domestic cats (CCR2; Esteves et al., 2007); (iii) one considered to be leopard cat (CAT; Prionailurus bengalensis) species-specific; (iv) four on the X chromosome; and (v) seven randomly selected intragenic SNPs (Supplementary Table 1). The later seven randomly intragenic SNPs were selected at the same time and with the same criteria as Kurushima et al. (2012). Briefly, SNPs were previously selected from the $1.9 \times$ coverage of the Abyssinian cat (Pontius et al., 2007), being heterozygous in the Abyssinian, within nonrepetitive regions and dispersed across the chromosomes. Only the SNPs that had strong Phred-like scores (>25) and were ~ 8 Mb apart in the genome assembly were chosen. Further selection of SNPs included proper and robust design for the GoldenGate assay using Illumina design tools (Gentrain score >0.55). Finally, each SNP had a call rate $>80\%$ and a minor allele frequency $>5\%$ across the entire data set, which is considered the inherent error rate in the assay efficiency.

Golden Gate Assay amplification and BeadXpress reads were performed following the manufacturer's protocol (Illumina Inc, San Diego, CA, USA) on 50–500 ng of DNA or whole-genome-amplified product (Kurushima et al., 2012). The BeadStudio software v. 3.1.3.0 with the Genotyping module v. 3.2.23 (Illumina Inc.) was used to analyze the data. This software provides an automated genotype calling and powerful quality-control features to assess reproducibility and Mendelian consistency.

Statistical analysis

Summary statistics were used to describe levels of genetic variability and differentiation on the wild and domestic subspecies. Minor allele frequency was calculated with FSTAT v. 2.9.3.2 (Goudet, 2001). To avoid any bias resulting from the inclusion of hybrid genotypes among the representatives of European wildcats, we performed comparative analyses with all domestic cats against the 82 putatively purest European wildcats (Table 1). All wildcats from Hungary and Scotland were excluded from these first analyses because of their high level of admixture proportions determined both from morphological presumption and genetical inference (Beaumont et al., 2001; Daniels et al., 2001; Pierpaoli et al., 2003; Lecis et al., 2006). The individuals excluded (48, see Table 1) were, afterwards, included to the data set for hybridization analyses. Significance of deviations from Hardy–Weinberg equilibrium, and the observed (H_O) and expected (H_E) heterozygosities (unbiased, Nei, 1978) were calculated for all locus–population combinations using Markov chain exact tests in ARLEQUIN 3.5.1.2 (Excoffier and Lischer, 2010), with a chain length of 100 000 and 3000 dememorization steps. FSTAT 2.9.3.2 was used to compute the Wilcoxon-signed rank test to evaluate differences in H_E between wild and domestic cats, accounting for differences in sample size (Goudet, 2001). Allelic richness (Ar) was computed for each group following a rarefaction method that compensates for uneven sample sizes, as implemented in the software HP-Rare 1.0 (Kalinowski, 2005). ARLEQUIN 3.5.1.2 (Excoffier and Lischer, 2010) was used to perform an analysis of molecular variance (AMOVA) of pairwise F_{CT} (wildcat versus domestic groups) for each polymorphic locus, testing the null hypothesis of no differentiation by permuting genotypes between populations (10 000 replicates; $P < 0.001$). AMOVA was also used for testing the H_E difference between the wild and domestic cat groups. Average values were calculated for autosomal SNPs alone.

The 158 SNPs were ranked for hybridization diagnostic value by computing (i) In (informativeness for assignment); (ii) Ia (informativeness for ancestry coefficients); and (iii) optimal rate of correct assignment, using INFOCALC (Rosenberg et al., 2003; Rosenberg, 2005). For each locus, average ranking values were determined. Moreover, the probability of identity was estimated with a correction for small sample size (P_{ID}unbiased; Paetkau et al., 1998) and the equivalent probability for a pair of siblings (P_{ID}sib; Waits et al., 2001) with GenAlEx 6.41 (Peakall and Smouse, 2006). These values were used to estimate the minimum number of loci required for describing unique individual genotypes.

Individual assignment and admixture analyses

To assign cats to populations and to test for admixture, 158 SNP genotypes from 274 cats were evaluated using two Bayesian clustering procedures. Assuming two main populations, European wildcat and domestic cats ($K = 2$), 10 independent runs of the Bayesian-based software STRUCTURE 2.3.3 (Pritchard et al., 2000; Falush et al., 2007; Hubisz et al., 2009) were computed. For each run, the average proportion of membership (Q) of the sampled populations, and the distribution of individual membership proportions (qi) to the two inferred clusters, with their 90% credibility intervals (CIs) were assessed. All computations were performed using the admixture model with correlated allele frequencies either without prior nongenetic information or considering the domestic cats as reference samples. Runs consisted of a burn-in of 105 cycles and 106 Markov Chain Monte Carlo iterations, and were averaged using CLUMPP version 1.1.1 (Jakobsson and Rosenberg, 2007) with the FullSearch algorithm and the G' pairwise matrix similarity statistics. Average assignments were plotted using DISTRUCT 1.1 (Rosenberg, 2004). The Bayesian model-based method implemented in the software NEWHY-BRIDS (Anderson and Thompson, 2002) was further applied to classify cats into discrete hybrid classes. NEWHYBRIDS estimates the posterior probability that individuals fall into each of six genotypic classes corresponding to hybrid categories (Hi): parental subspecies (domestic or wild), F1, F2 and the backcrosses. Uniform priors were chosen to downweight the influence of an allele that might be rare in one species and absent in the other. Ten independent runs were performed to test for stability.

The power of all SNPs to detect different hybrid classes was assessed by the analysis of the assignment accuracy obtained for simulated genotypes. One-hundred multilocus genotypes of each parental (wildcat × wildcat; domestic cat × domestic cat), F1 (wildcat × domestic cat), F2 (F1 × F1) and backcross (F1 × wildcat; F1 × domestic cat) categories were generated with the software HYBRIDLAB v1.0 (Nielsen et al., 2006, but see also Oliveira et al., 2008b) and, afterwards, analysed using STRUCTURE and NEWHYBRIDS under the same setting of the admixture analysis described above. Qi threshold values for all analyses where established by the minimum value for which all parental domestic cats could be correctly assigned. A complementary analysis was performed using a combined data set that include the simulated genotypes (600), plus the observed genotypes that displayed admixed genetic assignments or for which molecular assignments opposed their prior morphological identifications in the hybridization analyses of STRUCTURE and NEWHY-BRIDs. Analyses of all observed, simulated and both kind of genotypes prompted the elimination of 18 putative wildcat samples on all subsequent analyses, that were most likely included in the wildcat sampling group because of incorrect morphological identifications. Accordingly, the preliminary Bayesian inferences were re-run for the new data set of 256 cats, including 139 random-bred domestic cats, 112 putative European wildcats and five known hybrids. In addition, Bayesian analyses of simulated genotypes were performed for the best estimated minimum number of SNPs ($n = 35$), which accurately allow the evaluation of hybridization in individual cat samples ($n = 256$).

RESULTS
SNPs variability
The SNP genotype call rate was more than 80% per individual cat in all analysed cat samples ($n = 274$). Descriptive statistics are presented in Supplementary Tables 1 and 2. All SNPs were polymorphic among

domestic cats (minor allele frequency >5%, to ensure resolution with the domestics and avoid those unique for the Abyssinian). However, 22 SNPs (13.92%) were monomorphic among the wildcats, including the two phenotypic SNPs for *TYR* and *TYRP1*. Significant deviations from Hardy–Weinberg equilibrium, following Bonferroni correction ($P < 0.00016$), were detected in 16 SNP loci, eleven among the domestic population and five in the wildcat group. Although none of the 158 loci had alternative private alleles, a large proportion of SNP variability was significantly partitioned between wildcat and domestic cats (average $F_{CT} = 0.427$; AMOVA $P < 0.001$), with single-locus F_{CT} pairwise values ranging between 0 (ChrA3_159537633; ChrF2_78303221) and 0.891 (ChrE2_34027888). European wildcats proved to be significantly less variable than domestic cats, both at average values of expected heterozygosity ($H_{E(FCA)} = 0.340$;

$H_{E(FSI)} = 0.107$; $P < 0.001$) and Ar ($Ar_{(FCA)} = 1.738$; $Ar_{(FSI)} = 1.250$; $P < 0.001$). Exceptions to the lower wildcat's variability were found at 18 SNPs, for which wildcats exhibited higher H_E than domestic cats (Supplementary Tables 1 and 2). Ten SNPs had two to four times higher heterozygosity in wildcats, and five of the ten showed significant deviations from Hardy–Weinberg equilibrium.

The average informativeness scores of each locus (INFOCALC—Supplementary Table 2) revealed that SNPs with lowest values of H_E in both groups displayed the highest values of genetic differentiation and top rank numbers, as they represented high frequencies of the two possible alternate variants. For increasing SNP combinations based on the loci ranked list, $P_{(ID)unbiased}$ and $P_{(ID)sibling}$ at $P < 0.001$ were simultaneously obtained using 35 loci (Table 2). These 35 top-ranked SNPs had an average pairwise $F_{CT} = 0.74$ ($P < 0.001$).

Table 2 Genomic SNP panel of top ranked loci to detect European wildcat and domestic cat introgression

Chr	Position	Location	Ar		H_E		FIS		AMOVA	INFOCALC	
			FCA	FSI	FCA	FSI	FCA	FSI	pw-F_{CT}	Score	R
B4	255106	ADARB2	1.400	1.147	0.149	0.050	0.015	0.492	0.876	0.894	1
E2	34027888		1.274	1.275	0.098	0.097	−0.051	0.739	0.891	0.892	2
B1	158896635	FGF5	1.335	1.208	0.122	0.072	−0.066	-0.033	0.885	0.886	3
D2	91989307	TCF7L2	1.395	1.206	0.147	0.071	0.115	0.313	0.867	0.878	4
B4	149532846	TRIOBP	1.375	1.336	0.138	0.121	0.139	−0.061	0.847	0.855	5
C1	28702055	FAM77C	1.492	1.213	0.189	0.074	0.194	-0.034	0.832	0.850	6
A1	151648701	PDCH12	1.505	1.345	0.195	0.125	0.180	−0.066	0.800	0.825	7
E1	131587399	EIF4A3	1.495	1.365	0.191	0.133	0.193	−0.069	0.796	0.817	8
B1	10420438	ENPP6	1.472	1.404	0.180	0.149	0.139	−0.082	0.798	0.815	9
E2	38860686		1.672	1.239	0.279	0.084	0.442*	−0.039	0.750	0.789	10
A1	223501140		1.691	1.176	0.290	0.061	0.342	−0.026	0.757	0.787	11
F1	82716202		1.708	1.272	0.300	0.096	0.170	0.478	0.721	0.767	12
A2	201526186	DPP6	1.730	1.172	0.313	0.059	0.314	−0.025	0.732	0.764	13
D4	63622083	PALM2-AKAP2	1.648	1.352	0.266	0.128	0.203	0.128	0.734	0.763	14
C1	17428968	TTN	1.749	1.281	0.324	0.100	0.296	0.215	0.693	0.739	15
B1	176151181	KIT	1.803	1.239	0.359	0.084	−0.008	−0.039	0.669	0.723	16
E2	3147915	TNN13	1.814	1.187	0.367	0.064	0.049	0.385	0.668	0.719	17
A3	162208567	PLB1	1.809	1.148	0.363	0.051	0.191	−0.020	0.684	0.717	18
D1	18390852		1.730	1.447	0.313	0.168	0.192	-0.096	0.662	0.710	19
D1	15984279		1.856	1.140	0.398	0.048	0.222	0.492	0.648	0.696	20
D1	117527468	CD44	1.163	1.837	**0.056**	**0.381**	−0.026	0.351	0.747	0.695	21
B3	57141954		1.843	1.272	0.388	0.096	0.276	0.216	0.627	0.687	22
B3	77094074	LOC607552	1.275	1.847	**0.098**	**0.389**	0.436	0.077	0.696	0.678	23
E2	7580874	MYBPC	1.814	1.403	0.367	0.149	−0.018	0.511	0.609	0.672	24
F2	38395360		1.856	1.345	0.398	0.125	0.185	-0.066	0.594	0.655	25
E2	8422942	RPS11	1.373	1.861	**0.137**	**0.397**	0.140	0.391	0.667	0.653	26
A1	133621071		1.876	1.269	0.413	0.095	0.156	−0.046	0.592	0.653	27
F1	26100599	LAMC1	1.890	1.247	0.424	0.087	0.176	−0.041	0.583	0.651	28
A1	69424718	ABCC4	1.896	1.206	0.429	0.071	0.111	−0.032	0.589	0.646	29
A1	242150000	GHR	1.814	1.540	0.368	0.210	0.163	0.175	0.560	0.634	30
C1	52456776	C8B	1.918	1.202	0.448	0.070	0.112	−0.030	0.558	0.626	31
B1	202966562		1.911	1.304	0.442	0.109	0.216	−0.055	0.538	0.614	32
C2	106991233		1.875	1.487	0.413	0.186	0.136	0.153	0.526	0.608	33
D2	1020904	ACF	1.953	1.072	0.482	0.024	−0.011	−0.006	0.520	0.583	34
D1	116730000	CAT	1.952	1.108	0.481	0.036	0.144	−0.013	0.515	0.580	35
Average			1.670	1.320	0.296	0.123	0.161	0.141	0.74		

Abbreviations: AMOVA, analysis of molecular variance; Ar, allelic richness; H_E, expected heterozygosity; SNP, single-nucleotide polymorphism.
SNPs where wildcats displayed higher genetic diversity than domestic cats are shown in bold. Inbreeding coefficient (FIS); pairwise estimations of genetic differentiation between European wild and domestic cats (AMOVA pw-FCT); loci scores averaged across INFOCALC estimations (average) and resulting ranking values (rank). Gene acronyms indicate location of intragenic SNPs in the domestic cat genome. Intragenic SNPs are not known to be under any type of selection and are not associated with phenotypes. *Significant deviations from HWE at $P < 0.05$, Bonferroni corrected.

Detection of hybridization

Bayesian analyses with and without prior information for domestic samples yielded globally identical results (data not shown). Hence, all the presented results were performed without prior nongenetic information. Assuming two major populations in STRUCTURE ($K=2$), all domestic cats were clearly assigned to their expected cluster according to genetic variation at the 158 SNPs (Figure 2). However, as noted in the Materials and methods, 18 putative European wildcats showed qi values to the domestic cluster above 0.92 and very narrow CI ranges (0.745–1.00): seven from Portugal, four from Spain, four from Italy, one from Scotland and two from Hungary (Table 3; Figures 2 and 3). No sign of subdivision was detected in the studied wildcat populations (STRUCTURE from $K=1$ to 15, data not shown).

According to the Bayesian analyses of the SNP variability, 23 putative wildcats show genetic evidence of admixed ancestry both in STRUCTURE and NEWHYBRIDS computations (Table 3). The only exception was one cat from Scotland (ID 101), which was identified as a possible hybrid in STRUCTURE ($q_{FSI}=0.768$) and as a European wildcat in NEWHYBRIDS ($q_{FSI}=0.970$). Most of the 23 admixed cats found in the random sampling belong to Scotland ($n=7$) and Hungary ($n=8$). The other putative hybrid cats were recognized in Portugal ($n=1$), Germany ($n=1$), Italy ($n=5$), and Bosnia and Herzegovina ($n=1$). Known captive-bred hybrids clearly displayed signatures of admixture, with individual qi ranging from 0.289 to 0.734 in the wild genetic group (Table 3). Moreover, they were mostly assigned to their known hybrid category with high posterior probabilities ($qi>0.90$): ID 57 as F1, ID 60 as BxFSI, ID 61 as BxFCA and ID 63 as BxFSI (Table 3; Figure 3b). Overall, 89 of 130 (68.46%) putative wildcats are inferred to have no introgression with domestic cats. This number excludes the five known hybrids but includes Scottish cat 101 as possibly admixed.

SNP simulations for admixture analysis

The analyses, both in STRUCTURE and NEWHYBRIDS, of the combined data set including all 256 observed with the 600 simulated genotypes (using the HYBRIDLAB), globally revealed the same presumed misclassifications as obtained with real genotypes alone. A summary of the misclassifications expected in six simulated hybridization categories is presented in Table 4. Bayesian analyses of the simulated genotypes revealed that all parental, F1, F2 and backcrossed individuals could be correctly identified by the STRUCTURE algorithm using the 158 SNPs. Moreover, posterior probabilities of assignment to the different simulated categories of hybridization proved to be sufficiently discriminatory because as few as 1% F2, 28% BxFCA and 14% BxFSI of the properly assigned genotypes displayed CI values outside the expected range (Table 4). Assignment values for NEWHYBRIDS proved to be equally accurate for parental and first-generation hybrids; however, 4% F2, 3% BxFCA and 1% BxFSI were allocated to their own hybrid category with qi values lower than 0.85 (Table 4). Nevertheless, none of the referred cats were significantly ($qi>0.85$) allocated to one of the other remaining hybrid categories, preventing any case of misclassification. The simultaneous analysis of simulated and true genotypes confirmed results observed for real data alone (Table 4), both for what regards the probable misclassification of cats according to morphology and the detection of hybrids (Figure 3).

The performance for detecting hybridization of the 35 top-ranked SNPs (Table 2) was evaluated by simulations on the modified data set using STRUCTURE and NEWHYBRIDS (Table 5). The aim of this analysis was to see whether good discrimination could be obtained more economically by using only strongly differentiated SNPs. The high level of genetic differentiation in these SNPs allowed an overall clear distinction of simulated parental and hybrid genotypes, as most individuals were assigned to their expected cluster with high posterior probabilities ($qi>0.80$) using the reduced set of SNPs. STRUCTURE's misassignments were exclusively obtained for lower percentages of admixture, namely 8% of the simulated BxFCA and 4% of the simulated BxFSI. In only a few of the parental and first-generation hybrids, the ranges of the CI are outside the expected values but unsurprisingly less discriminatory CIs were noticed for backcrosses. For example, only 38% of BxFCA and 42% of BxFSI had CI ranges that never overlapped parental genotypes (Table 5). NEWHYBRIDS' clustering proved also to be highly efficient, with all parental, 98% F1, 90% F2, 90% BxFCA and 96% BxFSI being correctly allocated to their category with high posterior probabilities (Table 5). Only one of the unclassified genotypes (one BxFCA) would be incorrectly assigned to its correspondent parental group, with all of the other cases representing broad partitions among hybrid classes. Exceptionally, two simulated BxFSIs were identified as F1 and F2, whereas two F2 were classified as F1 and BxFSI.
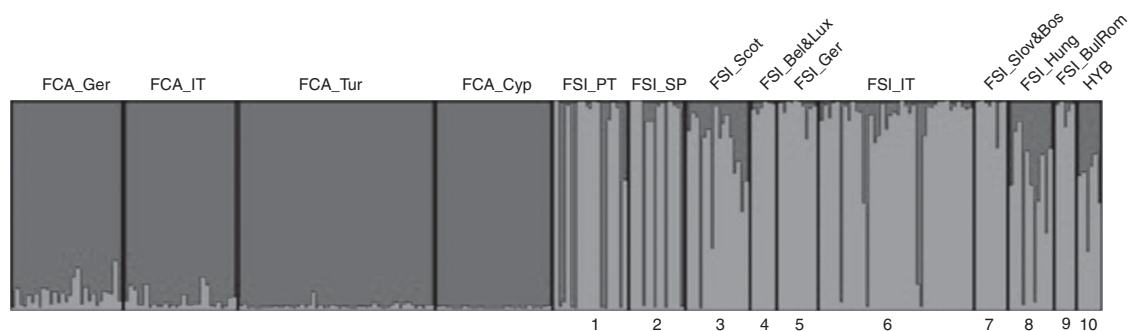

Figure 2 Average plot of the Bayesian admixture analyses performed in 10 independent STRUCTURE runs for $K=2$, using 158 SNPs on 139 known random-bred domestic cats (FCA) and 132 putative European wildcats (FSI). Each individual is represented by a single vertical bar divided into two genetic clusters, according to the proportion of their genome estimated to descend from each one of possible groups. Black vertical lines divide geographic groups of domestic and wild populations, which are labeled above the figure (Ger = Germany (5); IT = Italy (6); Tur = Turkey; Cyp = Cyprus; PT = Portugal (1); SP = Spain (2); Scot = Scotland (3); Bel&Lux = Belgium and Luxemburg (4); Slov&Bos = Slovenia and Bosnia (7); Hung = Hungary (8); BulRom = Bulgaria and Romania (9); HYB = Known hybrids (10)).

Table 3 Individual membership proportions (*qi*) of presumably misclassified and putatively admixed cats according to the Bayesian analyses performed in STRUCTURE and NEWHYBRIDS

ORIGIN	CAT	STRUCTURE *qi*				NEWHYBRIDS *qi*	
		FCA	CI	FSI	CI	Hi	
Portugal	711	0.957	(0.893–0.999)	0.043	(0.001–0.107)	FCA	1.000
	1024	0.935	(0.853–0.996)	0.065	(0.004–0.147)	FCA	1.000
	297	0.966	(0.908–0.999)	0.034	(0.001–0.092)	FCA	1.000
	298	0.984	(0.946–1.000)	0.016	(0.000–0.054)	FCA	1.000
	688	0.969	(0.908–1.000)	0.031	(0.000–0.092)	FCA	1.000
	689	0.976	(0.924–1.000)	0.024	(0.000–0.076)	FCA	1.000
	706	0.967	(0.906–1.000)	0.033	(0.000–0.094)	FCA	1.000
	712	0.383	(0.300–0.470)	0.617	(0.530–0.700)	F2/BxFSI	0.84/0.16
Spain	1027	0.941	(0.869–0.995)	0.059	(0.005–0.131)	FCA	1.000
	717	0.973	(0.921–1.000)	0.027	(0.000–0.079)	FCA	1.000
	728	0.976	(0.930–1.000)	0.024	(0.000–0.070)	FCA	1.000
	737	0.965	(0.905–1.000)	0.035	(0.000–0.095)	FCA	1.000
Scotland	101	0.232	(0.209–0.256)	0.768	(0.744–0.791)	FSI/BxFSI	0.97/0.03
	105	0.979	(0.929–1.000)	0.021	(0.000–0.071)	FCA	1.000
	106	0.189	(0.129–0.256)	0.811	(0.744–0.871)	BxFSI/FSI	0.60/0.40
	240	0.705	(0.616–0.791)	0.295	(0.209–0.384)	BxFCA/F2	0.96/0.04
	252	0.194	(0.124–0.273)	0.806	(0.727–0.876)	BxFSI/FSI	0.98/0.02
	268	0.352	(0.262–0.447)	0.648	(0.553–0.738)	F2/BxFSI	0.70/0.30
	269	0.300	(0.223–0.382)	0.700	(0.618–0.777)	BxFSI/F2	0.99/0.01
	272	0.532	(0.435–0.630)	0.468	(0.370–0.565)	F2	1.000
	273	0.392	(0.305–0.483)	0.608	(0.517–0.695)	F2/BxFSI	0.78/0.22
Germany	629	0.201	(0.132–0.277)	0.799	(0.723–0.868)	BxFSI/FSI	0.99/0.01
Italy	624	0.920	(0.839–0.990)	0.080	(0.010–0.161)	FCA	1.000
	580	0.487	(0.392–0.585)	0.513	(0.415–0.608)	F2	1.000
	671	0.963	(0.891–1.000)	0.037	(0.000–0.109)	FCA	1.000
	677	0.216	(0.148–0.292)	0.784	(0.708–0.852)	BxFSI	1.000
	678	0.182	(0.119–0.252)	0.818	(0.748–0.881)	BxFSI/FSI	0.69/0.31
	918	0.844	(0.745–0.937)	0.156	(0.063–0.255)	FCA	1.000
	992	0.491	(0.400-0.583)	0.509	(0.417-0.600)	F1/F2	0.91/0.09
	1006	0.953	(0.880–0.999)	0.047	(0.001–0.120)	FCA	1.000
	1009	0.178	(0.113–0.251)	0.822	(0.749–0.887)	BXFSI/FSI	0.89/0.11
Bosnia	1056	0.235	(0.166–0.312)	0.765	(0.688–0.834)	BxFSI	1.000
Hungary	211	0.413	(0.327–0.504)	0.587	(0.496–0.673)	BxFSI/F2	0.5/0.5
	214	0.160	(0.101–0.228)	0.840	(0.772–0.899)	BxFSI/FSI	0.69/0.31
	339	0.941	(0.847–0.999)	0.059	(0.001–0.153)	FCA	1.000
	352	0.257	(0.185–0.335)	0.743	(0.665–0.815)	BxFSI	1.000
	356	0.409	(0.325–0.499)	0.591	(0.501–0.675)	F2/BxFSI	0.97/0.03
	358	0.917	(0.832–0.990)	0.083	(0.010–0.168)	FCA	1.000
	361	0.479	(0.386–0.576)	0.521	(0.424–0.614)	F2	1.000
	613	0.265	(0.187–0.349)	0.735	(0.651–0.813)	BxFSI	1.000
	620	0.628	(0.528–0.728)	0.372	(0.272–0.472)	BxFCA/F2	0.61/0.39
	621	0.244	(0.174–0.320)	0.756	(0.680–0.826)	BxFSI	1.000
Known hybrids	57	0.400	(0.285–0.456)	0.600	(0.544–0.715)	F1/F2	0.91/0.09
	60	0.338	(0.257–0.423)	0.662	(0.577–0.743)	BxFSI/F2	0.93/0.7
	61	0.711	(0.616–0.804)	0.289	(0.196–0.384)	BxFCA	1.000
	62	0.321	(0.237–0.410)	0.679	(0.590–0.763)	F2/BxFSI	0.93/0.07
	63	0.266	(0.186–0.354)	0.734	(0.646–0.814)	BxFSI	1.000

Abbreviation: CI, credibility interval.
STRUCTURE *qi* values correspond to allocations with *K* = 2 to the domestic (FCA) and wild (FSI) inferred clusters, with their 90%CI. NEWHYBRIDS *qi* values reflect assignments to the six possible hybrid categories: domestic, European wild, F1, F2 and backcrosses. Presumed misclassified wildcats (18) that were eliminated from further analyses are shaded in light grey.
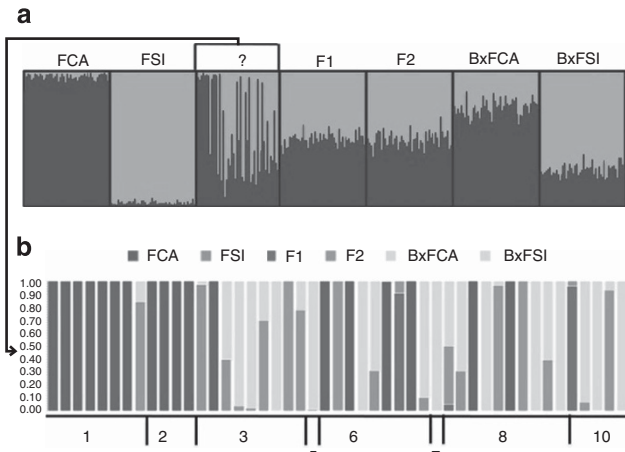
a

FCA | FSI | ? | F1 | F2 | BxFCA | BxFSI

b

■ FCA ■ FSI ■ F1 ■ F2 ■ BxFCA ■ BxFSI

Figure 3 Individual membership (q_i) values obtained using 158 SNPs under Bayesian-model computations. (**a**) STRUCTURE's plot of 100 simulated domestic (FCA), wild (FSI), F1, F2 and backcross (BxFCA; BxFSI) genotypes and 47 real individuals for which genetic data refute their straightforward allocation to the European wildcat subspecies (?); (**b**) NEWHYBRID's assignment of the same 47 dubious individuals to the different hybrid categories. Each individual is represented by a single vertical bar coloured according to the proportion of their genome descending from each of the inferred clusters (**a**) or hybrid class (**b**).

Table 4 Average membership proportion (*Q*) of simulated genotypes in the Bayesian analysis performed using STRUCTURE and NEWHYBRIDS

Simulated categories	Structure		NEWHYBRIDS
	FCA	FSI	
FCA	0.96	0.04	0.9996
$q_i > 0.85$	(0.77–1.00)	(0.00–0.23)	
WRONG	0 (4%)		0
FSI	0.02	0.98	0.9938
$q_i < 0.15$	(0.00–0.11)	(0.89–1.00)	
WRONG	0 (0%)		0
F1	0.50	0.50	0.9615
$0.4 < q_i > 0.6$	(0.30–0.70)	(0.30–0.70)	
WRONG	0 (0%)		0
F2	0.49	0.51	0.9050
$0.4 < q_i > 0.6$	(0.28–0.76)	(0.24–0.72)	
WRONG	0 (1%)		4%
BxFCA	0.73	0.27	0.9455
$0.15 < q_i > 0.85$	(0.51–0.92)	(0.08–0.49)	
WRONG	0 (28%)		3%
BxFSI	0.270	0.730	0.9455
$0.15 < q_i > 0.85$	(0.13–0.47)	(0.53–0.87)	
WRONG	0 (14%)		1%

Abbreviation: CI, credibility interval.
Results represent q_i values averaged over 10 independent runs. Minimum and maximum values of credibility intervals obtained in STRUCTURE are shown between brackets. The number of misclassified cats (WRONG), and the percentage of individuals for which CI ranges fallen outside the expected values are shown. NEWHYBRIDS' average posterior probability assignment to the correct category and percentages of unclassified simulated genotypes are indicated.

Table 5 Power to detect wildcat–domestic cat hybrids with 35 SNPs

Simulated categories	Structure		NEWHYBRIDS
	FCA	FSI	
FCA	0.934	0.945	0.9928
$q_i > 0.80$	(0.710–1.00)	(0.759–1.000)	
WRONG	0 (10%)		0
FSI	0.045	0.054	0.9998
$q_i < 0.20$	(0.000–0.226)	(0.000–0.257)	
WRONG	0 (2%)		0
F1	0.492	0.501	0.9836
$0.4 < q_i > 0.60$	(0.270–0.790)	(0.273–0.750)	
WRONG	0 (6%)		2%
F2	0.493	0.490	0.8804
$0.4 < q_i > 0.60$	(0.210–0.790)	(0.185–0.828)	
WRONG	0 (12%)		10%
BxFCA	0.707	0.721	0.8696
$0.2 < q_i > 0.8$	(0.388–0.950)	(0.375–0.998)	
WRONG	8% (62%)		10%
BxFSI	0.285	0.278	0.9293
$0.20 < q_i > 0.80$	(0.077–0.607)	(0.030–0.584)	
WRONG	4% (58%)		4%

Abbreviation: SNP, single-nucleotide polymorphism.

DISCUSSION

Introgression of domestic cat genes is a significant concern for the conservation of European wildcat populations. Hybridization can be either a widespread or localized event in wildcat populations. Hence, more precise detection of introgression levels is essential to prioritize habitats for wildcat preservation and to design efficient conservation strategies. Previous studies clearly show that the development of more powerful tools is still critical to accurately identify parental and hybrid individuals of this species because of the high similarity in morphology and genomes of wild and domestic forms. Although microsatellites have been the dominant markers in wildcat genetic studies (for example, Beaumont *et al.*, 2001; Randi *et al.*, 2001; Pierpaoli *et al.*, 2003; Lecis *et al.*, 2006; Germain *et al.*, 2008; Eckert *et al.*, 2010; O'Brien *et al.*, 2009), and recently mtDNA diagnostic SNPs have been suggested (Driscoll *et al.*, 2011), the increasing availability and numerous advantages of nuclear SNPs make them an appealing alternative and/or a complement to maternal and paternal lineage markers.

SNPs have been attracting a growing interest in a wide range of evolutionary applications and are becoming efficient tools among wildlife conservation-oriented studies (Brumfield *et al.*, 2003; Morin *et al.*, 2004; Seddon *et al.*, 2005; Morin *et al.*, 2009). Offering less variability per locus than STRs, SNPs provide a substantial number of advantages, namely: (i) reduced propensity for homoplasy due to lower mutation rates; (ii) higher density and more uniform distribution in genomes; (iii) suitability for successful high-throughput genotyping and straightforward comparability and transportability across laboratories and detection protocols; and (iv) highly successful application in fragmented DNA samples, for example, noninvasive and historical DNA (see Brumfield *et al.*, 2003; Morin *et al.*, 2004; Garvin *et al.*, 2010 for reviews). Nonetheless, the successful application of genome-wide batteries of nuclear SNPs in studies of wild populations is still limited to a few cases such as wolf-like species for studying their evolutionary history (vonHoldt *et al.*, 2011), wild sheep for detecting population structure and linkage disequilibrium (Miller *et al.*, 2011) and wild Atlantic salmon for the differentiation of farmed and wild individuals (Karlsson *et al.*, 2011). Recently, Monzón *et al.* (2013) used species-diagnostic SNPs to quantify the relative contributions of parental populations and better understand the complex hybrid

ancestry of the northeastern coyote. Here, we provide an analysis of nuclear SNPs in wildcats from a broad European range for applications in European wildcat conservation. Our main motivation was to improve molecular tools for detecting and quantifying hybridization, and testing the smaller and most informative set for it's potential use in noninvasive genetic samples.

Population variability

Genetic diversity, including SNPs Ar and H_E, showed marked differences between European wildcats and domestic cats. The wildcats, which were sampled from a broad proportion of their distribution across Europe, showed significantly lower genetic diversity. Generally, genetic variability is expected to be lower in the domesticated forms relatively to their wild counterparts, because of bottlenecks caused by low numbers of founder individuals and restricted gene flow imposed by human constrains (Doebley *et al.*, 2006). However, the selected SNPs were ascertained from the $1.9 \times$ genome sequence of an Abyssinian domestic cat and are highly polymorphic across all breeds of cats (Kurushima *et al.*, 2012). Thus, these SNPs cannot represent the spectrum of variability presumably present in the studied European wildcat populations and likely suffer from ascertainment bias. Variable SNP loci detected in the European wildcat samples will probably represent widespread ancestral polymorphism, and chances to identify population-specific alleles will be limited. Yet, we cannot exclude that extant European wildcat populations, which probably underwent repeated cycles of demographic fluctuations due to Pleistocene climate changes (Mattucci *et al.*, 2013), and have suffered recent population declines and fragmentation because of anthropogenic pressures, are actually less variable than domestic cats. Moreover, results obtained using other molecular markers, for example, STRs, also suggest that domestic cat may have higher genetic diversity than wildcats (for example, Pierpaoli *et al.*, 2003; Lecis *et al.*, 2006; Oliveira *et al.*, 2008a, b).

The pre-screening of SNPs for inclusion on arrays for this project implies that SNPs are polymorphic in the species in which they were ascertained. This could be the reason for not detecting fixed alleles between the domestic and the wildcat representatives in this study, notwithstanding the 22 monomorphic SNPs observed in wildcats. Although only a very small subset of the species genome was analyzed, a similar result could be expected for larger number of SNPs. Among dogs and wolves, no fixed SNPs have been detected in a 48-K panel from the Affymetrix Canine Mapping SNP 2.0 array (vonHoldt *et al.*, 2012). Even so, when two populations are subjected to different selective pressures, some levels of selection are expected to cause divergence in different parts of their genomes. Native California tiger salamanders (*Ambystoma californiense*) provide an excellent example of the benefit of SNPs in uncovering patterns of admixture (Fitzpatrick *et al.*, 2010). These authors were able to determine that only 3 out of 68 studied markers spread rapidly into native genomes, whereas the other 65 showed little evidence of introgression beyond the region where introductions of non-native barred tiger salamanders (*A. tigrinum mavortium*) occur. By demonstrating substantial evidence of heterogeneity in introgression rates among loci, this work highlighted the potential problems faced by those studies that only use a few neutral markers to detect hybridization (Allendorf *et al.*, 2010).

Bayesian clustering

The Bayesian clustering of the 274 individuals (139 random-bred cats, 130 putative European wildcats and 5 known hybrids) immediately revealed higher discriminative power of genotypes over phenotypes in identification of wildcats. Eighteen putative wildcats were allocated

with high posterior probabilities to the domestic cluster and therefore were excluded from the analysis. This was in agreement with previous reports for the species (for example, Oliveira *et al.*, 2008a, b), suggesting that morphological identification of European wildcat and domestic cats might not be as straightforward as some authors advocate (Ragni and Possenti, 1996; Daniels *et al.*, 1998; Kitchener *et al.*, 2005; Puzachenko, 2002; Yamaguchi *et al.*, 2004a, b; Krüger *et al.*, 2009; Platz *et al.*, 2011). A variety of issues could lead to misclassification, including (i) dead animals might have been highly degraded at the time of collection and discrimination of obvious morphological characters might not be possible; (ii) cats belong to past generations of admixture and demarked diagnostic traits are no longer expressed; (iii) samples were noninvasively collected (for example, scats and hairs) and morphological discrimination was not possible; (iv) overlap of morphological features; and (v) conservation biologist and naturalist bias of their morphological evaluation toward the collection of wild specimens. The fact that most, if not all, backcrosses remained undetected under morphological evaluation further confirms the higher efficiency of genotypes over phenotypes to identify past generation hybrids. The set of markers defined in this study should effectively circumvent many cases of wrong pre-classification and identify the origin of most unknown samples.

SNP power for admixture analysis

Any ancestral inference must strike a balance between economical, technical and statistical concerns (Rosenberg *et al.*, 2003). Ideally, the identification of recently introgressed hybrids, such as F1, F2 and first backcrosses, could be achieved with a minimum number of loci if the allele frequencies at these loci are sufficiently differentiated between the populations (Vähä and Primmer, 2006). The remarkable resemblance between European wildcats and domestic cats, the intricate history of sympatry and introgression that most probably influenced both the domestication (Driscoll *et al.*, 2007) and the expansion of domestic populations worldwide, might have created one of the most complicated frameworks to genetically discriminate parental groups of wild and domestic relatives. In the context of wildcat's conservation, genomic resources should be used to select the most informative ancestral markers among the huge number of loci available in DNA variants. Limited panels of 48 or 96 informative SNPs would be enough to design efficient and affordable applications, especially in cases of noninvasive sampling, or when analyses are performed to solve practical problems, such as assignment of unknown samples to parental categories, rather than complex population/introgression inferences. The identification of highly informative SNP loci from larger panels has already been proposed as a powerful approach to identify wolf (*Canis lupus lupus*) × dog (*Canis lupus familiaris*) hybrids, 24 loci proven to be informative for assignment to recent hybrid classes (vonHoldt *et al.*, 2012). If allocations are not definitive, a subsequent analysis of 100 loci has been suggested (vonHoldt *et al.*, 2012). In humans, subsets of informative SNPs delineate genetic relationships at the individual, parentage and population levels, namely for detecting human geographic structure (Liu *et al.*, 2005; Lao *et al.*, 2006). Similar studies in other species have also been conducted, such as for European bison (*Bison bonasus*; Tokarska *et al.*, 2009), Atlantic salmon (*Salmo salar*; Glover *et al.*, 2010), red fox (*Vulpes vulpes*; Sacks and Louie, 2008) and chicken breeds (Gärke *et al.*, 2012). Recently, Nussberger *et al.* (2013) developed a diagnostic marker set containing 48 SNPs that allows the identification of wildcats, domestic cats, their hybrids and backcrosses, and have demonstrated their accurate genotyping in single hairs (Nussberger *et al.*, 2014). However, these authors used a restricted set of reference

samples, and the choice of highly differentiated traits/loci from a small panel of parental individuals has been considered to possibly overlook population differentiation (Brumfield et al., 2003; Schlötterer, 2004; Morin et al., 2009). This is a concern among European wildcat populations because the genetic partition of the populations is still poorly known, and central European wildcats might not be as fragmented as other regions (Mattucci et al., unpublished). Studying just a reduced panel of parental individuals from very narrow areas might, then, under-represent wildcat variability at least in that specific population and may overestimate the level of genetic differentiation between wild and domestic cats that truly exists there. Considering that the knowledge on wildcat's genetic partition in Europe is still growing, the most accurate methodology is looking at European wildcats as an entire population that needs to be genetically differentiated from domestic cats, and try to find the most ancient variants that distinguish both forms. This would most likely prevent the advent of new variants in the future when adding more samples to the analyses. Therefore, to obtain the most powerful genetic tool for the analysis of hybridization/introgression dynamics, a combination of wide geographical samples with different types of markers from the entire genome should be evaluated (Driscoll et al., 2011), which preferably should represent both neutral and non-neutral variations (for example, Teeter et al., 2008).

To provide a similarly efficient panel of diagnostic markers for wildcat hybridization, the SNPs were ranked according to their utility in discriminating between wildcat and domestic cats. As few as 35 of the most differentiating SNPs provided correct admixture evidence for 99% of the cases, with as little as 8% of BxFCA and 4–5% of BxFSI remaining unclassified in STRUCTURE-based inferences. Therefore, the statistical power achieved with the 35 loci-based Bayesian clustering suggests that one can confidently accept the partition of individuals as European wildcat, domestic or first-generation hybrid cats (F1 and F2) with high confidence, whereas more cautious interpretations should be made when outlining admixed individuals (backcrosses). Even so, an underestimation of admixture rates in true populations is not expected, as the only case of missing hybrid identification was observed for a single simulated BxFCA. Although the 35 SNPs revealed outstanding success in hybridization inferences, a complete definition of all admixed cats in the different hybrid categories was fully obtained only with the entire set of 158 SNPs, even though 20% of loci had $F_{CT} < 0.10$.

Detection of hybridization in natural populations using SNPs
The inclusion of five known hybrids provided further evidence of the high accuracy of the assignment tests performed with the entire set of 158 loci, as all were assigned to their correct hybrid category. These results corroborated those obtained by simulation. However, the expectation of 100%, 96%, 97% and 99% identification of F1, F2, BxFCA and BxFSI hybrids, respectively, using NEWHYBRIDS might decrease with genotyping data. The panel of 158 SNPs successfully detected a putative hybrid class for all but one of the admixed cats identified by the same panel (ID 211). However, seven of the hybrids have been assigned with qi values between 0.60 and 0.78. These results confirm the high accuracy levels predicted by simulation analyses, but slightly increase the doubts in precisely identifying true hybrid genotypes. Globally these findings suggest that, although simulating hybrid classes might be a useful and indicative strategy for selecting informative loci and estimate the power of hybridization analyses, the inferences of introgression in true populations of European wildcats may be better refined by the inclusion of real genotypes of known hybrid categories in Bayesian clustering models. Simulation cannot account for novel and low-frequency alleles that could be discovered with additional sampling, and might provide an incomplete reflection of the true assignment power of our marker panel. Ideally, each inference should include simulation genotypes and several known hybrid individuals from different geographical locations and hybrid categories.

The highly discriminating loci discovered in this study may bring new insights to the study of European wildcat populations, specifically a powerful and efficient tool to detect and quantify hybridization with domestic cats. Further genotyping of additional populations should help to validate the selected SNPs. In addition, the possible combination of the SNPs described in our study with the ones developed by Nussberger et al. (2013) can eventually maximize the hybrid detection. Nevertheless, the new throughput technologies under development for domestic cats will soon allow the evaluation of the entire genome of F. silvestris species, supporting the identification of more diagnostic loci and potentially indicating areas of the genome involved with domestication (Montague et al., 2014; Tamazian et al., 2014). Limited X-linked SNPs were evaluated in this study and, because of its transmission pattern, X-linked genes are good candidates for selection during domestication and deserve further investigation. SNPs have already demonstrated the potential to equal or even outperform microsatellites for specific questions such as individual ancestry (Lao et al., 2008), population assignment (for example, Seddon et al., 2005; Narum et al., 2008; Smith and Seeb, 2008; Coates et al., 2009) and pedigree studies (Santure et al., 2010; Hauser et al., 2011), and proved to have large allele frequency differences among populations (Freamo et al., 2011). The inclusion of SNPs associated with specific known domestic cat phenotypes, particularly recessive traits such as melanism, hair types and gloving (for review see Lyons, 2010; 2012,) would likely increase the power for domestic cat introgression into wildcats. Combined repertoires of autosomal SNPs, X- and Y-linked markers and mtDNA variants should all help decipher the domestication of the cat and the dynamics of wildcat and domestic cat populations around the world. Ultimately, we consider that SNPs are the molecular markers of choice for hybridization studies, as they can provide an easier, cheaper and standardized method to be implemented in conservation programs.

for providing putative wildcat samples. We are also grateful to all anonymous veterinarians and biologists that assisted in sample collection.

Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001). The problems with hybrids: setting conservation guidelines. *Trends Ecol Evol* **16**: 613–622.

Allendorf FW, Hohenlohe PA, Luikart G (2010). Genomics and the future of conservation genetics. *Nat Rev Genet* **11**: 697–709.

Anderson EC, Thompson EA (2002). A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* **160**: 1217–1229.

Beaumont M, Barratt EM, Gottelli D, Kitchener AC, Daniels MJ, Pritchard JK et al. (2001). Genetic diversity and introgression in the Scottish wildcat. *Mol Ecol* **10**: 319–336.

Brumfield R, Beerli P, Nickerson DA, Edwards SV (2003). The utility of single nucleotide polymorphisms in inferences of population history. *Trends Ecol Evol* **18**: 249–256.

Brumfield RT (2010). Speciation genetics of biological invasions with hybridization. *Mol Ecol* **19**: 5079–5083.

Coates BS, Sumerford DV, Miller NJ, Kim Ks, Sappington TW, Siegfried BD et al. (2009). Comparative performance of single nucleotide polymorphism and microsatellite markers for population genetic analysis. *J Hered* **100**: 556–564.

Daniels MJ, Balharry D, Hirst D, Kitchener AC, Aspinall RJ (1998). Morphological and pelage characteristics of wild living cats in Scotland: implications for defining the 'wildcat'. *J Zool* **244**: 231–247.

Daniels MJ, Beaumont MA, Johnson PJ, Balharry D, Macdonald DW, Barratt E (2001). Ecology and genetics of wild-living cats in the north-east of Scotland and the implications for the conservation of the wildcat. *J Appl Ecol* **38**: 146–161.

Driscoll CA, Menotti-Raymond M, Roca AL, Hupe K, Johnson WE, Geffen E et al. (2007). The near eastern origin of cat domestication. *Science* **317**: 519–523.

Driscoll CA, MacDonald DW, O'Brien SJ (2009). From wild animals to domestic pets, an evolutionary view of domestication. *Proc Natl Acad Sci USA* **106**: 9971–9978.

Driscoll CA, Yamaguchi N, O'Brien SJ, Macdonald DW (2011). A suite of genetic markers useful in assessing wildcat (*Felis silvestris* ssp.)—domestic cat (*Felis silvestris catus*) admixture. *J Hered* **102**: S87–S90.

Eckert I, Suchentrunk F, Markov G, Hartl GB (2010). Genetic diversity and integrity of German wildcat (*Felis silvestris*) populations as revealed by microsatellites, allozymes, and mitochondrial DNA sequences. *Mamm Biol* **75**: 160–174.

Esteves PJ, Abrantes J, van der Loo W (2007). Extensive gene conversion between CCR2 and CCR5 in domestic cat (*Felis catus*). *Int J Immunogen* **34**: 321–324.

Excoffier L, Lischer HEL (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Res* **10**: 564–567.

Falush D, Stephens M, Pritchard JK (2007). Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol Ecol Notes* **7**: 574–578.

Fitzpatrick BM, Johnson JR, Kump DK, Smith JJ, Voss SR, Shaffer HB (2010). Rapid spread of invasive genes into a threatened native species. *Proc Natl Acad Sci USA* **107**: 3606–3610.

Freamo H, O'Reilly P, Berg PR, Lien S, Boulding EG (2011). Outlier SNPs show more genetic structure between two Bay of Fundy metapopulations of Atlantic salmon than do neutral SNPs. *Mol Ecol Res* **11**: 254–267.

Garvin MR, Saitoh K, Gharrett AJ (2010). Application of single nucleotide polymorphisms to non-model species: a technical review. *Mol Ecol Res* **10**: 915–934.

Gärke C, Ytournel F, Bed'hom B, Gut I, Lathrop M, Weigend S et al. (2012). Comparison of SNPs and microsatellites for assessing the genetic structure of chicken populations. *Anim Genet* **43**: 419–428.

Germain E, Benhamou S, Poulle ML (2008). Spatio-temporal sharing between the European wildcat, the domestic cat and their hybrids. *J Zool* **276**: 195–203.

Grabe H, Worel G (2001). Die Wildkatze - Zurück auf leisen Pfoten. *Buch & Kunstverlag Oberpfalz, Amberg.*

Glover KA, Hansen MM, Lien S, Als TD, Høyheim B, Skaala Ø (2010). A comparison of SNP and STR loci for delineating population structure and performing individual genetic assignment. *BMC Genet* **11**: 2–12.

Goudet J (2001). FSTAT, a program to estimate and test gene diversities and fixation indices version 2.9.3. Available from www.unil.ch/izea/softwares/fstat.html.

Hauser L, Baird M, Hilborn R, Seeb LW, Seeb JE (2011). An empirical comparison of SNPs and microsatellites for parentage and kinship assignment in a wild sockeye salmon (Oncorhynchus nerka) population. *Mol Ecol Res* **11**: 150–161.

Hertwig ST, Jungnickel A, Stepanow S, Böhle UR, Fischer MS (2009). Regionally high rates of hybridization and introgression in German wildcat populations (*Felis silvestris*, Carnivora, Felidae). *J Zool Syst Evol Res* **47**: 283–297.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009). Inferring weak population structure with the assistance of sample group information. *Mol Ecol Res* **9**: 1322–1332.

Jakobsson M, Rosenberg NA (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure'. *Bioinformatics* **23**: 1801–1806.

Johnson WE, Eizirik E, Pecon-Slattery J, Murphy WJ, Antunes A, Teeling E et al. (2006). The late Miocene radiation of modern Felidae: a genetic assessment. *Science* **311**: 73–77.

Kalinowski ST (2005). Hp-rare 1.0: a computer program for performing rarefaction on measures of allelic richness. *Mol Ecol Notes* **5**: 187–189.

Karlsson S, Moen T, Lien S, Glover KA, Hindar K (2011). Generic genetic differences between farmed and wild Atlantic salmon identified from a 7 K SNP-chip. *Mol Ecol Res* **11**: 247–253.

Kitchener AC, Yamaguchi N, Ward JM, Macdonald DW (2005). A diagnosis for the Scottish wildcat (Felis silvestris): a tool for conservation action for a critically-endangered felid. *Anim Con* **8**: 223–237.

Kohn MH, Murphy WJ, Ostrander EA, Wayne RK (2006). Genomics and conservation genetics. *Trends Ecol Evol* **21**: 629–637.

Krüger M, Hertwig ST, Jetschke G, Fischer MS (2009). Evaluation of anatomical characters and the question of hybridization with domestic cats in the wildcat population of Thuringia, Germany. *J Zool Syst Evol Res* **47**: 268–282.

Kurushima JD, Lipinski MJ, Gandolfi B, Froenicke L, Grahn JC, Grahn RA et al. (2012). Variation of cats under domestication: genetic assignment of domestic cats to breeds and worldwide random-bred populations. *Anim Genet* **44**: 311–324.

Lao O, van Duijn K, Kersbergen P, de Knijff P, Kayser M (2006). Proportioning whole-genome single-nucleotide-polymorphism diversity for the identification of geographic population structure and genetic ancestry. *Am J Hum Genet* **78**: 680–690.

Lao O, de Gruijter JM, van Duijn K, Navarro A, Kayser M (2008). Signatures of positive selection in genes associated with human skin pigmentation as revealed from analyses of single nucleotide polymorphisms. *Ann Hum Genet* **71**: 354–369.

Lecis R, Pierpaoli M, Biro ZS, Szemethy L, Ragni B, Vercillo F et al. (2006). Bayesian analyses of admixture in wild and domestic cats (*Felis silvestris*) using linked microsatellite loci. *Mol Ecol* **15**: 119–131.

Lipinski MJ, Froenicke L, Baysac KC, Billings NC, Leutenegger CM, Levy AM et al. (2008). The ascent of cat breeds: genetic evaluations of breeds and worldwide random-bred populations. *Genomics* **91**: 12–21.

Liu N, Chen L, Wang S, Oh C, Zhao H (2005). Comparison of singlenucleotide polymorphisms and microsatellites in inference of population structure. *BMC Genet* **6**: S26.

Lyons LA, Imes DL, Rah HC, Grahn RA. (2005a). Tyrosinase mutations associated with Siamese and Burmese patterns in the domestic cat (Felis catus). *Anim Genet* **36**: 119–126.

Lyons LA, Foe IT, Rah HC, Grahn RA.: Chocolate coated cats: TYRP1 mutations for brown color in domestic cats (2005b). *Mamm Genome* **16**: 356–366.

Lyons LA (2010). Feline genetics: clinical applications and genetic testing topics companion. *Anim Med* **25**: 203–212.

Lyons LA (2012). Genetic testing in domestic cats. *Mol Cell Probe* **26**: 224–230.

Mattucci F, Oliveira R, Bizzarri L, Vercillo F, Anile S, Ragni B et al. (2013). Genetic structure of wildcat (Felis silvestris) populations in Italy. *Ecol Evol* **3**: 2443–2458.

Miller J, Poissant J, Kijas J, Coltman Dand the International Sheep Genomics Consortium (2011). A genome-wide set of SNPs detects population substructure and long range linkage disequilibrium in wild sheep. *Mol Ecol Res* **11**: 314–322.

Montague MJ, Li G, Gandolfi B, Khan Z, Aken BL, SMJ Searle et al. (2014). Comparative analysis of the domestic cat genome reveals genetic signatures underlying feline biology and domestication. *Proc Natl Acad Sci USA* **111**: 17230–17235.

Monzón J, Kays R, Dykhuizen D. (2013). Assessment of coyote-wolf-dog admixture using ancestry-informative diagnostic SNPs. *Mol Ecol* **23**: 182–197.

Morin PA, Luikart G, Wayne RKSNP-Workshop Group (2004). SNPs in ecology, evolution and conservation. *Trends Ecol Evol* **19**: 208–216.

Morin PA, Martien KK, Taylor BL (2009). Assessing statistical power of SNPs for population structure and conservation studies. *Mol Ecol Res* **9**: 66–73.

Mullikin JC, Hansen NF, Shen L, Ebling H, Donahue WF, Tao W et al. (2010). Light whole genome sequence for SNP discovery across domestic cat breeds. *BMC Genomics* **11**: 406.

Narum S, Banks M, Beacham T, Bellinger M, Campbell M, Dekoning J et al. (2008). Differentiating salmon populations at broad and fine geographical scales with microsatellites and single nucleotide polymorphisms. *Mol Ecol* **17**: 3464–3477.

Nussberger B, Greminger MP, Grossen C, Keller LF, Wandeler P (2013). Development of SNP markers identifying European wildcats, domestic cats, and their admixed progeny. *Mol Eco Res* **13**: 447–460.

Nussberger B, Wandeler P, Camenisch G (2014). A SNP chip to detect introgression in wildcats allows accurate genotyping of single hairs. *Eur J Wildlife Res* **60**: 405–410.

Nei M (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**: 583–590.

Nielsen E.E., Bach L.A., Kotlicki P. (2006). Hybridlab (version 1.0): a program for generating simulated hybrids from population samples. *Mol Ecol Notes* **6**: 971–973.

O'Brien SJ, Devillard S, Say L, Vanthomme H, Léger F, Ruette S et al. (2009). Preserving genetic integrity in a hybridising world: are European Wildcats (*Felis silvestris silvestris*) in Eastern France distinct from sympatric feral domestic cats? *Biodiv Conserv* **18**: 2351–2360.

Oliveira R, Godinho R, Randi E, Ferrand N, Alves PC (2008a). Molecular analysis of hybridisation between wild and domestic cats (*Felis silvestris*) in Portugal: implications for conservation. *Conserv Genet* **9**: 1–11.

Oliveira R, Godinho R, Randi E, Alves PC (2008b). Hybridization vs conservation: are domestic cats threatening the genetic integrity of European wildcat (*Felis silvestris silvestris*) populations in Iberian Peninsula? *Phil Trans R Soc B* **363**: 2953–2961.

Paetkau D, Waits LP, Clarkson PL et al. (1998). Variation in genetic diversity across the range of North American brown bears. *Conserv Biol* **12**: 418–429.

Peakall R, Smouse PE (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* **6**: 288–295.

Pertoldi C, Wójcik JM, Matgorzata T, Kawalko A, Kristensen TN, Loeschcke V et al. (2010). Genome variability in European and American bison detected using the BovineSNP50 BeadChip. Conserv Genet **11**: 627–634.

Pierpaoli M, Biro ZS, Herrmann M, Hupe K, Fernandes M, Ragni B et al. (2003). Genetic distinction of wildcat (Felis silvestris) populations in Europe, and hybridization with domestic cats in Hungary. Mol Ecol **12**: 2585–2598.

Platz S, Hertwig ST, Jetscke G, Krüger M, Fishe MS (2011). Comparative morphometric study of the Slovakian wildcat population (Felis silvestris silvestris): evidence for a low rate of introgression? Mamm Biol **11**: 627–634.

Poissant J, Hogg JT, Davis CS, Miller JM, Maddox JF, Coltman DW (2010). Genetic linkage map of a wild genome: genomic structure, recombination and sexual dimorphism in bighorn sheep. BMC Genomics **11**: 524.

Pontius JU, Mullikin JC, Smith DR, Lindblad-Toh K, Gnerre S et al. (2007). Initial sequence and comparative analysis of the cat genome. Genome Res **17**: 1675–1689.

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

Puzachenko AY (2002). Hybrid syndrome and method for identification of hybrids in museum collections of Felis silvestris and Felis lybica. Säugetierk Inform **26**: 234–248.

Ragni B, Possenti M (1996). Variability of coat-color and markings system in Felis silvestris. Italian J Zool **63**: 285–292.

Randi E, Pierpaoli M, Beaumont M, Ragni B, Sforzi A (2001). Genetic identification of wild and domestic cats (Felis silvestris) and their hybrids using Bayesian clustering methods. Mol Biol Evol **18**: 1679–1693.

Robinson R (1977). Genetics for Cat Breeders, 2nd edn. Pergamon Press: Oxford.

Rosenberg NA, Li LM, Ward R, Pritchard JK (2003). Informativeness of genetic markers for inference of ancestry. Am J Hum Genet **73**: 1402–1422.

Rosenberg NA (2004). Distruct: a program for the graphical display of population structure. Mol Ecol Notes **4**: 137–138.

Rosenberg NA (2005). Algorithms for selecting informative marker panels for population assignment. J Comp Biol **12**: 1183–1201.

Sacks BN, Louie S (2008). Using the dog genome to find single nucleotide polymorphisms in red foxes and other distantly related members of the Canidae. Mol Ecol Res **8**: 35–49.

Santure AW, Staple J, Ball AD, Birkhead TR, Burke T, Slate J (2010). On the use of large marker panels to estimate inbreeding and relatedness: empirical and simulation studies of a pedigreed zebra finch population typed at 771 SNPs. Mol Ecol **19**: 1439–1451.

Schauenberg P (1969). L'identification du Chat forestier d'Europe Felis s. silvestris Schreber, 1777 par une méthode ostéometrique. Rev Suisse Zool **76**: 433–441.

Schauenberg P (1970). Note sur l'indice cranien du Chat domestique fèral. Rev Suisse Zool **78**: 209–215.

Schlötterer C (2004). The evolution of molecular markers – just a matter of fashion? Nat Rev Genet **5**: 63–69.

Seddon JM, Parker HG, Ostrander EA, Ellegren H (2005). SNPs in ecological and conservation studies: a test in the Scandinavian wolf population. Mol Ecol **14**: 503–510.

Smith CT, Seeb LW (2008). Number of alleles as a predictor of the relative assignment accuracy of short tandem repeat (STR) and single-nucleotide-polymorphism (SNP) baselines for chum salmon. Trans Am Fish Soc **137**: 751–762.

Tamazian G, Simonov S, Dobrynin P, Makunin A, Logachev A, Komissarov A et al. (2014). Annotated features of domestic cat - Felis catus genome. Gigascience **3**: 13.

Teeter KC, Payseur BA, Harris LW, Bakewell MA, Thibodeau LM, O'Brien JE et al. (2008). Genome-wide patterns of gene flow across a house mouse hybrid zone. Genome Res **18**: 67–76.

Tokarska M, Marshall T, Kowalczyk R, Wojcik JM, Pertoldi C, Kristensen TN et al. (2009). Effectiveness of microsatellite and SNP markers for parentage and identity analysis in species with low genetic diversity: the case of European bison. Hereditary **103**: 326–332.

Vähä JP, Primmer CR (2006). Efficiency of model- based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. Mol Ecol **15**: 63–72.

Vigne JD, Briois F, Zazzo A, Willcox G, Cucchi T, Thiébault S et al. (2012). First wave of cultivators spread to Cyprus at least 10,600 y ago. Proc Natl Acad Sci USA **109**: 8445–8449.

vonHoldt BM, Pollinger JP, Lohmueller KE, Han E, Parker HG, Quignon P et al. (2010). Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. Nature **464**: 898–902.

vonHoldt BM, Pollinger JP, Earl DA et al. (2011). A genome-wide perspective on the evolutionary history of enigmatic wolf-like canids. Genome Res **21**: 1294–1305.

vonHoldt BM, Pollinger JP, Earl DA, Parker HG, Ostrander EA, Wayne RK (2012). Identification of recent hybridization between gray wolves and domesticated dogs by SNP genotyping. Mamm Genome **24**: 80–88.

Waits LP, Luikart G, Taberlet P. (2001). Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. Mol Ecol **10**: 249–256.

Yamaguchi N, Driscoll CA, Kitchener AC, Ward JM, Macdonald DW (2004a). Craniological differentiation between European wildcats (Felis silvestris silvestris), African wildcats (F. s. lybica) and Asian wildcats (F. s. ornata): implications for their evolution and conservation. Biol J Linn Soc **83**: 47–63.

Yamaguchi N, Kitchener AC, Driscoll CA, Ward JM, Macdonald DW (2004b). Craniological differentiation amongst wild-living cats in Britain and southern Africa: natural variation or the effects of hybridisation? Anim Conserv **7**: 339–351.

Supplementary Information accompanies this paper on Heredity website (http://www.nature.com/hdy)