Molecular Ecology (2001) **10**, 319–336

Genetic diversity and introgression in the Scottish wildcat

M. BEAUMONT,*¶ E. M. BARRATT,*†† D. GOTTELLI,* A. C. KITCHENER,† M. J. DANIELS,‡** J. K. PRITCHARD§ and M. W. BRUFORD*‡‡

**Institute of Zoology, Zoological Society of London, Regents Park, London NW 1RY, UK, †Department of Geology and Zoology, National Museums of Scotland, Royal Museum, Chambers Street, Edinburgh EH1 1JF, UK, ‡Scottish Natural Heritage, 2 Anderson Place, Edinburgh EH6 5NP, UK, §Department of Statistics, University of Oxford, 1 South Parks Road, Oxford, OX1 3TG, UK*

Abstract

This paper describes a genetic analysis of wild-living cats in Scotland. Samples from 230 wild-living Scottish cats (including 13 museum skins) and 74 house cats from England and Scotland were surveyed for nine microsatellite loci. Pelage characteristics of the wild-living cats were recorded, and the cats were then grouped into five separate categories depending on the degree to which they conformed to the characteristics attributed to *Felis silvestris* Schreber, 1775. Allele frequency differences between the morphological groups are greater than those among the three house cat samples. Analysis of genetic distances suggests that more of the differences between individuals can be explained by pelage than geographical proximity, and that pelage and geographical location are not confounded. Ordination of the genetic distances suggests two main groups of wild-living cats, with intermediates, and one group is genetically very similar to the house cats, while the other group contains all cats taxonomically identified as wildcat based on morphology. A genetic mixture analysis gives similar results to the ordination, but also suggests that the genotypes of a substantial number of cats in the wildcat group are drawn from a gene pool with genotypes in approximately equilibrium proportions. We argue that this is evidence that these cats do not have very recent domestic ancestry. However, from the morphological data it is highly likely that this gene pool also contains a contribution from earlier introgression of domestic cat genes.

Keywords: admixture, domestic cat, hybridization, introgression, microsatellite, wildcat

Received 8 January 2000; revision received 7 August 2000; accepted 25 September 2000

Introduction

Regions where two genetically differentiated taxa meet, mate and hybridize and where introgression occurs have received considerable attention in recent years and have been studied in most major groups of organisms. Although many studies have been carried out on hybrid zones (Barton & Hewitt 1985, 1989; Goodman *et al*. 1999), others have been attempted where little spatial structure is evident or where one or both parental populations are absent (Gottelli *et al*. 1994; Roy *et al*. 1994; Grant & Grant 1997). These latter studies are important in a conservation context and the processes that occur in hybrid populations may prove to be an important force in the evolution of plants and animals in an increasingly disturbed landscape. These processes are of special concern if they involve endangered populations and have implications for the enforcement of legal protection for those species (O'Brien & Mayr 1991; Pennock & Dimmick 1997). We describe here one such study on the Scottish population of wild-living cats. A related study, using the same markers on a local population of cats is described in Daniels *et al*. (2001). Below we describe the known history of the population, summarize previous genetic surveys, and describe the aims of the present study.

Correspondence: Mark A. Beaumont. Present address: ¶School of Animal and Microbial Sciences, University of Reading, Whiteknights, PO Box 228, Reading RG6 6AJ, UK. Fax: + 44 118987 5123; E-mail: m.a.beaumont@reading.ac.uk

Present addresseses: ††3 Rutland Road, Hazel Grove, Stockport, Cheshire, SK7 6JD, UK, **Animal Biology Division, Scottish Agricultural College, West Mains Road, Edinburgh EH9 3JG, UK, ‡‡Cardiff School of Biosciences, Cardiff University, PO Box 915, Cathays Park, Cardiff CF10 3TL, UK.

Historical background to wildcats in Scotland

The taxonomic status and interpretation of the history of wild-living cats in Scotland is contentious. Some authors (e.g. Daniels *et al*. 1998) suggest that since the introduction of the domestic cat approximately 2000 years ago, there is no firm evidence regarding the identity of the Scottish wildcat, and that the use of this label presupposes a distinction that may not exist. However, for convenience, we will describe the history of wild-living cats in Scotland primarily based on the traditional view of earlier taxonomists (e.g. Miller 1912; Pocock 1951), and highlight the more contentious issues as they arise.

The wildcat, *Felis silvestris*, is the only native member of the cat family (Felidae) to survive in Britain today. It was present as part of the early Holocene fauna of mainland Britain (Yalden 1999) and was once widespread throughout the country (Easterbee *et al*. 1991). As sea levels rose during the early Holocene, the British Isles became isolated from mainland Europe 8000–9500 radiocarbon years ago (Yalden 1999). Traditionally, this isolated population is considered to be morphologically differentiated from European populations as *Felis silvestris grampia* (Miller 1912), on the basis of its darker colouration and bolder stripes on the legs and flanks. Controversy surrounds the taxonomic status of the wildcat in both Britain and mainland Europe (Daniels *et al*. 1998). Earlier taxonomists considered the European wildcat to be a distinct species *Felis silvestris* and the Scottish form has been described as a distinct subspecies *F. s. grampia*, defined by characteristic darker colouration than the mainland form plus well defined dark markings on the sides and legs (Miller 1912).

Since the middle ages the wildcat in Britain has experienced a steady reduction in range through hunting and habitat destruction. It probably disappeared from southern England during the 16th century and was lost from northern England and Wales by 1880 (Langley & Yalden 1977). This coincided with the last recorded sighting in southern Scotland and an increase in the level of persecution as a result of the establishment of large sporting estates. The population is thought to have reached its lowest level around 1914, when it is believed by some to have been confined to Inverness-shire, parts of Argyllshire, Ross-shire and Sutherland (Easterbee *et al*. 1991), although a recent review suggests that wildcats may have survived over a wider geographical area (A. Lamb, personal communication; Balharry & Daniels 1998). Freedom from persecution during and after the first World War (1914–18) then allowed the wildcat to expand its range, aided by the setting up of the Forestry Commission in 1919 and the resultant increased planting of coniferous woodland. However, recolonization of northern Scotland may have coincided with increased hybridization with domestic cats.

Current problems for the wildcat in Scotland

Since 1988 the Scottish wildcat has had full legal protection under Schedule 5 of the Wildlife and Countryside Act and it is also protected under Annex IV of the European Directive 92/43/EEC on the conservation of natural habitats and of wild fauna and flora (the Habitats and Species Directive), which requires strict protection within the EU.

The presence of feral domestic cats and the possibility of hybridization has, however, made this legislation ineffective because an expert witness in court was unable to distinguish wildcats from hybrids beyond reasonable doubt (Balharry & Daniels 1998). The legal protection of the Scottish wildcat is further complicated by the fact that domestic cats have been present in Britain since the Iron Age (more than 2000 years ago; Kitchener 1998), since when it has become established as a feral animal in many parts of Britain (Serpell 1988; Clutton-Brock 1989). There are differing views on the significance of the impact of this potential for hybridization. One view is that its contribution would have been slight until relatively recently (Kitchener 1998), but another view (Daniels *et al*. 2001), based on evidence of behavioural similarity between cats in a study area containing morphologically domestic and morphologically wildcat individuals, is that it may have made a substantial impact on the genetic composition of wild-living cats in Scotland.

In order to determine reliable molecular markers for the identification of wildcats, it is necessary to compare purebred wildcats with purebred domestic cats. However, wildcats and domestic cats have been in sympatry, and hence potentially hybridizing since before the wildcat was first described scientifically by Schreber in 1775. Therefore, we cannot be certain that introgression had not already occurred and obscured the discriminatory potential of both morphological and genetic markers. In its extensive geographical range in Eurasia and Africa the wildcat is always sympatric with domestic cats, so that unequivocally purebred wildcats are not available for comparison. Thus, the status of the wild-living cat population in Scotland and elsewhere is the subject of much debate.

Earlier studies using molecular markers

Molecular studies have been carried out by O'Brien and colleagues, in which the taxonomic status of the genus *Felis* has been determined within the Felidae as a whole (albumin immunological distance, Collier & O'Brien 1985; mitochondrial 12S rRNA and cytochrome *b*, Janczewski *et al*. 1995; Masuda *et al*. 1996; mitochondrial 16S rRNA and NADH-5 Johnson & O'Brien 1997). These studies have been unable to resolve relationships between the domestic cat, *F. silvestris* and *F. lybica* but have shown that these three

and the Sand Cat (*F. margarita*) form a distinct clade within the Felidae. These authors have suggested that these results support the proposal of Nowak (1991) and Randi & Ragni (1991) that *F. silvestris*, *F. lybica* and *F. catus* are better regarded as subspecies of a single polytypic species, *F. silvestris*. However, this species should be called *F. catus*, as this name is the earliest published of the three. The degree of divergence in mitochondrial sequence among these three taxa is approximately one fifth of the divergence between these and the jungle cat (*F. chaus*) and the black-footed cat (*F. nigripes*) (Masuda *et al*. 1996), and as the mitochondrial most recent common ancestor (MRCA) of all six species is placed from molecular evidence at around 6 Ma (Johnson & O'Brien 1997; see also Collier & O'Brien 1985), this would imply a mitochondrial MRCA among the four *silvestris*-like populations of around 1 Ma.

Studies based on nuclear markers support the conclusions drawn from mitochondrial markers that there is weak differentiation between wildcats and domestic cats. Allozyme electrophoresis of 54 loci by Randi & Ragni (1991) suggested that Italian populations of wild-living cats classed as wildcats showed reduced levels of genetic variability in comparison with domestic cats. Genetic distance analysis based on these markers indicated little differentiation between the wild-living cats and house cats. There was, however, evidence that the house cats were genetically more similar to the African wild cat (*F. libyca*) than to the local wild-living cat population. In a study of Scottish wild-living cats, Hubbard *et al*. (1992) examined differentiation between wild-living cats classed as wildcats and domestic cats using a number of different types of marker, including electrophoretic markers, and found frequency differences of alleles at some loci, but no fixed differences at any locus.

Aims of the present study

In order to clarify the legal situation and assist future conservation management, Scottish Natural Heritage (SNH) initiated a comprehensive study into the morphology, genetics and behavioural ecology of wild-living cats in Scotland. Traditionally, introgressed individuals can be identified by morphological examination but in the case of closely related taxa such as wildcats and domestic cats there is contention over the identification of unambiguous discriminatory characters. The genetic status of this population was uncertain at the outset of the study (Balharry & Daniels 1998), and three possibilities could be considered: (i) wild-living cats are genetically distinct from domestic cats, despite the overlap in morphology, and hybrids are rare and identifiable; (ii) wild-living cats consist of individuals that have differing mixtures of wildcat and domestic cat ancestry (a 'hybrid swarm', Mayr 1963); or (iii) wild-living cats are genetically indistinguishable from feral domestic cats. Using microsatellites (previously shown to be useful in such studies, e.g. Goostrey *et al*. 1998; Gottelli *et al*. 1994; Goodman *et al*. 1999), the aim of the project was to distinguish between these alternatives, and to determine if genetically unique populations are present which might require special consideration. In order to aid the analysis we used information on pelage characteristics of the sampled cats that have been used to identify wildcats in, for example, field guides (Easterbee 1991). Measurements of intestine length relative to body size, previously shown to be useful in classifying the wild-living cats in Scotland (Balharry & Daniels 1998; Daniels *et al*. 1998) were also available for a subsample of cats (Beaumont *et al*. 2001). Results are discussed in the context of the taxonomic controversy together with their implications for the continued protection of the Scottish wildcat.

Materials and methods

Samples localities and collection

A total of 230 skin and blood samples were analysed from both extant and museum specimens covering a range of sites across Scotland (see Fig. 1). Trapping and sampling, including collection from road kills, was carried out between 1989 and 1994 by the Scottish Natural Heritage (Balharry & Daniels 1998). Fifteen samples from museum specimens (University of Aberdeen, Zoology Department Museum; City of Dundee Museum) were included and of these, 13 provided genetic information used in this paper. Ordinance Survey (OS) grid references were recorded for the cats. These were converted to x and y coordinates measured from a common origin (position 000000 on the NR 100 km square in the OS coordinate system). The elevation of the sampling location of each cat was estimated from the OS grid references using Geographical Information System (GIS). Blood and tissue samples for comparison were collected from three domestic cat populations (house cats) through veterinary practices in Kirriemuir (20 miles north of Dundee) ($n = 18$), North Yorkshire ($n = 46$) and Surrey ($n = 10$). Samples were either stored immediately at −20 °C, or collected in tissue (5 M NaCl: 20% DMSO) (Amos & Hoelzel 1991), or blood buffer (0.1 M Tris-HCl, pH 8.0; 0.1 M EDTA; 10 mM NaCl; 0.5% SDS) (Bruford *et al*. 1998).

DNA extraction

Total DNA was isolated from tissue and hair samples using a standard proteinase K lysis and organic solvent purification method as described in Sambrook *et al*. (1989). Briefly 0.5–1 g of tissue was digested overnight with

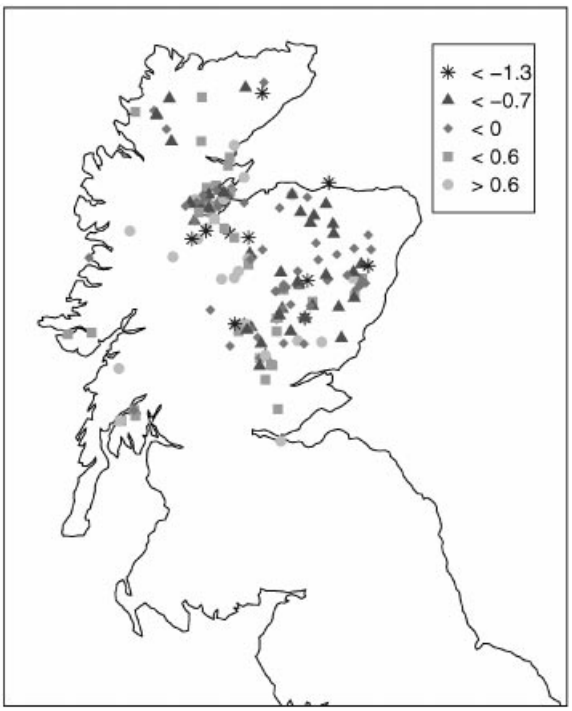

Fig. 1 Map of Scotland showing the locations of individuals and their scores on the first axis of the ordinations shown in Fig. 3 and described in the text. The scores have been banded into five groups, coded on a greyscale, as shown in the figure legend. Note that some individuals are masked by others caught at similar locations.

proteinase K at 50 °C. DNA was then extracted by treatment with phenol, phenol/chloroform/isoamyl alcohol (25:24:1) and chloroform/isoamyl alcohol (24:1). Precipitation was achieved by the addition of 1/10 volume 3 M sodium acetate and two volumes of absolute ethanol. The DNA was washed in 70% ethanol before being resuspended in sterile distilled water.

Microsatellite genotyping

Nine microsatellite loci, originally isolated in domestic cats, were used to screen the above samples and are listed in Table 1 (Menotti-Raymond & O'Brien 1995). Detection of microsatellite alleles was achieved by end-labelling one primer of the pair using [γ^{32}]-ATP (Amersham) and T4 polynucleotide kinase (NEB) at 37 °C for 1 h, and performing 30 cycles of PCR amplification in a 10-μL reaction volume containing 100–200 ng of genomic DNA, 1.5 pmoles of each primer. A 100 μM of each dNTP, 1.5–2.5 mM $MgCl_2$, 16 mM $(NH_4)_2SO_4$, 67 mM Tris-HCl (pH 8.8 at 25 °C), 0.01% Tween-20, 1 μL DMSO and 0.4 unit of *Taq*. Cycling was carried out in a Hybaid thermal cycler as follows: 94 °C for 3 min; (94 °C for 1 min; 55/56 °C for 1.5 min; 72 °C for 1 min); 72 °C for 10 min. Two microlitres of each product was then mixed with 2 μL of formamide loading dye (Amersham) and heated to 95 °C before being loaded onto a 6% sequencing gel (Sequagel™). An M13 control sequencing reaction (USB Sequenase kit) was run adjacent to the samples to provide an absolute size marker for the microsatellite alleles. Further samples of known allele size were used as additional standards. Gels were dried and exposed overnight at –70 °C. Due to poor quality of some of the material not all individuals could be genotyped for all nine loci and the distribution of numbers of cats genotyped for a particular number of loci is shown in Table 1. Full details of genotypes and the other variables measured in this study are available at http://www.rubic.rdg.ac.uk/~mab

Morphometrics and pelage characteristics

Photographs of each cat were submitted to the National Museums of Scotland for identification by A.C.K. using traditionally accepted morphological criteria (Ward & Kitchener, unpublished data). Cats were ranked as follows: 1 — wildcat; 2 — wildcat/hybrid; 3 — hybrid; 4 — domestic/hybrid; and 5 — domestic. For further analysis this variable is referred to as the traditional classification, and we will prefix reference to these groups with TC (e.g. TC-wildcat). Some of the cats were unclassified either because the photographs were not sufficiently clear to make a judgement, or because photographs were not available. This also applied to some of the museum skins, which, although judged to be wildcat by the museums were not independently checked by A.C.K.

In order to put the judgements on a more objective basis, approximately one year after the initial examination, the photographs were presented again, and were more explicitly classified using five pelage characters that are traditionally regarded as diagnostic (three of these characters are shown in Fig. 2). Each character was scored as 3 — wildcat, 2 — intermediate, 1 — domestic. The characters used are: tail shape (bushy = wildcat; tapering = domestic); dorsal stripe (stops at root of tail = wildcat; continues into tail = domestic); rump spots (stripes on rump = wildcat; spots on rump = domestic); tail tip colour (black = wildcat; any other colour = domestic); paw colour (white = domestic; other colour = wildcat). Again, this judgement was made without knowledge of the genotype of the cats. A composite variable summarizing these five measurements was calculated by standardizing all variables to have mean 0 and variance 1 and then calculating the mean score for each individual. Although the repeatability of these two scoring methods was not directly assessed, as shown in the results, the composite variable is highly correlated with the traditional classification score, which gives indirect evidence that each has reasonable repeatability.

Table 1(a) A description of the loci used and summary of the allelic variation in wild-living cats in Scotland and the house cats. The number of individuals that were scored, *n*, and the expected heterozygosity, H_E, are given for each locus. Mean and standard errors for the numbers of alleles and heterozygosity are shown in bold

			House cat			Wild-living (including Museum)		
Locus	No. Alleles (total)	Size range	*n*	No. Alleles	H_E	*n*	No. Alleles	H_E
Fca8	14	115–149	73	13	0.831	225	11	0.814
Fca23	13	132–156	74	13	0.886	221	13	0.722
Fca35	9	136–162	69	7	0.551	196	8	0.529
Fca43	9	118–134	74	9	0.653	227	9	0.650
Fca45	18	142–163	74	15	0.871	220	16	0.901
Fca77	11	134–152	73	8	0.649	225	12	0.689
Fca90	12	91–117	73	9	0.819	218	11	0.833
Fca96	15	184–230	74	13	0.576	225	10	0.749
Fca126	10	132–152	74	8	0.777	226	9	0.760
				10.6	**0.735**		**10.9**	**0.739**
				(0.97)	**(0.043)**		**(0.79)**	**(0.037)**

Table 1(b) The distribution of cats genotyped for a particular number of loci

	Number of loci typed								
	1	2	3	4	5	6	7	8	9
All cats	1	1	1	—	2	5	7	37	250
House cats	—	—	—	—	—	—	—	8	66
Wild-living (no museum)	—	—	1	—	2	1	4	25	184
Museum	1	1	—	—	—	4	3	4	—

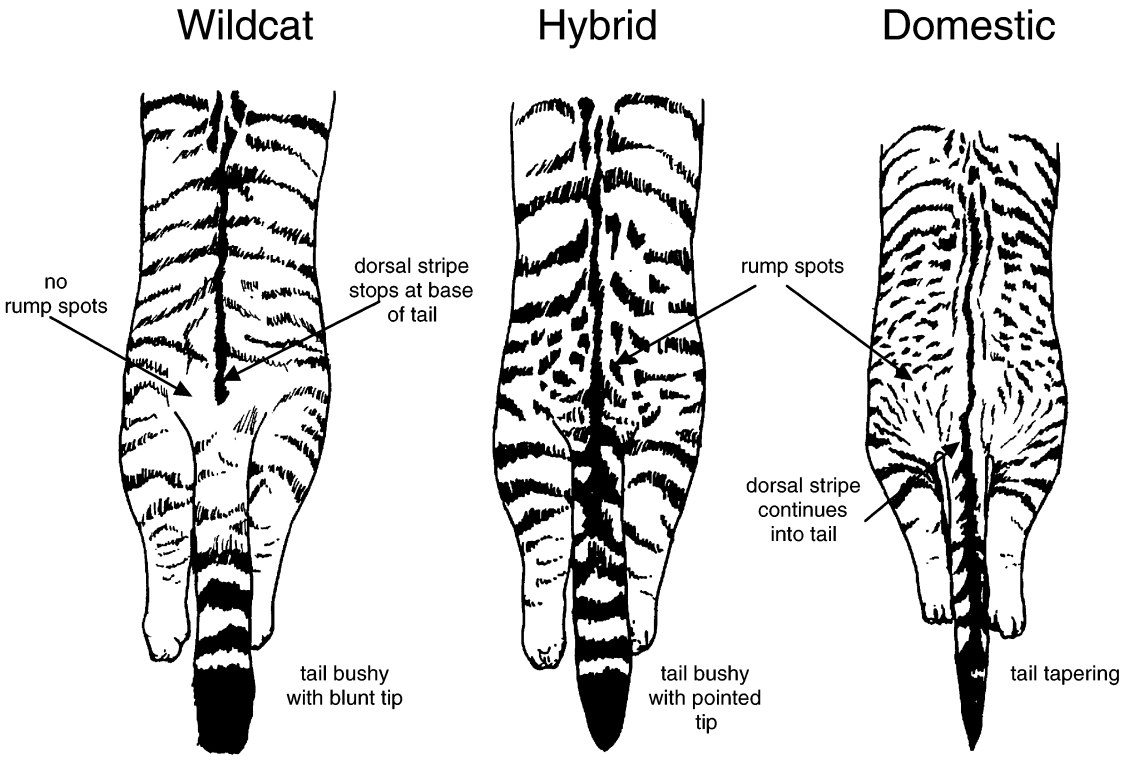

Fig. 2 Diagram showing three of the five diagnostic characters used to score individual cats morphologically. The three characters shown here are tail shape, dorsal stripe, and rump spots. Tail tip colour and paw colour are not illustrated.

Standard analyses of genetic variation

Allele and genotype frequencies were estimated for all wild-living cats as a separate group, these cats grouped by traditional classification, and also for the three groups of house cats. The presence of null alleles was tested by regressing the frequency of nonamplifying samples for each locus against the estimate of F_{IS}. Loci that have nulls at high frequency will have greater apparent homozygote excess, caused by null heterozygotes, and a greater number of nonamplifying samples due to null homozygotes (Goodman *et al*. 1999). The presence of private (group-specific) alleles in the wild-living and domestic cats was investigated. Departures from Hardy–Weinberg expectation were analysed, and genetic differentiation in allele frequencies quantified by θ of Weir & Cockerham (1984), using GENEPOP 3.1b (Raymond & Rousset 1995) and FSTAT (Jerome Goudet, http://www.unil.ch/izea/softwares/fstat.html). In addition, Goodman's (1997) estimator, ρ, was calculated for these groups using the package RSTCALC. Nei's genetic distance (Nei 1972, 1978) was estimated using the program MICROSAT (Eric Minch, http://human.stanford.edu/microsat). Although these estimators are normally used for discrete populations, they were estimated here only for the purpose of illustrating the magnitude of differentiation among groups on widely used scales.

Genetic, morphological, and environmental distances

Genetic distances among individual wild-living and house cats were estimated as standard allele-sharing distances, calculated as in Bowcock *et al*. (1995). In addition, average allele sharing distances were constructed between individual wild-living cats and all house cats, calculated as follows

$$ASD(x,-) = \frac{1}{2l(x)}\sum_{k=1}^{l(x)}\sum_{i=1}^{2}\sum_{j=1}^{a_k} p_{kj}d_{ijk}(x,-),$$

where the p_{kj} are the average allele frequencies pooled among the three house cat groups for the jth allele at the kth locus, a_k is the total number of alleles at the kth locus, $l(x)$ is the number of loci that can be compared in individual x. $d_{ijk}(x,-)$ is 1 if allele i at locus k in individual x is the same as allele j in the house cats; 0 otherwise. Genetic distances were calculated only when an individual had been typed for five or more loci. The reason for choosing this cut-off is that for pairwise comparisons, with fewer loci, some pairs of individuals could be scored for disjoint sets of loci, and therefore, no pairwise distance could be estimated.

The following morphological and environmental distances were constructed: the Euclidean distance between the coordinate points of the cat locations, described above; the absolute difference in elevation between two individuals; the absolute difference in ranking between two individuals based on the traditional classification; the Euclidean distance between the coordinate positions of two individuals based on the five morphological scores; the absolute difference in composite morphological score. These were used in the analyses described below to test whether there were correlations between the genetic distance between pairs of wild-living cats and distances based on geographical location of individuals, their elevation, and differences in morphology.

Regression analysis using distances among wild-living cats

To test whether significant relationships between the genetic distance between individuals and whether morphological distances were confounded with geographical effects, we employed multiple regression to identify confounding variables. Because the distances between individuals are nonindependent, we used matrix permutation methods to test for significance (Manly 1991). Although one and two variable methods have previously been described (Manly 1991), Legendre *et al*. (1994) have more recently described a general method. In their approach the dependent and independent variables are standardized with mean 0 and variance 1. The matrix of the dependent variable, represented in unfolded form (by row) as the vector **y**, is permuted by reordering the columns and rows. The matrix $(\mathbf{X}'\mathbf{X})^{-1}$ obtained from the matrix of the independent variables, **X**, need only be calculated once. For each permutation the vector of standardized partial regression coefficients, b, is then obtained as $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. R^2 was calculated as $((\mathbf{Xb})'(\mathbf{Xb}))/(n-1)$. A statistic T_i that is a linear function of the t statistic is then calculated as $b_i/\sqrt{1-R^2}$ for the ith regression coefficient. To find the best fitting model we used backward selection as described by Legendre *et al*. (1994), using a Bonferroni-corrected P-to-remove of 0.1, where the P-value of removing a parameter was calculated by estimating the proportion of permutations in which the statistic T_i was more extreme than that of the original observation. The P-value of R^2 could also be calculated in a similar way. A program that carries out the method, Permute!, has been written by P. Casgrain (Legendre *et al*. 1994). The program can be obtained from ftp://ftp.umontreal.ca/pub/casgrain/evolution. The original program (subsequently improved) was limited to a maximum of 29 objects. We, therefore, implemented a version that can use a larger number of objects (tested using the original Permute! program).

Regression analysis using distances between wild-living and house cats

The distance data of wild-living cats from the average house cat gene frequencies was analysed by the same

program, but in this case, because these are independent, the permutation of the **y**-vector was carried out by reordering all objects of the **y**-vector. Thus, the multiple regression analysis is the same as that described in Manly (1991) using randomization.

Ordination plots

The allele sharing distances among all cats (including house cats) were ordinated in two dimensions using standard metrical scaling (see, e.g. Manly 1994). In this method the individuals are represented in two dimensions in such a way that the Euclidean distance between individuals matches as closely as possible the allele sharing distance between individuals. For each of the categories in the traditional classification and for the three groups of house cats, the mean and standard error (square root of the variance divided by the sample size) of the scores of individual cats on the first two axes from the ordination were calculated. Centroids and ellipses based on the standard error were plotted for visual inspection. Within each group the score of individuals on the first axis was not significantly correlated with the score on the second axis, and, therefore, the ellipses were calculated assuming the correlation was zero: the radius of the ellipse along each axis is the standard error. These radii give an indication of the confidence of the location of the centroids, and tend to be smaller in groups with large sample sizes.

Density estimation

The distribution of scores on the first axis of the ordination was analysed using the density-fitting procedure of Kooperberg & Stone (1991), which estimates the logarithm of the density with splines. The splines used here are curves given by cubic equations over an interval, constrained to join smoothly to neighbouring curves at evenly placed points (knots). Thus, the log-density is approximated by a smooth curve composed of these splines, whose shape depends on the number knots. The method is fully automatic and places a number of spline knots evenly over the range of the data and then successively deletes individual knots (without changing their placement) to minimize $-2\,\mathrm{L} + \log(n)(\mathrm{k} - 1)$, where L is the maximum log-likelihood for a given number of knots, k, and n is the sample size. The second term in the equation is a penalty term; obviously the likelihood could be increasingly maximized by increasing the number of knots, and an optimal number can be found by use of some penalty against over-fitting the model. The formulation used here corresponds to the Bayesian Information Criterion (BIC) of Schwarz (1978) for comparing nested sets of models.

Genetic mixture analysis

In addition to the ordination analysis we have applied a more model-based (parametric) genetic mixture analysis, recently developed by Pritchard *et al*. (2000) (program available from http://www.stats.ox.ac.uk/~pritch/home.html). In this approach, it is assumed that there are K populations contributing to the gene pool of the sample population. For this analysis K is assumed to be 2. In the 2 population case, the two parental populations each have an unknown gene frequency distribution at each locus, $\mathbf{p}_{kl}$ for the $k = 1 \ldots 2$ populations and $l = 1 \ldots L$ loci. For the ith cat we assume that the proportion of its genotype that is drawn from population 1 or 2 is, respectively, q_i, or $1 - q_i$. Then the probability of each allele observed at locus l in individual i is modelled as $q_i p_{1l} + (1 - q_i)p_{2l}$. It is assumed that for each individual (conditional on q_i), the alleles at each locus are sampled independently of one another. Pritchard *et al*. (2000) have developed a Markov Chain Monte Carlo method that can allow the posterior probability distribution to be estimated separately for each parameter (in particular the q_i and p_{kl}), integrated over all the other parameters. This represents a significant extension to a methodology introduced by Rannala & Mountain (1997) because one does not need to specify in advance the gene frequencies of the contributing source populations. Probability intervals for all parameters can be easily calculated. For this analysis we used two modelling approaches. In one case we assumed uninformative priors on all the q_i (i.e. we assumed that, before applying the model to the data, all values of q_i were equally likely, for all i), and in the other case we assumed point priors of 0 for the q_i for the known domestic cats (i.e. we forced the model to consider them as pure domestics). The results from these two approaches were very similar, suggesting that the assumed prior on the domestics is reasonable, and we only present results from this latter model. We report the means for the individual admixture proportions, $\hat{q}_i$, and their 95% probability intervals. We also report the point estimates of gene frequencies in the parental populations. Using this we can calculate the statistic $\Delta\mu^2$ of Goldstein *et al*. (1995), and thereby estimate the divergence time of the parental populations.

Results

Analysis of genetic variation

A significant departure from Hardy–Weinberg expectations was observed in house cats, subdivided into the three samples, and wild-living cats grouped by the traditional classification (Table 2a). In an analysis where θ and F_{IS} were simultaneously estimated, F_{IS} was positive in both

Table 2(a) Summary of the *F*-statistics and analogues for the three house cat samples and the wild-living cats grouped according to the traditional classification criteria. *P* is the proportion of permutations of genotype frequency tables that are as extreme or more extreme as that observed, performed with Fstat as described in the text

	F_{IS}	θ	ρ	*P*
Classification				
House cats	0.137 (0.086–0.190)	0.003 (−0.009–0.015)	−0.016 (−0.009–0.046)	0.0088
Traditional (TC)	0.125 (0.073–0.173)	0.042 (0.026–0.061)	0.056 (0.034–0.120)	< 0.0001

Table 2(b) The distribution of θ (below diagonal) and Nei's D (above diagonal) in pairwise comparisons among all the five morphological groups and the three house cat samples. Genotypic permutations were performed on each pair as above, and the corresponding values of θ are underlined for $P < 0.01$, and marked in bold for $P < 0.001$

	n	Wild	Wild-Hybrid	Hybrid	Hybrid-Dom.	Domestic	House cat Scotland	House cat Surrey	House cat Yorkshire
Wild	27		0.075	0.005	0.045	0.172	0.4490	0.406	0.478
Wild-Hybrid	4	0.0261		0.080	0.045	0.246	0.3750	0.412	0.448
Hybrid	61	0.0028	0.0289		0.043	0.156	0.3860	0.336	0.399
Hybrid-Domestic	19	0.0195	0.0099	0.0187		0.084	0.1920	0.217	0.252
Domestic	91	**0.0599**	0.0719	**0.0526**	**0.0263**		0.0810	0.106	0.127
House cat, Scotland	18	**0.1158**	**0.0917**	**0.0975**	**0.0574**	**0.0231**		−0.007	0.022
House cat, Surrey	10	**0.1370**	**0.1120**	**0.1178**	**0.0615**	**0.0255**	−0.0099		0.015
House cat, Yorkshire	46	**0.1412**	**0.1283**	**0.1219**	**0.0793**	**0.0405**	0.0032	0.0051	

house cats and wild-living cats with a highly significant excess of homozygotes. To test for null alleles the correlation between the number of nonamplifying samples for each locus and F_{IS} was examined. The museum samples, whose failure to amplify could be attributed to the quality of the material, were left out of the analysis. The wild-living (non-museum) cats and house cats were examined separately. The correlation was 0.591 ($p \sim 0.090$) for the house cats and −0.418 ($p \sim 0.26$) for the wild-living cats. The locus that has the largest effect on both analyses is Fca35, for which no bands were obtained in five house cats, in 21 wild-living (non-museum) cats, and in all 13 museum specimens. In the house cat sample Fca35 has the largest value of F_{IS} (0.29), whereas in the wild-living cats F_{IS} is the lowest at 0.06. Thus, in conclusion, it seems more likely that Fca35 is generally more difficult to amplify, and sensitive to the quality of material, rather than that it contains specific null alleles.

The presence of private alleles as a tool to assess genetic introgression by one species with another has been used in a number of previous studies (e.g. Gottelli *et al*. 1994) and was assessed here. Because alleles at low frequency could by chance be sampled in one group but not the other, for descriptive purposes we have used a threshold frequency of 0.05 above which, based on the sample sizes, we can be confident of frequency differences between different groups. Within the house cat samples no private alleles at frequencies greater than 0.05 were present, whereas for wild-living cats two loci (Fca90 and Fca96) possessed private alleles at frequencies greater than 0.05. For locus 90, allele 91 (the shortest allele, Table 1) had frequency 0.062 (27/436) and for locus Fca96, allele 224 had frequency 0.32 (142/450). Interestingly, the frequency for this allele within the group whose TC score is closest to domestic cats is the lowest of the classes plotted. A comparison of heterozygosity (H_E), as shown in Table 1, shows no clear differences between the house cat and wild-living cat samples.

Analysis of the allele frequency differences among the three populations of house cat (summarized in Table 2a) shows little differentiation either as measured by θ or ρ although the genotype frequency differences are statistically significant. By contrast there is a larger degree of differentiation among the groupings of the wild-living cats based on the traditional classification. The value of θ, at around 0.04, is however, still low, but the information from pairwise comparisons (Table 2b) indicates that there is quite substantial differentiation present. For example pairwise θ-values between wild-living cats classed as domestics and those classed as wild, wild/hybrid, and hybrid are all greater than 0.05, while comparisons within the latter three groups are all less than 0.03. The pairwise values of Nei's D also follow a similar pattern. Pairwise values of θ between any of the three house cat groups and the wild, wild/hybrid, and hybrid categories are all greater than 0.09, while comparisons between the domestic category and the house cat samples are less than 0.05. The values of Nei's D reflect the same pattern. Comparisons among the

Table 3 (a) Mantel multiple regression analysis of the pairwise genetic distances among the wild-living cats in Scotland. The regression coefficients, b, and their probability value, p, are shown in the table. The abbreviations used for the variables are: **GD** — the Euclidean distance between the coordinate points of the cat locations; **ED** — the absolute difference in elevation between the two individuals; **TCD** — the absolute difference in ranking between two individuals based on the traditional classification; **CD** — the Euclidean distance between the coordinate positions of two individuals based on the five morphological scores; **MCD** — the absolute difference in composite morphological score. The sample size used in this table is 194 cats. The Bonferroni-corrected critical value for interpreting the marginal regressions is 0.01

		R^2	GD	ED	TCD	CD	MCD
Full Model	b	0.055	0.082	0.016	−0.016	0.175	0.056
	p	0.001	0.002	0.616	0.572	0.001	0.130
Final Model	b	0.053	0.083			0.210	
	p	0.001	0.001			0.001	
Marginals	b		0.095	0.008	0.096	0.214	0.191
	p		0.001	0.776	0.001	0.001	0.001

Table 3(b) Multiple regression analysis of the genetic distances between wild-living cats and the house cats. The abbreviations used for the variables are: **GX** and **GY** — geographical coordinates; **EL** — elevation; **TC** — traditional classification score; **MC** — composite morphological score; **TS** — tail shape; **DS** — dorsal stripe; **RS** – rump spots; **TT** — tail tip colour; **PC** — paw colour. The sample size used in this table is 173 cats. The Bonferroni-corrected critical value for interpreting the marginal regressions is 0.005

		R^2	GX	GY	EL	TC	MC	TS	DS	RS	TT	PC
Full Model	b	0.241	−0.021	0.161	0.233	−0.235		0.077	0.019	0.039	−0.062	0.056
	p	0.001	0.799	0.031	0.004	0.082		0.478	0.831	0.704	0.416	0.469
Final Model	b	0.214			0.190	−0.388						
	p	0.001			0.007	0.001						
Marginals	b		0.142	0.216	0.263	−0.423	0.372	0.365	0.300	0.289	0.130	0.194
	p		0.056	0.007	0.003	0.001	0.001	0.001	0.001	0.002	0.080	0.010

three house cat samples give very small values of θ and Nei's D.

Mantel multiple regression analysis of genetic distances between cats

The results of the Mantel multiple regression analysis of 194 cats for which there was no missing information with respect to the variables of interest is displayed in Table 3a. All variables, apart from difference in elevation, showed a significant relationship with allele sharing distance when analysed individually. The reduced model consisted of just two variables, the Euclidean morphological distance, with the highest regression coefficient, and the geographical distance. However, this model only explains a small proportion of the variation in genetic distances ($R^2 = 0.05$). Thus, although the effect is relatively weak, there is evidence that morphologically distinct cats tend to be genetically distinct and cats that are geographically separated are also genetically distinct but to a smaller degree. Importantly, however, the effect of morphology is not confounded by geographical effects. Indeed, the regression coefficients remain similar in the marginal, full, and reduced models indicating that there is little relationship between geography and morphology.

Multiple regression analysis of genetic distances of wild-caught cats from house cats

We also performed multiple regression analysis of the allele sharing distance between wild-caught cats and the house cats. We used the east–west and north–south coordinates of the wild-caught cats, their elevation, traditional classification score and the morphological scores as explanatory variables. A total of 173 cats were used that had no missing values for these variables. With a Bonferroni corrected significance level of 0.005, elevation, traditional classification, composite morphological score, tail shape, dorsal stripe, and rump spots were significantly correlated with the allele sharing distance. Because the composite morphological score is a linear combination of the five morphological scores it cannot be analysed together with them. Therefore, stepwise regression was carried out in two ways: (i) with the composite score and without the five individual scores; or (ii) with the five scores alone. The best fitting model for the allele sharing distance was $y = 0.19\text{EL} - 0.388\text{TC}$ ($R^2 = 0.214$, $P <= 0.001$) (Table 3b). This was obtained using either set of morphological variables. Thus, using this genetic distance, there is evidence that the wild-living cats most genetically distinct from house cats tend to live at higher elevations and tend

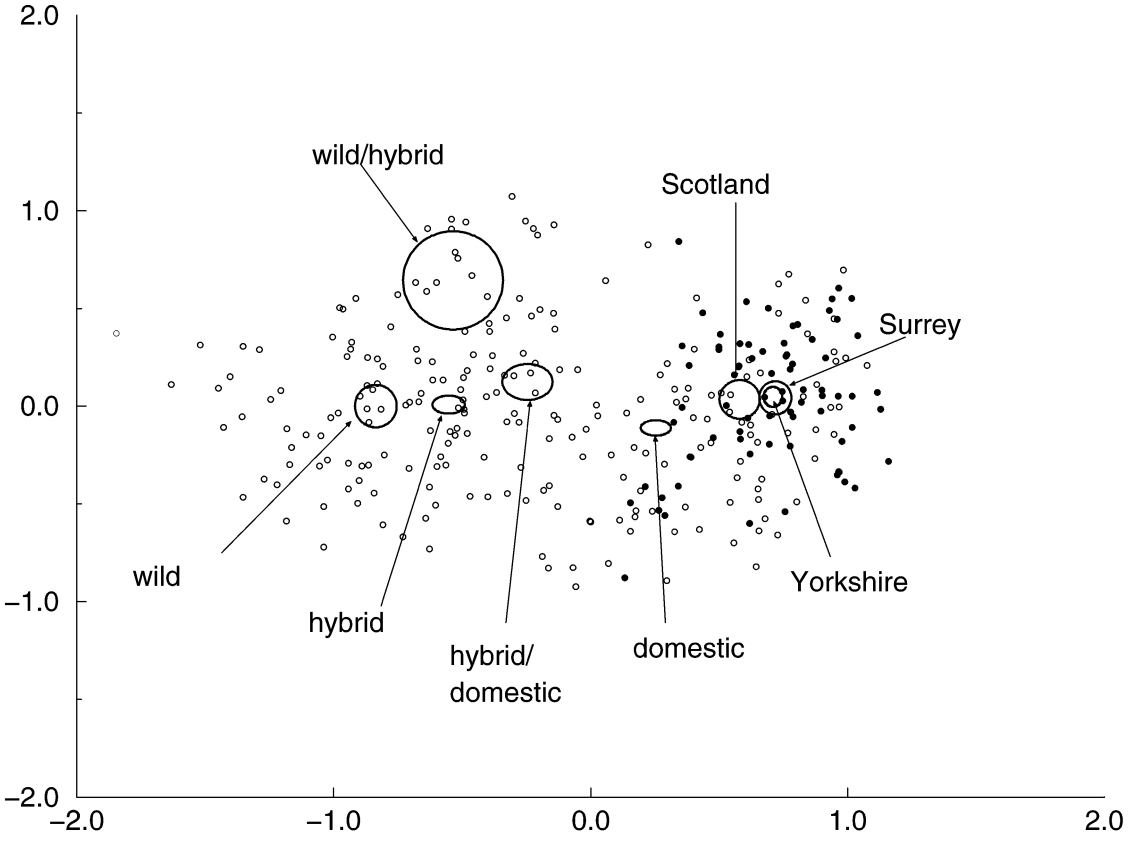

Fig. 3 Plot of the scores of individual cats (wild-living and house) on the first two axes of the ordinations of pairwise allele-sharing distances. The house cats are shown by filled circles. Also shown are standard error ellipses for the different groups of cats scored by the traditional classification on the first two axes of the ordinations.

to conform to the traditional morphological descriptions of wildcats. In this analysis it appeared that tail-tip colour and paw colour are poorly correlated with genetic distance.

Ordination

The coordinate positions of individual wild-living cats, derived from the ordination of their allele sharing distances, are highly correlated with their morphology (Fig. 3). The two principal axes explain only 7% of the variation. However, although this value is small, there is a strong separation of mean values between many of the groupings (Fig. 3), as judged by the standard errors. The three house cat samples occupy a very similar portion of the plot. With the exception of the wild-hybrid category, which is represented by only four individuals, and hence has large standard errors, there is a good match between traditional classification score and position on the *x*-axis. The Spearman's rank correlation between the score of individual wild-caught cats on the *x*-axis and their score by the traditional classification is 0.65 ($n = 199$; $P < 0.001$). The correlation with the composite morphological score is −0.64 ($n = 211$; $P < 0.001$). The correlation between the composite morphological and traditional classification scores is −0.79 ($n = 195$; $P < 0.001$). The geographical distribution of wild-living cats categorized according to the *x*-axis score is shown in Fig. 1.

Density estimation

There is evidence of two groups in the scores of individuals on the first axis of the ordination of wild-living cats (Fig. 4a,b). The scores on the first axis (*x*-axis) were divided into all wild-living cats (Fig. 4a) and all house cats (Fig. 4b) and these were analysed independently. Densities were estimated for these scores using the logspline method of Kooperberg & Stone (1991). The best-fitting model for the wild-living cats gives a bimodal distribution. The model uses five knots with log-likelihood −213. The next best fitting (by the BIC criterion) is unimodal with three knots and log-likelihood −234. Because these are nested models it is appropriate to test them by a likelihood ratio test, which gives $\chi^2 = 42$ with two degrees of freedom, $P < 0.001$. Thus, overall there is strong evidence that the

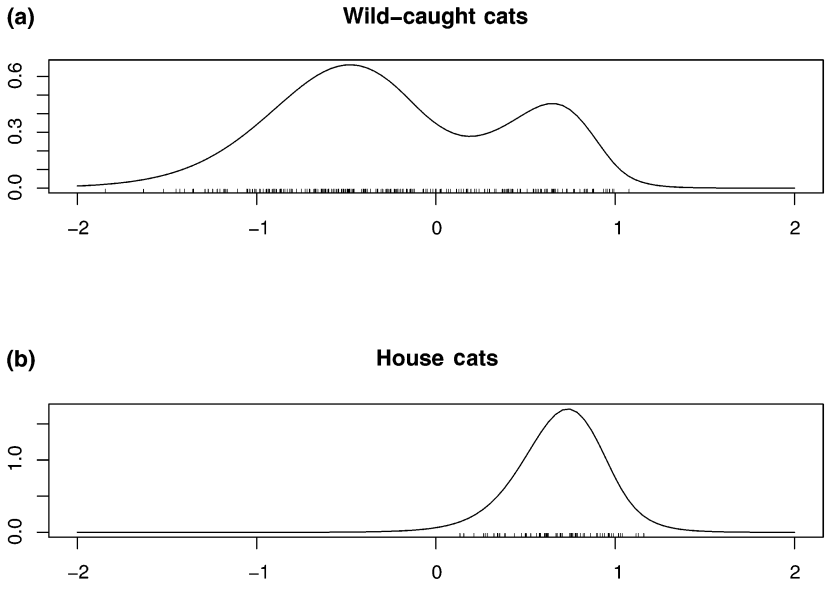

Fig. 4 (a) Density curve based on the scores of the wild-living cats on the first axis of the ordination. (b) Density curve based on the scores of the house cats. The positions of individual cats are shown along the *x*-axis as a 'rug' of points.

scores of the wild-living cats are bimodal. The scores of the house cats are unimodal, and the mode at higher scores of the wild-living cats coincides closely with that of the house cats.

The correlations between morphological scores and the position on the first axis are unlikely to be solely due to the mixture of discrete groups (wildcat and domestic) without the occurrence of hybrids. This is because there are still strong correlations between morphology and ordination score in the set of cats with a score on the first axis less than zero (TC score: Spearman's rho = 0.372, $n = 122$, $P < 0.01$; MC score: Spearman's rho = −0.377, $n = 132$, $P < 0.01$). Thus, overall, there is evidence of two groups of wild-living cats, but within these groups there are still strong correlations between morphology and genetics.

Genetic mixture analysis

Genetic mixture analysis (Pritchard *et al*. 2000) supports the results of the ordinations in suggesting the existence of two groups of wild-living cats, with intermediates. In Fig. 5 the means ($\hat{q}$) of the posterior distribution of q are plotted for each individual, where q is the proportion an individual's genome that comes from the wildcat population. The posterior distribution (density) is the probability distribution of q, given the data. This distribution is marginal to (integrated over) all the possible gene frequencies in the two parental populations. In other words, the method simultaneously estimates parental frequencies and q, and in order to obtain a posterior density for q that is independent of the parental frequencies it is necessary to calculate an average of the posterior density of q over all possible frequencies, weighted by their posterior density. Also shown in Fig. 5 are the 95% limits of q (i.e. an interval within which q lies with probability 0.95, given the data). Dashed lines indicate the 95% limits for the 13 museum skins. It can be seen that there is strong evidence of two groups. The proportion of wildcat individuals ($\hat{q} \sim 1$) appears to be larger than the proportion of domestics ($\hat{q} \sim 0$). The more intermediate individuals tend to have very wide probability limits for q, which often encompass at least one end and sometimes both ends of the 0–1 range. It is possible that these intermediate individuals are actually either domestic or wildcat, but have little information on q. Using the arbitrary bands, 0–0.1, 0.1–0.9, and 0.9–1, to define domestics, intermediates, and wildcats, we observe 39, 96, and 95 cats in these respective categories.

There is a strong correlation between $\hat{q}$ and the ordination score. A comparison of the logistic transformed q-values with the ordination scores on the first axis for each individual is shown in Fig. 6. The individuals have been shaded according to the traditional classification scores. In addition, wider open circles indicate the museum skin samples. Two of the skin samples were not included in the ordination analysis because they had too few loci to guarantee that all pairs of individuals could be compared (as explained in Methods). There is a broadly linear relationship between $\hat{q}$ and the ordination score, indicating that these parametric and nonparametric methods both pick out an underlying feature of the data. In addition, the traditional classification score of the cats correlates with their values of $\hat{q}$, as it does with the ordination score. Cats that have a TC-domestic phenotype tend to have low values of $\hat{q}$ and those that have a TC-wildcat phenotype have high values of $\hat{q}$. An interesting feature is that there appears to be an appreciable bias in the distribution of taxonomic types, whereby wild-living individuals judged to be wildcat

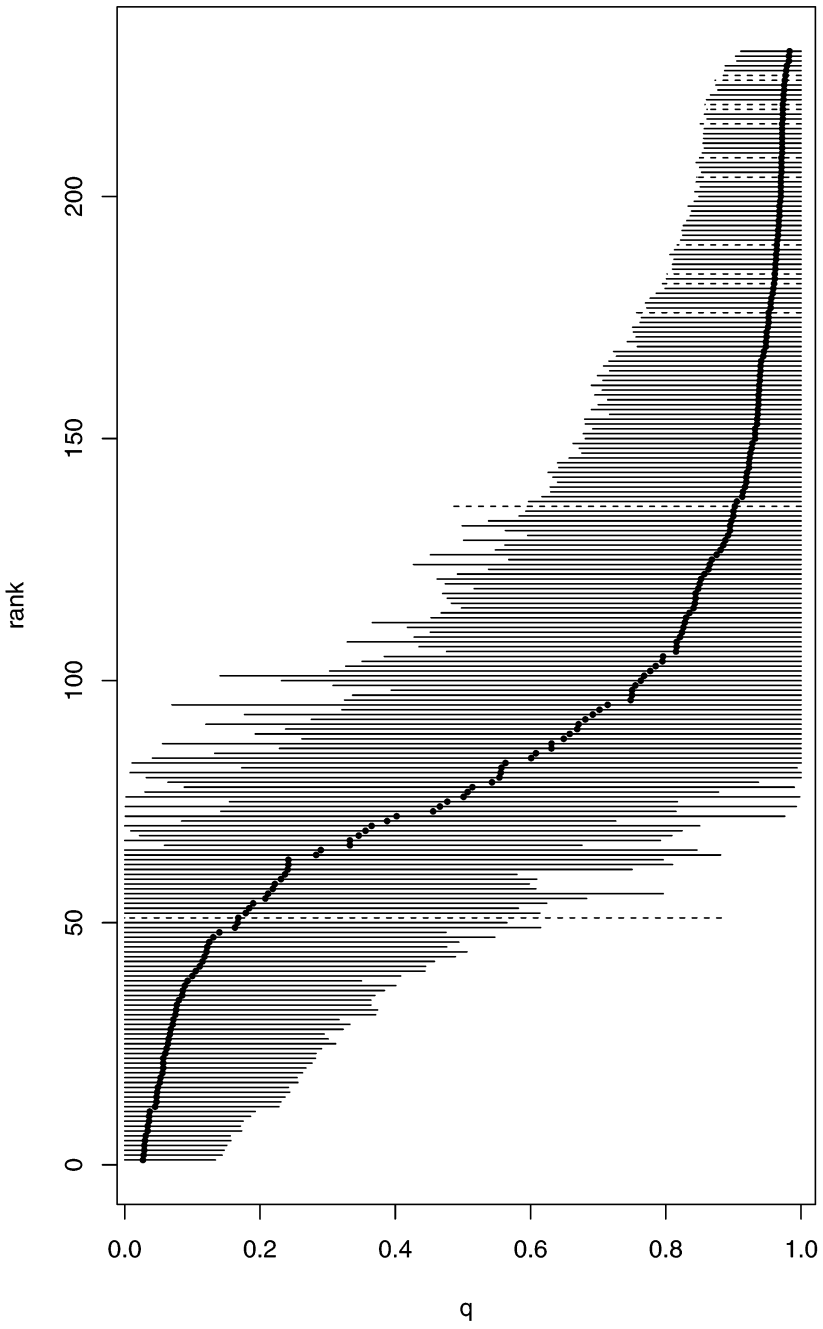

Fig. 5 This figure illustrates the distribution of $\hat{q}$ among individuals. The values of $\hat{q}$ have been ranked and the ranks are plotted against $\hat{q}$. Also shown are lines giving the 95% equal-tail posterior probability intervals for each individual. The dotted lines refer to museum specimens.

are always in the nondomestic genetic group, whereas wild-living cats that are judged to have a domestic phenotype are frequently found in the nondomestic genetic group. The majority of cats with values of $\hat{q} < 0.5$ are scored as TC-domestics with a few TC-hybrid/domestics and TC-hybrids. In contrast, those with values of $\hat{q}$ higher than 0.5 show a great diversity of taxonomic scores, including many scored as TC-domestics. Indeed, one individual scored as a TC-domestic has a higher value of $\hat{q}$ than any TC-wildcat cat.

There is tentative evidence of a decline in $\hat{q}$ over the last 50 years among the cats that conform to the wildcat phenotype. The earliest museum skin dates from 1945, a further seven date from around 1960, and three date post-1970 (two are undated). The number of museum skins is small, but we can also include the 17 wild-living cats obtained in the SNH survey that were judged by the traditional classification to be wildcat because these possess the same characteristics that would have been used to identify the museum specimens. We assumed a single date of 1990

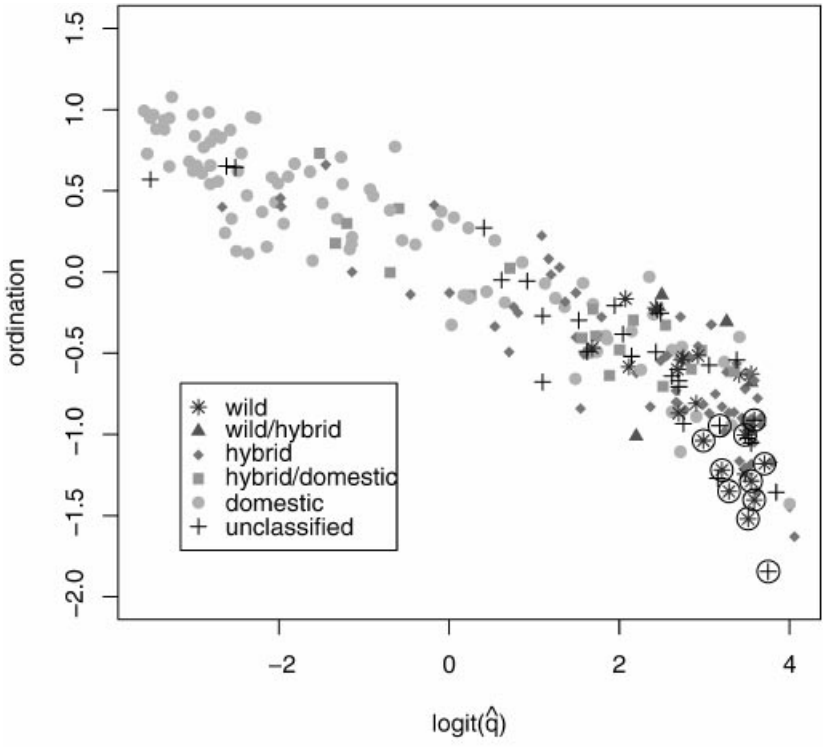

Fig. 6 Plot of the ordination scores against logit $\hat{q}$. The morphological types are indicated by colour shading, as given in the legend. The wider unfilled circles refer to museum specimens.

for these latter cats. The smallest value of $\hat{q}$ among the eight pre-1970 cats is 0.95, whereas of the 20 post-1970 cats only six have $\hat{q}$ greater than this. There is a significant correlation between $\hat{q}$ and date, using a conservative nonparametric test (Spearman's Rho = 0.442; p ~ 0.02). Thus, although the sample size is small, there is some evidence to suggest that there has been recent introgression among the cats that are categorized as wildcats by traditional taxonomic criteria.

There is little evidence of reduced variability in the posterior mean gene frequencies of the inferred wildcat parental group by comparison with the domestics. The mean heterozygosity (with standard error) over all loci for the domestic cat group is 0.749 (0.040), and for the wildcat group is 0.688 (0.038). The pairwise genetic differentiation between the two groups, using the posterior mean gene frequencies, is similar to that between the house cat samples and the wildcat samples defined by the traditional classification, given in Table 2b. The mean estimate of F_{ST}, Nei's D, R_{ST}, and $\Delta\mu^2$ (with 95% bootstrap confidence limits from resampling loci) are 0.12 (0.07–0.17), 0.42 (0.19–0.68), 0.19 (0.13–0.25), and 3.53 (1.67–5.80), respectively. The estimate of sampling error ignores the statistical error in inferring the parental gene frequencies using the assumed model of admixture, but this is likely to be smaller than the variance in the estimates among loci. By making a number of assumptions Nei's D and $\Delta\mu^2$ can be used to estimate divergence times, and these assumptions, the estimates, and their reliability are discussed below.

Discussion

An important feature of this study is that there is controversy over what constitutes a wildcat, and whether wildcats can be defined purely by morphological criteria (Daniels *et al*. 1998; Kitchener 1998). As a consequence we have no uncontroversial baseline population of wildcats with which to compare our measurements and estimate the degree of introgression between wildcats and domestics. This phenomenon, where an introduced population hybridizes introgressively with a native population, is relatively common (Rhymer & Simberloff 1996). As pointed out by Estoup *et al*. (1999), it is difficult to make estimates of the degree of admixture when the gene frequencies in the native population prior to admixture are unknown. An additional problem, peculiar to the wildcat, is that there is no consensus on the timescale of hybridization. Thus, we present here evidence of two groups of cats, and it is tempting, from the generally accepted historical record to suggest that the nondomestic group is a remnant of the postglacial *Felis silvestris* population identified from subfossil remains. However, given that the domestic cat has been present in Britain for over 2000 years, there is no clear historical evidence to rule out the possibility that the nondomestic group is actually derived largely or

completely from domestic cats, and the observed genetic differentiation is caused by drift over a short period in a small, previously isolated, feral group. Below we review the genetic results described in this paper, suggest the most likely explanations for the patterns observed, and discuss the implications for the conservation of the Scottish wildcat.

There are no alleles at frequencies greater than 5% in the 74 house cats that are not found in the 230 wild-living cats, and only two such alleles present in the wild-living cats that are not found in the house cats. The lack of private alleles in the house cats is easily explained if there has been substantial gene flow from the domestic to the wild-living population, because these alleles would then be in the wild-living population. As discussed below, the three house cat populations have very similar gene frequencies, with F_{ST} estimated to be 0.0027, and thus a similar explanation for the lack of private alleles in the wild-living cats is unlikely because otherwise we might expect more differentiation in the Kirriemuir sample. The existence of one allele in the wild-living cats that is at relatively high frequency but which is absent in all three samples of house cat suggests that the two groups are differentiated from each other with little gene flow from the wild-living cats to the house cats. However, it is possible that other house cat populations in northern Scotland possess this allele at higher frequencies, and further sampling is necessary before any firm conclusions can be drawn. This relative lack of private alleles is unusual for a population that has been assigned species or subspecies status. For comparison, studies of hybridization using microsatellites have shown substantial numbers of private alleles at the level of species (e.g. the Ethiopian wolf and dogs, Gottelli *et al*. 1994; red and sika deer, Goodman *et al*. 1999; coyote and grey wolf, Roy *et al*. 1994) or subspecies (e.g. taurine and zebu cattle, MacHugh *et al*. 1997). It is possible that the ancestral wildcat was not substantially differentiated from the domestic cat, or that it had less variation (i.e. with fewer alleles there is less of a chance of observing an allelic type that is not represented in the other group).

The observation of a high value of F_{IS} in the house cats is probably explained by the fact that many of the individual cats would come into the veterinary practices from a large catchment area, and would have relatively independent ancestries. In the wild-living cats the high F_{IS} is probably explained by two factors. First, as discussed later, there is evidence of two genetic groups of wild-living cats and the cats in most of the morphological categories, particularly those scored as domestics, hybrid-domestics, and hybrids, appear to be derived from both groups. Second, as also discussed later, there is evidence of geographical differences in gene frequencies of the cats, which will also lead to homozygote excess if geographically diverse individuals are grouped together.

The house cats, taken from three widely separated areas in the British Isles, have very similar gene frequencies, and this contrasts with the observation of large regional variation in frequencies of the classical pelage polymorphisms in the domestic cat (Pontier *et al*. 1995; Ruiz-Garcia 1997; Ruiz-Garcia 1999). In many studies of cat pelage polymorphisms, selection has often been inferred as the agent giving rise to the patterns (e.g. Metcalfe & Turner 1971; Hedrick 1985; Robinson 1987; Pontier *et al*. 1995), particularly as a consequence of human preferences and widespread neutering of pet cats. On a broader geographical scale, Todd (1977) has suggested that the distribution pattern of the coat colour alleles in Europe reflects ancient trading routes from the centres where the coat colour mutations arose and that the distribution did not reflect human preferences for particular coat colours. It could, however, be argued that the initial local increase in frequency of mutations was aided by human selection for novelties. Given this uncertainty, it is questionable whether studies of the classical pelage polymorphisms can predict patterns of variation observed at microsatellite loci, which are more likely to reflect the demographic history of populations with less influence of selection.

There is good evidence that wild-living cats that are grouped according to morphological criteria show appreciable differences in gene frequencies. We have used the statistics θ, ρ, and Nei's D to quantify these differences on a widely used scale. The overall degree of differentiation among all five morphological categories at around $\theta = 0.04$ is not large, and the pairwise difference between the two extreme morphological groups, TC-wildcats and TC-domestics, at $\theta = 0.06$ is not substantially larger. Yet this degree of differentiation is much greater than that observed between house cats sampled from different parts of Britain, and strongly suggests that there is a phenomenon in need of explanation. Furthermore, there are large gene frequency differences between the TC-wildcats and the house cats, with values of θ varying between 0.11 and 0.14, whereas similar comparisons for the TC-domestics vary from 0.02 to 0.04. Overall, however, the degree of differentiation observed between the morphological classes or between the different morphological classes and the house cat sample are not large, and are, for example, comparable with, and often smaller than, those observed in different dog breeds (Pihkanen *et al*. 1996; Morera *et al*. 1999; Zajc & Sampson 1999), and less than observed among populations of the channel island fox *Urocyon littoralis*, known to have been isolated from each other for thousands of years (Goldstein *et al*. 1999; Wayne *et al*. 1991). Generally, when populations become differentiated due to recent drift this is accompanied by detectable differences in heterozygosity and numbers of alleles between different populations (e.g. Lade *et al*. 1996), and there are no such discrepancies in the cat groups. However, it is difficult to

rule out the possibility that these differences have been lessened by recent introgression. The problem of interpretation of the observed differentiation is discussed in more detail below when the results of the genetic mixture analysis are considered.

There is a possibility that the correlations between morphology and gene frequencies are confounded with geographical effects. However, the Mantel tests show that, while there is a geographical effect (cats from the same area have smaller genetic distances), there is still a highly significant correlation between morphological distances between wild-living cats and genetic distances, independent of this. It should be noted that these genetic distances are a composite, averaged over a number of different loci, the number of which often varied in different comparisons due to missing data (see Table 1). However, while these missing data will increase the variance of any regression with genetic distance, it is unlikely to give rise to any specific bias. The consensus between the ordinations and the pairwise gene frequency analyses (e.g. Fig. 3 and Table 2) and between the ordinations and the genetic mixture analysis (Fig. 6) suggests that there are no systematic effects caused by missing values peculiar to the analyses with genetic distances.

In general, in the wild-living cats there are no clear-cut relationships between genetic distance from house cats and latitude, longitude or elevation. Regression of geographical coordinates with distance from house cats shows no strong result, although there is a weak suggestion of a latitudinal effect (cats genetically more different than house cats tend to be found further north). Regression of distance from house cats shows an effect of elevation. However, in the analysis of pairwise distances there was no evidence that cats that were more different in elevation were genetically more distinct. Discrepancies between results using the genetic distance between individuals and those based on distances from house cats can be explained if the differences between wild-living cats in terms of their genetic distance from house cats are only a small proportion of the total genetic distance between cats. This is consistent with the low variance explained in the first axis of the ordination. In addition, the significance level of results from the regressions using distances of wild-living cats from the house cats should be treated with a degree of caution given that the points are not independent because of spatial correlation. Regression will always be problematic with genetic data because of the nonindependence due to genealogical structure (although this is generally only appreciated with taxonomic data, Harvey & Pagel 1991). This problem is circumvented by the Mantel multiple regression, which can be viewed as an alternative to the use of the comparative method (Legendre *et al*. 1994).

The density analysis provides support for bimodality in the scores of wild-living cats on the first axis from the ordination of genetic distances. Initial evidence of two groups of wild-living cats was presented by Daniels *et al*. (1998), based on measurements of intestine length. The genetic mixture analysis of Pritchard *et al*. (2000) supports the conclusions drawn from the ordinations. The method requires that the number of populations is specified in advance, and we have chosen to specify two populations. Specifying the number of populations does not, however, imply the result. The genetic mixture analysis works by explaining the observed degree of intra- and interlocus gametic disequilibrium. If the entire sample was in equilibrium, then the observed genotype distribution could be obtained by drawing genotypes at random from one gene frequency distribution. In this case, where we have assumed two contributing populations, the marginal gene frequencies of the two parental populations would be very similar, and the inferred mixture proportions for each individual would have a very flat distribution with probability limits spanning the range 0–1. This is not observed in these data; in particular, the 95% credible limits for $\hat{q}$ plotted in Fig. 5 are, for many individuals, close to 0 or 1. The advantage of the genetic mixture analysis over the ordination is that it is based on a specific model, and allows estimation of probability intervals for the parameters. Nevertheless, ordination is a frequently used approach in population genetics, and it is useful to compare the results from both approaches. It is encouraging that there is a strong correlation between the value of $\hat{q}$ for each individual and their ordination score.

Both approaches suggest that there are two groups of wild-living cats. The case of the domestic group is probably straightforward to interpret. In the ordinations the peak of one mode from the density analysis coincides with that of the house cats. In the genetic mixture analysis we used point priors for the house cats, forcing them to have all their genotype sampled from this group. There are a number of wild-living cats that had 80–100% of their genotype sampled from this group, and therefore, it is reasonable to assume that their genotypes are exclusively or predominantly drawn from the domestic gene pool, with little introgression from the other group.

Interpretation of the second group is more contentious. For example, is it evidence of a separate population that has no history of introgression with domestic cats? If most cats showed evidence of hybridization they would have had intermediate values of $\hat{q}$ (see, e.g. Fig. 2 of Pritchard *et al*. 2000). The existence of around 50 wild-living cats with $\hat{q}$ greater than 0.95 and lower 95% limit greater than 0.8 strongly suggests that a large proportion of wild-living cats have been drawn at random from a gene pool that is distinct from the domestic cat pool, and in which the genotype frequencies are in approximately equilibrium proportions. However, this does not necessarily mean that this gene pool contains no introgressed domestic cat genes. The

microsatellite loci are either on separate chromosomes or are widely spaced on the same chromosome (Menotti-Raymond *et al*. 1999), and linkage disequilibrium will be negligible after a few generations of back-crossing (Goodman *et al*. 1999). Therefore, if hybridization has continued for many generations a sizeable portion of the wildcat gene pool could come from the domestic population. Thus, in answer to the question posed above, the existence of this second group of cats does not necessarily mean that we have identified a pure wildcat population with no introgression, but that we have identified a group of cats that show little evidence of very recent domestic cat ancestry.

The observation that many cats in this second group show a mixture of phenotypes is good evidence that they do contain introgressed domestic cat genes. There are significant correlations between morphology and ordination score within this second group. As can be seen from Fig. 6, cats that have $\hat{q}$ greater than 0.9 (~2 on a logit scale) show the full range of phenotypes, from TC-domestic to TC-wildcat. It is possible that many of these phenotypes may be due to the presence of a few introgressed coat colour genes, and that an individual homozygous for one of the many coat colour genes that are present in domestic cats (Robinson 1977) would automatically be scored as domestic. This biased effect from coat colour morphology may also explain why the majority cats with intermediate values of $\hat{q}$ are scored as domestics, particularly at $\hat{q} < 0.5$. It is possible that there is selection for particular coat colour phenotypes in the wild-living cats, and some of the coat colour genes, particularly the allele at the nonagouti locus that codes for black coat colour (Robinson 1977), may have always been present in the ancestral wildcat population. As discussed in Daniels *et al*. (1998), the presence of coat colour variants makes the definition of a wildcat based purely on pelage markings somewhat problematic. However, as demonstrated in this paper, in the current sample, strict possession of the classical markings, as judged by A.C.K, appears to be sufficient to place an individual in the nondomestic group. In future studies it may be worthwhile to include data from skulls including cranial volume, skull characters and measurements, which are highly correlated with the pelage characters used in this paper (Ward & Kitchener, unpublished data).

We have presented tentative evidence to suggest that there has been appreciable introgression over the last 50 years, based on the museum skins. The result is not highly significant ($p \sim 0.02$), although a Spearman's test is conservative. It is based on only eight samples collected earlier than 1970. These may have unusual leverage, and the result may not be supported with additional samples. Other than type I error, another alternative explanation for this observation is that the SNH sample has undue influence and there may be obscure reasons for expecting those cats classified as wildcats in this sample to have lower values of $\hat{q}$. Clearly, before stronger statements are made, it would be worthwhile making a more extensive study of museum specimens.

From the genetic data it is difficult to rule out the possibility that the genetic divergence between the groups is actually relatively recent, rather than reflecting postglacial isolation of the wildcat in Britain. Under the model assumption from which it was derived, $\Delta\mu^2$ (Goldstein *et al*. 1995; Zhivotovsky & Feldman 1995) has an expected value equal to twice the number of generations since the populations diverged multiplied by mutation rate multiplied by the variance in the number of repeats changed by each mutation. The expected value of Nei's D (Nei 1972, 1978) is the same, but without the variance term. Assuming a mutation rate of 5×10^{-4} (Goldstein *et al*. 1995), the single step mutation model, and a mean generation time of 3 years (i.e. the mean age of the parents of individuals when they are born) we obtain an inferred time of divergence of 10 590 years with 95% limits (5010–17400) for $\Delta\mu^2$. The corresponding value for Nei's D is 1260 (570–2040). There are a large number of caveats to consider. With uncertainty about the mutation rate, the limits are likely to be broader than these. The estimate based on Nei's D assumes that all mutations are novel and will underestimate divergence times when this is violated (Goldstein *et al*. 1995). It is generally assumed that there is a large degree of homoplasy in the microsatellite mutation process (Estoup & Cornuet 1998), but the degree to which the variance is greater than 1, as assumed here for $\Delta\mu^2$ is unclear. The model on which these calculations are based assumes that a population of size n splits instantaneously into two populations each of size n. With these data the model is likely to be violated in two ways: (i) the two populations are unlikely to be equal; and (ii) there is evidence that, even accounting for recent admixture, the inferred gene frequencies in the wildcat group have been affected by past introgression from domestic cats. The effect of unequal population size is to artificially increase the estimated divergence time (Charlesworth 1998; Hedrick 1999; Gaggiotti & Excoffier 2000). The effect of past gene flow, by making the gene frequencies more similar, will tend to reduce the estimated divergence time. Depending on the relative importance of these two processes the estimates could be biased upwards or downwards to a substantial degree.

In conclusion, we present here evidence that the wild-living cats of Scotland consist of a diverse set of individuals containing a mixture of domestic and wildcat genes. Many of these individuals appear to be drawn from a gene pool with genotypes at approximately equilibrium frequencies that are distinct from those of the domestic cat, but which have probably been influenced by past introgression. We can offer no diagnostic test of a true wildcat that contains

no domestic cat ancestry. In fact the evidence suggests that such cats are unlikely to exist. At the population level, however, there is strong evidence of a unique group of individuals that are different from domestic cats and which may be worthy of legal protection.

Acknowledgements

This study was funded by Scottish Natural Heritage and the Institute of Zoology, Zoological Society of London. Genetic analysis of the Scottish cats was initiated by Bob Wayne. David Balharry helped M.J.D. to collect the samples with the aid of people listed in Balharry & Daniels (1998). The house cat control samples were provided by Andrew Greenwood, the Thrums veterinary practice in Kirriemuir, and a practice in Surrey. Marylin Menotti-Raymond, National Cancer Institute, Frederick, Maryland, USA, kindly provided the microsatellite primer sequences. We thank Jennifer Ward for providing Fig. 2. The interpretations of the results presented in this paper are those of the authors and do not necessarily constitute SNH policy.

References

Amos B, Hoelzel AR (1991) Long-term preservation of whale skin for DNA analysis. *Report of the International Whaling Commission Special Issue*, **13**, 99–103.

Balharry D, Daniels MJ (1998) Wild living cats in Scotland. *Scottish Natural Heritage Research, Survey and Monitoring Report*, no. 23.

Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, **16**, 113–148.

Barton NH, Hewitt GM (1989) Adaptation, speciation and hybrid zones. *Nature*, **341**, 497–503.

Beaumont MA, Barratt EM, Gottelli D, Bruford MW (2001) An analysis of genetic differentiation in the wild-living cat population of Scotland. *Scottish Natural Heritage Research, Survey and Monitoring Report*, in press.

Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL (1995) High resolution human evolutionary trees with polymorphic microsatellites. *Nature*, **368**, 455–457.

Bruford MW, Hanotte O, Burke T (1998) Multi-locus and single-locus DNA fingerprinting. In: *Molecular Analysis of Populations: a Practical Approach*, 2nd edn (ed. Hoelzel AR), pp. 287–336. IRL Press, Oxford.

Charlesworth B (1998) Measures of divergence between populations and the effects of forces that reduce variability. *Molecular Biology and Evolution*, **15**, 538–543.

Clutton-Brock J (1989) A Natural History of Domesticated Mammals. British Museum (Natural History), London.

Collier GE, O'Brien SJ (1985) A molecular phylogeny of the Felidae: Immunological distance. *Evolution*, **39**, 473–487.

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'. *Journal of Zoology London*, **244**, 231–247.

Daniels MJ, Beaumont MA, Johnson PJ, Balharry D, Macdonald DW (2001) Ecology and genetics of wild living cats in the north east of Scotland and the implications for the conservation of the wildcat. *Journal of Applied Ecology*, in press.

Easterbee N (1991) The Wildcat. In: *The Handbook of British Mammals* (eds Corbet GB, Harris S), pp. 431–437. Blackwell Scientific Publications, Oxford.

Easterbee N, Hepburn LV, Jeffries DJ (1991) *Survey of the status and distribution of the wildcat in Scotland, 1983–1987*. Nature Conservancy Council for Scotland, Edinburgh.

Estoup A, Cornuet JM (1998) Microsatellite evolution: inferences from population data. In: *Microsatellites: Evolution and Applications* (eds Goldstein DB, Schlotterer C), pp. 49–65. Oxford University Press, Oxford.

Estoup A, Cornuet JM, Rousset F, Guyomard R (1999) Juxtaposed microsatellite systems as diagnostic markers for admixture: theoretical aspects. *Molecular Biology and Evolution*, **16**, 898–908.

Gaggioti OE, Excoffier L (2000) A simple method of removing the effect of a bottleneck and unequal population sizes on pairwise genetic distances. *Proceedings of the Royal Society of London, Series B*, **267**, 81–87.

Goldstein DB, Roemer GW, Smith DA, Reich DE, Bergman A, Wayne RK (1999) The use of microsatellite variation to infer population structure and demographic history in a natural model system. *Genetics*, **151**, 797–801.

Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW (1995) Genetic absolute dating based on microsatellites and the origin of modern humans. *Proceedings of the National Academy of Sciences of the USA*, **92**, 6723–6727.

Goodman SJ (1997) R_{ST} Calc: a collection of computer programs for calculating estimates of genetic differentiation from microsatellite data and determining their significance. *Molecula Ecology*, **6**, 881–885.

Goodman SJ, Barton NH, Swanson G, Abernethy K, Pemberton JM (1999) Introgression through rare hybridization: a genetic study of a hybrid zone between red and sika deer (genus *Cervus*) in Argyll, Scotland. *Genetics*, **152**, 355–371.

Goostrey A, Carss DN, Noble LR, Piertney SB (1998) Population introgression and differentiation in the great cormorant *Phalacrocorax carbo* in Europe. *Molecular Ecology*, **7**, 329–338.

Gottelli D, Sillero-Zubiri C, Applebaum GD, Roy MS, Girman DJ, Garcia-Moreno J *et al.* (1994) Molecular genetics of the most endangered canid — the Ethiopian wolf *Canis simensis*. *Molecular Ecology*, **3**, 301–312.

Grant PR, Grant BR (1997) Mating patterns of Darwin's finch hybrids determined by song and morphology. *Biological Journal of the Linnean Society*, **60**, 317–343.

Harvey PH, Pagel MD (1991) *The Comparative Method in Evolutionary Biology*. Oxford University Press, Oxford.

Hedrick PW (1985) Coat variants in cats: gametic disequilibrium between unlinked loci. *Journal of Heredity*, **76**, 127–131.

Hedrick PW (1999) Perspective: highly variable loci and their interpretation in evolution and conservation. *Evolution*, **53**, 313–318.

Hubbard AL, McOrist S, Jones TW, Boids R, Scott R, Easterbee N (1992) Is the survival of European wildcats *Felis silvestris* in Britain threatened by interbreeding with domestic cats? *Biological Conservation*, **61**, 203–208.

Janczewski DN, Modi WS, Stephens JC, O'Brien SJ (1995) Molecular evolution of mitochondrial 12S RNA and cytochrome *b* sequences in the pantherine lineage of Felidae. *Molecular Biology and Evolution*, **12**, 690–707.

Johnson WE, O'Brien SJ (1997) Phylogenetic reconstruction of the Felidae using 16S rRNA and NADH-5 mitochondrial genes. *Journal of Molecular Evolution*, **44**, S98–S116.

Kitchener AC (1998) The Scottish Wildcat — a cat with an identity crisis? *British Wildlife*, **9**, 232–242.

Kooperberg C, Stone CJ (1991) A study of logspline density estimation. *Computational Statistics and Data Analysis*, **12**, 327–347.

Lade JA, Murray ND, Marks CA, Robinson NA (1996) Microsatellite differentiation between Phillip island and mainland Australian

populations of the red fox *Vulpes vulpes*. *Molecular Ecology*, **5**, 81–87.

Langley PJW, Yalden DW (1977) The decline of the rarer carnivores in Great Britain during the nineteenth century. *Mammal Review*, **7**, 95–116.

Legendre P, Lapointe F-J, Casgrain P (1994) Modeling brain evolution from behaviour: a permutational regression approach. *Evolution*, **48**, 1487–1499.

MacHugh DE, Shriver MD, Loftus RT, Cunningham P, Bradley DG (1997) Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and Zebu cattle (*Bos taurus* and *Bos indicus*). *Genetics*, **146**, 1071–1086.

Manly BFJ (1991) *Randomisation and Monte Carlo Methods in Biology*. Chapman & Hall, New York.

Manly BFJ (1994) *Multivariate Statistical Methods: A Primer*, 2nd edn. Chapman & Hall, London.

Masuda R, Lopez JV, Slattery JP, Yuhki N, O'Brien SJ (1996) Molecular phylogeny of mitochondrial cytochrome *b* and 12S rRNA sequences in the Felidae: ocelot and domestic cat lineages. *Molecular Phylogenetics and Evolution*, **6**, 351–365.

Mayr E (1963) *Animal Species and Evolution*. Harvard University Press, Cambridge, MA.

Menotti-Raymond MA, David VA, Lyons LA *et al*. (1999) A genetic linkage map of microsatellites in the domestic cat (*Felis catus*). *Genomics*, **57**, 9–23.

Menotti-Raymond MA, O'Brien SJ (1995) Evolutionary conservation of ten microsatellite loci in four species of felidae. *Journal of Heredity*, **86**, 319–322.

Metcalfe JA, Turner JRG (1971) Gene frequencies in the domestic cats of York: evidence of selection. *Heredity*, **26**, 259–268.

Miller GS (1912) *Catalogue of the mammals of western Europe (Europe exclusive of Russia) in the collection of the British Museum*. British Museum (Natural History), London.

Morera L, Barba CJ, Garrido JJ, Barbancho M, de Andrés DF (1999) Genetic variation detected by microsatellites in five Spanish dog breeds. *Journal of Heredity*, **90**, 654–656.

Nei M (1972) Genetic distances between populations. *American Naturalist*, **106**, 283–292.

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

Nowak RM (1991) *Walker's Mammals of the World*, 5th edn. John Hopkins University Press, New York.

O'Brien SJ, Mayr E (1991) Bureaucratic mischief — recognizing endangered species and subspecies. *Science*, **251**, 1187–1188.

Pennock DS, Dimmick WW (1997) Critique of the evolutionarily significant unit as a definition for 'distinct population segments' under the US Endangered Species Act. *Conservation Biology*, **11**, 611–619.

Pihkanen S, Vainola R, Varvio S (1996) Characterizing dog breed differentiation with microsatellite markers. *Animal Genetics*, **27**, 343–346.

Pocock RI (1951) *Catalogue of the genus Felis*. British Museum (Natural History), London.

Pontier D, Rioux N, Heizmann A (1995) Evidence of selection on the orange allele in the domestic cat *Felis catus* the role of social structure. *Oikos*, **73**, 299–308.

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

Randi E, Ragni B (1991) Genetic variability and biochemical systematics of domestic and wild cat populations (*Felis Silvestris*: Felidae). *Journal of Mammalogy*, **72**, 79–88.

Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences of the USA*, **94**, 9197–9201.

Raymond M, Rousset F (1995) GENEPOP (Version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Rhymer JM, Simberloff D (1996) Extinction by hybridisation and introgression. *Annual Review of Ecology and Systematics*, **27**, 83–109.

Robinson R (1977) *Genetics for Cat Breeders*, 2nd edn. Pergamon, New York.

Robinson R (1987) Mutant gene frequencies in cats of the Greater London area. *Theoretical and Applied Genetics*, **74**, 579–583.

Roy MS, Geffen E, Smith D, Ostrander EA, Wayne RK (1994) Patterns of differentiation in North American wolf-like canids, revealed by analysis of microsatellite loci. *Molecular Biology and Evolution*, **11**, 553–570.

Ruiz-Garcia M (1997) Genetic relationships among some new cat populations sampled in Europe: a spatial autocorrelation analysis. *Journal of Genetics*, **76**, 1–24.

Ruiz-Garcia M (1999) Genetic structure of different cat populations in Europe and South America at a microgeographic level: importance of the choice of an adequate sampling level in the accuracy of population genetics interpretations. *Genetics and Molecular Biology*, **22**, 493–505.

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbour Laboratory Press, Cold Spring Harbour, New York.

Schwarz G (1978) Estimating the dimension of a model. *Annals of Statistics*, **6**, 461–464.

Serpell JA (1988) The domestication and history of the cat. In: *The Domestic Cat; the Biology of its Behaviour* (eds Turner DC, Bateson P), pp. 151–158. Cambridge University Press, Cambridge.

Todd NB (1977) Cats and commerce. *Scientific American*, **237**, 100–107.

Wayne RK, George SB, Gilbert D *et al*. (1991) A morphological and genetic-study of the island fox, *Urocyon littoralis*. *Evolution*, **45**, 1849–1868.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Yalden DW (1999) *The History of British Mammals*. Poyser, London.

Zajc I, Sampson J (1999) Utility of canine microsatellites in revealing the relationships of pure bred dogs. *Journal of Heredity*, **90**, 104–107.

Zhivotovsky LA, Feldman MW (1995) Microsatellite variability and genetic distances. *Proceedings of the National Academy of Sciences of the USA*, **92**, 11549–11552.

Mark Beaumont is primarily interested in developing and using statistical modelling methods for inferring demographic history from gene frequency information. Elizabeth Barratt is interested in the use of molecular methods to solve conservation problems, and has worked mainly on the population genetics of bats. Dada Gottelli is a specialist in microsatellite genotyping, with a particular interest in hybridization problems. Andrew Kitchener curates the bird and mammal collection in the National Museums of Scotland, and is interested in the worldwide taxonomic status of *Felis silvestris*. Mike Daniels has made detailed radiotelemetry studies of wild-living cats in Scotland and has contributed to much of the recent work on these cats. Jonathan Pritchard works in a statistical genetics group, and most of his research focuses on problems in population genetics. Mike Bruford is primarily interested in the fine-scale genetic structure of populations, especially those with complex social structure and/or threatened status.
