MOLECULAR ECOLOGY

On the use of genome-wide data to model and date the time of anthropogenic hybridisation: an example from the Scottish wildcat

Journal:	*Molecular Ecology*
Manuscript ID	MEC-21-0212
Manuscript Type:	Original Article
Date Submitted by the Author:	22-Feb-2021
Complete List of Authors:	Howard-McCombe, Jo; University of Bristol, School of Biological Sciences Ward, Daniel; University of Bristol, School of Mathematics Kitchener, Andrew; National Museums Scotland, Department of Natural Sciences Lawson, Dan; University of Bristol, School of Mathematics Senn, Helen; Royal Zoological Society of Scotland, RZSS WildGenes Lab Beaumont, Mark; University of Bristol, School of Biological Sciences
Keywords:	Hybridization, wildcat, admixture, approximate Bayesian computation, introgression

On the use of genome-wide data to model and date the time of anthropogenic hybridisation: an example from the Scottish wildcat

Howard-McCombe, J.[1], Ward, D.[2], Kitchener, A. C.[3], Lawson, D.[2], Senn, H. V.[4], Beaumont, M.[1]

[1]School of Biological Sciences, University of Bristol, Bristol, UK

[2]School of Mathematics, University of Bristol, Bristol, UK

[3]Department of Natural Sciences, National Museums Scotland, Edinburgh, UK

[4]RZSS WildGenes Lab, Royal Zoological Society of Scotland, Edinburgh, UK

Corresponding author:

Jo Howard-McCombe

School of Biological Sciences,

University of Bristol,

24 Tyndall Avenue,

Bristol

BS8 1TQ

j.howard-mccombe@bristol.ac.uk

Running title: Modelling hybridisation in Scottish wildcats

Abstract

While hybridisation has long been recognised as an important natural phenomenon in evolution, the conservation of taxa subject to introgressive hybridisation from domesticated forms is a subject of intense debate. Hybridisation of the Scottish wildcat, the UK's sole extant native felid, with the domestic cat is a good example in this regard. We develop a modelling framework to determine the timescale and mode of introgression using approximate Bayesian computation (ABC). Applying the model to ddRAD-seq data from 129 individuals, genotyped at 6,546 loci, we show that a population of wildcats genetically distant from domestic cats is still present in Scotland, though these individuals are found almost exclusively within the captive breeding program. Most wild-living cats sampled were introgressed to some extent. Additionally, we evaluate the effectiveness of current methods that are used to classify hybrids. We show that an optimised 35 SNP panel is a better predictor of the ddRAD-based hybrid score in comparison with a morphological method.

Keywords: hybridisation, wildcat, admixture, approximate Bayesian computation, introgression

Introduction

Hybridisation and introgression are important drivers of evolutionary change (Barton, 2001). Human-mediated hybridisation, however, is of increasing concern in conservation biology (Allendorf, Leary, Spruell, & Wenburg, 2001). Evolutionary processes may be disrupted by human activity, particularly when species distributions are altered by, for example, climate change, landscape use, or introduction of non-native species, leading to contact between populations that were originally allopatric. Whilst it is recognised this can generate a range of outcomes, some of which may be positive (e.g. genetic rescue; Johnson et al., 2010 or adaptive introgression; Pardo-Diaz et al., 2012), hybridisation and introgression are often considered threats to wild populations (Rhymer & Simberloff, 1996; Todesco et al., 2016). Loss of locally adaptive variation, reduction in fitness,

outbreeding depression or genetic swamping can all result in population or species extinction. Furthermore, introgressive hybridisation between domesticated species and wild populations increases the spread of potentially maladaptive, artificially selected variants in the wild (Randi, 2008).

The wildcat population in Scotland is an example of the threat of genetic extinction as a result of hybridisation (Mathews et al., 2018). The wildcat, *Felis silvestris*, is Britain's most endangered carnivoran and last remaining wild felid species. Wildcats have faced a long history of persecution and habitat loss and can hybridise with domestic cats to produce fertile offspring. Introgressive hybridisation is an increasingly serious threat to the dwindling population of this species in the Britain, which is now at risk of complete genetic replacement by hybrids in the wild (Breitenmoser, Lanz, & Breitenmoser-Würsten, 2019). Hybrids and feral domestic cats also compete with wildcats for habitat and resources and pose a disease transmission risk.

Modern domestic cats are derived from the Near Eastern wildcat species *Felis lybica*. The process of cat domestication was likely initiated as a result of their attraction to rodents, who themselves were attracted to grain stores associated with settled agriculture ~9,500 years ago (Driscoll et al., 2007). Though Driscoll *et al.* (2007) described just one wildcat species, *Felis silvestris,* distributed across Europe, Asia, and Africa, a recently revised Felidae taxonomy recognises two species of wildcat, *Felis silvestris* present in Europe, Caucasus and Turkey, and *Felis lybica* distributed in Africa and Asia (Kitchener et al., 2017).

Artificial selection has altered the morphology, behaviour, and rate of reproduction of domestic cats (Driscoll, Macdonald, & O'Brien, 2009). As a result, they are sufficiently diverged from wildcats to be considered a separate species, *Felis catus* (International Commission on Zoological Nomencalture, 2003). Domestic cats are widespread globally and found throughout the *Felis silvestris* range. Hybridisation between domestic cats and wildcats is variable across the wildcat

range in Europe (Yamaguchi, Kitchener, Driscoll, & Nussberger, 2015) and is particularly acute in Scotland for reasons that remain poorly understood.

The remaining Scottish wildcat population is believed to be small, whereas hybrid cats are prevalent in certain areas; in a 2017/18 survey of wildcat conservation "Priority Areas" (Littlewood et al., 2014) the ratio of un-neutered hybrids to wildcats was estimated at 6:1 (Breitenmoser et al., 2019). The wild-living population in Scotland now resembles a 'hybrid swarm' - a continuum of genetic backgrounds as a result of repeated back-crossing and mating between hybrids (Beaumont et al., 2001; Senn et al., 2019). A recent review of wildcat conservation in Britain by the IUCN concluded the population was "too small, with hybridisation too far advanced and the population too fragmented" to be considered viable (Breitenmoser et al., 2019).

Introgressive hybridisation, by definition, results in the movement of genes between species. However, the consequences of the introduction of domestic cat genes into wildcat populations, or the fitness of hybrid offspring, is poorly understood. It is unknown whether introduced domestic cat genes confer any selective advantage or disadvantage in hybrid populations. This is especially interesting in the context of a changing environment for wildcats, specifically habitat loss or change, and increased competition with, and spread of diseases from, feral domestic cats (Breitenmoser et al., 2019).

Methods to detect signals of natural selection commonly rely on identifying large differences in allele frequencies between populations (Lewontin & Krakauer, 1973). This is challenging for genetically continuous populations, such as the hybrid swarm observed in Scottish wildcats (Waples & Gaggiotti, 2006). Here we apply the tool *pcadapt* to perform scans for selection (Luu, Bazin, & Blum, 2017). *Pcadapt* uses a PCA-based approach to detect variants which are outliers with respect to population structure; it is especially appropriate for admixed individuals as it does not require population information *a priori*.

Uncertainty also surrounds the temporal patterns of hybridisation in Scotland. Domestic cats are thought to have become widespread during the Roman occupation of Britain ~2,000 years ago (Serpell, 2014), though cat remains have been found at Iron Age sites, including sites on the Orkney islands off the north coast of Scotland (Macdonald et al., 2010; Smith, 1994). The wildcat population dramatically declined during the 18th and 19th centuries due to hunting and habitat loss, and by the start of the 20th century wildcat range in the UK was limited to north-west Scotland. Significant introgression is believed to have occurred within the last 100 years, when the wildcat population expanded, increasing contact between the small remaining population of wildcats and domestic cats (Breitenmoser et al., 2019). Historic samples, collected over the last c. 100 years, support an acceleration of hybridisation in Scotland over this period (Senn et al., 2019).

Without a comprehensive understanding of hybridisation history or dynamics, or the impact of introgressive hybridisation on fitness, conservation of this species in Britain is not straightforward. Accurate population estimates are difficult to obtain due to the elusive nature of the species and limited ability to distinguish hybrids in the field based on morphology (Breitenmoser et al., 2019). This problem is compounded by the lack of a baseline reference for Scottish wildcats. The difficulties inherent in distinguishing wildcat and hybrid phenotypes results in haphazard protection, impedes accurate monitoring, and undermines the Scottish wildcat's legal status as a protected species.

The Scottish wildcat has served as a canonical example of domestic-wild hybridisation more generally. The aim of this study is, firstly, to clarify the population structure of wildcats in Scotland using a two-fold increase in the number of genetic markers compared to the most recent study (Senn et al., 2019). For this we use ddRAD-seq data; ddRAD-seq is an efficient way to sample thousands of markers for genome-wide estimates of hybridisation (Peterson, Weber, Kay, Fisher, & Hoekstra, 2012). Increasing the number of markers increases power to accurately identify complex hybrids and backcrosses (Boecklen & Howard, 1997), giving the greatest resolution to date of the hybrid swarm in Scotland.

Secondly, we use the expanded set of markers to evaluate the effectiveness of current tests to identify hybrid individuals. Finally, we obtain an estimate of the timescale of hybridisation using a model that predicts the observed pattern of population structure. A demographic model for Scottish wildcats was developed using an Approximate Bayesian Computational (ABC) framework (Beaumont, Zhang, & Balding, 2002), a model-based approach to parameter inference rooted in Bayesian statistics. We also apply the model to evaluating the performance of PCA-based methods to identify genes that are subject to natural selection in structured populations.

Methods

Data processing

ddRAD-seq data were generated for 129 individuals sampled between 1996 and 2017 (Senn et al., 2019). This included 71 individuals from the UK captive wildcat population (all sampled in 2017), 53 individuals from the wild in Scotland (22 Scottish Wildcat Action www.scottishwildcataction.org trapped cats, 31 roadkill samples) and five Scottish domestic cats, for full sample details see Supp. Table 1. Note that historical wildcat samples derived from museum specimens reported in Senn *et al.* (2019) could not be used for this study due to poor DNA quality.

Sequence reads were aligned using BWA (Li & Durbin, 2009) to the *Felis catus* reference genome v9.0 (GCF_000181335.3) (Pontius et al., 2007). Mapped reads were processed using STACKS v2.1 (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013). In STACKs a minimum of three reads were required to form a 'stack'. Resulting variants were filtered using a minimum allele frequency of 0.05 and a maximum proportion of heterozygous individuals of 0.7, treating the three sample sources (domestic, wild-living, and captive) as separate populations.

PLINK v1.9 (Chang et al., 2015) and VCFtools v1.15 (Danecek et al., 2011) were used to filter the data from STACKs. Specifically, the led to the removal of individuals with >30% missing data and

stringent subsequent filtering of loci to remove all sites with missing data. Closely related individuals were identified using IBD estimates calculated by PLINK, corrected to account for admixture using the method described by Morrison (2013). Corrected IBD estimates were used as input for PRIMUS (Staples et al., 2014) which uses genetic data to reconstruct pedigrees up to third degree relatives. Individuals were then removed from the dataset to limit relatedness (for the full list of excluded individuals see Supp. Table 1). Population genetic summary statistics (observed and expected heterozygosity, inbreeding coefficient and pairwise F_{ST}; Weir & Cockerham, 1984) were generated for the final dataset using PLINK and VCFtools.

Population structure

Principal component analysis (PCA) and ADMIXTURE (Alexander, Novembre, & Lange, 2009) were used to examine population structure. PCA was completed in R using *prcomp*. ADMIXTURE analyses were performed for seven values of K, ranging from two to eight, and included a calculation of cross-validation error to estimate the optimal value of K. All SNPs were included, the data were not considered dense enough to require thinning of markers (to minimise background linkage disequilibrium) prior to the analysis (Alexander et al., 2009).

Existing hybrid tests

Hybrid individuals are currently identified using a combination of genetic and morphological diagnostic tests: a seven-point pelage scoring system (Kitchener, Yamaguchi, Ward, & Macdonald, 2005) and a 35 SNP genetic test (Senn & Ogden, 2015). The pelage test (7PS) scores seven key morphological characteristics on an ordinal scale of 1,2,3 for domestic, hybrid or wildcat features, respectively. Putative wildcats score 19 or higher on this test (maximum score 21), though a lower threshold of 17 can be used to overcome possible recorder error, e.g., from poor quality camera-trap photos. The genetic test uses 35 SNPs that differentiate between wildcats and domestic cats (Nussberger, Greminger, Grossen, Keller, & Wandeler, 2013; Senn & Ogden, 2015). A 'hybrid score' is generated using STRUCTURE Q values between 0 and 1 (Pritchard, Stephens, & Donnelly, 2000);

higher values correspond to individuals with more wildcat ancestry. An LBQ score (i.e. the lower boundary of the Q value 90% CI) of 0.75 is proposed as the threshold to class individuals as putative wildcats, as distinct from hybrids (Senn & Ogden, 2015). Individuals with an LBQ≥0.75 are currently considered wildcats from a conservation management perspective.

We compared the performance of these hybrid tests using ADMIXTURE Q values from the ddRAD-seq data (6,546 SNPs) to determine hybrid status. None of the 35 SNPs from the genetic test were present in the ddRAD-seq data. Data were only included from individuals where both 35 SNP and pelage scores were available (n=59). The aim of this analysis was to compare the performance of these tests with diagnoses from a relatively dense marker set. Given the continuum of Q values observed in wild-living cats, a strict threshold (Q≥0.9) was used to select reference wildcat samples, but we recognise this threshold is somewhat arbitrary and does not necessarily denote 'true wildcat' status. Individuals with an ADMIXTURE Q score of 0.9 or more were classified as wildcat reference samples, and those below 0.9 as hybrids. Receiver operating characteristic (ROC) curves were then constructed to assess performance (Robin et al., 2011). Given the reference diagnosis, the true positive and false positive rates were calculated for both diagnostic tests at all possible threshold values. Plotting false positive rate against true positive rate (specificity vs sensitivity) for each classification threshold generated an ROC curve for each test. The area under the curve (AUC) is equivalent to the probability a test will rank a random positive instance higher than a random negative instance and is a useful metric to compare diagnostic tests. An AUC of 0.5 is essentially a random guess and an AUC of less than 0.5 is worse than random.

Outlier analysis

The data were screened for outliers using the R package *pcadapt* (Luu et al., 2017). The first three principal components were used in the analysis, following Cattell's Rule that eigenvalues relating to random variation lie on a straight line, and those relating to population structure depart from the

line (Cattell 1966). To reduce false positives, p-values < 1×10^{-6} were investigated as outliers (equivalent to 0.01 Bonferroni corrected).

To better understand the false positive rate of the outlying SNPs, simulated data (generated using a neutral model of evolution, described below) were also analysed using the same method in *pcadapt*. Ten simulated datasets were generated using a random sample of parameters values from the ABC posterior distribution (see below).

Demographic modelling

A demographic model for wildcats cats was developed within an ABC framework (Beaumont et al., 2002). ABC was developed as a rejection algorithm (Pritchard, Seielstad, Perez-Lezaun, & Feldman, 1999), in which simulated data are generated under a hypothesised model of evolution, with model parameters sampled from a known prior distribution. Summary statistics are taken from both the simulated data and observed data. An accepted sample of simulations (those with summaries closest to the observed data) are then used to estimate posterior distributions of the model parameters. Posterior estimates from the basic rejection algorithm can be improved with local linear (Beaumont et al., 2002) or non-linear regression (Blum & François, 2010).

Fig. 4A outlines the model developed for wildcat demography. Wildcat and domestic cat populations diverge, under a neutral model of evolution, for 500 generations. Generation time for a wildcat is estimated to be three years (Beaumont et al., 2001; Nussberger, Currat, Quilodran, Ponta, & Keller, 2018). The divergence of the two populations from a common ancestor is modelled using a computationally efficient two-stage approach; firstly, starting SNP frequencies for each population were simulated from a beta-binomial distribution, parameterised by F_{ST} (Balding & Nichols, 1995). These initialise an individual-based model of genetic inheritance in which at time T_1 gene-flow from domestics begins at a rate of mig_1 per generation. Gene-flow occurs at the same rate in every subsequent generation. At time T_2 the captive wildcat population is established from a random sample of wildcat individuals (referred to as the wild-living population from this point forward).

196 There is (limited) gene-flow (mig_2) from the wild-living population to the captive wildcats (reflecting
197 a number of wild-caught founders that have been incorporated into the captive population since it
198 was established). Population sizes remain constant throughout the simulation; we do not model any
199 fluctuations in wildcat population size (e.g., recent population expansion), and we do not model a
200 decline in the wildcat population as a direct result of hybridisation. Furthermore, unlike Quilodrán *et*
201 *al*. (2020), we do not consider a spatial model for hybridisation. Previous analysis indicates a
202 complex and patchy pattern of hybridisation, difficult to model on a large scale (Kilshaw et al., 2016;
203 Senn et al., 2019).

204 Data were simulated under this model using SLiM (Haller & Messer 2017), a toolkit for
205 evolutionary modelling. SLiM is individual-based, forward-simulating and, implements a Wright-
206 Fisher model of evolution (amongst others) in which generations are non-overlapping, individuals
207 are diploid, and offspring are generated through recombination and mutation of parental genotypes.
208 15,000 independent sites were modelled per individual (to replicate the observed SNP data from
209 ddRAD-seq). After 500 generations the genotypes of 46 captive wildcats, 45 wild-living and four
210 domestic cats were sampled at random, and summary statistics were calculated in R. Captive
211 individuals with a Q35 score of <0.9 (n=13) were filtered from the observed data. This functioned as
212 a proxy for the selection of putative wildcats for incorporation into the captive breeding programme,
213 in the model migrants are selected at random. The total number of simulations used for ABC was
214 509,070.

215 Prior distributions were chosen based on existing knowledge of the model system (for
216 details see Supp. Fig. 11). A wide prior was chosen for T_1, allowing hybridisation to begin at any
217 point in the simulation. A more informative prior was given to T_2 as we know the captive population
218 was established in 1960.

219 Given the strong separation of domestic cats and wildcats across the first principal
220 component (Fig. 1A), a set of PCA-based summaries was devised (measures of the distribution of

points across PC1 and PC2). Additional summaries included pairwise genetic distance (F_{ST}) and linkage disequilibrium measures, for full list see Supp. Table 2. The total number of summary statistics was 14. Owing to the correlation within and between parameters and summary statistics (Supp. Fig. 8), projection was used to reduce dimensionality and improve posterior estimates, following the approach of Fearnhead and Prangle (2012). Projection involves fitting a regression model between each parameter and the summary statistics. The regression model gives an estimate of the posterior mean for a given set of summary statistics. This prediction for each parameter can be viewed as a projection of the 14-dimensional summary statistics onto a 10-dimensional set of new summary statistics (Blum, Nunes, Prangle, & Sisson, 2013). To fit the regression model for the projection we chose 20% of simulated points that were closest to the observed set of summary statistics.

The final model parameters and summary statistics were decided via the process described in Supp. Figs. 5-7, which used goodness-of-fit test included in the R package *abc* (Csilléry, François, & Blum, 2012) and a novel method for dropping summary statistics (described in Supp. Box 1).

Parameter inference was carried out in R using the package *abc* (Csilléry et al., 2012). The closest 5,091 points (1%) were used to generate the posterior distributions, correcting for an imperfect match between the summary statistics and observed data using non-linear regression (neural network) (Blum et al., 2013; Raynal et al., 2019).

Results

The final dataset included 108 individuals: four Scottish domestic cats and 104 putative wildcats (45 wild individuals and 59 from the UK captive population), genotyped at 6,546 SNPs. 21 samples were excluded from the analysis to minimise relatedness in the dataset and/or as a result of stringent filtering of missing data. Population summary statistics are given in Table 1.

Population structure

Principal component analysis (Fig. 1A) showed a large proportion of the genotypic variation (23.9%) was explained by the first principal component (PC1). PC1 supports strong differentiation between domestic cats and a group of almost exclusively captive individuals, only two wild-living individuals are found at similarly extreme PC1 values. A large F_{ST} (0.446, Table 1) is observed between domestic cats and the captive wildcat population. The distinct PCA clustering and high F_{ST} values supports this as a cluster of putative wildcats. Most wild-living individuals are distributed across PC1, between these two groups, and are therefore considered putative hybrids. A much smaller proportion of the variance is explained by PC2 (2.8%) and PC3 (2.7%, Supp. Fig. 1).

An ADMIXTURE model with two ancestral populations (Fig. 1C, K=2) also supported distinct clustering of domestic cats and captive wildcats. The majority of wild individuals sampled had probable ancestry assigned to both groups, with varying amounts of 'domestic' ancestry. PC1 position was strongly correlated with ADMIXTURE Q values at K=2 (Spearman's r = 0.998, $p<0.001$; Supp. Fig. 2). Fig. 1B shows sampling locations for the wild individuals (where available), coloured by ADMIXTURE proportions at K=2. Individuals with domestic ancestry appear geographically widespread, with no clear single point of introgression. At K=3 further clustering within the putative wildcats is observed, including within the captive population. Cross-validation error indicated the most likely value of K for the whole dataset is 5 (Supp. Fig. 4).

Existing hybrid tests

ROC curves showed that both diagnostic tests performed well, with AUC values of 0.984 and 0.854 (Fig. 2). The 35 SNP test (LBQ≥0.75) outperformed the morphology-based test, with a low rate of both false positives and false negatives. Using a threshold of 17 the 7PS test showed nine false negatives and six false positives for the individuals analysed (i.e., individuals with few wildcat markings or features, but a high proportion of probable wildcat ancestry, and vice versa). At the

higher threshold of 19 there was only one instance of a false positive, but 19 false negatives. The 35 SNP test showed two false negatives and four false positives.

Evidence for natural selection

Pcadapt found three outlying SNPs that were reported to be most correlated with PC1 (Fig. 3B, for details see Supp. Table 3). Fig. 3B shows the PCA plot for the first two principal components, as in Fig. 1, with individuals coloured by genotype at each of the three positions (i.e., heterozygous, homozygous for allele 1 or homozygous for allele 2). For each SNP there was a clear difference in allele frequency between the domestic cat and captive wildcat populations. Notably, wild-living individuals had a high frequency of the domestic-type allele at these loci. This pattern does not seem to be an artefact of captive breeding, for each SNP shown in Fig. 3B the 'domestic' allele is at low frequency in wild individuals at similar PC1 positions as captive individuals, and at least one of these individuals was homozygous for the wildcat-type allele.

The SNPs are located on three different chromosomes. At the corresponding positions in the domestic cat genome SNPs 5147 and 5885 are found within protein-coding regions. SNP 5147 is found within the *SLC31A2* gene (chromosome D4, $p = 1.991 \times 10^{-7}$). In humans and mice *SLC31A2* has been shown to have copper ion transmembrane transporter activity (Okazaki et al. 2002; van den Berghe et al. 2007). SNP 5885 (chromosome E3 $p = 1.794 \times 10^{-7}$) is found within *ITGAX, ITGAX* is predicted to encode integrin subunit alpha X, orthologues of which are found in many other mammals, including humans and mice. Integrins generally are adhesion receptors, linking the extracellular matrix and cell cytoskeleton (Schnapp et al. 1995). They also interact with growth factor receptors to promote cell cycle progression and cell migration. SNP 2022 (chromosome B2, $p = 1.403 \times 10^{-11}$) is located 383bp downstream from the *TRAM2* gene, which encodes translocation associated membrane protein 2. In humans, *TRAM2* has been identified to have roles in collagen synthesis, protein transport and protein insertion into the membrane of the endoplasmic reticulum (Stefanovic et al. 2004).

Outlier SNPs are candidates for loci under selection, though extreme outliers can also be generated via neutral processes. Fig. 3A shows that outlying SNPs were generated under a neutral model of wildcat demography, a result of pre-existing population structure, emphasised by genetic drift. Even using a conservative threshold to minimise the false discovery rate, nine out of the ten sets of simulated data contained at least one SNP found to be outlying with respect to population structure across PC1 (see Table 2).

Demographic modelling

Our demographic model is capable of simulating data within the range of the observed data and the model fits these data well (Supp. Figs. 9 & 10). The first two axes of the posterior predictive PCA plots (Fig. 4C) show broadly the same patterns as the observed data, particularly with respect to the distribution of wild-living individuals across PC1. Prior and posterior distributions for the three parameters of interest (T_1, T_2, and mig_1) are shown in Fig 4B. The posterior mean for T_1, the time of onset of gene flow from domestics to wildcats, is 3.3 generations (95% HPD: 1.21– 5.). For T_2, the time the captive population was established, the mean is 19.3 generations (95% HPD: 9.4 – 30), respectively. Note that the estimate for T_1 is not constrained by the prior to any marked degree, whereas the historically informed prior for T_2 has a stronger effect. The migration rate of domestic cats into the wild-living population was estimated to be 0.13 (95% HPD: 0.076 – 0.19) i.e., for an individual selected at random from the wild-living population there is a 13% chance it is a domestic cat.

Discussion

Current status of the wildcat in Scotland

PCA and ADMIXTURE analysis (Fig. 1) demonstrated that a group of individuals genetically distinct from domestic cats (putative wildcats) persists in Scotland. Genetic differentiation between these

groups was supported by a high F_{ST}, as would be anticipated between two species (Hartl & Clark, 2007), and comparable to that between dogs and wolves (Cronin et al., 2015) or red and sika deer (McFarlane et al., 2020). This supports the findings of previous microsatellite (Beaumont et al., 2001) and SNP studies (Senn et al., 2019) that were able to differentiate between domestic cats and a group of putative wildcats in Scotland. Here we reanalyse the 76 samples used by Senn *et al.* (2019), with an additional 51 captive individuals and two additional wild individuals. We increase the resolution of this study with an additional 3,449 SNPs, and the data show the same broad patterns. Putative wildcats reported in this study were sampled almost exclusively from the UK captive population. Hybridisation in the wild appeared extensive. A continuum of genetic backgrounds is observed, the result of repeated hybridisation, backcrossing and mating between hybrids referred to as a 'hybrid swarm' (Mayr, 1963); almost all wild-living individuals sampled showed some evidence of introgression from domestic cats (Fig. 1). This supports the conclusion of Breitenmoser *et al.* (2019) that the wild population in Scotland is now too hybridised to be considered viable.

Demographic modelling supported a rapid emergence of the hybrid swarm effect and a recent crash in the Scottish wildcat population as a result of high geneflow from domestic cats. We take the generation time for wild-living cats to be around 3 years (Beaumont et al., 2001; Nussberger et al., 2018). The T_1 posterior mean (3.326 generations, or ~10 years) is implausibly recent, yet extensive model-checking (Fig. 4c, Supp. Figs. 5-10) suggests that the model generally fits well. The exact history of hybridisation in the Britain remains poorly understood (and is likely to show geographic variation) but hybridisation has been of increasing conservation concern since the 1980s (Hubbard et al., 1992, Kitchener et al. 1992, Easterbee et al. 1991) and is generally thought to be a consequence of wildcat range expansion in Scotland during the early 20th century coupled with continuing high levels of persecution, especially in eastern Scotland. This does not exclude the onset of significant introgression within the last few decades. Though no historical samples were included in this study, Senn *et al.* (2019) generated Q35 scores for 60 historic samples collected in Scotland

between 1895 and 1985. These are predominantly cats shot by gamekeepers and subsequently incorporated into museum collections, so there is potential bias towards individuals with wildcat features. Nonetheless, only five of the samples collected over this period are classified as hybrids, using the LBQ<0.75 threshold, and one as a domestic cat. In another example of hybridising species, Galaverni *et al.* (2017) date recent admixture between wolves and dogs in Italy to the 1940s, peaking in the 1990s.

The wildcat model is limited, however, by the ability of unlinked SNPs to detect ancient or complex patterns of admixture. Results presented here suggest our model is unable to detect signals of admixture beyond 30 generations or in this case, c. 100 years (Supp. Fig. 10). Haplotype and linkage disequilibrium information (from sequence data) are needed for accurate dating of admixture events, especially to separate historical admixture from the very recent (Hellenthal et al., 2014; Loh et al., 2013); this work in whole genome sequenced individuals is now underway.

Mattucci *et al.* (2019) used SNP array data to date admixture in continental European wildcat populations. Individuals were sampled from all five main biogeographic groups: Iberia, Central Europe, Central Germany, Italy and the Dinaric Alps (Mattucci, Oliveira, Lyons, Alves, & Randi, 2016). The study found hybridisation across all populations, occurring between six and 22 generations before present. The most recent admixture time reported by this study was 3.15 generations (though this date depended on the approach used). Mattucci *et al.* (2019) reported admixture times for individuals previously classified as true wildcats using microsatellite data, highlighting the power of a sequence-based approach to detecting historic and/or complex patterns of admixture (Gärke et al., 2012; Haasl & Payseur, 2011).

A recent hybridisation time for Scottish wildcats only seems likely in the face of high geneflow from domestic cats. Our model estimates gene flow to be 13% (95% HPD: 7-19%). In comparison, Quilodrán *et al.* (2020), using a forward simulating approach to model introgression in

368 the Swiss Jura wildcat population, estimated the rate of introgression to be 6%. At this lower rate of
369 introgression, it took 26 generations for the wildcat population to become 50% introgressed.

370 Quilodrán *et al.* (2020) use a spatial model to quantify introgression. Although this would be
371 challenging at the scale of the model presented here, especially considering the complex patterns of
372 introgression observed in the wild (Fig. 1B), it may be helpful in a future study to apply the approach
373 of Quilodrán *et al.* (2020), in conjunction with parameter estimates from the current model, to focus
374 on a geographical area of interest to better understand hybridisation dynamics in a priority area for
375 conservation management.

376 Tentative evidence is presented here that the 'hybrid swarm' effect can develop rapidly
377 following the breakdown of isolating mechanisms between two species, as has been observed in
378 other hybridising species such as deer (Smith, Carden, Coad, Birkitt, & Pemberton, 2014), loaches
379 (Kwan, Ko, & Won, 2014) and honey-bees (Pinto, Rubink, Patton, Coulson, & Johnston, 2005). Our
380 results may also support a recent acceleration of hybridisation in Britain. Though it is difficult to
381 conclude using the current model whether historical admixture has occurred (and to what extent), it
382 is clear there has been significant recent introgression within the last few decades.

383 An important feature of the model is the captive wildcat population. There is significant
384 interest surrounding this population, which comprises individuals that are among the last putative
385 wildcats in Britain, and especially regarding its value to continuing conservation efforts. It is
386 therefore important to understand the extent to which hybridisation has impacted this population.
387 It is clear from Fig. 1 that hybrids are present, though the number appears to be low. From the ABC
388 posterior distribution, T_2 (the time the captive population is established) occurs consistently before
389 gene-flow from domestic cats begins (T_1). This suggests the formation of the captive population in
390 the 1960s and 1970s may have occurred prior to significant recent admixture, and that this
391 population is an important reservoir of wildcat genes in Britain (probably aided in recent years by
392 accurate tests for hybrids, see below). How closely modern captive animals resemble the British

393 post-glacial population of wildcats, especially considering sympatry with domestic cats over the last
394 2000 years, remains to be determined.

395 Captive individuals have a wide distribution across PC2 and PC3 (though this explains only a
396 small proportion of the variation in the genetic data, 2.8% and 2.7%, respectively), and ADMIXTURE
397 plots show clustering within the captive population (Fig. 1C, K=3). The distribution of captive
398 individuals across PC2 was a difficult feature to replicate in the model (Fig. 4C). It is hard to
399 disentangle the impacts of maintaining a (historically small) captive breeding population, e.g.
400 inbreeding, genetic drift, or adaption to captivity (Frankham, 2018; Woodworth, Montgomery,
401 Briscoe, & Frankham, 2002), from genuine population structure. The presence of family groups was
402 limited following the identification of close relatives using PRIMUS. However, estimates of
403 relatedness are complicated by potential admixture (Morrison, 2013). Our results (Supp. Fig. 3)
404 imply the distribution of individuals across PC2 or PC3 is not a gradient of inbreeding across the
405 population.

406 Patterns relating to geographical origin in the wild samples were unclear due to the high
407 levels of introgression (Fig. 1B). In terms of introgression it seems clear there have been multiple
408 admixture events, possibly due to the pervasiveness of domestic cats in wildcat habitat in Scotland
409 and continuing high levels of persecution that maintained wildcat populations at low levels
410 (Kitchener & O'Connor 2010). The evidence presented here does not rule out that the observed
411 clustering in the captive population reflects biogeographic structure in the Scottish wildcat
412 population. The Great Glen, for example, has been suggested as a barrier to gene flow in the
413 Scottish red deer population (Pérez-Espona *et al.* 2008). The Great Glen is a ~100km long valley,
414 running along part of the Great Glen fault that bisects the Scottish Highlands. In red deer, strong
415 population differentiation is observed between the eastern and western sides of the Great Glen, and
416 it is possible that this is also a barrier to wildcat dispersal. However, wild-living individuals belonging

417 to a single cluster at K=3 were sampled from both sides of the Great Glen, so other geographical
418 barriers may need to be considered and tested with additional sampling and modelling.

419 A second possibility is that ADMIXTURE clustering at values of K greater than two reflect
420 temporal patterns of hybridisation, i.e., snapshots of the genetic composition of the wild-living
421 population at various points since the mid-20th century (a number of wild founders have been
422 incorporated into the captive population since it was founded in 1960). The value of K with the
423 lowest cross-validation error was five, this may be an effect of trying to break a continuum of
424 hybridisation levels into discrete units. It is interesting to note that captive individuals with probable
425 domestic ancestry at K=2 all belong to the same cluster at K=3.

426 Mattucci *et al.* (2016) suggest that strong population structure within wildcats in mainland
427 Europe (for example, between eastern and western Germany, Hertwig et al., 2009) represents
428 population expansion from five major mid-Pleistocene glacial refugia. Interestingly, PCA of the
429 microsatellite data collected for this study shows a similar 'anvil' shape, with *Felis silvestris* more
430 dispersed across PC2 than *Felis catus*. Population structure and expansion perhaps make this a
431 feature of wildcat genetics more generally (especially when compared to inbred domestic cats), and
432 we should avoid over-interpretation in the Scottish population (Lawson, van Dorp, & Falush, 2018).

433 *Evidence for natural selection*

434 The major application of outlier analyses is to detect loci under natural selection. There has
435 been some debate in the literature as to whether RAD-seq data are appropriate for this kind of
436 analysis (Catchen *et al.* 2017; Lowry *et al.* 2017; McKinney *et al.* 2017). Lowry *et al.* (2017) argue
437 that the sparsity of RAD-seq markers misses many candidate loci, especially in species where linkage
438 disequilibrium is low. This does not necessarily invalidate the small number of loci identified using
439 RAD-seq, though it would be useful to confirm these findings with sequence data when possible.

Confounding effects, such as population structure and demography, are more problematic for this study. Even at neutral loci the demographic history of a population can cause allele frequency to vary hugely in space due to genetic drift and/or migration (Hoban et al. 2016). For populations that are highly differentiated the variance in F_{ST} among neutral loci is large. Differences in allele frequencies between domestic cats and wildcats are therefore not surprising considering the genetic differentiation between the two populations, and do not necessarily correspond to deviations from neutrality. Population expansion can also produce the same signal as selection due to 'allele surfing', where populations at the leading edge of an expansion are small, and contribute disproportionately to the expanding population, accelerating the effects of drift. As discussed above, the wildcat population in Scotland is thought to have been expanding since the early 20th century (Breitenmoser et al., 2019).

Here we have applied *pcadapt* to detect selection, which is designed to be robust to demographic biases and handle genetically continuous, admixed populations (Luu et al., 2017). However, simulation results, based on our best-fitting demographic model for the wildcats, show evidence of a high number of false-positives in this setting (Table 2), even using the most conservative approach to controlling false discovery rate. Although simulation-based tests using *pcadapt* have often shown that it performs well (Luu et al., 2017), scenarios with high recent admixture have not been investigated.

Based on this finding it is difficult to make conclusive statements about natural selection in Scottish wildcats, or fitness consequences for hybrid populations. Mattucci *et al.* (2019) reported a number of genomic regions in wildcat x domestic hybrids with a high frequency of either wildcat or domestic alleles, and genes within these regions were found to be significantly enriched for specific gene ontology categories. A striking feature of Fig. 3B is the similarity in allele frequencies between domestic and hybrids cats, even in less introgressed individuals, which perhaps constitutes tentative evidence for adaptive introgression in Scotland. Adaptive introgression has been shown to occur in

other wild populations which hybridise with domesticates, such as goats and sheep (Barbato et al., 2017; Grossen et al., 2014). The SNP correlated with PC1 with the most extreme p-value reported by *pcadapt* (Table 2, Supp. Table 3) is found in the domestic cat genome near the TRAM2 gene. TRAM2 has also been identified in genome scans for loci linked to the severity of leukaemia virus infection in cattle (Carignano et al., 2018). This finding highlights disease transmission as a potential driver of selection in hybrid populations. Both wildcat and domestic-like regions identified by Mattucci *et al.* (2019) included genes involved in the immune system or associated with diseases or infection, including feline leukaemia virus. Feline leukaemia virus is potentially fatal to both wildcats and domestic cats, and has similar prevalence (~10%) in both species in Scotland (Daniels et al. 1999).

Existing tests for hybrids

Accurately identifying hybrids in the field is crucial to effective conservation of the wildcat in Scotland. In the absence of uncontroversial reference samples, we have used a score based on 6,546 ddRAD SNPs and investigated the relative effectiveness of field-based tests in recovering this. An ROC analysis (Fig. 2) showed both diagnostic tests to be informative in identifying hybrid individuals as judged by scores from the ddRAD SNPs. The pelage score was a less reliable indicator of wildcat ancestry; this is unsurprising as the characteristics scored by this test are likely to be controlled by a limited number of genes (Cieslak, Reissmann, Hofreiter, & Ludwig, 2011; Eizirik et al., 2010), the transmission of which is still poorly understood. Devillard *et al.* (2014) and Kitchener *et al.* (2005) reported a greater degree of accuracy when using anatomical characteristics (skull size and shape and intestinal length) as opposed to than pelage in order to identify hybrids. Mattucci *et al.* (2019) found genomics regions in hybrid individuals with a high frequency of wildcat-type alleles contained (amongst others) genes relating to morphology. If selection is acting on key morphological features, as this result suggests, pelage may not give an accurate picture of hybridisation across the genome. Using a more lenient threshold (7PS ≥17 for putative wildcats) pelage scoring appeared to give a number of false negatives and false positives, i.e., individuals with

probable wildcat ancestry that did not necessarily score highly for wildcat features and vice versa. A more conservative threshold of 7PS≥19 reduces the number of false positives but increases the false negative rate - a large number of individuals with high proportions of putatively wildcat ancestry are not classified as wildcats at this threshold.

We found the 35 SNP test to be a highly accurate predictor of the ddRAD SNP score; hybrids could be identified almost as well using the 35 SNPs as with a dense marker set of over 6000 SNPs. Four false positives and two false negatives were identified, though similar Q values were recovered using both marker sets for these individuals, so this may partly reflect the stringent threshold used to select reference wildcats from the ddRAD data.

Without accurate information on the history of hybridisation in Britain there is no uncontroversial baseline for Scottish wildcats with which to calibrate either diagnostic test. Therefore, we recommend the continued use of the pelage score and 35 SNP test in conjunction to identify hybrids, especially when considering individuals to be incorporated into the captive breeding programme.

Conclusion

We find a population of putative wildcats persists in Scotland. These individuals are almost exclusively found in the UK captive population, which appears to have been established prior to significant recent admixture and is supported by accurate tests for hybrids. It remains unclear to what extent historical admixture has affected the Scottish wildcat population, but divergence between domestic cats and putative wildcats remains high. The captive population is now an important resource for wildcat conservation in Britain. We find the wild-living population to be a hybrid swarm; almost all wild individuals sampled showed evidence of introgression from domestic cats. We predict a high rate of continuing gene-flow from domestic cats.

514

515 **Acknowledgements**

516 This study represents a further analysis of the ddRAD data set first published in Senn *et al.* (2019);
517 the authors of that study are thanked. The authors would like to thank again the many people who
518 have provided samples to this dataset over the years. Thanks to those who have handed wildcat
519 samples to the National Museum of Scotland; without these repeated individual efforts, these types
520 of study are not possible. We are also extremely grateful for the assistance of the wildcat captive
521 holding community in the UK and their participation in collecting samples. We thank Danielle
522 Gunn-Moore at the University of Edinburgh for access to domestic cat reference samples. We also
523 thank the staff, volunteers and collaborators of Scottish Wildcat Action for their participation in
524 sampling during this project and David Barclay for discussions relating to the studbook. The ddRAD
525 data creation was funded by RZSS and the Heritage Lottery Fund via grant to Scottish Natural
526 Heritage. JHM is support by a NERC Doctoral Training Partnership studentship from the Natural
527 Environment Research Council.

528 **References**

529 Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast Model-Based Estimation of Ancestry in Unrelated
530 Individuals, 1655–1664. https://doi.org/10.1101/gr.094052.109.vidual

531 Allendorf, F. W., Leary, R. F., Spruell, P., & Wenburg, J. K. (2001). The problems with hybrids: Setting
532 conservation guidelines. *Trends in Ecology and Evolution*, *16*(11), 613–622.
533 https://doi.org/10.1016/S0169-5347(01)02290-X

534 Balding, D. J., & Nichols, R. A. (1995). A method for quantifying differentiation between populations at multi-
535 allelic loci and its implications for investigating identity and paternity. *Genetica*, *96*(1–2), 3–12.
536 https://doi.org/10.1007/BF01441146

537 Barbato, M., Hailer, F., Orozco-Terwengel, P., Kijas, J., Mereu, P., Cabras, P., … Bruford, M. W. (2017). Genomic
538 signatures of adaptive introgression from European mouflon into domestic sheep. *Scientific Reports*,
539 *7*(1), 1–13. https://doi.org/10.1038/s41598-017-07382-7

540 Barton, N. H. (2001). The role of hybridization in evolution. *Molecular Ecology*, *10*(3), 551–568.
541 https://doi.org/10.1046/j.1365-294X.2001.01216.x

542 Beaumont, M. A., Zhang, W., & Balding, D. J. (2002). Approximate Bayesian Compututation in Population
543 Genetics. *Genetics*, *162*(4), 2025–2035.

544 Beaumont, M., Barratt, E. M., Gottelli, D., Kitchener, A. C., Daniels, M. J., Pritchard, J. K., & Bruford, M. W.
545 (2001). Genetic diversity and introgression in the Scottish wildcat. *Molecular Ecology*, *10*(2), 319–336.
546 https://doi.org/10.1046/j.1365-294X.2001.01196.x

Blum, M.G.B., Nunes, M. A., Prangle, D., & Sisson, S. A. (2013). A comparative review of dimension reduction methods in approximate bayesian computation. *Statistical Science*, *28*(2), 189–208. https://doi.org/10.1214/12-STS406

Blum, M. G.B., & François, O. (2010). Non-linear regression models for Approximate Bayesian Computation. *Statistics and Computing*, *20*(1), 63–73. https://doi.org/10.1007/s11222-009-9116-0

Boecklen, W. J., & Howard, D. J. (1997). Genetic Analysis of Hybrid Zones : Numbers of Markers and Power of Resolution Author (s): William J . Boecklen and Daniel J . Howard Published by : Wiley on behalf of the Ecological Society of America Stable URL : http://www.jstor.org/stable/2265918 REF. *Ecology*, *78*(8), 2611–2616.

Breitenmoser, U., Lanz, T., & Breitenmoser-Würsten, C. (2019). Conservation of the wildcat (Felis silvestris) in Scotland: Review of the conservation status and assessment of conservation activities, (February). Retrieved from http://www.scottishwildcataction.org/media/42633/wildcat-in-scotland-review-of-conservation-status-and-activities-final-14-february-2019.pdf

Carignano, H. A., Roldan, D. L., Beribe, M. J., Raschia, M. A., Amadio, A., Nani, J. P., … Miretti, M. M. (2018). Genome-wide scan for commons SNPs affecting bovine leukemia virus infection level in dairy cattle. *BMC Genomics*, *19*(1), 1–15. https://doi.org/10.1186/s12864-018-4523-2

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140. https://doi.org/10.1111/mec.12354

Catchen, J. M., Hohenlohe, P. A., Bernatchez, L., Funk, W. C., Andrews, K. R., & Allendorf, F. W. (2017). Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. *Molecular Ecology Resources*, *17*(3), 362–365. https://doi.org/10.1111/1755-0998.12669

Catell, R. B. (1966) The Scree Test For The Number Of Factors. Multivariate Behav Res. Apr 1;1(2):245-76. doi: 10.1207/s15327906mbr0102_10. PMID: 26828106.

Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1), 1–16. https://doi.org/10.1186/s13742-015-0047-8

Cieslak, M., Reissmann, M., Hofreiter, M., & Ludwig, A. (2011). Colours of domestication. *Biological Reviews*, *86*(4), 885–899. https://doi.org/10.1111/j.1469-185X.2011.00177.x

Cronin, M. A., Cánovas, A., Bannasch, D. L., Oberbauer, A. M., MeDrano, J. F., & Ostrander, E. (2015). Single nucleotide polymorphism (SNP) variation of wolves (Canis lupus) in Southeast Alaska and comparison with wolves, dogs, and Coyotes in North America. *Journal of Heredity*, *106*(1), 26–36. https://doi.org/10.1093/jhered/esu075

Csilléry, K., François, O., & Blum, M. G. B. (2012). Abc: An R package for approximate Bayesian computation (ABC). *Methods in Ecology and Evolution*, *3*(3), 475–479. https://doi.org/10.1111/j.2041-210X.2011.00179.x

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Daniels, M. J., Golder, M. C., Jarrett, O., & MacDonald, D. W. (1999). Feline Viruses in Wildcats from Scotland. *Journal of Wildlife Diseases*, *35*(1), 121–124. https://doi.org/10.7589/0090-3558-35.1.121

Devillard, S., Jombart, T., Léger, F., Pontier, D., Say, L., & Ruette, S. (2014). How reliable are morphological and anatomical characters to distinguish European wildcats, domestic cats and their hybrids in France? *Journal of Zoological Systematics and Evolutionary Research*, *52*(2), 154–162. https://doi.org/10.1111/jzs.12049

Driscoll, C. A., Macdonald, D. W., & O'Brien, S. J. (2009). From Wild Animals to Domestic Pets, and Evolutionary View of Domestication. In J. C. Avise & F. J. Ayala (Eds.), *In the Light of Evolution III: Two Centuries of Darwin* (pp. 89–109). Washington (DC): National Academies Press. https://doi.org/10.1016/B978-0-323-

60984-5.00062-7

Driscoll, C. A., Menotti-Raymond, M., Roca, A. L., Hupe, K., Johnson, W. E., Geffen, E., … Macdonald, D. W. (2007). The Near Eastern Origin of. *Middle East*, *317*(July), 519–523.

Eizirik, E., David, V. A., Buckley-Beason, V., Roelke, M. E., Schäffer, A. A., Hannah, S. S., … Menotti-Raymond, M. (2010). Defining and mapping mammalian coat pattern genes: Multiple genomic regions implicated in domestic cat stripes and spots. *Genetics*, *184*(1), 267–275. https://doi.org/10.1534/genetics.109.109629

Fearnhead, P., & Prangle, D. (2012). Constructing summary statistics for approximate Bayesian computation : semi-automatic approximate Bayesian computation [with Discussion] Author (s): Paul Fearnhead and Dennis Prangle Source : Journal of the Royal Statistical Society . Series B (Stati, *74*(3), 419–474.

Frankham, R. (2018). Conservation genetics. *Encyclopedia of Ecology*, 382–390. https://doi.org/10.1016/B978-0-12-409548-9.10559-7

Galaverni, M., Caniglia, R., Pagani, L., Fabbri, E., Boattini, A., & Randi, E. (2017). Disentangling timing of admixture, patterns of introgression, and phenotypic indicators in a hybridizing Wolf population. *Molecular Biology and Evolution*, *34*(9), 2324–2339. https://doi.org/10.1093/molbev/msx169

Gärke, C., Ytournel, F., Bed'Hom, B., Gut, I., Lathrop, M., Weigend, S., & Simianer, H. (2012). Comparison of SNPs and microsatellites for assessing the genetic structure of chicken populations. *Animal Genetics*, *43*(4), 419–428. https://doi.org/10.1111/j.1365-2052.2011.02284.x

Grossen, C., Keller, L., Biebach, I., Zhang, W., Tosser-Klopp, G., Ajmone, P., … Croll, D. (2014). Introgression from Domestic Goat Generated Variation at the Major Histocompatibility Complex of Alpine Ibex. *PLoS Genetics*, *10*(6). https://doi.org/10.1371/journal.pgen.1004438

Haasl, R. J., & Payseur, B. A. (2011). Multi-locus inference of population structure: A comparison between single nucleotide polymorphisms and microsatellites. *Heredity*, *106*(1), 158–171. https://doi.org/10.1038/hdy.2010.21

Haller, B. C., & Messer, P. W. (2017). SLiM 2: Flexible, interactive forward genetic simulations. *Molecular Biology and Evolution*, *34*(1), 230–240. https://doi.org/10.1093/molbev/msw211

Hartl, D. L. & Clark, A. G. (2007) Principles of population genetics (4th ed.) Oxford University Press, New York, USA

Hellenthal, G., Busby, G. B. J., Band, G., Wilson, J. F., Capelli, C., Falush, D., & Myers, S. (2014). A genetic atlas of human admixture history. *Science*, *343*(6172), 747–751. https://doi.org/10.1126/science.1243518

Hertwig, S. T., Schweizer, M., Stepanow, S., Jungnickel, A., Böhle, U. R., & Fischer, M. S. (2009). Regionally high rates of hybridization and introgression in German wildcat populations (Felis silvestris, Carnivora, Felidae). *Journal of Zoological Systematics and Evolutionary Research*, *47*(3), 283–297. https://doi.org/10.1111/j.1439-0469.2009.00536.x

Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., … Whitlock, M. C. (2016). Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. *The American Naturalist*, *188*(4), 379–397. https://doi.org/10.1086/688018

Hubbard, A. L., McOris, S., Jones, T. W., Boid, R., Scott, R., & Easterbee, N. (1992). Is survival of European wildcats Felis silvestris in Britain threatened by interbreeding with domestic cats? *Biological Conservation*, *61*(3), 203–208. https://doi.org/10.1016/0006-3207(92)91117-B

International Commission on Zoological Nomencalture. (2003). Opinion 2027 (Case 3010). *Bulletin of Zoological Nomenclature*, *60*, 81–84.

Johnson, W. E., Onorato, D. P., Roelke, M. E., Land, E. D., Cunningham, M., Belden, R. C., … O'Brien, S. J. (2010). Genetic Restoration of the Florida Panther. *Science*, *9*(1), 76–99. https://doi.org/10.1558/jsrnc.v4il.24

Kilshaw, K., Montgomery, R. A., Campbell, R. D., Hetherington, D. A., Johnson, P. J., Kitchener, A. C., … Millspaugh, J. J. (2016). Mapping the spatial configuration of hybridization risk for an endangered population of the European wildcat (Felis silvestris silvestris) in Scotland. *Mammal Research*, *61*(1), 1–11.

https://doi.org/10.1007/s13364-015-0253-x

Kitchener, A. C., Breitenmoser-Würsten, C., Eizirik, E., Gentry, A., Werdelin, L., Wilting, A., … Tobe, S. (2017). A revised taxonomy of Felidae. The final report of the Cat Classification Task Force of the IUCN/SSC Cat Specialist Group. *Cat News Special Issue 11*.

Kitchener, A. C., O'Connor, T. (2010) Wildcats, domestic cats and feral cats. In O'Connor, T., Sykes, N. (Eds.) Extinctions and invasions. A social history of British fauna (pp. 83-94). Windgather Press, Oxford.

Kitchener, A. C., Yamaguchi, N., Ward, J. M., & Macdonald, D. W. (2005). A diagnosis for the Scottish wildcat (Felis silvestris): A tool for conservation action for a critically-endangered felid. *Animal Conservation*, *8*(3), 223–237. https://doi.org/10.1017/S1367943005002301

Kwan, Y. S., Ko, M. H., & Won, Y. J. (2014). Genomic replacement of native Cobitis lutheri with introduced C. tetralineata through a hybrid swarm following the artificial connection of river systems. *Ecology and Evolution*, *4*(8), 1451–1465. https://doi.org/10.1002/ece3.1027

Lawson, D. J., van Dorp, L., & Falush, D. (2018). A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nature Communications*, *9*(1), 1–11. https://doi.org/10.1038/s41467-018-05257-7

Lewontin, R. C., & Krakauer, J. (1973). Distribution of gene frequency as a test of theory of the selective neutrality of polymorphisms. *Genetics*, *74*, 175–195.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Littlewood, N. A., Campbell, R. D., Dinnie, L., Gilbert, L., Hooper, R., Iason, G., … Ross, A. (2014). *Survey and scoping of wildcat priority areas. Scottish Natural Heritage Commissioned Report No. 768*.

Loader, C., (2013) locfit: Local regression, likelihood and density estimation. R package version 1.5-9.1. https://CRAN.R-project.org/package=locfit

Loh, P. R., Lipson, M., Patterson, N., Moorjani, P., Pickrell, J. K., Reich, D., & Berger, B. (2013). Inferring admixture histories of human populations using linkage disequilibrium. *Genetics*, *193*(4), 1233–1254. https://doi.org/10.1534/genetics.112.147330

Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., & Storfer, A. (2017). Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Molecular Ecology Resources*, *17*(2), 142–152. https://doi.org/10.1111/1755-0998.12635

Luu, K., Bazin, E., & Blum, M. G. B. (2017). pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, *17*(1), 67–77. https://doi.org/10.1111/1755-0998.12592

Macdonald, D. W., Yamaguchi, N., Kitchener, A. C., Daniels, M., Kilshaw, K., & Driscoll, C. (2010). Reversing cryptic extinction: the history, present, and future of the Scottish wildcat. *Biology and Conservation of Wild Felids*, (September 2016), 471–491.

Mathews, F., Kubasiewicz, L. M., Gurnell, J., Harrower, C. A., McDonald, R. A., & Shore, R. F. (2018). *A Review of the Population and Conservation Status of British Mammals: Technical Summary*. Retrieved from http://publications.naturalengland.org.uk/publication/5636785878597632

Mattucci, F., Galaverni, M., Lyons, L. A., Alves, P. C., Randi, E., Velli, E., … Caniglia, R. (2019). Genomic approaches to identify hybrids and estimate admixture times in European wildcat populations. *Scientific Reports*, *9*(1), 1–15. https://doi.org/10.1038/s41598-019-48002-w

Mattucci, F., Oliveira, R., Lyons, L. A., Alves, P. C., & Randi, E. (2016). European wildcat populations are subdivided into five main biogeographic groups: Consequences of Pleistocene climate changes or recent anthropogenic fragmentation? *Ecology and Evolution*, *6*(1), 3–22. https://doi.org/10.1002/ece3.1815

Mayr, E. (1963) Animal species and evolution. Harvard University Press, Cambridge, MA, USA.

McFarlane, S. E., Hunter, D. C., Senn, H. V., Smith, S. L., Holland, R., Huisman, J., & Pemberton, J. M. (2020).

Increased genetic marker density reveals high levels of admixture between red deer and introduced Japanese sika in Kintyre, Scotland. *Evolutionary Applications*, *13*(2), 432–441. https://doi.org/10.1111/eva.12880

McKinney, G. J., Larson, W. A., Seeb, L. W., & Seeb, J. E. (2017). RADseq provides unprecedented insights into molecular ecology and evolutionary genetics: comment on Breaking RAD by Lowry et al. (2016). *Molecular Ecology Resources*, *17*(3), 356–361. https://doi.org/10.1111/1755-0998.12649

Morrison, J. (2013). Estimation for Samples with Population Structure, *37*(6), 635–641. https://doi.org/10.1002/gepi.21737.Characterization

Nussberger, B., Currat, M., Quilodran, C. S., Ponta, N., & Keller, L. F. (2018). Range expansion as an explanation for introgression in European wildcats. *Biological Conservation*, *218*(November 2017), 49–56. https://doi.org/10.1016/j.biocon.2017.12.009

Nussberger, B., Greminger, M. P., Grossen, C., Keller, L. F., & Wandeler, P. (2013). Development of SNP markers identifying European wildcats, domestic cats, and their admixed progeny. *Molecular Ecology Resources*, *13*(3), 447–460. https://doi.org/10.1111/1755-0998.12075

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., … Hayashizaki, Y. (2002)Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature*, *420*(6915), 563–573. https://doi.org/10.1038/nature01266

Pardo-Diaz, C., Salazar, C., Baxter, S. W., Merot, C., Figueiredo-Ready, W., Joron, M., … Jiggins, C. D. (2012). Adaptive introgression across species boundaries in Heliconius butterflies. *PLoS Genetics*, *8*(6). https://doi.org/10.1371/journal.pgen.1002752

Pérez-Espona, S., Pérez-Barbería, F. J., Mcleod, J. E., Jiggins, C. D., Gordon, I. J., & Pemberton, J. M. (2008). Landscape features affect gene flow of Scottish Highland red deer (Cervus elaphus). *Molecular Ecology*, *17*(4), 981–996. https://doi.org/10.1111/j.1365-294X.2007.03629.x

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, *7*(5). https://doi.org/10.1371/journal.pone.0037135

Pinto, M. A., Rubink, W. L., Patton, J. C., Coulson, R. N., & Johnston, J. S. (2005). Africanization in the United States: Replacement of feral European honeybees (Apis mellifera L.) by an African hybrid swarm. *Genetics*, *170*(4), 1653–1665. https://doi.org/10.1534/genetics.104.035030

Pontius, J. U., Mullikin, J. C., Smith, D. R., Lindblad-Toh, K., Gnerre, S., Clamp, M., … McKernan, K. (2007). Initial sequence and comparative analysis of the cat genome. *Genome Research*, *17*(11), 1675–1689. https://doi.org/10.1101/gr.6380007

Pritchard, J. K., Seielstad, M. T., Perez-Lezaun, A., & Feldman, M. W. (1999). Population growth of human Y chromosomes: A study of y chromosome microsatellites. *Molecular Biology and Evolution*, *16*(12), 1791–1798. https://doi.org/10.1093/oxfordjournals.molbev.a026091

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959. https://doi.org/10.1111/j.1471-8286.2007.01758.x

Quilodrán, C. S., Nussberger, B., Macdonald, D. W., Montoya-Burgos, J. I., & Currat, M. (2020). Projecting introgression from domestic cats into European wildcats in the Swiss Jura. *Evolutionary Applications*, (October 2019), 1–12. https://doi.org/10.1111/eva.12968

Randi, E. (2008). Detecting hybridization between wild species and their domesticated relatives. *Molecular Ecology*, *17*(1), 285–293. https://doi.org/10.1111/j.1365-294X.2007.03417.x

Raynal, L., Marin, J. M., Pudlo, P., Ribatet, M., Robert, C. P., & Estoup, A. (2019). ABC random forests for Bayesian parameter inference. *Bioinformatics*, *35*(10), 1720–1728. https://doi.org/10.1093/bioinformatics/bty867

Rhymer, J. M., & Simberloff, D. (1996). Extinction By Hybridization and Introgression. *Annual Review of Ecology*

and Systematics, *27*(1), 83–109. https://doi.org/10.1146/annurev.ecolsys.27.1.83

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Miller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, *8*, 12–77. https://doi.org/10.1007/s00134-009-1641-y

Schnapp, L. M., Hatch, N., Ramos, D. M., Klimanskaya, I. V, Sheppard, D., & Pytela, R. (1995). The human integrin α8β1 functions as a receptor for tenascin, fibronectin, and vitronectin. *Journal of Biological Chemistry*, *270*(39), 23196–23202. https://doi.org/10.1074/jbc.270.39.23196

Senn, H., & Ogden, R. (2015). Wildcat hybrid scoring for conservation breeding under the Scottish Wildcat Conservation Action Plan.Royal Zoological Society of Scotland

Senn, H. V., Ghazali, M., Kaden, J., Barclay, D., Harrower, B., Campbell, R. D., … Kitchener, A. C. (2019). Distinguishing the victim from the threat: SNP-based methods reveal the extent of introgressive hybridization between wildcats and domestic cats in Scotland and inform future in situ and ex situ management options for species restoration. *Evolutionary Applications*, *12*(3), 399–414. https://doi.org/10.1111/eva.12720

Serpell, J. A. (2014). Domestication and history of the cat. In D. C. Turner & P. Bateson (Eds.), *The Domestic Cat: The Biology of its Behaviour* (3rd ed., pp. 83–100). Cambridge University Press. https://doi.org/10.1017/CBO9781139177177.011

Smith, B. B. (1994). Howe: four millennia of Orkney prehistory excavations 1978-1982. *Society of Antiquaries of Scotland Monograph Series Number 9*.

Smith, S. L., Carden, R. F., Coad, B., Birkitt, T., & Pemberton, J. M. (2014). A survey of the hybridisation status of Cervus deer species on the island of Ireland. *Conservation Genetics*, *15*(4), 823–835. https://doi.org/10.1007/s10592-014-0582-3

Staples, J., Qiao, D., Cho, M. H., Silverman, E. K., Nickerson, D. A., & Below, J. E. (2014). PRIMUS: Rapid reconstruction of pedigrees from genome-wide estimates of identity by descent. *American Journal of Human Genetics*, *95*(5), 553–564. https://doi.org/10.1016/j.ajhg.2014.10.005

Stefanovic, B., Stefanovic, L., Schnabl, B., Bataller, R., & Brenner, D. A. (2004). TRAM2 Protein Interacts with Endoplasmic Reticulum Ca2+ Pump Serca2b and Is Necessary for Collagen Type I Synthesis. *Molecular and Cellular Biology*, *24*(4), 1758–1768. https://doi.org/10.1128/mcb.24.4.1758-1768.2004

Todesco, M., Pascual, M. A., Owens, G. L., Ostevik, K. L., Moyers, B. T., Hübner, S., … Rieseberg, L. H. (2016). Hybridization and extinction. *Evolutionary Applications*, *9*(7), 892–908. https://doi.org/10.1111/eva.12367

van den Berghe, P. V. E., Folmer, D. E., Malingré, H. E. M., van Beurden, E., Klomp, A. E. M., van de Sluis, B., … Klomp, L. W. J. (2007). Human copper transporter 2 is localized in late endosomes and lysosomes and facilitates cellular copper uptake. *Biochemical Journal*, *407*(1), 49–59. https://doi.org/10.1042/bj20070705

Waples, R. S., & Gaggiotti, O. (2006). What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, *15*(6), 1419–1439. https://doi.org/10.1111/j.1365-294X.2006.02890.x

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure Author (s): B . S . Weir and C . Clark Cockerham Published by : Society for the Study of Evolution Stable URL : http://www.jstor.org/stable/2408641. *Evolution*, *38*(6), 1358–1370. https://doi.org/10.2307/2408641

Woodworth, L. M., Montgomery, M. E., Briscoe, D. A., & Frankham, R. (2002). Rapid genetic deterioration in captive populations: Causes and conservation implications. *Conservation Genetics*, *3*(3), 277–288. https://doi.org/10.1023/A:1019954801089

Yamaguchi, N., Kitchener, A., Driscoll, C., & Nussberger, B. (2015). Felis silvestris. *The IUCN Red List of Threatened Species 2015*, *8235*, e.T60354712A50652361. https://doi.org/10.2305/IUCN.UK.2015-2.RLTS.T60354712A50652361.en

Data Accessibility

All SNP data available from the Dryad Digital Repository [in progress]. Materials for demographic modelling available at [GitHub site, in progress]

Author Contributions

JHM designed the research, analysed the data, and wrote the paper. HS provided data for analysis. MB, DL and HS conceived the study and designed the research. DW and AK analysed the data. All authors critically reviewed the paper.

Tables and Figures

Table 1. Summary statistics for the three source populations: captive wildcats, wild individuals, and domestic cats. Weir & Cockerham (1984) estimates for population pairwise F_{ST} are shown on the right-hand side.

Summary	Population			Pairwise F_{ST}		
	Captive	Wild	Domestic		**Captive**	**Wild**
# Individuals	59	45	4	**Captive**		
# Loci	6546	6546	6546	**Wild**	0.130	
# Alleles	12258	13075	11448	**Domestic**	0.446	0.128
% missing data	0	0	0			
H_{Obs}	0.178	0.307	0.270			
H_{Exp}	0.285	0.285	0.285			
F	0.375	-0.077	0.055			

Table 2. *Pcadapt* using data simulated under a neutral model of evolution. The simulated data contain a number of outlying SNPs associated with PC1. For each of the 10 sets of simulated data the total number of SNPs is given, followed by the numbers of outlying SNPs associated with PC1 that are at least as small as the largest and smallest outlying p-values observed in the real data (unadjusted p-values). Following a Bonferroni correction (adjusted p-values), the number of outlying SNPs that were below a threshold of 0.01 is also reported.

Simulation No.	Total number of SNPs	Number of outlying SNPs associated with PC1		
		Unadjusted p-val≤1.991 x 10^{-7}	**Unadjusted p-val≤1.403 x 10^{-11}**	**Adjusted p-val<0.01**
1	7492	8	0	14
2	6858	3	0	15
3	7542	0	0	2
4	7358	5	0	5
5	7101	17	1	24
6	8208	1	0	1
7	7286	0	0	1
8	7570	4	0	3
9	7296	0	0	0

10	7502	14	4	16
Total	74213	52	5	81

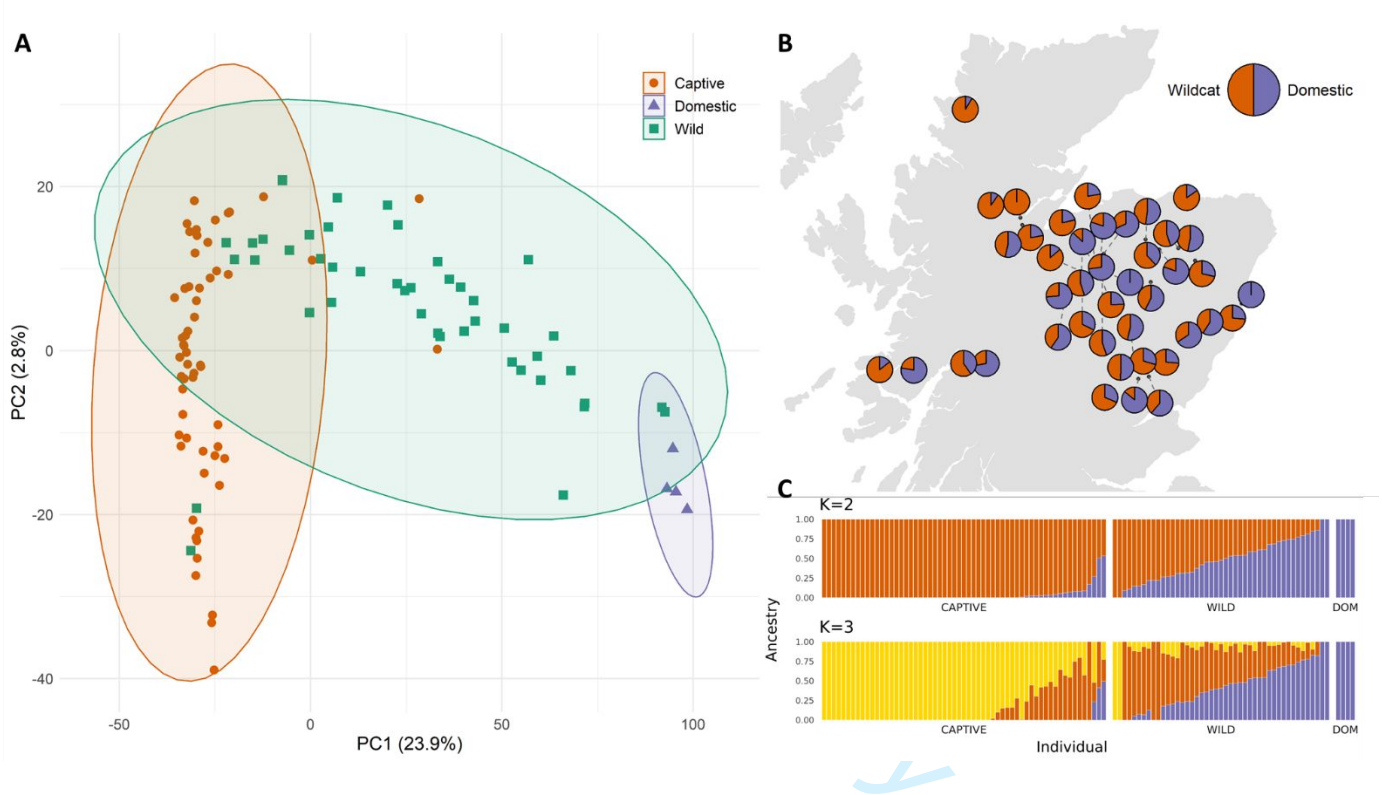

Figure 1. Population structure in the Scottish wildcat population. (A) Principal component analysis (PCA) shows a strong genetic differentiation between domestic cats and a group of putative wildcats across PC1. In the wild-living population a 'hybrid swarm' is observed, with a continuum of genetic backgrounds. (B) Sampling location of wild individuals (where known), pie charts show probable ancestry for each individual at K=2, as modelled using ADMIXTURE. (C) ADMIXTURE clustering (all individuals) at K=2 and K=3.

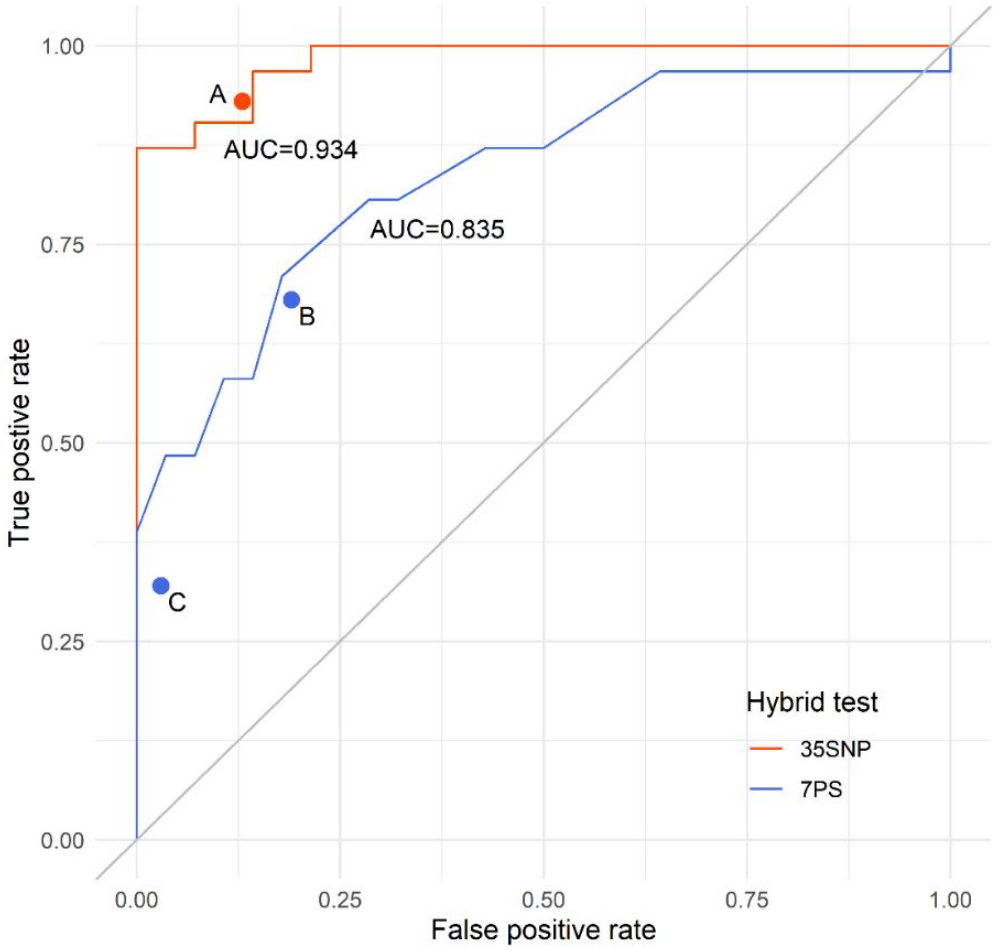

Figure 2. ROC curves for the current tests to identify wildcat/domestic hybrids: the 35 SNP genetic test (red) and seven-point pelage score (blue). True and false positive rates at the current thresholds for each test are shown using a point at the corresponding coordinate, (A) LBQ≥0.75, (B) 7PS≥17, (C) 7PS≥19.

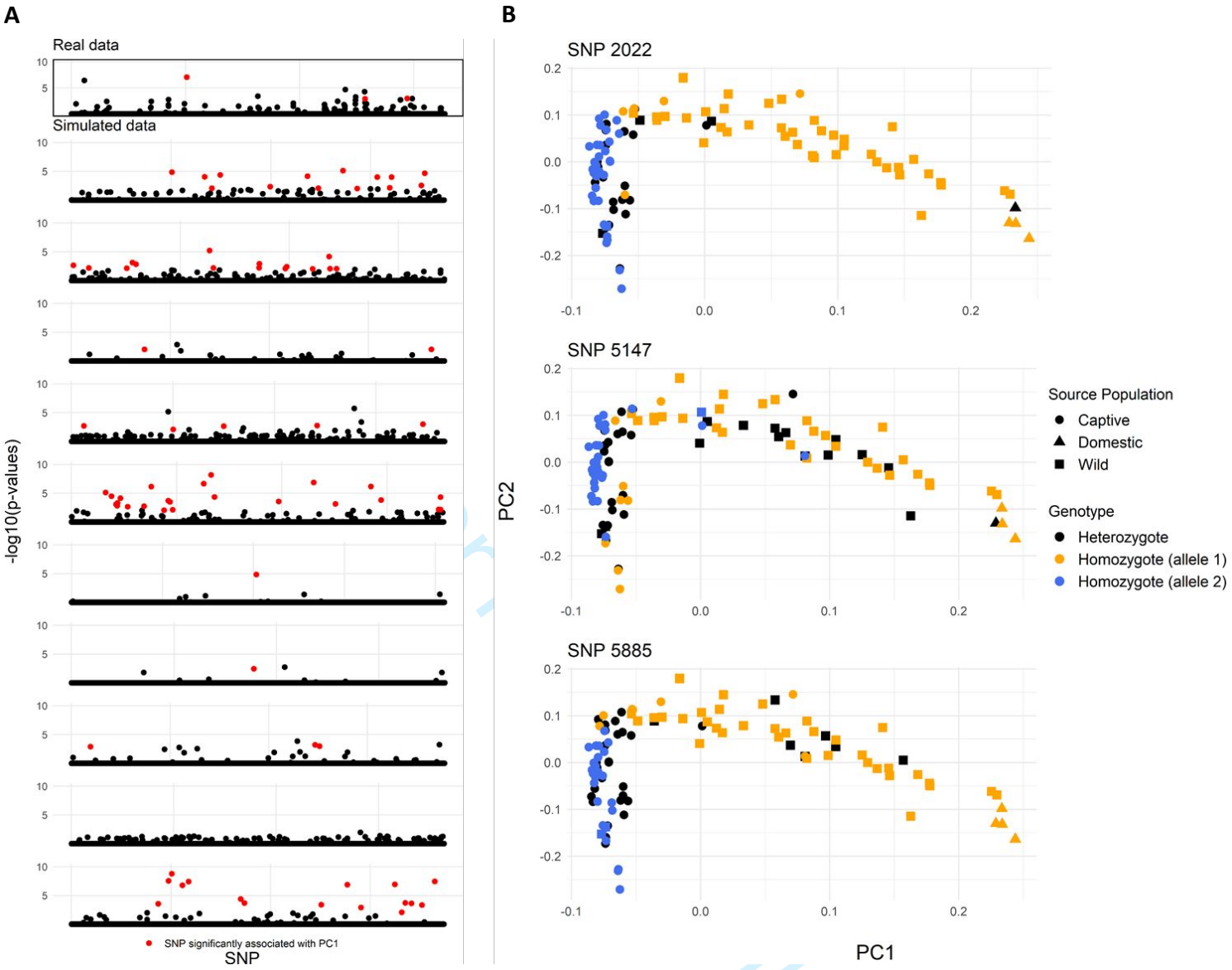

Figure 3. *Pcadapt* results for real and simulated data. (A) Manhattan plots for each set of SNPs analysed with *pcadapt*. The top row shows the real data, where these SNPs have been aligned to the domestic cat genome and are ordered by genomic position. The following rows are for simulated data. These data were simulated under a neutral model of evolution and generate a number of points classified as outliers by *pcadapt*. Red points correspond to outliers reported to be most correlated with PC1. (B) PCA plot coloured by genotype of the individual at each of the SNPs found to be significantly associated with PC1 in the real dataset.

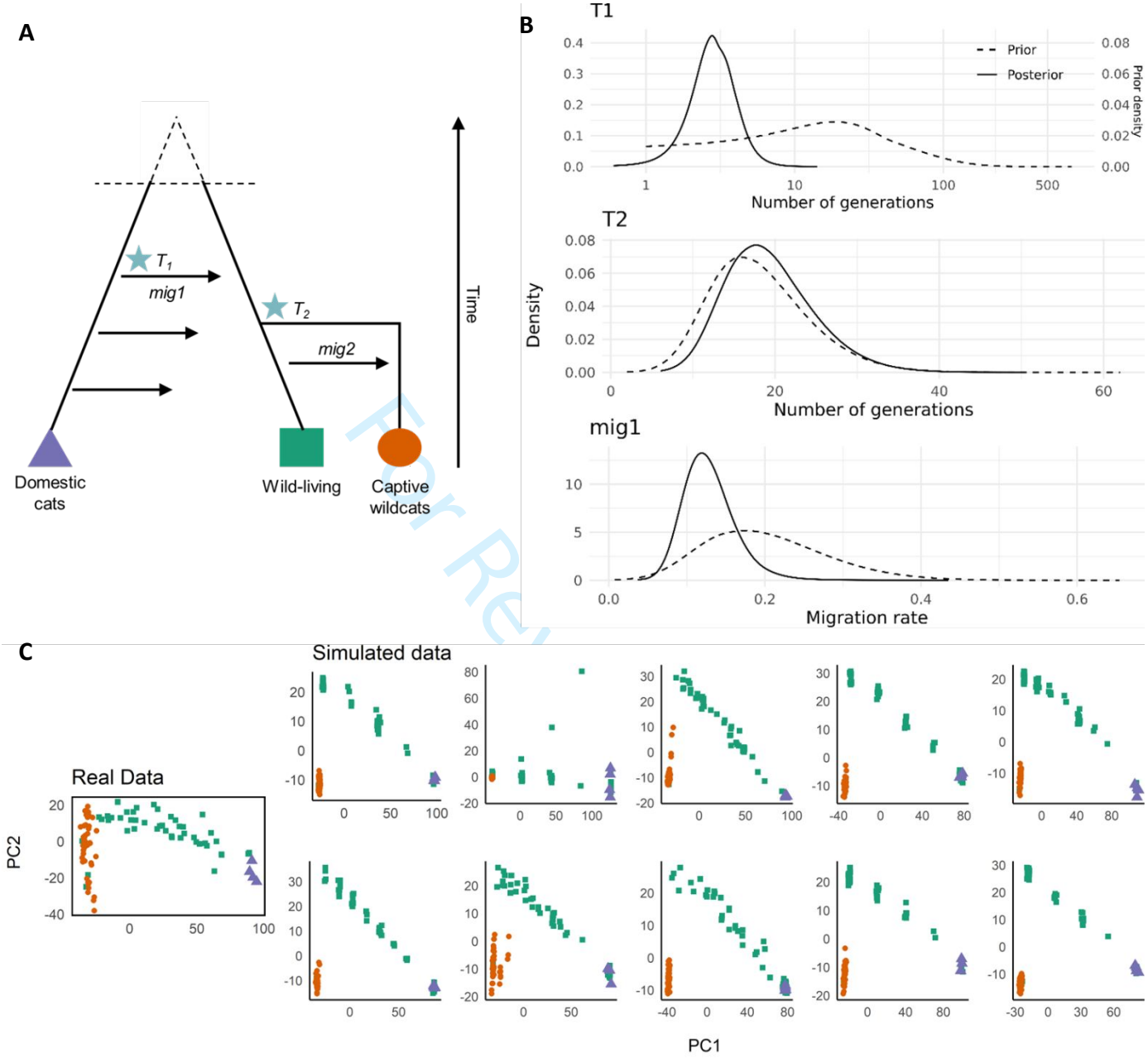

Figure 4. Modelling wildcat demography. (A) The model under which data were simulated; two parent populations (*F. catus* and *F. silvestris*) diverge under a neutral model of evolution. Gene-flow (introgression) from domestic cats begins at time T_1, at a rate of mig_1 for every subsequent generation. At time T_2 the captive population is formed from a random sample of wild-living cats. Limited gene-flow from the wild population into the captive population occurs at a rate of mig_2. (B) Prior and posterior distributions following ABC, dashed lines indicate the prior. Curves were fitted in R using *locfit* (Loader,2013). The model supports recent introgression in the Scottish wildcat population following high gene-flow from domestics. (C) PCA plots for the real data (left) and for random sample of simulated data from the posterior distribution (right). The model is broadly able to simulate the same patterns as we observe in the real data.