PhD

U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

Towards a wide genetic approach for the European wildcat (*Felis silvestris silvestris*) conservation

TOWARDS A WIDE GENETIC APPROACH FOR THE EUROPEAN WILDCAT (*Felis silvestris silvestris*) CONSERVATION:

IMPROVING NONINVASIVE MOLECULAR TECHNIQUES, POPULATION ANALYSIS AND ADMIXTURE INFERENCES

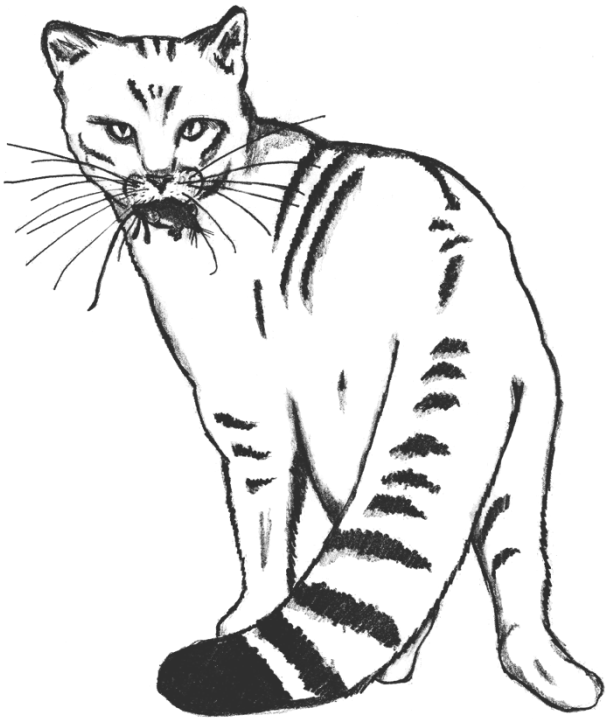
Rita Isabel Ribeiro Miranda de Oliveira
2012

Rita Isabel Ribeiro Miranda de Oliveira

2012

FC

PhD



TOWARDS A WIDE GENETIC APPROACH FOR THE EUROPEAN WILDCAT (*Felis silvestris silvestris*) CONSERVATION:

IMPROVING NONINVASIVE MOLECULAR TECHNIQUES, POPULATION ANALYSIS AND ADMIXTURE INFERENCES

Rita Isabel Ribeiro Miranda de Oliveira
Doutoramento em Biologia
Departamento de Biologia
2012

Orientador
Paulo Célio Alves, Professor Auxiliar, Faculdade de Ciências da Universidade do Porto

Coorientador
Ettore Randi, Professor, Università di Bologna

DECLARAÇÃO

Na elaboração desta dissertação, e nos termos do número 2 do Artigo 4º do Regulamento Geral dos Terceiros Ciclos de Estudos da Universidade do Porto e do Artigo 31º do D.L. 74/2006, de 24 de Março, com a nova redação introduzida pelo D.L. 230/2009, de 14 de Setembro, foi efetuado o aproveitamento total de um conjunto coerente de trabalhos de investigação já publicados ou submetidos para publicação em revistas internacionais indexadas e com arbitragem científica, os quais integram alguns dos capítulos da presente tese. Tendo em conta que os referidos trabalhos foram realizados com a colaboração de outros autores, o candidato esclarece que, em todos eles, participou ativamente na sua conceção, na obtenção, análise e discussão de resultados, bem como na elaboração da sua forma publicada.

A Faculdade de Ciências da Universidade do Porto foi a instituição de origem do candidato, tendo o trabalho sido realizado sob orientação do Doutor Paulo Célio Alves, Professor Auxiliar da Faculdade de Ciências da Universidade do Porto e investigador do Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO); e co-orientação do Doutor Ettore Randi, professor na Universidade de Bologna e director do departamento de genética da instituição de acolhimento do candidato, o Istituto Superiore per la Protezzione e la Ricerca Ambientale (ISPRA) em Itália. O trabalho laboratorial foi realizado no Centro de Testagem Molecular (CTM), unidade integrante do CIBIO, no departamento de genética do ISPRA e ainda na School of Veterinary Medicine (SVM), department of Population Health & Reproduction (PHR), University of California, em colaboração com a Doutora Leslie A. Lyons.

Este doutoramento foi apoiado fianceiramente pela Fundação para a Ciência e a Tecnologia (FCT) através da atribuição de uma bolsa individual com a referência SFRH/BD/2361/2005 e do projeto PTDC/CVT/71683/2006.

STATEMENT

In this dissertation, and in compliance with the number 2 of article 4º of the General Regulation of Third cycles of Studies in the University of Porto and with the article 31º from the DL 74/2006 from 24th March, with the new writing introduced by DL 230/2009, from 14th September, the results of published or submitted works were totally used and included in some of the Chapters of this thesis. In all these works, the candidate participated in obtaining, interpreting, analysing, discussing the results and writing the published forms.

The Faculdade de Ciências da Universidade do Porto was the home institution of the candidate, and the work was directed by Dr. Paulo Célio Alves, Auxiliary Professor of the Faculdade de Ciências da Universidade do Porto and researcher at Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO); and co-directed by Dr. Ettore Randi, Director of the Genetics Department of the Istituto Superiore per la Protezzione e la Ricerca Ambientale (ISPRA), in Italy. The laboratory work was performed in Centro de Testagem Molecular (CTM), a unit from CIBIO, in the genetics department of ISPRA and at the School of Veterinary Medicine (SVM), department of Population Health & Reproduction (PHR), University of California, in collaboration with Dra. Leslie A. Lyons.

This work was financially supported by Fundação para a Ciência e a Tecnologia through a PhD grant with reference SFRH/BD/24361/2005 and the project PTDC/CVT/71683/2006.

Aos meus pais

“In our age, when the term 'nature conservation' has become part of everyday life, and humanity is struggling with species' extinctions going faster than we can catalogue them, one can only look back with astonishment at the fact, that it has taken man more than two thousand years to realize that one is part of nature too, and, what is more, depends on it.”

Andrea Grill

“Anche il più piccolo dei felini, il gatto, è un capolavoro.”

Leonardo da Vinci

AGRADECIMENTOS / RINGRAZIAMENTI / ACKNOWLEDGEMENTS

My PhD thesis was, without any doubt, a long journey full of ups and downs. And because the years dealing with cat DNA accompanied the most demanding years of my personal life, I was particularly fortunate not to travel alone. In the following lines I acknowledge all of you that, with great dedication to the cause and/or the causer, supported me in this important step. If, by a lapse of memory (that characterize me so well!), I unfairly omitted your name, I leave you here my sincerest apologies…

I would like to start by acknowledging my supervisors Paulo and Ettore. Without you this thesis would have never been possible.

Em primeiro lugar agradeço ao Paulo, meu orientador e, acima de tudo, amigo. Obrigada por me teres introduzido no mundo da genética da conservação, quando há muitos anos te disse que o que eu queria era ajudar a conservar as espécies! Obrigada por me teres acompanhado todos estes anos com enorme disponibilidade e generosidade e por teres, desde muito cedo, acreditado nas minhas capacidades e no trabalho que desenvolvemos juntos. Com a tua enorme compreensão, as tuas palavras encorajadoras e o teu forte entusiasmo e dedicação dos últimos tempos és, juntamente com a minha família, a razão pela qual me encontro neste momento a escrever as últimas linhas desta tese. Muito obrigada.

To Ettore, my co-supervisor, I acknowledge for being so welcoming and for the great support you gave me when I first arrived in Bologna. Thank you for the opportunity of working in your great facilities at ISPRA, with you I used for the first time the sequencer and the extraction robot! Thank you for the total support and confidence in my work.

I would like also to thank Leslie Lyons, for welcoming me at your department in UC Davis and for sharing with me all your knowledge and experience in domestic cat's research. Thank you for making me staying in your house, for sharing your car, for making me feel at home. Thank you for your supporting words!

À Raquel Godinho, uma espécie de orientadora não oficial, agradeço a grande competência científica com que sempre acompanhaste e discutiste o meu trabalho, o inestimável apoio, a disponibilidade e a amizade que caracterizam os anos em que trabalhámos juntas. Obrigada pela grande consideração que sempre demonstraste por mim e pelo meu esforço.

I also express my sincere gratitude to all the institutions that supported and made this work possible: FCT, for the attribution of my PhD fellowship (SFRH/BD/24361/2005) and for supporting a research project that partially contributed to the work developed in this thesis (PTDC/CVT/71683/2006); CIBIO and Professor Nuno Ferrand for the opportunity of developing this study and learning so much with the group of outstanding researchers that work in this extraordinary research centre; ISPRA for welcoming me so well every 6 months and UC Davis, particularly the School of Veterinary Medicine at CCAH, for receiving me during 2,5 months of laboratory work.

All the molecular work presented in this thesis would not have been accomplished without the contribution of a great number of people or institutions that provided wild and domestic cat samples. I acknowledge Adriano De Faveri, ACENVA - Asociación para la Conservación y Estudio de la Naturaleza de Valladolid, Angelo Giuliani, Ângela Ribeiro, Andrea Sforzi, Aritz Ruiz-González and the University of Bask Country, Bernardino Ragni, BTVS/ICNB – Banco de Tecidos de Vertebrados Selvagens (in particular the effort from Armando Loureiro and Nuno Santos), Biró Zsolt, Claudio Monti, Direcção das clínicas veterinárias de Matosinhos, da Boa Nova e S. Francisco de Assis, Dagny Knauze, Francisco Álvares, Franz Suchentrunk, Francesca Vercillo, Gerardo Dominguez Penafiel, Giancarlo Cagnolati, Hubert Potocnik, Pablo Ferreras, Iris Eckert, João Rodrigues, José Carlos Brito, José Godoy and the EBD - Estación Biológica de Doñana, José Maria Fernández, José Luis Robles, José Vicente López-Bao, Juan Luis Ortega Arranz, Karsten Hupe, Lolita Bizzarri, Luca Lapini, Luigi Ricci, Manuela Malsaña, Martin Liberek, Matteo Visceglia, Mathias Herrmann, Marc Moes, Ellena Ballesteros-Duperón, Marcos Moleón, Emilio Virgós, José Miguel Barea-Azcón, José María Gil-Sánchez, Mauro Fabbri, Pedro Monterroso, Pirlot Romy, Sara Rocha and all anonymous people that contributed with any kind of cat samples to this thesis. I also acknowledge José Godoy and his team, especially Laura Soriano, from EBD for kindly teaching us their non-invasive procedures to deal with carnivore's scats.

I would like to thank all colleagues I met during my stay in the research laboratory of Leslie at UC DAVIS, in particular Monika Lipinski, Jennifer Kurushima, Robert Grahn, Jennifer Grahn and Leslie Bach for your support. It was a pleasure meeting you all.
Dedico um agradecimento especial às meninas Portuguesas que conheci em Davis: Catarina, Helena e Maria. Obrigada pelos excelentes momentos passados na vossa companhia e pela hospitalidade com que sempre me receberam.

Nel laboratorio di "Ozzangeles" ci sono passate tante persone e tante di queste sono diventate, oltre a splendidi colleghi di lavoro (grazie per l'aiuto in ogni cosa!), amici che tuttora fanno parte della mia vita. Ringrazio con enorme affetto gli amici che ho fatto durante gli anni passati a Bologna, una città meravigliosa che mi è rimasta nel cuore: Andrea Grill, Alice, Aritz, Barbara, Chiara, Claudia, De Faveri, Elena, Federica (grazie mille per il lavoro sul gatto che hai fatto negli ultimi mesi!), Ferri, Francesca, Ilaria, Mario, Marco, Marta, Max, Nadia, Romolo, Severino e Viglino. Romolix e Vitulano grazie per avermi coinvolto e fatto tanto ridere (col vostro migliore inglese!) nelle prime settimane passate ad Ozzano. Forse non vi siete mai resi conto di quanto siano state importanti quelle risate…"Ma Rita, what do you say in general?" and "who is the sillier, me or Severino?" mi vengono ancora in mente e mi fanno ancora ridere nei momenti più impensabili. Ringrazio specialmente i cari amici (sapete chi siete) con i quali ho fatto più e più volte l'aperitivo, sono andata in montagna o in collina, ho mangiato la pizza, la pasta, le granite, i gelati, ho visto il cinema all'aperto, i concerti in piazza, sono andata a nuotare nel fiume, in spiaggia, sono stata sgridata da Giuseppe (Oliveeeeeeeira è tardi!!!), ho combinato almeno una volta dei guai in lab, ho fatto la scalinata fino a San Luca avendo come premio una carota, ho confidato le novità della mia vita, ho chiacchierato del più e del meno senza annoiarmi, ho riso…ma riso tanto!

Ad Andrea Grill un ringraziamento speciale per tutto quello che mi ha dato il semplice fatto di aver passato del tempo insieme a te.

Os meus colegas e amigos do CIBIO e do CTM são uma espécie de família a que poucos se poderão dar ao luxo de pertencer. Eu dei. Obrigada a todos por me terem (mesmo que para alguns por breves momentos) acompanhado na concretização desta tese. Foi um prazer viver convosco as lides da genética e partilhar os mais inesquecíveis momentos de puro lazer.

Obrigada Sara, Sandra e Sara João pelo vosso empenho a resolver todos os pormenores burocráticos e técnicos e por estarem sempre disponíveis para facilitar o trabalho de todos nós. Susy, obrigada pela forma amiga com que sempre me apoiaste e pela grande consideração por mim e pelo meu trabalho. Sofia Mourão, obrigada por teres sempre encaixado nas tuas extrações os gatos que me foram aparecendo solitários. Diana e Teresa, obrigada pela revisão atenta de alguns capítulos desta tese; Sofia Silva, muito obrigada pela tua preciosa ajuda e discussão na análise de alguns dados. Às três, agradeço a constante disponibilidade e a preocupação para que tudo corresse pelo melhor.

À minha prima Sofia agradeço a ajuda indispensável para a formatação final desta tese e de algumas das suas figuras. Sem ti, teria sido muito mais difícil, demorado e, provavelmente, não concretizado!

Obrigada a todos os meus amigos, espalhados pelo mundo, que se mantiveram sempre tão próximos de mim. Um agradecimento especial dedico agora aos que, de uma maneira mais íntima, viveram comigo os altos e baixos desta viagem.

À Graça, ao Duarte e ao Pedro pelas excelentes saídas de campo que esporadicamente intercalavam com os dias passados no laboratório. Graças à vossa amizade e à vossa capacidade de me fazer rir constantemente (Graça e Pedro) ou de me fazer rir menos vezes mas tanto (Duarte), vivi e recordo esses momentos com enorme satisfação.

Ao Nuno, o "tubas" que me tirou do laboratório e me levou para o mar para nadar com peixes-lua; o formatador de figuras "upon request"; o pai que espero um dia igualar como mãe; o amigo sempre presente.

Às minhas mulaças do CTM, Diana, Helena, Teresa, Sandra, Sofia Mourão, Sofia Silva, sem as quais os dias de trabalho teriam sido decididamente muuuito menos divertidos. Obrigada pela tão presente entreajuda, pelo companheirismo, pela solidariedade e pelos agradáveis momentos de puro relaxe da pausa de almoço em formato piquenique. Helena, ainda que tenha tido a honra de partilhar contigo apenas alguns meses no laboratório, a tua boa disposição e dedicação aos outros deixou em mim marcas que trarei sempre comigo. Sandrolina, és uma fonte de frases que não lembram a ninguém e que nos fazem rir na altura em que as dizes e em todos os outros momentos em que nos lembramos delas. Obrigada pelo fim-de-semana transmontano, foi uma memorável lufada de ar fresco! Lindona, as tuas contagiantes tranquilidade e doçura foram pilares preciosos para o meu bem-estar no CTM. Pregui, este nome não te faz justiça! Obrigada pela tua tenacidade e dedicação que, principalmente nos últimos tempos, me deram tanta força. Té, o teu companheirismo e amizade de longos anos foram fundamentais. Obrigada por estares sempre presente. Di, que dizer das longas horas no laboratório que nos uniram durante anos de trabalho com os carnívoros... São a prova viva de que alegria no trabalho é possível e recomenda-se! Obrigada por isso e pelo teu enorme espírito de dedicação aos

outros do qual pude, afortunadamente, usufruir tantas e tantas vezes.

Paolo, "Vitinho", grazie mille per avermi sostituito nel lavoro della nostra cooperativa tutte le volte che la scritta della tesi mi ha fatto inchiodare davanti al computer e grazie per aver sempre cercato di spingermi a finirla.

Ao Ricardo e à família Freitas tenho tanto a agradecer...Ricardo, ainda que as nossas vidas tenham seguido rumos tão diferentes, não posso esquecer que começamos esta tese juntos e que, com o teu apoio incondicional, parti para outro país e estive tanto tempo "ausente" da nossa vida. O meu mais sincero agradecimento por teres sido um forte pilar. À tua cara família agradeço terem-me sempre feito sentir uma filha, uma irmã, uma dona (sim Gut, também tu fazes parte da família!). Permanecerão para sempre na minha memória...

Alla grandissima e altrettanto bella famiglia Bettinelli Pagani, che in un istante mi ha accolto e fatto sentire a casa mia. Grazie Rosy e Gianni per il vostro instancabile aiuto negli affari dell'agriturismo. È stato essenziale quando, per tante ore, ho dovuto sedermi davanti al computer.

Aos meus queridos pais e irmã, por todo o apoio que sempre me deram e pelo enorme esforço em proporcionar-me o melhor bem-estar nestes tempos de grandes mudanças pessoais e profissionais. Onde quer que estivesse senti sempre a vossa mão reconfortante colocada no meu ombro e soube sempre onde encontrar o meu mais fiel porto seguro. Pai, Mãe, é por e para vocês que hoje termino este projeto.

Aos meus sobrinhos Zé Miguel e Francisco, que dizer?! Frases como "Madagáscar é tão grande que deve ter metro!" e "Fui ver o Porto na zona vic!" são verdadeiras pérolas de boa disposição que me iluminaram a alma, principalmente quando mais oscilei. As mais dolorosas lágrimas de saudade soltaram-se à vossa memória. Obrigada pelo vosso tão genuíno amor.

Ao meu gato Primo (que tão bem me soube domesticar!) e ao meu cão Pinto (um companheiro para a vida!), obrigada pela vossa ternura e pela companhia que me fizeram nos últimos tempos.

Ao meu Aronne, infindo companheiro que me aconchega com enorme carinho e paciência. Agradeço-te com desmedido afeto as pequenas e grandes coisas que preencheram as nossas vidas e que me ajudaram a chegar até aqui. Obrigada por me teres acompanhado e vivido em Portugal, por leres estas palavras em Português e por me suportares nas horas de maior desânimo. Obrigada por sonhares comigo e, acima de tudo, me encorajares a concretizar esses sonhos.

Finalmente, agradeço ao meu pequeno rebento que, enquanto escrevo estas palavras, cresce silencioso dentro de mim e, ainda longe de saber o que é isto da vida e dos seus afazeres, se transformou já na minha mais comovente fonte de inspiração.

RESUMO

A introgressão de genes de gato doméstico em populações naturais do gato-bravo Europeu (*Felis silvestris silvestris*) é considerada um dos principais problemas de conservação para este ameaçado felino. Simultaneamente, a perda e fragmentação do habitat resultou numa evidente redução da distribuição original da espécie na Europa, causando importantes declínios demográficos e altos níveis de isolamento. Nesta dissertação, procuramos abordar questões relacionadas com os dois temas - hibridação artificial e fragmentação populacional – através de uma perspectiva baseada na genética da conservação.

Entre as populações Europeias, a Península Ibérica é comummente reconhecida como uma das regiões onde o gato-bravo mais se encontra em declínio pela ação concomitante da fragmentação do habitat e da hibridação. De modo a proceder à primeira descrição detalhada dos níveis de diversidade genética e níveis de miscigenação nesta região, foram analisados 12 microssatélites em 98 amostras provenientes de Portugal (*Artigo I*). Os resultados demonstraram a inexistência de uma diferenciação significativa entre as diversas populações selvagens. Por sua vez, gatos-bravos e domésticos demostraram níveis significativos de diferenciação. A análise consensual de diferentes metodologias bayesianas permitiu também a identificação de quatro indivíduos com ancestralidade híbrida. Posteriormente, este estudo foi alargado a toda a Península Ibérica, analisando um total de 184 amostras de gatos selvagens e domésticos (*Artigo II*). O nível de hibridação observado foi baixo (6.67%), embora amplamente disperso. No entanto, este valor poderá representar uma subestimativa do valor real, uma vez que a avaliação do poder estatístico das análises bayesianas efectuadas revelou 100% de probabilidade de identificar híbridos de primeira geração, mas apenas 91% dos genótipos F2 e 85% dos retrocruzamentos. Os resultados expostos ao longo deste capítulo reforçaram a necessidade de melhorar a capacidade de analisar geneticamente este problema, tanto relativamente ao incremento nos níveis de amostragem da espécie como no que diz respeito à identificação dos eventos de miscigenação.

Com o objectivo de aumentar significativamente o número de gatos analisados, alargamos a nossa abordagem à amostragem não-invasiva. Para desenhar a estratégia mais eficiente, superando em grande parte os obstáculos inerentes à análise deste tipo de amostras, procedemos à revisão exaustiva (*Artigo III*) das principais publicações científicas neste campo, não só à luz do estudo de populações selvagens, mas também nos contextos forense e médico lidando com baixa qualidade e quantidade de DNA. A revisão destes trabalhos revelou a existência de um grande potencial para o desenvolvimento de metodologias relativamente simples para responder a quase todas as questões genéticas até então realizadas apenas com amostras de boa qualidade. Desde a recolha de amostras à análise de dados, revimos e discutimos os mais marcantes avanços técnicos que podem contornar as principais desvantagens da genética não-invasiva, enquanto delineamos quais as principais necessidades de pesquisas adicionais.

A dificuldade da aplicação de estratégias moleculares não-invasivas na amostragem do gato-bravo, prendem-se não só com a qualidade dos genótipos produzidos mas também com a grande dificuldade de identificar morfologicamente os excrementos produzidos por esta espécie. Ao mesmo tempo, a identificação

de pêlos é frequentemente subjetiva e raramente específica. De modo a superar estas dificuldades e direcionar os esforços de genotipagem para a nossa espécie de interesse, foi desenvolvido um teste molecular de execução simples, rápida e económica, que permite, através da analise de variação existente em 221pb do gene IRBP, a distinção das 16 espécies de carnívoros que ocorrem no Sudoeste da Europa (*Artigo IV*). O teste revelou um sucesso de identificação em amostras não-invasivas de 78,57%, demonstrando a sua utilidade como ferramenta de identificação de amostras de gato-bravo para a realização de estudos específicos da espécie. A aplicação deste método permitirá ainda minimizar custos, tempo e erros em estudos não-invasivos dos carnívoros distribuídos pelo Sudoeste da Europa.

Por fim, procurou-se aprofundar, simultaneamente, o poder das análises de padrões de subestruturarão populacional e a detecção de eventos de miscigenação no continente europeu. Neste contexto, começamos por genotipar mais de 1000 gatos, amostrados em 19 localidades por toda a Europa, utilizando 38 microssatélites altamente polimórficos (*Artigo V*). Os resultados obtidos revelam que os três grupos taxonómicos presentes na Europa – gato-bravo Europeu, gato-bravo Africano e gato doméstico - se encontram bem diferenciados, e que as populações naturais de gato-bravo Europeu se encontram geneticamente estruturadas em 5 grandes áreas geográficas e 10 bem reconhecidas subpopulações. Entre as últimas destaca-se a região Este da Alemanha, que se diferencia totalmente das restantes e revela sinais de *bottleneck* populacional. A análise de hibridação confirma a documentada origem híbrida dos gatos na Hungria e na Escócia, e asserta o carácter globalmente não-miscigenado das restantes populações. No entanto, pôde detectar-se a presença de híbridos esporádicos em diversos locais e os resultados de simulações revelam que estes números podem ainda representar subestimativas dos valores reais, mesmo com o número elevado de *loci* e indivíduos analisados. Tal constatação impulsionou o trabalho desenvolvido no *Artigo VI*. Neste último trabalho procurou-se identificar novos polimorfismos que poderão melhorar, sobretudo, a nossa capacidade de identificar retrocruzamentos. Para tal, analisamos a diversidade genética em gatos-bravos e domésticos em 158 SNPs dispersos por todo o genoma do gato. Embora nenhum dos marcadores genéticos tenha demonstrado ser diagnóstico para a distinção das duas subespécies, diversos *loci* revelaram valores elevados de diferenciação genética. Por exemplo, 35 SNPs demonstraram um valor médio de diferenciação (F_{ST}) de 0.740. A análise do poder de afiliação usando métodos bayesianos demonstrou ainda que o conjunto de 158 SNPs apresenta as mais sólidas inferências de hibridação obtidas até ao momento, com 92 a 100% dos retrocruzamentos corretamente identificados. O desenvolvimento deste conjunto de SNPs (ou de outros) poderá ser também extremamente importante para a aplicação de metodologias não-invasivas, uma vez que representam uma ferramenta de elevado potencial para a amplificação de DNA de baixa qualidade e concentração. Além disso, a sua associação com microssatélites poderá melhorar significativamente o nível de discriminação entre os dois *taxa* e a capacidade de avaliação de miscigenação genética.

Palavras-chave: gato-bravo Europeu, gato doméstico, hibridação, fragmentação populacional, microssatélites, polimorfismos nucleotídicos simples (SNPs), genética não-invasiva, genética da conservação.

SUMMARY

Introgression of domestic cat genes into European wildcat (*Felis silvestris silvestris*) populations is considered one of the main conservation problems for this endangered feline. Concomitantly, habitat loss and fragmentation resulted in a strong reduction of wildcats' range in Europe, while causing important demographic declines and high levels of isolation. Both events may strongly jeopardize wildcat's fitness and evolutionary survival. In this thesis the questions related with the artificial hybridization and populations' fragmentation are addressed, using a conservation genetics perspective.

Among European populations, the Iberian Peninsula is commonly recognized as one of the regions where wildcats most are in decline by the concomitant action of habitat fragmentation and hybridization. In order to carry out the first detailed description of the levels of genetic diversity and admixture in this region, we analysed 12 microsatellites in 98 putative wildcats from Portugal (*Paper I*). The results showed no significant differentiation among the different wildcat populations. In turn, wild and domestic cats demonstrated significant levels of differentiation. The consensus analysis of different bayesian methodologies allowed the identification of four individuals with hybrid ancestry. Subsequently, we extended this work to the whole Iberian Peninsula, analysing a total of 184 samples of wild and domestic cats (*Paper II*). The degree of observed hybridization was low (6.67%), although widely dispersed. This scenario may, however, represent an underestimate of the true value, since the evaluation of the statistical power of the bayesian analyses revealed 100% probability of identifying first generation hybrids, but only 91% of the genotypes of F2, and 85% of the backcrosses. The results obtained reinforced the need to improve the ability to genetically analyse this problem, both in what concerns the collection of wildcat samples and the identification of admixture events.

In order to significantly increase the number of analysed cats and obtain samples from other areas of the species distribution we aimed at extending our approach to samples collected noninvasively. To be able to design the most successful and accurate scheme for wildcat's studies we extensively reviewed (*Paper III*) the major scientific publications not only in the light of wildlife non-invasive sampling but also in the context of forensic and medical work dealing with low quality and quantity of DNA. The number of studies is outstanding and the great potential to routinely use these sampling strategies to answer almost all genome-related questions achieved with good-quality samples is evident. From samples' collection to data analysis, we reviewed and discussed technical advances that may circumvent major weaknesses of non-invasive genetics, and outlined some needs for additional research.

The difficulty of applying molecular strategies in noninvasive wildcat samples is linked not only with the quality of the produced genotypes, but also with the great difficulty of morphologically identifying scats produced by this species. At the same time, the identification of hair is often subjective and rarely specific. To overcome these difficulties and direct genotyping efforts for our species of interest, we developed a simple, quick and inexpensive molecular test based on polymorphism at 221bp of the IRBP gene to distinguish all 16 carnivore species occurring in South-western Europe (*Paper IV*). The method proved to be successful in

identifying non-invasive samples of 78.57%, demonstrating its utility as a genetic tool for the identification of wildcat samples. Moreover, the application of this method will minimize costs, time and errors in noninvasive studies of carnivores distributed by Southwest Europe.

Finally we aimed at improving the power to detected population substructure in European wildcats and to individuate events of miscegenation between wild and domestic cats. In this context, we began by genotyping over 1000 cats, sampled at 19 locations throughout Europe, using 38 highly polymorphic microsatellite (*Paper V*). The results show that the three taxonomic groups present in Europe - European wildcat, African wildcat and domestic cat - are well differentiated, and that natural populations of the European wildcat are genetically structured into 5 large geographic areas 10 and well-recognized subpopulations. Among the latter, the Eastern Germany deserves particular attention, since it genetically differs from the other and shows signs of population bottleneck. The hybridization analysis confirmed the documented hybrid origin Hungarian and Scottish cats, and asserted the overall non-admixed character of the remaining populations. However, it could detect the presence of sporadic hybrids at various locations and results of simulations also show that these numbers could represent an underestimation of the actual values, even with the high number of analysed individuals and loci. This finding prompted the work developed in *Paper VI*. In this latter work we aimed at identifying novel polymorphisms that could improve our ability to identify backcrosses. To achieve this purpose, we analysed the genetic diversity in wild and domestic cats at 158 SNPs distributed throughout the genome of the cat. Although none of the genetic markers have proved to be diagnostic for distinguishing the two subspecies, several loci revealed high levels of genetic differentiation. For example, a set of 35 SNPs showed an average value of differentiation (F_{ST}) of 0.740. The analysis of the statistical power of our bayesian allocations further demonstrated that this set of 35 SNPs offer the strongest hybridization inferences obtained so far, with 92 to 100% of backcrosses correctly identified. The development highly diagnostic SNPs, as developed in this study, may also be very important for the application of noninvasive methodologies, since this type of loci represents a high potential tool for the amplification of low quality and quantity DNA. Moreover, its association with microsatellites might significantly improve the level of discrimination between wild and domestic cats and further increase our ability to evaluate admixture patterns.

Keywords: European wildcat, domestic cat, hybridization, fragmentation, microsatellites, single nucleotide polymorphisms (SNPs), non-invasive genetics, conservation genetics.

TABLE OF CONTENTS

LIST OF FIGURES:

CHAPTER 1

CHAPTER 2

Paper I

Paper IV

CHAPTER 4

Paper V

Paper VI

LIST OF TABLES:

Paper VI

CHAPTER 5

CHAPTER 1

General Introduction

"El gato, solo el gato apareció completo y orgulloso: nació completamente terminado, camina solo y sabe lo que quiere."

Pablo Neruda

It is believed that, as early as in the Upper Palaeolithic era, hunter-gatherers had already initiated the process of animal domestication. The earliest reliable evidence of this activity was obtained from archaeological remains in Israel: approximately 11,500 years ago (YA), a woman was buried with a puppy whose dental measurements proved to be compatible with the ones of a domestic dog (Davis and Valla, 1978). However, it was later on, in the beginning of the Neolithic period, that civilizations began to better recognize and well explore animals' qualities for their benefit. It all happened roughly 10,000 YA: a stout change from hunting-gathering to agricultural practices took place, resulting in pivotal restructuring of human societies, global alterations in biodiversity and profound reshaping of Earth's landscape (Zeder, 2008). At that time, humans brought animals under their care mainly to produce food, provide protection and help with servile labour, strongly dedicating themselves to one of the most important non-human relationships of all.

1. 1. Animal's domestication: a complex evolutionary process

The high number of archaeological and molecular studies focusing on domestication clearly illustrates the extraordinary complexity of this human-mediated evolutionary process (see Dobney and Larson 2006; Zeder 2006 for reviews). Not only it has been perpetrated in a number of independent places, but also several animals have been domesticated through a surprisingly high number of episodes. Among animals, pigs, cattle, chickens and horses provide exceptional evidences of such intricate process. The more studies explore the history of their domestication the more independent geographically separated events and/or ancestral wild subspecies are discovered to have been involved (see Jansen *et al.* 2002; Bruford *et al.* 2003; Larson *et al.* 2005; Eriksson *et al.* 2008 for details). From a genetic perspective, the dog is one of the best examples that unravelling domestication might be a complicated task. While Savolainen *et al.* (2002) have provided mitochondrial DNA evidence for an East Asian origin of domestic dogs, a more recent study using a much larger data set of nuclear loci points to the Middle East as the source of most genetic diversity in dogs and a more likely centre for their domestication (vonHoldt *et al.* 2010). On the other hand, rabbit domestication is apparently much simpler and the most probable scenario (based on mtDNA, Y and X Chromosomes and autosomes' variation) implies a recent and exclusive origin from French *Oryctolagus cuniculus cuniculus* (Carneiro *et al.* 2011). Today, the identification of wild ancestors, the precise time and location of the advent of modern domesticates and the processes leading to their current genetic diversity are still active areas of multidisciplinary research to which archaeogenetics is, more recently, significantly contributing (Zeder 2006).

Amongst the diverse spatio-temporal origins of animal domesticates, archaeological, cultural and genetic evidences point to the Late Pleistocene in the south-western Asia's Fertile Crescent as one of the earliest and most important centre of domestication (Diamond, 2002; Bruford *et al.* 2003; Figure 1).

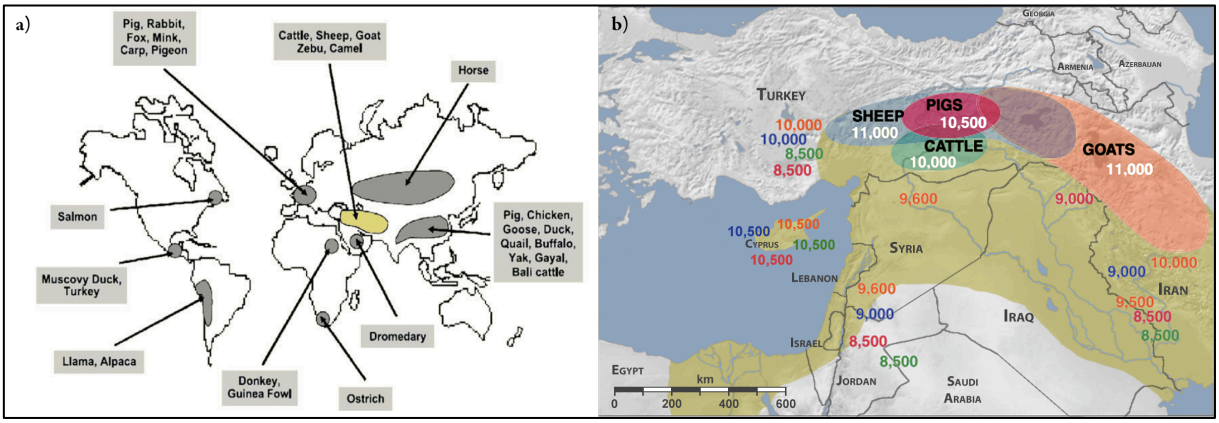


Figure 1. a) Hotspots of animals' domestication, mainly Southwest Asia (the Fertile Crescent and its eastern margin), East Asia (China and countries south of China including Indus valley) and South America (the Andean Chain). Adapted from Gregory (2009) b) Origin and dispersal of domesticates in the Fertile Crescent (adapted from Zeder 2008). Shaded areas and numbers represent, respectively, the general region and the approximate dates in which initial domestication expansion is thought to have taken place. The Fertile Crescent is part of present day Iran, Iraq, Syria and Turkey.

Among all animals that have been domesticated, the cat is one of the most fascinating ones. In one hand, domestic cats symbolize the human power to manipulate animals' evolution: from a single ancestral species, dozens of breeds have been selected into an exceptional variety of ways, seeking often opposite purposes (e.g. hairy *versus* hairless individuals). On the other hand, cat's domestication manifestly expresses the crucial role that the animal itself may have in creating such long lasting relationship with humans: cat's domestication most likely emerged from cats will to coexist, interact and profit from us, rather than being a simple result of human domain.

1.1.1. The self-domestication of the cat

For many years, the Nile Valley in Ancient Egypt was considered the centre of cat's domestication approximately 8,000-4,000 YA. As an alternative, some researchers had even proposed that cat domestication occurred in a number of different locations, with each event spawning a different cat breed (Clutton-Brock *et al.* 1999). However, archaeological and genetic findings revealed that this process started, almost certainly, long time before and in a single geographic area. The earliest evidence of a cat-human close relationship was found in Cyprus deposits, dated at 10,600 YA (Vigne *et al.* 2012), and molecular studies based on both mitochondrial genes and microsatellites suggest that cats' domestication likely began when humans started to build the first civilizations over the Fertile Crescent (Driscoll *et al.* 2007; Lipinski *et al.* 2008). But whereas other wild animals were voluntarily domesticated for specific tasks, cat's domestication most likely began as a commensalism-driven event. Since first Neolithic farmers started storing growing crops, rodent pests began infesting grain stores. At that time, neighbouring wildcat populations probably realized that human settlements were an ideal concentration of food resources, and farmers immediately took advantage of having

cats around eating the pests and started to encourage their presence with leftovers. In an evolutionary point of view, selection probably favoured those cats that were able to live in human-dominated environments and made them to proliferate. It is, therefore, reasonable to believe that humans did not try to influence breeding and behaviour of the first house cats, and they probably intermixed frequently with local wildcats, aspects that worked against rapid domestication (Driscoll *et al.* 2009). Mitochondrial DNA inferences corroborate this early admixture hypothesis: cat's domestication seems to have been perpetrated in multiple occasions and to have been influenced by the incorporation of at least five wildcats' matrilines, by means of repeated crossbreeding of recent domesticates with wild conspecific females (Driscoll *et al.* 2007).

Succeeding domestication, cats promptly colonized the entire world and became common in Europe and Asia, mainly by taking advantage of major land and sea trade routes. Subsequently, the first "natural" ancient breeds arose in specific geographic regions, where newly arrived cats experienced some level of isolation and gradually acquired and fixed distinctive alleles through genetic drift (Lipinski *et al*. 2008; Menotti-Raymond *et al.* 2008). But it was only by the time of the industrial revolution (late 18th–early 19th century) that pet cat owners were selectively mating to produce fancy breeds, with the majority of breeds dating back no further than 150 years (O'Brien *et al*. 2008). The first domestic cat show was held at the Crystal Palace in England in 1871. Because no selective breeding was perpetrated for long time, true domestication of cats arrived very recently and might still be an ongoing event. The modern timing of breeds' development and the allowable intercrossing between certain breeds in recent generations originate only modest phylogenetic and population genetic partitions of cat breeds when compared to other domesticated animals such as dogs. Nonetheless, there is a recognizable and diagnostic population structure among the domestic cat breeds (Lipinski *et al*. 2008; Menotti-Raymond *et al.* 2008).

While date and location of cat's domestication have been widely debated until recent years, domestic cats' ascendance soon seemed rather clear: archaeological (Kitchener, 1991; Yamaguchi *et al.* 2004a), morphologic (Yamaguchi *et al.* 2004a), and genetic (e.g. Randi and Ragni, 1991; Johnson and O´Brien, 1997; Driscoll *et al.* 2007) evidences concur that the domestic cat derived, very recently, from the Near East group of the wildcat species *Felis silvestris*: the *Felis silvestris libyca* subspecies.

1.2. Taking a walk on the wild side

The wildcat (*Felis silvestris*) is a medium-sized terrestrial carnivore from the FELIDAE family with one of the widest geographical distributions among all felids: the species range from Western Europe, throughout most of Africa, and from Arabia and southwest Asia to China (Nowell and Jackson, 1996; Sunquist and Sunquist, 2002). Archaeological remains suggest that the species probably appeared in Europe around 450,000-200,000 YA (Kitchener, 1995; Sommer and Benecke, 2006), descending from the Martelli's cat (*Felis lunensis* Martelli 1906), which was found in the continent during the early Pleistocene (Kitchener, 1991; Nowell and Jackson, 1996). Later on, most probably through several "out of Europe" migratory waves, wildcats colonized the entire Old World. Studies based on allozymes variation (Randi and Ragni, 2001), craniological

measurements (Yamaguchi *et al.* 2004a) and mitochondrial DNA diversity (Driscoll *et al.* 2007) concur that during the late Pleistocene European wildcats expanded suddenly and rapidly to the Middle East. Subsequently dispersed to east giving rise to the current Asian populations. Finally, expanded to south and west where colonized a good part of the African continent (Kitchener and Rees, 2009). This rapid expansion may have occurred even as recently as in the last c. 50,000 years, and European and African wildcats are supposed to have diverged c. 20,000 YA (Randi and Ragni, 1991).

Through their evolutionary history, characterized by at least three different range expansions punctuated by two differentiation events (Yamaguhi *et al.* 2004), wildcats have experienced several glacial–interglacial cycles (Kitchener and Dugmore, 2000), and the spatial changes they suffered on their ranges may explain current biogeographical patterns of morphological and molecular variation (Kitchener and Rees, 2009). Although several studies based on morphological characteristics suggest different wildcat species (e.g. Pocock, 1951; Haltenorth, 1953; Schauenberg, 1969, 1977; Kitchener, 1991), first molecular data promptly revealed that geographic groups of *Felis silvestris* corresponded not to separated species but to at least three ecologically, geographically and genetically divergent (but inter-fertile) subspecies: *F. silvestris silvestris* Schreber 1777, in Europe; *F. silvestris ornata* Gray 1832, in Asia; and *F. silvestris libyca* Forster 1780, in Africa (Randi and Ragni, 1991). This classification is currently recognized by The Council of Europe, by the World Conservation Union (IUCN) and by a number of wildcat researchers, but the most recent molecular study on wildcats' taxonomy strongly suggests that *Felis silvestris* comprises not only those but also two additional subspecies: *F. silvestris cafra* Desmarest 1822, which inhabits sub-Saharan Africa and is thus distinct from *libyca* (that is restricted to North Africa and the Middle East), and *F. silvestris bieti* Milne-Edwards 1892, the Chinese desert cat whose distribution is limited to the North East corner of the Tibetan Plateau (Driscoll *et al.* 2007; Figure 2). This polytypic species finally includes the most cosmopolitan of all felids: the domestic cat *F. silvestris catus*.

Since contemporary wildcats are considered widely distributed in Europe, Asia and Africa, the IUCN Red List classify them as "Least Concern". However, the low-risk category attributed to the species is clearly misleading. Even though the wildcat is the most widespread and probably the most numerous felid in the World, it is now critically endangered in several areas where its previous distribution was wider, and the global population trend is to decrease in future years (Driscoll and Nowell, 2009). Among the five wild subspecies of *Felis silvestris*, the European wildcat is probably the one that suffered the severest demographic changes and from which more populations are considered currently at risk.

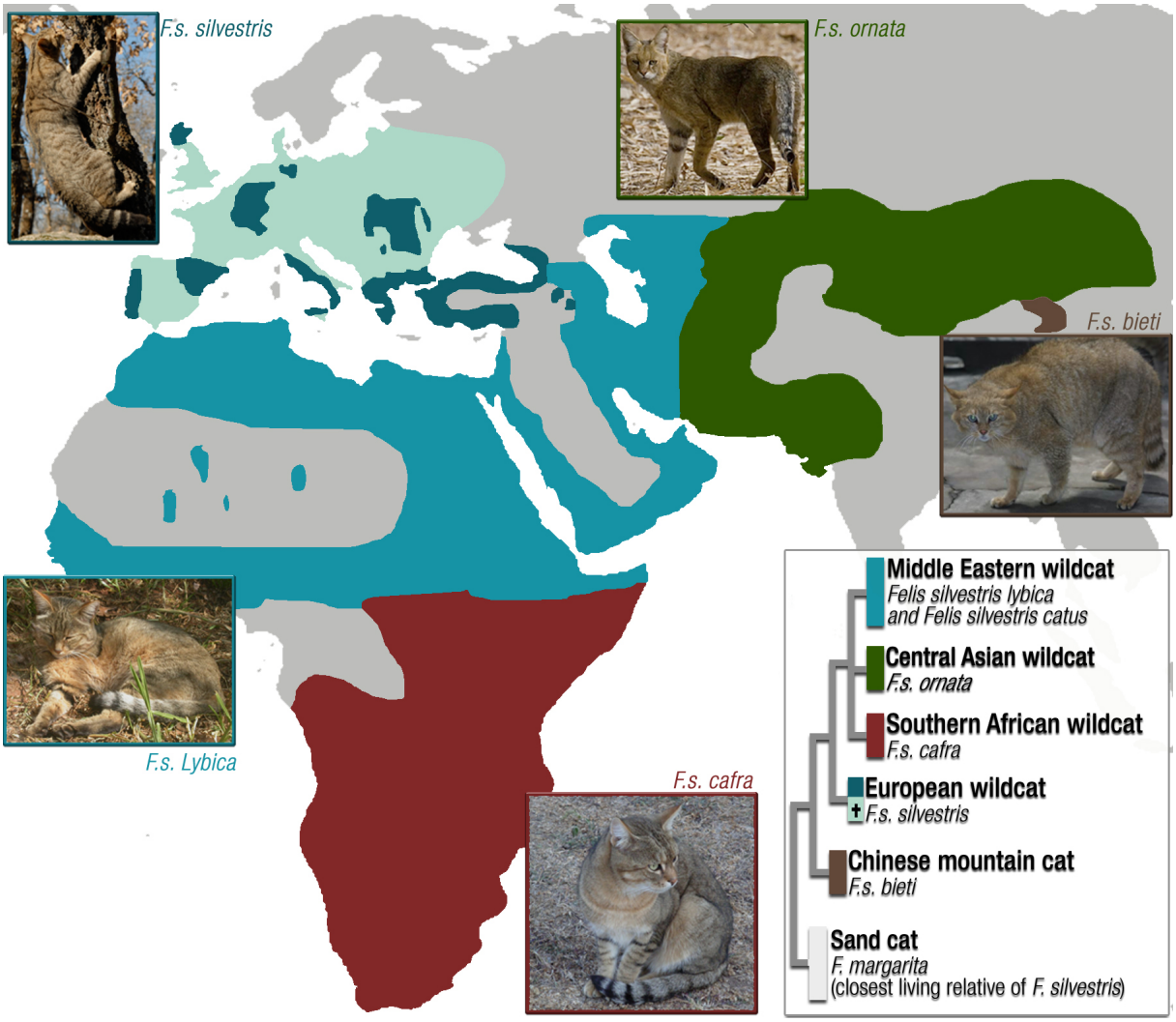


Figure 2. Old world distribution range of the *Felis silvestris* complex and phylogenetic relationship among the five recognized wild subspecies and domestic cats. Historical distribution of the European wildcat (*Felis silvestris silvestris*) is also represented in light green. Based on Driscoll *et al.* 2007.

1.2.1. Wildcats in Europe

According to the Atlas of European Mammals (Mitchell-Jones *et al.* 1999), *Felis silvestris silvestris* was formerly distributed throughout the whole Europe in continuous and stable populations, extending across the whole continent from the Iberian Peninsula to the Caucasus, from sea level to 2,250 m in the Pyrenees (Nowell and Jackson 1996; Palomo and Gisbert 2002). Across their entire distribution, while permanently needing a certain degree of landscape diversity and distance to human settlements and structures, wildcats have always shown an impressively variable habitat preference: in temperate bioclimatic areas, their presence seems strongly dependent on forests availability (Germain *et al.* 2008; Klar *et al.* 2008; Hertwig *et al.* 2009; Jerosch *et al.* 2010), while in Mediterranean landscapes they successfully occupy mosaic shrub-pasturelands

(Lozano *et al.* 2003; Monterroso *et al.* 2009). Additionally, this small felid is also particularly plastic in its prey selection. While specialized in rodents in most of its forestry distribution range (Nowell and Jackson, 1996; Sarmento, 2006; Germain *et al.* 2008), the wildcat majorly feed on wild rabbit (*Oryctolagus cuniculus*) when the species is present, namely in southern Iberian Peninsula. Wild rabbit abundance may be so important that it may represent the most valuable variable in explaining wildcat distribution in the Iberia (Malo *et al.* 2004; Lozano *et al.* 2006; Monterroso *et al.* 2009).

Despite of the great adaptability of the European subspecies to opposing niche and food resources, severe declines and local extirpations between the 18th and 19th centuries resulted in a highly fragmented relict distribution, and wildcats populations, albeit wide ranging, are represented today by isolated nuclei (Stahl and Artois 1991; Nowell and Jackson 1996; Peichocki 2001; Figure 3).

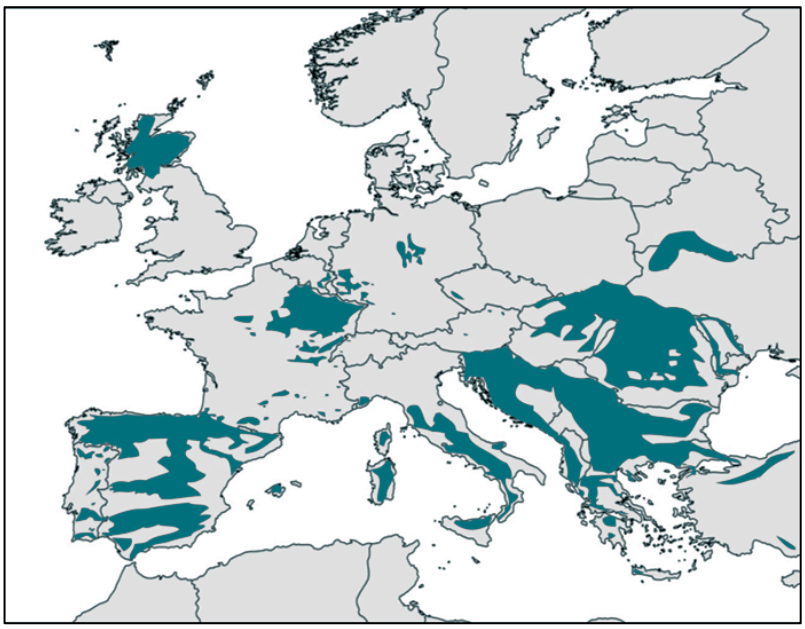


Figure 3. Fragmented distribution of the European wildcat (*F. s. silvestris*) in Europe. Adapted from Grabe and Worel (2001).

Current patterns of wildcat distribution in Europe are a reflection of past and strong demographic declines that have been documented for most of the species range. The great majority of wildcat's endemic populations are today considerably small and fragmented (e.g. France, O'Brien *et al.* 2009; Germany, Germain *et al.* 2008), and many of them are known to be critically decreasing (e.g. Poland, Wolsan *et al.* 2001; Scotland, Yamaguchi *et al.* 2004b). In Scotland, the population has been reduced to approximately 400 individuals (IUCN, 2007). Contrastingly, small increments on the species range have been detected in Switzerland and Belgium (IUCN, 2007), and although previously extinct in the Netherlands (Nowell and Jackson 1996), wildcats may be recolonizing from German populations in the Eifel or Ardennes forests (Canters *et al.* 2005). Different studies in central Europe also report that the species distribution seems stable or even expanding in the later years, due to some increment in forest cover as a result of massive agriculture

abandonment (Raimer 2006; Simon 2006 in Hertwig *et al.* 2009; O'Brien *et al.* 2009). Nevertheless, the recovery of these populations is expected to be slow and exceptionally vulnerable.

Sicily is the only Mediterranean island populated by European wildcats. The existent populations on other islands (including Sardinia and Corsica) may have two possible origins. They might be feral domestic cats, that were introduced by Neolithic navigators into these and other Mediterranean islands, about 8,000–6,000 YA, at an early stage of domestication (Gippoliti and Amori 2006, Kitchener *et al.* 2010), or populations derived from the involuntary introduction of the *Felis silvestris libyca* subspecies by the same navigators coming from North Africa (Ragni 1981; Pierpaoli *et al.* 2003). The earliest Sardinian record documents the presence of the species at least by 3,000 YA (Vigne 1992). The main questions concerning the relation of the Cretan population to others in continental Europe as well as how, when and from where the cat arrived in Crete, are still open questions. Cretan wildcats might even be the possible result of the transport of two wildcat subspecies to the island by humans: *F. s. libyca* from the Mediterranean Afro-Asiatic coast and *F. s. silvestris* from the Balkan Peninsula (Belardinelli, 2001).

Among today's European populations, the Iberian ones are recognized as the most threatened, and among the most poorly studied. In Portugal, wildcats are suspected to have decreased at a rate of more than 30% over three generations and are, consequently, listed as "Vulnerable" in the Red Data Book (Queiroz *et al.* 2006). The most detailed assessment of wildcat distribution suggests that the species is present in scattered and isolated nuclei (Fernandes, 2005), and seems to be dramatically decreasing (Sarmento *et al.* 2009). According to the Spanish Atlas of Terrestrial Mammals, the wildcat in Spain is "Near Threatened", being distributed across most of the country, except for the Balearic and Canary Islands, in fragmented populations (García-Perea 2002). Similarly to what has already occurred in Portugal, Spanish populations may be considered "vulnerable" in the near future (Palomo *et al*. 2007).

Despite the present global knowledge on European wildcat's distribution, most inventories done so far should be cautiously interpreted. Assessing the status of natural populations is remarkably difficult because wildcats are elusive and cryptic animals and, moreover, it is frequently challenging to distinguish between wild specimens, domestic free-ranging cats or hybrid individuals (Ragni, 1993; Nowell and Jackson 1996; see also bellow). Taking this into consideration, the current known European distribution of the species may well not correspond to the reality, especially in places where recent studies based on morphology, ecology and genetics were still not implemented. Yet, regional conservation status attributed in different countries reveals that European isolated populations struggle today with a number of serious threats.

1.3. Major risks faced by European wildcats

Since 1992, the European wildcat is under special protection in Europe. Not only the species is considered Near Threatened by the European Community and classified as Threatened or Vulnerable on a national basis in many countries, it is also protected by the Convention on International Trade in Endangered Species

(CITES Appendix II), listed on the EU Habitats and Species Directive from 12.05.1992 (Annex IV) and included at the Appendix II of the Bern Convention from 19.04.1979.

Intensive human persecution mainly for pest control, trophies or during (un)directed predator control campaigns (Stahl and Artois, 1991; Hertwig *et al*. 2009) has played a major role in the species extirpation in many places (Langley and Yalden, 1977; Duarte and Vargas, 2001). At the same time, due to their large home ranges and high mobility, wildcats are among the wild carnivores which are most affected by indirect poisoning and road kills (McOrist and Kitchener, 1994; Nowell and Jackson 1996; Lüps *et al.* 2002; Schulenberg 2005). One further threat to the wildcat, in particular in Portugal and Spain, is the drastic demographic decline of its main prey, the wild rabbits, caused mainly by fatal diseases as myxomatosis and viral hemorrhagic disease (e.g. Lozano *et al.* 2007; Monterroso *et al.* 2009). Lower densities of rabbits have been also pointed has consequence of the increment of large game herbivores in Spanish hunting estates (such as red deer and wild boar), areas where wildcats' density has been decreasing as much as 6-fold (Lozano *et al.* 2007).

But the European Council (Stahl and Artois 1991) underlines two other major extinction risks for the species today, which are considered the pillars for the future designing of management programs and development of wildcat's conservation strategies. First, in common with many other carnivores, wildcat populations severely decreased and may still be in decline due to loss and fragmentation of suitable habitat. Drastic changes of geographical ranges may result in high levels of genetic fragmentation, increasing the effects of genetic drift that may lead to the decline of effective population size (*N*e) and loss of genetic variability (Spielman *et al.* 2004a, b). Population size reduction and fragmentation are predicted to result in the concomitant decrease of adaptive genetic variation because: i) less established genetic variation is sustained in small populations; ii) fewer new mutations per unit time appear and the opportunity for recombination is reduced (even in large populations, beneficial new mutations rarely occur); iii) for the same reasons that selection fails to remove weakly detrimental mutations, the probability that a new adaptive mutation of small effect can be maintained by selection is reduced (Hedrick 2004; Kohn *et al.* 2006). Consequently, the potential of small populations to long-term adaptation is restricted and the probably of local extinction is absolute. Closely linked to this is the fact that habitat fragmentation usually reduces gene flow among populations and exposes them to the risk of reduced fitness and inbreeding depression. Therefore, it is not surprising that one of the most common rescue-strategies proposed by conservation geneticists includes the increase of gene flow among populations (Pertoldi *et al.* 2007). Until very recently, deforestation (Nowell and Jackson, 1996; Krüger *et al.* 2009), the reduction of Mediterranean scrubland for fire control (Lozano *et al.* 2003), and the loss of natural areas due to urbanization, extensive road networks and intensive agriculture (Easterbee *et al.* 1991; Stahl and Artois, 1991; McOrist and Kitchener, 1994; Nowell and Jackson, 1996; Klar *et al.* 2008, 2009) have simultaneously depleted and isolated natural European wildcat populations, and some of these threats persist in our days.

Second, wildcats evolutionary persistence is being compromised by the long-lasting and ever-increasing contact with free-roaming domestic cats, a threat that manifests itself in different aspects of the species survival. In one hand, even considering the actual level and potential impact of contagious diseases

from domestic cats has not yet been well-established in natural populations, there is high potential for disease transmission between domestic and wildcats (Daniels *et al*. 1999; Leutenegger *et al.* 1999), and some domestic infections proved to remain permanently in wild populations while otherwise they would die out (e.g. the feline immunodeficiency virus FIV found in French individuals; Fromont *et al*. 2000). At the same time, feral domestic cats compete with wild individuals for space and resources, occupying their home ranges and interfering with the demographic and reproductive dynamics of natural populations (Biró *et al.* 2004, 2005; Germain *et al.* 2008; Sarmento *et al.* 2009). Lastly, but possibly the most important, hybridization and introgression of domestic alleles into wildcats' genomes is blurring population boundaries and is potentially compromising the species' genetic purity, being thus pointed by most researchers as the foremost threat for wildcat's conservation (e.g. Stahl and Artois, 1991; Nowell and Jackson, 1996; Randi *et al.* 2001; Pierpaoli *et al.* 2003; Hertwig *et al.* 2009). In this context, further risks have been pointed out, namely the selective resistance of *silvestris* x *catus* hybrids to viral disease infection (Ragni and Possenti 1996, Ragni 1993), and the extra reproduction possibilities and territory competitions posed by admixed free-ranging cats (Biró *et al.* 2004, 2005; Germain *et al.* 2008).

1.3.1. Introgressive hybridization: the dodgy mating with domestic relatives

Hybridization involves the successful mating between individuals from two populations, or group of populations, which are distinguishable on the basis of one or more heritable characters, regardless of their taxonomic status (Arnold 1992). Biologists have long held contrasting views on the role of this phenomenon in evolution. While botanists soon recognized its importance in plant's diversification, zoologists where historically more sceptical and cautious about its outcomes in the evolutionary process (Barton 2001; Arnold 1997). However, the increasing number of hybrid species in several animal groups, by the combined use of modern molecular tools, imposed a re-evaluation of the weight of natural hybridization in animals' evolution and its importance in evolutionary novelties, speciation and adaptation is now well recognized for numerous species (Arnold 1997, 2004, 2006; Seehausen, 2004; Mallet 2005, 2007). But the most important source of genetic variation and major driver of speciation in sympatric *taxa* is not hybridization alone, but introgression. Introgressive hybridization implies the transfer of genes between the hybridizing groups, through interbreeding and repeated backcrossing, and originates a complex mixture of parental and admixed variants belonging to different generations of hybridization. Approximately 10% of species in major faunal groups hybridize, and much higher rates have been detected within some of the most rapidly diversifying subgroups (especially in birds and insects, Mallet 2005).

The revived interest in animal introgressive hybridization results also from the realization that anthropogenically driven changes on the spatial distribution of species are increasing the incidence of hybridization events (Reusch and Wood 2007). Even though natural hybridization and introgression in animals are today recognized as evolutionary movers, artificially events typically represent critical threats to native fauna (Wayne and Brown 2001; Randi 2008). This conservation-oriented view became stronger when hybridization (and introgression) started to be detected among endangered species, as a result of animals'

translocations and/or the free-ranging behaviour of domesticated forms (Rhymer and Simberloff, 1996; Allendorf *et al.* 2001). Especially in endangered *taxa*, interbreeding (with or without significant introgression) is generally considered disadvantageous as it promotes loss of genetic diversity and outbreeding depression, possibly leading to the disintegration of important traits that have arisen as local adaptations during the evolutionary process (Allendorf *et al.* 2001; Randi 2008). Rates of these types of hybridization and introgression are increasing dramatically worldwide also due to anthropogenic habitat modifications (fragmentation and loss); and the ever increasing pace of human activities suggests that this problem will aggravate everyday (Allendorf and Luikart 2007). Allendorf et al. (2001) identified three general outcomes of human-induced hybridization, with the proviso that situations are not exclusive in real biological but rather exhibit a continuum: i) hybridization without introgression; ii) widespread introgression; and iii) complete admixture. In the most extreme case, high frequency of hybridization events followed by backcrossing may lead to the formation of a hybrid swarm, and result in species replacement. Because the ecological and evolutionary implications of these categories can differ dramatically (Allendorf *et al.* 2001), correctly identifying the appropriate category for each scenario represents a critical first step towards developing effective conservation management strategies.

In vertebrates, hybridization often occurs in situations of local scarcity of conspecific mating partners and in cases where two sympatric *taxa*/populations/ESU's show evident asymmetry on the effective number of individuals. The strongest factor leading to this numerical imbalance is the introduction, voluntary or not, of domestic or exotic non-native populations (Mooney and Cleland, 2001). This introduction is even more problematic when the non-native *taxa* is familiar with human presence and activities, being that new colonizing feral groups are somehow best fitted to live in today's strongly changed environments. Among the endangered European carnivores, a number of species belonging to different families are affected by artificially mediated hybridization where domesticated forms are leading protagonists (see Gittleman *et al*. 2001). In Mustelidae, populations' decline of the European mink *Mustela lutreola* in Western Europe are thought to be related with hybridization with the American mink *Neovision vison* or the European polecat *M. putorius* (Lodé *et al.* 2005). Moreover, the first is being jeopardized by the overwhelming presence of domestic minks in the wild (Kidd *et al.* 2009), and the latter is currently threatened by sympatric domestic ferrets *M. p. furo*, with which it hybridizes since it was first introduced in nature for controling rabbit populations in Britain (Davison *et al.* 1999). Among canids, interbreeding between grey wolves (*Canis lupus*) and domestic dogs (*Canis familiaris*) has been reported worldwide, albeit studies have shown that these events are rare and there is no evidence for significant introgression of domestic genes into wild wolf populations (Vilà and Wayne, 1999). Even so, admixed individuals were identified in Spain (Blanco *et al.* 1992; Godinho *et al.* 2011; Italy (Randi and Lucchini 2002), and a number of eastern countries (e.g. Randi *et al.* 2000; Andersone *et al.* 2002; Vilà *et al.* 2003).

Already twenty years ago, Stahl and Artois reported that artificial interbreeding was probably a regular phenomenon in several wildcat populations and was a major conservation concern in 11 out of 17 European countries (Stahl and Artois, 1991). Today, historical, environmental and/or behavioral features that may influence processes of hybridization and introgression between wild and domestic cats remain still

to precisely define, and the actual extent of interbreeding is permanently under discussion (Beaumont *et al.* 2001; Pierpaoli *et al.* 2003; Kitchener *et al.* 2005).

1.3.2. Hybridization in European wildcat populations

Different ecological factors that influence wildcat populations may have been encouraging alone or concomitantly the probability of crossbreeding In Europe. Most likely, the fragmentation and loss of suitable habitat and the demographic decline of wildcat populations have been together increasing the risk of hybridization. Isolation and lower densities may encourage wildcats to make major shifts in their search for a reproductive partner. These movements increase the risk of meeting with domestic and hybrid individuals, which in turn become the closest and most accessible partners (Suminski 1962; Hubbard *et al.* 1992; Jaeger *et al.* 2005). Observations of wildcats' spatial activities in Portugal suggested, in fact, that most of admixture events are probably occurring between male wildcats and domestic females (Monterroso *et al*. 2009). Contrastingly, observations in Ardennes Mountains in France indicate that hybridization might not involve domestic females, but probably the low densities of wild males promote the crossbreeding between wild females and domestic males (Germain *et al.* 2008). Either way, these dynamics seem to most commonly happen on the edge of the species range, or in places where wildcats became rare and domestic feral cats are abundant. In Portugal, free-roaming domestic cats might have totally replaced wildcats in remote natural areas where wild individuals were formerly known to exist (Sarmento *et al.* 2009). Depending on their location and the rural/urban environment they live in, domestic cats form groups of different size characterised by dissimilar spatial and social structures, and mating systems, and different genetic structures (Pontier *et al.* 2009). This versatility might pose additional threats to native wildcat populations, which are probably much more dependent on prey and habitat availability, and susceptible to environmental changes.

The more similar the ecological necessities of two sympatric populations are, the more individuals from these populations have the chance to interact. Ecological requirements may, thus, also play a key role in the hybridization process. Recent studies conducted in Hungary (Birò *et al.* 2005) and in France (Germain *et al.* 2008), showed that feeding habits of hybrids, while intermediate between wild and domestic cats, overlap those of wildcats, suggesting that competition for food resources might exist and that wildcat populations may be negatively affected. Hybrids have also the broadest trophic niche and they search for food closer to human settlements than wildcats (Birò *et al.* 2005; Germain *et al.* 2008). Admixed animals may also be less susceptible to habitats changes and more proficient in colonizing new habitats (Germain *et al.* 2009). Finally, wild, domestic and hybrid cats do not reflect evident behavioral barriers in their space use and daily activity (Birò *et al.* 2004; Germain *et al.* 2009). Interestingly, Klar *et al.* (2008) report a fairly strict separation of ranges of wild and feral domestic cats in Germany, which they hypothesize to be one of the reasons behind the low frequency of hybridization detected in most wildcat populations in central Europe.

In cases of long-term sympatry, it has been hypothesized that, wildcats may exhibit an antagonistic behaviour towards domestic cats (Hubbard *et al.* 1992), and thus averting hybridization to a certain extend (Easterbee *et al.* 1991). Under the scenario of long-lasting simpatry in Europe, mating between wild and

domestic cats should occur only occasionally. If this is the case, after crossbreeding takes place, hybrids – rather than domestic forms – may be protagonists in maintaining and expanding hybridization (Germain *et al.* 2009). This reality poses the question if in places where putative wildcat populations have been apparently stable or even expanding that is a sign of populations "health" or, conversely, the result of frequent incidence of feral domestic and hybrid cats. Considering this, the growth and expansion of some wild populations, especially after World War II, could be partially attributed to hybridization (Stahl and Leger, 1992). At least in France, evidences show that wildcat's expansion should not be an artefact of hybridization, since crossbreeding was detected over the entire range of the species and does not congregate on newly occupied areas (Germain *et al.* 2009).

It is known that, once hybridization takes place, it is arduous to disrupt and block it, especially if admixed individuals are fertile and mate both among themselves and with the parental groups (Allendorf *et al.* 2001). In extreme situations, hybrids repeated backcrossing to one or both parental populations might lead to the disappearance of their "pure" parental genomes (Mallet 2005), and to the formation of a hybrid swarm (Allendorf *et al*. 2001). Being able to discriminate these scenarios strongly depends on the capable detection of wild, domestic and admixed individuals.

First studies on wildcats' populations based their ability to discriminate wild, domestic and hybrid cats on morphological characters. A combination of phenotypic features (Figure 4) and distinct anatomical variables (including pelage, limb bone measures, cranial variables, and indexes of intestinal length) were determined to distinguish both domestic and hybrid individuals from wildcats in Italy ((Ragni and Possenti, 1996), Scotland (Daniels *et al.* 1998; Kitchener *et al.* 2005), Germany (Krüger *et al.* 2009), Slovakia (Platz *et al.* 2011), as well as between domestic, European and African wild cats (Puzachenko 2002 and between European, African, and Asian wildcats (Yamaguchi *et al.* 2004a,b).

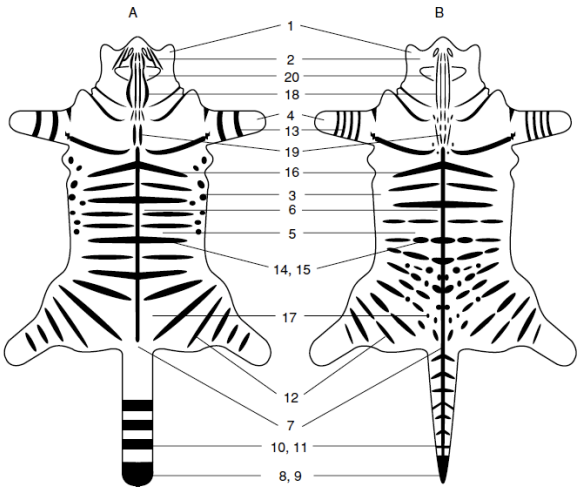


Figure 4. Diagram highlighting 20 diagnostic morphologic characters used to score cats as wild, domestic or hybrid (adapted from Kitchener *et al.* 2005). A=European wildcat; B=domestic cat

But visual distinction of free-ranging tabby domestic cats is frequently dubious because the two groups exhibit broadly overlapping characteristics for many morphological traits, and the diagnostic value of most characters seem to depend on geographic adaptations and hybridization levels (Platz *et al.* 2011). Moreover, morphological identifications are particularly erroneous for admixed cats, especially beyond the first generation of hybridization (Krüger *et al.* 2009). Being so, using morphological traits for distinguishing wild or domesticated states is most likely speculative and hardly universal.

On the light of these challenges, molecular biology soon became the discipline of choice for distinguishing wild and domestic cats and for detecting hybridization within wildcat populations.

1.3.3. Molecular advances in cat's admixture analysis and population structure

Recent advances over the past two decades in the development of molecular markers and bio-statistical techniques have led to a much higher capacity to study hybridization and, consequently, to a new understanding of the biogeographic patterns of admixture in wildcat populations. The first molecular studies, based on allozymes variability, showed that these protein loci were not able to unequivocally differentiate wildcats and their domestic counterparts (Randi and Ragni, 1991), probably due to mechanisms of natural selection that usually retain allele frequencies between closer *taxa/*populations in important functional loci, and to the general low mutation rate of these portions of the genome (Li, 1997). Thus, subsequent studies focused on both mtDNA sequences and microsatellites variation (e.g. Eckert *et al.* 2009; O'Brien *et al.* 2009; Hertwig *et al.* 2009), or, more frequently, solely on the analyses of microsatellite diversity.

First mitochondrial analyses showed that this genomic compartment might also be inadequate to sharply distinguish wild and domestic cat groups, as well as for identifying subspecies and geographic clusters, and to perform the correct allocation of individuals and detect hybridization events. The low capacity of mtDNA to clearly identify parental groups of these closely related *taxa* might be related to the early and long lasting gene flow between various wild and domestic cats lineages, which can endorse the introgression of mitochondrial haplotypes and dilution of signals of genetic structure (Driscoll *et al.* 2007; Hertwig *et al.* 2009). Alternatively, low cyto-nuclear resolution might result from incomplete lineage sorting between wild and domestic cats and not from interbreeding, since only 22% of discordances between morphological and mtDNA identifications showed evidence of recent hybridization in microsatellites while the others 78% might represent shared ancestral polymorphism (Driscoll *et al.* 2007). In summary, the weak and poorly resolved phylogenetic signals, the poor differentiation among lineages and the low-resolution network analyses suggest a limited utility of mtDNA to accurately identify hybrids and assign individuals of unknown origin to wild or domestic populations (Randi *et al.* 2001; Eckert *et al.* 2009; Hertwig *et al.* 2009). Furthermore, the use of mtDNA, being maternally inherited, does not provide any information about crosses between a female wildcat and a male domestic cat, which might be the main interaction between the two forms in some European wildcat populations (Germain *et al.* 2008). Finally, mtDNA is also of limited utility when hybrids backcross with individuals from the parental populations might be frequent.

The simultaneous score of several highly polymorphic microsatellite markers combined with new statistical methods has radically improved the assessment of populations' structure, individual assignments and admixture analyses. Although diagnostic loci have not been found for distinguishing wild, domestic cats and their hybrids, genetic analyses have been based upon loci at which the parental *taxa* have distinct allele frequencies. In the last decade, several admixture studies have been performed to determine if local wildcat populations are genetically distinct from sympatric free-ranging domestic cats, and thus assist defining in which way they are worth of legal protection. These studies have indicated diverse degrees of hybridization in Europe, suggesting that hybridization does not constitute a uniform threat throughout the entire range of the species (e.g. Beaumont *et al.* 2001; Randi *et al.* 2001; Pierpaoli *et al.* 2003; Lecis *et al.* 2006; O'Brien *et al.* 2009). Sharp differentiation and high genetic structure between phenotypically identified European wildcats and sympatric domestic cats were found in Italy, north-eastern France and Germany, suggesting either that hybridization levels are low (between 2-8% in Italy; Randi *et al.* 2001, Lecis *et al.* 2006) or that recent crossbreeding might occur but is somehow restrained (23,8% in France, O'Brien *et al.* 2009; maximum of 18,4% in Germany, Hertwig *et al.* 2009). On the other hand, other European populations seem to be severely affected by interbreeding and might represent hybrid swarms, namely in Scotland and Hungary. In Scotland it is even possible that very few genetically distinct wildcats remain, since populations show signs of significant amalgamation with domestic cats (Hubbard *et al.* 1992; Beaumont *et al.* 2001; Pierpaoli *et al.* 2003; Macdonald *et al.* 2004; Kitchner *et al.* 2005). Hungarian populations embrace up to 31 % of admixed cats (Lecis *et al.* 2006). In Germany, contrastingly results were reported on admixture levels: while 18,4% of admixture was recently detected across the whole country (with 43% of hybrids concentrated in the western population) by Hertwig *et al.* 2009, only 3% was detected by Eckert and colleagues (Eckert *et al.* 2009). These later findings had also been previously reported by Pierpaoli *et al.* (2003). Opposing results have been moreover reported for German wildcat population structure. Studies by Pierpaoli *et al.* 2003 and Eckert *et al.* 2009 suggest bottleneck events or reduced genetic diversity caused by genetic drift or inbreeding, while no evidences of such events were found by Hertwig *et al.* (2009). Still, the first two studies did not agree in what concerns German populations' genetic uniqueness and hybridization levels (see Eckert *et al.* 2009 for details). Overall European wildcats were assigned to distinct genetic clusters using a small battery of 12 unlinked microsatellites (southern versus central Europe clusters), suggesting that European wildcats include genetically differentiated subpopulations. Furthermore, analyses suggested that there has been an extended gene flow among populations in central Europe, with the notable exception of the Solling region in northern Germany (Pierpaoli *et al.* 2003).

In summary, recent molecular approaches have somehow unravelled the evolutionary history of *F. silvestris* (see Driscoll *et al.* 2007) and have given first insights on European wildcat's genetics (Pierpaoli *et al.* 2003). Nevertheless, a number of limitations still hinder the thoughtful knowledge of these populations. In the light of wildcats' conservation, hybridization is one of the most complex and controversial issues, and many of its pathways and outcomes are still to define. The great challenge begins with the apparently simple task of defining hybrid individuals.

1.4. Major drawbacks in contemporary wildcat conservation genetics

One of the most far-reaching conclusions that can be drawn from the above-mentioned genetic studies is that both number and type of analysed loci seem to be crucial for the resolution and robustness of admixture inferences and to the reasonable comparison between results, but cat's hybridization studies are still a lot below their optimum to provide such accurate output. Most recent wildcat's hybridization studies have been based mainly in the genotyping of less than 15 unlinked microsatellites (with the exception of Lecis *et al.* 2006 that analysed 27 linked and unlinked loci), and these markers seldom overlap among studies. At the same time, the analyses reported in the past decade cannot assert that "pure" wildcats persist, neither establish the effective number of specimens with hybrid ancestry, because number and type of genetic markers proved not to unequivocally discriminate between wild, domestic and backcrossed cats (Vähä and Primmer 2006; O'Brien *et al.* 2009; Hertwig *et al.* 2009). In fact, recent publications recommend the use of significantly larger sets of loci (Koskinen *et al.* 2004), suggesting at least 12 to 24 microsatellites for the correct identification of F1 hybrids with levels of genetic differentiation (F_{ST}) between hybridizing parental populations of 0,21 to 0,12, respectively (Vähä and Primmer 2006). Much higher number of molecular markers is, thus, needed to detect further generations of backcrossing. Considering the above assumptions, and since F_{ST} values estimated among European populations range between 0,10 to 0,16 (Randi *et al.* 2001; Pierpaoli *et al.* 2003; Lecis *et al.* 2006; Germain *et al.* 2008; Hertwig *et al.* 2009; O'Brien *et al.* 2009), the studies performed so far lack of statistical confidence and could only represent a very conservative estimation of admixture events. The advantages of using physically-linked loci to overcome this uncertainty is a matter of some controversy (Falush *et al.* 2003; Lecis *et al.* 2006; Vähä and Primmer 2006), and the only study applying linked microsatellites showed no significant improvement in the power of hybrids detection (Lecis *et al.* 2006).

One of the first works thoroughly exploring the important impact that the number of loci and the values of F_{ST} and H_E may have in individuals' assignment tests based on frequentist and Bayesian inferences was published by Manel *et al.* (2002). New genetic tools based in Bayesian analyses of population diversity/variability (STRUCTURE, Pritchard *et al.* 2000), proved to have more precision in identifying the origin of individuals based on both real and simulated multilocus genotypes (Manel *et al.* 2002; Figure 5). This work suggests that high number of loci together with high levels of polymorphism, is crucial for accurate inferences especially with decreasing levels of populations' differentiation. Nevertheless, this is expectably harder within hybridizing *taxa* and, moreover, in populations where several generation of hybrids occur (hybrid swarms).

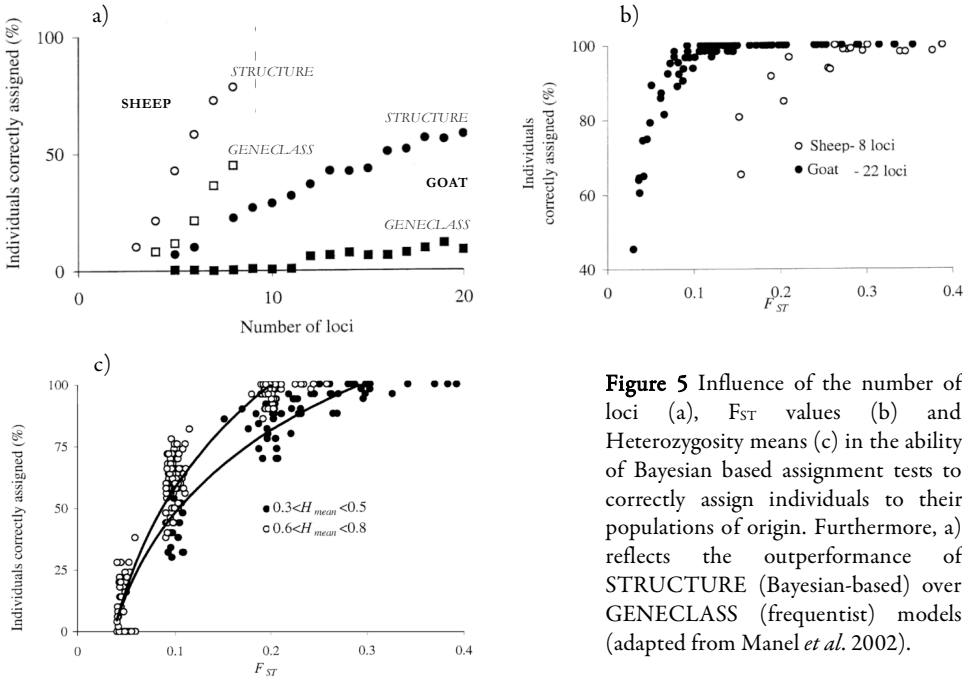


Figure 5 Influence of the number of loci (a), F_{ST} values (b) and Heterozygosity means (c) in the ability of Bayesian based assignment tests to correctly assign individuals to their populations of origin. Furthermore, a) reflects the outperformance of STRUCTURE (Bayesian-based) over GENECLASS (frequentist) models (adapted from Manel *et al.* 2002).

Not only quantity and quality of molecular loci but also sampling strategies vary greatly among studies and may explain, in part, differences in the reported genetic diversities and structures of wildcat populations (O'Brien *et al.* 2009; Hertwig *et al.* 2009). In addition, even considering that Bayesian models have been the analytical methods of choice among all recent publications, it is important to bear in mind that such modelling approaches are sensitive to a number of variables that varied among studies. In one hand, the estimation of distinct genetic clusters, the assignment of individuals to those clusters and the proportion of cats assigned as admixed or parental are sensitive to the genetic diversity and structure found in the studied populations, which in turn depend on markers number and choice, and on the number and proportion of analyzed wild, domestic and hybrid cats (Vähä and Primmer 2006; Randi 2008). On the other hand, the number of genetic clusters has been defined in some studies without using a priori population information (e.g. Lecis *et al.* 2006; O'Brien *et al.* 2009), whereas it was constrained in others (e.g. Randi *et al.* 2001; Pierpaoli *et al.* 2003).

High introgression levels reduce our ability to use, not only pelage and skeletal morphological traits, but also genetic markers to discriminate between wild, domestic and hybrids cats. The development of novel molecular markers and the use of a more comprehensive battery of loci are, thus, essential to improve our current knowledge on this endangered felid.

1.5. New genetic tools for cat studies: finding the way to distinctiveness

In general, domestication can be seen as a process by which humans control nearly all aspects of a domesticated species life, being reproduction the most important of all. Therefore, domestication is based in the process of selection. In its course, major impacts are perpetrated on the pattern and amount of animal's genetic diversity. Probably as a result of the variable domestication purposes and circumstances, the extent of those impacts is unpredictable, and the genetic consequences of domestication are extremely variable among species and genomic portions. For example, molecular analyses of mitochondrial and Y-chromosome DNA in domestic dogs documented extensive mtDNA diversity against much lower Y linked variability, which might be a result of a strong female-biased domestication process (Sundqvist *et al.* 2006). The same pattern was found in horses, with additional information on the X-chromosome (Lau *et al.* 2009). At the same time, although it is likely that wild animals were selected for traits like tameness and docility, selection for these traits have also been shown to confer additional morphological changes. For example, selective breeding experiences with silver foxes resulted in side-effects of delayed fear responses, changes in stress hormone levels, modified gene expression in the brain and coat colour variation (Belyaev, 1979; Trut 1999). Moreover, selection for hairlessness in dogs is associated to no development of teeth (Drogemuller *et al.* 2008), and in Ridgeback dogs the selection for hair ridge also leads to predisposition to dermoid sinus (Hillbertz *et al.* 2007). The decreased fertility in dairy cattle associated with high production is also another explicit example (Boichard *et al.* 2003). One way of looking to signatures of artificial selection is also comparing the genome of a domestic animal to its wild relative. For example, dN/dS ratio analyses in whole-genome SNPs suggest that dogs have accumulated deleterious mutations since the time of domestication (ratio 50% higher than in wolves), probably a result of relaxation of selective constraint (Cruz *et al.* 2008). A similar pattern has been observed between pigs and wild boar when investigated for the MC1R coat colour gene, showing that in one hand purifying selection against non-synonymous changes maintains conspicuous coloration in the wild, while on the other variable coat colour phenotypes in pigs result from direct human positive selection (Fang *et al.* 2009). The recent re-sequencing of the complete genomes from eight different populations of domestic chicken along with the genome of the wild ancestor Red Jungle fowl (*Gallus gallus*) revealed, through a selective sweep analysis, a candidate for a domestication locus in chicken, the thyroid stimulating hormone receptor *TSHR* (Rubin *et al.* 2010). All together, these studies show how powerful the joint analysis of the genome from a domestic species and its wild ancestor can be.

Past episodes of positive, balancing and/or purifying selection during cat's domestication are expected to have created distinct signatures in the genome of the domesticated versus wild populations, and to genetically tear domesticates apart from wild forms of the *F. silvestris* complex. Usually, genome-wide bottlenecks occur immediately in the beginning of the domestication process and also later at specific loci through breeds' development (Figure 7), respectively due to the small demographic size of the initiating domestic population and to the constant improvement (through strong selection) of purebred individuals with specific desired phenotypic traits (Bruford *et al.* 2003; Zeder 2006). Nevertheless the genetic diversity can vary between these two steps. For example in dogs, only a 5% reduction in nucleotide diversity was

observed as a result of domestication, whereas the loss of nucleotide diversity with breed formation averaged 35% (Gray *et al.* 2009). These scenarios of decreased effective population size (Ne) typically result in inbreeding depression and/or genetic drift, which produce evident reductions in neutral loci genetic diversity (Figure 6a). Moreover, positive selection in specific genes leads to genomic regions harbouring reduced levels of diversity, altered allele frequency spectrum and locally increased extent of linkage disequilibrium (Przeworski, 2002; McVean, 2007; Figure 6b). On the other hand, the relaxation of selective pressures at characteristics that are vital for wild living forms (and that are naturally maintained by purifying selection) but not for domesticated animals, along with the elevated variability at behavioural, physiological and phenotypic traits that may arise for better living under a human-controlled environment, might result in a significant increase in domesticates' genetic diversity when compared to their wild counterparts (Dobney and Larson 2006; Figure 6c).

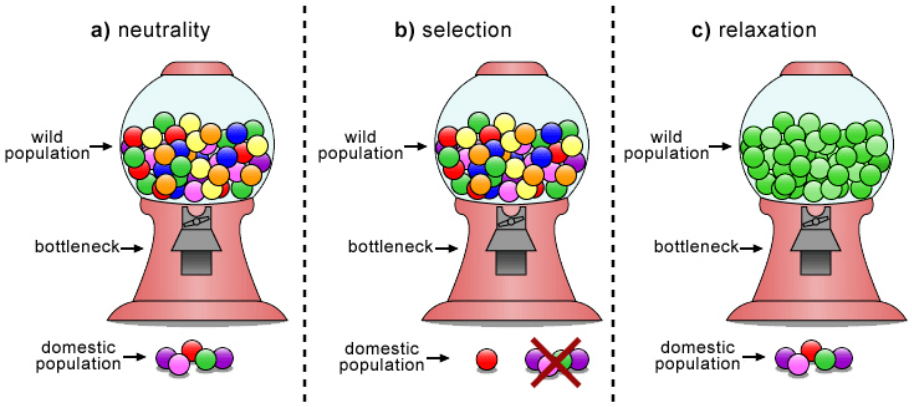


Figure 6. Diagram illustrating the major genetic outputs of domestication. Domestication causes, in one hand, genome-wide reductions of genetic diversity (resulting both from the early stages of domestication and from breeds improvement) at: (a) neutral loci due to demographic events and at (b) selected specific loci via positive selection of advantageous alleles. On the other hand, following the strong bottleneck effects, relaxation of selective pressures may increase domesticates' genetic diversity through the increment of post-domestication (*de novo*) mutational events (c). Gumballs of different colour represent different alleles. Adapted from Doebley *et al.* (2006).

Considering the relaxation scenario, multiple alleles in domesticates might be maintained overtime through balancing selection. However, it is important to recognize that this variability is not necessarily beneficial. For instance in dogs, the "domestication bottleneck" leading to relaxation of selective constrains (and the reduction in Ne at loci linked to those under positive breed selection) may have contributed to a greater accumulation of the above-mentioned nonsynonymous deleterious mutations (Björnerfeldt *et al.* 2006; Cruz *et al.* 2008). Food supplementation, lack of competitive mate selection, and the reduction or absence of predation risks in domesticates all result in a relaxation of selection that may cause the increase in frequency of deleterious genes and phenotypes (Zeder 2008). Furthermore, linkage between loci under strong positive selection and their flanking regions may diminish the efficiency of purifying selection, since there might be an increase of such mutations in these regions due to hitchhiking effects. The high incidence of genetic diseases in specific domestic breeds is most illustrative of this phenomenon. In the light of this

complexity in the genomic outcomes of domestication, fundamental questions regarding the basic definition and the processes underlying this process remain, today, still largely unanswered (Dobney and Larson 2006). Anyhow, it is clear that the wide genetic repercussions of domestication might be simultaneously detected over neutral loci and specifically selected genomic portions (Yamasaki *et al.* 2005).

In the course of cat's domestication, it is likely that both natural and artificial selections played crucial roles. At one hand, mutations outside human control may have created natural diversity that made some wildcats to be more compatible with human control. On the other hand, through artificial selection, humans significantly changed morphology, physiology and behaviour of domesticated cats (Diamond, 2002). More recently, the manipulation of domesticates gene pools caused tremendous phenotypic changes and created a variety of cat breeds for which diversifying selection is mostly related with morphologic aesthetic traits related with pelage types appreciated by humans. The immense phenotypic diversity that commonly segregates in domestic species provides exceptional opportunities to establish specific genotype/phenotype associations and to study the general mechanisms by which genetic variation governs biological functions. The recognition that a significant number of domestication traits are common to most domesticated animal forms (e.g. appearance of dwarf/giant varieties; piebald coat colour; changes in reproductive cycle; Trut 1999) prompted the breeding association studies to investigate the genetic basis of those characters. But although a number of genotype/phenotype correlations have been deciphered in cat's using candidate genes approaches (Lyons 2010, 2012), the increasing information on cat's genome demonstrated that some previously assumed associations are not straightforward (e.g. see Eizirik *et al.* 2010). Most of domestic cat studies have targeted genes based on well-studied biochemical and molecular pathways, and on their proven relation to a given trait in other animal models, mainly human, mouse and dog. The majority of identified genetic variants are related with coat patterns on the basis of both hair length (e.g. four different mutations at FGF5, Drogemuller *et al.* 2007, Kehler *et al.* 2007, and SNPs at KRT71 Gandolfi *et al.* 2010), and colour, in which genes like TYR (Lyons *et al.* 2005a; Schmidt-Küntzel *et al.* 2005), TYRP1 (Lyons *et al.* 2005b; Schmidt-Küntzel *et al.* 2005), MLPH (Ishida *et al.* 2006) or ASIP (Eizirik *et al.* 2003) are determinants. The same candidate gene strategy proved also extremely useful for identifying mutations related with inherited diseases (see Lyons 2012 for details). Concomitantly, classic linkage-mapping analyses suggested the association of KIT and, possibly, PDGFRA with white spotting in cats' colour patterns (Cooper *et al.* 2006), while allowed a detailed mapping of the X-linked orange locus (Schmidt-Küntzel *et al.* 2009).

The recent and rapid genesis of cat breeds from a limited number of individuals suggests that, in many cases, a small number of genes of large effect are responsible for breed characteristics (Pollinger *et al.* 2005). Today, approximately 33 genes containing 50 mutations are known to cause feline health problems or alterations in the cat's appearance (Lyons 2010). While most identified disease determinants are very specific to cat breeds and populations, coat mutations are common to the majority of cats and are, therefore, effective for global genetic typing (Table 1). Moreover, coat colour genes are suspected to have important pleiotropic effects upon morphology and behaviour (Pontier *et al.* 2009).

Table 1. Summary of the most relevant genetically determined coat traits in random-bred and pedigree cat breeds. All mutations are autosomal recessive.

GENE		FUNCTION	PHENOTYPE	BREED	MUTATION	REFERENCE
ASIP	Agouti signaling protein	Signaling protein in the melanin synthesis chemical pathway	Banded fur to solid	All	Del122to123	Eizirik *et al.* 2003
MC1R	Melanocortin 1 receptor	Signaling protein in the melanin synthesis chemical pathway	Brown colour variant (Amber)	Norwegian forest	G250A	Peterschmitt *et al.* 2009
TYRP1	Tyrosinase-related protein 1	Enzyme in the melanin synthesis chemical pathway	Brown, light brown colour variants	All	b=C8G; b'=C298T	Lyons *et al.* 2005a; Schmidt-Kuntzel *et al.* 2005
TYR	Tyrosinase	Enzyme in the melanin synthesis chemical pathway	Burmese, Siamese colour pattern; full albino	All	c^b=G715T; c^s=G940A; c=C975del	Lyons *et al.* 2005b; Schmidt-Kuntzel *et al.* 2005; Imes *et al.* 2006
MLPH	melanophilin	Carrier protein; tether melanosomes to the melanocytes' actin cytoskeleton.	Dilution: black to grey/blue, orange to cream	All	T83del	Ishida *et al.* 2006
KIT	tyrosine kinase c-Kit	Pigmentation of hair and skin; development of gastrointestinal tract, mast cells and sperm cells	Gloves: white feet	Birman	c.1035_1036delinsCA	Gandolfi *et al.* in press
KRT71	Keratin 71	Central role in hair formation: expressed in the inner root sheath of hair follicles	Hairless: Atrichia	Sphynx	c.816+1G>A c.1108-4_1184del;	Gandolfi *et al.* 2010
			Rexing: curly hair coat	Devon Rex	c.1184_1185insAGT TGGAG; c.1196insT	Gandolfi *et al.* 2010
FGF5	Long fur	Signaling factor controlling the hair follicle growth cycle	Long fur	All	c.356insT, C406T, c.474delT, A475C	Drogemuller *et al.* 2007; Kehler *et al.* 2007

The vast morphological, physiological and behavioural diversity occurring among domesticates is not usually observed in their wild counterparts, where phenotypic uniformity prevails (e.g. dogs and rabbits). Since basal morphology in European wildcats is unchanging, specific mutations determining the above-mentioned and other variable patterns in domestic cats are very rare or absent in natural populations, and alternative variants or random polymorphism are expected. At the same time, explicit genetic variants might benefit the way of living in nature and may have remained fixed in the wild populations (e.g. camouflage patterns crucial for hiding and hunting behaviour), while variable genetic expressions might be maintained in domestic individuals due to relaxation of selective pressures. Analysing levels of genetic diversity in genes known to have been under different types of selection (see Figure 6) during domestication and/or breeds' improvement may, thus, revolutionize wildcat hybridization studies. But unlike many other domesticated animals, which exhibit a huge range of sizes, shapes and temperaments, domestic cats are relatively homogeneous and less diverse. The reason for this relative lack of variability seems straightforward: cats experienced no such long-lasting selective breeding pressures as dogs or horses, because only around 60 breeds were selected in the last 150 years. Furthermore, random-bred feral cats are the ones representing true hybridization threats to the wild native populations, and among domestic cats they typify the individuals less targeted for artificial selection. Therefore, finding the ways of distinctiveness and detecting real levels of genetic admixture between wildcats and their domestic relatives might prevail as an exceptional defy for conservation geneticists. Conversely, a rural-urban effect at colour genes' variability has been already

suggested to occur even within random-bred cats (Pontier *et al.* 1995, 2009), which predicts the possibility of finding significant patterns of genetic partition among and between wildcats and free-ranging domestic cats at different genes. Anyhow, time is running against European wildcats survival as a naturally evolving species. Random-bred and cats of several breeds have conquered Europe and the world, human's respect and empathy, and they live today in most habitable places even in very remote areas (Figure 7).

Figure 7. Europe-wide stamps showing the broad distribution of domestic cats in this continent and explicitly demarking the human acceptance towards this felid.

1.5.1. The potential use of SNPs (Single Nucleotide Polymorphism)

Over the last decades, advances in molecular biology and the rapid development of DNA analysis technology have greatly increased our capacity to study the entire genome of a species, and we will soon be able to "genome-type" many individuals within those species (Allendorf *et al.* 2010; Ouborg *et al.* 2010). One of the immediate benefits of genomics was the use of such vast data in a few species for which genomic information and tools were available in related domestic species (e.g. red fox, Sacks and Louie 2008; wolf, vonHoldt *et al.* 2010; bison, Pertoldi *et al.* 2010; bighorn sheep, Poissant *et al.* 2010). Wildcats are among these "genome-enable" *taxa* (*sensu* Kohn *et al.* 2006), and thus will unquestionably benefit from the newly available genomic and biostatistical tools from the domesticat cat. The great amount of data on the domestic cat genome and the cross-species applicability of such data predict that new insights for genetic studies of related rare and endangered felids are forthcoming. Particularly, the recent availability of a light coverage of the domestic cat

genome (Pointius *et al.* 2007; Mullikin *et al.* 2010), along with the latest autosomal (Menotti-Raymond *et al.* 2009) and X genetic linkage maps (Schmidt-Küntzel *et al.* 2009), provide useful reference data for single nucleotide polymorphisms (SNPs) discovery and allows performing comparative map-based approaches to identify candidate loci. Furthermore, Davis *et al.* (2009) recently reported a high-resolution cat radiation hybrid (RH) map that constitutes a comprehensive framework for identifying genes controlling feline phenotypes of interest, and to aid in assembly of a higher coverage feline genome sequence. Finally, the release in 2011 of a 63K feline SNP array, including mutations found in the wildcat, promise to largely increase the number of informative loci available for wildcat research (L. Lyons, person. comm.).

One of the most promising outcomes of applying genome scanning of many markers to conservation genetics is the simultaneous estimation of neutral processes along with the identification of specific genomic regions responding to selection (Luikart *et al.* 2003; Primmer 2009; Allendorf *et al.* 2010; Ouborg *et al.* 2010). Multiple integrative approaches combining neutral genetic variability and diversity in specifically selected traits will soon become common practice for conservation geneticists under a wide-genomic perspective, rather than the traditional small-scale surveys of neutral marker variation that have been characterizing most studies on endangered species (Allendorf *et al.* 2010; Ouborg *et al.* 2010). The attempt to correlate neutral and non-neutral variability can be made by using a recent and very promising molecular tool in wildlife research, the single nucleotide polymorphisms (SNPs). SNPs are appropriate markers for the study of both artificial selection and neutral diversity in genome-wide analyses primarily because: i) constitute the most abundant type of genetic polymorphism in most, if not all, genomes (they are spread throughout the entire genome at high density); ii) are dispersed both in coding and non-coding regions, and can be either synonymous or non-synonymous; iii) are simple to score and less prone to errors compared to microsatellites, iv) have simple mutation models and powerful analytical methods, v) may be cost-effectively genotyped using many high-throughput technologies, which make them suitable for large-scale population level analyses using very large number of markers (Wayne and Morin 2004; and see Brumfield *et al*. 2003; Morin *et al*. 2004; Garvin *et al*. 2010 for reviews). Likewise, SNPs provide an alternative to the maternally inherited and linked mitochondrial markers (whose application is moreover affected by the existence pseudogenes in the nuclear genome, e.g Lopez *et al.* 1996; Antunes *et al.* 2007) and to the high homoplasic microsatellite loci, whose analyses are fraught with difficulties. In addition, the mutational processes of SNPs are much better understood and easier to model than microsatellites, providing the basis for more robust inferences using powerful statistic tools (Ryman and Palm 2006; see Morin 2009 for details). Finally, SNPs are universally applicable between laboratories and extremely flexible in detection protocols, thus not requiring standardization across detection platforms and challenging optimizations as in microsatellites (Coates *et al.* 2009). However, SNPs have the drawback of encompassing much lower polymorphism when compared to microsatellite diversity due to their mainly bi-allelic transmission (Schlotterer 2004). The substantial increase in the number of loci available today for different species when compared to microsatellites (Brumfield *et al.* 2003), and the consequent increase in the number of genotyped markers, may successfully compensate, though, for their lower inherent variation (Morin *et al.* 2004). This has been proved in several human genetic population structure studies (e.g. Liu *et a.l* 2005; Lao *et al.* 2007; Paschou *et al.* 2007), and in several animals. About 22 SNPs in wolves *Canis lupus* (Seddon *et al.* 2005); 37 in Chinook salmon *Oncorhynchus tshawytscha*

(Narum *et al.* 2008); and 51 in chum salmon *Oncorhynchus keta* (Smith and Seeb 2008) proved to detect population structure and provided high probability of correct population assignment as well or even outperforming 12, 13 and 15 microsatellites, respectively (Morin 2009). The overall ratio of SNPs to microsatellites that is needed for equivalent results varies among applications (Hess *et al.* 2011), but most evaluations of this relationship have been reporting values from approximately three (above references and Schopen *et al.* 2008 in poultry and cattle; Santure *et al.* 2010 in zebra finch, *Taeniopygia guttata*, pedigrees; Glover *et al.* 2010 in Atlantic salmon, *Salmo salar*, individual assignment) to twelve times greater (Liu *et al.* 2005 in humans' population structure analysis). SNPs have also been suggested as superior markers for F_{ST} estimates (Sacks and Louie 2008). Andersen *et al.* (2006) reported that 18 unlinked SNPs provide P_{ID} values comparable to the ones obtained from genotyping Italian wolves at 10 microsatellite loci. Also, recent studies show that there are twice as many triallelic SNPs as theoretically expected, which might represent an important increase in their utility (Casci *et al.* 2011). Finally, the combination of both SNPs and microsatellites proved to have greatest power than separated applications of both markers for fine-scale stock identification of Chinook salmon (Narum *et al.* 2008; Hess *et al.* 2011).

As genomic data is accumulating for several non-human animal species, SNPs are becoming efficient and highly available tools among wildlife conservation-oriented studies and have attracted growing interest for addressing questions in evolutionary biology and ecology (Morin *et al.* 2004; Seddon *et al.* 2005; see also Slate *et al.* 2009 and Garvin *et al.* 2010 for a summary of applications and technical issues). But despite all the above-mentioned advantages and cited applications, the use of SNPs in wildlife population genetics has been limited (Morin *et al.* 2009) and they have yet to be applied in wildcat's molecular studies. In what concerns hybridization, SNPs-based genomic approaches may provide exciting opportunities to assess differential rates of introgression across different genomic regions. Native California tiger salamanders (*Ambystoma californiense*) denote an excellent example of SNPs potential in studying detail admixture processes, since Fitzpatrick *et al.* (2010) were able to determine that only 3 out of 68 studied markers spread rapidly into native genomes, whereas the other 65 showed little evidence of introgression beyond the region where introductions of non-native barred tiger salamanders (*Ambystoma tigrinum mavortium*) occurred. By exposing such clear evidences of loci heterogeneity in introgression rates, this work further reflects the vulnerability of studies using a few neutral markers to detect hybridization (Allendorf *et al.* 2010).

An additional advantage of SNPs compared to microsatellites lays in the fact that the target DNA sequence in SNP-based genotyping is appreciably shorter (e.g. 50–70 bp) than that using microsatellites, thus improving our capacity to deal with poor quality samples, such as historical, noninvasive and otherwise degraded and low copy number DNA (Brumfield *et al.* 2003; Morin *et al.* 2004, 2009). Since noninvasive techniques are becoming one of the most popular ways of sampling critically threatened species, low quantity and quality DNA is becoming the major source of genetic data of such valuable *taxa* and developing effective tools to genetically characterize that DNA is imperative. Taking into consideration all mentioned facts, SNPs will probably replace microsatellites as the marker of choice in conservation genetics of endangered species (Morin *et al.* 2009).

1.5.2. The promise of noninvasive sampling for wildcat's research

As for many rare and endangered species, the specific and individual identification of wildcats and their relatives is also exceptionally difficult due to their low population densities, crepuscular/nocturnal activity, elusive behaviour, and to the logistical and ethical problems involved in their capture and handling during traditional capture-mark recapture approaches. Applying a two-pronged approach using both invasive (blood and tissue samples) and noninvasive samples would be of major importance to better study this endangered feline. Noninvasive sampling using faeces and hairs has still to be used in wildcat conservation studies and determining diagnostic visual features to discriminate wildcat noninvasive samples proved to be even harder than distinguishing phenotypic traits. Not only wildcat scats are impossible to distinguish from feral domestic cats and hybrid individuals, but also cat faecal material is often confounded with other sympatric and resource competing carnivores (e.g. red fox, *Vulpes vulpes*; Davison *et al.* 2002). Using molecular methods of scat identification may solve the problem of morphologic misidentifications and prevent significant biases in the subsequent estimation of population parameters. And since noninvasive approaches are becoming one of the most popular ways of sampling critically threatened species, low quantity and quality DNA is becoming the major source of genetic data of such valuable *taxa*. Developing adequate molecular tools for obtaining and analysing such data has been a full-time job for many wildlife geneticists and should also be a priority in wildcat's genetic research.

Nearly 20 years ago, new laboratory and analytical techniques have been developed, which allows us studying the biology of populations without even having to observe or capture individuals (Höss *et al.* 1992; Taberlet and Bouvet 1992). By using DNA extracted from biological traces left by the animal in the field, the noninvasive procedure avoids the side effect of impacting on individuals' survival and population dynamics, while providing a substantial increment in sampling numbers with lower costs (Kohn and Wayne, 1997). Among samples collected noninvasively, faeces (e.g. Creel *et al.* 2003; Fabbri *et al.* 2007; Ruiz-González *et al.* 2008), hairs (e.g. Kendall and McKelvey 2008), urine (e.g. Hausknecht *et al.* 2007), saliva (e.g. Sundqvist *et al.* 2008; Sastre *et al.* 2009) or scent marks (e.g. Lanyon *et al.* 2007) are the most explored material, being scats the most popular one.

The current availability of high-quality laboratory reagents combined with sophisticated analysis software allows, nowadays, a detailed study of nearly all kind of samples collected noninvasively in the field. Nevertheless, lower PCR amplification success and higher genotyping error rates are expected in comparison with high-quality invasive samples (such as fresh tissue and blood) and tend to limit the efficiency of this approach. To overcome, or at least to minimize these difficulties, several precautions and methodological guidelines should be always taken in consideration. The major drawbacks of using these samples that may lead to important failure and/or error rates are related with the fact that they usually provide DNA extracts characterized by low target DNA concentration and quality and by the presence of contaminants and various molecules that may disturb or inhibit PCR amplification (e.g. Taberlet *et al.* 1999; Roon *et al*., 2003; Broquet *et al.*, 2007). At the same time, amplification success is usually dependent on sample storage method, the collection season, and the species diet and the age of the scat (when dealing with stool samples), a number of

factors that are hardly simultaneously controlled at their maximum success levels. Even after being able to successfully extract and amplify the target DNA (while avoiding contamination), one should expect that data produced from noninvasive samples would contain substantial genotyping errors. Especially in microsatellite data, genotypes from noninvasive samples can be affected by two main errors that may be reflected on allele frequency estimates and genotypes discrimination: allelic dropout (ADO), which is the stochastic failure of one allele to amplify for heterozygous individuals, producing false homozygotes, and false alleles (FA), which are artefacts of amplification products generated during the first steps of PCR that can be misinterpreted as true alleles (Navidi *et al.* 1992; Taberlet *et al.*, 1996; Goossens *et al.*, 1998; Bradley and Vigilant, 2002). Since microsatellite genotyping is usually applied in individual identification, parentage and relatedness analyses, and global population genetic studies, genotyping errors might create an artificial excess of individuals (Creel *et al.* 2003), homozygotes and inbreeding rates (Taberlet *et al.* 1996; 1999), as well as false departures from Hardy-Weinberg equilibrium (Xu *et al.* 2002) and overall unreliable inferences of populations substructures and individuals relatedness (Miller *et al.* 2002). The key to deal with such demanding problems is to be able to detect, document and minimize errors through adequate analyses (e.g. Morin *et al.* 2001; Miller *et al.* 2002; Valière 2002). It is now common sense that dealing with error rates through final statistical analysis on produced data is cheaper and quickest (and anyway essential) than numerously replicate each genotype (e.g. Kalinowski *et al.* 2006; Johnson and Haydon (2007). At the same time, such approach will not depend on the amount of DNA we were able to extract for each sample. Choosing the most suitable attitude depends on our own believes and the data we have.

Although challenging, when overcoming the above-mentioned methodological issues the potential of noninvasive genetics is outstanding. In the last decade, many carnivores' conservation genetic studies based on noninvasive sampling have been published, providing numerous information on species/individuals identification, population parameters, behavioural patterns, and conservation and management strategies. Today, we are on the cusp of being able to unravel even more complicated scenarios related with long-standing ecological and evolutionary questions, since a growing number of noninvasive techniques yield good enough DNA and low enough genotyping error rates to allow researchers to address nearly all questions that can be addressed using traditional high-quality samples such as blood and tissue samples (e.g. Epps *et al.* 2006; Luikart *et al.*2008).

The role of molecular genetics in the biological conservation of rare, elusive and/or endangered species has been profiting from the rapid advances in noninvasive techniques. DNA-based methods using a wide variety of molecular markers can be reliably applied for detecting the presence and mapping the distribution of threatened species (Schwartz *et al.* 2004); determining individuals home-ranges and population sizes (e.g. Bellemain *et al.* 2005), gender (e.g. Lucchini *et al.* 2002; Kurose *et al.* 2005) and kinship (Gerloff *et al.* 1995; Rudnick *et al.* 2005); detecting important events such as gene flow (Broderick *et al.* 2003), population isolation or hybridization (e.g., Schwartz *et al.* 2004; Adams and Waits, 2007); monitoring population size (Frantz *et al.* 2004; Piggott *et al.* 2006); estimating species biological parameters (e.g. diet and habitat selection, Reed *et al.* 1997; Farrell *et al.* 2000); solving managing problems related to livestock attacks performed by wild protected animals (e.g. Blejwas *et al.* 2006; Sundqvist *et al.* 2008) or

identifying samples in forensic caseworks (e.g. Williams *et al.* 2003; Melton and Holland 2007; Wasser *et al.* 2007). On the light of these remarkable developments, the implementation of efficient noninvasive strategies may be seen as a singular opportunity to monitor populations of the European wildcat.

1.6. Objectives and thesis outline

According to the Council of Europe, the aims and priorities of the long-term, effective conservation and management of wildcats include: regular monitoring of their populations and distribution; research on hybridization and its effect; studies on the loss and destruction of habitats; and evaluation of mortality due to illegal hunting and road kills (Stahl and Artois 1991). In this thesis, we focused our research in contributing to better resolve the first two problematics. To accomplish this goal, we aimed at overcoming a number of important drawbacks in European wildcats' research, especially:

i) the extremely scarce knowledge on regional genetic diversity and purity of Iberian wildcats, one of the European populations that might be most critically endangered (IUCN, 2007);
ii) the difficulty in applying successful sampling and genotyping schemes in noninvasive DNA samples of European wildcats, especially scats, due to their ambiguous visual identification and their inherent low quality and quantity of amplifiable molecules;
iii) the low resolution achieved so far in the detection of hybrids between European wild and domestic cats and the extreme difficulty in discriminating between hybrid classes (F1, F2 and backcrosses), due to limitations in type and number of the molecular markers used to date.

The contents of this dissertation are organized in four chapters. In **Chapter 1** it is presented a *General Introduction* on the main topics that explain the current status of the European wildcat populations, while providing the necessary background that justifies the research presented afterwards in Chapter 2, 3 and 4. First we summarize the important empirical and conjectural knowledge on the domestication process of cats, and highlight the importance it may have had in the present day difficulty in identifying hybrid individuals and access accurate levels of admixture. We also provide relevant historical, environmental and conservation-oriented information on the species and describe some of the most important previous molecular studies curried out on European wildcats. We also focus this summary on the new molecular advances expected to significantly improve existing data, while reviewing major drawbacks of current analyses.

During the development of this thesis, we first started working in one of the most unknown and critically endangered populations of wildcats in Europe: the Iberian Peninsula. Even though wildcats are legally protected in Portugal and Spain on the basis of high hybridization risk and population fragmentation, the true impact of such phenomena was totally unknown and no one Knew in detail if a unique genetic identity of European wildcats (as we know them today) was still part of the native Iberian fauna. In **Chapter 2**, we provide for the first time a global genetic survey of Iberian wildcats aiming to determine the frequency, extension and the impact that domestic introgression might be having in endemic natural populations. We

address these questions through an integrated approach combining the use of highly polymorphic loci and Bayesian statistical inferences to i) investigate the extend of genetic variation and differentiation in Portuguese and Spanish wild and domestic cat populations; (ii) pinpoint hybridization and evaluate introgression of domestic alleles, and (iii) provide new insights and critical guidelines to the regional and global conservation of this threatened feline. The results are organized in two scientific papers already published in SCI journals:

Paper I. Oliveira R, Godinho R, Randi E, Ferrand N and Alves PC (2008) Molecular analysis of hybridization between wild and domestic cats (*Felis silvestris*) in Portugal: implications for conservation. *Conservation Genetics* **9**: 1-11.

Paper II. Oliveira R, Godinho R, Randi E and Alves PC (2008) Hybridization vs conservation: are domestic cats threatening the genetic integrity of European wildcat (*Felis silvestris silvestris*) populations in Iberian Peninsula? *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **363**: 2953-2961.

In the first manuscript we provide evidences that incongruences between phenotypic and genetic identifications of wildcats are common, which supports the idea that genetic identifications are essential tools in wildcat conservation. Similarly to other European populations considered markedly differentiated from the domestic form, differential allele frequencies, private alleles and significant values of genetic differentiation revealed a clear genetic distinction between Portuguese wild and domestic cats. But even though we found no genetic evidences of a constant and generalized gene flow between sympatric populations of wild and domestic cats, at least in most recent generations, admixture analysis revealed a significant proportion of hybrids (around 14%), distributed across the entire country. In *Paper II* we extended previous findings on the Portuguese population to the entire Iberian Peninsula, and discuss in detail the statistical confidence achieved using 12 unkinked microsatellites for cat's admixture analysis. Hybrid cats were exclusively identified in Portugal and closer genetic similarity was found between Portuguese wild and domestic cats, which might be an indication of higher levels of recent introgression when compared with Spain. We detected an increased power relatively to the first hybridization inferences (*Paper I*), an improvement achieved by the significant increase on the representatives of both parental *taxa* (wild versus domestic cats). Nevertheless, the number of hybrids identified in this study still represents a minimum number of admixed individuals, since a proportion of F2 and backcrosses cats remained undetected.

During the process of assembling data to produce the previous manuscripts, we immediately faced one of the major difficulties that most wildlife geneticists have to deal with when studying endangered and rare *taxa*: the low number of samples. This difficulty is particularly prejudicial when trying to accurately

describe subtle population structure and has been hampering the fast advance of wildcat genetic studies. Increasing sampling effort both at time and space levels seemed, thus, one of the most important drawbacks to overcome. Since mainly scat surveys provide time and cost-effective sampling efforts and significantly reduce anthropogenic pressures (e.g. animals capture and handling) that are frequently questioned on ethical bases, reinforcing our sampling scheme with noninvasive sampling procedures became a priority. However, no one had already applied such demanding approaches on wildcat's genetics and many decisions would have to be taken. At the same time, noninvasive genetics was (and probably still is) embracing years of total revolution and an immensurable number of innovative papers were being published. Bearing this in mind, we started our way to **Chapter 3**, where we present two already published papers in SCI journals:

Paper III. Beja-Pereira A and Oliveira R, Alves PC, Schwartz MK and Luikart G (2009) Advancing ecological understandings through technological transformations in non-invasive genetics. *Molecular Ecology Resources* **9** (4): 1279-1301.

Paper IV. Oliveira R, Castro D, Godinho R, Luikart G and Alves PC (2010) Species identification using a simple SSCP analysis of a nuclear gene: application to carnivores of southwest Europe. *Conservation Genetics* **11** (3): 1023-1032.

In *Paper III*, we reviewed in detail the most important steps in noninvasive genetic studies, from pre-PCR sampling to post-PCR data analysis, focusing in the most effective means to overcome major difficulties. We tried to answer important technical questions related with the production of the most reliable and accurate data, namely: i) which is the better way to obtain and preserve different types of noninvasive samples?; ii) how to extract and amplify DNA with the highest success rates and lower genotyping errors?; iii) how to be aware of the remaining inherent errors, and estimate, document and circumvent them? Among the most important perspectives we suggest that molecular ecologists should, on a permanent basis, explore the literature and cross-reference with the fields of forensics, human health and domestic animal health science, research areas that continuously generate technical improvements that can be applied in wildlife noninvasive genetics to improve both data production and analysis.

In the second manuscript (*Paper IV*) we directly deal with the challenges of using scats (and hair) samples in European wildcat's research. Applying noninvasive sampling schemes for studying this species - especially in places of sympatry with other carnivores with overlapping diet, activity patterns, behaviour and basic body shape (e.g. *Vulpes vulpes*, domestic cats and dogs) - might be hampered by important difficulties in visually identifying the specific origin of the sign. To correctly assign noninvasive samples to the wildcat, we developed a simple molecular test based on the high polymorphism detected among carnivores on the IRBP (*Interphoto-receptor Retinoid-Binding Protein*) gene. Identifications at the species level were achieve for all extant carnivores living in Iberian Peninsula through SSCP analysis of a small fragment of the gene's exon 1,

and high rates of PCR amplification and SSCP identification were retrieved both using faecal and hair material. With this simple test we provide a way to select putative wildcat samples for further genotyping in population and hybridization studies, not only because we rule out the possibility of having sampled other similar species but also because we use the amplification success of this fragment as pre-screening barometer of nuclear DNA quality and quantity.

But the molecular analyses performed in Chapter 2 clearly revealed that innovations in wildcat's research should not only focus on new sampling schemes (explored in Chapter 3), but also in the significant improvement of molecular tools, especially to overcome hybrids identification uncertainties. In **Chapter 4** we embrace this problem under two different perspectives, which allowed the design of two independent manuscripts that are currently under preparation:

Paper V. A dangerous return to nature: are free-ranging domestic cats threatening the genetic integrity of European wildcats in their genetically disrupted distribution? *In prep*

Paper VI. Nuclear Genome SNPs to Detect European wildcat *(Felis silvestris silvestris)* and domestic cats *(Felis s. catus)* hybridization**.** *In prep*

First, in *Paper V*, we increased almost four times the number of *loci* used so far to differentiate wild and domestic cat forms, and applied a total set of 38 microsatellites to genotype more than 1000 cat samples distributed across the entire distribution range of the species in Europe. Most important results indicate that past habitat fragmentation and demographic declines probably led to a detectable genetic diversification among wildcat populations. Our findings also confirm previous works on Scottish and Hungarian populations, which are clearly composed of a hybrid swarm with a variety of introgression degrees that might result from long-lasting hybridization. Cryptic hybrids were detected in low hybridizing populations (namely in Iberia, Italy, Germany, Slovenia and Bulgaria), suggesting that although crossbreeding may not be extensively occurring in these populations, at least in recent generations, it exists and might have important repercussions in the future if natural populations keep declining and fragmenting. Simulation analyses revealed an unsatisfying analytical power of the used loci to identify backcrossed hybrids among the analysed dataset, prompting the search of new informative molecular markers.

In *Paper VI*, we performed a wide genetic investigation of novel loci to establish a set of diagnostic tools suitable for accurately detect levels of introgressive hybridization between European wild and domestic cats. We based our investigation both in randomly dispersed variation found in the published domestic cat genome (Pointius *et al.* 2007) and in a candidate genes approach to detect evidence of artificial diversification. 139 random-bred cats, 133 European wildcats and 5 known hybrids (successfully genotyped using microsatellites in *Paper V*) were genotyped at 158 SNPs randomly dispersed in autosomes, in chromosome X, in genes related with previous described variation among domestic cats (ranging from

mutations related with coat colour patterns – e.g. in the *tyrosinase related protein 1* TYRP1 gene – to variation in genes regulating important diseases – e.g. *cardiac troponin* TNN13 gene), and candidate genomic regions that revealed at least one polymorphic position between European wild and domestic cats in previously published sequence data (Johnson *et al*. 2006; Esteves *et al*. 2007). We estimated levels of genetic variability and differentiation among wild and domestic cat populations, and evaluated the power of all loci to accurately identify admixture events and discriminate the different hybrid categories that might result from crossbreeding (F1, F2 and backcrosses). Results from Bayesian model-based computations of simulated and true genotypes showed that the entire set of 158 markers provides successful estimates of admixture, with just a few cases of hybrid cats remaining misclassified. Moreover, the use of reduced sets of highly informative content (n=35) provided a powerful approach to infer admixed ancestries. A simpler test could then be routinely applied, for example, in noninvasive studies of the species, reducing costs, labour and time of the analyses.

Finally, in **Chapter 5** we presented a *General Discussion* of the most relevant results obtained in this work and provided a detailed comparison with the most relevant works focusing artificial hybridization, genetic fragmentation and noninvasive genetics in natural (and in particular protected) populations. Additionally, we specifically discuss *Future Research and Perspectives* that may help clarifying the complexity of introgressive hybridization in European wildcat populations, and improve our capacity to effectively protect the species from irreversible genetic extinction. In **Chapter 6** we summarize the major *Conclusion*s obtained in the different chapters.

1.7. References

Adams JR and Waits LP (2007) An efficient method for screening faecal DNA genotypes and detecting new individuals and hybrids in the red wolf (*Canis rufus*) experimental population area. *Conservation Genetics* **8**: 123-131l.

Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001) The problems with hybrids: setting conservation guidelines. *Trends in Ecology and Evolution* **16**: 613-622.

Allendorf FW and Luikart G (2007*) Conservation and the Genetics of Populations*. Blackwell Publishing, Malden MA, USA.

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. *Nature Reviews Genetics* **11** (10): 697-709.

Andersen DH, Fabbri E, Santini A, Paget S, Cadieu E, Galibert F, André C, Randi E (2006) Characterization of 59 canine single nucleotide polymorphisms in the Italian wolf (*Canis lupus*) population. *Molecular Ecology Notes* **6** (4): 1184-1187.

Andersone Z, Lucchini V, Randi E, Ozolins J (2002) Hybridization between wolves and dogs in Latvia as documented using mitochondrial and microsatellite DNA markers. *Mammalian Biology* **67**: 79-90.

Antunes A, Pontius J, Ramos MJ, O'Brien SJ, Johnson WE (2007) Mitochondrial Introgressions into the Nuclear Genome of the Domestic Cat. *Journal of Heredity* **98**: 414–420.

Arnold ML (1992) Natural hybridization as an evolutionary process. *Annual Review of Ecology and Systematics* **23**: 237-261.

Arnold ML (1997) *Natural Hybridization and Evolution.* Oxford Series in Ecology and Evolution. Oxford University Press, Oxford.

Arnold ML (2006) *Evolution through genetic exchange* Oxford University Press, New York.

Arnold ML (2004) Natural hybridization and the evolution of domesticated, pest, and disease organisms. *Molecular Ecology* **13**: 997–1007.

Barton NH (2001) The role of hybridization in evolution. *Molecular Ecology* **10**: 551–568.

Beaumont M, Barratt EM, Gottelli D, Kitchener AC, Daniels MJ, Pritchard JK, Bruford MW (2001) Genetic diversity and introgression in the Scottish wildcat. *Molecular Ecology* **10**: 319–336.

Belardinelli A (2001) Distribution, Activity, Morphological and Morphometrical Characters of E*rinaceus concolor nesiotes*, *Mustela nivalis galinthias*, *Martes foina bunites*, *Meles meles arcalus*, *Felis silvestris cretensis* in Crete. *M.Sc Thesis in Management of Marine and Land-living Resources*. University of Crete.

Bellemain E, Swenson JE, Tallmon O, Brunberg S, Taberlet P (2005) Estimating population size of elusive animals with DNA from hunter-collected feces: four methods for brown bears. *Conservation Biology* **19**:150–161.

Belyaev D (1979) Destabilizing selection as a factor in domestication. *Journal of Heredity* **70**: 301-308.

Biró Z, Lanszki J, Szemethy L, Heltai M, Randi E (2005) Feeding habits of feral domestic cats (*Felis catus*), wild cats (*Felis silvestris*) and their hybrids: trophic niche overlap among cat groups in Hungary. *Journal of Zoology* **266**: 187–196.

Biró Z, Szemethy L, Heltai M (2004) Home range sizes of wildcats (*Felis silvestris*) and feral domestic cats (*Felis silvestris catus*) in a hilly region of Hungary. Mammalian Biology **69**: 302–310.

Björnerfeldt S, Webster MT, Vilà C (2006) Relaxation of selective constraint on dog mitochondrial DNA following domestication. *Genome research* **16**: 990-4.

Blanco JC and Cortés Y (2002) *Ecología, censos, percepción y evolución del lobo en España: análises de un conflicto*. SECEM, Málaga, 176 pp.

Blejwas KM, Williams CL, Shin GT, McCullough DR, Jaeger MM (2006) Salivary DNA evidence convicts breeding male coyotes of killing sheep. *Journal of Wildlife Management* **70**: 1087-1093.

Boichard D, Grohs C, Bourgeois F, Cerqueira F, Faugeras R, Neau A, Rupp R, Amigues Y, Boscher MY, Leveziel H (2003) Detection of genes influencing economic traits in three French dairy cattle breeds. *Genetics Selection Evolution* **35**: 77-101.

Bradley BJ and Vigilant L (2002) False alleles derived from microbial DNA pose a potential source of error in microsatellite genotyping of DNA from feces. *Molecular Ecology Notes* **2**: 602-605.

Broderick D, Idaghdour Y, Korrida A, Hellmich J (2003) Gene flow in great bustard populations across the Strait of Gibraltar as elucidated from excremental PCR and mtDNA sequencing. *Conservation Genetics* **4**: 793 – 800.

Broquet T, Menard N, Petit E (2007) Noninvasive population genetics: a review of sample source, diet, fragment length and microsatellite motif effects on amplification success and genotyping error rates. *Conservation Genetics* **8**: 249–260.

Bruford MW, Bradley DG, Luikart G (2003) DNA markers reveal the complexity of livestock domestication. *Nature Reviews Genetics* **4**: 900-910.

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

Canters KJ, Thissen JBM, van Diepenbeek A, Jansman HAH, Goutbeek K (2005) The wildcat (*Felis silvestris*) finally recorded in the Netherlands. *Lutra* **48**: 67-90.

Carneiro M, Afonso S, Geraldes A, Garreau H, Bolet G, Boucher S, Queney G, Tircazes A, Nachman MW and Ferrand N (2011) The genetic structure of domestic rabbits. *Molecular Biology and Evolution* **28**: 1801-1806.

Casci T (2011) Population genetics: SNPs that come in threes. *Nature Reviews Genetics* **11**(1): 8.

Coates BS, Sumerford DV, Miller NJ Kim Ks, Sappington TW, Siegfried BD, Lewis LC (2009) Comparative Performance of Single Nucleotide Polymorphism and Microsatellite Markers for Population Genetic Analysis. *Journal of Heredity* **100**: 556–564.

Clutton-Brock J (1999) *A Natural History of Domesticated Mammals*, Cambridge University Press.

Cooper MP, Fretwell N, Bailey SJ, Lyons LA (2006) *White spotting* in the domestic cat (*Felis catus*) maps near *KIT* on feline chromosome B1. *Animal Genetics* **37**: 163–165.

Creel S, Spong G, Sands JL, Rotella J, Zeigle J, Joe L, Murphy KM, Smith D (2003) Population size estimation in Yellowstone wolves with error-prone noninvasive microsatellite genotypes. *Molecular Ecology* **12**: 2003–2009.

Cruz F, Vila C, Webster MT (2008) The Legacy of Domestication: Accumulation of Deleterious Mutations in the Dog Genome. *Molecular Biology and Evolution* **25**: 2331-2336.

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* **244**: 231–247.

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

Daniels MJ, Golder MC, Jarrett O, MacDonald DW (1999) Feline viruses in wildcats from Scotland. *Journal of Wildlife Diseases* **35**: 121-4.

Davis SJ and Valla FR (1978) Evidence for domestication of the dog 12,000 years ago in the Natufian of Israel. *Nature* **276**: 608-610.

Davis BW, Raudsepp T, Wilkerson AJP, Agarwala R, Schäffer AA, Houck M, Chowdhary BP, Murphy WJ (2009) A high-resolution cat radiation hybrid and integrated FISH mapping resource for phylogenomic studies across Felidae. *Genomics* **93**: 299–304.

Davison A, Birks JDS, Griffiths, Kitchener AC, Biggins D (1999) Hybridization and the phylogenetic relationship between polecats and domestic ferrets in Britain. *Biological Conservation* **87**: 155-161.

Davison A, Birks JDS, Brookes RC, Braithwaite TC, Messenger JE (2002) On the origin of faeces: morphological versus molecular methods for surveying rare carnivores from their scats. *Journal of Zoology (London)* **257**: 141–143.

Diamond J (2002) Evolution, consequences and future of plant and animal domestication. *Nature* **418**: 700-707.

Dobney K and Larson G (2006) Genetics and animal domestication: new windows on an elusive process. *Journal of Zoology* **269**: 261-271.

Doebley JF, Gaut BS, Smith BD (2006) The molecular genetics of crop domestication. *Cell* **127**: 1309-1321.

Driscoll CA, Menotti-Raymond M, Roca AL, Hupe K, Johnson WE, Geffen E, Harley EH, Delibes M, Pontier D, Kitchener AC, Yamaguchi N, O'Brien SJ, MacDonald DW (2007) The near eastern origin of cat domestication. *Science* **317**: 519–523.

Driscoll CA, MacDonald DW, O'Brien SJ (2009) From wild animals to domestic pets, an evolutionary view of domestication. *Proceedings of the National Academy of Sciences USA* **106**: 9971-9978.

Driscoll C and Nowell K (2009) *Felis silvestris*. In: IUCN 2010. IUCN Red List of Threatened Species. Version 2010.4. <www.iucnredlist.org>.

Drogemuller C, Rufenacht S, Wichert B, Leeb T (2007) Mutations within the FGF5 gene are associated with hair length in cats. *Animal Genetics* **38**: 218-221.

Drogemuller C, Karlsson EK, Hytonen MK, Perloski M, Dolf G, Sainio K, Lohi H, Lindblad-Toh K, Leeb T (2008) A mutation in hairless dogs implicates FOXI3 in ectodermal development. *Science* **321**: 1462-1462.

Duarte J, Vargas JM (2001) Son selectivos los controles de predadores en los cotos de caza? *Galemys* **13**: 1–9.

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

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

Eizirik E, Yuhki N, Johnson WE, Menotti-Raymond M, Hannah SS, O'Brien SJ *et al*. (2003) Molecular genetics and evolution of melanism in the cat family. *Current Biology* **13**: 448-53.

Eizirik, E, David V, Buckley-Beason V, Roelke ME, Schäffer A, Hannah SS, Narfström K, O'Brien SJ, Menotti-Raymond M (2010) Defining and mapping mammalian coat pattern genes: multiple genomic regions implicated in domestic cat stripes and spots. *Genetics* **184**: 267-75.

Epps CW, Palsbøll PJ, Wehausen JD, Roderick GK, McCullough DR (2006) Elevation and connectivity define genetic refugia for mountain sheep as climate warms. *Molecular Ecology* **15**: 4295–4302.

Eriksson J, Larson G, Gunnarsson U, Bed'hom B, Tixier-Boichard M, Stromstedt L, Wright D, Jungerius A, Vereijken A, Randi E, Jensen P, Andersson L (2008) Identification of the yellow skin gene reveals a hybrid origin of the domestic chicken. *PLoS Genetics* **4**(2): e1000010.

Fabbri E, Miquel C, Lucchini V, Santini A, Caniglia R, Duchamp C, Weber JM, Lequette B, Marucco F, Boitani L, Fumagalli L, Taberlet P, Randi E (**2007**) From the Apennines to the Alps: colonization genetics of the naturally expanding Italian wolf (*Canis lupus*) population. *Molecular Ecology* **16:** 1661-1671.

Falush D, Stephens M, Pritchard J (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**: 1567–1587.

Fang MY, Larson G, Ribeiro HS, Li N, Andersson L (2009) Contrasting mode of evolution at a coat color locus in wild and domestic pigs. *Plos Genetics* **5**(1): e1000341.

Farrel LE, Romant J, Sunquist ME (2000) Dietary separation of sympatric carnivores identified by molecular analysis of scats. *Molecular Ecology* **9**:1583–1590.

Fernandes ML (2005) O gato-bravo no Nordeste Trasmontano. João Azevedo Editor. Mirandela. Portugal.

Fitzpatrick BM, Johnson JR, Kump DK, Smith JJ, Voss SR, Shaffer HB (2010) Rapid spread of invasive genes into a threatened native species. *Proceedings of the National Academy of Sciences USA* **107**: 3606–3610.

Frantz AC, Schaul M, Pope LC, Fack F, Schley L, Muller CP, Roper TJ (2004) Estimating population size by genotyping remotely plucked hair: the Eurasian badger. *Journal of Applied Ecology* **41**: 985–995.

Fromont E, Sager A, Léger F, Bourguemestre F, Jouquelet E, Stahl P, Pontier D, Artois M (2000) Prevalence and pathogenicity of retroviruses in wildcats in France. *Veterinary Record* **146**: 317-9.

Gandolfi B, Bach L, Beresford L, *et al.*: Off with the gloves: mutation in KIT implicated for the unique white spotting phenotype of Birman cats. (in press)

Gandolfi B, Outerbridge CA, Beresford LG, Myers JA, Pimentel M, Alhaddad H, Grahn JC, Grahn RA, Lyons LA (2010) The naked truth: Sphynx and Devon Rex cat breed mutations in KRT71. *Mammalian Genome* **21**: 509-515.

García-Perea R (2002) Andean mountain cat, oreailurus jacobita: morphological description and comparison with other felines from the altiplano. *Journal of Mammalogy* **83**: 110-124.

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

Gerloff U, Schlötterer C, Rassmann K, Rambold I, Hohmann G, Frutth B, Tautz D (1995) Amplification of hypervariable simple sequence repeats (microsatellites) from excremental DNA of wild living Bonobos (*Pan paniscus*). *Molecular Ecology* **4**: 515–518.

Germain E, Benhamou S, Poulle M-L (2008) Spatio-temporal sharing between the European wildcat, the domestic cat and their hybrids. *Journal of Zoology* **276**: 195–203.

Germain E, Ruette S, Poulle M-L (2009) Likeness between the food habits of European wildcats, domestic cats and their hybrids in France. *Mammalian Biology Zeitschrift fur Saugetierkunde* **74**: 412-417.

Gippoliti S and Amori G (2006) Ancient introductions of mammals in the Mediterranean Basin and their implications for conservation. *Mammal Review* **36**: 37-48.

Gittleman JL, Funk SM, Macdonald DW, Wayne RK (2001) Carnivore conservation (Conservation Biology, Vol. 5). Lavoisier, Cambridge.

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

Godinho R, Llaneza L, Blanco L, Lopes S, Álvares F, García E, Palacios V, Cortés Y, Talegón J and Ferrand N (2011) Genetic evidence for multiple events of hybridization between wolves and domestic dogs in the Iberian Peninsula. *Molecular Ecology* **20**: 5154–5166.

Goossens B, Waits LP, Taberlet P (1998) Plucked hair samples as a source of DNA: reliability of dinucleotide microsatellite genotyping. *Molecular Ecology* **7**: 1237-1241.

Grabe H and Worel G (2001) *Die Wildkatze. Zurück auf leisen Pfoten*, pp. 14-27. Buch and Kunstverlag Oberpfalz, Amberg.

Gray MM, Granka JM, Bustamante CD, Sutter NB, Boyko AR, Zhu L, Ostrander EA, Wayne RK (2009) Linkage disequilibrium and demographic history of wild and domesticated canids. *Genetics* **18**: 1493–1505.

Gregory T (2009) Artificial Selection and Domestication: Modern Lessons from Darwin's Enduring Analogy. *Evolution Education and Outreach* **2**: 5–27.

Haltenorth T (1953) *Die wildkatzen der alten welt*. Leipzig: Geest and Portig.

Hausknecht R, Gula R, Pirga B, Kuehn R (2007) Urine-A source for noninvasive genetic monitoring in wildlife. *Molecular Ecology Notes* **7**: 208–212.

Hedrick PW (2004) Foxy MHC selection story. *Heredity* **93**: 237-238.

Helyar SJ, Hemmer-hansen J, Bekkevold D, Taylor MI, Ogden R, Limborg MT, Cariani A, Maes GE, Diopere E, Carvalho GR, Nielsen EE (2011) Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Molecular Ecology Resources* **11**: 123–136.

Hertwig ST, Jungnickel A, Stepanow S, Böhle UR, Fischer MS (2009) Regionally high rates of hybridization and introgression in German wildcat populations (*Felis silvestris*, Carnivora, Felidae). *Journal of Zoological Systematics and Evolutionary Research*. **47**: 283–297.

Hess JE, Matala AP, Narum SR (2011) Comparison of SNPs and microsatellites for fine-scale application of genetic stock identification of Chinook salmon in the Columbia river basin. *Molecular Ecology resources* **11**: 137–149.

Hillbertz N, Isaksson M, Karlsson EK, Hellmen E, Pielberg GR, Savolainen P, Wade CM, Von Euler H, Gustafson U, Hedhammar K, Nilsson M, Lindblad-Toh K, Andersson L, Andersson G (2007) Duplication of *FGF3*, *FGF4*, *FGF19* and *ORAOV1* causes hair ridge and predisposition to dermoid sinus in Ridgeback dogs. *Nature Genetics* **39**: 1318-1320.

Höss M, Kohn M, Pääbo S, Knauer F, Schröder W (1992) Excrement analysis by PCR. *Nature* **359**: 199-199.

Hubbard AL, McOris S, Jones TW, 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**: 203–208.

Imes DL, Geary LA, Grahn RA, Lyons LA (2006) Albinism in the domestic cat (*Felis catus*) is associated with a tyrosinase (TYR) mutation. *Animal Genetics* **37**: 175-178.

Ishida Y, David VA, Eizirik E, Schäffer AA, Neelam BA, Roelke ME, Hannah SS, O'brien SJ, Menotti-Raymond M (2006) A homozygous single-base deletion in MLPH causes the dilute coat color phenotype in the domestic cat. *Genomics* **88**: 698-705.

Jaeger JAG, Bowman J, Brennan J, Fahrig L, Bert D, Bouchard J, Charbonneau N, Frank K, Gruber B, von Toschanowitz KT (2005) Predicting when animal populations are at risk from roads: an interactive model of road avoidance behavior. *Ecological Modelling* **185**: 329–348.

Jansen T, Forster P, Levine MA, Oelke H, Hurles M, Renfrew C, Weber J, Olek K (2002) Mitochondrial DNA and the origins of the domestic horse. *Proceedings of the National Academy of Sciences USA* **99**: 10905-10910.

Jerosch S, Götz M, Klar N, Roth M (2010) Characteristics of diurnal resting sites of the endangered European wildcat (*Felis silvestris silvestris*): Implications for its conservation. *Journal for Nature Conservation* **18**: 45-54.

Johnson PCD and Haydon DT (2007) Maximum-likelihood estimation of allelic dropout and false allele error rates from microsatellite genotypes in the absence of reference data. *Genetics* **175**: 827–842.

Johnson WE and O'Brien SJ (1997) Phylogenetic reconstruction of the Felidae using 16S rRNA and NADH-5 mitochondrial genes. *Journal of Molecular Evolution* **44** (Suppl. 1): S98–S116.

Kalinowski ST, Taper ML, Marshall TC (2006) Revising how the computer program Cervus accommodates genotyping error increases success in paternity assignment. *Molecular Ecology* **16**: 1099–1106.

Kehler JS, David VA, Schaffer AA, Bajema K, Eizirik E, Ryugo DK, Hannah SS, O'Brien SJ, Menotti-Raymond M (2007) Four independent mutations in the feline fibroblast growth factor 5 gene determine the long-haired phenotype in domestic cats. *Journal of Heredity* **98**: 555-566.

Kendall KC and McKelvey KS (2008) Hair collection. In: *Noninvasive Survey Methods for North American Carnivores* (eds Long RA, MacKay P, Ray JC, Zielinski WJ), pp. 135–176. Island Press, Washington, DC.

Kidd AG, Bowman J, Lesbarrères D, Schulte-hostedde AI (2009) Hybridization between escaped domestic and wild American mink (*Neovison vison*). *Molecular Ecology* **18**: 1175-1186.

Kitchener AC (1991) *The natural history of the wild cats*. Cornell University Press, New York. 280 p.

Kitchener AC and Dugmore AJ (2000) Biogeographical change in the tiger, *Panthera tigris*. *Animal Conservation* **3**: 113-124.

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

Kitchener AC and Rees EE (2009) Modelling the dynamic biogeography of the wildcat: implications for taxonomy and conservation. *Journal of Zoology* **279**: 144–155.

Kitchener AC, Daniels M, Yamaguchi N, Kilshaw K, Driscoll C (2010) Reversing cryptic extinction: the history, present and future of the Scottish Wildcat. In: D.W. Macdonald and Loveridge A.C. (eds), *Biology and conservation of wild felids*, Oxford University Press, Oxford, pp. 471-491.

Klar N, Fernández N, Kramer–Schadt S, Herrmann M, Trinzen M, Büttner I, Niemitz C (2008) Habitat selection models for European wildcat conservation. *Biological Conservation* **141**: 308–319.

Klar N Herrmann M, Kramer-Schadt S (2009) Effects and Mitigation of Road Impacts on Individual Movement Behavior of Wildcats. *Journal of Wildlife Management* **73** (5): 631-638.

Kohn MH, Murphy WJ, Ostrander EA, Wayne RK (2006) Genomics and conservation genetics. *Trends in Ecology and Evolution* **21**: 629-37.

Koskinen MT, Hirvonen H, Landry PA, Primmer CR (2004) The benefits of increasing the number of microsatellites utilized in genetic population studies: an empirical perspective. *Hereditas* **141**: 61–67.

Krüger M, Hertwig ST, Jetschke G, Fischer MS (2009) Evaluation of anatomical characters and the question of hybridization with domestic cats in the wildcat population of Thuringia, Germany. *Journal of Zoological Systematics and Evolutionary Research* **47**: 268–282.

Kurose N, Masuda R, Tatara M (2005) Fecal DNA Analysis for Identifying Species and Sex of Sympatric Carnivores: A Noninvasive Method for Conservation on the Tsushima Islands, Japan. *Journal of Heredity* **96**: 688-697.

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

Lanyon CV, Rushton SP, O'Donnell AG, Goodfellow M, Ward AC, Petrie M, Jensen SP, Morris Gosling L, Penn DJ (2007) Murine scent mark microbial communities are genetically determined. FEMS *Microbiology Ecology* **59**: 576–583.

Lao O, de Gruijter JM, van Duijn K, Navarro A, Kayser M (2007) Signatures of positive selection in genes associated with human skin pigmentation as revealed from analyses of single nucleotide polymorphisms. *Annals of Human Genetics* **71:** 354-369.

Larson G, Dobney K, Albarella U, Fang M, Matisoo-Smith E, Robins J, Lowden S, Finlayson H, Brand T, Willerslev E, Rowley-Conwy P, Andersson L, Cooper A (2005) Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science* **307**: 1618-1621.

Lau AN, Peng L, Goto H, Chemnick L, Ryder OA, Markova KD (2009) Horse domestication and conservation genetics of Przewalski's horse inferred from sex chromosomal and autosomal sequences. *Molecular Biology and Evolution* **26**: 199–208.

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

Leutenegger CM, Hofmann-Lehmann R, Riols C, Liberek M, Worel G, Lups P, Fehr D, Hartmann M, Weilenmann P, Lutz H. (1999) Viral infections in free-living populations of the European wildcat. *Journal of Wildlife Diseases* **35**: 678-86.

Li WH (1997) *Molecular Evolution*. Sinauer Associates, Sunderland, MA.

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

Liu N, Chen L, Wang S, Oh C, Zhao H (2005) Comparison of singlenucleotide polymorphisms and microsatellites in inference of population structure. *BMC Genetics* **6**(Suppl. 1): S26.

Lodé T, Guiral G, Peltier D (2005) European Mink–Polecat Hybridization Events: Hazards from Natural Process? *Journal of Heredity* **96**: 89–9.

Lopez JV, Cevario S and O'Brien SJ (1996) Complete nucleotide sequences of the domestic cat (*Felis catus*) mitochondrial genome and a transposed mtDNA tandem repeat (Numt) in the nuclear genome. *Genomics* **33**: 229–246.

Lozano J, Virgós E, Malo AF, Huertas DL, Casanovas JG (2003) Importance of scrub-pastureland mosaics for wild-living cats occurrence in a Mediterranean area: implications for conservation of the wildcat (*Felis silvestris*). *Biodiversity and Conservation* **12**: 921–935.

Lozano J, Moleón M, Virgós E (2006) Biogeographical patterns in the diet of the wildcat, *Felis silvestris* Schreber, in Eurasia: factors affecting the trophic diversity. *Journal of Biogeography* **33**: 1076-1085.

Lozano J, Virgós E, Cabezas–Díaz S, Mangas J G (2007) Increase of large game species in Mediterranean areas: is the European wildcat (*Felis silvestris*) facing a new threat? *Biological Conservation* **138:** 321–329.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews in Genetics* **4**: 981–94.

Luikart G, Zundel S, Rioux D, Miquel C, Keating KA, Hogg JT, Steele B, Foresman K, Taberlet P (2008) Low genotyping error rates for microsatellite multiplexes and noninvasive fecal DNA samples from bighorn sheep. *Journal of Wildlife Management* **72**: 299–304.

Lüps P, Flückiger PF, Peier D, Schmidt P (2002) Fund einer Waldkatze *Felis silvestris* bei Oberbuchsiten. *Mitteilungen der Naturforschenden Gesellschaft des Kantons Solothurn* **39**: 41-45.

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

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

Lyons L (2010) Feline Genetics: Clinical Applications and Genetic Testing Topics in Companion Animal. *Medicine* **25**: 203-212.

Lyons L (2012) Genetic testing in domestic cats. *Molecular and Cellular Probes* **26**(6): 224-30.

MacDonald DW, Daniels MJ, Driscoll C, Kitchener A, Yamaguchi N (2004) *The Scottish Wildcat. Analyses for Conservation and an Action Plan*. Wildlife Conservation Research Unit, Oxford.

Mallet J (2005) Hybridization as an invasion of the genome. *Trends in Ecology and Evolution* **20** (5): 229-237.

Mallet J (2007) Hybrid speciation. *Nature* **446**: 279-283.

Malo AF, Lozano J, Huertas DL, Virgós E (2004) A change of diet from rodents to rabbits (*Oryctolagus cuniculus*). Is the wildcat (*Felis silvestris*) a specialist predator? *Journal of Zoology (London)* **263**: 401–407.

Manel, S, Berthier P, Luikart G (2002) Detecting Wildlife Poaching: Identifying the Origin of Multilocus Genotypes. *Conservation Biology* **16**: 650-659.

McOrist S and Kitchener AC (1994) Current threats to the European wildcat, *Felis silvestris*, in Scotland. *Ambio* **23**: 243–245.

McVean G (2007) The structure of linkage disequilibrium around a selective sweep. *Genetics* **175**: 1395–1406.

Melton T and Holland C (2007) Routine Forensic Use of the Mitochondrial 12S Ribosomal RNA Gene for Species Identification. *Journal of Forensic Sciences*, **52**: 1305–1307.

Menotti-Raymond M, David VA, Schäffer AA, Stephens R, Wells D, Kumar-Singh R, O'Brien SJ, Narfström K (2007) Mutation in CEP290 discovered for cat model of human retinal degeneration. *Journal of Heredity* **98**: 211-20.

Miller C, Joyce P, Waits LP (2002) Assessing allelic dropout and genotype reliability using maximum likelihood. *Genetics* **160**: 357–366.

Mitchell-Jones AJ, Amori G, Bogdanowicz W, Krystufek B, Reijnders PJH, Spitzenberger F, Stubbe M, Thissen JBM, Vohralik V, Zima J (1999) *The Atlas of European mammals*. 484 pp. Poyser Natural History, London, United Kingdom.

Monterroso P, Brito JC, Ferreras P, Alves PC (2009) Spatial ecology of the European wildcat in a Mediterranean ecosystem: dealing with small radio–tracking datasets in species conservation. *Journal of Zoology* **279**: 27–35.

Mooney HA and Cleland EE (2001) The evolutionary impact of invasive species. *Proceedings of the National Academy of Sciences USA* **10**: 5446–5451.

Morin PA, Chambers KE, Boesch C, Vigilant L (2001) Quantitative polymerase chain reaction analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (*Pan troglodytes verus*). *Molecular Ecology* **10**: 1835–1844.

Morin PA, Luikart G, Wayne RK, SNP-Workshop Group (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution* **19**: 208–216.

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

Mullikin JC, Hansen NF, Shen L, Ebling H, Donahue WF, Tao W, Saranga DJ, Brand A, Rubenfield MJ, Young AC, Cruz P; NISC Comparative Sequencing Program, Driscoll C, David V, Al-Murrani SW, Locniskar MF, Abrahamsen MS, O'Brien SJ, Smith DR, Brockman JA (2010) Light whole genome sequence for SNP discovery across domestic cat breeds. *BMC Genomics* **11**: 406.

Navidi W, Arnheim N, Waterman MS (1992) A multiple-tubes approach for accurate genotyping of very small DNA samples by using PCR: statistical considerations. *American Journal of Human Genetics* **50**: 347–359.

Narum S, Banks M, Beacham T, Bellinger M, Campbell M, Dekoning J, Elz A, Guthrie III C, Kozfkay C, Miller K, Moran P, Phillips R, Seeb l, Smith C, Warheit K, Young S, Garza J (2008) Differentiating salmon populations at broad and fine geographical scales with microsatellites and single nucleotide polymorphisms. *Molecular Ecology* **17**: 3464–3477.

Nowell K and Jackson P (1996) European wildcat, *Felis silvestris, silvestris* group Schreber, 1775. In *Wild cats: status survey and conservation action plan*: 110–113. Nowell, K. and Jackson, P. (Eds). Gland, Switzerland: IUCN.

O'Brien SJ, Johnson W, Driscoll C, Pontius J, Pecon-Slattery J, Menotti-Raymond M (2008) State of cat genomics. *Trends in Genetics* **24**: 268–279.

O'Brien SJ, Devillard S, Say L, Vanthomme H, Léger F, Ruette S, Pontier D (2009) Preserving genetic integrity in a hybridising world: are European Wildcats (*Felis silvestris silvestris*) in Eastern Francedistinct from sympatric feral domestic cats? *Biodiversity and Conservation* **18**: 2351-2360.

Palomo LJ and Gisbert J (2002) *Atlas de los mamíferos terrestres de España*. Dirección General de Conservación de la Naturaleza. SECEM-SECEMU, Madrid, Spain.

Palomo LJ, Gisbert J, Blanco JC (2007) *Atlas y libro rojo de los mamíferos terrestres deE spaña.* Dirección general para la biodiversidad-SECEM-SECEMU, Madrid. 564 pp.

Peichocki R (2001) Lebensräume. Die Verbreitung der Wildkatze in Europa. In: H. Grabe and G. Worel (eds), *Die Wildkatze. Zurück auf leisen Pfoten*, pp. 14-27. Buch and Kunstverlag Oberpfalz, Amberg.

Paschou P, Ziv E, Burchard EG, Choudhry S, Rodriguez-Cintron W, Mahoney MW, Drineas P (2007) PCA-correlated SNPs for structure identification in worldwide human populations. *PLoS Genetics* **3**: 1672–1686.

Pertoldi C, Bijlsma R, Loeschcke V (2007) Conservation genetics in a globally changing environment: present problems, paradoxes and future challenges. *Biodiversity and Conservation* **16**: 4147–4163.

Pertoldi C, Wójcik JM, Matgorzata T, Kawalko A, Kristensen TN, Loeschcke V, Gregersen VR, Coltman D, Wilson GA, Randi E *et al.* (2010) Genome variability in European and American bison detected using the BovineSNP50 BeadChip. *Conservation Genetics* **11**: 627–634.

Peterschmitt M, Grain F, Arnaud B, Deléage G, Lambert V (2009) Mutation in the melanocortin 1 receptor is associated with amber colour in the Norwegian forest cat. *Animal Genetics* **40**: 547-552.

Pierpaoli M, Biro ZS, Herrmann M, Hupe K, FernandesM, Ragni B, Szemethy L, Randi E (2003) Genetic distinction of wildcat (*Felis silvestris*) populations in Europe, and hybridization with domestic cats in Hungary. *Molecular Ecology* **12**: 2585–2598.

Piggott MP, Banks SC, Stone N, Banffy C, Taylor AC (2006) Estimating population size of endangered brush-tailed rockwallaby (*Petrogale Penicillata*) colonies using faecal DNA. *Molecular Ecology* **15**: 81–91.

Piry S, Alapetite A, Cornuet J-M, Paetkau D, Baudouin L, Estoup A (2004) GeneClass2: A Software for Genetic Assignment and First-Generation Migrant Detection. *Journal of Heredity* **95**:536-539.

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

Pocock RI (1951) *Catalogue of the genus Felis*. London: British Museum.

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

Pollinger JP, Bustamante CD, Fledel-Alon A, Schmutz S, Gray MM, Wayne RK (2005) Selective sweep mapping of genes with large phenotypic effects. *Genome Resources* **15**: 1809-1819.

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.

Pontier D, Fouchet D, Bahi-Jaber N, Poulet H, Guiserix M, Natoli E, Sauvage F (2009) When domestic cat (*Felis silvestris catus*) population structures interact with their viruses. *Comptes rendus biologies* **332**: 321-328.

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

Primmer CR (2009) From Conservation Genetics to Conservation Genomics. *Annals of the New York Academy of Sciences* **1162**: 357–368.

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

Przeworski M (2002) The signature of positive selection at randomly chosen loci. *Genetics* **160**: 1179–1189.

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

Raimer F (2006) *Die Wildkatzenpopulationen in Hessen und Niedersachsen seit dem 18*. Jahrhundert. In: Kleine K, Grosse R (ed), Naturschutz-Akademie Hessen; Bund für Umwelt- und Naturschutz Deutschland; Institut für Tierökologie und Naturbildung. NAH Akedemie-Berichte 5, NZH Verlag, Wetzlar, pp 39–44.

Ragni B (1981) *Gatto selvatico, Felis silvestris Schreber, 1777. Distribuzione e biologia di 22 specie di mammiferi in Italia*. Consiglio Nazionale delle Ricerche, Rome.

Ragni B (1993) *The crucial problem of in vivo identification of wildcat and recognition of hybrids with domestic cats*. Peruglia, Italy: Institute of Zoology, Peruglia University.

Ragni B and Possenti M (1996) Variability of coat-colour and markings system in *Felis silvestris*. *Italian Journal of Zoology* **63**: 285–292.

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

Randi E, Lucchini V, Christensen MF, Mucci N, Funk SM, Dolf G, Loeschcke V (2000) Mitochondrial DNA variability in Italian and east European wolves: detecting the consequences of small population size and hybridization. *Conservation Biology* **14**: 464–473.

Randi E and Lucchini V (2002) Detecting rare introgression of domestic dog genes into wild wolf (*Canis lupus*) populations by Bayesian admixture analysis of microsatellite variation. *Conservation Genetics* **3**: 31–45.

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

Randi E (2008) Detecting hybridization between wild species and their domesticated relatives. *Molecular Ecology*, **17**: 285-293.

Reed JZ, Tollit D, Thompson P, Amos W (1997) Molecular scatology: the use of molecular genetic analysis to assign species, sex, and individual identity to seal faeces. *Molecular Ecology* **6**: 225–234.

Reusch TBH and Wood TE (2007) Molecular ecology of global change. *Molecular Ecology* **16**: 3973–3992.

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

Roon DA, Waits LP, Kendall KC (2003) A quantitative evaluation of two methods for preserving hair samples. *Molecular Ecology Notes* **3**: 163–166.

Rubin CJ, Zody MC, Eriksson J, Meadows JRS, Sherwood E, Webster MT, Jiang L, Ingman M, Sharpe T, Ka S, Hallbook F, Besnier F, Carlborg O, Bed'hom B, Tixier-Boichard M, Jensen P, Siegel P, Lindblad-Toh K, Andersson L (2010) Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* **464**: 587-591.

Rudnick JA, Katzner TE, Bragin EA, Rhodes EO Jr, Dewoody JA (2005) Using naturally shed feathers for individual identification, genetic parentage analyses, and population monitoring in an endangered Eastern imperial eagle (*Aquila heliaca*) population from Kazakhstan. *Molecular Ecology* **14**: 2959–2967.

Ruiz-González A, Rubines A, Berdi°on O, Gómez-Moliner BJ (2008) A noninvasive genetic method to identify the sympatric mustelids pine marten (*Martes martes*) and stone marten (*Martes foina*): preliminary distribution survey on the northern Iberian Peninsula. *European Journal of Wildlife Research* **54**: 253–261.

Ryman N anf Palm S (2006) POWSIM: a computer program for assessing statistical power when testing for genetic differentiation. *Molecular Ecology Notes* **6**: 600–602.

Sacks BN and Louie S (2008) Using the dog genome to find single nucleotide polymorphisms in red foxes and other distantly related members of the Canidae. *Molecular Ecology Resources* **8**: 35-49.

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

Sarmento P (1996) Feeding ecology of the European wildcat *Felis silvestris* in Portugal. *Acta Theriologica* **41**:409–414.

Sarmento P, Cruz J, Eira C, Fonseca C (2009) Spatial colonization by feral domestic cats *Felis catus* of former wildcat *Felis silvestris silvestris* home ranges. *Acta Theriologica* **54**: 31–38.

Sastre N, Francino O, Lampreave G, Bologov VV, López-Martín JM, Sánchez A, Ramírez O (2009) Sex identification of wolf (*Canis lupus*) using non-invasive samples. *Conservation Genetics* **10**: 555–558.

Savolainen P, Zhang Y, Luo J, Lundeberg J, Leitner T (2002) Genetic Evidence for an East Asian Origin of Domestic Dogs. *Science* **298**: 1610-1613.

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

Schauenberg P (1977) Intestine length of *Felis silvestris* Schreber. *Mammalia* **41**: 356–360.

Schlötterer, C (2004) The evolution of molecular markers – just a matter of fashion? *Nature Review Genetics* **5**: 63–69.

Schmidt-Küntzel A, Eizirik E, O'Brien SJ, Menotti-Raymond M (2005) Tyrosinase and tyrosinase related protein 1 alleles specify domestic cat coat color phenotypes of the albino and brown loci. *Journal of Heredity*, **96**: 289-301.

Schmidt-Küntzel A, Nelson G, David VA, Schäffer AA, Eizirik E, Roelke ME, Kehler JS, Hannah SS, O'Brien SJ, Menotti-Raymond M (2009) A domestic cat X chromosome linkage map and the sex-linked *orange* locus-mapping of *orange*, multiple origins, and epistasis over *non-agouti*. *Genetics* **181**: 1415-1425.

Schopen GCB, Bovenhuis H, Visker MHPW, Van Arendonk JAM (2008) Comparison of information content for microsatellites and SNPs in poultry and cattle. *Animal Genetics* **39**: 451–453.

Schulenburg J (2005) Säugetiere (Mammalia). In: Günther A., Nigmann U. and Achtziger R. (eds.), Analyse der Gefährdungsursachen planungsrelevanter Tiergruppen in Deutschland zur Ergänzung der bestehenden Roten Listen gefährdeter Tiere. *Naturschutz und Biologische Vielfalt 21, Bonn-Bad Godesberg:* 70–124.

Schwartz MK, Pilgrim KL, Pilgrim KL, McKelvey KS, Lindquist EL, Claar JJ, Loch S, Ruggiero LF (2004) Hybridization between Canada lynx and bobcats: genetic results and management implications. *Conservation Genetics* **5**: 349–355.

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

Seehausen O (2004) Hybridization and adaptative radiation. *Trends in Ecology Evolution* **19**: 198–207.

Slate J, Gratten J, Beraldi D, Stapley J, Hale M, Pemberton JM (2009) Gene mapping in the wild with SNPs: guidelines and future directions. *Genetica* **136**: 97–107.

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

Sommer RS and Benecke N (2006) Late Pleistocene and Holocene development of the felid fauna (Felidae) of Europe: a review. *Journal of Zoology* **269**: 7-19.

Spielman D, Brook BW, Briscoe DA, Frankham R (2004a) Does inbreeding and loss of genetic diversity decrease genetic resistance? *Conservation Genetics* **5**: 439–448.

Spielman D, Brook BW, Frankham R (2004b) Most species are not driven to extinction before genetic factors impact them. *Proceedings of the National Academy of Sciences USA* **101**:15261–15264.

Stahl P and Artois M (1991) Status and conservation of the wildcat (Felis silvestris) in Europe and around the Mediterranean rim. Convention on the Conservation of European Wildlife and Natural habitats Standing Committee. *Council of Europe Nature and Environment Series* **69**: 1-76.

Suminski P (1962) Characteristics of the pure form of the wildcat *Felis silvestris* Schreber. *Archaeological Science* **15**: 277–296.

Sundqvist A-K, Bjornerfeldt S, Leonard JA, Hailer F, Hedhammar A, Ellegreen H, Vilà C (2006) Unequal contribution of sexes in the origin of dog breeds. *Genetics* **172:** 1121-1128.

Sundqvist A-K, Ellegren H, Vilà C (2008) Wolf or dog? Genetic identification of predators from saliva collected around bite wounds on prey. *Conservation Genetics* **9**: 1275-1279.

Sunquist M, Sunquist F (2002) *Wild Cats of the World*. The University of Chicago Press, Chicago.

Taberlet P, Bouvet J (1992) Bear conservation genetics. *Nature* **358**: 197.

Taberlet P, Griffin S, Goossens B *et al.* (1996) Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research* **24**: 3189–3194.

Taberlet P, Waits LP, Luikart G (1999) Noninvasive genetic sampling: look before you leap. *Trends in Ecology and Evolution* **14**: 323–327.

Trut L (1999) Early Canid Domestication: the Farm-Fox Experiment. *American Scientist* **87**: 160-169.

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

Valière N (2002) Gimlet: a computer program for analysing genetic individual identification data. *Molecular Ecology Notes* **2**: 377-379.

Vigne J-D (1992) Zooarchaeology and the biogeographical history of the mammals of Corsica and Sardinia since the last ice age. *Mammal Review* **22**: 87–96.

Vigne JD, Briois F, Zazzo A, Willcox G, Cucchi T, Thiébault S, Carrère I, Franel Y, Touquet R, Martin C, Moreau C, Comby C, Guilaine J (2012) First wave of cultivators spread to Cyprus at least 10,600 y ago. *Proceedings of the National Academy of Sciences of the United States of America* **109**(22): 8445-8449.

Vilà C and Wayne R (1999) Hybridization between wolves and dogs. *Conservation Biology* **13**: 195-198.

Vilà C, Walker C, Sundqvist A, Flagstad Ø, Andersone Z, Casulli A *et al.* (2003) Combined use of maternal, paternal and bi-parental genetic markers for the identification of wolf-dog hybrids. *Heredity* **90**: 17-24.

vonHoldt, Bridgett M, John P Pollinger, Kirk E Lohmueller, Eunjung Han, Heidi G Parker, Pascale Quignon, Jeremiah D Degenhardt, *et al.* (2010) Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature* **464**: 898-902.

Wasser SK, Mailand C, Booth R, Mutayoba B, Kisamo E, Clark B, Stephens M (2007) Using DNA to track the origin of the largest ivory seizure since the 1989 trade ban. *Proceedings of the National Academy of Sciences USA* **104**: 4228–4233.

Wayne RK and Morin PA (2004) Conservation genetics in the new molecular era. *Frontiers in Ecology and the Environment* **2**: 89–97.

Wayne RK and Brown DM (2001) Hybridization and conservation of carnivores. In: *Carnivore Conservation* (eds Gittleman JL, Funk S, Macdonald DW, Wayne RK). Cambridge University Press, Cambridge.

Williams CL, Blejwas K, Johnston JJ, Jaeger MM (2003) A coyote in sheep¢s clothing: predator identification from saliva. *Wildlife Society Bulletin* **31**: 926–932.

Wolsan M and Okarma H (2001) *Felis silvestris* Schreber, 1775. In: Glowacinski (ed.), *Polish red data book of animals*, pp. 94-95, 83-84. Warszawa.

Yamaguchi N, Driscoll CA, Kitchener AC, Ward JM, Macdonald DW (2004a) Craniological differentiation between European wildcats (*Felis silvestris silvestris*), African wildcats (*F. s. libyca*) and Asian wildcats (*F. s. ornata*): implications for their evolution and conservation. *Biological Journal of the Linnean Society* **83**: 47-63.

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

Yamasaki M, Tenaillon MI, Bi IV, Schroeder SG, Sanchez-Villeda H, Doebley JF, Gaut BS, McMullen MD (2005) A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement. *Plant Cell* **17**: 2859-2872.

Zeder MA (2006) Central questions in the domestication of plants and animals. *Evolutionary Anthropology* **15**:105-117.

Zeder MA (2008) Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proc Natl Acad Sci USA* **105**:11597–11604.

CHAPTER 2

European wildcats in Iberian Peninsula

"If this is not done, future ages will certainly look back upon us as a people so immersed in the pursuit of wealth as to be blind to higher considerations."

Alfred Russel Wallace

PAPER I. Oliveira R, Godinho R, Randi E, Ferrand N and Alves PC (2008)

MOLECULAR ANALYSIS OF HYBRIDIZATION BETWEEN WILD AND DOMESTIC CATS (*Felis silvestris*) IN PORTUGAL: IMPLICATIONS FOR CONSERVATION.

Conservation Genetics 9: 1-11.

PAPER II. Oliveira R, Godinho R, Randi E and Alves PC (2008)

HYBRIDIZATION VS CONSERVATION: ARE DOMESTIC CATS THREATENING THE GENETIC INTEGRITY OF EUROPEAN WILDCAT (*Felis silvestris silvestris*) POPULATIONS IN IBERIAN PENINSULA?

Philosophical Transactions of the Royal Society of London B: Biological Sciences 363: 2953-2961.

Molecular analysis of hybridisation between wild and domestic cats (*Felis silvestris*) in Portugal: implications for conservation

Rita Oliveira[#†], Raquel Godinho[#], Ettore Randi[§], Nuno Ferrand[#†] and Paulo Célio Alves[#†]

#CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Campus Agrário de Vairão, 4485-661 Vairão, Portugal
†Departamento de Zoologia e Antropologia, Faculdade de Ciências da Universidade do Porto, 4099-002 Porto, Portugal
§ISPRA, Istituto Superiore per la Protezione e la Ricerca Ambientale, Via Cà Fornacetta 9, 40064 Ozzano dell'Emilia, Bologna, Italia

ABSTRACT

The endangered European Wildcat (*Felis silvestris silvestris*) is represented, today, by fragmented and declining populations whose genetic integrity is considered to be seriously threatened by crossbreeding with widespread free-ranging domestic cats. Extensive and recent hybridisation has been described in Hungary and Scotland, in contrast with rare introgression of domestic alleles in Italy and Germany. In Portugal, the wildcat is now listed as VULNERABLE in the Red Book of Portuguese Vertebrates. Nevertheless, genetic diversity of populations and the eventual interbreeding with domestic cats remain poorly studied. We surveyed genetic variation at 12 autosomal microsatellites for 34 wild and 64 domestic cats collected across Portugal. Wild and domestic cats were significantly differentiated both at allele frequencies and sizes (F_{ST}=0.11, R_{ST}=0.18, P<0.001). Population structure and admixture analyses performed using Bayesian approaches also showed evidence of two discrete groups clustering wild and domestic populations. Results did not show significant genetic divergence among Northern, Central and Southern wildcats. Six morphologically identified wildcats were significantly assigned to the domestic cluster, revealing some discrepancy between phenotypic and genetic identifications. We detected four hybrids (approximately 14%) using a consensus analysis of different Bayesian model-based software. These hybrids were identified throughout all sampled areas, suggesting that hybridisation is of major concern for the appropriate implementation of wildcat conservation strategies in Portugal.

Keywords: wildcat, domestic cat, hybridisation, microsatellites, admixture analysis, Bayesian clustering, conservation genetics.

INTRODUCTION

Although globally distributed across Europe and South-western Asia, the European wildcat (*Felis silvestris silvestris*) is currently represented by fragmented and declining populations. Even though legally protected by important Directives (as Habitat Directive, Bern Convention and CITES) in most European countries, wildcat populations are considerably threatened mainly due to the concomitant habitat destruction and fragmentation, poison and road kills, proliferation of viral diseases and hybridisation with its domestic counterpart (Stahl and Artois, 1994; Nowel and Jackson, 1996; Beaumont *et al*, 2001; Randi *et al*, 2001).

Crossbreeding with widespread free-ranging domestic cats is one of the main threats for wildcat survival underlined by the European Council (Stahl and Artois, 1994). Therefore, it became imperative to study differentiation between wild, domestic cats and cryptic hybrids and to evaluate the rate and impact of hybridisation. The problematic definition of morphological criteria allowing unambiguous distinction between the three forms, along with the particularly challenging identification of hybrids beyond first generation (Daniels *et al*, 1998; Allendorf *et al*, 2001), prompted the initiation of genetic studies into diagnostic molecular traits. A number of European wildcat studies have used microsatellites with much more accurate results than former works using mitochondrial DNA (Hubbard, 1992) and allozymes (Randi and Ragni, 1991), especially when combining highly polymorphic markers and recently developed Bayesian clustering models. Among European populations, results suggest variable rates of domestic genes introgression, with wide and recently hybridising populations in Hungary and Scotland (Beaumont *et al*, 2001; Daniels *et al*, 2001; Pierpaoli *et al*, 2003; Lecis *et al*, 2006) contrasting with a low admixture scenario in Italy and Germany (Randi *et al*, 2001; Pierpaoli *et al*, 2003; Eckert and Hartl, 2005; Lecis *et al*, 2006). Although reasons for the observed variability remain unidentified, the anthropogenic-mediated dispersion of domestic cats throughout wildcat distribution and the unknown effects of long-term simpatry raised a global concern regarding both genetic and taxonomic status of the European wildcat (McOrist and Kitchener, 1996; Daniels *et al*, 1998).

With the exception of littoral areas, wildcats were formerly widespread in Portugal (Nowell and Jackson, 1996). However, its present distribution appears to be considerably smaller. Similarly to other European populations, massive habitat loss and landscapes fragmentation, progressive and invasive urbanization, and scarce availability of prey (as a result of the severe decrease of wild rabbit, the main natural prey in Mediterranean landscapes, Gil-Sánchez *et al*, 1999; Lozano *et al*, 2003) may have led to population decline and, eventually, promoted reproductive interactions with domestic cats. A few ecological studies were implemented in Portuguese protected areas (Sarmento, 1996; Fernandes, 1996; Monterroso *et al*, 2005; Ferreira *et al*, 2005), documenting an evident versatility in food ecology and habitat selection. Nevertheless, ecological, ethological and, particularly, genetic features of the wildcat population are still poorly explored. A first molecular approach was performed by Fernandes (1996); however, the analysis of a small number of samples and loci prevented obtaining consistent results. More recently, Pierpaoli *et al* (2003), in a broad European study, identified one individual with

hybrid ancestry among 13 Portuguese wildcats. Nevertheless, frequency, extension and impact of domestic genes introgression remain unknown.

In this study, we present the first integrated approach combining the use of highly polymorphic loci and Bayesian statistical approaches to i) investigate the extend of genetic variation and differentiation in Portuguese wild and domestic cat populations; ii) pinpoint hybridisation and evaluate introgression of domestic alleles, iii) provide new insights and critical guidelines to the regional and global conservation of this threatened feline. This work represents a first-step to clarify central population-level questions for wildcat management and long-term protection in Portugal, producing reference molecular data for future studies on historical and recent patterns of genetic diversity and for monitoring populations' demography, gene flow and genetic structure.

MATERIALS AND METHODS

Sampling and DNA extraction

We analysed a total of 98 tissue, blood and swab samples comprising 34 wild and 64 domestic cats (of which 16 are purebred and 48 are mutt/feral individuals). Wildcat samples were provided by BTVS-ICN (Wild Animal Tissue Bank, Portuguese Conservation Institute), and were distributed across the North (4), the Centre (4) and the South (26) of Portugal (Fig. 1). The low population density of wildcats in Portugal associated with their elusive behaviour difficult obtaining larger sample sizes from this feline. Wildcats were taxonomically identified by collectors according to their coat-colour pattern (Ragni and Possenti, 1996), biometrics (Schauenberg, 1977) and geographical location, independently from any genetic information. In order to survey potentially divergent domestic cat gene pools and obtain a representative sampling of the domestic subspecies, we collected samples from Northwest and South-east of Portugal (Fig. 1). We directed sampling effort to areas where human settlements are known to overlap with wildcat distribution. We extracted total genomic DNA using salting-out and phenol-chloroform procedures, both adapted from Sambrook *et al*, 1989.

Figure 1. Geographical location and number of sampled individuals (W = wildcats, D = domestic cats).

Microsatellites typing and data analysis

I. Individual genotyping

We assessed individual multilocus genotypes using 12 neutral unlinked microsatellites, formerly isolated and characterized in domestic cat (Menotti-Raymond and O'Brien, 1995; Menotti-Raymond *et al*, 1999). Specific choice of this battery is justified by its prior successful and informative use in recent wildcat studies (Beaumont *et al*, 2001; Daniels *et al*, 2001, Randi *et al*, 2001; Pierpaoli *et al*, 2003; Eckert and Hartl, 2005, Lecis *et al*, 2006). Polymerase chain reaction (PCR) amplifications of individual microsatellites followed Randi *et al* (2001). PCR products were separated by electrophoresis on a 6% denaturing polyacrylamide gel and visualized by silver staining.

II. Analysis of Genetic diversity

Allele frequencies, standard diversity indices and observed (H_O) and expected (H_E) heterozigosities for each locus and population were calculated using GENETIX 4.05 (Belkhir *et al*, 1996-2004). We estimated allelic richness (*AR*) using FSTAT 2.9.3.2 (Goudet, 2001). Guo and Thompson's (1992) Markov chain method (*MCMC*) was implemented in GENEPOP 3.4 (Raymond and Rousset, 1995) to evaluate significant deviations from Hardy-Weinberg Equilibrium (*HWE*) for all locus-population combinations and statistically infer pairwise Linkage Equilibra (*LE*) among loci. We adjusted significance levels using sequential Bonferroni correction for multiple comparisons in the same data set (Rice, 1989). GENEPOP 3.4 and FSTAT 2.9.3.2 were used to compute single and multilocus *F* (Weir and Cockerham, 1984) and *R* statistics (Slatkin, 1995), accounting for variation in population sizes. We estimated the genetic relationship between wild and domestic populations through a hierarchical Analysis of Molecular Variance (*AMOVA*; Excoffier *et al*, 1992), implemented in ARLEQUIN 3.01 (Excoffier *et al*, 2005) using Φ_{ST} and R_{ST}. We used the same analysis to estimate the genetic differentiation between geographically separated cats sampled across Portugal, within both wild and domestic populations. In order to increase the number of individuals per wildcat geographical group, we assembled cats from North and Centre and compared with the ones from South. The significance of genetic differentiation was tested by random permutation, under the null hypothesis that all individuals belong to a single global population. Using FSTAT 2.9.3.2, we computed Wilcoxon signed rank test to evaluate differences in allelic diversity (*AD*), allelic richness (*AR*) and H_E between pairs of geographical groups.

III. Population structure and admixture analyses using multilocus genotype data

Population structure, individual assignments and admixture proportions were estimated through different Bayesian-based statistical techniques using: i) the clustering procedure described by Pritchard *et al* (2000) and recently updated in STRUCTURE 2.1 (Falush *et al*, 2003); ii) the method developed by Anderson and Thompson (2002) and performed in NEWHYBRIDS and iii) a model-based software described by Wilson and Rannala (2003) and carried out in BAYESASS 1.2. Both

STRUCTURE and NEWHYBRIDS were implemented providing prior non-genetic classification for all known domestic cats, since we had the confident reference that all domestic individuals were true domestic without any recent ancestry in the wild population. The use of this type of information frequently results in Bayesian inference improvement and is strongly supported by STRUCTURE's authors in cases of unequivocal preclassification (Pritchard and Wen, 2003). We included or not that information for the wildcats and the putative hybrids detected without non-genetic information. NEWHYBRIDS was used to achieve a more detailed analysis of admixture proportions and hybrids ancestry, by inferring the posterior probability assignment (Q) of each sampled individual to six genotype frequency classes: Pure I; Pure II; F1; F2; Backcross I and Backcross II. We also used BAYESASS to estimate recent migration rates between wild and domestic populations. In this software results are presented as the number of times each individual is assigned to each population and were transformed, in this study, into probabilistic values.

We assessed the power of admixture analysis to detect parentals and F1, F2 and backcross hybrids by simulation of parental and hybrid genotypes in the program HYBRIDLAB (Nielsen *et al* 2001), as recently described by Barilani *et al* (2006). Briefly, in our original dataset, we selected a subset of 20 wild and 40 domestic cats that revealed, in STRUCTURE, an individual proportion of membership (qi) >0.90 to their parental cluster, in order to exclude possible hybrids. Starting from this sampling, we simulated 100 genotypes of each parental and hybrid classes, procedure that was repeated 10 times. The simulated genotypes were then used in STRUCTURE with K=2 and no prior population information, in order to evaluate the efficiency of admixture analysis to study our population and define the appropriate threshold value that should be used for the individual assignment to one single population cluster or hybrid class. Following simulations data, we defined a threshold of 0.80 (see Results) for all methods and each genotype was assigned to each group based on its qi. In the case of STRUCTURE, we also evaluated the 90% credibility intervals (CI) of individual's qi. According to each model features and their previous use in population structure analyses, we computed all programs using the profiles described in Table 1.

Table 1. Programs profiles defined to analyse population structure using three Bayesian clustering methods.

Profile	Program		
	STRUCTURE	**NEWHYBRIDS**	**BAYESASS**[c]
MCMC iterations	10^5	10^5	$3.0x10^6$
Burn-in period	10^4	10^4	10^6
Inference of K (populations)	MAXPOP =1-5		
Others	Independent runs = 5 Model = *Admixture model*[a]	*Uniform priors*[b]	Sampling frequency = 2000

a allows individuals to have mixed ancestry and was performed using two model options: correlated (F model) and independent (I model) allele frequencies between populations.

b uniform priors consider that at least one copy of each allele has been found in both populations. This approach reduces the influence of low frequency alleles, preventing sampling and genotyping errors in closely related populations.

c convergence of MCMC algorithm was firstly confirmed using different initial values of migration and inbreeding levels (0.10 and 0.15 for both parameters).

RESULTS

A first exploratory Bayesian analysis revealed that six morphologically preclassified wildcats, named Fs2, Fs6, Fs9, Fs10, Fs21 and Fs23, were significantly assigned to the domestic cluster according to their multilocus genotypes (e.g. q_1>0.94 in STRUCTURE; P ≤0.80). Based on these results and on documented errors for the unequivocal phenotypic distinction between European wildcats, tabby domestic cats and their hybrids (Ragni, 1993), wrong morphological identification was considered the most plausible explanation for this incongruence. In fact, Fs2 was identified as a "strange" colour pattern wildcat by the collector and Fs21 and Fs23 were found particularly damaged in the field, preventing a complete analysis of morphological traits or hiding some phenotypic signs of domestication. Consequently, these six individuals were excluded from the analysis and the new sampling profile became constituted of 28 wildcats, 21 from the South, four from the Centre and three from the North of Portugal.

Analysis of Genetic diversity

All loci were polymorphic in both wild and domestic cats, showing from seven (Fca077) to 16 (Fca026) alleles per locus. Although differential frequency distribution of alleles was the most significant parameter of distinction between both populations, we found a total of 12 private alleles, six in wild and six in domestic cats (in order to prevent sampling and/or genotyping errors we only considered alleles with frequency higher than 5%; Table 2). A significant deficit of heterozygotes was detected in domestic cats (F_{IS}=0.09; P<0.05). None of the combinations between pairs of loci disclosed a significant deviation from Linkage Equilibra (*LE*) (P<0.0041, Bonferroni-corrected for 12 independent replications). A significant departure from *HWE* was observed in only two over 24 possible locus-population combinations, Fca126 in wildcats (P=0.0004; F_{IS}=0.21) and Fca088 in domestic cats (P=0.0000; F_{IS}=0.37; Table 3).

Table 2. Allelic frequencies at 12 polymorphic microsatellites among Portuguese wild and domestic cat populations.

Locus	N	Population	Allelic frequencies															
		Allele size (bp)	123	125	131	135	137	139	141	143	145	147	149					
Fca008	63	domestic	0.01	0.15	0.01	0.02	0.22	0.07	0.11	0.15	0.23	0.01	0.02					
	28	wild	0.00	0.02	0.00	0.00	0.02	0.11	0.29	0.46	0.07	0.03	0.00					
		Allele size (bp)	132	134	136	138	140	142	144	146	148	150	152	154				
Fca023	63	domestic	0.01	0.13	0.00	0.02	0.59	0.05	0.05*	0.01	0.02	0.07*	0.03	0.02				
	28	wild	0.42	0.27	0.07*	0.00	0.12	0.03	0.00	0.00	0.02	0.00	0.02	0.05				
		Allele size (bp)	130	132	134	138	140	142	144	146	148	150	152	154	156	158	160	162
Fca026	64	domestic	0.05	0.00	0.00	0.01	0.03	0.00	0.02	0.16	0.09	0.27	0.18	0.07	0.05	0.02	0.04	0.01
	28	wild	0.04	0.02	0.02	0.00	0.00	0.07*	0.12	0.16	0.18	0.07	0.05	0.14	0.13	0.00	0.00	0.00
		Allele size (bp)	118	120	122	124	126	128	130	132	134	138	142	148	150			
Fca043	64	domestic	0.02	0.02	0.32	0.05	0.06*	0.37	0.08	0.00	0.02	0.01	0.01	0.03	0.01			

	28	wild	0.07	0.09	0.63	0.05	0.00	0.05	0.04	0.05*	0.00	0.02	0.00	0.00	0.00	
	Allele size (bp)		147	151	152	153	154	155	156	157	158	159	160	161	162	164
Fca045	64	domestic	0.04	0.02	0.01	0.21	0.08	0.24	0.00	0.01	0.04	0.28	0.01	0.06	0.00	0.01
	28	wild	0.03	0.00	0.05	0.07	0.04	0.04	0.05*	0.02	0.23	0.07	0.27	0.09	0.04	0.00
	Allele size (bp)		211	219	221	223	225	227	229	231	233	235				
Fca058	64	domestic	0.00	0.02	0.01	0.23	0.08	0.04	0.06	0.51	0.02	0.03				
	28	wild	0.02	0.04	0.05	0.07	0.05	0.11	0.41	0.25	0.00	0.00				
	Allele size (bp)		143	145	147	149	151	153	155							
Fca077	61	domestic	0.01	0.39	0.05	0.22	0.16	0.15	0.02							
	28	wild	0.00	0.20	0.14	0.13	0.21	0.32	0.00							
	Allele size (bp)		111	113	115	117	119	121	123	125	127	129				
Fca088	62	domestic	0.01	0.04	0.22	0.08	0.17	0.06	0.16	0.24*	0.00	0.02				
	28	wild	0.00	0.00	0.02	0.20	0.32	0.18	0.26	0.00	0.02	0.00				
	Allele size (bp)		185	209	211	213	215	217	219	221	223	225	227	229	231	233	237
Fca096	61	domestic	0.02	0.03	0.67	0.03	0.01	0.01	0.10	0.02	0.08	0.00	0.00	0.02	0.00	0.01	0.00
	28	wild	0.03	0.02	0.14	0.05	0.07	0.04	0.05	0.20	0.05	0.04	0.05	0.04	0.20*	0.00	0.02
	Allele size (bp)		137	139	141	143	145	147	149	151	153	155	161			
Fca126	64	domestic	0.00	0.03	0.10	0.38	0.09	0.20*	0.04	0.12	0.02	0.01	0.01			
	27	wild	0.35*	0.07	0.17	0.20	0.15	0.00	0.00	0.04	0.02	0.00	0.00			
	Allele size (bp)		138	140	142	144	146	148	150	152	154	156	158	160		
Fca132	60	domestic	0.24	0.17	0.03	0.01	0.04	0.01	0.10	0.24	0.03	0.11	0.01	0.01		
	28	wild	0.04	0.02	0.09	0.11	0.02	0.00	0.12	0.28	0.23	0.09	0.00	0.00		
	Allele size (bp)		122	124	128	130	132	134	138							
Fca149	61	domestic	0.00	0.23	0.33	0.17	0.17	0.10*	0.00							
	28	wild	0.02	0.02	0.09	0.32	0.42	0.00	0.13*							

* Private alleles (P≥0.05)

Table 3. Summary of diversity indices for each locus-population combination: allelic diversity (*AD*), allelic richness (*AR*), observed (H_O) and expected heterozygosities (H_E) and inbreeding coefficient (F_{IS}). F_{ST} (coefficient of genetic differentiation) and R_{ST} (F_{ST} analogue accounting for allelic size variation) estimations between wild and domestic populations are also presented for each locus. SE= standard error.

	DOMESTIC CATS					WILDCATS						
LOCUS	*AD*	*AR*	H_O	H_E	F_{IS}	*AD*	*AR*	H*o*	H_E	F_{IS}	F_{ST}	R_{ST}
Fca008	11	8.88	0.78	0.83	0.08	7	6.93	0.61	0.68	0.13	0,11	0,12
Fca023	11	9.10	0.54	0.62	0.14	8	7.93	0.61	0.73	0.19	0,22	0,30
Fca026	13	11.10	0.82	0.85	0.03	11	10.93	0.86	0.88	0.04	0,04	0,01
Fca043	12	9.30	0.73	0.74	0.01	8	7.97	0.68	0.59	-0.14	0,12	0,20
Fca045	12	9.22	0.78	0.80	0.04	12	11.97	0.75	0.85	0.13	0,11	0,13
Fca058	9	7.60	0.61	0.67	0.10	8	7.97	0.71	0.75	0.06	0,13	0,00
Fca077	7	6.11	0.77	0.75	-0.02	5	5.00	0.86	0.78	-0.09	0,04	0,06
Fca088	9	8.19	0.52	0.83	0.38*	6	5.93	0.71	0.75	0.07	0,08	-0,02
Fca096	11	8.28	0.42	0.53	0.20	14	13.93	0.96	0.88	-0.08	0,21	0,42
Fca126	10	8.49	0.66	0.78	0.17	7	7.00	0.63	0.78	0.21*	0,11	0,39
Fca132	12	9.58	0.80	0.83	0.04	9	8.93	0.96	0.82	-0.16	0,06	0,13
Fca149	5	5.00	0.77	0.77	-0.01	6	5.93	0.89	0.69	-0.28	0,12	0,25
Average	10.08	8.40	0.69	0.75	0.09	8.41	8.37	0.77	0.76	0.01	0,11	0,18
(SE)			(0.13)	(0.10)				(0.13)	(0.09)			

*significant departures from *HWE* (P<0.0041; Bonferroni-corrected for 12 independent comparisons)

An average F_{ST}=0.11 over all loci revealed a significant genetic differentiation between wild and domestic Portuguese populations (P <0.001; Table 3). Multilocus R_{ST} was also highly significant (R_{ST}=0.18; P <0.001; Table 3). These results reflect distinct gene pools for both groups, differing simultaneously in allele frequencies and sizes, and suggest that new mutations are also contributing to the allelic diversity found in both populations.

The hierarchical *AMOVA* among different geographical groups revealed a non significant differentiation between localities (North+Centre versus South), with 96.33% of genetic diversity explained by interindividual differences within groups (Φ_{ST}=0.04; P ≤0.05; Table 4). Partition of microsatellites variability between Northern and Southern domestic cats also disclosed a non-significant value (Φ_{ST}=0.02; P ≤0.05; Table 4). According to R_{ST} statistic, allelic size variation is also not significantly partitioned among wild and domestic cat groups (R_{ST}=-0.03 and R_{ST}=0.02, respectively; P ≤0.05). Moreover, Wilcoxon signed rank tests corroborated these results, showing no significant differences in H_E, allelic richness (*AR*) and allelic diversity (*AD*) between pairs of geographical sites. These results encouraged the analysis of Portuguese wild and domestic cats as two global clusters, each one comprising all geographically separated individuals from each subspecies.

Table 4. Hierarchical Analysis of Molecular Variance (AMOVA) for wild (North+Centre and South) and domestic cat (North and Centre) geographical groups computed in ARLEQUIN, using Φ_{ST}.

	Source of variation	Variance	% of variation	Φ_{ST}
WILDCAT	Among groups	0.169	3.67	0.037 (P ≤0.05)*
	Within groups	4.445	96.33	
	Total	4.614		
DOMESTIC CAT	Among groups	0.077	1.91	0.019 (P ≤0.05)*
	Within groups	3.946	98.09	
	Total	4.023		

*P=significance level, after 15 000 permutations

Bayesian inference of population structure and admixture analyses

We first used STRUCTURE to identify the best performing model for admixture analysis and, as a result, we defined allele frequencies correlated model (*F* model) as the one that better explains the observed population structure, providing the most accurate assignment of all unequivocally preclassified domestic cats. This model also provided a better assignment in other European studies (Pierpaoli *et al*, 2003: Lecis *et al*, 2006) and is frequently more efficient to detect genetic structure in closely related groups (Pritchard and Wen, 2003). Using *F* model without any prior non-genetic information, we inferred the most probable number of genetic clusters (*K*) presented in the sample by estimating *Ln* posterior probabilities of the data and choosing the smallest value of *K* that captures the major structure in the data set (Pritchard *et al*, 2000). Maximum increase in *Ln* posterior probabilities was observed for *K*=2 and the conversion of likelihood values obtained for all inferred *K* (1-6) into probabilities following Pritchard and Wen (2003) revealed an approximately 100% probability of

having two distinct clusters in the dataset against almost 0.00% for all other values of K. These results suggest that pooled individuals might be subdivided in two genetically discrete populations. With K=2 and using only genetic information, we estimated the average membership proportions (Q) of each predefined group (wild and domestic cats) into both clusters inferred by the program. Results showed a clear partition of both predefined populations, by the separation of two distinct genetic clusters grouping domestic (Cluster I; Q_I= 0.89) and wild (Cluster II; Q_{II}= 0.97) individuals. Nevertheless, a q value of 0.11 in Cluster II coming from the domestic population predicted domestic genes introgression into wildcat population.

The admixture analysis performed on simulated genotypes was able to efficiently recognise 100% of the parental individuals at a threshold of qi=0.80 (the minimum q_I value was 0.802) and all the F1 hybrids were correctly identified as admixed cats. However, twelve F2 (12%) and 20 backcross (20%) genotypes showed a qi>0.80 to one single cluster and could not be distinguished from parental individuals. All hybrids detected by simulations revealed very wide 90% CI, ranging between 0.18 and 0.80.

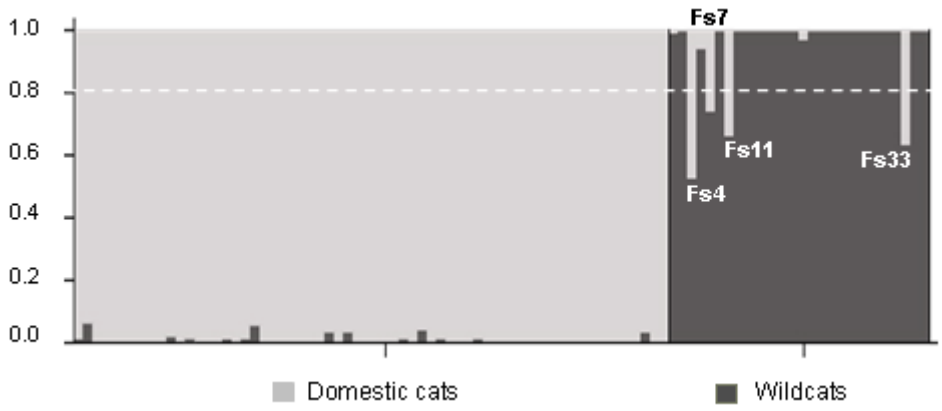


Figure 2. Posterior probability assignments of Portuguese wild and domestic cats, using prior non-genetic information for all domestic individuals. Each cat is represented by a vertical bar fragmented in K sections of specific length, according to their membership proportion in both genetic clusters inferred by STRUCTURE. Fs4, Fs7, Fs11 and Fs33 revealed a significant ancestry in the domestic cluster and are most likely admixed (horizontal white line = Q threshold).

Using STRUCTURE with a cutpoint of 80% and without information on wildcats' prior identification, all domestic cats were grouped in Population I (average membership coefficient of individuals (q_I) = 0.99) and Population II grouped approximately 86% of the preclassified wildcats (24/28) with average q_{II}=0.98 and 90% CI between 0.81 and 1.00. Putative wildcats Fs4, Fs7, Fs11 and Fs33 were genetically identified as hybrids, demonstrating cumulative individual q_{II}<0.80 distributed between the two sampled groups (Fig. 2), and with 90% CI in cluster I and II ranging between 0.21 and 0.79 (Table 5). In a second performance of the model using prior morphological identification for all sampled wildcats (USEPOPINFO=1) and including or not that information for

the putative hybrids formerly identified (POPFLAG=0 or 1), all posterior probabilities were congruent: the four admixed cats also revealed a wild assignment and 90% CI <0.80, disclosing significant values of domestic ancestry. The ancestral class of individuals whose genotypes indicate a hybrid ancestry can be assessed, either in current or first and second past generations. However, none of these individuals presented posterior probabilities >0.80 for only one of the past hybrid generations. Even though Fs4 presented a considerably superior probability of being an F1 hybrid, his membership proportion was lower than the 80% threshold considered in this study (q_{F1}=0.72).

Table 5. Individual proportions membership (q) of the 4 putative hybrids Fs4, Fs7, Fs11 and Fs33 using prior non-genetic information for all domestic individuals, both in STRUCTURE and NEWHYBRIDS. In STRUCTURE, individuals were assigned into two clusters corresponding to the domestic and wild groups and, in NEWHYBRIDS, into different genotype classes: pure domestic cat, pure wildcat, F1, F2, Bx I (backcross with domestic cat) and Bx II (backcross with wildcat). STRUCTURE 90% credibility intervals (CI) are shown in brackets.

CAT	STRUCTURE		NEWHYBRIDS					
	Domestic	Wild	Domestic	Wild	F1	F2	Bx I	Bx II
Fs4	0.526 (0.383-0.765)	0.474 (0.235-0.617)	0.099	0.033	0.678	0.083	0.04	0.067
Fs7	0.303 (0.213-0.591)	0.697 (0.409-0.787)	0.288	0.073	0.224	0.197	0.078	0.140
Fs11	0.392 (0.307-0.603)	0.608 (0.397-0.693)	0.632	0.039	0.070	0.097	0.014	0.149
Fs33	0.366 (0.244-0.729)	0.634 (0.271-0.756)	0.223	0.209	0.165	0.200	0.083	0.119

Using prior individual non-genetic classification for all domestic cats in NEWHYBRIDS, we obtained a sharp distinction between individual membership proportions of domestic and wild individuals. All domestic cats were probabilistically assigned to the same genotype frequency class, Pure I (average Q=0.98), while Pure II class grouped approximately 82% of the wildcats (23/28), with an average posterior probability of 0.95. Five phenotypically identified wildcats – Fs4, Fs5, Fs7, Fs11 and Fs33 – were only partially assigned to the wild population (individual Q<0.80) and revealed posterior probabilities clearly distributed among different hybrid frequency classes. Similar results were achieved for all but Fs5 when prior non-genetic information was included for all domestic and wildcats or when excluding that information for the five putative hybrids. A detailed analysis of individual membership revealed that none of the genetically admixed cats was assigned to a single hybrid class with Q>0.80, hindering the clear definition of their admixed ancestry (Fig. 3).

The estimation of recent migration rates in BAYEASS revealed a potential introgression of domestic alleles in wildcat population (m=0.064; SD=0.027) corresponding to a migration proportion of 4.1 ± 1.73 individuals per generation. A negligible migration of 0.14 ± 0.14 wildcats was detected into the domestic population (m=0.005; SD=0.005). According to the probability distributions of individual migrant ancestries in three possible states - non-migrant, migrant or offspring of a migrant and a non-migrant - all domestic cats were correctly assigned to the domestic cluster with posterior

probabilities higher than 99%. Among the 28 analysed wildcats, 82% (23/28) were significantly allocated to the wildcat cluster (P>0.86). In agreement with inferences made with Pritchard *et al*'s (2000) and Anderson and Thompson's (2002) approaches, putative wildcats Fs4, Fs7, Fs11 and Fs33 were assigned to both wild and domestic clusters, disclosing a significant posterior probability of being a second generation migrant (0.91, 0.95, 0.87, 0.91, respectively). As observed in NEWHYBRIDS estimates without incorporating prior population information, wildcat Fs5 revealed a significant posterior probability of having hybrid ancestry, however, a consensus analysis of all procedures did not allow this individual to be identified as an admixed cat, considering that five out of seven methodological options performed in this study resulted in its significant wild assignment (P>0.87). A consensus evaluation of all Bayesian analyses consistently identified four individuals with hybrid ancestry among the 28 putative wildcats analysed (Table 5). The admixed cats where collected throughout the sampling area: one in North (Fs33), one in Centre (Fs7) and two in South of Portugal (Fs4 and Fs11).

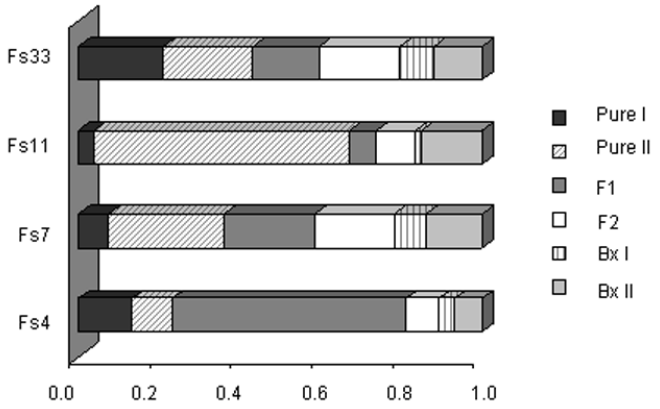


Figure 3. Posterior probabilities of genotype frequency classes performed in NEWHYBRIDS for the four admixed cats (Fs4, Fs7, Fs11 and Fs33), inferred without including their prior phenotypic information. Each cat corresponds to a horizontal bar divided in six segments representing the probability of each individual into the different genotype classes: Pure I (pure domestic cat), Pure II (pure wildcat), F1, F2, Bx I (backcross with domestic cat) and Bx II (backcross with wildcat).

DISCUSSION

Phenotypic versus genetic identifications

Six morphologically preclassified wildcats were identified as domestic according to genetic data, showing some discrepancy between phenotypic and molecular identifications. We know that morphometric identification is often difficult when characters are close to indicative thresholds, especially under the uncertain definition of diagnostic traits and the possibility of resemblance between wild, domestic and hybrid cats due to natural variation (Daniels *et al*, 1998). Furthermore, some of the analysed samples were found extremely deteriorated in the field disabling a detailed identification and their sympatric location to wildcat populations certainly complicated their

classification. Individuals considered to be erroneously pre-identified were collected in protected areas where wildcats inhabit, which confirm an effective overlap between wild and domestic cats in these areas. Considering it is expected that wildlife protection actively occurs in natural parks, control of free-ranging domestic cats should be questioned in these regions. Incongruence between phenotypic and genetic classifications supports the idea that genetic identifications are essential tools in conservation issues, especially in cases where morphological identifications are dubious. Accordingly, one important feature of this work is the construction of reference genetic compositions for Portuguese wild and domestic cats, which, based on genetic clustering comparisons, will allow the future allocation of unidentified samples of this endangered species.

Genetic diversity

Differential allele frequencies, private alleles and significant F_{ST} and R_{ST} values reveal a clear genetic distinction between Portuguese wild and domestic cats (Table 3). Similarly, high genetic variability disclosed by allelic diversity (*AD*), allelic richness (*AR*) and H_E (Table 3) is in concordance with published data in genetically viable wildcat populations, such as Italian, German and Slovenian ones (Randi *et al*, 2001, Pierpaoli *et al*, 2003, Lecis *et al*, 2006). In contrast, genetic diversity observed in this study clearly opposes results obtained in highly admixed populations from Hungary and Scotland (Beaumont *et al*, 2001; Daniels *et al*, 2001; Pierpaoli *et al*, 2003 and Lecis *et al*, 2006). Accordingly, we may infer that Portuguese population of European wildcat maintains its genetic identity, despite some recent introgression of domestic genes.

Analyses of Molecular Variance performed among wild and domestic cat geographical groups suggest the absence of genetic substructure in both subspecies (table 4), which coincide with the low genetic differentiation observed in domestic populations across Europe (F_{ST}≈3% and R_{ST}≈1%; Beaumont *et al* 2001; Pierpaoli *et al*, 2003; Eckert and Hartl, 2005). This genetic continuity is certainly related to the anthropogenic character of domestic reproduction, which hinders the definition of isolated and panmictic populations. On the other hand, low genetic divergence between geographically separated wildcats opposes documented values for other European populations, such as German ones, where Western and Eastern populations disclosed a F_{ST}=0.19 (Eckert and Hartl, 2005). Although our results should be taken with caution due to the low number of samples from Northern and Central Portugal, they indicate that widely separated Portuguese wildcats might have maintained gene flow in the past. However, the increasing habitat fragmentation and the destruction of important ecological corridors might lead to a considerable geographic isolation and differentiation in the future.

Population structure of wild and domestic populations and admixture analysis

Sample partition obtained using STRUCTURE has an obvious biological sense, since it corresponds to the split of wild and domestic cats in two discrete genetic clusters. Among the 28 putative wildcats analysed in this study, we identified four genetically admixed individuals through a consensus

evaluation of all model-based Bayesian approaches and specific methodological options. However, hybrids ancestry remained undisclosed since the global analysis of all clustering methods did not statistically define a single hybrid class assignment for any of the admixed cats. Analysis of the simulation results revealed that the 12 microsatellites used in this study are able to detect 100% of parentals and F1 hybrids using a threshold of 80%, while only 88% of F2 and 80% of backcrosses were detected. These simulated hybrids revealed wide ranges of 90% CI, which are known to occur in admixed genotypes (Pritchard *et al*, 2000; Barilani *et al*, 2006). In our population, while most of the 90% CI ranged between 0.80 and 1.00 in wildcats, the four putative hybrids showed values ranging from 0.21 to 0.79, as expected. These findings suggest that our analyses are reliable in the identification of the four admixed cats, but might represent an underestimation of the true number of existing hybrids, since a few F2 and backcross genotypes can remain undetected. The cutpoint of 80% selected in our study is in agreement with previous works focusing on wild and domestic cat hybridisation (Pierpaoli *et al*, 2003; Lecis *et al*, 2006). At the same time, high performance has been attributed to this q-value for the accurate detection of purebred and hybrid groups in both STRUCTURE and NEWHYBRIDS, when using 12 loci to study populations with $F_{ST} \approx 0.12$ (for details see Barilani *et al*, 2006 and Vähä and Primer, 2006). The uncertainties in the detection of past generations admixture and in the definition of hybrid classes highlight the inherent difficulty to deal with closely related (sub)species and might be explained by the need of a strong genetic differentiation and an increased number and type of loci for the clear allocation of wild x domestic cat hybrids to a single genotypic class (Wilson and Rannala, 2003; Lecis *et al*, 2006). In fact, at least 48 unlinked loci might be needed to detect hybrids beyond first generation, even in cases of clearly divergent parental populations (Vähä and Primer, 2006). Even though improving admixture analysis with linked loci did not significantly improve its power in population studies of Italian and Hungarian cats (Lecis *et al*, 2006), genotyping a large number of unlinked and linked microsatellites, combined with novel molecular markers, may enable better statistical estimates of hybridisation further back in the past (Falush *et al*, 2003; Lecis *et al*, 2006). Although we should carefully interpret our results, Fs4 might be an F1 hybrid, even though its association to this class was not statistically supported by the threshold used in this study. Fs7, Fs11 and Fs33 may have a more ancient ancestry in the domestic population. Wild and domestic cat populations revealed asymmetrical migration rates, suggesting only a possible introgression of domestic alleles into the wildcat population (m= 0.064) and not a bidirectional gene flow.

Implications for conservation and management

The endangered European wildcat has a central importance in Portuguese wildlife protection, since it might be the only resident wild feline after the probable extinction of reproductive populations of Iberian-lynx (Pires and Fernandes, 2003). Accordingly, results of this molecular study should be used as guidelines by Portuguese conservation authorities, in order to effectively preserve and monitor the long-term genetic integrity of wildcat populations. Even though we found no genetic evidence for a constant and generalized gene flow between sympatric populations of wild and domestic cats, at least in most recent generations, admixture analysis revealed a significant proportion of hybrids (around 14%), distributed in all regions analysed, and migration rates documented an effective negative impact

on wildcats' genetic composition caused by hybridisation. The extensive geographical distribution of admixed cats reveals domestic introgression clearly not restricted to a particular area, while alerts for a possible decrease in differentiation between Portuguese wild and domestic cats. Accordingly, we suggest that regional and global management strategies should recognize the prevention of crossbreeding between European wildcat and domestic cat as high conservation priority (Randi *et al*, 2001; Wolf *et al*, 2001). To avoid the risk of genetic admixture, outbreeding depression, reduced fitness and lowered genetic variability three main actions should be promoted, including: *i*) public campaigns to inform authorities and local human populations on Portuguese wildcat status and threats; *ii*) the legal control of domestic cats by capturing and neutering free-ranging animals; *iii*) and the effective protection of large suitable habitats, mainly preventing the creation of environmental obstacles for wildcat dispersal (Stahl and Artois, 1991).

ACKNOWLEDGEMENTS

We thank Armando Loureiro from BTVS-ICN (Wild Animal Tissue Bank, Portuguese Conservation Institute), Pedro Monterroso and Dra Maria Antonieta Ferreira for providing samples. We also thank to all veterinarians and biologists that assisted in sample collection. We are grateful to Massimo Pierpaoli for helpful comments on data analysis. Rita Oliveira is supported by Fundação para a Ciência e a Tecnologia (FCT) through a PhD grant SFRH/BD/24361/2005.

REFERENCES

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

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

Barilani M, Sfougaris A, Giannakopoulos A, Mucci N, Tabarroni C and Randi E (2006) Detecting introgressive hybridisation in rock partridge populations (Alectoris graeca) in Greece through Bayesian admixture analyses of multilocus genotypes. *Conservation Genetics*.

Beaumont M, Barratt EM, Gottelli D, Ktchener AC, Daniels MJ, Pritchard JK, Bruford MW (2001) Genetic diversity and introgression in the Scottish wildcat. *Molecular Ecology* 10: 319-336.

Belkhir K, Borsout P, Goudet J, Chikhi L, Bonhomme (1996-2004) Genetix, logiciel sous Windows™ pour la génétique des populations. Laboratoire Génome et Population, CNRS UPR 9060, Université de Montpellier II, Montpellier, France.

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* 244: 231-247.

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

Eckert and Hartl (2005) Conservation genetics of the European wildcat (*Felis silvestris silvestris*) in Germany. Poster presentation in Symposium of Biology and Conservation of the European Wildcat (*Felis silvestris silvestris*). Germany January 21st-23rd 2005.

Excoffier L, Simouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitocondrial DNA restriction data. *Genetics* 131: 479-491.

Excoffier, L. G. Laval, S. Schneider (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1: 47-50.

Falush D, Stephens M, Pritchard J (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164: 1567-1587.

Fernandes M (1996) Estudo genético do gato-bravo europeu: abordagem ao problema da hibridação. Dissertação de mestrado em Conservação da Diversidade Animal. Faculdade de Ciências da Universidade de Lisboa, Portugal.

Ferreira PJ, Pereira I, Marques JT, Grilo C, Baltazar C, Santos-Reis M (2005) Modelling wild cat (*Felis silvestris*) distribution in a Portuguese Natura 2000 Site (SPA Moura Barrancos, SE Portugal) Poster presentation in Symposium of Biology and Conservation of the European Wildcat (*Felis silvestris silvestris*). Germany January 21st-23rd 2005.

Gil-Sánchez JM, Valenzuela G, Sánchez JF (1999) Iberian wild cat *Felis silvestris* tartessia predation on rabbit Oryctolagus cuniculus: functional response and age selection. *Acta Theriologica* 44: 421-428.

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

Guo SW and Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportions for multiple alleles. *Biometrics* 48: 361-372.

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

Lecis R, Pierpaoli M, Birò S, Szemethy L, Ragni B, Vercillo F, Randi E (2006) Bayesian analyses of admixture in wild and domestic cats (*Felis silvestris*) using linked microsatellite loci. *Molecular Ecology* 15: 119-131.

Lozano J, Virgos E, Malo A, Huertas DL, Casanovas JG (2003) Importance of scrub-pastureland mosaics on Wildliving cats occurrence in a Mediterranean area: implications for the conservation of the wildcat (*Felis silvestris*). *Biodiversity and Conservation* 12 (5): 921-935.

McOrist S and Kitchener AC (1996) Current threats to the European wildcat in Scotland. *Biological Conservation* 76 (2): 212.

Mennoti-Raymond MA and O´Brien S (1995) Evolutionary conservation of ten microssatellites loci in four species of FELIDAE. *Journal of Heredity* 86: 319-322.

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

Monterroso P, Sarmento P, Ferreras P, Alves PC (2005) Spacial distribution of the European wildcat (*Felis silvestris*) in Vale do Guadiana Natural Park, South Portugal. Poster presentation in Symposium of Biology and Conservation of the European Wildcat (*Felis silvestris silvestris*). Germany January 21st-23rd 2005.

Nielsen EE, Bach LA, Kotlicki P (2006) Hybridlab (version 1.0): a program for generating simulated hybrids from population samples. *Molecular Ecology Notes* 6 (4): 971–973.

Nowell K and Jackson P (1996) The wild cats: status survey and conservation action plan. International Union for Nature Conservation/Cat Specialist Group, Gland, Switzerland.

Pierpaoli M, Birò ZS, Herrmann M, Hupe K, Fernandes M, Ragni B, Szemethy L, Randi E (2003) Genetic distinction of wildcat (*Felis silvestris*) populations in Europe, and hybridization with domestic cats in Hungary. *Molecular Ecology* 12: 2585-2598.

Pires AE and Fernandes ML (2003) Last lynxes in Portugal? Molecular approaches in a pre-extinction scenario. *Conservation Genetics* 4: 525–532.

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

Pritchard J and Wen W (2003) Documentation for structure software: Version 2. Department of Human genetics, University of Chicago, USA.

Ragni B (1993) The crucial problem of in vivo identification of wildcat and recognition of hybrids with domestic cats. Institute of Zoology, Perugia University, Italy.

Ragni B and Possenti M (1996) Variability of coat-colour and markings system in *Felis silvestris*. Italian *Journal of Zoology* 63: 285–292.

Randi E and Ragni B (1991) Genetic variability and biochemical systematics of domestic and wild cat populations (*Felis silvestris*: FELIDAE). *Journal of Mammology* 72 (1): 79-88.

Randi E, Pierpaoli M, Beaumont M, Ragni B (2001) Genetic identification of wild and domestic cats (*Felis silvestris*) and their hybrids using Bayesian Clustering Methods. Molecular Biology and *Evolution* 18 (9): 1679-1693.

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

Rice WR (1989) Analyzing tables of statistical tests. *Evolution* 43: 223-225.

Sambrook E, Fritsch F, Maniatis T (1989) Molecular Cloning. Cold Spring Harbour Press, Cold Spring Harbour, New York.

Sarmento P (1996) Feeding ecology of the European wildcat *Felis silvestris* in Portugal. *Acta Theriologica* 41: 409-414.

Stahl P and Artois M (1994) Status and conservation of the wildcat (*Felis silvestris*) in Europe and around the Mediterranean rim. Convention on the Conservation of European Wildlife and Natural Habitats Standing Committee. *Council of Europe Nature and Environment Series* 69: 1–76.

Schauenberg P (1977) La stature du Chat forestiere *Felis silvestris* Schreber, 1777, et la variabilité morphologique de l'espèce. *Revue Suisse de Zoologie* 84: 323-337.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetic*s 139: 457-462.

Vähä JPK and Primmer CR (2006) Detecting hybridization between individuals of closely related populations – a simulation study to assess the efficiency of model-based Bayesian methods to detect hybrid individuals. *Molecular Ecology* 15: 63–72.

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

Wilson GA and Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* 163: 1177-1191.

Yamaguchi N, Driscoll C, Kitchener A, Ward JM, MacDonald DW (2004) Craniological differentiation between European wildcats (*Felis silvestris silvestris*), African wildcats (*F. s. libyca*) and Asian wildcats (*F. s. ornata*): implication for their evolution and conservation. *Biological Journal of the Linnaean Society* 83: 47-63.

Hybridization *vs* conservation: are domestic cats threatening the genetic integrity of European wildcat (*Felis silvestris silvestris*) populations in Iberian Peninsula?

Rita Oliveira[#†], Raquel Godinho[#], Ettore Randi[§] and Paulo Célio Alves[#†]

#CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Campus Agrário de Vairão, 4485-661 Vairão, Portugal
†Departamento de Zoologia e Antropologia, Faculdade de Ciências da Universidade do Porto, 4099-002 Porto, Portugal
§ISPRA, Istituto Superiore per la Protezione e la Ricerca Ambientale, Via Cà Fornacetta 9, 40064 Ozzano dell'Emilia, Bologna, Italia

ABSTRACT

Crossbreeding between wild and free-ranging domestic species is one of the main conservation problems for some threatened species. The documented situation of the wildcat in some European populations is a good example of this critical phenomenon. Extensive hybridisation was described in Hungary and Scotland, in contrast with occasional interbreeding in Italy and Germany. First analyses in Portugal revealed a clear genetic differentiation between wild and domestic cats; however, four cryptic hybrids were detected. Here, we extended the approach to Iberian Peninsula by performing multivariate and Bayesian analyses of multilocus genotypes for 44 Portuguese wildcats, 31 Spanish wildcats and 109 domestic cats. Globally, wild and domestic cats were significantly differentiated (F_{ST}=0.20, P<0.001) and clustered in two discrete groups. Diverse clustering methods and assignment criteria identified an additional hybrid in the Centre of Portugal, performing a total of 5 admixed individuals. Power of admixture analyses was assessed by simulating hybrid genotypes, which revealed that the used markers were able to detect 100% of first generation hybrids, 91% of F2 genotypes and 85% of backcrosses. These findings suggest that true proportion of admixture can be higher and that the ongoing improvement of genetic tools for hybrids detection is crucial for wildcat conservation.

Keywords: *Felis silvestris*, hybridisation, microsatellites, Bayesian admixture analysis, conservation genetics

INTRODUCTION

The anthropogenic-mediated dispersion of free-ranging domestic cats (*Felis silvestris catus*) and their contact with natural populations of European wildcats (*Felis silvestris silvestris*) is considered one of the main threats for the survival of the endangered wildcat populations throughout all Europe. The unknown effects of long-term sympatry between both wild and domestic cats resulted in a global concern regarding the genetic and taxonomic status of the European wildcat (McOrist and Kitchener, 1994; Daniels *et al*. 1998; Beaumont *et al*. 2001). The main problems that lead to artificial hybridisation are related to habitat fragmentation and home range changes; scarce availability of prey; and the increasing structuring of small and isolated natural populations (Rhymer and Simberloff, 1996; Allendorf *et al*. 2001). These factors may have been promoting a more frequent and large-scale contact between wild and domestic cats and a continuous backcrossing of hybrid individuals to parental populations may eventually culminate in a deep and irreversible genetic pollution of wild populations. For example, the swamping of domestic alleles into the wildcat genome over successive generations resulted in a high admixture level in Hungarian and Scottish populations (Daniels *et al*. 2001; Beaumont *et al*. 2001; Pierpaoli *et al*. 2003; Lecis *et al*. 2006). Interestingly, only occasional interbreeding was found in Italy, Germany and Portugal, with wildcat populations clearly differentiated from domestic cats and still preserving their genetic singularity (Randi *et al*. 2001; Pierpaoli *et al*. 2003; Eckert and Hartl, 2005; Lecis *et al*. 2006; Oliveira *et al*. 2007).

Preventing introgression in wild populations depends on the efficient detection of admixed individuals. Several genetic approaches have been extensively and successfully used to address the identification of hybrids in different taxa (Rhymer and Simberloff, 1996; Allendorf *et al*. 2001), especially in cases where phenotypical classifications of hybrid classes or even parental groups are dubious, as occur between wild, domestic cats and their hybrids (Daniels *et al*. 1998; Beaumont *et al*. 2001). The ability to genetically distinguish admixed individuals within sympatric populations of closely related (sub)species can provide invaluable resources for wildlife management and protection, and has proven to be essential in studies of population structure and admixture in wildcat populations (Daniels *et al*. 1998; Oliveira *et al*. 2007). Similarly to most organisms, intraspecific distinction among *Felis silvestris* is barely based on diagnostic genetic differences and, consequently, the identification of parental and hybrid individuals is based on probabilistic assessments (Nielsen *et al*. 2006). The most promising mechanism to study artificial hybridisation is the combination of highly informative molecular markers with model-based Bayesian software, mainly because Bayesian admixture analyses are powerful to assess levels of population differentiation, even when reference parental groups cannot be sampled and, at the same time, results are not influenced by the proportion of hybrids in the sample (Pritchard *et al*. 2000; Anderson and Thompson 2002; Corander and Marttinen, 2006; Vähä and Primmer, 2006). In the case of wildcat hybridisation studies, it is possible to sample domestic cats of "pure" origin but not to genotype baseline samples of "pure" wildcats (Beaumont *et al*. 2001). On the other hand, the high variability of admixture rates documented so far reflects the need to use advanced methods not sensitive to those variables.

In Iberian Peninsula, the European wildcat is considered VULNERABLE (VU) in Portugal and NEAR THREAT (NT) in Spain. These classifications are based on the low density and fragmented status of the populations, and also on the consequent high risk of extinction through hybridisation with the copious and pervasive domestic form (Cabral *et al*. 2005). The conservation status of the species reflects the importance and urgency to understand the structure and dynamics of the Iberian populations; nevertheless, many ecological and genetic features are still unknown. Admixture analyses performed in our first study of Portuguese wildcats revealed that hybridisation is not frequent and widespread at least in most recent generations. However, four cryptic hybrids were identified in different geographic areas and an evident sympatry between wildcats and its domestic counterpart was detected (Oliveira *et al*. 2007).

Starting from reference molecular data in Oliveira *et al*. (2007), we extent the admixture analyses to other areas in Iberian Peninsula by improving both sample size and distribution. In this study, we present a first descriptive step to aid the regional and global conservation of this small feline in Iberia by investigating, for the first time, the differentiation between Iberian wild and domestic cats; and further evaluating the degree and extend of introgressive hybridisation across different areas in the Peninsula. Additionally, we infer the power and limits of the Bayesian admixture analyses to identify admixed genotypes in our dataset, in order to design new pathways of research. Thus, we discuss the utility of this study as a model to continue the development of DNA-based tools to detect and monitor hybridisation.

METHODS

Sampling and individual multilocus genotyping

We analysed a total of 184 tissue, blood and swab samples from domestic and putative wildcats that were collected in Iberian Peninsula in the period of 1993-2006. Sampling comprised 44 wildcats from North (4), Centre (9) and South (31) of Portugal; 31 wildcats from Granada (22), Asturias (3) and Basque Country (6); and 109 feral and purebred domestic cats distributed across Iberian Peninsula (Figure 1). These new sample set corresponds to an increase of approximately 50% relatively to the one analysed in Oliveira *et al*. 2007. Putative wildcat samples were opportunistically collected or were obtained from collaborative ecological studies. Since wildcats live in low densities and fragmented populations in Iberia, and considering that are elusive and rare animals, obtaining a larger sample size from this feline is a difficult task and implies long and extensive and persistent efforts. These samples were mainly identified by collectors according to their wild coat phenotype (Ragni and Possenti, 1996), biometrics and geographic location. All wildcats were morphologically identified independently from any genetic information. Four of them were indicated as possible hybrids by the collectors: FSI711, FSI719, FSI725 and FSI878. Domestic cat samples were obtained from cat pounds, private owners and road killed animals, including individuals that are sympatric with wildcats (collected in small isolated rural villages).

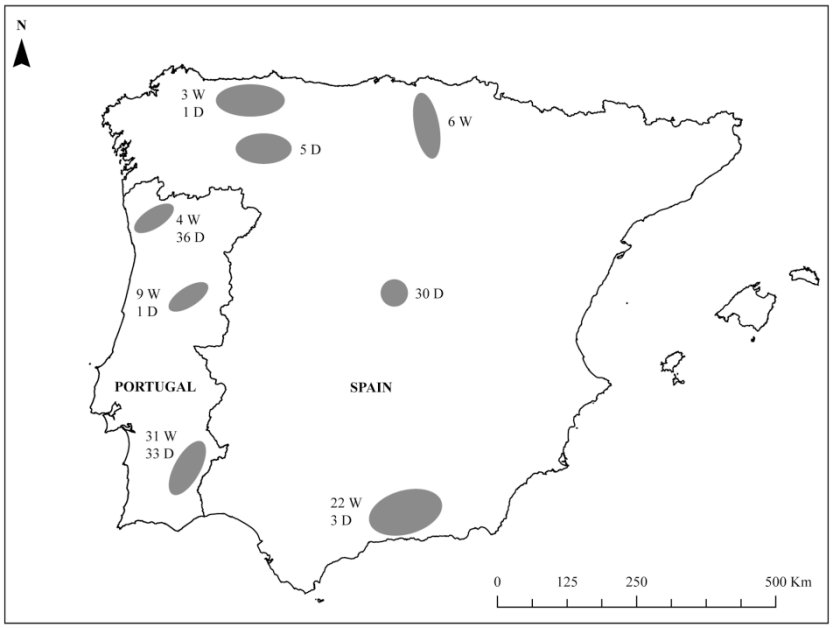


Figure 1. Sampling locations of the studied cats in Iberian Peninsula (w = wildcats, d = domestic cats).

Total genomic DNA was extracted using salting-out procedure (Sambrook *et al*., 1989) and guanidine thiocyanate (Gerloff *et al*. 1995). A battery of 12 autosomal unlinked microsatellites, formerly isolated and characterized in the domestic cat (Menotti-Raymond and O'Brien 2005; Menotti-Raymond *et al*. 1999), was used to obtain individual multilocus genotypes. PCR amplifications for each locus were performed following Randi *et al*. (2001). Fragments were separated by size on an ABI 3130*xl* sequencer and genotypes were analysed using GENEMAPPER 4.0 (Applied Biosystems).

Analysis of genetic partition and multivariate clustering

Allele frequencies, allelic richness (*AR*), standard diversity indices and expected (*HE*) heterozigosities for each locus were calculated using GenAlEx 6b4 (Peakall and Smouse, 2006). Guo andThompson (1992) exact test was implemented in GENEPOP 3.4 (Raymond and Rousset, 1995) in order to statistically evaluate deviations from Hardy-Weinberg Equilibrium for all locus-population combinations and infer pairwise Linkage Disequilibrium for all loci. Significance levels were adjusted using the sequential method of Bonferroni for multiple comparisons in the same dataset (Rice, 1989). GenAlEx 6b4 and GENEPOP 3.4 were used to compute F_{ST} (Weir and Cockerham, 1984) between wild and domestic cat populations. Partition of microsatellite diversity between and within wild and domestic populations was estimated through an Analysis of Molecular Variance (AMOVA) on Euclidean pairwise genetic distances, using Φ analogues of Wright's F-statistics. Significance testing was done by random permutation. Principal Component Analysis (PCA) applied to individual multilocus genotypes, was also computed in GenAlEx 6b4 in order to describe genetic variation among populations. Wildcats were grouped in two different ways for substructure analysis: i) wildcats from Portugal (Fsi PT) vs. wildcats from Spain (Fsi SP), and wildcats from Northern Iberia (North and

Centre of Portugal, Asturias, León and Basque Country) vs wildcats from Southern Iberia (South of Portugal and Granada).

Population structure and Bayesian admixture analyses

Bayesian-based analyses of population structure and admixture proportions were performed using STRUCTURE version 2.1. (Pritchard *et al*. 2000; Falush *et al*. 2003). All analyses were computed using the admixture model and assuming that allele frequencies are correlated among populations. Using software settings previously described in Oliveira *et al*. 2007, the number of discrete genetic clusters (K) present in the total sample was estimated with K=1-8, by analysing the probability of the data using the formula $\mathrm{LnP(D)}_K$-$\mathrm{LnP(D)}_{k\text{-}1}$ (Garnier *et al*. 2004) and by converting likelihood values obtained for all inferred K into probabilities (Pritchard and Wen, 2003). For the selected K values, we estimated the membership proportion (Q) of the sampled populations into the detected clusters, and the individual membership proportion q (the proportion of each individual genome that has ancestry in those clusters) was used as a metric of cats sorting into each genetic group. Following simulation results (see below) and previous studies (Oliveira *et al*. 2007; Lecis *et al*. 2006; Pierpaoli *et al*. 2003), we used an inclusive threshold of q and its 90% confidence interval (CI) >0.80 to assign each individual genotype to one single cluster. Admixed genotypes were detected when an individual proportion of membership was partitioned and lower than 80% to each genetic group (for details on computation and models interpretation see Oliveira *et al*. 2007).

The inherent drawback of the Bayesian approach is that validity of the assumed priors and efficiency of analysed loci cannot be statistically assessed; consequently, simulations have to be implemented for each empirical dataset in order to evaluate the statistical limit of that particular study (Nielsen *et al*. 2006). We assessed the power of the markers and models used in the admixture analyses to distinguish among parental and hybrid classes, and we established the range of q values expected for all possible admixture generations by simulating both parental and hybrid genotypes in HYBRIDLAB 1.0 (Nielsen *et al*. 2006). Based on individual multilocus genotypes, the program initially estimates locus by locus allele frequencies for each of the parental wild and domestic populations. Then multilocus F1 hybrid genotypes are created by randomly selecting one allele from each of the two populations, according to their frequency distribution (Nielsen *et al*. 2006). Simulations of other hybrid classes (F2 and backcrosses genotypes) can be computed by the successive use of simulated genotypes as starting point populations. Briefly, we selected 40 parental domestic and 40 parental wildcats to generate 100 genotypes of each hybrid class: F1, F2 and backcrosses. Parental genotypes were selected among individuals that revealed qi + 90% CI >0.90 in STRUCTURE, in order to exclude possible undetected hybrids. With K=2, simulated genotypes were then used in STRUCTURE without any prior non-genetic information, aiming to assess the efficiency of the admixture analyses in estimating the proportion of hybrids in the simulated dataset (see Barilani *et al*. 2006 for further details).

RESULTS

In a first exploratory Bayesian analysis, three phenotipically preclassified wildcats – FSI 728, 729 and 737 – were significantly assigned to the domestic cluster (individuals q_w>0.95; P <0.80). All individuals were collected in Granada province in areas of wildcat distribution. Based on previous results and the considerable error documented for unequivocal phenotypic distinction between European wild and tabby domestic cats (Ragni, 1993; Oliveira *et al*. 2007), wrong morphological identification was considered a plausible explanation for this incongruence and the most conservative way of dealing with this problem in order to avoid introducing errors in our analyses (see Discussion). Consequently, these three individuals were excluded from the analysis and the total dataset became constituted of 181 samples: 109 domestic cats and 72 wildcats.

Genetic diversity and multivariate clustering of individual genotypes

All diversity estimations and differentiation values were corrected after excluding the individuals whose hybrid ancestry could be detected in the Bayesian analysis presented below. All loci were polymorphic in both European wild and domestic cats, showing a mean of 5.25 alleles per locus. None of the combinations between pairs of loci disclosed a significant deviation from Linkage Equilibrium. High levels of expected heterozygosity were found among Portuguese wild, Spanish wild and domestic cats (He=0.759±0.025; He=0.707±0.035; He=0.771 ±0.028, respectively). Although most of the variation was found within populations (80%), results reflect distinct gene pools among the sampled groups. Over all loci, a highly significant proportion of the total genetic variation was partioned between both wild and domestic populations ($F_{ST\ (Fca\ vs\ Fsi)}$=0.20; P <0.001). Multilocus pairwise interpopulation differences was also significant between Portuguese and Spanish wildcats: $F_{ST\ (Fca\ vs\ Fsi\ PT)}$=0.20; $F_{ST\ (Fca\ vs\ Fsi\ SP)}$=0.24; $F_{ST\ (Fsi\ PT\ vs\ Fsi\ SP)}$=0.11 (P <0.001; Figure 2). Significant genetic differentiation when grouping wildcats from Northern (North and Centre of Portugal, Asturias, León and Basque Country) and Southern Iberia in two different populations was also found (F_{ST}=0.10; P <0.001; AMOVA), however, genetic closeness between samples was independent of their geographic proximity and genotypes partition was more random than the one observed when grouping Portuguese vs Spanish cats (data not shown). Simultaneously, the partition of wildcats in two different clusters roughly corresponding to the division of Portuguese and Spanish cats was obtained in the Bayesian analyses (see below). Accordingly, we maintained the political nomination of samples origin. Principal Component Analysis scores of all individuals were graphically presented in a plan defined by two principal axes, which explain, cumulatively, 55.58% of the total genetic variability (Figure 2a). The plotting disclosed an evident separation between wild (Fsi SP + Fsi PT) and domestic cat populations, revealing a clear genetic differentiation between them. A closer proximity was found between Portuguese wildcats and domestic cats when compared to the genetic proximity between Spanish wild and domestic cats (Figure 2b and c). Some putative Portuguese wildcats plotted towards the domestic

group, corresponding to outlier individuals that might have admixed ancestry (Figure 2b; see Bayesian analysis below for outlier individuals' identification).

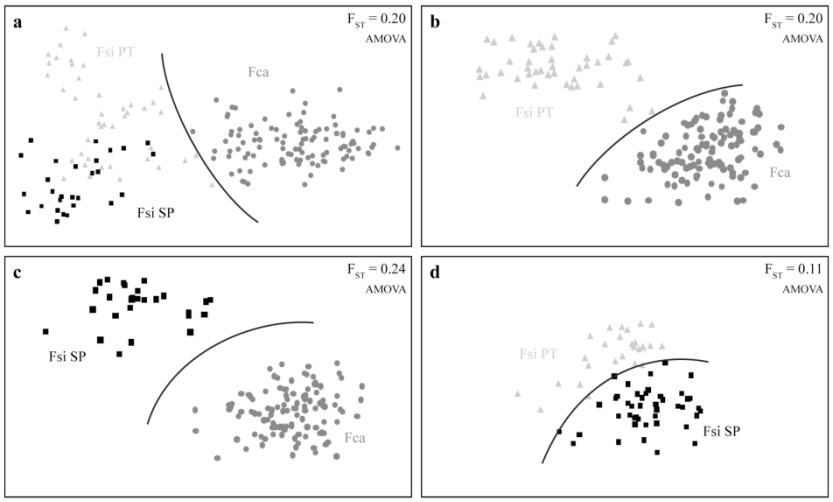


Figure 2. Principal Component Analysis of individual multilocus genotypes, computed using GenAlEx 6b4 (Peakall and Smouse, 2006) and genetic divergence (FST) among the different wild and domestic populations, assessed by Analysis of Molecular Variance (Fca = domestic cats, Fsi PT = wildcats from Portugal, Fsi SP = wildcats from Spain): a) Fca versus Fsi PT+Fsi SP; b) Fca versus Fsi PT; c) Fca versus Fsi SP; d) Fsi PT versus Fsi SP. Approximate area of expected admixed genotypes and possible hybrids are identified in b).

Bayesian inference of population structure and admixture

Bayesian admixture analyses using only genetic information clearly suggest the presence of two or three sharply differentiated groups in Iberian Peninsula, since the probability of the data increased steadily for K=2 and K=3. Afterwards, $LnP(D)_K$ - $LnP(D)_{k-1}$ reached a plateau. At the same time, the conversion of likelihood values into probabilities following Pritchard and Wen (2003) revealed an approximately 100% probability of having two or three distinct clusters in the dataset against almost 0.00% for all other values of *K*. With *K*=2 and using only genetic information, we estimated the average membership proportions (*Q*) of each predefined group (wild and domestic cats) into both clusters genetically inferred. All domestic cats were probabilistically assigned to cluster I with Q_I=0.99, while wildcats were mostly assigned to cluster II with Q_{II}=0.96. Accordingly, splitting the samples in two clusters allowed assigning individuals to their biological partition as wild and domestic cats. For K=3, all domestic cats were equally assigned to cluster I while wildcat samples were further subdivided between clusters II and III, with Portuguese cats clustering with Q_{II}= 0.683 and Q_{III}=0.250 and Spanish wildcats assigning with Q_{II}= 0.169 and Q_{III}=0.817 (Figure 3 a and b). Partition assignment of some Spanish genotypes in cluster II and Portuguese cats in cluster III does not reflect a closer geographic origin to the other nominal population.

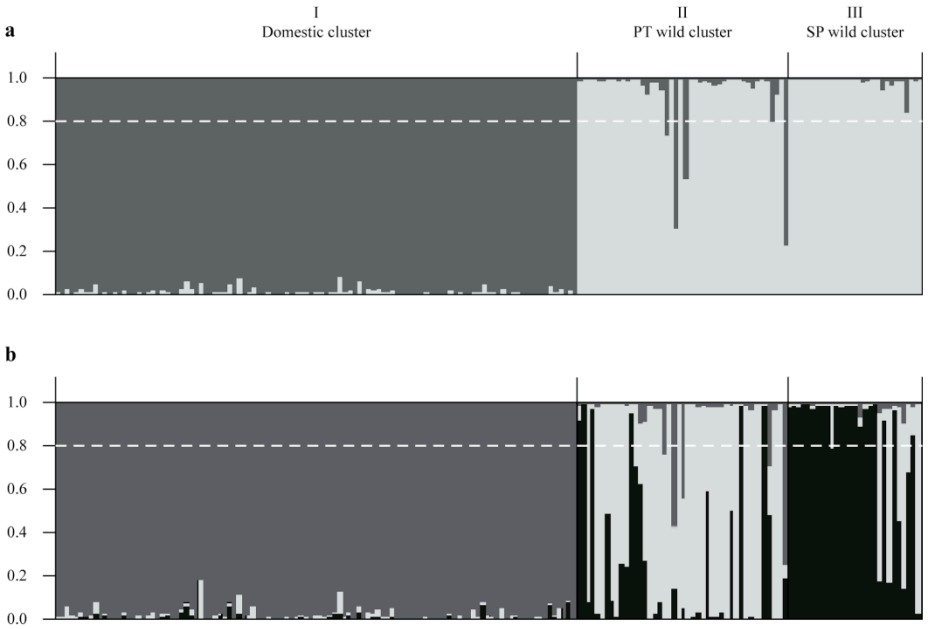


Figure 3. Plot of the probabilistic assignment of wild and domestic cats to the genetic clusters inferred by the Bayesian analysis performed in STRUCTURE, with K=2 (a) and K=3 (b). Each cat is represented by a vertical bar fragmented in *K* coloured sections which are relative to their membership proportion in the diverse genetic clusters: I - domestic (dark grey); II - Portuguese wildcat (light grey) and III - Spanish wildcat (black). The horizontal white line represents the threshold probability of 80% used to assign each individual to a single population.

At a probabilistic threshold of $qi>0.80$, the admixture analysis performed on simulated genotypes was able to efficiently recognise 100% of the parental individuals, with 90% CI higher than 0.88. All the F1 hybrids were correctly identified as admixed cats, however, a proportion of 9% F2 and 15% backcross genotypes showed a $qi>0.80$ to one single cluster and could not be distinguished from their parentals (Figure 4). All hybrids detected by simulations revealed very wide 90% CI, ranging between 0.20 and 0.80. Accordingly, we performed admixture analyses using the cutpoint of $qi>0.80$.

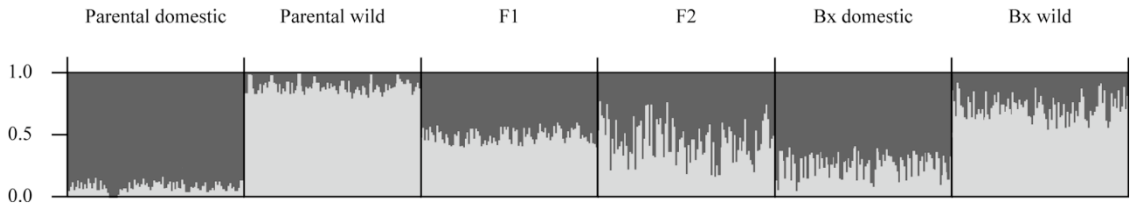


Figure 4. Plot of the Bayesian analyses performed in STRUCTURE using simulated parental, F1, F2 and backcrosses genotypes. The program was computed for K=2 under admixture model. Each individual is represented by a vertical bar coloured according to the proportion of its genome descending from *K* clusters. Genotypes were simulated using HYBRIDLAB 1.0 (Nielsen *et al.* 2006).

Table 1. Individual assignment and inferred ancestry of sampled wild and domestic cats, indicating the individuals identified as admixed cats . Individual *qi* values were calculated in STRUCTURE, supplying or not prior population information (ancestry in the other population: migrant = sampled generation, F1 = first generation hybrid, F2 = second generation hybrid). . Cluster I – Domestic cat; II Wildcat. Values in brackets = 90% confidence intervals.

Modelation	Samples	Cluster I	Cluster II		
Without prior non-genetic Information	**Domestic cats (n=109)**	**0.987 (0.890-1.000)**	**0.013 (0.000-0.210)**		
	Wildcats (n=67)	**0.041 (0.000-0.130)**	**0.959 (0.870-1.000)**		
	FSI682	0.695 (0.587-0.760)	0.305 (0.240-0.413)		
	FSI685	0.448 (0.345-0.701)	0.552 (0.299-0.655)		
	FSI689	0.547 (0.270-0.698)	0.453 (0.302-0.730)		
	FSI711	0.696 (0.384-0.755)	0.304 (0.245-0.616)		
	FSI878	0.709 (0.597-0.759)	0.291 (0.241-0.403)		
With information for:		**Population of Origin**	**Other population**		
			migrant	**F1**	**F2**
All individuals	**Domestic cats (n=109)**	**0.978**	**0.000**	**0.001**	**0.021**
	Wildcats (n=67)	**0.981**	**0.000**	**0.004**	**0.015**
	FSI682	0.048	0.100	0.317	0.320
	FSI685	0.362	0.001	0.766	0.174
	FSI689	0.191	0.138	0.265	0.405
	FSI711	0.160	0.034	0.503	0.304
	FSI878	0.025	0.091	0.112	0.772
All except hybrids	**Domestic cats (n=109)**	**0.979**	**0.000**	**0.001**	**0.020**
	Wildcats (n=67)	**0.972**	**0.000**	**0.007**	**0.021**
	FSI682	0.576 (0.383-0.765)	0.424 (0.235-0.617)		
	FSI685	0.503 (0.313-0.691)	0.497 (0.309-0.687)		
	FSI689	0.502 (0.313-0.697)	0.498 (0.303-0.687)		
	FSI711	0.536 (0.344-0.729)	0.464 (0.271-0.656)		
	FSI878	0.771 (0.245-0.799)	0.229 (0.755-0.211)		

Values of individual proportion of membership *qi* and their 90% CI computed with K=2-3 showed that all domestic cats had q_d>0.80 and minimum 90% CI of 0.89. Cluster II grouped approximately 93% of the phenotipically classified wildcats (90% CI between 0.87 and 1.00). Among Portuguese wildcats, we found five exceptions to this sharp differentiation, since putative wildcats FSI 682, 685, 689, 711 and 878 revealed a wild assignment and 90% CI <0.80, disclosing significant values of domestic ancestry (Table I; Figure 3a). At the same time, while most of the 90% CI values ranged between 0.87 and 1.00, putatively admixed cats revealed wider credibility intervals, ranging from 0.27 to 0.73. Individuals FSI 682, 685, 689 and 878 had been identified as outliers in PCA plotting (Figure 2b). In a more stringent modelation using prior morphological identification for all sampled wildcats (USEPOPINFO=1) and including or not that information for the putative hybrids formerly identified (POPFLAG=0 or 1), all posterior probabilities confirmed the results based only on genetic classification and the five admixed cats also revealed a q_w and 90% CI <0.80 (Table I). Even though the ancestral class of hybrid genotypes can be assessed, either in current or first and second past generations, none of these individuals presented posterior probabilities >0.80 for a single past hybrid generations. In any case, putative wildcat FSI 682 presented a considerably superior probability of

being an F1 hybrid (q_{F1}=0.77) and FSI 878 of being a backcross with the domestic gene pool (q_{Bxd}=0.78). Assuming that probabilistic assignments below threshold value indicate admixed cats, a minimum of 2.8% of the Iberian wildcats sampled in this study showed signals of introgressive hybridisation.

DISCUSSION

The preservation of wildlife populations has always been a controversial issue, mainly because of all variables that must be taken in consideration to design management programs. Nevertheless, it is a general agreement that conservation measures should focus on preserving healthy and outbred populations essentially by maintaining sufficiently large and suitable habitats that allow genetic exchange. The protection of some wildlife environments not only benefits that particular species, but also assists the preservation of important ecosystems and other co-habitant species. Unfortunately, the European continent has undergone significant habitat loss and fragmentation over the years, impeding the natural range of most wildlife species. A variety of wildcat populations has now extremely limited natural ranges and lives in low densities.

Along with habitat preservation, maintenance of genetically unique and "pure" wild populations is recognized as high conservation priority. Artificial hybridization between a species and its domesticated equivalent can severely influence the conservation status of threatened species and their legal protection. Introgression of alien domestic alleles has even led to extinction of many populations and species (see Rhymer and Simberloff 1996; Allendorf *et al*. 2001). The risk of introgression of domestic cat genes into wildcat gene pools is a big concern to conservation biologists since most wildcat populations are now in juxtaposition with the urban ranges of feral domestic cats (Stahl and Artois, 1994). Crossbreeding with domesticated forms may culminate in the homogenization of gene pools and result in outbreeding depression, reduced fitness and, consequently, severe population declines of wild populations (Barilani *et al*. 2006). Specially for closely related sub(species), identifying the ecological and biological driving forces of this phenomenon can be exceptionally challenging. Being able to understand these factors or even to identify such admixture events is particularly complex in domestic and wild cat subspecies, considering their significant genetic proximity when compared to other hybridising *taxa* (e.g. grey wolf and dog, Verardi *et al*. 2006; coyote and red wolf, Adams *et al*. 2007).

In order to develop population management programs for European wildcats, the uniqueness and "genetic purity" of populations need to be evaluated. Here, we successfully performed Bayesian admixture analyses of empirical and simulated datasets using multilocus microsatellites genotypes from wild and domestic cats across Iberian Peninsula. Our findings confirm the conclusions documented in previous hybridisation studies, showing that for populations with F_{ST} values around 0.12-0.20, 12 to 24 loci are sufficient to detect first generation hybrids (Barilani *et al*. 2006; Vähä and Primmer, 2006, Oliveira *et al*. 2007). However, detectability of hybrids decreases exponentially with repeated backcross

into the parental groups and beyond the second generation of hybridisation some individuals classified as pure wildcats might actually result from repeated backcrosses of admixed cats with wild individuals. Even though we were able to improve our first analyses by increasing the representatives of both parental populations (see Oliveira *et al*. 2007), only a slight increase in F2 (88% to 91%) and backcrosses (80% to 85%) identification was achieved in the simulation tests, and it was not enough to unambiguously identify all hybrid classes. Accordingly, it is essential to find the means to increase the power of Bayeasian analyses to differentiate and detect admixture between wild and domestic cats. For genotypes with no missing data, it is expected that confidence intervals are wider in admixed individuals, mainly if the parental populations are not sampled (Pritchard *et al*. 2000, Barilani *et al*. 2006). The wide 90% CI observed for all simulated hybrids confirm these findings. The correct estimation of membership of all the unequivocally preclassified domestic cats also corroborates the accuracy of the analyses.

Hybridisation can be overall widespread or locally very rare, which can be related with specific circumstances under what crossbreeding occurs. Mapping levels of introgression across European wildcat populations can then be used to prioritize areas for preservation and perform focused and efficient conservation strategies. Empirical population structure analyses showed that Iberian wild and domestic cats have high average posterior probabilities of assignment to their parental clusters, belonging to two clearly separated gene pools. Thus, we may assume that genetically distinct European wildcats remain in Iberian Peninsula and populations are scarcely hybridised in the most recent generations. Nevertheless, using both a stringent procedure where prior population information is given and without using any non-genetic information, at least 6.9% (5/72) of the Iberian wildcats probably have hybrid ancestry (2.8% of all samples). These findings add an admixed cat to the four already detected in the Portuguese wildcat population by Oliveira *et al*. (2007), individual that was sampled in a natural park in the Centre of Portugal. These admixed individuals probably represent diverse levels of hybridisation, suggesting that crossbreeding exists and should be regarded as a real threat to the wild population. Hybrid individuals were exclusively identified in Portugal and closer genetic proximity was found between Portuguese wild and domestic cats, which might be an indication of higher levels of recent introgression when compared with Spain. Small localized populations are known to be more susceptible to decline through hybridisation and Portuguese populations are thought to be decreasing, increasingly fragmented and isolated. According to our simulations on F2 and backcrosses detection, we regard this number of hybrids as a minimum value of admixed cats in Portugal, since past events of crossbreeding might have remained undetected if hybrids are crossbreeding with individuals that belong to the parental populations. Geographically separated wildcat populations from Portugal and Spain revealed a genetic divergence that suggests they should be considered singular units of study. However, some of the Portuguese and Spanish cats were assigned to the other cluster without any apparent biological/ecological reason, such as translocation or geographical proximity of animals. In our perspective, this can be explained by two simultaneous reasons. In one hand, we are dealing with populations that are genetically very close and, although, significant divergence can already be found (significant Fst values in AMOVA and K=3 in STRUCTURE), splitting of genotypes is still not complete and not sufficient to divide wildcats in

completely non-overlapping separated groups. On the other hand, and as a consequence of the genetic proximity mentioned, the number and type of molecular markers need to be increased to be able to perform fine-substructure analysis across wildcat populations.

According to genetic data, three of the morphologically preclassified Spanish wildcats were significantly assigned to the domestic cluster. Only two of the genetically admixed cats were phenotipically identified as possible hybrids (FSI711 and FSI878) and two morphological hybrids were genetically classified as wildcats with very high membership probabilities (FSI719 and FSI725; qw>0.96, minimum 90% CI of 0.93). In a wildcat population with admixture, there would be the possibility that the putative wildcats that were genetically assigned to the domestic cluster were in fact backcrosses that we weren't able to detect due to the referred lower power of our analyses for F2 and backcrosses identification. However, no hybrids that could over-time backcross with wild or domestic cats were identified in most recent generations among the Spanish cats. Particularly from the Granada population where these three cats were collected, we have genotyped 19 more individuals and no evidences of recent admixture were found. Accordingly, although we can't discard this possibility, we think that these results most likely reveal discrepancies between phenotypic and molecular identifications as described in previous studies (Ragni, 1993; Lecis *et al*. 2006) and we excluded them from the analyses. These findings highlight the importance of molecular tools for wild, domestic cats and cryptic hybrids identification.

Ongoing development of wildcat molecular studies

Many demographic, ecological and historical reasons might be involved in the diverse hybridisation and introgression rates found across European wildcat population: i) it is possible that habitat changes and fragmentation may have had higher impact in original forest landscapes (Central Europe) than in mosaic Mediterranean landscapes (Southern Europe); ii) tradition to have house cats or to feed feral domestic cats can also be an important variable; iii) the different demographic declines that wildcat populations might have undergone may have allowed feral domestic cats to crossbreed differently in the past, when populations comprised just a few reproductive individuals; iv) the multi domestication events recently described by Driscoll and co-authors (2007) which might have resulted in diverse domestic cat gene pools that might have introgressed differently into the wild populations across Europe.

Different methodological developments may contribute for a substantial improvement of the population analysis presented in this study and for answering several question regarding hybridisation rates across European wildcat populations. A two-pronged molecular approach (using both invasive and noninvasive sampling procedures) would be of major importance to monitor populations of this endangered and elusive feline. Scat surveys allow a time and cost-effective sampling effort in inconspicuous populations and significantly reduce anthropogenic pressures still frequently related to ecological and genetic studies. At a molecular level, it is crucial to overcome identification uncertainties searching for more powerful diagnostic traits. The ability to identify hybridisation further back in the past using neutral unlinked microsatellites would imply a significant higher genotyping effort, as

suggested by Rosenberg *et al*. (2003), and more recently by Vähä and Primmer (2006) Accordingly, a simultaneous increase in number and type of analysed loci would be necessary to discriminate between hybrid classes and to develop high-resolution inferences, specially if combined with recently available statistical methods based in Bayesian assumptions (Pritchard *et al*. 2000; Vähä and Primmer, 2006). The use of a large number of unlinked and linked microsatellites may allow better estimates of individual cats' proportion of membership into inferred clusters and genotyping microsatellites located in linkage groups might enable better statistical estimates of hybridisation further back in the past (Lecis *et al*., 2006). On the other hand, a wide-genomic investigation of novel molecular markers and the establishment of new diagnostic loci is our current field of investigation. Domestication produced obvious changes in domestic cat reproduction, coat colour, size, disease resistance and behaviour, as compared to the ancestral wildcat. Therefore, we will perform a detailed analysis of polymorphism at candidate genes underlying several domestic traits, determine current patterns of diversity in such genes and search for genetic footprints in cat's genome, i.e., signatures of selection at these loci that may have happened during domestication events. This analysis will focus on candidate genes identified as having major functional roles in mammal species, namely the ones most likely involved in litter sizes, fertility and coat color patterns diversity. Following this line of research, we aim to identify single nucleotide polymorphism (*SNPs*), molecular markers that might overcome some technical errors inherent to microsatellites (e.g. size homoplasy) and that have been revealing high efficiency, genotyping facility and analytical simplicity in their gradual application in population structure and admixture analyses (Zhang and Hewitt, 2003). The utility of all newly identified markers can also be evaluated and rated so a simple and rapid protocol can be designed (using the most informative ones) as a routine DNA-based test to detect and monitor hybridization in the wild.

It is also important to point out that consolidation of molecular inferences should include an extensive ecological knowledge of wildcat populations. More focused conservation policies might be achieved through the identification of historical and recent ecological features that could be related to and promote admixture. A more extensive study should aim to relate habitat variables with hybridisation by a comparative analysis of scarcely admixed versus largely hybridised populations in both less modified and disturbed landscapes across al Iberian Peninsula.

ACKNOWLEDGMENTS

We thank Elena Ballesteros Duperón, Marcos Moleón, Manuela Malsaña, Gerardo Dominguez Penafiel, J.M. Fernández and J.L. Robles for providing wild and domestic cat samples from Spain. We also acknowledge Portuguese National Tissues Bank / National Conservation Institute (BTVS/ICN) and Pedro Monterosso for providing Portuguese wildcat samples, and several biologists and veterinarians that assisted domestic cat sampling across Portugal. Rita Oliveira is supported by Fundação para a Ciência e a Tecnologia (FCT) through a PhD grant SFRH/BD/24361/2005. Raquel Godinho is supported by the Project PTDC/CVT/71683/2006 and the Post-Doc grant SFRH/BPD/12723/2003, both from FCT.

REFERENCES

Adams JR, Lucash C, Achutte L and Waits LP (2007) Location hybrid individuals in the red wolf (*Canis rufus*) experimental population area using spatially targeted sampling strategy and faecal DNA genotyping. *Molecular Ecology* 16: 1823-1834.

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

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

Barilani M, Sfougaris A, Giannakopoulos A, Mucci N, Tabarroni C and Randi E (2006) Detecting introgressive hybridisation in rock partridge populations (*Alectoris graeca*) in Greece through Bayesian admixture analyses of multilocus genotypes. Conservation Genetics.

Beaumont M, Barratt EM, Gottelli D, Kitchener AC, Daniels MJ, Pritchard JK, Bruford MW (2001) Genetic diversity and introgression in the Scottish wildcat. *Molecular Ecology* 10: 319-336.

Cabral MJ, Almeida J, Almeida PR, Dellinger T, Ferrand de Almeida N, Oliveira ME, Palmeirim JM, Queiroz AI, Rogado L, Santos-Reis M (eds) (2005) Livro Vermelho dos Vertebrados de Portugal. Instituto de Conservação da Natureza, Lisboa.

Corander and Martinnen (2006) Bayesian identification of admixture events using multilocus molecular markers. *Molecular Ecology* 15: 2833-2843.

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* 244: 231-247.

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

Driscoll CA, Menotti-Raymond M, Roca AL, Hupe K, Johnson WF, Geffen E, Harley E, Delibes M, Pontier D, Kitchener AC, Yamaguchi N, O'Brien SJ, Macdonald D(2007), The Near Eastern origin of cat domestication. *Science* 317: 519-23.

Eckert and Hartl (2005) Conservation genetics of the European wildcat (*Felis silvestris silvestris*) in Germany. Poster presentation in Symposium of Biology and Conservation of the European Wildcat (*Felis silvestris silvestris*). Germany January 21st-23rd 2005.

Falush D, Stephens M, Pritchard J (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164: 1567-1587.

Garnier S, Alibert P, Audiot P, Prieur B, Rasplus J-Y (2004) Isolation by distance and sharp discontinuities in gene frequencies: implications for the phylogeography of an alpine insect species, Carabus solieri. *Molecular Ecology* 13: 1883-1897.

Gerloff U, Schlotterer C, Rassmann K, Rambold I, Hohmann G, Frutth B and Tautz D (1995) Amplification of hypervariable simple sequence repeats (microsatellites) from excremental DNA of wild living Bonobos (*Pan paniscus*). *Molecular Ecology* 4: 515-518.

Guo SW and Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportions for multiple alleles. *Biometrics* 48: 361-372.

Lecis R, Pierpaoli M, Birò S, Szemethy L, Ragni B, Vercillo F, Randi E (2006) Bayesian analyses of admixture in wild and domestic cats (*Felis silvestris*) using linked microsatellite loci. *Molecular Ecology* 15: 119-131.

McOrist S and Kitchener AC (1994) Current threats to the European wildcat in Scotland. *Biological Conservation* 76 (2): 212.

Menotti-Raymond MA and O´Brien S (1995) Evolutionary conservation of ten microssatellites loci in four species of FELIDAE. *Journal of Heredity* 86: 319-322.

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

Nielsen EE, Bach LA, Kotlicki P (2006) Hybridlab (version 1.0): a program for generating simulated hybrids from population samples. *Molecular Ecology Notes* 6 (4): 971–973.

Oliveira R, Godinho R, Randi E, Ferrand N and Alves PC (2007) Molecular Analysis of hybridisation between wild and domestic cats (*Felis silvestris*) in Portugal: implications for conservation.

Peakall R and Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6: 288-295.

Pierpaoli M, Birò ZS, Herrmann M, Hupe K, Fernandes M, Ragni B, Szemethy L, Randi E (2003) Genetic distinction of wildcat (*Felis silvestris*) populations in Europe, and hybridization with domestic cats in Hungary. *Molecular Ecology* 12: 2585-2598.

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

Pritchard J and Wen W (2003) Documentation for structure software: Version 2. Department of Human genetics, University of Chicago, USA.

Ragni B (1993) The crucial problem of in vivo identification of wildcat and recognition of hybrids with domestic cats. Institute of Zoology, Peruglia University, Italy.

Ragni B and Possenti M (1996) Variability of coat-colour and markings system in *Felis silvestris*. *Italian Journal of Zoology* 63: 285–292.

Randi E, Pierpaoli M, Beaumont M, Ragni B (2001) Genetic identification of wild and domestic cats (*Felis silvestris*) and their hybrids using Bayesian Clustering Methods. *Molecular Biology and Evolution* 18 (9): 1679-1693.

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

Rhymer JM and Simberloff D (1996) Extinction by hybridisation and introgression. *Annual Review in Ecology and Systematics* 27: 29-44.

Rice WR (1989) Analyzing tables of statistical tests. *Evolution* 43: 223-225.

Rosenberg NA, Li LM, Ward R and Pritchard JK (2003) Informativeness of genetic markers for inference of ancestry. *American Journal of Human Genetics* 73: 1402-1422.

Sambrook E, Fritsch F, Maniatis T (1989) Molecular Cloning. Cold Spring Harbour Press, Cold Spring Harbour, New York.

Stahl P and Artois M (1994) Status and conservation of the wildcat (*Felis silvestris*) in Europe and around the Mediterranean rim. Convention on the Conservation of European Wildlife and Natural Habitats Standing Committee. *Council of Europe Nature and Environment Series* 69: 1–76.

Vähä JPK and Primmer CR (2006) Detecting hybridization between individuals of closely related populations – a simulation study to assess the efficiency of model-based Bayesian methods to detect hybrid individuals. *Molecular Ecology* 15:63–72.

Verardi A, Lucchini V and Randi E (2006) Detecting introgressive hybridisation between free-ranging domestic dogs and wild wolves (*Canis lupus*) by admixture linkage disequilibrium analysis. *Molecular Ecology* 15: 2845-2855.

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

Zhang D-X and Hewitt GM (2003) Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular Ecology* 12: 563-584.

CHAPTER 3

Noninvasive genetics

"Conservation biology should be, by definition, noninvasive."

DeSalle & Amato

PAPER III. Beja-Pereira A, Oliveira R, Alves PC, Schwartz MK and Luikart G (2009)

ADVANCING ECOLOGICAL UNDERSTANDINGS THROUGH TECHNOLOGICAL TRANSFORMATIONS IN NONINVASIVE GENETICS.

Molecular Ecology Resources 9(4): 1279-1301.

PAPER IV. Oliveira R, Castro D, Godinho R, Luikart G and Alves PC (2010)

SPECIES IDENTIFICATION USING A SIMPLE SSCP ANALYSIS OF A NUCLEAR GENE: APPLICATION TO CARNIVORES OF SOUTHWEST EUROPE.

Conservation Genetics 11(3): 1023-1032.

Advancing ecological understandings through technological transformations in noninvasive genetics

Albano Beja-Pereira*[1], Rita Oliveira*†[1], Paulo C Alves*†, Michael K Schwartz‡, Gordon Luikart#[1]

*CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, Campus Agrário de Vairão, 4485-661 Vairão, Portugal
†Departamento de Zoologia e Antropologia, Faculdade de Ciências da Universidade do Porto, 4099-002 Porto, Portugal
‡ USDA Forest Service, Rocky Mountain Research Station, Missoula, MT 59801, USA

[1]these authors contributed equally to this work

ABSTRACT

Noninvasive genetic approaches continue to improve studies in molecular ecology, conservation genetics and related disciplines such as forensics and epidemiology. Noninvasive sampling allows genetic studies without disturbing or even seeing the target individuals. Although noninvasive genetic sampling has been used for wildlife studies since the 1990s, technological advances continue to make noninvasive approaches among the most used and rapidly advancing areas in genetics. Here, we review recent advances in noninvasive genetics and how they allow us to address important research and management questions thanks to improved techniques for DNA extraction, preservation, amplification and data analysis. We show that many advances come from the fields of forensics, human health and domestic animal health science, and suggest that molecular ecologists explore literature from these fields. Finally, we discuss how the combination of advances in each step of a noninvasive genetics study, along with fruitful areas for future research, will continually increase the power and role of noninvasive genetics in molecular ecology and conservation genetics.

Keywords: Conservation Biology, Molecular Ecology, Conservation Genetics, Ecological Genetics, Population Ecology, Population Genetics, Forensics, Wildlife Management

INTRODUCTION

Noninvasive genetic sampling was first used in wild animals nearly two decades ago (Höss *et al.* 1992; Taberlet & Bouvet 1992). The main advantage of noninvasive genetics is that it allows biologists to study many individuals and populations without contacting, disturbing, or even seeing the organisms. Samples collected noninvasively include faeces, hairs, urine, saliva from chewed material, feathers, scent marks, eggshells, sloughed skin, and even menstrual fluid (Table 1). The largest contributions of noninvasive approaches are to studies that focus on (i) identification of individuals for studies of population size and individual movement, (ii) wildlife forensic cases, (iii) delineation of populations and population genetic parameters (structure, gene flow and demographic his- tory such as bottleneck detection), and (iv) assessment of mating systems and behavioral ecology (Table 1).

A growing number of noninvasive techniques yield good enough DNA and low enough genotyping error rates to allow researchers to address nearly all questions that can be addressed using traditional high-quality samples such as blood (e.g. Epps *et al.* 2006; Luikart *et al.*2008a). This is exciting because noninvasive studies 5–10 years ago were generally more limited in scope by high genotyping error rates and low polymerase chain reaction (PCR) amplification success (reviewed in Taberlet *et al.* 1999; Waits & Paetkau 2005). In this review, we report recent advances from different research fields, hoping to open communication channels and diffuse information among disciplines.

Rapid advancements in forensic science, human medical research, and livestock disease studies, and ancient DNA techniques continuously generate improved techniques that can be applied in noninvasive genetics to improve both data production and analysis. Unfortunately, these scientific communities seldom cross-reference each other. To continually improve molecular ecology and conservation genetic studies, we recommend that researchers occasionally search for novel approaches in journals from diverse fields including forensics (e.g. Journal of Forensic Sciences), human and animal health (Avian Disease, New England Journal of Medicine), microbiology (e.g. Journal of Applied Microbiology), biochemistry and biotechniques (Analytical Biochemistry and Nature Methods), and bioinformatics, e.g. Biometrika (see also our literature cited).

Table 1. A list of different biological samples, taxa and purposes for which noninvasive sampling have been used in wild animal populations.

	Group	Species	Purpose	Study
Blood in snow	Mammals	Wolf *(Canis lupus*)	Species identification	(Scandura, 2005)
Buccal and cloacal swab	Amphibians and Reptiles	Tuatara (*Sphenodon punctatus*)	Methodology	(Miller, 2006)
Buccal swab	Amphibians and Reptiles	Alpine newt (*Triturus alpestris)* and green tree frog (*Hyla arborea)*	Methodology	(Broquet *et al.*, 2007a)
Buccal swab	Birds	Black-capped chickadee (*Poecile atricapillus*) and boreal chickadee (*P. hudsonica*)	Methodology, individual and gender identification	(Handel *et al.*, 2006)
Eggshell	Birds	Greater snow goose (Chen caerulescens atlantica)	Methodology	(Lecomte *et al.*, 2006)
Eggshell	Birds	Domestic chicken (*Gallus gallus*)	Chicken anemia virus detection	(Miller *et al.*, 2003)
Eggshell, feathers, buccal swab	Birds	Sage-Grouse (*Centrocercus uropihasianus*)	Gender determination	(Bush *et al.*, 2005)
Eggshell	Birds	Caspian tern (*Sterna caspia*) and herring gull (*Larus argentatus*)	Methodology	(Schmaltz *et al.*, 2006)
Faeces	Birds	chick-rearing macaroni penguin (*Eudyptes chrysolophus*)	Diet determination	(Deagle *et al.*, 2007)
Faeces	Birds	European stonechat (*Saxicola torquata rubicola*)	Hormones monitoring	(Goymann, 2005)
Faeces	Birds	Dhole (*Cuon alpinus*)	Population genetics and phylogeography	(Iyengar *et al.*, 2005)
Faeces	Birds	Capercaillie (*Tetrao urogallus*)	Population structure, gene flow	(Regnaut *et al.*, 2006)
Faeces	Mammals	Red wolf (*Canis rufus*)	Hybridization monitoring	(Adams and Waits, 2007)
Faeces	Mammals	Brown bear (*Ursus arctos*)	Population size estimation	(Bellemain *et al.*, 2005)
Faeces	Mammals	Wild western gorilla (*Gorilla g. gorilla*)	kinship associations	(Bradley *et al.*, 2007)
Faeces	Mammals	Multiple fur seal species (*Arctocephalus sp*)	Diet determination	(Casper *et al.*, 2007)
Faeces	Mammals	Wolf (*Canis lupus*)	Population density	(Creel *et al.*, 2003)
Faeces	Mammals	Desert bighorn sheep (*Ovis canadensis nelsoni*)	Gene flow estimation	(Epps *et al.*, 2006)
Faeces	Mammals	Multiple carnivore species	Diet determination	(Farrell *et al.*, 2000)
Faeces	Mammals	Eurasian badger (*Meles meles*)	Population size estimation	(Frantz *et al.*, 2003)
Faeces	Mammals	Atlantic spotted dolphins (*Stenella frontalis*)	Methodology	(Green *et al.*, 2007)
Faeces	Mammals	Amur tiger (Panthera tigris altaica)	Individual identification by scent-marking dogs	(Kerley and Salkina, 2007)
Faeces	Mammals	Coyote (*Canis latrans*)	Population size estimation	(Kohn *et al.*, 1999)
Faeces	Mammals	Otter (*Lutra lutra*)	Methodology	(Lampa *et al.*, 2008)
Faeces	Mammals	Bighorn sheep (*Ovis canadensis canadensis*)	Host genetic diversity and parasitism	(Luikart *et al.*, 2008a)
Faeces	Mammals	Chimpanzee (*Pan troglodytes verus*)	Methodology	(Morin *et al.*, 2001)
Faeces	Mammals	Iberian lynx (*Lynx pardinus*)	Species identification	(Palomares *et al.*, 2002)
Faeces	Mammals	Rock wallaby (*Petrogale penicillata*)	Population density	(Piggott *et al.*, 2006)
Faeces	Mammals	Lesser horseshoe bat (*Rhinolophus hipposideros*)	Methodology	(Puechmaille *et al.*, 2007)
Faeces	Mammals	Kit Fox (*Vulpes macrotis mutica*)	Population genetics	(Smith *et al.*, 2005)
Faeces	Mammals	Western gorilla (*Gorilla g. gorilla*) and barbary macaque (*Macaca sylvanus*)	Methodology	(Vallet *et al.*, 2008)
Faeces and hair	Mammals	European pine marten (*Martes martes*) and stone marten (*Martes foina*)	Species identification	(Ruiz-González *et al.*, 2008)
Faeces and hair	Mammals	Black bear (Ursus americanus)	Spatial behaviour	(Schwartz *et al.*, 2006)
Faeces and hair	Mammals	Wolverine (Gulo gulo)	Species and individual identification	(Ulizio *et al.*, 2006)
Faeces and urine	Mammals	Wolverine (*Gulo gulo*)	Individual identification and gender determination	(Hedmark *et al.*, 2004)
Faeces, sloughed skin and eggshell	Amphibians and Reptiles	Common European viper (*Vipera berus*), ringed snake (*Natrix natrix*) and smooth snake (*Coronella austriaca*)	Methodology	(Jones *et al.*, 2008)

Feathers	Birds	Greater flamingo (Phoenicopterus roseus)	Gender determination	(Balkiz *et al.*, 2007)
Feathers	Birds	Powerful owl (Ninox strenua)	Methodology	(Hogan *et al.*, 2008)
Feathers	Birds	Lesser spotted eagle (*Aquila pomarina*)	Social Organization	(Meyburg *et al.*, 2007)
Feathers	Birds	Eastern imperial eagle (Aquila heliaca)	Species identification	(Rudnick *et al.*, 2007)
Feathers	Birds	Eastern imperial eagle (*Aquila heliaca*)	Population genetics	(Rudnick *et al.*, 2008)
Feathers	Birds	Capercaillie (*Tetrao urogallus*)	Methodology	(Segelbacher, 2002)
Feathers (including museum specimens)	Birds	Spanish imperial eagle (*Aquila adalberti*)	Methodology	(Horvath *et al.*, 2005)
Feathers and eggshell	Birds	47 bird species	Gender determination	(Jensen *et al.*, 2003)
Foot mucus	Invertebrates	Multiple terrestrial snails	Methodology	(Palmer *et al.*, 2008)
Foot mucus	Invertebrates	Multiple intertidal snails	Methodology	(Kawai *et al.*, 2004)
Fresh water	Amphibians and Reptiles	American bullfrog (*Rana catesbeiana*)	Species identification	(Ficetola *et al.*, 2008)
Hair	Mammals	Domestic dog *(Canis familiaris*)	Methodology	(Bjornerfeldt and Vilà, 2007)
Hair	Mammals	San Joaquin kit fox (*Vulpes macrotis mutica*)	Methodology	(Bremner-Harrison *et al.*, 2006)
Hair	Mammals	Black bear (*Ursus americanus*)	Population density	(Dreher *et al.*, 2007)
Hair	Mammals	Giant panda (*Ailuropoda melanoleuca*)	Gender determination	(Durnin *et al.*, 2007)
Hair	Mammals	Orang-utan (*Pongo spp*)	Methodology	(Goossens *et al.*, 2004)
Hair	Mammals	Multiple North American carnivores	Population genetics	(Kendall and McKelvey, 2008)
Hair	Mammals	Mountain pygmy-possum (*Burramys parvus)*	Genetic diversity	(Mitrovski *et al.*, 2007)
Hair	Mammals	Brown bear (*Ursus arctos*)	Individuals abundance	(Mowat and Strobeck, 2000)
Hair	Mammals	Eurasian lynx (*Lynx lynx*)	Population monitoring	(Schmidt and Kowalczyk, 2006)
Hair	Mammals	Southern hairy-nosed wombat (*Lasiorhinus latifrons*)	Spatial distribution and habitat use	(Walker *et al.*, 2008)
Hair	Mammals	Ocelot (*Leopardus pardalis*)	Species, gender individual identification	(Weaver *et al.*, 2005)
Hair	Mammals	Multiple carnivore species	Methodology	(Zielinski *et al.*, 2006)
Hair / Faeces / Urine / Tooth / Saliva	Mammals	Wolf (*Canis lupus*)	Gender determination	(Sastre *et al.*, 2008)
Insect exuviae / frass	Invertebrates	Multiple butterfly species	Species identification	(Feinstein, 2004)
Ivory	Mammals	African Elephant *(Loxodonta africana spp*)	Forensic cases	(Wasser *et al.*, 2007)
Menstrual bleeding	Mammals	Taiwan macaque (*Macaca cyclopis*)	Methodology	(Chu *et al.*, 1999)
Museum specimen	Birds	*Gallinago spp*	Methodology	(Lee and Prys-Jones, 2008)
Museum specimen	Mammals	Brown bear (*Ursus arctos*)	Phylogeography	(Leonard *et al.*, 2000)
Museum specimen	Mammals	Stoat (*Mustela erminea*)	Methodology	(Martinkova and Searle, 2006)
Museum specimen	Mammals	Wolverine (*Gulo gulo*)	Evolutionary significant units	(Schwartz *et al.*, 2007)
Regurgitate	Mammals	Wolf (*Canis lupus*)	Individuals dispersion	(Valière and Taberlet, 2000)
Saliva	Birds	Common marmoset (*Callithrix jacchus*)	Cortisol levels and behavioural stress	(Cross *et al.*, 2004)
Saliva	Mammals	Wild chimpanzee (*Pan troglodytes verus*)	Individual identification	(Inoue *et al.*, 2007)
Saliva	Mammals	Wolf (*Canis lupus*)	Predator identification	(Sundqvist *et al.*, 2008)
Saliva	Mammals	Coyote (*Canis latrans*)	Predator identification	(Blejwas *et al.*, 2006)
Scent mark	Mammals	Multiple murine species	Microbial parasite communities identification	(Lanyon *et al.*, 2007)
Skin, blubber and meat	Mammals	Pacific minke whale (*Balaenoptera acutorostrata* spp)	Forensic cases	(Baker *et al.*, 2007)
Sloughed / Shed skin	Mammals	Humpback whale (Megaptera novaeangliae)	Methodology	(Elphinstone *et al.*, 2003)
Sloughed / Shed skin	Mammals	Humpback whale (*Megaptera novaeangliae*)	Individuals abundance	(Palsboll *et al.*, 1997)

Sloughed / Shed skin	Mammals	Ringed seal (*Phoca hispida*)	Methodology	(Swanson *et al.*, 2006)
Urine	Mammals	Japanese macaques (*Macaca fuscata*)	Methodology	(Hayakawa and Takenaka, 1999)
Urine	Mammals	Wolverine (*Gulo gulo*)	Methodology	(Hedmark *et al.*, 2004)
Urine	Mammals	Wolf (*Canis lupus*)	Population monitoring	(Hausknecht *et al.*, 2007)
Urine	Mammals	Multiple canid species	Species and individual identification	(Valière and Taberlet, 2000)
Urine	Mammals	Wolf (*Canis lúpus*)	Molecular sexing	(Sastre *et al.*, 2008)

This review is structured around the steps in a noninvasive study, from pre-PCR sampling to post-PCR data analysis, and concludes with perspectives for future research. Noninvasive studies should not be seen as a one-step process, but as a chain of steps that should be monitored independently. The chain starts in the living animal and ends only when the statistical analyses of the final data provide convincing evidence that results and conclusions are reliable. We consider five major steps to be monitored and how to avoid pitfalls and improve non-invasive studies (Fig. 1). Accordingly, this review is structured around steps and techniques, not research questions (e.g. paternity analysis, population structure), which allow readers to quickly go to the step or technique of interest (pre-PCR to post-PCR) to find information.

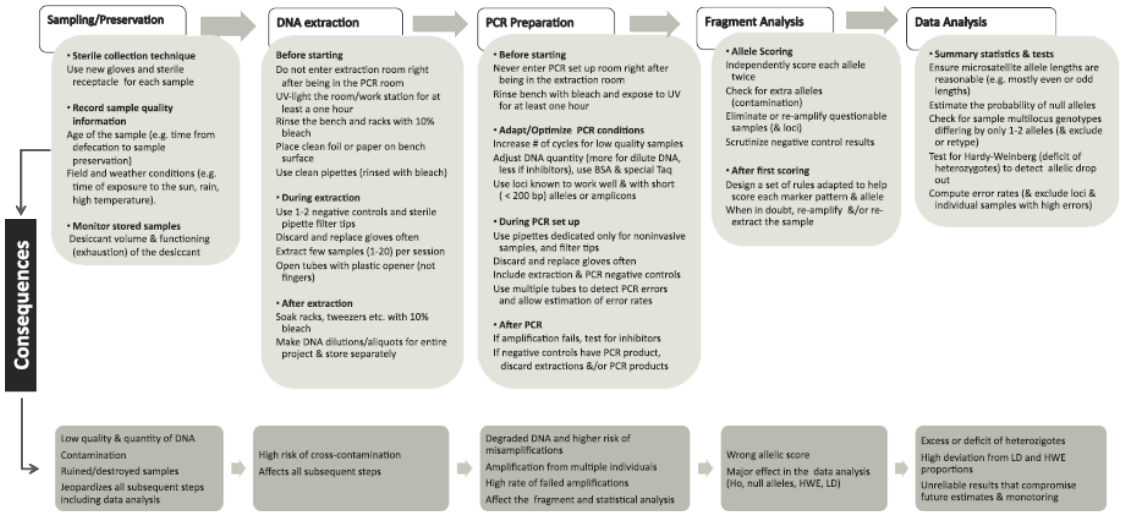


Figure. 1. Schematic representation of some critical points (light grey lists) that should be checked at each of the five main steps (white rectangles on top) of the noninvasive samples processing. Below (arrow boxes) are some likely consequences of not correctly following and monitoring these points. Some points are common sense and widely known but nonetheless are often violated. HWE, Hardy–Weinberg equilibrium proportions; LD, linkage or genotypic-disequilibrium.

PRE-PCR

a) Obtaining samples

Creative ways to noninvasively obtain DNA from numerous types of samples are continuously being developed, improved and evaluated (Table 1). The collection of everything from menstrual fluid to mucus trails left by snails has been used to identify species and individuals noninvasively (Table 1). Several sample types can be obtained by following a trail of an animal on natural surfaces such as snow or sand, without ever seeing the target animals. For example, Ulizio *et al.* (2006) collected 169 hair samples and 58 scat samples on 54 wolverine backtracks. One creative study reported the noninvasive detection of species (a frog, Rana catesbeiana) in natural wetlands by PCR testing for mtDNA in water samples (Ficetola *et al.* 2008).

Faeces are one of the most commonly used noninvasive materials because, for many species, it is the easiest to find in the wild and it provides more information (e.g. diet, stress hormone status, reproductive hormones, parasite infection and parasite DNA) than other sample types (Kohn & Wayne 1997; Goymann 2005; Luikart *et al.*2008a; Schwartz & Monfort 2008). Faeces in some species (e.g. ungulates, arboreal primates, macropods, etc.) can be collected just after observing individuals defaecate without disturbing the animals. An enormous advantage of observing the target animals is that the faeces are fresh and DNA is relatively little degraded. It can also help avoid collecting from nontarget species and determine sex (by observation) and thus avoid need for DNA-based species identification and sexing (Epps *et al.* 2006; Luikart *et al.* 2008b). In secretive or elusive species such as forest ungulates, bears, fishers, mountain lions and tigers, domestic dogs can be trained to find faeces (reviewed in McKay *et al.* 2008). Dogs can also identify individual animals, as was shown in a study of faeces from known tigers (Kerley & Salkina 2007). Hair is another widely collected material (Table 1). In apes (e.g. orangutans, chimpanzees), individuals build a new nest every night and hair that is shed during the night can be found in the nests. Researchers recommend using only hairs with visible root bulbs as many shed hairs do not contain large bulbs with DNA (Goossens *et al.* 2004). In a study of wolves, hair (along with faeces, urine and saliva) allowed highly successful DNA amplification (93% of samples) for noninvasive sexing of individuals using sex chromosome markers (Sastre *et al.* 2008). Hair is also often recovered frozen in the snow tracks of felids and canids and in bed sites of ungulates.

Many hair snare devices have been invented (e.g. Bremner-Harrison *et al.* 2006; Zielinski *et al.* 2006) for noninvasive sampling. Hair snares are used to sample bears (e.g. Immell & Anthony 2008; Kendall *et al.* 2009), felids (e.g. Weaver *et al.* 2005; Schmidt & Kowalczyk 2006) and mustelids Mowat & Paetkau 2002. Barbed wire or sticky tape is also often strung around bait stations or draped across animal burrow entrances to pluck hairs when animals pass by (Pauli *et al.* 2008; Tóth 2008; Walker *et al.* 2008). Along with hair snaring devices, com- mercial lures (such as catnip and valerian oils, among other attractants) have been successfully used to attract and elicit cheek-rubbing behavior in felid species (e.g. *Lynx canadensis*, McDaniel *et al.* 2000). For hair snares, a potential advantage is that they obtain plucked hairs, which generally contain more and larger root bulbs (with cells and DNA) than shed hairs. However, it might be difficult avoiding cross-contamination between individuals because multiple individuals can be sampled before hairs are recovered from the snare. As birds use mammal hair to strengthen the structure of their nest, recently Tóth (2008) used bird nests as sources of hair samples and identify mammals that occupy or migrate through a specific area.

Feathers have repeatedly been shown to be a good source of DNA. Shed feathers can be collected from nests. Feather snares (e.g. sticky tape) potentially could help obtain feather samples, but to our knowledge have not been reported in the literature. A particularly informative and recent study (Hogan *et al.* 2008) showed that different feather types (down, semi plume, contour or remige / rectrice) yield useful DNA. However, feather condition (as estimated from physical appearance) strongly influenced PCR amplification success.

For eggshells, a recent study used cotton swabs to obtain DNA from the external shells of herring gull (*Larus argentatus*) and Caspian tern (*Sterna caspia*) eggs (Handel *et al.* 2006). Researchers

verified that the DNA samples were maternal (not the chick's) by comparing microsatellite profiles with those obtained from adults and chicks from the same nests. In all of 28 tests, the egg swabs matched the maternal microsatellite genotype. In a screening of many nests of both species, microsatellite markers were successfully amplified from egg swabs. Eggshells are also used as a source of DNA in veterinary health and disease studies (e.g. Miller *et al.* 2003), from which molecular ecologists might learn new and useful techniques (e.g. improved DNA extraction or PCR techniques). Eggshells, feathers and mouth swabs from sage- grouse (*Centrocercus urophasianus*) were compared for PCR success in one of the few studies directly comparing multiple sample types (Bush *et al.* 2005). These authors found hatched eggshell membranes yield useful DNA (better than predated eggshells), as did plucked body contour feathers, chick down feathers and mouth swabs. However, allelic dropout rates of approximately 10% were observed for eggshells, and moulted feathers had only 60% PCR amplification success (Bush *et al.* 2005).

Saliva is also a good source of DNA. It is often used in forensics, for example, to recover DNA from bite marks found in homicides, assault and other criminal cases (Anzai-Kanto *et al.* 2005). In wildlife, Williams *et al.* (2003) used saliva collected from sheep bite wounds to identify the canid species responsible for attacks on domestic sheep; the authors identified the predator species (coyote) and determined the sex of the individual. Saliva is also used to solve cases of livestock attacks in which wolves and dogs are the main suspects (Sundqvist *et al.* 2008). For some sample types, a new swab sampling technique reported in the forensics literature could improve the quality of genotyping Pang & Cheung, 2007. The double swab technique, using a wet cotton swab followed by a dry cotton swab, was compared with the classical technique (one wet swab) for recovering DNA from evidence collected at crime scenes. Swab techniques could potentially improve noninvasive sampling studies involving material such as eggshells or any surfaces that animals come into contact, rub against, lick or bite (e.g. rocks, sticks). Further evaluation of this and other sampling methods is needed.

b) Preserving DNA in noninvasive samples

A growing diversity of protocols exists for preserving DNA in samples. This makes it difficult for researchers to understand which protocol is most reliable, most thoroughly validated, or requires further development and testing. For noninvasive samples, it is essential to conduct a pilot study using the exact target material, preservation method and extraction technique to ensure recovery of sufficient DNA (Bhagavatula & Singh 2006; Valière *et al.* 2006; Schwartz & Monfort 2008).

The preservation of DNA in a noninvasive sample is a race to inhibit enzymes that degrade DNA, i.e. nucleases. There are three main approaches used to preserve samples: deactivation of nucleases via removal of water, deactivation of nucleases via the elimination of cations (e.g. $MgCl_2$; Thomas & Gilbert 2006) and inhibition of nuclease activity via storage of samples at low temperatures. Removal of water is achieved using drying agents (e.g. ethanol, silica gel) or drying techniques (e.g. vacuum spinning, lyophilization, oven heating). Removal of cations is achieved using chelators such as EDTA or resin (e.g. Chelex®). Insufficient volumes of preservatives (ethanol or silica)

or failure to freeze samples quickly often leads to DNA degradation. Several published studies comparing different preservation protocols can help researchers choose the best protocol according to their samples (Roon *et al.* 2003; Hajkova *et al.* 2006; Broquet *et al.* 2007b; Santini *et al.* 2007). Nonetheless, there are inconsistencies among some studies, and even some suggestion that there is an interaction between preservation techniques and extraction methods (Piggott & Taylor 2003).

A potentially improved preservation approach is to combine use of silica and ethanol (ETOH) protocols, however, the ETOH (90%) performed similar to the combined two-step method when using lower quality samples (Roeder *et al.* 2004). This combined approach was repeated on gorilla and chimpanzee faeces and yielded more DNA than silica alone or RNAlater alone (Nsubuga *et al.* 2004). Nonetheless, this approach has not been extensively compared with other methods (e.g. ETOH 97%) in a wide variety of species. Long-term preservation might be improved by the addition of trehalose as a preservative agent (Smith & Morin 2005), although this method has not been independently evaluated. There is a great need for comparative evaluations of most preservation methods.

For faeces preservation, it is difficult to decide which desiccant (e.g. ETOH, silica, salts) should be used. A large amount of any desiccant should be used per sample (e.g. 5–10 parts of desiccant per part of sample) to rapidly and completely dry the sample material. Given the wide use and success, we recommend the use of ETOH in large volumes (5–10 times the sample volume) and in high concentration (≥95% ETOH). With faecal material, ETOH has advantages over silica in that ETOH prevents formation of faecal powders (thus cross-contamination by aerosol). It also keeps the external mucous layer containing cells packed against faecal material, whereas silica can be abrasive and can remove mucus and cells from outer surface of the faeces (e.g. during transportation and shaking of samples). ETOH has a notable disadvantage in being flammable and therefore potentially dangerous and more expensive to ship via airplane. As an alternative, silica is useful and widely tested but again requires large volumes of this mineral to ensure rapid drying and to avoid exhausting the desiccation function. For faeces, RNA later® might be a better preservative than ETOH or silica (http://www.aim.uzh.ch/orangutannetwork/GeneticSamplingProtocol.html#18; Nsubuga *et al.* 2004). RNA later® is a solution meant for preserving RNA in tissue. However, the solution is expensive (US$2–4 depending on volume needed per sample), and further research is needed to formally test and compare it with other preservatives.

For hair samples, the most common storage method is simply to store it (shed or plucked) in a dry envelope often with silica gel granules at room temperature (Jeffery *et al.* 2007). A fairly thorough comparative study of freezing (-20°C) vs. silica desiccant found that freezing gives slightly higher (though nonsignificant) amplification success for both microsatellites and mtDNA from brown bear hair, Ursus arctos (Roon *et al.* 2003). Amplification success was above approximately 90% up to 6 months of storage but dropped below approximately 80% between 6 and 12 months for both the 1000-bp mtDNA fragment and three microsatellite loci. More comparative studies are needed using different preservation methods including a combination of freezing and silica gel, and perhaps storing hairs immediately into a lysis or storage buffer solution. Sorting hairs based on root bulb size and

quality should also be conducted to maximize amplification success and perhaps improve accuracy of comparisons among preservation methods (Jeffery *et al.* 2007).

For feathers, storage in paper envelopes at —20o allowed successful amplification of mtDNA and nDNA from powerful owls (Ninox strenua; Hogan *et al.* 2008). In this study, the paper envelopes containing 637 shed feathers were stored in plastic bags in dry and dark conditions for up to 7 months. Amplification success was 80–90% for mtDNA and microsatellites on feathers in good condition but only 30–40% for feathers in poor condition (with visible physical degradation of calamus and barbs on the vane). Feather type had no effect on amplification success. We recommend against using plastic bags as humidity can potentially build up inside, unless silica desiccant is inside the bag. Feathers from adult eagles (Rudnick *et al.* 2007) stored dry at room temperature yielded microsatellite genotypes using a pre-amplification PCR method (PCR section below), although nearly 10% of samples yielded no PCR product. In the same study, developing chick feathers were stored at room temperature in a lyses buffer (EDTA, SDS) for several months before being ultimately stored at -80°C up to several years before yielding microsatellite genotypes.

Saliva samples are generally preserved by freezing at -20°C (Anzai-Kanto *et al.* 2005). For example, Anzai- Kanto *et al.* (2005) published a study using human saliva in which they estimate that 0.3 mL of saliva is enough to provide DNA for genotyping 15 loci. Swabs are the most general method to sample buccal / oral DNA, and these swabs are generally dried at room temperature followed by freezing at -20°C or even colder temperatures (e.g. see Sundqvist *et al.* 2008).

Urine samples as a source of DNA have been increasingly used in recent years. Urine can be collected using a swab to swipe the surface location where the animal urinated (e.g. rocks, sticks, leaves). The swab will absorb the urine together with the cells. Another method, used in winter, is the collection of urine in snow (yellow snow). Researchers have melted yellow snow in a 15-mL tube, which will contain urine, cells and DNA (Hausknecht *et al.* 2007). This method has been tested in carnivores, in particular the wolf. Urine samples can also be collected from soil samples. We have collected fresh ungulate urine from dirt, which becomes mud (G. Luikart, unpublished). We stored the urine mud in six volumes of 95% ETOH, similar to faecal samples, until extraction in the laboratory using stool extraction kits or soil kits (see below). While urine can be a useful material, it often has a lower amplification success rate as compared with other noninvasive samples (Hedmark *et al.* 2004). Hedmark *et al.* noticed a decline in microsatellite amplification success of wolverine urine (40% success) as compared with faeces (65% success).

c) Extracting DNA from noninvasive samples

DNA extraction is a crucial step, because all subsequent steps in a genetic project hinge upon extraction quality. Phenol/chloroform extraction methods were the most widely used 10–15 years ago, but now are seldom used, mostly because the chemicals are hazardous, the approach is time-consuming, and sometimes PCR inhibitors remain after extraction. As alternatives, different methods have appeared, most of them imported from forensic genetics (e.g. see book by Morling 2008).

Resin-based (e.g. Chelex®) extractions are widely used for noninvasively collected samples. Chelex is useful for extracting DNA from hair follicles (Mitrovski *et al.* 2005; Koukoulas *et al.* 2008), stains at crime scenes, and even for formalin-fixed archived tissues (Chakraborty *et al.* 2006). Its main advantages are speed and low cost (http://bugs.bio.usyd.edu.au/DNA/DNAextrn.html). The main disadvantages are that (i) DNA extracts are not always highly pure, (ii) DNA can degrade after several months, and (iii) Chelex itself can inhibit the PCR amplification (Willard *et al.* 1998).

Commercial kits for extracting DNA are also widely used. Among these, the most common are silica-based spin column kits. The working principle of this method involves the lysis of the cell membranes (e.g. by deter- gents and proteinase K), followed by purification using silica-based compounds in spin columns that bind and then allow washing of DNA (Boom *et al.* 1990). The great success of these kits results from their ease of use and adaptability to a wide range of biological samples (e.g. plant tissues, bacteria growing media, skin, muscle, bone, faeces, urine, blood, museum skins, ancient bone) with minimum changes.

When comparing five DNA extraction methods, the extracted samples from which a fragment of 149 bp of the mtDNA was successfully PCR amplified using a commercial kit (QIAGEN Stool DNA extraction kit) was 100%, followed by 88% using guanidinium thiocyanate- silica, 75% for the digest buffer / phenol–chloroform, 38% for chelex-100 and 25% for the lyses buffer / column purification method (Bhagavatula & Singh 2006).

For pellet-form faeces, which are amenable to a surface wash, the wash technique combined with commercial extraction kits [e.g. DNeasyTM Blood Kit (QIAGEN)] has been highly successful. The washing step is a simple 10–15 min incubation of a faecal pellet in a buffer solution followed by extraction of DNA from the buffer using a blood DNA extraction kit (Luikart *et al.* 2008b). The surface-wash liquid contains relatively few PCR inhibitors and therefore does not always require use of the more expensive and time-consuming 'stool kits' with additional steps to remove inhibitors. This approach yields high amplification success, low genotyping error rates and large quantities of DNA. For faeces, a cell enrichment method has been reported to recover large quantities of high molecular weight DNA (Wan *et al.* 2006). The cell enrichment based protocol is so far the only one that deals with large quantity of faeces, and is based on the soaking in a large volume of buffer to disperse the faecal material completely. A commercial company (Noninvasive Technologies) offers a kit for a similar extraction, but it costs over US$200 for the extraction of two individual samples. With faecal (& urine) samples, it is difficult to quantify the amount of extracted DNA using conventional methods (e.g. spectrophotometer) because these are inefficient with trace quantities of DNA, they cannot estimate DNA degradation, nor can they differentiate between DNA from the target species or microbes often in faecal (& urine) DNA extractions. To cope with these limitations, several assays have been developed using real-time quantitative (RTQ) PCR (Morin *et al.* 2007). Unfortunately, RTQ-PCR still is not affordable for all laboratories and alternative low-cost methods can be used to quantify the DNA extracted from some noninvasive samples. For example, Ball *et al.* (2007) used a method based on PicogreenTM (Molecular probes), a fluorescent dye, to measure the amount of double-stranded DNA extracted from noninvasive samples (e.g. faeces). PicogreenTM binds double-

stranded DNA and when excited by laser releases a florescent signal that is proportional to the amount of double-stranded DNA present in the tested aliquot. However, unlike RTQ-PCR, fluorescent dye methods cannot differentiate between the target species vs. microbial DNA.

The urine samples can be collected either by using a swab across the surface where the animal urinated (e.g. rocks, bush leafs) or in winter from snow. One extraction method involves centrifuging cells (sloughed off from the epithelium of the urinary tract). Once the cells are collected in a pellet, standard DNA extraction protocols can be used (Hausknecht *et al.* 2007). This approach is also valid for buccal-mouth wash (nondestructive) sampling in humans (Mayntz-Press & Ballantyne 2007). Another extraction method directly precipitates DNA from the sample (e.g. snow) containing the urine (Valière & Taberlet 2000). Direct precipitation would be advantageous when cells burst and DNA is free. DNA from the urine deposited in the soil (mud) can be obtained using stool DNA extraction kits or soil DNA extraction kits (e.g. Thakuria *et al.* 2009). Comparative evaluations of extraction kits on humans suggest that some kits (miniMAG) yield far better DNA than others, including DNA from pathogens being monitored noninvasively (Tang *et al*. 2005). Noninvasive wildlife studies might benefit from testing and using kits used in human studies.

For hairs, an improved extraction method reported use of Ca^{+} to increase digestion and release of DNA of hair shafts. In a forensic-based study of hairs from 170 dogs from different breeds, the quantity of DNA extracted increased 100% when compared to the well- established QIAGEN tissue kit (Pfeiffer *et al.* 2004).

Finally, it is important to mention that plastic tubes may have a strong effect of reducing DNA quantities when the amount of DNA in the sample is very low (fewer than 1000 target copies) because of DNA adhering to the plastic walls of the tube. A recent study showed that use of low-retention plastic tubes significantly reduce DNA loss, but DNA from nontarget species added to prevent the loss of target DNA had no effects (Ellison *et al*. 2006). As this problem becomes better understood, we imagine that low-retention plastic tubes will drop in price; more research is needed on changes in DNA yield caused by tube choice.

POLYMERASE CHAIN REACTION

Here we review approaches to improve PCR amplification of DNA, including pre-PCR treatments (for inhibitors and broken DNA fragments), amplification of smaller fragments (mini-STRs/microsatellites and SNPs), nested PCR techniques, different *Taq* polymerase enzymes and genotype scoring criteria.

a) Overcoming PCR inhibitors

Inhibition of PCR can cause low amplification rates, even in samples with abundant DNA and apparently suitable for PCR (Kontanis & Reed 2006). For example, faeces contain compounds

that can be strong PCR inhibitors, including complex polysaccharides, products from food degradation (e.g. acids, secondary plant compounds, enzymes, lipids and proteins), RNA and bacteria. As previously discussed, DNA extraction protocols combined with washes for DNA purification are essential to remove inhibitors. However, some inhibitors may still remain and result in amplification failure.

Dilution of the DNA extracts is the simplest way to reduce inhibitors (dilution is the solution to pollution). For example, Thornton & Passen (2004) diluted approximately 256-fold the DNA extract obtained from 10 mg of bovine faeces to achieve amplification inhibited by phytic acid (present in plants). Dilution also increased amplification efficiency of Iberian lynx (*Lynx pardinus*) mtDNA from 92.6% to 99%, equivalent to the benefit of performing a second amplification for each sample (Palomares *et al.* 2002). However, genotyping errors can be caused by low target DNA quantity or the presence of PCR inhibitors (or both interacting). Accordingly, a balance between diluting PCR inhibitors and over-diluting the DNA in the extract often must be established.

Precipitation of DNA (e.g. with ETOH) also removes inhibitors (and increase DNA concentration). This involves a washing step of the DNA pellet before re-dis- solving the DNA precipitant in water or buffer. Addition of PCR adjuvants such as bovine serum albumin (BSA), dimethyl sulfoxide, or nonionic detergents (e.g. Tween 20 and Triton X-100) often binds inhibitors and improves amplification specificity. Most noninvasive studies include an additive in PCR protocols. BSA is the most widely used adjuvant (from 0.1 to 1.2 lg/lL in concentration) because it seldom interferes with PCR in the absence of an inhibitor.

b) Overcoming DNA degradation and fragmentation

Using very short fragments such as mini short tandem repeats (mini-STRs, also called mini-microsatellies) or single nucleotide polymorphisms (SNPs) can help overcome difficulties amplifying degraded DNA (e.g. Campbell & Narum 2008). In several noninvasive studies, long amplicons (>200–300 bp) produced significantly higher allelic dropout rates than short amplicons (Broquet & Petit 2004; Buchan *et al.* 2005). Several studies have redesigned primers to produce shorter amplicons and improve microsatellite analysis in forensic research (e.g. Butler *et al.* 2003; Chung *et al.* 2004). In fact, studies using historical or ancient DNA typically amplify multiple small (100 bp) regions, instead of one large region as is typical with high-quality DNA (Schwartz *et al.* 2007).

Single nucleotide polymorphism studies can achieve higher amplification success and lower error rates than microsatellites, because SNP amplicons are generally shorter (<100 bp) than microsatellite amplicons (100–300 bp). For example, Musgrave-Brown *et al.* (2007) showed that a 52-plex SNP assay performed better than STR (microsatellite) typing on degraded samples. How- ever, the biallelic nature (and thus limited heterozygosity) of SNPs must be compensated by typing a larger number of SNP loci (Morin *et al.* 2004, 2009a, b). Thus, even if there is a lower error rate per SNP, the amplification of many more SNPs may cumulatively increase the overall (multilocus) genotyping error rates. More research is needed to quantify the increase in multilocus error rates when adding more loci

because the increase can be unpredictable given that errors are often not randomly distributed among PCRs, alleles and loci (Pompanon *et al.* 2005).

The benefit of amplifying shorter SNP fragments is likely to outweigh the lower variation and need to include more loci when using SNPs. For example, Camp- bell & Narum (2008) genotyped chinook salmon samples of varying quality with 13 microsatellite and 29 SNP assays and the average genotyping success for good, intermediate and poor quality samples was 98%, 97% and 79% for SNPs but only 96%, 24% and 24% for micro- satellite loci respectively. Few studies have quantified genotyping error rate using SNPs in noninvasive or historical samples. Morin & Mccarthy (2007) used 19 SNPs in a study using historical samples of bowhead whales; they found a 0.1% genotyping error rate, which is lower than most noninvasive studies.

During PCR, broken DNA fragments may anneal to each other and form priming sites needed for amplification, resulting in different sized fragments and the scoring of false alleles. To prevent this unwanted production of chimeric alleles (e.g. DNA fragments that anneal together giving the appearance of another allele) and to avoid the occurrence of jumping PCR (recombination between similar DNA sequences during PCR that is promoted in damaged/fragmented DNA), Čuljković *et al.* (2003) described a pretreatment of DNA fragments before PCR by adding a poly(A) tail at the 3¢ prime end of templates to eliminate homology between fragments. This has been successfully used in ancient DNA studies, but not to our knowledge in noninvasive studies.

c) Overcoming low DNA quantity

Several PCR-based strategies to overcome problems associated with low-quantity DNA have been proposed recently. Pre-amplification (i.e. double amplification) is an efficient procedure to increase the amount of low copy number template because products from a first amplification are used as templates for a subsequent PCR; this pre- amplification increases the DNA available for the second desired amplification (e.g. Lau *et al.* 2003). A second PCR with internal (nested) primers can also increase genotyping success and specificity to amplify only the target locus because the internal primers (as well initial external primers) can be locus specific. The same is true when using only one internal primer in the second PCR (Bellemain & Taberlet 2004). A semi-nested or second PCR can be especially useful to improve amplification of certain difficult loci. A second PCR is also useful after whole genome pre-amplification or multiplex pre-amplification.

Whole genome amplification is the production of amplicons across an entire genome to increase the amount of template DNA available for subsequent locus- specific genotyping (Kittler *et al.* 2002). This approach has been successfully applied before genotyping micro-satellites, although preferential amplification of the shorter alleles might occur. Similarly, whole-genome amplification with degenerate primers (i.e. mixtures of similar, but not identical, primers) has been successfully used for large-scale SNP genotyping despite a detectable loss in genotype accuracy (Grant *et al.* 2002). In some studies, as the one reported by Vigilant (1999) in genotyping shed chimpanzee hairs, this strategy was ineffective for improving microsatellite genotyping.

Pre-amplification of multiple loci in a multiplex can improve microsatellite genotyping from noninvasive samples (Box 1). This method can increase the quantity of target DNA fragments for each locus while minimizing consumption of the initial DNA extract. In this approach, an initial large-volume PCR with all primer pairs is per- formed followed by a second or nested PCR of each genetic marker (Piggott *et al.* 2004). The use of this two-step PCR approach revealed significant improvements in efficiency relative to standard PCR (Piggott *et al.* 2004; Hedmark & Ellegren 2005; Arandjelovic *et al.* in press). Because it requires less DNA extract, multiplex pre- amplification allows typing more loci, which is often a limitation in noninvasive genetics.

Box 1. The promise of real-time quantitative PCR

Real-time quantitative PCR quantifies the amount of target-specific, 'amplifiable' DNA from an extraction. This is important because DNA might exist in a sample (e.g. quantified by fluorometry), but not be amplifiable because of PCR inhibitors, extreme DNA fragmentation, and/or the DNA is from nontarget species. RTQ-PCR differs from regular PCR in that the PCR product is quantified as the PCR is occurring, using a fluorescent dye. In each PCR cycle, the amount of the target locus DNA doubles and so does the fluoresce intensity. An RTQ-PCR machine is a PCR machine with a fluorometer. Advantages of RTQ-PCR are its sensitivity (it is the most sensitive PCR method for low quantity of DNA) and that there is no post-PCR manipulation of samples (gel electrophoresis); this saves time and money, and avoids contamination as post-PCR tubes are never opened in the laboratory.

Real-time quantitative PCR has enormous (largely untapped) potential to improve noninvasive studies by identifying samples with enough nuclear DNA to avoid genotyping errors. The amount of DNA necessary to avoid genotyping errors (allelic dropout) has been estimated to be approximately 100–600 pg by theoretical and empirical studies (e.g. Taberlet *et al.* 1996; Morin *et al.* 2001). RTQ-PCR could improve noninvasive studies by excluding extremely low quality samples and identifying samples at risk of having genotyping errors.

A single RTQ-PCR can identify species in addition to quantifying amplifiable DNA (Berry & Sarre 2007). Species identification is possible if species-specific primers are used or if the targeted PCR product has a different melting curve (Berry & Sarre 2007). RTQ-PCR could replace species identification methods, which often involve mtDNA analysis and that currently are the standard first step in many noninvasive studies (e.g. Swango *et al.* 2006).

The first paper using RTQ-PCR on noninvasive samples was Morin *et al.* (2001). Subsequently, the same RTQ-PCR was used on ape faeces to identify factors (e.g. temperature) and sample preservation methods (ethanol and silica) that improve PCR amplification. Several recent papers report successful RTQ-PCR of DNA from faeces and urine, although most papers involve testing for cancer genes or disease pathogens in humans or livestock (e.g. Inglis & Kalischuk 2004; Queipo-Ortuño *et al.* 2006; Itzkowitz *et al.* 2007). These recent papers are highly encouraging and suggest that RTQ-PCR from faeces and urine is highly feasible and efficient.

We expect that RTQ-PCR will be widely used in future noninvasive studies because the methods have become easier (e.g. with commercial kits), less expensive, and clearly work on noninvasive samples

(Hausknecht *et al.* 2007). An RTQ-PCR reaction can cost as little as approximately US$1 per PCR (e.g. Berry & Sarre 2007). The cheapest RTQ-PCR method (SYBR green) is also often highly reliable (e.g. Smith *et al.* 2002). An RTQ-PCR machine costs approximately US$15 000–30 000 and prices are likely continue to fall (e.g. see http://www.biocompare.com/matrix/2838/Real-Time-PCR-ermalCyclers(Thermocyclers).html).

However, multiplex pre-amplification has drawbacks. Allelic dropout can occur more frequently than for conventional PCR, suggesting that this type of error is often generated during the first-step multiplex (Lampa *et al.* 2008). In addition, a multiplex might increase the proportion of nonamplifiable loci because of the competition between loci (Lampa *et al.* 2008). Alternatively, genotypes can be obtained by performing additional single standard PCR whenever single locus amplification remains the most suitable approach to satisfy efficiency and accuracy (Parsons 2001). Although nested PCR increases the efficacy and sensitivity for amplifying target genomic fragments, it has the drawback of increasing the risk of contamination, because it requires two PCR reactions and, consequently, doubles the handling of materials. This problem might be particularly prominent for noninvasive studies.

d) Overcoming non-specific amplification and contamination

Co-amplification of nonspecific products and contamination can be major problems in noninvasive genetics. PCR with low quality and quantity target DNA can increase the probability of amplifying nontarget regions. It also increases the probability that contaminant DNA is at similar or higher concentrations than target DNA (Pompanon *et al.* 2005). Navidi *et al.* (1992) estimated that sporadic contamination could cause up to 7% error in large-scale studies, and Buchan *et al.* (2005) estimated that 1.3% of the baboon DNA analyzed and 1.2% of the negative controls of their study were contaminated with human DNA.

Hot start PCR is one of the most effective means to improve specificity, fidelity and sensitivity in DNA amplifications. Effective protocols are now widely avail- able thanks to the use of engineered thermostable polymerases (whether using an inhibitor antibody or chemical modification) that require heat activation prior to PCR cycling, and because of the use of high-performance PCR buffers with optimized combinations of salts and additives (e.g. Radstrom *et al.* 2008). *Taq* polymerases such as AmpliTaq GoldTM (Applied Byosistems), Fast-Start *Taq* DNA Polymerase (Roche), Platinum® *Taq* DNA polymerase (Invitrogen), TrueStart *Taq* DNA Polymerase (Fermentas), AccuSure DNA Polymerase (Bioline), Phusion High-Fidelity DNA Polymerase (Finnzymes) are a list of good examples (see Box 2).

DNA is present everywhere in a laboratory, especially where PCRs are frequently performed because amplified fragments persist as aerosols. Design of species-specific primers reduces the risk of amplification of nonspecific fragments and external DNA from human, prey items or bacteria (particularly in faecal material). Primers that do not amplify nontarget species (e.g. humans) can be

designed. This is increasingly feasible thanks to increasing availability of sequence data from many species and software programs to align and compare multiple sequences.

Improved primer design with conventional software, such as Primer 3 and a number of later adaptations (Rozen & Skaletsky 2000; Kim & Lee 2007; Koressaar & Remm 2007), PERLPRIMER (Marshall 2007) or SNPBOX (Weckx *et al.* 2005) and highly specific multiplex primer design tools are now available on the web. The server Primerstation for the human genome (http://ps.cb.k.u-tokyo.ac.jp Yamada *et al.* 2006), the program MULTIPLX (Kaplinski *et al.* 2005) and the packages PRIMO (from BioToolKit 320; Chang Bioscience) and PrimerPremier (PREMIER Biosoft) are examples of effective ways for designing specific primers in large-scale analyses.

Box 2. Polymerase enzymes for PCR

Presently, there are several hundred companies selling over 20 kinds of polymerase enzymes. There are two main characteristics that a polymerase enzyme must have that are crucial for amplifying small amounts of DNA: fidelity and 3' → 5' exonuclease activity (proofreading). Fidelity is particularly important when sequencing to detect SNP's. Heterozygous nucleotide sites must be unambiguously identified (in diploid individuals) or, for example, the false discovery rate of SNPs might be high.

Proofreading with 3' → 5' exonuclease activity is lacking in some polymerases [e.g. in *Thermus aquaticus* (*Taq*)] and sequencing error rates are higher than for polymerases with exonuclease activity [e.g. isolated from *Pyrococcus furiosus* (*Pfu*), *Thermococcus litoralis* (*Vent*), *Pyrococcus woesei* (*Pwo*)], which are often designated as high-fidelity polymerases. Studies comparing regular *Taq* polymerase vs. high-fidelity polymerases, such as the *Pfu*, report far lower error rates for the high-fidelity enzymes (Hansen *et al.* 2001).

Microsatellite genotyping with high-fidelity polymerases also gives lower error rates (Hite *et al.* 1996). When genotyping microsatellite loci (mostly dinucleotide), annoying stutter products are often formed during the PCR amplification. The primary cause of 'stutter' bands is a change in the number of repeat units because of slip-strand extension by *Taq* DNA polymerase. However, the use of high-fidelity polymerases (e.g. *Pfu*, *Vent*) reduces the formation of stutters as 3' → 5' exonuclease activity removes 3' nontemplate nucleotides (Hite *et al.* 1996).

A study testing different polymerase enzymes (Spitaleri *et al.* 2004), showed that, for low template quantities, the regular *Taq* polymerases perform poorly and, for example, can increase allele dropout rates. However, in the same study, the engineered polymerases (e.g. AmpliTaq Gold) maintained high fidelity and sensitivity at very low DNA concentrations.

Amplification performance is another important characteristic. In this respect, it is well demonstrated that engineered DNA polymerases perform much better with low quality DNA. This is mainly because engineered DNA polymerases allow for the PCR hot-start technique. Hot start greatly increases the specificity and sensitivity of DNA amplification by avoiding competing side reactions during pre-PCR setup that can be initiated the moment that all reactants have been mixed and misprimming occurs.

At least two kinds of inactive polymerases are presently commercialized and often used in noninvasive studies: (i) recombinant DNA polymerase (e.g. AmpliTaq Gold® *Taq* DNA Polymerase; Roche Molecular Systems) engineered to be activated at temperatures higher than 90°C, and (ii) Anti-*Taq* DNA polymerase antibodies, which inhibit polymerase activity at room temperature (e.g. Platinum® *Taq* DNA Polymerase; Invitrogen).

Precautions such as those in ancient DNA laboratories should be followed to prevent and monitor for contamination. Gilbert *et al.* (2005) describe nine criteria for working with ancient DNA and categorize risk factors associated with different projects. The criteria include isolation of work areas, use of negative controls for extractions and amplifications, amplification of only small segments, reproducibility, use of cloning of products to assess damage and contamination, independent replication, preservation of co-occurring biomolecules, quantification of DNA and evaluation of associated remains. They also consider hominid projects being the riskiest, followed by projects on cultivars and domestic animals, with low-risk projects involving projects on other wildlife species. Among the most important precautions, PCR set up should never be performed in the same day or just after conducting PCR or entering a room with PCR machines or post-PCR samples (see Fig. 1). Amplifying additional loci that work in possible contaminant species might also allow identifying contamination that remains undetected in the analysis of the target markers (e.g. Wandeler *et al.* 2003). For example, because of the high copy number of mitochondrial molecules, using mitochondrial specific primers in both samples and controls may be a sensitive way to monitor for contamination when working with nuclear DNA (Pusch *et al.* 1998).

Design of PCR protocols that minimize manipulation can reduce contamination risk. One could, for example, develop multilocus assays to successfully work using the minimum number of single-tube reactions, as it would imply less manipulation for higher quantity of data produced per sample. RTQ-PCR has no post-PCR handling (e.g. gel electrophoresis) and so tubes are not opened after PCR, which minimizes DNA molecules in the laboratory (Nazarenko *et al.* 2002). RTQ-PCR also allows the real- time monitoring of target DNA amplification (Box 1) as well as direct scoring of the desired results (e.g. melt curve analysis, which can detect nontarget amplification).

Negative controls are essential to monitor contamination. Several blanks should be placed in the beginning (to monitor for environmental and/or reagents contaminations) and in the middle and end (to detect cross-contamination) of a series of samples (Borst *et al.* 2004). Minimizing PCR cycles (e.g. to 35 cycles) can reduce contamination risks because tiny amounts of contamination would unlikely lead to visible PCR products on electropherograms or gels. Human forensic laboratories typically limit their PCR cycles to <35. However, this can be problematic for degraded DNA samples, which can require 40–45 PCR cycles.

Mixed samples can cause errors in noninvasive genetics but can be detected and avoided using recent techniques and software (Roon *et al.* 2005). Great efforts are made to solve problems of DNA mixtures because more than one donor is frequently responsible for the material recovered from a

forensic scene (e.g. in a rape, DNA from the victim and the aggressor might be collected simultaneously). In this context, novel computational programs have been developed to separate admixed genotypes, such as PENDULUM (Bill *et al.* 2005) or MAIES (Cowell *et al.* 2006) that are based on different models to analyze peak area values on electropherograms. DNA mixture should not be regarded as a major limitation, because, if > 6–8 highly polymorphic microsatellites are genotyped, it is likely that some loci will have three alleles, which is impossible for diploid species, and thus would indicate possible contamination. Many wildlife and conservation based studies that identify mixed samples simply discard these samples in favour of those that indicate only one animal deposited the sample.

POST-PCR AND GENOTYPING ERRORS

The most insidious problem in noninvasive genetics is genotyping errors. We define a genotyping error as a difference between the true genotype and the inferred genotype (Pompanon *et al.* 2005; Luikart *et al.* 2008b), which does not include failed PCRs or failed DNA extractions. Amplification failure (no PCR product) is not as problematic as a genotyping error (erroneous genotype) because mistakes in data interpretation are less likely from failed PCRs. Genotyping error detection and avoidance (e.g. by using the multi-tubes approach) have been thoroughly reviewed elsewhere (Pompanon *et al.* 2005), and so below we summarize and update the available information, and highlight the main problems and ways to avoid them.

Three main kinds of genotyping errors are generally reported as: (i) allelic dropout (stochastic detection of false homozygotes at heterozygous loci because of failure of one allele to amplify), (ii) false alleles (creation of new alleles caused by slippage events of *Taq* polymerase during early cycles of PCR, that may reach a concentration similar to the authentic alleles when limited template exists), and (iii) Human error, the incorrect identification of alleles as a result of cross-contamination in the field or in the laboratory or database manipulation errors (Hoffman & Amos 2005; Pompanon *et al.* 2005). Human errors in data entry and manipulation (e.g. in spread sheets) are often the most frequent cause of genotyping errors (Paetkau 2003; Schwartz *et al.* 2006). Among the nonhuman-induced errors, allelic dropout is usually the most common error. Extremely dissimilar error rates (depending on species, season of the year and sample type) have been documented, ranging from as low as 0–2% in faecal analysis (Bonin *et al.* 2004; Maudet *et al.* 2004) and 10% in human buccal samples (Whitaker *et al.* 2001) to approximately 24% in some carnivore faeces (Johnson & Haydon 2007), and over 30% in shed hairs (Gagneux *et al.* 1997). However, comparison of rates is challenging as some laboratories are more conservative in discarding samples, while others readily discard samples that show even the slight-est sign of failure. These decisions dramatically change the reported error rate.

There are four main approaches used to handle genetic errors from noninvasive samples. The first and the most common is called the multiple tubes approach first developed by Navidi *et al.* (1992) and Taberlet *et al.* (1996), which suggests that 6–10 similar genotypes should be obtained for a locus to define an individual as homozygous or heterozygous (see also Miller *et al.* 2002). Here, each sample

at each locus is run multiple times to ensure genotype consistency. Some form of this approach is used in almost every noninvasive study. However, while multi-tubing will detect genotyping errors, it can exhaust the DNA extracted and is fiscally expensive. In addition, the multi-tube approach may increase errors as samples are handled more often (inducing human error) and there are more chances to produce false alleles, which can be interpreted as a missing allele (allelic dropout). In addition, multi-tubing does nothing to prove that the existing database is error free. A second approach is to quantify the amount of target, amplifiable nuclear DNA in the sample (Morin *et al.* 2001). Once this quantity is known, the appropriate number of multi-tube re-runs can be conducted. Morin *et al.* (2001) recommended that if a sample has <25 pg (of amplifiable DNA) per reaction, it should be discarded; if it has 26–100 pg per reaction, then seven repeat genotypings of the sample are necessary; if it has 101–200 pg per reaction then four repeats are required; and if >200 pg per reaction, only two repeats are necessary (see also Box 1). A third approach has been to use computer algorithms to detect genotyping errors. Depending on the data and goal of the study, various algorithms have been suggested (Ewen *et al.* 2000; Miller *et al.* 2002; Valière 2002; Van Oosterhout *et al.* 2004; McKelvey & Schwartz 2005; Kalinowski 2006). Some of these examine deviations from Hardy–Weinberg proportions, others use pedigree information to catch errors, while others use the number of mismatches in recaptures (i.e. genotypes identified more than once and differing by only one or two alleles; McKelvey & Schwartz 2005) as an error signal. A recent paper suggests that sample-specific errors (only a few poor quality individual samples) can cause significant deviations from Hardy–Weinberg proportions; such samples should be identified and often discarded (Miquel *et al.* 2006). Some of the most widely used soft- ware tools for detecting and avoiding genotyping errors are provided in Table 2.

Table 2: Some examples of the most widely used methods and software programs developed mainly for detecting and preventing genotyping errors.

	Main Functions					
Software	*Identifying problematic samples*	*Estimating the number of multitube repeats*	*Identifying problematic loci (allelic dropout, null alleles)*	*Testing for HWE departures*	*Identifying mixed samples*	**References**
Quality Indexes		✓	✓			Miquel *et al* 2006 (1)
Gemini (2)	✓	✓	✓		✓	Valiére *et al* 2002
Hw- Quickcheck				✓		Kalinowski, 2006
Pedmanager (3)	✓		✓	✓		Ewen *et al* 2000
Cervus				✓	✓	Marshall *et al* 1998
Gimlet	✓		✓		✓	Valiére, 2002
Reliotype	✓		✓		✓	Miller *et al* 2002
Micro-Checker			✓	✓	✓	Van Oosterhout *et al* 2004
Dropout (4)	✓		✓		✓	Mckelvey & Schwartz 2005

(1) Program available upon request from the authors. (2) Simulation based method to detect consensus genotypes; (3) When pedigree information is available; (4) Bimodal test (for loci that cause many samples to differ by only one allele)

The fourth error handling approach is to model various error rates in the final statistical analysis. For example in capture–mark–recapture studies, Lukacs & Burnham (2005) derived a

method to incorporate the probability of genotyping error into the closed-population models of Otis *et al.* (1978), Huggins (1989) and Pledger (2000) using the disproportionate number of genotypes collected once relative to genotypes collected more frequently to estimate error. These approaches have been developed for estimating animal abundance, but are relatively rare in population genetic studies. Another example is in parentage studies where accommodating genotyping errors during likelihood computations can improve paternity analyses, as has been shown using the software Cervus (Kalinowski *et al.* 2006). In a related study, Wang (2004) developed likelihood methods to infer full- and half-sibships from marker data with a high error rate and to identify typing errors at each locus in each reconstructed sib family.

It is important to note that blood and tissue samples are too often assumed to always yield low genotyping error rates. However, error rates can be substantial if these (normally high quality) samples are poorly preserved (Hoffman & Amos 2005). Comparative analysis of genotyping errors for noninvasive and assumed good quality DNA is helpful and needed (Soulsbury *et al.* 2007), but should be interpreted with caution.

Regarding this, Johnson & Haydon (2007) developed a maximum-likelihood-based method for estimating error rates from a single replication of a sample of genotypes. Simulations show it to be accurate and robust. It is implemented in a computer program, PENDANT, which estimates allelic dropout and false allele error rates with 95% confidence regions from microsatellite genotype data and performs power analysis. Finally, as mentioned in the previous section, mixed samples (with DNA from more than one individual) can be identified and computational programs have been developed to resolve genotypes, such as PENDULUM (Bill *et al.* 2005) or MAIES (Cowell *et al.* 2006).

PERSPECTIVES

The most promising areas for future research and development in noninvasive genetic studies involve large-scale PCR multiplexing techniques, massively parallel sequencing technologies, and more holistic studies including diet and parasite or disease analyses. Future multiplexing techniques should allow analysis of tens to hundreds of loci (Porreca *et al.* 2007; Meyer *et al.* 2008) on noninvasive samples (see also Box 3). This would vastly increase the statistical power of noninvasive approaches and facilitate use of massively parallel sequencing while making possible the targeted sequenc- ing of interesting segments of the genome (e.g. exons under selection).

New SNP multiplex genotyping systems use tiny volumes (nanolitres) for SNP genotyping assays (e.g. TaqMan; ABI), which reduces the costs of reaction chemicals by nearly 98%, while automating and speeding up the genotyping process. For example, a new multiplex system using SNP chips from Fluidigm at BioMark (http://www.fluidigm.com/applications/genotype-profiling.html) allows simultaneous genotyping of 48 or 96 SNP loci on each of 48 or 96 individuals at a cost of only US$0.10–0.20 per SNP (Perkel 2008). These systems, however, also require an initial investment in equipment often of the order of US$50 000–300 000.

Massively parallel sequencing technologies, e.g. 454 pyrosequencing by synthesis, and sequencing by ligation (Ellegren 2008; Shendure & Li 2008), should improve noninvasive studies because they work well on short DNA fragments typical of difficult and ancient DNA (Green *et al.* 2006). The main disadvantage of these sequencing technologies is that they do not allow easy sequencing of many individuals (samples), and the cost per sequencing run is thousands of dollars. However, costs are declining and clever study design can allow an entire study to be conducted on a single sequencing run, thereby minimizing total costs.

Box 3. Multiplex PCR techniques

Multiplex PCR amplification has great untapped potential to improve noninvasive sampling by reducing cost, increas- ing speed and reducing consumption of DNA from typically low quantity sources (Henegariu *et al.* 1997; Butler 2005). Reducing manipulation and handling (fewer PCRs per individual sample) also minimizes the possibility of contami- nation and error during reaction setup.

Optimization of multiplex assays generally requires more time and effort than standard single plexes, because it involves designing primer pairs that do not interact and at the same time anneal under the same conditions. Optimiza- tion also sometimes requires, adjusting primer pair concentrations to give similar amounts of PCR product, choosing fluorescence labels for sets of loci according to their allele or size range, and combining all these aspects in an efficient and low-cost protocol (e.g. Guo & Milewicz 2007). Whenever possible, loci more difficult to amplify should be labelled with the highest energetic labels (e.g. blue fluoresces brighter than red). Once obtained, multiplexes greatly facilitate genotyping of large population samples rapidly and at reduced cost.

In forensics and noninvasive genetic studies, multiplex PCR is being used more for both microsatellites and SNPs (Morin & McCarthy 2007). Rapid and economical multiplex assays also exist for monitoring the international trade of protected species; for example, a multiplex of several species-specific primers allows the distinction among shark spe- cies (Shivji *et al.* 2005; Magnussen *et al.* 2007). Multiplexes have also been designed to study natural animal popula- tions, e.g. a multiplex of 14 microsatellites in one PCR was developed for racoon, *Procyon lotor*, Fike *et al.* (2007).

Three main issues can facilitate multiplex PCR on noninvasive samples: (i). Recently developed commercial kits can facilitate co-amplification of 5–10 loci or more (Luikart *et al.* 2008b). These kits include a new buffer that reduces com- petition among loci and improves primer annealing. Multiplex PCR can be >30% cheaper than standard singleplex (Mukherjee *et al.* 2007); (ii) The use of algorithms and software to design improved primer sets with no primer interac- tions (Kaderali *et al.* 2003; Vallone & Butler 2004); and (iii) The use of universal fluorescent tails on the 5¢ end of primers to label PCR products (Oetting *et al.* 1995; Neilan *et al.* 1997). Fluorescent labelling of one primer in a pair is expensive, ranging between US$100 and 150 (Schuelke 2000).

To reduce costs, Oetting *et al.* (1995) developed a single reaction nested PCR that allows easy and consistent geno- typing and more homogeneous PCR amplification among loci. For each locus, PCR includes three different primers: a reverse primer, a forward primer with a 5' tail (e.g. M-13 sequence), and the universal M-13 primer with fluores- cent-labelling. During the first PCR cycles, the

forward primer with tail hybridizes with the target DNA fragments and is incorporated into the products, and then temperature is lowered (53ºC) to allow the universal tail to anneal and incorporate fluorescence to the subsequent PCR products. With this technique, one can synthesize and use one labelled forward primer (M-13) for each of several loci in a multiplex PCR (Missiaggia & Grattapaglia 2006). At the same time, PCR multiplex amplification will be facilitated as the same forward primer (M-13) can give more even amplification among loci and provide better results for low template DNA (Schuelke 2000). Laboratories studying many species can benefit a lot from using a common universal labelled tail or tails. A cost reduction of 40% can be achieved in the amplification of 10 microsatellites when compared with conventional methods (Missiaggia & Gratt- apaglia 2006).

Most studies use the M13 sequence as the universal tail, but any sequence with no complementarity to target gen- ome could be used (Neilan *et al.* 1997). For multiplexing several loci where some of them have overlapping size ranges, one can optimize the PCR reaction using different fluorescent tails (Missiaggia & Grattapaglia 2006; Guo & Milewicz 2007).

Single nucleotide polymorphism multiplex assays can work well using low quantity DNA, for example, 50 pg (Onofri *et al.* 2006). Mini-STRs (up to 150 bp) have also been penta-plexed revealing detection limits of 12.5 pg for arti- ficially degraded human DNA (Meissner *et al.* 2007). In noninvasive wildlife studies, multiplex PCR is not widely used. However, Mukherjee *et al.* (2007) developed a multiplex protocol to identify tiger species from faeces using three small mtDNA fragments. The multiplex had a significant decrease in the number of false negatives compared with conventional PCR (especially in old faeces).

Increasingly holistic noninvasive genetic studies are possible. They combine multiple kinds of information (e.g. on diet, parasite load, parasite population genetics, as well as host genetics and physiological status [stress and reproductive hormone secretions]) allowing more valuable studies addressing multiple questions or by providing more complete information on individuals allowing new questions to be addressed. Valentini *et al.* (2009) used 454 pyrosequencing on chloroplast DNA from faecal samples to determine the diet of bears (as well as birds, snails and grasshoppers). They showed that DNA-based faecal diet analysis using universal primers (e.g. DNA barcoding) and pyrosequencing can help determine what plant species are consumed by an individual. In the future, noninvasive collection of spatially referenced faeces from across a landscape could allow a comprehensive study of a species (e.g. bears) in an area by the enumeration of individuals, identification of gender, examination of diet, estimation of parasite load and parasite transmission patterns among individuals and geographic areas. This type of information could become crucial to the management of species and their habitat.

CONCLUDING REMARKS

Application of noninvasive genetic approaches is exciting and promising. The power and role of noninvasive genetics in molecular ecology and conservation genetics will continually increase, thanks to the advancements in each step of a noninvasive study (Fig. 1) including new technologies (e.g.

massively parallel sequencing) and advancements from different disciplines (e.g. human and livestock health, and forensics). Nonetheless, noninvasive genetic studies still usually require more funding and efforts in the laboratory, compared with traditional genetic studies with high-quality DNA, to ensure low genotyping error rates. Monitoring the efficacy and error rate associated with each of the multiple steps in a noninvasive study is crucial to ensure success.

Among the greatest needs for additional research is to directly compare the relative performance of new and improved methods (e.g. for sample storage, DNA extraction and amplification) in multiple independent laboratories, taxa and sample types. The lack of independent and quantitative comparisons of techniques makes it difficult to provide advice on which methods are best for a given species, sample type and sample conditions (but see Schwartz & Monfort 2008, p. 242). Some techniques might be species-specific and environment dependent, but more studies are needed to assess this issue.

Research questions, including those that could be addressed previously only using high-quality samples, can now be addressed using noninvasive genetics, thanks to lower error rates and our ability to analyze more loci and more samples. For example, in many natural populations, it is increasingly feasible to estimate relatedness, infer parentage and reconstruct pedigrees, all of which require many loci and low genotyping error rates. Genetic monitoring (Schwartz *et al.* 2007), defined as the quantification of temporal changes in DNA-based estimators (e.g. population abundance or effective size), is also becoming more feasible because more samples can be genotyped with more loci, thereby increasing statistical power to detect reduced variation, changes in population size and immigration. In addition, noninvasive genetics continually improves the ability of law enforcement to detect illegal trafficking of animals (e.g. Manel *et al.* 2002) by providing more representative samples across populations and increasing recovery of DNA from confiscated samples.

We are on the cusp of answering long-standing ecological and evolutionary questions in rare and elusive species, thanks to improved noninvasive sampling and new technologies for analyzing short DNA fragments (Morin & McCarthy 2007; Millar *et al.* 2008). This includes questions about the genetic basis of local adaptation that can be addressed by using genome-wide scans (Wiehe *et al.* 2006) and population genomic approaches (Luikart *et al.*, 2003) requiring genotyping of many loci, which is becoming feasible in noninvasive genetics. It also includes questions about how landscape features influence gene flow and dispersal in natural populations, which is a main goal of land- scape genetics, an emerging approach that combines landscape ecology and population genetics (Manel *et al.* 2003). Landscape genetics typically requires analyzes of hundreds of samples widely dispersed across landscapes; this is feasible only via noninvasive genetic approaches in some taxa.

In disease ecology, we will be able to estimate trans- mission rates and address questions about landscape features or environmental variables influencing disease spread, by noninvasively sampling of parasites (or parasite DNA) from hosts (Archie *et al.* 2009). For example, many microparasites (bacteria and viruses) and macro- parasites (helminthes) are environmentally transmitted (shed into the environment) and can be obtained from faeces, urine or saliva. We can even conduct population

genomic studies on parasites (e.g. to identify genes influencing transmission or virulence) for wildlife disease that are notoriously difficult to study because they require capture of many individuals, which is difficult or impossible, as described earlier for elusive, rare or dangerous wildlife species.

Overall, the recent boom in technological advances is rapidly advancing the relatively new field of noninvasive genetics. These new technologies are often derived from human-based fields such as medicine and genomics. The challenge for molecular ecologists will be keeping up with and integrating these rapidly changing fields and technologies to aid in the study and monitoring of wild populations.

ACKNOWLEDGEMENTS

This review was developed during weekly international journal club using Skype (Skype.com) for discussions between researchers at the University of Montana in the USA and the University of Porto in Portugal. We thank to Kristy Pilgrim, Steve Amish and three anonymous reviewers for constructive comments on this manuscript. AB-P and RO were supported by FCT grants SFRH/BPD/17822/2004) and (SFRH/BD/24361/2005), respectively. GL was supported by the Portuguese American Foundation for Development, CIBIO, and UP. MKS was supported to work on this project by a Presidential Early Career Award for Science and Engineering.

REFERENCES

Adams J, Waits L (2007) An efficient method for screening faecal DNA genotypes and detecting new individuals and hybrids in the red wolf (*Canis rufus*) experimental population area. *Conservation Genetics* 8, 123-131.

Anzai-Kanto E, Hirata MH, Hirata RDC*, et al.* (2005) DNA extraction from human saliva deposited on skin and its use in forensic identification procedures. *Brazilian Oral Research* 19, 216-222.

Baker CS, Cooke JG, Lavery S*, et al.* (2007) Estimating the number of whales entering trade using DNA profiling and capture-recapture analysis of market products. *Molecular Ecology* 16, 2617-2626.

Balkiz O, Dano S, Barbraund C*, et al.* (2007) Sexing Greater Flamingo Chicks from Feather Bulb DNA. *Waterbirds* 30, 450-453.

Bellemain E, Swenson JE, Tallmon D, Brunberg S, Taberlet P (2005) Estimating population size of elusive animals with DNA from hunter-collected feces: four methods for brown bears. *Conservation Biology* 19, 150-161.

Bellemain E, Taberlet P (2004) Improved noninvasive genotyping method: application to brown bear (*Ursus arctos*) faeces. *Molecular Ecology Notes* 4, 519-522.

Berry O, Sarre SD (2007) Gel-free species identification using melt-curve analysis. *Molecular Ecology Notes* 7, 1-4.

Bhagavatula J, Singh L (2006) Genotyping faecal samples of Bengal tiger *Panthera tigris tigris* for population estimation: A pilot study. *BMC Genetics* 7, 48.

Bill M, Gill P, Curran J, *et al.* (2005) PENDULUM-a guideline-based approach to the interpretation of STR mixtures. *Forensic Science International* 148, 181-189.

Björnerfeldt S, Vilà C (2007) Evaluation of methods for single hair DNA amplification. *Conservation Genetics* 8, 977-981.

Blejwas KM, Williams CL, Shin GT, McCullough DR, Jaeger MM (2006) Salivary DNA Evidence Convicts Breeding Male Coyotes of Killing Sheep. *Journal of Wildlife Management* 70, 1087-1093.

Bonin A, Bellemain E, Eidensen PB, *et al.* (2004) How to track and assess genotyping errors in population genetic studies. *Molecular Ecology* 13, 3261-3273.

Boom R, Sol CJ, Salimans MM, *et al.* (1990) Rapid and simple method for purification of nucleic acids. *Journal of Clinical Microbiology* 28, 495-503.

Borst A, Box ATA, Fluit AC (2004) False-positive results and contamination in nucleic acid amplification assays: suggestions for a prevent and destroy strategy. *European Journal of Clinical Microbiology and Infection Diseases* 23, 289-299.

Bradley BJ, Doran-Sheehy DM, Vigilant L (2007) Potential for female kin associations in wild western gorillas despite female dispersal. *Proceedings of the Royal Society of London B* 274, 2179-2185.

Bremner-Harrison S, Harrison SWR, Cypher BL, Murdock JD, Maldonado J (2006) Development of a single-sampling noninvasive hair snare. *Wildlife Society Bulletin* 34, 456–461.

Broquet T, Berset-Braendli L, Emaresi G, Fumagalli L (2007a) Buccal swabs allow efficient and reliable microsatellite genotyping in amphibians. *Conservation Genetics* 8, 509-511.

Broquet T, Menard N, Petit E (2007b) Noninvasive population genetics: a review of sample source, diet, fragment length and microsatellite motif effects on amplification success and genotyping error rates. *Conservation Genetics* 8, 249-260.

Broquet T, Petit E (2004) Quantifying genotyping errors in noninvasive population genetics. *Molecular Ecology* 13, 3601-3608.

Buchan JC, Archie EA, Van Horn RC, Moss CJ, Alberts SC (2005) Locus effect and sources of error in noninvasive genotyping. *Molecular Ecology Notes* 5, 680-683.

Bush K, Vinsky M, Aldridge C, Paszkowski C (2005) A comparison of sample types varying in invasiveness for use in DNA sex determination in an endangered population of greater Sage-Grouse (*Centrocercus uropihasianus*). *Conservation Genetics* 6, 867-870.

Butler J, Shen Y, McCord B (2003) The development of reduced size STR amplicons as tools for analysis of degraded DNA. *Journal of Forensic Science* 48, 1054-1064.

Butler JM (2005) Constructing STR multiplex assays. *Methods in Molecular Biology* 297, 53-66.

Casper RM, Jarman SN, Deagle BE, Gales NJ, Hindell MA (2007) Detecting prey from DNA in predator scats: a comparison with morphological analysis, using Arctocephalus seals fed a known diet. *Journal of Experimental Marine Biology and Ecology* 347, 144-154.

Chakraborty A, Sakai M, Iwatsuki Y (2006) Museum fish specimens and molecular taxonomy: A comparative study on DNA extraction protocols and preservation techniques. *Journal of Applied Ichthyology* 22, 160-166.

Chu J-H, Wu H-Y, Yang J, Takenaka O, Lin Y-S (1999) Polymorphic microsatellite loci and low-invasive DNA sampling in *Macaca cyclopis*. *Primates* 40, 573-580.

Chung DT, Drábek J, Opel KL, Butler JM, McCord BR (2004) A study on the effects of degradation and template concentration on the amplification efficiency of the STR miniplex primer sets. *Journal of Forensic Science* 49, 733-740.

Cowell RG, Lauritzen SL, Mortera J (2006) MAIES: A tool for DNA mixture analysis 22nd Conference on Uncertainty in Artificial Intelligence (UAI), 90-97.

Creel S, Spong G, Sands JL*, et al.* (2003) Population size estimation in Yellowstone wolves with error-prone noninvasive microsatellite genotypes. *Molecular Ecology* 12, 2003-2009.

Cross N, Pines MK, Rogers LJ (2004) Saliva sampling to assess cortisol levels in unrestrained common marmosets and the effect of behavioral stress. *American Journal Primatology* 62, 107-114

Čuljković B, Savić D, Stojković O, Romac S (2003) Poly(A) tailing of ancient DNA: a method for reproducible microsatellite genotyping. *Analytical Biochemistry* 318, 124-131.

Deagle BE, Gales NJ, Evans K*, et al.* (2007) Studying Seabird Diet through Genetic Analysis of Faeces: A Case Study on Macaroni Penguins (*Eudyptes chrysolophus*). *PLoS ONE* 2, e831.

Dreher BP, Winterstein SR, Scribner KT*, et al.* (2007) Noninvasive Estimation of Black Bear abundance Incorporating Genotyping Errors and Harvested Bear. *Journal of Wildlife Management* 71, 2684-2693.

Durnin ME, Palsbøll PJ, Ryfer OA, McCullough DR (2007) A reliable genetic technique for sex determination of giant panda (*Ailuropoda melanoleuca*) from non-invasively collected hair samples. *Conservation Genetics* 8, 715-720.

Ellegren H (2008) Sequencing goes 454 and takes large-scale genomics into the wild. *Molecular Ecology* 17, 1629-1631.

Ellison SR, English CA, Burns MJ, Keer JT (2006) Routes to improving the reliability of low level DNA analysis using real-time PCR. *BMC-Biotechnology* 6, 33.

Elphinstone MS, Hinten GN, Anderson MJ, Nock CJ (2003) An inexpensive and high-throughput procedure to extract and purify total genomic DNA for population studies. *Molecular Ecology Notes* 3, 317-320.

Epps CW, Palsbøll PJ, Wehausen JD, Roderick GK, McCullough DR (2006) Elevation and connectivity define genetic refugia for mountain sheep as climate warms. *Molecular Ecology* 15, 4295-4302.

Ewen KR, Bahlo M, Treloar SA*, et al.* (2000) Identification and analysis of error types in high-throughput genotyping. *American Journal of Human Genetics* 67, 727-736.

Farrell LE, Roma J, Sunquist ME (2000) Dietary separation of sympatric carnivores identified by molecular analysis of scats. *Molecular Ecology* 9, 1583-1590.

Feinstein J (2004) DNA sequence from butterfly frass and exuviae. *Conservation Genetics* 5, 103-104.

Ficetola GF, Miaud C, Pompanon F, Taberlet P (2008) Species detection using environmental DNA from water samples. *Biology letters* 4, 423-425.

Fike JA, Drauch AM, Beasley JC, Dharmarajan G, Rhodes OE (2007) Development of 14 multiplexed microsatellite loci for raccoons *Procyon lotor*. *Molecular Ecology Notes* 7, 525-527.

Frantz AC, Pope LC, Carpenter P*, et al.* (2003) Reliable microsatellite genotyping of the Eurasian badger (*Meles meles*) using faecal DNA. *Molecular Ecology* 12, 1649-1661.

Gagneux P, Boesch C, Woodruff DS (1997) Microsatellite scoring errors associated with noninvasive genotyping based on nuclear DNA amplified from shed hair. *Molecular Ecology* 6, 861-868.

Goossens B, Abdullah ZB, Sinyor JB (2004) Which Nests to Choose: Collecting Shed Hairs from Wild Orang-Utans. *Folia Primatologica* 75, 23-26.

Goymann W (2005) Noninvasive Monitoring of Hormones in Bird Droppings: Physiological Validation, Sampling, Extraction, Sex Differences, and the Influence of Diet on Hormone Metabolite Levels. *Annals of the New York Academy of Science* 1046, 35-53.

Grant SFA, Steinlicht S, Nentwich U*, et al.* (2002) SNP genotyping on a genome-wide amplified DOP-PCR template. *Nucleic Acids Research* 30, e125.

Green ML, Herzing DL, Baldwin JD (2007) Noninvasive methodology for the sampling and extraction of DNA from free-ranging Atlantic spotted dolphins (*Stenella frontalis*). *Molecular Ecology Notes* 7, 1287-1292.

Green RE, Krause J, Ptak SE*, et al.* (2006) Analysis of one million base pairs of Neanderthal DNA. *Nature* 444, 330-336.

Guo D-C, Milewicz DM (2007) Pyrosequencing protocols. In: *Methods in Molecular Biology* eds. Walker JM, Marsh S), pp. 57-62. Humana Press.

Hajkova P, Zemanova B, Bryja J, *et al.* (2006) Factors affecting success of PCR amplification of microsatellite loci from otter faeces. *Molecular Ecology Notes* 6, 559-562.

Handel CM, Pajot LM, Talbot SL, Sage GK (2006) Use of buccal swabs for sampling DNA from nestling and adult birds. *Wildlife Society Bulletin* 34, 1094-1100.

Hansen AJ, Willerslev E, Wiuf C, Mourier T, Arctander P (2001) Statistical evidence for miscoding lesions in ancient DNA templates. *Molecular Biology and Evolution* 18, 262-265.

Hausknecht R, Gula R, Pirga B, Kuehn R (2007) Urine - a source for noninvasive genetic monitoring in wildlife. *Molecular Ecology Notes* 7, 208-212.

Hayakawa S, Takenaka O (1999) Urine as another potential source for template DNA in polymerase chain reaction (PCR). *American Journal of Primatology* 48, 299-304.

Hedmark E, Ellegren H (2005) A test of the multiplex pre-amplification approach in microsatellite genotyping of wolverine faecal DNA. *Conservation Genetics* 7, 289-293.

Hedmark E, Flagstad Ø, Segerstrom P, *et al.* (2004) DNA-based individual and sex identification from wolverine (*Gulo gulo*) faeces and urine. *Conservation Genetics* 5, 405-410.

Henegariu O, Heerema NA, Dlouhy SR, Vance GH, Vogt PH (1997) Multiplex PCR: Critical parameters and step-by-step protocol. *Biotechniques* 23, 504-511.

Hite JM, Eckert KA, Cheng KC (1996) Factors affecting fidelity of DNA synthesis during PCR amplification of d(C-A)n.d(G-T)n microsatellite repeats. *Nucleic Acids Research* 24, 2429-2434.

Hoffman D, Amos W (2005) Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. *Molecular Ecology* 14, 599-612.

Hogan FE, Cooke R, Burridge CP, Norman JA (2008) Optimizing the use of shed feathers for genetic analysis. *Molecular Ecology Resources* 8, 561-567.

Horvath MB, Martinez-Cruz B, Negro JJ, Kalmar L, Godoy JA (2005) An overlooked DNA source for non-invasive genetic analysis in birds. *Journal of Avian Biology* 36, 84-88.

Huggins RM (1989) On the statistical analysis of capture experiments. *Biometrika* 76, 133-140.

Inglis GD, Kalischuk LD (2004) Direct quantification of *Campylobacter jejuni* and *Campylobacter lanienae* in Feces of Cattle by Real-Time Quantitative PCR. *Applied Environmental Microbiology* 70, 2296-2306.

Inoue E, Miho I-M, Osamu T, Toshisada N (2007) Wild Chimpanzee infant urine and saliva sampled noninvasively usable for DNA analyses. *Journal of Primatology* 48, 156-159.

Itzkowitz SH, Jandorf L, Brand R, *et al.* (2007) Improved fecal DNA test for colorectal cancer screening. *Clinical Gastroenterology and Hepatology* 5, 111-117.

Iyengar A, Babu VN, Hedges S, *et al.* (2005) Phylogeography, genetic structure, and diversity in the dhole (*Cuon alpinus*). *Molecular Ecology* 14, 2281-2297.

Jeffery KJ, Abernethy KA, Tutin CEG, Bruford MW (2007) Biological and environmental degradation of gorilla hair and microsatellite amplification success. *Biological Journal of the Linnean Society* 91, 281-294.

Jensen T, Pernasetti FM, Durrant B (2003) Conditions for rapid sex determination in 47 avian species by PCR of genomic DNA from blood, shell-membrane blood vessels, and feathers. *Zoo Biology* 22, 561-571.

Johnson PCD, Haydon DT (2007) Maximum-likelihood estimation of allelic dropout and false allele error rates from microsatellite genotypes in the absence of reference data. *Genetics* 175, 827-842.

Jones R, Cable J, Bruford MW (2008) An evaluation of non-invasive sampling for genetic analysis in northern European reptiles. *Herpetological Journal* 18, 32-39.

Kaderali L, Deshpande A, Nolan JP, White PS (2003) Primer-design for multiplexed genotyping. *Nucleic Acids Research* 31, 1796-1802.

Kalinowski ST (2006) HW-QuickCheck: a computer program for checking genotypes for agreement with Hardy-Weinberg expectations. *Molecular Ecology Notes* 6, 974-979.

Kaplinski L AR, Puurand T, Remm M. (2005) MultiPLX: automatic grouping and evaluation of PCR primers. *Bioinformatics* 21, 1701-1702. .

Kawai K, Shimizu M, Hughes RN, Takenaka S (2004) A non-invasive technique for obtaining DNA from marine intertidal snail. *Journal of Marine Biology Association U.K.* 84, 773-774.

Kendall KC, McKelvey KS (2008) Hair Collection. In: *Noninvasive survey methods for North American carnivores* (eds. Long RA, MacKay P, Ray JC, Zielinski WJ), pp. 135-176. Island Press, Washington D.C.

Kendall KC, Stetz JB, Boulanger J*, et al.* (2009) Demography and genetic structure of a recovering brown bear population. *Journal of Wildlife Management* In Press.

Kerley LL, Salkina GP (2007) Using scent-matching dogs to identify individual Amur tigers from scats. *Journal of Wildlife Management* 71, 1349-1356.

Kim N, Lee C (2007) QPRIMER: a quick web-based application for designing conserved PCR primers from multigenome alignments. *Bioinformatics* 23, 2331-2333.

Kittler R, Stoneking M, Kayser C (2002) A whole genome amplification method to generate long fragments from low quantities of genomic DNA. *Analytical Biochemistry* 300, 237-244.

Kohn MH, Wayne RK (1997) Facts from feces revisited. *Trends in Ecology and Evolution* 6, 223-227.

Kohn MH, York EC, Kamradt DA*, et al.* (1999) Estimating population size by genotyping faeces. *Proceedings of the Royal Society of London B* 266, 657-663.

Kontanis EJ, Reed FA (2006) Evaluation of real-time PCR amplification efficiencies to detect PCR inhibitors. *Journal of Forensic Science* 51, 795-804.

Koressaar T, Remm M (2007) Enhancements and modifications of primer design program Primer3. *Bioinformatics* 23, 1289-1291.

Koukoulas I, O'Toole FE, Stringer P, van Oorschot RAH (2008) Quantifiler™: observations of relevance to forensic casework. *Journal of Forensic Science* 53, 135-141.

Lampa S, Gruber B, Henle K, Hoehn M (2008) An optimisation approach to increase DNA amplification success of otter faeces. *Conservation Genetics* 9, 201-210.

Lanyon CV, Rushton SP, O'Donnell AG*, et al.* (2007) Murine scent mark microbial communities are genetically determined. *FEMS Microbiology Ecology* 59, 576-583.

Lau LT, Fung YW, Wong FP*, et al.* (2003) A real-time PCR for SARS-coronavirus incorporating target gene pre-amplification. *Biochemical and Biophysical Research Communications* 312, 1290-1296.

Lecomte N, Gauthier G, Bernatchez L, Giroux J-F (2006) A nondamaging blood sampling technique for waterfowl embryos. *Journal of Field Ornithology* 77, 67-70.

Lee PLM, Prys-Jones RP (2008) Extracting DNA from museum bird eggs, and whole genome amplification of archive DNA. *Molecular Ecology Resources* 8, 551-560.

Leonard JA, Wayne RK, Cooper A (2000) Population genetics of Ice Age brown bears. *Proceedings of the National Academy of Sciences from the USA* 97, 1651-1654.

Luikart G, Pilgrim K, Visty J, Ezenwa VO, Schwartz MK (2008a) Candidate gene microsatellite variation is associated with parasitism in wild bighorn sheep. *Biology Letters* 4, 228-231.

Luikart G, Zundel S, Rioux D*, et al.* (2008b) Low genotyping error rates for microsatellite multiplexes and noninvasive fecal DNA samples from bighorn sheep. *Journal of Wildlife Management* 72, 299–304.

Lukacs PM, Burnham KP (2005) Review of capture-recapture methods applicable to noninvasive genetic sampling. *Molecular Ecology* 14, 3909-3919.

Magnussen JE, Pikitch EK, Clarke SC*, et al.* (2007) Genetic tracking of basking shark products in international trade. *Conservation Genetics* 10, 199-207.

Marshall O (2007) Graphical design of primers with PerlPrimer. *Methods in Molecular Biology* 402, 403-414.

Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology* 7, 639-655.

Martinkova N, Searle JB (2006) Amplification success rate of DNA from museum skin collections: a case study of stoats from 18 museums. *Molecular Ecology Notes* 6, 1014-1017.

Maudet C, Luikart G, Dubray D, Von Hardenberg A, Taberlet P (2004) Low genotyping error rates in wild ungulate faeces sampled in winter. *Molecular Ecology Notes* 4, 772-775.

Mayntz-Press KA, Ballantyne J (2007) Performance characteristics of commercial Y-STR multiplex systems*. *Journal of Forensic Science* 52, 1025-1034.

McKelvey KS, Schwartz MK (2005) dropout: a program to identify problem loci and samples for noninvasive genetic samples in a capture-mark-recapture framework. *Molecular Ecology Notes* 5, 716-718.

Meissner C, Bruse P, Mueller E, Oehmichen M (2007) A new sensitive short pentaplex (ShoP) PCR for typing of degraded DNA. *Journal of Forensic Science* 166, 121-127.

Meyburg B-U, Meyburg C, Francj-Neumann F (2007) Why do female Lesser Spotted Eagles (*Aquila pomarina*) visit strange nests remote from their own? *Journal of Ornithology* 148, 157-166.

Miller C, Joyce P, Waits LP (2002) Assessing allelic dropout and genotype reliability using maximum likelihood. *Genetics* 160, 357-366.

Miller HC (2006) Cloacal and buccal swabs are a reliable source of DNA for microsatellite genotyping of reptiles. *Conservation Genetics* 7, 1001-1003.

Miller MM, Ealey KA, Oswald WB, Schat KA (2003) Detection of chicken anemia virus DNA in embryonal tissues and eggshell membranes. *Avian Disease* 47, 662-671.

Miquel C, Bellemain E, Poillot C, Taberlet P (2006) Quality indexes to assess the reliability of genotypes in studies using noninvasive sampling and multiple-tube approach. *Molecular Ecology Notes* 6, 985- 988.

Missiaggia A, Grattapaglia D (2006) Plant microsatellite genotyping with 4-color fluorescent detection using multiple-tailed primers. *Genetics and Molecular Resources* 5, 72-78.

Mitrovski P, Heinze DA, Broome L, Hoffmann AA, Weeks AR (2007) High levels of variation despite genetic fragmentation in populations of the endangered mountain pygmy-possum, *Burramys parvus*, in alpine Australia. *Molecular Ecology* 16, 75-87.

Mitrovski P, Heinze DA, Guthridge K, Weeks AR (2005) Isolation and characterization of microsatellite loci from the Australian endemic mountain pygmy-possum, *Burramys parvus* Broom. *Molecular Ecology Notes* 5, 395-397.

Morin P, Mccarthy M (2007) Highly accurate SNP genotyping from historical and low quality samples. *Molecular Ecology Notes* 7, 937-946.

Morin PA, Chambers KE, Boesh C, Vigilant L (2001) Quantitative polymerase chain reaction analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (*Pan troglodytes verus*). *Molecular Ecology* 10, 1835-1844.

Morin PA, Hedrick NM, Robertson KM, Leduc CA (2007) Comparative mitochondrial and nuclear quantitative PCR of historical marine mammal tissue, bone, baleen, and tooth samples. *Molecular Ecology Notes* 7, 404-411.

Morin PA, Luikart G, Wayne RK, The SNP Workshop Group (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution* 19, 208-216.

Morling N (2008) Handbook of Forensic Genetics. In: *Forensic Science and Medicine*, p. 400. Humana Press, New York.

Mowat G, Strobeck C (2000) Estimating population size of grizzly bears using hair capture, DNA profiling, and mark-recapture analysis. *Journal of Wildlife Management* 64, 183-193.

Mukherjee N, Mondol S, Andheria A, Ramakrishnan U (2007) Rapid multiplex PCR based species identification of wild tigers using non-invasive samples. *Conservation Genetics* 8, 1465-1470.

Musgrave-Brown E, Ballard D, Balogh K*, et al.* (2007) Forensic validation of the SNPforID 52-plex assay *Forensic Science International* 1, 186-190.

Navidi W, Arnheim N, Waterman MS (1992) A multiple-tubes approach for accurate genotyping of very small DNA samples by using PCR: Statistical considerations. *American Journal of Human Genetics* 50, 347-359.

Nazarenko I, Lowe B, Darfler M*, et al.* (2002) Multiplex quantitative PCR using self-quenched primers labelled with a single fluophore. *Nucleic Acids Research* 30, e37.

Neilan BA, Wilton AN, Jacobs D (1997) A universal procedure for primer labelling of amplicons. *Nucleic Acids Research* 25, 2938-2939.

Oetting WS, Lee HK, Flanders DJ*, et al.* (1995) Linkage analysis with multiplexed short tandem repeat polymorphisms using infrared fluorescence and M13 tailed primers. *Genomics* 30, 450-458.

Onofri V, Alessandrini F, Turchi C*, et al.* (2006) Development of multiplex PCRs for evolutionary and forensic applications of 37 human Y chromosome SNPs. *Forensic Science International* 157, 23-35.

Otis DL, Burnham KP, White GC, Anderson DR (1978) Statistical inference from capture data on closed animal populations. *Wildlife Monographs* 62, 178.

Paetkau D (2003) An impirical exploration of data quality in DNA-based population inventories. *Molecular Ecology* 12, 1375-1387.

Palmer A, Styan C, Shearman D (2008) Foot mucus is a good source for non-destructive genetic sampling in Polyplacophora. *Conservation Genetics* 9, 229-231.

Palomares F, Godoy JA, Piriz A, O´Brien SJ, Johnson WE (2002) Faecal genetic analysis to determine the presence and distribution of elusive carnivores: design and feasibility for Iberian lynx. *Molecular Ecology* 11, 2171-2182.

Palsboll PJ, Allen J, Berub M*, et al.* (1997) Genetic tagging of humpback whales. *Nature* 388, 767-769.

Pang BC, Cheung BK (2007) Double swab technique for collecting touched evidence. *Legal Medicine* 9, 81-84.

Parsons KM (2001) Reliable microsatellite genotyping of dolphin DNA from faeces. *Molecular Ecology Notes* 1, 341-344.

Pauli JN, Hamilton MB, Crain EB, Buskirk SW (2008) A Single-Sampling Hair Trap for Mesocarnivores. *Journal of Wildlife Management* 72, 1650-1652.

Pfeiffer I, Völkel I, Täubert H, Brenig B (2004) Forensic DNA-typing of dog hair: DNA-extraction and PCR amplification. *Forensic Science International* 141, 149-151.

Piggott M, Bellemain EA, Taberlet P, Taylor AC (2004) Multiplex pre-amplification method that significantly improves microsatellite amplification and error rates for faecal DNA in limiting conditions. *Conservation Genetics* 5, 417-420.

Piggott MP, Banks SC, Stone N, Banffy C, Taylor AC (2006) Estimating population size of endangered brush-tailed rockwallaby (*Petrogale Penicillata*) colonies using faecal DNA. *Molecular Ecology* 15, 81-91.

Piggott MP, Taylor AC (2003) Extensive evaluation of faecal preservation and DNA extraction methods in Australian native and introduced species. *Australian Journal of Zoology* 51, 341-355.

Pledger S (2000) Unified maximum likelihood estimates for closed capture-recapture models using mixtures. *Biometrics* 56, 434-442.

Pompanon F, Bonin A, Bellemain E, Taberlet P (2005) Genotyping errors: Causes, consequences and solutions. *Nature Reviews Genetics* 6, 847-846.

Porreca GJ, Zhang K, Li J*, et al.* (2007) Multiplex amplification of large sets of human exons. *Nature Methods* 4, 931-936.

Puechmaille SJ, Mathy G, Petit EJ (2007) Good DNA from bat droppings. *Acta Chiropterologica* 9, 269-276.

Pusch CM, Giddings I, Scholz M (1998) Repair of degraded duplex DNA from prehistoric samples using Escherichia coli DNA polymerase I and T4 DNA ligase. *Nucleic Acids Research* 26, 857-859.

Queipo-Ortuño MI, Colmenero JD, Muñoz N*, et al.* (2006) Rapid diagnosis of Brucella epididymo-orchitis by real-time polymerase chain reaction assay in urine samples. *The Journal of Urology* 176, 2290-2293.

Radstrom P, Lofstrom C, Lovenklev M, Knutsson R, Wolffs P (2008) Strategies for overcoming PCR inhibition. *Cold Spring Harbor Protocols* 2008.

Regnaut S, Lucas FS, Fumagalli L (2006) DNA degradation in avian faecal samples and feasibility of non-invasive genetic studies of threatened capercaillie populations. *Conservation Genetics* 7, 449-453.

Roeder AD, Archer FI, Poinar HN, Morin PA (2004) A novel method for collection and preservation of faeces for genetic studies. *Molecular Ecology Notes* 4, 761-764.

Roon DA, Thomas ME, Kendall K, Waits LP (2005) Evaluating mixed samples as a source of error in noninvasive genetic studies using microsatellites. *Molecular Ecology* 14, 195-201.

Roon DA, Waits LP, Kendall KC (2003) A quantitative evaluation of two methods for preserving hair samples. *Molecular Ecology Notes* 3, 163-166.

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (eds. Krawetz S, Misener S), pp. 365-386. Humana Press, Totowa, NJ, USA.

Rudnick JA, Katzner TE, Bragin EA, DeWoody JA (2007) Species identification of birds through genetic analysis of naturally shed feathers. *Molecular Ecology Notes* 7, 757-762.

Rudnick JA, Katzner TE, Bragin EA, DeWoody JA (2008) A non-invasive genetic evaluation of population size, natal philopatry, and roosting behavior of non-breeding eastern imperial eagles (*Aquila heliaca*) in central Asia. *Conservation Genetics* 9, 667-676.

Ruiz-González A, Rubines A, Berdión O, Gómez-Moliner BJ (2008) A non-invasive genetic method to identify the sympatric mustelids pine marten (*Martes martes*) and stone marten (*Martes foina*): preliminary distribution survey on the northern Iberian Peninsula. *European Journal of Wildlife Research* 54, 253-261.

Santini A, Lucchini V, Fabbri E, Randi E (2007) Ageing and environmental factors affect PCR success in wolf (*Canis lupus*) excremental DNA samples. *Molecular Ecology Notes* 7, 955-961.

Sastre N, Francino O, Lampreave G*, et al.* (2008) Sex identification of wolf (*Canis lupus*) using non-invasive samples. *Conservation Genetics* 10: 555-558.

Scandura M (2005) Individual sexing and genotyping from blood spots on the snow: A reliable source of DNA for non-invasive genetic surveys. *Conservation Genetics* 6, 871-874.

Schmaltz G, Somers CM, Sharma P, Quinn JS (2006) Non-destructive sampling of maternal DNA from the external shell of bird eggs. *Conservation Genetics* 7, 543-549.

Schmidt K, Kowalczyk R (2006) Using scent-marking stations to collect hair samples to monitor Eurasian lynx populations. *Wildlife Society Bulletin* 34, 462-466.

Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* 18, 233-234.

Schwartz MK, Aubry KB, McKelvey KS*, et al.* (2007) Inferring geographic isolation of wolverine in California using historical DNA. *Journal of Wildlife Management* 71, 2170-2179.

Schwartz MK, Cushman SA, McKelvey KS, Hayden J, Engkjer C (2006) Detecting genotyping errors and describing black bear movement in North Idaho. *Ursus* 17, 138-148.

Segelbacher G (2002) Noninvasive genetic analysis in birds: testing reliability of feather samples. *Molecular Ecology Notes* 2, 367-369.

Shivji MS, Chapman DD, Pikitch EK, Raymond PW (2005) Genetic profiling reveals illegal international trade in fins of the great white shark, *Carcharodon carcharias*. *Conservation Genetics* 6, 1035-1039.

Smith DA, Ralls K, Cypher BL, Maldonado JE (2005) Assessment of scat-detection dog surveys to determine kit fox distribution. *Wildlife Society bulletin* 33, 897-904.

Smith S, Vigilant L, Morin PA (2002) The effects of sequence length and oligonucleotide mismatches on 5¢ exonuclease assay efficiency. *Nucleic Acids Research* 30, e111.

Soulsbury C, Iossa G, Edwards K, Baker P, Harris S (2007) Allelic dropout from a high-quality DNA source. *Conservation Genetics* 8, 733-738.

Spitaleri S, Piscitello D, Di Martino D (2004) Experimental procedures comparing the activity of different *Taq* polymerases. *Forensic Science International* 146, S167-S169.

Sundqvist A-K, Ellegren H, Vilà C (2008) Wolf or dog? Genetic identification of predators from saliva collected around bite wounds on prey. *Conservation Genetics* 9(5): 1275-1279.

Swango KL, Hudlow WR, Timken MD, Buoncristiani MR (2006) A quantitative PCR assay for the assessment of DNA degradation in forensic samples. *Forensic Science International* 158, 14-26.

Swanson BJ, Kelly BP, Maddox CK, Moran JR (2006) Shed skin as a source of DNA for genotyping seals. *Molecular Ecology Notes* 6, 1006-1009.

Taberlet P, Griffin S, Goossens B*, et al.* (1996) Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research* 24, 3189-3194.

Taberlet P, Mattock H, Dubois-Paganon C, Bouvet J (1993) Sexing free-ranging brown bears *Ursus arctos* using hairs found in the field. *Molecular Ecology* 2, 399-403.

Taberlet P, Waits LP, Luikart G (1999) Noninvasive genetic sampling: look before you leap. *Trends in Ecology and Evolution* 14, 323-327.

Tang Y-W, Sefers SE, Li H, Kohn DJ, Procop GW (2005) Comparative evaluation of three commercial systems for nucleic acid extraction from urine specimens. *Journal of Applied Microbiology* 43, 4830-4833.

Thomas M, Gilbert P (2006) Postmortem damage of mitochondrial DNA. In: *Human Mitochondrial DNA and the Evolution of Homo sapiens* (eds. Bandelt H-J, Macaulay V, Richards M), pp. 91-115. Springer-Verlag Berlin.

Thornton CG, Passen S (2004) Inhibition of PCR amplification by phytic acid, and treatment of bovine fecal specimens with phytase to reduce inhibition. *Journal of Microbiological Methods* 59, 43-52.

Tobal K, Layton DM, Mufti GJ (1989) Non-invasive isolation of constitutional DNA for genetic analysis. *The Lancet* 2, 1281-1282.

Tóth M (2008) A New Noninvasive Method for Detecting Mammals From Birds Nests. *Journal of Wildlife Management* 72, 1237-1240.

Ulizio T, Squires JR, Petscher DH*, et al.* (2006) The efficacy of obtaining genetic-based identifications from putative wolverine snow tracks. *Wildlife Society Bulletin* 34, 1326-1332.

Valentini A, Miquel C, Nawaz MA*, et al.* (2009) New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the trnL approach. *Molecular Ecology Resources* 9(1): 51-60.

Valiere N (2002) Gimlet: a computer program for analysing genetic individual identification data. *Molecular Ecology Notes* 2, 377-379.

Valière N, Taberlet P (2000) Urine collected in the field as a source of DNA for species and individual identification. *Molecular Ecology* 9, 2150-2152.

Vallet D, Petit EJ, Gatti S, Levréro F, Ménard N (2008) A new 2CTAB/PCI method improves DNA amplification success from faeces of Mediterranean (Barbary macaques) and tropical (lowland gorillas) primates. *Conservation Genetics* 9, 677-680.

Vallone PM, Butler JM (2004) AutoDimer: a screening tool for primer-dimer and hairpin structures. *Biotechniques* 37, 226-231.

Van Oosterhout C, Hutchinson WF, Willis DPM, Shipley P (2004) Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* 4, 535–538.

Vigilant L (1999) An evaluation of techniques for the extraction and amplification of DNA from naturally shed hairs. *Biological Chemistry* 380, 1329 - 1331.

Waits LP, Paetkau D (2005) Noninvasive genetic sampling tools for wildlife biologists: A review of applications and recommendations for accurate data collection. *Journal of Wildlife Management* 69, 1419-1433.

Walker FM, P S, Taylor AC (2008) Evidence for habitat fragmentation altering within-population processes in wombats. *Molecular Ecology* 7, 1674-1684.

Wan Q-H, Zhu L, Wu HUA, Fang S-G (2006) Major histocompatibility complex class II variation in the giant panda (*Ailuropoda melanoleuca*). *Molecular Ecology* 15, 2441-2450.

Wandeler P, Smith S, Morin PA, Pettifor RA, Funk SM (2003) Patterns of nuclear DNA degeneration over time - a case study in historic teeth samples. *Molecular Ecology* 12, 1087-1093.

Wasser SK, Mailand C, Booth R*, et al.* (2007) Using DNA to track the origin of the largest ivory seizure since the 1989 trade ban. *Proceedings of the National Academy of Sciences from the USA* 104, 4228-4233.

Weaver JL, Wood P, Paetkau D, Laack LL (2005) Use of scented hair snares to detect ocelots. *Wildlife Society Bulletin* 33, 1384-1391.

Weckx S, De Rijk P, Van Broeckhoven C, Del-Favero J (2005) SNPbox: a modular software package for large-scale primer design. *Bioinformatics* 21, 385-387.

Whitaker JP, Cotton EA, Gill P (2001) A comparison of the characteristics of profiles produced with the AMPFlSTR(R) SGM Plus(TM) multiplex system for both standard and low copy number (LCN) STR DNA analysis. *Forensic Science International* 123, 215-223.

Willard JM, Lee DA, Holland MM (1998) Recovery of DNA for PCR Amplification from Blood and Forensic Samples Using a Chelating Resin. In: *Forensic DNA Profiling Protocols* (eds. Lincoln PJ, Thomson J), pp. 9-18. Humana Press, Totowa, NJ.

Williams CL, Blejwas K, Johnston JJ, Jaeger MM (2003) A coyote in sheep´s clothing: predator identification from saliva. *Wildlife Society Bulletin* 31, 926-932.

Yamada T, Soma H, Morishita S (2006) PrimerStation: a highly specific multiplex genomic PCR primer design server for the human genome. *Nucleic Acids Research* 34, W665-W669.

Zielinski WJ, Schlexer FV, Pilgrim KL, Schwartz MK (2006) The efficacy of wire and glue hair snares in identifying mesocarnivores. *Wildlife Society Bulletin* 34, 1152-1161.

Species identification of sympatric wild carnivores from South-western Europe: a straightforward method using a nuclear gene

R. Oliveira[1,2], D. Castro[1,2], R. Godinho[1], G. Luikart[1,2,3], P. C. Alves[1,2]

[1]CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, Campus Agrário de Vairão, 4485-661 Vairão, Portugal
[2]Departamento de Zoologia e Antropologia, Faculdade de Ciências da Universidade do Porto, Rua Campo Alegre s/n, 4169 - 007 Porto, Portugal
[3]Division of Biological Sciences, University of Montana, USA

ABSTRACT

Species identification is essential for noninvasive studies of elusive and rare animals, and for detecting illegal harvest or trade of wildlife species. However, most molecular tests identify only a limited number of species or require multiple laboratory steps to distinguish many *taxa*. Additionally, most protocols use mitochondrial DNA being, therefore, especially prone to problems such as nuclear insert copies, high intraspecific diversity or heteroplasmy. Here, we developed a straightforward molecular test based on the polymorphism detected on a 221bp exon fragment of the IRBP (*Interphotoreceptor Retinoid-Binding Protein*) nuclear gene. This fragment revealed 51 variable sites (including 12 non-synonymous and 19 species-specific sites), which enabled the successful distinction of all 16 carnivore species from South-western Europe. A SSCP (Single-strand Conformational Polymorphism) gel electrophoresis technique was also optimized to allow the simple and inexpensive application of this test. Sequences and SSCP profiles were consistent in identifying a total of 387 samples, including faeces (172) and hairs (17) collected noninvasively in the field. Due to its low cost, simplicity, and wide range of identifiable species, this test shows great promise to facilitate studies in molecular ecology, conservation genetics, and forensic analysis, as well as DNA bar-coding projects.

Keywords: Species ID, nuclear gene IRBP, carnivores, Single-strand Conformational Polymorphism, conservation genetics, noninvasive genetics

INTRODUCTION

Reliable identification of species is fundamental in molecular ecology, conservation biology, forensic sciences and wildlife management because many studies in these disciplines crucially depend on species identification for a wide range of applications, such as the definition of geo- graphic distributions, estimation of densities, and the analysis of biological and behavioural parameters (Long *et al.* 2008). However, detection and identification of species is difficult for rare and threatened *taxa*, mainly due to their low densities, nocturnal and elusive behaviour, and to the logistical and ethical difficulties involved in their capture and handling during traditional capture-mark- recapture approaches. This has prompted the development of noninvasive tools applied to samples such as faeces, hair, feathers, urine, saliva, scent marks, sloughed skin and animal products illegally trafficked (e.g., Taberlet *et al*. 1999; Dalèn *et al.* 2004, Arrendal *et al*. 2007; Hogan *et al*. 2008; Jones *et al*. 2008; Sundqvist *et al*. 2008; Sastre *et al*. 2009). However, noninvasive samples seldom allow species identification based on morphology alone (Davison *et al*. 2002). When sympatric carnivores have similar body features, behaviour and feeding habits, visual discrimination of scats and hairs are subjective and error-prone (Riddle *et al*. 2003; Gómez-Moliner *et al*. 2004; Kurose *et al*. 2005; Nagata *et al*. 2005; Pilot *et al*. 2007). Therefore, noninvasive studies of carnivore species should include genetic identification.

During the last decade, a number of molecular protocols have been designed to identify mammal carnivores. Direct sequencing of a diagnostic DNA fragment (e.g., Farrell *et al*. 2000; Murakami 2002; Adams and Waits 2007; Karlsson and Holmlund 2007; Kitano *et al*. 2007) is still relatively expensive for large-scale wildlife surveys, and DNA mixtures (e.g., contamination with prey DNA) are complicated to decipher. Therefore, alternative protocols have been developed, like mtDNA RFLP (Restriction Fragment Length Polymorphism) assays (e.g., Hoss *et al*. 1992; Paxinos *et al*. 1997; Hansen and Jacobsen 1999; Mills *et al*. 2000; Williams *et al*. 2003; Lucentini *et al*. 2007; Ruiz-González *et al*. 2008) and tests based on positive versus negative PCR amplifications using species- specific primers (e.g., Palomares *et al*. 2002; Dalén *et al*. 2004; Kurose *et al*. 2005; Fernandes *et al*. 2007; Tobe and Linacre 2008). Most of the protocols target few *taxa*, frequently endangered species and sympatric ones, or species from which noninvasive signs are usually similar. In this context, Fernandes *et al*. (2007) reported the broadest DNA-based method specifically designed for carnivore discrimination, using species-specific mtDNA primers to identify 15 species in the Iberian Peninsula. More recently, real-time PCR and melt-curve analysis protocols have also been developed to distinguish a few carnivore species (Berry and Sarre 2007; O'Reilly *et al*. 2008).

Despite the recent advances in developing new molecular assays, important limitations still exist. For instance, although real-time PCR presents relatively low cost and is less time-consuming when compared with methods requiring post-PCR manipulations, the protocol developed by Berry and Sarre (2007) showed that intraspecific polymorphism or melt temperature overlap may occur between closely related species. Additionally, DNA-based methods specifically designed for carnivore species discrimination (e.g., Fernandes *et al*. 2007) are based solely on mtDNA polymorphism, presenting some limitations: mtDNA evolves rapidly, which may result in intraspecific diversity too

high for species discrimination proposes (e.g., O'Reilly *et al.* 2008); heteroplasmy, which has been documented for several mammalian mitochondrial genomes (Hsieh *et al.* 2001; Paneto *et al.* 2007); nuclear copies (NUMTS), which are known to occur in different species (Zhang and Hewitt 1996; Lopez *et al.* 1996; Kim *et al.* 2006; Antunes *et al.* 2007); and mtDNA introgression which also is common in mammals (e.g., Ballard and Whitlock 2004; Alves *et al.* 2006). Therefore, new methodologies based on nuclear genes could be more informative or provide extra information to complement mtDNA-based protocols.

In South-western Europe, 16 wild carnivore species are sympatrically distributed, including animals from six different families: Mustelidae (weasel *Mustela nivalis*, western polecat *Mustela putorius*, European mink *Mustela lutreola*, American mink *Mustela vison*, stoat *Mustela erminea*, stone marten *Martes foina*, pine marten *Martes martes*, European otter *Lutra lutra*, Eurasian badger *Meles meles*); Canidae (wolf *Canis lupus*, red fox *Vulpes vulpes*); Felidae (European wildcat *Felis silvestris*, Iberian lynx *Lynx pardinus*); Viverridae (common genet *Genetta genetta*); Ursidae (brown bear *Ursus arctos*) and Herpestidae (Egyptian mongoose *Herpestes ichneumon*). For most of these species, especially the ones with high conservation status such as the Iberian lynx, the European wildcat, the wolf and the European mink, it is urgent to clarify their precise distribution, to elucidate behavioural, ecological and genetic features, and to evaluate biological relationships among populations and subspecies. Here we present a simple, rapid and inexpensive technique to identify all carnivore species in South-western Europe based on the high discriminatory power of a short fragment of the nuclear gene Interphotoreceptor Retinoid-Binding Protein (IRBP), which can also be successfully applied to the identification of noninvasive and forensic samples. Based on its simplicity, high efficiency and reliability, and low cost, we further discuss its usefulness for carnivore management and biodiversity conservation.

MATERIAL AND METHODS

This study was divided in three main steps, including the development, laboratory testing and practical implementation of a novel DNA-based technique.

a) Developing the assay for species discrimination: carnivores IRBP sequences analysis

A fragment of approximately 1,040 bp of the IRBP gene was PCR amplified with carnivore universal primers, following Flynn and Nedbal (1998). Between two to five individuals belonging to each of the 16 species representing all natural extant carnivores from South-western Europe, plus the dog and the domestic cat (Table 1), were sequenced for both strands on a 3130xl Genetic Analyser Sequencer (Applied Biosystems/HITACHI). Sequences were aligned together with previously published ones using the software SEQ- SCAPE 2.0 (Applied Biosystems). All new sequences were submitted to NCBI GenBank (accession numbers from GQ214060 to GQ214077). After analysing the entire fragment variability in MEGA 3.1 (Kumar *et al.* 2004), we selected a short fragment of the first exon, which revealed high interspecific diagnostic polymorphism, thus showing great potential to easily and

reliably identify all species. This region was also selected because lower intraspecific variation is likely to occur in exons, when compared to intronic regions or other non-coding fragments.

Table 1. Number of tissue (and noninvasive) samples genotyped from each carnivore species and their geographic location. The number of samples from each different country is indicated.

Species	Samples	Geographic location
Canis sp.	24 (11)	Portugal (14); Spain (7); Romania (3)
Felis spp.	108 (85)	Portugal (85); Spain (9); Italy (5); Germany (5); Bulgaria (2); Romania (2)
Genetta genetta	14	Portugal (10); Spain (4)
Herpestes ichneumon	10	Portugal (9); Spain (1)
Lutra lutra	13	Portugal (4); Austria (9)
Lynx pardinus	4	Spain (4)
Martes foina	31 (18)	Portugal (19); Spain (11); Greece (1)
Martes martes	37 (17)	Portugal (1); Italy (7); Spain (27); Finland (2)
Meles meles	21 (6)	Portugal (6); Denmark (2); Spain (11); Italy (2)
Mustela erminea	12	Portugal (2); Czech Republic (1); Denmark (2); Great Britain (1); Luxemburg (1); Ireland (1); Italy (2); Norway (1); Spain (1)
Mustela lutreola	5	Spain (5)
Mustela nivalis	18	Portugal (5); Austria (1); Denmark (1); Italy (6); Spain (2); Finland (3)
Mustela putorius	19	Portugal (9); Spain (5); Italy (3); Denmark (2)
Mustela vison	8	Portugal (2); Spain (4); Denmark (2)
Vulpes vulpes	89 (80)	Portugal (29); Spain (60)
Ursus arctos	14 (12)	Italy (12); Spain (2)

b) Optimizing the routine protocol: sequencing and an alternative PCR-SSCP technique

To amplify the small selected fragment, we designed PCR primers targeting conserved regions among all species, which resulted in a PCR product of 221 bp: the forward primer, IRBPex1.Fw (5'-GAGAAAGCMCTGGCCATCCT-3'), and the reverse primer, IRBPex1.Rw (5'-ACCAGGAGCCTGGGGTCCTC-3') have their 50 positions at nucleotides 940 and 1160, respectively, of the human IRBP gene (accession number X53044). The optimal PCR profile was identified by testing a range of hybridization temperatures (55 - 65°C) using the following conditions: 4 min at 94°C followed by 30 cycles at 94°C for 30 s, 60°C for 30 s and 72°C for 30s, and a final extension at 72°C for 5 min. PCR amplifications were performed for all the individuals sequenced in a), in a total volume of 15 μl containing: 19 PCR buffer (Ecotaq), 1.8 mM $MgCl_2$, 0.3 μM of each primer, 0.3 mM of each dNTP, 1 μl of DMSO (Dimethylsulfoxide), 0.5 U of *Taq* polymerase

(Ecotaq) and approximately 10 ng of genomic DNA. In order to optimize an alternative protocol to the direct sequencing of this fragment, the resulting PCR products were analysed based on SSCP gel electrophoresis using a vertical electrophoresis system apparatus (BIO-RAD Protean® II xiCell).

After optimization of the experimental conditions, the best discrimination of all SSCP patterns was achieved using a 12% polyacrylamide gel (59:1 acrylamide: methylbisacrylamide) ran with 19 TBE buffer at constant voltage of 450 V and temperature of 20°C, during 6h30. Samples were genotyped as follows: 1 μl of amplified product was mixed with 4 μl of denaturing loading buffer (95% deionized formamide, 10 mm NaOH, 0.01% bromophenol blue and 0.01% xylenecyanol) and, then, denatured for 5 min at 95°C. Mixtures were kept on ice until 4 μl were loaded on the gel. Visualization was made by silver staining (SILVER SEQUENCE™ Staining Reagents, Promega Corporation, USA).

c) Validating the PCR-SSCP method for species discrimination: testing robustness and reliability

Total DNA from 198 tissue samples belonging to the 16 carnivore species was extracted using a standard salting- out procedure (Sambrook *et al.* 1989). Sampling was distributed throughout different geographical locations in Europe, in order to span potential genetic variants within species (Table 1). Additionally, ten Iberian samples from domestic cats (5) and dogs (5) were used to test the possibility of discriminating between them and their wild counterparts.

To evaluate the applicability of the method in non- invasive studies, faecal DNA of known origin was extracted for some species (Table 1) using a guanidine thiocyanate protocol (adapted from Gerloff *et al.* 1995; Godoy, personal communication). Sampling was per- formed in different field sites across Iberian Peninsula, under a variety of exposure times and environmental conditions (e.g., different temperatures and humidity levels), given that both fresh (less than 3 - 4 day old) and medium aged scats (maximum estimated age of approximately 1 month) were collected in both Mediterranean and Atlantic habitats. Noninvasive samples comprised 212 scats from *Felis* spp. (80), *Canis* sp. (11), *Vulpes vulpes* (80), *Meles meles* (6), *Martes martes* (17) and *Martes foina* (18), from which 143 (67.5%) were fresh and 69 (32.5%) were medium aged samples. One individual in captivity was noninvasively sampled for each of the above mentioned species (except for pine marten), in order to additionally analyse a fresh and less exposed scat sample as positive control of the noninvasive experiments. DNA from remotely plucked hairs of brown bear (12) and European wildcat (5) was also extracted, using the Qiagen QIamp DNA Micro Kit according to manufacturer's instructions. Until DNA extraction, both scat and hair samples were stored in absolute ethanol (5 parts of ethanol per part of sample) at room temperature.

In order to reduce possible cross-amplification of prey items, only the external surface and tip of each scat (where intestinal epithelium cells of the predator accumulate) were used for DNA extraction. Nevertheless, contamination with prey and human DNA was monitored by extracting, amplifying, and SSCP identifying DNA from the lab technicians and from a comprehensive array of potential preys (Appendix). Laboratory procedures for noninvasive samples were conducted in a separate and autonomous facility, under sterile conditions. A maximum of seven samples and one

negative control were handled in each extraction round. For both invasive and noninvasive samples, pre and post-PCR manipulations were conducted in physically separated rooms.

All samples were genotyped with the technique described above in b). However, for noninvasive samples, minor modifications were made to the original PCR-SSCP protocol: two independent PCRs of 40 cycles were performed, the PCR reaction volume was increased to 25 μl (all reagents proportionally) and 4 μl of DNA extract were used. Positive and negative controls (from both extraction and PCR) were loaded on each gel in order to compare the noninvasive results with the expected species-specific conformations and for contamination monitoring, respectively.

RESULTS

Highly robust and specific PCR amplifications of the IRBP-exon 1 fragment were obtained for all target carnivore species. The highest annealing temperature without compromising amplification yield was selected to reduce unwanted PCR products and maximize specificity (60°C), although lower temperatures can be used to facilitate amplification for poorer quality samples (57 - 60°C; e.g., scats). The analysis of the partial IRBP-exon 1 showed a total of 51 variable sites from which 12 are non-synonymous and at least 19 are species-specific, thus revealing potentially interspecific diagnostic polymorphism (Table 2).

A specific and unique SSCP band pattern was obtained for each of the 16 species in the 198 tissue samples (Fig. 1) and all individuals from the same species showed consistent SSCP patterns, even for samples from distant geo- graphic origins. Intra-specific polymorphism was identified for *Felis* spp., *Canis* sp., *Martes foina*, *Mustela erminea* and *M. nivalis* (accession numbers from GQ214054 to GQ214059; Table 2), although this variation never over- lapped with the patterns observed for other species. This IRBP locus did not distinguish between wolf and dog or between wild and domestic cats.

We also verified the usefulness of this protocol using noninvasive DNA from some of the studied carnivores. Considering only the scats where it was possible to extract DNA, we successfully amplified and identified 81,13% (172/212) (79.72% fresh and 84.06% medium aged scats), showing that the method is both reliable and robust when applied to samples with low DNA quantity and quality (Fig. 2). Amplification success proved also not to be species sensitive (data not shown). All of the 17 plucked hair samples from bear and wildcat were successfully identified.

Although PCR amplifications were obtained for some non-target species that may be part of carnivores' diet in South-western Europe (32.69% of the potential prey spe- cies analysed), their SSCP profiles were easily distin- guishable from all the carnivore species. For all other investigated prey species, including several mammals, birds, fishes, amphibians, reptiles and a few invertebrates, no amplification was detected (Appendix).

Table 2. Interspecific polymorphic positions of a 221bp fragment from IRBP exon 1 in wild carnivores from South-western Europe.

		2 1	2 4	2 5	3 0	3 1	3 3	4 2	4 3	4 5	4 6	4 8	5 1	5 3	6 0	6 5	6 9	7 2	7 3	7 4	7 6	7 8	7 9	8 7	8 8	9 0	1 0 0 5	1 0 0 6	1 0 0 7	1 0 0 8	1 1 2 5	1 1 2 8	1 1 2 9	1 1 3 2	1 1 4 1	1 1 4 4	1 1 4 7	1 1 5 3	1 1 5 4	1 1 5 5	1 1 5 9	1 1 6 8	1 1 6 9	1 1 7 1	1 1 7 4	1 1 8 0	1 1 8 3	1 1 8 6	1 1 8 9	1 1 9 5	1 1 9 7	1 2 1
European wildcat	*Felis s. silvestris*	**C**	**T**	**C**	**C**	**C**	**G**	**A**	**G**	**A**	**G**	**T**	**C**	**A**	**G**	**A**	**G**	**G**	**C**	**A**	**A**	**C**	**T**	**C**	**C**	**T**	**C**	**A**	**G**	**C**	**M**	**A**	**C**	**G**	**C**	**A**	**A**	**G**	**G**	**A**	**G**	**C**	**A**	**C**	**A**	**C**	**T**	**C**	**A**	**T**	**T**	**T**
Iberian lynx	*Lynx pardinus*	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Egyptian mongoose	*Herpestes ichneumon*	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	G	G	.	.	.	.	.	.	C	.	C	.	.	.	G	C	T	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.
Common Genet	*Genetta genetta*	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	A	.	A	G	G	.	.	T	.	.	.	C	.	C	.	.	.	G	.	T	.	.	.	.	.	T	.	T	.	.	.	.	.	.	.	C
Brown bear	*Ursus arctos*	.	.	.	.	.	.	.	.	G	A	A	G	G	.	.	C	.	.	.	G	.	.	G	.	.	.	.	C	.	C	.	.	.	T	G	G	T	C	.	.	.	.	G	G	.	.	.	.	.	.	C	.
European otter	*Lutra lutra*	.	.	T	.	A	.	T	.	.	.	G	G	G	.	G	.	C	.	.	.	.	.	A	.	C	.	G	C	.	C	.	T	.	T	.	G	T	.	.	.	.	G	G	G	.	A	T	G	.	C	C
Pine marten	*Martes martes*	.	.	T	.	.	.	C	.	.	.	G	T	G	.	G	.	C	.	.	G	.	C	G	.	C	.	G	C	.	C	.	.	.	T	G	G	T	.	.	.	.	G	G	G	.	A	T	.	.	C	C
Stone marten	*Martes foina*	.	.	T	.	.	.	C	.	.	.	G	.	G	.	G	.	C	.	.	G	.	C	G	.	C	.	G	C	.	C	.	.	S	T	G	G	C	.	.	.	.	G	G	G	.	A	T	G	.	C	C
Eurasian badger	*Meles meles*	.	.	T	.	.	.	C	.	.	.	G	G	G	.	G	.	C	.	.	G	.	.	G	.	C	.	G	C	.	C	.	.	.	T	G	G	C	.	.	.	A	G	G	G	.	A	T	G	.	C	C
Polecat	*Mustela putorius*	.	.	T	T	.	.	T	.	.	.	G	G	G	K	G	C	C	.	.	G	.	.	A	.	.	.	G	C	T	C	.	T	.	T	G	G	T	.	.	A	.	G	G	G	.	A	T	G	.	C	C
Stoat	*Mustela erminea*	.	.	T	.	.	.	T	.	.	.	G	G	G	.	G	C	C	.	.	G	.	.	A	.	.	A	G	C	T	C	.	T	.	T	G	G	T	.	.	.	.	G	G	G	Y	A	T	G	.	C	.
Weasel	*Mustela nivalis*	.	.	T	T	.	.	T	.	.	.	G	G	G	.	G	C	C	.	.	G	.	.	A	.	.	Y	G	C	T	C	.	T	.	T	G	G	T	.	.	.	.	G	G	G	.	A	T	G	.	C	C
European mink	*Mustela lutreola*	.	.	T	T	.	.	T	.	.	.	G	G	G	.	G	T	C	.	.	G	.	.	A	.	.	.	G	C	T	C	.	T	.	T	G	G	T	.	.	.	T	G	G	G	.	A	T	G	.	C	C
American mink	*Mustela vison*	.	.	T	T	.	.	T	.	.	.	G	G	G	.	G	.	C	.	.	G	.	.	.	T	C	.	G	C	T	C	.	T	.	T	G	G	T	.	.	.	.	G	G	G	.	A	T	G	.	C	C
Wolf	*Canis lupus*	G	C	.	.	.	C	C	A	C	A	A	G	.	.	.	C	.	.	.	G	T	.	R	.	C	.	G	C	.	C	G	T	.	.	C	G	C	A	G	.	.	.	.	G	T	.	T	G	C	.	.
Red fox	*Vulpes vulpes*	G	C	.	.	.	C	C	.	C	.	G	A	G	.	.	C	.	.	.	G	.	.	G	.	C	.	G	C	.	C	G	.	.	.	C	G	T	A	G	.	.	.	.	G	T	.	T	G	C	.	C

Position 1 corresponds to position 940 of the Human IRBP exon 1 (accession n. X53044). Boxes represent species-specific nucleotide variations and shaded positions correspond to non-synonymous variation.

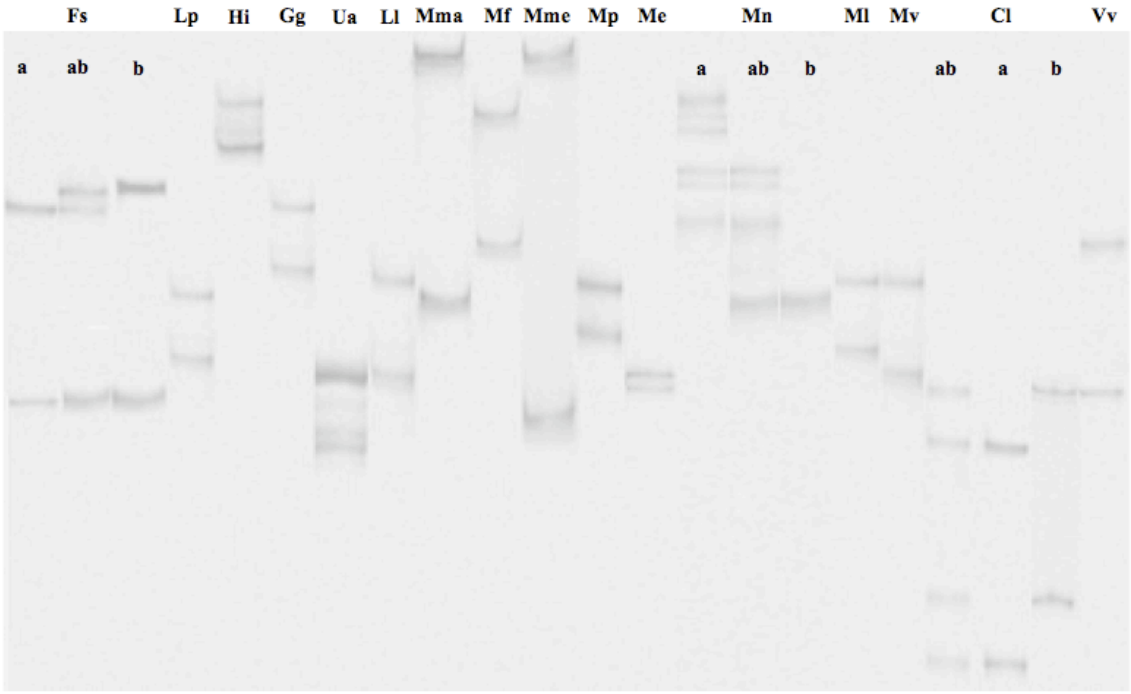


Figure 1. Separation of variants of the IRBP gene (221bp) in Iberian wild carnivores by SSCP analysis on 12% polyacrylamide gels. Visualization was done by silver staining. Fs – *Felis silvestris*; Lp – *Lynx pardinus*; Hi – *Herpestes ichneumon*; Gg – *Genetta genetta*; Ua – *Ursus arctos*; Ll – *Lutra lutra*; Mma – *Martes martes*; Mf – *Martes foina*; Mme – *Meles meles*; Mp – *Mustela putorius*; Me – *Mustela erminea*; Mn – *Mustela nivalis*; Ml – *Mustela lutreola*; Mv – *Mustela vison*; Cl – *Canis lupus*; Vv - *Vulpes vulpes* (a, b and ab represent intraspecific polymorphism found in *Felis silvestris*, *Mustela nivalis* and *Canis lupus*).

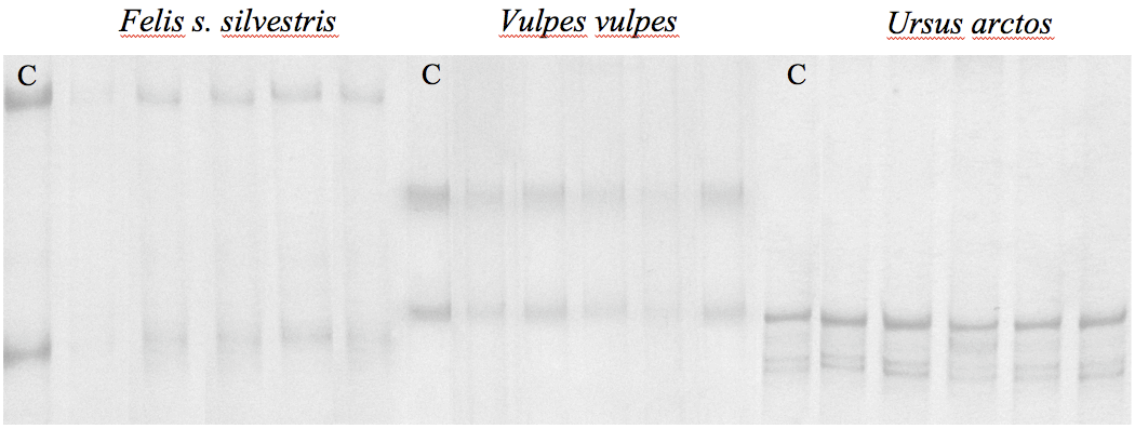


Figure 2. SSCP patterns obtained for five European wildcat and red fox scats, and five brown bear hairs samples. The first lane for each species corresponds to a positive control (C).

DISCUSSION

Development of DNA-based technologies to identify the species of origin of unknown samples has been the focus of many conservation, wildlife, and forensic biologists. Yet, financial and technical limitations still hinder their wide and frequent use in conservation research and management plans (Broquet *et al.* 2007). In noninvasive studies, repeated PCR analyses and continuous monitoring of data quality is essential to ensure accurate data (Taberlet *et al.* 1996; Beja-Pereira *et al.* 2009), therefore, it is crucial to decrease the time and cost involved in each step of data production. In this context, our relatively inexpensive, easy and quick PCR-SSCP approach can be considerably useful, especially in studies of carnivores from South-western Europe. With a simple PCR and SSCP gel analysis, a minimum of 40 samples can be simultaneously compared, depending on the electrophoresis apparatus used. This methodological approach allows reducing 5 times the costs relatively to direct sequencing. Thus, for long-term and/or large-scale monitoring programs, this method can be both faster and cheaper than sequencing, while much more straightforward than possible RFLP tests that require several steps (e.g., PCR, restriction endonuclease digestion etc.) for distinguishing all species. Nevertheless, sequencing the small (221 bp) IRBP fragment might be preferred in lower scale projects or when this kind of electrophoresis apparatus is not available.

Although mitochondrial assays are often thought to be more efficient than nuclear ones for noninvasive molecular studies (due to the higher copy number per cell), the IRBP nuclear fragment showed high amplification rates using both faecal (average 81.13%) and plucked hair (100%) samples, while overcoming limitations inherent to the use of mtDNA. Concomitantly, this method has several other potential advantages: (1) all samples under analysis can be sequenced or subjected to a PCR-SSCP without any prior assumptions based on morphologic identifications, which are often applied to noninvasive samples before PCR in order to reduce cost and time in multiple-amplifications approaches; (2) it helps solving the problem of false negatives in species-specific PCRs without the need to co-amplify an additional fragment to control PCR success (e.g., Zaidi *et al.* 1999; Mukherjee *et al.* 2007); (3) while in species-specific assays negative results for the target species do not allow the immediate identification of the true predator and further experiments are needed (Palomares *et al.* 2002; Fernandes *et al.* 2007; O'Reilly *et al.* 2008), in the described assay a single positive PCR-SSCP will directly give the final species identification and (4) DNA mixtures can be easily detected if multiple SSCP patterns appear in a single sample, while species-specific protocols or sequencing will not detect or easily decipher possible contaminations, respectively. Although the presence of contaminating DNA does not influence species ID when specific primers are used, further amplifications using other molecular markers might be sensitive to the non-detected contaminations and result in wrong genotyping. Accordingly, we consider that the capacity of this assay to individuate target species from prey DNA should be viewed also as a tool for screening DNA samples for following research. When compared to protocols based on mtDNA polymorphism, another advantage of our method, that may significantly reduce the time and cost of carnivore monitoring programs, is the fact that it can also be used as a pre-screening tool to identify samples with amplifiable nuclear DNA. In particular, the amplification success of this fragment allows selecting samples with sufficient nuclear DNA for further genetic studies requiring, for example, the amplification of microsatellites or other nuclear SNPs.

Species identification using partially degraded DNA, as the one extracted from noninvasive samples, should rely on the amplification of short DNA targets (Taberlet *et al.* 1999; Morin *et al.* 2001; Broquet *et al.* 2007). Unfortunately, most nuclear genes have low mutation rates and therefore long fragments (500 - 1000 bp) are usually required to detect enough variable sites for several species identification. Given the small size of this highly variable nuclear fragment (<250 bp), analysis of DNA from noninvasive samples, such as faeces or hairs, was notably efficient (81.13% of identification success) and in the same level to success rates reported for mtDNA fragments (e.g., 72%, Fernandes *et al.* 2007). The amplification was successful in samples subject to different degradation factors in the field (e.g., high vs. low temperature and humidity) since samples were collected across the entire Iberian Peninsula in distinct Mediterranean and Atlantic landscapes. These findings demonstrate the applicability of this molecular method in a range of conditions that are common in noninvasive studies. The small size of the fragment also made it suitable for SSCP analysis, since the optimal amplicon size for detection of point mutations is around 200 bp (Orita *et al.* 1989; Ortí *et al.* 1997). Due to its relatively short length and high variability, this genomic portion is also highly promising for routine forensic applications.

Among noninvasive samples, faeces are often the most widely used but contain the highest concentration of non- target DNA, namely from bacteria and prey items (Bradley and Vigilant 2002). At the same time, PCR inhibitors are present in scats in higher concentrations than other sample types, e.g., hair samples. In faecal samples, inhibitors may also vary according to predator's diet (Murphy *et al.* 2003). Nevertheless, our assay did not show variation related to species or samples types. No confusion was introduced in species identification in cases of prey items contamination, which suggests that even if prey DNA is co-extracted with the predator DNA, we will still correctly identify the predator.

The SSCP is usually considered a simple molecular technique (e.g., Sunnucks *et al.* 2000). In fact, due to its straightforwardness, high efficiency and low price, this PCR-SSCP test provides an excellent universal protocol for identifying all sympatric carnivores from South-western Europe, and likely other continental regions around the world. Both analysed under direct sequencing or PCR- SSCP assays, this diagnostic tool may help to improve our knowledge on carnivores' distribution and population status in areas such as the Iberian Peninsula, which represents an important hotspot of biodiversity where many species are still poorly studied. As an example, it has already been successfully applied in the study of Portuguese populations of European wildcat and red fox, revealing high efficiency in solving erroneous identifications based on scats morphology alone (Castro *et al.* unpublished data). In that study, a total of 95 noninvasive samples from both species were submitted to IRBP identification, with an average success of 76.52%.

Molecular species ID is also an important means to identify species protected by legislation such as the Convention for International Trade of Endangered Species (CITES), and for detecting and monitoring poaching, illegal harvest and trade of protected species. DNA-based species identification has helped detect fraudulent trade of several species and illegal wild animal hunting (An *et al.* 2007; for more examples see Allendorf and Luikart 2007, Chap. 20). In this context, the IRBP fragment here analysed can be easily applied for controlling and monitoring illegal trade and hunting of the endangered carnivore species from South-western Europe. Finally, its high level of polymorphism and discriminatory power to distinguish species from the same family (and also its capacity to differentiate some prey species included in our study,

data not shown), suggest that IRBP constitute a good candidate gene to further distinguish other species around the world, in order to improve the description of biodiversity within the barcoding framework, which still presents the limitation of being based solely on mtDNA (Moritz and Cicero 2004; Rubinoff 2006).

ACKNOWLEDGMENTS

We thank N. Ferrand and R. Faria for com- ments on earlier versions of this manuscript. We are grateful to Portuguese National Tissues Bank/National Institute for Nature and Biodiversity Conservation (BTVS/ICN-B), P. Monterroso, F. Alvares, B. Gomez-Moliner, C. Gortazar, J. Searle, F. Suchentrunk, E. Randi, J. Godoy, P. Ferreras, A. Ruiz-Gonza´lez, C. Ferreira, J. Vicente, Z. Boratynski, T. Perez and J. Paupério for providing samples. This work was partially financed by the research project PTDC/CVT/71683/2006, from FCT (Fundaçãoo para a Ciência e a Tecnologia). R. Oliveira and R. Godinho worked under PhD (SFRH/BD/24361/2005) and Post-Doc (SFRH/BPD/36021/2007) grants, respectively, both financed by FCT. D. Castro was supported by a CIBIO grant and G. Luikart was supported by the Portuguese-American Foundation for Development, CIBIO and UP.

APPENDIX

Appendix 1. PCR-SSCP results obtained for the 52 potential carnivore preys analysed in this study.

Group	Common name	Species	N	PCR-SSCP
Mammals	Wild boar	*Sus scrofa scrofa*	4	Yes
	Domestic pig	*Sus scrofa domesticus*	4	Yes
	Domestic cattle	*Bos taurus*	4	No
	Domestic sheep	*Ovis aries*	5	No
	European roe deer	*Capreolus capreolus*	2	Yes[a]
	Wild rabbit	*Oryctolagus cuniculus*	6	Yes[a]
	Iberian hare	*Lepus granatensis*	4	Yes[a]
	Brown hare	*Lepus europaeus*	4	Yes[a]
	Broom hare	*Lepus castroviejoi*	4	Yes[a]
	Horse	*Equus caballus*	5	Yes
	Donkey	*Equus asinus*	2	Yes
	Wild goat	*Capra pyrenaica*	2	No
	Domestic goat	*Capra hircus*	2	No
	Mouse	*Mus sp.*	2	Yes
	Wood mouse	*Apodemus sylvaticus*	2	Yes
	Greater white-toothed shrew	*Crocidura russula*	2	Yes
	Crowned shrew	*Sorex coronatus*	1	Yes
	Pygmy shrew	*Sorex minutus*	2	Yes
	Spanish shrew	*Sorex granarius*	1	Yes
	Pygmy white-toothed shrew	*Suncus etruscus*	1	Yes
	Water shrew	*Neomys sp.*	1	Yes

Amphibians and reptiles	Lataste's viper	*Vipera latastei*	2	No
	Asp viper	*Vipera aspis*	2	No
	Iberian adder	*Vipera seoanei*	2	No
	Schreiber's green lizard	*Lacerta schreiberi*	5	No
	Ocellated lizard	*Lacerta lepida*	1	No
	Viperine water snake	*Natrix maura*	1	No
	Common toad	*Bufo bufo*	1	No
	Natterjack toad	*Bufo calamita*	1	No
	Sharp-ribbed newt	*Pleurodeles waltl*	1	No
	Large psammodromus	*Psammodromus algirus*	1	No
Birds	Red-legged partridge	*Alectoris rufa*	2	No
	Chukar partridge	*Alectoris chuckar*	2	No
	Rock partridge	*Alectoris graeca*	2	No
	Common quail	*Coturnix coturnix*	2	No
	Japanese quail	*Coturnix japonica*	2	No
	Eurasian scops owl	*Otus scops*	2	No
	Common barn Owl	*Tyto alba*	5	No
	Eurasian woodcock	*Scolopax rusticola*	5	No
	Robin	*Erithacus rubecula*	5	No
	Domestic chicken	*Gallus gallus*	5	No
	Common blackbird	*Turdus merula*	1	No
	Song thrush	*Turdus philomelos*	1	No
	Redwing	*Turdus iliacus*	1	No
	Rock pigeon	*Columba livia*	2	No
	Turtle dove	*Streptopelia turtur*	2	No
Fishes and crustaceous	Rainbow trout	*Oncorhynchus mykiss*	5	No
	Pike	*Esox lucius*	2	No
	Gudgeon	*Gobio gobio*	2	No
	Carp	*Cyprinus carpio*	2	No
	Pumpkinseed	*Lepomis gibbosum*	2	No
	Louisiana crayfish	*Procambarus clarkii*	2	No

[a] Faint amplification that did not produce a clear SSCP pattern

REFERENCES

Adams J, Waits L (2007) An efficient method for screening faecal DNA genotypes and detecting new individuals and hybrids in the red wolf (*Canis rufus*) experimental population area. *Conserv Genet* **8**:123-131.

Allendorf FW, Luikart G (2007) Conservation and the genetics of populations. Blackwell, Oxford

Alves PC, Harris DJ, Melo-Ferreira J, Branco M, Suchentrunk F, Boursot P, Ferrand N (2006) Hares on thin ice: introgression of mitochondrial DNA in hares and its implications for recent phylogenetic analyses. *Mol Phylogenet Evol* **40**:640-641.

An J, Lee M, Min M, Lee M, Lee H (2007) A molecular genetic approach for species identification of mammals and sex determination of birds in a forensic case of poaching from South Korea. *Forensic Sci Int* **167**:59 - 61.

Antunes A, Pontius J, Ramos MJ, O'Brien S, Jand Johnson WE (2007) Mitochondrial introgressions into the nuclear genome of the domestic cat. *J Hered* **98**(5):414-420.

Arrendal J, Vilá C, Bjorklund M (2007) Reliability of noninvasive genetic census of otters compared to field censuses. *Conserv Genet* **8**:1097-1107.

Ballard JWO, Whitlock M (2004) The incomplete natural history of mitochondria. *Mol Ecol* **13**:729-744.

Beja-Pereira A, Oliveira R, Alves PC, Schwartz MK and Luikart G (2009) Advancing ecological understandings through technological transformations in noninvasive genetics. *Mol Ecol Resour* **9**(4): 1279-1301.

Berry O, Sarre SD (2007) Gel-free species identification using melt-curve analysis. *Mol Ecol Notes* **7**:1-4.

Bradley BJ, Vigilant L (2002) False alleles derived from microbial DNA pose a potential source of error in microsatellite genotyping of DNA from faeces. *Mol Ecol Notes* **2**:602-605.

Broquet T, Ménard N, Petit E (2007) Noninvasive population genetics: a review of sample source, diet, fragment length and microsatellite motif effects on amplification success and genotyping error rates. *Conserv Genet* **8**:249-260.

Dalén L, Gotherstrom A, Angerbjorn A (2004) Identifying species from pieces of faeces. *Conserv Genet* **5**:109-111.

Davison A, Birks JDS, Brookes RC, Braithwaite TC, Messenger AE (2002) On the origin of faeces: morphological versus molecular methods for surveying rare carnivores from their scats. *J Zool* **257**:141-143.

Farrell LE, Romant J, Sunquist ME (2000) Dietary separation of sympatric carnivores identified by molecular analysis of cats. *Mol Ecol* **9**:1583-1590.

Fernandes CA, Ginja C, Pereira I, Tenreiro R, Bruford MW, Santos-Reis M (2007) Species-specific mitochondrial DNA markers for identification of non-invasive samples from sympatric carnivores in the Iberian Peninsula. *Conserv Genet* **9**(3):681-690.

Flynn JJ, Nedbal MA (1998) Phylogeny of the Carnivora (Mammalia): congruence vs incompatibility among multiple data sets. *Mol Phylogenet Evol* **9**:414-426.

Gerloff U, Schlotterer C, Rassmann K, Rambold I, Hohmann G, Fruth B, Tautz D (1995) Amplification of hypervariable simple sequence repeats (microsatellites) from excremental DNA of wild living bonobos (*Pan paniscus*). *Mol Ecol* **4**:515-518.

Gómez-Moliner BJ, Cabria MT, Rubines J, Garin I, Madeira MJ, Elejalde A, Aihartza J, Fournier Pand Palazón S (2004) PCR-RFLP identification of mustelid species: European mink (*Mustela lutreola*),

American mink (*M. vison*) and polecat (*M. putorius*) by analysis of excremental DNA. *J Zool* **262**:311-316.

Hansen MM, Jacobsen L (1999) Identification of mustelid species: otter (*Lutra lutra*), American mink (*Mustela vison*) and polecat (*Mustela putorius*), by analysis of DNA from faecal samples. *J Zool* **247**:177-181.

Hogan FE, Cooke R, Burridge CP, Norman JA (2008) Optimizing the use of shed feathers for genetic analysis. *Mol Ecol Resour* **8**:561-567.

Hoss M, Kohn M, Paabo S (1992) Excremental analysis by PCR. *Nature* 359:199.

Hsieh H-M, Chiang H-L, Tsai L-C, Lai S-Y, Huang N-E, Linacre A, Lee JC-I (2001) Citochrome b gene for species identification of the conservation animals. *Forensic Sci Int* **122**(1):7-18.

Jones R, Cable J, Bruford MW (2008) An evaluation of non-invasive sampling for genetic analysis in northern European reptiles. *Herpetol J* **18**:32-39.

Karlsson AO, Holmlund G (2007) Identification of mammal species using species-specific DNA pyrosequencing. Forensic *Sci Int* **173**:16-20.

Kim JH, Antunes A, Luo SJ, Menninger J, Nash WG, O'Brien SJ, Johnson WE (2006) Evolutionary analysis of a large mtDNA translocation (numt) into the nuclear genome of the Panthera genus species. *Gene* **366**(2):292-302.

Kitano T, Umetsu K, Tian W, Osawa M (2007) Two universal primer sets for species identification among vertebrates. *Int J Legal Med* **121**:423-427.

Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular volutionary genetics analysis and sequence alignment briefings. *Bioinformatics* **5**:150-163.

Kurose N, Masuda R, Tatara M (2005) Fecal DNA Analysis for identifying species and sex of sympatric carnivores: a noninva- sive method for conservation on the tsushima islands, Japan. *J Hered* **96**(6):688-697.

Long RA, Mackay P, Ray J, Zielinski WJ (2008) Noninvasive survey methods for carnivores. Island Press, Washington, DC.

Lopez JV, Cevario S, O'Brien SJ (1996) Complete nucleotide sequences of the domestic cat (*Felis catus*) mitochondrial genome and a transposed mtDNA tandem repeat (Numt) in the nuclear genome. *Genomics* **33**:229-246.

Lucentini L, Vercillo F, Palomba A, Panara F, Ragni B (2007) A PCR- RFLP method on faecal samples to distinguish *Martes martes, Martes foina, Mustela putorius* and *Vulpes vulpes*. *Conserv Genet* **8**(3):757-759.

Mills LS, Pilgrim KL, Schwartz MK, McKelvey K (2000) Identifying lynx and other North American felids based on mtDNA analysis. *Conserv Genet* **1**:285-288.

Morin PA, Chambers KE, Boesch C, Vigilant L (2001) Quantitative PCR analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzes (*Pan troglodytes verus*). *Mol Ecol* **10**:1835-1844.

Moritz C, Cicero C (2004) DNA barcoding: promise and pitfalls. *PLoS Biol* **2**(10):e354.

Mukherjee N, Mondol S, Andheria A, Ramakrishnan U (2007) Rapid multiplex PCR based species identification of wild tigers using non- invasive samples. *Conserv Genet* **8**(6):1465-1470.

Murakami T (2002) Species identification of mustelids by comparing partial sequences on mitochondrial DNA from fecal samples. *J Vet Med Sci* **64**(4):321-323.

Murphy MA, Waits LP, Kendall KC (2003) The influence of diet on faecal DNA amplification and sex identification in brown bears (*Ursus arctos*). *Mol Ecol* **12**:2261-2265.

Nagata J, Aramilev VV, Belozor A, Sugimoto A, McCullough DR (2005) Fecal genetic analysis using PCR-RFLP of cytochrome b to identify sympatric carnivores, the tiger *Panthera tigris* and the leopard *Panthera pardus*, in far eastern Russia. *Conserv Genet* **6**:863-865.

O'Reilly C, Statham M, Mullins J, Turner PD, O'Mahony D (2008) Efficient species identification of pine marten (*Martes martes*) and red fox (*Vulpes vulpes*) scats using a 50 nuclease real-time PCR assay. *Conserv Genet* **9**:735-738.

Orita M, Iwahana H, Kanazawa H, Hayashi K, Sekiya T (1989) Detection of polymorphisms of human DNA by gel electropho- resis as single-strand conformation polymorphisms. *Proc Natl Acad Sci USA* **86**:2766-2770.

Ortí G, Hare MP, Avise JC (1997) Detection and isolation of nuclear haplotypes by PCR-SSCP. *Mol Ecol* **6**:575-580.

Palomares F, Godoy JA, Piriz A, O'Brien J, Jonson E (2002) Faecal genetic analysis to determine the presence and distribution of elusive carnivores: design and feasibility for the Iberian lynx. *Mol Ecol* **11**:2171-2182.

Paneto GG, Martins JA, Longo LV, Pereira GA, Freschi A, Alvarenga VL, Chen B, Oliveira RN, Hirata MH, Cicarelli RM (2007) Heteroplasmy in hair: differences among hair and blood from the same individuals are still a matter of debate. *Forensic Sci Int* **173**:117-121.

Paxinos E, McIntosh C, Ralls K, Fleischer R (1997) A noninvasive method for distinguishing among canid species: amplification and enzyme restriction of DNA from dung. *Mol Ecol* **6**:483-486.

Pilot M, Gralak B, Goszczyn′ ski J, Posłuszny M (2007) A method of genetic identification of pine marten (*Martes martes*) and stone marten (*Martes foina*) and its applications to faecal samples. *J Zool* **271**:140-147.

Riddle AE, Pilgrim KL, Mills LS, Mckelvey KS, Ruggiero LF (2003) Identification of mustelids using mitochondrial DNA and non-invasive sampling. *Conserv Genet* **4**:241-243.

Rubinoff D (2006) DNA barcoding evolves into the familiar. *Conserv Biol* **20**:1548-1549.

Ruiz-González A, Rubines J, Berdión O, Gómez-Moliner BJ (2008) A non-invasive genetic method to identify the sympatric mustelids pine marten (*Martes martes*) and stone marten (*Martes foina*): preliminary distribution survey on the northern Iberian Peninsula. *Eur J Wildl Soc* **54**:253-261.

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, vol 2. Cold Spring Harbour Laboratory, Cold Spring Harbor.

Sastre N, Francino O, Lampreave G, Bologov VV, López-Martín JM, Sánchez A, Ramírez O (2009) Sex identification of wolf (Canis lupus) using non-invasive samples. *Conserv Genet* **10**(3):555-558.

Sundqvist A-K, Ellegren H, Vilá C (2008) Wolf or dog? Genetic identification of predators from saliva collected around bite wounds on prey. *Conserv Genet* **9**(5):1275-1279.

Sunnucks P, Wilson ACC, Beheregaray LB, Zenger K, French J, Taylor AC (2000) SSCP is not so difficult: the application and utility of single-stranded conformation polymorphism in evolu- tionary biology and molecular ecology. *Mol Ecol* **9**:1699-1710.

Taberlet P, Griffin S, Goossens B, Questiau S, Manceau V, Escaravage N, Waits LP, Bouvet J (1996) Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Res* **24**:3189-3194.

Taberlet P, Waits LP, Luikart G (1999) Noninvasive genetic sampling: look before you leap. *Trends Ecol Evol* **14**(8):323-327.

Tobe SS, Linacre AM (2008) A multiplex assay to identify 18 European mammal species from mixtures using the mitochondrial cyto- chrome b gene. *Electrophoresis* **29**(2):340-347.

Williams CL, Blejwas K, Johnston JJ, Jaeger MM (2003) A coyote in sheep's clothing: predator identification from saliva. *Wildl Soc Bull* **31**:1-7.

Zaidi RH, Jaal Z, Hawkes NJ, Hemingway J, Symondson WOC (1999) Can multiple-copy sequences of prey DNA be detected amongst the gut contents of invertebrate predators? *Mol Ecol* **8**:2081-2087.

Zhang DX, Hewitt GM (1996) Nuclear integrations: challenges for mitochondrial DNA markers. *Trends Ecol Evol* **11**:247-251.

CHAPTER 4

Improving the molecular toolbox

"There are those who claim that we carry useless DNA, but they're wrong. If there is something in our genes, there's a reason for it. We don't let things grow on us. I have tried to put irrelevant gene sequences into things as simple as bacteria. If it doesn't serve some purpose, the bacteria get rid of it right away. I assume that my body is at least as smart as bacteria when it comes to things like DNA."

Kary Mullis

Paper V. Oliveira R, Mattucci F, Lyons LA, Alves PC and Randi E

A DANGEROUS RETURN TO NATURE: ARE FREE-RANGING DOMESTIC CATS THREATENING THE GENETIC INTEGRITY OF EUROPEAN WILDCATS IN THEIR GENETICALLY DISRUPTED DISTRIBUTION? *In prep*

Paper VI. Oliveira R, Randi E, Mattucci F, Kurushima JK, Lyons LA and Alves PC

NUCLEAR GENOME SNPS TO DETECT EUROPEAN WILDCAT *(Felis silvestris silvestris)* AND DOMESTIC CATS *(Felis s. catus)* HYBRIDIZATION. *In prep*

A dangerous return to nature: are free-ranging domestic cats threatening the genetic integrity of European wildcats in their genetically disrupted distribution?

Rita Oliveira [1,2,†], Federica Mattucci[3†], Ettore Randi[3], Leslie Lyons[4] and Paulo Célio Alves[1,2]

[1] CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBio - Laboratório Associado, Campus Agrário de Vairão, 4485-661 Vairão, Portugal
[2] Departamento de Biologia, Faculdade de Ciências da Universidade do Porto, Porto, Portugal
[3] ISPRA, Istituto Superiore per la Protezione e la Ricerca Ambientale, Via Cà Fornacetta 9, 40064 Ozzano dell'Emilia, Bologna, Italia
[4] Department of Population Health and Reproduction, School of Veterinary Medicine, University of California-Davis, 1114 Tupper Hall, Davis, CA 95616, USA

[†] these authors contributed equally to this work

ABSTRACT

Since domestic cats have been introduced throughout all Europe they soon became feral and live, today, in sympatry with most wildcat populations. But while domestic cats expanded in number and range, habitat loss and fragmentation resulted in important demographic declines and high levels of isolation of the wild populations. In an evolutionary perspective, genome-wide consequences of both populations' fragmentation and artificial hybridization may strongly jeopardize wildcats' survival. Therefore, the development of adequate conservation management polices on wildcat populations implies evaluating their genetic variability and "purity". In this work we surveyed genetic variation at 38 unlinked microsatellites (37 autosomal and one X-located) in 1128 wild-living domestic, African and European wildcats from Portugal to Romania, representing most of the distribution range of wildcats in Europe. Through a variety of descriptive statistics, and multivariate and Bayesian analyses we: a) evaluated the differences between wild and domestic cats; b) estimated levels of population fragmentation; c) identified genetic signatures of past and recent bottleneck and d) described the frequency and geographical distribution of hybridization with domestic cats. Results confirmed that wild and domestic genotypes form two well-differentiated entities across most of the analysed locations, with the exception of Hungarian and Scottish cats. Genetic evidences of hybridization suggested, nevertheless, that cryptic hybrids might be present in several non-hybridizing subpopulations. Moreover, a strong genetic diversification was described among European wildcat populations, with as much as 10 independent clusters being identified. Using a simulation approach we provided compelling evidences that European-wide signals of crossbreeding can be confidently achieved with the analysed set of loci for what regards first generation of hybrids, but the slight difficulty in accurately identifying backcrosses with domestic cats suggest that we might still be underestimating hybridization levels. The evidences pointed out in this study may result from different historical, demographic and ecological conditions; therefore, future studies should focus on a better understanding of such variables.

Keywords: European wild and domestic cats, population structure, hybridization, microsatellites, clustering admixture analysis, conservation genetics

INTRODUCTION

The wildcat *Felis silvestris* is a polytypic species comprising six ecologically, geographically and genetically differentiated subspecies that inhabit the entire Old World (see Driscoll *et al.* 2007 for details). In Europe, three of them coexist: The European Wildcat *F. silvestris silvestris*, Schreber 1777, from Portugal to Romania; The African Wildcat *F. silvestris libyca*, Forster 1780, in the Mediterranean islands of Sardinia, Corsica and Crete (Randi and Ragni, 1991; Driscoll *et al*. 2007); and the domestic descendant of *libyca* cats, the domestic cat *F. silvestris catus,* that populates the entire continent. Archaeological remains suggest that the European subspecies probably appeared in the continent around 450,000-200,000 YA (Kitchener, 1991; Sommer & Benecke, 2006), descending from the Martelli's cat (*Felis lunensis* Martelli 1906), which was found in Europe during the early Pleistocene (Kitchener, 1991; Nowell & Jackson, 1996). The presence of African wildcats in Mediterranean islands is instead believed to be a consequence of human translocations at very early stages of domestication, probably less than 11o00 YA by Neolithic navigators (Vigne *et al*. 2012). Molecular studies based on both mitochondrial and microsatellite variation suggest that cats' domestication likely began when humans started to build the first civilizations over the Fertile Crescent (Driscoll *et al.* 2007; Lipinski *et al.* 2008), and the earliest evidence of a cat-human close relationship was found in Cyprus deposits from 10,600 YA (Vigne *et al.* 2012). Succeeding domestication, cats promptly colonized the entire world and became common in Europe mainly by taking advantage of major land and sea trade routes of Romans, Etruscans and Greeks (Clutton-Brock 1999).

Current patterns of European wildcats distribution and genetic variability are most probably a reflection of both natural and anthropogenic events. In one hand, the range shifts suffered during the climatic oscillations of the Pleistocene contributed to shape wildcat's demographic history and genetic diversity (Kitchener and Rees, 2009). On the other hand, strong demographic declines in the 18th and 19th centuries due to deforestation, intensive human persecution (Stahl & Artois, 1994), road kills (Nowell & Jackson 1996) and scarcity of major preys (e.g. Lozano *et al.* 2007; Monterroso *et al.* 2009) have been documented for most of the species range (Stahl & Artois, 1991; Nowell & Jackson 1996). Such drastic demographic changes may result in high levels of genetic fragmentation, exposing populations to reduced fitness, inbreeding depression and reduced capacity of adapting to environmental changes. It is therefore not surprising that one of the most common rescue-strategies proposed by conservation geneticists includes the increase of gene flow among populations (Pertoldi *et al.* 2007).

Concomitantly, the everlasting and widespread diffusion of free-ranging domestic cats in sympatry with wildcat ranges created the ideal conditions for crossbreeding and introgression of domestic alleles into wildcats' genomes, compromising the evolutionary resilience of European wildcats (Stahl & Artois, 1994; Nowel & Jackson 1996). Today, historical, environmental and behavioural issues that may influence processes of hybridization and introgression in European wildcat populations remain still to define, and the actual extent of interbreeding is controversial. During the last decade, several studies focused on the analysis of genetic diversity of European wildcats and in the assessment of hybridization with their domestic relatives, documenting genetic variation at mtDNA, allozymes and microsatellites loci (e.g. Beaumont *et al*. 2001, in

Scotland; Randi *et al*. 2001, in Italy; Pierpaoli *et al*. 2003, in 9 European Countries; Lecis *et al.* 2006, in Italy and Hungary; Oliveira *et al*. 2008a,b, in Iberian Peninsula; Eckert *et al*. 2010 and Hertwig *et al*. 2009, in Germany; O'Brien *et al*. 2009, in France). The levels of hybridization detected in the different studies varied from rare and limited in central and southern Europe (e.g. Germany, Italy and Iberia) to widespread and frequent in Scotland and Hungary, suggesting that wild-living cat populations might considerably vary in admixture proportions and any European-wide generalization must be avoided. One of the most far-reaching conclusions that can be drawn from the published studies is that the number of molecular markers is crucial for the resolution and robustness of admixture inferences, and they are still under the optimum for the precise identification of admixed cats (e.g. Oliveira *et al*. 2008). Moreover, the variability in sampling strategies, genotyping schemes and data analysis hampers the reasonable comparison among studies and, consequently, among populations. To effectively establish conservation actions for European wildcat populations is, therefore, mandatory to improve the molecular tools for the understanding of the entire genetic architecture and variation of European populations, while credibly identifying admixture events. Here, through multivariate clustering analysis and Bayesian-model approaches, we report the first range wide European study on wildcat populations with more than 1000 cats genotyped at 38 unlinked microsatellites. With such broad-scale analysis we intend to: i) accurately infer subtle population structure among European populations and understand in what extend are these populations genetically fragmented or isolated; ii) assess hybridization between domestic and European wild cats across the entire range of the European wildcat and estimate the level of introgression in natural populations; iii) improve the accuracy to detect hybrids between both subspecies.

MATERIAL AND METHODS

Sampling and laboratory procedures

We analysed a total of 1128 biological samples (tissue, blood, buccal swabs, hair and skin samples) from 686 putative European wildcats, 26 African wildcats, 329 domestic cats, 80 (possible) hybrids and 7 captive-bred hybrids from Italy. The European wildcat sample represents the majority of the species range in Europe, from Portugal to Romania, and was obtained by taking advantage of long-term and active European wide collaborations (see Table 1 and Figure 1 for details). All putative wildcats were previously morphologically identified by collectors, taking in consideration some or all of the following characteristics: wildcat phenotype, life history, cranial and intestinal indexes, stomach content and/or biometric indices (Schauenberg 1969, 1977; French *et al.* 1988; Ragni and Possenti 1996). Since random bred cats are the ones hypothetically crossbreeding with wild individuals, we restrained our domestic cat sampling to: i) feral cats that had no feeding or housing assistance from humans, ii) random bred cats having some influence by humans, and iii) cats owned and cared for by humans but not specifically from a breed. The 80 wild-living hybrid cats include all samples that have been indicated as possible hybrids according to morphologic data. African wildcats have been sampled in Sardinia, Corsica and North Africa. Total genomic DNA was isolated using standard phenol-chloroform (Sambrook and Russell 2006), high-salt methods (Sambrook *et al*. 1989)

or the QIAamp® DNA Micro Kit (Qiagen, California, USA), depending on the available quality and quantity of each sample. Two negative controls were included for each extraction run.

Table 1. Sampling size and location of all genotyped cats. Hybrid cats that have not been identified through phenotypic evaluation and that have been initially included in the parental populations are reported between brackets ("admixed genotypes").

Subspecies	Sampling locations	Acronym	n
Domestic cats	Italy	FCA	74(1)
F.s. catus, n=326	Poland	FCA	17
+(admixed genotypes; n=3)	Slovenia	FCA	7(1)
	Germany	FCA	31
	Switzerland	FCA	3
	Scotland	FCA	3
	Portugal	FCA	83
	Spain	FCA	79(1)
	Hungary	FCA	28
	Sardinia & Corsica	FCA	3
Putative European wildcats	Italy: NE Alps	NE Alps	75(1)
F.s. silvestris, n=628	Slovenia + Bosnia & Herz.	SLO+B&H	36(1)
+(admixed genotypes; n=58)	Bulgaria + Romania	BUL+ROM	12+1(3)
	Poland	POL	16(2)
	Italy: Tuscany	IT-Tus	24(12)
	Italy: C & S	IT-CS	98(13)
	Italy: Sicily	IT-Sic	11(2)
	Germany: E	Ger-E	49
	Germany: SW	Ger-SW	177(15)
	Belgium: Wallonia	BEL	16
	Luxembourg	LX	10(1)
	Switzerland	SWI	3
	Scotland	SCO	7
	Portugal: N & C	IP-N	11
	Spain: N& C	IP-N	27(2)
	Portugal: S	IP-SW	17(5)
	Spain: SE	IP-SE	27(1)
	Hungary	HUN	11
African wildcats	Africa: N	FLI	9
F.s. libyca, n=26	Sardinia & Corsica	FLI	17
Hybrids (or possible Hybrids)	NE Alps		6
F.s. catus x *F.s. silvestris,* n=87	Captivity (Italy)		7
	Italy: C & S		1
	Germany: SW		2
	Spain: N & C		1
	Hungary		59
	Scotland		11

Acronym = symbol used, in this study, to identify the different populations

All cats were individually genotyped at 37 autosomal and one X-located domestic cat microsatellites (Menotti-Raymond *et al*. 2003) that where chosen according to the assortment made by Lipinski *et al*. (2008)

based on the criteria of high heterozygosity, high polymorphism information content (P_{IC}) and wide chromosomal distribution. All loci were amplified in eight PCR multiplex reactions using the QIAGEN Multiplex PCR kit following the manufacturer's protocol (Supplementary Table S1). All forward primers sequences were modified to include an additional universal tail fluorescently labelled with 6-FAM, VIC, PET or NED dyes (Applied Biosystems, Foster City, CA, USA). A touch-down thermocycling protocol was performed as following: 95ºC for 15 min; 6 cycles of denaturation at 94ºC for 30s, primer annealing with temperatures between 62-57ºC, decreasing 1ºC every cycle for 60s, and sequence extension at 72ºC for 60s; followed by 25 cycles (35 for hair and museum samples) of the previous cycling protocol but with a permanent annealing temperature of 57ºC; eight additional cycles at 53ºC for labelled tails' incorporation and a final extension step at 60ºC for 30 min. All amplifications were performed in a total volume of 10μl using both an Applied Biosystems thermal cycler (GeneAmp® *PCR* System *9700)* and Bio-Rad thermal cyclers (MyCycler and iCycler). Hair and museum skin samples were amplified in four replicates and in separate rooms dedicated to low DNA-content samples. PCR products were, afterwards, separated by size on an ABI3130xl Genetic Analyzer (Applied Biosystems Inc.) and genotypes were determined in GENEMAPPER 4.1 (Applied Biosystems Inc.) by comparison with size standard fragments of Genescan-500 LIZ, ABI. All genotyping steps included the negative controls for extraction and PCR. Additionally, a reference positive control was always included to infer PCR success and to calibrate independent runs. The software MICROCHECKER (van Oosterhout *et al.* 2004) was used to assess the potential presence of null alleles, after Bonferroni correction. Additional laboratory details are available upon request.

Analyses of genetic diversity and differentiation

Microsatellites diversity was estimated separately for European wild (FSI), African wild (FLI) and domestic cats (FCA), excluding all admixed genotypes detected in the hybridization analyses and all cats from the highly hybridizing populations of Scotland and Hungary (see bellow). We also evaluated genetic diversity within each cluster that split the European wildcats in different subpopulations (see bellow). The ARLEQUIN software package version 3.5.1.2 (Schneider *et al* 2000; Excoffier and Lischer 2010) was computed to estimate allele frequencies, mean number of alleles per locus (Na) and the observed (H_O) and expected heterozygosities (H_E). The same software was used to assess significance of deviations from Hardy–Weinberg equilibrium (Markov chain length of 100 000 and 3000 dememorization steps) and pairwise linkage disequilibrium (100 initial conditions followed by 16000 permutations) for all locus-subpopulation combinations, based on the exact test of Guo & Thompson (1992). P-values were adjusted for multiple tests using a sequential Bonferonni correction (Rice 1989). Allelic richness (*Ar*) and Private alleles richness (*PAr*) were computed for each population following a rarefaction method that compensates for uneven sample sizes, as implemented in the software HP-Rare (Kalinowski 2005). Genetic differentiations among the predefined taxonomic groups and between pairs of subpopulations were investigated with pairwise F_{ST} (Weir & Cockerham's 1984) and R_{ST} (Slatkin 1995) measures as implemented in GENEPOP 4.1 and FSTAT 2.9.3, respectively. An analysis of molecular variance (AMOVA) on Euclidean pairwise genetic distances was computed using analogues of Wright's *F*-statistics, as implemented in GENALEX (Peakall and Smouse, 2006). AMOVA was conducted

among groups and within groups and the significance of these parameters were estimated by 10 000 permutations of the distance matrix.

In addition to low gene flow and fragmentation, genetic differentiation among populations can also be caused by local bottlenecks and founder events. We used two methodological approaches representing different temporal resolution and sensitivity to characterize timing and magnitude of possible genetic bottlenecks: i) the first is most effective at detecting very recent bottlenecks of low magnitude and postulate: a) reductions in effective population size cause a temporary excess of heterozygote genotypes relative to the number of alleles in the population (heterozygote excess test, Cornuet and Luikart 1997), and b) alleles with intermediate frequency will be most abundant because of rare allele loss during a bottleneck event (mode-shift test, Luikart *et al*. 1998); and ii) the second follows Garza and Williamson's (2001) *M*-ratio test, which is best suited for detecting more severe, older bottlenecks (up to 100 generations ago), where *M* is the ratio of the number of alleles (*K*) over the range in fragment sizes (*r*), which is predicted to decline in a bottleneck because the number of alleles should decrease faster than the range in fragment sizes. For the first method, the software BOTTLENECK 1.2.02 (Cornuet and Luikart, 1997) was computed assuming a two-phase mutational model (TPM; Luikart *et al* 1998) with 90% one-step mutations. The two-tailed Wilcoxon signed rank test was used to determine significance of the observed deviations (Cornuet and Luikart 1997). For the second, *M*-ratio was assessed using the software M_P_Val for each locus and averaging the value of *M* over loci (Garza and Williamson 2001). The significance of *M* was determined by comparing to a critical value (*Mc*), calculated from hypothetical populations in mutation-drift equilibrium using the program CRITICAL_*M* with 10,000 simulation replicates (Garza and Williamson 2001). We used a two-phase mutation model with 10% multi-step mutations, an average size of non one-step mutations of 3.5, and theta of 5 and 10 to evaluate the sensitivity of the method to this variable.

Population structure, individuals' assignment and admixture analyses

Bayesian analyses of population structure were implemented in STRUCTURE 2.3.3 (Pritchard *et al* 2000; Falush *et al* 2007; Hubisz *et al* 2009) to simultaneously: i) infer the level of genetic structure in the data (number of *K* clusters) and estimate the average proportion of membership (*Qi*) of all sampled populations to each inferred clusters; and ii) assign each multilocus genotype to one or more *K* clusters according to their posterior probability membership assignments (*qi* coefficients) and their 90% credibility intervals (CI). Each run of the program was replicated five times, with 10^4 burn-in followed by 10^5 MCMC iterations, using a *qi* threshold of 0.85 (see bellow) to assign each cat to a single cluster. The optimal number of clusters (*K*) was identified using ΔK and ΔF_{ST} statistics (Evanno *et al*. 2005) as implemented in CORRSIEVE 1.6.1 (Campana *et al*. 2011). All computations focusing on hybrids detection were performed combining the admixture and *F* models and both with or without any prior non-genetic information, while analyses of European wildcat populations' structure were run using the sampling location model (Hubisz *et al*. 2009). Non-genetic information was included: a) for all individuals, b) for all cats except the presumed hybrids detected in a).

To avoid any bias in the analyses of divergence between taxonomic groups and among European wildcat sub-populations, we started our computations by preliminarily identifying all genotypes with possible

hybrid ancestry. STRUCTURE was run with two different datasets to assign individuals into two possible populations (K=2): putative European wild cats (FSI) *versus* domestic cats (FCA), and putative African wildcats (FLI) *versus* domestic cats. At this point, all putatively admixed cats (n=74) were excluded from the dataset. Divergence among subspecies (FCA, n=294; FSI, n=610 and FLI, n=26) was, then, estimated using the full resulting database (n=930), while population sub-structuring in the European wildcat was investigated only for *F.s. silvestris* specimens (n=610). To identify the partition of FSI genotypes in main macroareas, the number of *K* possible clusters was forced to vary from 1 to 20. Within each of the detected macroareas (see Results), further substructure was assessed by varying K between 1 and 6. Patterns of genetic differentiation among cat subspecies and European wildcat populations (excluding all hybrids) were also explored by Discriminant Analysis of Principal Components (DAPC) and Principal Component Analysis (PCA) implemented in the ADEGENET package (Jombart 2008).

Finally, to overcome a possible bias in admixture results due to within-species population structuring, we performed thorough admixture analysis within each of the European wildcat macro areas. Cats' ancestry was computed using K=2 with prior population information (usepopinfo activated) for the domestic and wildcats that were genetically pre-identified in the first runs of STRUCTURE. Putatively admixed cats were investigated without any prior information. The same computations were used to examine evidences of admixture within African wildcats. Concomitantly, a more detailed analysis of hybrids proportion was accomplished with the Bayesian model-based method implemented in NEWHYBRIDS 1.1 (Anderson & Thompson, 2002), by inferring the posterior probability that admixed genotypes identified by STRUCTURE belong to six genotype frequency classes: wildcat (FSI); domestic cat (FCA); first (F_1) and second (F_2) generation hybrids; backcrosses with wildcat (BxFSI) and with domestic cats (BxFCA). The software was run with a burn-in period of 10^4 iterations followed by 10^5 iterations of the Monte Carlo Markov Chains, with "Uniform" priors for both mixing proportions and allele frequencies.

Bayesian approaches lack of any statistical legitimacy for evaluating the efficiency of models, loci and priors to analyse empirical data, so that simulations are needed to validate the inferences made in each particular study (Nielsen *et al*. 2006). To assess the power of our set of microsatellites for allocating individuals to populations and identifying admixed individuals, and to establish the range of *q* values that assign each genotype into one of the six different possible admixed generations, we simulated multilocus parental and hybrid genotypes using HYBRIDLAB 1.0 (Nielsen *et al.* 2006). For each macroarea, fifty domestic cats randomly selected in the entire dataset and fifty wildcats from that specific region were selected to generate 100 simulated genotypes of each parental type (FCA and FSI). Using these simulated genotypes as starting point, we then simulated 100 genotypes of each hybrid class: F1, F2 and respective first-generation backcrosses. Resulting genotypes were, afterwards, used to carry out admixture analyses with STRUCTURE using K= 2, under the same computation variables described above.

RESULTS

Genetic diversity at 37 autosomal unlinked microsatellites

All loci were polymorphic in the total sample of European wild (FSI, n=610), African wild (FLI, n=26) and domestic (FCA, n=294) cats, showing between 4 (locus FCA035, for FLI) and 32 alleles (locus FCA628, for FCA). Monomorphic loci were detected within some of the European wildcat subpopulations discriminated under clustering and Bayesian analyses (see bellow), namely FCA310 for FSI-1.1.2; FCA035, FCA088 and FCA678 for FSI-1.2.3; FCA305 for FSI-4.1 and FSI-4.2 and FCA310 for FSI-4.3. Across all loci and populations values of observed and expected heterozygosities ranged between 0.042 (locus FCA305, for FSI) to 0.885 (locus FCA023 for FLI) and 0.064 (locus FCA305, for FSI) to 0.911 (locus FCA628, for FLI), respectively. Global genetic diversity showed marked differences between the different subspecies, with FSI showing the lowest values for all measures analyzed except the average number of alleles, which was lower for FLI (see Supplementary Table 2 for details on variability per locus). *Ar* and *Par* estimates for 52 genes (correction accounting for FLI low sample size of 26 individual) varied between 8.07 (FSI), 9.25 (FCA) and 9.84 (FLI), and 1.39 (FSI), 1.19 (FCA) and 1.99 (FLI), respectively. The mean values of observed and expected heterozygosities were similar between domestic (H_O = 0.664 ±0.103; H_E = 0.778 ±0.098) and African wildcats (H_O = 0.701±0.132; H_E = 0.811±0.100), but lower for European wildcats (H_O = 0.582±0.158; H_E = 0.741±0.180). European wildcats and domestic cats both showed average H_O values significantly lower than expected, with F_{IS} values significantly higher than zero (0.206 and 0.147, respectively; $P<0.001$). Although few loci proved to significantly deviate from HW equilibrium even when each wildcat genetic cluster (Table 2) was analyzed separately, none of them showed significant deviations for all subpopulations. The same pattern was observed for LE estimates. The explanation for these sporadic departures might be population-related rather than locus specific effects. Both inbreeding or further population structure than the one statistically confirmed by our analyses are expected to influence European wildcat populations. In addition, domestic cats are inherently artificial populations that violate any assumption of panmitic populations. African wildcats were globally in HWE and LE, with only one locus (FCA649) showing significant deviations from equilibrium (Table2, Supplementary Table S1).

Table 2. Variability at 37 autosomal microsatellites for the three cat subspecies analysed (FCA, FSI and FLI) and for the European wildcat subpopulations detected in clustering analyses. All putative hybrids and admixed populations were excluded. *n* = sample size; N_A = mean number of alleles per locus; *Ar* and *PAr*= Allelic and Private Allelic richness; H_O, H_E = observed and expected heterozygosities; F_{IS} = inbreeding coefficient (* significant departures from HWE at $p < 0.001$, Bonferroni corrected); HWE and LE = number of tests out of equilibrium at 37 loci and 666 pairwise comparisons, respectively.

***Spp* or Population**[a]		**Code**[b]	***N***	N_A	***Ar***[c]	***Par***[c]	H_O	H_E	F_{IS}	**HWE**	**LE**
All	*F. s. catus*	FCA	294	14.950 (±4.770)	8.93	1.19	0.664 (±0.103)	0.778 (±0.098)	0.147*	23	8
	F. s. silvestris	FSI	610	14.108 (±3.211)	8.07	1.39	0.582 (±0.158)	0.741 (±0.180)	0.206*	37	207
	F. s. libyca	FLI	26	9.838 (±2.820)	9.84	1.99	0.701 (±0.132)	0.811 (±0.100)	0.134	1	0
Macro Area 1	*F. s. catus*	FCA-1	98	12.459 (±4.018)	7.18	0.6	0.684 (±0.110)	0.778 (±0.101)	0.121	6	2
	F. s. silvestris	FSI-1	273	11.60 (±2.630)	5.98	0.89	0.593 (±0.164)	0.724 (±0.172)	0.157*	25	23
	NE Alps, SLO+ B&H, BUL, POL	FSI-1.1	140	9.811 (±2.283)	5.87	0.49	0.621 (±0.171)	0.700 (±0.176)	0.109	3	2
	NE Alps, SLO+B&H	FSI-1.1.1	111	8.595 (±2.166)	5.27	0.10	0.625 (±0.177)	0.687 (±0.174)	0.086	2	1
	BUL+ROM, POL	FSI-1.1.2	29	7.167 (±2.396)	5.66	0.17	0.601 (±0.186)	0.728 (±0.149)	0.148	1	0
	IT no alps	FSI-1.2	133	9.622 (±2.265)	5.95	0.46	0.562 (±0.184)	0.697 (±0.176)	0.170*	15	2
	IT-Tsc	FSI-1.2.1	24	6.324 (±1.717)	5.33	0.11	0.543 (±0.203)	0.676 (±0.168)	0.182*	1	0
	IT-CS	FSI-1.2.2	98	8.784 (±2.225)	5.48	0.15	0.578 (±0.191)	0.691 (±0182)	0.157*	10	0
	IT-Sic	FSI-1.2.3	11	3.794 (±1.274)	3.57	0.12	0.500 (±0.229)	0.564 (±184)	0.122	0	0
Macro Area 2	*F.s. catus*	FCA-2	34	8.595 (±2,522)	6.32	0.40	0.617 (±0.137)	0.742 (±0.111)	0.166	0	1
	F.s. silvestris	FSI-2	255	10.514 (±3.106)	5.74	0.89	0.577 (±0.170)	0.712 (±0.189)	0.183*	30	116
	Ger-NE	FSI-2.1	49	5.917 (±2.298)	4.21	0.13	0.534 (±0174)	0.637 (±0.173)	0.100	5	3
	Ger-SW, LX, BEL, SWI	FSI-2.2	206	9.919 (±3.022)	5.64	0.13	0.590 (±0.180)	0.707 (±0.186)	0.281*	16	38
Macro Area 4	*F.s. catus*	FCA-4	162	12.216 (±4.008)	6.66	0.37	0.644 (±0.132)	0.764 (±0.094)	0.157*	15	1
	F. s. silvestris	FSI-4	96	10.622 (±2.890)	6.20	0.51	0.559 (±0.160)	0.751 (±0.157)	0.213*	18	14
	IP-N	FSI-4.1	38	8.139 (±2.072)	6.21	0.24	0.614 (±0.173)	0.757 (±0.122)	0.178*	7	3
	IP-SW	FSI-4.2	17	5.306 (±1.582)	4.61	0.1	0.568 (±0.212)	0.637 (±0.144)	0.096	0	0
	IP-SE	FSI-4.3	27	6.472 (±2.535)	5.29	0.18	0.549 (±0.182)	0.697 (±0.175)	0.203*	3	0

[a] Populations' acronyms are used as inTable 1;

[b] Codes correspond to the symbols used to discriminate the different genetic clusters (see below) that subdivide European subpopulations;

[c] *Ar* and *PAr* were obtained for 52 and 22 genes when comparing subspecies and macro areas, respectively.

The hierarchical AMOVA was performed by subdividing the non-hybridizing set of samples into the three taxonomic groups (European wildcats, African wildcats, and domestic cats) and by grouping European wildcat samples into subsequent partitions as detailed in Table 3. Although most of the variation was found within groups (between 84% and 92%), results reflect high genetic divergence. Genetic variability was significantly partitioned among taxonomic groups (Φ_{ST} = 0.158; F_{ST} = 0.115; R_{ST}=0.387) and among all genetically identified locations (Φ_{ST} = 0.127; F_{ST} = 0.078; R_{ST}=0.126), indicating that wildcats are subdivided into clearly distinct gene pools in Europe. Over all loci and populations, a substantial proportion of genetic variation was attributed to mutation (as measured by R_{ST}) especially when comparing the three cat subspecies (R_{ST}= 0.387; Table 3).

Table 3. Hierarchical Analysis of Molecular Variance (AMOVA) computed in ARLEQUIN using Φ_{ST} and values of F_{ST} and R_{ST} estimated under different perspectives of samples grouping.

Grouping	Populations	Source of	Variance	%	Φ_{ST} *	F_{ST}	R_{ST}
by cat subspecies		Among groups	6.887	16%	0.158	0.115	0.387
		Within groups	36.591	84%			
	FCA/ FSI/ FLI						
European wildcats by macro areas		Among groups	2.877	8%	0.077	0.045	0.081
		Within groups	34.534	92%			
	FSI-1/ FSI-2/ FSI-4						
European wildcats by major partition within macro areas		Among groups	4.314	12%	0.116	0.072	0.124
		Within groups	33.008	88%			
	FSI-1.1/ FSI-1.2/ FSI-2.1/ FSI-2.2/ FSI-4						
European wildcats by subpopulations		Among groups	4.728	13%	0.127	0.078	0.126
		Within groups	32.462	87%			
	FSI-1.1.1 / FSI-1.1.2 / FSI-1.2.1 / FSI-1.2.2 / FSI-1.2.3						
European wildcats within FSI-1		Among groups	4,693	13%	0.131	0,077	0,149
		Within groups	31.008	87%			
	FSI-1.2.1 / FSI-1.2.2 / FSI-1.2.3						
European wildcats within FSI-2		Among groups	5.647	15%	0.148	0,098	0,164
		Within groups	32,512	85%			
	FSI-2.1 / FSI-2.2						
European wildcats within FSI-4		Among groups	4.159	11%	0.109	0.076	0.078
		Within groups	33.977	89%			
	FSI-4.1 / FSI-4.2 / FSI-4.3						

* all Φ_{ST} values were highly significant at P<0.001

Pairwise F_{ST} and R_{ST} estimations of genetic differentiation between pairs of populations revealed a substantial partition of the European wildcat population into subpopulations (Table 4), with most of the comparisons resulting in divergence estimates above 0.05. R_{ST} values between wildcat subpopulations were in average 2-3 times the divergence obtained with F_{ST}, reflecting the importance of allele size differences in splitting the wildcat group. Results reflect also the higher genetic proximity between African wildcats and domestic cats, for which values of F_{ST} (from 0.046 to 0.074) and R_{ST} (from 0.107 to 0.132) were globally lower than values between African and European wildcats (F_{ST}=0.117-0.212; R_{ST}= 0.316-0.552; Table 4).

Table 4. Genetic divergence parameters (F_{ST}, below diagonal, and R_{ST}, above diagonal) for pairwise comparison between all the non-hybridizing European wildcat subpopulations and domestic cats sampled in our study. The last two columns report the genetic differentiation between each wildcat subpopulations and all domestic cats.

	FCA1	1.1.1	1.1.2	1.2.1	1.2.2	1.2.3	FCA2	2.1	2.2	FCA4	4.1	4.2	4.3	FLI	FCA F_{ST}	FCA R_{ST}
FCA1		0,407	0,338	0,228	0,310	0,282	0,040	0,351	0,447	0,006	0,279	0,258	0,252	0.107	-	-
1.1.1	0,144		0,078	0,188	0,159	0,264	0,485	0,140	0,119	0,432	0,147	0,293	0,189	0.518	0.146	0.387
1.1.2	0,120	0,039		0,125	0,140	0,277	0,404	0,090	0,076	0,379	0,118	0,266	0,201	0.448	0.122	0.345
1.2.1	0,127	0,095	0,066		0,037	0,106	0,290	0,166	0,165	0,275	0,046	0,145	0,090	0.316	0.128	0.245
1.2.2	0,130	0,078	0,058	0,042		0,095	0,401	0,170	0,160	0,346	0,061	0,136	0,088	0.410	0.131	0.304
1.2.3	0,202	0,170	0,151	0,138	0,098		0,316	0,259	0,200	0,335	0,178	0,279	0,219	0.355	0.201	0.299
FCA2	0,032	0,163	0,134	0,150	0,149	0,224		0,431	0,503	0,065	0,358	0,345	0,329	0.132	-	-
2.1	0,160	0,109	0,089	0,151	0,118	0,203	0,189		0,128	0,386	0,144	0,266	0,165	0.492	0.158	0.347
2.2	0,128	0,058	0,046	0,097	0,079	0,163	0,140	0,093		0,476	0,158	0,301	0,186	0.552	0.130	0.423
FCA4	0,011	0,160	0,135	0,140	0,144	0,215	0,046	0,172	0,144		0,317	0,300	0,299	0.137	-	-
4.1	0,110	0,057	0,040	0,070	0,054	0,148	0,130	0,095	0,061	0,122		0,089	0,058	0.397	0.112	0.284
4.2	0,170	0,142	0,143	0,179	0,139	0,266	0,194	0,195	0,160	0,185	0,104		0,107	0.343	0.172	0.272
4.3	0,142	0,098	0,086	0,123	0,081	0,159	0,165	0,122	0,102	0,154	0,041	0,108		0.378	0.143	0.262
FLI	0.046	0.162	0.132	0.146	0.144	0.212	0.074	0.195	0.147	0.053	0.117	0.163	0.154		0.048	0.121

Comparisons between observed and expected heterozygosities in BOTTLENECK provided no evidences of recent genetic bottlenecks among European wildcats, with loci fitting mutation-drift equilibrium under the TPM model (Table 5). The M-ratio test showed instead that the population from Eastern Germany (FSI-2.1) had an average M-value significantly lower than the critical values estimated for θ=5 (0.764) and θ=10 (0.746).

Table 5. Analyses of bottleneck signatures for each of the 10 wildcat subpopulations according to M-ratio (Garza and Williamson, 2001) and BOTTLENECK (Cornuet and Luikart, 1997) probability tests. Following author's suggestion, populations with less than 15 individuals were not analysed for recent bottlenecks (N/A).

Wildcat Population	N	M-ratio M	Critical M (θ=5)	Critical M (θ=10)	Bottleneck P<0.05
1.1.1	111	0,826	0,780	0,772	1,000
1.1.2	29	0,795	0,748	0,720	0,957
1.2.1	24	0,745	0,742	0,710	0,995
1.2.2	98	0,814	0,779	0,769	1,000
1.2.3	11	0,765	0,707	0.652	N/A
2.1	**49**	**0,711**	**0,764**	**0,746**	**0,928**
2.2	206	0,853	0,786	0,785	1,000
4.1	38	0,875	0,758	0,734	0,987
4.2	17	0,753	0,729	0,687	0,996
4.3	27	0,852	0,747	0,718	0,631

Population structure: subspecies and European populations' clustering

Bayesian clustering analyses performed in STRUCTURE to assign individuals into two possible populations clearly suggested the presence of two well-differentiated genetic clusters that sharply split putative European wildcats and domestic cats. All domestic cats were assigned to cluster I (hereafter referred as FCA) with an average proportion of membership Q_{FCA} = 0.967, while European wildcats sampled across the entire continent were assigned to cluster II (hereafter referred as FSI) with variable Q_{FSI} values: Portugal = 0.827; Spain = 0.866; Scotland = 0.465; Belgium, Luxembourg and Switzerland = 0.962; Germany = 0.954; Italy = 0.897; Slovenia = 0.965; Hungary = 0.460; and Bulgaria and Poland = 0.925. Average proportions of membership noticeably demonstrate the admixed nature of Scottish and Hungarian populations, while indicate that other European areas might be regarded as representatives of non-admixed wild populations. Among non-hybridizing populations, a total of 74 individuals have shown evidences of possible hybrid ancestry. Moreover, all 7 captive bred hybrids were confirmed as admixed cats (see below the detailed analysis of admixture patterns for these individuals). When comparing putative African wildcats (FLI) and domestic cats, distinction between the two biological groups was clear for K=2, with domestic cats assigning with average Q_{FCA}= 0.755 and African wildcats clustering with Q_{FLI}=0.812. However, no admixture inferences have been made for the *libyca* subspecies in this study, since no straightforward threshold value for individuals' assignment could be accurately inferred (data not shown).

After excluding all putatively admixed cats from the dataset, sub-structuring of European wildcats' populations was better achieved by the partition of the data into five major genetic clusters, which separate the subspecies into five main geographical macroareas: FSI-1 (South-Central and Eastern Europe); FSI-2 (Central Europe); FSI-3 (Scotland); FSI-4 (Iberian Peninsula) and FSI-5 (Hungary). By exploring further subdivision within the non-hybridizing macroareas (FSI-1, FSI-2 and FSI-4), additional substructuring could be markedly detected by first splitting the 3 regions in 5 differentiated clusters that segregate macroarea FSI-1 into two smaller groups and separates the Eastern German population from the other wildcats from central Europe: FSI-1.1 (Alps NE, Slovenia, Bulgaria and Polonia); FSI-1.2 (Italy); FSI-2.1 (Germany-NE); FSI-2.2 (Belgium, Luxembourg, Switzerland and Germany-SW) and FSI-4 (Iberia). Ultimately, European wildcat genotypes could be partitioned into 10 well-differentiated genetic clusters. Population structure inferences are summarized in Table 6 and Figure 1.

Table 6. Summary of the Bayesian analysis performed in STRUCTURE for the best *K* value obtained either a) analysing all populations together, either b) analysing each of the wildcat macroareas defined across Europe.

		Inferred clusters									
	Pop	1	2	3	4	5	1	2	3	4	5
	dataset	a) among all wildcat populations					b) within macroareas 1, 2 and 4				
FSI-1	Alps NE	0	0	0	**1.000**	0	0	0	0	**1.000**	0
	Slovenia	0	0	0	**1.000**	0	0	0	0	**1.000**	0
	Bulgaria	0.227	0.08	0	**0.693**	0	0	0	0	0.001	**0.999**
	Polonia	0.010	0.103	0	**0.887**	0	0	0	0	0.001	**0.999**
	Italy (Tuscany)	0	**0**	0	0	**1.000**	0	**0.916**	0.084	0	0
	Italy CS	0	0	0	0	**0.979**	0.023	0.020	**0.936**	0.020	0
	Italy (Sicily)	0.009	**0**	0	0	**1.000**	**1.000**	0	0	0	0
FSI-2	Germany E	0	0	**1.000**	0	0	**0.980**	0.020			
	Belgium	**0.992**	0	**0**	0	0	0	**1.000**			
	Luxembourg	**0.932**	0.068	0	0.002	0	0	**1.000**			
	Switzerland	**0.998**	0	0	0	0	0	**1.000**			
	Germany SW	**0.994**	0	0	0.006	0	0	**1.000**			
FSI-4	Iberia N	0	**1.000**	0	0	0	0	**1.000**	0		
	Iberia SW	0	**0.970**	0	0.030	0	**0.985**	0.006	0.009		
	Iberia SE	0	**1.000**	0	0	0	0	0.004	**0.996**		

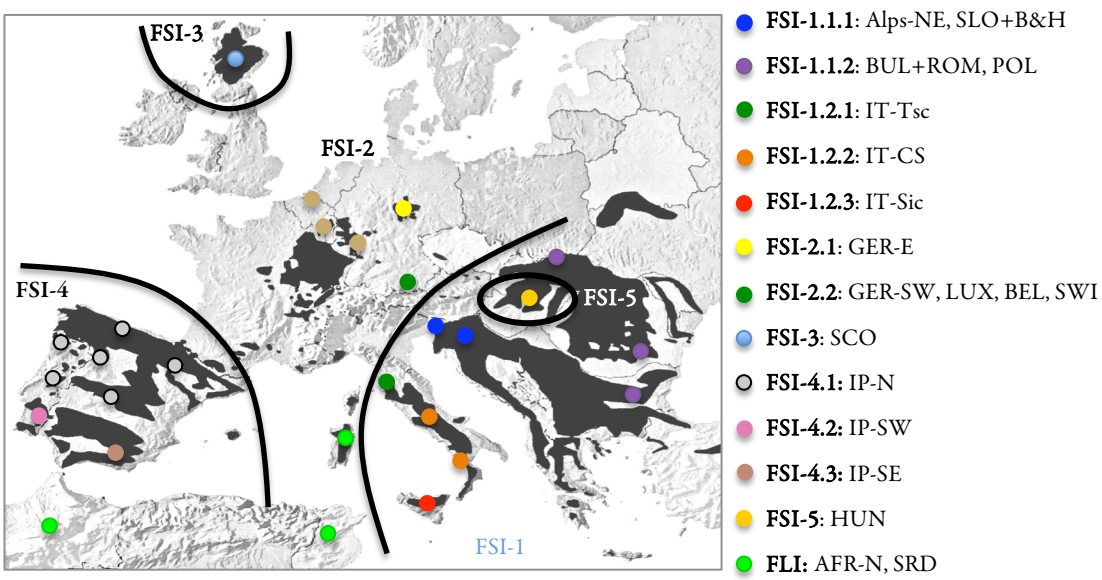


Figure 1. Approximate sampling locations of wildcats surveyed across Europe and North Africa. Colours represent the highest partition of samples into genetic clusters and black lines divide the major differentiated groups (macroareas) in the European wildcat, as identified by multivariate and Bayesian analyses. Dark areas in the map correspond to the approximate current distribution of *Felis silvestris* in Europe (adapted from Grabe and Worel 2001).

Patterns of genetic variation graphically summarized by DAPC scatter-plots (Figure 2) sharply distinguished *catus*, *silvestris* and *libyca* subspecies, reflected the closer genetic similarity between African wild and domestic cats and exposed a broader partition in the space for the African subspecies (Figure 2a). Moreover, multivariate clustering confirmed the evident divergence among European wildcat sub-populations (Figure 2b, c and d). The first Principal Component (PC) describes most of the genetic diversity among cat subspecies (Figure 2a) and NE Alps, Slovenia+Bosnia & Herzegovina, Bulgaria, Poland and Italy (Figure 2b). The first two PCs proved to be important in diversifying Iberian wildcat samples (Figure 2c) and subpopulations within the Italian sample (Figure 2b).

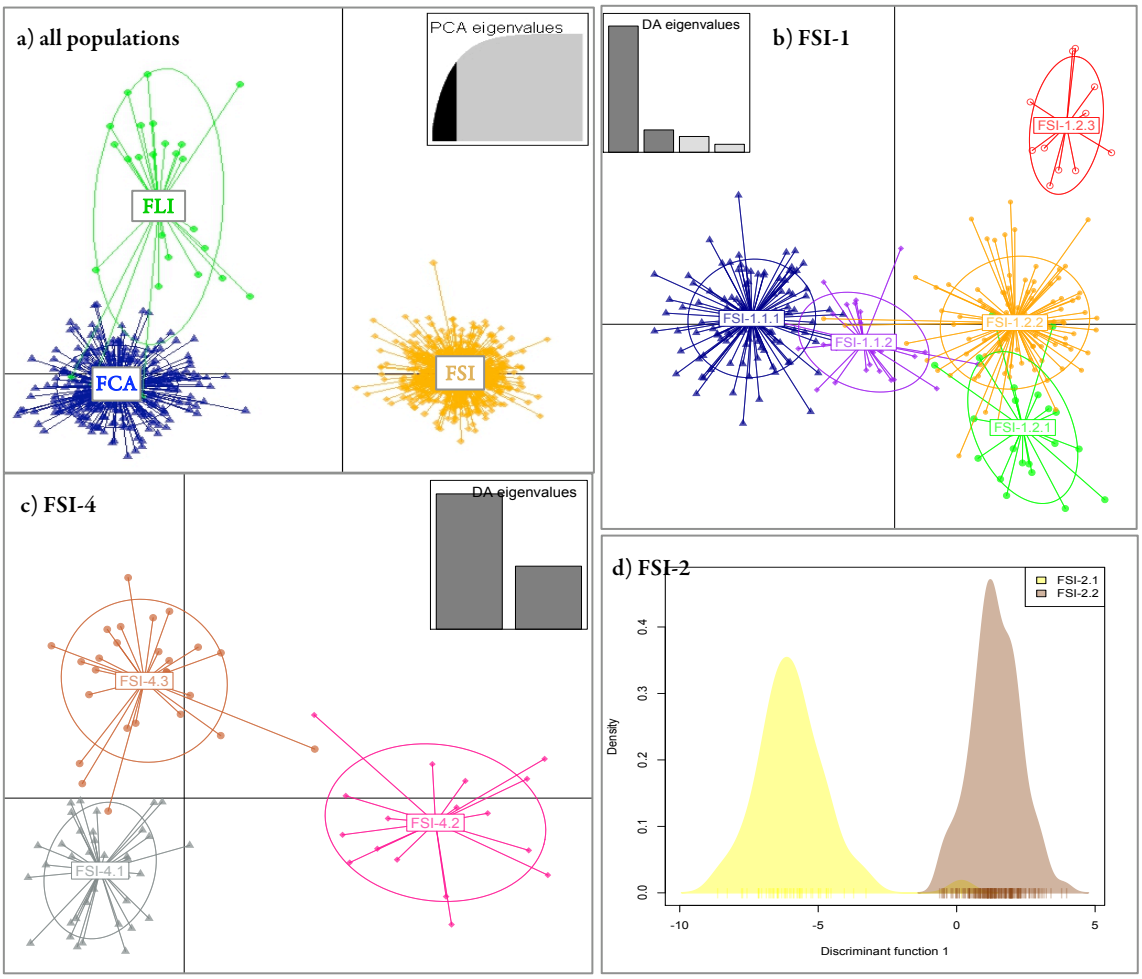


Figure 2. Plot of Discriminant Analysis of Principal Components (DAPC) obtained with ADEGENET, displaying genetic diversity among: a) European *Felis silvestris* subspecies; b) populations in South-Central and Eastern Europe; c) South-western populations (Iberian Peninsula) and d) Central European wildcats. Individuals (dots) and populations (coloured ellipses) are positioned on the space maximizing separation between groups.

Admixture analysis

The admixture pattern among European wildcat subpopulations was first accessed for the non-hybridizing ones, by computing the "admixture" model in STRUCTURE with $K = 2$ for each of the main macroareas 1, 2 and 4. NEWHYBRIDS' analyses were compared with these runs in order to obtain the most accurate estimates of hybridization. A conservative evaluation was performed to identify admixed genotypes by considering that only the samples partially assigned to both clusters in at least three of the four calculations (STRUCTURE without prior identifications, with prior identifications for all cats or with prior information for all cats except presumed hybrids; and NEWHYBRIDS without non-genetic information) could be considered true hybrids. For simplification, we discuss the computations without prior non-genetic information. Domestic and European wildcats were clearly partitioned between two genetic groups in STRUCTURE with: a) Q_{FCA} = 0.999 and Q_{FSI}= 0.929 for macroarea 1; b) Q_{FCA} = 1.000 and Q_{FSI}= 0.990 for macroarea 2; c) Q_{FCA} = 0.998 and Q_{FSI}= 0.984 for macroarea 4. At threshold qi < 0.85 (see below), 44 samples out of 434 were partially assigned to both clusters within macroarea 1, showing qi from 0.181 to 0.819 (90% CI between 0.048 – 0.952) to the wildcat cluster. From these individuals, 11 have been identified in NE Alps, Slovenia and Bosnia & Herzegovina; five in Bulgaria, Poland and Romania; and 28 in Italy (12 in Tuscany, 14 in the region of centre and south of the peninsula and 2 in Sicily. Among macroarea 2, a total of 18 out of 310 cats were probabilistically assigned as admixed cats, with *qi* values varying between 0.201 and 0.799 (range of 90% CI = 0.066-0.934) to the wildcat cluster. These hybrid genotypes have been identified in South-western Germany (n=17) and Luxembourg (n=1). Among macroarea 4, a total of 12 out of 258 samples have shown evidences of hybridization with domestic cats, showing *qi* between 0.157 and 0.843 (90% CI variable between 0.040 and 0.960). From these, six belong to the North and Centre of Iberian Peninsula; five to the South-western area (Portugal) and one to the South-eastern subpopulation (Spain). The entire sample of admixed genotypes included three individuals phenotypically identified as domestic cats, three samples for which hybridization was considered possible, seven cats identified as admixed and 61 individuals that were morphologically identified as European wildcats. All seven captive hybrids were confirmed as admixed with *qi* values varying between 0.195 and 0.707 (90% confidence intervals between 0.016 and 0.865) to the wildcat cluster. In total, 74 wild-living cats were identified as putative hybrids. NEWHYBRIDS' analyses partially confirmed STRUCTURE results, but nine of the putatively admixed cats where instead assigned to the parental classes when a threshold of 0.70 was considered: seven to the wild group and two to the domestic one. From the other 65 admixed genotypes, six have been identified as F1, 24 as F2, and five as BxFSI, while most of the hybrids ancestry remained unknown (30 out of 65, 46.15%; see Supplementary Table S2 for details).

Bayesian analyses of simulated genotypes within FSI-1, FSI-2 and FSI-4 in STRUCTURE, revealed that parental individuals should always be correctly assigned to their cluster of origin with an average Q of 0.952-0.962 to the domestic cluster and of 0.965-0.973 to the wildcat cluster, and that 90% CI should be above 0.85 in all cases (Supplementary Table S3a). Therefore, qi>0.85 was stipulated as our statistical limit to assign individuals to a single cluster. Average membership coefficients of simulated F1 and F2 genotypes was Q_{FCA}=0.503-0.516 (0.355<CI<0.659) and Q_{FCA}=0.494-516 (0.352-0.659), respectively, while backcrosses-simulated individuals have shown higher assignments to the respective parental clusters and wider 90% CI.

BxFCA showed an average membership coefficient of 0.745-0.768 to the domestic cluster, and CI between 0.601 and 0.890, and BxFSI were assigned to the wildcat cluster with Q_{FSI}= 0.740-0.756 and 0.611<CI<0.874. Individual assignment results suggest that our set of loci and samples provide 100% correct identifications of parental individuals, but that their accuracy power decreases for admixed genotypes. Anyway, F1 and F2 simulated genotypes could be correctly identified in 92-97% and 72-73% of the samples, respectively, and none of the non-identified cats was assigned to a different class. The identification of backcrosses proved to be less straightforward, with 12-14 BxFCA being identified as domestic cats and 7-12 BxFSI being identified as wildcats. The statistical confidence for the assignment of hybrid cats into specific classes performed in NEWHYBRIDS confirmed the high confidence in identifying parental wild and domestic cats, while revealing similar difficulties in precisely individuate hybrid classes of admixed genotypes (Supplementary Table S3b). Even with a less stringent threshold value 0.70, several individuals remained unclassified, with a maximum of 8 F1, 17 F2, 10 BxFCA and 7 BxFSI showing *qi* values subdivided between at least two possible hybrid classes. Misclassification of hybrids was only evident among simulated BxFCA, for which a maximum fraction of 4% could be assigned to the domestic cat group. BxFSI proved to be possibly assigned to the F1 and F2 hybrid classes, but none of the simulated genotypes was wrongly identified as a true wildcat. These findings suggest that the six cats displaying incongruent results between STRUCTURE and NEWHYBRIDS analyses, should most probably represent true wildcats weakly assigned by STRUCTURE analyses rather than be introgressed cats from past generations that could not be identified using our sampling and molecular schemes in NEWHYBRIDS. According to this interpretation, we might conclude that admixed evidences could be confirmed for 65 samples.

The genetic composition of cats in Hungary and Scotland, especially when compared to the other European realities, confirmed the particularly admixed nature of these populations (Figure 3 d, e and f). Individual assignment values were frequently intermediate between the wild and domestic clusters, with as much as 47.62% (n=10) and 37.76% (n=37) of the samples showing *qi* values between 0.15 and 0.85 in Scotland (n_{total}=21) and Hungary (n_{total}=98), respectively. These findings clearly contrast with the *qi* values observed among non-hybridizing populations, for which high proportions of membership to the wild ($0.00<q_{FCA}<0.15$) and domestic ($0.85<q_{FCA}<1.00$) clusters where obtained for 85.25% of the samples in macroarea 1, 92.26% in macroarea 2 and 95.35% in macroarea 4 (Figure 3 a, b and c). The plot of individual scores in the first two principal variables of a Principal Component Analysis (Figure 3f) also indicated the admixed nature of Hungarian and Scottish cats, with a high proportion of the samples plotting in an intermediate position between the European wild and domestic cat clusters.

As a result of their highly admixed composition, prior non-genetic identification of wildcats in Hungary and Scotland is strongly compromised. For example, among the 59 Hungarian samples morphologically identified as possible hybrids only 37 were genetic identified as admixed. It seems, therefore, that the accurate simulation of parental and hybrid genotypes within these populations would be highly problematic. Because of the documented hybridizing nature of these populations, most samples have also been georeferenced and morphologically identified according to detailed phenotypic indexes (B. Zirò, personal communication). Therefore, the examination of hybrids' ancestry will be extensively and multi-disciplinarily discussed elsewhere.

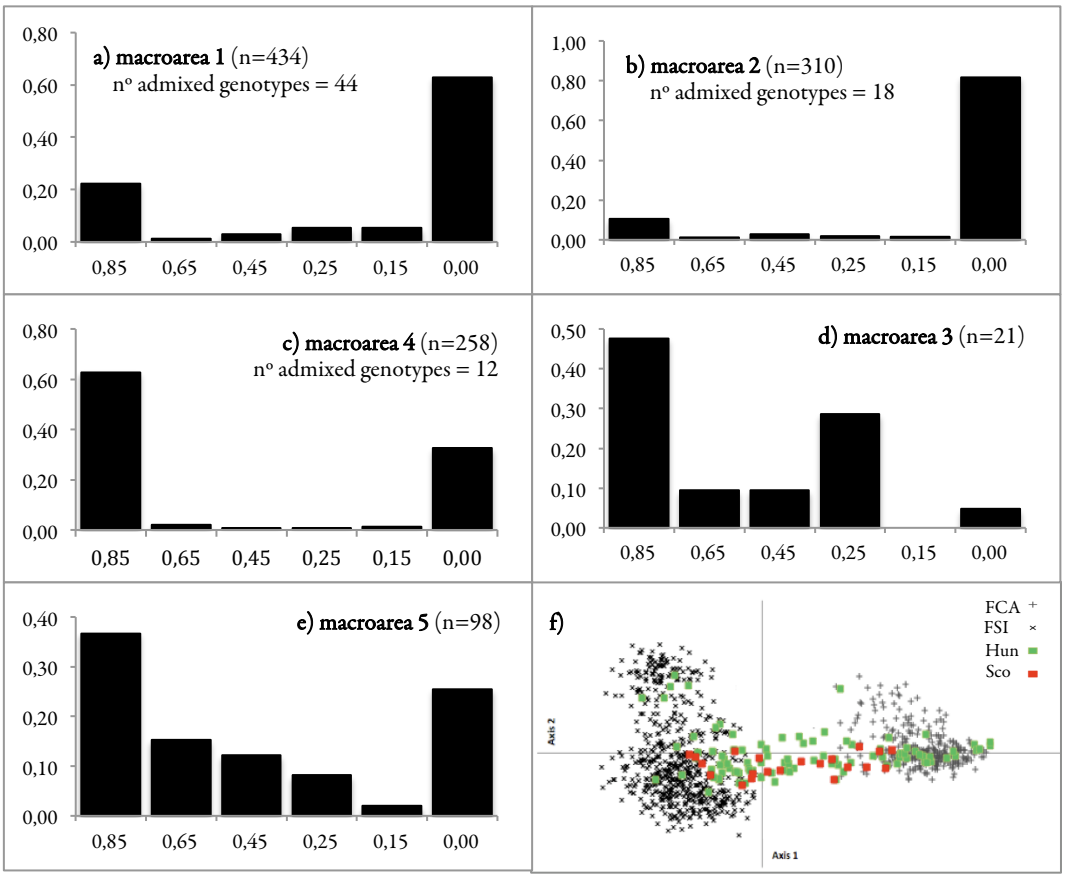


Figure 3. Plot of the percentage of individuals (y-axis) showing different membership proportions (x-axis) to the domestic cluster across the different macroareas identified in Europe: a) macroarea 1 (South-Central and Eastern Europe); b) macroarea 2 (Central Europe); c) macroarea 4 (South-western populations, i.e, Iberian Peninsula); d) Scotland (macroarea 3) and e) Hungary (macroarea 5). Values 0.85; 0.65; 0.45; 0.25; 0.15 and 0.00 identify cat samples showing, respectively, *qi* values to the domestic cluster between 1.00 and 0.85; 0.85 and 0.65; 0.65 and 0.45; 0.45 and 0.25; 0.25 and 0.15; and 0.15 and 0.00. f) Principal component analysis displaying multivariate clustering of Hungarian (Hun) and Scottish (Sco) cats relatively to all sampled European wild (FSI) and domestic cats (FCA).

DISCUSSION

In the last decade, a number of molecular studies have focused in the study of genetic diversity and hybridization patterns among European wildcats. Most recent research has taken advantage of the revolutionary advances in molecular and statistical technics faced by conservation genetics, and most populations across Europe have now been investigated. Examples of detailed analyses can be found for Iberian Peninsula (Oliveira *et al.* 2008), France (O'Brien *et al.* 2009), Italy (Randi *et al.* 2001), Germany (Hertwig *et al.* 2009; Eckert *et al.* 2010), Hungary (Lecis *et al.* 2006) and Scotland (Beaumont *et al.* 2001), where the analysis of mitochondrial variation and/or microsatellites diversity has suggested varying degrees of hybridization between wild and domestic cats. Until now, only Pierpaoli and colleagues have tried to draw the complete picture for the species across its European range by analyzing a total of 12 polymorphic

microsatellites in 336 cats from 9 different countries (Pierpaoli *et al*. 2003). However, the low number of putative wildcat samples in some of the analyzed regions (eg. Iberian Peninsula n=13, Hungary n=17) may have limited the deep analysis of population structure and demographic variation. Furthermore, the number of loci used to analyze admixture patterns (n=12) remained far from the advised numbers according to convincing simulations by Vähä and Primmer (2006), since as much as 48 loci with average F_{ST}=0.21 might be needed to distinguish F1, F2, backcrosses and parental individuals. While not exactly achieving these numbers, we used in this study 37 autosomal microsatellite loci (with average F_{ST} varying between 0.12 and 0.20 in European subpopulations) and one locus in the X Chromosome (FCA240) to: a) describe genetic variation at 1128 cats and evaluate population structure within European wildcats; b) investigate admixture patterns and geographical distribution of hybrids within each of the main macroareas inhabited by genetically diverse wildcats in Europe.

Genetic diversity and genetic consequences of populations' declines

Patterns of genetic diversity among European wild, African wild and domestic cats have shown that the three subspecies represent highly variable taxa, with the *libyca* cats displaying the higher values of allelic richness, private alleles and levels of heterozygosity (Table 2). Subsequent higher diversity was found for domestic cats, with the exception of private alleles richness that was higher for European wildcats. The high genetic variability found for African wildcats might be a results of one or both of the following intrinsic characteristics of the evolutionary history of the subspecies: a) in one hand *libyca* cats are known to occur in an extremely wide distributional range (Driscoll *et al.* 2007) and to display a very broad habitat tolerance (Driscoll & Nowell, 2010), which might have protected the species against strong past population declines and promote gene flow among populations; and b) past episodes of crossbreeding between domestic cats and their wild ancestors might have occurred in multiple occasions and for a long time during the process of domestication (Driscoll *et al.* 2007) and might have continuously support the maintenance of high genetic diversity. These same reasons might also in part explain the levels of variability found within European domestic cats, for which patterns of high diversity may, in fact, reflect a diversity of evolutionary origins and possible admixture. Results from AMOVA and pairwise-F_{ST} and R_{ST} estimates further confirm the close genetic proximity of *libyca* and *catus* subspecies, for which a clearly significant divergence was hard to obtain based on our set of loci. Instead, significant differentiation was detected between domestic cats and ten divergent subpopulations of European wildcats, showing that, despite of long co-existence, two entities are well differentiated in Europe and introgressive hybridization has not been the major factor shaping both wild and domestic cats' gene pools (but see Hungarian and Scottish populations). Since our sampling is composed of substantially different numbers of individuals per subpopulation (e.g. 11 samples in Sicily – FSI-1.2.3 - against 206 wildcats in Ger-SW, Luxembourg, Belgium and Switzerland – FSI-2.2), and deviations from HWE and LE for two of the detected subpopulations (North and Centre of Iberian Peninsula – FSI-4.1 – and FSI-2.2) might indicate that further substructure could be present at least in these regions, any comparison of subpopulations' "purity" could be misleading. Globally, our results confirm previous findings reported by Pierpaoli *et al*. 2003 and by studies within some European countries (Portugal and Spain,

Oliveira *et al*. 2008; France, O'Brien *et al*. 2009; Italy, Randi *et al*. 2001; and Germany, Hertwig *et al*. 2009 and Eckert *et al*. 2010), increasing today's confidence in the existence of true non highly hybridizing populations in Europe.

Since the amount of genetic variation is high within populations, but the populations are small and isolated, demographic declines and hybridization seem to pose a higher threat to the populations' persistence than strong inbreeding depression and low genetic variation. Carnivores such as the European wildcat are, in fact, considered to be particularly sensitive to population decline and local extinction (Gittleman *et al*. 2001), especially due to their low population densities, high generation times, evident sensibility to ecosystems modifications, and current exposition to important survival threats resulting from anthropogenic changes (Schipper *et al*. 2008). Among all the studied subpopulations, genetic signatures of population bottleneck were identified only for the Eastern population of Germany (FSI-2.1), confirming predictions by Pierpaoli *et al*. 2003 and Eckert *et al.* 2010 and contrasting the absence of such event advocated by Hertwig *et al*. 2009. Although no significant signs of recent heterozygosity excess could be detected in our study (as assessed by the procedure of Cornuet and Luikart, 1997), the 49 samples genotyped from this region displayed generally lower values of genetic diversity than other subpopulations (lower number of alleles, allelic richness, private alleles richness and heterozygosities; Table 2), and the ratio between the number of alleles and the range in allele size (M-ratio) suggested a significant reduction in effective population size. These results suggest that this population might have suffered past demographic declines that left detectable footprints on its genetic diversity. Although large population declines are known to have occurred across the entire species' range (Driscoll & Nowell, 2010), no evidences could be found for any of the other sampled subpopulations.

The loss of population fitness associated with a loss of genetic diversity is expected and well known from many studies of rare species (Frankham and Ballou 2003). Furthermore, a loss of genetic diversity may not be a problem over the short term but may reduce the ability of the population to evolve following future changes in environmental conditions or directly in the target species, as may occur under the scenario of introgressive hybridization. Therefore, the wildcat populations from Eastern German should be regarded as a conservation priority.

Range-wide population structure of European wildcats

Habitat fragmentation may disrupt original patterns of gene flow and lead to drift-induced differentiation among local population units. Top predators such as the wildcat may be particularly susceptible to this effect, given their low population densities, leading to small effective sizes in local fragments. Wildcats have a high dispersal rate and often disperse over long distances, suggesting that they may counteract this process and that there was probably little differentiation within European populations in the past. It is, however, evident that a significant decrease in gene flow and a recent increase in population fragmentation occurred. Our results indicate a clearly detectable genetic diversification among wildcat populations, with the entire sampling of non-hybridizing wildcats being partitioned at least in 10 well-defined genetic clusters. The first partition of the data in five macroareas (Table 6 and Figure 1) reflects the Bayesian clustering reported by Pierpaoli *et al*. (2003), that subdivided the European population in southern

and central Europe and separated the Eastern German population from all the other wildcat populations. However, our analysis clearly defined higher fragmentation of the European sampling, with the sharp separation of subpopulations within those clusters. Southern Europe could now be well differentiated in two main macroareas (Iberian Peninsula: FSI-4) and Italy, Slovenia and Bosnia&Herzegovina (FSI-1) with their respective further subdivisions (see Table 6 and Figures 1 and 2 for details). These results clearly demonstrate that larger sampling and better molecular definition was able to reveal additional information on populations' structure. It is, therefore, clear that Europe does not constitute a unique biogeographical unit for wildcats and at least five major evolutionary significant units (ESU's) and ten minor subpopulations should be recognized. It is however clear that some areas of the species distribution have maintained certain levels of reproductive contact, especially in Eastern Europe (Figure 1). Studying which mechanisms control gene flow among European wildcat populations would be particularly important for a better understanding of the results found in this study and, in general, for a better knowledge of the processes shaping wildcat's evolutionary history. While one may predict that a considerable proportion of today's fragmentation might result from habitat degradation and direct persecution (among others), many other geographical, historical and ecological factors may contribute to explain genetic differentiation among local wildcat populations. For example, ecological factors (climate, habitat types and diet composition) proved to strongly influence the amount of gene flow among European grey wolf populations, rather than topographic barriers or historical populations' fragmentation (Pilot *et al*. 2006). This example of a carnivore species that is also widely mobile, wide distributed and suffers from similar conservation threats (e.g. hybridization with domestic relatives) highlights the importance of further studies aimed at understanding the direct mechanism that links population ecology and population genetic structure in wildcats (Pilot *et al*. 2006).

For a better and more comprehensive evaluation of the reported genetic partition we further recommend the European-wide evaluation of phylogeographic patterns based on mtDNA diversity and nuclear sequencing in the near future. Although using mtDNA for studying hybridizing taxa might be prone to errors due to its uniparental inheritance, the fact that wildcats from the entire Europe are today genotyped for 38 microsatellites provides great confidence to select the "purest" wildcat samples for subsequent mtDNA analyses.

Detection of hybridization

For conservation biologists, discriminating the domestic cat from its wild progenitor has been and still is a long-standing dilemma. Due to the high similarity in morphology and genomes of wild and domestic cat forms, high number and type of molecular tools are needed for this discrimination. The high difficulty in differentiating sympatric wild, domestic and hybrid cats may affect our understanding of the species, and compromise conservation actions and the species legal protection. At the same time, the early detection of hybrids is especially relevant because once hybridisation takes place and introgressed cats appear in the population they act as a bridge between wild and domestic cats, keeping, or even increasing, the degree of introgression (Germain *et al*. 2008; Hertwig *et al*. 2009). The results of this study confirmed that wild and domestic cats are genetically distinct in Southern and Central Europe, nonetheless 58 out of 686 putative

wildcat samples (8,5%) have shown evidences of admixed ancestry, together with 3 phenotypically-identified domestic cats and 13 morphological hybrids. When evaluating admixture patterns within subpopulations and comparing them with the estimates reported by previous studies, it is evident that higher sample and loci numbers increased the number of detected hybrids namely within Iberia (n=5, Oliveira *et al.* 2008; n=12, present study) and Italy (n=2, Pierpaoli *et al*. 2003; n=30, present study). Although we might be very confident in the global classification of European populations as hybridizing or non-hybridizing ones, our findings suggest that the continuous improvement of sampling schemes and molecular procedures may continue to result in the discrimination of higher number of admixed genotypes and, consequently, in a better definition of true introgressive hybridization patterns. Moreover, simulation and real genotypes analyses have shown that classifying the different admixed genotypes as F1, F2 or backcrossed hybrids might still be speculative and further development in molecular tools and statistical approaches is crucial to circumvent this drawback. Finally, we may predict that the observed asymmetry in backcross assignment probabilities for BxFCA and BxFSI should be a result of the higher genetic diversity found among domestic cats (Godinho *et al.* 2011). Because our data does not allow the precise discrimination between domestic and backcrossed cats (Supplementary Table S3), we might still be reporting slightly lower rates of hybridization than the ones occurring in nature.

The composition of the Hungarian and Scottish samples analysed in this study clearly corroborates the broadly documented hybrid nature of these populations, both through molecular (Beaumont *et al.* 2001; Pierpaoli *et al.* 2003; Lecis *et al*. 2006) and morphological (Kitchener *et al*. 2005) studies. Here, we have demonstrated that a better coverage of the species genome (by genotyping 38 unlinked microsatellites) undoubtedly confirms that these populations comprise an assemblage of hybrid genotypes with variable admixed ancestry, and that introgressive hybridization might be widespread among wild-living cats. According to Kitchener *et al.* 2005, 88% of wild-living Scottish cats may be hybrids or feral domestic cats.

Today, paired with continuous attempts to limit in space and time the occurrence of hybridization, it is necessary to shift the focus from a static idea of species to a more comprehensive concept of the wildcat, as a species that clearly still exist but in a different form to their ancestors, inhabit various and changing habitats, evolve within their contemporary environment, and should be protected by effective conservation management actions. Artificial hybridization should be a priority in the development and implementation of conservation strategies in Europe and long-term genetic monitoring of admixture should be used to identify critical areas of introgression and implement the most appropriate conservation strategies. Moreover, understanding the links between admixture distribution and (historical and) ecological factors would largely improve the efficacy of any conservation measure taken to preserve wildcat populations in Europe.

Future Perspectives

Wildcats are fully protected across most of its range in Europe, are listed on the EU Habitats and Species Directive (Annex IV), protected by the Bern Convention (Appendix II), included on CITES Appendix II and classified as threatened at the national level in many European states. However, there have been no recent large-scale surveys of the species that provide a global picture of its European status (Driscoll & Nowell,

2010). Because population fragmentation and introgressive hybridization have been pointed out as major threats to the survival of wildcat's natural populations, we have focused our research in European-wide analyses of both menaces and provided new insights into the knowledge of species. Nevertheless, many questions persist and arise especially for what regards the mechanisms and dynamics of introgressive hybridization. For example, a detailed analysis of the geographical location of hybrids relatively to a precise map of distribution for each wildcat subpopulation would help understanding if crossbreeding between wild and domestic cats has been restricted to peripheral areas of wildcat range as previously advocated by Oliveira *et al*. (2008) in Iberia and Randi *et al*. (2001) in Italy, or if admixture events might also take place in the core of the species distribution. Furthermore, identifying the direction of hybridization through the additional genotyping of uniparentally inherited loci would be essential for the proper understanding of hybridization dynamics in natural populations. For example, while observations of wildcats' spatial activities suggested that most of admixture events are probably occurring between male wildcats and domestic females (Birò *et al*. 2004), observations in Ardennes Mountains in France indicate that hybridization might not involve domestic females, but probably the low densities of wild males promote the crossbreeding between wild females and domestic males (Germain *et al*. 2008).

Although we were able to discuss European wide evidences and rates of hybridization, it is premature to consider that the numbers reported in this study strictly predict the true proportion of hybrids across populations. Among the dataset used in this study, unbalanced efforts and sampling strategies have been applied for each subpopulation, with geographical areas being represented by samples collected during long-lasting ecological studies of the species (e.g. Hungary or Italy) and others for which cat samples were obtained only opportunistically (e.g. North and Centre of Iberia). The unbalanced number of domestic and wildcat sampled in FSI-1 or FSI-2 and FSI-4, that is well noticeable in Figure 3, is most representative of this heterogeneity. Moreover, in areas other than Hungary and Scotland, it is possible that samples displaying domestic phenotypes would not be intentionally collected, but they could still be backcrosses with domestic cats and not pure domestic sympatric individuals (a problem that as also been underlined for the study of wolf x dog hybridization by Godinho *et al*. 2011). The fact that none of the admixed genotypes found in this study has been identified as a backcross with a domestic cat might not only be a consequence of possible errors in our hybridization inferences but also a consequence of this partial behavior in samples collection. We, therefore, recommend future range-wide works to be based in an European agreement for homogenizing sampling strategies, with the collection of all putatively wild, domestic and hybrid phenotypic samples.

Finally, wildcat hybridization studies should soon take advantage of the increasing knowledge of the domestic cat genome and of the resulting availability of new informative markers other than microsatellites. For example, the analysis of multiple SNPs (Single Nucleotide Polymorphisms) and the selection of the most informative ones for the differentiation between wild and domestic cats might soon help to improve our understanding of hybridization at a genome-wide level. In this context, the analysis of genetic variants that are responsible for the vast morphological, physiological and behavioural diversity occurring among domesticates might be particularly interesting. For example, since basal morphology in European wildcats is unchanging, specific mutations determining variable patterns in domestic cats are very rare or absent in natural populations, and alternative variants or random polymorphism are expected. At the same time,

explicit genetic variants might benefit the way of living in nature and may have remained fixed in the wild populations (e.g. camouflage patterns crucial for hiding and hunting behaviour), while variable genetic expressions might be maintained in domestic individuals due to relaxation of selective pressures. Analysing levels of genetic diversity in mutations known to have been under different types of selection during domestication and/or breeds' improvement may, thus, revolutionize wildcat hybridization studies.

ACKNOWLEDGEMENTS

We thank M. Herrman, F. Suchentrunk, M. Liberek, Natural Museum of Scotland, P. Romy, B. Szolt, A. Sforzi, B. Ragni, L. Lapini, A. de Faveri, K. Hupe, I. Eckert, H. Potocnik, M. Moes, G. Cagnolati, F. Vercillo, J.Godoy, M. Malsaña, N. Mejias, J.M. Fernández, J.L. Robles, G.D. Penafiel, E.B. Duperón, M. Moleón, P. Monterroso, F. Álvares, J. Rodrigues, BTVS/ICNB: Portuguese National Tissue Bank/National Institute for Nature and Biodiversity Conservation, P Lyberakis and their collaborators for providing putative wildcat samples. We are also grateful to all anonymous veterinarians and biologists that assisted in samples collection. Rita Oliveira was supported by Fundação para a Ciência e a Tecnologia (FCT) through the PhD grant SFRH/BD/24361/2005 and the research project PTDC/CVT/71683/2006.

REFERENCES

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

Beaumont M, Barratt EM, Gottelli D, Kitchener AC, Daniels MJ, Pritchard JK, Bruford MW (2001) Genetic diversity and introgression in the Scottish wildcat. *Molecular Ecology* **10**: 319–336.

Biró Z, Szemethy L, Heltai M (2004) Home range sizes of wildcats (*Felis silvestris*) and feral domestic cats (*Felis silvestris catus*) in a hilly region of Hungary. *Mammalian Biology* **69**: 302–310.

Cornuet JM and Luikart G (1997) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144**: 2001-2014.

Campana, MG, HV Hunt, H Jones & J White. 2011. *CorrSieve*: software for summarizing and evaluating Structure output. *Molecular Ecology Resources* **11**: 349-352.

Clutton-Brock J (1999) A Natural History of Domesticated Mammals, Cambridge University Press.

Driscoll CA, Menotti-Raymond M, Roca AL, Hupe K, Johnson WE, Geffen E, Harley EH, Delibes M, Pontier D, Kitchener AC, Yamaguchi N, O'Brien SJ, MacDonald DW (2007) The near eastern origin of cat domestication. *Science* **317**: 519–523.

Driscoll C & Nowell K (2010) *Felis silvestris*. In: IUCN 2012. IUCN Red List of Threatened Species. Version 2012.2.

Eckert I, Suchentrunk F, Markov G, Hartl GB (2010) Genetic diversity and integrity of German wildcat (*Felis silvestris*) populations as revealed by microsatellites, allozymes, and mitochondrial DNA sequences. *Mammalian Biology - Zeitschrift fur Saugetierkunde* **75**(2): 160-174.

Excoffier L, Simouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitocondrial DNA restriction data. *Genetics* **131**: 479-491.

Excoffier L and Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**: 564-567.

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

Frankham R, Ballou JD (2003) Introduction to conservation genetics. University of Cambridge, Cambridge, UK.

French D, Corbett LK, Easterbee N (1988) Morphological discriminants of Scottish wildcats (*Felis silvestris*), domestic cats (*F. catus*) and their hybrids. *Journal of Zoology, London*, **161**, 75–123.

Garza JC, Williamson EG (2001) Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology* **10**: 305–318.

Germain E, Benhamou S, Poulle M-L (2008) Spatio-temporal sharing between the European wildcat, the domestic cat and their hybrids. *Journal of Zoology* **276**: 195–203.

Gittleman JL, Funk SM, Macdonald DW, Wayne RK (2001) Carnivore conservation (Conservation Biology, Vol. 5). Lavoisier, Cambridge.

Godinho R, Llaneza L, Blanco L, Lopes S, Álvares F, García E, Palacios V, Cortés Y, Talegón J and Ferrand N (2011) Genetic evidence for multiple events of hybridization between wolves and domestic dogs in the Iberian Peninsula. *Molecular Ecology* **20**: 5154–5166.

Guo SW and Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportions for multiple alleles. *Biometrics* **48**: 361-372.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. Molecular Ecology Resources **9**(5): 1322-1332.

Jombart T (2008) Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**(11): 1403-1405.

Kitchener AC (1991) *The natural history of the wild cats*. Cornell University Press, New York. 280 p.

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

Kitchener AC and Rees EE (2009) Modelling the dynamic biogeography of the wildcat: implications for taxonomy and conservation. *Journal of Zoology* **279**: 144–155.

Hertwig ST, Schweizer M, Stepanow S, Jungnickel A, Böhle U-R, Fischer MS (2009) Regionally high rates of hybridization and introgression in German wildcat populations (*Felis silvestris*, Carnivora, Felidae). *Journal of Zoological Systematics and Evolutionary Research* **47**: 283–297.

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

Lecis R, Pierpaoli M, Birò ZS, Szemethy L, Ragni B, Vercillo F, Randi E (2006) Bayesian analyses of admixture in wild and domestic cats (*Felis silvestris*) using linked microsatellite loci. *Molecular Ecology* **15**: 119–131.

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

Lozano J, Virgós E, Cabezas–Díaz S, Mangas J G (2007) Increase of large game species in Mediterranean areas: is the European wildcat (*Felis silvestris*) facing a new threat? *Biological Conservation* **138:** 321–329.

Luikart G, Allendorf FW, Cornuet J-M and Sherwin WB (1998) Distortion of allele frequency distributions provides a test for recent population bottlenecks. *Journal of Heredity* **89** (3): 238-247.

Menotti-Raymond M, *et al*. (2003) Second-generation integrated genetic linkage/radiation hybrid maps of the domestic cat (*Felis catus*). *Journal of Heredity* **94**: 95–106.

Monterroso P, Brito JC, Ferreras P, Alves PC (2009) Spatial ecology of the European wildcat in a Mediterranean ecosystem: dealing with small radio–tracking datasets in species conservation. *Journal of Zoology* **279**: 27–35.

Nielsen EE, Bach LA, Kotlicki P (2006) Hybridlab (version 1.0): a program for generating simulated hybrids from population samples. *Molecular Ecology Notes* **6**: 971–973.

Nowell K & Jackson P (1996) European wildcat, *Felis silvestris, silvestris* group Schreber, 1775. In *Wild cats: status survey and conservation action plan*: 110–113. Nowell, K. & Jackson, P. (Eds). Gland, Switzerland: IUCN.

O'Brien J, Devillard S, Say L, Vanthomme H, Léger F, Ruette S, Pontier D (2009) Preserving genetic integrity in a hybridising world: are European Wildcats (*Felis silvestris silvestris*) in Eastern France distinct from sympatric feral domestic cats? *Biodiversity and Conservation* **18** (9): 2351-2360.

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

Oliveira R, Godinho R, Randi E and Alves PC (2008b) Hybridization vs conservation: are domestic cats threatening the genetic integrity of European wildcat (*Felis silvestris silvestris*) populations in Iberian Peninsula? *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **363**: 2953-2961.

Peakall R and Smouse PE (2006) Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**: 288–295.

Pertoldi C, Bijlsma R, Loeschcke V (2007) Conservation genetics in a globally changing environment: present problems, paradoxes and future challenges. *Biodiversity and Conservation* **16**(14): 4147-4163.

Pierpaoli M, Biro ZS, Herrmann M, Hupe K, FernandesM, Ragni B, Szemethy L, Randi E (2003) Genetic distinction of wildcat (*Felis silvestris*) populations in Europe, and hybridization with domestic cats in Hungary. *Molecular Ecology* **12**: 2585–2598.

Pilot M, Jedrzejeski W, Branicki W, Sidorovich VE, Jedrejewska B, Stachura K and Kunk SM (2006) Ecological factors influence population genetic structure of European grey wolves. *Molecular Ecology* **15**: 4533–4553.

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

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

Randi E and Ragni B (1991) Genetic variability and biochemical systematics of domestic and wildcat populations (*Felis silvestris*, Felidae). *Journal of Mammology* **72**: 79–88.

Ragni B and Possenti M (1996) Variability of coat-colour and markings system in *Felis silvestris*. Italian *Journal of Zoology* **63**: 285–292.

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

Sambrook J and Russell DW (2006) Purification of Nucleic Acids by Extraction with Phenol:Chloroform. *Cold Spring Harbor Protocols*. doi:10.1101/pdb.prot4455.

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

Schauenberg P (1977) La stature du Chat forestiere *Felis silvestris* Schreber, 1777, et la variabilité morphologique de l'espèce. *Revue Suisse de Zoologie* **84**: 323-337.

Schipper J, Hoffmann M, Duckworth JW and Conroy J (2008) The 2008 IUCN red listings of the world's small carnivores. *Small Carnivore Conservation* **39**: 29–34

Schneider S, Roessli D, Excoffier L (2000) Arlequin: a software for population genetics data analysis, *Genetics and Biometry Laboratory, Department of Anthropology*, University of Geneva, Geneva.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139:** 457-462.

Sommer RS and Benecke N (2006) Late Pleistocene and Holocene development of the felid fauna (Felidae) of Europe: a review. *Journal of Zoology* **269**: 7–19.

Stahl P and Artois M (1994) Status and conservation of the wildcat (*Felis silvestris*) in Europe and around the Mediterranean rim. Convention on the Conservation of European Wildlife and Natural habitats Standing Committee. *Council of Europe Nature and Environment Series* **69**: 1-76.

Vähä JPK and Primmer CR (2006) Detecting hybridization between individuals of closely related populations – a simulation study to assess the efficiency of model-based Bayesian methods to detect hybrid individuals. *Molecular Ecology* 15: 63–72.

Van Oosterhout C, Hutchinson WF Wills DPM and Schipley P (2004) Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**: 535–538.

Vigne JD, Briois F, Zazzo A, Willcox G, Cucchi T, Thiébault S, Carrère I, Franel Y, Touquet R, Martin C, Moreau C, Comby C, Guilaine J (2012) First wave of cultivators spread to Cyprus at least 10,600 y ago. *Proceedings of the National Academy of Sciences of the United States of America* **109**(22): 8445-8449.

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

Supplementary Table S1. Description of the 38 microsatellites used to genotype all cat (*Felis silvestris*) samples excluding admixed genotypes (see Results). Locus identifications (ID) and chromosome assignments (Chr) are from Menotti-Raymond *et al.* (2003). Primer tails were labelled to fit the design of eight multiplexes. The number of repetitions (R), the allelic range (in base pairs), the observed number of alleles (Na), the allelic richness (*Ar*) based on 52 genes and the expected values of heterozygosity (H_E) are reported at each locus for domestic cats (FCA, n=294), European wildcats (FSI, n=610) and African wildcats (FLI, n=26).

						FCA			FSI			FLI		
Multiplex	**ID**	**Label**	**R**	**Chr**	**Allele**	**Na**	***Ar***	**HE**	**Na**	***Ar***	**HE**	**Na**	***Ar***	**HE**
M19	**Fca058**	NED	2	E2	118-234	11	9.285	0.711	11	6.546	0.659	10	10	0.829
	Fca077	VIC	2	C2	132-162	10	7.587	0.747	11	8.184	0.797	7	7	0.752
	Fca088	FAM	2	B3	99-125	13	10.927	0.845	11	5.532	0.651	12	12	0.854
	Fca126	NED	2	B1	103-157	14	10.326	0.817	13	6.475	0.755	10	10	0.852
	Fca453	FAM	4	A1	174-206	9	7.102	0.658	10	5.067	0.657	6	6	0.719
M32	**Fca023**	FAM	2	B1	124-154	14	12.956	0.784	14	6.333	0.752	11	11	0.799
	Fca045	NED	2*	D4	135-165	21	17.798	0.916	22	11.947	0.854	14	14	0.865
	Fca080	FAM	2	A3	210-250	18	11.517	0.751	16	8.572	0.821	11	11	0.879
	Fca094	NED	2	F2	207-249	16	13.41	0.883	16	9.042	0.754	14	14	0.89
	Fca097	VIC	2	B1	124-154	12	9.88	0.841	15	10.543	0.861	12	12	0.817
M34	**Fca005**	NED	2	E1	124-152	13	9.342	0.762	11	8.063	0.799	7	7	0.808
	Fca035	VIC	2	D2	116-168	15	8.953	0.533	18	11.414	0.897	4	4	0.322
	Fca090	FAM	2	A1	82-118	16	13.717	0.808	14	8.097	0.792	12	12	0.853
	Fca262	VIC	2	D2	163-195	15	12.153	0.839	16	10.229	0.855	12	12	0.828
M50	**Fca008**	NED	2	A1	112-148	16	12.474	0.863	18	8.692	0.807	9	9	0.82
	Fca043	VIC	2	C2	106-146	13	9.689	0.714	13	7.297	0.656	9	9	0.833
	Fca096	NED	2	E2	96-235	21	14.354	0.517	18	11.079	0.884	8	8	0.715
	Fca293	VIC	2	C1	175-255	10	8.116	0.747	11	6.955	0.76	9	9	0.842
	Fca649	FAM	2	C1	114-152	19	12.989	0.841	18	7.546	0.69	12	12	0.882
M51	**Fca026**	FAM	2	D3	126-162	16	11.758	0.836	17	9.833	0.842	10	10	0.799
	Fca132	NED	2	D3	127-161	24	17.436	0.878	15	9.69	0.845	11	11	0.858
	Fca391	NED	4	B3	219-267	13	9.87	0.708	12	9.913	0.887	6	6	0.741
	Fca628	VIC	2	E3	77-155	32	23.204	0.91	22	10.281	0.834	16	16	0.911
M52	**Fca105**	FAM	2	A2	169-209	18	12.863	0.841	13	9.256	0.863	15	15	0.897
	Fca123	VIC	2*	A1	125-155	13	9.663	0.837	11	7.182	0.802	8	8	0.771
	Fca211	NED	2	B1	98-120	9	7.537	0.644	12	7.565	0.744	8	8	0.845
	Fca305	NED	2	B2	174-224	12	9.038	0.681	14	2.587	0.064	8	8	0.772
	Fca698	FAM	2	D1	210-272	17	15.311	0.893	16	8.576	0.775	14	14	0.884
PP1	**Fca069**	FAM	2	B4	86-124	14	11.375	0.833	16	8.569	0.78	8	8	0.738
	Fca075	NED	2	E2	104-142	16	10.47	0.83	13	9.852	0.876	10	10	0.873
	Fca220	FAM	2	F2	202-224	11	7.987	0.594	12	8.472	0.813	7	7	0.821
	Fca229	NED	2	A1	142-176	14	11.933	0.774	16	9.37	0.723	8	8	0.776
	Fca441	FAM	4	D3	127-173	16	10.711	0.744	15	8.747	0.754	8	8	0.78
PP2	**Fca149**	FAM	2	B1	116-134	9	7.234	0.768	9	5.597	0.697	6	6	0.775
	Fca223	PET	2	F1	196-240	22	17.196	0.862	13	6.001	0.61	11	11	0.909
	Fca310	VIC	2	C2	106-140	13	10.4	0.777	11	2.612	0.068	13	13	0.909
	Fca678	FAM	2	A1	196-236	8	6.855	0.81	9	6.987	0.739	8	8	0.784
	FCA240	FAM	2	X	142-186	11	6.719	0.780	11	3.786	0.195	6	6	0.685

* indicates imperfect dinucleotides showing intermediate alleles

Supplementary Table S2. Individual assignment (STRUCTURE) and inferred ancestry (NEWHYBRIDS) of all genotypes considered to be possibly admixed among the total sample of European wild and domestic cats. Results are shown for cats that were partially assigned to the domestic (FCA) and wild (FSI) clusters at least in three of the admixture analyses performed. Computations were performed within macro areas 1, 2 and 4. Values in bold indicate the most likely hybrid class (*qi*>0.70).

				STRUCTURE (*qi* values)				NEWHYBRIDS (*qi* values)					
Cluster	**Location**	**Prior ID**	**Code**	**FCA**	**FSI**	**90% Confidence intervals**		**FCA**	**FSI**	**F1**	**F2**	**BxFCA**	**BxFSI**
	NE Alps	Domestic	H1.1	0.182	0.818	(0.669,0.957)	(0.043,0.331)	0,657	0,000	0,000	0,001	0,342	0,000
	NE Alps	Hybrid	H1.2	0.810	0.190	(0.641,0.966)	(0.034,0.359)	0,564	0,000	0,000	0,184	0,252	0,000
	NE Alps	Hybrid	H1.3	0.496	0.504	(0.338,0.656)	(0.344,0.662)	0,000	0,000	**0,849**	0,119	0,001	0,030
	NE Alps	Hybrid	H1.4	0.562	0.438	(0.402,0.719)	(0.281,0.598)	0,000	0,000	**0,963**	0,032	0,004	0,001
	NE Alps	Hybrid	H1.5	0.559	0.441	(0.397,0.719)	(0.281,0.603)	0,000	0,000	**0,937**	0,056	0,004	0,003
1.1.1	NE Alps	Hybrid	H1.6	0.518	0.482	(0.362,0.674)	(0.326,0.638)	0,000	0,000	**0,853**	0,125	0,001	0,020
	NE Alps	Hybrid	H1.7	0.450	0.550	(0.295,0.609)	(0.391,0.705)	0,000	0,000	**0,844**	0,093	0,000	0,063
	NE Alps	Wild	H1.8	0.433	0.567	(0.284,0.586)	(0.414,0.716)	0,000	0,000	0,158	0,489	0,000	0,352
	SLO	Domestic	H1.9	0.819	0.181	(0.677,0.952)	(0.048,0.323)	0,566	0,000	0,000	0,105	0,329	0,000
	SLO	Wild	H1.10	0.395	0.605	(0.236,0.566)	(0.434,0.764)	0,000	0,000	0,001	**0,926**	0,000	0,073
	B&H	Wild	H1.11	0.430	0.570	(0.264,0.602)	(0.398,0.736)	0,000	0,000	0,026	0,587	0,000	0,387
	BUL	Wild	H1.12	0.573	0.427	(0.404,0.743)	(0.257,0.596)	0,000	0,000	0,006	**0,983**	0,007	0,004
	BUL	Wild	H1.13	0.414	0.586	(0.260,0.575)	(0.425,0.740)	0,000	0,000	0,000	**0,962**	0,000	0,037
1.1.2	POL	Wild	H1.14	0.252	0.748	(0.105,0.414)	(0.586,0.895)	0,000	0,289	0,001	0,024	0,000	0,686
	POL	Wild	H1.15	0.243	0.757	(0.106,0.399)	(0.601,0.894)	0,000	**0,821**	0,000	0,022	0,000	0,158
	ROM	Wild	H1.16	0.296	0.704	(0.153,0.453)	(0.547,0.847)	0,242	0,000	0,000	0,690	0,068	0,000
	IT-Tus	Wild	H1.17	0.233	0.767	(0.099,0.383)	(0.617,0.901)	0,000	0,155	0,001	0,009	0,000	**0,835**
	IT-Tus	Wild	H1.18	0.289	0.711	(0.141,0.453)	(0.547,0.859)	0,000	0,092	0,001	0,156	0,000	**0,752**
	IT-Tus	Wild	H1.19	0.389	0.611	(0.233,0.553)	(0.447,0.767)	0,000	0,010	0,001	**0,722**	0,000	0,267
	IT-Tus	Wild	H1.20	0.311	0.689	(0.160,0.475)	(0.525,0.840)	0,000	0,153	0,000	0,647	0,000	0,199
	IT-Tus	Wild	H1.21	0.684	0.316	(0.520,0.841)	(0.159,0.480)	0,004	0,000	0,331	0,417	0,248	0,000
1.2.1	IT-Tus	Wild	H1.22	0.569	0.431	(0.410,0.726)	(0.274,0.590)	0,000	0,000	**0,730**	0,249	0,018	0,003
	IT-Tus	Wild	H1.23	0.169	0.831	(0.059,0.302)	(0.698,0.941)	0,000	**0,907**	0,000	0,001	0,000	0,093
	IT-Tus	Wild	H1.24	0.183	0.817	(0.028,0.349)	(0.651,0.972)	0,000	**0,806**	0,000	0,003	0,000	0,192
	IT-Tus	Wild	H1.25	0.447	0.553	(0.283,0.616)	(0.384,0.717)	0,000	0,000	0,000	**0,966**	0,000	0,034
	IT-Tus	Wild	H1.26	0.463	0.537	(0.297,0.633)	(0.367,0.703)	0,000	0,000	0,000	**0,998**	0,000	0,002
	IT-Tus	Wild	H1.27	0.525	0.475	(0.337,0.715)	(0.285,0.663)	0,000	0,000	0,000	**0,988**	0,001	0,011
	IT-Tus	Wild	H1.28	0.239	0.761	(0.063,0.420)	(0.580,0.937)	0,000	**0,757**	0,000	0,016	0,000	0,227
	IT-CS	Hybrid	H1.29	0.287	0.713	(0.149,0.435)	(0.565,0.851)	0,000	0,670	0,000	0,183	0,000	0,147
	IT-CS	Wild	H1.30	0.278	0.722	(0.111,0.458)	(0.542,0.889)	0,000	0,646	0,000	0,151	0,000	0,203
	IT-CS	Wild	H1.31	0.238	0.762	(0.092,0.399)	(0.601,0.908)	0,000	**0,705**	0,000	0,038	0,000	0,257
	IT-CS	Wild	H1.32	0.264	0.736	(0.118,0.426)	(0.574,0.882)	0,000	0,523	0,000	0,088	0,000	0,389
1.2.2	IT-CS	Wild	H1.33	0.395	0.605	(0.238,0.562)	(0.438,0.762)	0,000	0,002	0,002	**0,770**	0,000	0,226
	IT-CS	Wild	H1.34	0.193	0.807	(0.048,0.356)	(0.644,0.952)	0,000	0,599	0,000	0,005	0,000	0,396
	IT-CS	Wild	H1.35	0.253	0.747	(0.115,0.408)	(0.592,0.885)	0,000	0,483	0,000	0,017	0,000	0,500
	IT-CS	Wild	H1.36	0.216	0.784	(0.064,0.374)	(0.626,0.936)	0,000	**0,870**	0,000	0,005	0,000	0,125
	IT-CS	Wild	H1.37	0.375	0.625	(0.200,0.565)	(0.435,0.800)	0,000	0,029	0,000	**0,847**	0,000	0,125
	IT-CS	Wild	H1.38	0.326	0.674	(0.184,0.479)	(0.521,0.816)	0,000	0,002	0,000	0,184	0,000	**0,814**

	IT-CS	Wild	H1.39	0.296	0.704	(0.156,0.451)	(0.549,0.844)	0,000	0,023	0,000	0,349	0,000	0,628
	IT-CS	Wild	H1.40	0.211	0.789	(0.066,0.371)	(0.629,0.934)	0,000	0,647	0,000	0,009	0,000	0,344
	IT-CS	Wild	H1.41	0.296	0.704	(0.152,0.452)	(0.548,0.848)	0,000	0,022	0,003	0,149	0,000	**0,827**
	IT-CS	Wild	H1.42	0.290	0.710	(0.147,0.447)	(0.553,0.853)	0,000	0,096	0,001	0,051	0,000	0,853
1.2.3	IT-Sic	Wild	H1.43	0.330	0.670	(0.180,0.494)	(0.506,0.820)	0,000	0,042	0,000	0,541	0,000	0,417
	IT-Sic	Wild	H1.44	0.608	0.392	(0.449,0.763)	(0.237,0.551)	0,000	0,000	0,026	**0,930**	0,043	0,001
2.2	GER-SW	Hybrid?	H2.1	0.691	0.309	(0.517,0.864)	(0.136,0.483)	0,000	0,002	0,000	**0,958**	0,000	0,040
	GER-SW	Hybrid?	H2.2	0.478	0.522	(0.249,0.715)	(0.285,0.751)	0,024	0,000	0,000	**0,969**	0,007	0,001
	GER-SW	Wild	H2.3	0.568	0.432	(0.407,0.726)	(0.274,0.593)	0,000	0,000	0,000	**1,000**	0,000	0,000
	GER-SW	Wild	H2.4	0.328	0.672	(0.188,0.482)	(0.518,0.812)	0,003	0,000	0,000	**0,849**	0,148	0,000
	GER-SW	Wild	H2.5	0.427	0.573	(0.251,0.612)	(0.388,0.749)	0,000	0,000	0,092	**0,750**	0,157	0,000
	GER-SW	Wild	H2.6	0.309	0.691	(0.167,0.466)	(0.534,0.833)	0,034	0,000	0,000	0,497	0,469	0,000
	GER-SW	Wild	H2.7	0.498	0.502	(0.334,0.670)	(0.330,0.666)	0,000	0,000	0,531	0,434	0,033	0,001
	GER-SW	Wild	H2.8	0.509	0.491	(0.329,0.699)	(0.301,0.671)	0,000	0,000	0,434	0,527	0,035	0,004
	GER-SW	Wild	H2.9	0.450	0.550	(0.294,0.615)	(0.385,0.706)	0,000	0,000	0,000	**0,986**	0,013	0,000
	GER-SW	Wild	H2.10	0.506	0.494	(0.325,0.696)	(0.304,0.675)	0,000	0,000	0,002	**0,990**	0,007	0,001
	GER-SW	Wild	H2.11	0.454	0.546	(0.298,0.615)	(0.385,0.702)	0,000	0,000	0,000	**1,000**	0,000	0,000
	GER-SW	Wild	H2.12	0.757	0.243	(0.571,0.985)	(0.015,0.429)	0,000	0,101	0,000	0,655	0,000	0,243
	GER-SW	Wild	H2.13	0.449	0.551	(0.284,0.620)	(0.380,0.716)	0,000	0,000	0,562	0,329	0,109	0,000
	GER-SW	Wild	H2.14	0.652	0.348	(0.461,0.845)	(0.155,0.539)	0,000	0,001	0,000	**0,994**	0,000	0,005
	GER-SW	Wild	H2.15	0.215	0.785	(0.063,0.381)	(0.619,0.937)	0,631	0,000	0,000	0,067	0,302	0,000
	GER-SW	Wild	H2.16	0.579	0.421	(0.378,0.778)	(0.222,0.622)	0,000	0,000	0,000	**0,998**	0,000	0,001
	GER-SW	Wild	H2.17	0.350	0.650	(0.142,0.577)	(0.423,0.858)	0,176	0,000	0,001	**0,707**	0,116	0,000
	LX	Wild	H2.18	0.201	0.799	(0.066,0.359)	(0.641,0.934)	**0,907**	0,000	0,000	0,048	0,045	0,000
4.1	IP-N	Domestic	H4.1	0.827	0.173	(0.619,1.000)	(0.000,0.381)	**0,774**	0,000	0,000	0,107	0,000	0,120
	IP-N	Hybrid?	H4.2	0.780	0.220	(0.641,0.909)	(0.091,0.359)	0,474	0,000	0,000	0,138	0,000	0,388
	IP-N	Wild	H4.3	0.342	0.658	(0.165,0.536)	(0.464,0.835)	0,000	0,005	0,000	**0,988**	0,007	0,000
	IP-N	Wild	H4.4	0.191	0.809	(0.000,0.381)	(0.619,1.000)	0,000	**0,718**	0,000	0,100	0,182	0,000
	IP-N	Wild	H4.5	0.823	0.177	(0.578,1.000)	(0.000,0.422)	0,603	0,000	0,000	0,292	0,000	0,106
	IP-N	Wild	H4.6	0.325	0.675	(0.163,0.502)	(0.498,0.837)	0,000	0,000	0,000	**0,924**	0,075	0,000
4.2	IP-SW	Wild	H4.7	0.743	0.257	(0.564,0.934)	(0.066,0.436)	0,000	0,285	0,000	0,392	0,000	0,323
	IP-SW	Wild	H4.8	0.249	0.751	(0.111,0.401)	(0.599,0.889)	0,293	0,000	0,000	0,160	0,547	0,000
	IP-SW	Wild	H4.9	0.592	0.408	(0.437,0.741)	(0.259,0.563)	0,000	0,000	0,000	**0,998**	0,000	0,002
	IP-SW	Wild	H4.10	0.255	0.745	(0.063,0.445)	(0.555,0.937)	0,384	0,000	0,004	0,083	0,529	0,000
	IP-SW	Wild	H4.11	0.157	0.843	(0.040,0.292)	(0.708,0.960)	0,633	0,000	0,000	0,001	0,366	0,000
4.3	IP-SE	Wild	H4.12	0.628	0.372	(0.454,0.793)	(0.207,0.546)	0,000	0,000	0,000	**0,980**	0,000	0,020
Captivity		Hybrid	57	0.434	0.566	(0.278,0.594)	(0.406,0.722)	0,000	0,000	0,019	**0,874**	0,000	0,107
		Hybrid	59	0.594	0.406	(0.441,0.745)	(0.255,0.559)	0,000	0,000	0,176	**0,765**	0,057	0,001
		Hybrid	60	0.299	0.707	(0.156,0.452)	(0.548,0.865)	0,000	0,000	0,000	0,510	0,0000	0,490
		Hybrid	61	0.805	0.195	(0.640,0.984)	(0.016,0.360)	0,001	0,000	0,000	0,481	0,518	0,000
		Hybrid	62	0.308	0.692	(0.159,0.462)	(0.538,0.841)	0,000	0,000	0,000	0,413	0,000	0,587
		Hybrid	63	0.461	0.539	(0.302,0.625)	(0.375,0.698)	0,000	0,000	0,000	**0,844**	0,000	0,156
		Hybrid	64	0.305	0.695	(0.147,0.479)	(0.521,0.853)	0,000	0,041	0,008	0,162	0,000	**0,789**

Supplementary Table S3. Assignment of simulated parental genotypes (P_{FCA}, P_{FSI}), F_1, F_2 and first generation of backcrosses with domestic cat (BxFCA) and wildcats (BxFSI) generated using HYBRIDLAB. Computations were performed within macro areas 1, 2 and 4. A) Average proportion of membership *Qi* obtained in STRUCTURE (average 90% CI are shown in brackets). N summarizes the number of individuals correctly assigned to their cluster of origin (upper row) and the number of hybrids misclassified as parentals, either domestic (P_{FCA}) either wild (P_{FSI}; lower row). b) Number of genotypes assigned at q_i = 0.70, 0.80 and 0.90 to their expected clusters in NEWHYBRIDS. The number of genotypes allocated to another class (other) or for which *qi* values were distributed among classes with *qi*<0.70 (no) are shown.

a)	Macroarea 1			Macroarea 2			Macroarea 4			Average	
	Q_{FCA}	Q_{FSI}	N	Q_{FCA}	Q_{FSI}	N	Q_{FCA}	Q_{FSI}	N	Q_{FCA}	Q_{FSI}
P_{FCA} *qi*>0.85	0.954 (0.872-0.998)	0.046 (0.002-0.128)	100	0.953 (0.869-0.998)	0.047 (0.002-0.131)	100	0.962 (0.887-0.999)	0.038 (0.001-0.113)	100	0.956 (0.876-0.998)	0.044 (0.002-0.124)
P_{FSI} *qi*>0.85	0.035 (0.001-0.104)	0.965 (0.896-0.999)	100	0.035 (0.001-0.105)	0.965 (0.895-0.999)	100	0.027 (0.000-0.089)	0.973 (0.911-1.000)	100	0.032 (0.001-0.099)	0.968 (0.901-0.999)
F_1 *0.40<qi<0.60*	0.516 (0.372-0.659)	0.484 (0.341-0.628)	92 0 P	0.503 (0.355-0.654)	0.497 (0.346-0.645)	92 0 P	0.506 (0.362-0.651)	0.494 (0.349-0.638)	97 0 P	0.508 (0,363-0,655)	0.492 (0,345-0,637)
F_2 *0.40<qi<0.60*	0.516 (0.372-0.659)	0.484 (0.341-0.628)	72 0 P	0.507 (0.359-0.657)	0.493 (0.343-0.641)	73 0 P	0.494 (0.352-0.638)	0.506 (0.362-0.648)	73 0 P	0.506 (0.361-0.651)	0.494 (0.349-0.639)
BxFCA *0.85>qi>0.60*	0.745 (0.601-0.866)	0.255 (0.134-0.399)	84 12 P_{FCA}	0.759 (0.617-0.887)	0.241 (0.113-0.383)	80 14 P_{FCA}	0.768 (0.630-0.890)	0.232 (0.110-0.370)	80 13 P_{FCA}	0.757 (0.634-0.881)	0.243 (0.119-0.366)
BxFSI *0.85>qi>0.60*	0.260 (0.135-0.389)	0.740 (0.611-0.865)	88 7 P_{FSI}	0.246 (0.126-0.382)	0.754 (0.618-0.874)	87 12 P_{FSI}	0.244 (0.127-0.377)	0.756 (0.623-0.873)	87 7 P_{FSI}	0.250 (0,129-0,382)	0.750 (0,617-0,871)

b)	Macro area 1					Macro area 2					Macro area 4				
	0.70	**0.80**	**0.90**	**other**	**no**	**0.70**	**0.80**	**0.90**	**other**	**no**	**0.70**	**0.80**	**0.90**	**other**	**no**
P_{FCA}	100	97	91	0	0	100	89	86	0	0	100	98	88	0	0
P_{FSI}	100	100	99	0	0	100	100	97	0	0	100	100	100	0	0
F_1	90	81	72	0	8	92	84	73	0	8	94	91	86	1 F2	5
F_2	80	70	60	3 B_1	17	83	73	58	1 B1	16	81	77	70	1 F1 2 B1 1 B2	15
BxFCA (or B_1)	81	73	70	3 FCA 6 F2	10	82	78	74	2 FCA 1 F1 3 F2	12	86	85	78	4FCA 1 F1 1 FCA	8
BxFSI (or B_2)	91	83	80	1 F1 1 F2	7	92	89	86	1 F1 1 F2	6	91	89	85	2 F1 1 F2	6

Nuclear Genome SNPs to detect European wildcat *(Felis silvestris silvestris)* and domestic cats *(Felis s. catus)* hybridization.

Rita Oliveira[1,2], Ettore Randi[3], Federica Mattucci[3], Jennifer D. Kurushima[4], Leslie A. Lyons[4], Paulo C. Alves[1,2,5]

[1] CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBio - Laboratório Associado, Campus Agrário de Vairão, 4485-661 Vairão, Portugal
[2] Departamento de Biologia, Faculdade de Ciências da Universidade do Porto, Porto, Portugal
[3]Laboratorio di Genetica, Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA), Ozzano dell'Emilia, Bologna, Italy
[4]Department of Population Health and Reproduction, School of Veterinary Medicine, University of California - Davis, Davis, CA USA
[5]University of Montana, Wildlife Biology Program, College of Forestry and Conservation, Missoula, Montana, USA

ABSTRACT

Introgression of domestic cat (*Felis s. catus*) genes into European wildcat (*Felis s. silvestris*) populations is a challenging complication for endangered wildcat conservation. STRs and mtDNA variation has estimated the rate of introgression and of identification of hybrids. Newer genomic technologies support the use of nuclear markers to estimate introgression. In this study, genetic variation at 158 unlinked, autosomal single nucleotide polymorphisms (SNP) were analyzed in 277 European cats samples, including, 139 domestic cats, 130 putative European wildcats and 5 captive-bred hybrids. On average, SNP heterozygosity values were higher among domestic cats versus wildcats H_E=0.340 and H_E= 0.222, respectively. Although no diagnostic SNP markers were fixed or private in any population, pairwise F_{CT} distance values revealed marked differences between wildcat and domestic cat groups, with 35 loci providing an average F_{CT} value above 0.74. The power of the analysed loci to accurately identify admixture events and discriminate the different hybrid categories was evaluated. Results from Bayesian model-based computations of simulated and real genotypes show that the 158 SNPs provide successful estimates of admixture, with 100% hybrid individuals (up to 2-3 generations in the past) being correctly identified in STRUCTURE analyses and over 92% using the NEWHYBIDS' algorithm. None of the unclassified cats was wrongly allocated to another hybrid class. Thirty-five SNPs provided the most parsimonious panel for robust inferences of parental and first generations of admixed ancestries. This approach may now be used to further reconstruct both the historical and recent evolution of wildcat populations and, hopefully, to develop sound conservation guidelines for its legal protection in Europe.

Keywords: feline, hybridization, introgression, single nucleotide polymorphisms.

INTRODUCTION

The survival and conservation of indigenous European wildcat (*Felis silvestris silvestris*) populations might be locally threatened by introgressive hybridization with feral domestic cats (*Felis silvestris catus*). Over the last decade, the genotyping of several highly polymorphic molecular markers, specifically microsatellites (short tandem repeats - STR), and partial mitochondrial DNA sequences, combined with new Bayesian statistical tools have radically improved knowledge of the genetics of European wildcats. The studies have provided new insights into populations' structure and admixture (e.g. Beaumont *et al.* 2001; Randi *et al.* 2001; Pierpaoli *et al.* 2003; Kitchener *et al.* 2005; Lecis *et al.* 2006; Oliveira *et al.* 2008; Eckert *et al.* 2010; O'Brien *et al.* 2009; Hertwig *et al.* 2009). Wildcats have been domesticated recently, the first evidence of taming dated from ca 10,600 years ago in Cyprus (Vigne *et al.* 2012) and wild and domesticated forms have always remained fully interfertile (Robinson 1977, Ragni 1993). Admixture between both sub-species is thought to have been occurring since feral domestic cats started their expansion across the entire range of the wildcat species (Driscoll *et al.* 2009). In narrow contact areas, where *taxa* boundaries are probably maintained, introgression may be minimal and the cat populations remain distinct (e.g., Pierpaoli *et al.* 2003; Kitchener *et al.* 2005; Lecis *et al.* 2006). However, regions with widespread admixture might produce hybrid swarms, likely leading to the genetic extinction of the parental populations (Allendorf *et al.* 2001; Beaumont *et al.* 2001; Brumfield 2010; Fitzpatrick *et al.* 2010). European wildcats have apparently experienced both extremes. Wildcats in Scotland and Hungary show widespread hybridization and deep genetic introgression with domestic cats (Beaumont *et al.* 2001; Lecis *et al.* 2006), while only sporadic hybridization or no detectable introgression have been observed in Italy, Iberia and northeast France (Pierpaoli *et al.* 2003; Lecis *et al.* 2006; Oliveira *et al.* 2008; O'Brien *et al.* 2009).

Although wildcat and domestic cat hybridization has been addressed in several studies, detecting hybrids and introgressed individuals, and understanding the causes limiting or favouring introgression, are still complex and controversial issues affecting wildcat research and conservation. The recent domestication may implicate an overall shallow differentiation between the wild and domestic cat sub-species, and thus be most readily detectable by changes in allelic frequencies of non-functional markers. The maternal inheritance of mtDNA can track the female contributors of recent hybrids, if diagnostic variants have been identified (McEwing *et al.* 2011), however, after several generations of admixture and backcrosses, gene flow is convoluted and the extent of introgression is difficult to evaluate within individuals or populations with the current limited battery of genetic markers. Combinations of markers, such as STRs and mtDNA (Driscoll *et al.* 2011) have improved hybrid detection, however, resolution remains limited. The development of a larger suite of molecular tools, applicable in invasive and noninvasive samples, is essential to increase the power of admixture analyses, and thus imperative to promote the adequate conservation planning of European wildcat populations.

High-throughput technologies that use limited quantities of DNA, as well as improved genomic resources, such as SNP arrays and sequence assemblies, have enabled the genome-wide genotyping of many individuals in closely related species, such as wild species for which genomic information is available for their

domesticated relatives (e.g. red fox, Sacks and Louie 2008; wolf, vonHoldt *et al.* 2010; bison, Pertoldi *et al.* 2010; bighorn sheep, Poissant *et al.* 2010). Among these "genome-enabled" *taxa* (*sensu* Kohn *et al.* 2006), European wildcats studies will benefit from the cross-species applicability of domestic cat data. Specifically, the recent light coverage of the domestic cat genome (Pontius *et al.* 2007; Mullikin *et al.* 2010) that has included SNP discovery in the African wildcat sub-species (*Felis silvestris libyca*), provides useful reference data for the discovery of new nuclear markers for assessing the introgression of domestic cat genes in the wild counterparts.

During the different phases of the domestication process, different portions of the genome are sculpted by artificial selection, from the early steps of initial taming until the devoted formation and improvement of breeds (Wiener and Wilkinson, 2011). Consequently, specific genomic variants should contrast between wild ancestors and their domesticated relatives. In the silver fox (*Vulpes vulpes*), white spotting and tail and ear carriage differences are associated with tamed individuals (Trut, 1999). Body morphology and coat color patterns in European wildcats show little variation, however, their domestic counterparts have distinct coat color, fur type, behaviour, breeding cycle, and morphologic variants (see Lyons 2010 and Lyons 2012 for a summary of most relevant phenotype/genotype associations among domestic cat breeds). Specific mutations determining variable patterns in domestic cats are predictably absent in natural wildcat populations, and are therefore strong diagnostic candidate genetic variants for the distinction of wild and domestic relatives.

This study examines the power of anonymous and domestic cat phenotypic-conferring SNPs to estimate introgression in conspecific wildcats. The quantity of SNPs needed for an accurate inference of individuals assignment to the wild or domestic population and, furthermore, for the determination of hybrids' admixture ancestry was also examined. Accurate introgression estimates of wildcat populations will promote and prioritize conservation efforts for the species, protecting important habitats and natural resources for small wild felids.

MATERIALS AND METHODS

European Wildcats and Domestic Cats

Morphologically identified wildcats (FSI, n = 130) were selected from ISPRA and CIBIO tissue bank collections taking in consideration the natural distribution of European wildcat (*Felis silvestris silvestris)* and the fragmentation of its populations in Europe (Pierpaoli *et al.* 2003). Sampling was performed across diverse geographic localities by randomly selecting a few available samples from each location (Table 1, Figure 1). Five known wildcat by domestic cats hybrids obtained in captivity were included in the analyses (Pierpaoli *et al.* 2003). Random bred cats (FCA, n = 139) living in regions sympatric to the wildcats but in urban areas were also genotyped (Table 1; Lipinski *et al.* 2008; Kurushima *et al*. 2012). DNA from tissue samples was extracted as described by Pierpaoli *et al.* (2003). Buccal swabs from domestic cats from Cyprus were obtained from the Cyprus Malcolm Cat Sanctuary and were prepared as previously described (Kurushima *et al*. 2012).

DNA from putative European wildcats and captive bred hybrids was whole genome amplified (WGA) according to manufacturer's recommendations using the REPLI-g® Midi Kit (Qiagen).

Table 1. Demographics of cats used for SNP analyses of introgression.

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

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

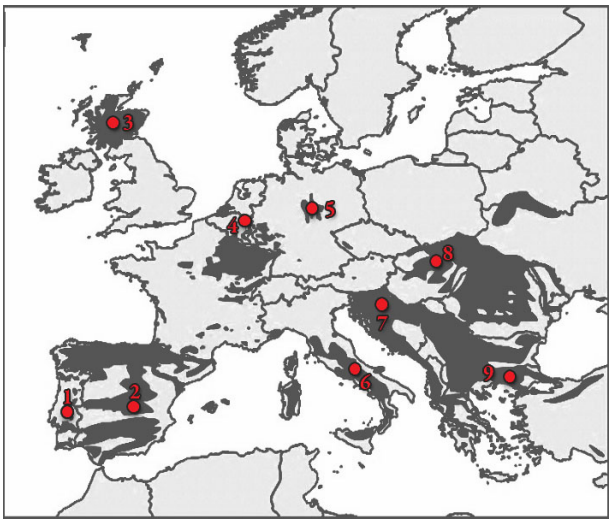


Figure 1. Sampling locations of putative European wildcats. Shaded areas correspond to the approximate current distribution of *Felis silvestris* in Europe (adapted from Grabe and Worel 2001).

SNPs genotyping

SNPs from the 19 cat autosomes (n = 154) and X chromosome (n = 4) were selected and used to genotype all cats, including: i) 138 Abyssinian SNPs randomly dispersed across the cat genome (Kurushima *et al.* 2012); ii) ten in morphologic and disease candidate genes with presumed phenotype/genotype correlation in domestic cats (Lyons 2012); iii) nine SNPs in candidate genomic regions that revealed at least one polymorphic position between European wild and domestic cats (Johnson *et al.* 2006) or for which high variability was known among domestic cat (CCR2 in Esteves *et al.* 2007); iv) four on the X chromosome; and v) one species-specific (Supplementary Table 1). Golden Gate Assay amplification and BeadXpress reads were performed following the manufacturer's protocol (illumina Inc.) on 50 - 500 ng of DNA or whole genome amplified product. BeadStudio software v. 3.1.3.0 with the Genotyping module v. 3.2.23 (illumina Inc.) was used to analyze the data (Kurushima *et al.* 2012).

Statistical Analysis

Minor allele frequency (MAF) was calculated with FSTAT v. 2.9.3.2 (Goudet 2001). Summary statistics were used to describe levels of genetic variability and differentiation on the wild and domestic subspecies. To avoid any bias resulting from the inclusion of hybrid genotypes among the representatives of European wildcats, comparative analysis with all domestic cats against the 82 putatively purest European wildcats were performed (see Results). All wild living cats from Hungary and Scotland were excluded from these first analyses due to their high level of admixture proportions determined both from morphologically presumption and genetically inference (Beaumont *et al*. 2001; Daniels *et al*. 2001; Pierpaoli *et al.* 2003; Lecis *et al.* 2006). The individuals excluded were, afterwards, included to the dataset for hybridization analyses. Significance of deviations from Hardy–Weinberg equilibrium (HWE), and the observed (H_O) and expected (H_E) heterozygosities (unbiased, Nei 1978) were calculated for all locus-population combinations using Markov chain exact tests in the ARLEQUIN 3.5.1.2 (Excoffier and Lischer 2010) package, with a chain length of 100,000 and 3,000 dememorization steps. FSTAT 2.9.3.2 was used to compute the Wilcoxon signed rank test to evaluate differences in H_E between wild and domestic cats, accounting for differences in sample size (Goudet 2001). Allelic richness (Ar) was computed for each group following a rarefaction method that compensates for uneven sample sizes, as implemented in the software HP-Rare 1.0 (Kalinowski 2005). The ARLEQUIN 3.5.1.2 (Excoffier and Lischer 2010) package was used to perform an AMOVA of pairwise F_{CT} (wildcat versus domestic groups) for each polymorphic locus, testing the null hypothesis of no differentiation by permuting genotypes between populations (10,000 replicates; $P<0.001$). Average values were calculated for autosomal SNPs alone. The 158 SNPs were ranked for hybridization diagnostic value by computing: i) *In* (informativeness for assignment), ii) *Ia* (informativeness for ancestry coefficients) and iii) *ORCA* (optimal rate of correct assignment), using INFOCALC (Rosenberg *et al.* 2003; Rosenberg 2005). For each locus, average ranking values were determined. Moreover, the probability of identity was estimated with a correction for small sample size ($P_{ID}unbiased$; Paetkau *et al.* 1998) and the equivalent probability for a pair of siblings ($P_{ID}sib$; Waits *et al.* 2001) with GenAlEx 6.41 (Peakall and Smouse 2006). These values were considered the estimated minimum number of loci required for describing unique individual genotypes.

Individual Assignment and Admixture Analyses

To assign cats to populations and to test for admixture, 158 SNP genotypes from 274 cats were evaluated using two Bayesian clustering procedures. Assuming two main populations, European wildcat and domestic cats (K = 2), ten independent runs of the Bayesian-based software STRUCTURE 2.3.3 (Pritchard *et al.* 2000; Falush *et al.* 2007; Hubisz *et al.* 2009) were computed. For each run, the average proportion of membership (*Q*) of the sampled populations and the distribution of individual membership proportions (*qi*) to the two inferred clusters, with their 90% credibility intervals (CI) were assessed. All computations were performed using the admixture model with correlated allele frequencies either without prior non-genetic information or considering the domestic cats as reference samples. Runs consisted of a burn-in of 10^5 cycles and 10^6 Markov Chain Monte Carlo (MCMC) iterations, and were averaged using CLUMPP version 1.1.1 (Jakobsson and Rosenberg 2007) with the FullSearch algorithm and the G' pairwise matrix similarity statistics. Average assignments were plotted using DISTRUCT 1.1 (Rosenberg 2004). The Bayesian model-based method implemented in the software NEWHYBRIDS (Anderson and Thompson, 2002) was further applied to classify hybrid generations of admixed cats, under the same computational parameters referred for

STRUCTURE. NEWHYBRIDS estimates the posterior probability that individuals fall into each of six genotypic classes corresponding to hybrid categories (Hi): parental subspecies (domestic or wild), F1, F2, and the backcrosses. Uniform priors were chosen to down-weight the influence of an allele that might be rare in one species and absent in the other. Ten independent runs were performed to test for stability.

The power of all SNPs to detect different hybrid classes was assessed by the analysis of the assignment accuracy obtained for simulated genotypes. One-hundred multi-locus genotypes of each parental (wildcat x wildcat; domestic cat x domestic cat), F1 (wildcat x domestic cat), F2 (F1 x F1) and backcross (F1 x wildcat; F1 x domestic cat) categories were generated with the software HYBRIDLAB v1.0 (Nielsen *et al.* 2006) and, afterwards, analysed using STRUCTURE and NEWHYBRIDS under the same setting of the admixture analysis described above. *Qi* threshold values for all analyses where established by the minimum value for which all parental domestic cats could be correctly assigned. Observed genotypes that displayed admixed genetic assignments or for which molecular assignments opposed their prior morphological identifications in the hybridization analyses of STRUCTURE and NEWHYBRIDS were also analysed together with the 600 simulated genotypes. Analyses of all observed, simulated and both kind of genotypes prompted the elimination of 18 putative wildcat samples that were most likely included in the wildcat sampling group due to incorrect morphological identifications. Accordingly, the preliminary Bayesian inferences were re-run for the new dataset of 256 cats, including, 139 random-bred domestic cats, 112 putative European wildcats and five known hybrids. In addition, Bayesian analyses of simulated genotypes where performed for the best 35 SNPs estimated to accurately allow individual genotyping of cat samples.

RESULTS

SNPs variability

The SNP genotype call rate was greater than 80% in all analysed cat samples (n = 274). Descriptive statistics are presented in Supplementary Table 1 and 2. All SNPs were polymorphic among domestic cats, implying a MAF >5%, however, 22 SNPs (13.92%) were monomorphic amongst the wildcats. Significant deviations from HWE, following Bonferroni correction ($P < 0.00016$), were detected in 16 SNP loci, eleven among the domestic population and five in the wildcat group. Although none of the 158 loci had alternative private alleles, a large proportion of SNP variability was significantly partitioned between wildcat and domestic cats (average $F_{CT} = 0.427$; $P < 0.001$), with single-locus F_{CT} pairwise values ranging between 0.000 (ChrA3_position159537633; ChrF2_78303221) and 0.891 (ChrE2_34027888; AMOVA $P < 0.001$). The mean value of H_E was 0.223 (±0.134), ranging from 0.043 (at locus RASA2) to the possible maximum of 0.500 (ChrB2_45093345; ChrB4_21098349; ChrD4_41078218) in the domestic group, and ranging between 0.000 and 0.499 (chrA3_159537633) in the wildcat group. Altogether, European wildcats proved to be significantly less variable than domestic cats, both at average values of expected heterozygosity ($H_{E\,(FCA)} = 0.340$; $H_{E(FSI)} = 0.107$; $P < 0.001$) and allelic richness ($Ar_{(FCA)} = 1.738$; $Ar_{(FSI)} = 1.250$; $P < 0.001$). Exceptions to the lower wildcat's variability were found at 18 SNPs, for which wildcats exhibited higher H_E than domestic cats (Supplementary Table 1 and 2). Ten SNPs had 2 - 4 times higher heterozygosity in wildcats, being that five of the ten showed significant deviations from HWE.

The average informativeness scores of each locus (INFOCALC - Supplementary Table 2) revealed that SNPs with lowest values of H_E in both groups displayed the highest values of genetic differentiation and top rank numbers, as they represented high frequencies of the two possible alternate variants. Loci for which wildcats showed no variation and domestic cats revealed high H_E values (e.g. ChrB1_88148379) might also be highly useful (Supplementary Table 2). For increasing SNPs combinations based on the loci ranked list, $P_{(ID)unbiased}$ and $P_{(ID)sibling}$ at $P < 0.001$ were simultaneously obtained using 35 loci (Table 2). These 35 top-ranked SNPs had an average pairwise F_{CT} =0.74 (P<0.001).

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

			Ar		H_E‡		FIS		AMOVA	INFOCALC	
Chr	**POSITION**	**Location**	**FCA**	**FSI**	**FCA**	**FSI**	**FCA**	**FSI**	**pw-F_{CT}**	**score**	**R**
B4	255106	ADARB2	1.400	1.147	0.149	0.050	0.015	0,492	0.876	0.894	1
E2	34027888		1.274	1.275	0.098	0.097	-0.051	0,739	0.891	0.892	2
B1	158896635	FGF5	1.335	1.208	0.122	0.072	-0.066	-0,033	0.885	0.886	3
D2	91989307	TCF7L2	1.395	1.206	0.147	0.071	0.115	0,313	0.867	0.878	4
B4	149532846	TRIOBP	1.375	1.336	0.138	0.121	0.139	-0,061	0.847	0.855	5
C1	28702055	FAM77C	1.492	1.213	0.189	0.074	0.194	-0,034	0.832	0.850	6
A1	151648701	PDCH12	1.505	1.345	0.195	0.125	0.180	-0,066	0.800	0.825	7
E1	131587399	EIF4A3	1.495	1.365	0.191	0.133	0.193	-0,069	0.796	0.817	8
B1	10420438	ENPP6	1.472	1.404	0.180	0.149	0.139	-0,082	0.798	0.815	9
E2	38860686		1.672	1.239	0.279	0.084	0.442*	-0,039	0.750	0.789	10
A1	223501140		1.691	1.176	0.290	0.061	0.342	-0,026	0.757	0.787	11
F1	82716202		1.708	1.272	0.300	0.096	0.170	0,478	0.721	0.767	12
A2	201526186	DPP6	1.730	1.172	0.313	0.059	0.314	-0,025	0.732	0.764	13
D4	63622083	PALM2-AKAP2	1.648	1.352	0.266	0.128	0.203	0,128	0.734	0.763	14
C1	17428968	TTN	1.749	1.281	0.324	0.100	0.296	0,215	0.693	0.739	15
B1	176151181	KIT	1.803	1.239	0.359	0.084	-0.008	-0,039	0.669	0.723	16
E2	3147915	TNN13	1.814	1.187	0.367	0.064	0.049	0,385	0.668	0.719	17
A3	162208567	PLB1	1.809	1.148	0.363	0.051	0.191	-0,020	0.684	0.717	18
D1	18390852		1.730	1.447	0.313	0.168	0.192	-0,096	0.662	0.710	19
D1	15984279		1.856	1.140	0.398	0.048	0.222	0,492	0.648	0.696	20
D1	117527468	CD44	1.163	1.837	**0.056**	**0.381**	-0.026	0,351	0.747	0.695	21
B3	57141954		1.843	1.272	0.388	0.096	0.276	0,216	0.627	0.687	22
B3	77094074	LOC607552	1.275	1.847	**0.098**	**0.389**	0.436	0,077	0.696	0.678	23
E2	7580874	MYBPC	1.814	1.403	0.367	0.149	-0.018	0,511	0.609	0.672	24
F2	38395360		1.856	1.345	0.398	0.125	0.185	-0,066	0.594	0.655	25
E2	8422942	RPS11	1.373	1.861	**0.137**	**0.397**	0.140	0,391	0.667	0.653	26
A1	133621071		1.876	1.269	0.413	0.095	0.156	-0,046	0.592	0.653	27
F1	26100599	LAMC1	1.890	1.247	0.424	0.087	0.176	-0,041	0.583	0.651	28
A1	69424718	ABCC4	1.896	1.206	0.429	0.071	0.111	-0,032	0.589	0.646	29
A1	242150000	GHR	1.814	1.540	0.368	0.210	0.163	0,175	0.560	0.634	30
C1	52456776	C8B	1.918	1.202	0.448	0.070	0.112	-0,030	0.558	0.626	31
B1	202966562		1.911	1.304	0.442	0.109	0.216	-0,055	0.538	0.614	32
C2	106991233		1.875	1.487	0.413	0.186	0.136	0,153	0.526	0.608	33
D2	1020904	ACF	1.953	1.072	0.482	0.024	-0.011	-0,006	0.520	0.583	34
D1	116730000	CAT	1.952	1.108	0.481	0.036	0.144	-0,013	0.515	0.580	35
Average			**1.670**	**1.320**	**0.296**	**0.123**	**0.161**	**0.141**	**0.74**		

Allelic richness (Ar); expected heterozygosity (H_E); Inbreeding coefficient (F_{IS}); pairwise estimations of genetic differentiation between European wild and domestic cats (AMOVA pw-F_{CT}); loci scores averaged across INFOCALC estimations (average) and resulting ranking values (rank). Genes' acronyms indicate location of SNPs in the domestic cat genome. Voluntarily screened SNPs, based on previously known polymorphisms, are underlined.

Detection of hybridization

Bayesian analyses with and without prior information for domestic samples yielded globally identical results (data not shown). Hence, all the presented results were performed without prior non-genetic information. Assuming two major populations in STRUCTURE (K = 2), all domestic cats were clearly assigned to their expected cluster according to genetic variation at the 158 SNPs (Figure 2). However, 18 putative Europeans wildcats showed *qi* values to the domestic cluster above 0.92 and very narrow CI ranges (0.745 – 1.00): seven from Portugal, four from Spain, four from Italy, one from Scotland and two from Hungary (Table 3; Figure 2 and 3).

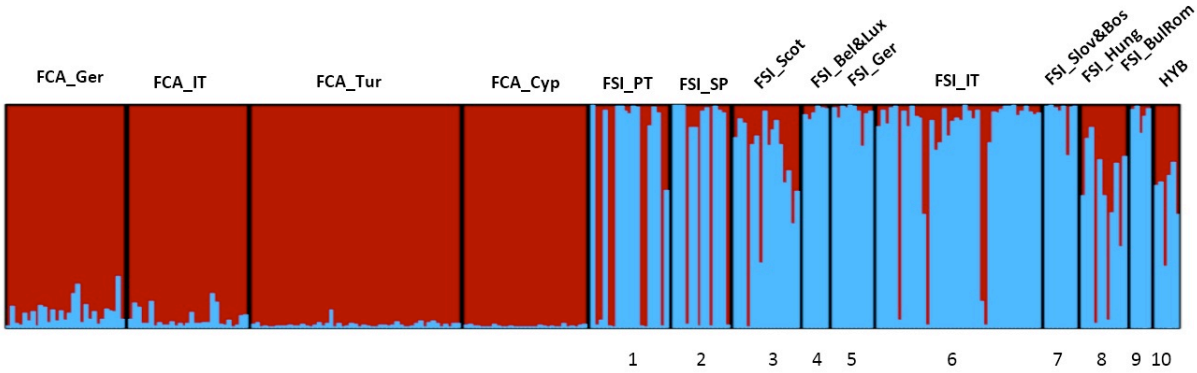


Figure 2. Average plot of the Bayesian admixture analyses performed in 10 independent STRUCTURE runs for K = 2, using 158 SNPs on 139 known random-bred domestic cats (FCA) and 132 putative European wildcats (FSI). Each individual is represented by a single vertical bar divided into two genetic clusters,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

According to the Bayesian analyses of the SNP variability, 23 putative wildcats have show genetic evidence of admixed ancestry both in STRUCTURE and NEWHYBRIDS computations (Table 3). The only exception was one cat from Scotland (ID 101), which was identified as a possible hybrid in STRUCTURE (q_{FSI}=0.768) and as a European wildcat in NEWHYBRIDS (q_{FSI} = 0.970). As predictable, most of the 23 admixed cats found in the random sampling belong to Scotland (n = 7) and Hungary (n = 8), which are well known for their high rates of hybridization (Beaumont *et al.* 2001; Pierpaoli *et al.* 2003; Lecis *et al.* 2006). Putative hybrid cats were moreover recognized in Portugal (n = 1), Germany (n = 1), Italy (n = 5), and Bosnia & Herzegovina (n = 1). Known captive-bred hybrids clearly displayed signatures of admixture, with individual *qi* ranging from 0.266 in the domestic cluster (0.186 < CI < 0.354) to 0.734 in the wild genetic group (0.646 < CI < 0.814). Moreover, they were mostly assigned to their known hybrid category with high posterior probabilities (*qi* > 0.90): ID 57 as F1, ID 60 as BxFSI, ID 61 as BxFCA and ID 63 as BxFSI (Table 3; Figure 3b). Overall, 89 of 130 (68.46%) putative wildcats, excluding the five known hybrids

but including Scottish cat 101 as possibly admixed, are suggested as having no introgression with domestic cats.

Table 3. Individual membership proportions (*qi*) of presumable misclassified and putatively admixed cats according to the Bayesian analyses performed in STRUCTURE and NEWHYBRIDS. STRUCTURE *qi* values correspond to allocations with K=2 to the domestic (FCA) and wild (FSI) inferred clusters, with their 90% credibility intervals (CI). NEWHYBRIDS *qi* values reflect assignments to the six possible hybrid categories: domestic, European wild, F1, F2 and backcrosses. Presumed misclassified cats that were eliminated from further analyses are shaded in light grey.

		STRUCTURE *qi*				NH *qi*	
ORIGIN	CAT	FCA	CI	FSI	CI	Hi	
Portugal	711	0.957	(0.893-0.999)	0.043	(0.001-0.107)	FCA	1.000
	1024	0.935	(0.853-0.996)	0.065	(0.004-0.147)	FCA	1.000
	297	0.966	(0.908-0.999)	0.034	(0.001-0.092)	FCA	1.000
	298	0.984	(0.946-1.000)	0.016	(0.000-0.054)	FCA	1.000
	688	0.969	(0.908-1.000)	0.031	(0.000-0.092)	FCA	1.000
	689	0.976	(0.924-1.000)	0.024	(0.000-0.076)	FCA	1.000
	706	0.967	(0.906-1.000)	0.033	(0.000-0.094)	FCA	1.000
	712	0.383	(0.300-0.470)	0.617	(0.530-0.700)	F2/BxFSI	0.84/0.16
Spain	1027	0.941	(0.869-0.995)	0.059	(0.005-0.131)	FCA	1.000
	717	0.973	(0.921-1.000)	0.027	(0.000-0.079)	FCA	1.000
	728	0.976	(0.930-1.000)	0.024	(0.000-0.070)	FCA	1.000
	737	0.965	(0.905-1.000)	0.035	(0.000-0.095)	FCA	1.000
Scotland	101	0.232	(0.209-0.256)	0.768	(0.744-0.791)	FSI/BxFSI	0.97/0.03
	105	0.979	(0.929-1.000)	0.021	(0.000-0.071)	FCA	1.000
	106	0.189	(0.129-0.256)	0.811	(0.744-0.871)	BxFSI/FSI	0.60/0.40
	240	0.705	(0.616-0.791)	0.295	(0.209-0.384)	BxFCA/F2	0.96/0.04
	252	0.194	(0.124-0.273)	0.806	(0.727-0.876)	BxFSI/FSI	0.98/0.02
	268	0.352	(0.262-0.447)	0.648	(0.553-0.738)	F2/BxFSI	0.70/0.30
	269	0.300	(0.223-0.382)	0.700	(0.618-0.777)	BxFSI/F2	0.99/0.01
	272	0.532	(0.435-0.630)	0.468	(0.370-0.565)	F2	1.000
	273	0.392	(0.305-0.483)	0.608	(0.517-0.695)	F2/BxFSI	0.78/0.22
Germany	629	0.201	(0.132-0.277)	0.799	(0.723-0.868)	BxFSI/FSI	0.99/0.01
Italy	624	0.920	(0.839-0.990)	0.080	(0.010-0.161)	FCA	1.000
	580	0.487	(0.392-0.585)	0.513	(0.415-0.608)	F2	1.000
	671	0.963	(0.891-1.000)	0.037	(0.000-0.109)	FCA	1.000
	677	0.216	(0.148-0.292)	0.784	(0.708-0.852)	BxFSI	1.000
	678	0.182	(0.119-0.252)	0.818	(0.748-0.881)	BxFSI/FSI	0.69/0.31
	918	0.844	(0.745-0.937)	0.156	(0.063-0.255)	FCA	1.000
	992	0.491	(0.400-0.583)	0.509	(0.417-0.600)	F1/F2	0.91/0.09
	1006	0.953	(0.880-0.999)	0.047	(0.001-0.120)	FCA	1.000
	1009	0.178	(0.113-0.251)	0.822	(0.749-0.887)	BXFSI/FSI	0.89/0.11
Bosnia	1056	0.235	(0.166-0.312)	0.765	(0.688-0.834)	BxFSI	1.000
Hungary	211	0.413	(0.327-0.504)	0.587	(0.496-0.673)	BxFSI/F2	0.5/0.5

	214	0.160	(0.101-0.228)	0.840	(0.772-0.899)	**BxFSI/FSI**	0.69/0.31
	339	0.941	(0.847-0.999)	0.059	(0.001-0.153)	**FCA**	1.000
	352	0.257	(0.185-0.335)	0.743	(0.665-0.815)	**BxFSI**	1.000
	356	0.409	(0.325-0.499)	0.591	(0.501-0.675)	**F2/BxFSI**	0.97/0.03
	358	0.917	(0.832-0.990)	0.083	(0.010-0.168)	**FCA**	1.000
	361	0.479	(0.386-0.576)	0.521	(0.424-0.614)	**F2**	1.000
	613	0.265	(0.187-0.349)	0.735	(0.651-0.813)	**BxFSI**	1.000
	620	0.628	(0.528-0.728)	0.372	(0.272-0.472)	**BxFCA/F2**	0.61/0.39
	621	0.244	(0.174-0.320)	0.756	(0.680-0.826)	**BxFSI**	1.000
Known Hybrids	**57**	0.400	(0.285-0.456)	0.600	(0.544-0.715)	**F1/F2**	0.91/0.09
	60	0.338	(0.257-0.423)	0.662	(0.577-0.743)	**BxFSI/F2**	0.93/0.7
	61	0.711	(0.616-0.804)	0.289	(0.196-0.384)	**BxFCA**	1.000
	62	0.321	(0.237-0.410)	0.679	(0.590-0.763)	**F2/BxFSI**	0.93/0.07
	63	0.266	(0.186-0.354)	0.734	(0.646-0.814)	**BxFSI**	1.000

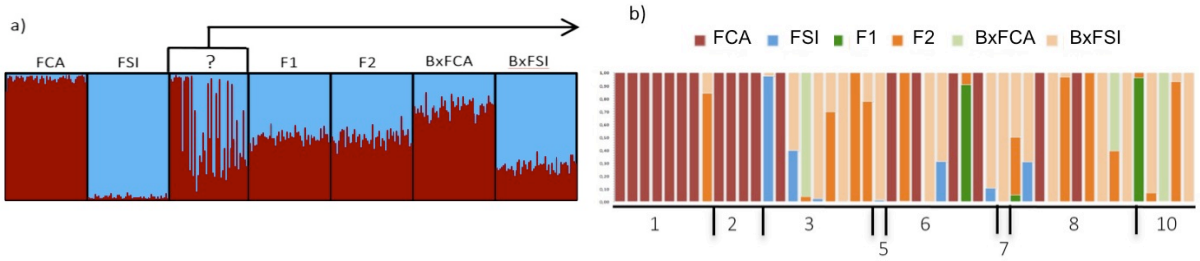


Figure 3. Individual membership (*qi*) values obtained using 158 SNPs under Bayesian-model computations: a) STRUCTURE's plot of 10 representatives of simulated domestic (FCA), wild (FSI), F1, F2 and backcross (BxFCA; BxFSI) genotypes and 47 real individuals for which genetic data refutes their straightforward allocation to the European wildcat subspecies (?); b) NEWHYBRID's assignment of the same 47 dubious individuals to the different hybrid categories. Each individual is represented by a single vertical bar coloured according to the proportion of their genome descending from each of the inferred clusters (Figure 3a) or hybrid class (Figure 3b).

SNPs simulations for Admixture Analysis

The parallel analyses of all 274 observed and 600 simulated genotypes, both in STRUCTURE and NEWHYBRIDS, globally revealed the same presumed misclassifications as obtained with true genotypes alone. A summary of the misclassifications expected in six simulated hybridization categories is presented in Table 4. Bayesian analyses of the simulated genotypes revealed that all parental, F1, F2 and backcrossed individuals could be correctly identified by the STRUCTURE algorithm using the 158 SNPs. Moreover, posterior probabilities of assignment to the different simulated categories of hybridization proved to be significantly demarcated since as few as 1% F2, 28% BxFCA and 14% BxFSI of the properly assigned genotypes displayed CI values outside the expected range (Table 4). Assignment values for NEWHYBRIDS proved to be equally accurate for parental and first-generation hybrids, but 4% F2, 3% BxFCA and 1% BxFSI were allocated to their own hybrid category with *qi* values lower than 0.85 (Table 3). Nevertheless, none of

the referred cats were significantly (qi > 0.85) allocated to one of the other remaining hybrid categories, preventing any case of misclassification. The simultaneous analysis of simulated and true genotypes confirmed results observed for real data alone (Table 4), both for what regards the probable miss-classification of cats according to morphology and the detection of hybrids (Figure 3).

Table 4 Average membership proportion (Q) of simulated genotypes in the Bayesian analysis performed using STRUCTURE and NEWHYBRIDS. Results represent *qi* values averaged over 10 independent runs. Minimum and maximum values of credibility intervals obtained in STRUCTURE are shown between brackets. The percentage of misclassified cats (WRONG), and the percentage of individuals for which CI ranges fallen outside the expected values (as shown in the first column) are shown. NEWHYBRIDS' average posterior probability assignment to the correct category and percentages of unclassified simulated genotypes are shown.

SIMULATED CATEGORIES	STRUCTURE FCA	STRUCTURE FSI	NEWHYBRIDS
FCA q_I>0,85*	0,959 (0,774-1,000)	0,040 (0,00-0,226)	0,9996
WRONG	0 (4%)		0
FSI q_{II}<0,15	0,025 (0,00-0,108)	0,975 (0,892-1,000)	0,9938
WRONG	0 (0%)		0
F_1 0,4<qi>0,6	0,499 (0,300-0,695)	0,501 (0,305-0,700)	0,9615
WRONG	0 (0%)		0
F_2 0,4< qi >0,6	0,494 (0,280-0,761)	0,506 (0,239-0,720)	0,9050
WRONG	0 (1%)		4%
B X FCA 0,15< qi >0,85	0,731 (0,509-0,924)	0,269 (0,100-0,491)	0,9455
WRONG	0 (28%)		3%
B X FSI 0,15< qi >0,85	0,270 (0,129-0,469)	0,730 (0,531-0,900)	0,9455
WRONG	0 (14%)		1%

The performance for detecting hybridization of the 35 top-ranked SNPs (**Table 2**) was evaluated by simulations on the modified dataset using STRUCTURE and NEWHYBRIDS (**Table 5**). The high differentiation power of some of the SNPs allowed an overall clear distinction of simulated parental and hybrid genotypes, as most individuals were assigned to their expected cluster with high posterior probabilities (qi > 0.80) using the reduced set of SNPs. STRUCTURE's miss-assignments were exclusively obtained for older generations of admixture, namely 8% of the simulated BxFCA and 4% of the simulated BxFSI. Very few parental and first-generation hybrids have shown ranges of CI outside the expected values but, as expected, less stringent credibility intervals were noticed for backcrosses, being that only 38% of BxFCA and 42% of BxFSI displayed CI ranges that never overlapped parental genotypes (Table 5). NEWHYBRIDS' clustering proved also to be highly efficient, with all parental, 98% F1, 90% F2, 90% BxFCA and 96% BxFSI being

correctly allocated to their category with high posterior probabilities (Table 5). It is moreover evident that only one of the unclassified genotypes (one BxFCA) would be incorrectly assigned to its correspondent parental group, with all of the other cases representing broad partitions among hybrid classes. Exceptionally, two simulated BxFSI were identified as F1 and F2, while two F2 were classified as F1 and BxFSI.

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

SIMULATED CATEGORIES	STRUCTURE FCA	STRUCTURE FSI	NEWHYBRIDS
FCA q_I>0,85*	0,959 (0,710-1,000)	0,040 (0,00-0,290)	0,9996
WRONG	0 (4%)		0
FSI q_{II}<0,15	0,025 (0,00-0,226)	0,975 (0,774-1,000)	0,9938
WRONG	0 (0%)		0
F_1 0,4<*qi*>0,6	0,499 (0,295-0,713)	0,501 (0,287-0,705)	0,9615
WRONG	0 (0%)		0
F_2 0,4< *qi* >0,6	0,494 (0,210-0,730)	0,506 (0,270-0,790)	0,9050
WRONG	0 (1%)		4%
B X FCA 0,15< *qi* >0,85	0,731 (0,388-0,950)	0,269 (0,050-0,612)	0,9455
WRONG	0 (28%)		3%
B X FSI 0,15< *qi* >0,85	0,270 (0,077-0,607)	0,730 (0,393-0,923)	0,9455
WRONG	0 (14%)		1%

DISCUSSION

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

SNPs have been attracting growing interest in a wide range of evolutionary applications and are becoming efficient tools among wildlife conservation-oriented studies (Brumfield *et al.* 2003; Morin *et al.* 2004; Seddon *et al.* 2005; Morin *et al.* 2009). Offering less variability per locus than STRs, SNPs provide a substantial number of advantages, namely: i) reduced propensity for homoplasy due to lower mutation rates; ii) higher density and more uniform distribution in genomes; iii) suitability for successful high-throughput genotyping and straightforward comparability and transportability across laboratories and detection protocols, and iii) highly successful application in fragmented DNA samples, e.g. non-invasive and historical DNA (see Brumfield *et al*. 2003; Morin *et al*. 2004; Garvin *et al*. 2010 for reviews). Nonetheless, the successful application of genome-wide batteries of nuclear SNPs in studies of wild populations is still limited to a few cases such as wolf-like species for studying their evolutionary history (vonHoldt *et al.* 2011), wild sheeps for detecting population structure and linkage disequilibrium (Miller *et al.* 2011), and wild Atlantic salmon for the differentiation of farmed and wild individuals (Karlsson *et al.* 2011). The present study provides the first analysis of nuclear SNPs for applications in European wildcat conservation and is motivated by the primary goal of improving molecular tools for detecting and quantifying hybridization.

Population Variability

SNPs were ascertained from the 1.9x genome sequence of an Abyssinian cat, suggesting that the selected SNPs do not represent the most variable loci among wildcat. Since the Abyssinian is one of the oldest foundation cat breeds (Lipinski *et al.* 2008; Menotti-Raymond *et al.* 2008), these SNPs would be expected to generally diverse amongst random bred cats and exhibit high H_E values. These same SNPs have been shown to be highly polymorphic across a variety of different cat breeds and populations (Kurushima *et al.* 2012), thus a breed-associated ascertainment bias appears minimal. Genetic diversity, including *Ar* and H_E, showed marked differences between European wildcats and domestic cats, with wildcats revealing significantly lower global genetic diversity, especially considering a broader geographic sampling in the wildcats. Generally, genetic variability is expected to be lower in domesticated forms relatively to their wild counterparts, due to the domestication bottleneck caused by the low number of founder individuals and the restricted gene flow imposed by human constrains (Doebley *et al.* 2006). However, the progenitor of the domestic cat is considered the *Felis s. lybica* subspecies (Driscoll *et al*. 2007; Lipinski *et al*. 2008), thus a direct comparison between the proposed wild progenitor and domesticate cannot be evaluated in this study. A striking example of the problematic inflating of H_E in the populations for which SNPs were initially ascertained was shown by Schuster *et al.* (2010), where African individuals genotyped for a large scale SNP chip developed mainly from European SNP data expressed lower heterozygosity than Europeans. Overall, some level of ascertainment bias likely originates as part of the divergence between random bred domestic cats and European wildcats, as well as the relatively low numbers of genotyped individuals and populations (Morin *et al.* 2004; see Helyar *et al.* 2011). A comparison of SNPs ascertained from domestic cats versus SNPs ascertained specifically from wildcats, such as the ~5,000 SNPs on the illumina Infinium Feline 63K iSelect DNA array, may help resolve the methodological causes of the increased genetic diversity observed in domestic cats as compared to wildcats.

No diagnostic alleles were identified between the domestic and the wild representatives in this study. Although only a very small subset of the species genome was analyzed, a similar result could be expected for larger number of SNPs. Amongst dogs and wolves, no diagnostic SNPs have been detected in a 48K panel from the Affymetrix Canine Mapping SNP 2.0 array (vonHoldt *et al.* 2012). Even so, when two populations are subjected to different selective pressure, some level of natural and artificial selections are expected to cause divergence in different parts of their genome. SNPs-based genomic approaches may provide exciting opportunities to assess differential rates of introgression across different genomic regions. Native California tiger salamanders (*Ambystoma californiense*) denote an excellent example of SNPs potential in studying detail admixture processes, since Fitzpatrick *et al.* (2010) were able to determine that only 3 out of 68 studied markers spread rapidly into native genomes, whereas the other 65 showed little evidence of introgression beyond the region where introductions of non-native barred tiger salamanders (*Ambystoma tigrinum mavortium*) occurred. By exposing such clear evidences of loci heterogeneity in introgression rates, this work reflected the vulnerability of studies using a few neutral markers to detect hybridization (Allendorf *et al.* 2010).

Bayesian Clustering

Ideally, the identification of recently introgressed hybrids, such as F1, F2 and first backcrosses, could be achieved with a minimum number of loci if the loci significantly disclose high levels of differentiation between the populations (Vähä and Primmer, 2006). The remarkable resemblance between European wildcats and domestic cats, and the intricate history of sympatry and introgression that most probably influenced both the domestication (Driscoll *et al.* 2007) and the expansion of domestic populations worldwide, might have created one of the most complicated frameworks to genetically discriminate parental groups of wild and domestic relatives. The Bayesian clustering of the 274 individuals (139 random-bred cats, 130 putative European wildcats and 5 known hybrids) immediately revealed higher discriminative power of genotypes over phenotypes in wildcats' identification. Eighteen putative wildcats were allocated with high posterior probabilities to the domestic cluster and therefore excluded from the analysis, in agreement with previous reports for the species (e.g. Oliveira *et al.* 2008a,b), suggesting morphological identification of European wildcat and domestic cats might not be as straightforward as some authors advocate (Ragni and Possenti, 1996; Daniels *et al.* 1998; Kitchener *et al.* 2005; Puzachenko 2002; Yamaguchi *et al.* 2004a,b; Krüger *et al.* 2009; Platz *et al.* 2011). A variety of issues could lead to mis-classification, including, i) dead animals might have been highly degraded at the time of collection and discrimination of obvious morphological characters might not be possible; ii) cats belong to past-generations of admixture and demarked diagnostic traits are no longer expressed; iii) samples were non-invasively collected (e.g. scats and hairs) and morphological discrimination were not possible; and iv) conservation biologist and naturalist bias their morphological evaluation towards the collection of wild specimens, namely out of urban areas, which is obviously much more attractive. The fact that most, if not all, backcrosses remained undetected under morphological evaluation further confirms the highest efficient of genotypes over phenotypes to identify past

generation hybrids. The set of markers defined in this study should effectively circumvent many cases of wrong pre-classification and steadily identify the origin of most unknown sample.

SNP Power for Admixture Analysis

Any ancestry inference must balance among economical, technical and statistical power concerns (Rosenberg *et al.* 2003). In the context of wildcat's conservation, the use of the relatively large number of loci made available in this study might not be beneficial for all applications, especially in cases of non-invasive sampling, or when analyses are performed to solve minor problems, such as assignment unknown samples to parental categories, rather than complex population/introgression inferences. Substantial reductions of genotyping effort could be realized if a subset of loci produced estimates of nearly the same quality as the full data set. The identification of highly informative SNP loci from larger panels has already been proposed as a powerful approach to identify wolf (*Canis lupus lupus*) x dog (*Canis lupus familiaris*) hybrids, twenty-four loci proven to be informative for assignment to recent hybrid classes (vonHoldt *et al.* 2012). If allocations are not definitive, a subsequent analysis of 100 loci is suggested (vonHoldt *et al.* 2012). In humans, subsets of informative SNPs delineate genetic relationships at the individual, parentage and population levels, namely for detecting human geographic structure (Liu *et al.* 2005; Lao *et al.* 2006). Similar studies in other species have also been conducted, such as for European bison (*Bison bonasus*; Tokarska *et al.* 2009), Atlantic salmon (*Salmo salar*; Glover *et al.* 2010), red fox (*Vulpes vulpes*) (Sacks and Louie 2008), and chicken breeds (Gärke *et al.* 2012). However, the choice of highly differentiated traits/loci from a small panel of individuals has been considered a possible reason for overlooking differential population (Brumfield *et al.* 2003; Morin *et al.* 2009; Schlotterer 2004), distorting the perception of the actual levels of introgressive hybridization in nature (Yuri *et al.* 2009). To obtain the most unbiased view of the genetic structure of introgression dynamics, different types of markers from the entire genome should be evaluated (Driscoll *et al.* 2011) and preferably represent both neutral and non-neutral variation (e.g. Teeter *et al.* 2008).

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

Specific SNPs that confer phenotypes that are well defined in domestic cats were genotyped in the wildcats, including the SNPs associated with Type B blood type, longhair, and coat color. The blood type SNP (at gene CMAH) has a high frequency in the random bred population of Turkey (Giger), the potential seat of cat domestication (Lipinski *et al.* 2008). Long hair is also suggested as an old variant (at gene FGF5) occurring near the time of cat domestication, likely originating in cats from Persian (LA Lyons, personal

communication). *TYR* color variants have been historically documented in random bred cats of South Asia. *TYRP1* SNPs cause variation in the amount of eumelanin, leading to variations in brown tones and hues of the pelage, which could likely be tolerated and potentially advantageous in the wildcat population. Although variation was detected in the random bred cats for each of these SNPs, only the longhair variant was present in six of the genetically identified wildcats, and always in the heterozygous state. Interestingly, the SNP variant associated with hypertrophic cardiomyopathy in Maine Coons cat (at gene MYBPC) has been detected in seven wildcats. From these SNPs, the ones located in the genes *FGF5* and *MYBPC* have been ranked among the top 35 loci, the 3rd and 24th most informative SNPs, respectively, due to disruptive distribution of allele's frequencies between the domestic and the wildcat groups. Additionally, mutations in *KIT*, the gene known to cause white spotting in several species such as pig and horses, a phenotype associated with domestication, was ranked as a highly diagnostic marker for hybridization, 16th of the 158 SNPs. Interestingly, Belyaev's experiments on the domestication of silver foxes showed that coat color changes, particularly white spotting, might be a by-product of selection for taming (Trut, 1999).

Detection of hybridization in natural populations using SNPs

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

The highly discriminating loci discovered in this study may bring new insights to the study of European wildcat populations, specifically a powerful and efficient tool to detect and quantify hybridization with domestic cats. Further genotyping of additional populations should help valid the selected SNPs. New throughput technologies under development for domestic cats will soon allow the evaluation of the entire genome of *Felis silvestris* species, supporting the identification of more diagnostic loci and potentially indicating areas of the genome involved with domestication. Limited X-linked SNPs were evaluated in this study and because of its transmission pattern, X-linked genes are good candidates for selection during domestication and deserve further investigation. SNPs have already demonstrated the potential to equal or even outperform microsatellites for specific questions such as individual ancestry (Lao *et al.* 2008), population assignment (e.g. Seddon *et al.* 2005; Narum *et al.* 2008; Smith and Seeb 2008; Coates *et al.* 2009) and pedigrees studies (Santure *et al.* 2010, Hauser *et al.* 2011), and proved to strongly segregate among populations (Freamo *et al.* 2011). However, both types of markers are useful in population genetic studies, a

combination of SNPs and microsatellites may be the most effective suite of loci (e.g. Narum *et al.* 2008; Hess *et al.* 2011) and could be powerful for cat introgression studies. Other phenotypic polymorphisms might be equally highly informative. For example, the Illumina technology proved to be poor technology for indels, therefore, specific known domestic cat phenotypes, such as melanism at the *Agouti* locus (*ASIP*), (Eizirik *et al.* 2003) would likely be an important diagnostic for domestic cat introgression into wildcats. Combined, repertoires of autosomal SNPs and STRs, X and Y-linked markers and mtDNA variants should all help decipher the domestication of the cat and the dynamics of wildcat and domestic cat populations around the world.

ACKNOWLEDGEMENTS

Funding for this study was provided by Fundação para a Ciência e a Tecnologia (FCT) through a PhD grant SFRH/BD/24361/2005 (RO) and in part by National Geographic Expedition Grant (EC0360-07), National Institutes of Health - National Center for Research Resources (NCRR) grant R24 RR016094R24, now the Office of Research Infrastructure Programs (ORIP) grant R24OD010928, the University of California - Davis, Center for Companion Animal Health, the Winn Feline Foundation, and a gift from Illumina, Inc., (LAL), and the University of California - Davis Wildlife Health Fellowship (JDK). We thank M. Herrman, F. Suchentrunk, M. Liberek, Natural Museum of Scotland, P. Romy, B. Szolt, A. Sforzi, B. Ragni, L. Lapini, A. de Faveri, K. Hupe, I. Eckert, H. Potocnik, M. Moes, G. Cagnolati, F. Vercillo, M. Malsaña, N. Mejias, J.M. Fernández, J.L. Robles, G.D. Penafiel, E.B. Duperón, M. Moleón, P. Monterroso, F. Alvares, J. Rodrigues, BTVS/ICNB: Portuguese National Tissue Bank/National Institute for Nature and Biodiversity Conservation, P Lyberakis and their collaborators for providing putative wildcat samples. We are also grateful to all anonymous veterinarians and biologists that assisted in samples collection.

REFERENCES

Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001) The problems with hybrids: setting conservation guidelines. *Trends in Ecology and Evolution* **16**: 613-622.

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. *Nature Reviews Genetics* **11** (10): 697-709.

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

Andersson L(2001) Genetic dissection of phenotypic diversity in farm animals. *Nat. Rev. Genet.* **2**: 130–138.

Beaumont M, Barratt EM, Gottelli D, Kitchener AC, Daniels MJ, Pritchard JK, Bruford MW (2001) Genetic diversity and introgression in the Scottish wildcat. *Molecular Ecology* **10**:319–336.

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

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

Coates BS, Sumerford DV, Miller NJ Kim Ks, Sappington TW, Siegfried BD, Lewis LC (2009) Comparative Performance of Single Nucleotide Polymorphism and Microsatellite Markers for Population Genetic Analysis. *Journal of Heredity* **100**: 556–564.

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

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

Doebley JF, Gaut BS, Smith BD (2006) The molecular genetics of crop domestication. *Cell* **127**: 1309-1321.

Driscoll CA, Menotti-Raymond M, Roca AL, Hupe K, Johnson WE, Geffen E, Harley EH, Delibes M, Pontier D, Kitchener AC, Yamaguchi N, O'Brien SJ, MacDonald DW (2007) The near eastern origin of cat domestication. *Science* **317**: 519–523.

Driscoll CA, MacDonald DW, O'Brien SJ (2009) From wild animals to domestic pets, an evolutionary view of domestication. *Proceedings of the National Academy of Sciences USA* **106**: 9971-9978.

Driscoll CA, Yamaguchi N, O'Brien SJ, and Macdonald DW (2011) A Suite of Genetic Markers Useful in Assessing Wildcat *(Felis silvestris* ssp.)— Domestic Cat (*Felis silvestris catus*) Admixture. *Journal of Heredity* **102**: S87-S90.

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

Eizirik E, Yuhki N, Johnson WE, Menotti-Raymond M, Hannah SS, O'Brien S (2003) Molecular Genetics and Evolution of Melanism in the Cat Family. *Current Biology* **13**: 448-453.

Esteves PJ, Abrantes J and van der Loo W (2007) Extensive gene conversion between CCR2 and CCR5 in domestic cat (*Felis catus*) *International Journal of Immunogenetics* **34**: 321-324.

Excoffier L and Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 10: 564-567

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

Fitzpatrick BM, Johnson JR, Kump DK, Smith JJ, Voss SR, Shaffer HB (2010) Rapid spread of invasive genes into a threatened native species. *Proceedings of the National Academy of Sciences USA* **107**: 3606–3610.

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

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

Gärke C, Ytournel F, Bed'hom B, Gut I, Lathrop M, Weigend S and Simianer H (2012) Comparison of SNPs and microsatellites for assessing the genetic structure of chicken populations. *Animal Genetics* **43**: 419-428.

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

Goudet J (2001) FSTAT, a Program to Estimate and Test Gene Diversities and Fixation Indices Version 2.9.3. Available from www.unil.ch/izea/softwares/fstat.html.

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

Helyar SJ, Hemmer-hansen J, Bekkevold D, Taylor MI, Ogden R, Limborg MT, Cariani A, Maes GE, Diopere E, Carvalho GR, Nielsen EE (2011) Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Molecular Ecology Resources* **11**: 123–136.

Huson DH and Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23**: 254–267.

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

Jonhson WE, Eizirik E, Pecon-Slattery J, Murphy WJ, Antunes A, Teeling E, O'Brien S (2006) The late Miocene radiation of modern Felidae: A genetic assessment. *Science* **311**: 73-77.

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

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

Kauer MO, Dieringer D, Schlötterer C (2003) A microsatellite variability screen for positive selection associated with the "Out of Africa" habitat expansion of Drosophila melanogaster. *Genetics* **165**: 1137-1148.

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

Krüger M, Hertwig ST, Jetschke G, Fischer MS (2009) Evaluation of anatomical characters and the question of hybridization with domestic cats in the wildcat population of Thuringia, Germany. *Journal of Zoology systematics and evolutionary research* **47**: 268–282.

Kurushima JD, Lipinski MJ, Gandolfi B, Froenicke L, Grahn JC, Grahn RA, Lyons LA (2012) Variation of cats under domestication: genetic assignment of domestic cats to breeds and worldwide random-bred populations. *Animal Genetics*. doi: 10.1111/age.12008.

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

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

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

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

Liu N, Chen L, Wang S, Oh C, Zhao H (2005) Comparison of singlenucleotide polymorphisms and microsatellites in inference of population structure. *BMC Genetics* **6**(Suppl. 1): S26.

Lyons L (2010) Feline Genetics: Clinical Applications and Genetic Testing Topics in Companion Animal. *Medicine* **25**: 203-212.

Lyons L (2012) Genetic testing in domestic cats *Molecular and Cellular Probes* **26**(6): 224-30.

McEwing R, Kitchener AC, Holleley C, Kilshaw K, O'Donoghue P (2011) An allelic discrimination SNP assay for distinguishing the mitochondrial lineages of European wildcats and domestic cats. *Conservation Genetics Resources*. doi: 10.1007/s12686-011-9499-z.

Menotti-Raymond M, David VA, Pflueger SM, Lindblad-Toh K, Wade CM, O'Brien SJ, and Johnson WE (2008) Patterns of molecular genetic variation among cat breeds. *Genomics* **91**: 1–11.

Miller J, POissant J, Kijas J, Coltman D and the International Sheep Genomics Consortium (2011), A genome-wide set of SNPs detects population substructure and long range linkage disequilibrium in wild sheep. *Molecular Ecology Resources* **11**: 314–322.

Morin PA, Luikart G, Wayne RK, SNP-Workshop Group (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution* **19**: 208–216.

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

Mullikin JC, Hansen NF, Shen L, Ebling H, Donahue WF, Tao W, Saranga DJ, Brand A, Rubenfield MJ, Young AC, Cruz P; NISC Comparative Sequencing Program, Driscoll C, David V, Al-Murrani SW, Locniskar MF, Abrahamsen MS, O'Brien SJ, Smith DR, Brockman JA (2010) Light whole genome sequence for SNP discovery across domestic cat breeds. *BMC Genomics* **11**: 406.

Narum S, Banks M, Beacham T, Bellinger M, Campbell M, Dekoning J, Elz A, Guthrie III C, Kozfkay C, Miller K, Moran P, Phillips R, Seeb l, Smith C, Warheit K, Young S, Garza J (2008) Differentiating salmon populations at broad and fine geographical scales with microsatellites and single nucleotide polymorphisms. *Molecular Ecology* **17**: 3464–3477.

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

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

O'Brien SJ, Devillard S, Say L, Vanthomme H, Léger F, Ruette S, Pontier D (2009) Preserving genetic integrity in a hybridising world: are European Wildcats (*Felis silvestris silvestris*) in Eastern Francedistinct from sympatric feral domestic cats? *Biodiversity and Conservation* **18**: 2351-2360.

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

Oliveira R, Godinho R, Randi E and Alves PC (2008b) Hybridization vs conservation: are domestic cats threatening the genetic integrity of European wildcat (*Felis silvestris silvestris*) populations in Iberian Peninsula? *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **363**: 2953-2961.

Peakall R and Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**: 288-295.

Pertoldi C, Wójcik JM, Matgorzata T, Kawalko A, Kristensen TN, Loeschcke V, Gregersen VR, Coltman D, Wilson GA, Randi E *et al.*. (2010) Genome variability in European and American bison detected using the BovineSNP50 BeadChip. *Conservation Genetics* **11**: 627–634.

Pierpaoli M, Biro ZS, Herrmann M, Hupe K, FernandesM, Ragni B, Szemethy L, Randi E (2003) Genetic distinction of wildcat (Felis silvestris) populations in Europe, and hybridization with domestic cats in Hungary. *Molecular Ecology* **12**: 2585–2598.

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

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

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

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

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

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

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

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

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

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes* **4**: 137–138.

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

Sacks BN and Louie S (2008) Using the dog genome to find single nucleotide polymorphisms in red foxes and other distantly related members of the Canidae. *Molecular Ecology Resources* **8**: 35-49.

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

Schlötterer, C (2004) The evolution of molecular markers – just a matter of fashion? *Nature Review Genetics* **5**: 63–69.

Schuster SC, Miller W, Ratan A, Tomsho LP, Giardine B, Kasson LR, Harris RS, Petersen DC, Zhao F, Qi J, *et al.*. (2010) Complete Khoisan and Bantu genomes from southern Africa. *Nature* **463**: 943–947.

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

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

Teeter KC, Payseur BA, Harris LW, Bakewell MA, Thibodeau LM, O'Brien JE, Krenz JG, Sans-Fuentes M, Nachman MW, Tucker PK (2008) Genome-wide patterns of gene flow across a house mouse hybrid zone. *Genome Research* **18**: 67-76.

Tokarska M, Marshall T, Kowalczyk R, Wojcik JM, Pertoldi C, Kristensen TN, Loeschcke V, Gregersen VR, Bendixen C (2009) Effectiveness of microsatellite and SNP markers for parentage and identity analysis in species with low genetic diversity: the case of European bison. *Heredity* **103**: 326-332.

Trut L (1999) Early Canid Domestication: the Farm-Fox Experiment. *American Scientist* **87**: 160-169.

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

Vigne JD, Briois F, Zazzo A, Willcox G, Cucchi T, Thiébault S, Carrère I, Franel Y, Touquet R, Martin C, Moreau C, Comby C, Guilaine J (2012) First wave of cultivators spread to Cyprus at least 10,600 y ago. *Proceedings of the National Academy of Sciences of the United States of America* **109**(22): 8445-8449.

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

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

vonHoldt BM, Pollinger JP, Earl DA, Parker HG, Ostrander EA, Wayne RK (2012) Identification of recent hybridization between gray wolves and domesticated dogs by SNP genotyping. *Mammalian Genome* doi: 10.1007/s00335-012-9432-0.

Wiener and Wilkinson (2011) Deciphering the genetic basis of animal domestication. *Proceedings of the Royal Society B* **278**: 3161-3170.

Yamaguchi N, Driscoll CA, Kitchener AC, Ward JM, Macdonald DW (2004a) Craniological differentiation between European wildcats (*Felis silvestris silvestris*), African wildcats (*F. s. lybica*) and Asian wildcats (*F. s. ornata*): implications for their evolution and conservation. *Biological Journal of the Linnean Society* **83**: 47-63.

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

Yuri T, Jernigan RW, Brumfield RT, Bhagabati NK, and Braun MJ (2009) The effect of marker choice on estimated levels of introgression across an avian (Pipridae: Manacus) hybrid zone. *Molecular ecology* **18**: 4888-903.

SUPPLEMENTARY MATERIAL

Supplementary Table 1. Summary of genetic variability at 154 autosomal SNPs for the domestic cats; for the European wildcats excluding identified hybrids, individuals strongly assigned to the domestic cluster and admixed populations from Hungary and Scotland; for all the putative wildcats including identified hybrids; for the identified hybrids alone and for 4 of the captive known hybrids (one F1, one F2, one BxFCA and one BXFSI).

	Domestic cats n=139	European wildcats n=82	Putative wildcats n=112	Identified Hybrids n=23	Captive known Hybrids n=4
Nº polymorphic loci	154	132	149	143	116
Ar	1.738	1.250	1.350	1.570	1.640
H_O	0.295	0.104	0.120	0.239	0.402
H_E	0.340	0.107	0.145	0.248	0.291
F_{IS}	0.131	0.169	0.197	0.104	-0.043
F_{CT}		0.427	0.395	0.220	0.161

Ar = allelic richness; H_O and H_E observed and expected heterozygosities; F_{IS}= inbreeding coefficient; F_{CT}= estimates of genetic differentiation to the domestic cat group (AMOVA).

Supplementary Table 2. Genetic description of all SNP loci used in this study: Allelic richness (Ar); expected heterozygosity (H_E); Inbreeding coefficient (F_{IS}); pairwise estimations of genetic differentiation between European wild and domestic cats (AMOVA pw-F_{CT}); loci scores averaged across INFOCALC estimations (score) and resulting ranking values (R). Genes' acronyms indicate SNPs that could be located in the domestic cat genome (underlined ones indicate location of SNPs that were voluntarily screened based on previously known polymorphisms).

			Ar		H_E†		FIS		AMOVA	INFOCALC	
Chr	**POSITION**	**Location**	**FCA**	**FSI**	**FCA**	**FSI**	**FCA**	**FSI**	**pw-F_{CT}**	**score**	**R**
B4	255106	ADARB2	1.400	1.147	0.149	0.050	0.015	0,492	0.876	0.894	1
E2	34027888		1.274	1.275	0.098	0.097	-0.051	0,739	0.891	0.892	2
B1	158896635	FGF5	1.335	1.208	0.122	0.072	-0.066	-0,033	0.885	0.886	3
D2	91989307	TCF7L2	1.395	1.206	0.147	0.071	0.115	0,313	0.867	0.878	4
B4	149532846	TRIOBP	1.375	1.336	0.138	0.121	0.139	-0,061	0.847	0.855	5
C1	28702055	FAM77C	1.492	1.213	0.189	0.074	0.194	-0,034	0.832	0.850	6
A1	151648701	PDCH12	1.505	1.345	0.195	0.125	0.180	-0,066	0.800	0.825	7
E1	131587399	EIF4A3	1.495	1.365	0.191	0.133	0.193	-0,069	0.796	0.817	8
B1	10420438	ENPP6	1.472	1.404	0.180	0.149	0.139	-0,082	0.798	0.815	9
E2	38860686		1.672	1.239	0.279	0.084	0.442*	-0,039	0.750	0.789	10
A1	223501140		1.691	1.176	0.290	0.061	0.342	-0,026	0.757	0.787	11
F1	82716202		1.708	1.272	0.300	0.096	0.170	0,478	0.721	0.767	12
A2	201526186	DPP6	1.730	1.172	0.313	0.059	0.314	-0,025	0.732	0.764	13
D4	63622083	PALM2-AKAP2	1.648	1.352	0.266	0.128	0.203	0,128	0.734	0.763	14
C1	17428968	TTN	1.749	1.281	0.324	0.100	0.296	0,215	0.693	0.739	15
B1	176151181	KIT	1.803	1.239	0.359	0.084	-0.008	-0,039	0.669	0.723	16
E2	3147915	TNN13	1.814	1.187	0.367	0.064	0.049	0,385	0.668	0.719	17
A3	162208567	PLB1	1.809	1.148	0.363	0.051	0.191	-0,020	0.684	0.717	18
D1	18390852		1.730	1.447	0.313	0.168	0.192	-0,096	0.662	0.710	19
D1	15984279		1.856	1.140	0.398	0.048	0.222	0,492	0.648	0.696	20
D1	117527468	CD44	1.163	1.837	**0.056**	**0.381**	-0.026	0,351	0.747	0.695	21
B3	57141954		1.843	1.272	0.388	0.096	0.276	0,216	0.627	0.687	22

B3	77094074	LOC607552	1.275	1.847	**0.098**	**0.389**	0.436	0,077	0.696	0.678	23
E2	7580874	MYBPC	1.814	1.403	0.367	0.149	-0.018	0,511	0.609	0.672	24
F2	38395360		1.856	1.345	0.398	0.125	0.185	-0,066	0.594	0.655	25
E2	8422942	RPS11	1.373	1.861	**0.137**	**0.397**	0.140	0,391	0.667	0.653	26
A1	133621071		1.876	1.269	0.413	0.095	0.156	-0,046	0.592	0.653	27
F1	26100599	LAMC1	1.890	1.247	0.424	0.087	0.176	-0,041	0.583	0.651	28
A1	69424718	ABCC4	1.896	1.206	0.429	0.071	0.111	-0,032	0.589	0.646	29
A1	242150000	GHR	1.814	1.540	0.368	0.210	0.163	0,175	0.560	0.634	30
C1	52456776	C8B	1.918	1.202	0.448	0.070	0.112	-0,030	0.558	0.626	31
B1	202966562		1.911	1.304	0.442	0.109	0.216	-0,055	0.538	0.614	32
C2	106991233		1.875	1.487	0.413	0.186	0.136	0,153	0.526	0.608	33
D2	1020904	ACF	1.953	1.072	0.482	0.024	-0.011	-0,006	0.520	0.583	34
D1	116730000	CAT	1.952	1.108	0.481	0.036	0.144	-0,013	0.515	0.580	35
F1	38051725	PFKFB2	1.955	1.150	0.484	0.051	0.149	-0,020	0.491	0.565	36
E1	4114158		1.954	1.236	0.483	0.083	0.106	0,259	0.472	0.544	37
B4	105706694		1.951	1.294	0.480	0.105	0.341*	-0,053	0.467	0.541	38
A1	68485376	GPC6	1.571	1.910	**0.227**	**0.440**	0.173	0,490*	0.492	0.538	39
C2	20000000	APP	1.955	1.298	0.483	0.106	0.536*	-0,053	0.473	0.538	40
A3	75156179		1.943	1.408	0.471	0.151	0.212	0,084	0.447	0.533	41
C2	156491175		1.942	1.543	0.471	0.212	0.158	0,109	0.397	0.503	42
E1	5453028		1.968	1.234	0.498	0.082	0.124	-0,038	0.410	0.485	43
A1	27523501		1.955	1.506	0.483	0.194	0.218	0,016	0.378	0.473	44
D1	104941557		1.669	1.935	**0.278**	**0.462**	0.117	0,356	0.388	0.461	45
B1	88148379	FNIP2	1.969	1.000	0.498	0.000	-0.015	NA	0.416	0.457	46
C2	147124460	MYRIP	1.561	1.953	0.496	0.000	-0.019	0,048	0.408	0.457	47
B1	80161671	PALLD	1.967	1.000	**0.222**	**0.481**	0.288	NA	0.398	0.456	48
D4	41078218		1.970	1.147	0.500	0.050	-0.028	-0,020	0.387	0.453	49
A1	175780586	TMEM171	1.968	1.077	0.498	0.026	0.328*	-0,007	0.386	0.452	50
B4	21098349		1.970	1.276	0.500	0.098	0.269	0,254	0.343	0.438	51
B2	45093345		1.970	1.349	0.500	0.127	-0.034	0,323	0.331	0.426	52
B3	13666494		1.967	1.174	0.496	0.060	-0.064	-0,026	0.348	0.423	53
E2	7950477		1.964	1.108	0.494	0.036	0.160	0,664	0.358	0.425	54
A3	38781591	PLCB1	1.143	1.964	**0.049**	**0.492**	-0.022	0,446*	0.444	0.412	55
B2	3940000	CMAH	1.958	1.000	0.487	0.000	0.379*	NA	0.362	0.413	56
X	4696293		1.968	1.319	0.497	0.115	0.050	0,578	0.311	0.404	57
B4	40319102		1.956	1.072	0.484	0.024	0.099	-0,006	0.339	0.400	58
B2	138312489		1.955	1.072	0.484	0.024	0.193	-0,006	0.336	0.399	59
B4	47638578	CHD4	1.163	1.965	**0.056**	**0.492**	-0.026	0,061	0.444	0.398	60
A3	91058022		1.959	1.174	0.488	0.060	0.196	-0,026	0.316	0.393	61
D1	18570323		1.953	1.106	0.482	0.036	-0.011	-0,013	0.319	0.387	62
B3	76202196	ACTC	1.955	1.203	0.484	0.070	0.040	-0,032	0.296	0.375	63
A1	8742286		1.945	1.106	0.474	0.036	-0.045	-0,013	0.298	0.367	64
A2	152258936		1.944	1.147	0.473	0.050	0.082	-0,020	0.284	0.359	65
E2	22632289	CEP89	1.933	1.037	0.462	0.012	0.018	0,000	0.300	0.356	66
B3	39203469	SCAMP2	1.935	1.074	0.464	0.025	0.095	-0,006	0.291	0.355	67
B4	143006494		1.941	1.145	0.470	0.049	0.152	-0,020	0.281	0.352	68
C1	190502133		1.880	1.933	**0.416**	**0.461**	0.040	0,238	0.211	0.335	69
C2	126240000	RASA2	1.125	1.932	**0.043**	**0.459**	-0.019	0,243	0.365	0.326	70
X	6976318		1.925	1.143	0.455	0.049	0.106	-0,019	0.254	0.326	71
D3	25530000	GNAZ	1.931	1.206	0.461	0.071	0.034	0,313	0.246	0.324	72
E3	36044809		1.913	1.038	0.444	0.013	0.091	0,000	0.268	0.321	73
F1	565223		1.940	1.275	0.469	0.097	0.074	-0,048	0.237	0.320	74
D1	101321498		1.946	1.375	0.475	0.137	0.353*	-0,074	0.217	0.315	75
B4	147206961	TOM1	1.902	1.038	0.434	0.013	0.057	0,000	0.261	0.312	76
C1	215441574		1.915	1.140	0.446	0.048	0.148	0,492	0.245	0.312	77
B1	12214271		1.927	1.263	0.457	0.093	0.070	-0,045	0.221	0.305	78
B4	144693308		1.921	1.206	0.452	0.071	0.265	-0,032	0.230	0.302	79
A2	19001000	CCR2	1.129	1.912	**0.044**	**0.442**	0.321	0,432*	0.325	0.298	80
A2	202225770		1.894	1.106	0.428	0.036	0.079	-0,013	0.230	0.286	81

D2	717969	SGMS1	1.875	1.037	0.413	0.012	0.352*	0,000	0.233	0.285	82
C1	396397		1.901	1.213	0.433	0.074	0.279	-0,034	0.201	0.280	83
E2	35914023		1.872	1.073	0.410	0.024	0.126	-0,006	0.217	0.276	84
D1	128010000	MYBPC3	1.854	1.037	0.396	0.012	0.081	0,000	0.216	0.266	85
D1	10789012		1.867	1.108	0.406	0.036	0.188	-0,013	0.203	0.265	86
E3	55434272		1.884	1.210	0.419	0.073	0.083	0,313	0.187	0.262	87
A3	130195244		1.856	1.074	0.397	0.025	0.178	-0,006	0.207	0.261	88
C1	123164748		1.852	1.073	0.395	0.024	0.121	-0,006	0.203	0.258	89
B3	81790000	MYH7	1.891	1.269	0.425	0.095	0.317	0,217	0.174	0.258	90
A3	12082294	PTPRT	1.841	1.000	0.386	0.000	-0.075	NA	0.219	0.255	91
C1	116355295		1.879	1.266	0.415	0.094	0.012	-0,046	0.165	0.246	92
C2	262401		1.833	1.076	0.380	0.025	0.083	-0,007	0.186	0.243	93
D1	118901000	RAG1	1.858	1.172	0.399	0.059	0.028	-0,025	0.177	0.240	94
E2	39211557		1.942	1.615	0.471	0.247	0.235	0,252	0.117	0.240	95
B1	54775572		1.804	1.000	0.360	0.000	-0.023	NA	0.193	0.236	96
B4	146486983		1.797	1.000	0.356	0.000	0.290	NA	0.191	0.231	97
C2	5215469		1.803	1.037	0.360	0.012	0.070	0,000	0.183	0.229	98
B1	68520000	CLU	1.808	1.076	0.363	0.025	0.284	1.000	0.173	0.226	99
B1	195678303		1.788	1.000	0.350	0.000	0.178	NA	0.186	0.226	100
C1	216852686		1.797	1.045	0.355	0.015	0.377*	0,000	0.173	0.224	101
D2	105772916		1.797	1.038	0.356	0.012	0.206	0,000	0.178	0.219	102
B2	6949528		1.817	1.141	0.369	0.048	0.322	-0,019	0.158	0.214	103
B2	41509834		1.766	1.000	0.336	0.000	0.083	NA	0.174	0.213	104
B3	51317931		1.776	1.037	0.342	0.012	0.201	0,000	0.167	0.213	105
A3	99507784		1.828	1.234	0.377	0.082	0.093	-0,038	0.137	0.210	106
B3	104483970		1.809	1.141	0.364	0.048	0.256	-0,019	0.153	0.209	107
B4	20001848		1.186	1.823	**0.064**	**0.372**	0.199	0,606*	0.199	0.208	108
D1	16242433		1.791	1.143	0.351	0.049	-0.102	-0,019	0.146	0.197	109
D1	126256993		1.733	1.000	0.315	0.000	0.048	NA	0.160	0.196	110
A1	10141047		1.757	1.110	0.330	0.037	-0.030	-0,013	0.136	0.190	111
F2	68572596		1.721	1.000	0.308	0.000	0.054	NA	0.154	0.190	112
F1	27124984		1.743	1.074	0.321	0.025	-0.020	1.000	0.139	0.189	113
B4	142658074		1.742	1.073	0.320	0.024	0.186	1.000	0.141	0.188	114
E1	48228153		1.742	1.072	0.320	0.024	0.234	-0,006	0.142	0.182	115
D1	125811329		1.767	1.185	0.336	0.063	0.095	-0,023	0.104	0.181	116
A1	208054462		1.740	1.113	0.319	0.038	0.132	-0,013	0.126	0.180	117
C1	24148281		1.761	1.176	0.332	0.061	0.174	-0,026	0.117	0.180	118
X	5142294		1.736	1.108	0.316	0.036	0.219	0,664	0.126	0.179	119
X	30335088		1.713	1.075	0.303	0.025	0.361	1.000	0.133	0.174	120
E2	36986631		1.957	1.956	0.486	0.483	0.116	0,341	0.055	0.173	121
A1	223506906		1.669	1.000	0.278	0.000	0.038	NA	0.136	0.167	122
F1	21799641		1.721	1.140	0.308	0.048	0.102	0,492	0.111	0.166	123
F2	46855978		1.871	1.965	**0.409**	**0.492**	-	-	0.045	0.153	124
D3	24823793	CABIN1	1.612	1.000	0.247	0.000	0.194	NA	0.114	0.144	125
B3	111000326		1.613	1.037	0.248	0.012	0.127	0,000	0.103	0.139	126
A1	225057933		1.806	1.479	0.362	0.182	-0.020	0,172	0.054	0.138	127
E3	67006512	PRSS27	1.590	1.000	0.236	0.000	0.424*	NA	0.110	0.137	128
D2	74293444		1.669	1.178	0.278	0.061	-0.033	-0,026	0.080	0.136	129
E1	130875919	HRNBP3	1.587	1.000	0.235	0.000	0.507*	NA	0.109	0.136	130
F2	8427817		1.955	1.971	**0.483**	**0.498**	-0.003	0,356	0.023	0.132	131
C1	44520932	RSP01	1.586	1.037	0.234	0.012	0.168	0,000	0.096	0.129	132
D1	126847301		1.613	1.072	0.248	0.024	0.203	-0,006	0.094	0.127	133
F2	74863327		1.727	1.319	0.311	0.115	-0.044	0,154	0.064	0.127	134
D1	105498119		1.588	1.073	0.235	0.024	0.159	-0,006	0.085	0.124	135
E1	3912105		1.569	1.072	0.226	0.024	-0.014	-0,006	0.080	0.118	136
E2	65436639		1.899	1.963	**0.432**	**0.491**	0.047	0,318	0.024	0.118	137
F1	91517402	RGS5	1.785	1.912	**0.348**	**0.442**	0.222	0,032	0.023	0.115	138
A1	235579538		1.537	1.041	0.210	0.014	0.088	0,000	0.078	0.113	139
B1	105520000	SLC7A11	1.529	1.037	0.206	0.012	0.225	0,000	0.080	0.111	140

F2	26886470	PAG1	1.507	1.000	0.196	0.000	0.105	NA	0.084	0.110	141
B1	199564532		1.522	1.037	0.203	0.012	-0.053	0,000	0.078	0.109	142
D1	66177762		1.540	1.073	0.212	0.024	0.005	-0,006	0.071	0.108	143
C2	187325	ADARB1	1.497	1.000	0.191	0.000	0.298	NA	0.075	0.107	144
A1	7429296		1.561	1.129	0.222	0.044	-0.072	-0,015	0.058	0.105	145
A2	554046	MEX3D	1.499	1.039	0.193	0.013	0.107	0,000	0.070	0.102	146
D3	122502120		1.765	1.501	0.335	0.192	0.179	0,281	0.033	0.102	147
D4	42000379		1.464	1.000	0.176	0.000	0.066	NA	0.075	0.098	148
C2	150774106		1.442	1.000	0.167	0.000	0.079	NA	0.071	0.092	149
D2	1752007		1.453	1.037	0.171	0.012	0.341	0,000	0.061	0.089	150
E1	48700963		1.398	1.000	0.148	0.000	0.015	NA	0.061	0.080	151
B4	1687419		1.397	1.000	0.147	0.000	-0.083	NA	0.061	0.080	152
D4	28094000	TYRP1	1.364	1.000	0.134	0.000	0.142	NA	0.053	0.072	153
D1	70545000	TYR	1.360	1.000	0.132	0.000	0.380	NA	0.053	0.071	154
B4	3093827		1.487	1.203	0.187	0.070	0.349	0,314	0.026	0.069	155
D3	24565823		1.351	1.037	0.128	0.012	0.043	0,000	0.040	0.063	156
A3	159537633	GPR113	1.960	1.971	**0.489**	**0.499**	0.042	0,387	0.000	0.049	157
F2	78303221		1.966	1.969	**0.495**	**0.497**	0.031	0,187	0.000	0.005	158
Mean (autosomal SNPs)			**1.738**	**1.250**	**0.340**	**0.107**	**0.131**	**0.169**	**0.427**		

* Significant deviations from HWE at P<0.05. Bonferroni corrected

+Significant deviations from LE at P<0.05. Bonferroni corrected

‡Monomorphic loci are underlined and cases where wildcats displayed higher genetic diversity than domestic cats are shown in bold

NA = not analysed

CHAPTER 5

General Discussion

"No tame animal has lost less of its native dignity or maintained more of its ancient reserve. The domestic cat might rebel tomorrow."

William Conway

The human-mediated dispersal of the domestic cat pan-globally together with the past demographic decline of wildcats' populations and the fragmentation of suitable habitat have increased the risk of artificial hybridization and have promoted the extinction of some natural populations of the European wildcat. This work contributed mainly to better understand the current state of introgressive hybridization between wild and domestic cats and the population structure of wildcats in Europe, and trailed a forceful birth of new dilemmas and hindrances that progressively led to a crescendo in data production and methodological choices. First we assessed levels of hybridization in the Iberian Peninsula (*Papers I* and *II*), an area where the knowledge on wildcat populations is very scarce despite the high ecological relevance of the species, as being the only felid besides the Iberian lynx (*Lynx pardinus*), which is in turn critical endangered (cit iucn). When we started to assess the hybridization status of Iberian populations, two main problems arisen: a) the huge difficulty in obtaining DNA samples from this elusive and rare species, and b) the considerable uncertainty when determining hybrid cats. The elusive behaviour and the low population numbers of this endangered felid are the main causes of the first problem, while the intriguing process of cat domestication, the probable long-standing admixture between wild and domestic forms and the high similarity in their morphology and genomes contribute together to the second. To overcome both difficulties, we examined the possibility of using non-invasive sampling strategies (*Papers III* and *IV*), increased our initial battery of neutral microsatellites (*Paper V*) and investigated new molecular tools to establish a set of highly informative SNPs of differentiation between the two forms (*Paper VI*). The flowchart of crucial steps that guided the pathways of our research and of manuscripts' writing is represented in Figure 1.

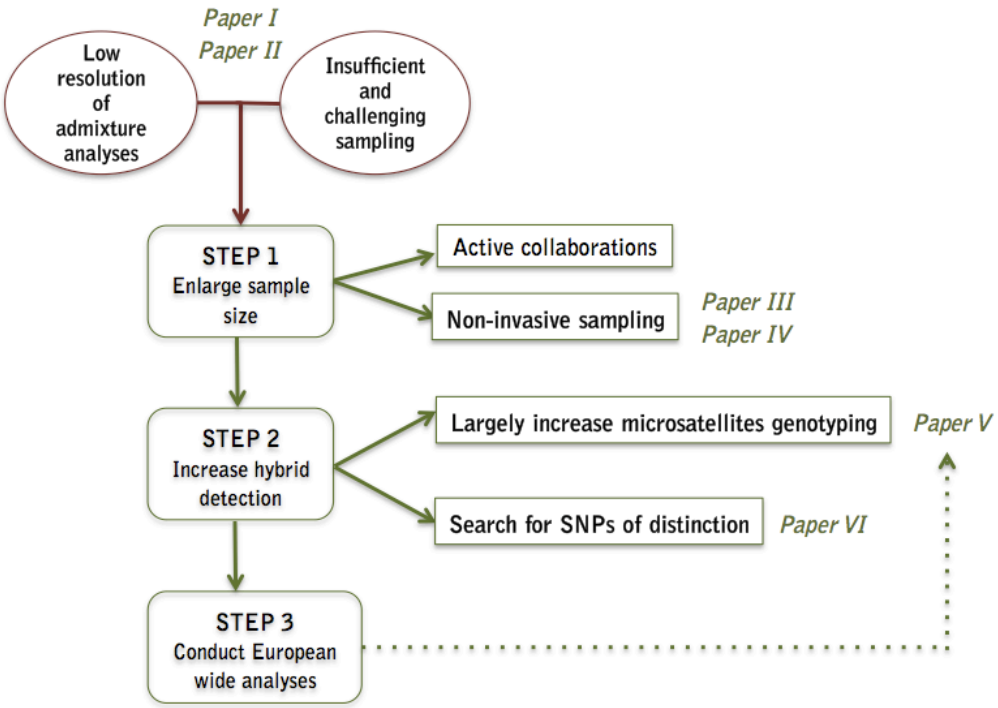


Figure 1. Simplified workflow of this thesis.

We largely focused our research in the study of introgressive hybridization because we believe that it is one of the most significant influence faced by wildcat populations nowadays. Throughout all Europe, domestic populations outnumber the wild ones by several orders of magnitude (Driscoll and Nowell 2009). This reality clearly increases the chances of contact and crossbreeding, especially for natural populations that suffered severe demographic decreases in the past (or that are still suffering), and which are highly fragmented. The outcomes of hybridization with domestic relatives might be numerous and more or less severe, but it is clear that the spread of domestic genes into natural populations may lead to progressive losses of locally adapted alleles and to the disruption of co-adapted gene complexes (Rhymer and Simberloff 1996; Allendorf and Luikart 2007). Hybridized animals may not be as capable and fit as the naturally evolved local-specific wild ancestors, compromising the long-term survival of the threatened species and promoting further population declines (Rhymer and Simberloff, 1996; Allendorf *et al.* 2001; Randi, 2008). Moreover, excessive introgression of domestic genes may further change the functional nature of native populations in the community, for example, by altering interactions with co-evolved prey populations or competitors (Ryan *et al.* 2009). Thus hybridization may have strong implications both in the biology of wildcat native populations and in the policies related to their protection. But if a population has not become a hybrid swarm and still contains a reasonable number of parental genotypes it could be potentially recovered, a concept that motivated the detailed molecular analysis of European wildcat populations.

Given that the results produced in this thesis have been extensively reported and discussed in the manuscripts of Chapters 2, 3 and 4, in this section the main findings will be summarily highlighted and integrated. Major focus will be given to future research that seems to be demanding in wildcats' conservation genetics.

5.1 Hybridization in Iberian Wildcats

The first strong evidence of crossbreeding between wild and domestic cats in Portugal where obtained in 2003, where one hybrid individual as been detected over 13 putative wildcat samples by Pierpaoli *et al.* (2003). The identification of one admixed cat in such a reduced sampling opened the suspicion that hybridization in Portugal (and possibly Spain) could be an important threat to the species. On top of this putative problem, the Iberian wildcat populations are known to be fragmented, are considered regionally threatened (wildcats are listed as vulnerable - VU - in Portugal and near threat – NT - in Spain; Cabral *et al.* 2005) and due to the critical situation of the Iberian lynx situation, the wildcat might play an additional importance in the community of the Iberian mesocarnivores. To further understand the impact of hybridization in the Iberian Peninsula, we initially analysed a total of 181 samples (72 putative wildcats and 109 domestic cats) using Multivariate and Bayesian Clustering methods (*Paper I and II*). We this first approach we have shown that Iberian wild and domestic cats belong to two clearly divergent parental clusters with high genetic diversity. Hence, both populations seem to be scarcely hybridized, and introgressive hybridization apparently did not act as the major factor shaping the genetic pool of the Iberian wildcats. At the same time, possible strong demographic declines in the past seems to not result in important population

bottleneck events, that could left evident traces in the genetic markers analysed in this first study. Nevertheless, bayesian admixture analyses of empirical and simulated datasets using multilocus genotypes, revealed that at least 6.9% (5/72) of the Iberian wildcats probably have hybrid ancestry. These admixed individuals probably represent diverse levels of hybridization, suggesting that crossbreeding exists and should be regarded as a real threat to the wild population. The levels of hybridization detected in these two first studies, by means of 12 microsatellite genotyping, can be compared with other European studies for which levels of introgression have been inferred using between 8 and 27 STRs (Table 1).

Table 1. Comparison between hybridization levels documented in different European populations. Levels of genetic differentiation (F_{ST}) and hybridization rates (HYB %) are reported for microsatellites (STRs) data.

Population	**N**total	**mtDNA**	**STRs**	**F_{ST}**	**HYB %**	**Reference**
Iberian Peninsula	184	No	12	0,20	6.9%	Oliveira *et al.* 2008a,b
Scotland	304	No	9	0,12	hybrid swarm	Beaumont *et al.* 2001
Italy	128	Yes	12	0,13	2%	Randi *et al.* 2001
Italy & Hungary	182	No	27	0,14	8 & 25-31%	Lecis *et al.* 2006
EU countries (9)	336	No	12	0,11	Variable	Pierpaoli *et al.* 2003
NE France	209	No	13	0,16	23.8%	O'Brien *et al.* 2009
Germany	149	Yes	11	0,12	18.4%	Hertwig *et al.* 2009
Germany	244	Yes	8	0,19	3%	Eckert *et al.* 2010

Estimates of crossbreeding in the Iberian Peninsula could be approximately equalled to the low estimates found in Italy (Randi *et al.* 2001), and in one study in Germany (Eckert *et al.* 2010; but see Hertwig *et al.* 2009), while they contrast with the higher values attributed to North-eastern France (O'Brien *et al.* 2009), Western Germany (Hertwig *et al.* 2009), Hungary (Lecis *et al.* 2009) and Scotland (Beaumont *et al.* 2001). These estimates corresponded, though, to approximate scenarios. Although the use of hypervariable microsatellite markers, combined with Bayesian-based assignment tests, has radically improved the analytical power of admixture analyses in many hybridizing groups (e.g. Quintela *et al.* 2010), difficulties in the precise identification of hybrids persisted among the referred studies. According to our and Ecket *et al.* (2009) simulations on the detection of F1, F2 and backcrosses, and following the indications of published studies on simulated genotypes (Vähä and Primmer 2006), we must regard the number of detected hybrids as the minimum observed value. Especially beyond the second generation of hybridization, some individuals classified as "pure" could actually result from repeated backcrosses of admixed cats with parental individuals, and the small number of analysed loci did not allow the precise identification of hybrids. According to Vähä and Primmer (2006), as much as 48 loci are needed to separate backcrosses from purebred parental individuals, even when F_{ST} values between parental populations are as high as 0.21. Godinho *et al.* (2011) have shown that 42 autosomal microsatellites displaying F_{ST}=0.16 may also provide very confortable conclusions, achieving assignment posterior probabilities above 90% for a single hybrid class in 6 out of 8 Iberian wolf x dog hybrids. Although, the number of loci used by Godinho *et al.* 2011 clearly exceed the one

used in the analyses performed in *Papers I* and *II* (n=12), these preliminary results represented the first Iberian-scale study of European wildcats and provided a solid and scientific background to proceed with further molecular studies on the wildcat in the Iberian Peninsula.

The situation of wildcats in Portugal seems to be particularly problematic and to contrast the first evidences of expansion of the species in central Europe (e.g. Raimer 2006; Simon 2006 in Hertwig *et al.* 2009; O'Brien *et al.* 2009). Evidences of local disappearance have been recently reported. Sarmento *et al.* (2009) have shown that of regions where wildcats used to be widespread and no domestic cats were present (as the Malcata Natural Park, between 1998 and 2001) are now mainly occupied by typical domestic phenotype individuals (between 2005 and 2007). Additionally, recent efforts to capture and radio-track wild individuals in the South of Portugal have proven that the species might be disappearing (P. Monterroso, person. comm.) from areas where they have been formerly camera-trapped and captured (as Guadiana Valley, Monterroso *et al.* 2009). These evidences need, however, further validation, especially because the secretive behaviour of wildcats, along with low population densities, make data collection complicated (Monterroso *et al.* 2009).

Under the framework of the potential impact of hybridization and the risk of local disappearance, we considered that applying non-invasive sampling procedures would be essential for the long-term monitoring of this endangered and elusive feline in the Iberian Peninsula. Therefore, the second stage of our research was concentrated in the search of new non-invasive genetic methodologies, which should substantially increase the power to detect wildcats (or their hybrids).

5.2 Non-invasive genetic approaches: first outcomes in Iberian Studies

The detailed scrutiny of the literature on non-invasive genetics performed in *Paper III* allowed us to perform a broad-scale review of the available methods, which may yield good enough DNA and low enough genotyping error rates, used to address nearly all questions that can be addressed by traditional high-quality samples (e.g. blood and tissue samples). In a near future, the field of non-invasive genetics will be prevalently composed by new genomic approaches, namely through large-scale PCR multiplexing techniques, chip arrays and massive parallel sequencing technologies. However, these advances are now being applied for the first time in invasive cat's material, and we are still taking the first steps into the possibility of largely applying such techniques to wildcat non-invasive samples. It was crucial to design routine laboratory analyses for non-invasive cat samples, as this could immediately provide great advances to the current knowledge of the species. Since faecal and hair material have been considered two of the most efficient survey tools to study the distribution, abundance, diet, reproductive status and behavioural features (among others) of mesocarnivores, we focused our work in the development of new practical, fast and low-cost molecular methods to perform scat and hair analyses (*Paper IV*).

The identification of Iberian carnivore species through the analysis of polymorphism at a small fragment of the IRBP gene developed in *Paper IV* provided a valuable tool to identify our target species and

has been already applied in a number of recent studies (Piñeiro and Barja, 2012; Piñeiro *et al.* 2012; Monterroso *et al*. 2012). The identification of scat samples from presumptive Iberian wildcats has been performed by this method to allow the subsequent study of the scent mark deposing places preferred by European wildcats in the North-western region of Spain (Piñeiro and Barja 2012). Scat samples were first species-identified (as described in *Paper IV*), and further individually identified using 12 microsatellite loci (as described in *Paper I* and *II*). This procedure assured the accurate analysis of wildcat samples by both preventing any misclassification of faeces (IRBP polymorphism) and any samples' duplication (individual genotyping). From 41 analysed samples, 26 (63.4%) were identified as wildcat samples and none of them was identified as belonging to other carnivore species (including feral domestic cats and hybrids). The 26 samples were attributed to 16 different individuals, a crucial *a priori* information for the exact and precise evaluation of wildcats' behavioural preferences (Piñeiro and Barja, 2012). The same 26 genetically identified samples were further applied in wildcat physiological studies (Piñero *et al*. 2012). These authors found that increased physiological stress levels in wildcats were both a response to the level of tourism within different zones in the Natural Park Montes do Invernadeiro (Spain) and a response to the seasonal reproductive state of wildcats (Piñeiro *et al.* 2012). Furthermore, the species identification method designed on *Paper IV* and the species assignment technique described in Palomares *et al.* (2002), were simultaneously applied by Monterroso *et al.* (2012) to: i) identify variable accuracy rates in the species identification of carnivore scat samples, and ii) to predict the major factors affecting that accuracy. Accordingly to this study, putative wildcat faeces displayed much higher species identification error rates (88%) than the ones from red foxes (14%) and stone marten (22%), highlighting the necessity to base wildcat studies in *a priori* molecular identifications. Finally, our approach also inspired the use of IRBP variability to discriminate Iberian rodent species through non-invasive sampling by Barbosa *et al.* (2012).

The methodology we have developed provides major advances in studies of European wildcats and other Iberian carnivores. It may contribute to the genetic analysis of higher number of individuals from different locations and encourage the regular census and monitoring of populations over time. But, to better understand the impact of hybridization and reduce rates of this phenomenon in nature, other management actions were still demanding at this point. Hybridization studies still lacked of full accuracy and we still had to focus our research in the improvement of the efficiency of molecular tools' to assess populations' genetic structure and introgressive hybridization rates.

5.3. Improving the molecular tools in wildcat studies

5.3.1. Increasing the number of STRs in wildcat studies: an European approach to detect population structure and hybridization

To contribute to the accuracy power of admixture analyses we first started with the range-wide analyses of genetic variability and admixture proportions of 1128 cat samples genotyped at 38 unlinked microsatellites (Chapter 4, *Paper V*). This work provide the most complete molecular scenario obtained so far for the wildcats in Europe, both for what regards the sampling scheme and the number of analysed loci. Outputs of this work might be particularly important considering that independent studies have recently focused on analogous questions, yet with contrasting results. German studies are most illustrative of this subjectivity, since opposing scenarios were described for what concerns both wildcat's demographic changes and levels of interbreeding. First, while 18.4% of admixture was detected across the whole country by Hertwig *et al.* 2009 (especially in South-Western Germany), very low levels were reported by Eckert *et al.* 2010 (3%) and no hybrids were identified by Pierpaoli *et al.* 2003. Results from our study confirm the presence of hybrids in South-western Germany, with 8.47% (15 out of 177) of admixed genotypes found among the sampled population. This variability among studies remarkably highlight the need to carefully evaluate all estimations of hybridization, and suggests that we should pass from a static idea of absolute values to a more dynamic notion that current estimations might be strongly dependent on sampling, molecular and statistical strategies. Second, bottleneck events or reduced genetic diversity caused by genetic drift or inbreeding were suggested by Pierpaoli *et al.* 2003 and Eckert *et al.* 2010, while no evidences of such events were found by Hertwig and colleagues (Hertwig *et al.* 2009). Although no recent bottlenecks could be detected in the entire European sample (as measured by BOTTLENECK, Cornuet and Luikart, 1997), our analysis of 49 samples from Eastern Germany corroborates evidences of a past genetic bottleneck, as identified by M-ratio test analyses. In this area, the prevalence of just few scattered habitat spots populated by few individuals might have acted as a barrier to individuals' dispersal and, thus, limited gene flow and promoted allelic diversity loss due to genetic drift (Pierpaoli *et al.* 2003; Eckert *et al.* 2010). Overall, it is evident that Eastern German deserves special attention in the European context, due to the geographical and genetic isolation of the population. Globally across Europe we reported high levels of genetic diversity, both within and among wildcat populations, which might indicate that the demographic decline did not produce a noticeable reduction of genetic variation at the analyzed STRs. It is however crucial to bear in mind that population bottlenecks might generate significant detectable signals only in extreme cases when effective population sizes decreases rapidly to tens rather than to hundreds of individuals (Luikart and Cornuet 1998). As a result, many studies have failed to detect bottleneck footprints from genetic data even when demographic data indicate that the population has gone through size collapse (e.g. Busch *et al.* 2007; Mardulyn *et al.* 2008). Therefore, we should not look at our results as confident evidences that no genetic bottlenecks occurred among other European countries, but as strong indications that populations seem to have, today, no significant signs of genetic depletion. Nevertheless, in our opinion the significant population structure found

across the entire European distribution range suggests that present subpopulations might represent evolutionary significant units (ESU's) worth of specific oriented conservation, since accelerated genetic drift could have taken place in relict populations and contributed to a significant genetic partition of wildcats in at least 10 subpopulations.

Admixture analysis confirmed that Scottish and Hungarian populations are most probably composed of a hybrid swarm, with a variety of introgression degrees and intermediate phenotypes difficult to identify. This might result from long lasting hybridization that seems not to have occurred in any of the other European regions (Beaumant *et al.* 2001; Pierpaoli *et al.* 2003; Lecis *et al.* 2006). The non-hybridizing areas are, nevertheless, influenced by sporadic hybridization and as more samples are analysed as more admixed genotypes are found. As discussed in *Paper V*, the fact that sampling strategies are not specifically design for admixture studies might strongly influence the inferences made so far, and low sampling sizes in many areas might not allow the objective estimate of admixture rates. Therefore, although this study has provided important contributions to the knowledge of the *Felis silvestris* complex inhabiting Europe, future hybridization analyses might continuously improve our understanding of the species.

According to the parallel examination of real and simulated genotypes, the analysis of 38 microsatellites across the entire Europe (*Paper V*) seems to considerably increase our capacity to discriminate between hybrid genotypes and parental individuals when compared to the analysis performed for 12 loci among Iberian samples (*Paper II*). First, the probabilistic threshold value to assign each individual to one single class could be straighten from $qi>0.80$ to $qi>0.85$, reflecting that individual membership assignments were higher for parental genotypes. Second, although 9% of F2 hybrids could be missidentified as parental individuals in Paper *II*, none of the F2 genotypes where wrongly classified as wild or domestic samples in *Paper V*. On the other hand, no significant improvements were obtained for what regards the identification of backcrossed genotypes, with as much as 14% backcrosses with domestic cats and 12% backcrosses with wildcats still being classified as parental cats using 38 loci-based analyses in STRUCTURE. Results obtained with NEWHYBRIDS revealed a slightly better assignment of past-generation hybrids, with a maximum of 4% backcrosses with domestic cats being wrongly identified as domestic samples (*Paper V*). These findings are in agreement with the previously nominated simulations performed by Vähä and Primmer (2006), since our numbers of loci are still beneath the ideal. Moreover, they confirm a marginal better performance of NEWHYBRIDS for detecting backcrossed individuals when a considerably high number of loci is used (Vähä and Primmer, 2006).

Although a great methodological effort has been made in this study to triplicate (n=38) the number of loci previously used to analyse hybridization rates at European levels (n=12, Pierpaoli *et al.* 2003), it was still evident that achieving accurate inferences of admixture proportions require further advances in molecular tools. While one possibility could be increasing again the number of unlinked microsatellites to approximately 50 loci or/and to add a subset of linked STRs to investigate their capacity to better discriminate old backcrosses (Lecis *et al.* 2006), it became evident that an alternative methodological choice could be more advantageous in our case: the analysis of SNP loci. As thoroughly discussed in the *General Introduction*, SNPs provide a number of advantages relatively to microsatellites for what regards statistical and laboratory procedures and their potential use in conservation genetics is today widely recognized.

However, methods for discovering large numbers of SNPs typically rely on relatively expensive research projects, which has limited their development in many wildlife *taxa*. But since wildcats are among the privileged species that may benefit from the advanced research performed for their domestic relatives, we could successfully joint one of the first SNP genotyping arrays performed to analyze worldwide domestic cat populations. The results from our study represent a first step towards developing rapid, inexpensive and reliable SNP-based assays to assess introgression in the wild.

5.3.2. In the search of new loci of distinctiveness between European wild and domestic cats: the use of SNPs in cat's conservation genetics

Ideally, the identification of hybrids and their categories (F1, F2 and first backcrosses) could be achieved with a minimum number of loci if they disclose significantly high levels of differentiation between the studied groups (Vähä and Primmer, 2006). But the remarkable resemblance between European wild and domestic cats, and the intricate history of sympatry and introgression that most probably influenced both the domestication (Driscoll *et al.* 2007) and the expansion of domestic populations worldwide, might have created one of the most complicated frameworks to genetically discriminate parental groups of wild and domestic relatives. Even so, when two populations are subjected to different selective pressures, which has certainly been the case of wild and domestic cats, some level of natural and artificial selections are expected to cause divergence in different parts of their genome. Therefore, with the ultimate goal of finding highly differentiated (and methodologically and economically advantageous) molecular tools to identify hybridization between European wild and domestic cats, we assessed the genetic variability of 158 SNPs in 139 domestic, 130 putative wild and 5 captive-bred hybrid cats, comprising: a) mutations encompassing known phenotypic characteristics predictably divergent between wild and domestic cats and b) randomly genome-sparse SNP variation across the genome. With this work we aimed at addressing, for the first time, important questions regarding the successful application of SNPs in hybridization studies of European wildcat's, more precisely: i) are there SNPs of distinctiveness between domestic and wild forms and, if not, are there loci showing significant differences in allele frequencies that will help determining hybridization events?; ii) how many SNPs are needed for an accurate inference of individuals assignment to the wild or domestic population and, furthermore, for the determination of hybrids' admixture ancestry?

Although no diagnostic fixed differences were detected, our findings revealed high genetic differentiation between wild and domestic cats, with 35 SNPs showing values of genetic divergence (F_{ST}) between 0.515 and 0.891. Although random-bred cats might not represent entities that suffered strong artificial selective pressures, past episodes of positive, balancing and/or purifying selection during cat's domestication appear in fact to have created distinct signatures in the genome that tear them apart from *Felis silvestris silvestris*. Divergent distribution of the two possible alleles and low levels of heterozygosity characterized the most informative SNP loci, but understanding if it results from disruptive selection during the domestication process deserves further analysis. While the random-bred domestic cats used in this study

should represent most of the genetic variability that can be found among the domestic cat gene pool (since they are not constrained by breed standards), European wildcats do not readily represent the ancestral wild population from which domestic cats have been domesticated and results from this work do not allow to make considerations about loci selected during the domestication process itself. Instead, domestication is used as the starting mechanism that made us predict that significant differentiation between wild and domestic cats should exist across their genomes and that genotyping schemes of such variation could be applied in conservation studies of the endangered European wildcat. Understanding the role of these mutations in cat's domestication imply their detailed survey in specimens of *Felis silvestris lybica*, which is the ancestor of all domestic cats (Driscoll *et al.* 2007).

Having information about several independent loci across the genome, we could have tried to statistical identify SNPs displaying atypical patterns of non-neutral diversity. Using SNPs data, a number of statistical tests of neutrality are today commonly used to detect such outlier loci (e.g. Kauer *et al.* 2003; Beaumont and Nichols 1996; Foll and Gaggiotti 2008; Excoffier *et al.* 2009). In the search of molecular tools displaying genomic signals of selection between wild and domestic cats we must, however, be very critical about the methods to apply and inferences to be made. In our sample, wild and domestic cats represent hierarchically substructured populations, that suffered significant demographic changes and that share a common recent genetic evolution. Moreover, they represent just small subsets of the entire population of both subspecies in Europe and identifying genetic signs of selection among these samples could be misleading. Finally, although we are confident that the number of loci used in this study is remarkably valuable for the analyses of admixture, it is just an extremely low proportion of variable positions in cat's genome and a forthcoming broader coverage of SNPs variability in several wild and domestic cat samples (on-going research) will surely provide better estimates of selection at specific loci or genomic regions. For the above-mentioned reasons, we decided not to apply selection methods in our data.

5.3.3. SNPs *versus* STRs or SNPs + STRs?

The high density of SNPs in the genome and the increasing feasibility of high-throughput genotyping technologies have been catalysing the shift in marker preference towards SNPs. However, SNPs and STRs markedly differ in mutation rate and mechanism, and the simultaneous consideration of polymorphism at both genetic markers may provide insights that are difficult to obtain from the analysis of one of the markers alone. SNPs and STRs may, thus, convey complementary information and "anchor" the large rate of polymorphism at STRs with the low-homoplasy at SNPs. In fact, one of the questions that prevail from the SNPs research developed in *Paper VI* is if these loci may outperform microsatellites in admixture analyses or if both markers should be combined for finest estimations. In the following paragraphs we provide a preliminary discussion on the performance of both markers in hybrids' detection, by comparing results from *Paper VI* with the estimates obtained using a set of 39 STRs (38 used in *Paper V*; Supplementary Table 1). We have analysed the same 274 cat individual genotypes following the methods described in *Paper VI*,

estimated their information content (INFOCALC), and determined their ability to assign cat samples to parental groups and hybrid classes (using STRUCTURE and NEWHYBRIDS algorithms).

5.3.3.1. The informativeness of both markers for admixture analysis

Rankings of informativeness determined as explained in *Paper VI*, proved that SNPs generally provide greater power than STRs for discriminating between the wild and domestic groups (Figure 2, Supplementary Table 1). Nevertheless, two STRs outperformed SNP loci, the autosomal FCA 224 (with F_{CT} =0.454; $H_{E(FCA)}$=0.594 and $H_{E(FSI)}$ = 0.447) and the X-located FCA651 (with F_{CT}=0.333; $H_{E(FCA)}$=0.570 and $H_{E(FSI)}$ = 0.767).

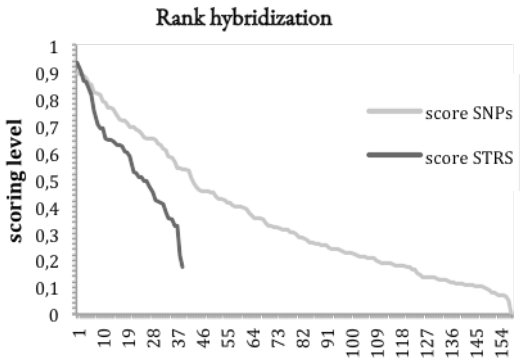


Figure 2. Scoring levels (averaged across informativeness values) for discriminating between wild and domestic groups.

The ranking lists of the two marker types, either alone or together, does not precisely reflect the single locus estimates of F_{CT} values between European wild and domestic cats. This nonconformity should be a consequence of the inherent properties of informativeness scores, which not only take in consideration the divergence between populations (F_{CT} values) but also the diversity values that characterize both groups (namely *Ar* and H_E). Therefore, a STR locus with high *Ar* in one population but low in the other (e.g. FCA224; F_{CT}=0.507) outperforms a SNP locus with higher F_{CT} values but with less opportunity to segregate allele frequencies due to the biallelic nature of the markers (e.g. ChrB4_pos255106; F_{CT}=0.877). In a simpler way, signals of differentiation between wild and domestic cats are stronger when one population shows one or few alleles while the other exhibits many (STR loci), in opposition to when each population shows different frequencies of only two possible alleles (SNPs loci). Basely it follows the rule "one against many is better than one against another". Lower F_{ST} and higher informativeness has been similarly documented for STRs rather than SNPs in human genome-wide comparisons (Payseur and Jing, 2009). But why are F_{CT} values globally lower for STRs? One of the reasons might be their inherent homoplasy, which is responsible for the deflating of F_{ST} values as mutations in different populations are identical by state but not by descent (Queney *et al.* 2001). Moreover, when mutation rates are high, the multi-allele nature of STRs might originate values of

within-population and total heterozygosity that approach one, constraining the maximum possible value of F_{ST} to be small, even when populations share few alleles. Recent studies have proposed a series of new estimators of genetic diversity using microsatellites (Jost 2008; Heller and Siegismund, 2009; Meirmans and Hedrick 2011), however, the advantages and drawbacks of each estimator are still under discussion. Since F_{ST} estimations should be the most appropriate for biallelic markers (Meirmans and Hedrick 2011) and SNPs were the loci of major interest in our study, we decided to analyse our data according to F_{ST}.

5.3.3.2. Confidence in SNPs and STR-based assignments using Bayesian models

Bayesian analyses of the simulated genotypes in *Paper VI* revealed that all parental, F1, F2 and backcrossed individuals could be correctly identified by the STRUCTURE's algorithm using the new set of 158 SNPs (*Paper VI* and Table 3a). Using a stringent threshold value of $qi>0.85$ to assign individuals to a single class, assignment values for NEWHYBRIDS proved to be equally accurate for parental and first-generation hybrids, although 4% F2, 3% BxFCA and 1% BxFSI were allocated to their own hybrid category with $qi<0.85$. Nevertheless, none of the referred cats were significantly assigned to one of the other remaining hybrid categories, preventing any case of misclassification also with the NEWHYBRIDS' algorithm. As a result, true backcross genotypes obtained for the analysed SNPs should never be misclassified as belonging to a different class, but, in the worst-case scenario, a very few of them might remain as past-generation hybrids of old but uncertain ancestry. The set of 39 STRs proved also to be highly effective in STRUCTURE's clustering analyses (Table 3b), although several of the results provide solid evidences that the new SNP loci might outperform admixture inferences done with this set of STRs. First, *Qi* threshold values for the assignment to a single cluster/class where lower for the analyses performed only using STR-genotypes ($0,20>qi<0,80$), when compared to SNP-based allocations ($0,15<qi>0,85$). Second, not all simulated genotypes where correctly identified since 21% of BxFCA and 36% of BxFSI showed significant posterior probabilities of assignment to the respective parental clusters. Third, NEWHYBRIDS' membership proportions were less clear than the values obtained for SNPs-simulated genotypes, with 1% FSI, 6% F1, 29% F2, 9% BxFCA and 8% BxFSI remaining unclassified due to their high partition in more than one category (Table 2). Moreover, a few misclassifications might occur for 9% F2 and 1% BxFSI cats, since they might be allocated to other hybrid categories. Nevertheless, STRs seem not to neglect true levels of admixture for this set of samples and loci, since none of the misclassified hybrids could be wrongly identified as "pure" domestic or European wildcat.

The results now reported for the STRs are somewhat different from the ones obtained in the European-wide study (*Paper V*), where only a maximum of 14% backcrosses in STRUCTURE and 4% in NEWHYBRIDS could be wrongly identified as parental individuals. The observed asymmetry is probably related with the highly different representation of parentals' genetic diversity between studies. In fact, the representation of parental genotypes is considered to influence the power of Bayesian algorithms, especially for particularly unstable individuals' allocations (Falush *et al.* 2003) as can be the case of backcrossed individuals. Analyses performed in *Paper V* are based on genotypes from 294 domestic cats and 610 wildcats divided in 3 macroareas, meaning that the initial use of true parental individuals to create simulated

genotypes has been based in a much more precise sampling from each region than in *Paper VI* (for which true parental genotypes were available for 82 wild and 139 domestic cats from the entire Europe analysed together). These findings again reflect the case-specific nature of all presented estimates, and alert to the importance of looking at these results as relative rather than absolute. For this reason, the results from the comparison between SNPs and STRs should be seen as guidelines for other cases, even within the same species and populations. It is possible that also the capacity of SNPs for identifying hybrid cats in larger and partitioned sample sets will suffer slight differences, and simulations to estimate the assignment power within that specific dataset should be made.

Table 2. Average membership proportion (Q) of simulated genotypes in the Bayesian analysis performed using STRUCTURE and NEWHYBRIDS. Results represent *qi* values averaged over 10 independent runs. Minimum and maximum values of credibility intervals obtained in STRUCTURE are shown between brackets. Analyses correspond to three different datasets: a) all SNPs; b) all STRs; c) combining SNPs and STRs. The percentage of misclassified cats (WRONG), and of individuals for which CI ranges fallen outside the expected values of *qi* (as shown in the first column) are in parenthesis. NEWHYBRIDS' average *qi* assignment to the correct category and percentages of unclassified simulated genotypes are shown for the same three analyses.

	a) 158 SNPs		**b) 39 STRs**		**c) 197 SNPs & STRs**		NEWHYBRIDS		
SIMULATED CATEGORIES	**FCA**	**FSI**	**FCA**	**FSI**	**FCA**	**FSI**	158 SNPs	39 STRs	SNPs & STRs
FCA q_I>0,85*	0,959 (0,774-1,000)	0,040 (0,00-0,226)	0,968 (0,783-1,000)	0,032 (0,000-0,217)	0,978 (0,837-1,000)	0,022 (0,000-0,163)	0,9996	0,9985	1,0000
WRONG	0 (4%)		0 (1%)		0(1%)		0	0	0
FSI q_{II}<0,15	0,025 (0,00-0,108)	0,975 (0,892-1,000)	0,028 (0,000-0,251)	0,972 (0,749-1,000)	0,013 (0,000-0,099)	0,987 (0,901-1,000)	0,9938	0,9935	1,0000
WRONG	0 (0%)		0 (2%)		0(0%)		0	1%	0
F_1 0,4<*qi*>0,6	0,499 (0,300-0,695)	0,501 (0,305-0,700)	0,481 (0,244-0,771)	0,519 (0,229-0,756)	0,495 (0,327-0,670)	0,505 (0,330-0,673)	0,9615	0,9552	1,0000
WRONG	0 (0%)		0 (6%)		0(0%)		0	6%	0
F_2 0,4< *qi* >0,6	0,494 (0,280-0,761)	0,506 (0,239-0,720)	0,470 (0,167-0,803)	0,530 (0,197-0,833)	0,484 (0,317-0,667)	0,516 (0,333-0,683)	0,9050	0,7856	1,0000
WRONG	0 (1%)		0 (15%)		0 (0%)		4%	29%	0
B X FCA 0,15< *qi* >0,85	0,731 (0,509-0,924)	0,269 (0,100-0,491)	0,751 (0,405-0,997)	0,249 (0,003-0,543)	0,747 (0,570-0,923)	0,253 (0,077-0,430)	0,9455	0,9374	0,9990
WRONG	0 (28%)		21% (86%)		0 (22%)		3%	9%	0
B X FSI 0,15< *qi* >0,85	0,270 (0,129-0,469)	0,730 (0,531-0,900)	0,224 (0,004-0,424)	0,776 (0,576-0,996)	0,255 (0,130-0,420)	0,745 (0,580-0,870)	0,9455	0,9382	0,9998
WRONG	0 (14%)		36% (93%)		0 (8%)		1%	8%	0

* *qi* values for STRs datasets were lowered to 0,80 based on the correct allocation of all known domestic cats

Although it is clear that the relative performance of SNPs and STRs favours the first, the best results can be undoubtedly achieved for the combination of both markers. When using both STRs and SNPs marker (total of 197 loci), all admixed simulated cats were correctly defined both using STRUCTURE and NEWHYBRIDS (Table 3c). According to these findings, we suggest that the better admixture analyses will be achieved in the future by combining the most informative set of SNPs and STRs. One promising strategy

would be also to identify microatellites linked to the most informative SNPs, or vice-versa, and create a database of compound genetic markers commonly known as SNPSTRs. This combination of co-inherited markers evolving at different rates may offer the possibility of gaining better-resolved insights into past demographic and admixture events, and might provide better basis to detect the effects of selection (Mountain *et al.* 2002).

5.3.3.3. Selection of a smaller set of markers for routine diagnosis

Since this work was largely motivated by the need to routinely identify admixture events in conservation studies of European wildcat populations, we have also investigated the efficiency of as few as 35 SNPS to identify hybrid cats (*Paper VI*). The selection of highly diagnostic SNPs from large panels has the ability to provide confident assignments (and, therefore, admixture inferences), as has been demonstrated in human research of geographic population structure and ancestry (Lao *et al.* 2006), fish self-assignment accuracy Glover *et al.* (2010) or in the present study. Comparatively to the total set of markers, the smaller panel of 35 SNPs still provided exceptionally valuable results for such a low number of loci, with only 8% backcrosses with domestic cats and 4% of backcrosses with wildcats being wrongly classified as parental individuals. Nevertheless, to obtain the most unbiased view of introgression dynamics we should still base our analyses in a large number of molecular tools, preferably the maximum 197 loci analysed in this study. The importance of using a considerable number of loci to accurately detect admixed populations has also been outlined in the analysis of chicken breeds, where a minimum number of 100-250 SNPs proved to be indispensible to detect their hybrid origin (Gärke *et al.* 2012). Moreover, the choice of highly differentiated traits/loci from just a small panel of individuals and genomes has been referred as a possible reason to overlook differential population structure in non-diagnostic/less differentiating traits (Brumfield *et al.* 2003; Schlotterer 2004; Morin *et al.* 2008), and to distort our perception of the actual levels of introgressive hybridization in nature (Yuri *et al.* 2009). Concluding, although we may advise wildcats monitoring works to initiate with the genotyping at 35 SNPs especially when labour and costs must be reduced, we recommend the analyses of more loci whenever possible.

5.4. Future perspectives and lines of research

Over the past ten years, the rapid technological developments in high-throughput genotyping and sequencing have heralded a true era of population genomics. The high-resolution datasets that suddenly became available accelerated the development of analytical approaches for exploring questions surrounding signatures of population history (e.g. linkage disequilibrium, population structure, admixture). But while high-density SNP maps have been already used for many animal species, including human, mouse or dog, the domestic cat genome was for long less explored. However, a feline SNP chip (Illumina 63K Infinium feline iSelect DNA

array) has been already released in 2011 and a number of array-base studies are now in progress. Ongoing wildcat research will take full advantage of this knowledge, especially because this new array includes approximately 4000 SNPs that were identified in the wildcat (L. Lyons, person. comm.). The possibility for using large numbers of SNPs may enable the detection of nuclear genotypes that may be associated with introgression and/or phenotypic selection that occurred during the domestication process. Those SNPs would definitely better differentiate domestic and wild cats.

Conceptually related techniques could, then, be successfully applied in conservation genomics of wildcats, such as selective sweep and linkage disequilibrium (LD) mapping approaches (Kohn *et al.* 2006). Selective sweep mapping has proven extremely powerful in the identification of genes underlying specific phenotypes in dogs (Pollinger *et al.* 2005; Akey *et al.* 2010). In addition, the selective sweeps have been identified in many domestic and wild populations suggesting a general approach for finding genes under intense selection during domestication (Schloetterer and Harr 2002; Luikart *et al.* 2003; Akey *et al.* 2004; Storz 2005; Wright *et al.* 2010). Under this perspective, large genomic scans can potentially be used to identify several regions containing genes that have a major influence on domestic cats' diversifying selection, improving our capacity to discriminate between them and wild relatives. This would be especially true for regions under selection shared among different breeds and that tear them apart from the wild populations. The great advantage of this approach over traditional candidate gene surveys is that it can be performed using molecular markers alone, without the prior knowledge about traits under selection (Shikano *et al.* 2010; Shimada *et al.* 2010). Therefore, it is especially suitable for organisms such as the wildcat, for which phenotypic, physiologic and behavioural similarity with feral domestic cats hinder the obvious targeting of selected loci.

A complementary approach would be LD mapping, which utilizes the principle that genome regions under selection will have marker loci that are inherited in a non-random fashion, that is, they are in LD. In animals, some of the most interesting examples were reported in dogs for studying breed-specific phenotypes, including hereditary pathologies (e.g. Ostrander and Wayne 2005; Goldstein *et al.* 2006). The first LD mapping in the cat identified causative mutations in colour patterns at the TYRP1 locus (Shmidt-Kuntzel *et al.* 2005), and a new genomic location for silver or hypopigmentation in cats has been recently mapped (Menotti-Raymond *et al.* 2009). Because a change of the organization of the genetic diversity across the genome is typically found in domesticated species when compared to wild relatives, similar LD strategies could account for variation inside wildcat's genome, and portions of the genome rather that single locus variants could then be applied in differentiation and introgressive hybridization studies. For example, linkage disequilibrium in natural populations of wild canids (Gray *et al.* 2009), mice (Laurie *et al.* 2007) and rabbits (Carneiro *et al.* 2011) proved to decay faster when compared to their domestic populations. European wildcats are not the known ancestral of domestic cats, but in turn they are the historical wild progenitors of *Felis silvestris lybica*, the wild subspecies domesticated in the Fertile Crescent (Driscoll *et al.* 2007). Therefore, even higher levels of genomic differentiation can be expected and higher discrimination between wild and domestic forms might be achieved for admixture inferences.

It is interesting to realize that some of the SNP loci analysed in our study display complete monomorphy among European wildcats, while significantly vary among domesticates (e.g. mutations at

MYRIP, FNIP2 or CMAH gene). Nevertheless, some of the analysed SNPs show that some regions of the wildcat genome maintained high levels of variability. But while some of those vary considerably also among random-bred cats (e.g. mutation at GPR113), others indicate a very small variation among the domestic cats (e.g. CCR2). A better understanding of these and many other mutations could be achieved in the future, by combining a wider knowledge of genetic variation in different populations with gene expression analyses. Presumably, gene expression differences can be neutral, detrimental or adaptive, but if adaptive, they might identify ecologically unique populations. A shift towards the study of functional genetic variation may thus improve our understanding of the processes affecting genetic variation in this rare and endangered species (Kohn *et al.* 2006).

To date, molecular research on wildcats' hybridization (including the work performed throughout this thesis) focused in distinguishing admixed individuals and assess hybridization levels, but no efforts have been made to specifically define how introgressive hybridization is truly shaping wild populations. The genetic basis of hybridization is surprisingly complex: hybridization is not synonymous of introgression and many questions remain on its effective impact. Nowadays, not only little is known on the specific genes that are being incorporated, but also the role of these introgressive genes in shaping the ecological features (e.g. habitat selection, feeding ecology, individuals' home ranges) of populations. At the same time we do not known if ecological characteristics are influencing themselves the way genes introgress. Under this line of research, wildcats may be used as models to answer the key question: how is domestication shaping wild species genomes? After being able to clearly distinguish parental individuals and hybrid cats, more specific genomic questions might thus be addressed: i) how likely is it that different domestication alleles will transfer from cats into their wild ancestors?; ii) is there a sexually biased introgression of domestic alleles?; iii) are population ecological determinants being shaped by (and shaping) the genomic nature of introgression? Native California tiger salamanders (*Ambystoma californiense*) denote an excellent example of SNPs potential in studying detail dynamics of introgression, since Fitzpatrick and collaborators (2010) were able to determine that only 3 out of 68 studied markers spread rapidly into native genomes, whereas the other 65 showed little evidence of introgression beyond the region where introductions of non-native barred tiger salamanders (*Ambystoma tigrinum mavortium*) occurred. Similar studies among wildcats' populations could remodel our view of the hybridization process.

Artificial hybridization is thought to influence the conservation status of threatened species, since it may result in outbreeding depression, reduced fitness and, thus, severe population declines. This conservation-oriented view is for sure the guiding framework of this thesis. But are we sure that all hybridization is an *a priori* weakness for all wild populations? The exotic invasion of native genomes through intraspecific hybridization is not universally seen as a conservation concern because some authors argue that populations of the same species usually share alleles, and introgression would not cause outbreeding depression. Moreover, the introduction of new genetic variation could be beneficial through the improvement of population fitness (Allendorf *et al.* 2001; Candille *et al.* 2007; Anderson *et al.* 2009). A remarkable example of this is the high frequency of dark coat-colour in North-American wolves, caused by a domestic dog variant of the beta-defensin gene (Candille *et al.* 2007). Genetic data revealed that this mutation was introduced in wolf populations through crossbreeding with domestic dogs and has been

positively selected as an advantageous trait in North American forest habitats. This is indeed the first example of introgressive hybridization between a domesticate and representatives of its wild ancestor that may originate additional adaptive variation in the wild (Anderson *et al.* 2009). There may be other similar cases, but this has been considered for now the exception rather than the rule. If future studies find similar patterns among wildcat populations, the study of introgression might also move towards understanding how crossbreeding with domestic cats is at some extend helping natural population to adapt to the highly humanized European environments.

5.5. Final remarks

The work presented in this thesis provides the first low scale genome-wide characterization of nucleotide and microsatellites diversity in wild, domestic and hybrid cats and yields important insights for the species conservation. By reconciling STRs and SNPs data, we significantly improved the power of admixture analysis accomplished to date. This approach may now be used and further improved to solve both evolutionary and recent questions on wildcats' hybridization and, hopefully, guide its legal protection in Europe. In the future, dramatic enhancements in genotyping efficiency will lower the cost of high-resolution genotyping so that gains in resolution will be possible at a very small fraction of the current cost. This will make it possible to evaluate much larger numbers of loci across tens of thousands of samples, enormously increasing both the power and the resolution of evolutionary and admixture analysis.

Although our study represents a notable effort on wildcat's genetic research, it can only give us a glimpse at how European wildcat genome is being affected by ongoing hybridization and fragmentation. We hope that a continuous work on wildcat genetics will identify all critical areas of admixture, while understanding in detail which and how domestic genes introgress and actually threaten the species survival.

The conservation of the wildcat in the constantly changing, human-altered landscapes is mandatory, but one cannot forget the critical role that hybrids have in maintaining hybridization levels. Therefore, it is obligatory to include management of hybrids in future conservation strategies for the species. The eco-ethology of hybrids and their relationships among themselves and with individuals of parental populations has been poorly studied (but see Birò *et al.* 2004, 2005; Germain *et al.* 2008, 2009), although hybrids are directly involved in the phenomenon of introgression in wild populations. All progeny of a hybrid is a hybrid, which means that the proportion of admixed individuals in a population may increase even if admixture proportions (i.e. the proportion of alleles that come from each parental taxa in hybridizing populations) are constant (Allendorf and Luikart 2007). It has also been suggested that free-ranging domestic cats may have an important advantage over their wild counterparts, since the supplementary feeding by humans does not reduce their hunting skills and motivation, while causing their numbers not to be influenced by prey's population densities (Coleman *et al.* 1997). In the light of this, further studies should not only focus in a better understanding of wildcat's ecological-genetic dynamics, but also fully explore domestic and hybrid cats' characteristics.

5.6. References

Akey JM, Eberle MA, Rieder MJ, Carlson CS, Shriver MD, Nickerson DA, Kruglyark L (2004) Population History and Natural Selection Shape Patterns of Genetic Variation in 132 Genes. *Plos Biology* **2**(10): e286.

Akey, J. M., A. L. Ruhe, D. T. Akey, A. K. Wong, C. F. Connelly, J. Madeoy, T. J. Nicholas, and M. W. Neff. 2010. Tracking footprints of artificial selection in the dog genome. *Proceedings of the National Academy of Sciences of the USA* **107**: 1160-1165.

Allendorf FW, Leary RF, Spruell Paul, Wenburg JK (2001) The problems with hybrids: setting conservation guidelines. *Trends in Ecology and Evolution* **16**(11): 613–622.

Allendorf FW and Luikart G (2007*) Conservation and the Genetics of Populations*. Blackwell Publishing, Malden MA, USA.

Anderson TM, vonHoldt BM, Candille SI, Musiani M, Greco C (2009) Molecular and Evolutionary History of Melanism in North American Gray Wolves. *Science* **323** (5919), 1339.

Barbosa S, Paupério J, Searle JB, Alves PC (2012) Genetic identification of Iberian rodent species using both mitochondrial and nuclear loci: application to noninvasive sampling. *Molecular Ecology Resources.* doi: 10.1111/1755-0998.12024.

Beaumont M, Barratt EM, Gottelli D, Ktchener AC, Daniels MJ, Pritchard JK, Bruford MW (2001) Genetic diversity and introgression in the Scottish wildcat. *Molecular Ecology* **10**: 319-336.

Beaumont MA and Nichols RA (1996) Evaluating Loci for Use in the Genetic Analysis of Population Structure. *Proceedings of the Royal Society B: biological sciences* **263**(1377): 1619-1626.

Biró Z, Lanszki J, Szemethy L, Heltai M, Randi E (2005) Feeding habits of feral domestic cats (*Felis catus*), wild cats (*Felis silvestris*) and their hybrids: trophic niche overlap among cat groups in Hungary. *Journal of Zoology* **266**: 187–196.

Biró Z, Szemethy L, Heltai M (2004) Home range sizes of wildcats (*Felis silvestris*) and feral domestic cats (*Felis silvestris catus*) in a hilly region of Hungary. *Mammalian Biology* **69**: 302–310.

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

Busch JD, Waser PM, Dewoody JA (2007) Recent demographic bottlenecks are not accompanied by a genetic signature in banner-tailed kangaroo rats (*Dipodomys spectabilis*). *Molecular Ecology* **16**: 2450–2462.

Cabral MJ, Almeida J, Almeida PR, Dellinger T, Ferrand de Almeida N, Oliveira ME, Palmeirim JM, Queiroz AI, Rogado L, Santos-Reis M (eds) (2005) Livro Vermelho dos Vertebrados de Portugal. Instituto de Conservação da Natureza, Lisboa.

Candille SI, Kaelin CB, Cattanach BM, Yu B, Thompson DA, Nix MA, Kerns JA, Schmutz SM, *et al*. (2007) A B-defensin mutation causes black coat color in domestic dogs. *Science* **318**: 1418-1423.

Carneiro M, Blanco-Aguiar JA, Villafuerte R, Ferrand N, Nachman MW (2010) Speciation in the European rabbit (*Oryctolagus cuniculus*): islands of differentiation on the X chromosome and autosomes.

Evolution **64**: 3443-3460.

Coleman JS, Temple SA and Craven SR (1997) *Cats and Wildlife: a conservation dilemma*. University of Wisconsin, Madison 199–216. Serpell, J (Ed.) Cambridge: Cambridge University Press.

Cornuet JM and Luikart G (1997) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144**: 2001-2014.

Driscoll CA, Menotti-Raymond M, Roca AL, Hupe K, Johnson WE, Geffen E, Harley EH, Delibes M, Pontier D, Kitchener AC, Yamaguchi N, O'Brien SJ, MacDonald DW (2007) The near eastern origin of cat domestication. *Science* **317**: 519–523.

Driscoll C and Nowell K (2009) *Felis silvestris*. In: IUCN 2010. IUCN Red List of Threatened Species. Version 2010.4. <www.iucnredlist.org>.

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

Excoffier L, Hofer T and Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity* **103**: 285-98.

Falush D, Stephens M, Pritchard J (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**: 1567-1587.

Fitzpatrick BM, Johnson JR, Kump DK, Smith JJ, Voss SR and Shaffer HB (2010) Rapid spread of invasive genes into a threatened native species. *Proceedings of the National Academy of Sciences of the USA* **107**(8): 3606-3610.

Foll M and Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A bayesian perspective. *Genetics* **180**: 977–993.

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**: 419–428

Germain E, Benhamou S, Poulle M-L (2008) Spatio-temporal sharing between the European wildcat, the domestic cat and their hybrids. *Journal of Zoology* **276**: 195–203.

Germain E, Ruette S, Poulle M-L (2009) Likeness between the food habits of European wildcats, domestic cats and their hybrids in France. *Mammalian Biology Zeitschrift fur Saugetierkunde* **74**: 412-417.

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

Godinho R, Llaneza L, Blanco L, Lopes S, Álvares F, García E, Palacios V, Cortés Y, Talegón J, Ferrand N (2011) Genetic evidence for multiple events of hybridization between wolves and domestic dogs in the Iberian Peninsula. *Molecular Ecology* **20**: 5154–5166.

Goldstein O, Zangerl B, Pearce-Kelling S, Sidjanin DJ, Kijas JW, Felix J, Acland GM, Aguirre GD (2006) Linkage disequilibrium mapping in domestic dog breeds narrows the progressive rod–cone

degeneration interval and identifies ancestral disease-transmitting chromosome. *Genomics* **88**: 541–550.

Gray MM, Granka JM, Bustamante CD, Sutter NB, Boyko AR, Zhu L, Ostrander E, Wayne R. (2009) Linkage disequilibrium and demographic history of wild and domestic Canids. *Genetics* **181**: 1493-1505.

Heller R, Siegismund HR (2009) Relationship between three measures of genetic differentiation G_{ST}, D_{EST} and G'_{ST}: how wrong have we been? *Molecular Ecology* **18**: 2080–2083.

Hertwig ST, Jungnickel A, Stepanow S, Böhle UR, Fischer MS (2009) Regionally high rates of hybridization and introgression in German wildcat populations (*Felis silvestris*, Carnivora, Felidae). *Journal of Zoological Systematics and Evolutionary Research* **47**: 283–297.

Jost L (2008) G_{ST} and its relatives do not measure differentiation. *Molecular Ecology* **17**: 4015–4026.

Kauer MO, Dieringer D, Schlötterer C (2003) A microsatellite variability screen for positive selection associated with the "Out of Africa" habitat expansion of Drosophila melanogaster. *Genetics* **165**: 1137-1148.

Kohn MH, Murphy WJ, Ostrander EA, Wayne RK (2006) Genomics and conservation genetics. *Trends in Ecology and Evolution* **21**: 629-37.

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

Laurie CC, Nickerson DA, Anderson AD, Weir BS, Livingston RJ, et al. (2007) Linkage disequilibrium in wild mice. *PLoS Genetics* **3**: e144.

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

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics* **4**: 981–994.

Mardulyn P, Vaesen M-A, Milinkovitch MC (2008) Controlling population evolution in the laboratory to evaluate methods of historical inference. *PLoS ONE* **3**:e2960.

Meirmans PG, Hedrick PW (2011) Assessing population structure: F_{ST} and related measures. *Molecular Ecology Resources* **11**: 5–18.

Menotti-Raymond M, David VA, Eizirik E, Roelke ME, Ghaffari H, O'Brien, SJ (2009) Mapping of the domestic cat "SILVER" coat color locus identifies a unique genomic location for silver in mammals. *Journal of heredity* **100** Suppl 1, S8-13.

Monterroso P, Brito JC, Ferreras P, Alves PC (2009) Spatial ecology of the European wildcat in a Mediterranean ecosystem: dealing with small radio–tracking datasets in species conservation. *Journal of Zoology* **279**: 27–35.

Monterroso P, Castro D, Silva TL, Ferreras P, Godinho R, Alves PC (2012) Factors affecting the (in)accuracy of mammalian mesocarnivore scat identification in South-western Europe. *Journal of Zoology*. doi: 10.1111/jzo.12000.

Mountain JL, Knight A, Jobin M, Gignoux C, Miller A, Lin AA, Underhill PA (2002) SNPSTRs: Empirically derived, rapidly typed, autosomal haplotypes for inference of population history and mutational processes. *Genome Research* **12**: 1766-1722.

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

O'Brien SJ, Devillard S, Say L, Vanthomme H, Léger F, Ruette S, Pontier D (2009) Preserving genetic integrity in a hybridising world: are European Wildcats (*Felis silvestris silvestris*) in Eastern Francedistinct from sympatric feral domestic cats? *Biodiversity and Conservation* **18**: 2351-2360.

Ostrander EA and Wayne RK (2005) The canine genome. *Genome Research* **15**: 1706-1716.

Palomares F, Godoy JA, Piriz A, O´Brien SJ, Johnson WE (2002) Faecal genetic analysis to determine the presence and distribution of elusive carnivores: design and feasibility for the Iberian lynx. *Molecular Ecology* **11**: 2171-2182

Pierpaoli M, Birò ZS, Herrmann M, Hupe K, Fernandes M, Ragni B, Szemethy L, Randi E (2003) Genetic distinction of wildcat (*Felis silvestris*) populations in Europe, and hybridization with domestic cats in Hungary. *Molecular Ecology* **12**: 2585-2598.

Piñeiro A, Barja I, Silván G and Illera JC (2012) Effects of tourist pressure and reproduction on physiological stress response in wildcats: management implications for species conservation. *Wildlife Research* **39**(6) 532-539.

Piñeiro A and Barja I (2012) The plant physical features selected by wildcats as signal posts: an economic approach to fecal marking. *Naturwissenschaften* **99**(10): 801-809.

Pollinger JP, Bustamante CD, Fledel-Adon A, Schmutz S, Gray MM and Wayne RK (2005) Selective sweep mapping of genes with large phenotypic effects. *Genome Research* **15**: 1809–1819.

Queney G, Ferrand N, Weiss S, Mougel F, Monnerot M (2001) Stationary distributions of microsatellite loci between divergent population groups of the European rabbit (*Oryctolagus cuniculus*). *Molecular Biology and Evolution*, **18**, 2169–2178.

Quintela M, Thulin C-G, Höglund J (2010) Detecting hybridization between willow grouse (*Lagopus lagopus*) and rock ptarmigan (*L. muta*) in Central Sweden through Bayesian admixture analyses and mtDNA screening. *Conservation Genetics* **11**: 557-569.

Randi E, Pierpaoli M, Beaumont M, Ragni B (2001) Genetic identification of wild and domestic cats (*Felis silvestris*) and their hybrids using Bayesian Clustering Methods. *Molecular Biology and Evolution* **18** (9): 1679-1693.

Randi E (2008) Detecting hybridization between wild species and their domesticated relatives. *Molecular Ecology*, **17**: 285-293.

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

Ryan ME, Johnson JR, Fitzpatrick BM (2009) Invasive hybrid tiger salamander genotypes impact native amphibians. *Proceedings of the National Academy of Sciences of USA* **106**: 11166–11171.

Sarmento P, Cruz J, Eira C, Fonseca C (2009) Spatial colonization by feral domestic cats *Felis catus* of former wildcat *Felis silvestris silvestris* home ranges. *Acta Theriologica* **54**: 31–38.

Schlötterer C and Harr B (2002) Single nucleotide polymorphisms derived from ancestral populations show no evidence for biased diversity estimates in *Drosophila melanogaster*. *Molecular Ecology* **11**: 947–950.

Schmidt-Küntzel A, Eizirik E, O'Brien SJ & Menotti-Raymond M (2005) Tyrosinase and tyrosinase related protein 1 alleles specify domestic cat coat color phenotypes of the albino and brown loci. *Journal of Heredity* **96**: 289-301.

Shikano T, Ramadevi TJ, Merila J (2010) Identification of Local- and Habitat-Dependent Selection: Scanning Functionally Important Genes in Nine-Spined ticklebacks (*Pungitius pungitius*). *Molecular Biology and Evolution* **27**: 2775-2789.

Shimada Y, Shikano T, Merila J (2011) A high incidence of selection on physiologically important genes in the three--spined stickleback, *Gasterosteus aculeatus*. *Molecular Biology and Evolution* **20**: 181-193.

Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology* **14**: 671–688.

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

Wright D, Rubin C-J, Martinez Barrio A, Schütz K, Kerje S, Brändström H, Kindmark A, Jensen P, Andersson L (2010) The genetic architecture of domestication in the chicken: effects of pleiotropy and linkage. *Molecular ecology* **19**(23): 5140-5156.

Yuri T, Jernigan RW, Brumfield RT, Bhagabati NK, Braun MJ (2009) The effect of marker choice on estimated levels of introgression across an avian (Pipridae: Manacus) hybrid zone. *Molecular ecology* **18**: 4888-903.

Supplementary Table 1. Genetic description of all compared SNP and STR loci (Allelic richness, Ar; expected heterozygosity, H_E; Inbreeding coefficient, F_{IS}; and pairwise estimations of genetic differentiation between European wild and domestic cats, AMOVA pw-F_{CT}) and ranking values averaged across INFOCALC estimations for each type of loci independently, Rank a, and in the global set of markers, Rank b.

		Ar		H_E‡		FIS		AMOVA	Rank a	Rank b
Chr	POSITION Mb	FCA	FSI	FCA	FSI	FCA	FSI	F_{CT}	Hyb	Hyb
A1	7429296	1.561	1.129	0.222	0.044	-0.072	-0.015	0.058	145	184
A1	8742286	1.945	1.106	0.474	0.036	-0.045	-0.013	0.298	64	97
A1	10141047	1.757	1.110	0.330	0.037	-0.030	-0.013	0.136	111	149
A1	27523501	1.955	1.506	0.483	0.194	0.218	0.016	0.378	44	71
A1	68485376	1.571	1.910	**0.227**	**0.440**	0.173	0.490*	0.492	39	59
A1	69424718	1.896	1.206	0.429	0.071	0.111	-0.032	0.589	29	41
A1	133621071	1.876	1.269	0.413	0.095	0.156	-0.046	0.592	27	38
A1	151648701	1.505	1.345	0.195	0.125	0.180	-0.066	0.800	7	12
A1	175780586	1.968	1.077	0.498	0.026	0.328*	-0.007	0.386	50	77
A1	208054462	1.740	1.113	0.319	0.038	0.132	-0.013	0.126	117	155
A1	223501140	1.691	1.176	0.290	0.061	0.342	-0.026	0.757	11	17
A1	223506906	1.669	1.000	0.278	0.000	0.038	NA	0.136	122	161
A1	225057933	1.806	1.479	0.362	0.182	-0.020	0.172	0.054	127	166
A1	235579538	1.537	1.041	0.210	0.014	0.088	0.000	0.078	139	178
A1	242150000	1.814	1.540	0.368	0.210	0.163	0.175	0.560	30	44
A2	554046	1.499	1.039	0.193	0.013	0.107	0.000	0.070	146	185
A2	19001000	1.129	1.912	**0.044**	**0.442**	0.321	0.432*	0.325	80	117
A2	152258936	1.944	1.147	0.473	0.050	0.082	-0.02	0.284	65	98
A2	201526186	1.730	1.172	0.313	0.059	0.314	-0.025	0.732	13	19
A2	202225770	1.894	1.106	0.428	0.036	0.079	-0.013	0.230	81	118
A3	12082294	1.841	1.000	0.386	0.000	-0.075	NA	0.219	91	128
A3	38781591	1.143	1.964	**0.049**	**0.492**	-0.022	0.446*	0.444	55	86
A3	75156179	1.943	1.408	0.471	0.151	0.212	0.084	0.447	41	61
A3	91058022	1.959	1.174	0.488	0.060	0.196	-0.026	0.316	61	93
A3	99507784	1.828	1.234	0.377	0.082	0.093	-0.038	0.137	106	144
A3	130195244	1.856	1.074	0.397	0.025	0.178	-0.006	0.207	88	125
A3	159537633	1.960	1.971	**0.489**	**0.499**	0.042	0.387	0.000	157	196
A3	162208567	1.809	1.148	0.363	0.051	0.191	-0.02	0.684	18	25
B1	10420438	1.472	1.404	0.180	0.149	0.139	-0.082	0.798	9	15
B1	12214271	1.927	1.263	0.457	0.093	0.070	-0.045	0.221	78	115
B1	54775572	1.804	1.000	0.360	0.000	-0.023	NA	0.193	96	133
B1	68520000	1.808	1.076	0.363	0.025	0.284	1.000	0.173	99	136
B1	80161671	1.967	1.000	**0.222**	**0.481**	0.288	NA	0.398	48	75
B1	88148379	1.969	1.000	0.498	0.000	-0.015	NA	0.416	46	73
B1	105520000	1.529	1.037	0.206	0.012	0.225	0.000	0.08	140	179
B1	158896635	1.335	1.208	0.122	0.072	-0.066	-0.033	0.885	3	5
B1	176151181	1.803	1.239	0.359	0.084	-0.008	-0.039	0.669	16	23
B1	195678303	1.788	1.000	0.350	0.000	0.178	NA	0.186	100	137
B1	199564532	1.522	1.037	0.203	0.012	-0.053	0.000	0.078	142	181
B1	202966562	1.911	1.304	0.442	0.109	0.216	-0.055	0.538	32	49
B2	3940000	1.958	1.000	0.487	0.000	0.379*	NA	0.362	56	87
B2	6949528	1.817	1.141	0.369	0.048	0.322	-0.019	0.158	103	141
B2	41509834	1.766	1.000	0.336	0.000	0.083	NA	0.174	104	142
B2	45093345	1.970	1.349	0.500	0.127	-0.034	0.323	0.331	52	80
B2	138312489	1.955	1.072	0.484	0.024	0.193	-0.006	0.336	59	91
B3	13666494	1.967	1.174	0.496	0.060	-0.064	-0.026	0.348	53	81
B3	39203469	1.935	1.074	0.464	0.025	0.095	-0.006	0.291	67	101
B3	51317931	1.776	1.037	0.342	0.012	0.201	0.000	0.167	105	143
B3	57141954	1.843	1.272	0.388	0.096	0.276	0.216	0.627	22	32
B3	76202196	1.955	1.203	0.484	0.070	0.040	-0.032	0.296	63	96
B3	77094074	1.275	1.847	**0.098**	**0.389**	0.436	0.077	0.696	23	33
B3	81790000	1.891	1.269	0.425	0.095	0.317	0.217	0.174	90	127

B3	**104483970**	1.809	1.141	0.364	0.048	0.256	-0.019	0.153	107	145
B3	**111000326**	1.613	1.037	0.248	0.012	0.127	0.000	0.103	126	165
B4	**255106**	1.400	1.147	0.149	0.050	0.015	0.492	0.876	1	3
B4	**1687419**	1.397	1.000	0.147	0.000	-0.083	NA	0.061	152	191
B4	**3093827**	1.487	1.203	0.187	0.070	0.349	0.314	0.026	155	194
B4	**20001848**	1.186	1.823	**0.064**	**0.372**	0.199	0.606*	0.199	108	146
B4	**21098349**	1.970	1.276	0.500	0.098	0.269	0.254	0.343	51	79
B4	**40319102**	1.956	1.072	0.484	0.024	0.099	-0.006	0.339	58	90
B4	**47638578**	1.163	1.965	**0.056**	**0.492**	-0.026	0.061	0.444	60	92
B4	**105706694**	1.951	1.294	0.480	0.105	0.341*	-0.053	0.467	38	58
B4	**142658074**	1.742	1.073	0.320	0.024	0.186	1.000	0.141	114	152
B4	**143006494**	1.941	1.145	0.470	0.049	0.152	-0.02	0.281	68	102
B4	**144693308**	1.921	1.206	0.452	0.071	0.265	-0.032	0.23	79	116
B4	**146486983**	1.797	1.000	0.356	0.000	0.290	NA	0.191	97	134
B4	**147206961**	1.902	1.038	0.434	0.013	0.057	0.000	0.261	76	113
B4	**149532846**	1.375	1.336	0.138	0.121	0.139	-0.061	0.847	5	9
C1	**396397**	1.901	1.213	0.433	0.074	0.279	-0.034	0.201	83	120
C1	**17428968**	1.749	1.281	0.324	0.100	0.296	0.215	0.693	15	22
C1	**24148281**	1.761	1.176	0.332	0.061	0.174	-0.026	0.117	118	156
C1	**28702055**	1.492	1.213	0.189	0.074	0.194	-0.034	0.832	6	10
C1	**44520932**	1.586	1.037	0.234	0.012	0.168	0.000	0.096	132	171
C1	**52456776**	1.918	1.202	0.448	0.070	0.112	-0.03	0.558	31	47
C1	**116355295**	1.879	1.266	0.415	0.094	0.012	-0.046	0.165	92	129
C1	**123164748**	1.852	1.073	0.395	0.024	0.121	-0.006	0.203	89	126
C1	**190502133**	1.880	1.933	**0.416**	**0.461**	0.040	0.238	0.211	69	104
C1	**215441574**	1.915	1.140	0.446	0.048	0.148	0.492	0.245	77	114
C1	**216852686**	1.797	1.045	0.355	0.015	0.377*	0.000	0.173	101	138
C2	**187325**	1.497	1.000	0.191	0.000	0.298	NA	0.075	144	183
C2	**262401**	1.833	1.076	0.380	0.025	0.083	-0.007	0.186	93	130
C2	**5215469**	1.803	1.037	0.360	0.012	0.070	0.000	0.183	98	135
C2	**20000000**	1.955	1.298	0.483	0.106	0.536*	-0.053	0.473	40	60
C2	**106991233**	1.875	1.487	0.413	0.186	0.136	0.153	0.526	33	50
C2	**126240000**	1.125	1.932	**0.043**	**0.459**	-0.019	0.243	0.365	70	107
C2	**147124460**	1.561	1.953	0.496	0.000	-0.019	0.048	0.408	47	74
C2	**150774106**	1.442	1.000	0.167	0.000	0.079	NA	0.071	149	188
C2	**156491175**	1.942	1.543	0.471	0.212	0.158	0.109	0.397	42	66
D1	**10789012**	1.867	1.108	0.406	0.036	0.188	-0.013	0.203	86	123
D1	**15984279**	1.856	1.140	0.398	0.048	0.222	0.492	0.648	20	28
D1	**16242433**	1.791	1.143	0.351	0.049	-0.102	-0.019	0.146	109	147
D1	**18390852**	1.730	1.447	0.313	0.168	0.192	-0.096	0.662	19	26
D1	**18570323**	1.953	1.106	0.482	0.036	-0.011	-0.013	0.319	62	94
D1	**66177762**	1.540	1.073	0.212	0.024	0.005	-0.006	0.071	143	182
D1	**70545000**	1.360	1.000	0.132	0.000	0.380	NA	0.053	154	193
D1	**101321498**	1.946	1.375	0.475	0.137	0.353*	-0.074	0.217	75	112
D1	**104941557**	1.669	1.935	**0.278**	**0.462**	0.117	0.356	0.388	45	72
D1	**105498119**	1.588	1.073	0.235	0.024	0.159	-0.006	0.085	135	174
D1	**116730000**	1.952	1.108	0.481	0.036	0.144	-0.013	0.515	35	55
D1	**117527468**	1.163	1.837	**0.056**	**0.381**	-0.026	0.351	0.747	21	29
D1	**118901000**	1.858	1.172	0.399	0.059	0.028	-0.025	0.177	94	131
D1	**125811329**	1.767	1.185	0.336	0.063	0.095	-0.023	0.104	116	154
D1	**126256993**	1.733	1.000	0.315	0.000	0.048	NA	0.16	110	148
D1	**126847301**	1.613	1.072	0.248	0.024	0.203	-0.006	0.094	133	172
D1	**128010000**	1.854	1.037	0.396	0.012	0.081	0.000	0.216	85	122
D2	**717969**	1.875	1.037	0.413	0.012	0.352*	0.000	0.233	82	119
D2	**1020904**	1.953	1.072	0.482	0.024	-0.011	-0.006	0.52	34	54
D2	**1752007**	1.453	1.037	0.171	0.012	0.341	0.000	0.061	150	189
D2	**74293444**	1.669	1.178	0.278	0.061	-0.033	-0.026	0.08	129	168
D2	**91989307**	1.395	1.206	0.147	0.071	0.115	0.313	0.867	4	6
D2	**105772916**	1.797	1.038	0.356	0.012	0.206	0.000	0.178	102	139

D3	**24565823**	1.351	1.037	0.128	0.012	0.043	0.000	0.04	156	195
D3	**24823793**	1.612	1.000	0.247	0.000	0.194	NA	0.114	125	164
D3	**25530000**	1.931	1.206	0.461	0.071	0.034	0.313	0.246	72	109
D3	**122502120**	1.765	1.501	0.335	0.192	0.179	0.281	0.033	147	186
D4	**28094000**	1.364	1.000	0.134	0.000	0.142	NA	0.053	153	192
D4	**41078218**	1.970	1.147	0.500	0.050	-0.028	-0.02	0.387	49	76
D4	**42000379**	1.464	1.000	0.176	0.000	0.066	NA	0.075	148	187
D4	**63622083**	1.648	1.352	0.266	0.128	0.203	0.128	0.734	14	20
E1	**3912105**	1.569	1.072	0.226	0.024	-0.014	-0.006	0.08	136	175
E1	**4114158**	1.954	1.236	0.483	0.083	0.106	0.259	0.472	37	57
E1	**5453028**	1.968	1.234	0.498	0.082	0.124	-0.038	0.41	43	69
E1	**48228153**	1.742	1.072	0.320	0.024	0.234	-0.006	0.142	115	153
E1	**48700963**	1.398	1.000	0.148	0.000	0.015	NA	0.061	151	190
E1	**130875919**	1.587	1.000	0.235	0.000	0.507*	NA	0.109	130	169
E1	**131587399**	1.495	1.365	0.191	0.133	0.193	-0.069	0.796	8	13
E2	**3147915**	1.814	1.187	0.367	0.064	0.049	0.385	0.668	17	24
E2	**7580874**	1.814	1.403	0.367	0.149	-0.018	0.511	0.609	24	34
E2	**7950477**	1.964	1.108	0.494	0.036	0.160	0.664	0.358	54	82
E2	**8422942**	1.373	1.861	**0.137**	**0.397**	0.140	0.391	0.667	26	37
E2	**22632289**	1.933	1.037	0.462	0.012	0.018	0.000	0.3	66	100
E2	**34027888**	1.274	1.275	0.098	0.097	-0.051	0.739	0.891	2	4
E2	**35914023**	1.872	1.073	0.410	0.024	0.126	-0.006	0.217	84	121
E2	**36986631**	1.957	1.956	0.486	0.483	0.116	0.341	0.055	121	160
E2	**38860686**	1.672	1.239	0.279	0.084	0.442*	-0.039	0.75	10	16
E2	**39211557**	1.942	1.615	0.471	0.247	0.235	0.252	0.117	95	132
E2	**65436639**	1.899	1.963	**0.432**	**0.491**	0.047	0.318	0.024	137	176
E3	**36044809**	1.913	1.038	0.444	0.013	0.091	0.000	0.268	73	110
E3	**55434272**	1.884	1.210	0.419	0.073	0.083	0.313	0.187	87	124
E3	**67006512**	1.590	1.000	0.236	0.000	0.424*	NA	0.11	128	167
F1	**565223**	1.940	1.275	0.469	0.097	0.074	-0.048	0.237	74	111
F1	**21799641**	1.721	1.140	0.308	0.048	0.102	0.492	0.111	123	162
F1	**26100599**	1.890	1.247	0.424	0.087	0.176	-0.041	0.583	28	39
F1	**27124984**	1.743	1.074	0.321	0.025	-0.020	1.000	0.139	113	151
F1	**38051725**	1.955	1.150	0.484	0.051	0.149	-0.02	0.491	36	56
F1	**82716202**	1.708	1.272	0.300	0.096	0.170	0.478	0.721	12	18
F1	**91517402**	1.785	1.912	**0.348**	**0.442**	0.222	0.032	0.023	138	177
F2	**8427817**	1.955	1.971	**0.483**	**0.498**	-0.003	0.356	0.023	131	170
F2	**26886470**	1.507	1.000	0.196	0.000	0.105	NA	0.084	141	180
F2	**38395360**	1.856	1.345	0.398	0.125	0.185	-0.066	0.594	25	35
F2	**46855978**	1.871	1.965	**0.409**	**0.492**	-0.399*	-0.778*	0.045	124	163
F2	**68572596**	1.721	1.000	0.308	0.000	0.054	NA	0.154	112	150
F2	**74863327**	1.727	1.319	0.311	0.115	-0.044	0.154	0.064	134	173
F2	**78303221**	1.966	1.969	**0.495**	**0.497**	0.031	0.187	0.000	158	197
X	**4696293**	1.968	1.319	0.497	0.115	0.050	0.578	0.311	57	89
X	**5142294**	1.736	1.108	0.316	0.036	0.219	0.664	0.126	119	157
X	**6976318**	1.925	1.143	0.455	0.049	0.106	-0.019	0.254	71	108
X	**30335088**	1.713	1.075	0.303	0.025	0.361	1.000	0.133	120	159
Mean (autosomal SNPs)		**1.738**	**1.250**	**0.340**	**0.107**	**0.131**	**0.169**	**0.427**		
A1	*FCA008*	12.215	10.379	0.875	0.794	0.091	0.111	0.045	34	99
A1	*FCA090*	11.249	7.82	0.8	0.773	0.118	0.339*	0.074	30	84
A1	*FCA123*	9.676	9.479	**0.808**	**0.815**	0.014	0.197	0.088	26	68
A1	*FCA229*	9.951	10.236	0.749	0.697	0.072	0.195	0.254	3	7
A1	*FCA453*	6.357	5.995	0.685	0.668	0.027	0.079	0.148	28	78
A1	*FCA678*	7.577	7.44	0.815	0.717	0.227	0.470*	0.133	23	64
A2	*FCA105*	16.236	9.51	0.866	0.857	0.005	0.219	0.036	37	106
A3	*FCA080*	10.402	10.684	**0.79**	**0.839**	0.071	0.197	0.092	21	62
A3	*FCA224*	11.392	4.322	0.594	0.447	0.295	0.454*	0.454	1	1
B1	*FCA023*	11.562	6.363	0.772	0.672	0.118	0.196	0.248	4	8
B1	*FCA097*	10.53	9.255	0.829	0.797	0.02	0.348	0.114	15	45

B1	*FCA126*	8.944	7.221	**0.71**	**0.761**	0.248	0.169	0.086	27	70
B1	*FCA149*	7.556	6.617	0.79	0.711	0.086	0.04	0.059	35	103
B1	*FCA211*	7.546	8.856	**0.695**	**0.778**	0.11	0.14	0.053	33	95
B2	*FCA305*	10.277	3.773	0.805	0.094	0.394*	0.488*	0.353	10	31
B3	*FCA088*	12.374	6.136	0.856	0.615	0.541*	0.115	0.175	19	52
B3	*FCA391*	6.652	10	**0.678**	**0.883**	0.236	0.43	0.127	14	43
B4	*FCA069*	8.738	7.466	0.829	0.691	0.073	0.086	0.074	31	85
C1	*FCA293*	6.858	8.577	**0.714**	**0.767**	0.032	0.143	0.048	36	105
C1	*FCA649*	13.98	8.853	0.863	0.714	0.139	0.193	0.123	18	51
C2	*FCA043*	8.578	7.609	0.747	0.66	0.043	0.19	0.175	11	36
C2	*FCA077*	6.961	7.95	**0.642**	**0.76**	0.014	0.145	0.116	29	83
C2	*FCA310*	8.94	4.524	0.805	0.084	0.145	-0.024	0.311	16	46
D1	*FCA698*	14.667	9.652	0.897	0.781	0.026	0.183	0.085	20	53
D2	*FCA035*	6.998	15.763	**0.506**	**0.904**	0.154	0.624*	0.212	17	48
D2	*FCA262*	10.991	11.864	**0.841**	**0.846**	0.236	0.429*	0.055	25	67
D3	*FCA026*	10.111	12.539	0.851	0.814	0.129	0.16	0.051	32	88
D3	*FCA132*	13.888	12.571	0.896	0.831	0.017	-0.06	0.062	24	65
D3	*FCA441*	9.898	11.04	**0.74**	**0.776**	-0.006	0.096	0.005	39	158
D4	*FCA045*	18.519	10.493	0.885	0.825	0.078	0.196	0.097	12	40
E1	*FCA005*	10.046	8.672	0.79	0.753	0.079	0.167	0.015	38	140
E2	*FCA058*	7.69	6.941	**0.659**	**0.692**	0.018	0.187	0.239	9	30
E2	*FCA075*	9.019	10.924	**0.789**	**0.86**	0.036	0.235	0.123	8	27
E2	*FCA096*	11.367	12.136	**0.496**	**0.883**	0.034	0.145	0.282	5	11
E3	*FCA628*	20.123	12.775	0.869	0.82	0.083	0.27	0.128	6	14
F1	*FCA223*	16.988	7	0.908	0.646	0.057	0.142	0.14	7	21
F2	*FCA220*	10.428	9.437	**0.574**	**0.816**	0.084	0.147	0.163	22	63
X	*FCA240*	8.871	3.502	0.79	0.159	0.581*	0.074	0.333	13	42
X	*FCA651*	4.727	6.652	**0.57**	**0.767**	0.675*	0.817*	0.324	2	2
Mean (autosonal STRs)		**10.484**	**8.744**	**0.763**	**0.736**	**0.135**	**0.258**	**0.148**		

* Significant deviations from HWE at P<0.05, Bonferroni corrected
+Significant deviations from LE at P<0.05, Bonferroni corrected
‡Monomorphic loci are underlined and cases where wildcats displayed higher genetic diversity than domestic cats are shown in bold

CHAPTER 6

Conclusions

The work developed in this thesis brings new insights into the application of noninvasive molecular techniques and into the improvement of population and admixture analyses of wildcat populations. In the following paragraphs, the main conclusions are summarized.

1. Wildcats from Iberian Peninsula were among the less studied across all Europe, and molecular studies were urgent to clarify the current status of the species in this area. The initial analysis of domestic and wildcats based on 12 microsatellite loci (STRs), first in Portugal and latter in the entire Peninsula, provided the first solid evidences that Iberian wildcats represent a genetic entity clearly differentiated from their domestic relatives, and that introgressive hybridization, although present, should not have been a major factor shaping the current diversity of their genomes. A maximum of five individuals have shown evidences of recent admixture with domestic cats.

2. The assignment power of the 12 loci to identify hybrid genotypes was statistically evaluated. Results revealed that, although valuable, these STRs were below the ideal number of markers needed to identify second and third generation hybrids in our sample. In fact, the number of admixed genotypes might be actually higher, since 9% of F2 and 15% of backcrosses remained undetected. Moreover, sampling sizes should be largely increased for better describing European wildcats throughout their entire range.

3. Wildcats are endangered carnivores that live in low densities and have nocturnal and elusive behaviour. Therefore, the best knowledge on the species might be achieved through the collection of data using noninvasive sampling methods. We focused subsequent research efforts in the review of the most relevant works of noninvasive genetics and summarized their major advantages, drawbacks and future applications. The fast advances especially in the fields of forensics and medical research provide an outstanding background to the parallel application of such progresses in the study of endangered wildlife. At the same time, the great technical advances in laboratory practices and statistical analyses that are conquering the field of animal genetics promise to revolutionize wildlife conservation in a very close future.

4. Scats and hairs are considered the most effective means for the indirect study of mesocarnivores as the wildcat, but they lack from straightforward identification based on morphologic characteristics and are, thus, prone to errors in species assignment. To overcome these difficulties and promote the use of noninvasive sampling strategies in future studies of this endangered felid, we established a molecular test to identify the 16 species of carnivores that inhabit South-western Europe. This method allowed the unambiguous identification of samples through a simple PCR-SSCP protocol, and noninvasive DNA provided reasonable identification success rates. In alternative or combined to other species-specific and individual-identifier markers, this test can be widely used in the study of European wildcat's distribution, abundance, trophic preferences, pathologies, and population and introgressive hybridization dynamics.

5. A range-wide evaluation of population structure and hybridization events was performed for wildcats in Europe, based on genetic variation at 38 STRs in a total of 1128 cats. A sharp differentiation of *taxa* (European wildcat, African wildcat and domestic cat) and European wildcat's subpopulations was obtained, revealing the great utility of this type of markers to deal with closely related subspecies and populations. The 686 analysed European wildcats proved to be significantly divided into 5 major macroareas, and at least in ten minor genetic clusters, which we suggest to be regarded as independent evolutionary units for conservation purposes. Evidences of genetic bottlenecks were identified for a single subpopulation in Eastern Germany. The evaluation of hybridization rates at the different macroareas revealed that most regions are not strongly shaped by introgressive hybridization, with the exception of Hungarian and Scottish cats. Nevertheless, it was clear that analysing higher number of loci and samples from different regions resulted in the detection of higher number of hybrids relatively to previous studies (e.g. Iberian Peninsula and Italy). Although the number of loci applied here has greatly outnumbered (2-3 times) all the previous molecular studies of the species, we were still lacking full confidence in the identification of backcrossed cats since 12-14% of backcrosses with domestic cats and 7-12% of backcrosses with wildcats could be wrongly identified as parental individuals.

6. The final work of this thesis aimed at overcoming this difficulty by developing a new set of informative molecular markers. In this context, 158 SNPs were evaluated in their capacity to differentiate between European wild, domestic and hybrid cats. Results provided outstanding levels of differentiation between both subspecies (F_{ST} values between 0.515 and 0.891 for the 35 most informative SNPs) and simulations indicated that all the admixed genotypes up to 2-3 generations in the past could be confidently identified. We further suggested the selection of a smaller set of highly differentiating loci for routine, inexpensive and fast diagnosis of cat samples, and the use of more loci when justified by incongruences in morphological and genetic identifications and in particularly admixed populations. Importantly, this new set of loci might be regarded as having great value for the future analysis of noninvasive samples, due to SNPs predictable success in the amplification of low quality and quantity DNA.

Overall, the analysis of a great number of cat samples, the evaluation of numerous and different kinds of molecular markers and the development of a noninvasive molecular procedure provided great advances in the study of wildcat populations. At the same time, this work set the stage for more profound questions that need to be addressed in the future. Especially in what concerns hybridization dynamics it is crucial to understand i) what proportion of the genome introgress and model wildcat populations?; ii) which ecological and genetic variables may influence and be influenced by admixture?; iii) is all hybridization jeopardizing wild populations or may it also help natural population to adapt to the highly humanized European environments? Today, we are still in the way of fully determining the mechanisms that can support the long-term persistence of wildcat populations, but regardless of that, this dissertation leave an important contribution to wildcat's research and to the successful conservation of this endangered species.

PhD