



HAL
open science

The European wildcat as a model for the study of wildlife: focus on hybridization and the circulation of viruses

Marie-Pauline Beugin

► **To cite this version:**

Marie-Pauline Beugin. The European wildcat as a model for the study of wildlife : focus on hybridization and the circulation of viruses. Ecology, environment. Université de Lyon, 2017. English. NNT : 2017LYSE1275 . tel-01807665

HAL Id: tel-01807665

<https://theses.hal.science/tel-01807665>

Submitted on 5 Jun 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



N° d'ordre NNT : 2017LYSE1275

THESE de DOCTORAT DE L'UNIVERSITE DE LYON

opérée au sein de

l'Université Claude Bernard Lyon 1

Ecole Doctorale N° 341

Evolution, Ecosystème, Microbiologie, Modélisation

Spécialité de doctorat :

Biologie évolutive, Biologie des populations, écophysiologie

Soutenue publiquement le 15/12/2017, par :

Marie-Pauline Beugin

The European wildcat as a model for the study of wildlife: focus on hybridization and the circulation of viruses

Devant le jury composé de :

GILOT-FROMONT Emmanuelle
MICHALAKIS Yannis
NUSSBERGER Béatrice
ROUSSET François
VITALIS Renaud
PONTIER Dominique
QUENEY Guillaume

Pr/LBBE/Vet-Agro Sup/LYON
DR1/MIVEGEC/MONTPELLIER
PhD/ University of Zurich/ ZURICH
DR1/ISEM/MONTPELLIER
DR2/CBGP/MONTPELLIER
PU Université Lyon 1/LBBE/LYON
PhD/Président d'Antagene/LA TOUR DE SALVAGNY

UNIVERSITE CLAUDE BERNARD - LYON 1

Président de l'Université

Président du Conseil Académique

Vice-président du Conseil d'Administration

Vice-président du Conseil Formation et Vie Universitaire

Vice-président de la Commission Recherche

Directrice Générale des Services

M. le Professeur Frédéric FLEURY

M. le Professeur Hamda BEN HADID

M. le Professeur Didier REVEL

M. le Professeur Philippe CHEVALIER

M. Fabrice VALLÉE

Mme Dominique MARCHAND

COMPOSANTES SANTE

Faculté de Médecine Lyon Est – Claude Bernard

Directeur : M. le Professeur G.RODE

Faculté de Médecine et de Maïeutique Lyon Sud – Charles Mérieux

Directeur : Mme la Professeure C. BURILLON

Faculté d'Odontologie

Directeur : M. le Professeur D. BOURGEOIS

Institut des Sciences Pharmaceutiques et Biologiques

Directeur : Mme la Professeure C. VINCIGUERRA

Institut des Sciences et Techniques de la Réadaptation

Directeur : M. X. PERROT

Département de formation et Centre de Recherche en Biologie Humaine

Directeur : Mme la Professeure A-M. SCHOTT

COMPOSANTES ET DEPARTEMENTS DE SCIENCES ET TECHNOLOGIE

Faculté des Sciences et Technologies

Directeur : M. F. DE MARCHI

Département Biologie

Directeur : M. le Professeur F. THEVENARD

Département Chimie Biochimie

Directeur : Mme C. FELIX

Département GEP

Directeur : M. Hassan HAMMOURI

Département Informatique

Directeur : M. le Professeur S. AKKOUCHE

Département Mathématiques

Directeur : M. le Professeur G. TOMANOV

Département Mécanique

Directeur : M. le Professeur H. BEN HADID

Département Physique

Directeur : M. le Professeur J-C PLENET

UFR Sciences et Techniques des Activités Physiques et Sportives

Directeur : M. Y.VANPOULLE

Observatoire des Sciences de l'Univers de Lyon

Directeur : M. B. GUIDERDONI

Polytech Lyon

Directeur : M. le Professeur E.PERRIN

Ecole Supérieure de Chimie Physique Electronique

Directeur : M. G. PIGNAULT

Institut Universitaire de Technologie de Lyon 1

Directeur : M. le Professeur C. VITON

Ecole Supérieure du Professorat et de l'Education

Directeur : M. le Professeur A. MOUGNIOTTE

Institut de Science Financière et d'Assurances

Directeur : M. N. LEBOISNE

AKNOWLEDGEMENTS/REMERCIEMENTS

Tout d'abord, je tiens à remercier les membres du jury pour avoir accepté de juger de la qualité de mon travail de thèse. Merci aux rapporteurs de cette thèse pour leurs remarques extrêmement intéressantes sur le manuscrit.

Merci également à Dominique et Guillaume pour m'avoir prise en thèse alors que j'étudiais l'évolution culturelle chez l'Homme, que je m'apprêtais à aller étudier la méiose de la levure et que je ne savais pas que les chats sauvages existaient. Merci vraiment de m'avoir donnée l'opportunité de travailler sur un sujet aussi intéressant et plein de questionnements. Merci également d'avoir respecté mon rythme de travail (surtout mon nombre d'heures de sommeil) et pour la bonne humeur qu'il y a eu dans tous nos échanges.

Merci également à David qui m'a encadrée pour toutes les parties méthodologiques et mathématiques et qui m'a été d'un grand secours scientifique mais aussi dans les moments de stress avec sa bonne humeur et son ironie que j'ai pas toujours tout de suite saisie :D

Merci également à Stéphanie pour tes conseils, ta bonne humeur et tous ces moments où on a échangé. Cela m'a fait super plaisir de t'avoir dans l'équipe et de pouvoir te rencontrer.

Je voudrais également remercier toutes les personnes du LBBE et du labex Ecofect avec qui j'ai partagé ces années de thèse. Les Kikis, merci pour toutes ces batailles de lance-bonhomme, de ne pas m'avoir attaquée avec les fléchettes Nerf et d'avoir bien pourri mon bureau à chaque fois que j'étais absente :D Désolée d'avoir baissé le rythme avec les dessins ces derniers mois. Merci aussi à tous les autres doctorants/post-doc/permanents (on va faire bureau par bureau pour oublier personne : Seb, Miko, Elise, Salomé, Etienne, Jenn, Célia, Morgane, Laura, Tim, Valentine, Fred, Rebecca, Elodie, Louise, Victor, Eliane, Sylvie, Pierre, Pierre, Vérane, Sylvia, Jeanne, Marion, Marion) avec qui j'ai partagé des repas et plein d'autres bon moments et mauvais parfois (Seb, merci de m'avoir laissée me plaindre :D).

Merci à mon autre équipe, à Antagene, merci de m'avoir toujours accueillie et intégrée dans l'équipe comme si j'étais souvent là. Merci tout particulièrement à Anne et Cécile qui ont tout fait pour trouver comment optimiser mes aller-venues, surtout en cette fin de thèse pour ménager mon stress ☺

Merci à Thibaut pour tout ce que tu m'as appris professionnellement ou pas (:D).

Mourrrraahh

Merci à tous mes copains de prépa ou pas, qui étaient sur Lyon ou pas : Papa Gauthier, Marine, Gabi, Zean-Zean, Guigui, Ulyyyyyysse, Laetitia, Jean, Maelys, Mathilde, Céline, Loulou, Marion, Colin, Benji, Diddy, Lily, Nico...pour toutes ces années de rire, de blagues incomprises, de tapons et de pickapoo... Et merci de toujours être là ☺

Merci à ma famille qui est très sympa et qui m'a toujours soutenue quoi que je décide de faire bien que ce soit parfois assez difficile à suivre. J'espère vous amuser plus que je ne vous stresse ☺

Je suis aussi obligée de remercier Bob. Tu sais que j'ai envie de te vanner pour ce que tu as écrit dans tes remerciements de thèse mais je vais rester sobre. Merci d'avoir partagé ces quelques années avec moi, j'espère en passer encore quelques-unes encore à tes côtés où que ce soit, baignée de ton amour à contempler tes yeux assortis à la piscine. Après on verra ce qu'on fait :p

Finalement, je me dois de remercier le chaton, Lady Fluffy, qui a désespérément tenté de m'étouffer dans mon sommeil pendant deux ans et demi. Grâce à elle, j'ai pu constater à quel point les chats ont une manière bien à eux de penser, d'agir et de faire tomber toutes sortes d'objet au sol en ayant toujours l'air aussi surpris. Je ne sais pas si elle m'a appris grand-chose sur le chat sauvage, mais en tout cas, elle m'a permis de toujours aimer les chats même dans les moments les plus difficiles.

Une grosse pensée également pour Pilou qui m'a écouté réviser mes leçons et préparer mes oraux pendant bien des années...

Cette thèse s'est réalisée dans le cadre d'une convention CIFRE (ANRT N° 2014/0504) entre la compagnie Antagene et le Laboratoire de Biométrie et Biologie Evolutive de Lyon, et dans le cadre des travaux du LABEX ECOFECT (ANR- 11- LABX-0048) de l'université de Lyon.

ABSTRACT

Hybridization and infectious diseases are two major issues for wildlife conservation worldwide. The European wildcat *Felis silvestris silvestris*, through its interactions with its close relative the domestic cat *Felis silvestris catus*, represents a valuable model for the study of these two issues and their interactions. The European wildcat is both threatened by hybridization and infectious diseases. This, combined with the high diversity of environments where it lives throughout Europe, allows to perform comparative studies and to understand which environmental determinants impact gene and pathogen flows. Here we propose two new methodological developments for the detection of hybrids based on genetic markers allowing for a better comparability between studies and leading to a fast detection of hybrids respectively. Hybrid detection and assessment of spatial relatedness pattern were carried out in two local populations of European wildcats differing mostly on the level of fragmentation of their environment. For one of these populations, we conducted a serological survey to investigate whether domestic cats and wildcats exchange some of the most common viruses of the domestic cat (FPV, FHV, FCV, FIV). We found a higher rate of hybridization in the most fragmented environment. There, the wildcat population, in spite of the surrounding domestic cats that were infected at high prevalence for the four viruses, was not infected by any of the viruses. The presence of genetic or behavioral barriers may explain this result. The local sampling achieved in this work allowed us to investigate mechanisms behind hybridization and viruses' circulation. In both environments, the European wildcat does not seem threatened by domestic cats. However, measures should be taken to prevent a future increase in frequency of the phenomenon both for the control of gene and virus flows.

Keywords: population genetics, modelling, infectious diseases, STRUCTURE, local sampling, felids

RESUME

L'hybridation et les maladies infectieuses sont deux problématiques majeures pour la conservation de la faune sauvage à travers le monde. Le Chat sauvage européen *Felis silvestris silvestris*, à travers ses interactions avec son proche apparenté le Chat domestique *Felis silvestris catus*, représente un modèle intéressant pour l'étude de ces deux problématiques et de leur interaction. Le fait que le Chat sauvage soit touché par ces deux phénomènes, combiné à la variabilité des habitats dans lesquels il vit en Europe, permet de conduire des études comparées et de comprendre quels déterminants environnementaux influencent les flux de gènes et de pathogènes. Ici, nous proposons deux nouvelles approches méthodologiques basées sur l'analyse de marqueurs génétiques, pour une meilleure comparabilité entre études et une détection rapide des hybrides respectivement. Nous avons recherché les hybrides et regardé la distribution spatiale des individus apparentés dans deux populations locales divergeant principalement sur le niveau de fragmentation de l'environnement. Dans l'une de ces populations, nous avons également conduit une étude sérologique pour déterminer si les chats sauvages et domestiques échangeaient certains des virus communs du Chat domestique (PVF, HVF, CVF, VIF). Nous avons observé un taux d'hybridation plus élevé dans l'environnement le plus fragmenté. Malgré la ceinture de chats domestiques infectés à haute prévalence autour d'elle, la population de chats sauvages de ce même environnement n'était infectée par aucun des virus. La présence de barrières génétiques et/ou comportementales expliquerait ce résultat. L'échantillonnage local présenté ici nous a permis de mieux comprendre les mécanismes à la base de l'hybridation et de la circulation des virus. Dans les deux environnements, le Chat sauvage européen ne semble pas menacé par le Chat domestique. Toutefois, des mesures préventives devraient être adoptées pour éviter que cela ne devienne le cas.

Mots-clé : génétique des populations, modélisation, maladies infectieuses, STRUCTURE, échantillonnage local, félidés

Table of contents

ACKNOWLEDGEMENTS/REMERCIEMENTS.....	VII
ABSTRACT.....	IX
RESUME.....	XI
LIST OF ILLUSTRATIONS.....	XVII
LIST OF TABLES	XXI

Prologue	1
1. GENERAL INTRODUCTION	3
1.1. Studying wildlife.....	5
1.1.1. Defining wildlife	5
1.1.2. The reasons for the study of wildlife	5
1.1.3. A multidisciplinary field	6
1.2. Hybridization as a major evolutionary process	7
1.2.1. From a myth to a widespread reality.....	7
1.2.2. Adaptive vs threatening hybridization	12
1.3. The increasing place of infectious diseases	17
1.3.1. Definition and incidence of infectious diseases worldwide	17
1.3.2. The study and management of infectious diseases	18
1.3.3. A wide diversity of infectious diseases dynamics	19
1.3.4. Zoonosis and emerging infectious diseases.....	20
1.4. The wildcat as a model organism for wildlife.....	23
1.4.1. Settlement history of European wildcat populations.....	23
1.4.2. Threatening interactions with the domestic cat.....	29
1.5. Goals of the PhD work	35
1.5.1. Hybridization.....	35
1.5.2. Infectious diseases	35

2.	MATERIAL AND METHODS	37
2.1.	Genetic markers.....	39
2.1.1.	What is a genetic marker?.....	39
2.1.2.	Genetic markers available for wildlife study	40
2.1.3.	Genetic markers chosen.....	42
2.1.4.	Non-invasive vs invasive genetics	43
2.2.	Epidemiological modelling	46
2.2.1.	The different categories of model.....	46
2.2.2.	Example of Mathematical formalization.....	49
2.2.3.	The viruses considered in the PhD work	54
2.3.	Areas of study and sampling method.....	56
2.3.1.	North-Eastern France	56
2.3.2.	French Pyrenees.....	58
3.	DETECTING HYBRIDS, A METHODOLOGICAL CHALLENGE	61
	Foreword	63
3.1.	Iterative determination of representative parents for detecting hybrids: application to wild and domestic cat populations	65
3.2.	A fast likelihood solution to the genetic clustering problem	107
	Addendum	141

4. OPENING THE ECOLOGICAL BLACK BOX TO UNDERSTAND HYBRIDIZATION.....	143
Foreword.....	145
4.1. Female in the inside, male in the outside: insights into the spatial organization of a European wildcat population.....	147
Interlude.....	181
4.2. Hybridization pattern of a Pyrenean population of European wildcats	183
Addendum	207
5. CIRCULATION OF VIRUSES BETWEEN DOMESTIC CATS AND EUROPEAN WILDCATS.....	209
Foreword.....	211
Are European wildcats threatened by the common viruses of the domestic cats?.....	213
Addendum	241
6. GENERAL DISCUSSION AND PERSPECTIVES	243
6.1. Methodological guidelines for the study	246
6.1.1. The genetic approach: are NGS unavoidable?.....	246
6.1.2. Choosing a method adapted to the question addressed	251
6.2. The sampling scheme: a primordial step for studies' comparability	254
6.2.1. Limits of the opportunistic sampling scheme	254
6.2.2. Local sampling schemes to understand the heterogeneity of hybridization patterns.....	255
6.2.3. How may local sampling help for hybrid management?.....	258

6.3. How may epidemiology be included in hybrid management guidelines.....	259
6.3.1. The problem with traditional hybrid management.....	259
6.3.2. Using epidemiology as a criterion for the management of hybrids	260
6.3.3. Assessing if hybrids are an epidemiological threat	264
6.4. Is the European wildcat threatened by domestic cats?	265
6.4.1. Do we fully understand the role of fragmentation on hybridization?	265
6.4.2. Do Pyrenean wildcats need particular protection measures?	267
6.4.3. Is the European wildcat threatened in northeastern France?	269
6.4.4. Future studies required to better understand the system	270
Conclusion	273
Epilogue.....	275
Résumé de la thèse en français	277
References	287
Appendix 1: references for the figure 1.2	327
Appendix 2: A single multiplex of twelve microsatellite markers for the simultaneous study of the brown hare (<i>Lepus europaeus</i>) and the mountain hare (<i>Lepus timidus</i>)	331
Appendix 3: A set of 20 multiplexed single nucleotide polymorphism (SNP) markers specifically selected for the identification of the wild boar (<i>Sus scrofa scrofa</i>) and the domestic pig (<i>Sus scrofa domesticus</i>)	341

LIST OF ILLUSTRATIONS

Figure 1.1: Attic kylix medallion (around 515 BC) representing the Minotaur.....	7
Figure 1.2: Importance of hybridization across mammal genders.....	11
Figure 1.3: Phylogenetic relationships between the sub-species of the gender <i>Felis</i> built from mitochondrial genes ND5 and ND6 (Driscoll et al. 2007).....	24
Figure 1.4: Morphological characterization of the European wildcat (Kitchener et al. 1995).....	26
Figure 1.5: Examples of phenotypes in domestic cats including striped phenotypes closed to the European wildcat phenotype.....	31
Figure 2.1: Representation of the different steps of the genotyping of a microsatellite locus in a diploid individual.....	45
Figure 2.2: Example of transition diagram.....	50
Figure 2.3: Representation of the different steps of the simulation of a stochastic epidemiological dynamics using the events presenting in Figure 2.2.....	53
Figure 2.4: Landscape configuration in North-eastern France with alternating of agricultural fields and forests.....	56
Figure 2.5: Two domestic cats trapped in a farm from North-eastern France.....	57
Figure 2.6: Landscape configuration in French Pyrenees with small villages in the middle of continuous mountainous areas.....	58
Figure 3.1.1: Description of the iterative algorithm.....	74
Figure 3.1.2: Number of parents, hybrids and suspect individuals detected in wild and domestic morphologically determined populations using the iterative algorithm either with a π value at initialization of 1%, 20% or 50%.....	84
Figure 3.1.3: Average proportion over all simulations of parental individuals and first generation backcross individuals categorized as parent, suspect and hybrid.....	88
Figure 3.1.4: Impact of the π value at initialization on the thresholds' values calculated with the ' <i>n best</i> ' method or the iterative algorithm.....	89
Figure S3.1.1: Illustration of the conditions required for individuals to be integrated in the representative pool.....	97
Figure S3.1.2: Representation of the different categories (Parent, Suspect, Hybrid) according to the q -values distributions and thresholds obtained	98

Figure S3.1.3: Convergence of the algorithm for the definition of the representative pool in domestic cats and wildcats depending on the initialization.....	99
Figure S3.1.4: Impact of the value of π at initialization on the categorization of domestic cats and wildcats depending on the initialization and the method (conservative vs. relaxed) with the 'n best' method.....	100
Figure S3.1.5: Comparison of the results obtained for the categorization of cats either as parent, hybrid or suspect with the iterative algorithm, the 'n best' method and the <i>threshold</i> method for the domestic cats and the wildcats.....	101
Figure S3.1.6: Convergence of the algorithm for both the <i>ideal</i> and the <i>continuum</i> scenario for the small and the large populations.....	102
Figure S3.1.7: Impact of the π value at initialization on the proportion of known parents and hybrids integrated in the representative pool for the <i>continuum</i> scenario.....	103
Figure S3.1.8: Impact of the π value at initialization on the proportion of known parents and hybrids integrated in the representative pool for the <i>ideal</i> scenario.....	104
Figure S3.1.9: Impact of the value of π at initialization on the categorization made by the iterative algorithm.....	105
Figure S3.1.10: Variability of the categorization across simulations using the relaxed approach.....	106
Figure 3.2.1: Comparison of the various methods on simulated genetic clusters.....	125
Figure 3.2.2: Comparison of <i>snapclust</i> and NEWHYBRIDS for the identification of hybrids using simulated data.....	127
Figure 3.2.3: Illustration of <i>snapclust</i> using simulated hybrids from cattle breed microsatellite data.....	129
Figure S3.2.1: Selection of the optimal number of clusters for the microbov example.....	139
Figure S3.2.2: Output of the cross-validation analysis carried out to determine the number of principal components to retain in the DAPC led on the microbov example.....	140
Figure 4.1.1: Distribution of the villages surrounding the wildcat population.....	153
Figure 4.1.2: Body weight and age according to the sex of individuals and their sub-species.....	158
Figure 4.1.3: Spatial organization of the wildcat population.....	164
Figure S4.1.1: Quality verification of the radio-tracking dataset.....	172
Figure S4.1.2: Number of locations per month for the eight individuals taken into account in the statistical analyses of the home-range.....	174
Figure S4.1.3: Results obtained with STRUCTURE HARVESTER using Evanno's method for the determination of the optimal number of clusters.....	175

Figure S4.1.4: Structure membership probabilities for domestic cats and wildcats.....	176
Figure S4.1.5: Radio-tracking locations obtained for each individual.....	177
Figure S4.1.6: Locations of the wildcats and distance to villages.....	180
Figure 4.2.1: Localization of the study area.....	195
Figure 4.2.2: Locations of the samples according to the sex of individuals.....	196
Figure 4.2.3: Scatter plot of the discriminant analysis of principal components.....	198
Figure 4.2.4: Results from the STRUCTURE analysis for domestic cats and wildcats.....	199
Figure S4.2.1: Linear regression and 95% confidence interval of the distance between two fresh samples according to the coefficient of relationship.....	205
Figure S4.2.2: Results from the <i>snappclust</i> analysis for domestic cats and wildcats.....	206
Figure 5.1: Transfer diagram representing the dynamics of the FPV, FIV, FCV and FHV.....	222
Figure 5.2: Theoretical illustration of the kind of results expected through our modelling approach.....	229
Figure 5.3: Differential number of effective contacts between domestic cats and wildcats (φ) and basic reproductive number ratio (ρ) for which a null prevalence is expected.....	232
Figure S5.1: Combination of φ and ρ for which a null prevalence is expected for different wildcat population sizes and all cats transmitting the virus.....	237
Figure S5.2: Distribution of the ages in the entire population of domestic cats.....	238
Figure S5.3: Distribution of the ages of the infected individuals for each of the four viruses considered in the study.....	239
Figure 6.1: Graphic interpretation of the decision tree proposed by Wayne and Schaffer (2016) and decision tree implemented with an epidemiological criterion.....	263

LIST OF TABLES

Table 2.2.1: Events of the SI model presented and associated transition rates.....	52
Table 3.1.1: Sample sizes simulated for each category of individuals.....	85
Table S3.1.1: Distribution of the 22 microsatellite markers in the different multiplexes.....	97
Table 3.2.1: Parameters used in the simulations using the computer program QuantiNEMO.....	121
Table S3.2.1: Runtimes of the different methods for the clustering and hybridization analyses.....	134
Table S3.2.2: Effect of different variables on the simulation results in the absence of hybrids on the True Positive Rate (TPR).....	135
Table S3.2.3: Effect of different variables on the simulation results in the absence of hybrids on the True Negative Rate (TNR).....	136
Table S3.2.4: Effect of different variables on the simulation results in the presence of hybrids on the probability of correct assignment.....	137
Table S3.2.5: Effect of different variables on the simulation results in the presence of hybrids on the support to the real class.....	138
Table 4.1.1: Classification of the wildcats according to their kinship relationships.....	160
Table S4.1.1: Microsatellite markers used for the species identification and the parentage analysis.....	170
Table S4.1.2: Probabilities of belonging to the assigned category compared to the three others for the retained pairwise kinships.....	171
Table S4.1.3: Wildcat individuals used for each analysis performed in the study.....	178
Table S4.1.4: Criteria used to keep or discard individuals in the different analyses performed in the study.....	179
Table S4.2.1: Indices of diversity for the domestic cats and the European wildcats sampled in the Pyrenees.....	203
Table 5.1: Models chosen for each virus and associated parameters.....	225
Table 5.2: Transitions and associated transition rates in the stochastic model.....	225
Table 5.3: Number of infected individuals over the total number of individuals tested for the different hybrid categories.....	230

Prologue

All along my cursus, I have mixed applied studies during my Engineering School and more fundamental research, choosing all my internships in academic laboratories. The possibility to carry out a PhD thesis, thanks to a CIFRE convention, both in the private laboratory Antagene and the academical laboratory of Biometry and Evolutionary biology has been an amazing opportunity to carry on my studies remaining connected both to the fundamental sciences, and to practical considerations which are often more easily understandable by large audience.

Antagene is an animal genomics laboratory that historically achieved genetic tests to detect hereditary diseases, and for genetic identification in pets. More recently, Antagene proposed new services for the genetic study of wildlife, including the development of markers and panels of markers, and genetic analyses for a wide range of species (partridges, cats, wild boars, wolves, hares, rodents, etc). My PhD consisted in focusing on the European wildcat in order to acquire a competence for the study of hybridization that could be used in other species as part of Antagene's activities. The methods used and/or developed for the study of the European wildcat were then applied in these other species. I also participated in bio-informatic analyses and literature monitoring in these other species.

In this introduction, I introduce first wildlife and the problematics attached to its study. I present then the two problematics I have been focusing on during my PhD, hybridization and the circulation of viruses, and finally present the European wildcat.

1. GENERAL INTRODUCTION

1.1. STUDYING WILDLIFE

1.1.1. DEFINING WILDLIFE

The definition of wildlife has been strongly consistent over time and based on the opposition between animals which can be owned by humans and animals which cannot. In the *History of animals* (343 BC), Aristote describes thus private (domestic) and non-private (wild) animals and specifies that some non-private animals will never become private like the wolf for instance. Subsequent studies led twenty centuries later about dog domestication processes showed that the wolf may not have been such a good example (Vila et al. 1997, Leonard et al. 2002). In the XVIIIth century, Buffon also dedicates a substantial part of its *Natural History* (1749) to the description of wildlife and described it again as “free”, still in opposition to owned animals, but also as “happy” and “innocent”, thus anchoring deeply his definition in the “good savage” myth conveyed by the *Essais* from Montaigne or the *Discours sur l'origine et les fondements de l'inégalité des hommes* written by Buffon's contemporary, Rousseau. Today, in legal texts, wildlife remains defined in opposition to owned and exploited animals, pets and farm animals, which do not live in natural environment.

1.1.2. THE REASONS FOR THE STUDY OF WILDLIFE

The study of wildlife is intimately linked to its management and preservation which themselves are deeply anchored in hunting practices for which the long-term stock sustainability is a key parameter. Traces of such wildlife management can be found since the thirtieth century in China. However, the very first work associated with wildlife management is the *Game management* of Leopold (Leopold 1933). The tight connection between hunting and wildlife management is still going strong. In France, for instance, the ONCFS (National Office for Hunting and Wildlife) plays an important role for the management both of game species for stock sustainability purposes (Saint-Andrieux and Barboiron 2015, Mauvy et al. 2010)

and more recently of species which do not belong to the game species list and whose conservation is threatened (Piédallu et al. 2016).

In recent years, another phenomenon has triggered the development of wildlife studies. Human societies have indeed come to realize that the resources on Earth are limited (Odum and Barrett 2005) and that they need to be preserved. These resources include the biodiversity and the species, especially wild animal species, living on earth.

1.1.3. A MULTIDISCIPLINARY FIELD

The study of wildlife is a very general notion that encompasses many issues which call for the use of a wide variety of techniques and concepts belonging to different disciplines (Bury et al. 2006). Then, wildlife research can consist in the investigation of on-going mechanisms at the individual level such as stress response (Sheriff et al. 2011) or personality traits (Dingemanse et al. 2004), as much as processes involving several individuals either from the same population (for instance social patterns, Armitage et al. 2017) or from different populations and/or species (like hybridization or inter-specific competition). All these levels of focus as well as the issues they convey are closely intertwined and hardly understandable when taken independently. Lately, the impact of men's activities on all these different issues has been increasingly studied in order to better assess our impact on the environment.

During my PhD, I firstly focused on hybridization, an important issue that is often promoted by human's activities (Crispo et al. 2011). Secondly, I focused on the factors which can promote or prevent the circulation of viruses between two hosts, a topical point for wildlife with strong interactions with hybridization issues.

1.2. HYBRIDIZATION AS A MAJOR EVOLUTIONARY PROCESS

1.2.1. FROM A MYTH TO A WIDESPREAD REALITY



Figure 1.1: Attic kylix medallion (around 515 BC) representing the Minotaur. National museum of archeology of Madrid

Hybridization is a concept well apprehended by the collective mind as the mixing of individuals belonging to different species. Such hybrids are notably widely represented in mythology (Figure 1.1) with, for instance, the Minotaur, son of the human Pasiphae and a bull. This view of hybridization actually fits with the historical definition of the phenomenon which defined hybrids as the sterile progeny of two individuals belonging to different species (Mayr 1942). According to this definition, hybrids are systematically sterile due notably to genetic incompatibilities between the hybridizing populations (Burke and Arnold 2001, Dobzhansky-Muller Model). The Mule and the Hinny, which result from the crossing of horses (*Equus caballus*) and donkeys (*Equus asinus*) are good

illustrations of such hybridizing process (Benirschke et al. 1962). This definition of hybridization implies that hybridizing individuals will leave less offspring than non-hybridizing ones and thus have a lower fitness. Outbreeding behaviors would then be selected against by natural selection and hybrids would result from ‘mistakes’. A definition of the species solely based on a sterility criterion is however debated (Mayr 1942) notably because this sterility is far from being systematic in experimental studies. Darwin (1859) provided many examples, mostly drawn from plants, where the hybridization between different species led to fertile hybrids and, on the contrary, examples where hybridization between varieties led to sterile progeny. Several propositions have been made to overcome this limitation.

The main difficulty for the definition of hybridization can be illustrated by this question in Harrison (1993):

Is there a middle ground between hybrids as the product of interspecific crosses and hybrids as offspring of any pair of genetically distinct individuals?

Different definitions of hybridization, differing mostly regarding the level of differentiation to be considered between the parental individuals (Stebbins 1959, Harrison 1990, Gompert and Buerkle 2016) have been proposed. However, and in spite of the technological advances and notably the development of genetic tools, Harrison’s question remains unanswered and definitions pile up. During my PhD thesis, according to Rhymer and Simberloff (1996), I considered hybridization as the reproduction of individuals belonging to genetically differentiated populations. Given that this definition does not include any threshold, the term ‘hybridization’ could refer to any mixing wherever it occurred between species, sub-species or lineages.

Under this definition, many more situations correspond to hybridizing events and hybrids can display different levels of fertility (Arnold and Hodges 1995, Arnold and Martin 2010): only one sex hit by sterility (Svedin et al. 2008), both sexes partially sterile or as fit as parental individuals (Arnold and Hodges 1995),

hybrids fitter than their parents although such hybrid vigor may be bound to particular environmental conditions (Grant & Grant 1992, Arnold & Hodges 1995, Fitzpatrick and Shaffer 2007). The ability of hybrids to produce offspring may also be regulated by behavioral features. For instance, in seals (*Arctocephalus spp.*) of Macquarie Island, hybrids tend to have less offspring while they are equally fertile as parents because females do not recognizing them as suitable mates (Lancaster 2006). Such impact of sexual selection on the reproductive success of hybrids has also been reported in birds and more especially in flycatchers (*Ficedula spp.*) (Svedin et al. 2008) where 80% of the loss of fertility was attributed to sexual selection (ability to mate, sperm competition).

Even if more animal populations have been considered to undergo hybridization under this new definition, hybridization in animal species has remained poorly acknowledged for years while it was largely recognized in plants as an important phenomenon (Harrison 1993). Given the reproductive system of plants which relies on wind, water or animals for seed dispersal, 'mistakes' leading to hybridization were expected to occur more often (Mayr 1942). This delay for the recognition of hybridization as a frequent and major evolutionary phenomenon in animals has widely been attributed to the influence on his contemporaries of the zoologist Ernst Mayr who predicted hybridization to be possible but highly marginal in animal species (Mayr 1942, Mallet 2005, Dowling and Secor 1997). However, the fact that plants are easier to manipulate experimentally as well as the absence of reliable genetic tools until recently may also explain this late recognition. The development of genetic tools has indeed considerably expanded our capacity to detect hybridization because hybrids are often difficult to distinguish from parental forms solely based on morphology (Mallet 2005). Notably, genetics have led to the discovery of cryptic species, i.e. species that cannot be differentiated morphologically (Clare 2011, Trigo et al. 2013), and which can hybridize.

These recent advances have thus led to a wide recognition of hybridization by the scientific community as a major evolutionary process in animals, and this

process has been documented in many animal taxa including mammals (Figure 1.2). Its role in the evolution of animal genomes remains however complex as it is both considered as a driver for the formation and evolution of species and a threat for their persistence.

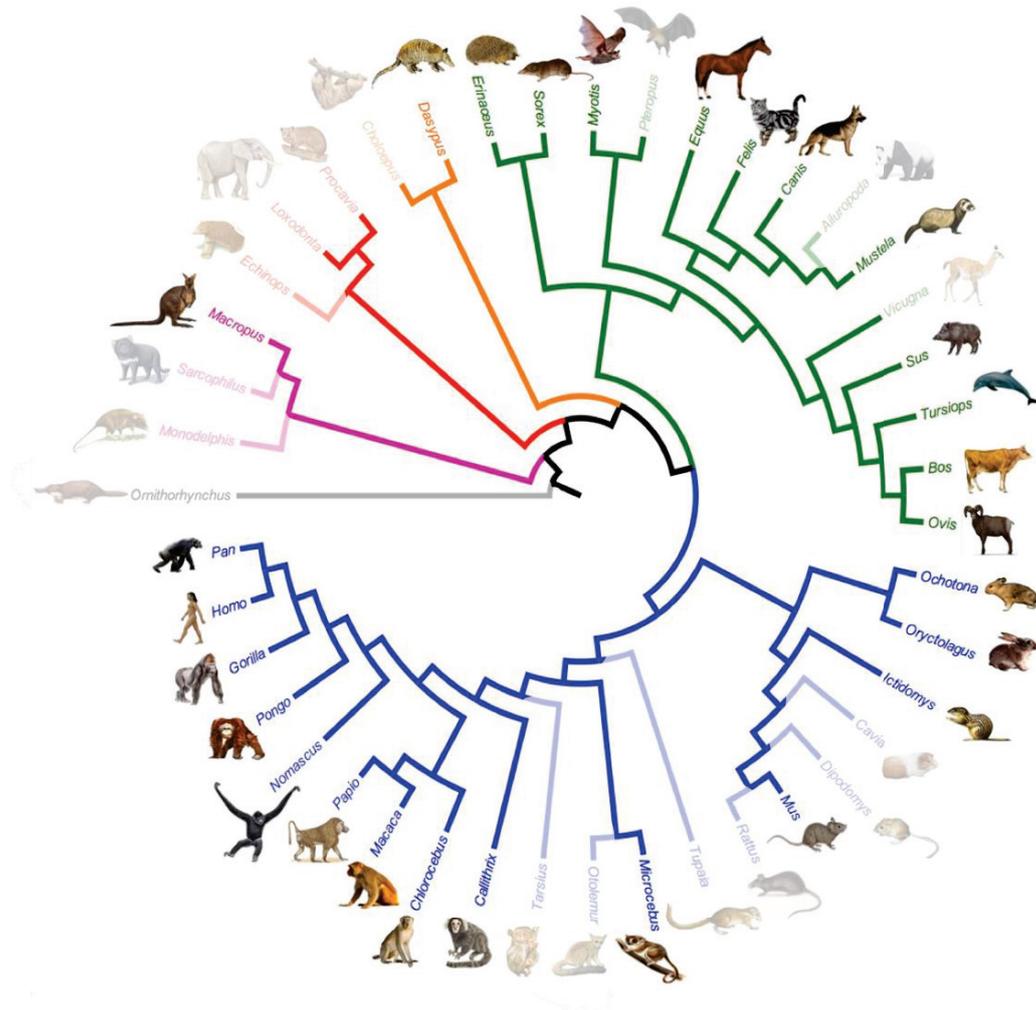


Figure 1.2: Importance of hybridization across mammal genders. Genders in transparency are those for which we found no study reporting natural and/or anthropogenic hybridization published over the last thirty years (adapted from OrthoMaM database, Appendix 1 for references).

1.2.2. ADAPTIVE VS THREATENING HYBRIDIZATION

a. Natural hybridization as a major evolutionary process

More and more evidence suggest that the species as we know them have been shaped through a process of reticulate evolution (Arnold 1992), that is with substantial gene flows between species (Arnold and Kunte 2017, Abbott et al. 2013). Arnold (2006) even proposed to consider the representation of species according to a 'web of life' instead of the more simplistic 'tree of life'. In animals, these gene flows would have occurred through hybridization and introgression defined as the durable integration of genes acquired by hybridization (Mallet et al. 2016, Larsen et al. 2010).

One way these gene flow may have shaped animal genomes is through the formation of hybrid lineages (Seehausen 2004, Abbott et al. 2013). Such hybrid speciation would depend on hybrid fitness and the availability of niches where hybrids are fitter than their parents (Abbott et al. 2013, Seehausen 2004). This speciation process is considered rare in animals but has been reported in several occasions (Larsen et al. 2010, vonHoldt et al. 2011). For instance, vonHoldt et al. (2011) showed that in North Carolina, endemic species of wolves may actually be the descendants of hybrids resulting from the reproduction of coyotes (*Canis latrans*) and gray wolves (*Canis lupus*).

The role of hybridization in the evolution of genomes is however believed to have mainly occurred *via* the transfer of genes from one population to another (Arnold and Kunte 2017, Jeong et al. 2014) during natural hybridizing events. Hybridization can indeed happen naturally, i.e., independently from human activities (Mallet 2005, Malukiewics et al. 2015), either between populations which usually live in sympatry or between populations which move and meet while they used to live in allopatry. Classically, natural hybridization is considered as mainly adaptive, the gene transfer it conveys allowing to acquire beneficial adaptations to the environment. Examples of adaptive hybridization between sympatric populations are principally reported in amphibians which would switch their

mating preference to hetero-specifics depending on environmental conditions (Pfennig 2007, Qi et al. 2014). For example, females Spadefoot toads (*Spea spp.*) prefer to mate hetero-specific males when the year is dry, which leads to a loss of fertility of the offspring but also to a faster development period necessary to get out of the pods before they are completely empty (Pfennig 2007). Transfer of advantageous adaptations through population mixing is also documented in Human populations. For instance, the admixture of Han Chinese and Sherpa would have allowed individuals to adapt to high altitude conditions (lack of oxygen, cold temperatures) and thus settle in a new environment (Jeong et al. 2014).

Examples of hybridization during secondary contacts are, for a large part, related to contacts driven by the last glacial period, which gathered many species in refuges in Spain, Italy, and in the Balkans (Stamatis et al. 2009, Ferrero et al. 2011). During these episodes of temporary sympatry, hybridization can be the simple result of a higher frequency of contacts or it can be adaptive. Such adaptive behavior has been proposed notably for the Brown hare *Lepus europaeus* (Alves et al. 2003, Melo-Ferreira et al. 2007), which presents strong introgression of mitochondrial DNA originating from the mountain hare *Lepus timidus* that lives in northern Europe and high altitude. It was proposed that, given the role of mitochondria in energetic metabolism, this introgression was adaptive and allowed brown hares to adapt to colder temperatures (Alves et al. 2008).

b. The threat of anthropogenic hybridization

Hybridization can thus promote the appearance of new species and shape existing ones. Hybridization can also lead to the collapse of several species into one (Seehausen et al. 2008) and be a source of extinction and a threat for the conservation of wildlife species (Allendorf et al. 2001).

This threatening potential stands both for low and high levels of outbreeding depression. When the outbreeding depression is substantial, i.e. when hybrids have a low fitness (sterility or low viability), the extinction of parental individuals can occur through *demographic swamping* (Wolf et al. 2001). In this case, the extinction of the parental form results from a waste of energy as it is allocated into a vain reproduction with an incompatible partner, which leads to a decline of the population's growth rate. Such extinction pathway has been proposed for trouts (*Oncoshynchus spp.*, Leary et al. 1995) but is expected to be rare (Prentis et al. 2007, Todesco et al. 2016). When the outbreeding depression is not severe and when hybrids show good levels of viability and fertility, they can lead to the disruption of local adaptations of genes (Allendorf et al. 2001) and promote the appearance in the parental population of new combinations of alleles that are deleterious and lead to the extinction of this population. When the hybrids are fit enough, they can replace the extinct parental population. Such process is classically referred to as *genetic swamping* (Wolf et al. 2001, Todesco et al. 2016). Genetic swamping is expected to happen more often than demographic swamping and has been reported in several occasions in animals (Roberts et al. 2010, Coleman et al. 2014).

These past decades, the role of hybridization in the extinction of species has been particularly investigated given that the impact of humans on their environment (fragmentation, species introductions, climate change) has created new opportunities for hybridization and, in some cases, led to the increase in frequency of the phenomenon where it already existed (Todesco et al. 2016).

This anthropogenic hybridization can result directly from the action of human when hybridization is induced by the release of exotic individuals or hybrids. For instance, restocking policies of game species are widely conducted in Galliforms

(Sokos et al. 2008) and have led, notably for the partridge (*Alectoris spp.*) or the quail (*Coturnix spp.*) to the release of hybrids in the environment (Sanchez-Donoso et al. 2012, Casas et al. 2016). The individuals bred in captivity to restock natural populations proved in these situations to be hybrids of the target species to restock and another species, reared for its meat, that, in the case of the partridge, diverged from the native species million years ago (Randi et al. 1992). The release of these domesticated individuals leads then to the introduction of alleles and genes that may be deleterious for the survival of the species.

Anthropogenic hybridization can also occur indirectly, as a result of the changes caused by human activities on their environment. For instance, climate change would have promoted the hybridization between North American flying squirrels (*Glaucomys spp.*), which had never been reported to outbreed, next to a shift in their distribution range in relation with the global warming (Garroway et al. 2010). Human activities may also impact physico-chemical properties of the environment. In Lake Victoria, the cichlid fishes (*Haplochromis spp.*), which can interbreed but do not in natural conditions due to mate choice based on coloration, were indeed observed to hybridize next to the increase in turbidity of the water (Seehausen 1997). In this case, humans promoted hybridization by making impossible the recognition of conspecifics. Finally, the fragmentation, urbanization and destruction of natural habitats create new opportunities of encounters (Crispo et al. 2011) and notably between domestic forms and their wild counterparts. Hybridizing events involving the domestic cats (*Felis silvestris catus*) and the domestic dogs (*Canis lupus familiaris*) are particularly investigated (Oliveira et al. 2007, 2008, Nussberger et al. 2013, Le Roux et al. 2015, Godinho et al. 2011, Newsome et al. 2013), notably because their popularity as pets has led to the explosion of their demography (FACCO 2016). In addition, the process of domestication and breed selection undergone by the domestic cat and the domestic dog has led to the selection of specific traits including behavioral traits, like tameness (Lindberg et al. 2007, Montague et al. 2014, Tamazian et al. 2014), and morphological traits (Lipinski et al. 2008, Anderson et al. 2009); these traits can be

particularly problematic for the survival of individuals in natural habitats or create conflicts with humans (tameness of hybrids between wolves and dogs for instance, Monzón et al. 2014).

Hybridization between wild and domestic forms may thus represent a serious threat for the conservation of the wild populations it affects. This particular situation is particularly studied, and even more given that this connection between wild and domestic populations can represent a threat for human and veterinary health if it came to promote the circulation of infectious diseases between habitats.

1.3. THE INCREASING PLACE OF INFECTIOUS DISEASES

1.3.1. DEFINITION AND INCIDENCE OF INFECTIOUS DISEASES WORLDWIDE

Infectious diseases are defined as diseases caused by an infectious agent (virus, bacteria, fungi mostly) which enters and spreads in a human or animal body and can be transmitted by infected individual to a susceptible one (Barreto 2006, WHO). This transmission occurs when susceptible and infected individuals interact in a way that allows the exchange of the agent (simple contact, sexual contact, bites). These diseases have strongly marked human demographic history with drastic declines notably after the plague pandemic in the 14th century (Morens et al. 2008) or later, during the 1918 influenza pandemic (Morens and Fauci 2013). Nowadays, their impact on human societies remains a major issue as millions of deaths resulting from infectious diseases are identified every year (WHO). For instance, respiratory infections result in the death of more than four million people worldwide every year (Fauci and Morens 2012) while AIDS, a major pandemic currently affecting human societies, affected more than 36 million people in 2015 and caused the death of more than 1 million this same year (WHO, UNAIDS 2016). Infectious diseases are also of major importance for other animal species, both domestic and wild. Only these past years, several infectious diseases have resulted in major economic loss in farm animals, for instance in duck sector in southwestern France due to the avian flu (Briand et al. 2017). Regarding wildlife, infectious diseases have proven to be a serious threat with drastic demographic decreases reported. For instance, in 2001, the outbreak of Ebola in human populations has been associated with drastic declines in great apes' populations (Leroy 2004) and particularly in Gorillas *Gorilla gorilla* (indices of presence reduced by 50 to 80% depending on the species after the epidemic). In Europe, the Iberian Lynx *Lynx pardinus* is the most threatened felid in the world (Delibes et al.

2000) mostly because of the decline of its main prey the wild European Rabbit *Oryctolagus cuniculus* whose populations have been decimated by infectious diseases (Moreno et al. 2008). Overall, 37% of the extinction in the wild of this past century are at least partly attributable to infectious diseases (Smith 2009).

1.3.2. THE STUDY AND MANAGEMENT OF INFECTIOUS DISEASES

The management of infectious diseases classically involves the use of epidemiological models that allow to understand the dynamics of the infectious diseases, predict their future evolution and also establish control and prevention measures and assess their efficiency (Ferrari et al. 2008, Haydon et al. 2006, Fulford et al. 2002). Some particular thresholds are widely used to characterize infectious diseases: the basic reproductive number R_0 , which corresponds to the number of secondary cases caused by the introduction of an infected individual in a population entirely susceptible (Diekmann and Heesterbeek 2000), and the critical community size (CCS) which is the minimal population size necessary to sustain a pathogen (Lloyd-Smith et al. 2005, Bartlett 1957). The R_0 is classically used to assess whether a pathogen can propagate in the environment. If R_0 is equal or greater than one, this means that an infected case causes the infection of at least another individual and thus that the pathogen is expected to propagate. If R_0 is below one, the pathogen is not expected to spread. The R_0 has been widely used to assess for instance the percentage of a population that must be vaccinated to stem the propagation of a pathogen (Anderson and May 1985, Fitzpatrick et al. 2012). However, when the transmission is stochastic, infectious diseases may increase in frequency even when R_0 is below the unity (Antia et al. 2003). A critical community size is expected for the viruses whose transmission depends on the total population size (Lloyd-Smith et al. 2005). The existence of this CCS would explain why threatened species, which are distributed into small populations, have less parasites compared to larger populations (Woolhouse 2001, Altizer et al. 2007).

More and more, the management of infectious diseases combines epidemiological and ecological approaches (Johnson et al. 2016). The genetic diversity of the populations (Altizer et al. 2003), their social organization (Nunn et al. 2015), dispersal behavior (Craft 2015) or inter-specific interactions (Caron et al. 2015), are thus investigated in regard to the potential role they have on the propagation of infectious diseases. New methodological approaches are thus increasingly proposed to combine ecology and epidemiology.

1.3.3. A WIDE DIVERSITY OF INFECTIOUS DISEASES DYNAMICS

The range of infectious diseases dynamics that can be modelled, studied and managed is very wide as infectious diseases can take many different forms. For some of them, like the seasonal flu, the infection will last for one or two weeks, while for others like AIDS, the infection persists until the death of the individual. Death can be induced by infectious diseases or not, in the long term or in the short term. For instance, the Marburg virus has a case fatality rate that can reach 88%, while almost all infected individuals recover from the Chikungunya (World Health Organization). These different patterns are the result of a co-evolutionary process between the host and the pathogen, which defined over time levels of virulence of the pathogen (Ebert and Hamilton 1996), or ability to develop some immune response for the host (Best and Kerr 2000).

Infectious diseases can also differ regarding the number of different hosts they affect which can go from one specific host (Kuiken et al. 2006) to more than a dozen (Bruning-Fann et al. 2001, Nugent 2011, Woodroffe et al. 2016, Roex et al. 2013, Krajewska et al. 2014, Gortazar et al. 2008). In the first case, pathogens are referred to as specialist (or single-host). The exact range of hosts infected by specialist strains is not clear across studies but ranges from one (strict specialism) to a few (Woolhouse et al. 2001), nor is defined any minimum differentiation required between two hosts to identify them as two types. Herpesviruses are

typically considered highly adapted to one specific type of host (Tischer & Osterrieder 2010) as well as the immunodeficiency viruses, which show specific strains in humans, monkeys, and cats for instance. On the contrary, pathogens able to infect a wide range of hosts are generalist pathogens (or multi-host pathogens). This type of pathogen can be illustrated by the bovine tuberculosis (Bruning-Fann et al. 2001, Nugent 2011) but also the canine distemper virus, which has been observed in a wide range of species from felids, mustelids to seals (Beineke et al. 2015). Such generalist strains and the modalities of their circulation have become the focus of many studies as they are responsible for zoonoses, i.e. infectious diseases that spread from animal species to humans (Estrada-Peña et al. 2014), and constitute an important source of emerging infectious diseases (Greger 2007, Morand et al. 2014).

1.3.4. ZONOSIS AND EMERGING INFECTIOUS DISEASES

Emerging diseases are considered as the third epidemiological transition in human history after the development of agriculture which favored pathogen exchanges between newly domesticated animals and humans (Wolfe et al. 2007), and the improvement in hygiene and nutrition at the beginning of the XXth century that resulted in a drastic decrease of the infectious diseases' incidence (Greger 2007). They correspond to diseases that have recently increased in frequency or geographical range or recently expanded their host range to a new species or result from the appearance of a new pathogen (Daszak et al. 2001). Of all the types of pathogens that can induce infectious diseases, viruses are particularly represented (Woolhouse & Gowtage-Sequeria 2005).

The emergence of an infectious disease can occur in different situations. First, the emergence can occur in a host which could not be infected by the pathogen. In that case, the emergence of the disease requires a “jump” from another host to the new host and the successful passing of many consecutive steps (Parrish 2008). Because of these many steps, the jump from one species to another is expected to

remain a rare event. Two main factors have been proposed as facilitators for such passage: their phylogenetic proximity and their contact rate. There is no clear evidence that phylogenetic proximity favours cross-species transmission as such events have happened between species phylogenetically distant (Parrish et al. 2008, Allison et al. 2012, Zhuo and Feschotte 2015). Furthermore, studies focusing on the impact of environmental changes on the emergence of infectious diseases tend to strengthen the role of the contact rate in the acquisition of a new host. The fragmentation, urbanization, climate change, etc, induced by humans' activities create new interfaces and contact opportunities (Crispo et al. 2011) and have been proposed to promote jumps and infectious diseases emergence (Estrada-Peña et al. 2014, Woolhouse et al. 2005, Murray et Daszak 2013). For instance, Murray and Daszak (2013) estimated that 20% of the emerging diseases since 1940 are due to land use changes by humans.

Second, the emergence of an infectious disease can occur in a host that can already be infected by the pathogen involved. In this case, the absence of this pathogen was due either to a lack of exposure or a size of population below the critical community size that did not allow the persistence of the pathogen. The emergence of the pathogen in this situation is then the result of a change in the interactions between the host and its environment, including other hosts present in this environment that may be infected by the pathogen and maintain it. These other hosts may then act as a reservoir for the pathogen and introduce it in new hosts when environmental conditions create contact opportunities. The frequency of these reintroductions combined with the capacity of propagation of the pathogen in the new host will then define whether there is a real emergence of the disease in the new host population (Fenton and Pedersen 2005). Several major reservoirs have been identified like rodents (Han et al. 2015). More recently, bats have been increasingly suspected as one of the major reservoir of emerging diseases including Ebola (Leroy et al. 2005, Calisher et al. 2006). Domestic canids and felids have also been proposed as major reservoirs especially for wildlife diseases (Pedersen et al. 2007) as up to 80% of their pathogens are considered to be generalists that can

infect wildlife species (Cleaveland et al. 2001). For instance, domestic dogs may constitute a reservoir for the Serengeti ecosystem (Cleaveland et al. 2000). Domestic animals also played an important part in the evolution of human infectious diseases (Salkeld et al. 2016, Wolfe et al. 2007) and still exchange pathogens with human populations. For instance, in Saudi Arabia, an outbreak of Middle-East Respiratory Syndrome (SARS) that caused the death of more than 300 people in 2012, is believed to originate from domestic camels (*Camelus spp.*, Salkeld et al. 2016, Rasmusen et al. 2015). Domestic animals are thus involved in the transmission of infectious diseases due to their central place at the interface between wildlife and anthropic areas. The study of how wildlife species may come to interact and share viruses with domestic animals encloses many key questions in relation with the emergence of infectious diseases both in wildlife, in domesticated animals, and in humans. One particular domesticated species widely represented in urban areas is the domestic cat *Felis silvestris catus*, which may interact with wildlife species and especially with the European wildcat *Felis silvestris silvestris*, on which I focused during my PhD.

1.4. THE WILDCAT AS A MODEL ORGANISM FOR WILDLIFE

1.4.1. SETTLEMENT HISTORY OF EUROPEAN WILDCAT POPULATIONS

a. Who is the European wildcat?

The European wildcat (*Felis silvestris silvestris* Schreber 1777) is a medium-size carnivore belonging to the *Felidae* family, which would have appeared 35 million years ago (Driscoll et al. 2007). Two sub-families can be distinguished inside this family: the *panterinae* which include the Lion *Panthera leo* or the Tiger *Panthera tigris* for instance, and the *felinae* which include many species including the *Felis* gender. The *Felis* gender includes five sub-species of cats in addition to the European wildcat (Figure 1.3, Driscoll et al. 2007): the northern African wildcat *F.s.lybica* (Forster 1780), the central Asian wildcat *F.s. ornata* (Gray 1830), the southern African wildcat *F.s. cafra* (Desmarest 1822), the Chinese desert cat *F.s. bieti* (Milne-Edwards 1872) and the domestic cat *F.s. catus*.

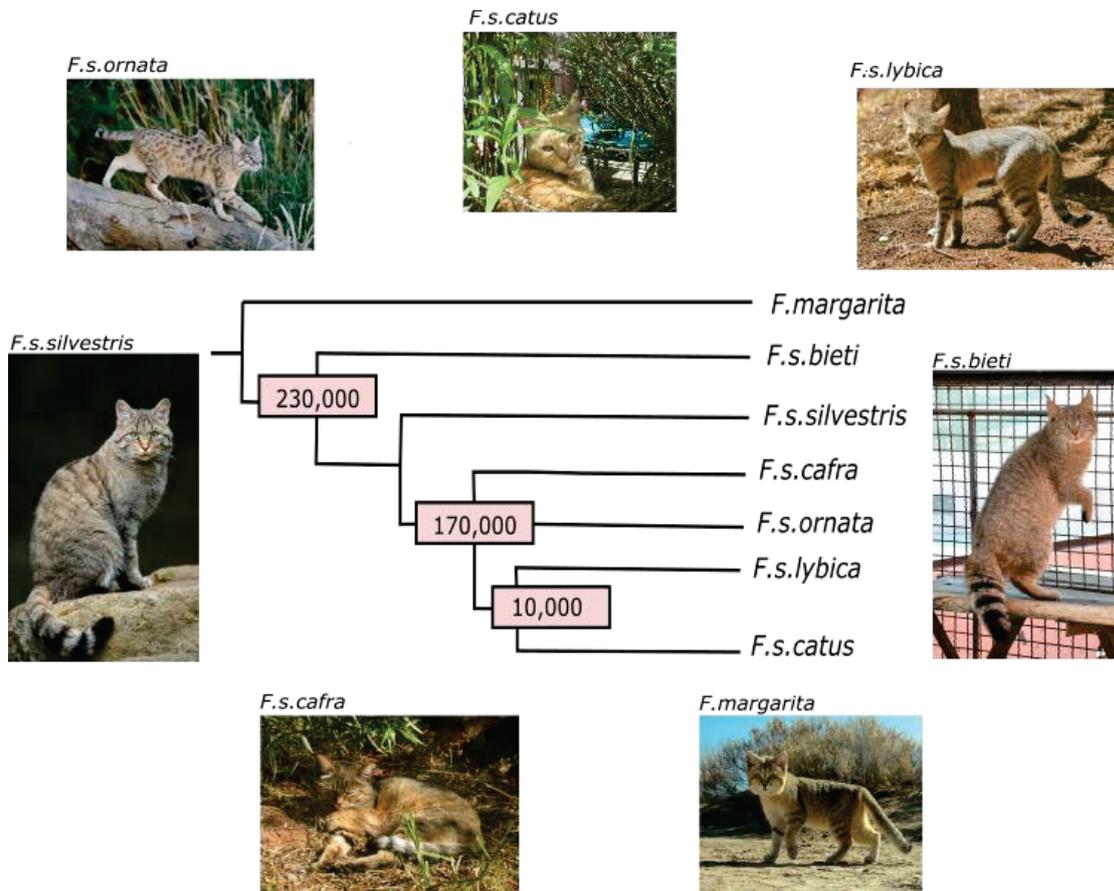


Figure 1.3: Phylogenetic relationships between the sub-species of the gender *Felis* built from mitochondrial genes ND5 and ND6 adapted from Driscoll et al. (2007). Photography credits: Terry Whittaker (*F. s. ornata*), A. Sliwa (*F. s. lybica*), XiNing Wild zoo (*F. s. bieti*), Payman Sazeh (*F. margarita*), Sonelle (*F. s. cafra*) and Luc Viatour (*F. s. silvetris*). Personal photography for the domestic cat.

Early representations of the European wildcat can be found in the *Natural History* of Georges Buffon (1749, Figure 1.4.A.). However, this representation does not include most of the morphological and anatomic criterions used to identify European wildcats. Kitchener et al. (2005) proposed an exhaustive list of twenty pelage criterions for the distinction between European wildcats and domestic cats (Figure 1.4.B.). The European wildcat is generally brown-gray or dark-gray with tabby patterns. A black line runs along its spine and stops at the basis of the tail while the dorsal line generally continues on the tail in domestic cats. The shape of the tail is also one of the main pelage difference between wild and domestic cats. The European wildcat presents a thick tail ending with a broadly rounded black tip, preceded by two to three black bands. Other pelage characteristics have been identified but often less discriminant like white fur on the chin or a reddish back of the ear. Pelage criterions can be complemented by craniological and anatomical criterions such as the dimensions of the skull of the intestine length (Yamaguchi et al. 2004, Devillard et al. 2014). These criterions proved efficient for the identification of individuals from the two sub-species. Some proved however more efficient than the others. Thus, among the anatomical criterions, the intestinal length and the cranial volume prove to be the most discriminant (Krüger et al. 2009).

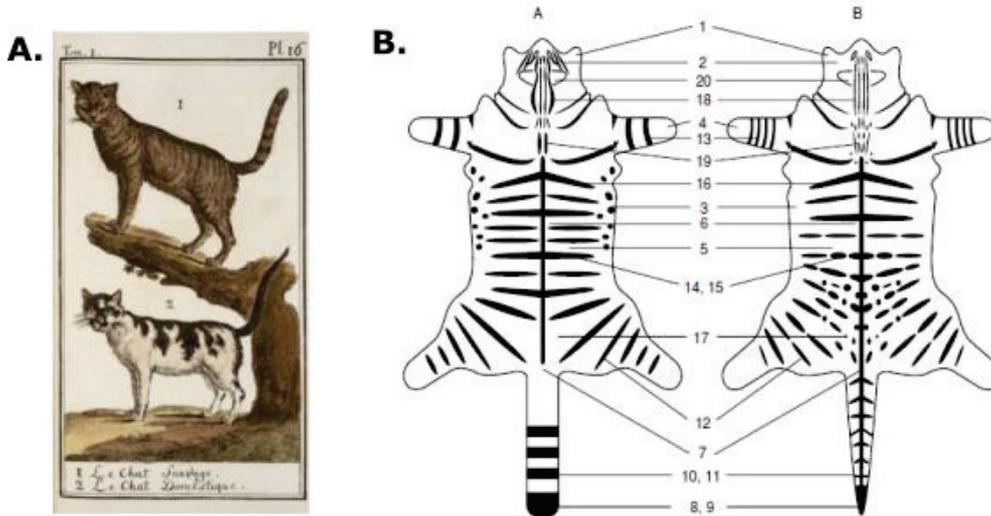


Figure 1.4: Morphological characterization of the European wildcat. A. Representation of the domestic and the wildcat from the Natural History from Georges Buffon (1749). B. representation of the twenty pelage criteria proposed by Kitchener et al. (2005). Each line and number corresponds to one of the criteria.

The European wildcat was originally believed to live exclusively in forests. However, recent studies showed that it preferentially settles in mixed habitats where both closed and open areas are available, for rest and hunting, respectively (Lozano et al. 2003, Jerosh et al. 2010, Silva et al. 2013). Wildcats mostly prey on rodents but can integrate other preys in their diet such as rabbits, which is preferred when available (Lozano et al. 2006), birds or lizards (Malo et al. 2004, Sarmiento 1996). Their diet does not include any food coming from humans as wildcats keep their distance from human infrastructures (Germain et al. 2009).

The European wildcat is a solitary and highly territorial felid like most representatives of its family (Sunquist and Sunquist 2002). For this reason, almost no overlapping is observed between the home-ranges of different individuals, and especially between females (Corbett 1979, Biró et al. 2004). Females tend to select

home ranges that gives them access to large resources to achieve reproduction and weaning of the cubs in better conditions (Sarmiento et al. 2009, Urra 2003, Stahl et al. 1988) while males' home-ranges often encompass several females' territories (Corbett et al. 1979, Stahl et al. 1988). An increase of movements across home-ranges have been reported during the mating season when individuals are looking for a suitable mate (Corbett 1979). The season of reproduction in wildcats typically starts in November and ends in February with most of the mating occurring between January and February (Condé and Shauenberg 1974). Only one litter per year is expected in Northeastern France (Condé and Shauenberg 1974) although two litters per year were suggested in Scottish wildcats (Corbett 1979). A litter is classically composed of three to five kittens, and the mother provides all the parental care.

b. Evolution of its distribution range

The first fossil records of the European wildcat correspond to a period of the Pleistocene going from 74,000 years to 15,000 years BC (Sommer and Benecke 2006). The fossils retrieved from this period and until recently show that the European wildcat has been widely distributed in Europe in the past, from Scandinavia to the Mediterranean border. The fluctuations in the distribution of the fossils reflect past expansions of the populations of European wildcats from forested habitat to mixed ones where forest alternates with more open areas, but also a sensitivity to climate conditions and particularly to too thick snow covers (Sommer and Benecke 2006, Mermod and Liberek 2002).

Today, the European wildcat is considered as a threatened species by the International Union for the Conservation of Species (IUCN) that listed it as "Least Concern" (Yamaguchi et al. 2015). The convention of Bern, which aims at promoting cooperation to ensure the conservation of natural resources and which came into effect in 1982, lists the European wildcat in the Annex II of the

convention among the “strictly protected fauna species”. At the European level, it is also listed as a “European protected species of animal” in the Annex IV for the Habitats and Species directive; and most European countries consider it as a protected species at the national level. The decline of the populations of European wildcats that led to these conservation measures results from several factors.

First, the European wildcat has been directly persecuted and hunted. In Scotland, it is still considered as a pest and persecuted (MacDonald 2004). Also, European wildcat populations are strongly impacted by traffic and road-kills (Yamaguchi et al. 2015). Besides their impact on wildcats’ mortality, roads are part of the human infrastructures that contribute to the destruction and fragmentation of the European wildcat’s habitat by imposing new barriers to wildcats movements (Klar et al. 2009, Hartmann et al. 2013). This fragmentation appears at the European scale as the distribution area of the European wildcat is discontinuous (O’Brien et al. 2008) but also more locally with the report of many populations in fragmented environments (Millán and Rodríguez 2009 in Spain, Hertwig et al. 2009 in Germany, Léger et al. 2008 in France). This fragmentation has a direct impact on the size of European wildcats’ populations and on the genetic variability they can sustain, which may impact directly their viability (Spielman et al. 2004). Fragmentation of European wildcats populations is then a major threat for the conservation of this sub-species. However, it is not considered as the major risk of extinction for the European wildcat. The domestic cat represents the major threat for the conservation of the European wildcat because these two sub-species can hybridize but also because they may exchange pathogens.

1.4.2. THREATENING INTERACTIONS WITH THE DOMESTIC CAT

a. Who is the domestic cat?

Genetic analyses classify the domestic cat *Felis silvestris catus* in the same genetic cluster than the northern African wildcat *Felis silvestris lybica* (Driscoll et al. 2007, Ottoni et al. 2017, Figure 1.3). The finding of cat remaining in a human sepulture from Cyprus dating from 9,500 years BP (Vigne et al. 2004) along with other indices like the representation of house cats in tombs in Egypt from 3,600 years BP (Driscoll et al. 2009) led to the conclusion that the domestic cats result from a domestication process undergone by northern African wildcats less than 10,000 years BP. This process of domestication is still difficult to understand because wildcats make poor candidates for domestication if we consider that their obligatory carnivore diet prevent them from digesting efficiently anything but proteins, that they are solitary and that they do not provide any clear services to the humans with whom they live (Driscoll et al. 2009). The most probable explanation for cat domestication right after the origin of agriculture 12,000 years BP, is that the stocking of grains in human settlements provided a stable resource for cats and that individuals acquired a higher tolerance for human presence that gave them access to this resource. Genes associated with tameness and response to reward would be involved in this self-domestication process (Tamazian et al. 2014, Montague et al. 2014). In exchange, cats would have kept grains from rodents (Driscoll et al. 2009). The dispersion of these tamed wildcats that became later the domestic cats we know, was achieved through human terrestrial and maritime commercial roads since prehistoric times, 6500 years BP (Lipinski et al. 2008, Ottoni et al. 2017). Upon their arrival they may have met and hybridized with local sub-species of wildcats, including the European wildcat which was then confined to Europe (Ottoni et al. 2017). The wide diversity of domestic cats breeds we observe today (Figure 1.5) is far more recent as breed selection programs would have

started 150 to 200 years ago (Lipinski et al. 2008, Driscoll et al. 2009). This selection of breeds based on morphological characters, and notably pelage color, explains the wide diversity of phenotypic patterns in the cat, which still include striped tabby patterns close to the pelage of European wildcats (Figure 1.5). Today, domestic cats are increasingly popular as pets. In France alone, the number of owned cats is measured in millions of individuals (FACCO 2016). However, this represents only a fraction of the domestic cats that are actually present as many more are expected to live under a feral way of life, i.e. without an owner, feeding opportunistically on human resources or hunting. Due to their exploding demography, domestic cats have become an issue for the conservation of several species, mostly rodents and birds (van Heezik et al. 2010, Medina et al. 2011), and also for the conservation of the European wildcat.



Figure 1.5: examples of phenotypes in domestic cats including striped phenotypes close to the European wildcat phenotype (top left and bottom left).

b. Hybridization and introgression

The phylogenetic proximity between wild and domestic cats, which would have diverged 200,000 years ago (Driscoll et al. 2007), has led to the presence of weak reproductive barriers between the two sub-species and allows hybridization (Oliveira et al. 2008, Pierpaoli et al. 2003, Beaumont et al. 2001, Hertwig et al. 2009, Lecis et al. 2006, Randi et al. 2001, Say et al. 2012, O'Brien et al. 2009, Nussberger et al. 2013). Their phylogenetic proximity is even more important bearing in mind that hybridization, and thus gene exchanges, between the two sub-species is expected to have occurred when domestic cats were brought to Europe by humans and natural hybridization between European and African wildcats would have occurred in refugia during the last glacial period (Ottoni et al. 2017).

At a recent time-scale, hybridization between the European wildcat and its domestic counterpart has been widely observed across the entire distribution range of European wildcats. In the United-Kingdom, 'pure' European wildcats may have gone extinct and replaced by a swarm of hybrids, presenting however features close to the European wildcats' (Daniels et al. 2002), notably due to persecutions (Beaumont et al. 2001). Such swarm of hybrids has also been reported in Hungary (Lecis et al. 2006). In other countries, hybridization has been reported at lower levels (from 0 to 30%, Randi et al. 2001, Oliveira et al. 2008, Hertwig et al. 2009, O'Brien et al. 2009, Nussberger et al. 2013, Ballesteros-Duperón et al. 2015). Hybridization between sub-species appears to be widespread but in highly heterogeneous proportions. As already mentioned, this phenomenon is not expected to be new (Ottoni et al. 2017) and gene flows and introgression are expected to have occurred for years between wildcats and domestic cats (Hertwig et al. 2009, French et al. 1988, Hubbard et al. 1992). Some barriers, like 'agonistic' behaviors, are thus expected to exist and to have prevented a total merge of the two sub-species (Corbett et al. 1979, French and Easterbee 1988). The increasing fragmentation and destruction of the wildcat habitat may disrupt these barriers along with demographic dynamics and promote encounters with hetero-specific mates (French and Easterbee 1988). Furthermore, domestic cats and wildcats can

share territories (Sarmiento et al. 2009). The exploding demography of the domestic cats may thus change inter-specific interactions and notably competitive interactions.

The identification of hybrids between the domestic cat and the wildcat has been classically performed based on morphological data, using an extension of the rules proposed by Kitchener et al. (2005, Figure 1.4). However efficient for the distinction of domestic cats and European wildcats, this morphological tool becomes far less efficient when it comes to distinguish hybrids from the parental sub-species (Devillard et al. 2014, Ballesteros-Duperón et al. 2015). The rise of genetic techniques has allowed to circumvent this low resolution issue and molecular markers such as microsatellite markers or single nucleotide polymorphism markers have been widely used for the study of hybridization in cats. A large tool-box is then available to investigate hybridization between these two sub-species (Nussberger et al. 2013, 2014a, Oliveira et al. 2015).

c. Pathogens sharing

The domestic cat carries a wide range of pathogens including nematodes (Fisher 2003, Traversa et al. 2009), protozoans (Lélu et al. 2010), bacteria (Handt et al. 1994) and viruses (Fromont et al. 1997, Duarte et al. 2012). These pathogens can be shared with the European wildcat (Duarte et al. 2012) and the European wildcat may even be a reservoir for some of them (e.g. *Troglostrongylus brevior*, Falsone et al. 2014).

Many viruses have been proposed to infect European wildcats. The wide majority of the serological studies conducted on the European wildcat viruses were however performed on individuals morphologically identified (McOrist et al. 1991, Daniels et al. 1999, Leutenegger et al. 1999, Millán and Rodríguez 2009) and there is a chance that positive individuals were in fact hybrids. The viruses infected and potentially threatening the European wildcat thus remain poorly known along with

the role hybrids may play in their circulation between sub-species and their evolution from a sub-species to another and notably from the domestic cat to the European wildcat. Two studies (Račnik et al. 2008, Duarte et al. 2012) were however led with a genetic identification of individuals and have shown that some viruses known to be extremely dangerous for wild felid conservation, such as the Canine Distemper Virus (Duarte et al. 2012, Roelke and Parker 1996), could infect the European wildcat. The assessment of the conditions under which viruses circulate between wild and domestic species is then a pressing issue to better understand how threatened by viruses the European wildcat may be.

d. Why is the European wildcat a relevant model?

The European wildcat, similarly to the Lynx *Lynx pardinus* for the study of the impact of roads on wildlife (Litvaitis et al. 2015), represents an interesting model organism for the study of hybridization and viruses' circulation between domestic and wild populations as it allows for the study of these phenomenon under a wide range of conditions (climatic, human persecution, fragmentation and urbanization...). Working on the domestic cat/European wildcat system also comes with a methodological and conceptual advantage. The increasing popularity of domestic cats as pets has indeed led to the soaring of veterinary practices adapted to this species. In relation with these practices, knowledge about the cat has accumulated as much regarding its anatomy, physiology, genetic, viruses (Pontius et al. 2007). The two sub-species being close, a substantial part of this knowledge and the tools resulting from it are transposable through often minor adaptation far less costly than for other species. Thus, for genetics for example, many genetic markers used for wildcats' studies have been adapted from the domestic cat markers (O'Brien et al. 2008) and many markers developed for the domestic cats are still available to improve the current molecular tools (Menotti-Raymond et al. 2003)

1.5. GOALS OF THE PHD WORK

The general objective of my PhD work was to master concepts, and methodologies encountered for the study of wildlife and especially regarding hybridization issues. For this purpose, I focused on one particular organism, the European wildcat *Felis silvestris silvestris* which may be threatened by hybridization in some parts of its distribution. The epidemiological aspects addressed in my PhD work were not directly used and applied as part of Antagene activities but represent primordial bases for a better understanding of the threat raised by hybridization events and more especially the threat it represents through its impact on virus circulation.

1.5.1. HYBRIDIZATION

More precisely, regarding hybridization, my PhD aimed both at understanding which ecological mechanisms may impact on the probability of hybridizing for wild and domestic cats, and develop new methods for the detection of hybrids. Regarding this methodological development, two different approaches were used. First, we worked on a rationalization of existing methods which define thresholds allowing to distinguish parental individuals from hybrid ones, and propose an iterative algorithm which would lead to a better comparability between studies (Chapter 3, Part 1). Second, we developed a new method, much faster than existing ones, allowing the categorization of individuals in different hybrid classes (Parents, F1, F2, backcrosses) with an efficiency comparable to what is obtained with existing programs (Chapter 3, Part 2). Regarding the ecological mechanisms, I tried to understand how the habitat fragmentation can modulate the level of hybridization of wildcats and domestic cats by studying two different populations of wildcats which evolve in very different landscape configuration: a highly fragmented environment (Chapter 4, Part 1) and in a continuous environment (Chapter 4, Part 2).

1.5.2. INFECTIOUS DISEASES

Regarding infectious diseases, the aim of the PhD was to assess, at a local scale, the level of infections of European wildcats and hybrids by four viruses

commonly found in domestic cat populations and to discuss possible underlying processes explaining the observed prevalence. No European wildcat was infected by any of the four viruses. This result led us to wonder whether this truly reflected an absence of the virus in the population or whether this absence of infected wildcats reflected something else. We thus developed a statistic-dynamic model taking into account different parameters (observed prevalence in the domestic cat populations, effective contacts between domestic cats and wildcats, propagation of the virus in the wildcat population) to assess under which conditions we could expect to observe zero infected wildcats. We then discussed whether this absence of infected wildcats could result from the presence of barriers, either genetic or behavioral (Chapter 5).

2. MATERIAL AND METHODS

GENERAL CONSIDERATIONS

2.1. GENETIC MARKERS

2.1.1. WHAT IS A GENETIC MARKER?

Genetic markers are a category of molecular markers, i.e. portions of a molecule from which information can be drawn, which have been soaring since 1980s (Schlötterer 2004). They consist in a sequence of DNA which can be present under various allelic forms in the same individuals or in the different individuals present in a population. The length of this sequence strongly varies depending on the marker from one base to several dozens. These genetic markers can be inherited from parents to offspring following Mendelian rules of inheritance for most of them (but we will see the case of mitochondrial markers and Y-linked markers). They can also undergo mutational events and thus pass from an allelic form to another. The different markers do not necessarily follow the same mutational scheme (Ellegren 2004, Morin et al. 2004). For population genetics studies, genetic markers are expected to be neutral and thus, not to be affected by selection. Thus, their allelic distribution across individuals only results from population's dynamics and reproduction patterns. Classically, several genetic markers are used jointly in order to define a genotype which will correspond to the ensemble of alleles one individual display for all of the genetic markers considered. The number of markers used is an important parameter as it impacts the resolution power with which questions can be addressed (Vahä and Primmer 2006).

Since the eighties, several molecular markers have been developed and abandoned in favor of more effective markers. Today, the genetic markers widely used are the microsatellite markers and the Single Nucleotide Polymorphism (SNP) markers. Many studies foresee SNPs as the next choice genetic marker while microsatellites would disappear (Morin et al. 2004). However, both types of markers remain widely used.

2.1.2. GENETIC MARKERS AVAILABLE FOR WILDLIFE STUDY

We propose here to review the advantages and disadvantages associated with each of these genetic markers.

Allozymes are the first molecular marker developed (Schlötterer 2004). They rely on enzymatic, i.e. protein, polymorphism which can be detected by electrophoresis due to differences in size and charge of the different forms of a considered enzyme. Allozymes are still used in recent studies, mainly for species identification (Eckert et al. 2010, Djan et al. 2014, Lodé et al. 2005). Their cheap cost is one of their most attractive features but they suffer from major flaws such as a low polymorphism, compared to microsatellites (Djan et al. 2014). In addition, their use requires fresh material and thus lacks practicality compared to DNA-based markers, DNA being more stable than proteins (Schlötterer 2004).

In early eighties, restricted length polymorphism markers (RFLP) were developed. They consist in the digestion of DNA by restriction enzyme then amplification of the restricted fragments. Depending on the individuals, the fragments will vary in size and thus, migrate differentially. The amplified fragments are not characterized using RFLPs, thus we cannot know if the markers used correspond or not to neutral areas of the genomes, which is an important requirement in many population genetics analyses (Khan 2015). Yet, RFLPs have been used in recent studies, for instance to develop markers on mtDNA (Chazara et al. 2010, Lancaster et al. 2006).

A close type of markers, the amplified fragment length polymorphism (AFLP), is used both to detect species and assess admixture between populations (Dodson et al. 2015). The use of AFLPs relies on the digestion of DNA by restriction enzymes and then includes the use of adaptors before a step of amplification (Vos et al. 1995). Although efficient (Selkoe and Toonen 2006), AFLPs require much more markers than microsatellites to reach the same resolution and thus have been largely dethroned by microsatellites.

The nineties also match with the period when random amplification of polymorphic DNA (RAPD) markers were developed (Schlötterer 2004). The use of RAPD consists in random amplifications of a genome. Although cheap, RAPD

markers are not much used nowadays for several reasons. First, they are mainly dominant, which means that heterozygotes will be seen as homozygotes. Thus, only phenotypic frequencies can be drawn from their use while allelic frequencies are essential for analyses such as admixture (Sunnucks 2000, Barilani et al. 2007a). In addition, like RFLPs, the nature of the amplified DNA being unknown, some key hypotheses in population genetics and especially neutrality may be violated and the interpretation of RAPD results difficult (Barilani et al. 2007a). RAPDs have also been discarded because of a lack of repeatability (Bonin et al. 2004).

Microsatellites have become increasingly popular along the years notably thanks to their high polymorphism that allows to use much less markers relative to other methods, to achieve more efficient analyses (Selkoe and Toonen 2006). They correspond to repeats of DNA patterns (typically 2 to 6 bases) present in high frequency in the nuclear genome. The alleles of a microsatellite, which result from slippage and errors during replication, correspond to different numbers of repeats. Typically, microsatellites display between five to forty repeats (Selkoe and Toonen 2006). Beside their high polymorphism, microsatellites are markers of choice because they are expected to be neutral with a Mendelian inheritance. Also, the flanking regions surrounding the microsatellite can be widely conserved across species, thus allowing the use of the same markers to study several species (Huang et al. 2005, Coughlan et al. 2006, Barbará et al. 2007). They have been widely used to answer many different issues from population structure (Piertney et al. 1998, Hamill et al. 2006), to hybrid detection (Oliveira et al. 2008, Godinho et al. 2011) passing by relatedness estimates (Janečka et al. 2007). However, they are far from being ideal markers. First, their mutation scheme is complex (Ellegren 2004), and the same allele may result from several evolutionary trajectories due to homoplasy processes (Selkoe and Toonen 2006). For this reason, microsatellites are rarely used to reconstruct phylogenies (Petren et al. 1999, Orisini et al. 2004), the risk of homoplasy increasing in distantly related species. Second, their analysis is not possible to automate (Schlötterer 2004, Guichoux et al. 2011), thus making time-consuming the use of an important number of

markers. Finally, the microsatellites suffer from their lack of transposability between labs (Morin et al. 2004, Schlötterer 2004, Nussberger et al. 2013).

This particular drawback of microsatellites does not stand for SNP markers which are, on the contrary, highly transposable with existing data repositories (Schlötterer 2004). They also present a rather straight forward mutation scheme, which makes easier interpretations, and are expected to be more representative of the entire genome compared to microsatellites (Morin et al. 2004). They consist in a polymorphism at one base and have generally only two alleles. Because of this low polymorphism, more SNP markers relative to microsatellites are systematically required to reach the same resolution (typically 4 to 10 more SNPs than microsatellites, Morin et al. 2004). In many studies, nuclear markers are combined with mitochondrial DNA. Mitochondrial DNA is not inherited according to Mendelian rules as it is transmitted uniparentally (typically by the mother). The mtDNA is expected to evolve faster than the nuclear genome and thus become more quickly diagnostic of taxa (Sunnuck 2000). That is why this type of marker is highly popular for phylogenetic and phylogeographic studies (Searle et al. 2009, Grill et al. 2009, Corrales et al. 2014). Mitochondrial DNA is also used in hybridization studies, in complement to nuclear markers, as it provides interesting data about the direction of hybridization (are females from one populations mating more with the males of the other population compared to the other way round?). In complement, and in order to cover both sexes, genetic markers only transmitted by the male can also be used. These markers are Y-linked markers and can, in the same way, help for the study of hybridization (Wirtz 1999, Nussberger et al. 2014b, Randi et al. 2014). It also can convey clues of ancient events of hybridization thanks to its absence of recombination between loci (Alves et al. 2003, Melo-Ferreira et al. 2007).

2.1.3. GENETIC MARKERS CHOSEN

During my PhD, I worked with a panel of 31 microsatellites, amplified in three multiplexes, which were developed in the domestic cat (Menotti-Raymond et al. 1999) and adapted to wildcats. These markers have been used in previous

studies (O'Brien et al. 2009, Say et al. 2012). We expect this number of microsatellites to allow for a high resolution for the detection of hybrids between wild and domestic cats (Oliveira et al. 2008). We chose microsatellites over SNP (Nussberger et al. 2014b, Oliveira et al. 2015 for an efficient detection of hybrids between wild and domestic cats) because, in addition to the detection of hybrids, we wanted to carry out relatedness analyses, for which microsatellites provide a higher resolution (Morin et al. 2004).

The genotyping of microsatellites requires several steps (Figure 2.1.). First the DNA is extracted from the biological material sampled. Then, the regions of the DNA that enclose the targeted microsatellite loci are amplified. Dyes of different colors (usually four colors) are used to allow for the simultaneous analysis of several loci that produce fragments of similar length. These fragments are separated according to their length by electrophoresis in a sequencer. The results of the migration are then read in a dedicated computer programs in which the intensity of the band obtained by electrophoresis is represented according to their size. Alleles for the locus are then deduced from this representation.

2.1.4. NON-INVASIVE VS INVASIVE GENETICS

The biological material on which genetic markers are analyzed can be of different natures. Classically, analyses have been performed on high quality material, such as blood and tissues, which are easy to collect on domesticated animals during veterinary controls. On wild animals, such biological material can be collected notably through protocols of capture-mark-recaptures but they imply an importance disturbance of the environment that one wants to be minimal during field studies, especially for endangered species. In addition, some species can prove hard to trap in the first place either due to their elusive behavior or because only a few individuals are left in the wild.

For these reasons, in the last twenty years, non-invasive genetic sampling has been developed. It consists in the collection of material without the need to trap or even observe the targeted individual (Taberlet and Bouvet 1992, Höss 1992, Morin et al. 1993). The material collected may correspond to feces, hairs,

urine, feather, etc. Such material can be present in large quantities on the field but it must be sampled in optimal conditions (weather, temperature, hour of the day, time since deposit) and then processed in the lab with precautions (avoidance of contamination, several genotyping) because the DNA contained in these samples is typically present in low quantity and/or quality (Taberlet and Luikart 1999, Maudet et al. 2004, Panasci et al. 2011, Wedrowicz et al. 2013). During my PhD, both invasive and non-invasive genetic samples have been used.

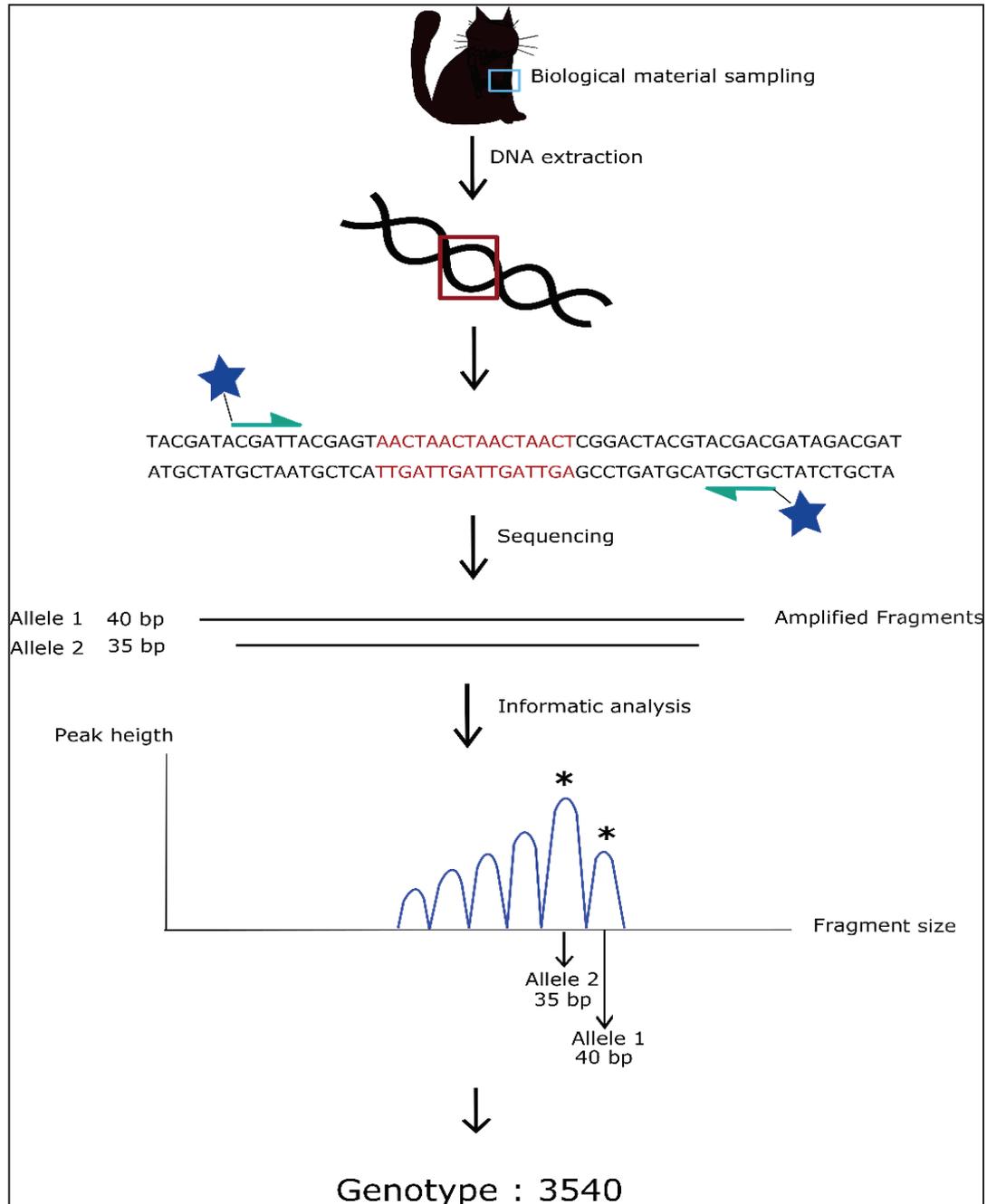


Figure 2.1: Representation of the different steps of the genotyping of a microsatellite locus of a diploid individual. In this example, the microsatellite is a tetranucleotide, i.e. the repeated motif counts four base pairs (AACT). We represent the case of a heterozygote individual. The stars represent the dye used for the amplification and sequencing step. Here, it corresponds to a blue dye.

2.2. EPIDEMIOLOGICAL MODELLING

2.2.1. THE DIFFERENT CATEGORIES OF MODEL

Mathematical modelling consists in a simplified representation of a global process by mathematical equations. The building of these equations requires a clear identification of the major mechanisms at the basis of the process we want to model (De Jong et al. 1995). The selective inclusion of only the main mechanisms involved in the global process implies that we consider the mechanisms not taken into account either as marginal or as conservative (their integration would only reinforce the conclusions of the modelling approach).

In epidemiology, the global process we want to describe is the propagation and, if necessary, the extinction of pathogens for control and prevention purposes but also to predict future evolution of the host-pathogen system. The models mainly used in epidemiology differ according to three aspects: the scale at which the propagation of the virus is considered, whether or not stochasticity is included, and the timescale at which events (infections, recovery, etc) occur.

a. Compartment models versus individual-centered models

The epidemiologic dynamics according to which a pathogen propagates in a population of hosts can be firstly described at the scale of groups of individuals. This type of models is referred to as compartment models. They rely on the description of a limited number of states that describe the hosts. For instance, if we imagine a pathogen that infects individuals for a week before it is eliminated by the host and the host to go back to being susceptible after the infection. From an epidemiological point of view, hosts can be either susceptible (available for an infection) or infected. The population studied can thus be described by these two compartments. All individuals from the same compartment are supposed to be

strictly identical. If a difference (for instance sex) has to be taken into account in the model, new compartments must be considered in the model. Individuals can transit from a compartment to another according to rates of transitions which are defined according to the system specificities. One particular rate requires to be carefully chosen: the rate of pathogen transmission. Classically, the transmission rate can be described either as dependent on the total population size of the system (action mass transmission or density-dependent transmission) or independently (proportionate mixing or frequency-dependent transmission). The relevance of these two ways of describing a transmission event remain strongly debated (De Jong et al. 1995, McCallum et al. 2001, Guiserix 2009). The epidemiological dynamics can also be described at the individual level. This scale of study allows to take into account individuals specificities (sex, age, etc) which may impact significantly the dynamics of the system under study (Bansal et al. 2007, Sieber et al. 2014, White et al. 2015). Contrary to compartment models, in such individuals-centered models, each individual is unique and cannot be taken as a surrogate of another.

In my PhD, I focused on compartment models because I worked at the population level and because the use of individual-centered models would have complicated my models and made the interpretation of the results more hazardous.

b. Deterministic versus stochastic models

Transitions from one state to another (susceptible to infected for instance) can be considered to happen at every time step in a constant manner. When this is the case, the model is considered deterministic and the results obtained with this model will be identical for one given set of initial conditions. Deterministic models can represent a good approximation for systems of infinite or very large size (Nasell 1995). For populations of limited sizes however (which correspond in fact

to any real population), it may not represent faithfully the dynamics observed on the field. Stochastic models may thus be preferred to account for real dynamics. In such models, the transition from one state to another is drawn from a probabilistic law. The process of transition from one state to another is considered without memory, this means that the state of the system at the time $t+1$ only depends on its state at time t . The probabilistic law that allows to model such process is the exponential law. Thus, the rates of transition for a stochastic model are drawn from a negative exponential law whose rate corresponds to the rate of transition used in the deterministic model. Stochastic models can thus lead to different results even when started with the same set of initial conditions. They are expected to be more precise than deterministic models (Nasell 1995) and to better capture processes of extinction of the pathogen (Bartlett 1956).

European wildcats are observed in populations in a fragmented distribution that does not allow for large population sizes, stochastic models were then implemented during my PhD thesis.

c. Discrete versus continuous time models

Finally, models can be described in discrete (e.g. Oli et al. 2006) or continuous (e.g. Berthier et al. 2000), time. In the first case, events (infection, recovery) are supposed to occur at a given frequency (every year for instance) while in continuous time models, events can occur at any frequency. In epidemiology, continuous-time models are often preferred, and this is also the type of model we chose to implement.

2.2.2. EXAMPLE OF MATHEMATICAL FORMALIZATION

Let us consider again the model that includes only two compartments: susceptible and infected. If we consider the transmission to be independent for the total size of the population, then the flow of individuals that transit from S to I at each time step can be written as $\beta SI/N$ where S is the number of susceptible individuals at the time t , I the number of infected individuals at time t , and N the total size of the population that equals at any time $S+I$. If we had considered the transmission to follow an action mass pattern, than the flow between S and I would have been βSI . The flow of individuals that recover from the virus and that goes from the compartment I to the compartment S can be expressed as dI with d the rate of individuals recovering at every step of time and I the number of infected individuals in the compartment at the time considered. At this point, it is worth noting that in compartment models, because individuals from the same compartment cannot be distinguished, the time spent in a compartment by an individual is not taken into account to determine when this individual will transit from this compartment to another. This is however not true, individuals which are sick first are expected to recover also among the first. This is then another approximation made in compartments models.

In addition to the epidemiological flows, we can add demographic flows that will correspond to the processes of death and birth in the population. If we consider b the birth rate and m the mortality rate, then, provided that both susceptible and infected reproduce similarly, the flow entering the S compartment through birth can be designed as $b(S+I)$ while the flows of individuals disappearing from the compartments S and I because they die will be mS and mI respectively.

All the compartments and flows can be reported in a diagram in order to better appreciate the model implemented (Figure 2.2.).

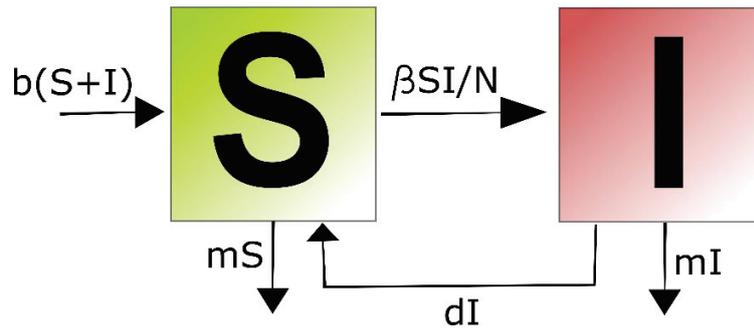


Figure 2.2: example of transition diagram. The boxes S and I correspond to the compartments « Susceptible” and “Infected” respectively. The arrows represent the flows that exist between the different compartments or coming into and from the different compartments. The expressions given beside the arrow correspond to the value of the flows. The letters b, m, d and β correspond to the birth rate, the death rate, the recovery rate and the transmission rate respectively.

Once the compartments and flows have been determined, equations can be built to represent the system.

In deterministic models, these equations are built according to an elementary law: the number of individuals in a compartment at time $t+h$ (h being the small time laps separating two steps) is equal to the number of individuals that were present in the compartment at time t minus the individuals that have left the compartment during h and increased by the number of individuals that entered the compartment during h .

For instance, if we consider the compartment S:

Individuals present at time $t = S(t)$

Individuals present at time $t+h = S(t+h)$

Individuals that left the compartment during h : $h\beta SI/N + hmS$

Individuals that entered the compartment during h : $hb(S+I) +hdI$

Then, we obtain:

$$S(t + h) = S(t) - \frac{h\beta S(t)I(t)}{N} - hmS(t) + hb(S(t) + I(t)) + dhI(t)$$

Which can be rewritten as:

$$\frac{S(t + h) - S(t)}{h} = \frac{-\beta S(t)I(t)}{N(t)} - mS(t) + dI(t) + b(S(t) + I(t))$$

Because we are in continuous time, h is expected to be very small. When h tends towards zero, this equation becomes a differential equation:

$$\frac{dS}{dt} = \frac{-\beta SI}{N} - mS + dI + b(S + I)$$

For simplification, we write S instead of $S(t)$ and I instead of $I(t)$. In a similar way, we can describe the dynamic in time of the compartment I with the following differential equation:

$$\frac{dI}{dt} = \frac{\beta SI}{N} - mI - dI$$

From this system of two differential equations, several parameters can be determined like the number of susceptible and infected individuals at equilibrium, i.e. when these number stop varying, which corresponds to the values of S and I when $\frac{dS}{dt} = 0$ and $\frac{dI}{dt} = 0$. The basic reproductive number R_0 can also be calculated

from this system as the ratio between the flow entering the infected compartment and the flow leaving the infected compartment for $S \simeq N$ and $I = 1$. In this particular example, we may find

$$R_0 = \frac{\beta}{m + d}$$

The same diagram of transitions can be used for a deterministic and a stochastic model. Then, the events considered in the stochastic model and the transition rates associated will be as reported in table 2.1.

Table 2.2.1: Events of the SI model presented and associated transition rates used at parameters of the exponential exponential law from which the probability of realization of the events is drawn.

TRANSITION	EVENT	TRANSITION RATE
$(S,I) \rightarrow (S+1, I)$	Birth	$b(S+I)$
$(S,I) \rightarrow (S-1, I)$	Death of a susceptible	mI
$(S,I) \rightarrow (S-1, I+1)$	Infection	$\beta SI/N$
$(S,I) \rightarrow (S+1, I-1)$	Recovery	dI
$(S,I) \rightarrow (S, I-1)$	Death of an infected	mI

Stochastic models and their associated transition rates can be used in two different ways. First the model can be characterized by a system of probabilistic equations that can be analytically resolved. Second, a simulation study can be performed to study the system. The simulation process used in this second situation is represented in Figure 2.3.

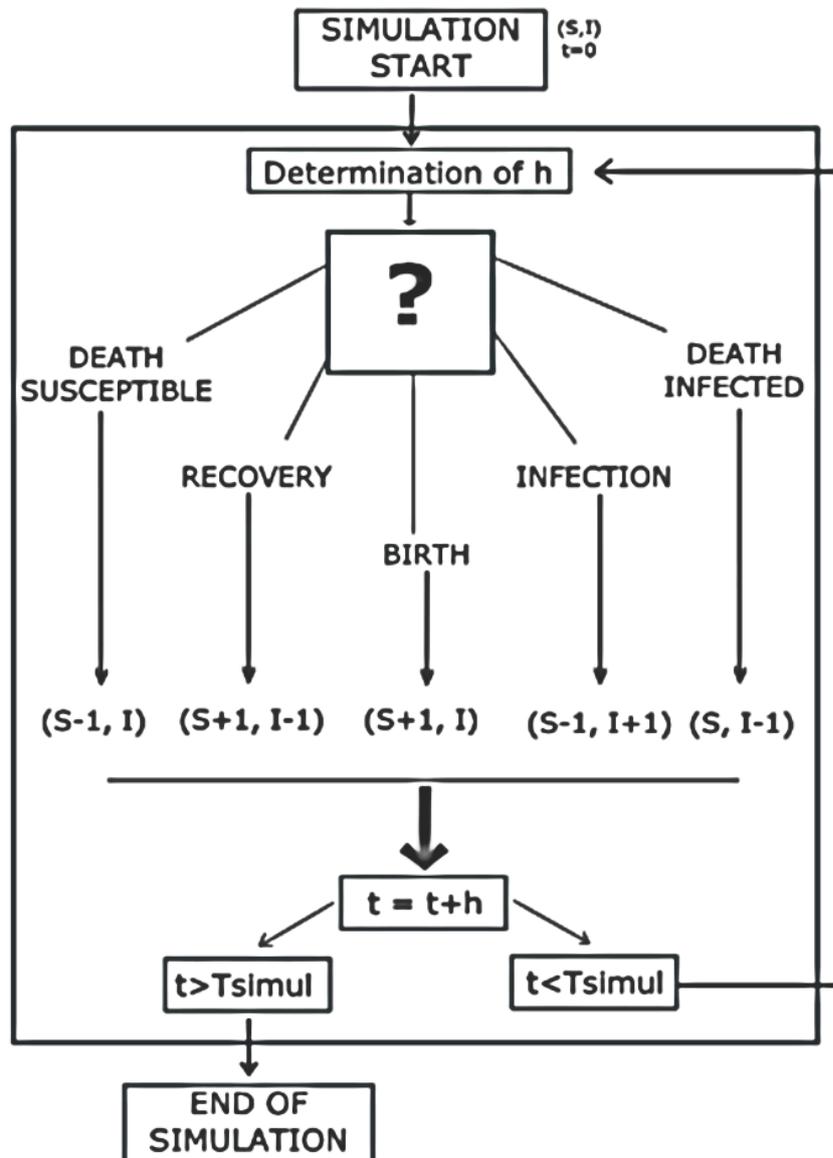


Figure 2.3: Representation of the different steps of the simulation of a stochastic epidemiological dynamics using the events presented in the Figure 2.2. At time $t=0$, the system is at a given state (S, I) . Then, according to a negative exponential law, the timelaps h between two events is determined along with the event associated with this timelaps. The number of individuals is implement in consequence as well as the time that is increase of h . Then, if the time is larger than the maximal time fixed for the simulation (T_{simul}), the program stops. If not, the simulation carries on with the determination of a new timelaps until the next event, etc.

2.2.3. THE VIRUSES CONSIDERED IN THE PHD WORK

a. The Feline Herpes Virus

The feline herpes virus (FHV) is a DNA virus from the *varicellovirus* gender. Closely related viruses from this gender have been observed in other carnivores such as dogs (*Canis lupus familiaris*, Papageorgiou et al. 2016) or in seals (*Phoca vitulina*, Bodewes et al. 2015). The FHV infection occurs mostly via nasal, oral and conjunctival mucous membranes and results, after a short period of incubation (2 to 6 days in average), first in an acute phase of disease during which most of the virus shedding happens then, in many cases, by cycles of latency/reactivations during which the virus can also be spread (Gaskell et al. 2007). The FHV is a well-known pathogen of domestic cats which has been observed in other felids such as the Cheetah *Acinonyx jubatus* (Munson et al. 2004) or the Lion *Panthera leo* (Hofmann-Lehmann et al. 1996).

The latency/reactivation cycles led us to consider the FHV as a virus which persists all along the lifetime. Thus, the FHV can be modelled using three compartments: Susceptible, Exposed and Infected. However, given the very short time of incubation compared to a lifetime, a SI compartment model was used.

b. The Feline Panleukopenia Virus

The Feline Panleukopenia Virus (FPV) is DNA virus which is observed in wild and domestic felids but also in minks (*Mustela vison*) and raccoons (*Procyon lotor*, Parrish 1995) and which appears to spread between these different hosts (Hoelzer and Parrish 2010). Though the FPV can be transmitted directly, a major part of its transmission is believed to happen indirectly given the high resistance of the virus in the environment (Hoelzer and Parrish 2010). In most cases, the individuals infected recover from the infection within ten days after the infection. Cats recovering from the infection can develop immunity (Scott and Geissinger 1997). Four compartments may thus be required to model the FPV epidemiological

dynamics: susceptible, infected, recovered and an environmental compartment (Berthier et al. 2000).

c. The Feline Calicivirus

The Feline Calicivirus (FCV) belongs to the family of the *Calicividae* which contains some major pathogens in Humans (viruses responsible for the gastroenteritis) and in wildlife (hemorrhagic syndrome of the Hare for instance). The FCV has been an important pathogen of felids for more than 40 years and is observed under different strains (Radford et al. 2007). The most common form leads to oral and upper respiratory tract disease from which individuals recover after two to three weeks. Some cats do not eliminate spontaneously the virus but these cases remain rare. Focusing on the main form of the virus, two compartments seem to be enough to model the FCV: susceptible and infected with a possibility of recovering to be susceptible again.

d. The Feline Immunodeficiency Virus

The Feline Immunodeficiency virus (FIV) is very close to the Human Immunodeficiency Virus (HIV) and causes an immunodeficiency syndrome close to one observed in humans with the same three steps: Acute phase, clinically asymptomatic phase and terminal phase. In felids however, the risks for opportunistic infections, neurological diseases and tumors are increased (Hartmann 2011). The mode of transmission is also different as bites and aggressive contacts are the main transmission path in felids which makes aggressive individuals more at risk (Sparger 1993). Given the absence of recovery from the FIV, a SI model, similar to the theoretical model presented above as an illustration, is usually used (Fouchet et al. 2009).

2.3. AREAS OF STUDY AND SAMPLING METHOD

2.3.1. NORTH-EASTERN FRANCE

From 2007 to 2011, wild and domestic cats were sampled in the French department of Meuse, in north-eastern France which corresponds to the historical area distribution of the wildcat in France (Léger et al. 2008). This part of France benefits from a degraded oceanic climate with hot summers (mean temperatures in July between 18 and 19°C) and cold winters (1.5°C in average in January), with significant falls of rain (800 mm per year in average).

The area of study corresponds to 400 km² where an alternating between small villages (about 25 in the area of study, between 30 and 600 inhabitants per village), agricultural fields and forests is observed (Figure 2.4). The mean density over the area is of 35 inhabitants per kilometer square, which is low compared to the national average of 103 inhabitants per kilometer square (INSEE, 2011). The high number of villages leads to the presence of numerous roads between them and, consequently, between forests, resulting in a fragmented environment. The fragmentation in this part of France remains however moderate with an effective size of natural environment mesh between 12 and 30 km² (CEMAGREF 2010).



Figure 2.4: landscape configuration in North-Eastern France with alternating of agricultural fields and forests – personal photography

Invasive sampling in north-eastern France

Except for 10 wild individuals which were collected dead on the roads, domestic and wild cats were trapped using baited two-door cages in order to collect blood (Figure 2.5). For domestic cats, the cages were hidden in farms and baited with sardines. For wildcats, all cages were hidden in the forest and baited with valerian. In both cases, once trapped, cats were weighted and put asleep using ketamine and Imalgène in quantity relative to the weight of the cat. Once the cat was asleep, blood was sampled from the jugular vein and different measurements were taken regarding the sex, the global body condition or the phenotype of the individual. Cats were then left in a calm place until they woke up. Both genetic and serological analyses were performed on the blood collected.



Figure 2.5: Two domestic cats trapped in a farm from North-eastern France –
Personal photography

2.3.2. FRENCH PYRENEES

From 2014 to 2017, the natural reserves of Jujols and Nohèdes from eastern French Pyrenees, were sampled. Indices of the presence have been observed for decades but this region of France has aroused little interest for scientific study. Natural reserves correspond to regulated territories which aim at protecting natural heritage such as flora, fauna, geological formations, etc.

Jujols and Nohèdes display elevations between 760 and 2459m and benefits therefore from a mountainous climate with Mediterranean influences. The summer (19°C in July) and winter (3°C in January) average temperatures are close to those observed in north-eastern France. Precipitations in this area of France are abundant with an average of 913 mm per year.

Nowadays, the presence of men is lower in the Pyrenees compared to north-eastern France, however, in the past, men pressure may have been far more important. The area of study we considered corresponds approximately to 325 km² and counts 10 villages which gather from 10 to 380 inhabitants, leading to an average density of 9.1 inhabitant per kilometer square. The landscape is not considered as fragmented (Figure 2.6, CEMAGREF 2010).



Figure 2.6: landscape configuration in French Pyrenees with small villages in the middle of continuous mountainous areas – personal photography

Non-invasive sampling in the Pyrenees

The status of the “Natural reserve” did not allow us to capture wildcats in the Pyrenees. Then, we used non-invasive sampling to gather genetic information. Feces were collected opportunistically across the area of study by agents of the natural reserves. Domestic cats from the villages in direct contact with the reserves were also sampled either collecting hairs or using oral smear. Only cats which had access to the outside, were born in one of the villages were sampled in order to assess the genetic profile of the local population of domestic cats, likely to have reproduced with wildcats. Only genetic analyses were performed on those samples.

3. Detecting hybrids, a methodological challenge

Foreword

The detection of hybridization is complex especially when the genetic differentiation between the populations which hybridize is low. Given the rise of studies focusing on hybridization, many methodological developments have been proposed to assess how individuals may be admixed between several populations.

The methods actually available for the detection of hybrids present several flaws and notably regarding (1) the comparability they allow between studies, and (2) their speed of execution.

In this chapter, I present two different methodological developments I worked on during my PhD. In the first part, I present a method which helps rationalizing the use of existing computer programs (STRUCTURE) and which, I expect, will help improving the comparability between studies dealing with hybridization in the future. In the second part, I present a new method. Because of its high speed and good efficiency, this method represents a valuable alternative for hybridizing studies especially in an industrial context where computing facilities are not as important as in universities.

3.1. Iterative determination of representative parents for detecting hybrids: application to wild and domestic cat populations



Domestic cat or ancient hybrid? – Personal photography

Abstract

Hybridizing events are increasingly observed across taxa and, in spite of their adaptive potential are mainly considered as a threat for the conservation of many species. Important methodological developments have been achieved to improve the detection of hybrid individuals, however, some aspects of these methods still need further investigation. This is particularly true when the detection of hybrids must be achieved without known reference individuals, in which case, a pool of representative parental individuals must be built from the dataset. The process by which this pool is built varies importantly across studies. Here, we describe an iterative algorithm which allows to define a representative pool using a rationalized approach. We show that this algorithm leads to a more representative pool of parental individuals compared to the building approaches usually found in the literature. The further classification of individuals either as parent, hybrid or suspect was similar according to the three methods. The application to a real dataset of wild and domestic cats allowed us to detect low rates of hybridization between the two sub-species. This algorithm represents a first step for the rationalization of the entire hybrid detection process in absence of reference individuals, which is expected to increase the comparability of studies on hybridization.

Keywords: *Felis silvestris catus*, *F s. silvestris*, computer simulations, threshold determination, STRUCTURE software, hybridization

BEUGIN Marie-Pauline, FOUCHET David, QUENEY Guillaume, ROUSSET François,
PONTIER Dominique (in preparation)

3.1.1. INTRODUCTION

Hybridization, defined as the reproduction of individuals belonging to two genetically distinct populations (Morgan-Richards *et al.* 2009) is recognized as a major evolutionary process involved in both extinction and adaptive dynamics (Arnold 1992, Melo-Ferreira *et al.* 2007; McDevitt *et al.* 2009; Monzón *et al.* 2014; Qi *et al.* 2014). It would happen in 25% of the animal species on average (Mallet 2005) and yet be sufficiently rare to allow for the persistence of differentiated species. Current studies tend to focus on hybridization as a threat for the conservation of species and, recently, it was shown that hybridization is promoted by global changes (anthropogenic modification of the environment, Oliveira *et al.* 2007; Galov *et al.* 2015; climate change, De La Torre *et al.* 2015; Taylor *et al.* 2015). Hybridization is then expected to increase in frequency in coming years. In line with these studies, the International Union for the Conservation of Nature (IUCN, 2016) has pointed out hybridization as the major threat for the persistence of a wide variety of species. Hence, assessing to what extent a population undergoes hybridization has become a central issue for the establishment of the conservation status of many species (Brennan *et al.* 2014) but also represents a major phenomenon to take into account in order to evaluate their adaptive potential (Hoffmann and Sgrò 2011; Sackman and Rokyta 2013).

Morphology often provides limited resolution for the detection of hybrids because they are cryptic or result from the hybridization of cryptic species (Stuart-Fox *et al.* 2009; Devillard *et al.* 2014). Genetics have then become the main approach for the study of hybridization (Barilani *et al.* 2007a, b; Oliveira *et al.* 2008; Olano-Marin *et al.* 2014; Vuillaume *et al.* 2015) and computer programs implementing powerful statistical methods have been developed to analyze the genotypes (Pritchard *et al.* 2000, Corander *et al.* 2006, Anderson & Thompson

2002, Alexander et al. 2009). The most popular computer programs are based on Bayesian inference. They differ in several ways: for instance, the type of markers used, the number of parental populations, the detection of hybrid categories (F1, F2, backcrosses...). The basic idea behind these programs is to infer, for each individual of the sample, the proportion of its genes that comes from the K different clusters (assignment probability expressed as q -values) included in the analysis. These clusters can be the two parental populations ($K=2$) or parental populations and categories of hybrids (F1, F2, backcrosses, etc) depending on the program used. The q -values range between zero and one, and these values sum to 1 for one particular individual. In an ideal situation, the q -values obtained for parental individuals and hybrids of different categories, would belong to distinct, separated distributions. However, because most of the time analyses rely on a limited number of genetic markers (Oliveira et al. 2008; Vuillaume *et al.* 2015) and concern hybridization between loosely differentiated species for which diagnostic markers are not always available (Barilani *et al.* 2007a,b; Godinho *et al.* 2011; Mengoni *et al.* 2015), these distributions tend to overlap widely. Interpreting such data is then challenging due the potentially complex distribution of different levels of hybridity among individuals. To deal with this complexity, an important decision is to determine optimal threshold values in order to establish for which assignment probabilities values an individual belongs most probably to a given cluster.

Several methods have been developed for the determination of this threshold (Le Roux *et al.* 2015; Godinho *et al.* 2011; Oliveira *et al.* 2007). A method commonly used, especially when no reference individuals (i.e., individuals for which we know if they are hybrids or not) are available, consists in selecting a limited number of individuals from the original dataset that had been unequivocally assigned to the parental populations to simulate F1 hybrids, F2 hybrids and backcrosses using

dedicated programs (Hybridize function Jombart 2008, HYBRIDLAB, Nielsen et al. 2006). This simulated sample is then analyzed to determine a threshold using either the highest assignment probability reached by a hybrid or the lowest assignment probability reached by a parent, using one of the available software packages (Oliveira *et al.* 2007; Barilani *et al.* 2007b; O'Brien *et al.* 2009; Hertwig *et al.* 2009; Godinho *et al.* 2011, 2015). The limited set of individuals initially selected is considered to contain reliable representatives of the parental species. However crucial, this first selection step varies from one study to another and the rationale behind the selection of parental individuals remains opaque. Some authors choose which individuals are representative using a threshold (0.9 in Oliveira et al. 2007; 0.95 in Bohling and Waits 2011). Any individual above it is then integrated in the representative pool of parents. Others choose the n individuals best assigned to each parental population (37.5% in Hertwig et al. 2009; about 65% in Oliveira et al. 2007). This proportion of selected individuals over the total number of sampled individuals varies a lot among studies and its determination is unclear. A general method is then required to determine an appropriate pool of representative individuals for each parental population.

In this study, we propose a new method to select the number of individuals representative of each parental population. Our method relies on an iterative algorithm that integrates step by step individuals in the parental pool. This algorithm can be applied to any assignment probability distribution generated by available computer programs. Our approach is illustrated and compared to existing methods for hybrid identification using (i) simulated datasets with different compositions in hybrids and levels of differentiation, and (ii) a real dataset composed of wildcats (*Felis silvestris silvestris*) and domestic cats (*F. s. catus*) from Northeastern France. Hybridization between these two sub-species has been

observed in several European countries (Beaumont *et al.* 2001; Pierpaoli *et al.* 2003; Oliveira *et al.* 2007; Hertwig *et al.* 2009; Nussberger *et al.* 2014; Le Roux *et al.* 2015, Beugin *et al.* 2016) and is presented as the major threat for the conservation of the European wildcat (Driscoll *et al.* 2007, Driscoll & Nowell 2010). Therefore, assessing the proportion of hybrids in wildcat populations is currently a topic of crucial interest.

3.1.2. MATERIAL AND METHODS

a. Defining representative parental individuals

In previous studies, the identification of thresholds for the detection of hybrids is based on the selection in one step of a pool of representative individuals from the original dataset according to different rules (e.g., the best n assigned parental individuals, use of a fixed threshold, Oliveira *et al.* 2007, Hertwig *et al.* 2009, Bohling and Waits 2011). With such an approach, only part of the individuals categorized *in fine* as parents are considered for the simulations. This might be a problem as not taking these individuals into account may lead to ignore part of the genetic diversity of the parental species. To overcome this limitation, we propose an iterative algorithm that allows at each step to integrate individuals above the threshold in the pool of representative parents and to continue iterations until the number of individuals integrated reaches a plateau. We chose to use the clustering computer program STRUCTURE (Pritchard *et al.* 2000; Falush *et al.* 2003; Hubisz *et al.* 2009) to determine the individual probabilities of assignment. Several reasons explain this choice: this program is widely used in hybridization studies with more than 25,500 citations referenced in Google Scholar in early 2017 for the three references. In addition, this program can deal both with microsatellite and SNP

data (Haasl and Payseur 2010). Furthermore, compared with other programs such as BAPS (Corander *et al.* 2003), STRUCTURE is less conservative, *i.e.* it detects more easily individuals as hybrids including true hybrids and thus detects more efficiently older generations of hybrids as hybrids (Bohling *et al.* 2013). Simulations with STRUCTURE were performed in the Linux Environment using the computing facilities of the CC LBBE/PRABI. In this study, only the case of hybridization between two parental populations is considered.

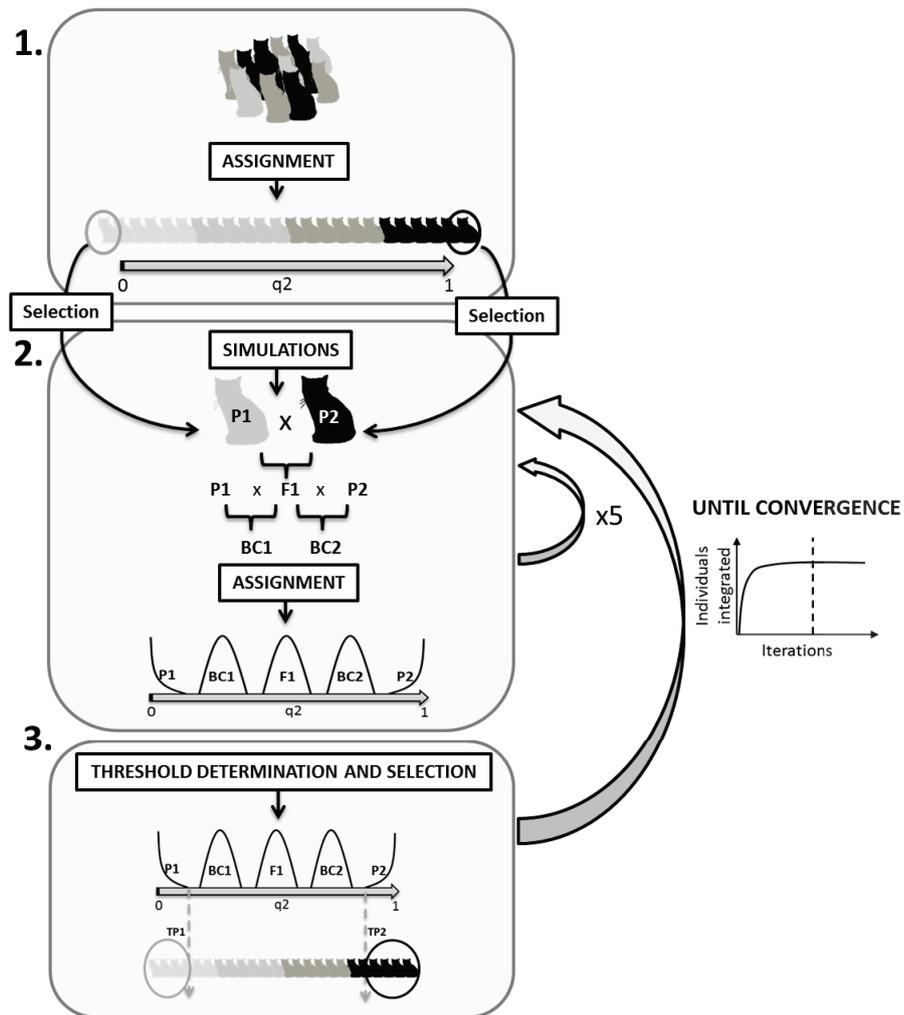


Figure 3.1.1: Description of the iterative algorithm. The figure represents the three main steps of the algorithm. The first frame corresponds to the initialization of the algorithm. Individuals from the studied population are analyzed with STRUCTURE in order to get q -values for each cluster. Then, a few are selected to be the first representative parents. The second frame correspond to the simulation part of the algorithm. Individuals previously selected are used to simulate more hybrids and parents for which we get q -values with STRUCTURE analysis. This step is repeated 5 times and all the q -values obtained are pooled together. At the end of these repetitions, thresholds corresponding to the lowest q -value reached by a parental individual are calculated at a 1% risk. Then, we go back to step 2, these new selected parents being used for the simulations. The step 2 and 3 are repeated until convergence, i.e., until a plateau is obtained, for the number of individuals integrated to the pool of representative parents. The dashed line on the convergence plot corresponds to the iterations taken into account to determine the final pool of representative parents.

b. Description of the iterative algorithm

The algorithm can be split into three steps (Fig. 3.1.1).

In the first step, the initialization step, individuals from the study sample are analyzed in order to obtain individual probabilities of assignment to each parental species as well as 90% credibility intervals for these probabilities. Calculations were carried out under the admixture model assuming correlated allele frequencies and $K = 2$, which corresponds to the assumption of two species contributing to the gene pool of the sample. No prior species information is used. Preliminary analyses allowed us to determine optimal lengths of burn-in and MCMC chains, *i.e.* length allowing for the convergence of the chain on the one hand, and a restrained calculation time on the other hand. They led to set a burn-in of 100,000 steps followed by 300,000 iterations. Individuals were then ranked according to the value of their probability of assignment in the parental species 1, the $\pi\%$ individuals best ranked were then selected as initial representatives of the parental species 1. The same step was performed in order to obtain representative individuals of the parental species 2. Three π values were tested: 1%, 20% and 50%.

This initial pool of individuals thus built at the end of step 1 was then used in step 2. Two hundred individuals from each parental cluster were simulated using the function *hybridize* from the package *adegenet* (Jombart 2008). As the composition of the dataset may influence the outputs of STRUCTURE (Bohling *et al.* 2013), we added simulated hybrids (200 F1, 200 F1*P1, 200 F1*P2 with P1 and P2 being the parental species 1 and 2) to the dataset in order to get parental q -values in presence of hybrids. All these simulated individuals were then analyzed using STRUCTURE with the same parameters as before in order to obtain the q -values for each of them. This second step was repeated 5 times in order to check for results

consistency; at each repetition, the q -values were stored along with the category of each simulated individual in order to be used in step 3.

In step 3, the q -values were used to define two thresholds that corresponded to the 1% quantile of the distributions of q -values for each parental species (TP1 and TP2 thresholds). The threshold values obtained were then used to select a new pool of representative parents. The representative parents were, in this case, all individuals for which the lower bound of the 90% credibility interval initially calculated in step 1 was above the threshold defined by the lowest value reached by simulated parents (TP1 and TP2). We chose to determine the threshold based on the mean probability of assignment provided by STRUCTURE and to select individuals based on the credibility interval in order to limit the integration of individuals in the parental pool at each iteration (Appendix, Figure S3.1.1.). Based on this new pool of representative individuals, we repeated steps 2 and 3 five times iteratively. The number of repeats was set in order to reach a plateau for the number of individuals integrated. The final pool of representative parents obtained at the end of the iterations could then be used to further categorize individuals.

c. Comparison with two other methods

The pool of representative parents was also determined using two methods widely used in the literature. Both rely on the probability of assignment calculated for the initial sampled population (thus corresponding to the results obtained at the end of the step 1 of the algorithm). The first method consists in selecting individuals based on their q -value. A threshold commonly used to differentiate hybrids and parents is 0.9 (Lancaster *et al.* 2006; De La Torre *et al.* 2015; Le Roux *et al.* 2015; Guildea *et al.* 2015). Following the method used in Oliveira *et al.* (2008), representative parents were then defined as the individuals presenting both a mean q -value and the lower bound of credibility interval above this threshold

value. The second method consists in selecting the n individuals best assigned to each cluster. The proportion of individuals chosen varies among studies with authors selecting from 20 to 75% of the sampled dataset to build the pool (Oliveira *et al.* 2007, 2008; O'Brien *et al.* 2009; Hertwig *et al.* 2009). Here, we chose for the comparative study to establish the representative pool using π values of 1%, 20% or 50% best assigned individuals in each population, similarly to what was done with the iterative algorithm.

d. Detection and classification of hybrids

Once the representative pool of parents was built, individuals (parental individuals and hybrids of category F1, F2 and backcrosses of first generation) were simulated for each method using the function *hybridize* from the package *adegenet* (Jombart 2008). Two hundred individuals were simulated for each category and the simulated dataset was analyzed with STRUCTURE program using the same parameters as in the algorithm. The probabilities of assignment obtained as output were then used to define four thresholds. As in the implementation of the iterative algorithm, the thresholds were calculated with a 1% risk. The two first thresholds corresponded to the 1% quantile of the distributions of q -values of the simulated parental individuals, for each cluster. The two other thresholds corresponded to 99% quantile of the distribution of q -values of first generation backcrosses. Consistently with the denominations used in the algorithm description, the first two thresholds will be thereafter denoted as TP1 and TP2, and the last two as TH1 and TH2. These four thresholds allowed for the definition of zones allowing for the classification of individuals as either parent, hybrid or suspect. The zone corresponding to suspect individuals, hereafter called grey area, included all the individuals presenting a q -value that could belong to several distributions (Appendix, Figure S3.1.2). Individuals were assigned to one of these zones either in

a conservative way (to allow more easily an individual to be detected as parent; conservative approach), considering their mean probability of assignment to the clusters, or in a more relaxed way (to allow more easily an individual to be detected as hybrid; relaxed approach), considering the lower bound of the credibility interval.

e. Application to datasets

Sampling and genotyping of cat populations in North-Eastern France

The study took place in Northeastern France between April 2008 and May 2011. Three hundred and seventy-one domestic cats and 42 presumptive wildcats were captured and measured according to the protocol described in Beugin *et al.* (2016). The fieldwork was conducted by qualified people according to French legislation. Accreditation has been granted to the UMR-CNRS5558 (accreditation number 692660703). Cats were classified as domestic or wild based on morphologic criterions (Léger *et al.* 2008; Devillard *et al.* 2014).

All cats were genotyped using 22 microsatellites amplified in three multiplexes (Appendix, Table S3.1.1). DNA extraction was performed using a purification column kit (Nucleospin 96 Tissue kit, Macherey-Nagel) following the manufacturer protocol. PCR reactions were performed step by step following a unidirectional workflow starting in a clear room with positive air pressure where sensitive reagents, enzymes and primers, were prepared. Then, DNA and reagents were assembled in a pre-PCR room. PCR amplifications happened in 96-well microplates in a post-PCR area with negative air pressure. The PCR reaction occurred in a final volume of 10 μ l that contained 5 μ l of Mastermix Taq polymerase (Type-it, QIAGEN), 1.35 μ l of primers pairs at a final concentration comprised between 0.08 and 0.6 μ M, and 30ng of DNA. Each pair of primers was coupled with a fluorescent dye. The reaction started by a denaturation step at 95°C for five

minutes. This step was followed by thirty PCR cycles (denaturation step = 95°C, 30s; annealing step = 55.9°C, 90s; elongation step = 72°C, 30s) and a final elongation step at 60°C during 30 minutes. PCR products were resolved on a capillary sequencer ABI PRISM 3130 XL (Applied Biosystem) in denaturing conditions (formamide) and an internal size marker in one migration for each multiplex. All these steps were performed using filtered tips. Finally, the electrophoregrams were analyzed using GENEMAPPER 4.1 (Applied Biosystem/Life Technologies) independently twice. Results were then confronted and ambiguous loci were set to missing data. The resulting genotypes were used to check for Hardy-Weinberg equilibrium and linkage disequilibrium using FSTAT 2.9.3.2 (Goudet 1995), respectively, with a 5% Bonferroni correction. The software Genetix 4.05.2 (Belkhir *et al.* 1996) was furthermore used to estimate Weir & Cockerham's F_{st} (1984) between domestic and wildcats. The frequencies of null alleles in both sub-species were calculated according to Brookfield (1996) and the statistical significance of these frequency assessed using a binomial exact test following De Mêus *et al.* (2002). Finally, we estimated the rate of migration per generation using BAYESASS 3.0.3 (Wilson and Rannala 2003), i.e., the proportion of gene flow from domestic cats to wildcats and from wildcats to domestic cats. We run BAYESASS with a MCMC chain length of 5,000,000 after a burn-in period of 1,000,000, with a sampling interval of 2,000. Our iterative algorithm was run three times to check for results consistency.

Simulated datasets

The simulated datasets fulfilled three goals. First, we assessed whether the iterative algorithm converged to a final pool of individuals truly representative of the parental populations whatever the value of π at initialization. This convergence was assessed based on the composition of the final pool of representative parents

both in terms of F_{ST} (according to Weir and Cockerham 1984) and individuals integrated in the pool (number and hybrid category). Second, we evaluated the reliability of the results obtained with the iterative algorithm for the real dataset compared to those obtained with the two other methods. For this purpose, the simulated datasets included several of the main characteristics of the real dataset: number of loci (22 loci), polymorphism (5 alleles per locus) and level of differentiation (F_{ST} close to 0.17, using Weir & Cockerham's estimator, see Results section) and the relative proportion of the two populations. Thirdly, we evaluated to what extent the iterative algorithm provided results comparable to the ones obtained with the *threshold* and the *n best* methods in terms of percentage of correct identifications, type 1 error (proportion of parents categorized as hybrids and reciprocally) and composition of the grey area. For this aim, we generated additional datasets with F_{ST} equal to 0.30 and 0.5 to widen the situations of genetic structure considered. Also, we compared the thresholds obtained with the iterative algorithm and the *n best* method for the three values of π at initialization, as well as the resulting categorization of individuals, in order to assess the final impact of the π value.

For computation time purposes, we simulated only a limited number of datasets and could not explore the whole range of dispersal and mutational models classically proposed by simulations programs (e.g. QuantiNEMO, Neuenschwander et al. 2008). Given the complexity of the domestic cat genetic structures, it was difficult to select some "best" model from these computer programs. We chose then to simulate datasets using a simple approach that does not use coalescence and does not require the selection of a particular model, and that allows to build many different genetic structures, possibly including structures reachable with classical models. The simulated datasets were built as follows: for each locus, the allelic

frequencies of the A ($A=5$, see Results section) alleles were drawn from a Dirichlet distribution. Then, each allele from each locus and each individual, was randomly sampled among the A possible according to the obtained frequency for the population. Only simulated datasets with a F_{ST} value equal to the expected F_{ST} (equal to 0.17 as in the real cat populations, or 0.3 and 0.5) were used for further analyses. For each simulated dataset of parental individuals, we generated hybrid individuals using the function *hybridize*. Two different scenarios, corresponding to two hybrid distributions, were analyzed. In the first one called “*Ideal*” scenario, we simulated only first generation backcrosses as hybrids in the proportion obtained for the real dataset (using the relaxed approach). In the second scenario called “*Continuum*” scenario, we simulated hybrid categories up to the third generation of backcross; the proportion of first generation backcrosses being identical to the proportion observed for the first scenario. Given the number of markers and the F_{ST} value between wild and domestic cats (see Results section), and the fact that we considered only first generation backcrosses for the determination of the thresholds in our approach, we did not expect to detect hybrids passed the first generation of backcrosses (Vähä and Primmer 2006). The aim of these two scenarios was to test the impact of further generations of hybrids on the detection of first generation backcrosses. The iterative algorithm was run three times for each simulated dataset in order to test for consistency across repeats. The impact of the value of π at initialization, of the scenario and the repeat on the composition of the representative pool was tested using a mixed linear model with the simulated dataset (60 datasets simulated) as random parameter. In addition, we assessed the percentage of divergence between two representative pools as the percentage of individuals that were not shared by the different representative pools compared.

3.1.3. RESULTS

a. Population structure and hybrid detection in cat sample

All the 22 microsatellites were polymorphic in the 42 presumptive wildcats and 371 domestic cats. Neither the wildcat nor the domestic cat populations presented any deviation from Hardy-Weinberg equilibrium nor did they show any linkage disequilibrium or significant null allele frequencies. Overall, domestic cats displayed a higher number of alleles per locus than wildcats (mean = 11.77, sd = 0.69 vs. mean = 6.91, sd = 0.43, respectively). However, when considering only alleles observed with a frequency higher than 5%, the mean number of alleles per locus was similar between the two subspecies (mean = 4.82, sd = 0.23 vs. mean = 4.50, sd = 0.27, respectively). The number of private alleles was higher in domestic cats than in wildcats, however, relative to their sample sizes, the wildcats presented more private alleles than domestic cats with a frequency higher than 5%. A significant genetic differentiation between the two sub-species was found (Weir and Cockerham's F_{ST} = 0.154; 90% confidence interval: 0.119-0.196). Finally, we did not detect any significant gene flow from domestic cats to wildcats (m = 0.0009; 95% credibility interval between 0 and 0.0034) while we found a low but significant gene flow from wildcats to domestic cats (m = 0.0077; 95% credibility interval between 0.00019 and 0.028).

As expected, the iterative algorithm reached a plateau for the number of individuals integrated in the final representative pool of parents (Appendix, Figure S3.1.3). Thirty-six individuals were systematically included in the representative pool of wildcats whatever the repeat or the value of π at initialization. An additional individual was integrated in a single repeat. Regarding the domestic representative pool, we found between 0 and 0.2% of divergence between repeats, which correspond to a difference of one individual. Depending on the value of π at

initialization, the representative pool counted between 342 and 349 individuals, the largest number of integrated individuals being obtained with the π equal to 20% and 50%. The conservative approach allowed us to detect 2 hybrids among the morphologically domestic cats and none among the morphologically wild population. Using the relaxed approach, 16 and 6 hybrids were detected, respectively. This would correspond to rates of hybridization between 0 and 14% in the wildcat population, and between 0.5 and 4.3% in the domestic population. Both for domestic and wildcats, the number of individuals used for the initialization did not change the final number of individuals detected either as parent, hybrid or suspect while it impacted strongly the results obtained with the “*n best*” method (Figure 3.1.2, Appendix - Figure S3.1.4). The “*n best*” method and the iterative algorithm led to similar results when the initialization was made with the 50% best individuals in the “*n best*” method. The threshold method also yielded comparable results to what was obtained with the iterative algorithm but recognized only five hybrids, the sixth individual being categorized as suspect (Appendix, Figure S3.5).

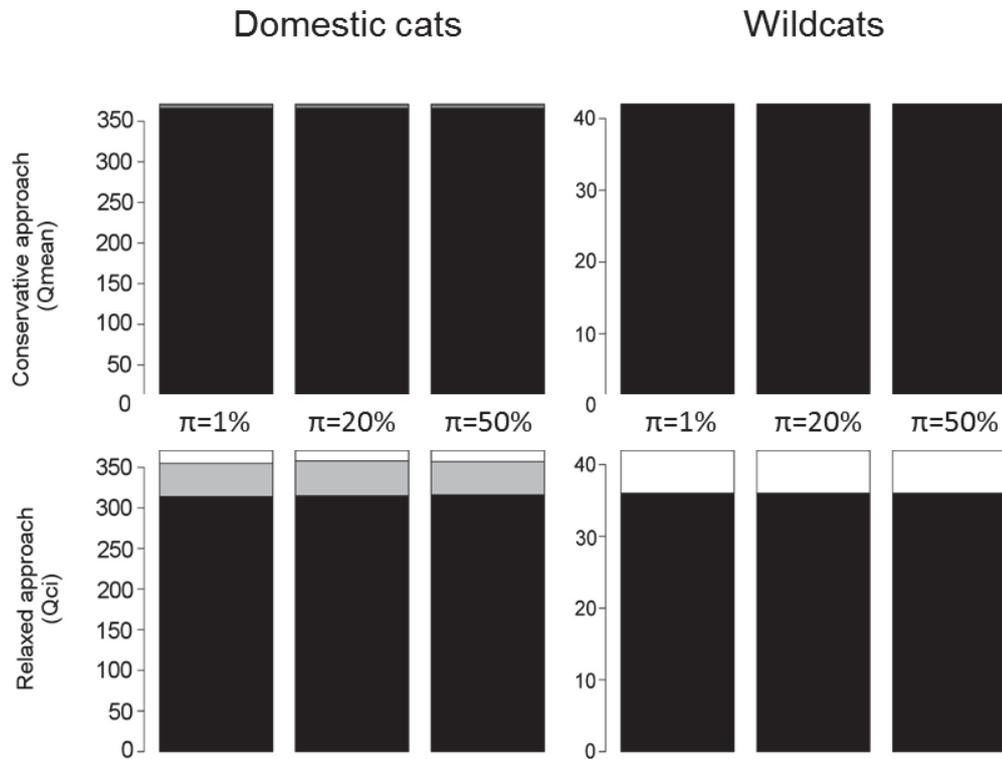


Figure 3.1.2: Number of parents (in black), hybrids (in white) and suspect individuals (in grey) detected in wild and domestic morphologically determined populations using the iterative algorithm either with a π value at initialization of 1%, 20% or 50%. Q_{mean} is the mean q -value of individuals and Q_{ci} is the low boundary of the CI 90%.

b. Simulations and performance of the iterative algorithm

The sample sizes simulated for each category of individuals and each scenario ('*Ideal*' and '*Continuum*') are given in Table 1.

Scenario	P1	P2	1st	2nd	3rd	F_{ST}	A	Large/ Small
Real dataset	42	371	6/15	0	0	0.154	4.5/4.82	8.83
<i>Ideal</i>	110	1070	20/50	0	0	0.155/0.274/0.465	5/5	8.61
<i>Continuum</i>	70	970	20/50	20/50	20/50	0.146/0.259/0.442	5/5	8.61

Table 3.1.1: Sample sizes simulated for each category of individuals. P1 represents the smallest parental population (wildcats in the real dataset) and P2 the largest (domestic cats in the real dataset). 1st, 2nd and 3rd are the three generations of backcrosses which can be simulated depending on the scenario. The F_{ST} value given is the mean of the values obtained over the different simulations corresponding to each scenario and each level of F_{ST} tested. A is the mean number of alleles per locus for small population/large population. The last column is the ratio between the number of individuals from P2 over the number of individuals from P1.

We found a plateau in both scenarios (Figure S3.1.6 in supplementary material). Lower values of π at initialization led to more differentiated pools of representative individuals, especially in the *continuum* scenario ($\chi^2 = 30.96$, $df = 1$, $p < 0.001$). However, all F_{ST} values remained very close to the expected values of 0.17 (0.174 on average for *ideal* scenario, 0.172 for *continuum*), 0.3 (0.303 on average for *ideal* scenario, 0.283 for *continuum*) and 0.5 (0.506 on average for *ideal*

scenario, 0.473 for *continuum*). In addition, less individuals were integrated in the pool for low values of π , again especially with the *continuum* scenario ($\chi^2 = 9.01$, $df = 1$, $p = 0.0027$). However, at least 91% of the true parental individuals were integrated in the pool on average whatever the value of π at initialization, and less than 1% of the representative pool corresponded to hybrids. Finally, the repeat did not impact significantly the differentiation of the pool. The method showed thus a high repeatability. We assessed that the final pools of representative parents diverged by 4.5% (sd = 0.032) between cases initialized with $\pi = 1\%$ on the one hand, and $\pi = 20\%$ and $\pi = 50\%$ on the other hand, these two last initial conditions diverging in an order of magnitude similar to what was observed between repeats starting with the same value of π (divergence by 1.1%, sd = 0.088). This state with a majority of parents integrated in the final representative pool of individuals led to similar thresholds and categorization whatever the value of π (Figure S3.1.9 and S3.1.10, supplementary material). Then, although the composition of the representative pool changed depending on the value of π , the iterative algorithm seemed to converge *in fine* to similar results whatever the initial conditions.

These results contrasted strongly with what was obtained with the '*n best*' method for which the value of π at initialization had a major impact on the categorization. The lowest π value led indeed to a worse recognition of parents as such (Mann Whitney tests: $p < 0.0001$ both for conservative and relaxed approaches when comparing $\pi = 1\%$ with $\pi = 20\%/50\%$; $p = 0.048$ with a unilateral test on the conservative approach and $p = 0.0013$ with a bilateral test on the relaxed approach $\pi = 20\%$ and $\pi = 50\%$) and, in parallel, to a higher proportion of parents recognized as hybrids (Figure 3.1.3, Figure S3.1.10 of the appendix) compared to initializations with $\pi = 20\%$ and $\pi = 50\%$. This can be explained by the fact that, contrary to what was obtained with the iterative algorithm, the threshold values are deeply

impacted by the initialization value and especially between $\pi = 1\%$ and $\pi = 20\%$ and 50% (Figure 3.1.4). Such difference can, in turn, be directly linked to the representative pool of parents selected using this method which changes drastically in terms of number of individuals integrated by construction, but also in terms of F_{ST} value ($\chi^2 = 134.52$, $df = 1$, $p < 0.001$) with a mean size effect of 0.86 ($CI_{95}=0.78-0.94$) which was stronger than the effect size obtained with the iterative algorithm (0.999 with a CI_{95} of 0.993-1.005). Using the '*n best*' approach, we reached on average 0.26 for an expected F_{ST} value of 0.17, 0.38 for an expectation of 0.3 and 0.6 for an expected value of 0.5. Similar categorization results could be obtained with the iterative algorithm and the '*n best*' method for $\pi=50\%$ for this last method. We chose to consider this π value of initialization for the '*n best*' method to compare it to the '*threshold*' method along with the iterative algorithm as it seemed the most efficient initial condition for this method. We found that the final categorization was quite similar between methods (Figure 3.1.3). However, parental individuals were more likely to be categorized as hybrids with the '*n best*' method compared with the '*threshold*' method (Mann-Whitney test, $p < 0.001$) and the iterative algorithm (Mann-Whitney test, $p < 0.001$) both with the relaxed and conservative approaches, while less hybrids were categorized as suspects with the '*n best*' method ($p = 0.002$ for both methods) with the conservative approach. There was no significant difference for any other aspect of the categorization.

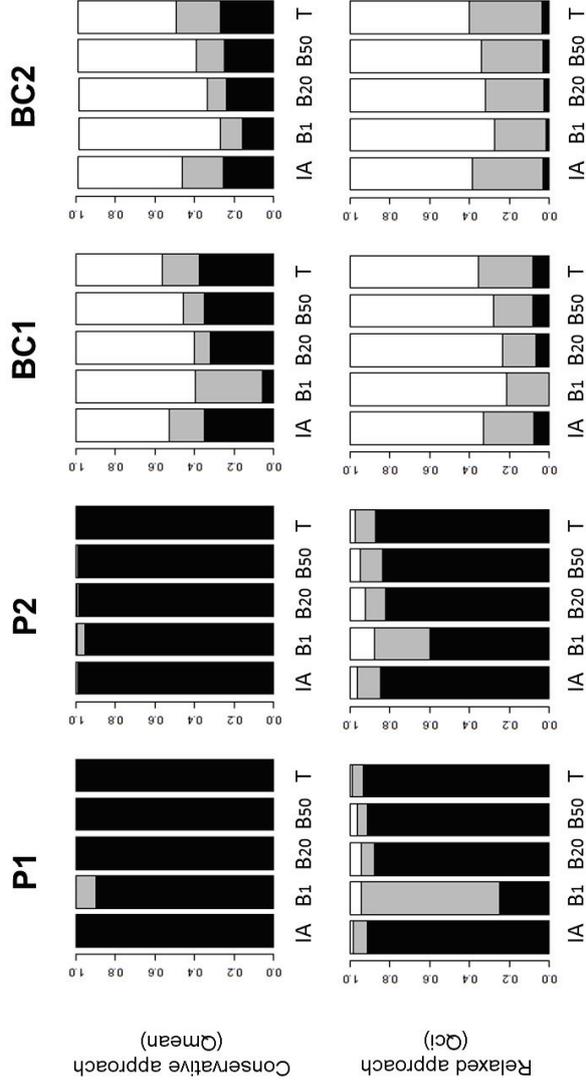


Figure 3.1.3: Average proportion over all simulations of parental individuals (P1 and P2) and first generation backcross individuals (BC1 and BC2) categorized as parent (in black), suspect (in grey) and hybrid (in white); either using the conservative or relaxed approach and for the different methods of selection of the representative pool (IA = Iterative Algorithm, B₁, B₂₀ and B₅₀ = *n Best* approach with π values at initialization of 1%, 20% and 50% respectively, T = Threshold approach).

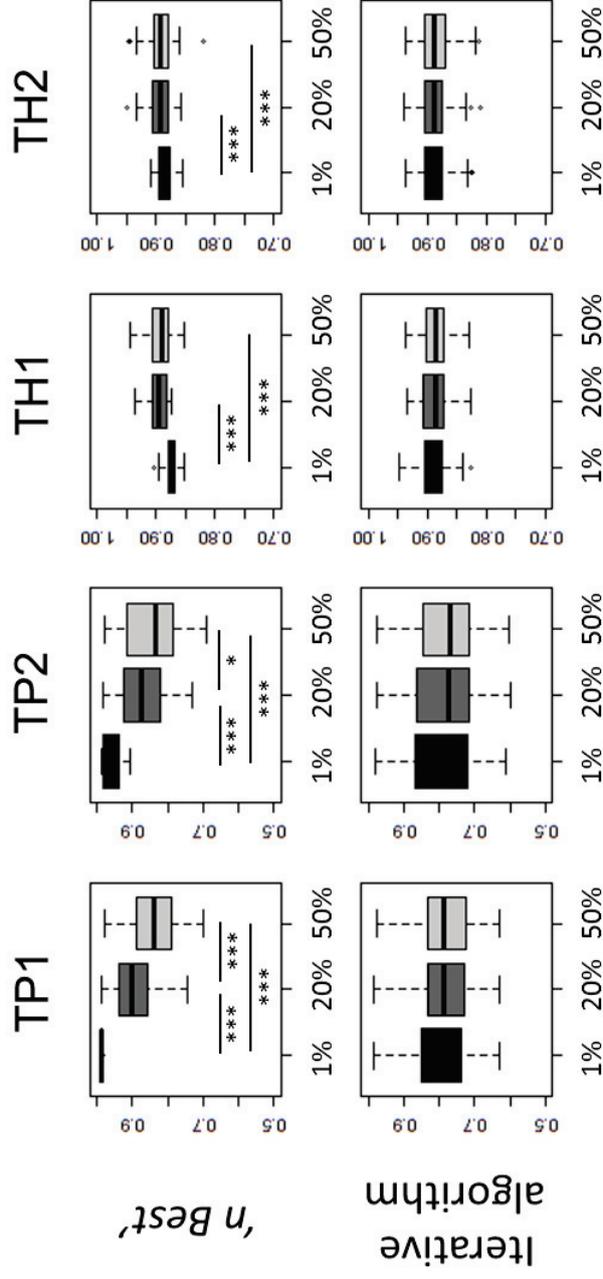


Figure 3.1.4: Impact of the π value at initialization on the thresholds 'values (TP1, TP2, TH1 and TH2) calculated with the 'n best' method (on top) or the iterative algorithm (at the bottom). Only significant differences are represented (***) corresponds to p -values lower than 0.001, ** to $p < 0.01$ and * to $p < 0.05$).

The composition of the grey area depended on the method used although the impact of the method remained marginal with low effect sizes (effect sizes below 1.5). On average, hybrids and parental individuals were equally represented in the grey area. The *Continuum* scenario led however to lower proportions of parents in the grey area ($\chi^2 = 1436.32$, $df = 1$, $p < 0.0001$). The proportion of parental individuals in the grey area did not depend on the differentiation of the simulated population nor the value of π at initialization for the iterative algorithm. On the contrary, more parental individuals were present in the grey area for high differentiation levels ($\chi^2 = 8.66$, $df = 1$, $p = 0.0033$) and low values of π at initialization ($\chi^2 = 11.36$, $df = 1$, $p = 0.00075$) with the ‘*n best*’ method.

3.1.4. DISCUSSION

This study extends previous work that has evaluated the power of approaches relying on microsatellite markers to assign individuals to hybrid categories. Here, starting with a particular situation where populations are weakly differentiated and genotyped at a substantial number of loci ($n = 22$) to more differentiated populations, we proposed an iterative algorithm which allowed for the selection of a representative and realistic pool of representative parental individuals. The constitution of such pool is expected to lead to a better assessment of hybridization rates and of the uncertainty attached to the calculation of these rates as well as a better comparability between studies.

a. Efficiency of the algorithm to build a realistic and representative pool

We showed that the iterative algorithm allowed building representative pools made mostly of parental individuals. According to Vahä and Primmer (2006), the correct identification of first generation backcross would require 48 markers for an

F_{ST} (calculated according to Weir and Cockerham 1984) value of 0.21 while the detection of F1 would require only 12 loci for the same F_{ST} . Thus, we did not expect the 22 microsatellites used here to be powerful enough to detect all generations of backcrosses individuals or even to detect all first-generation backcrosses. The limited integration of backcrossed individuals in the representative pool may then result directly from the discriminatory power of the microsatellite panel used and not from the methodology itself. Such a conclusion is supported by the results of Oliveira (2012) as first generation backcrosses between domestic and wildcats in the Iberian Peninsula could not be correctly identified (from 12 to 14% of backcrosses identified as parental individuals) using a panel of 38 microsatellite markers for an F_{ST} value of 0.2. In addition to being relatively hybrid-free, the representative pool obtained with the iterative algorithm encompassed most of the individuals which can be identified as parents, and thereby, it contained most of the genetic diversity of the two parental populations and presented a realistic differentiation level between parental populations. Regarding the detection of hybrids, we found that the three methods compared led to very similar results for the simulated datasets we considered in this study. The main advantage of the iterative algorithm is its rationality and transferability between studies. It can be started with a very small, statistically irrelevant, number of individuals and lead to the same result as would be done with a larger number of initializing individuals unlike the '*n best*' method. This is still an arbitrary step but this allows to initialize the algorithm with the best assigned individuals to each cluster instead of taking into account the *n* best assigned.

Finally, we chose to explicitly take into account the uncertainty of the classification by considering a suspect class. The benefit of this approach appears particularly when the detection of hybrids relies on the credibility interval and thus

on a relaxed approach. Such approach would indeed, without the grey area, lead to an over-detection of hybrids but a very high confidence in a parental class. The addition of a grey area, leads to the optimization of both the confidence in a parental and a hybrid classification. Furthermore, it allows to assess a range of plausible values for the rate of hybridization considering all suspect individuals as parent for the lower value and all suspects as hybrids as a upper value.

b. Hybridization between wildcat and domestic cats in northeastern France

The classification as hybrid by the iterative algorithm appears reliable according to the simulation study. The rates of hybridization of the cat populations studied reach values up to 14.29% for the morphologically wildcat population, and at least (as hybrids can be present in the grey area) 4% in the morphologically domestic population. The overall rate of hybridization (5%) corresponds to the lower range of values observed in neighboring countries (from 2-8% in Italy Randi et al. 2001 and Lecis et al. 2006, 3-18.4% in Germany in Eckert et al. 2010 and Hertwig et al. 2009, 14% in Portugal Oliveira et al. 2008, 25-31% in Hungary in Lecis et al. 2006 and 45% in Scotland Beaumont et al. 2001 based on 8 to 21 microsatellites and using the *n best* approach for Hertwig et al. 2009, Oliveira et al. 2008). Even when considering all suspect individuals as hybrids, the overall rate of hybridization we obtain (13.8%) corresponds to moderate intensities of hybridization. Compared with estimations obtained from French samples, the rate of hybridization we found is surprisingly low (25% on average in Germain *et al.* (2008,2009), O'Brien *et al.* 2009, Say *et al.* 2012). A possible explanation for this difference may be linked to different sampling strategies. While our sampling is local, focused on one population of living wildcats and on neighboring populations of living domestic cats, previous studies in France relied upon opportunistic

sampling of road-killed animals, over a much larger area. (Germain *et al.* (2008, 2009) showed that hybrids tend to live in intermediary environments, between forests and villages. This would expose them to road mortality more often than wildcat living in the forest or domestic cats in the villages and thus, sampling schemes based on road-killed animals may be biased towards a higher proportion of hybrids. Nussberger *et al.* (2014b) tested the impact of such different sampling strategies and did not find any significant difference in the rate of hybridization obtained either with a systematic or an opportunistic sampling. However, this comparison was led between hairs non-invasively sampled and carcasses while in this study, almost all wildcats were trapped (10 road-kills on 42 wildcats). Then, it would be interesting to see if we would have retrieved a difference when the systematic sampling consisted of trapping.

Another surprising result, especially regarding the simulation study outputs, is the total absence of suspect individuals in the wildcat population. This result may be due to the unbalanced sampling between wildcats and domestic cats, which is known to generate biases in STRUCTURE (Puechmaille 2016). We thus expected in the domestic cat population more individuals classified in the suspect category due to a wider range of q -values. This was observed in most simulations, and, interestingly, a suspect zone was always detected in the simulations for both populations. To go further, we resampled the dataset several times in order to obtain similar population sizes for the two cat populations. These resampled datasets ($n = 10$) were then analyzed in order to classify individuals into parent, hybrid and suspect. While the number of suspect individuals did not differ based upon the mean probability of assignment, we found a significantly higher number of suspect individuals in the domestic population (Mann-Whitney Test, $w = 49$, $p = 0.00074$) when the classification was carried out with the relaxed approach. If the

unbalanced sampling size has an impact, it is not the only reason for the observed difference in the proportion of suspect individuals between wild and domestic cat populations. This could be due to the fact that the domestic population is more introgressed than the wild one, thus the presence of suspect individuals is a manifestation of this introgression. This hypothesis is strengthened by the fact that the only weak but significant gene flow in our system was from wildcats to domestic cats. A similar gene flow was reported by Hertwig et al. (2009) in eastern Germany, in a fragmented area notably marked by extensive cultivated landscapes. Our results are consistent with the hypothesis of asymmetrical hybridization formulated in Beugin et al. (2016) for this particular population of wildcats. In this study, eighteen of the wildcats were equipped with radio-collars in order to investigate the spatial organization of the population. The authors showed that the population seemed to be organized with a core of wildcat females surrounded by wildcat males. It was then proposed that this organization may promote hybridization between wild males and domestic females and thus, introgression would indeed be expected in higher proportions in the domestic population.

c. Future developments

The simulated datasets we generated presented one major difference compared to real datasets, they did not present any private allele. Although this may introduce biases, we would expect them to lead to lower distinction efficiency of hybrid and parents, and we have no reason to think that this would change the relative performance of the three methods. Currently, the iterative algorithm method rests upon STRUCTURE program, which requires a substantial computation time (hours are necessary for a single analysis using the iterative algorithm). Although the algorithm can be adapted to any program delivering assignment probabilities (NEWHYBRIDS, BAPS), this time limitation remains as

none of these programs run quickly and as efficiently as STRUCTURE for now (pers. Observation, Bohling et al. 2013). The adaptation of the algorithm to faster approaches would permit to consolidate current results given the low but existing variability between repetitions, by allowing a higher number of repetitions. Our algorithm is also expected to help clarifying the process of defining thresholds for the classification of individuals. However, more work is still needed to complete its rationalization. In particular, the simulation step during which individuals of different categories are simulated is a key step which would need more attention. Some authors have indeed mentioned that the composition in hybrids of the dataset changes the output of STRUCTURE (Bohling et al. 2013). This raises the question of how many individuals we should simulate for each category and if some proportions should be observed for these simulations.

3.1.5. CONCLUSION

The rationalization of the building of the representative pool proposed here represents an important first step towards a rationalization of the entire process of threshold determination and a better comparability between studies on hybridization. We detected substantial on-going hybridization between wild and domestic cats. However, the question as to know whether this hybridization is a threat for the conservation of the wildcat population has to be debated taking account that the core of the forest does not appear affected by this genetic mixing up. These results make all the more pressing to conduct more studies at a local scale to investigate whether such pattern, with a central nucleus of the population apparently protected from external influences, is observed elsewhere. Indeed, while studies assessing hybridization rates have been multiplying in wildcats, they have not been associated with spatial and behavioral data. We believe that to

further understand hybridization between cat sub-species and in other species, studies combining different approaches are strongly needed.

AKNOWLEDGEMENTS

MPB, DP and DF are supported by the LabEx ECOFECT (ANR-11-LABX-0048)

3.1.6. APPENDICES

Table S3.1.1: Distribution of the 22 microsatellite markers in the different multiplexes.

Multiplex 1	Multiplex 2	Multiplex 3
Fca126 Fca675	Fca8 Fca26 Fca45 Fca58 Fca96 Fca124 Fca577 Fca668	Fca69 Fca105 Fca201 Fca220 Fca229 Fca293 Fca310 Fca453 Fca678 Fca75 Fca149 Fca441

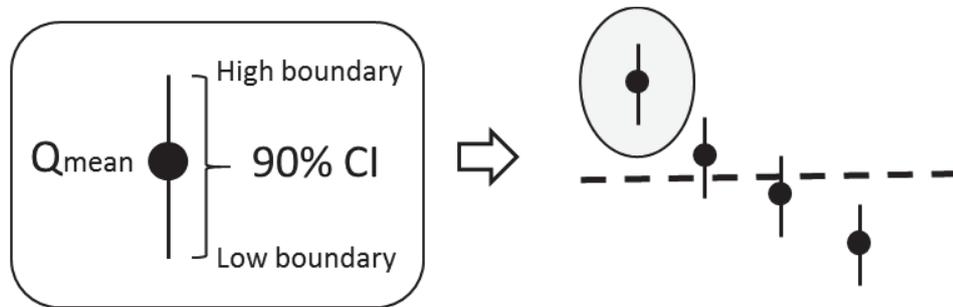


Figure S3.1.1: Illustration of the conditions required for individuals to be integrated in the representative pool. The dashed line represents the threshold determined at the considered iteration and the four points correspond to the four positions possible regarding this threshold for individual assignments. The only position retained to be integrated in the representative pool is the situation circled in grey where both the mean probability of assignment and its 90% credibility interval lowest boundary are above the threshold.

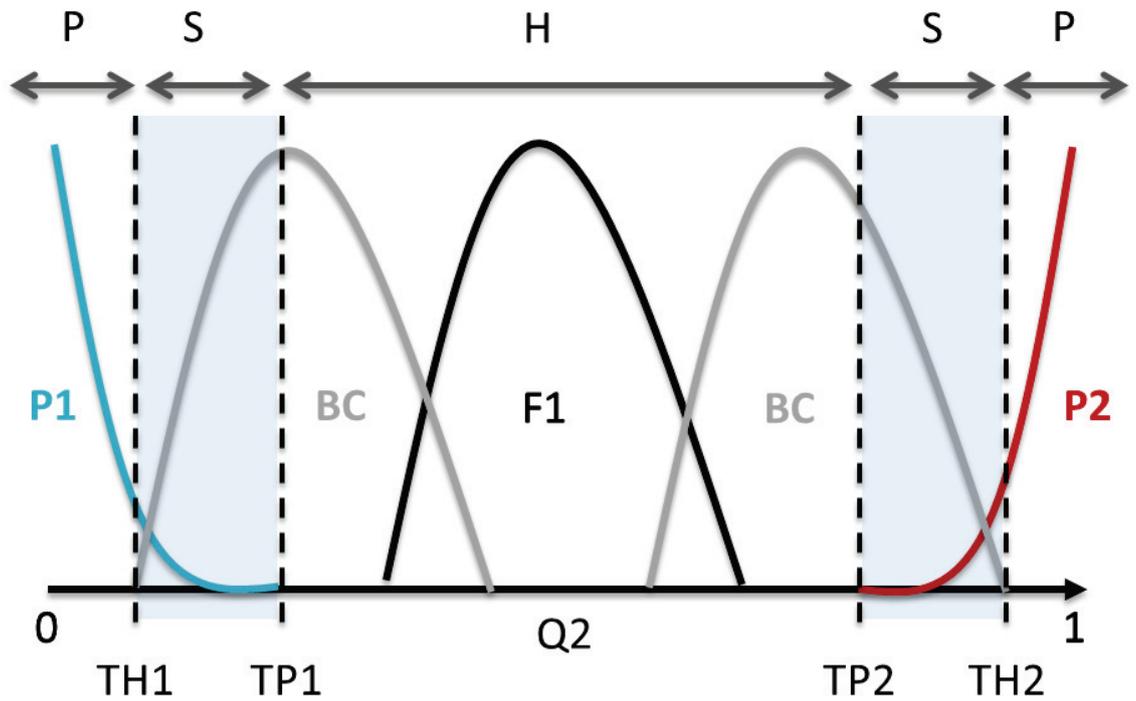


Figure S3.1.2: Representation of the different categories (P = Parent, S = Suspect, H = hybrid) according to the q -values distributions and thresholds obtained.

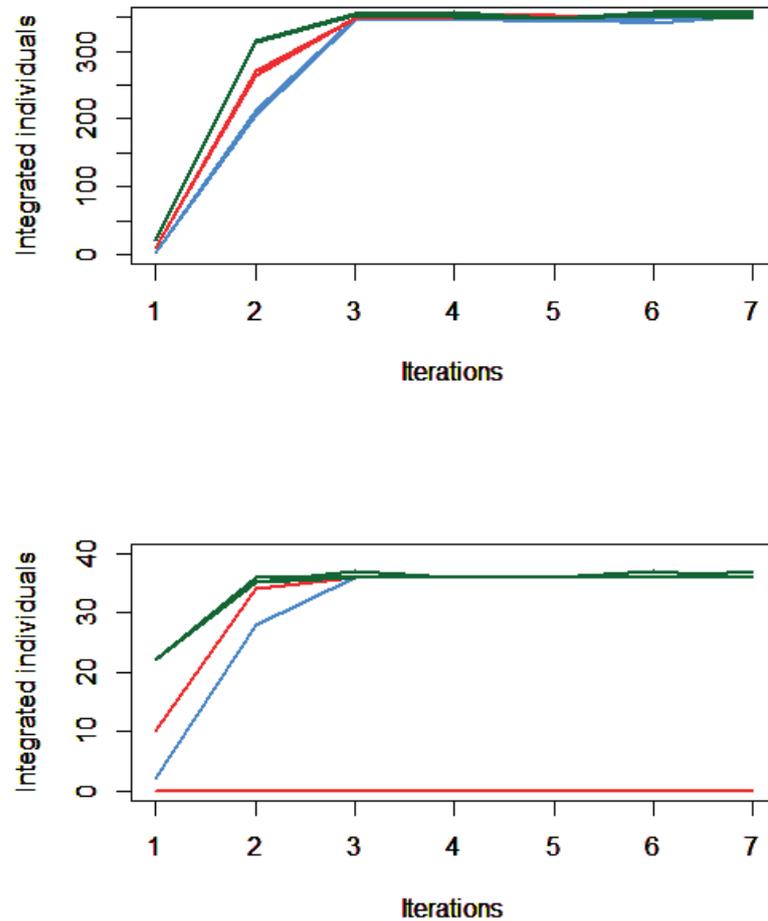


Figure S3.1.3: Convergence of the algorithm for the definition of the representative pool in domestic cats (on top) and wildcats (at the bottom) depending on the initialization. Blue lines correspond to the repeats made with $\pi = 1\%$, red lines to the repeats made with $\pi = 20\%$ and green lines to repeats made with $\pi = 50\%$.

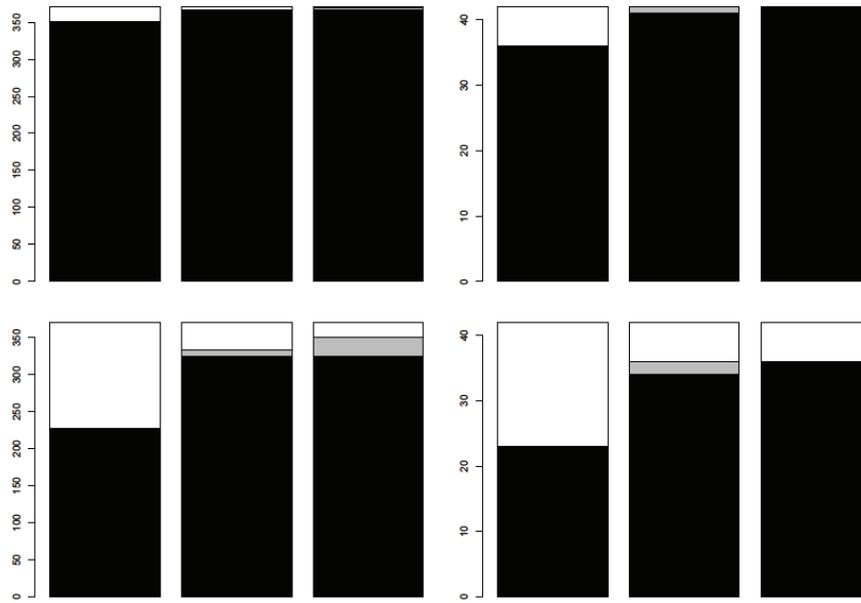


Figure S3.1.4: impact of the value of π at initialization on the categorization of domestic cats (on the left) and wildcats (on the right) depending on the initialization and the method (conservative on top vs relaxed at the bottom) with the *n best* method. Parents are in black, suspects in grey and hybrids in white.

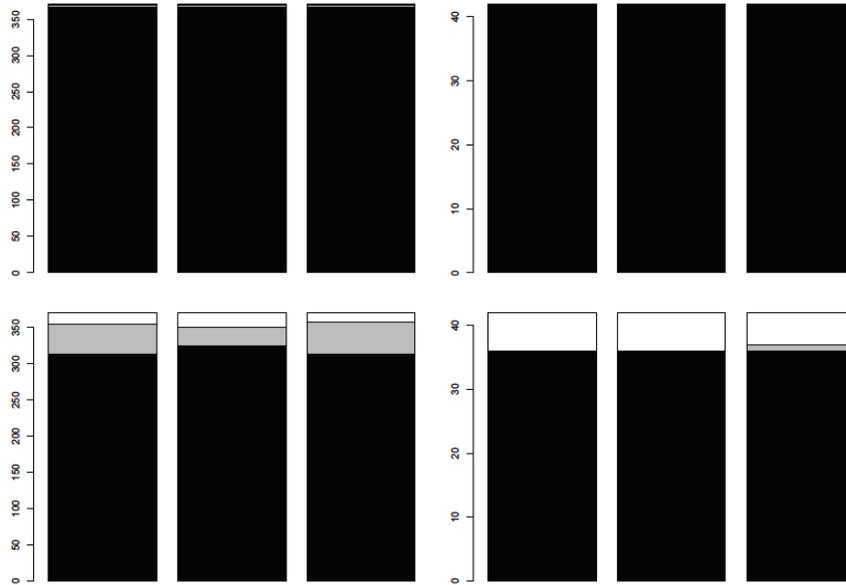


Figure S3.1.5: Comparison of the results obtained for the categorization of cats as either parents, hybrid or suspect with the Iterative algorithm ($\pi = 1\%$), the “*n best*” method ($\pi = 50\%$) and the *threshold* method for the domestic cats (on the left) and the wildcats (on the right) with the conservative (at the top) and the relaxed (at the bottom) approaches. Parents are in black, suspects in grey and hybrids in white.

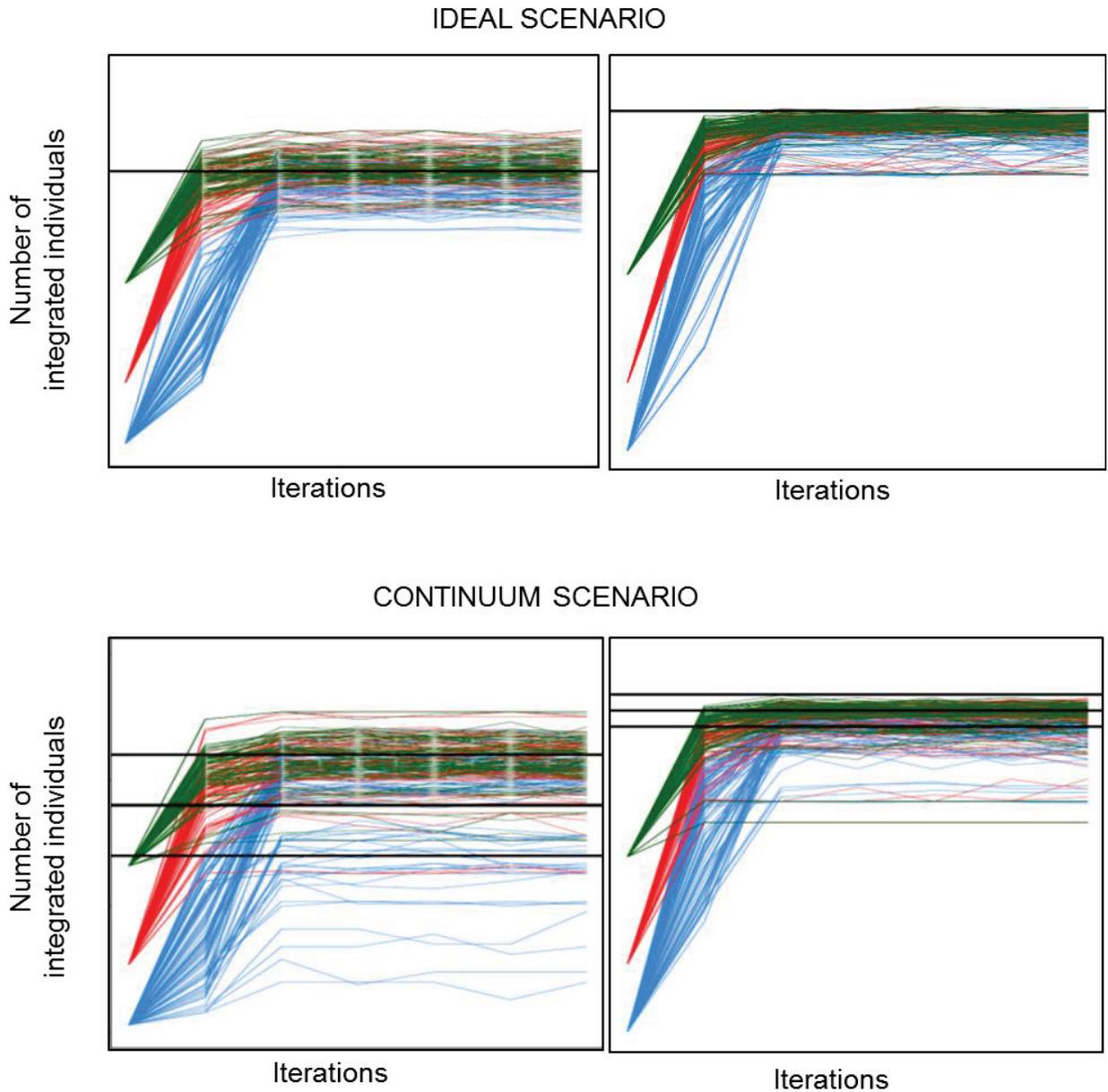


Figure S3.1.6: Convergence of the algorithm for both the *ideal* and the *continuum* scenario for the small (on the left) and the large (on the right) populations for $\pi=1\%$ (blue), $\pi = 20\%$ (red) and $\pi=50\%$ (green). For the *ideal* scenario, the black line corresponds to the actual number of parents while for the *continuum* scenario, the three lines correspond to 90% of the parental individuals, the exact number of parental individuals, and the number of parental individuals increased by the second backcrosses which are not expected to be detected with our panel resolution (from bottom to top).

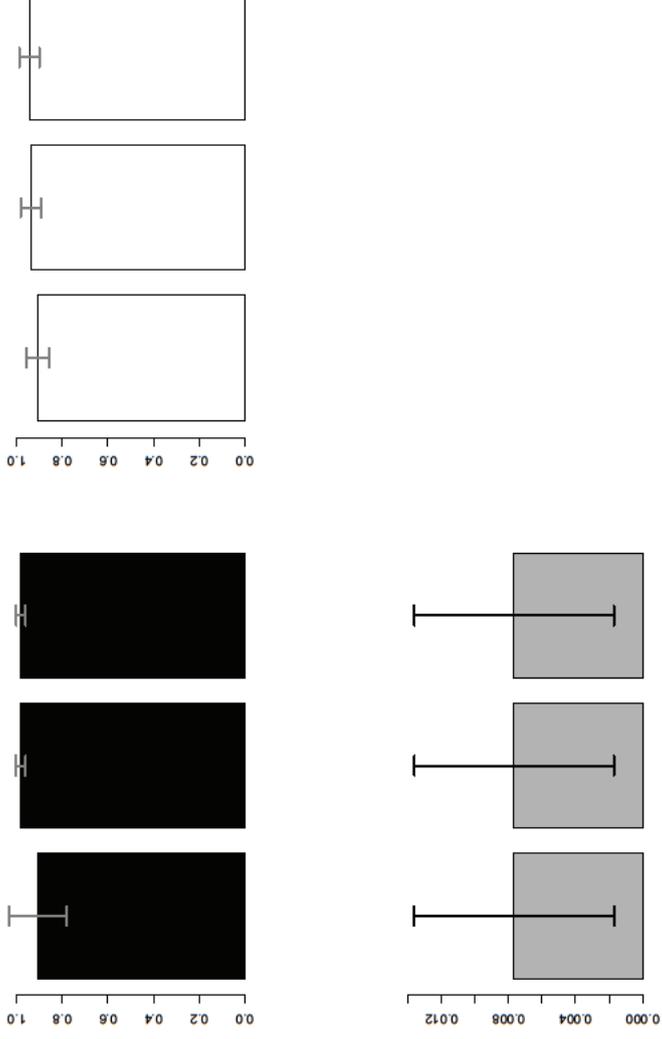


Figure S3.1.7: Impact of the value of π at initialization on the proportion of known parents integrated in the representative pool (P1 and P2 in black and white respectively) and the proportion of hybrids (all categories confounded) in grey for simulations under the *continuum* scenario.

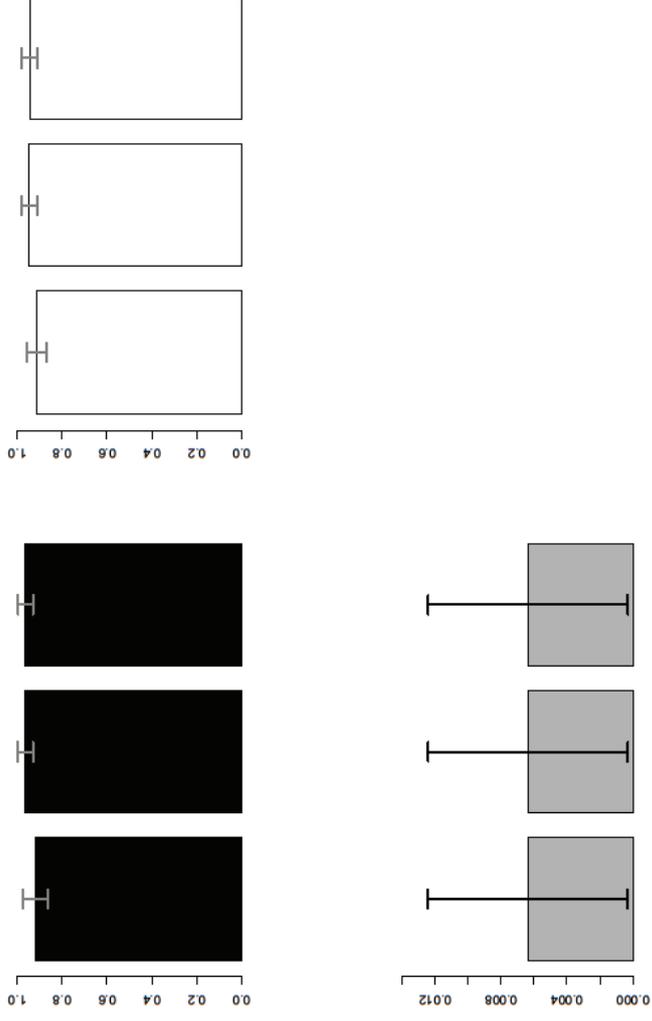


Figure S3.1.8: Impact of the value of π at initialization on the proportion of known parents integrated in the representative pool (P1 and P2 in black and white respectively) and the proportion of hybrids (all categories confounded) in grey for simulations under the *ideal* scenario.

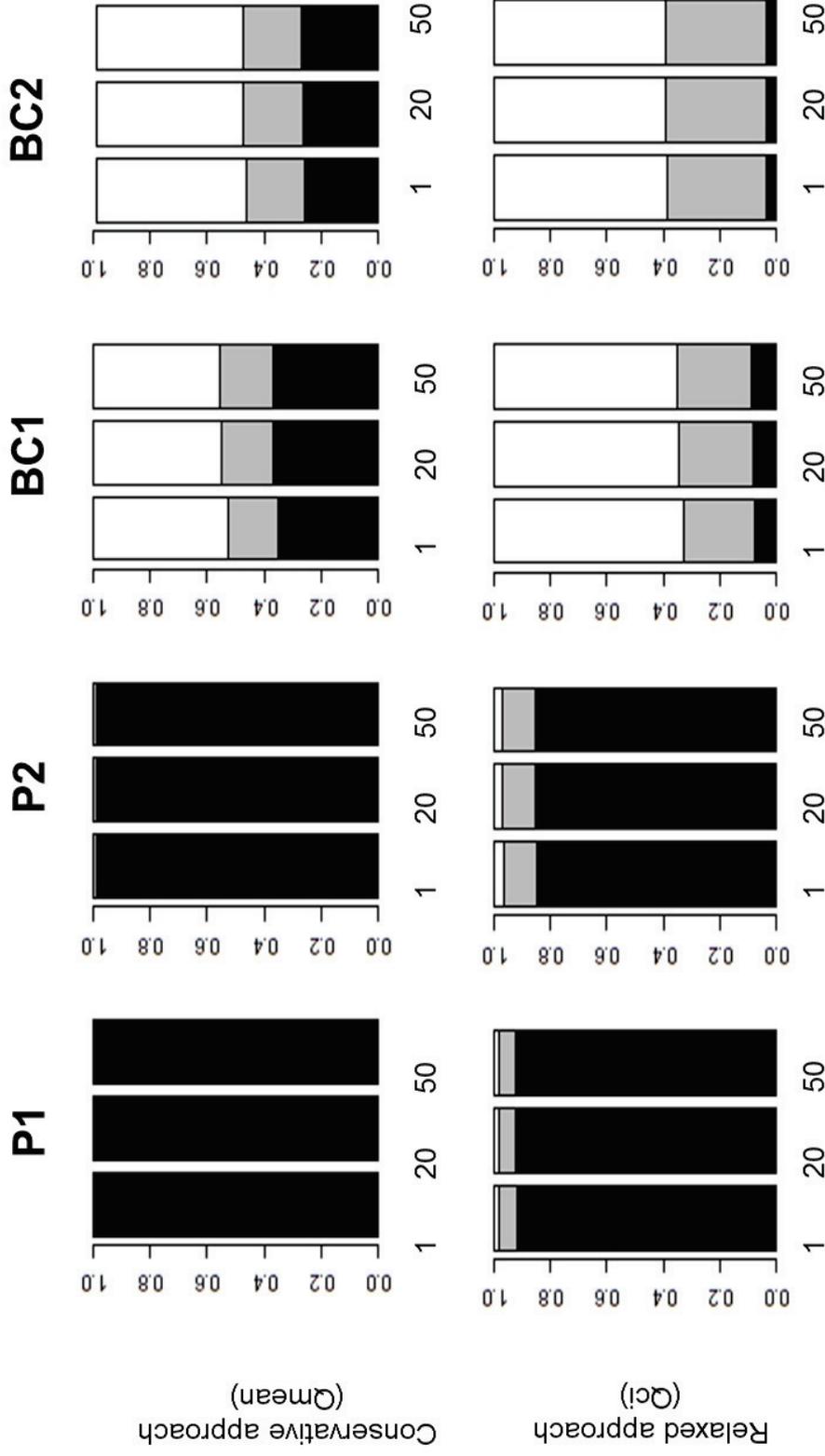


Figure S3.1.9: Impact of the value of π at initialization on the categorization made by the iterative algorithm for all known hybrid categories and both approaches over the different scenarios (*ideal* and *continuum*). Parents are in *black*, suspects in grey and hybrids in white.

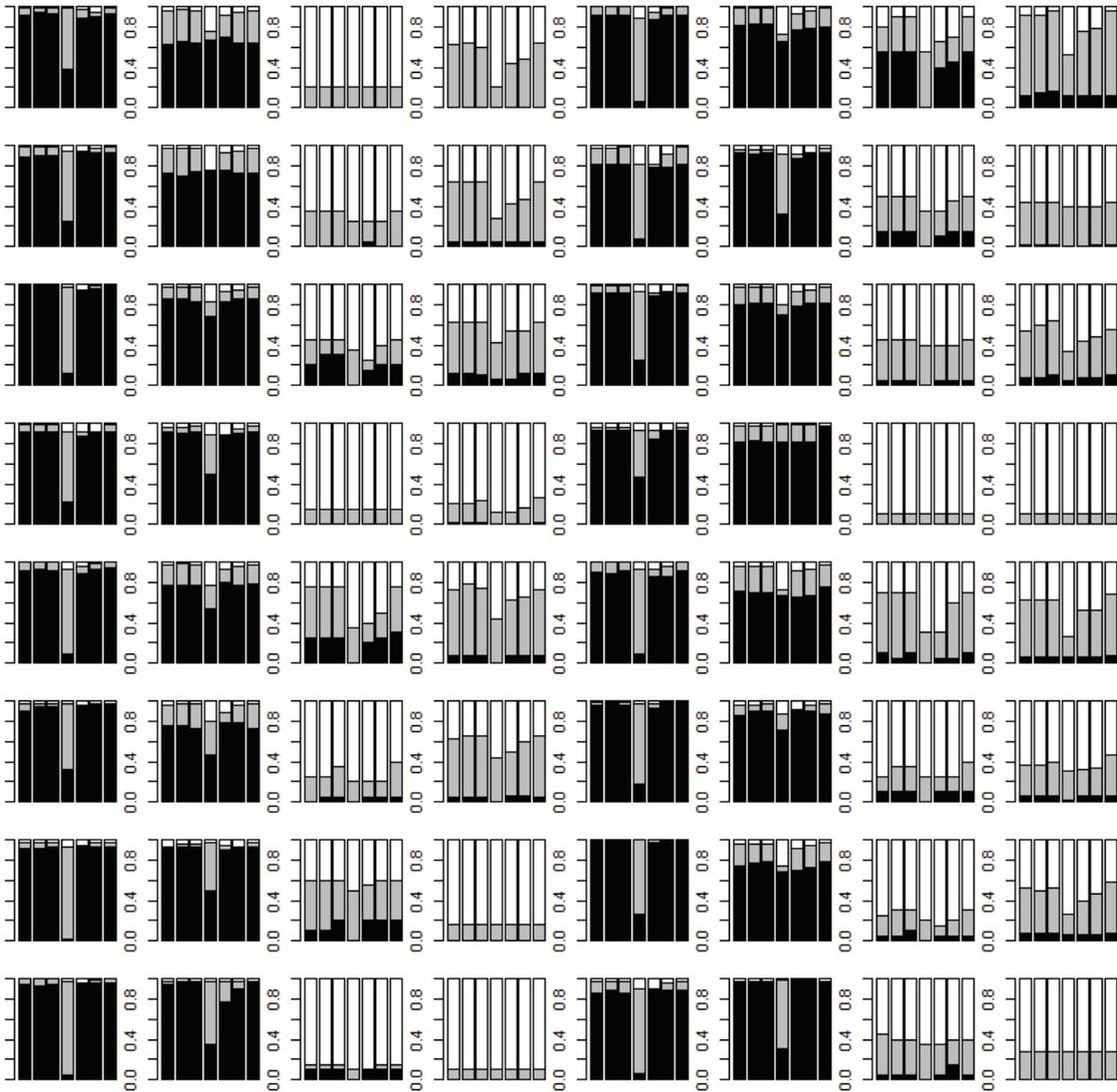


Figure S3.1.10: variability of the categorization across simulations using the relaxed approach. Each simulation is represented by 4 graphics (P1, P2, BC1, BC2) in line. Two simulations are then presented by line. On each graph, seven bars are represented, the three first bars correspond to results obtained with the iterative algorithm either with $\pi = 1\%/20\%/50\%$ at initialization. The three next bars correspond to the results obtained with the '*n best*' method also depending on the three values of π . Finally, the last bar corresponds to the results obtained with the *threshold* method. Parents are in black, suspects in grey and hybrids in white.

3.2. A fast likelihood solution to the genetic clustering problem



Domestic cat or hybrid? Personal photography from Sardinia in the Pyrénées
Orientales

ABSTRACT

The investigation of genetic clusters in natural populations is an ubiquitous problem in a range of fields relying on the analysis of genetic data, such as molecular ecology, conservation biology and microbiology. Typically, genetic clusters are defined as distinct panmictic populations, or parental groups in the context of hybridisation. Two types of methods have been developed for identifying such clusters: model-based methods, which are usually computer-intensive but yield results which can be interpreted in the light of an explicit population genetic model, and geometric approaches, which are less interpretable but remarkably faster.

Here, we introduce *snapclust*, a fast maximum-likelihood solution to the genetic clustering problem, which allies the advantages of both model-based and geometric approaches. Our method relies on maximising the likelihood of a fixed number of panmictic populations using a combination of geometric approach and fast likelihood optimization using the Expectation-Maximization (EM) algorithm. It can be used for assigning genotypes to populations and optionally identify various types of hybrids between two parental populations. Several goodness-of-fit statistics can also be used to guide the choice of the retained number of clusters.

Using extensive simulations, we show that *snapclust* performs comparably to current gold-standards for genetic clustering as well as hybrid detection, with some advantages for identifying hybrids after several backcrosses, while being orders of magnitude faster than other model-based methods. We also illustrate how *snapclust* can be used for identifying the optimal number of clusters, and subsequently assign individuals to various hybrid classes simulated from an empirical microsatellite dataset.

snapclust is implemented in the package *adegenet* for the free software R, and is therefore easily integrated into existing pipelines for genetic data analysis. It can be applied to any kind of codominant markers, and can easily be extended to more complex models including, for instance, varying ploidy levels. Given its flexibility and computer-efficiency, it provides a useful complement to the existing toolbox for the study of genetic diversity in natural populations.

Keywords: genetic clustering, EM algorithm, hybridization, genetic assignment, population membership, relative performances, SNP, microsatellites

Marie-Pauline BEUGIN, Thibault GAYET, Dominique PONTIER, Sébastien DEVILLARD and Thibaut JOMBART (in revision in *Methods in Ecology and Evolution*)

3.2.1. INTRODUCTION

The identification of groups of genetically related individuals within a population, *sensu* population subdivision, is an ubiquitous problem in most fields in which genetic data analysis plays an important role including molecular ecology, evolutionary and conservation genetics. Quantifying the magnitude of the population subdivision, assessing whether the genetic differentiation matches with the spatial repartition of subpopulations or not, and, identifying from which genetic units individuals belong or come have been the focus of attention of population geneticist from the inception of population genetics (Wright 1951). Specific applications include for example the definition of panmictic groups (Pritchard, Stephens & Donnelly 2000; Corander, Waldmann & Sillanpää 2003; Falush, Stephens & Pritchard 2003), the classification of isolates into distinct lineages in microbiology (Maiden *et al.* 1998; Feil *et al.* 2004), the investigation of social or ecological units in molecular ecology (Sugg *et al.* 1996; Jombart, Devillard & Balloux 2010), and the identification of various types of hybrids in conservation genetics (Allendorf *et al.* 2001; Anderson & Thompson 2002; Vähä & Primmer 2006). Because of this wealth of applications, genetic clustering has received considerable interest from the methodologists community. Seeking the number of genetic clusters from a set of individual genotypes and assigning individuals into clusters has become a gold standard in population genetics, and, a large number of statistical methods have been developed and used routinely for nearly two decades (Pritchard, Stephens & Donnelly 2000; Anderson & Thompson 2002; Corander, Waldmann & Sillanpää 2003; Falush, Stephens & Pritchard 2003; Jombart, Devillard & Balloux 2010).

While there is no single taxonomy of methods, a natural separation can be made between '*model-based*' approaches, which use a population genetics model to compute a likelihood, including maximum-likelihood (ML) and Bayesian methods (Pritchard, Stephens & Donnelly 2000; Anderson & Thompson 2002; Dupanloup,

Schneider & Excoffier 2002; Corander, Waldmann & Sillanpää 2003; Falush, Stephens & Pritchard 2003), and ‘*geometric*’ approaches, which cluster individuals based on their distances in the genetic space spanned by allelic data, without assuming a specific population genetics model (Feil *et al.* 2004; Jombart, Devillard & Balloux 2010). In genetic clustering problems, the likelihood is defined as the probability that the set of genotypes under consideration was generated under a given population structure and model of evolution. As such, these methods are more readily interpretable: individual group membership probabilities genuinely reflect the probability that the individual ‘belongs’ to the different groups. Unfortunately, these methods are typically computer-intensive, as they involve the exploration of a high-dimensional parameter space using optimisation procedures (Dupanloup, Schneider & Excoffier 2002) or Markov Chain Monte Carlo (MCMC) techniques (Pritchard, Stephens & Donnelly 2000; Corander, Waldmann & Sillanpää 2003; Falush, Stephens & Pritchard 2003; Vähä & Primmer 2006). While more efficient implementations have been developed (Tang *et al.* 2005; Alexander, Novembre & Lange 2009; Raj, Stephens & Pritchard 2014), geometric approaches remain an appealing alternative, as they are typically orders of magnitude faster, while producing comparably accurate results under a range of simulation scenarios (Jombart, Devillard & Balloux 2010). The main limitation of geometric approaches lies in the fact that their results are harder to interpret biologically. Indeed, these methods typically identify clusters from pairwise genetic distances, without providing group membership probabilities (Jombart, Devillard & Balloux 2010; Legendre & Legendre 2012), so that weak separation between clusters or admixture patterns cannot be distinguished from strong, clear-cut population structure. To some extent, this issue can be addressed using exploratory approaches such as the DAPC (Jombart, Devillard & Balloux 2010), to visualise cluster diversity in a reduced space and even estimate group assignment probabilities, but these probabilities merely reflect genetic proximities, and cannot be interpreted as probabilities that an individual belongs to a given population.

Here, we combine both types of approaches to formulate a new clustering method called '*snapclust*', which retains the advantages of both worlds. Our method relies on the most common population genetics model which underlies Hardy-Weinberg (HW) equilibrium to compute the likelihood of a given clustering solution. Rapid convergence to ML estimates of clusters is achieved by combining geometric approaches (Jombart, Devillard & Balloux 2010; Legendre & Legendre 2012) and the Expectation-Maximisation (EM) algorithm (Dempster, Laird & Rubin 1977). In practice, our method allows to select the optimal number of clusters within a set of genotypes, and provides results where group assignment scores are genuine probabilities that a given genotype was generated in various populations under HW model, while remaining essentially as fast as geometric approaches (Jombart, Devillard & Balloux 2010). Our method can also be used for identifying various types of hybrids between two parental populations. Besides, being a ML estimation method, *snapclust* can also be combined with goodness-of-fit statistics such as Akaike information criterion (AIC, Akaike 1998) or the Bayesian information criterion (BIC, Schwarz 1978) to guide the choice of the optimal numbers of clusters.

In this paper, we describe the model underlying *snapclust* and its implementation, and then compare the performance of our method with current gold-standards for genetic clustering (STRUCTURE (Pritchard, Stephens & Donnelly 2000; Falush, Stephens & Pritchard 2003), BAPS, *adegenet's find.cluster* (Jombart, Devillard & Balloux 2010) and hybrid identification (Newhybrids (Anderson & Thompson 2002))). Using a large number of simulations, we assessed the impact of the number of loci, the dispersal model, the level of genetic differentiation between populations, and the number of populations (when looking at multiple clusters without hybrids), on the performance of the different methods. We also provide a worked example based on the analysis of a simulated dataset to illustrate typical results provided by the method. *snapclust* is implemented in the package *adegenet* (Jombart 2008; Jombart & Ahmed 2011) for the R software (R Core Team 2017), thus being readily compatible with a wealth of tools for genetic data analysis in R

(Goudet 2005; Paradis 2010; Popescu, Huber & Paradis 2012; Kamvar, Tabima & Grünwald 2014; Jombart *et al.* 2017).

3.2.2. MATERIAL AND METHODS

a. Rationale of snapclust

Model likelihood

We consider a dataset of allelic profiles $x = \{x_{ij}\}$ where i indexes individuals ($i = 1, \dots, N$) and j indexes loci ($j = 1, \dots, J$), so that x_{ij} is a vector of allele counts for individual i at locus j . The likelihood of our model is defined as the probability of observing these data given a clustering solution $g = \{g(i)\}$, where $g(i)$ defines the group of individual i , with groups indexed by $k = 1, \dots, K$. Under the HW model, this likelihood is defined as:

$$p(x_{ij} | f_{g(i),j}, \boldsymbol{\pi}) = M(x_{ij}, f_{g(i),j}, \boldsymbol{\pi})$$

where M is the probability mass function of the multinomial distribution, $f_{g(i),j}$ is the vector of allele frequencies in group $g(i)$ at locus j , and $\boldsymbol{\pi}$ is the ploidy of the organism considered. Allele frequencies within a group are directly computed as the relative frequencies of each allele in this group. Assuming independence between loci, the likelihood term for the genotype i is given by:

$$p(x_i | f_{g(i)}, \boldsymbol{\pi}) = \prod_j p(x_{ij} | f_{g(i),j}, \boldsymbol{\pi})$$

where $f_{g(i)} = \{f_{g(i),1} \dots f_{g(i),j}\}$ and $x_i = \{x_{i,1}, \dots, x_{i,j}\}$. If we further assume independence of individuals conditional on their group memberships, the general likelihood is given by:

$$p(x | f, g, \boldsymbol{\pi}) = \prod_i p(x_i | f_{g(i)}, \boldsymbol{\pi})$$

where $f = \{f_1 \dots f_K\}$. In practice, we will consider the log-likelihood of a clustering solution defined as:

$$LL(g) = \sum_i \sum_j M(x_{ij}, f_{g(i),j}, \boldsymbol{\pi})$$

Note that while the current implementation of *snapclust* considers a constant ploidy across individuals and loci, the formula above can readily be extended to varying ploidy, in which case $\boldsymbol{\pi}$ will become an individual- or locus- specific term.

Assuming that all clusters have been sampled, the probability $p(g(i) = k)$ that an individual i belongs to a group k is defined by the standardised likelihood:

$$p(g(i) = k) = p(x_i | g(i) = k, f_k, \boldsymbol{\pi}) / \sum_q p(x_i | g(i) = q, f_q, \boldsymbol{\pi})$$

Modelling hybridisation

The clustering model above can be readily extended to accommodate the presence of hybrids. For simplicity, we consider a case where hybrids are obtained from two parental populations A and B . The allelic composition $f_{H,j}(w)$ of a hybrid population H at locus j is defined as a mixture of the allele frequencies of two parental populations, $f_{A,j}$ and $f_{B,j}$. This mixture is defined by the hybridisation coefficient w , which indicates the proportion of the genomes of the hybrid population coming from the parental population A , so that:

$$f_{H,j}(w) = w f_{A,j} + (1 - w) f_{B,j}$$

Modelling of hybridisation through the coefficient w is very flexible, as it enables the specification of any kind of hybrids between A and B . For instance, first-generation hybrids (F1) correspond to $w = 0.5$, while first and second generations backcrosses with A respectively correspond to $w = 0.25$ and $w = 0.125$. The likelihood of a hybrid is defined as before, but using the allele frequencies mixture:

$$p(x_i | g(i) = H, f_A, f_B, w, \boldsymbol{\pi}) = \prod_j p(x_{i,j} | f_{H,j}, \boldsymbol{\pi})$$

Optimisation procedure

snapclust achieves fast likelihood maximisation using the EM algorithm (Dempster, Laird & Rubin 1977), in which the vector of group membership g is treated as a latent variable. In this respect, our approach is closely related to K-

means clustering, except that *snappclust* maximises a log-likelihood rather than between-group distances (Jombart, Devillard & Balloux 2010). The EM algorithm proceeds by alternating computation of the likelihood, and assignment of individuals to their most likely cluster. Allele frequencies are updated at each iteration using their maximum likelihood estimation, i.e. the mean frequencies of alleles in individuals of a given group. The algorithm, adapted from the use of EM for maximizing likelihood in mixed distribution problems (Fraley & Raftery 2002), can be formalised through the following steps:

1. define initial group assignments g (see ‘starting point’ below)
2. (*expectation* step) update allele frequencies f within each group, computed as the relative frequencies of alleles amongst individuals of this group; compute group membership probabilities $p(g(i) = k)$ for all individuals i and groups k
3. (*maximisation* step) update the group definition g : based on group membership probabilities computed in step 2, assign each individual to their most likely group
4. return to step 2 until convergence

We assume convergence when the difference in log-likelihoods in two successive iterations becomes negligible, i.e. is less than an arbitrary threshold (set to 10^{-10} by default).

Starting point

The EM algorithm typically converges very fast, generally within 10 iterations in the simulated and empirical datasets described here. Unlike some other optimisation procedures and MCMC, it is a deterministic algorithm, so that it always converges to the same solution for a given starting point (step 1). As a consequence, it is unfortunately also prone to being trapped in local maxima, yielding sub-optimal results for some starting points. To avoid this issue, we implemented several options to define the initial clusters used as starting point of the algorithm. The first strategy, borrowed from the original implementation of *K*-means in R (R Core Team 2017), is a ‘brute force’ approach in which the algorithm

is run multiple times, using each time a randomly defined group assignment, and retaining the solution with the highest likelihood. The second strategy which we introduce here is to use fast geometric approaches such as Ward's clustering (Legendre & Legendre 2012) or K -means after dimension reduction (Jombart, Devillard & Balloux 2010) to set up the initial clusters. Based on our simulated datasets, random initial groups with 50 independent replicates, K -means, and Ward initialisation all gave similar results. By default, we recommend using Ward as it will be faster for for most datasets. The three methods are available in the implementation of the algorithm, as well as any other user-defined initial clusters.

Finding the optimal number of clusters

The advantage of using a ML approach is that different models can be compared using classical goodness-of-fit statistics. While a full comparison of model selection techniques for genetic clustering is beyond the scope of the present paper, we have implemented four different information criteria shown to be useful for selecting the true number of clusters in the case of mixtures of distributions (Akogul & Erisoglu 2016). These statistics all rely on measuring the lack of fit of the model (deviance), and use different penalties for the complexity of the model (number of free parameters). The first, Akaike's Information Criterion (AIC, Akaike 1998), is probably the most frequently used for models comparison. Noting L' the estimated maxima of $LL(g)$, the AIC of our model is computed as:

$$AIC = -2L' + 2(K(P - J))$$

where the first term is the deviance of the model, and the second term corresponds to the complexity of the model, with P being the total number of alleles in the dataset across J loci. The complexity reflects the fact that for each of the K groups, $(P - J)$ independent allele frequencies are estimated, so that the total number of free parameters of the model is $(K(P - J))$. We also implemented the variant of the AIC for small sample sizes, defined as (Akogul & Erisoglu 2016):

$$AICc = -2L' + 2(K(P - J)N) / (N - KP + KJ - 1))$$

A popular alternative to AIC and AICc is the Bayesian Information Criterion (BIC, Schwarz 1978), which also relies on a penalised deviance, albeit putting a stronger cost on complexity:

$$BIC = -2L' + \ln(N) (K(P - J))$$

Finally, we also implemented the Kullback Information Criterion (KIC, Cavanaugh 1999), which gave the best overall results for detecting the number of clusters from mixtures of multivariate normal distributions (Akogul & Erisoglu 2016):

$$KIC = -2L' + 3(K(P - J) + 1)$$

All these statistics have similar behaviours in that the lower values typically indicate better fits. In practice, a sharp decrease in the statistics values with increasing numbers of clusters is most likely to reveal the optimal numbers of clusters (Jombart, Devillard & Balloux 2010).

Implementation and availability

snapclust is implemented in the R package *adegenet* (Jombart 2008; Jombart & Ahmed 2011) version 2.1.0, available via R's native package installation system as well as on github (<https://github.com/thibautjombart/adegenet>). The function `snapclust.em` implements the basic method, including different options for defining the initial state of the EM algorithm and the model for hybrids classification. The functions `AIC`, `AICc`, `BIC` and `KIC` implement the respective goodness of fit statistics. The function `snapclust.em.choose.k` derives clustering solutions for increasing numbers of clusters and computes the associated goodness of fit statistics, so that it can guide the choice of the optimal number of clusters. The method is documented in a dedicated online tutorial available by typing `adegenet::adegenetTutorial('snapclust')` in a R session. Code and documentation are released under GPL >= 2 license.

b. Simulations

Simulated datasets without hybrids

The datasets were simulated using QuantiNEMO (Neuenschwander *et al.* 2008) with the parameters indicated in Table 1. We chose to simulate Single Nucleotide Polymorphism (SNP) markers and explored a wide range of possible configurations by varying four simulation parameters: the number of loci, the dispersal model, the rate of dispersion, and the number of populations. The different rates of dispersal led to different levels of differentiation between populations. All combinations of dispersal rate and number of loci were tested as the number of loci and the differentiation level are expected to define jointly the resolution of a panel of genetic markers (Vähä & Primmer 2006). This led to 36 combinations of parameters. Ten independent random replicates were obtained for each combination leading to 360 simulated datasets. For each replicate, the number of populations and the dispersal model were chosen randomly. The number of individuals per population was fixed to 100.

Simulated datasets with hybrids

The simulated datasets used for the clustering of hybrids were derived from the previous simulations, by sampling two parental populations (P1, P2) at random in each of the 360 simulated datasets described before. For each, hybrids were simulated using the function *hybridize* of the *adegenet* package to obtain F1 hybrids (P1 x P2), first generation back-crosses (BC1: F1 x P1 and F1 x P2), and second generation back-crosses (BC2: (F1 x P1) x P1 and (F1 x P2) x P2). Each simulated dataset was formed by 100 individuals from P1 and P2 each, and 10 individuals from each hybrid class (i.e. 50 hybrids in total). While arbitrary, these sample sizes yielded a sufficient number of hybrids to analyse while retaining enough individuals to characterize the genetic makeup of parental populations.

Analyses of simulated datasets without hybrids

Our simulation study focussed on comparing *snapclust* to existing standard for the assignment of individual genotypes to groups (rather than inferring the true number of clusters). Therefore, the number of clusters was fixed to the known number of populations within the simulated dataset for all presented analyses. The clustering of individuals in absence of hybrids was performed using the *snapclust*, STRUCTURE 2.3.4 (Pritchard, Stephens & Donnelly 2000; Falush, Stephens & Pritchard 2003; Hubisz *et al.* 2009), BAPS 5.4 (Corander, Waldmann & Sillanpää 2003; Tang *et al.* 2009; Cheng *et al.* 2013), and *adegenet*'s `find.clusters` (Jombart, Devillard & Balloux 2010). The *snapclust* analysis was carried out using default parameters (group assignment initialization using the 'ward' option). STRUCTURE analyses were carried out using an admixture model with correlated allele frequencies between populations and no *a priori* information on population membership. The program was run ten times for result consistency purposes, with MCMC length of 500,000 after a burn-in of 100,000 iterations. Individuals were assigned to the cluster for which their posterior assignment probability was the highest. For BAPS, we performed a 'mixture clustering' analysis. Finally, we ran the function `find.clusters` retaining 90% of the total variation in the initial dimension reduction step.

As clusters identified in these previous analyses are not labelled, it was impossible to judge if individuals were assigned to their true cluster. To assess the quality of the results and compare the different methods, we used pairwise comparisons of individuals instead, examining whether pairs of individuals were adequately placed in the same, or different clusters. We used two complementary measures to do so calculated on each of the 360 simulated dataset analysed. The *True Positive Rate (TPR)* was defined as the proportion of individuals belonging to the same population which were indeed clustered together by the method. The *True Negative Rate (TNR)* was defined as the proportion of individuals which did not belong to the same population and were adequately placed in different groups

by the method. Note that the Rand index (Rand 1971), which can be used for comparing unlabelled clusters, is proportional to $(TPR + TNR)$, so that the present analyses should give a more detailed account of clustering results than the Rand index alone. The impact of the different simulation parameters on TPR and TNR was assessed using separate multivariate linear regressions. As classical linear regression is designed to predict a response variable which can take any positive or negative values, a logit transformation was applied to the proportions, so that $\log(TPR / 1 - TPR)$ and $\log(TNR / 1 - TNR)$ were used as response variables.

We tested for the effects of the number of loci, the dispersal model, the overall Fst between simulated populations, the number of populations, and the clustering method. We also investigated potential two-way interactions between the clustering method and the four simulation parameters we varied (Table 3.2.1), as well as between the number of loci and the Fst . Backward stepwise model selection based on AIC was used to retain significant predictors, and confirmed using classical likelihood ratio tests. Bonferroni correction was used to account for multiple testing with a target type 1 error of 1%. When assessing the overall differences between methods across all simulations, we compared TPR and TNR predicted by the respective models by transforming predicted logit rates back to their original scale.

Table 3.2.1: Parameters used in the simulations using the computer program QuantiNEMO.

Parameter	Value	Parameter	Value
generations	10000	Population size	100
Number of loci	[20, 50, 80, 150, 300, 500]	Number of alleles	2
Dispersal model	Migrant-pool island model or 1-D stepping stones model	Mutational model	K-allele model
Dispersal rate	[0, 0.001, 0.002, 0.003, 0.005, 0.01]	Mutation rate	0
Number of populations	2 - 15	Mating system	random mating

Analyses of simulated datasets with hybrids

The clustering of individuals in presence of hybrids was carried out using *snapclust* and the computer program NEWHYBRIDS (Anderson & Thompson 2002). The *snapclust* analysis was carried out using the default parameters and specifying hybridisation coefficients for F1, first (BC1) and second (BC2) generation backcrosses (hybrid.coef values: 0.5, 0.25, 0.125). The NEWHYBRIDS analysis was carried out using Jeffreys's prior and setting the burn-in period to 100,000, with a MCMC length of 500,000 iterations. Ten repetitions were carried out for each simulated dataset. Unlike the previous comparison, parental and hybrid classes are labelled, so that it was possible to compare the methods by directly examining how well they assigned individuals to their actual hybrid group, using the *mean correct group assignment*, computed as the proportion of individuals whose type (parental, F1, BC1 and BC2) was correctly identified. In addition, we also examined the group membership probability calculated by each method for the true group, later referred to as the '*support*' for the true group. As before, the impact of the different simulation parameters on the performance of the methods was assessed using multiple linear regression on logit probabilities, with separate models for the mean correct group assignment, and the support to the true group. In both cases, the following predictors were included: number of loci, dispersal model, *Fst*, as well as the hybrid class (parental, F1, first or second back-cross), and the clustering method. Interaction were investigated between the method and the simulation parameters, and between the number of loci and the *Fst*. As before, variable selection was achieved using backward stepwise selection based on AIC and likelihood ratio tests, using a Bonferroni correction to account for multiple testing with a target type 1 error of 1%.

c. Illustration using microsatellite data

To complement the simulation study which assessed the overall performances of our method, we illustrated its practical application by reproducing a typical analysis of microsatellite markers data, starting with the identification of the most likely number of clusters, followed by the assignment of individuals to groups, and the description of relationships between groups. We simulated hybrids from an empirical dataset of 30 microsatellite markers typed for 15 breeds (Laloë *et al.* 2007), distributed as the ‘microbov’ dataset in *adegenet*. Parental populations were obtained by sampling 30 individuals from the Lagunaire and 30 from the Salers populations. Hybrids were simulated using the function *hybridize*, to obtain 30 F1 hybrids, and then 30 of each first and second backcrosses, resulting in 210 individuals. While arbitrary, these numbers replicate a situation where hybrids are more numerous than parental populations, as could be the case in nature when studying large hybridization zones.

We first carried out a global clustering analysis on this dataset, looking for the optimal number of clusters using AIC (function *snapclust.em.choose.k*) in order to confirm that K=2 parental populations was the optimal solution. We then looked for potential hybrids (function *snapclust.em*), using hybridisation coefficients corresponding to F1 (0.5), first generation backcross (0.25, 0.75) and second-generation backcross individuals (0.125, 0.875). Group membership probabilities were visualised using the function *compoplot*. As a complement, we also explored the diversity between hybrid classes using a discriminant analysis of principal components (DAPC, Jombart, Devillard & Balloux 2010), employing cross-validation to determine the optimal number of principal components to retain.

3.2.3. RESULTS

a. Clustering of individuals without hybrids

All four different methods exhibited very good performances in terms of *TPR* (most results above 90%) and near perfect *TNR*, showing that clusters present in the simulated dataset were overall well recovered by all approaches (Figure 3.2.1). Runtime analysis showed that *snapclust* was on average 27 times faster than BAPS and about 120,000 times faster than STRUCTURE, with an average analysis time below a second (Table S3.2.1).

Multivariate linear regression captured a large fraction of the variation in $\text{logit}(TPR)$ values (Adjusted R^2 : 61.45%; $p < 2.2 \times 10^{-16}$; Table S3.2.2). It revealed some significant differences between methods, with larger *TPR* for BAPS ($t = 4.095$; $p = 4.46 \times 10^{-5}$) compared to STRUCTURE and *snapclust* (Table S3.2.2). Results for *find.clusters* were more difficult to interpret because of negative interactions with increasing *Fst* ($t = -4.35$; $p = 1.43 \times 10^{-5}$) and numbers of populations ($t = -5.35$; $p = 9.98 \times 10^{-8}$), but were overall lower than in other methods across all simulations (Figure 3.2.1). Increased number of loci and *Fst* values generally led to improved *TPR* values, although a saturation effect was observed, so that large numbers of loci and stronger *Fst* effectively cancelled out ($t = -8.47$; $p = 5.88 \times 10^{-17}$). We note, however, that while significant the effect sizes were all very small, so that the actual differences in the methods' performance across all simulations remained negligible. Indeed, the mean predicted *TPR* varied only marginally, with 97% for *find.clusters*, 98% for *snapclust* and STRUCTURE, and 99% for BAPS.

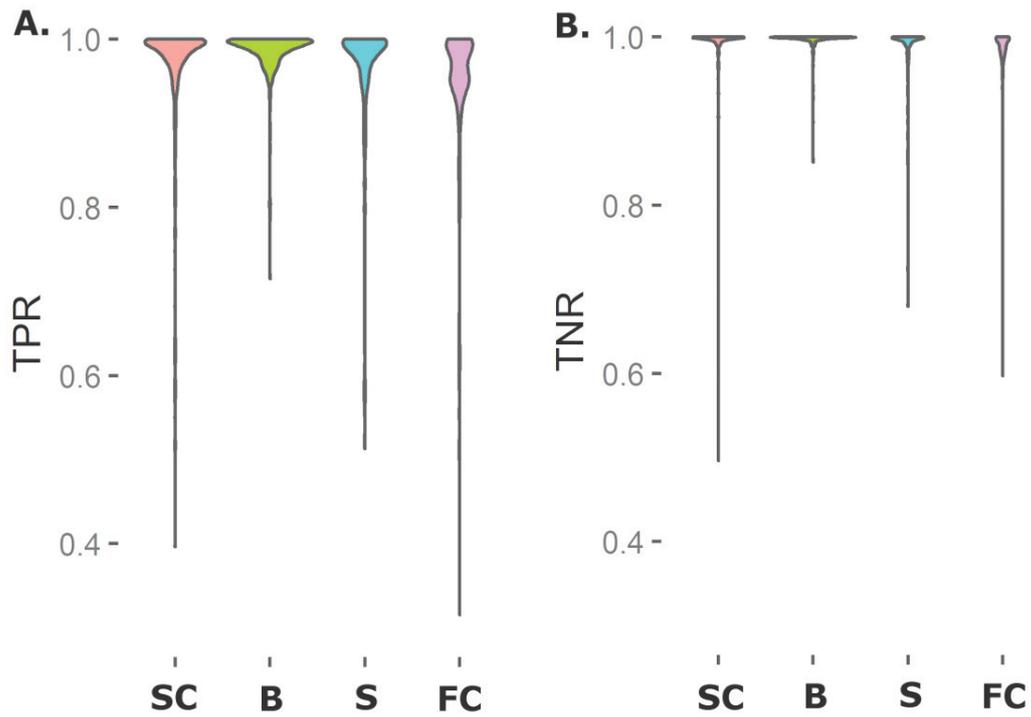


Figure 3.2.1: comparison of the various methods on simulated genetic clusters. This figure shows the distribution of A) the true positive rates (*TPR*) and B) true negative rates (*TNR*) obtained over all the 360 simulations for the four different methods: *snaphclust* (SC), BAPS (B), STRUCTURE (S) and *find.clusters* (FC) for the clustering of individuals in absence of hybrids. This width of the envelopes reflects the density of points.

Similar results were observed for $\text{logit}(TNR)$ values, although the model explained a smaller fraction of the variance in TNR values (Adjusted R^2 : 43.98%; $p < 2.2 \times 10^{-16}$; Table S3.3.3). Increased number of loci ($t = 12.60$; $p = 1.33 \times 10^{-34}$) and larger Fst ($t = 23.27$; $p = 7.79 \times 10^{-102}$) generally improved TNR values, although the same saturation effect as for TPR was observed ($t = -9.08$; $p = 3.45 \times 10^{-19}$). In addition, larger number of populations led to improved TNR ($t = 10.16$; $p = 1.91 \times 10^{-23}$). The function *find.clusters* also exhibited significantly lower TNR than the other methods ($t = -5.14$; $p = 3.08 \times 10^{-7}$). Again, while these differences were significant, the overall results across all simulations showed negligible variation across the different methods, and excellent performances overall, with average predicted TNR above 99% for all methods.

b. Clustering of individuals with hybrids

Results based on the proportion of correct assignment and the support to the true group both showed similar patterns, with stark contrast between *snapclust* and NEWHYBRIDS (Figure 3.2.2, Tables S3.2.4-5). The final model of the proportion of correct assignment explained most of the variation in the results (adjusted $R^2 = 63.78\%$; $p < 2.2 \times 10^{-16}$). Increased number of loci ($t = 23.32$; $p = 2.74 \times 10^{-110}$) and stronger Fst ($t = 31.28$; $p = 5.094 \times 10^{-185}$) generally improved group prediction, although a significant yet negligible saturation effect was observed between the two ($t = -5.046$; $p = 4.78 \times 10^{-7}$). While hybrid classes were on average harder to identify than parental populations, with the lowest success observed for deeper backcrosses, the two methods behaved very differently: NEWHYBRIDS seemed to recover parental populations more efficiently, but *snapclust* exhibited improved performances for the identification of hybrids in deeper levels of hybridization (Figure 3.2.2, Table S3.2.4-5). This contrast was strongest for BC2, in which the odd ratio of accurate group predictions averaged to 4.80 in *snapclust* (95% CI: [2.09 - 11.01]) compared to NEWHYBRIDS. Results were qualitatively identical when examining the support to the true group (Figure 3.2.2, Table S3.2.5), although the

difference in odd ratio for BC2 was smaller, with an average of 1.74 (95% CI: [1.05 - 2.87]). As for the clustering comparison, *snapclust* also proved more computer efficient, being on average 525,000 times faster than NEWHYBRIDS, with an average runtime of 0.54s.

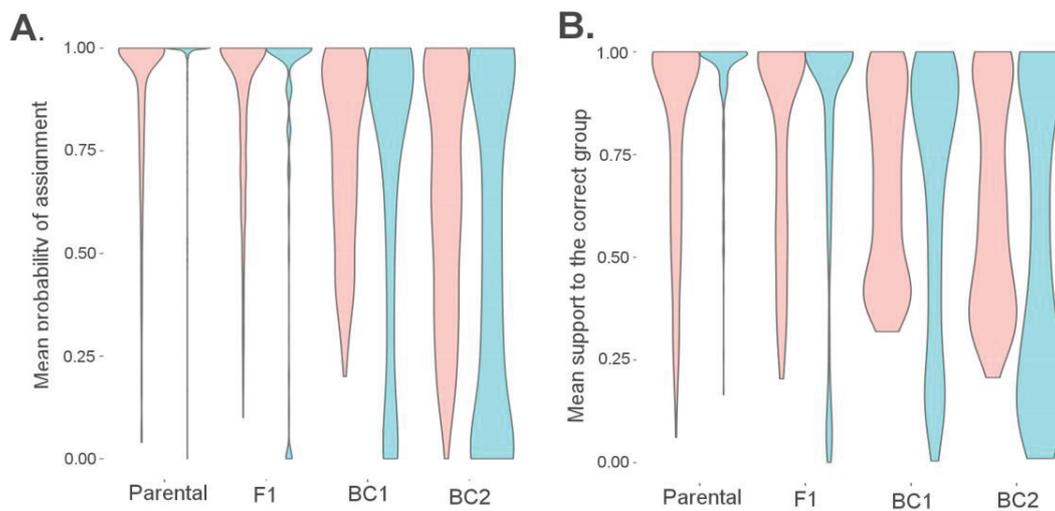


Figure 3.2.2: comparison of *snapclust* (red) and NEWHYBRIDS (blue) for the identification of hybrids using simulated data. This figure shows the distributions of A) the mean proportion of correct group assignment and B) the support (i.e. group membership probability) for the true class across all simulated datasets. Three hybrid classes are considered in the simulations in addition to the parental class: first generation hybrids (F1), first generation backcrosses (BC1) and second-generation backcrosses (BC2). This width of the envelopes reflects the density of points.

c. Illustration on the microbov data

AIC values computed for increasing values of K showed a sharp decrease at K=2, with only marginal improvements for K=3, hinting to the existence of two major clusters (Figure 3.2.3A) here formed by the parental populations (Salers and Lagunaire). Other goodness-of-fit statistics (AICc, BIC, KIC) also pointed to K=2, but AIC showed the most clear-cut result (Appendix, Figure S3.2.1). Subsequent analysis with *snapclust* including F1 hybrids as well as first and second-generation backcrosses shows well-identified parental clusters, as well as a large number of individuals assigned to the hybrid classes (Figure 3.2.3B). Parental and F1 hybrids groups were well identified, with 98.3% and 93.3% of successful individual assignment, respectively. Deeper hybrid classes were much harder to recover, with 51.7% of the BC1 and only 16.7% of BC2 correctly identified. This result is however in line with expectations in the presence of weak genetic differentiation. Indeed, while moderate genetic differentiation was observed between parental populations ($F_{st} = 0.157$), the average differentiation between BC2 and the ‘neighbouring’ groups (closest parent and BC1) was negligible ($F_{st} < 0.01$). This lack of differentiation was confirmed by a DAPC retaining 20 dimensions (Figure S3.2.2), which showed that individuals were structured along a cline of genetic differentiation between the two parental populations, with considerable overlap between ‘neighbouring’ groups (Figure 3.2.3C).

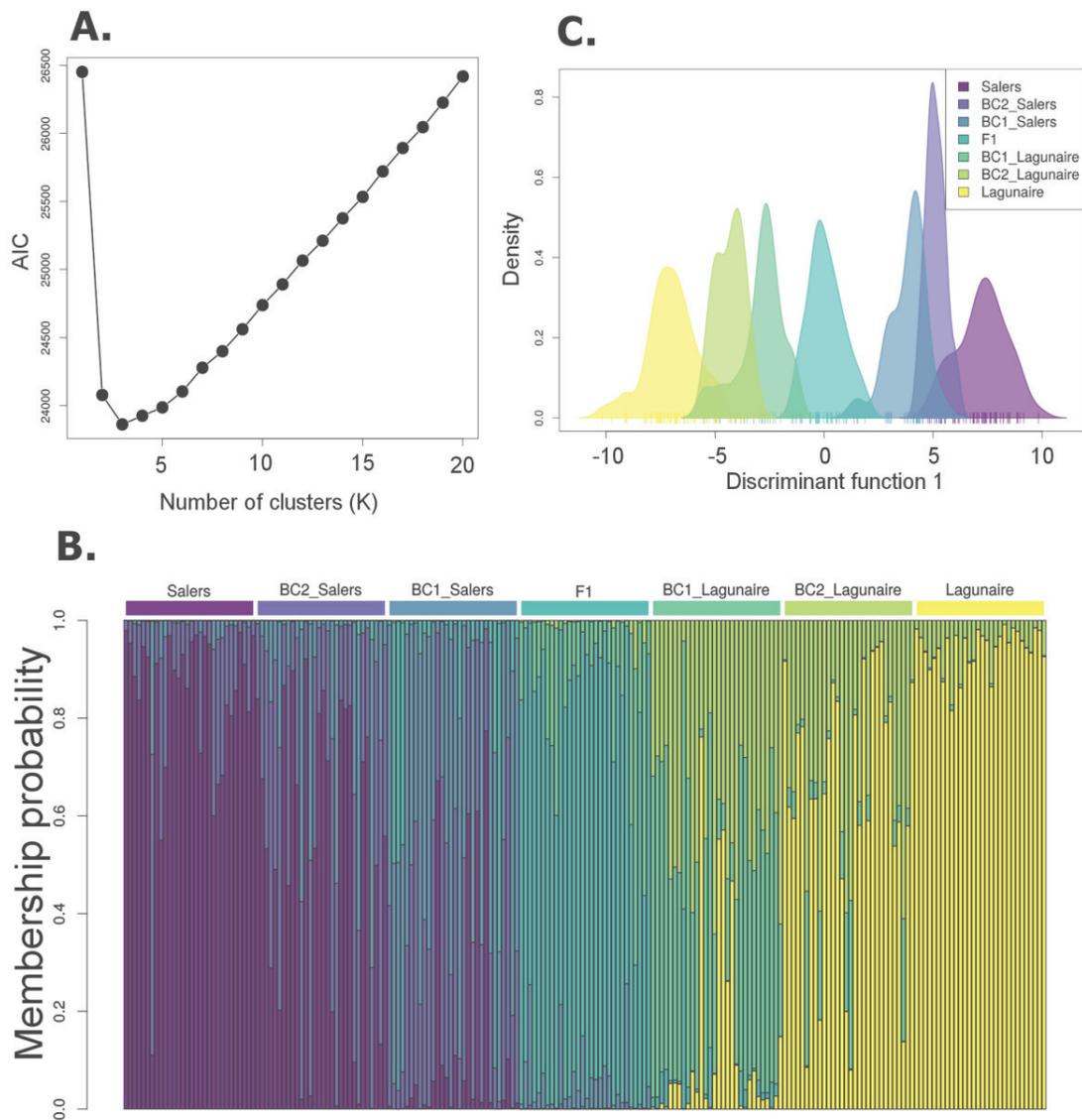


Figure 3.2.3: illustration of *snapclust* using simulated hybrids from cattle breed microsatellite data. A. Representation of the Akaike Criterion value according to the number of populations (K) considered. B. Representation of the individual probability of assignment obtained with the function *snapclust.em* for the different types of individuals present in the dataset. C. Representation of the first axis of the discriminant analysis of principal components carried out on the hybrid groups found using the *snapclust* analysis.

3.2.4. DISCUSSION

We have introduced ‘*snapclust*’, a new genetic clustering method which achieves fast maximum likelihood identification of the optimal number of clusters within a set of genotypes, assignment of individuals to panmictic populations, and can also be used to detect various classes of hybrids. The analyses of simulated data show that our method performs as well as current gold-standards for genetic clustering under the investigated models. Indeed, while statistically significant differences were observed in *TPR* and *TNR* across methods with BAPS exhibiting the best results, these differences were in fact negligible in terms of absolute performance: predicted *TPR* was 97% for *snapclust* compared to 98% for BAPS, and predicted *TNR* exceeded 99% for both methods. When used to detect hybrids, *snapclust* exhibited different performances from NEWHYBRIDS, being less accurate for identifying parental populations but better at recovering deeper hybrid classes such as second generation back-crosses, while being again tremendously more computer efficient. The combination of likelihood estimation and EM algorithm for cluster detection is not new (Fraleley & Raftery 2002), and has been used successfully as a fast yet powerful alternative to more complex likelihood-based methods in other fields than population genetics (Fraleley & Raftery 1998). As such, we believe the kind of approach introduced here offers exciting prospects for extending previous efforts for making model-based genetic clustering methods more computer-efficient (Tang *et al.* 2005; Alexander, Novembre & Lange 2009; Raj, Stephens & Pritchard 2014).

The fact that *snapclust* is orders of magnitude faster than other model-based approaches gives it a substantial practical advantage, especially when the analysis needs to be run multiple times, as is the case when investigating different values of *K*, when conducting a simple simulation study, or when using resampling strategy to assess statistical uncertainty. This latter aspect in particular is worth investigating, as our method does not, unlike Bayesian approaches (e.g. BAPS, STRUCTURE) include a natural measure of uncertainty in the form of distributions

of group membership probabilities for each individual. For *snapclust*, an alternative approach to assess statistical uncertainty may be to use bootstrap, in which case the method would be run a large number of times (e.g. 100) on datasets obtained by random re-sampling (with replacement) of the loci. Such approach would provide a distribution of group membership probabilities for each individual (one per run), and thereby a measure of uncertainty. Bootstrap on loci can readily be implemented using existing tools for genetic data handling (Jombart 2008; Jombart & Ahmed 2011; Kamvar, Tabima & Grünwald 2014). It would be relatively easy to apply in the case of hybridization between two parental clusters, in which case clusters are labelled, and therefore comparable across different runs. In the general case of unlabelled clusters, however, the difficulty of matching clusters across different runs will first need to be overcome for this approach to be applied.

While our simulation study required substantial computational resources, there are undoubtedly many more scenarios and methods to explore, involving a wider range of population genetics models, optimization procedures, and potentially various types of genetic markers. The relative effects of selection, recombination, and linkage disequilibrium remain to be investigated. The latter may be of first concern, as it would break the assumption of independence between loci, in which case our model only approximates the actual, unknown likelihood. This said, the very same assumption underpins maximum likelihood phylogenetic reconstruction, which has nonetheless proved tremendously useful over the past decades (Felsenstein 1981, 2004). We also note that our simulation study compared assignment of individuals to groups across different methods, assuming the true number of clusters was known. Examining performances for inferring the optimal number of clusters would have led to prohibitive computational times, and was beyond the scope of the present study. Further work dedicated to investigating this specific issue would undoubtedly be useful. In particular, the choice of the adequate measure of goodness-of-fit, and the potential impact of maximum likelihood approximation through the EM algorithm should be given further consideration. With this in mind, we implemented four different statistics

measuring the goodness-of-fit of clustering solutions, which should hopefully provide the needed flexibility for future investigations of the 'true K'.

In our simulations, the number of loci and levels of genetic differentiation varied independently, so that the *resolution* of the datasets may not have been sufficient for detecting some of the hybrid classes, especially the second generation backcrosses (Vahä & Primmer 2006). While this was not a problem for comparing the relative performances of *snapclust* and NEWHYBRIDS, ensuring sufficient resolution should be a primary concern in empirical studies. Ideally, further work will formulate guidelines for defining the minimum resolution required for recovering specific hybrid classes. As a pragmatic alternative, we suggest comparing clustering solutions involving different degrees of hybridisation, and selecting the model providing the best fit of the data (e.g., sensu AIC).

The approach described here is flexible, as it can accommodate any type of codominant markers including microsatellites and SNPs, and can readily be extended to varying ploidy levels. Interestingly, it can also be extended to other genetic models as well, including potentially more complex ones. Contrary to Bayesian approaches which can need hundreds of thousands or even millions of iterations to reach mixing and provide a representative sample from the posterior distribution, our fast likelihood maximisation using the EM algorithm converges in a few iterations - typically less than 10 in our simulations. As a consequence, our approach could have great potential for addressing more complex population genetics model, as long as their likelihood is tractable or can be reasonably approximated. One potential obstacle to such extensions is that group memberships need to be treated as a discrete variable, where individuals essentially belong to one group. This will exclude mixture models in which individuals effectively have multiple origins. A workaround for this issue may be to model 'mixed groups' explicitly, as we have done in our hybridisation model.

Our method is implemented in the R package *adegenet* (Jombart 2008; Jombart & Ahmed 2011), which supports a wide range of data including microsatellites, SNPs, and amino-acid sequences, and implements several methods for exploring genetic data (Jombart, Pontier & Dufour 2009), revealing spatial patterns (Jombart *et al.* 2008), or investigating genetic clusters (Jombart, Devillard & Balloux 2010). Interoperability between different tools has been a long standing issue in genetic data analysis (Excoffier & Heckel 2006). We hope the availability of *snapclust* in the same environment as a wealth of other tools for population genetics (Goudet 2005; Paradis 2010; Kamvar, Tabima & Grünwald 2014; Archer, Adams & Schneiders 2017) and phylogenetics (Bortolussi *et al.* 2006; Schliep 2011; Revell 2012; Popescu, Huber & Paradis 2012; Jombart *et al.* 2017) will enhance its usefulness for the community.

3.2.5. ACKNOWLEDGEMENTS

TJ is funded by the Medical Research Council Centre for Outbreak Analysis and Modelling and the National Institute for Health Research - Health Protection Research Unit for Modelling Methodology. MPB and DP are supported by the LabEx ECOFECT (ANR- 11- LABX- 0048). We thank Jérôme Goudet for some insightful discussions on genetic data simulations. Simulation analyses with STRUCTURE and NEWHYBRIDS were performed using the computing facilities of the CC LBBE/PRABI.

3.2.6. APPENDICES

Table S3.2.1: runtimes of the different methods for the clustering and hybridisation analyses. Values indicate the median time computed across all simulations, with the 95% percentile interval indicated between square brackets. Computing times indicated for Structure and Newhybrids correspond to the computing time for one repetition. Time is indicated either in seconds ('s') or in hours ('h').

	Clustering without hybrids	Clustering with hybrids
snapclust	0.89s [0.11s ; 6.45s]	0.54s [0.16s ; 2.27s]
BAPS	24s [7s ; 89.05s]	NA
STRUCTURE	29.36h [2.17h ; 377.25h]	NA
find.clusters	0.15s [0.02s ; 3.09s]	NA
Newhybrids	NA	78.87h [10.15h ; 503.24h]

Table S3.2.2: effect of different variables on the simulation results in the absence of hybrids on the True Positive Rate (*TPR*). This table provides the summary of the final multivariate regression carried out on $\text{logit}(TPR)$. Bold font indicates significant results after Bonferroni correction with a target type 1 error of 1%. Contrasts were set for 'Method' with STRUCTURE as the intercept.

	Estimate	Std. Error	t value	Pr(> t)
Intercept	1.94	0.18	10.74	6.54e-26
Number of loci	0.0043	0.00040	9.93	1.68e-22
Number of populations (K)	-0.038	0.016	-2.44	0.015
Fst	6.61	0.26	25.16	7.74e-116
Method_snapclust	0.023	0.25	0.094	0.93
Method_BAPS	1.0045	0.25	4.095	4.46e-05
Method_find.clusters	0.89	0.25	3.62	0.00031
Number of loci : Fst	-0.0058	0.00070	-8.47	5.88e-17
Fst : Method_snapclust	-0.036	0.33	-0.11	0.91
Fst : Method_BAPS	-0.95	0.33	-2.91	0.0037
Fst : Method_find.clusters	-1.43	0.33	-4.35	1.43e-05
Nb_Loci : Method_snapclust	-0.00010	0.00050	-0.11	0.91
Nb_Loci : Method_BAPS	-0.0014	0.00050	-2.88	0.0041
Nb_Loci : Method_find.clusters	-0.00020	0.00050	-0.32	0.75
Nb_populations : Method_snapclust	0.0074	0.022	0.34	0.74
Nb_populations : Method_BAPS	0.0085	0.022	0.39	0.70
Nb_populations : Method_find.clusters	-0.12	0.022	-5.35	9.98e-08

Table S3.2.3: effect of different variables on the simulation results in the absence of hybrids on the True Negative Rate (*TNR*). This table provides the summary of the final multivariate regression carried out on $\text{logit}(TNR)$. Bold font indicates significant results after Bonferroni correction with a target type 1 error of 1%. Contrasts were set for 'Method' with STRUCTURE as the intercept.

	Estimate	Std. Error	t value	Pr(> t)
Intercept	3.59	0.12	30.79	4.04e-160
Number of loci	0.0038	0.00030	12.60	1.33e-34
Number of populations (K)	0.075	0.0074	10.16	1.91e-23
Dispersal model_stepping stones	-0.095	0.11	-0.86	0.39
Fst	3.77	0.16	23.27	7.79e-102
Method_snapclust	0.014	0.11	0.13	0.90
Method_BAPS	0.33	0.11	3.02	0.0025
Method_find.clusters	-0.56	0.11	-5.14	3.08e-07
Number of loci : Fst	-0.0059	0.00060	-9.08	3.45e-19
Stepping stones : Method_snapclust	-0.0026	0.15	-0.017	0.99
Stepping stones : Method_BAPS	-0.048	0.15	-0.31	0.75
Stepping stones : Method_find.clusters	-0.49	0.15	-3.16	0.0016

Table S3.2.4: effect of different variables on the simulation results in the presence of hybrids, on the probability of correct assignment. This table provides the summary of the final multivariate regression carried out on $\logit(p)$, where ‘ p ’ is the probability of correct group assignment. Bold font indicates significant results after Bonferroni correction with a target type 1 error of 1%. Contrasts were set for ‘Method’ with NEWHYBRIDS as the intercept.

	Estimate	Std. Error	t value	Pr(> t)
Intercept	0.65	0.17	3.85	0.00012
Number of loci	0.011	0.00050	23.32	2.74e-110
Dispersal model	0.12	0.045	2.73	0.0063
Fst	7.96	0.25	31.28	5.094e-185
Hybrid Class F1	-1.26	0.18	-7.18	8.56e-13
Hybrid Class BC1	-4.19	0.18	-23.87	4.82e-115
Hybrid Class BC2	-5.71	0.18	-32.58	2.077e-198
Method_snapclust	-1.43	0.18	-8.13	6.32e-16
Number of loci : Fst	-0.0052	0.0010	-5.046	4.78e-07
Hybrid Class F1 : Method_snapclust	1.63	0.25	6.55	6.59e-11
Hybrid Class BC1 : Method_snapclust	1.98	0.25	7.99	2.014e-15
Hybrid Class BC2 : Method_snapclust	2.99	0.25	12.07	9.019e-33

Table S3.2.5: effect of different variables on the simulation results in the presence of hybrids, on the support to the real class. This table provides the summary of the final multivariate regression carried out on $\text{logit}(p)$, where ' p ' is the support to the true group. Bold font indicates significant results after Bonferroni correction with a target type 1 error of 1%. Contrasts were set for 'Method' with NEWHYBRIDS as the intercept.

	Estimate	Std. Error	t value	Pr(> t)
Intercept	0.40	0.10	3.94	8.26e-05
Number of loci	0.0077	0.00030	26.51	9.85e-139
Dispersal model	0.11	0.027	4.082	4.60e-05
Fst	5.61	0.15	36.50	6.71e-240
Hybrid Class F1	-0.64	0.11	-5.99	2.29e-09
Hybrid Class BC1	-3.033	0.11	-28.62	1.023e-158
Hybrid Class BC2	-3.81	0.11	-36.00	2.14e-234
Method_snapclust	-1.34	0.11	-12.69	6.27e-36
Number of loci : Fst	0.0033	0.00060	5.37	8.72e-08
Hybrid Class F1 : Method_snapclust	0.69	0.15	4.61	4.20e-06
Hybrid Class BC1 : Method_snapclust	1.24	0.15	8.30	1.56e-16
Hybrid Class BC2 : Method_snapclust	1.90	0.15	12.67	7.96e-36

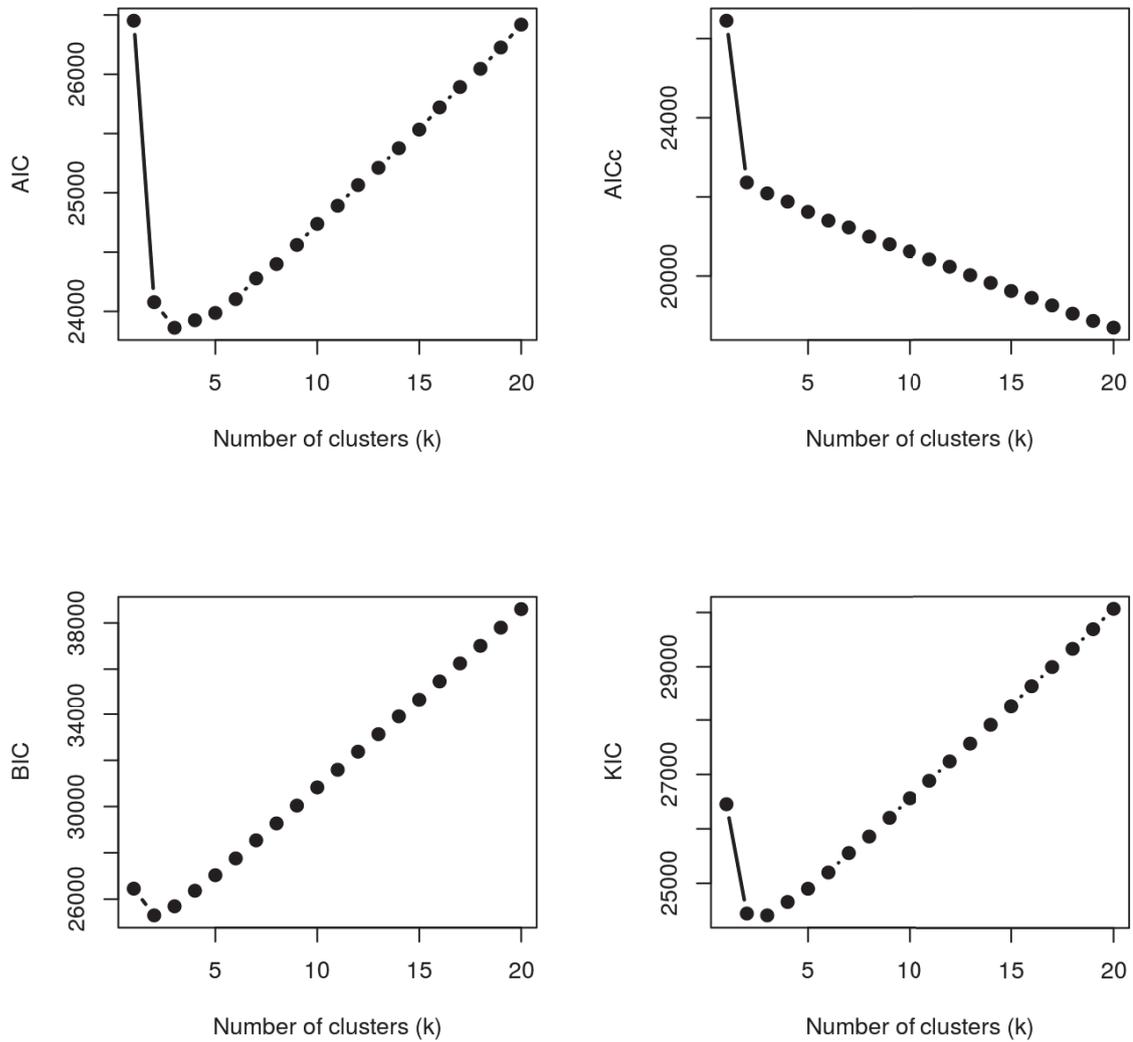


Figure S3.2.1: selection of the optimal number of clusters for the microbov example. This figure shows results of the 4 goodness-of-fit statistics implemented for *snaplust* (AIC, AICc, BIC, and KIC), identifying 2 optimal clusters in from the microbov example dataset.

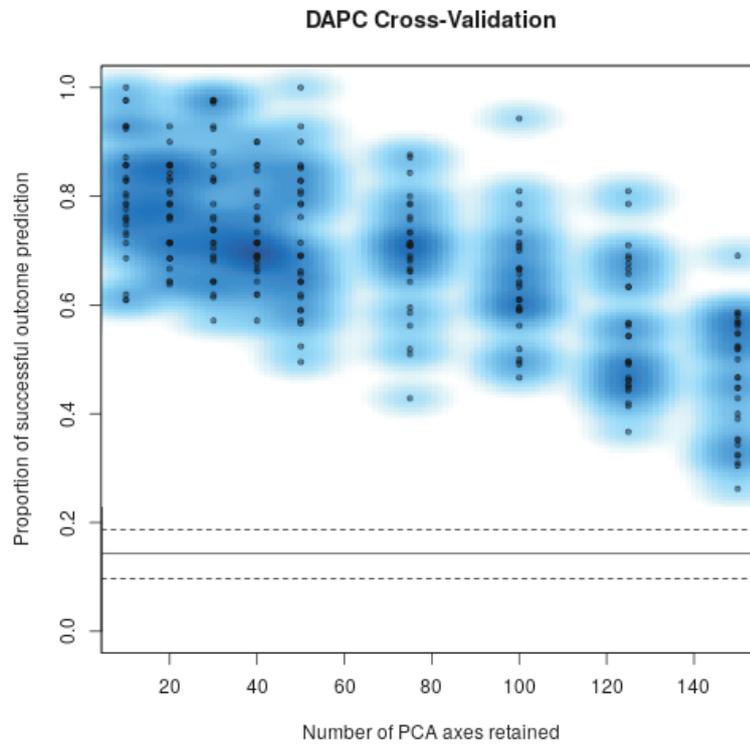


Figure S3.2.2: Output of the cross-validation analysis carried out to determine the number of principal components to retain in the DAPC led on the microbov example. This figure shows the proportion of successful group re-assignment via cross-validation for varying numbers of PCA axes retained, using 90% of individuals for the training set. For each number of PCA axes, 30 independent replicates were obtained. The density of observation is indicated in blue shades. The plain and dashed lines indicate the mean expectation from a random classifier, and its 95% confidence interval.

Addendum

The two methods presented in this chapter address two different issues: how can we improve existing methodologies and which new methods can be developed to better cope with current issues.

The improving of existing method is a never-ending process. In this chapter, we worked on the statistical processing of data but other aspects can be considered to improve the detection of hybrids. Notably, the panel of genetic markers on which rely the statistical analysis can be designed specifically for the study of hybridization through the integration of widely differentiated markers in the panel (Godinho et al. 2011, Nussberger et al. 2013). During my PhD, I came to work on a similar panel of selected markers, developed by Antagene for the detection of hybrids between the wild boar *Sus scrofa scrofa* and the domestic pig *Sus scrofa domesticus* (Beugin et al. 2017, Appendix 3).

The method *snapclust* allows for a quicker detection of hybrids and works efficiently both with SNP and microsatellite markers while recent methodological developments mainly aim at increasing the speed of calculation for SNP data coming from next-generation sequencing. An algorithm close to the *snapclust* algorithm, FLOCK, has been proposed by Duchesne and Turgeon (2009). Although the description of the algorithm lack details, FLOCK seems also to rely on an expectation-maximization (EM) process and to make the assumption of Hardy-Weinberg equilibrium and linkage equilibrium (Paetkau et al. 1995). In addition, the FLOCK algorithm includes a “leave-one-out” step that is not implemented in *snapclust* but that could be interesting as a future development of the method. FLOCK has been proposed as an efficient and much faster alternative to STRUCTURE (Duschene and Turgeon 2009, Anderson and Barry 2015) for the clustering of individuals into sub-populations, similarly to *snapclust* in this chapter (part 2). *Snapclust* offers additional functionalities as it allows for the categorization of individuals into hybrid classes, thus providing a fast alternative to NEWHYBRIDS (Anderson and Thompson 2002). In addition, its implementation in R makes it readily accessible.

The method proposed here leads to assignment probabilities on which we did not apply thresholds. However, the same issue of overlapping between hybrid categories when the differentiation is low exists with this method and thus, the algorithm described in the first part may be, in the future, applied to the *snapclust* method in order to reinforce the reliability of the categorization of individuals into hybrid categories.

4. Opening the ecological black box to understand hybridization

Foreword

Genetic and statistical analysis of biological material allows, as seen before, to assess which individuals are most probably hybrids and which ones are most probably parents (either domestic cats or wildcats in our study case). The rates of hybridization obtained with these methods allow to know if hybridization occurs and in which proportion. However, it does not allow to understand why, under which conditions it occurs and does not allow to assess how changes in the environment may promote hybridization in the future, especially given the on-going global changes.

In order to better understand the determinants of hybridization in the local population from northeastern France, we combined genetic data with radio-tracking to assess the spatial organization of the wildcat population. The first part of this chapter presents in detail this study.

4.1. Female in the inside, male in the outside: insights into the spatial organization of a European wildcat population



Feral domestic cat from a farm in northeastern France – personal photography

ABSTRACT

Hybridization between the European wildcat, *Felis silvestris silvestris*, and the domestic cat, *Felis silvestris catus*, has been found in several European countries with different landscape structures and in various proportions. In this study, we focus on a local population of European wildcats in forests fragmented by agricultural lands in northeastern France. Our aim is to better understand how the spatial organization of the wildcats in this particular type of environment might impact the proportion of hybridization. We combined radio-tracking and genetics through the use of microsatellite markers in order to assess both the spacing pattern and the level of hybridization of this wildcat population. Hybridization is rare in this wildcat population with only one putative hybrid (most likely backcrossed) detected out of 42 putative wildcats. We found that most females were concentrated inside the forest while males stood in the periphery or outside the forest. Furthermore, many males and females resulted related. Such a spacing pattern might limit contacts between male domestic cats and female wildcats and can be one of the causes that explain the low level of hybridization in the wildcat population in this environment. We could not exclude the possibility of hybrid presence in the neighboring domestic cat populations. Our results yield new insights on the influence that the landscape configuration and the spacing pattern can have on genetic flow between the populations of the two subspecies.

Keywords: *Felis silvestris silvestris*, hybridization, microsatellites, kinship, habitat fragmentation, radio-tracking

BEUGIN, M. P., LEBLANC, G., QUENEY, G., NATOLI, E., & PONTIER, D. (2016). Female in the inside, male in the outside: insights into the spatial organization of a European wildcat population. *Conservation Genetics*, 17(6), 1405-1415.

4.1.1. INTRODUCTION

The European wildcat is a medium-sized carnivore that belongs to one of the five *Felis silvestris* sub-species, with a wide geographical distribution (Driscoll et al. 2007). For long time, humans have had conflictual relationships with the European wildcat (Stahl and Artois 1994; Inskip and Zimmermann 2009) that led to the reduction of their number. Nowadays, its range goes from Portugal (Oliveira et al. 2008) to Bulgaria (Petrov et al. 1992; Randi 2008) passing through Scotland (Daniels et al. 2001), and the species is classified as 'Least Concern' by the International Union for the Conservation of Nature (Driscoll and Nowell 2010). The most threatening species for the wildcat is the domestic cat *Felis silvestris catus*, which descends from the African wildcat *F. s. lybica* (Vigne et al. 2004; Driscoll et al. 2007). The domestic cat is increasingly appreciated as pet and always for its ability to control rodents (Loss et al. 2013) in rural environment, as it was in the past. In France, it is estimated that eleven millions of domestic cats are owned in 2012 (FACCO 2012) for 65 million inhabitants. This estimation does not take into account feral and stray cat populations. Both are domestic cats, the former are born in or have reverted to a wild state, while the latter have not had socialization with humans but live around rural properties (Bradshaw et al. 1999; Medina et al. 2014). The growing popularity of the domestic cat as pet increases the population of feral and stray cats directly through uncontrolled births that in turn increase uncontrolled adult feral and stray cats (e.g., Hellard et al. 2012). In rural habitats, both stray and feral cats might favor hybridization between the wild and the domestic cat.

Hybridization between these two sub-species has been found in several European countries in different proportions (Beaumont et al. 2001; Pierpaoli et al. 2003; Lecis et al. 2006; Oliveira et al. 2008; Hertwig et al. 2009; Say et al. 2012; Nussberger et al. 2014b) and has probably led to the extinction of 'true' European wildcats in Scotland (Beaumont et al. 2001; Daniels et al. 2002). The heterogeneity of hybridization importance suggests that genetic compatibility is not enough to explain the occurrence of hybridizing events and that

environmental parameters such as the landscape configuration might play a substantial role.

For hybridization to happen, individuals from the two sub-species have to meet. In France, the areas of distribution of the two sub-species are considered to overlap almost entirely (Léger et al. 2008). Encounters and hybridization are expected to occur during the mating season when both feral cats and wildcats have been observed to roam outside their usual home range in order to find a sexual partner (Corbett 1979; Ferreira et al. 2011). However, the assessment of the sub-species areas does not contain information on their spacing pattern and/or on the spatial relationships between males and females (and on the resulting mating system; Corbett 1979). Moreover, most studies focused on individual movements in their environment (Biró et al. 2004; Germain et al. 2008; Jerosch et al. 2010) but few considered the whole population and investigated how individuals interact (e.g., Klar et al. 2008).

Here, we analyzed the spacing pattern of a population of wildcats at a fine geographic scale in Lorraine (France) where the landscape is structured with fields and forests and which constitutes the historical area of distribution of the European wildcat in France (Léger et al. 2008). We combined genetic analyses and radio-tracking in order to, firstly, distinguish wildcats, domestic cats and hybrids and, secondly, to analyze the spacing pattern of the wildcats. Furthermore, we analyzed the genetic relationships of wildcats in order to understand if the kinship influenced their spatial distribution. Then, matching these results our aim was to assess whether the spacing pattern of this wildcat population may explain the observed rate of hybridization.

4.1.2. MATERIALS AND METHODS

a. Study area and data collection

The study took place in Northeastern France (5°45'51.0"E, 48°31'04.3"N) between April 2008 and May 2011. All the cats of the study were trapped and radio-tracked in an area of 130 km². We further used this area to assess the density of the wildcat population. The landscape consists of an alternation

between forests, agricultural fields and permanent grasslands with an altitude comprised between 300 and 400m in forest and between 250 and 300m in the villages surrounding it. A total of sixteen villages, with a mean density of 35 inhabitants per km² (less than 600 inhabitants per village, 228 on average), were in direct proximity with the central forest where wildcats were sampled (Figure 4.1.1). The fragmentation in this region was estimated as being substantial (Cemagref 2010). The local climate is semi-continental. The precipitations are abundant and regular all along the year. The average temperature in summer is 18.5-19°C. The average winter temperatures approach 1°C in January.

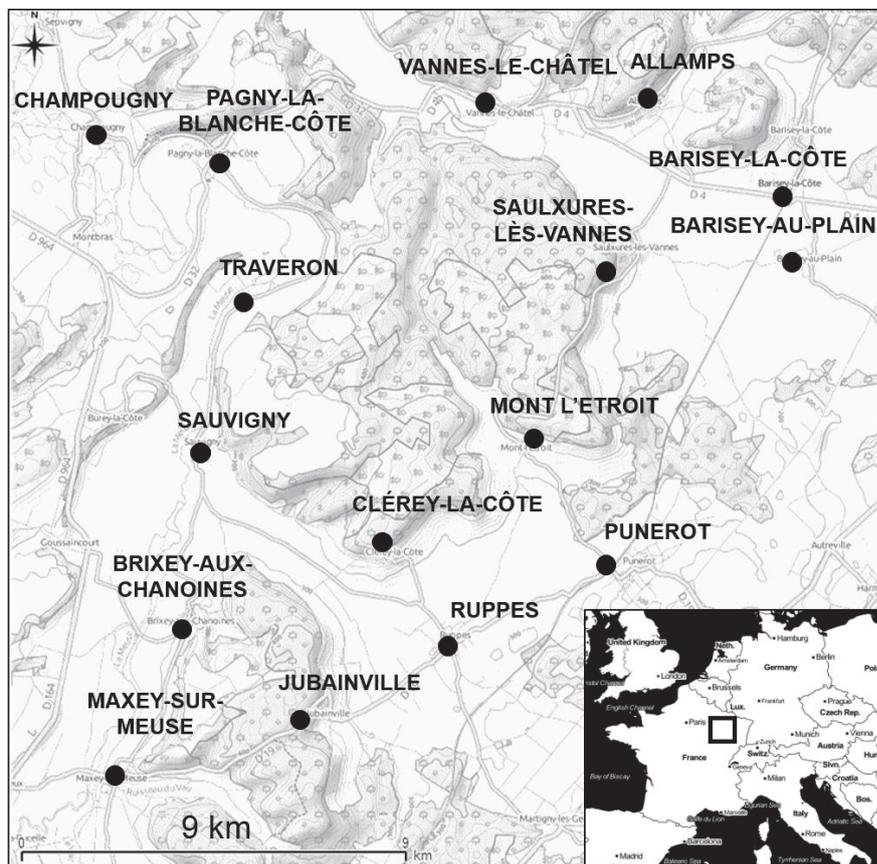


Figure 4.1.1: Distribution of the villages surrounding the wildcat population (forests in grey) where all the domestic cats were trapped (data from QGIS).

The fieldwork has been conducted by qualified people according to current French legislation. Accreditation has been granted to the UMR-CNRS5558 (accreditation number 692660703) for the program. Cats were captured using trapping cages containing crushed valerian roots (*Valeriana officinalis*), a common attractant for cats. As it mimics cats' sexual secretions, valerian's use is bound to the mating season (Kilshaw et al. 2015). All captures were then made from November to February of each year. Trapped individuals were anaesthetized with ketamine chlorohydrate (Imalgène 1000, 15mg/kg, Merial) and acepromazine (Vétranquil 5.5%, 0.5 mg/kg, Ceva). A permanent subcutaneous electronic device (transponder Trovan, AEG & Telefunken Electronic, UK) was injected in each cat to aid subsequent identification of each individual. Sex, age, pelage and morphological characteristics including body weight (using a hanging scale, $\pm 0,3\%$) and cranial measurements were assessed. A blood and hair samples were collected for further genetic analysis. We determined wildcats' age based on teeth and morphological characteristics. All cats presenting the typical wildcat phenotype (specific coat color; Leger et al. 2008) were classified as wildcats and this classification was confirmed or rejected based on the result of the genetic analysis. Individuals were released at the site of capture. Thirty-two wildcats (20 females, 12 males) were caught alive and the eighteen first caught individuals (14 females, 4 males) were equipped with VHF radio transmitters (Biotrack, UK). The dead bodies of ten wildcats (3 females, 7 males) were collected on the road following car crush accidents. Thirty cats with a typical domestic phenotype were captured using baited traps in the villages surrounding the forests. All cats were feral (n=19, 14 females, 5 males) or domestic with free access to the outside (n=11, 6 females, 5 males). The same protocol followed for wildcats was applied for domestic cats.

b. Genetic analysis: subspecies characterization and kinship determination

The forty-two wildcats were genotyped using thirty-one microsatellites and the thirty domestic cats using twenty-five (Menotti-Raymond et al. 1999; O'Brien et al. 2009, Appendix, Table S4.1.1). DNA extraction was performed using a purification column kit (Nucleospin 96 Tissue kit, Macherey-Nagel). DNA was then amplified and analyzed using an ABIPRISM 3130XL Applied Biosystem DNA sequencer. Results of sequencing were read using GeneMapper v.4.1 (Applied Biosystem/Life Technology).

The genetic identification of the species to which each individual belonged was performed by employing the Bayesian analysis implemented in STRUCTURE v.2.3 (Pritchard et al. 2000; Falush et al. 2003) based on the twenty-five microsatellites genotyped for both the wild and domestic cats. The analysis was performed under a model allowing admixture and using no prior information of phenotypic classification. Individuals were then assigned to the *F.s. silvestris* or to the *F.s. catus* clusters according to the posterior probabilities obtained. We performed twenty independent runs, each with values of K assumed genetic groups ranging from 1 to 4. The expected value of K was 2 but we ran K to higher values to discard the eventuality of a third potential hybrid cluster. According to Gilbert et al. (2012)'s suggestions, we performed 500,000 Markov Chain Monte Carlo iterations after a burn-in period of 100,000. The convergence of the algorithm was checked visually using STRUCTURE run-time plots. We determined the number of clusters following the method of Evanno et al. (2005) using STRUCTURE HARVESTER online web 0.6.94 (Earl and VonHoldt 2011) after running CLUMPP v.1.1.2 (Jakobsson and Rosenberg 2007) in order to obtain an average matrix over the twenty simulations. We considered as hybrid any individual for which the posterior assignment probability was lower than 0.91 according to the maximum threshold used for the detection of hybrids in wildcats (between 0.80 and 0.91; Mattucci et al. 2013; O'Brien et al. 2009; Oliveira et al. 2008). Since our aim was to check whether individuals morphologically classified as wildcats were truly wildcats, we chose the highest threshold in order to be conservative. For each population, deviations from

Hardy-Weinberg equilibrium and linkage disequilibrium were tested using FSTAT v.2.9.3.2 (Goudet 1995). The presence of null alleles or other scoring errors was evaluated using MICROCHECKER v.2.2.3. (Van Oosterhout et al. 2004). We estimated F_{st} values according to Weir and Cockerham's (1984) version of Wright's F-statistic using GENETIX v.4.05.2 (Belkhir et al. 1996).

We used ML-RELATE (Kalinowski et al. 2006) software and estimated genetic relationships between all dyads of wildcats, based on 31 microsatellites. This program uses the maximum likelihood estimate of relatedness between pairs of individuals to discriminate between the relationships: Unrelated (U), Parent-Offspring (PO), Full sib (FS), Half sib (HS). We tested all assignments to one of these categories compared to the alternative ones using 1,000 simulations, which gave us a probability that the determined category is the correct one (Appendix, Table S4.1.2). We discarded any pairwise relationship for which the p -value of the determined category against unrelated (U) was higher than 5%. When two individuals were categorized as FS and when their age differed by at least one year, kinship was assumed to be PO. In order to test whether relatives are geographically closer together compared to non-relatives, we calculated the pairwise geographic distance between individuals using the program SpaGeDi (Hardy and Vekemans 2002). Then, we statistically assessed if the pairwise distance between relatives was significantly different from the pairwise distance between non-related individuals using a Mann-Whitney test. For this spatial analysis, we discarded seven individuals out of forty-two for which we did not record the coordinates with a GPS.

c. Spatial organization and home-range analysis

We located each of the 18 equipped wildcats at least once per week (mean = 1.24, sd = 0.18) by triangle procedure using hand-held antenna. Individuals were monitored all along the year, including during the mating season (November to February). For each individual, we had a number of locations over a year comprised between 52 and 81. Locations were positioned on a map using Quantum GIS (Quantum GIS Development Team 2012) and the plugin open layers available in the software. In order to assess the minimal distance

between each village and the locations recorded, we calculated the distance matrix with QGIS between each location and the periphery of urban areas. Urban areas were previously defined by building a polygon circling all the infrastructures inside each village. A random distribution of 1546 locations was simulated. A distance matrix was also calculated and further compared to the distances obtained with our dataset with a Mann-Whitney test. We also estimated the number of locations monitored within the forests. Only the forests, and not the wooded local areas such as field borders, were taken into account for this calculation. We used the package `adehabitatHR` (R Development Core Team 2010; Calenge 2011) to estimate the annual home-range size using the Minimum Convex Polygon estimator (MCP). We set the percentage parameter to 95% as usually recommended in order to overcome possible bias due to an occasional displacement of the individual outside its home-range. Only individuals spotted in the same area all along their monitoring were taken into account. We did not have enough locations to establish the accurate home-range size, which would be comparable with other studies, except from one male (M2, Appendix Figure S4.1.1). However, the estimations obtained allowed us to compare home-range sizes of individuals within this study. We chose to compare only adults over a period of one year (Appendix, Figure S4.1.2).

4.1.3. RESULTS

a. Genetic characterization of the two subspecies

More than 50% of the loci was amplified in all domestic and wildcats, consequently they were all included in STRUCTURE analysis. The optimal value of K was 2 (Appendix, Figures S4.1.3, S4.1.4). All individuals classified as wildcats according to their external phenotypic characteristics were assigned to one cluster with posterior probabilities greater than 0.8 while all presumed domestic cats were allocated to the second cluster with probabilities greater than 0.97. Then, the phenotypic classification coincided to the wildcat genotype. The two clusters were highly differentiated with an F_{st} value of 0.16 (Confidence

Interval at 95%: 0.12 – 0.2). All domestic cats were trapped outside the forest, at the edge or inside the villages (Appendix, Figure S4.1.5).

Among the forty-two wildcats, only one was identified as hybrid (F9C) with a posterior probability of 0.82 for the wild cluster in STRUCTURE. This individual was discarded from the following analyses (Appendix, Table S4.1.3 and Table S4.1.4 for detailed information about individuals used in the different analyses). Therefore, this population of wildcats was mostly constituted of ‘true’ wildcats and was not a swarm of hybrids. We did not detect neither departure from Hardy-Weinberg equilibrium for any of the loci nor linkage disequilibrium both in the domestic and wildcat populations. Analysis with MICRO-CHECKER did not indicate the presence of null alleles in the wildcat population, but it suggested the presence of null alleles at 3 loci (Fca45, Fca96 and Fca577 with estimated null allele frequencies of 0.077, 0.13 and 0.10, respectively) in the domestic cat population.

We found no significant difference in body weight between domestic and wildcats (Figure 4.1.2). The age of trapped domestic cats ranged between 1 and 7 years old while for wildcats, the estimated age varied between 1 and 4. These two variables, body weight and age, were strongly correlated (Spearman rank correlation test, $Rho = 276.46$, $p = 0.00085$).

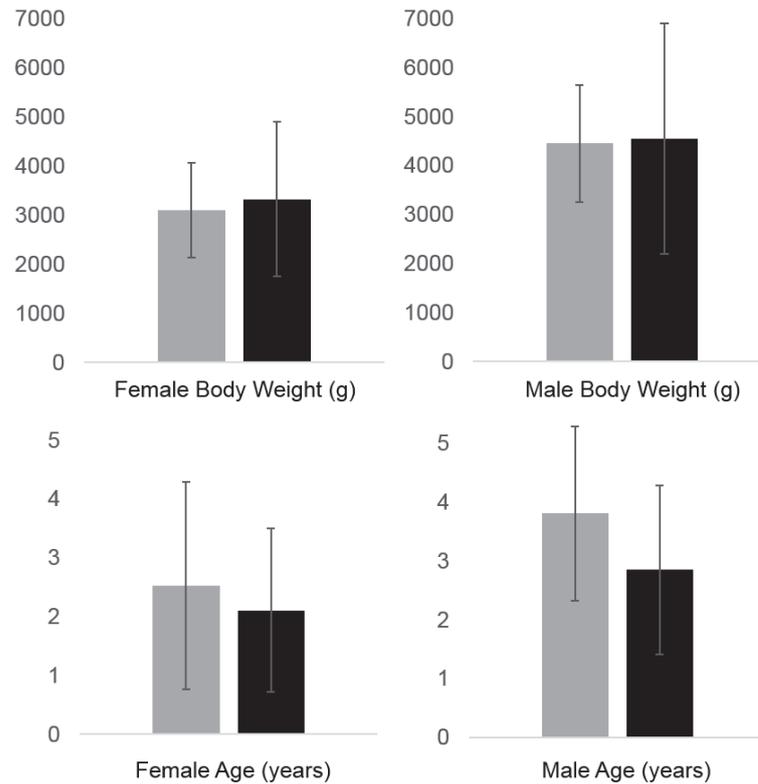


Figure 4.1.2: Body weight and age according to the sex of individuals and their sub-species (Grey: *Felis s. catus*; Black: *Felis s. silvestris*). Errors bars stand for standard deviation.

b. Kinship analysis

We found that 90% of the wildcats were related to at least another individual (Table 4.1.1). Most of the pairwise relationships identified (79.2%) were poorly defined (corresponding to HS) and the remaining ones were equally distributed between PO and FS relationships. Fifty per cent of the pairwise kinships involved two females, 34.9% a male and a female, and 15.1% involved two males. Males were related with less than two individuals on average (mean = 1.89, sd = 1.633) while females had three relatives on average (mean = 3.23, sd = 1.66). The lower number of relatives for males was significant (Mann-Whitney test: $w = 309.5$, $p = 0.0075$). We found three males (M22, M29, M33) and one female (F13) who were not related to any other individual.

Table 4.1.1: Related individuals classified depending on the kind of kinship relation for each wildcat by the software ML_RELATE. The sex of each individual is indicated by F for females and M for males. For each individual, we indicate if he is linked by a PO, FS or HS relationship with any individual. The direction of the PO relationships is given in parenthesis: C mean “Child”, P means “Parent”, “?” means that the determination of the age was not precise enough to draw any conclusion. F15 and F18 individuals were initially assigned to FS by ML-RELATE, we set their kinship to PO as one was fully adult (F15) and the other one juvenile (F18).

Ind.	PO	FS	HS
M1C		M4	
M2C			F19
M3C	F5C (P)	F4C	F28, F5, M11
M6C			M14
M7C		M10C	M8C
M8C			F28, M10C, M7C
M10C		M7C	M8C
M2		F31	F32, F5
M4		M1C	
M11			M3C, F4C, F3
M14			M6C, M27, M30, F25
M16			F24
M17	F12 (C)		
M22			
M26			F5C, F32, F19
M27			F23, M14
M29			
M30			F25, M14
M33			

Ind.	PO	FS	HS
F4C	F5C (P)	M3C	F28, F5, F7, M11
F5C	M3C (C), F4C (C)		F28,F5, M26
F3		F8	M11
F5			M2, M3C, F4C, F5C, F28, F32
F6	F7(C),F9(?)		F20, F21
F7	F6 (P)		F9, F4C, F31
F8		F3	F24
F9	F6(?)		F7, F21, F31
F10			F15, F28, F31
F12	M17 (P)		F28
F13			
F15	F18 (C)		F20, F10
F18	F15 (P)		F24,
F19			M2C, M26
F20			F19, F6, F15
F21			F6, F9
F23			F25, M27
F24			F8, F18
F25			M30, F23, M14
F28			F5C,F5, F10, F12, M3C, F4C, M8C
F31		M2	F7, F9, F10
F32			F5, M2, M26

Relatedness appeared to be correlated with the pairwise geographic distance only when considering female-female related couples. The existence of parentage between two females (PO, FS and HS taken altogether) was indeed correlated with a smaller geographic distance between those females ($w = 935.5$, $p = 0.0029$) while it was not significant when considering related couples involving males ($p = 0.69$). This resulted in the occurrence of 'related units' made of related females with neighboring home-ranges (see Figure 4.1.3.a.). Two individuals belonged to the same related unit when, in addition to being geographically close, they were strongly related (PO/FS). HS relationships allowed an individual to be considered as part of a related unit when the HS kinship was retrieved for several individuals of the unit. No related units that included adult males were observed.

c. Spatial organization of the population

The density in our area was estimated to 0.32 cats per km². Wildcats were not homogeneously distributed across the massif. They remained significantly further from the villages than by chance ($w = 167$, $p < 0.0001$). On average, the minimal distance between a cat and a village was of 1km ($sd = 0.73$, $min = 52m$, see appendix, Figure S4.1.6). Also, 92% of the locations were monitored inside the forests, forests that occupied about 28% of the area of study. All the fourteen females radio-tracked had settled inside the forest with some overlapping between their home ranges; some of the home-ranges included the ecotone between the forest and the crops (Figure 4.1.3.b.). Only one female (F23) was caught clearly outside the forest (see Figure 4.1.3.c.). Home ranges of males were systematically on the border of the forest, between two forests (M2) or outside the forest (M11). No male wildcat was captured in the heart of the forest. During the time of the study two males (M4, M17, see Figure 4.1.3.d.) changed their home range from year to year because they were probably dispersing.

One male over the four died during the monitoring so only three equipped males had enough locations to allow home-range size estimation, but two of them (M4, M17) were dispersing during the study. For male M2, both a global

home-range (over the whole period of radio-tracking) and an annual home-range (year 2009) were calculated, the latter for comparison with those of females. Eight females matched the requirements defined (i.e., enough locations over a complete year, see appendix Table S4.1.3, S4.1.4). The annual home-range size estimated for the male M2 (422.11 ha) was larger than all estimated female annual home-range sizes (mean = 130.82, sd = 39.82, see Figure 3.b.). The global home-range size of the male M2 was 347.57 ha.

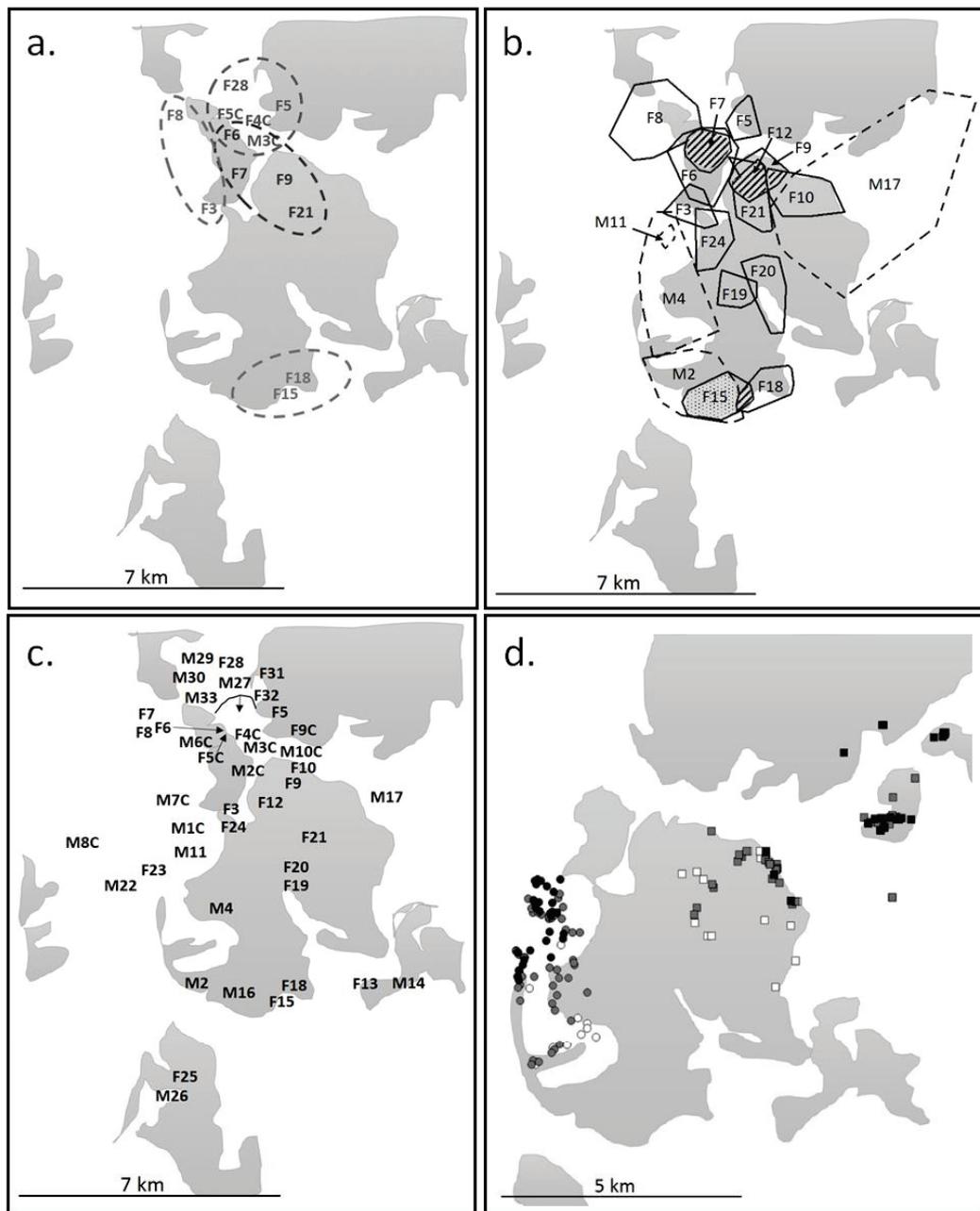


Figure 4.1.3: Spatial organization of the wildcat population. Forests are represented in grey. White areas are composed of fields and urban areas. a. Representation of the four main related units identified by the kinship analysis. Only the thirteen females or kittens defining these units are represented. Each unit is shown by the dotted circle that encloses the individuals of the unit. The four units are: F3-F8, F6-F7-F9-F21, F5-F28-F5C-F4C-M3C, F18-F15. b. Home-range represented by 95% minimum convex polygons for individuals (15 individuals) that did not die during the monitoring for males (dotted lines) and

females (solid lines). Substantial overlapping between females (hatching areas) and between a male (observed only for the male M2) and females (dotted areas) are represented. c. Trapping locations of all the wildcats. d. Movement per semester of the two migrating males. One (M4) is shown by filled circles, the other one (M17) with filled squares. For both the monitoring lasted three semesters. White spots represent the first semester, grey spots the second semester and black spots the last one.

4.1.4. DISCUSSION

The combination of genetic analyses and radio-tracking of the wildcats allowed us to estimate the level of hybridization in this environment and shed light on remarkable features of the spacing pattern that might explain the low proportion of hybrids.

a. Density and individual spacing pattern

Our wildcat population density (0,32 cats.km⁻²) falls within the values found in other areas (0.16-0.44 cats.km⁻², Dimitrijevic 1980; 0.17-0.25 cats.km⁻², Heller 1992; Okarma et al. 2002, 0.1-0.13 cats.km⁻²) and it is coherent with studies carried out in similar landscapes in France (0.2-0.5 cats/km²; Stahl & Léger 1992) and in Switzerland (0.35 cats.km⁻²; Weber et al. 2008). This density might be underestimated, as it was not possible to trap all individuals in spite of an important capture effort.

Our results showed substantial differences in the spacing pattern between females and males. We found that all the females concentrate inside the forest. Females are known to be highly territorial (Biró et al. 2004) and the rarity of cases of overlapping in the population studied here confirms this statement. Most female home ranges allow direct access to agricultural fields or grasslands (see Figure 4.1.3.b., Figure S4.1.5). While forests provide a shelter against many different threats, fields are food-rich areas attracting rodents that cats can hunt. Such alternating between resting sites and hunting grounds has been previously proposed to be the optimal environment for wildcats (Lozano et al. 2003).

On the contrary, we found that all the males were trapped and/or located with telemetry outside or in the periphery of the forest all year round.

Unfortunately, we do not have enough data on male home ranges, but the fact that no males were captured or detected within the forest raises questions about the observed population spatial structure. The greater dependence of females on forest compared to males has also been reported in a quite similar habitat in Germany by Klar et al. (2008). The pattern we observe might be related to different requirements between sexes. Availability of prey might be more important in males than the protection provided by the forest cover. On the opposite, females that do all the rearing of offspring might privilege shelters and forest ecotones often richer in small prey abundance and diversity than interior forest (Doyle 1990; Gomez and Anthony 1998; Osbourne et al. 2005). Finally, as expected (Daniels et al. 2001; Biró et al. 2004), males have probably larger home-range than females and may include several female home-ranges, but the shortage of data on males does not allow us to confirm this pattern (see Figure 4.1.3.b.). The home-range (347.57 ha over the whole monitoring period - 422.11 ha over one year) of the only male (M2) included in our analyses was three times larger than that of females and overlapped part of the home-ranges of two females (F15, F18). Its size falls within the range (from 170 to 1000 ha) found in the literature (Corbett 1979; Stahl 1986; Daniels et al. 2001; Biró et al. 2004; Piñeiro & Barja 2011; Kilshaw et al. 2015).

b. Kinship structure

The wildcat population exhibited strong relatedness but the structure differed in males and females. If most females resulted to be related among them and to males, males were not related among them. In addition, related females tended to remain in the same area while kinship did not affect the spatial distribution of males: related and unrelated males were found at comparable distances. This pattern might reflect a tendency in males to disperse more and over larger distances than females (if they disperse). The hypothesis is reinforced first by the high percentage of males (70%) killed from collisions with vehicles, then by the fact that the two individuals who changed their home range year after year were both males. Such a male-biased dispersal behavior is a common feature in felids (Pusey and Packer 1987; Janečka et al. 2007; Croteau

et al. 2010) and in particular in domestic cats (Devillard et al. 2004; Hansen et al. 2007).

We did not find a clear pattern between relatedness and home ranges overlapping in females. We observed four cases of substantial overlapping between females. In one case the two females (F9 and F10, four and three years old, body weight of 4 kg and 3.1 kg, respectively) were not related. The overlapping lasted only one spring during which two other neighboring females were located outside the usual area where they had been monitored until then. This might reflect a temporarily environment disturbance more than a relevant behavioral pattern. In the three other cases home range overlap was observed between related females. Two of them occurred most probably between a mother and her daughter (F6-F7, F15-F18) but the kinship for the last one was unclear (F9-F21). In each case the daughter was a young adult (one year old) and a kitten (not trapped) was observed close to one of the daughters. These observations suggest that prey might be abundant enough in this habitat resulting in greater tolerance for overlap between females and their relatives.

c. Hybridization

We found only one putative hybrid, most probably backcrossed, out of the 42 putative wildcats, suggesting a negligible impact of hybridization on this area (2.3%). No hybrid individuals were detected in the 30 domestic cats. Our local hybridization rate confirmed what was found at a larger scale in France (O'Brien et al. 2009), and is among the lowest rates found in Europe (from 2.1% in Italy, Randi et al. 2001, to 26.1% in Hungary, Pierpaoli et al. 2003). Such a low hybridization rate has been found in similar environments in eastern Germany (4%, Hertwig et al. 2009) and Italy (8%, Lecis et al. 2006). Previous studies have found various levels of differentiation between wildcats and domestic cats (F_{st} varying from 0.11-0.12: Hertwig et al. 2009, Beaumont et al. 2001, Pierpaoli et al. 2003, Randi et al. 2008, to 0.20-0.22: Oliveira et al. 2008, Mattucci et al. 2013). The differentiation ($F_{st}=0.16$) we found in our study between wildcats and domestic cats is then substantial when compared to what has been

observed in various locations and confirms that hybridization is rare in our study area.

In this type of environment, crops act as “corridors” between the forest and the villages, allowing physical meeting between domestic and wildcats. During the same three-year study (2008-2011), 474 domestic feral and stray cats have been captured in the different villages surrounding the forests (Hellard et al. 2012). Both subspecies share the same diet based on rodents (Sarmiento 1996; Malo et al. 2004; Germain et al. 2009; Piñeiro and Barja 2011) and both utilize agricultural fields as hunting grounds (G. Leblanc, personal observations). Furthermore, since the breeding season of the two sub-species partially overlap (November-February for the wildcats; February-September for the domestic cats, Condé & Schauenberg 1974; Gagnon & Dantzer 2013), conditions for extensive hybridization would be met in this environment.

The specific wildcat population spacing pattern, with a sex-biased distribution of the wildcats inside/outside the forest (females inside and males on the edges) might limit mating between female wildcats and male domestic cats. In addition, male wildcats are known to be strongly territorial with large territories encompassing smaller female ones (Klar et al. 2008). Although we did not have enough male wildcats monitored to confirm this pattern, male wildcats can be effective to limit the possibility of mating between male domestic cats and female wildcats living in their territory. On the contrary, we cannot exclude that male wildcats can mate with female domestic cats. Such asymmetry in the hybridization pattern has been suggested for wildcats in Swiss Jura (Nussberger et al. 2014b) based on mtDNA analysis. No wildcat phenotype was recognized in the captured 474 domestic cats, but genetic analyses are going to be achieved to test for this hypothesis in the near future (Beugin et al. in preparation).

4.1.5. CONCLUSION

To conclude, the radio-tracking of a local population of wildcats combined with genetic analysis allowed us to shed light on the spacing pattern of the wildcat in a mixed forest/crop habitat, never described previously. Males' kinship was uneasy to trace, probably due to their high dispersal over long distances, while we identified related females living in close proximity, suggesting a phylopatric behavior. Although the data is not sufficient to draw definitive conclusions, the pattern found can partly explain why the level hybridization is low in France, despite the claimed complete overlapping of wild and domestic cat areas. Our results suggest that to understand the underlying mechanisms responsible for hybridization in a species it is essential to analyze the situation at a fine geographic scale, which put in evidence possible barriers between the sub-species due, for example in this case, to habitat choice. More studies combining genetics and animal radio-tracking should be led in order to assess how unique this spatial organization is, how much it depends on the landscape structure and to investigate further if it limits the risk of hybridization in the wildcat.

4.1.6. ACKNOWLEDGMENTS

We thank Elodie Antoine for her help in monitoring wildcats and Dominique Allainé for helpful comments. This study was funded by the FIV-VAX Lyon Biopôle project.

4.1.7. APPENDICES

Table S4.1.1: Microsatellite markers used for the species identification and the parentage analysis (Menotti-Raymond et al. 1999; O'Brien et al. 2009).

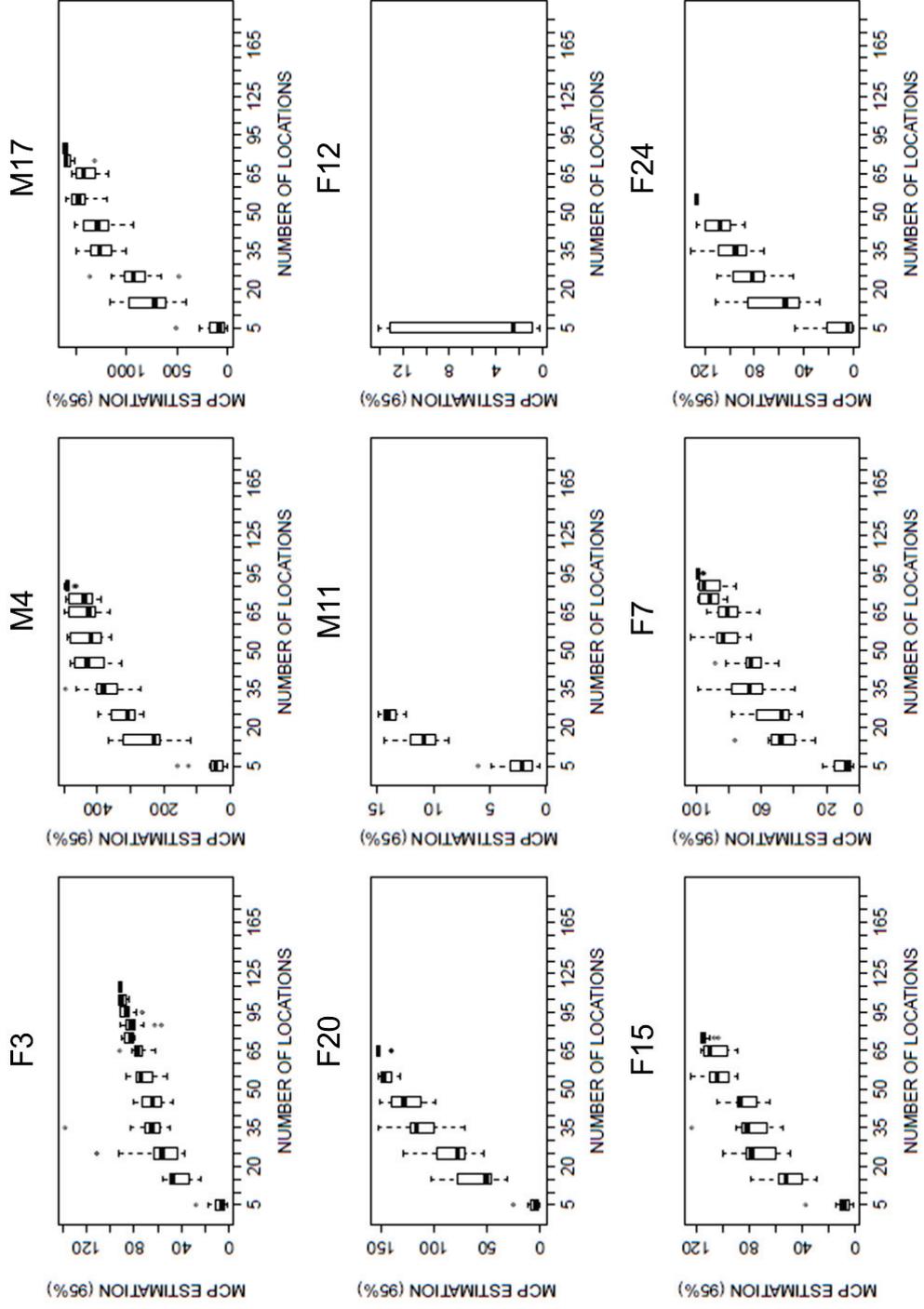
MARKER NAMES	SPECIES IDENTIFICATION	PARENTAGE ANALYSIS
Fca8	yes	yes
Fca26	yes	yes
Fca43	yes	yes
Fca45	yes	yes
Fca58	yes	yes
Fca77	yes	yes
Fca96	yes	yes
Fca124	yes	yes
Fca126	yes	yes
Fca547	yes	yes
Fca577	yes	yes
Fca668	yes	yes
Fca675	yes	yes
Fca031	no	yes
Fca024	no	yes
Fca078	no	yes
F37	no	yes
Fca023	no	yes
Fca085	no	yes
Fca069	yes	yes
Fca105	yes	yes
Fca201	yes	yes
Fca220	yes	yes
Fca229	yes	yes
Fca293	yes	yes
Fca310	yes	yes
Fca453	yes	yes
Fca678	yes	yes
Fca075	yes	yes
Fca149	yes	yes
Fca441	yes	yes

Table S4.1.2: Probabilities of belonging to the assigned category compared to the three other relationship categories proposed by ML-RELATE for the retained pairwise kinships.

Ind 1	Ind 2	Assessed relationship	vs PO	vs FS	vs HS	vs U
F12	M17	PO	-	0	0.003	0
F4C	F5C	PO	-	0	0	0
F5C	F4C	PO	-	0	0.001	0
F5C	M3C	PO	-	0.001	0	0
F6	F7	PO	-	0	0	0
F6	F9	PO	-	0	0	0
F7	F6	PO	-	0.002	0.001	0
F9	F6	PO	-	0	0	0
M17	F12	PO	-	0.002	0.001	0
M3C	F5C	PO	-	0	0	0
F15	F18	FS	0	-	0.022	0
F18	F15	FS	0	-	0.026	0
F31	M2	FS	0	-	0.007	0
F4C	M3C	FS	0	-	0.001	0
F8	F3	FS	0	-	0.064	0
M10C	M7C	FS	0	-	0	0
M1C	M4	FS	0	-	0	0
M2	F31	FS	0	-	0.011	0
M3C	F4C	FS	0	-	0.004	0
M4	M1C	FS	0	-	0	0
M7C	M10C	FS	0	-	0	0
F3	F8	FS	0	-	0.059	0.001
F10	F15	HS	0	0	-	0
F18	F24	HS	0	0.025	-	0
F24	F18	HS	0	0.027	-	0
F28	F5C	HS	0	0	-	0
F5C	F28	HS	0.005	0	-	0
F7	F9	HS	0	0.001	-	0
F9	F7	HS	0	0	-	0
M10C	M8C	HS	0	0.001	-	0
M27	M14	HS	0	0.006	-	0
M7C	M8C	HS	0	0.002	-	0
M8C	M10C	HS	0	0	-	0
M8C	M7C	HS	0	0.004	-	0
F15	F20	HS	0	0.011	-	0.001
F21	F9	HS	0	0.009	-	0.001
F31	F7	HS	0	0.001	-	0.001
F15	F10	HS	0	0	-	0.002
F19	M2C	HS	0	0.005	-	0.002
F5C	F5	HS	0	0.002	-	0.002
M14	M27	HS	0	0.01	-	0.002
F10	F31	HS	0	0.003	-	0.003
F28	F4C	HS	0	0	-	0.003
F31	F10	HS	0	0.02	-	0.003
F5	F5C	HS	0	0.003	-	0.003
F9	F21	HS	0	0.01	-	0.003
M2C	F19	HS	0	0.007	-	0.003
F20	F15	HS	0	0.013	-	0.004
M11	M3C	HS	0	0.035	-	0.004
M3C	M11	HS	0	0.033	-	0.004
F25	F23	HS	0	0.009	-	0.006
F4C	F28	HS	0	0.021	-	0.006
F20	F6	HS	0	0.002	-	0.007

Ind 1	Ind 2	Assessed relationship	vs PO	vs FS	vs HS	vs U
F31	F9	HS	0	0	-	0.008
F6	F20	HS	0	0.003	-	0.008
M14	M30	HS	0	0.031	-	0.009
M26	F19	HS	0	0.01	-	0.01
F7	F31	HS	0	0	-	0.011
F9	F31	HS	0	0.003	-	0.011
M2	F5	HS	0	0	-	0.011
F3	M11	HS	0	0.095	-	0.012
F4C	M11	HS	0	0.005	-	0.012
F5	M2	HS	0	0	-	0.012
F23	F25	HS	0	0.004	-	0.013
F19	M26	HS	0	0.012	-	0.014
M30	M14	HS	0	0.027	-	0.014
F5	M3C	HS	0	0.001	-	0.016
M11	F3	HS	0	0.001	-	0.016
F5	F4C	HS	0	0	-	0.018
F4C	F5	HS	0	0	-	0.019
F25	M30	HS	0	0	-	0.021
F28	F5	HS	0	0.012	-	0.022
M11	F4C	HS	0	0.027	-	0.025
M6C	M14	HS	0	0	-	0.026
F5C	M26	HS	0	0.001	-	0.029
F32	M26	HS	0	0	-	0.03
M26	F32	HS	0	0	-	0.03
M3C	F5	HS	0	0	-	0.03
M3C	F28	HS	0	0	-	0.031
F5	F28	HS	0	0.003	-	0.032
M30	F25	HS	0	0	-	0.032
M2	F32	HS	0	0	-	0.033
F10	F28	HS	0	0.004	-	0.034
F28	F10	HS	0	0.002	-	0.034
M14	F25	HS	0	0	-	0.034
M14	M6C	HS	0	0.002	-	0.034
M27	F23	HS	0	0	-	0.034
F25	M14	HS	0	0	-	0.035
M8C	F28	HS	0	0.001	-	0.036
F28	M3C	HS	0	0	-	0.038
F28	M8C	HS	0	0	-	0.039
F32	F5	HS	0	0	-	0.039
F5	F32	HS	0	0.001	-	0.039
F7	F4C	HS	0	0.001	-	0.04
M26	F5C	HS	0	0.001	-	0.04
F23	M27	HS	0	0	-	0.041
F24	F8	HS	0	0.001	-	0.044
F28	F12	HS	0	0.001	-	0.044
F20	F19	HS	0	0.013	-	0.046
F32	M2	HS	0	0	-	0.047
M16	F24	HS	0	0	-	0.047
F4C	F7	HS	0	0	-	0.048
F12	F28	HS	0	0	-	0.049
F6	F21	HS	0	0	-	0.049
F8	F24	HS	0	0	-	0.049
F21	F6	HS	0	0	-	0.05

Figure S4.1.1: Quality verification of the dataset. For each individual, we randomly select an increasing number of locations in the total set of locations sampled for it. Each sampling was performed 15 times and plotted. We estimated the number of locations sufficient for a reliable estimation of the home-range size when a plateau was reached (Martin et al. 2013). In our data, it was the case for only one individual (M2).



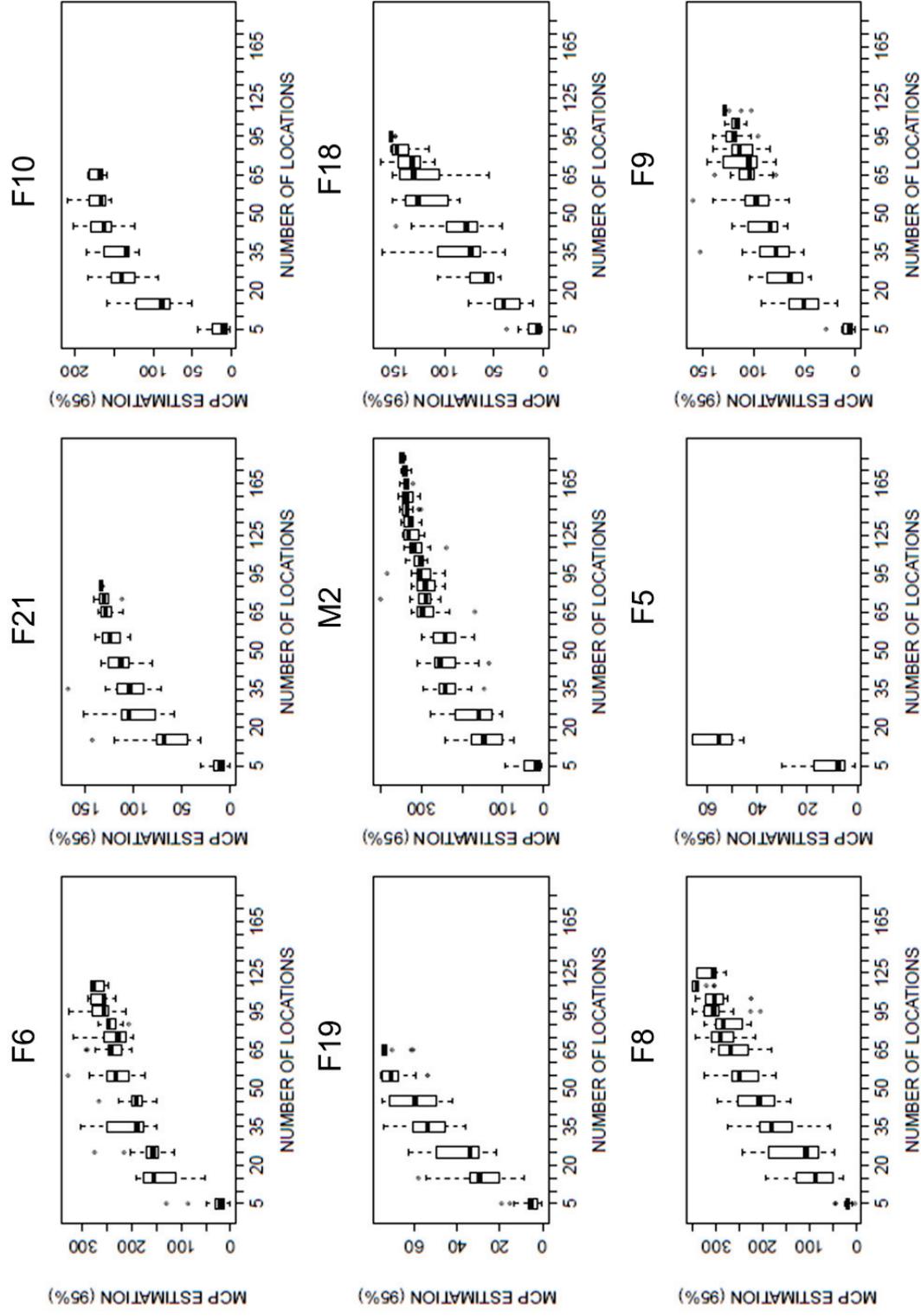


Figure S4.1.2: For the eight individuals taken into account in the statistical analysis, we plotted the number of locations sampled per month in order to evaluate the homogeneity of the sampling between individuals. Locations were homogeneously sampled between individuals followed the same year. The year being taken into account in our statistical model, our data allow us to compare and draw conclusions about the parameters affecting the home-range size. We also represented the only male for which the home-range size could be estimated (M2) over one year.

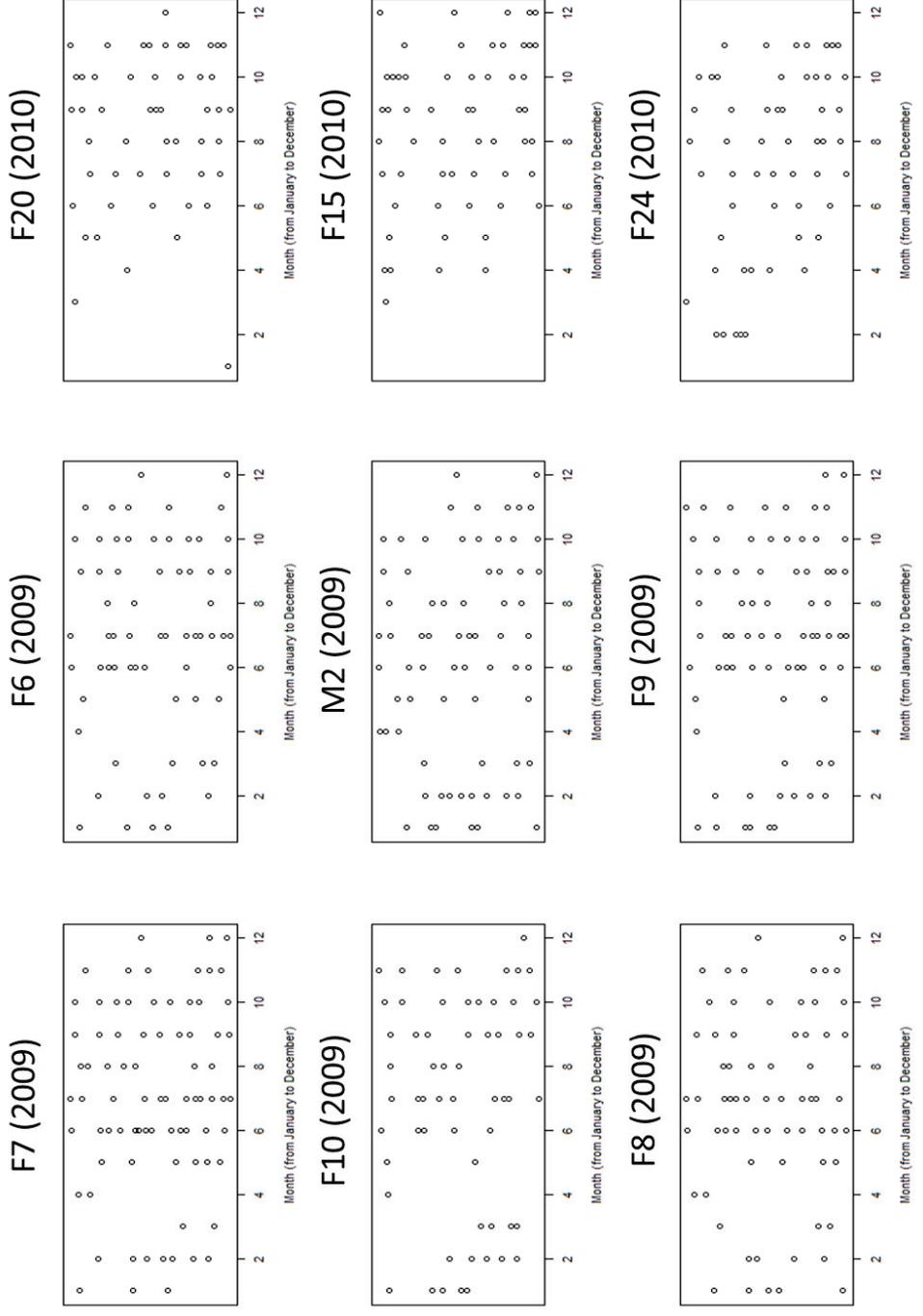


Figure S4.1.3: Results obtained in STRUCTURE HARVESTER (Earl and VonHoldt 2011) using Evanno's (Evanno et al. 2005) method for the determination of the optimal number of clusters K.

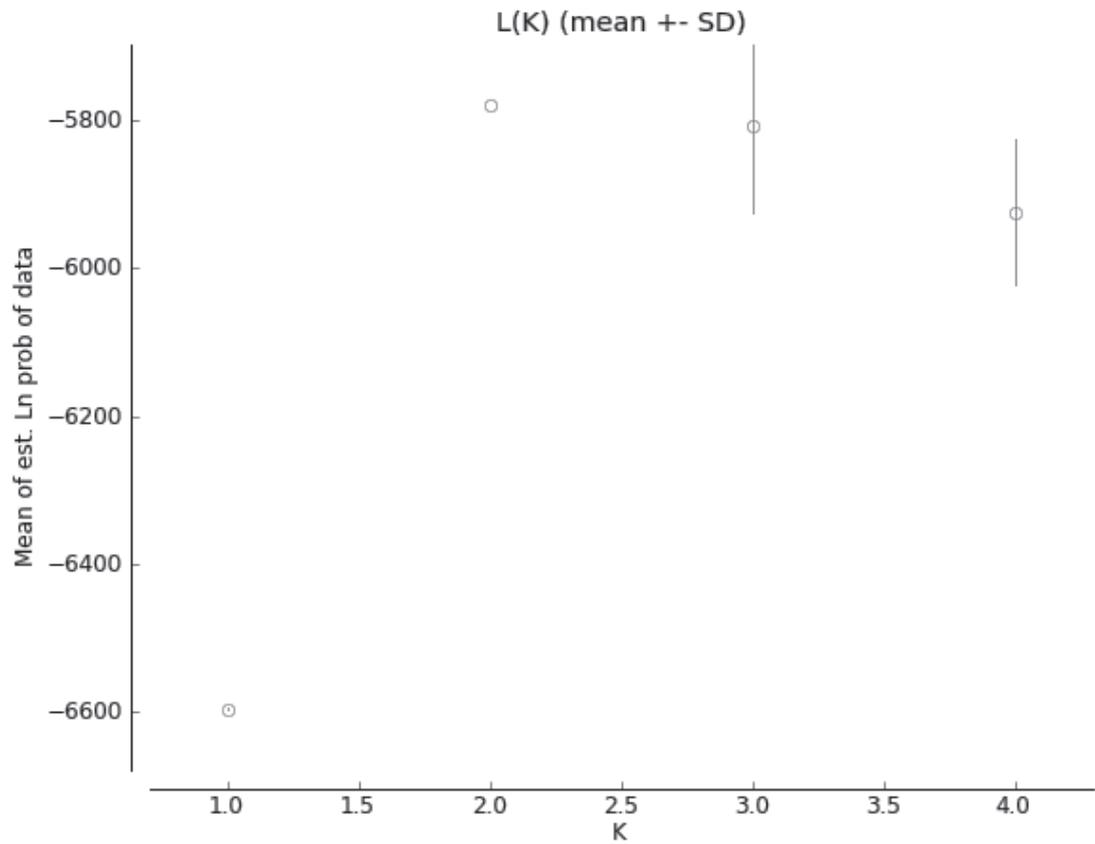


Figure S4.1.4: STRUCTURE output. The posterior probabilities of the *Felis s. catus* are represented in grey while the posterior probabilities of the cluster *Felis s. silvestris* are in black. The dotted white lines represent the threshold of 0.91. Any individual between the two lines is considered to be hybrid.



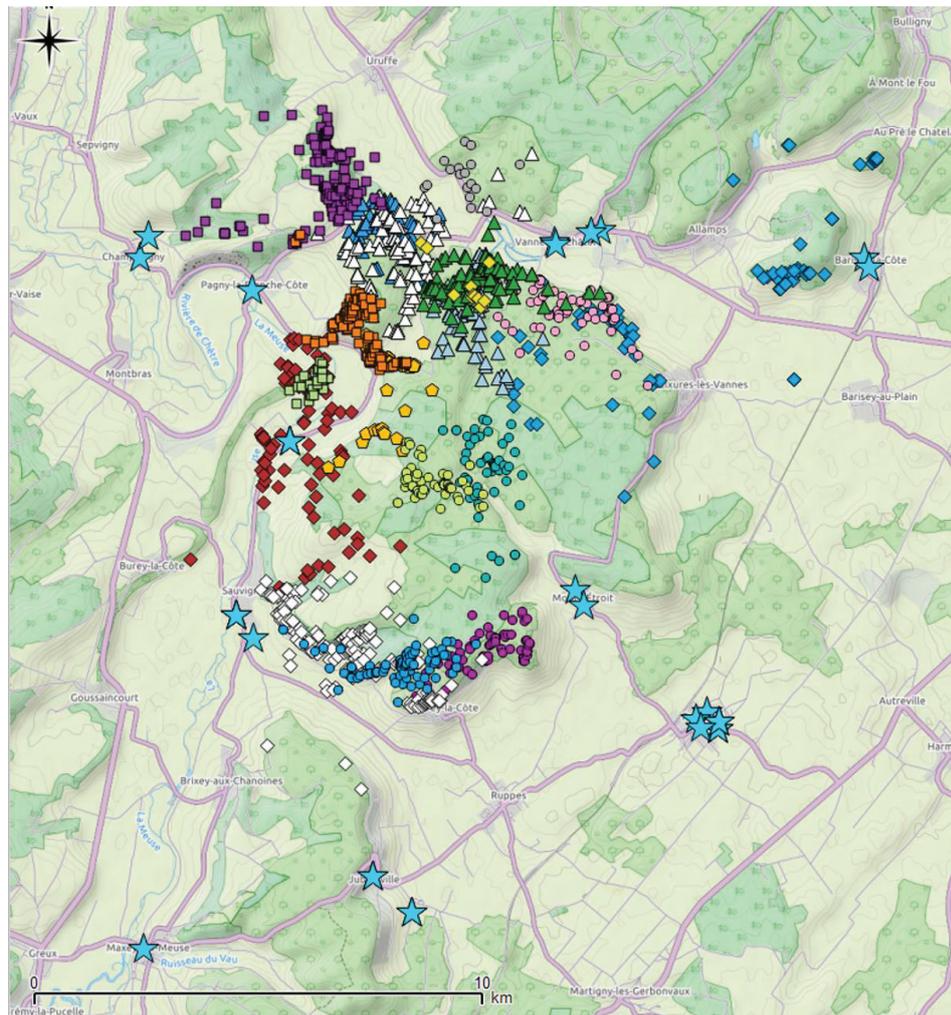


Figure S4.1.5: Radio-tracking locations obtained for each individual. Domestic cat trapping locations are represented by blue stars. Each wildcat monitored by radio-tracking is represented by a different combination of color and shape (white diamond: M2; blue circles: F18; violet circles: F15; red diamonds: M4; light green circles: F19; turquoise circles: F20; green squares: M11; yellow pentagons: F24; light blue triangles: F21; blue diamonds: M17; orange squares: F3; green triangles: F9; pink circles: F10; white triangle : F6 ; yellow diamonds: F12; blue triangles: F7; grey circles: F5; violet squares: F8). Data are from QGIS software (open layer plugin).

Table S4.1.3: Wildcat individuals used for each analysis performed in the study. “yes” means that the individual was taken into account for the analysis, “no” that the individual was discarded for this analysis.

NAME	RADIO-TRACKING	STRUCTURE ANALYSIS	AGE/WEIGHT CORRELATION	ML-RELATE KINSHIP ANALYSIS	KINSHIP-DISTANCE ANALYSIS	HOME-RANGE SIZE ANALYSIS
M2	yes	yes	yes	yes	yes	yes
M4	yes	yes	yes	yes	yes	no
M11	yes	yes	yes	yes	yes	no
M14	no	yes	yes	yes	yes	no
M16	no	yes	yes	yes	yes	no
M17	yes	yes	yes	yes	yes	no
M22	no	yes	yes	yes	yes	no
M26	no	yes	yes	yes	yes	no
M27	no	yes	no	yes	no	no
M29	no	yes	no	yes	no	no
M30	no	yes	no	yes	no	no
M33	no	yes	no	yes	no	no
M1C	no	yes	yes	yes	yes	no
M2C	no	yes	yes	yes	yes	no
M3C	no	yes	yes	yes	yes	no
M6C	no	yes	yes	yes	yes	no
M7C	no	yes	yes	yes	yes	no
M8C	no	yes	yes	yes	yes	no
M10C	no	yes	no	yes	yes	no
F3	yes	yes	yes	yes	yes	no
F5	yes	yes	yes	yes	yes	no
F6	yes	yes	yes	yes	yes	yes
F7	yes	yes	yes	yes	yes	yes
F8	yes	yes	yes	yes	yes	yes
F9	yes	yes	yes	yes	yes	yes
F10	yes	yes	yes	yes	yes	yes
F12	yes	yes	yes	yes	yes	no
F13	no	yes	yes	yes	yes	no
F15	yes	yes	yes	yes	yes	yes
F18	yes	yes	yes	yes	yes	no
F19	yes	yes	yes	yes	yes	no
F20	yes	yes	yes	yes	yes	yes
F21	yes	yes	yes	yes	yes	no
F23	no	yes	yes	yes	yes	no
F24	yes	yes	yes	yes	yes	yes
F25	no	yes	yes	yes	yes	no
F28	no	yes	no	yes	no	no
F31	no	yes	no	yes	no	no
F33	no	yes	no	yes	no	no
F4C	no	yes	yes	yes	yes	no
F5C	no	yes	yes	yes	yes	no
F9C	no	yes	no	no	no	no

Table S4.1.4: Criteria used to keep or discard individuals in the different analyses performed in the study.

ANALYSIS	CRITERION
RADIO-TRACKING	The first 18 individuals captured.
STRUCTURE ANALYSIS	Be amplified at more than 50% of the loci.
AGE/WEIGHT CORRELATION	All individuals identified as "pure" wildcats.
ML-RELATE KINSHIP ANALYSIS	All individuals identified as "pure" wildcats.
KINSHIP-DISTANCE ANALYSIS	All individuals identified as "pure" wildcats and for which we have GPS coordinates.
HOME-RANGE SIZE ANALYSIS	Individuals monitored by radio-tracking, identified as "pure" wildcat, for which locations are available over a complete year and that do not show any sign of dispersion.

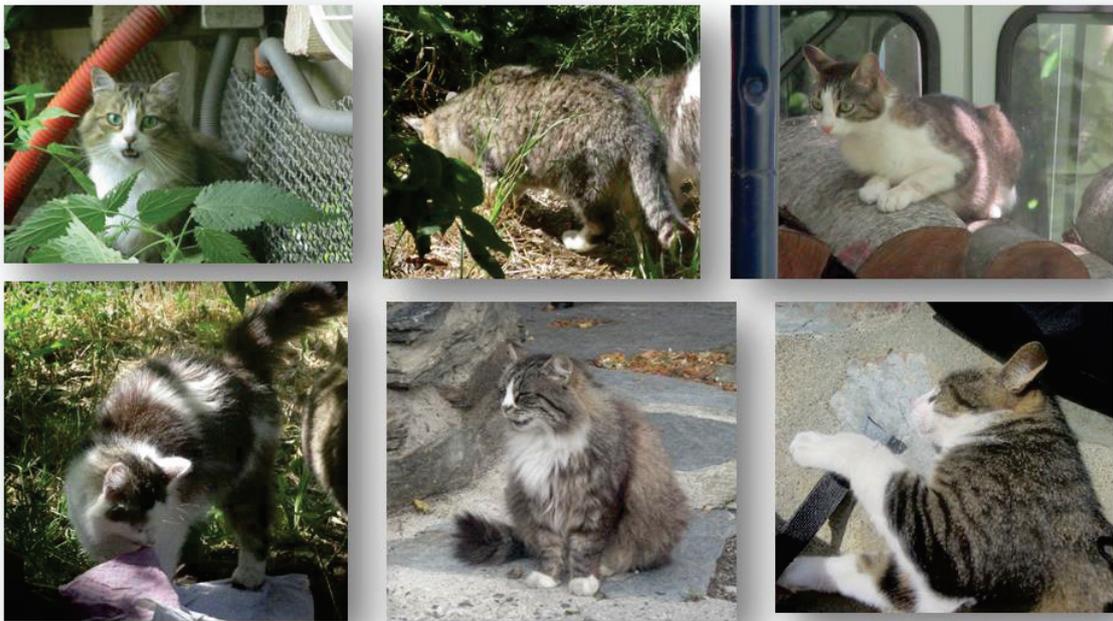
Interlude

If we consider the results obtained in chapter 3 (Part 1) and chapter 4 so far, the rate of hybridization in the local population of European wildcats remains moderate (between 0 and 14%) compared to the rates observed in Europe.

Interestingly, the low but significant gene flow from wildcats to domestic cats detected in Chapter 3 is consistent with an asymmetrical hybridization pattern probably related to the spatial organization of wildcats (between wildcat males and domestic females). The different results we obtained combining different approaches are thus consistent and seem to support the existence of a low gene flow between cat sub-species in Northeastern France.

In the second part of this chapter, we studied a second local population of European wildcats from southern France in the Pyrénées Orientales. We compared the features observed in this second population (parentage, hybridization) to the observations from northeastern France in order to understand whether the environmental conditions, and notably different levels of fragmentation, lead to different hybridization and population structure patterns.

4.2. Hybridization pattern of a Pyrenean population of European wildcats



Domestic cats from Nohèdes – personal photographs

ABSTRACT

The European wildcat *Felis silvestris silvestris* is threatened across the totality of its area of distribution by hybridization, in varying degrees, with the domestic cat *F.s. catus*. The underlying ecological processes promoting hybridization, remain largely unknown. In France, wildcats are mainly present in the North-East but signs of their presence in the French Pyrenees have been recently provided. However, no studies have been carried out in the French Pyrenees to assess the genetic status of wildcats and the extent of their hybridization with domestic cats. We have compared a local population of wildcats living in a continuous forested habitat in the French Pyrenees and a local population of wildcats living in a fragmented forested habitat in Northeastern France to evaluate how habitat fragmentation and the presence of agricultural fields influence the population structure of European wildcats. Seventy-one wildcat-like scats and 27 hair samples of domestic cats were collected in the Pyrenees and genotyped using 31 microsatellite markers to assess hybridization and relatedness patterns. We show that close kin were not found in the same geographic location contrary to what was observed for females in the Northeastern wildcat population. Furthermore, there was no evidence of hybridization in the Pyrenean wildcats and only one domestic cat raised suspicions in spite of a very close proximity between the two subspecies. The two wildcat populations were significantly differentiated ($F_{st} = 0.08$). In addition, the genetic diversity of the Pyrenean wildcats was lower than that of other wildcat populations in France and in Europe. Taken together, these results suggest that habitat fragmentation, and in particular the absence of agricultural fields, may play an important role in lowering the probability of hybridization by reducing the likelihood of contact with domestic cats. Moreover, our results suggest that the French Pyrenean wildcat populations is isolated and may be threatened by a lack of genetic diversity.

Keywords: *Felis silvestris silvestris*, *Felis silvestris catus*, relatedness, microsatellites, non-invasive sampling

BEUGIN Marie-Pauline, SALVADOR Olivier, QUENEY Guillaume, NATOLI Eugenia, PONTIER Dominique (In prep).

4.2.1. INTRODUCTION

Hybridization is especially common between subspecies, due to incomplete reproductive isolation and therefore a higher likelihood of successful interbreeding (Rhymer and Simberloff 1996; Levin 2002; Randi 2008). This is the case for the European wildcat *Felis silvestris silvestris*, a medium-sized carnivore widely spread in Europe (O'Brien et al. 2008), which is highly threatened over its entire distribution area by its closely related domestic counterparts *F. s. catus* (Yamaguchi et al. 2015). Interbreeding between wildcats and domestic cats may lead to introgressive hybridization, followed by disruption of local genetic adaptations, and then to a loss of the European wildcat genetic integrity, and even to the extinction of the sub-species (Allendorf et al. 2001).

Studies across the area of distribution of the European wildcat have shown that there is a high degree of variability in the extent of admixture with domestic cats. High levels (up to 45%) of hybridization have been reported in Hungary and Scotland (Beaumont et al. 2001; Daniels et al. 2001; Pierpaoli et al. 2003; Lecis et al. 2006; Randi 2008), while low levels (between 0 and 2%) of interbreeding with domestic cats have been shown in Germany, Italy, and Portugal (Randi et al. 2001; Pierpaoli et al. 2003; Lecis et al. 2006; Oliveira et al. 2008). The direction of the gene flow also varied, some studies reporting a gene flow from domestic cats to wildcats (Oliveira et al. 2007, Nussberger et al. 2014) while others showed the opposite with a detected flow from wildcats to domestic cats (Hertwig et al. 2009, Beugin et al. *in preparation*). Such high degree of heterogeneity in hybridization modalities and subsequent introgression may result from differences in methodological approaches (Beugin et al. *in preparation*) but it probably reflects different environmental conditions (e.g., habitat fragmentation, urban pressure). Characterizing the patterns and processes of hybridization in nature is crucial to the introduction of measures designed to prevent hybridization and then plan efficient conservation guidelines for European wildcats.

The different underlying processes leading to hybridization have been underinvestigated in the past, since this requires to focusing on the interacting populations of wildcats and domestic cats at a local scale. To our knowledge, only one study combining genetics and radio-tracking of wildcats (Beugin et al. 2016) has been conducted at a local scale in an area of ancient sympatry in Northeastern France. This study was conducted in a fragmented landscape, where forests, crops and villages alternate. A spatial sexual segregation was observed, with females living mostly inside the forest with an access to crops, and males remaining at the edge of the forest. The localization of male wildcats was proposed to be a factor promoting hybridization between them and female domestic cats. Furthermore, the home-ranges of related females were spatially close - even overlapping - suggesting that prey might be abundant enough in this habitat to result in a greater tolerance for overlap between females and their relatives (Beugin et al. 2016). However, this single study may not have accurately captured all the diversity in the hybridization processes and more local studies are urgently needed to deepen our understanding of the factors promoting hybridization in the European wildcat.

In this paper, we have conducted a genetic study of wildcats at a local scale within a protected area in the French part of the Pyrenees. The French Pyrenean wildcat population is suspected to be relatively isolated within the species' distribution range in France and northern Europe (Say et al. 2012), and has never been genetically characterized, particularly regarding hybridization. In this area the forest landscape is highly continuous, contrary to the fragmented forests of Northeastern France (CEMAGREF 2010). Here, we have endeavored to assess the genetic status of this wildcat population and the impact of spatial proximity of wildcats to human influences on their genetic admixture by comparing results of relatedness and hybridization pattern with the wildcat population of northeastern France studied in Beugin et al. (2016). Given the continuous forest habitat with few interfaces between forests and villages in this area, we may expect hybridization to be rare or inexistent in Pyrenean European wildcats. We did not expect female

natal philopatry when food resources are less abundant in the study area due to the absence of agricultural fields.

4.2.2. MATERIAL AND METHODS

a. Study area and non-invasive sampling

Fresh feces of wildcats were collected opportunistically from 2010 to 2016 in the nature reserves of Jujols and Nohèdes in the eastern part of the French Pyrenees (Figure 4.2.1). The Nohèdes Nature Reserve presents elevations ranging from 760-2,459m while the elevation of the Jujols Nature Reserve ranges between 1,100 and 2,172m. The study area covers a total surface of 325 km² of continuous forest (oak, maple, ash, pines, beech). Experienced field agents from the nature reserves have collected evidence for additional occurrences of European wildcats since 1993 based on camera-trapping surveys, direct observations or feces, which has allowed us to associate the sampling of fresh feces to the overall presence of the European wildcat in the study area. For domestic cats, hairs were sampled in 2010 and 2017 in the villages of Nohèdes (N = 20), Conat (N=4), and Serdinya (N = 3), located on the edge of the reserves (see Figure 4.2.1). Only individuals born in the villages, sterilized or not, were included in our sampling. A total of 71 feces and 27 hair samples were thus collected. We also included in the analysis 42 European wildcats (Beugin et al. 2016) as well as 31 domestic cats originating from different villages in northeastern France.

b. DNA extraction and microsatellite genotyping

All cats were genotyped using 31 autosomal microsatellites and one marker of sex, amplified in three multiplexes. DNA extraction was performed using a purification column kit (Nucleospin 96 Tissue kit, Macherey-Nagel) following the manufacturer protocol. PCR reactions were performed step-by-step following a unidirectional workflow starting in a clear room with positive air pressure where sensitive reagents, enzymes and primers, were prepared. DNA and reagents were

then assembled in a pre-PCR room. PCR amplifications were made in 96-well microplates in a post-PCR area with negative air pressure. The PCR reaction occurred in a final volume of 10 μ l that contained 5 μ l of Mastermix Taq polymerase (Type-it, QIAGEN), 1.35 μ l of primer pairs at a final concentration between 0.08 and 0.6 μ M, and 30ng of DNA. Each pair of primers was coupled with a fluorescent dye. The reaction started with a denaturation step at 95°C for five minutes. This step was followed by thirty PCR cycles (denaturation step = 95°C, 30s; annealing step = 55.9°C, 90s; elongation step = 72°C, 30s) and a final elongation step at 60°C during 30 minutes. PCR products were resolved on a capillary sequencer ABI PRISM 3130 XL (Applied Biosystem) under denaturing conditions (formamide) and an internal size marker in one migration for each multiplex. All these steps were performed using filtered tips. Finally, the electrophoregrams were analyzed using GENEMAPPER 4.1 (Applied Biosystem/Life Technologies) twice independently. Ambiguous loci were classed as missing data. Only the individuals with an amplification success higher than 70% were included in the subsequent analyses.

c. Consensus genotypes and population genetics analyses

Consensus genotypes were built as follows. Two genotypes were considered to represent the same individual when (1) they were identical, (2) they only differed by missing data and these missing data did not represent more than ten microsatellite markers, (3) they only differed by missing data below the threshold of ten markers and a single difference that could be explained by allelic dropout. For each locus, the frequency of null alleles was assessed following Brookfield's (1996) method, and tested for significance using binomial tests according to De Mêeus et al. (2002). All loci exhibiting significant evidence of null alleles were discarded from further analyses. Deviations of loci from both Hardy-Weinberg equilibrium (HWE) and linkage equilibrium were both tested using FSTAT v 2.9.3.2. (Goudet 1995) with a Bonferroni correction and a 5% risk for all populations. Loci showing a departure from HWE were also discarded from the analysis. The software FSTAT v2.9.3.2. was also used for estimating Weir and Cockerham's F_{ST}

between wildcat and domestic cat populations from the Pyrenees and from Northeastern France as well as allelic richness. Expected (H_E) and observed (H_o) heterozygosities were calculated using GenALEx 6.501 (Peakall and Smouse 2012). Finally, a discriminant analysis of principal components (DAPC, Jombart et al. 2010) was used in order to visualize the differentiation between domestic and European wildcat populations from northeastern France and the Pyrenees.

d. *Spatial structure and relatedness*

The 52 fresh fecal samples for which the sampling location was recorded were displayed on a map using QGIS v2.8.1., together with the other indices of presence of European wildcats (feces and photo trapping). The program ML-Relate (Kalinowski et al. 2006) was used to calculate pairwise relatedness between all wildcat individuals. Using a linear model, we tested whether sex or relatedness were a significant predictor of the pairwise geographical distance between individuals. Geographical distances between individuals were calculated with QGIS v2.8.1. We considered the mean pairwise distance between samplings for individuals sampled several times. Statistical analysis was performed in R 2.14.2 (R core development team 2016).

e. *Admixture analysis*

The Bayesian clustering method implemented in STRUCTURE v2.3.4 (Pritchard et al. 2000, Falush et al. 2003) was used to identify wildcats, domestic cats and possible hybrids by applying the *admixture model* with correlated allele frequencies. The optimal number of clusters K was determined using the method described by Evanno et al. (2005). STRUCTURE was used to assess membership proportions (q_i) to the inferred K clusters, which correspond to the proportion of each individual's multilocus genotype belonging to each of the inferred K clusters. The threshold level to differentiate wildcats and domestic cats from hybrids was determined by selecting individuals which were believed to be representative of the parental populations using the iterative algorithm described in Beugin et al. (in

preparation). This algorithm consists of the repetition of a simulation-selection process allowing for the building of a representative pool of parents (representative regarding the differentiation between parental population and their genetic diversity). At each step, individuals from different hybrid classes (parental classes, F1, F2, first generation backcrosses) were simulated using the function *hybridize* from the package *adegenet* (Jombart 2008). At initialization, these simulations were carried out based on the 5% top-ranked (based on q -values) individuals from each cluster. These simulated individuals were then analyzed with STRUCTURE using no prior information in order to get individual q -values. All the individuals characterized by a lower bound of their 90% credibility interval higher than the 1% quantile of the distribution of q -values of the parental simulated individuals were then integrated into the representative parental pool and used to carry on the simulations at the next iteration. This simulation-selection process was then repeated until the number of individuals integrated in the representative parental pool stabilized. This way, all individuals that can be considered as a parental individual are used to define the threshold that allows hybrids to be distinguished from parents.

In order to carry out such categorization, we simulated 200 individuals (domestic cats, wildcats, F1, F2, F1 x domestic and F1 x wild) using the function *hybridize* from the R package *adegenet* (Jombart 2008) based on the pool of representative individuals built and ran STRUCTURE ten times in order to determine four thresholds corresponding to the lowest q -values reached by parental individuals on one hand (thresholds TP1 and TP2) and the highest q -values reached by hybrid individuals in the second hand (thresholds TH1 and TH2). This defined the borders of the parental zone, the hybrid zone and the grey zone. All STRUCTURE analyses (iterative algorithm and threshold determination) were run for a burn-in period of 100,000 and MCMC length of 100,000 iterations according to Gilbert et al. (2012) and following graphical verifications of the convergence of the algorithm based on the parameter alpha. We ran the function *snapclust* (Beugin et al. submitted) from the *adegenet* package in order to

determine the category of detected hybrids. Given the number of microsatellites and the expected order of magnitude of the differentiation between domestic cats and wildcats (F_{st} = 0.11 to 0.20 – Beaumont et al. 2001; Oliveira et al. 2008; Hertwig et al. 2009; Mattucci et al. 2013; Beugin et al. 2016) we considered only F1 hybrids and first-generation backcrosses (F1 x Domestic cat and F1 x Wildcat) in the *snappclust* analysis. Additionally, the direction of the gene flow between wild and domestic cats was assessed by estimating the rate of migration per generation with the computer program BAYESASS 3.0.3. (Wilson and Rannala 2003) with a MCMC chain of 5,000,000 after a burn-in period of 1,000,000 with a sampling interval of 2,000. All other parameters were left to default.

4.2.3. RESULTS

a. Genotyping and markers' selection

Feces and/or direct observations of European wildcats have been reported all over the nature reserves of Jujols and Nohèdes up to an elevation of 2430m. Samples genotyped were collected over the entire area where signs of the presence of the European wildcat had been reported (see Figure 4.2.1). Forty-five out of the 71 fresh feces collected, and 22 out of the 27 hair samples collected were successfully amplified at more than 70% of the loci. On average, 74.1% (sd = 0.31) of the loci were successfully amplified using feces while 81.5% (sd = 0.32) of the loci were successfully amplified on average from hairs.

We did not detect any significant null allele frequencies, nor did we detect significant deviations from HWE or any linkage disequilibrium. Subsequent analyses were thus performed on the 31 microsatellites. We identified 39 unique genotypes including 21 domestic cats (15 females and 6 males) and 18 European wildcats (10 females and 8 males). Six (3 females and 3 males) out of the 18 European wildcats were sampled several times (from two to 6 times). Feces from the same individual were found at a maximal distance of 5.3 km for a male, and 3 km for a female. Different feces from the same individual were found within 540 m

of elevation for one male, whereas the other individuals were sampled within 100-200 m of elevation. These recaptures did not allow us to establish home-ranges due to the lack of locations per individual. We detected wildcats genetically confirmed as such in elevations up to 2250 m. Contrary to what was observed in northeastern France (Beugin et al. 2016), no spatial sex segregation was observed (Figure 4.2.2).

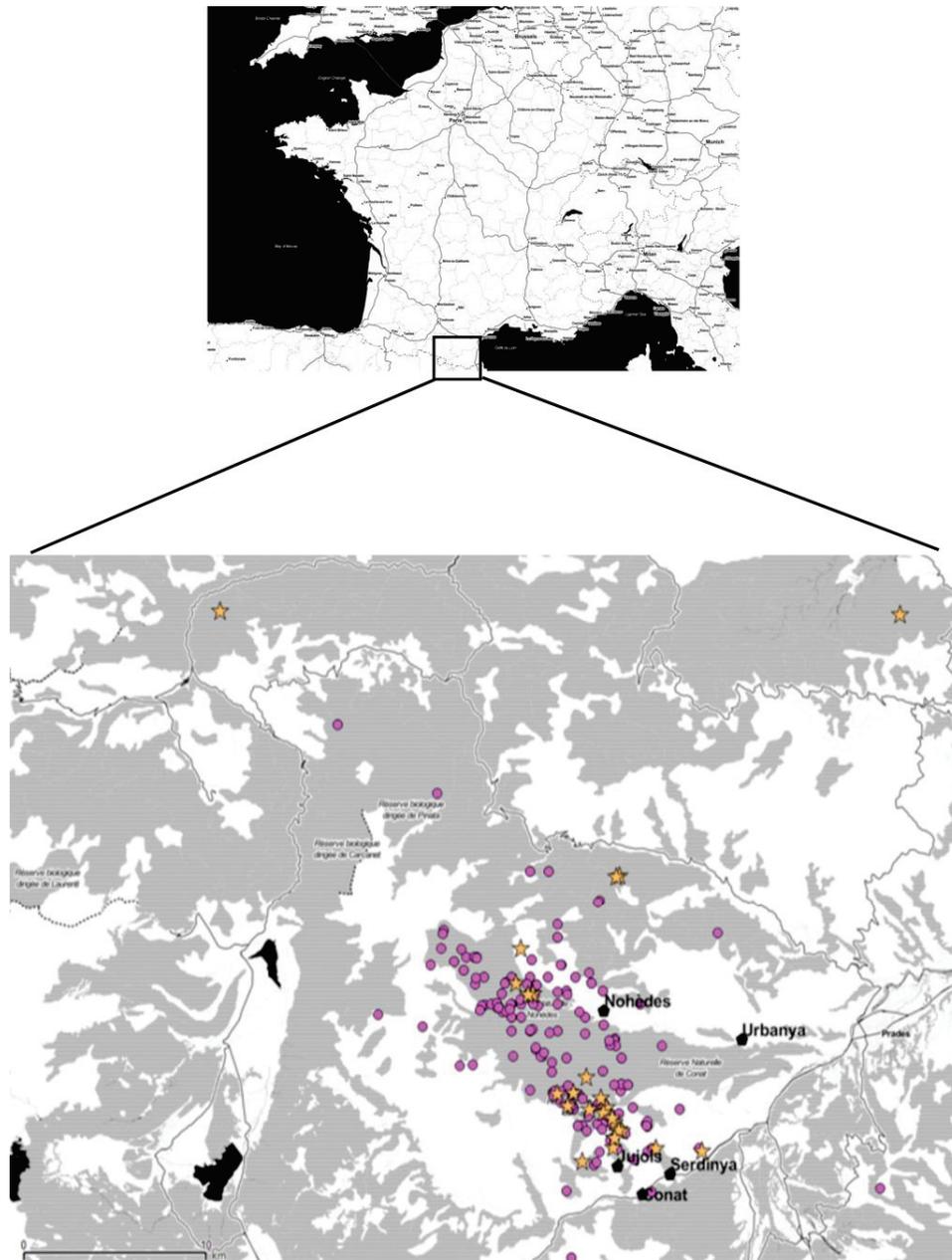


Figure 4.2.1: Localization of the study area. The pink circles represent all the locations where feces, camera-trapping or direct observations attested for the presence of the European wildcat. The yellow stars correspond to the locations of the fresh feces. The villages where domestic cats were sampled (except Urbanya) are indicated in black and are located at the edge of the natural reserves of Jujols and Nohèdes.

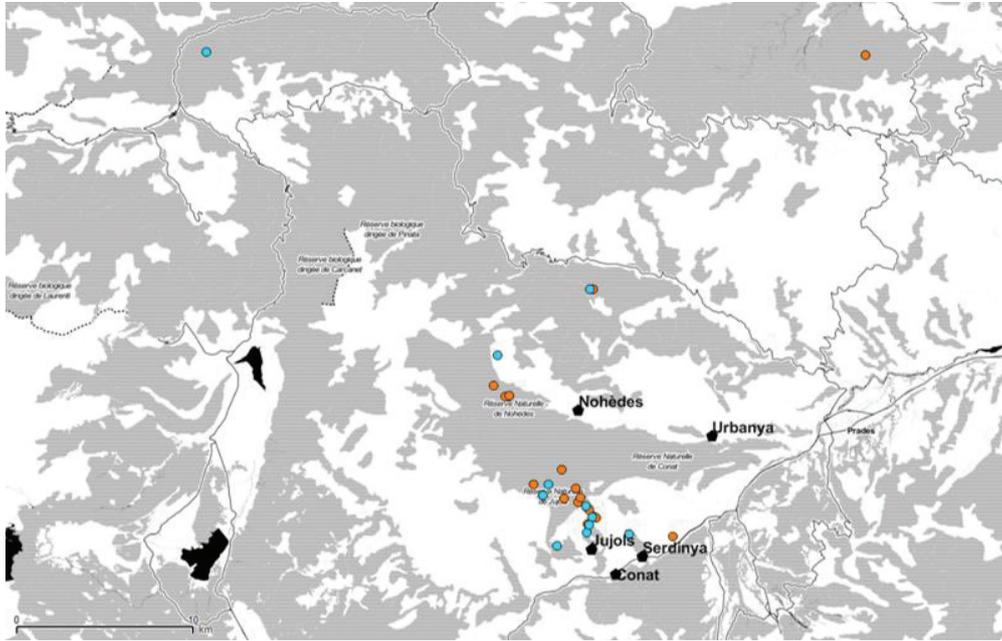


Figure 4.2.2: locations of the samples according to the sex of individuals. Females are indicated in orange and males in blue

b. Genetic diversity and kinship pattern

The domestic cats showed a higher allelic richness (mean = 6.01, sd = 1.28, Appendix, Table S4.2.1) than European wildcats (mean = 4.55, sd = 1.39) in the Pyrenees. Northeastern European wildcats also had an allelic richness higher than their Pyrenean conspecifics (mean = 5.38, sd = 1.37). Average values of heterozygosity were slightly higher in the Pyrenean domestic cats ($H_o = 0.696$; $H_E = 0.712$) than in European wildcats ($H_o = 0.659$; $H_E = 0.619$). The Pyrenean populations of European wildcats and domestic cats were significantly differentiated with a F_{ST} value of 0.18. A significant differentiation between Pyrenean wildcats and Northeastern domestic cats ($F_{ST} = 0.19$), and wildcats ($F_{ST} = 0.0835$), was also found, while Northeastern domestic cats were significantly differentiated from Pyrenean domestic cats ($F_{ST} = 0.0414$). The DAPC scatter-plot (Figure 4.2.3) confirmed the sharp distinction between the two European wildcat populations.

In both sexes, related wildcat individuals were not sampled significantly closer together compared to unrelated individuals in the Pyrenees according to the linear mixed model (Appendix, Figure S4.2.1). On the contrary, in Northeastern France, related females were captured significantly closer together than unrelated females ($F = 6.88$, $df = 1$, $p = 0.0095$; Appendix, Figure S4.2.1).

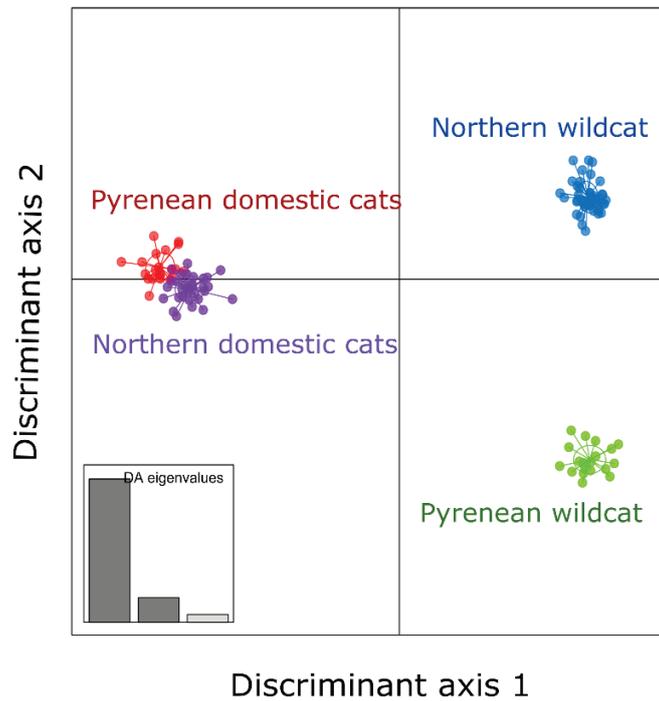


Figure 4.2.3: Scatter plot of the discriminant analysis of principal components (DAPC). The domestic cats are represented in red (Pyrenean population) and purple (Northeastern population), and the European wildcats in green (Pyrenean population) and blue (northeastern population). The first PC describes 81% of the genetic diversity and the second axis 14%.

c. Admixture analysis

Evanno's method showed that $K=2$ best described our data, one cluster corresponding to the domestic cats, the other corresponding to the European wildcats (Figure 4.2.4). The iterative algorithm allowed us to define hybrids as individuals presenting a mean probability of assignment (conservative approach) or a lower bound of credibility interval (relaxed approach) below 0.79 for domestic cats, and below 0.83 for European wildcats.

Using the conservative approach, we did not detect any hybrid and *snappclust* confirmed the absence of hybrids. Using the relaxed approach, one of the domestic cats sampled in Nohèdes was detected as being hybrid and this individual detected as being hybrid by the relaxed approach was substantially assigned to the first-generation backcross category by *snappclust* (Appendix, Figure S4.2.2). The absence of gene flow from domestic cats to wildcats ($m = 0.0148$ – CI95: 0-0.042) as well as from wildcats to domestic cats ($m = 0.0160$ – CI95: 0-0.046) was consistent with the absence of hybridization between the two subspecies.

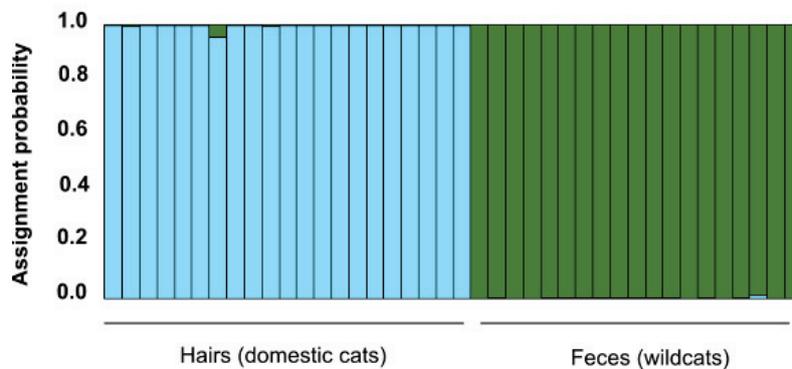


Figure 4.2.4: Results from the STRUCTURE analysis. Each cat genotype is represented by a vertical bar split into $K=2$ colored sections, according to its relative assignment to the genetic cluster: domestic cats in blue, European wildcats in green. The proportion of the bar in a given color represents the assignment probability of the individual for the corresponding cluster.

4.2.4. DISCUSSION

Our study has confirmed the presence of wildcats in the French Pyrenees within a large area of the nature reserves of Nohèdes and Jujols up to 2,430 m. Moreover, our study has provided the first genetic characterization of a local population of the French Pyrenean wildcats, despite their presence being acknowledged since 1993 (Observations collected by trained personnel of the nature reserve, Say et al. 2012).

The primary interesting result is that males as well as females in close proximity are not kin related suggesting that both males and females disperse in this continuous forest landscape, i.e., related females did not tend to remain in the same area contrary to the wildcat population of northeastern France (Beugin et al. 2016). The dispersal pattern may directly reflect the level of food resource availability. In fragmented environments as observed in Northeastern France, with forest alternating with field crops, large areas rich in resources are available for wildcats (Lozano et al. 2003, Silva et al. 2013). With carnivores, food distribution has been suggested to be the major determinant of species spatial distribution (MacDonald 1983). The importance of resource distribution on the spacing pattern of wildcat females has already been proposed (Stahl et al. 1988, Sarmiento et al. 2009) and was supported by the study in northeastern France (Beugin et al. 2016). Thus, although the European wildcat is acknowledged to live solitarily (e.g., Corbett 1979, Biró et al., 2004), its dispersal pattern may show more variability than has been described up to now.

The second important result is that no hybrid was categorically found neither in the Pyrenean wildcats nor in the Pyrenean domestic cat population. The absence of hybrids in the Pyrenean population contrasts with the situation in Northeastern France, where six wildcats out of 42 showed signs of hybridization (Beugin et al. 2016, Beugin et al. in preparation). The absence of hybrids in the continuous Pyrenean forest landscape may imply that the absence of crops frequented by both wild and domestic cats for hunting purposes may impede encounters between the two subspecies. Actually, domestic cats, even feral domestic cats, do not enter the

forest environment (during the last seven years no pictures of domestic cats have been taken by camera trapping). Unfortunately, the sample size of this study was limited and more extensive studies will be required to confirm the absence of hybridization in this environmental context.

Finally, the Pyrenean wildcats showed values of genetic diversity (4.55) lower than other wildcat or domestic cat populations in France and in Europe (4.84 alleles per locus on average in the Pyrenean wildcat population while between 3 and 11.8 can be found in the literature with rare populations below 6; Oliveira et al. 2007, O'Brien et al. 2009, Eckert et al. 2010, Say et al. 2012, Mattucci et al. 2013), suggesting that wildcats in the Pyrenees may be threatened by a lack of genetic diversity. Not surprisingly, the genetic differentiation between Pyreneans domestic cats and their Northeastern counterparts was moderate ($F_{ST}=0.04$). In contrast, the Pyrenean wildcat population was significantly differentiated from the northeastern wildcats with a higher F_{ST} value (0.08). This genetic divergence could result from a classic genetic process of isolation by distance (IBD, Wright 1943), which supposes continuity in the distribution of the European wildcat on French territory; such a cline has been described in the northeastern wildcat population (Say et al. 2012, Würstlin et al. 2016). Alternatively, some gap in the distribution of wildcats may exist in France as suggested by O'Brien et al. (2008) and Say et al. (2012). The lower genetic diversity of the Pyrenean wildcat population is supportive of the isolation of this population and thus of the existence of a fragmented pattern of distribution of the European wildcat across France.

4.2.5. CONCLUSION

Results in this study have added novel information to the European wildcat population structure in France. They provided further information about the relationship between environmental conditions and hybridization risks in French wildcat populations. Conservation strategies of wildcats should take into account the local habitat such as the existence of a fragmented or continuous forest

environment, and the presence of agricultural fields. Further investigation should also focus on the spatial distribution of French wildcats - in particular we need to confirm whether the French Pyrenean population is isolated from the main area of wildcats in France (Say et al. 2012) but connected to the Spanish Pyrenean wildcat population. Depending upon the answer, the usefulness of wildlife corridors to enhance connectivity between the different wildcat populations should be addressed to ensure the long-term viability of the French Pyrenean wildcat population.

4.2.6. ACKNOWLEDGEMENTS

We thank everybody from the reserves and villages of Nohèdes, Conat, Jujols and Serdinya who collaborated in providing wildcat and domestic cat samples and sharing their observations. We also thank Robin Buckland for his careful check of English language.

4.2.7. APPENDIX

Table S4.2.1: Indices of diversity for the domestic cats and the European wildcats sampled in the Pyrenees per locus. The observed and expected heterozygosities are reported as *Ho* and *He*, respectively. *Na* stands for the number of alleles per locus, *N* corresponds to the number of individuals successfully amplified at each locus. *Null* reports the null allele frequency estimated using the Brookfield estimator (Brookfield 1996).

Locus	Domestic cats					European wildcat				
	Ho	He	Na	N	Null	Ho	He	Na	N	Null
F37	0.917	0.840	8	12	-0.042	0.250	0.403	3	11	0.109
Fca8	0.846	0.805	7	13	-0.023	0.789	0.747	7	18	-0.025
Fca023	0.538	0.754	6	13	0.123	0.737	0.609	3	18	-0.079
Fca024	0.917	0.788	6	12	-0.072	0.462	0.701	5	13	0.141
Fca031	0.619	0.799	8	21	0.100	0.667	0.667	7	15	0.000
Fca43	0.571	0.585	7	21	0.009	0.895	0.798	6	18	-0.054
Fca45	0.714	0.797	8	21	0.046	0.684	0.704	5	18	0.011
Fca58	0.905	0.788	6	21	-0.065	0.368	0.320	4	18	-0.037
Fca77	0.524	0.468	5	21	-0.038	0.737	0.648	6	18	-0.054
Fca078	0.500	0.758	7	20	0.147	0.526	0.524	5	18	-0.002
Fca085	0.714	0.641	5	21	-0.045	0.316	0.266	2	18	-0.039
Fca96	0.450	0.451	6	20	0.001	1.000	0.793	7	14	-0.115
Fca124	0.857	0.798	7	21	-0.033	0.889	0.826	7	17	-0.035
Fca126	0.778	0.869	10	18	0.049	0.632	0.726	6	18	0.055

Locus	Domestic cats					European wildcat				
	Ho	He	Na	N	Null	Ho	He	N	Null	
<i>Fca547</i>	0.667	0.782	7	21	0.065	0.789	0.723	5	18	-0.039
<i>Fca577</i>	0.688	0.771	7	16	0.047	0.667	0.656	4	11	-0.006
<i>Fca668</i>	0.714	0.709	6	21	-0.003	0.833	0.691	4	17	-0.084
<i>Fca675</i>	0.905	0.829	10	21	-0.042	0.737	0.787	6	18	0.028
<i>Fca26</i>	0.667	0.739	6	21	0.042	0.737	0.684	6	18	-0.031
<i>Fca069</i>	0.810	0.799	7	21	-0.006	1.000	0.763	5	18	-0.134
<i>Fca075</i>	0.714	0.718	5	21	0.002	0.778	0.718	5	17	-0.035
<i>Fca105</i>	0.667	0.660	7	21	-0.004	0.842	0.796	7	18	-0.025
<i>Fca149</i>	0.810	0.718	5	21	-0.053	0.556	0.566	4	17	0.007
<i>Fca201</i>	0.857	0.828	8	21	-0.016	0.737	0.544	4	18	-0.125
<i>Fca220</i>	0.714	0.637	6	21	-0.047	0.750	0.732	5	15	-0.010
<i>Fca229</i>	0.619	0.621	5	21	0.001	0.474	0.511	4	18	0.025
<i>Fca293</i>	0.667	0.787	6	21	0.067	0.833	0.662	3	17	-0.103
<i>Fca310</i>	0.571	0.642	6	21	0.043	0.053	0.051	2	18	-0.001
<i>Fca441</i>	0.714	0.735	5	21	0.012	0.706	0.529	5	16	-0.115
<i>Fca453</i>	0.714	0.677	6	21	-0.022	0.538	0.500	3	12	-0.026
<i>Fca678</i>	0.714	0.719	5	21	0.003	0.833	0.687	5	17	-0.087

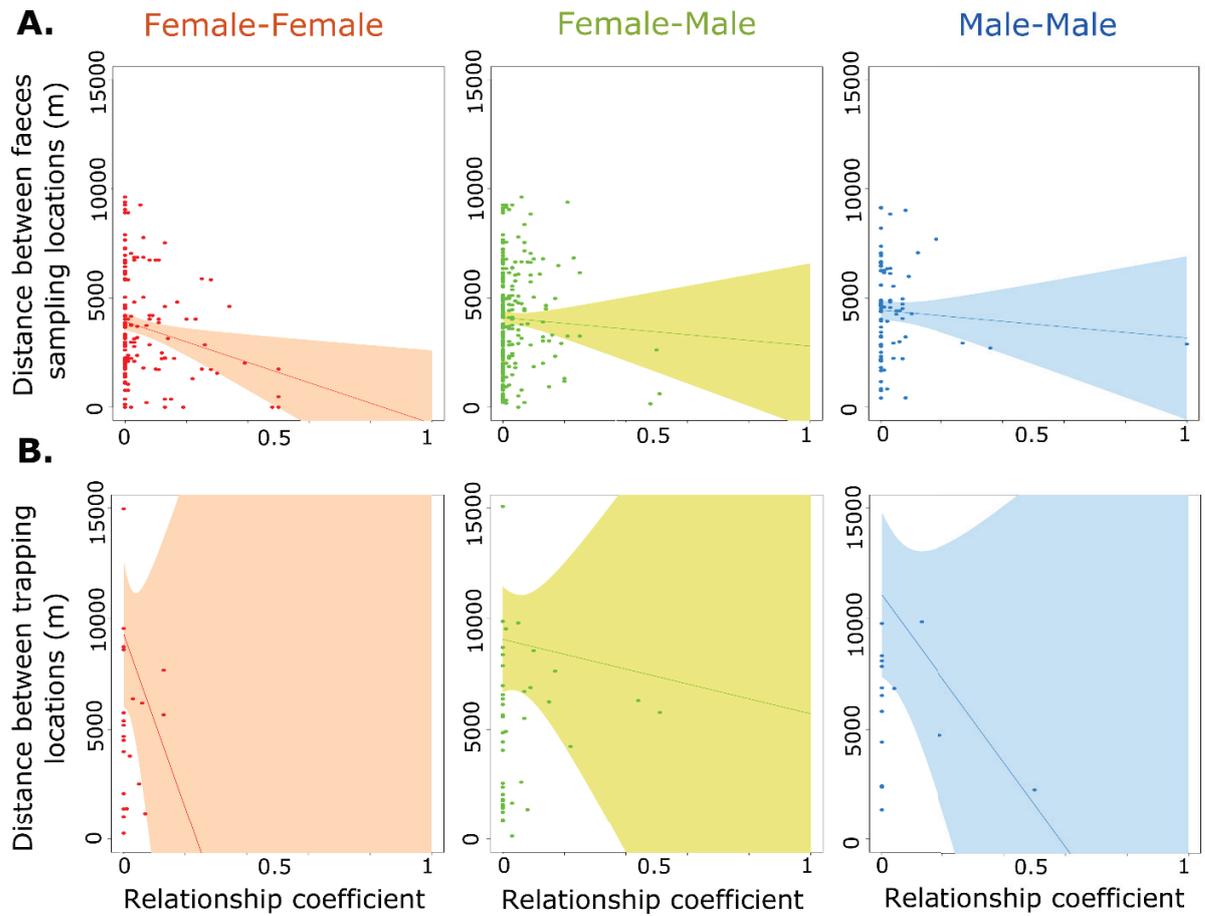


Figure S4.2.1: Linear regression and 95% confidence interval (represented using different colors) of the distance between two fresh feces samples according to the coefficient of relationship. A. in Northeastern France and B. in the French Pyrenees.

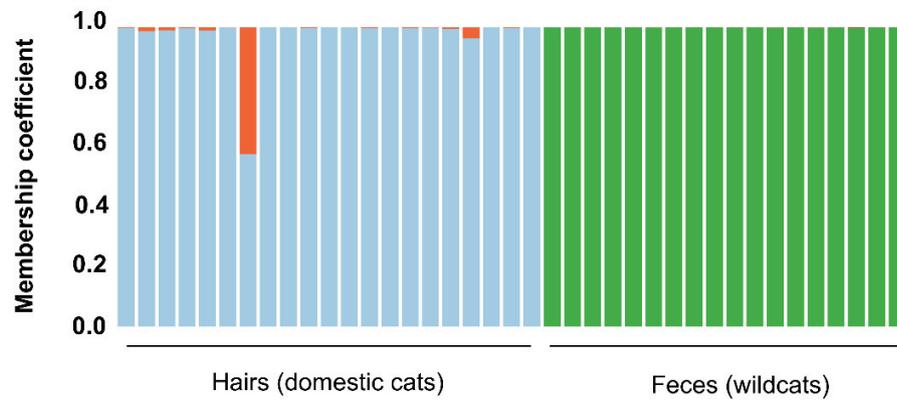


Figure S4.2.2: Results from the *snaphclust* analysis. Each vertical bar represents an individual. The colors represent the different hybrid categories included in the analysis: domestic cats in blue, wildcats in green, F1 in purple, F1xdomestic cats in orange, F1xwildcats in brown. The proportion of the bar in a given color represents the probability for the individual belonging to the corresponding category.

Addendum

Studies combining genetics and radio-tracking at a local scale led us to suggest that a high heterogeneity in European wildcats' populations 'spatial organization exists. The understanding of this variability is of major importance when a population is threatened by hybridization. We are still lacking crucial information about the Cat system to fully understand the threat possibly raised by hybridization. Notably, no study was conducted regarding any possible mate choice that could limit gene flows.

During my PhD, I worked mainly on two other systems: the brown hare *Lepus europeus* and the mountain hare *Lepus timidus* in one hand (Beugin et al. 2016, Appendix 2), and bats from the *Pteronotus parnellii* cryptic species complex on the other hand (Filippi-Codaccioni et al. *in prep*). In both cases, hybrids were detected but the ecological mechanisms at the basis of such events of hybridization differ completely. For the hares, hybridization seems driven by the global warming and the species displacements it induces. In bats, most of the hybridization is detected through the analysis of mitochondrial DNA and interspecific reproduction remains exceptionally rare even if the species live in sympatry, in the same caves. Strong mechanisms of reproductive isolation are then expected to prevent hybridization in spite of frequent opportunities of encounters.

Through these different systems, the Cat, the Hare and the Bat, we could clearly see that the presence of hybrids does not reflect the same biological processes according to the species. The ecological study of each species thus remains essential to understand and predict the evolution of hybridization.

5. Circulation of viruses between domestic cats and European wildcats

Foreword

The observation of hybridization between the domestic cat and the European wildcat reveals that individuals from the two sub-species encounter and are thus likely to exchange pathogens. In order to assess to what extent such pathogen flow exist between domestic cats and wildcats, we conducted a serological survey on the European wildcat population from north-eastern France. Results regarding the viruses that spread in domestic cats' populations have been published in Hellard et al. (2011).

In this chapter, we report the results of this serological survey which allowed us to assess the presence of the feline immunodeficiency virus (FIV), the feline calicivirus (FCV), the feline herpesvirus (FHV) and the feline parvovirus (FPV) in the domestic cats' population, the wildcats' population and their hybrids according to the categorization established in Chapter 3, Part 1. This serological survey indicated that none of the European wildcat is infected by any of these four viruses. If we can expect wildcats to be unable to sustain a virus because of their small population size and their solitary way of life, we would expect the epidemiological pressure maintained by domestic cats to be sufficient to lead to a non-null number of infected wildcats.

Several hypotheses may explain this null prevalence. First, it could be due to the presence of a strain which has not been detected by the tests used. However, laboratory procedures were established in order to detect all the strains of each virus. Second, given the small population sizes, the observation of a null prevalence may result from sampling fluctuations. The individuals sampled would then not capture all the diversity of epidemiological states present in the population. Thirdly, a behavioral barrier may prevent the transmission of these four viruses from the domestic cats to the European wildcats. Finally, the viruses observed in the domestic cat population may be specialist viruses which cannot infect the European wildcats and their absence may reflect the existence of a genetic or physiological barrier between the two sub-species.

In this chapter, we intend to assess under which values conditions of between sub-species and between wildcats transmission the null prevalence we observed is statistically expected. Then, we confront these conditions to what we know of the ecology of the wildcat to assess whether these conditions are consistent with the known parameters of the population.

Are European wildcats threatened by the common viruses of the domestic cats?



Domestic cat in the urban environment – personal photography

ABSTRACT

The European wildcat is threatened by hybridization with the domestic cat. The domestic cats may also represent an epidemiological threat for the wildcats as they carry several viruses. Serological surveys aiming at assessing whether or not wildcats were susceptible to domestic cats' viruses have been led these past years and many viruses have been detected in wildcats. However, almost none of these studies combined genetics and serology in order to ensure that the wildcats tested are true wildcats and not hybrids. Here, we propose a serological survey of a wild population from North-eastern France, where all wildcats and domestic cats have been genetically identified as pure individuals or hybrids, for four common viruses of the domestic cat: FCV, FHV, FPV and FIV. We found a null prevalence for these four viruses in the European wildcat population while all were present in domestic cats and in presumed hybrids sampled in villages. Using a statistical model, we found that the absence of infected individuals was plausible and could only be observed under particular conditions of circulation of the viruses compatible with the existence of a barrier – either behavioral or physiological – between sub-species. For all four viruses, the existence of such barrier matched indeed with the ecology of the European wildcat. Such results suggest that this particular population of wildcats is not threatened by these common viruses of domestic cats.

Keywords: cross-species transmission, dynamic-statistical model, behavioral barrier, infectious diseases

BEUGIN MP, FOUCHET D, LEBLANC G, REGIS C, POULET H, NATOLI E, QUENEY G,
PONTIER D
(In preparation)

5.1.1. INTRODUCTION

Infectious diseases represent a major concern for the conservation of wildlife (Pedersen et al. 2007). Domestic animals, and especially domestic dogs and cats, are among the primary vectors of infectious diseases for wildlife (Cleaveland et al. 2001). Their high densities allow viruses to persist within their populations when wild populations cannot maintain pathogens because of their small size. For example, domestic dogs in the Serengeti ecosystem have been identified as a reservoir for rabies (Lembo et al. 2008) and the Canine Distemper Virus (Cleaveland 2000), which has led to a severe demographic decline (one third of the population was killed) in the Lion *Panthera leo* population in 1994 (Roelke-Parker et al. 1996). Wild felids may also be endangered by viruses transmitted by free-ranging domestic cats *Felis silvestris catus* (Millán and Rodríguez 2009) including the European wildcat *F. s. silvestris*, a close relative of the domestic cat (Driscoll et al. 2007), whose solitary behavior combined with small population sizes can restrain the spread and persistence of viruses (Lozano and Malo 2012). Several of the most common viruses of the domestic cats have been identified in wildcats' populations (McOrist 1991, Daniels 1999, Leutenegger 1999, Račnick 2008, MacDonald 2015, Wasieri 2009, Millán and Rodríguez et al. 2009, Duarte et al. 2012) like the Feline Calicivirus (FCV), the Feline Herpes Virus (FHV) and the Feline Panleukopenia Virus (FPV). These viruses can be (re)introduced in wildcats through contacts with their domestic relatives, and thus domestic cats may represent a threat for the conservation of the European wildcat from an epidemiological point of view.

Domestic cats are also largely acknowledged as a threat for the conservation of the European wildcat from a genetic point of view and, more precisely, for the conservation of its genetic integrity (Yamaguchi et al. 2015). The two sub-species of cats have been reported to hybridize all over the area of repartition of the European wildcat (Randi 2001, Oliveira et al. 2007, Hertwig et al. 2009, Beaumont et al. 2001, Nussberger et al. 2014b, O'Brien et al. 2009, Lecis et al. 2006), and recent studies suggest that such widespread hybridization events already led to the introgression of

domestic genes into the wildcat genome (Hertwig et al. 2009). Such hybrid genome may represent an epidemiological threat because their intermediary genome as well as their roaming behavior (Germain et al. 2008, 2009) may constitute a bridge for pathogens to adapt to the wildcat genome and vice versa (Floate and Whitham 1993). Because hybridization is promoted by global changes such as habitat fragmentation (Crispo et al. 2011, Todesco et al. 2016), it is expected to increase in frequency in the future. A deeper knowledge of the conditions of sharing of pathogens between wildcats and domestic cats may thus allow to better assess the epidemiological threat associated with the presence of hybrids.

Wildcats infected by feline pathogens (FCV, FPV, FIV, FHV) have been reported in several studies (McOrist 1991, Daniels 1999, Leutenegger 1999, Racnick 2008, MacDonald 2015, Wasieri 2009, Millan et al. 2009). In all but one (Račnik et al. 2008), no genetic identification of individuals was conducted. Given that hybrid individuals cannot be reliably distinguished from parental individuals based on morphological traits and coat color alone (Ballesteros-Duperón et al. 2015), positive individuals identified phenotypically as wildcats without genetic confirmation may actually be hybrids and not true European wildcats. The extent to which European wildcats are threatened by the viruses of the domestic cats thus remains poorly documented.

In this study, we conducted a serological survey regarding four of the most common viruses of the domestic cat, the FPV, the FHV, the FCV and the Feline Immunodeficiency Virus (FIV), in a population of European wildcats from Northeastern France. This population of wildcats is surrounded by 15 populations of domestic cats infected by all of these four viruses with prevalence ranging from 20% to 81% depending on the virus (Hellard et al. 2011). Previous studies (Beugin et al. 2016) showed that the European wildcats in this population hybridize with the domestic cats meaning that contacts occur between the two sub-species at least during the breeding period. We thus expected that these interactions between the individuals could lead to cross-species disease transmission. Surprisingly, none of the 23 wildcats tested were infected by any of the four viruses. As we did not expect this null prevalence to result from the absence of detection of the viruses, we built a

dynamic and stochastic model in order to assess (1) if a null prevalence was expected given the prevalence of the four viruses in the 15 domestic cat populations under study, (2) under which transmission conditions between the two sub-species as well as within wildcat population a null prevalence was expected, and finally (3) if these conditions could be consistent with the ecology of the European wildcat and whether it may reflect the existence of a barrier preventing the transmission of the pathogen between the two subspecies.

5.1.2. MATERIAL AND METHODS

a. Sampling and serology

The fieldwork has been conducted by qualified people according to current French legislation. Accreditation has been granted to the UMR-CNRS5558 (accreditation number 692660703) for the program. Fifteen populations of domestic cats corresponding to 15 villages are included in our study. These populations largely varied in size, from 13 individuals in Clérey-la-Côte to 71 in Sauvigny, and were distant from each other by 1.2 to 4 km. A total of 511 domestic cats, among which 285 feral cats (with no owner), were sampled in those villages and tested for the FPV, the FCV, the FHV and the FIV (Hellard et al. 2011). None of the domestic cats was vaccinated and all roam freely. Overall, the estimated prevalence of the FIV, FHV, FCV and FPV were on average 15.99%, 59.82%, 81.75% and 21%, respectively (Hellard et al. 2011). However, the prevalence varied according to the population and different risk factors such as sex, age, way of life (owned vs. unowned), and phenotype (orange vs. non orange cats). Unowned feral cats were on average more infected by FIV (21.97% vs. 10%), FHV (67.8% vs. 54.15%) and FCV (86.64% vs. 77.56%) than owned cats, and less infected by FPV (15.91% vs. 36.59%).

The villages alternated with agricultural fields and forests where European wildcats lived. In the European wildcat population, females preferentially lived inside the forest with an access to the agricultural fields while males lived at the periphery of

the forest (Beugin et al. 2016). Twenty-five wildcats were captured using trapping cages treated with valerian roots (*Valeriana officinalis*), a common attractant for cats. All captures were made from November to February of each year between April 2008 and May 2011. Trapped individuals were anaesthetized with ketamine chlorohydrate (Imalgène 1000, 15mg/kg, Merial) and acepromazine (Vétranquil 5.5%, 0.5 mg/kg, Ceva). A blood sample was taken before the individual was released. Sex and age class (juvenile, sub-adult and adult) were determined based on morphometric characters (Pascal and Castanet 1978). The twenty-five wildcats corresponded to 4 juveniles (<1 year old, 4 females), 10 sub-adults (1-3 years old, 7 females and 3 males) and 11 adults (>3 years old, 6 females and 5 males).

The ELISA method (SNAP Combo +, Idexx) was used to detect FIV-antibodies. According to the manufacturer, this test's sensitivity and specificity are both 99.9% (Idexx). Specific antibodies against FHV, FCV or FPV were measured by a specific blocking ELISA (Poulet, 2007, see Hellard *et al.* 2011 for details). The FCV being highly antigenic, the ELISA test for this virus was conceived in order to measure antibodies whatever the antigenic variant of the sampling (Poulet et al. 2008). Antibodies can be detectable a few month to several years (1 to 7 years) after the infection (Scott and Geissinger 1999, Mouzin et al. 2004) depending on the virus.

b. Prevalence in the European wildcats and in hybrids

According to the relaxed approach used in Beugin et al. (in preparation), our dataset was composed of 23 European wildcats, and 2 hybrids that may correspond to wild backcrosses (Wild x F1). Among the 511 domestic cats tested for the four viruses by Hellard et al. (2011), 371 were genotyped (see Beugin et al. in preparation) and 16 of them were identified as domestic backcrosses (Domestic x F1). We tested whether hybrids were more infected by the four viruses compared to parental individuals with a binomial linear model in which infected cases were explained by the hybrid status of the individual (0 for parent, 1 for hybrid) and different risk factors (sex, age, way of life, phenotype) known to influence the susceptibility to these virus (Hellard et al.

2011). Due to the low number of wild backcrosses, this test was only performed to compare prevalence in domestic cats and in domestic backcrosses.

c. Description of the dynamic-statistical model

Intuitively, the probability to observe i infected individuals in a population of size N can be modelled by a binomial distribution with the parameters N and p , the probability of being infected. The use of this binomial model supposes however that all individuals are equally likely to be infected and that their probability of being infected is independent from the other individuals' epidemiologic status. For transmissible diseases, these two hypotheses are not verified. Susceptible individuals do not have the same probability of being infected than infected ones, and the probability of being infected depends on the status of the other individuals and notably whether they are infected or not. One possibility to account for this dependency is to model the probability of observing i infected individuals by a dynamical process of between-host transmission (Fouchet et al. 2009). These types of models are widely used for the study of infectious diseases (e.g. Berthier et al. 2000, Keeling and Eames 2005).

The dynamic-statistical model we propose is a compartment model in which individuals are distributed in different compartments that correspond to the different states in which they can be observed. We considered three states: susceptible, infected and recovered. Individuals could transit from one compartment to another according to epidemiological parameters (transmission rate, recovery rate) specific to each virus, and demographic parameters (mortality rate). These three compartments were all included for the FPV (SIR model). Individuals infected by the FPV become immune to this virus for several years after the infection (Scott and Geissinger 1997) while no such immunity is observed for the FCV, FHV and FIV. Furthermore, given the short life expectancy of domestic cats in this environment (four years, Hellard et al. 2011), we considered immunity to be acquired for life (Berthier et al. 2000). FPV can remain in *feces*-contaminated ground for one year (Csiza et al. 1971). Adding the

environmental contamination compartment would allow to better reproduce the dynamics of the FPV. However, to simplify calculations, we considered that individuals directly transmit the virus for one year before they recovered. Individuals can also recover from the FCV but, as they do not develop any immunity against the virus, they become susceptible again (SIS model) when the infection ends after one month on average (Radford et al. 2007). On the contrary, individuals infected with the FIV and the FHV do not recover once infected (SI model).

The models corresponding to the different viruses can be represented graphically with a transfer diagram (Figure 5.1). The infection of European wildcats was expected to occur according to two different paths: between-wildcat transmission (transmission rate $\beta'(N')$) or domestic-wildcat transmission (transmission rate τ). The FPV, FCV and FHV being transmitted by simple contact, we considered their transmission to be density-dependent, i.e. to depend on the total size of the European wildcat population. For these three viruses then, $\beta'(N') = \beta'$. The FIV is transmitted through bites and classically modelled using a frequency-dependent transmission (Fouchet et al. 2009), for this virus then, $\beta'(N') = \beta'/N'$.

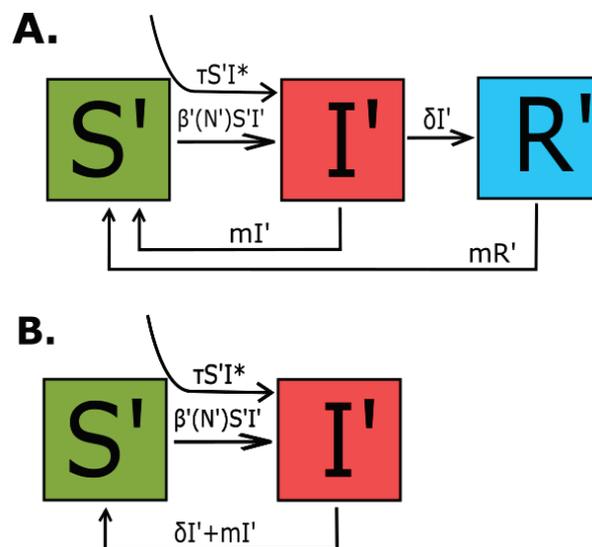


Figure 5. 1: Transfer diagram representing the dynamics of the FPV (A) and the FIV, FHV and FCV (B with $\delta=0$ for FIV and FHV).

Differential equations can be drawn from these diagrams to represent mathematically the dynamics of the between-host transmission in the European wildcat population. For the FCV, FHV and FIV, a system of two equations can describe the dynamics ($\delta = 0$ for FHV and FIV):

$$\begin{aligned}\frac{dS'}{dt} &= -\beta'(N')S'I' - \tau S'I^* + mI' + \delta I' \\ \frac{dI'}{dt} &= \beta'(N')S'I' + \tau S'I^* - mI' - \delta I'\end{aligned}\tag{1}$$

For the FPV, the system includes three equations:

$$\begin{aligned}\frac{dS'}{dt} &= -\beta S'I' - \tau S'I^* + m(I + R) \\ \frac{dI'}{dt} &= \beta S'I' + \tau S'I^* - mI' - \delta I' \\ \frac{dR'}{dt} &= \delta I' - mR'\end{aligned}\tag{2}$$

Where N' is the size of the European wildcat population, S' the number of susceptible wildcats, I' the number of infected wildcats, I^* the number of infected domestic cats per population of domestic cats at equilibrium. This number could be calculated by the resolution of a similar system of differential equations representing the dynamics in the domestic cat populations when all equations were considered equal to zero, and the system at equilibrium. Given that the viruses considered are common and have been circulating for years, we assumed that the domestic cat populations were at equilibrium:

$$\begin{aligned}\frac{dS}{dt} &= -\beta(N)SI + mI + \delta I \\ \frac{dI}{dt} &= \beta(N)SI - mI - \delta I\end{aligned}\tag{3}$$

It could also be efficiently averaged by $N*P$ where N is the size of a population of domestic cats and P the prevalence reported on average in domestic cat populations by Hellard et al. (2011). This model applied to the domestic population was also used in order to estimate the basic reproduction number R_0 associated with each virus, the transmission rate β by confronting the expression of the expected prevalence with the prevalence observed. The expression of β and R_0 depends on the compartment model considered and are reported in Table 5.1. The transmission parameter β' was expressed according to the parameters inferred from the domestic cat populations. To this end, we defined ρ as

$$\rho = \frac{R'_0}{R_0} \quad (4)$$

According to this definition, $\beta'(N')$ could then be expressed as $\rho N\beta(N)/N'$ in the equations. The equations presented so far allow for a deterministic study of the between-host transmission of the virus and do not allow to capture virus extinction processes. The stochastic version of these models must then be used in order to study the conditions under which a null prevalence is expected. These stochastic models include the same compartments and transitions between compartments that the deterministic model but the transition rates are drawn from a negative exponential law (Table 5.2).

Table 5. 1: models chosen for each virus and parameters (m : mortality, δ : recovering rate) associated with each of them. The prevalence (P) considered are the mean true prevalence found in Hellard et al. (2011) for these populations. N corresponds to the number of individuals present in a population of domestic cats.

Virus	Model	m	δ	β	P	R_0
FCV	SIS	0.0007	0.03	$\frac{m + \delta}{N(1 - P)}$	0.81	$\frac{\beta N}{m + \delta}$
FHV	SI	0.0007	0	$\frac{m}{N(1 - P)}$	0.60	$\frac{\beta N}{m}$
FIV	SI	0.0007	0	$\frac{m}{1 - P}$	0.16	$\frac{\beta}{m}$
FPV	SIR	0.0007	0.003	$\frac{m + \delta}{N(1 - P)}$	0.20	$\frac{\beta N}{m + \delta}$

Table 5. 2: Transitions and associated transitions rates in the stochastic model

Event	Transition	Transition rate
Between-wildcat infection	$(S,I,R) \rightarrow (S-1, I+1, R)$	$\beta(N)S'I'$
Domestic-wild infection	$(S,I) \rightarrow (S-1, I+1)$	$\tau S'I^*$
Recovery	$(S,I) \rightarrow (S+1, I-1)$	$\delta I'$
Recovery (FPV)	$(S,I,R) \rightarrow (S, I-1, R+1)$	
Death of an infected	$(S,I,R) \rightarrow (S+1, I-1, R)$	mI'
	$(S,I) \rightarrow (S+1, I-1)$	
Death of a recovered	$(S,I,R) \rightarrow (S+1, I, R-1)$	mR'

Given that we considered that populations were independent and of constant size, this model, based on a continuous-time Markov process, can be described by the following ordinary differential equations (Fouchet et al. 2009) for FHV, FIV (with $\delta = 0$) and FCV:

$$\begin{aligned} \frac{dp_i(t)}{dt} = & - \left[\tau I^* (N - I') + \frac{\rho \beta (N) N}{N'} (N' - I') I' + m I' + \delta I' \right] p_i(t) \\ & + \left[\tau I^* (N - (I' - 1)) + \frac{\rho \beta (N) N}{N'} (N' - (I' - 1)) (I' - 1) \right] p_{i-1}(t), \\ & + [m(I' + 1) + \delta(I' + 1)] p_{i+1}(t) \end{aligned} \quad (5)$$

where $p_i(t)$ is the probability of observing exactly i infected individuals at a time t . We tested two different values of N' ($N' = 30$ and $N' = 50$) based on the 42 wildcats sampled in Beugin et al. (2016). As N' value did not change our conclusions (Appendix S5.1), we presented results for $N' = 30$. Two situations corresponding to two values of N and P were considered: a situation where all domestic cats (owned and unowned) can transmit the virus ($N = 50$, $P =$ average over all cats), and the other one where only unowned cats transmit the viruses ($N = 25$ as owned and unowned cats were equally represented in each population, and P corresponding to the prevalence in unowned cats).

For the FPV, we decided to build a simplified model where we expressed the number of recovered wildcats R' as $\frac{\delta}{m} I'$ according to equation (2) when we considered the system at equilibrium. Using this simplification, the equation becomes:

$$\begin{aligned}
 \frac{dpi(t)}{dt} = & - \left[\tau I^* \left(N - I' - \frac{\delta}{m} I' \right) + \frac{\rho \beta(N)N}{N'} \left(N' - I' - \frac{\delta}{m} I' \right) I' + mI' + \delta I' \right] p_i(t) \\
 & + \left[\tau I^* \left(N - (I' - 1) - \frac{\delta}{m} (I' - 1) \right) + \frac{\rho \beta(N)N}{N'} \left(N' - (I' - 1) - \frac{\delta}{m} (I' - 1) \right) (I' - 1) \right] p_{i-1}(t) \\
 & + [m(I' + 1) + \delta(I' + 1)] p_{i+1}(t)
 \end{aligned} \tag{6}$$

Given that the four viruses we considered have been infecting domestic cats for years and that the wildcats and domestic cats have been living in sympatry for many years (Stahl et al. 1988, O'Brien et al. 2009, Beugin et al. 2016), we assumed that the populations had reached equilibrium and that the distribution of the probability to observe i infected individuals was stable. We could thus assess the probability of observing zero infected individuals according to the value of the parameter τ and ρ by resolving the system of equations when all of them were equal to zero using the computer program MATLAB with a Gaussian elimination method and the condition:

$$\sum_{i=0}^{N+1} p_i = 1 \tag{7}$$

We used two parameters as measures of the between sub-species transmission and between wildcat transmission respectively. The parameter φ , defined as the ratio between the number of effective contacts occurring between domestic cats $\beta(N)N$ and the number of effective contacts occurring between domestic cats and wildcats $\tau N'$, was used as a measure of the between-species transmission. Increasing values of φ corresponded to effective contacts far more frequent between domestic cats than between domestic cats and wildcats, and thus to a loss of contacts between sub-species. The parameter ρ , described previously, was used as a measure for the propagation loss within the wildcat population. This parameter was expected to vary

between 0 (no propagation in the wildcat population) and 1 (similar propagation within wildcats and within domestic cats). The limit beyond which the probability to observe zero infected individuals was below 5% was drawn according to these two parameters. All combinations of these two parameters corresponding to the area below the line could statistically lead to the observation of a null prevalence (Figure 2). We have varied the number of domestic populations transmitting the viruses, and we represented the corresponding lines. All 15 populations of domestic cats were likely to transmit viruses to the wildcat population. However, they did not have the same size or prevalence for all four viruses, and are not all as close to the wildcat population, their combined effect was not expected to be fifteen times the effect of one neighboring population. To understand how the number of transmitting domestic populations can impact the limit, we added the parameter π in the previous equations, which corresponded to the number of populations exchanging viruses with the wildcat population. For instance, for FCV, FIV and FHV, we added the parameter π to equations (1) as follows:

$$\begin{aligned} \frac{dpi(t)}{dt} = & - \left[\tau I^* (N - I') * \pi + \frac{\rho \beta (N) N}{N'} (N' - I') I' + m I' + \delta I' \right] p_i(t) \\ & + \left[\tau I^* (N - (I' - 1)) * \pi + \frac{\rho \beta (N) N}{N'} (N' - (I' - 1)) (I' - 1) \right] p_{i-1}(t) \\ & + [m(I' + 1) + \delta(I' + 1)] p_{i+1}(t) \end{aligned} \quad (8)$$

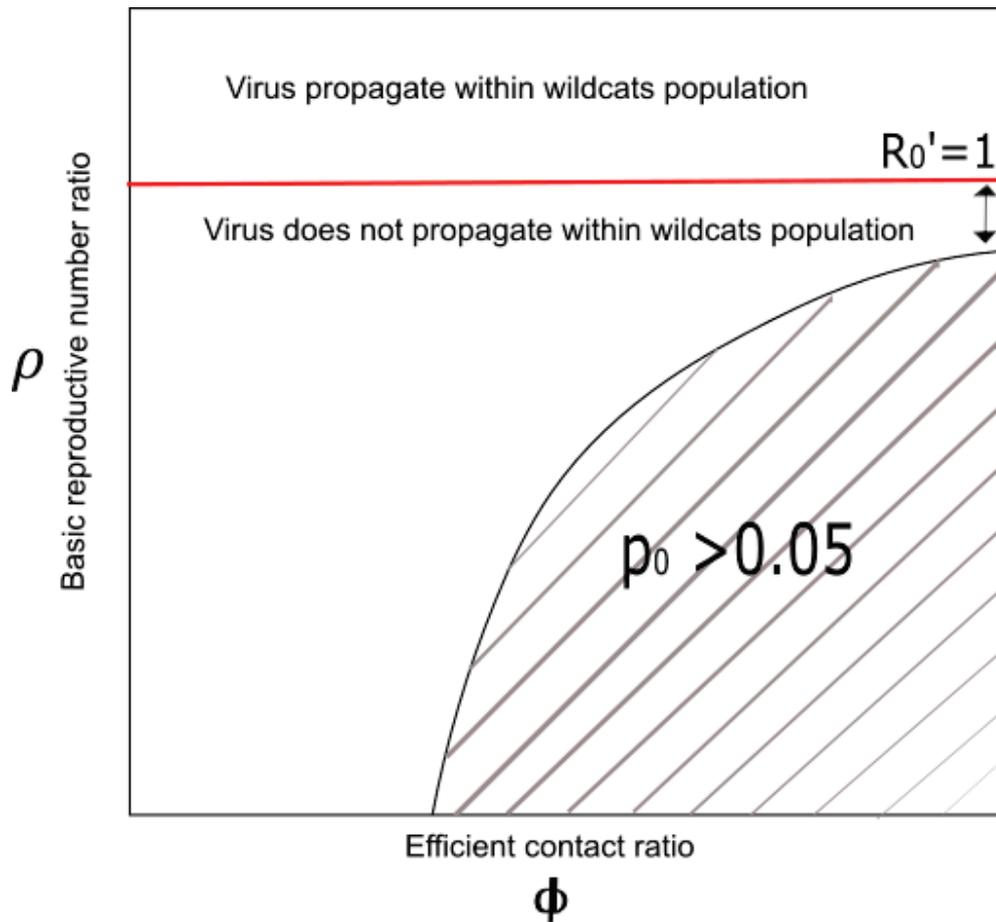


Figure 5. 2: Theoretical illustration of the kind of results expected through our modelling approach. The hatched zone corresponds to area where the observation of a null prevalence is probable at a risk of 5%. The vertical red line corresponds to the reintroduction delay between domestic cats' populations Φd and the horizontal line to the value of ρ corresponding to $R'_0 = 1$. Below this line, the virus is not expected to propagate in absence of external reintroduction, above it, it can circulate.

5.1.3. RESULTS

a. Prevalence in domestic cats, wildcats and hybrids

Pure domestic cats and domestic backcrosses did not differ regarding the prevalence of FIV, FHV and FPV. On the contrary, domestic backcrosses were significantly more infected than pure domestic cats by FCV ($F = 5.03$, $df = 1$, $p = 0.026$).

Table 5.3.: Number of infected individuals over the total number of individuals tested for three hybrid categories: pure wildcats (wildcat), wild backcrosses (F1 x Wildcat) and domestic backcrosses (F1 x Domestic cat).

Genetic Status	FIV	FHV	FCV	FPV
Wildcat	0/23	0/23	0/23	0/23
Wild backcrosses	0/2	0/2	0/2	0/2
Domestic backcrosses	5/16	9/16	16/16	7/16

b. Conditions of circulation and null prevalence

The results presented here were obtained for a wildcat population containing 30 individuals and with only unowned domestic cats transmitting the viruses. The results obtained with all the domestic cats and 50 European wildcats are presented in the appendix (Figure S5.1).

For all four viruses, the observation of zero infected individuals in the European wildcat population was expected under variable conditions depending on the virus (Figure 5.3). A null prevalence could be expected for φ values ranging from 1.2 (FPV) to 50 (FCV) for $R'_0=0$, and from 2 (FIV) to 100 (FCV) for $R'_0=1$. A lower number of contacts was then expected between domestic cats and European wildcats compared to between domestic cats to explain the observation of zero infected wildcats. Increasing the number of domestic cat populations applying an epidemiological pressure on the European wildcat population increased the limit values of φ compatible with the observation of a null prevalence. This increase in π affected more the FHV and the FCV for which the minimal value of φ compatible with a null prevalence, for $R'_0=1$, was multiplied by a factor 13 between the case with $\pi=1$ and $\pi=15$. For the FPV and the FIV, factors of 11 and 10 were found respectively.

For the FCV and the FHV, we found a maximal value for R'_0 beyond which we do not expect to observe a null prevalence in the European wildcat population (2.31 and 2.5); this corresponded respectively to 30% and 70% of the R_0 value characterizing the domestic cat population. We did not observe such a plateau for the FIV and the FPV, which displayed a wide range of R'_0 values compatible with a null prevalence.

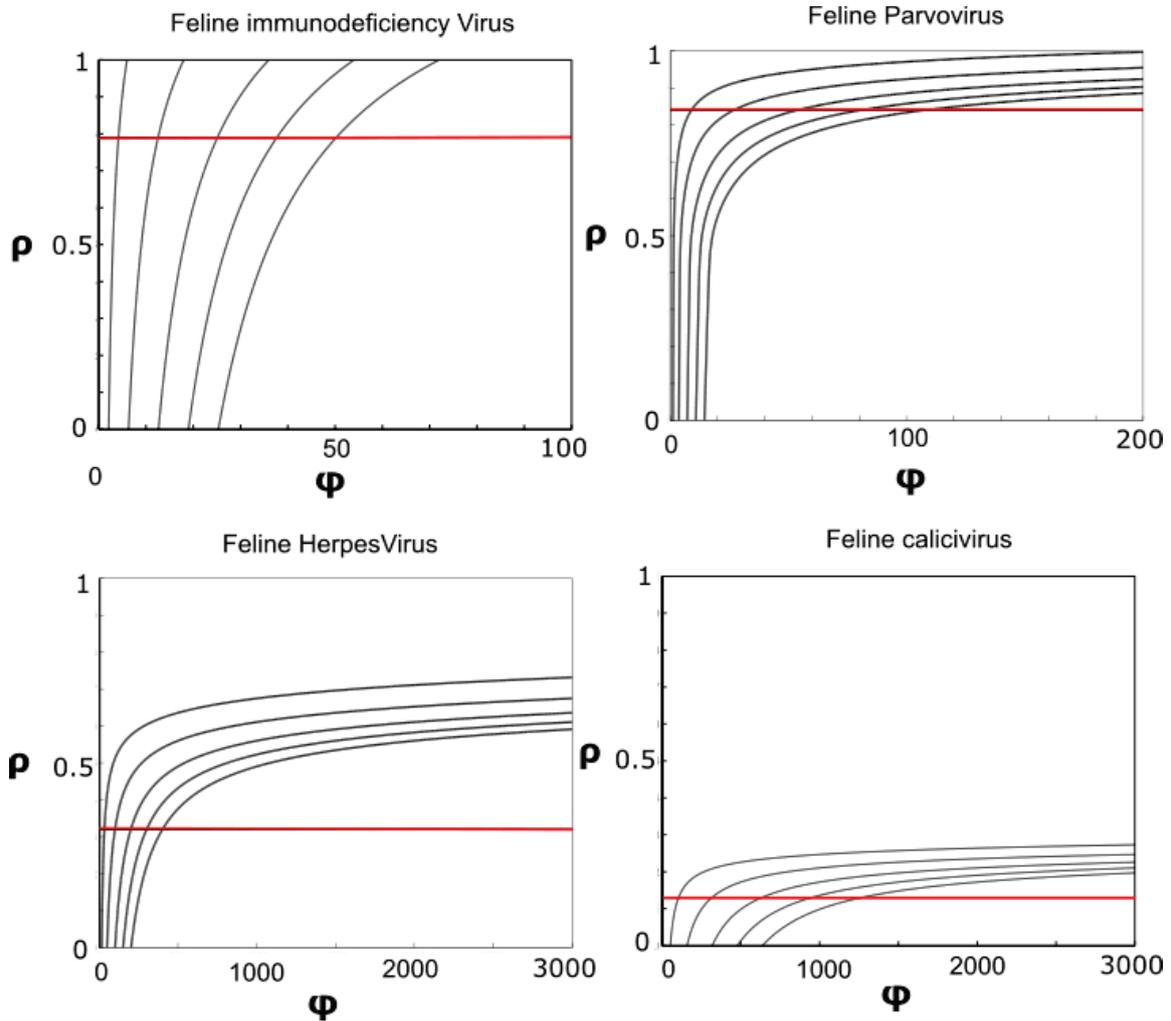


Figure 5. 3: Differential number of efficient contacts between domestic cats and European wildcats (φ), and the basic reproductive number ratio (ρ) for which a null prevalence is expected in the wildcat population, for a number of domestic populations transmitting the virus from 1 to 15 (lines correspond to 1, 3, 6, 9 and 15 populations) with a 5% risk. The horizontal line states for which ratio ρ the basic reproductive number of the wildcat population is below 1 and thus for which ratio ρ the virus does not propagate in the wildcat population in the absence of reintroduction from the domestic cat populations.

5.1.4. DISCUSSION

The present study reports the absence of wildcat contact with all four viruses (FIV, FCV, FHV, FPV) even for the oldest wildcats (4 years old) despite the endemicity of these viruses (antibodies are detected in all age classes in the domestic cat populations, with prevalence of antibodies reaching up to 81%; Appendix, Figures S5.2 and S5.3) in the surrounding populations of domestic cats. The only two serological surveys associated with genetic identification of the wildcats (Duarte et al. 2012, Račnick et al. 2008) did not report any infected individual for FIV and FPV for a similar number of individuals tested ($N = 15-28$ in Duarte et al. 2012, $N = 17$ for Račnik et al. 2008). In both cases, the European wildcats were sampled in a protected area, poorly fragmented, and with a low urban pressure, which may be associated with a low domestic cat pressure. On the contrary, the European wildcats of our study live in a highly fragmented environment with a high domestic cat pressure. The absence of viruses in this local wildcat population is thus particularly intriguing.

a. A barrier for FCV and FHV

A null prevalence for FCV and FHV reflected a loss of propagation from 30% (FHV) to 70% (FCV) in the wildcat population and thus, probably the existence of a barrier to the transmission of these viruses. In addition, less frequent between sub-species effective contacts compared to the between domestic contacts were associated with a null prevalence. This difference in the number of effective contacts between sub-species was more important for FCV ($\varphi = 100$ for FCV vs. $\varphi=30$ for FHV for $R'_0 = 1$) than for FHV. Thus, the observation of a null prevalence for FCV may more likely reflect the existence of a barrier between the two sub-species. Such a loss of propagation and effective contacts can be explained either by a loss of contacts involving wildcats or a cost for viral strains infecting wildcats. Current knowledge does not allow us to further discuss the existence of a cost. On the contrary, FCV and FHV being transmitted during friendly contacts (Gaskell and Povey 1982), their transmission may be lowered by low levels of sociality (Hellard et al. 2012). The

solitary way of life of the European wildcat (Sunquist and Sunquist 2002) is thus expected to significantly reduce transmission opportunities.

b. An inconclusive pattern for FIV and FPV

For FIV and FPV, our model failed to show categorically that the null prevalence observed reflects the existence of barriers. However, this epidemiological model relies on several assumptions that tend to underestimate either the epidemiological pressure undergone by the wildcat population (no contact with other wildcat populations taken into account) or the circulation of viruses (cross-immunity, no co-infection). Our model is then conservative. Thus, the conditions associated with a null prevalence are expected to be even more restrictive if we relax these assumptions and make our model closer to the reality.

The parvovirus is a virus known to easily jump the species barrier (Kerr et al. 1995, Allison et al. 2012) and have been reported in a European wildcat from a wildlife park (Wasieri et al. 2009). We may then expect R'_0 to be close to 1. In this case ($R'_0 = 1$), a null prevalence is expected for effective contacts between domestic cats 10 to 110 times more frequent than effective contacts between domestic cats and wildcats. Several hypotheses may explain these results: the two sub-species may adjust their behavior to avoid contact with each other; the two sub-species live in different habitats ; finally, it may also result from the fact that feces (the main transmission path for FPV) are used to mark territories and are not distributed randomly in the environment (Piñeiro and Barja 2012). As European wildcats and domestic cats occupy the same territories in a limited way (Ferreira et al. 2011, Sarmiento et al. 2009), individuals may be confronted to the feces of hetero-specifics only occasionally. On the contrary, for the FIV, a strain specialist of the domestic cat can be expected. This virus has indeed never been detected in European wildcats (Duarte et al. 2012, Račnik et al. 2008) and phylogenetic analyses comfort the existence of rare cross-transmission events between felid species (Troyer et al. 2008). For low R'_0 , which would reflect a low propagation of the domestic strain in the

wildcat population, a null prevalence is expected for domestic cats having effective contacts 3 to 25 times more frequently with their conspecific than with wildcats. A behavioral barrier may be involved. FIV is indeed transmitted through bites from males to females during reproductive events or during fights between males, for the monopolization of females or territories (Sparger 1993). A previous study led on this population of wildcats suggested that hybridization events are mostly expected between wild males and domestic females (Beugin et al. 2016). Females are not involved in the transmission of the virus (Courchamp et al. 2000), thus cross-transmission of FIV during reproductive events is probably a rare event. Furthermore, the periods of reproduction of the domestic cats and the European wildcats poorly overlap (Condé and Schauenberg 1974, Gagnon and Dantzer 2013), thus, aggressive contacts between males for the monopolization of females may not occur on a frequent basis. Lastly, home-ranges of the two sub-species do not overlap (Germain et al. 2008, 2009, Ferreira et al. 2011), competitive interactions for territories are then probably limited.

c. Infectious status of hybrids

Hybrids were more frequently found infected by FCV. This result suggests that hybrids could be differentially infected by viruses and may, if frequent in the environment, act as a hybrid bridge (Floate and Whitham 1993) promoting the transmission of pathogens from domestic cats to European wildcats. This bridging capacity would result from intermediary behavioral traits (Biró et al. 2005, Germain et al. 2009, Kilshaw et al. 2015). In our study, hybrids were equally distributed among owned and unowned individuals in the domestic population and thus did not appear behaviorally intermediary, while in the wildcat population, they were captured outside the forest, suggesting an intermediary behavior. The epidemiological threat represented by hybrids remains thus to be investigated. Also, the investigation of the evolutionary history of the viruses, and notably of the genetic flows that may have shaped current strains, should draw more attention. It should allow to better

understand the molecular processes involved in the acquisition of resistance and shed light on mechanisms attached to the existence of a genetic barrier and to the disruption of this barrier due to host hybridization.

5.1.5. CONCLUSION

The local population of European wildcats presented in this study does not appear threatened by the domestic cats' viruses. In addition, hybridization remains limited and may not represent an issue for the transmission of viruses between the two sub-species. However, preventive measures limiting hybridization and virus circulation, such as neutering and vaccination, should be taken in order to prevent future outbreaks mediated or not by hybrids in the European wildcat population.

5.1.6. APPENDICES

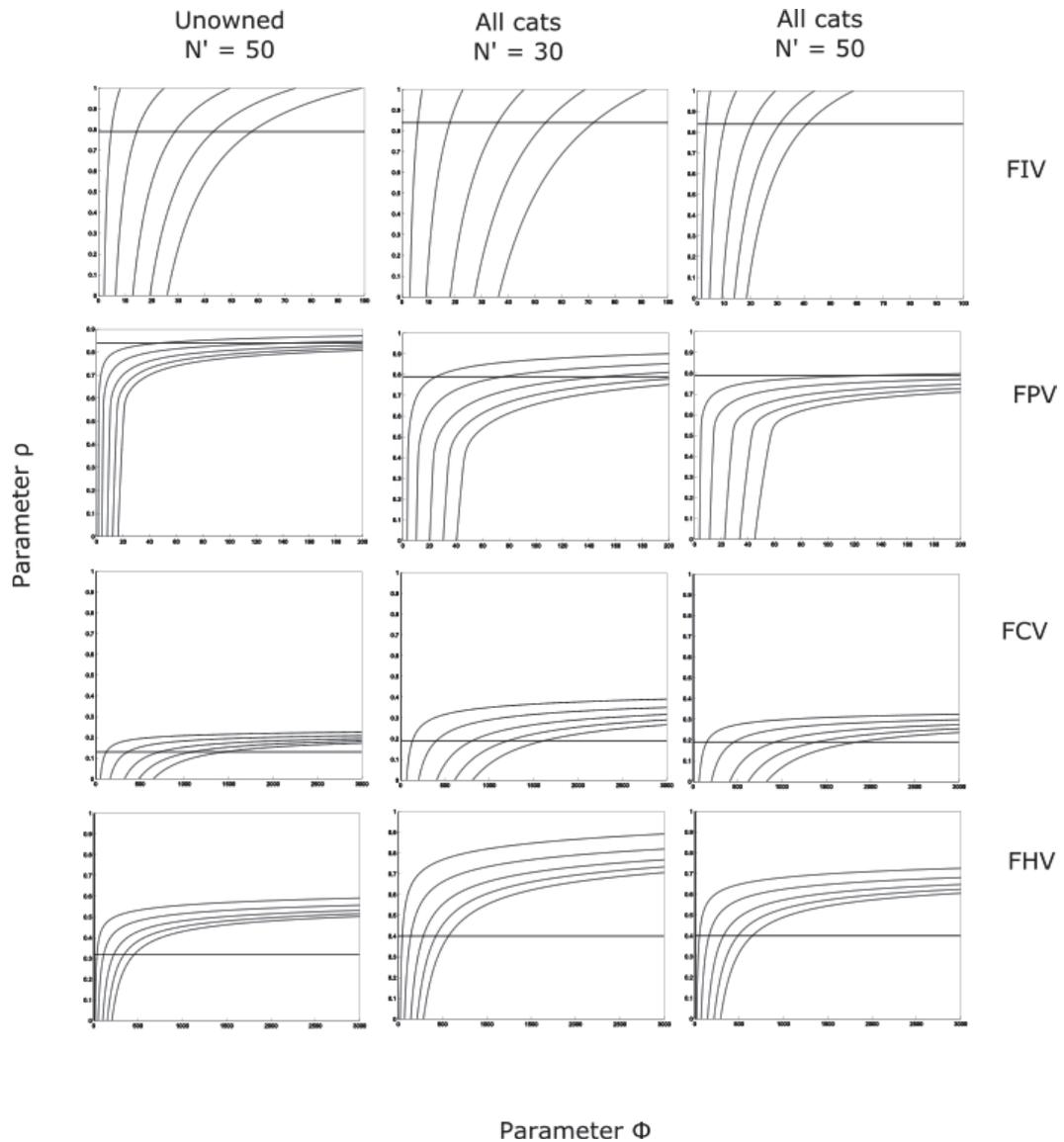


Figure S5. 1: Combination of φ and ρ for which a null prevalence is expected in the wildcat population, for a number of domestic populations transmitting the virus from 1 to 15 (lines correspond to 1, 3, 6, 9 and 15 populations) with a 5% risk and indifferent conditions (all domestic cats considered or only unowned cats, 30 of 50 wildcats). The horizontal line states for which ratio ρ the basic reproductive number of the wildcat population is below 1 and thus for which ratio ρ the virus does not propagate in the wildcat population in the absence of reintroduction from the domestic cat populations.

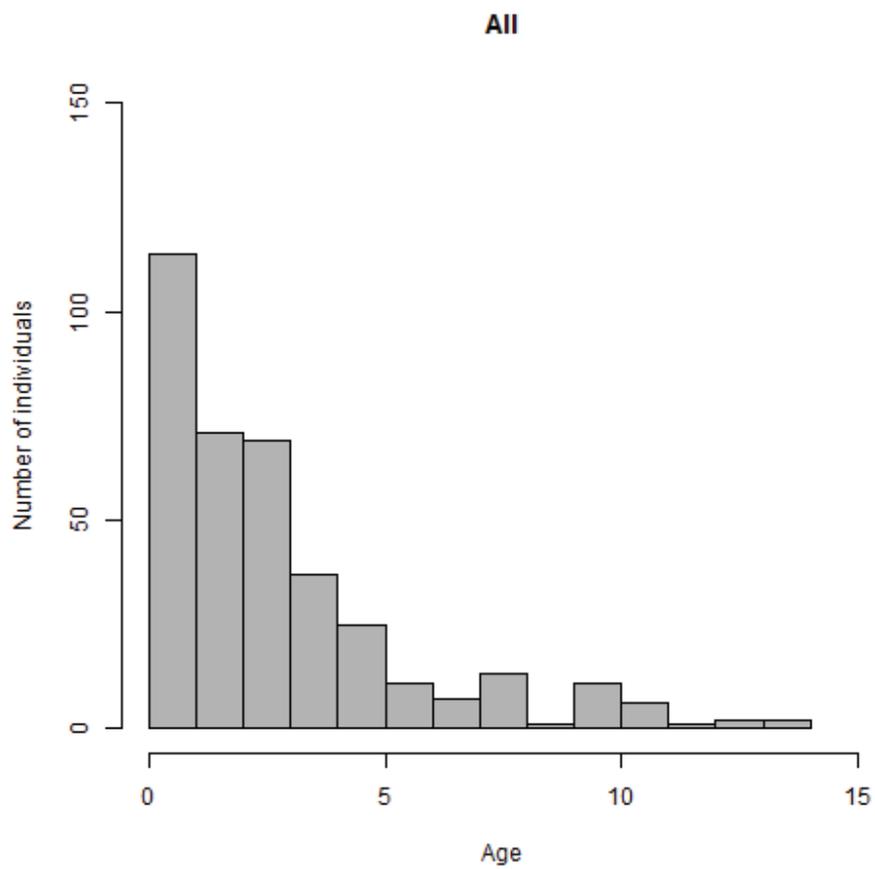


Figure S5. 2: Distribution of the ages in the entire population of domestic cats.

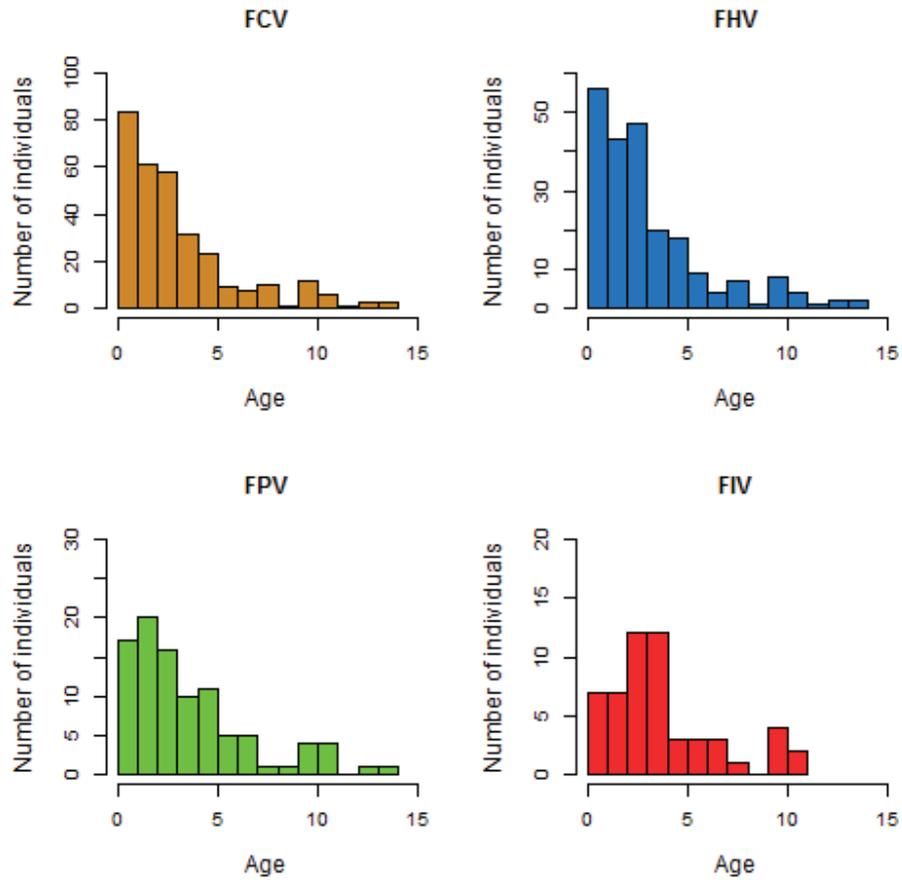


Figure S5. 3: Distribution of the ages of the infected individuals for each of the four viruses considered in the study.

Addendum

The null prevalence observed in the local population of European wildcats led to the formulation of two hypotheses: either the strains observed, sometimes at very high prevalence, in the domestic cats' populations could not infect wildcats because the differentiation between the two cat sub-species was enough to impose a prohibitive cost to such generalist transmission; or, a behavioral barrier existed between wildcats and domestic cats which prevented the passage of pathogens from domestic cats to wildcats.

Our work showed that the existence of a barrier to the transmission of the virus between the two sub-species is plausible. The behavioral differences between wild and domestic cats appearing to be of crucial importance. Such differences may be partly coded by regions of the genome that have evolved during the domestication of the domestic cat (Montague et al. 2014, Tamazian et al. 2014) or originate from its ancestor the African wildcat (Vigne et al. 2004, Driscoll et al. 2007). Thus, the mixing and introgression of genes may soften these differences and weaken the barrier protecting wildcats from the pathogens of domestic cats. In this particular population, the epidemiological risk seems contained at present but an increase of hybridization in the future may disrupt current epidemiological equilibrium.

The absence of a behavioral barrier would mean that generalist strains circulate in the area of study. In this area, domestic cats and wildcats are distributed according to a meta-population configuration that has been shown to impact the persistence of generalist strategies (Dennis et al. 2011, Hesse et al. 2015). In order to test if a generalist strain, able to infect both the domestic cats and the wildcats could, persist in our area of study, we built a stochastic and dynamic epidemiological model. Preliminary results showed a high variability in the persistence of a generalist strain depending on the epidemiological dynamics but that this persistence is expected in many cases.

6. General discussion and perspectives

Less than sixty-five years separate us from the groundbreaking discovery of the double-strand structure of the DNA molecule by James Watson and Francis Crick (Watson and Crick 1953). Yet, genetic markers are now widely used notably in ecology where the rise of genetic tools has, among other advances, completely revolutionized our vision of the tree of life, of the species it contains and the evolutionary history that led to their existence today (Delsuc et al. 2005). Specifically, it allowed to capture the importance of processes such as hybridization which were traditionally discarded because considered as minor.

The field of hybridization studies has been growing these past decades as new methodological and genetic approaches were developed. Hybridization has now been detected in many different species and sub-species thanks to the gain of resolution allowed by these new developments (Grant and Grant 1992, Lodé et al. 2005, Harrison et al. 2014, Vuillaume et al. 2015, Godinho et al. 2015, Pongracz et al. 2017). The importance of this evolutionary process has made essential the transfer of concepts and methods from academical environments to applied structures, which are more closely involved in everyday monitoring of wildlife species.

The general objective of my PhD was to master methods and concepts associated with hybridization studies, in order to re-use them in the context of Antagene's activities. Ideally, I wanted to establish a routine easily applicable for any situation. However, given the many questions attached to the study of hybridization, instead of establishing a routine, I switched towards identifying several guidelines to help choose the most appropriate method according to the question addressed. Hereafter, I discuss these different guidelines. Then, I discuss another category of guidelines, those proposed regarding the management of hybrids and propose to integrate epidemiological criterions in a decision tree otherwise composed on ecological considerations only. Finally, to conclude on the European wildcat, I address the question as to know whether the domestic cat represents a threat for its conservation in the populations

studied here, and propose future developments for the study of the hybridization and viruses' circulation between domestic cats and European wildcats.

6.1. METHODOLOGICAL GUIDELINES FOR THE STUDY

The sampling scheme is of primary importance as all the analyses depend on this very first step. A full part will be dedicated to this essential step. Here, I will thus discuss guidelines for the selection of a genetic and statistical approach.

6.1.1. THE GENETIC APPROACH: ARE NGS UNAVOIDABLE?

Next-generation sequencing (NGS), which allow for a faster and cost-effective sequencing of genomes (Metzker 2010), are considered as the future of hybridization and introgression studies (Twyford and Ennos 2012). Recently, their application to hybridization issues have largely focused on human species and showed how modern humans hybridized with *Homo neandertalis* and which neandertal genes introgressed and are today part of our genome (Sankararaman et al. 2014). It has also been used in other species, for example in Canids (*Canis spp.*, Anderson et al. 2009, VonHoldt 2011, 2016), Butterflies (*Lycaeides spp.*, Gompert et al. 2014) or Suckers (*Catostomidae spp.*, Mandeville et al. 2015). These studies mostly aimed at understanding the importance of reticulated evolution in the formation of species, or at characterizing introgression shedding light on regions that introgress preferentially.

In spite of this rise of NGS techniques, medium-sized panels (MSP) of genetic markers are still widely used for the study of hybridization (Khosravi et al. 2013, Francisco et al. 2014, Pongracz et al. 2017). Classically, the ability to detect efficiently hybrid categories, i.e. the resolution of a panel of genetic

markers, is considered to depend on a trade-off between the number of markers it contains and the differentiation between the hybridizing populations (Vahä and Primmer 2006). The ability to detect hybridization using MSPs may thus be questioned, especially between loosely differentiated populations, such as wild species and their domestic counterparts. If we cannot change the evolutionary history that shaped the genomes of the populations nor the actual genetic differentiation between them, we can select genetic markers that are not representative of the entire genome of the species but allow to inflate the differentiation between them (Nussberger et al. 2013, Godinho et al. 2015, Oliveira et al. 2015) and move along the number of loci – differentiation trade-off. For single nucleotide polymorphism markers, two alleles are generally observed per locus (Nussberger et al. 2013), thus, the selection of highly differentiated markers can be achieved by the inclusion in the panel of markers for which alternative alleles are fixed in the hybridizing populations. Panels of highly differentiated SNP have been built to assess hybridization between the European wildcat and the domestic cat and reached F_{st} value of 0.96 (Nussberger et al. 2013) and 0.74 (Oliveira et al. 2015) while the differentiation between the two sub-species are usually assessed to range between 0.12 and 0.22 (Randi 2008, Hertwig et al. 2009, Oliveira et al. 2008). The identification of diagnostic markers is not so easy with microsatellites given their high polymorphism but highly differentiated panels of markers can be proposed. Godinho et al. (2015) thus proposed a panel of 13 microsatellites allowing to raise the F_{st} value from 0.19 to 0.34. These panels proved efficient for the detection of first and second- generation hybrids (F1 and F2) and on average 80% of the first-generation backcrosses.

For now, the selection of these markers is mostly based on the differentiation between hybridizing populations. However, other parameters such as the number of alleles per locus, or the observed heterozygosity may impact the resolution of a panel of markers and may be relevant parameters to take into account, specifically for microsatellite markers selection. In many

situations, population do not have the same genetic diversity and the backcrosses to the population that shows a smallest number of alleles and heterozygosity are usually better detected (Godinho et al. 2015, Beugin et al. 2017). Further investigations may thus be led to see how such asymmetry may impact the detection of hybrids and may be used in the selection of microsatellite markers.

In the future however, the use of microsatellite markers for the detection of hybrids and the study of introgression may be compromised (Nussberger et al. 2013); SNP panels may be favored, and even more given the need to multiply local studies (see hereafter) and thus collaborations between laboratories. However, microsatellite markers remain highly relevant for population structure and parentage studies which can provide crucial information to understand hybridization processes (Chapter 4). A combination of markers may thus provide the best results in the future.

Globally, the efficient detection of hybrids passed the second-generation backcrosses with current MSPs remains limited (Godinho et al. 2015, Lancaster et al. 2006, but see Nussberger et al. 2014 for a larger panel composed of 68 diagnostic SNP markers). Two related questions arise from this statement: is this a limitation for the use of MSPs? And when does a hybrid stops being one (Daniels and Corbett 2003)? Second generation backcrosses correspond to hybrid with 87.5% of their genome that belong to one of the parental populations on average, third generation backcrosses to hybrids with 93.5% belonging to one population. Many authors consider 90% as an efficient threshold to distinguish parental individuals from hybrids (Lancaster *et al.* 2006; De La Torre *et al.* 2015; Le Roux *et al.* 2015; Guildea *et al.* 2015). We may expect some second generation backcrosses and most third generations to exceed this threshold. Their management as hybrids thus does not seem obvious. Given the recent recommendations for hybrid management (Daniels and Corbett 2003, Wayne and Schaffer 2016) which favor ecosystemic function

over genetic purity, a race for the detection of old generations of backcrosses does not seem relevant. In addition, for the regular monitoring of species, actions conducted on young generations of hybrids allow *de facto* a control of older generations and introgression to some extent. The use of medium-sized panels thus remains relevant for applied study of hybridization thanks to the compromise they represent between scientific and economic requirements. Next-generation sequencing represents however a huge source of information notably for the construction of these panels (see appendix 3) as it gives access to hundreds, even thousands, of markers (Twyford and Ennos 2012, Loughnan et al. 2015, Maduna et al. 2017) and thus increases the chances of finding highly differentiated or diagnostic markers.

Table 6. 1: Summary of the issues encountered across studies dealing with hybridization and methods proposed to deal with them. MSP corresponds to Medium-Sized panel, AIM to Ancestry Informative markers (acronym used in the literature, corresponds to MSP made of highly differentiated markers).

	QUESTION	METHOD	BIASES/LIMITS
SAMPLING SCALE	Is there hybridization?	Large/Local sampling	Hybrids relative to which parents?
	Hybridization determinants?	Local sampling	Extensive sampling
	Individual identification of hybrids	Local sampling	Hybrids relative to which parents?
	Characterization of introgression (past dynamics, gene identification)	Large sampling	Economic cost
SAMPLING MATERIAL	Hybridization rate	Blood/feces/Hairs/oral swab/skin	-
	Hybrid susceptibility to pathogens	Blood/Oral swab	-
GENETIC APPROACH	Identification of hybrids for monitoring	AIM based on SNP	Biased general population genetics
	Population genetics and hybridization	AIM based on SNP+microsatellites	Cost/use of several panels
	Characterization of the introgression	Next-generation sequencing (NGS)	Economical cost
	Collaboration between laboratories	SNP rather than microsatellites	More markers required
	Optimal identification of parents	Relaxed approach	Parent misclassification
STATISTICAL APPROACH	Optimal identification of hybrids	Conservative approaches	Hybrid misclassification
	Range of possible hybridization rate	Combination of methods	-
	Categorization of the individuals	NEWHYBRIDS, <i>snaplust</i>	Risk of being too greedy

6.1.2. CHOOSING A METHOD ADAPTED TO THE QUESTION ADDRESSED

The capacity to detect old generations of backcrosses depends on the panel of genetic markers used for the analysis but also on the computer programs and the associated statistics chosen to analyze the data. Depending on the statistical routine, the same individuals will not be detected as hybrids. The computer programs and the processing of the data delivered by these programs are indeed more or less conservative, this means that they are more or less likely to categorize an individual as a hybrid. Similarly to what was presented in Chapter 3 (Part 1), conservative methods will tend to detect less hybrids while relaxed methods will, on the contrary, detect more easily individuals as hybrids. Both approaches can be useful and the extent of their usefulness is once again strongly dependent on the question addressed regarding hybridization.

When the objective is to detect parental individuals, for instance for restocking programs such as those which are commonly performed in Galliforms (Sokos et al. 2008), one wants to maximize the chances that the detected parents are true parents in order to restock the wild population with parental genes. In this case, relaxed methods are the most indicated as any doubtful individual will be categorized as hybrid. The iterative algorithm described in Chapter 3 (Part 1) can be used under a relaxed approach and the *snapclust* method can be seen as a relaxed counterpart of the program NEWHYBRIDS as it identifies limit individuals more easily as hybrids (Chapter 3, Part 2). On the contrary, when killing policies are conducted (Landelle 2005, Jackiw et al. 2015), one wants to eliminate as much as possible hybrids, especially when the hybrids result from the cross-reproduction of an endangered species. In such cases, the conservative methods will allow to detect as hybrids only individuals with strong cues of hybridization. Again, in Chapter 3 (Part 1), we proposed a conservative use of the iterative algorithm

that identifies only the most striking hybrids. Other computer programs allow to detect hybrids in a conservative way like BAPS (Corander and Marttinen 2006) or the method described by Bohling et al. (2011) which consists in assigning individuals as hybrids when the credibility interval provided by STRUCTURE along with the assignment probability, does not contain zero or one. The advantage of these two methods is that they are faster than the one proposed in Chapter 3. For theoretical purposes, several methods, conservative and relaxed methods but also geometrical (Discriminant Analysis of principal components for instance) and population genetics' based methods (STRUCTURE, *snappclust*), are usually combined to have a range of possible rates of hybridization, and to provide a general characterization of the population (Randi 2008, Bohling et al. 2011, Couch et al. 2016).

Finally, one may want to identify the hybrid category (F1, F2, etc) of the individuals. For this purpose, two computer programs have been specifically designed: NEWHYBRIDS (Anderson and Thompson 2002) and *snappclust* described in Chapter 3. The efficient classification of an individual in its hybrid category is more difficult to achieve than the mere distinction between parental individuals and hybrids, and requires the use of more markers (Vahä and Primmer 2006). In those programs, the user needs to determine which hybrid categories he wants to consider. The definition of hybrid categories indistinguishable from the parental populations is a source of mistakes and wrong assignments of parents as hybrids and, reciprocally, of hybrids as parents. Thus, further development of a rigorous method allowing to assess which hybrid categories one can reasonably consider is required given the panel of genetic markers used for the study. We expect such method to rely on the same parameters (number of loci, differentiation, number of alleles, heterozygosity) than those used to build MSP of highly differentiated markers mentioned earlier (Nussberger et al. 2013, Godinho et al. 2015, Oliveira et al. 2015)

Taking time evaluating which approach is the most relevant regarding the question we intend to answer may lead to save hours of statistical analysis given that some computer programs are highly time-consuming. Then, for efficiency purposes as much as scientific relevance, this decisional step must not be neglected.

6.2. THE SAMPLING SCHEME: A PRIMORDIAL STEP FOR STUDIES' COMPARABILITY

6.2.1. LIMITS OF THE OPPORTUNISTIC SAMPLING SCHEME

The sampling schemes in hybridization studies are much diversified: opportunistic sampling of road-killed individuals (Germain et al. 2009, Say et al. 2012, Galov et al. 2015), individual trapping (Lodé et al. 2005, Garroway et al. 2010, Chapter 4-Part1), non-invasive sampling (Schwartz et al. 2004, Godinho et al. 2015, Chapter 4-Part 2). All these different approaches allow for the sampling of genetic material in enough quantity and quality to perform genetic analyses. However, they may not allow to sample the same individuals and thus to detect similarly hybridization. The opportunistic sampling of carcasses of road-killed animals for the study of hybridization is particularly widespread in wildcat studies (Germain et al. 2009, Say et al. 2012, Randi et al. 2001) but also in other species (Godinho et al. 2011, Galov et al. 2015). In wildcats, road kills are a major cause of mortality (Klar et al. 2008) and thus provide biological material in large quantity without the setting of heavy sampling protocols (trapping, systematic collect of feces, etc). However, the individuals killed by traffic may not be representative of the whole population and thus, carcasses may lead to biased results for the characterization of hybridization (Nussberger et al. 2014b) compared to what would be assessed with an exhaustive trapping of live individuals.

First, carcasses may mainly belong to dispersers because they are more likely to cross roads to join their new territory. As hybrids may be perceived differentially by parental individuals (Lancaster et al. 2006), behaviors of exclusion may lead hybrids to disperse more than pure individuals. Then, the study of road-killed animals may lead to an over-estimation of the number of hybrids. Second, in cats, and in several other mammals (Pusey and Packer 1987,

Höner et al. 2007, Lawson Handley and Perrin 2007), dispersers are mostly males due to a sex-biased dispersal (Devillard et al. 2004, Hansen et al. 2007). Then, if hybridization affects differentially males and females, we may not be able to characterize all the hybrids and capture their diversity. Third, in some habitats, like in our area of study in the Pyrenees, wildcats are not impacted by traffic and no carcasses are found on the roads. Studies led solely on road-killed individuals may thus miss part of the local heterogeneity that exist in the rate of hybridization. Comparison between opportunistic and systematic sampling using hair traps locally was made by Nussberger et al. (2014b) and they did not find major differences between the results obtained with the two sampling schemes. A more systematic comparison of different sampling schemes may however be relevant in future studies given the number of possible biases attached to opportunistic sampling.

Opportunistic samplings all over the area of repartition of the European wildcat reported highly contrasted levels of hybridization (18% in Hertwig et al. 2009, 30% in Say et al. 2012, 50% in Pierpaoli et al. 2003). Even if the biases attached to opportunistic sampling prove in the future to have only a marginal impact on the rate of hybridization estimated, this sampling scheme remains poorly adapted to the study of the variability of hybridization patterns because of the biases listed earlier. In order to do so, local studies between interacting individuals are required.

6.2.2. LOCAL SAMPLING SCHEMES TO UNDERSTAND THE HETEROGENEITY OF HYBRIDIZATION PATTERNS

Local sampling studies aiming at understanding why hybridization occurs in cats, instead of assessing to what extent it occurs, are extremely rare. Local studies have been conducted in Swiss Jura (Nussberger et al. 2014b) or in north-eastern France, close to our area of study (Germain et al. 2008) but did not allow to link ecological features with the presence or absence of hybrids.

Nussberger et al. (2014b), reported a high resolution genetic study focused on the detection of hybrids. The panel of genetic markers used was however not relevant for the study of population structure, or parentage analysis, and was not combined with other techniques and notably spatial approaches that could have allowed to understand how individuals may meet. On the other hand, Germain et al. (2008) combined genetics and radio-tracking but used a number of microsatellite markers ($N = 13$) that did not allow to detect reliably hybrids passed the first generation (F1, O'Brien et al. 2009). Thus, given the rarity of F1 hybrids, and the absence of difference between wildcats and hybrids, we may question the reliability of the results reported in this study. Also, only 9 wildcats (5 wildcats and 4 putative hybrids) were monitored in a forest twice the size of the central forest in our area of study and thus that probably contains far more individuals. The results presented in Germain et al. (2008) may thus not be representative of wildcat behaviors.

To our knowledge, the local study conducted in north-eastern France is the first study including exclusively interacting individuals from both sub-species and combining an exhaustive sampling of the populations (42 wildcats and 511 domestic cats trapped in an area of about 250 km²), the use of 22 microsatellite markers allowing for the detection of most of the first generation backcrosses (67 to 98% depending on the method) and a spatial monitoring of more than half the individuals trapped alive (18 wildcats equipped over 32 wildcats trapped alive). The same sampling effort could not be deployed in the Pyrenees for practical reasons (no authorization for live trapping, lack of time during the PhD to set a rigorous protocol of feces sampling allowing to assess home-ranges from capture-recapture of feces). However, we did collect a substantial number of feces in proximity of the villages where domestic cats live and where hybridization front is expected to stand.

In the case of the hybridization between wildcats and domestic cats, the presence of agricultural fields seemed to be of major importance and to promote the encounter of sub-species due to the density of rodents they

contain (Chapitre 2). For other hybridization systems like the wolf and the dog, or like the wild boar and the domestic pig, the presence of rodents is not expected *a priori* to play a major part given that they are not majority in the diet of these species (Gade-Jørgensen and Stegegaard 2000, Herrero et al. 2006, Van Dijk et al. 2008). Other processes may thus be at the basis of hybridization. In wolves, the need for local studies has also been pointed out (Pacheco et al. 2017). Recently, a population scale study in the Iberian Peninsula mixing hybridization detection, spatial data and relatedness estimates, showed that, contrary to classic expectations, hybrids may result from multiple events of hybridization over a large spatial area and not events spatially restricted (Pacheco et al. 2017). This study also led to propose social group disruptions events to promote hybridization. This was also proposed for a wolf-Coyote (*Canis rufus* x *Canis latrans*) system in North California (Bohling and Waits 2015) that was studied at a population scale and for which ecological and demographic data (pedigree, breeder experience...) were linked to hybrid detection. The social group stability may then be an important factor shaping hybridization patterns in canids.

We probably do not have yet a full assessment of the potential of local sampling schemes to understand hybridization. More local studies are required in the future to fully grasp the diversity of hybridization patterns and of the mechanisms that shape them. Beyond a variability in the extent of hybridization and introgression, these local studies may allow to study in more details which genes introgress and if some introgress more than others. The introgression of different genes according to the populations may lead some populations to be more threatened by introgressive hybridization than others. For instance, if the fact that both the African wildcat and the domestic cat are susceptible to FIV (Troyer et al. 2008) results from a sharing of susceptibility genes, and if these genes introgress in some populations but not all, some may be more endangered than the others from an epidemiological point of view.

6.2.3. HOW MAY LOCAL SAMPLING HELP FOR HYBRID MANAGEMENT?

Local sampling studies provide insights into the causes of hybridization and thus important information to plan measures aiming at avoiding situations promoting it. In line with Chapter 4's results and the studies led by Pancheco et al. (2017) and Bohling and Waits (2015), measures may for example consist in green engineering to shape interfaces in order to make them less prone to promote encounters, or in a change of hunting practices in order to limit their impact of social structure equilibria. The multiplication of local studies may allow to refine these measures over time.

The studies about hybridization are classically led at a larger scale, over several non-interacting populations (Verardi et al. 2006, Pierpaoli et al. 2003, Vuillaume et al. 2015). This sampling scale allows to detect hybrids relative to a "genetic core" that corresponds to the genetic material shared by all individuals from the same sub-species whatever their local environment or history. The individuals identified as hybrids using this large sampling scheme may be individuals from populations that went through intense episodes of introgressive hybridization but may not be hybridizing anymore. Hybrids may thus not result from recent reproduction events between sub-species but be the remnant of a past intensive introgressive hybridization. Large sampling may thus not provide real-time information about hybridizing processes and lead to the elimination of individuals that are in fact not conveying nor genes nor pathogens at the time of the study. Alternatively, hybrids detected locally may be hybrids between hybrid forms and local populations may not be worth managing against hybridization. As a consequence, local scale may not allow to identify hybrids in a satisfactory way for conservationists. To what extent local sampling is relevant for the identification of hybrids and their management in the short-term remains an open question that may not find an answer before guidelines for the management of hybrids are clarified.

6.3. HOW MAY EPIDEMIOLOGY BE INCLUDED IN HYBRID MANAGEMENT GUIDELINES

6.3.1. THE PROBLEM WITH TRADITIONAL HYBRID MANAGEMENT

Recent studies have largely focused on anthropogenic hybridization which is considered as systematically deleterious for the persistence of species (Mallet 2005). In line with this view, policies for the conservation of species mainly preconize to stop hybridization either by stemming processes leading to the production of hybrids or by the elimination of hybrid individuals, considered as “exotic” individuals threatening native populations (Landelle 2005, Jackiw et al. 2015). The few legal texts dealing with hybrids mostly address hybrid killing measures, certainly because the absence of local studies has led to a lack of information regarding the ecological mechanisms at the basis of hybridization, and thus limit our capacity to prevent hybridization before it happens.

These policies are based on the premise that clear rules allow to designate which individuals are hybrids and to assess whether or not they represent a threat. However, this premise is far from being true and even the use of the term “hybridization” is controversial in some situations (Gompert and Buerkle 2016). Genetic studies have indeed shown the major impact of reticulated evolution and that many genomes contain to some extent genes originating from other species (Mallet et al. 2016, Medugorac et al. 2017, Sankararaman et al. 2014). The elimination of all individuals carrying traces of introgression would then be a non-sense and may even become damageable for the species. Alternatively, the determination of a threshold delimiting when individuals stop being hybrids is hardly achievable (Harrison 1993, Daniels and Corbett 2003), and thus it can be difficult to distinguish hybridization from conspecific

reproduction when populations are loosely differentiated. Typically, for the European wildcat and the domestic cat, their phylogenetic proximity (O'Brien et al. 2008) and past history of hybridization and introgression (Hertwig et al. 2009, Ottoni et al; 2017) may lead to wonder whether we should really consider the reproduction between these sub-species as a hybridizing process (Fredriksen et al. 2015). However, because domestic cats are a domesticated sub-species, their hybrids are more easily considered as a threat.

The genetic criterions being of limited relevance for the definition of which hybrids should be protected, eliminated or simply tolerated, these past years, it has been more and more replaced by functional criterions (Daniels and Corbett 2003, Jackiw et al. 2015, Galaverni et al. 2017). According to these new criterions, if hybrids and parental individuals are ecosystem surrogates, then hybrids may not need to be removed from the environment (Galaverni et al. 2017). While classical approaches have a static view of the conservation of species where species must remain as they are or even return to their state before human impacted the environment, these ecosystem-based guidelines take into consideration the necessary evolution of species (Fredriksen 2015) which can require hybridization, which proved to be at times beneficial even when anthropogenic (Anderson et al. 2009).

6.3.2. USING EPIDEMIOLOGY AS A CRITERION FOR THE MANAGEMENT OF HYBRIDS

In Chapter 5, we proposed behavioral features to explain, for a substantial part, the absence of infected wildcats in the population from North-Eastern France and proposed that hybrids, if they increased in number, may disrupt the current equilibrium and lead to non-null prevalence among wildcats. Then, the presence of hybrids may clearly become a threat for the conservation of the European wildcat. So, why not consider the role hybrids play in the circulation

of pathogens as an additional criterion to decide whether or not hybrids should be protected?

This criterion has never been proposed as far as I know and, more generally, such association of epidemiology and ecology in management guidelines is difficult to find in the literature and possibly inexistent in spite of the tight connections existing between ecology and epidemiology, and their respective importance for wildlife conservation. Recently, Wayne and Schaffer (2016) proposed a tree of decision that relies on three rules allowing to decide in a specific context whether or not hybrids should be protected (Figure 6.2). According to their first rule, if hybrids result from a natural hybridizing process, they must be protected, else, more criterions must be examined. Given that the impact of men on the environment is global, natural hybridization has become very difficult to separate from anthropogenic hybridization (Mallet 2005, Malukiewicz et al. 2015). In practice, the only hybridizing events that can be attributed to the action of humans without questions are those resulting from the release of hybrids in nature, for instance for restock purposes in Galliforms (Sokos et al. 2008). For all other situations, natural and anthropogenic mechanisms are most likely combined in the hybridizing process. In most cases then, this first rule will not allow to take a decision. The second rule consists in defining if hybrids are the ecological surrogates of the parental individuals. In many cases, especially for backcrosses of advanced generation, we can expect this to be true but the extent to which two organisms are surrogate is, once again, a complicated matter (Wayne and Schaffer 2016). A third rule is then proposed by Wayne and Schaffer: if the restoration of the natural habitat leads to the selection of native genes, then hybrids may be protected, at least temporarily, until the restoration succeeds in stopping hybridization. I propose to add an epidemiological criterion before this third rule (Figure 6.2) that consists in assessing whether hybrids may represent a threat for the parental populations notably by promoting the emergence of new pathogens in the

endangered population. Like the first two rules of Wayne and Schaffer's decision tree, the epidemiological threat represented by hybrids is quantifiable based on present information. In addition, the effects of the restoration of the habitat may be in the long term while emerging viruses can lead to rapid decline of endangered populations (Roelke-Parker et al. 1996, Leroy et al. 2004). If hybrids, due to their intermediary genome, do promote pathogen jumps and the emergence of new diseases (Parrish and Kawaoka 2005), which remains poorly investigated as detailed hereafter, it makes then more sense to act regarding epidemiology before acting regarding habitat restoration. Habitat restoration may however be considered to prevent future hybridization and infectious diseases outbreaks even if hybrids represent an epidemiological threat and must be eliminated in the short term (Figure 5.2).

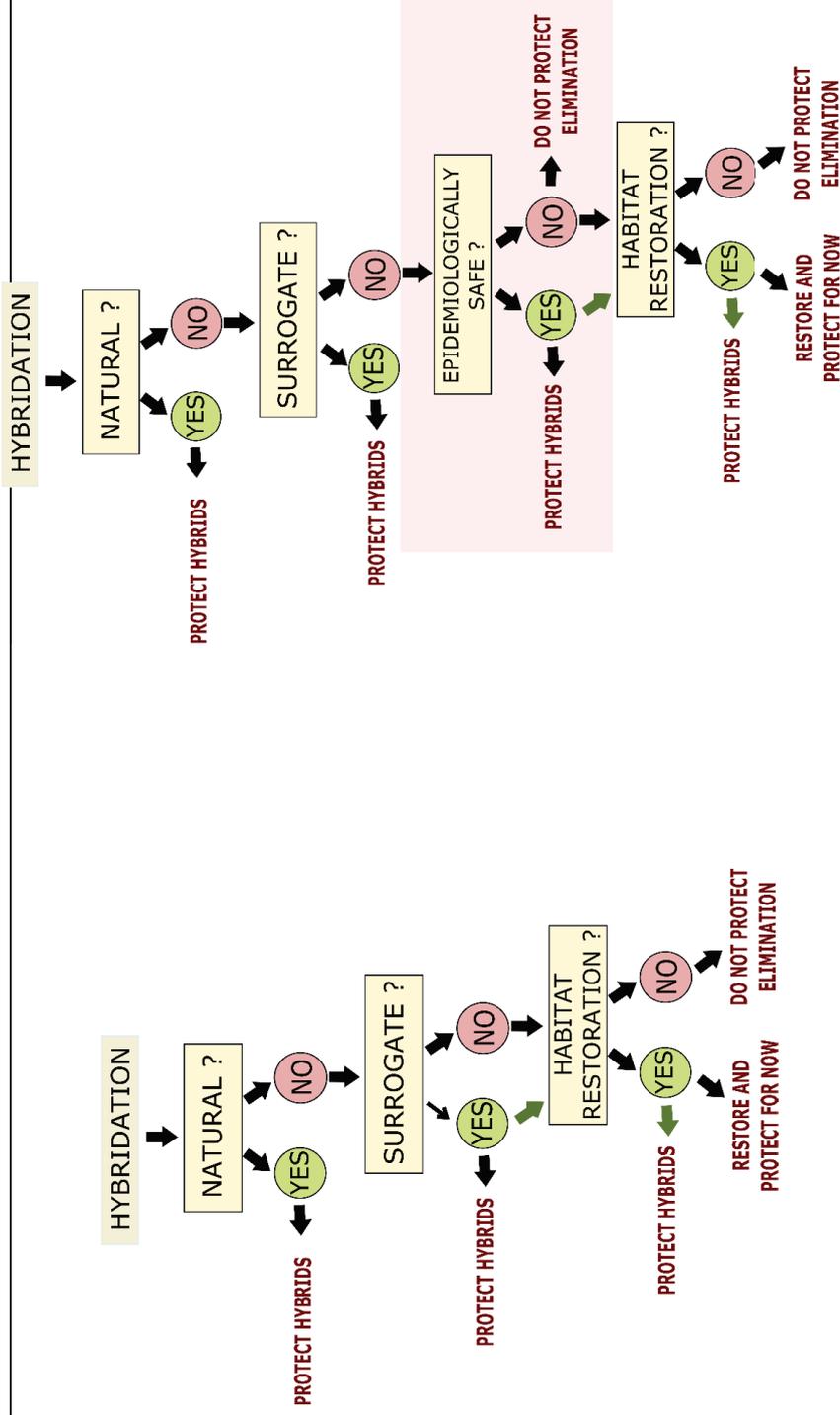


Figure 6. 1: Graphic interpretation of the decision tree of Wayne and Schaffer (on the left) and decision tree implemented with an epidemiological criterion proposed (on the right). The additional step is highlighted with a pink frame.

6.3.3. ASSESSING IF HYBRIDS ARE AN EPIDEMIOLOGICAL THREAT

The hybrid may represent a threat for the conservation of parental species if they promote the transfer of parasites from a parental population to another and thus act as a bridge (Floate and Whitham 1993, Lebrun et al. 1992). Le Brun et al. (1992) reported indeed a correlation between hybrid status (hybrid class) and level of prevalence to a ectoparasite (*Diplozoon gracile*) in Barbels (*Barbus spp.*) and proposed based on these observations that hybrids may be eco-ethological bridges allowing the transfer of parasites from one species to another. The “Hybrid-bridge hypothesis” was clearly stated only one year later in plant systems (Floate and Whitham 1993). Since, the hybrid bridge hypothesis has remained mainly studied in plants. For instance, a recent study in Oaks (*Quercus spp.*, Tovar-Sanchez and Oyama 2006) showed that oak hybrids were susceptible to more parasites and may, because of a complete morphological and genetic continuum, favor shifts from one plant to another. The hybrid bridge remains however poorly investigated, especially in animals and regarding the evolution of infectious agents. More investigations are thus required to assess how important this process may be for the spread of infectious diseases.

Pending additional information, the assessment of the epidemiological threat represented by hybrids may be achieved using two criteria usually mentioned in studies associated with the hybrid bridge hypothesis: the relative diversity of parasites in hybrids and parental forms, and the continuum of prevalence along hybrid classes. If hybrids are susceptible to more parasites, measures against them or the absence of measures protecting them should be considered. If the diversity is equivalent and the prevalence in hybrids is similar to the prevalence in parental forms or if the diversity in hybrids is lower than the diversity in parental individuals, then hybrids may be temporarily protected and the third rule of Wayne and Schaffer (2016) implemented. In parallel, the prevalence of the considered

parasites should be expressed as a function of the hybrid class in order to assess if there is a clear relationship between these two variables as expected if hybrids act as a bridge (Lebrun et al. 1992). This particular measure raises again the issue of the detection of old generations of backcrosses as this relationship would be better estimated with a large number of hybrid classes. However, in the first instance, hybrids' protection measures may be avoided when a relation appears considering only the F1, the F2 and the two first generations of backcrosses which are accessible with current panels of markers (Nussberger et al. 2013, 2014).

6.4. IS THE EUROPEAN WILDCAT THREATENED BY DOMESTIC CATS?

The European wildcat proved to be a useful model for the study of hybridization and regarding the circulation of viruses. Beyond the methodological developments its study allowed, we could gather information which lead us to wonder whether or not the domestic cat threatens locally the European wildcats, if measures should be taken in order to limit hybridization and, necessary, what kind of measures.

6.4.1. DO WE FULLY UNDERSTAND THE ROLE OF FRAGMENTATION ON HYBRIDIZATION?

Habitat fragmentation is expected to be a threat for the conservation of the European wildcat (Crispo et al. 2011). Some of our results comfort such deleterious effect. The comparison between north-eastern France and Pyrenean populations led us to conclude that the presence of agricultural fields rich in food resources around forests, in fragmented agricultural landscapes, promoted encounters between sub-species and hybridization (Chapitre 4). However, our results also provided insights that may moderate the negative impact of fragmentation on the

persistence of the European wildcat. First, the spatial segregation of males and females (Chapter 2, Part 1) in the fragmented environment seemed to promote hybridization between wild males and domestic females (Chapter 4, Part 1) and introgression of wildcat genes into domestic genomes (Chapter 3, Part 1). The use of mtDNA markers and/or Y-linked markers is however necessary to confirm this pattern. If confirmed, this means that this spatial segregation may actually play in favor of the wildcat genetic integrity. Second, the genetic diversity in the Pyrenean wildcat population was significantly lower than the genetic diversity in north-eastern France. While this difference may be due to the isolation of the wildcat population in this mountainous environment (Say et al. 2012, Chapter 4, Part 2), we may wonder whether the absence of hybridization is partly responsible for this difference. Genetic rescue, i.e. the increase of population fitness owing to the immigration of new alleles (Whiteley et al. 2015), is one of the possible beneficial impacts of hybridization (Todesco et al. 2016). This mechanism may have played a major part in the evolution of other felids from the genus *Panthera*. Figueiró et al. (2017) conducted indeed a genome-wide study in this genus and identified post-speciation admixture between species of this genus has probably been a recurrent route of evolutionary rescue when populations reached low effective sizes and were thus at risk of extinction. So, we may wonder to what extent hybridization participates in the persistence of a viable genetic diversity in European wildcats depending on the environment and isolation of the population.

Habitat fragmentation may also have led to the formation of rudimentary social structures in the wildcat population from north-eastern France (Chapter 2). This is the first time that a non-solitary way of life is reported for European wildcats. This shows that this sub-species of cats is more flexible in terms of social organization than previously thought (Sunquist and Sunquist 2002). Given that the social structure of a population impacts the dynamics of infectious diseases, this observation is crucial for future modelling approaches of wildcats' infectious diseases spread (Pontier et al. 2009).

The effect of fragmentation thus appears more complex than usually thought. Its effect may probably need to be decomposed into several aspects, like the nature of the “anthropic” habitat (agricultural fields, roads, grassland, villages...), in order to investigate with more details how it impacts hybridization between wildcats and domestic cats, but more largely, hybridization in wildlife. This better understanding of fragmentation and human impact on hybridization could notably be achieved thanks to next-generation sequencing. Using whole-genome sequencing (WGS), Medugorac et al. (2017) identified episodes of intense introgression between yaks and cattle that coincided with the Medieval Climate Anomaly and the Dzungar-Qing War, two periods during which high mortality rates in cattle led to hybridization events to restore herds. A similar study led on wildcats and domestic cats may allow to understand the introgressive history between the two sub-species and to relate it to archives about the impact of men on the environment in order to better understand the part human played in past introgression between cat sub-species.

6.4.2. DO PYRENEAN WILDCATS NEED PARTICULAR PROTECTION MEASURES?

The low genetic diversity of the European wildcat population studied in the Pyrenees (Chapter 4, Part2) is worrisome and may be a problem for the persistence of the sub-species in this area in the future (Spielman et al. 2004). More investigations are required to assess whether Pyrenean wildcats are threatened and if specific conservation measures should be considered in this area. In particular, more investigations should focus on the characterization of the factors that contribute to the isolation of the population. Controlled genetic rescues are increasingly considered to prevent extinction of populations marked by low genetic diversity (Tallmon et al. 2004, Hedrick and Fredrickson 2010). Could such genetic rescue be applied on Pyrenean wildcats if their decline is confirmed? If so, should

neutering policies of domestic cats be reassessed in the Pyrenees in order to allow hybridization between the two sub-species to some extent?

Furthermore, the only individual we may suspect to be a hybrid lived in the village of Nohèdes. The presence of hybrids in the domestic cat population may actually be expected. According to Wirtz (1999), hybridization is expected to occur mainly between the female of the population less represented and the males of the most common population, in response of a lack of conspecific males. This direction of hybridization was for instance observed for 90% of the hybridizing events between coyotes (*Canis latrans*) and red wolves (*Canis rufus*) in North Carolina (Bohling and Waits 2015). Generally, the less represented population is the endangered population. In Nohèdes however, the situation may be reversed, the village being surrounded by mountains where wildcats live (Chapter 4, Part 2). Hybridization may then be expected to occur between domestic females and wild males under Wirtz' hypothesis. Because females take care alone of the cubs, hybrids would then preferentially live near the village. In Nohèdes, most but not all domestic cats had been neutered by local associations. This partial neutering must have led to a diminution of the domestic cats available for mating. Could the scarcity of available mates among domestic cats lead them to accept wildcat males as sexual partners and hybridize? Is partial sterilization policy better than no sterilization in the Pyrenees?

These questions are for now impossible to answer without more investigations but they open interesting perspectives for the future.

6.4.3. IS THE EUROPEAN WILDCAT THREATENED IN NORTH-EASTERN FRANCE?

As it is, I think that it is very complicated to conclude and to decide whether or not hybridization is a problem for the conservation of European wildcats in northeastern France. Past studies implied that hybrids of recent generations had an intermediary way of life reflected by an intermediary diet (Biró et al. 2005, Germain et al. 2009) or spatial behavior (Kilshaw et al. 2015). Hybrids may thus not be ecological surrogates of the European wildcats. However, the identification of hybrids was not achieved genetically (Kilshaw et al. 2015) or led on 12 to 13 markers that did not allow to retrieve reliably first generation backcrosses (O'Brien et al. 2009). Thus, these results may need to be interpreted with caution. Our results (Chapter 5) may not comfort these results. In our study, the 16 putative hybrids were indeed equally distributed between owned ($N = 9$) and unowned ($N = 7$) cats and some ($N=6$) were even tame enough to be sampled without trapping. These sixteen hybrids being detected with a relaxed approach (Chapitre 3, Part 1), we may suspect some of them to be in fact domestic cats. However, even when we considered the two hybrid detected with a conservative approach, both ways of life (owned and unowned) were represented. Regarding hybrids detected in the wildcat populations, all but one were captured outside the forest, which may be due to an intermediary behavior; however, we cannot discard the hypothesis that they are in fact wildcats because these hybrids were detected with the relaxed approach. There are thus no categorical results showing an intermediary way of life of hybrids neither in the literature nor in our results. Furthermore, in spite of generations living in sympatry and probable on-going introgression (Hertwig et al. 2009), the two sub-species remain clearly distinguishable both on an ecological, morphological and genetic point of view. Barriers preventing too high rates of hybridization are thus expected (French and Easterbee 1988, Hubbard et al. 1992) and may still hold.

Regarding the circulation of viruses, the current rate of hybridization in north-eastern France does not appear to promote the transfer of domestic cat viruses to the wildcats. However, some viruses (FIV specifically) may not be shared by domestic cats and wildcats. Thus, hybrids may represent a potential threat. Only two putative wild backcrosses (F1 x Wildcat) were tested during the serological survey and none was found positive to any of the four viruses. While these two individuals are clearly not statistically relevant to draw any conclusion, they question the existence of a continuity of infection across hybrid classes and of a hybrid bridge.

In conclusion, measures against hybrids appear currently difficult to justify as they do not represent a categorical threat. However, given that the proliferation of domestic cats is a general issue, measures aiming at controlling their demography (neutering) should be applied regardless of the actual threat raised by hybrids. The impact of sterilization policies on hybridization should however be tested in order to better predict the effects of such policies. Furthermore, vaccination policies may be conducted in order to decrease the epidemiological pressure domestic cats apply on the wildcat population and thus prevent future epidemiological threats due to the presence of hybrids. Given that hybrids are part of the domestic cat population, such policies would also disrupt any existing hybrid bridge.

6.4.4. FUTURE STUDIES REQUIRED TO BETTER UNDERSTAND THE SYSTEM

The European wildcat remains mysterious in many aspects and a lot remains to be investigated notably to understand its interactions with domestic cats. In Chapter 4 (Part 1), we based our results on radio-tracking monitoring. This monitoring allowed us to have an approximation of the interactions we

may expect between sub-species. More precision is required to better assess to what extent wildcats meet domestic cats. More extensive radio-tracking (Martin et al. 2013) or the use of contact collars (Craft and Caillaud 2011) may allow to build a contact network useful for hybridization but also in epidemiological studies. Notably, it may allow to build individual-centered models where individuals could be characterized by their connections with their neighbors. Such network models have been used in lions to understand how different patterns of dispersal can impact the spread of a virus (Craft et al. 2011, 2015) and may bring a new level of precision in the wildcat-domestic cat system. Such studies remain for now rare due to the difficulty to assess contact networks (Craft et al. 2015). The use of a more intensive radio-tracking would also be interesting to know if there is a spatial temporal sharing between domestic and wildcats, but also between wildcats to precise the structure of the related units we observed (Chapter 4, Part 1). Finally, investigations at the population scale should be associated in the future with investigations at the individual scale in order to give insights into behavioral and ecological features of European wildcats. For instance, hybridization being based on the reproduction of individuals, it would be crucial to understand wildcats can discriminate hybrids and reject them from their habitat or as mates during reproduction.

Conclusion

The European wildcat as organism model did not allow me to establish a unique routine applicable to any species and any question for the simple reason that such unique routine is impossible to establish. The study of hybridization encloses indeed many questions which make impossible the use of a single analytical path. Instead, I could draw crucial questions to ask and some guidelines to follow from the sampling schemes to the statistical approach when confronted to the study of hybridization.

Particularly, one must be aware of the major importance of the sampling scheme because it defines the questions to which we will be able to provide answers. This is of course true for any experimental design, but given the multiplicity of questions raised by hybridization, questions which may seem similar, I think it is important to press the importance of spending time on the sampling design.

My PhD work allowed me to answer some questions and to identify even more issues. Then, more investigation is still to come to understand hybridization and virus circulation between European wildcats and domestic cats and more largely, across species.

Epilogue

During the three years that lasted my PhD, I experienced both life in the academic world and in the private sector. These two environments proved extremely different and both shaped my way of thinking, making me some kind of “cultural hybrid”.

Working in an academic laboratory taught me how to study hybridization with logic and neutrality, knowing that results are true up to a certain probability. Antagene showed me under which circumstances hybridization is studied and monitored: for captive-bred individual releases, for conservation of endangered species, for legal expertises, etc, and that in these applications, the study of hybridization is not neutral but based on the premise that hybrids are deleterious for the species, a view largely shared by the society.

Knowing the paradoxical place of hybridization in the evolution of genomes, it has been really difficult for me to understand why there was this gap, why hybrids were so despised by the society. Then, I wondered whether or not people knew about this paradox and I realized that, during my years at school, hybridization was defined as the mixing between individuals from different species, that evolutionary ecology was not taught and that, if I had not pursued my studies in ecology, I would have never known that hybridization could be beneficial for species. Then, I realized that our educational system along with the way academic laboratories communicate with the society, does not allow people to truly understand the ecological challenges of our time nor the measures undertaken to face them.

At a time when ecological measures are taken in order to stem our impact on the planet, I think that a better exchange between academic laboratories and the society is required to help people understand why some measures are taken but

also to calm the debates associated with them and make them more objective and constructive.

During my engineering school, I specialized in the management of the environment. Now I realize how important it is to understand both academic results, methods and approaches as well as applied concerns and societal context, to manage efficiently.

Hybridization can be an adaptive motor in some cases. I think this is true for cultural hybridization and more particularly here, I believe that the introgression of some of the academic knowledge to the society along with an introgression of a better appreciation of the societal context in the academic labs, will benefit the future management of wildlife, and more widely the management of our planet.

Résumé de la thèse en français

La faune sauvage a toujours suscité l'intérêt des hommes et plus particulièrement depuis l'éveil de la pensée écologique. au milieu du XXe siècle. Aujourd'hui menacée par les changements globaux et par les activités de l'Homme, elle est au cœur de nombreuses études scientifiques. Deux problématiques apparaissent aujourd'hui majeures concernant la conservation des espèces sauvages : l'hybridation, qui correspond à la reproduction d'individus appartenant à des populations génétiquement différenciables, et la circulation des maladies infectieuses qui sont une importante cause de mortalité à travers le monde tant pour les espèces sauvages que pour l'Homme.

L'hybridation est reconnue comme un processus évolutif majeur chez les animaux seulement depuis quelques décennies. Elle aussi bien participer à la survie des espèces qu'à leur extinction. L'essor des outils génétiques depuis les années 1980 a permis de mieux détecter ce processus et a conduit à l'identification d'individus hybrides dans de nombreux taxa. La détection de l'hybridation est donc devenue un enjeu majeur pour les gestionnaires de la faune sauvage et la transposition des outils développés dans le milieu académique est devenue essentielle pour organiser un suivi de l'hybridation dans les meilleures conditions. Ma thèse, financée par une Convention Industrielle de Formation par la REcherche (CIFRE) a consisté à permettre ce transfert méthodologique du monde de la recherche académique vers le laboratoire privé Antagene qui propose des prestations (génotypage, développement de marqueurs, diagnostic génétique) dans le secteur de la génomique animale.

Dans ce but, je me suis focalisée sur un organisme en particulier, le Chat sauvage Européen *Felis silvestris silvestris*, qui est une espèce menacée en Europe, et qui est touché par ces phénomènes d'hybridation. En effet, son homologue domestique, *Felis silvestris catus*, est une sous-espèce de Chat suffisamment proche génétiquement du Chat sauvage pour que les deux sous-espèces puissent se reproduire. Les hybrides découlant de cette union, constituent potentiellement une menace pour l'intégrité génétique du Chat sauvage car, en cas de croisement de ces hybrides avec des individus « purs », ces hybrides pourraient faciliter l'intégration

de gènes domestiques, qui sont potentiellement délétères dans un habitat sauvage, dans le génome des chats sauvages et les mener à l'extinction. Ces hybrides pourraient également constituer une menace d'un point de vue épidémiologique. Les chats domestiques sont en effet porteurs de nombreux virus dont certains ont déjà été impliqués dans des catastrophes démographiques chez d'autres carnivores sauvages. Les chats sauvages, quant à eux, vivent en petites populations fragmentées et ne peuvent probablement pas maintenir ces virus. Ils ne sont donc potentiellement pas armés pour leur faire face, faute d'une coévolution entre ces virus et les défenses immunitaires des chats sauvages. L'introduction de virus des chats domestiques dans les populations de chats sauvages pourrait donc s'avérer dangereuse pour la conservation du Chat sauvage. Les hybrides pourraient aider au passage de virus d'une sous-espèce à l'autre en faisant office de navette du fait de leur comportement intermédiaire, ou bien en procurant aux virus un milieu favorisant leur adaptation à un nouvel hôte. Au cours de ma thèse, et afin de pleinement appréhender quelle menace peuvent représenter les hybrides pour les chats sauvages, j'ai donc travaillé à comprendre à la fois la menace génétique et la menace épidémiologique que représentent les hybrides. Deux populations locales de chats sauvages nous ont servi à mener à bien ce travail : une population dans le Nord-Est de la France, et une population dans les Pyrénées Orientales.

L'impact des hybrides sur la conservation des chats sauvages ne peut être traité que si l'on est en mesure de détecter les hybrides. L'augmentation du nombre de cas d'hybridation rapportés dans la littérature a naturellement conduit au développement de nombreuses approches méthodologiques assurant cette détection. Plusieurs aspects méthodologiques restent toutefois à améliorer. Au cours de ma thèse, je me suis concentrée sur deux aspects. Le logiciel STRUCTURE est le plus utilisé dans les études portant sur l'hybridation. Il permet de déterminer dans quelles proportions un individu appartient à l'une ou l'autre des populations. En l'absence d'individus dont on sait s'ils sont « purs » ou bien hybrides, il est coutumier de déterminer des seuils sur ces proportions pour établir ce qu'est un hybride et ce qu'est un parent à partir de simulations faites sur la base d'un

échantillon du jeu de données censé représentatif des individus « purs ». La sélection de ces individus représentatifs est un procédé souvent obscur et non justifié dans les études. Le but de notre première approche méthodologique a donc été de répondre à ce manque de rationalité pour la définition du groupe d'individus représentatifs. Pour cela, nous avons mis en place un algorithme itératif permettant de ne partir que d'un ou deux individus, et de construire le groupe représentatif uniquement sur la base de l'information contenue dans les génotypes et non en fonction de critères arbitraires. Nous avons appliqué cette approche aux deux populations de chats sauvages étudiées. Dans le Nord-Est de la France, nous avons détecté six individus potentiellement hybrides dans la population de chats sauvages et seize autres dans les villages où vivent les chats domestiques. Dans cette population, un flux de gènes des chats sauvages vers les chats domestiques a été détecté. Au contraire, dans la population pyrénéenne, un seul chat est apparu potentiellement hybride, au cœur d'un village particulièrement isolé dans la montagne. Aucun hybride n'a en revanche été détecté dans l'habitat des chats sauvages. Cet individu potentiellement hybride ressort suspect avec la deuxième approche méthodologique, appelée *snappclust*, sur laquelle j'ai travaillé au cours de ma thèse. Il s'agit d'une méthode rapide de détection de groupes d'individus qui permet de catégoriser les individus en tant que pur ou bien en tant qu'hybride de classe donnée (première génération, rétrocroisement). Cette méthode se base sur une approche d'espérance-maximisation qui est une méthode permettant de rapidement converger vers un état final et qui, de par sa vitesse, se distingue d'autres méthodes telles que les méthodes bayésiennes implémentées dans STRUCTURE ou NEWHYBRIDS par exemple. Dans cette approche, les individus sont répartis dans différents groupes aléatoirement ou bien selon une méthode de regroupement géométrique (distance de Ward, K-means). A partir de ce premier regroupement, les fréquences alléliques pour chaque locus dans chaque groupe sont calculées. A partir de ces fréquences alléliques, on peut calculer la vraisemblance qu'un individu appartienne à un groupe plutôt qu'à un autre. Les individus sont alors redistribués vers le groupe auquel ils appartiennent le plus

vraisemblablement. De nouvelles fréquences alléliques sont alors calculées, et ainsi de suite jusqu'à ce que la vraisemblance globale du système se stabilise. Des probabilités d'appartenance aux différents groupes inclus dans l'analyse (par exemple on peut définir la classe des chats sauvage, la classe des chats domestiques et la classe des F1 nés de la reproduction entre un chat sauvage et un chat domestique) sont alors calculées à partir des vraisemblances normalisées individuelles d'appartenir à chacune de ces classes étant donné les fréquences alléliques associées à chacune de ces classes. Cette méthode se base sur les hypothèses classiques en génétique des populations que sont l'équilibre de Hardy-Weinberg et l'absence de déséquilibre de liaison. Une étude par simulation ainsi que l'étude d'un jeu de données réel nous ont permis de comparer les résultats obtenus avec cette méthode par rapport aux résultats obtenus avec des méthodes déjà décrites. En ce qui concerne la catégorisation des individus dans les différentes classes hybrides, notre méthode donne des résultats sensiblement différents de ceux que l'on peut obtenir avec NEWHYBRIDS, logiciel permettant également une catégorisation en classes d'hybrides mais beaucoup plus gourmande en temps de calcul. Tandis que notre méthode tend à catégoriser les individus à la limite de définition entre les parents et les hybrides dans des classes hybrides, NEWHYBRIDS tend à les catégoriser en tant qu'individus purs. Dans les deux cas, un biais existe donc pour les classes hybrides les plus difficiles à étudier. Selon la question posée, ces biais peuvent être utilisés à l'avantage de l'utilisateur. *Snapclust* représente donc une méthode intéressante pour la détection d'hybrides, en particulier pour une utilisation dans des structures ne disposant pas de logistique poussée pour effectuer des calculs sur ordinateur.

Dans l'optique de comprendre les mécanismes sous-jacents à l'existence ou l'absence d'individus hybrides dans les populations de chats sauvages du nord-est de la France et des Pyrénées Orientales, nous avons combiné ces approches de détection de l'hybridation avec d'autres approches génétiques ou bien faisant appel à d'autres techniques de suivi de la faune sauvage. Dans le nord-est de la France,

dix-huit chats sauvages, principalement des femelles, ont été équipés de collier de radiopistage et suivis sur une période allant de six mois à trois ans à hauteur d'une localisation par semaine en moyenne. Cette étude nous a permis de suivre les mouvements des femelles et de caractériser leur territoire. Il s'est avéré que les femelles étaient présentes au cœur de la forêt avec un accès quasi-systématique aux champs agricoles qui entourent le massif forestier où vivent les chats sauvages. Les mâles quant à eux, semblaient présents plutôt en périphérie du massif. Une telle configuration pourrait favoriser une hybridation asymétrique entre les mâles sauvages et les femelles domestiques. Une telle asymétrie serait cohérente avec le flux de gènes des chats sauvages vers les chats domestiques mentionné plus tôt. Une étude de parenté a également été conduite sur cette population et mise en relation avec la localisation des territoires des individus. Des groupes de femelles apparentées vivant très proches les unes des autres ont ainsi été observés. De tels groupes ont été décrits chez le chat domestique dans des milieux riches en ressources alimentaires. Le regroupement de femelles apparentées chez les chats sauvages pourrait correspondre à un mécanisme similaire et résulter de la présence en abondance de rongeurs à chasser dans les champs agricoles à proximité.

En l'absence d'autorisation de capture des chats sauvages dans les Pyrénées, une étude de radiopistage n'a pas pu être mise en place à l'instar de ce qui a été fait en Lorraine. En revanche, à partir de fèces et de poils, des génotypes permettant de détecter des hybrides mais aussi d'établir des relations de parenté ont été dressés. L'analyse des relations de parenté en lien avec les localisations d'échantillonnage, a permis de mettre en lumière un patron différent de celui observé en Lorraine. Dans les Pyrénées, les individus des deux sexes semblent disperser dans la mesure où l'on n'a pas échantillonné des individus apparentés dans des zones géographiquement proches. L'absence de groupes de femelles apparentées dans le milieu pyrénéen où on ne retrouve pas de champs agricoles riches en proies conforte notre hypothèse d'un effet majeur de la présence de champs dans l'organisation sociale des chats sauvages.

Ces champs pourraient également être la raison pour laquelle les individus s'hybrident davantage en Lorraine que dans les Pyrénées. En effet, les proies dont ils regorgent sont aussi bien prisées des chats sauvages que des chats domestiques. Ces zones d'abondance constituent donc des zones potentielles de rencontre entre les deux sous-espèces et donc des zones facilitant les processus d'hybridation.

Les rencontres permises par ces champs peuvent également permettre l'échange de virus entre chats domestiques et chats sauvages. Afin de vérifier si des virus communs du chat domestique présents à haute prévalence dans les populations voisines de la population de chats sauvages, étaient effectivement échangés entre les deux sous-espèces de chats, nous avons conduit une étude sérologique. Que ce soit pour le parvovirus félin (PVF), le calicivirus félin (CVF), l'herpesvirus félin (HVF) ou le virus de l'immunodéficience féline (VIF), aucun chat sauvage n'a été détecté comme étant infecté. Cette prévalence nulle est surprenante étant donné la forte pression épidémiologique potentiellement exercée par les populations domestiques sur la population de chats sauvages. Par ailleurs, ces virus ont été rapportés dans des populations de chats sauvages dans des études précédentes. Néanmoins, ces études ne se basaient pas sur une identification génétique des individus. Ainsi, étant donné la difficulté avec laquelle les hybrides et les chats sauvages peuvent être différenciés seulement sur la base de critères morphologiques, il est possible que les individus infectés dans ces études ne soient pas des chats sauvages. L'absence d'individu infecté peut être expliquée de différentes manières. Tout d'abord, les chats sauvages auraient pu être infectés par une souche particulière non comprise dans le spectre de souches détectables par le protocole de laboratoire mis en place. Nous pouvons rejeter cette hypothèse car des précautions ont été prises pour détecter toutes les souches correspondant à chaque famille de virus. Deuxièmement, les souches présentes chez les chats domestiques peuvent être spécifiques du chat domestique et, à cause d'une barrière génétique et/ou physiologique, ne pas pouvoir être transmises au Chat sauvage. Troisièmement, la souche peut être capable d'infecter les chats sauvages mais une absence de contact entre les chats domestiques et les chats sauvages,

conjugée à la petite taille de la population sauvage, empêche son maintien dans la population sauvage. Nous avons construit un modèle statistico-dynamique permettant de tester pour quelles valeurs de paramètres (différentiel de contacts efficaces et perte de propagation dans la population sauvage par rapport à la population domestique), nous pouvions statistiquement espérer observer une prévalence nulle. Nos résultats ont montré que pour l'ensemble des virus, une diminution de la transmission entre espèces (par rapport à de la transmission entre chats domestiques) et/ou entre chats sauvages, est associée à l'absence de chat sauvage infecté. Pour le VIF, l'existence d'une barrière génétique est fort probable dans la mesure où ce virus n'a jamais été détecté dans les populations de chats sauvages. Par ailleurs, la configuration spatiale des chats sauvages qui favorise la reproduction de mâles sauvages avec les femelles domestiques pourrait ajouter une barrière comportementale. Le VIF est en effet principalement transmis lors de la reproduction ou lors de conflits pour l'accès aux femelles. Les femelles étant moins porteuses du virus et les saisons de reproduction étant décalées, la transmission du virus pourrait se faire à moindre fréquence entre sous-espèces même en l'absence d'une barrière génétique. Pour les trois virus restants, une barrière comportementale est favorisée. Elle pourrait être due à la moindre fréquence des contacts sociaux entre chats sauvages étant donné la territorialité de cette sous-espèce (CVF, HVF) ou bien à la différence d'habitat et d'emplacement des territoires (PVF). Dans tous les cas, on constate que des barrières protègent les chats sauvages des virus domestiques. Ces barrières pourraient être mises en péril par une augmentation du nombre d'hybrides car, de par leur comportement ou leur génome intermédiaire, ils pourraient les affaiblir,

Enfin, au cours de ma thèse, j'ai pu développer des méthodes permettant une détection efficace des hybrides chez le Chat mais aussi chez d'autres espèces. A travers le Chat et ces autres espèces, j'ai pu me poser des questions sur la pertinence de choisir tel ou tel type de marqueur génétique et sur l'impact du protocole d'échantillonnage sur les résultats et surtout sur la

comparabilité d'études menées indépendamment avec des protocoles différents. Globalement, étant donné la complexité des problématiques en lien avec l'hybridation, il est apparu essentiel de bien choisir ses outils en fonction de sa question et de bien délimiter ce que ce choix d'outil permet ou non de conclure. Cette remarque est vraie pour n'importe quelle étude scientifique et est apparue particulièrement cruciale pour l'étude de l'hybridation étant donné la multitude d'approches disponibles dans la littérature.

Concernant le Chat sauvage, il ne semble pour l'heure que peu menacé par l'hybridation à la fois dans le nord-est de la France et dans les Pyrénées. Toutefois, dans le nord-est de la France, des mesures préventives empêchant l'augmentation en fréquence du nombre d'hybrides devraient être envisagées, principalement pour empêcher l'effondrement des barrières épidémiologiques qui semblent en place. Se positionner quant à la menace génétique que représente l'hybridation est plus complexe de par le rôle ambigu de l'hybridation sur l'évolution des espèces. , L'utilisation de nouvelles techniques de séquençage pourrait permettre de statuer sur la menace génétique que représente l'hybridation entre chats domestiques et chats sauvages pour la conservation de ce dernier.

References

References

References

- Abbott R, Albach D, Ansell S, et al (2013) Hybridization and speciation. *Journal of Evolutionary Biology* 26:229–246. doi: 10.1111/j.1420-9101.2012.02599.x
- Akaike H (1988) Information theory and an extension of the maximum likelihood principle. *Selected papers of Hirotugu Akaike* (eds E. Parzen, K. Tanabe & G. Kitagawa) pp199-213. Springer Series in Statistics. Springer New-York.
- Akogul S, Erisoglu M (2016) A Comparison of information criteria in clustering based on mixture of multivariate normal distributions. *Mathematical and Computational Applications* 21:34. doi: 10.3390/mca21030034
- Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* 19:1655–1664. doi: 10.1101/gr.094052.109
- Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001) The problems with hybrids: Setting conservation guidelines. *Trends in Ecology and Evolution* 16:613–622. doi: 10.1016/S0169-5347(01)02290-X
- Allison AB, Harbison CE, Pagan I, et al (2012) Role of multiple hosts in the cross-species transmission and emergence of a pandemic parvovirus. *Journal of Virology* 86:865–872. doi: 10.1128/JVI.06187-11
- Altizer S, Harvell D, Friedle E (2003) Rapid evolutionary dynamics and disease threats to biodiversity. *Trends in Ecology and Evolution* 18:589–596. doi: 10.1016/j.tree.2003.08.013
- Altizer S, Nunn CL, Lindenfors P (2007) Do threatened hosts have fewer parasites? A comparative study in primates. *Journal of Animal Ecology* 76:304–314. doi: 10.1111/j.1365-2656.2007.01214.x
- Alves PC, Ferrand N, Suchentrunk F, Harris DJ (2003) Ancient introgression of *Lepus timidus* mtDNA into *L. granatensis* and *L. europaeus* in the Iberian Peninsula. *Molecular Phylogenetics and Evolution* 27:70–80. doi: 10.1016/S1055-7903(02)00417-7
- Alves PC, Melo-Ferreira J, Freitas H, Boursot P (2008) The ubiquitous mountain hare mitochondria: multiple introgressive hybridization in hares, genus *Lepus*. *Philosophical Transactions of the Royal Society B: Biological Sciences* 363:2831–2839. doi: 10.1098/rstb.2008.0053
- Anderson EC, Thompson EA (2002) A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* 160:1217–29.
- Anderson RM, May RM (1985) Vaccination and herd immunity to infectious diseases. *Nature* 318:323–329. doi: 10.1038/318323a0

References

- Anderson TM, Bridgett M, Candille SI, et al (2009) Molecular and evolutionary history of melanism in North American gray wolves. *Science* 323:1339–1343. doi: 10.1126/science.1165448.Molecular
- Anderson EC, Barry PD (2015) Interpreting the FLOCK algorithm from a statistical perspective. *Molecular Ecology Resources* 15: 1020:1030. doi: 10.1111/1755-0998.12417
- Antia R, Regoes RR, Koella JC, Bergstrom CT (2003) The role of evolution in the emergence of infectious diseases. *Nature* 426:658–661. doi: 10.1038/nature02177.1.
- Archer FI, Adams PE, Schneiders BB (2017) stratag: An r package for manipulating, summarizing and analysing population genetic data. *Molecular Ecology Resources* 17:5–11. doi: 10.1111/1755-0998.12559
- Armitage KB (2017) Hibernation as a major determinant of life-history traits in marmots. *Journal of Mammalogy* 98:321–331. doi: 10.1093/jmammal/gyw159
- Arnold ML (1992) Natural hybridization as an evolutionary process. *Annual Review of Ecology and Systematics* 23:237–261.
- Arnold ML, Hodges SA (1995) Are natural hybrids fit or unfit relative to their parents? *Trends in Ecology and Evolution* 10:67–71. doi: 10.1016/S0169-5347(00)88979-X
- Arnold ML, Kunte K (2017) Adaptive genetic exchange: A tangled history of admixture and evolutionary innovation. *Trends in Ecology and Evolution* 32:601–611. doi: 10.1016/j.tree.2017.05.007
- Arnold ML, Martin NH (2010) Hybrid fitness across time and habitats. *Trends in Ecology and Evolution* 25:530–536. doi: 10.1016/j.tree.2010.06.005
- Arnold ML, Meyer A (2006) Natural hybridization in primates: ne evolutionary mechanism. *Zoology* 109:261–76. doi: 10.1016/j.zool.2006.03.006
- Ballesteros-Duperon E, Virgos E, Moleon M, et al (2015) How accurate are coat traits for discriminating wild and hybrid forms of *Felis silvestris*. *Mammalia* 79:101–110.
- Bansal S, Grenfell BT, Meyers LA (2007) When individual behaviour matters: homogeneous and network models in epidemiology. *Journal of the Royal Society, Interface* 4:879–891. doi: 10.1098/rsif.2007.1100
- Barbará T, Palma-Silva C, Paggi GM, et al (2007) Cross-species transfer of nuclear microsatellite markers: Potential and limitations. *Molecular Ecology* 16:3759–3767. doi: 10.1111/j.1365-294X.2007.03439.x

References

- Barilani M, Bernard-Laurent A, Mucci N, et al (2007a) Hybridisation with introduced chukars (*Alectoris chukar*) threatens the gene pool integrity of native rock (*A. graeca*) and red-legged (*A. rufa*) partridge populations. *Biological Conservation* 137:57–69. doi: 10.1016/j.biocon.2007.01.014
- Barilani M, Sfougaris A, Giannakopoulos A, et al (2007b) Detecting introgressive hybridisation in rock partridge populations (*Alectoris graeca*) in Greece through Bayesian admixture analyses of multilocus genotypes. *Conservation Genetics* 8:343–354. doi: 10.1007/s10592-006-9174-1
- Barreto ML (2006) Infectious diseases epidemiology. *Journal of epidemiology and Community Health* 60:192–195. doi: 10.1136/jech.2003.011593
- Bartlett MS (1956) Deterministic and stochastic models for recurrent epidemics. In: *Proceedings on the third Berkeley symposium on mathematical statistics and probability*. Vol.4, n°81, p109
- Bartlett MS (1957) Measles periodicity and community size. *Journal of the Royal Statistical Society: Series A* 120:48–70. doi: 10.2307/2342553
- Beaumont M, Barratt EM, Gottelli D, et al (2001) Genetic diversity and introgression in the Scottish wildcat. *Molecular Ecology* 10:319–336.
- Beineke A, Baumgärtner W, Wohlsein P (2015) Cross-species transmission of canine distemper virus-an update. *One Health* 1:49–59. doi: 10.1016/j.onehlt.2015.09.002
- Belkhir K, Borsa P, Chikhi L, et al (2004) GENETIC 4.05, Windows TM software for population genetics. Laboratoire génome, populations, interactions, CNRS UMR5000.
- Benirschke K, Brownhill LE, Beath MM (1962) Somatic chromosomes of the horse, the donkey and their hybrids, the mule and the hinny. *Journal of Reproduction and Fertility* 4:319–26. doi: 10.1530/jrf.0.0040319
- Berthier K, Langlais M, Auger P, Pontier D (2000) Dynamics of a feline virus with two transmission modes within exponentially growing host populations. *Proceedings Biological Sciences* 267:2049–56. doi: 10.1098/rspb.2000.1248
- Best SM, Kerr PJ (2000) Coevolution of host and virus: The pathogenesis of virulent and attenuated strains of myxoma virus in resistant and susceptible European rabbits. *Virology* 267:36–48. doi: 10.1006/viro.1999.0104
- Biró Z, Szemethy L, Heltai M (2004) Home range sizes of wildcats (*Felis silvestris*) and feral domestic cats (*Felis silvestris f. catus*) in a hilly region of Hungary. *Mammalian Biology* 69:302–310. doi: 10.1078/1616-5047-00149
- Bodewes R, Sánchez Contreras GJ, García AR, et al (2015) Identification of DNA

References

- sequences that imply a novel gammaherpesvirus in seals. *Journal of General Virology* 96:1109–1114. doi: 10.1099/vir.0.000029
- Bohling JH, Adams JR, Waits LP (2013) Evaluating the ability of Bayesian clustering methods to detect hybridization and introgression using an empirical red wolf data set. *Molecular Ecology* 22:74–86. doi: 10.1111/mec.12109
- Bohling JH, Waits LP (2011) Assessing the prevalence of hybridization between sympatric *Canis* species surrounding the red wolf (*Canis rufus*) recovery area in North Carolina. *Molecular Ecology* 20:2142–2156. doi: 10.1111/j.1365-294X.2011.05084.x
- Bohling JH, Waits LP (2015) Factors influencing red wolf-coyote hybridization in eastern North-Carolina, USA. *Biological Conservation* 184:108–116 doi: 10.1016/j.biocon.2015.01.013
- Bonin A, Bellemain E, Eidesen PB, et al (2004) How to track and assess genotyping errors in population genetics studies. *Molecular Ecology* 13:3261–3273. doi: 10.1111/j.1365-294X.2004.02346.x
- Bortolussi N, Durand E, Blum M, François O (2006) apTreeshape: Statistical analysis of phylogenetic tree shape. *Bioinformatics* 22:363–364. doi: 10.1093/bioinformatics/bti798
- Bradshaw JWS, Horsfield GF, Allen JA, Robinson IH (1999) Feral cats: Their role in the population dynamics of *Felis catus*. *Applied Animal Behaviour Science* 65:273–283. doi: 10.1016/S0168-1591(99)00086-6
- Brennan AC, Woodward G, Seehausen O, et al (2014) Hybridization due to changing species distributions: Adding problems or solutions to conservation of biodiversity during global change? *Evolutionary Ecology Resources* 16:475–491.
- Briand FX, Schmitz A, Ogor K, et al (2017) Emerging highly pathogenic H5 avian influenza viruses in France during winter 2015/16: Phylogenetic analyses and markers for zoonotic potential. *Eurosurveillance* 22:1–11. doi: 10.2807/1560-7917.ES.2017.22.9.30473
- Brookfield JFY (1996) A simple new method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology* 5:453–455. doi: 10.1046/j.1365-294X.1996.00098.x
- Bruning-Fann CS, Schmitt SM, Fitzgerald SD, et al (2001) Bovine tuberculosis in free-ranging carnivores from Michigan. *Journal of wildlife diseases* 37:58–64. doi: 10.7589/0090-3558-37.1.58
- Burke JM, Arnold ML (2001) Genetics and the fitness of hybrids. *Annual Review of Genetics* 35:31–52. doi: 10.1146/annurev.genet.35.102401.085719

References

- Bury BR (2006) Natural history, field ecology, conservation biology and wildlife management: time to connect the dots. *Herpetological Conservation and Biology* 1:56–61.
- Calenge C (2011) Home Range Estimation in R: the adehabitatHR Package. *Office* 1–60.
- Calisher CH, Childs JE, Field HE, et al (2006) Bats: Important reservoir hosts of emerging viruses. *Clinical Microbiology Reviews* 19:531–545. doi: 10.1128/CMR.00017-06
- Caron A, Cappelle J, Cumming GS, et al (2015) Bridge hosts, a missing link for disease ecology in multi-host systems. *Veterinary Research* 46:1–11. doi: 10.1186/s13567-015-0217-9
- Casas F, Arroyo B, Viñuela J, et al (2016) Are farm-reared red-legged partridge releases increasing hunting pressure on wild breeding partridges in central Spain? *European Journal of Wildlife Research* 62:79–84. doi: 10.1007/s10344-015-0975-8
- Cavanaugh JE (1999) A large-sample model selection criterion based on Kullback's symmetric divergence. *Statistics and Probability Letters* 42:333–343. doi: 10.1016/S0167-7152(98)00200-4
- Cemagref, Chéry J-P, Deshayes M (2010) La fragmentation des espaces naturels par région forestière départementale.
- Chazara O, Minvielle F, Roux D, et al (2010) Evidence for introgressive hybridization of wild common quail (*Coturnix coturnix*) by domesticated Japanese quail (*Coturnix japonica*) in France. *Conservation Genetics* 11:1051–1062. doi: 10.1007/s10592-009-9951-8
- Cheng L, Connor TR, Sirén J, et al (2013) Hierarchical and spatially explicit clustering of DNA sequences with BAPS software. *Molecular Biology and Evolution* 30:1224–1228. doi: 10.1093/molbev/mst028
- Clare EL (2011) Cryptic species? patterns of maternal and paternal gene flow in eight Neotropical bats. *PLoS One*. doi: 10.1371/journal.pone.0021460
- Cleaveland S, Appel MGJ, Chalmers WSK, et al (2000) Serological and demographic evidence for domestic dogs as a source of canine distemper virus infection for Serengeti wildlife. *Veterinary Microbiology* 72:217–227. doi: 10.1016/S0378-1135(99)00207-2
- Cleaveland S, Laurenson MK, Taylor LH (2001) Diseases of humans and their domestic mammals: pathogen characteristics, host range and the risk of emergence. *Philosophical Transactions of the Royal Society B: Biological Sciences* 356:991–999. doi: 10.1098/rstb.2001.0889

References

- Coleman RR, Gaither MR, Kimokeo B, et al (2014) Large-scale introduction of the Indo-Pacific damselfish *Abudefduf vaigiensis* into Hawai'i promotes genetic swamping of the endemic congener *A. abdominalis*. *Molecular Ecology* 23:5552–5565. doi: 10.1111/mec.12952
- Condé B, Schauenberg P (1974) Reproduction du Chat forestier (*F. silvestris* Schr.) dans le nord-est de la France. *Revue Suisse de Zoologie* 81:45–52.
- Corander J, Marttinen P (2006) Bayesian identification of admixture events using multilocus molecular markers. *Molecular Ecology* 15:2833–2843. doi: 10.1111/j.1365-294X.2006.02994.x
- Corander J, Waldmann P, Sillanpää MJ (2003) Bayesian analysis of genetic differentiation between populations. *Genetics* 163:367–374.
- Corbett LK (1979) Feeding ecology and social organisation of wildcats (*Felis silvestris*) and domestic cats (*Felis catus*) in Scotland. Aberdeen. PhD Thesis.
- Corrales C, Pavlovska M, Höglund J (2014) Phylogeography and subspecies status of Black Grouse. *Journal of ornithology* 155:13–25. doi: 10.1007/s10336-013-0982-z
- Couch AJ, Unmack PJ, Dyer FJ, Lintermans M (2016) Who's your mama? Riverine hybridization of threatened freshwater trout cod and Murray cod. *PeerJ*, 4:e2593
- Coughlan J, Mirimin L, Dillane E, et al (2006) Isolation and characterization of novel microsatellite loci for the short-beaked common dolphin (*Delphinus delphis*) and cross-amplification in other cetacean species. *Molecular Ecology Notes* 6:490–492. doi: 10.1111/j.1471-8286.2006.01284.x
- Courchamp F, Say L, Pontier D (2000) Transmission of Feline Immunodeficiency Virus in a population of cats (*Felis catus*). *Wildlife Research* 27:603-611
- Craft ME (2015) Infectious disease transmission and contact networks in wildlife and livestock. *Transactions of the Royal Society B: Biological Sciences* 370:1–12. doi: 10.1098/rstb.2014.0107
- Craft ME, Volz E, Packer C, Meyers LA (2011) Disease transmission in territorial populations: the small-world network of Serengeti lions. *Journal of the Royal Society, Interface* 8:776–786. doi: 10.1098/rsif.2010.0511
- Craft, M. E., & Caillaud, D. (2011). Network models: an underutilized tool in wildlife epidemiology? *Interdisciplinary perspectives on infectious diseases, 2011*.
- Crispo E, Moore J-S, Lee-Yaw JA, et al (2011) Broken barriers: Human-induced changes to gene flow and introgression in animals. *BioEssays* 33:508–518. doi: 10.1002/bies.201000154

References

- Croteau EK, Heist EJ, Nielsen CK (2010) Fine-scale population structure and sex-biased dispersal in bobcats (*Lynx rufus*) from southern Illinois. *Canadian Journal of Zoology* 88:536–545. doi: 10.1139/Z10-024
- Csiza CK, De Lahunta A, Scott FW, Gillepsie JH (1971) Pathogenesis of feline panleukopenia virus in susceptible newborn kittens II. Pathology and immunofluorescence. *Infection and Immunity* 3(6):838:846.
- Daniels MJ, Beaumont MA, Johnson PJ, et al (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. doi: 10.1046/j.1365-2664.2001.00580.x
- Daniels MJ, Corbett L (2003) Redefining introgressed protected mammals: When is a wildcat a wild cat and a dingo a wild dog? *Wildlife Research* 30:213–218. doi: 10.1071/WR02045
- Daniels MJ, Golder MC, Jarrett O, MacDonald DW (1999) Feline viruses in wildcats from Scotland. *Journal of Wildlife Diseases* 35:121–124.
- Daniels MJ, Wright TCM, Bland KP, Kitchener AC (2002) Seasonality and reproduction in wild-living cats in Scotland. *Acta Theriologica (Warsz)* 47:73–84. doi: 10.1007/BF03193568
- Darwin C (1859) *On the Origin of the species by means of natural selection or the preservation of favoured races in the struggle for life.*
- Daszak P, Cunningham AA, Hyatt AD (2001) Anthropogenic environmental change and the emergence of infectious diseases in wildlife. *Acta Tropica* 78:103–116. doi: 10.1016/S0001-706X(00)00179-0
- de Jong MCM, Diekmann O, Heesterbeek H (1995) How does transmission of infection depend on population size. In: *Epidemic Models: their structure and relation to data.* p84–94.
- De La Torre A, Ingvarsson PK, Aitken SN (2015) Genetic architecture and genomic patterns of gene flow between hybridizing species of *Picea*. *Heredity* 115:153–164. doi: 10.1038/hdy.2015.19
- Delibes M, Rodríguez A, Ferreras P (2000) Action plan for the conservation of the Iberian lynx in Europe (*Lynx pardinus*). *Nature and Environment* 111:44.
- Delsuc F, Brinkmann H, Philippe H (2005) Phylogenomics and the reconstruction of the tree of life. *Nature Reviews Genetics* 6:361–375. doi: 10.1038/nrg1603
- Dempster AP, Laird NM, Rubin DB (1977) Maximum Likelihood from Incomplete

References

- Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (methodological)*, 1-38.
- Devillard S, Jombart T, Léger F, et al (2014) How reliable are morphological and anatomical characters to distinguish European wildcats, domestic cats and their hybrids in France? *Journal of Zoological Systematics and Evolutionary Research* 52:154–162. doi: 10.1111/jzs.12049
- Devillard S, Say L, Pontier D (2004) Molecular and behavioural analyses reveal male-biased dispersal between social groups of domestic cats. *Ecoscience* 11:175–180.
- Diekmann O, Heesterbeek JAP (2000) *Mathematical epidemiology of infectious diseases: Model building, analysis and interpretation*. Vol.5. John Wiley and Sons.
- Dimitrijevic S (1980) Ecomorphological characteristics of the wildcats (*Felis silvestris* Schreber 1777) on the territory of Vojvodina. *Nauk Beogr* 58:219–282.
- Dingemanse NJ, Both C, Drent PJ, Tinbergen JM (2004) Fitness consequences of avian personalities in a fluctuating environment. *Transactions of the Royal Society B: Biological Sciences* 271:847–852. doi: 10.1098/rspb.2004.2680
- Djan M, Popović D, Veličković N (2014) Comparison of allozyme and microsatellite variability in brown hare populations (*Lepus europaeus Pallas*) from Vojvodina (Serbia). *Biologia Serbica*, 35(1-2).
- Dodson JJ, Bourret A, Barrette MF, et al (2015) Intraspecific genetic admixture and the morphological diversification of an estuarine fish population complex. *PLoS One* 10:e0123172. doi: 10.1371/journal.pone.0123172
- Dowling TE, Secor and CL (1997) the Role of Hybridization and Introgression in the Diversification of Animals. *Annual Review of Ecology and Systematics* 28:593–619. doi: 10.1146/annurev.ecolsys.28.1.593
- Doyle A (1990) Use of riparian and upland habitats by small mammals. *Journal of Mammalogy* 71:14–23.
- 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 of the United States of America* 106 Suppl:9971–9978. doi: 10.1073/pnas.0901586106
- Driscoll CA, Menotti-Raymond M, Roca AL, et al (2007) The Near Eastern origin of cat domestication. *Science* 317:519–523. doi: 10.1126/science.1139518
- Driscoll CA, Nowell K (2010) *Felis silvestris*. The IUCN Red List of Threatened

References

Species. Version 2014.3. www.iucnredlist.org. Accessed 5 May 2015

- Duarte A, Fernandes M, Santos N, Tavares L (2012) Virological Survey in free-ranging wildcats (*Felis silvestris*) and feral domestic cats in Portugal. *Veterinary Microbiology* 158:400–4. doi: 10.1016/j.vetmic.2012.02.033
- Duchesne P, Turgeon J (2009) FLOCK: A method for a quick mapping of admixture without source samples. *Molecular Ecology Resources* 9: 1333-1344. doi: 10.1111/j.1755-0998.2009.02571.x
- Dupanloup I, Schneider S, Excoffier L (2002) A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology* 11:2571–2581. doi: 10.1046/j.1365-294X.2002.01650.x
- Earl DA, VonHoldt BM (2011) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4:359–361. doi: 10.1007/s12686-011-9548-7
- Ebert D, Hamilton WD (1996) Sex against virulence: The coevolution of parasitic diseases. *Trends in Ecology and Evolution* 11:79–82.
- 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* 75:160–174. doi: 10.1016/j.mambio.2009.07.005
- Ellegren H (2004) Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics* 5:435–445. doi: 10.1038/nrg1348
- Estrada-Peña A, Ostfeld RS, Peterson AT, et al (2014) Effects of environmental change on zoonotic disease risk: an ecological primer. *Trends in Parasitology* 30:205–214. doi: 10.1016/j.pt.2014.02.003
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14:2611–20. doi: 10.1111/j.1365-294X.2005.02553.x
- Excoffier L, Heckel G (2006) Computer programs for population genetics data analysis: a survival guide. *Nature Reviews Genetics* 7:745–758. doi: 10.1038/nrg1904
- Falsone L, Brianti E, Gaglio G, et al (2014) The European wildcats (*Felis silvestris silvestris*) as reservoir hosts of *Troglostrongylus brevior* (Strongylida: Crenosomatidae) lungworms. *Veterinary Parasitology* 205:193–8. doi: 10.1016/j.vetpar.2014.06.024

References

- Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* 164:1567–1587.
- Fauci AS, Morens DM (2012) *The perpetual challenge of infectious diseases*. 5:454–461. doi: 10.1056/NEJMra1108296
- Feil EJ, Li BC, Aanensen DM, et al (2004) eBURST: Inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *Journal of Bacteriology* 186:1518–1530. doi: 10.1128/JB.186.5.1518-1530.2004
- Felsenstein J (2004) *Inferring Phylogenies*. Sunderland, MA: Sinauer Associates.
- Felsenstein J (1981) Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution* 17:368–376. doi: 10.1007/BF01734359
- Fenton A, Pedersen AB (2005) Community epidemiology framework for classifying disease threats. *Emerging Infectious Diseases* 11:1815–1821. doi: 10.3201/eid1112.050306
- Ferrari MJ, Grais RF, Bharti N, et al (2008) The dynamics of measles in sub-Saharan Africa. *Nature* 451:679–684. doi: 10.1038/nature06509
- Ferreira JP, Leitão I, Santos-Reis M, Revilla E (2011) Human-related factors regulate the spatial ecology of domestic cats in sensitive areas for conservation. *PLoS One* 6:1–10.
- Ferrero ME, Blanco-Aguiar JA, Lougheed SC, et al (2011) Phylogeography and genetic structure of the red-legged partridge (*Alectoris rufa*): More evidence for refugia within the Iberian glacial refugium. *Molecular Ecology* 20:2628–2642. doi: 10.1111/j.1365-294X.2011.05111.x
- Figueiró H V., Li G, Trindade FJ, et al (2017) Genome-wide signatures of complex introgression and adaptive evolution in the big cats. *Science advances* 3:e1700299. doi: 10.1126/sciadv.1700299
- Fisher M (2003) *Toxocara cati*: An underestimated zoonotic agent. *Trends in Parasitology*. 19:167–170.
- Fitzpatrick BM, Shaffer HB (2007) Hybrid vigor between native and introduced salamanders raises new challenges for conservation. *Proceedings of the National Academy of Sciences of the United States of America* 104:15793–15798. doi: 10.1073/pnas.0704791104
- Fitzpatrick MC, Hampson K, Cleaveland S, et al (2012) Potential for rabies control through dog vaccination in wildlife-abundant communities of Tanzania. *PLoS*

References

- Neglected Tropical Diseases* 6:6–11. doi: 10.1371/journal.pntd.0001796
- Floate KD, Whitham TG (1993) The “Hybrid Bridge” hypothesis: Host shifting via plant hybrid swarms. *The American Naturalist* 141:651–662.
- Fouchet D, Leblanc G, Sauvage F, et al (2009) Using dynamic stochastic modelling to estimate population risk factors in infectious disease: The example of FIV in 15 cat populations. *PLoS One* 4:e7377. doi: 10.1371/journal.pone.0007377
- Fraley C (1998) How many clusters? Which clustering method? Answers via model-based cluster analysis. *The Computer Journal* 41:578–588. doi: 10.1093/comjnl/41.8.578
- Fraley C, Raftery AE (2002) Model-based clustering, discriminant analysis and density estimation. *Journal of the American Statistical Association* 97:611–631.
- Francisco FO, Santiago LR, Brito RM, et al (2014) Hybridization and asymmetric introgression between *Tetragonisca angustula* and *Tetragonisca fiebrigi*. *Apidologie* 45:1–9. doi: 10.1007/s13592-013-0224-7
- Fredriksen A (2016) Of wildcats and wild cats: Troubling species-based conservation in the Anthropocene. *Environment and Planning D: Society and Space* 34:689–705. doi: 10.1177/0263775815623539
- French DD, Easterbee N (1988) Morphological discriminants of Scottish wildcats. 235–259. *Journal of Zoology*, 214(2), 235-259.
- Fromont E, Bernard UC, et al (1997) Modelling the feline leukemia virus (FeLV) in natural populations of cats (*Felis catus*). *Theoretical Population Biology* 52:60–70.
- Fulford GR, Roberts MG, Heesterbeek JAP (2002) The metapopulation dynamics of an infectious disease: tuberculosis in possums. *Theoretical Population Biology* 61:15–29. doi: 10.1006/tpbi.2001.1553
- Gade-Jorgensen I, Stagegaard R (2000) Diet composition of wolves *Canis lupus* in east-central Finland. *Acta Theriologica (Warsz)* 45:537–547. doi: 10.4098/AT.arch.00-52
- Gagnon A-C, Dantzer R (2013) Le comportement reproducteur. In: *Comportement du chat : Biologie et Clinique*. Les éditions du point vétérinaire, Maison Alfort, p 151
- Galaverni M, Caniglia R, Pagani L, et al (2017) Disentangling timing of admixture, patterns of introgression, and phenotypic indicators in a hybridizing wolf population. *Molecular Biology and Evolution* 34:2324–2339. doi: 299

References

10.1093/molbev/msx169

- Galov A, Fabbri E, Caniglia R, et al (2015) First evidence of hybridization between golden jackal (*Canis aureus*) and domestic dog (*Canis familiaris*) as revealed by genetic markers. *Royal Society Open Science* 2:150450. doi: 10.1098/rsos.150450
- Garroway CJ, Bowman J, Cascaden TJ, et al (2010) Climate change induced hybridization in flying squirrels. *Global Change Biology* 16:113–121. doi: 10.1111/j.1365-2486.2009.01948.x
- Gaskell R, Dawson S, Radford A, Thiry E (2007) Feline herpesvirus. *Veterinary Research* 38:337–354. doi: 10.1051/vetres:2006063
- Gaskell R, Povey R (1982) Transmission of feline viral rhinotracheitis. *Veterinary Record* 111:359–362. doi: 10.1136/vr.111.16.359
- 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. doi: 10.1111/j.1469-7998.2008.00479.x
- Germain E, Ruetten S, Poulle ML (2009) Likeness between the food habits of European wildcats, domestic cats and their hybrids in France. *Mammalian Biology* 74:412–417. doi: 10.1016/j.mambio.2009.05.008
- Gilbert KJ, Andrew RL, Bock DG, et al (2012) Recommendations for utilizing and reporting population genetic analyses: The reproducibility of genetic clustering using the program structure. *Molecular Ecology* 21:4925–4930. doi: 10.1111/j.1365-294X.2012.05754.x
- Godinho R, Llaneza L, Blanco JC, et al (2011) Genetic evidence for multiple events of hybridization between wolves and domestic dogs in the Iberian Peninsula. *Molecular Ecology* 20:5154–66. doi: 10.1111/j.1365-294X.2011.05345.x
- Godinho R, López-Bao JV, Castro D, et al (2015) Real-time assessment of hybridization between wolves and dogs: combining noninvasive samples with ancestry informative markers. *Molecular Ecology Resources* 15:317–328. doi: 10.1111/1755-0998.12313
- Goldberg TL, Grant EC, Inendino KR, et al (2005) Increased infectious disease susceptibility resulting from outbreeding depression. *Conservation Biology* 19:455–462. doi: 10.1111/j.1523-1739.2005.00091.x
- Gomez D, Anthony R (1998) Small mammal abundance in riparian and upland areas of five seral stages in western Oregon. *Northwest Science* 72:293–302.
- Gompert A, Lucas LK, Buerkle CA et al. (2014) Admixture and the organization of genetic diversity in a butterfly species complex revealed through common and

References

- rare genetic variants. *Molecular Ecology* 23(18): 4555-4573.
- Gompert Z, Buerkle CA (2016) What, if anything, are hybrids: enduring truths and challenges associated with population structure and gene flow. *Evolutionary applications* 9(7):909-923.
- Gortázar C, Torres MJ, Vicente J, et al (2008) Bovine tuberculosis in Doñana Biosphere Reserve: The role of wild ungulates as disease reservoirs in the last Iberian lynx strongholds. *PLoS One*. doi: 10.1371/journal.pone.0002776
- Goudet J (1995) FSTAT (Version 1.2): A Computer Program to Calculate F-Statistics. *Journal of Heredity* 86:485-486.
- Goudet J (2005) HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes* 5:184-186. doi: 10.1111/j.1471-8278.2004.00828.x
- Grant PR, Grant BR (1992) Hybridization of bird species. *Science* 256:193-7. doi: 10.1126/science.256.5054.193
- Greger M (2007) The human/animal interface: Emergence and resurgence of zoonotic infectious diseases. *Critical Reviews in Microbiology* 33:243-299. doi: 10.1080/10408410701647594
- Grill A, Amori G, Aloise G, et al (2009) Molecular phylogeography of European *Sciurus vulgaris*: Refuge within refugia? *Molecular Ecology* 18:2687-2699. doi: 10.1111/j.1365-294X.2009.04215.x
- Guichoux E, Lagache L, Wagner S, et al (2011) Current trends in microsatellite genotyping. *Molecular Ecology Resources* 11:591-611. doi: 10.1111/j.1755-0998.2011.03014.x
- Guildea C, Hitchen Y, Duffy R, et al (2015) Introgression threatens the survival of the critically endangered freshwater crayfish *Cherax tenuimanus* (Decapoda: Parastacidae) in the Wild. *PLoS One* 10:e0121075. doi: 10.1371/journal.pone.0121075
- Guiserix M (2009) Mécanismes d'émergence des maladies infectieuses: étude par la modélisation du rôle de la protection de groupe, dans des populations hôtes homogènes ou structurées spatialement. Université Claude Bernard-Lyon 1. PhD Thesis.
- Haasl RJ, Payseur BA (2010) Multi-locus inference of population structure: a comparison between single nucleotide polymorphisms and microsatellites. *Heredity* 106:158-171. doi: 10.1038/hdy.2010.21

References

- Hall RJ (2016) Hybridization helps colonizers become conquerors. *Proceedings of the National Academy of Sciences* 113:9963–9964. doi: 10.1073/pnas.1611222113
- Hamill RM, Doyle D, Duke EJ (2006) Spatial patterns of genetic diversity across European subspecies of the mountain hare, *Lepus timidus* L. *Heredity* 97:355–365. doi: 10.1038/sj.hdy.6800880
- Han BA, Schmidt JP, Bowden SE, Drake JM (2015) Rodent reservoirs of future zoonotic diseases. *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.1501598112
- Handt L, Fox J, Dewhirst F (1994) *Helicobacter pylori* isolated from the domestic cat: public health implications. *Infection and Immunity*, 62(6) :2367–2374.
- Hansen H, Hess SC, Cole D, Banko PC (2007) Using population genetic tools to develop a control strategy for feral cats (*Felis catus*) in Hawai'i. *Wildlife Research* 34:587–596. doi: 10.1071/WR07043
- Hardy OJ, Vekemans X (2002) SPAGeDi : a versatile computer program to analyse spatial genetic structure at the individual or populational levels. *Molecular Ecology Notes* 2:618–620.
- Harrison RG (1993) Hybrids and hybrid zones: historical perspective. In: *Hybrid zones and the evolutionary process*. pp 3–12
- Harrison RG (1990) Hybrid zones: windows on evolutionary process. *Oxford Surveys in Evolutionary Biology* 7:69–128.
- Harrison HB, Feldheim KA, Jones GP et al. (2014) Variation of microsatellite multiplexes for parentage analysis and species discrimination in two hybridizing species of coral reef fish (*Plectropomus* spp. , Serranidae) *Ecology and Evolution* 4: 2046-2057
- Hartmann K (2011) Clinical aspects of feline immunodeficiency and feline leukemia virus infection. *Veterinary Immunology and Immunopathology* 143:190–201. doi: 10.1016/j.vetimm.2011.06.003
- Hartmann S a., Steyer K, Kraus RHS, et al (2013) Potential barriers to gene flow in the endangered European wildcat (*Felis silvestris*). *Conservation Genetics* 14:413–426. doi: 10.1007/s10592-013-0468-9
- Haydon DT, Randall DA, Matthews L, et al (2006) Low-coverage vaccination strategies for the conservation of endangered species. *Nature* 443:692–5. doi: 10.1038/nature05177
- Hedrick PW, Fredrickson R (2010) Genetic rescue guidelines with examples from Mexican wolves and Florida panthers. *Conservation Genetics* 11:615–626. doi:

References

10.1007/s10592-009-9999-5

- Hellard E, Fouchet D, Santin-Janin H, et al (2011) When cats' ways of life interact with their viruses: A study in 15 natural populations of owned and unowned cats (*Felis silvestris catus*). *Preventive Veterinary Medicine* 101:250–264. doi: 10.1016/j.prevetmed.2011.04.020
- Hellard E, Pontier D, Sauvage F, et al (2012) True versus false parasite interactions: A robust method to take risk factors into account and its application to feline viruses. *PLoS One* 7:e29618. doi: 10.1371/journal.pone.0029618
- Heller M (1992) Status and conservation of the wildcat (*Felis silvestris*) in the Badem-Württemberg, Southwest Germany. In: *Seminary on the biology and conservation of the wildcat*. pp 42–45
- Herrero J, García-Serrano A, Couto S et al. (2006) Diet of wild boar *sus scrofa* L. and crop damage in an intensive agroecosystem. *European Journal of Wildlife Research* 52(4):245-250 doi: 10.1007/s10344-006-0045-3
- Hertwig ST, Schweizer M, Stepanow S, et al (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. doi: 10.1111/j.1439-0469.2009.00536.x
- Hoelzer K, Parrish CR (2010) The emergence of parvoviruses of carnivores. *Veterinary Research* doi: 10.1051/vetres/2010011
- Hoffmann A, Sgrò C (2011) Climate change and evolutionary adaptation. *Nature* 470:479–485. doi: 10.1038/nature09670
- Hofmann-Lehmann R, Fehr D, Grob M, et al (1996) Prevalence of antibodies to feline parvovirus, calicivirus, herpesvirus, coronavirus and immunodeficiency virus and of feline leukemia virus antigen and the interrelationship of these viral infections in free-ranging lions in east Africa. *Clinical and Diagnostic Laboratory Immunology* 3:554–562. doi: 1071-412X/96/\$04.00+0
- Höner OP, Wachter B, East ML, et al (2007) Female mate-choice drives the evolution of male-biased dispersal in a social mammal. *Nature* 448:798–801. doi: 10.1038/nature06040
- Höss M, Kohn M, Pääbo S, et al (1992) Excrement analysis by PCR. *Nature* 359:199.
- Huang Y, Tu J, Cheng X, et al (2005) Characterization of 35 novel microsatellite DNA markers from the duck (*Anas platyrhynchos*) genome and cross-amplification in other birds. *Genetics, Selection, Evolution* 37:455–472. doi: 10.1186/1297-

References

9686-37-5-455

- Hubbard AL, McOris S, Jones TW, et al (1992) Is survival of European wildcats *Felis silvestris* in Britain threatened by interbreeding with domestic cats? *Biological Conservation* 61:203–208. doi: 10.1016/0006-3207(92)91117-B
- Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources* 9:1322–1332. doi: 10.1111/j.1755-0998.2009.02591.x
- Inskip C, Zimmermann A (2009) Human-felid conflict: a review of patterns and priorities worldwide. *Oryx* 43:18. doi: 10.1017/S003060530899030X
- IUCN (2016) IUCN 2016. The IUCN red list of threatened species. In: Version 2016.2. www.iucnredlist.org. Accessed 4 Sep 2016
- Jackiw RN, Mandil G, Hager HA (2015) A framework to guide the conservation of species hybrids based on ethical and ecological considerations. *Conservation Biology* 29:1040–1051. doi: 10.1111/cobi.12526
- Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23:1801–6. doi: 10.1093/bioinformatics/btm233
- Janečka JE, Blankenship TL, Hirth DH, et al (2007) Evidence for male-biased dispersal in bobcats *Lynx Rufus* using relatedness analysis. *Wildlife Biology* 13:38–47. doi: 10.2981/0909-6396(2007)13[38:EFMDIB]2.0.CO;2
- Jeong C, Alkorta-Aranburu G, Basnyat B, et al (2014) Admixture facilitates genetic adaptations to high altitude in Tibet. *Nature Communications* 5:3281. doi: 10.1530/ERC-14-0411.Persistent
- 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. doi: 10.1016/j.jnc.2009.02.005
- Johnson PTJ, Roode JC De, Fenton A (2016) Why infectious disease research needs community ecology. *Science* 349:1–20. doi: 10.1126/science.1259504.Why
- Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24:1403–5. doi: 10.1093/bioinformatics/btn129
- Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* 27:3070–1. doi: 10.1093/bioinformatics/btr521
- Jombart T, Archer F, Schliep K, et al (2017) Apex: Phylogenetics With Multiple

References

- Genes. *Molecular Ecology Resources* 17:19–26. doi: 10.1111/1755-0998.12567
- Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* 11:94. doi: 10.1186/1471-2156-11-94
- Jombart T, Devillard S, Dufour A-B, Pontier D (2008) Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity* 101:92–103. doi: 10.1038/hdy.2008.34
- Jombart T, Pontier D, Dufour A-B (2009) Genetic markers in the playground of multivariate analysis. *Heredity*, 102:330–341. doi: 10.1038/hdy.2008.130
- Kalinowski ST, Wagner AP, Taper ML (2006) ML-RELATE: A computer program for maximum likelihood estimation of relatedness and relationship. *Molecular Ecology Notes* 6:576–579. doi: 10.1111/j.1471-8286.2006.01256.x
- Kamvar ZN, Tabima JF, Grünwald NJ (2014) *Poppr*: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2:e281. doi: 10.7717/peerj.281
- Keeling MJ, Eames KTD (2005) Networks and epidemic models. *Journal of the Royal Society, Interface* 2:295–307. doi: 10.1098/rsif.2005.0051
- Khan F (2015) Molecular markers : An Excellent tool for genetic analysis. *Journal of Molecular Biomarkers & Diagnosis*, 6(3), 1.
- Khosravi R, Rezaei HR, Kaboli M (2013) Detecting hybridization between Iranian wild wolf (*Canis Lupus Pallipes*) and free-ranging domestic dog (*Canis Familiaris*) by analysis of microsatellite markers. *Zoological Science* 30:27–34. doi: 10.2108/zsj.30.27
- Kilshaw K, Johnson PJ, Kitchener AC, Macdonald DW (2015) Detecting the elusive Scottish wildcat *Felis silvestris silvestris* using camera trapping. *Oryx* 49:207–215. doi: 10.1017/S0030605313001154
- 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.
- Klar N, Fernandez N, Kramer-schadt S, et al (2008) Habitat selection models for European wildcat conservation. *Biological Conservation* 141:309–318.
- Klar N, Herrmann M, Kramer-Schadt S (2009) Effects and mitigation of road impacts on individual movement behavior of wildcats. *Journal of Wildlife Management* 73:631–638. doi: 10.2193/2007-574
- Krajewska M, Lipiec M, Zabost A, et al (2014) Bovine tuberculosis in a wild boar

References

- (*Sus scrofa*) in Poland. *Journal of Wildlife Diseases* 50:1001–2. doi: 10.7589/2013-07-187
- 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. doi: 10.1111/j.1439-0469.2009.00537.x
- Kuiken T, Holmes EC, McCauley J, et al (2006) Host species barriers to influenza virus infections. *Science* 312:394–397. doi: 10.1126/science.1122818
- Laloë D, Jombart T, Dufour A-B, Moazami-Goudarzi K (2007) Consensus genetic structuring and typological value of markers using multiple co-inertia analysis. *Genetics Selection Evolution* 39:545–567. doi: 10.1051/gse:2007021
- Lancaster ML, Gemmell NJ, Negro S, et al (2006) Ménage à trois on Macquarie Island: hybridization among three species of fur seal (*Arctocephalus spp.*) following historical population extinction. *Molecular Ecology* 15:3681–92. doi: 10.1111/j.1365-294X.2006.03041.x
- Landelle P (2005) Le statut juridique des espèces hybrides. *Faune Sauvage* 265 :70-75.
- Larsen PA, Marchan-Rivadeneira MR, Baker RJ (2010) Natural hybridization generates mammalian lineage with species characteristics. *Proceedings of the National Academy of Sciences* 107:11447–11452. doi: 10.1073/pnas.1000133107
- Lawson Handley LJ, Perrin N (2007) Advances in our understanding of mammalian sex-biased dispersal. *Molecular Ecology* 16(8):1559-1578
- Le Brun N, Renaud F, Berrebi P, Lambert A (1992) Hybrid zones and host-parasite relationships: effects on the evolution of parasitic specificity. *Evolution* 46:56–61. doi: 10.1017/CBO9781107415324.004
- Roex N, Koets AP, van Helden PD, Hoal EG (2013) Gene polymorphisms in African buffalo associated with susceptibility to bovine tuberculosis Infection. *PLoS One* 8:1–6. doi: 10.1371/journal.pone.0064494
- Le Roux JJ, Foxcroft LC, Herbst M, MacFadyen S (2015) Genetic analysis shows low levels of hybridization between African wildcats (*Felis silvestris lybica*) and domestic cats (*F. s. catus*) in South Africa. *Ecology and Evolution* 5:288–299. doi: 10.1002/ece3.1275
- Leary RF, Allendorf FW, Sage GK (1995) Hybridization and introgression between introduced and native fish. In: *American Fisheries Society Symposium*. pp 91–101

References

- Lecis R, Pierpaoli M, Birò ZS, et al (2006) Bayesian analyses of admixture in wild and domestic cats (*Felis silvestris*) using linked microsatellite loci. *Molecular Ecology* 15:119–31. doi: 10.1111/j.1365-294X.2005.02812.x
- Legendre P, Legendre L (2012) *Numerical ecology*. Elsevier
- Léger F, Stahl P, Ruelle S, Wilhelm J-L (2008) La répartition du chat forestier en France : évolutions récentes. *Faune Sauvage* 280:24–38.
- Lélu M, Langlais M, Poulle ML, Gilot-Fromont E (2010) Transmission dynamics of *Toxoplasma gondii* along an urban-rural gradient. *Theoretical Population Biology* 78:139–147. doi: 10.1016/j.tpb.2010.05.005
- Lembo T, Hampson K, Haydon DT, et al (2008) Exploring reservoir dynamics: A case study of rabies in the Serengeti ecosystem. *Journal of Applied Ecology* 45:1246–1257. doi: 10.1111/j.1365-2664.2008.01468.x
- Leonard JA (2002) Ancient DNA evidence for old world origin of new world dogs. *Science* 298:1613–1616. doi: 10.1126/science.1076980
- Leopold A (1933) *Game Management*. Charles Scribner's sons, New-York. pp481.
- Leroy EM (2004) Multiple Ebola virus transmission events and rapid decline of central African wildlife. *Science* 303:387–390. doi: 10.1126/science.1092528
- Leroy EM, Kumulungui B, Pourrut X, et al (2005) Fruit bats as reservoirs of Ebola virus. *Nature* 438:575–576. doi: 10.1038/438575a
- Leutenegger CM, Hofmann-Lehmann R, Riols C, et al (1999) Viral infections in free-living populations of the European wildcat. *Journal of Wildlife Diseases* 35:678–686. doi: 10.7589/0090-3558-35.4.678
- Levin DA (2002) Hybridization and extinction in protecting rare species, conservationists should consider the dangers of interbreeding, which compound the more well-known threats to wildlife. *American Scientist* 90(3): 254-261.
- Lindberg J, Björnerfeldt S, Bakken M, et al (2007) Selection for tameness modulates the expression of heme related genes in silver foxes. *Behavioral and Brain Functions* 3:18. doi: 10.1186/1744-9081-3-18
- Lipinski MJ, Froenicke L, Baysac KC, et al (2008) The ascent of cat breeds: Genetic evaluations of breeds and worldwide random-bred populations. *Genomics* 91:12–21. doi: 10.1016/j.ygeno.2007.10.009
- Litvaitis JA, Reed GC, Carroll RP, et al (2015) Bobcats (*Lynx rufus*) as a model

References

- organism to investigate the effects of roads on wide-ranging carnivores. *Environment Management* 55:1366–1376. doi: 10.1007/s00267-015-0468-2
- Lloyd-Smith JO, Cross PC, Briggs CJ, et al (2005) Should we expect population thresholds for wildlife disease? *Trends in Ecology and Evolution* 20:511–519. doi: 10.1016/j.tree.2005.07.004
- Lodé T, Guiral G, Peltier D (2005) European mink-polecat hybridization events: Hazards from natural process? *Journal of Heredity* 96:89–96. doi: 10.1093/jhered/esi021
- Loss SR, Will T, Marra PP (2013) The impact of free-ranging domestic cats on wildlife of the United States. *Nature Communications* 4:1396. doi: 10.1038/ncomms2380
- Loughnan SR, Beheregaray LB, Robinson NA (2015) Microsatellite marker development based on next-generation sequencing for the smooth marron (*Cherax cainii*, Austin) and cross-species amplification in other *Cherax* species. *BMC Research Notes* 8:370. doi: 10.1186/s13104-015-1345-z
- Lozano J, Malo AF (2012) Conservation of the European Wildcat (*Felis silvestris*) in Mediterranean Environments: A reassessment of current threats. In: Williams GS (ed) *Mediterranean Ecosystems*. Nova Science Publishers, Inc, pp 1–31
- 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, Malo AF, et al (2003) Importance of scrub – pastureland mosaics for wild- living cats occurrence in a Mediterranean area : implications for the conservation of the wildcat (*Felis silvestris*). *Biodiversity and Conservation* 12:921–935. doi: 10.1023/A:1022821708594
- MacDonald DW (2004) The Scottish wildcat: analyses for conservation and an action plan. Wildlife Conservation Research Unit.
- Macdonald DW (1983) The ecology of carnivore social behaviour. *Nature* 301:379–384.
- Macdonald V (2015) Scottish wildcats threatened by feline form of HIV, study finds. <http://www.scottishwildcataction.org/latest-news/2015/september/news-scottish-wildcats-threatened-by-feline-form-of-hiv-study-finds/>.
- Maduna SN, Rossouw C, da Silva C, et al (2017) Species identification and comparative population genetics of four coastal houndsharks based on novel NGS-mined microsatellites. *Ecology and Evolution* 7:1462–1486. doi: 10.1002/ece3.2770

References

- Maiden MC, Bygraves JA, Feil E et al. (1998) Multilocus sequence typing : a portable approach to the identification of clones within populations of pathogenetic microorganisms. *Proceedings of the National Academy of Sciences of the United States of America* 95:3140-3145.
- Mallet J (2005) Hybridization as an invasion of the genome. *Trends in Ecology and Evolution* 20:229–37. doi: 10.1016/j.tree.2005.02.010
- Mallet J, Besansky N, Hahn MW (2016) How reticulated are species? *BioEssays* 38:140–149. doi: 10.1002/bies.201500149
- 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* 263:401–407. doi: 10.1017/S0952836904005448
- Malukiewicz J, Boere V, Fuzessy LF, et al (2015) Natural and Anthropogenic Hybridization in Two Species of Eastern Brazilian Marmosets (*Callithrix jacchus* and *C. penicillata*). *PLoS One* 10:e0127268. doi: 10.1371/journal.pone.0127268
- Mandeville, E. G., Parchman, T. L., McDonald, D. B., & Buerkle, C. A. (2015). Highly variable reproductive isolation among pairs of *Catostomus* species. *Molecular Ecology*, 24(8), 1856-1872.
- Martin J, Rey B, Pons JB, et al (2013) Movements and space use of feral cats in Kerguelen archipelago: A pilot study with GPS data. *Polar Biology* 36:1531–1536. doi: 10.1007/s00300-013-1365-x
- Mattucci F, Oliveira R, Bizzarri L, et al (2013) Genetic structure of wildcat (*Felis silvestris*) populations in Italy. *Ecology and Evolution* 3:2443–2458. doi: 10.1002/ece3.569
- Maudet C, Luikart G, Dubray D, et al (2004) Low genotyping error rates in wild ungulate faeces sampled in winter. *Molecular Ecology Notes* 4:772–775. doi: 10.1111/j.1471-8286.2004.00787.x
- Mauvy B, Péroux R, Guitton J-S (2010) La gestion « en deux temps » des prélèvements de lièvres. *Faune Sauvage* 23–29.
- Mayr E (1942) *Systematics and the origin of species, from the viewpoint of a zoologist*. Harvard University Press.
- McCallum H, Barlow N, Hone J (2001) How should pathogen transmission be modeled? *Trends in Ecology and Evolution* 16:295–300. doi: 10.1016/S0169-5347(01)02144-9
- McDevitt AD, Mariani S, Hebblewhite M, et al (2009) Survival in the Rockies of an endangered hybrid swarm from diverged caribou (*Rangifer tarandus*) lineages.

References

- Molecular Ecology* 18:665–679. doi: 10.1111/j.1365-294X.2008.04050.x
- McOrist S, Boid R, Jones TW, et al (1991) Some viral and protozoal diseases in the European wildcat (*Felis silvestris*). *Journal of Wildlife Diseases* 27:693–696.
- Medina FM, Bonnaud E, Vidal E, et al (2011) A global review of the impacts of invasive cats on island endangered vertebrates. *Global Change Biology* 17:n/a-n/a. doi: 10.1111/j.1365-2486.2011.02464.x
- Medina FM, Bonnaud E, Vidal E, Nogales M (2014) Underlying impacts of invasive cats on islands: Not only a question of predation. *Biodiversity and Conservation* 23:327–342. doi: 10.1007/s10531-013-0603-4
- Medugorac I, Graf A, Grohs C, et al (2017) Whole-genome analysis of introgressive hybridization and characterization of the bovine legacy of Mongolian yaks. *Nature Genetics* 49:470–475. doi: 10.1038/ng.3775
- Meeüs T de, Béati L, Delaye C, et al (2002) Sex-Biased Genetic Structure in the Vector of Lyme Disease, *Ixodes Ricinus*. *Evolution* (N Y) 56:1802–1807. doi: 10.1111/j.0014-3820.2002.tb00194.x
- Melo-Ferreira J, Boursot P, Randi E, et al (2007) The rise and fall of the mountain hare (*Lepus timidus*) during Pleistocene glaciations: expansion and retreat with hybridization in the Iberian Peninsula. *Molecular Ecology* 16:605–18. doi: 10.1111/j.1365-294X.2006.03166.x
- Mengoni C, Mucci N, Randi E (2015) Genetic diversity and no evidences of recent hybridization in the endemic Italian hare (*Lepus corsicanus*). *Conservation Genetics* 16:477–489. doi: 10.1007/s10592-014-0674-0
- Menotti-Raymond M, David VA, Chen ZQ, et al (2003) Second-Generation Integrated Genetic Linkage/Radiation Hybrid Maps of the Domestic Cat (*Felis catus*). *Journal of Heredity* 94:95–106. doi: 10.1093/jhered/esg008
- Menotti-Raymond M, David VA, Lyons LA, et al (1999) A genetic linkage map of microsatellites in the domestic cat (*Felis catus*). *Genomics* 57:9–23. doi: 10.1006/geno.1999.5743
- Mermod CP, Liberek M (2002) The role of snowcover for European wildcat in Switzerland. *Zeitschrift für Jagdwissenschaft* 48:17–24. doi: 10.1007/BF02192388
- Metzker ML (2010) Sequencing technologies – the next generation. *Nature Reviews. Genetics* 11(1):31.
- Millán J, Rodríguez A (2009) A serological survey of common feline pathogens in

References

- free-living European wildcats (*Felis silvestris*) in central Spain. *European Journal of Wildlife Research* 55:285–291. doi: 10.1007/s10344-008-0246-z
- Montague MJ, Li G, Gandolfi B, et al (2014) Comparative analysis of the domestic cat genome reveals genetic signatures underlying feline biology and domestication. *Proceedings of the National Academy of Sciences* 1–6. doi: 10.1073/pnas.1410083111
- Monzón J, Kays R, Dykhuizen DE (2014) Assessment of coyote-wolf-dog admixture using ancestry-informative diagnostic SNPs. *Molecular Ecology* 23:182–197. doi: 10.1111/mec.12570
- Morand S, McIntyre KM, Baylis M (2014) Domesticated animals and human infectious diseases of zoonotic origins: Domestication time matters. *Infection Genetics Evolution* 24:76–81. doi: 10.1016/j.meegid.2014.02.013
- Moreno S, Beltrón JF, Cotilla I, et al (2008) Long-term decline of the European wild rabbit (*Oryctolagus cuniculus*) in south-western Spain. *Wildlife Research* 34:652–658.
- Morens DM, Fauci AS (2013) Emerging infectious diseases: Threats to human health and global stability. *PLoS Pathogens* 9:7–9. doi: 10.1371/journal.ppat.1003467
- Morens DM, Folkers GK, Fauci AS (2008) Emerging infectious: a perpetual challenge. *Lancet Infectious Diseases* 8:710–719. doi: 10.1016/S1473-3099(08)70256-1.Emerging
- Morgan-Richards M, Smissen RD, Shepherd LD, et al (2009) A review of genetic analyses of hybridisation in New Zealand. *Journal of the Royal Society of New Zealand* 39:15–34. doi: 10.1080/03014220909510561
- Morin PA, Luikart G, Wayne RK (2004) SNPs in ecology, evolution and conservation. *Trends Ecol Evol* 19:208–216. doi: 10.1016/j.tree.2004.01.009
- Morin PA, Wallis J, Moore JJ, et al (1993) Non-invasive sampling and DNA amplification for paternity exclusion, community structure, and phylogeography in wild chimpanzees. *Primates* 34:347–356. doi: 10.1007/BF02382630
- Mouzin D, Lorenzen M, Haworth J, King V (2004) Duration of serologic response to three viral antigens in cats. *Journal of the American Veterinary Medical Association* 224:61–66.
- Munson L, Marker L, Dubovi E, et al (2004) Serosurvey of viral infections in free-ranging Namibian cheetahs (*Acinonyx jubatus*). *Journal of Wildlife Diseases* 40:23–31. doi: 10.7589/0090-3558-40.1.23

References

- Murray KA, Daszak P (2013) Human ecology in pathogenic landscapes: Two hypotheses on how land use change drives viral emergence. *Current Opinion in Virology* 3:79–83. doi: 10.1016/j.coviro.2013.01.006
- Nasell I (1995) The threshold concept in stochastic epidemic and endemic models. In: *Epidemic model: their structure and relation to data*. p 71
- Neuenschwander S, Hospital F, Guillaume F, Goudet J (2008) quantiNemo: An individual-based program to simulate quantitative traits with explicit genetic architecture in a dynamic metapopulation. *Bioinformatics* 24:1552–1553. doi: 10.1093/bioinformatics/btn219
- Newsome TM, Stephens D, Ballard G-A et al. (2013) Genetic profile of dingoes (*Canis lupus dingo*) and free-roaming domestic dogs (*C.l.familiaris*) in the Tanami desert, Australia. *Wildlife Research* 40(3):196-206
- 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. doi: 10.1111/j.1471-8286.2006.01433.x
- Nugent G (2011) Maintenance, spillover and spillback transmission of bovine tuberculosis in multi-host wildlife complexes: A New Zealand case study. *Veterinary Microbiology* 151:34–42. doi: 10.1016/j.vetmic.2011.02.023
- Nunn CL, Jordan F, McCabe CM, et al (2015) Infectious disease and group size: more than just a numbers game. *Philosophical Transactions of the Royal Society B: Biological Sciences* 370:20140111–20140111. doi: 10.1098/rstb.2014.0111
- Nussberger B, Greminger MP, Grossen C, et al (2013) Development of SNP markers identifying European wildcats, domestic cats, and their admixed progeny. *Molecular Ecology Resources* 13:447–460. doi: 10.1111/1755-0998.12075
- Nussberger B, Wandeler P, Camenisch G (2014a) A SNP chip to detect introgression in wildcats allows accurate genotyping of single hairs. *European Journal of Wildlife Research* 60:405–410. doi: 10.1007/s10344-014-0806-3
- Nussberger B, Wandeler P, Weber D, Keller LF (2014b) Monitoring introgression in European wildcats in the Swiss Jura. *Conservation Genetics* 15:1219–1230. doi: 10.1007/s10592-014-0613-0
- O'Brien J, Devillard S, Say L, et al (2009) Preserving genetic integrity in a hybridising world: Are European Wildcats (*Felis silvestris silvestris*) in eastern France distinct from sympatric feral domestic cats? *Biodiversity and Conservation* 18:2351–2360. doi: 10.1007/s10531-009-9592-8
- O'Brien SJ, Johnson W, Driscoll C, et al (2008) State of cat genomics. *Trends in Genetics* 24:268–79. doi: 10.1016/j.tig.2008.03.004

References

- Odum E, Barrett G (2005) *Fundamentals of ecology*. 5th, Thomson Br. Belmont, CA
- Okarma H, śnieżko S, Olszańska A (2002) The occurrence of wildcat in the Polish Carpathian Mountains. *Acta Theriologica (Warsz)* 47:499–504. doi: 10.1007/BF03192474
- Olano-Marin J, Plis K, Sönnichsen L, et al (2014) Weak population structure in European roe deer (*Capreolus capreolus*) and evidence of introgressive hybridization with Siberian roe deer (*C. pygargus*) in northeastern Poland. *PLoS One* 9:e109147. doi: 10.1371/journal.pone.0109147
- Oli MK, Venkataraman M, Klein PA, et al (2006) Population dynamics of infectious diseases: A discrete time model. *Ecological Modelling* 198:183–194. doi: 10.1016/j.ecolmodel.2006.04.007
- Oliveira R (2012) Towards a wide genetic approach for the European Wildcat (*Felis silvestris silvestris*) conservation: improving noninvasive molecular techniques, population analysis and admixture inferences. Universidade de Porto. PhD thesis.
- Oliveira R, Godinho R, Randi E, et al (2007) Molecular analysis of hybridisation between wild and domestic cats (*Felis silvestris*) in Portugal: implications for conservation. *Conservation Genetics* 9:1–11. doi: 10.1007/s10592-007-9297-z
- Oliveira R, Godinho R, Randi E, Alves PC (2008) Hybridization versus conservation: are domestic cats threatening the genetic integrity of wildcats (*Felis silvestris silvestris*) in Iberian Peninsula? *Philosophical Transactions of the Royal Society B: Biological Sciences* 363:2953–2961. doi: 10.1098/rstb.2008.0052
- Oliveira R, Randi E, Mattucci F et al. (2015) Toward a genome-wide approach for detecting hybrids: informative SNPs to detect introgression between domestic cats and European wildcats (*Felis silvestris*). *Heredity* 1-11.
- Orsini L, Huttunen S, Schlötterer C (2004) A multilocus microsatellite phylogeny of the *Drosophila virilis* group. *Heredity* 93:161–165. doi: 10.1038/sj.hdy.6800487
- Osbourne J, Anderson J, Spurgeon A (2005) Effects of habitat on small-mammal diversity and abundance in West Virginia. *Wildlife Society Bulletin* 33:814–822.
- Otoni C, Van Neer W, De Cupere B, et al (2017) The palaeogenetics of cat dispersal in the ancient world. *Nature Ecology & Evolution* 1:139. doi: 10.1038/s41559-017-0139
- Pacheco C, López-Bao JV, García EJ et al. (2017) Spatial assessment of wolf-dog hybridization in a single breeding period? *Scientific Reports* 7:42475

References

doi:10.1038/srep42475

- Paetkau D, Calvert W, Stirling I, Strobeck C (1995) Microsatellite analysis of population structure in Canadian polar bears. *Molecular Ecology* 4(3): 347-354. doi: 10.1111/j.1365-294X.1995.tb00227.x
- Panasci M, Ballard WB, Breck S, et al (2011) Evaluation of fecal DNA preservation techniques and effects of sample age and diet on genotyping success. *Journal of Wildlife Management* 75:1616–1624. doi: 10.1002/jwmg.221
- Papageorgiou K V., Suárez NM, Wilkie GS, et al (2016) Genome sequence of canine herpesvirus. *PLoS One* 11:1–11. doi: 10.1371/journal.pone.0156015
- Paradis E (2010) pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics* 26:419–20. doi: 10.1093/bioinformatics/btp696
- Parrish CR (1995) Pathogenesis of feline panleukopenia virus and canine parvovirus. *Baillière's clinical haematology* 8(1):57-71.
- Parrish CR, Holmes EC, Morens DM, et al (2008) Cross-species virus transmission and the emergence of new epidemic diseases. *Microbiology and Molecular Biology Reviews* 72:457–470. doi: 10.1128/MMBR.00004-08
- Parrish CR, Kawaoka Y (2005) The origins of new pandemic viruses: the acquisition of new host ranges by canine parvovirus and influenza A viruses. *Annual Review of Microbiology* 59(1):553-586
- Peakall R, Smouse PE (2012) GenALEX 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28:2537–2539.
- Pedersen AB, Jones KE, Nunn CL, Altizer S (2007) Infectious diseases and extinction risk in wild mammals. *Conservation Biology* 21:1269–1279. doi: 10.1111/j.1523-1739.2007.00776.x
- Petren K, Grant BRB, Grant PRP (1999) A phylogeny of Darwin's finches based on microsatellite DNA length variation *Proceedings of the Royal Society B* 266:321–329. doi: 10.1098/rspb.1999.0641
- Petrov I, Nikolov H, Gerasimov S (1992) Craniometrical sex determination of wild cat *Felis silvestris* in Bulgaria. *Acta Theriologica (Warsz)* 37:381–396.
- Pfennig KS (2007) Facultative mate choice drives adaptive hybridization. *Science* 318:965–967. doi: 10.1126/science.1146035

References

- Piédallu B, Quenette PY, Mounet C, et al (2016) Spatial variation in public attitudes towards brown bears in the French Pyrenees. *Biological Conservation* 197:90–97. doi: 10.1016/j.biocon.2016.02.027
- Pierpaoli M, Birò ZS, Herrmann M, et al (2003) Genetic distinction of wildcat (*Felis silvestris*) populations in Europe, and hybridization with domestic cats in Hungary. *Molecular Ecology* 12:2585–2598. doi: 10.1046/j.1365-294X.2003.01939.x
- Piertney SB, MacColl ADC, Bacon PJ, Dallas JF (1998) Local genetic structure in red grouse (*Lagopus lagopus scoticus*): Evidence from microsatellite DNA markers. *Molecular Ecology* 7:1645–1654. doi: 10.1046/j.1365-294x.1998.00493.x
- Piñeiro A, Barja I (2011) Trophic strategy of the wildcat *Felis silvestris* in relation to seasonal variation in the availability and vulnerability to capture of *Apodemus* mice. *Mammalian Biology* 76:302–307. doi: 10.1016/j.mambio.2011.01.008
- Pongracz JD, Paetkau D, Branigan M, Richardson E (2017) Recent hybridization between a polar bear and grizzly bears in the Canadian Arctic. *Arctic* 70:151–160. doi: 10.14430/arctic4643
- Pontier, D., Fromont, E., Courchamp, F., Artois, M., & Yoccoz, N. G. (1998). Retroviruses and sexual size dimorphism in domestic cats (*Felis catus* L.). *Proceedings of the Royal Society of London B: Biological Sciences*, 265(1392), 167-173.
- Pontier D, Fouchet D, Bahi-Jaber N et al. (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, et al (2007) Initial sequence and comparative analysis of the cat genome. *Genome Research* 17:1675–89. doi: 10.1101/gr.638007
- Popescu AA, Huber KT, Paradis E (2012) Ape 3.0: New tools for distance-based phylogenetics and evolutionary analysis in R. *Bioinformatics* 28:1536–1537. doi: 10.1093/bioinformatics/bts184
- Poulet H (2007) Alternative early life vaccination programs for companion animals. *Journal of Comparative Pathology* 137:S67–S71. doi: 10.1016/j.jcpa.2007.04.020
- Poulet H, Jas D, Lemeter C, et al (2008) Efficacy of a bivalent inactivated non-adjuvanted feline calicivirus vaccine: Relation between in vitro cross-neutralization and heterologous protection in vivo. *Vaccine* 26:3647–3654. doi: 10.1016/j.vaccine.2008.04.082

References

- Prentis PJ, White EM, Radford IJ, et al (2007) Can hybridization cause local extinction: A case for demographic swamping of the Australian native *Senecio pinnatifolius* by the invasive *Senecio madagascariensis*? *New Phytologist* 176:902–912. doi: 10.1111/j.1469-8137.2007.02217.x
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.
- Puechmaille SJ (2016) The program STRUCTURE does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. *Molecular Ecology Resources* 16:608–627. doi: 10.1111/1755-0998.12512
- Pusey AE, Packer C (1987) The evolution of sex-biased dispersal in lions. *Behaviour* 101:275–310.
- Qi Y, Lu B, Gao H, et al (2014) Hybridization and mitochondrial genome introgression between *Rana chensinensis* and *R. kukunoris*. *Molecular Ecology* 23:5575–88. doi: 10.1111/mec.12960
- Quantum GIS Development Team (2012) Quantum GIS Geographic Information System. In: *Open Source Geospatial Found.* Proj. <http://qgis.osgeo.org>.
- Račnik J, Skrbinšek T, Potočnik H, et al (2008) Viral infections in wild-living European wildcats in Slovenia. *European Journal of Wildlife Research* 54:767–770. doi: 10.1007/s10344-008-0202-y
- R Development Core Team (2010-2017) R: A language and environment for statistical computing. R Found Stat Comput Vienna, Austria.
- Radford AD, Coyne KP, Dawson S, et al (2007) Feline calicivirus. *Veterinary Research* 38:319–335. doi: 10.1051/vetres:2006056
- Raj A, Stephens M, Pritchard JK (2014) FastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics* 197:573–589. doi: 10.1534/genetics.114.164350
- Rand WM (1971) Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association* 66:846–850.
- Randi E (2008) Detecting hybridization between wild species and their domesticated relatives. *Molecular Ecology* 17:285–293. doi: 10.1111/j.1365-294X.2007.03417.x
- Randi E, Hulva P, Fabbri E, et al (2014) Multilocus detection of wolf x dog hybridization in Italy, and guidelines for marker selection. *PLoS One* 9:e86409. doi: 10.1371/journal.pone.0086409

References

- Randi E, Pierpaoli M, Beaumont M, et al (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, Meriggi A, Lorenzini R et al. (1992) Biochemical analysis of relationships of mediterranean *Alectoris* partridges. *The American Ornithologists'Union* 109(2):358-367
- Ranwez V, Delsuc F, Ranwez S, et al (2007) OrthoMaM: A database of orthologous genomic markers for placental mammal phylogenetics. *BMC Evolutionary Biology* 7:241. doi: 10.1186/1471-2148-7-241
- Rasmussen SA, Gerber SI, Swerdlow DL (2015) Middle east respiratory syndrome coronavirus: Update for clinicians. *Clinical Infectious Diseases* 60:1686–1689. doi: 10.1093/cid/civ118
- Revell LJ (2012) phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3:217–223. doi: 10.1111/j.2041-210X.2011.00169.x
- Rhymer JM, Simberloff D (1996) Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics* 27:83–109. doi: 10.1146/annurev.ecolsys.27.1.83
- Roberts DG, Gray CA, West RJ, Ayre DJ (2010) Marine genetic swamping: Hybrids replace an obligately estuarine fish. *Molecular Ecology* 19:508–520. doi: 10.1111/j.1365-294X.2009.04501.x
- Roelke-Parker ME, Munson L, Packer C, et al (1996) A canine distemper virus epidemic in Serengeti lions (*Panthera leo*). *Nature* 379:441:445
- Sackman AM, Rokyta DR (2013) The adaptive potential of hybridization demonstrated with bacteriophages. *Journal of Molecular Evolution* 77:221–230. doi: 10.1007/s00239-013-9586-8
- Saint-Andrieux C, Barboiron A (2015) Tableaux de chasse ongulés sauvages Saison 2014-2015. *Faune Sauvage* 308.
- Salkeld DJ, Stapp P, Tripp DW, et al (2016) Ecological traits driving the outbreaks and emergence of zoonotic pathogens. *Bioscience* 66:118–129. doi: 10.1093/biosci/biv179
- Sanchez-Donoso I, Vilà C, Puigcerver M, et al (2012) Are farm-reared quails for game restocking really common quails (*Coturnix coturnix*)?: A genetic approach. *PLoS One*. doi: 10.1371/journal.pone.0039031
- Sankararaman S, Mallick S, Dannemann T et al. (2014) The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* 507(7492):354-357.

References

- Sarmiento P (1996) Feeding ecology of the European wildcat *Felis silvestris* in Portugal. *Acta Theriologica* (Warsz) 41:409–414. doi: 10.4098/AT.arch.96-39
- Sarmiento 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 Theriol* (Warsz) 54:31–38.
- Say L, Devillard S, Léger F, et al (2012) Distribution and spatial genetic structure of European wildcat in France. *Animal Conservation* 15:18–27. doi: 10.1111/j.1469-1795.2011.00478.x
- Schliep KP (2011) phangorn: Phylogenetic analysis in R. *Bioinformatics* 27:592–593. doi: 10.1093/bioinformatics/btq706
- Schlötterer C (2004) Opinion: The evolution of molecular markers — just a matter of fashion? *Nature Reviews Genetics* 5:63–69. doi: 10.1038/nrg1249
- Schwartz MK, Pilgrim KL, McKelvey KS, et al (2004) Hybridization between Canada lynx and bobcats: Genetic results and management implications. *Conservation Genetics* 5:349–355. doi: 10.1023/B:COGE.0000031141.47148.8b
- Schwarz G (1978) Estimating the Dimension of a Model. *The Annals of Statistics* 6:461–464. doi: 10.1214/aos/1176344136
- Scott FW, Geissinger C (1997) Duration of immunity in cats vaccinated with an inactivated feline panleukopenia, herpesvirus and calicivirus vaccine. *Feline Practice*
- Scott FW, Geissinger C (1999) Long-term immunity in cats vaccinated with an inactivated trivalent vaccine *American Journal of Veterinary Research* 60:652–658
- Searle JB, Kotlík P, Rambau R V, et al (2009) The Celtic fringe of Britain: insights from small mammal phylogeography. *Proceedings of the Royal Society B* 276:4287–94. doi: 10.1098/rspb.2009.1422
- Seehausen O (1997) Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science* 277:1808–1811. doi: 10.1126/science.277.5333.1808
- Seehausen O (2004) Hybridization and adaptive radiation. *Trends in Ecology and Evolution* 19:198–207. doi: 10.1016/j.tree.2004.01.003
- Seehausen O, Takimoto G, Roy D, Jokela J (2008) Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Molecular Ecology* 17:30–44. doi: 10.1111/j.1365-294X.2007.03529.x
- Selkoe KA, Toonen RJ (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters* 9:615–29. doi:

References

- 10.1111/j.1461-0248.2006.00889.x
- Sheriff MJ, Dantzer B, Delehanty B, et al (2011) Measuring stress in wildlife: Techniques for quantifying glucocorticoids. *Oecologia* 166:869–887. doi: 10.1007/s00442-011-1943-y
- Sieber M, Robb M, Forde S et al. (2014) Dispersal network structure and infection mechanism shape diversity in a coevolutionary bacteria-phage system. *The ISME journal* 8(3):504-14
- Silva AP, Kilshaw K, Johnson PJ, et al (2013) Wildcat occurrence in Scotland: Food really matters. *Diversity and Distributions* 19:232–243. doi: 10.1111/ddi.12018
- Simková A, Dávidová M, Papoušek I, Vetešník L (2013) Does interspecies hybridization affect the host specificity of parasites in cyprinid fish? *Parasites & Vectors* 6:95. doi: 10.1186/1756-3305-6-95
- Smith KF, Acevedo-Whitehouse K, Pedersen AB (2009) The role of infectious diseases in biological conservation. *Animal Conservation* 12:1–12. doi: 10.1111/j.1469-1795.2008.00228.x
- Sokos CK, Birtsas PK, Tsachalidis EP (2008) The aims of galliforms release and choice of techniques. *Wildlife Biology* 14:412–422. doi: 10.2981/0909-6396-14.4.412
- Sommer RS, Benecke N (2006) Late Pleistocene and Holocene development of the felid fauna (*Felidae*) of Europe: A review. *Journal of Zoology* 269:7–19. doi: 10.1111/j.1469-7998.2005.00040.x
- Sparger EE (1993) Current thoughts on feline immunodeficiency virus infection. *Veterinary Clinics of North America: Small Animal Practice* 23:173–191. doi: 10.1016/S0195-5616(93)50011-6
- Spielman D, Brook BW, Frankham R (2004) Most species are not driven to extinction before genetic factors impact them. *Proceedings of the National Academy of Sciences* 101:15261–15264. doi: 10.1073/pnas.0403809101
- Stahl P (1986) Le chat forestier d'Europe (*Felis silvestris*, Schreber 1777): Exploitation des ressources et organisation spatiale. Nancy. PhD Thesis.
- Stahl P, Artois M (1994) Status and conservation of the wildcat (*Felis Silvestris*) in Europe and around the Mediterranean Rim (Vol. 69). Council of Europe
- Stahl P, Artois M, Aubert MFA (1988) Organisation spatiale et déplacements des chats forestiers adultes (*Felis silvestris*, Schreber, 1777) en Lorraine. *Rev Ecol (Terre Vie)* 43:113–132.
- Stahl P, Léger F (1992) Le chat sauvage d'Europe (*Felis silvestris* Schreber, 1777).

References

- In: *Encyclopédie des carnivores de France*. SFEPM, Paris, p 50
- Stamatis C, Suchentrunk F, Moutou KA, et al (2009) Phylogeography of the brown hare (*Lepus europaeus*) in Europe: a legacy of south-eastern Mediterranean refugia? *Journal of Biogeography* 36:515–528. doi: 10.1111/j.1365-2699.2008.02013.x
- Stebbins GL (1959) The role of hybridization in evolution. *Proceedings of the American Philosophical Society* 103:231–251.
- Stuart-Fox D, Godinho R, de Bellocq JG, et al (2009) Variation in phenotype, parasite load and male competitive ability across a cryptic hybrid zone. *PLoS One*. doi: 10.1371/journal.pone.0005677
- Sugg DW, Chesser RK, Dobson FS, Hoogland JL (1996) Population genetics meets behavioral ecology. *Trends in Ecology and Evolution* 11:338–342. doi: 10.1016/0169-5347(96)20050-3
- Sunnucks P (2000) Efficient genetic markers for population biology. *Trends in Ecology and Evolution* 15:199–203. doi: 10.1038/mt.2008.212
- Sunquist M, Sunquist F (2002) *Wildcats of the world*. University of Chicago Press, Chicago
- Svedin N, Wiley C, Veen T, et al (2008) Natural and sexual selection against hybrid flycatchers. *Proceedings of the Royal Society of Sciences B* 275:735–744. doi: 10.1098/rspb.2007.0967
- Taberlet P, Bouvet J (1992) Bear conservation genetics. *Nature*. 358:197.
- Taberlet P, Luikart G (1999) Non-invasive genetic sampling and individual identification. *Biological Journal of the Linnean Society* 68:41–55. doi: 10.1111/j.1095-8312.1999.tb01157.x
- Tallmon DA, Luikart G, Waples RS (2004) The alluring simplicity and complex reality of genetic rescue. *Trends in Ecology and Evolution* 19:489–496. doi: 10.1016/j.tree.2004.07.003
- Tamazian G, Simonov S, Dobrynin P, et al (2014) Annotated features of domestic cat - *Felis catus* genome. *Gigascience* 3:13. doi: 10.1186/2047-217X-3-13
- Tang H, Peng J, Wang P, Risch NJ (2005) Estimation of individual admixture: Analytical and study design considerations. *Genetic Epidemiology* 28:289–301. doi: 10.1002/gepi.20064
- Taylor SA, Larson EL, Harrison RG (2015) Hybrid zones: Windows on climate change. *Trends in Ecology and Evolution* 30:398–406. doi: 10.1016/j.tree.2015.04.010

References

- Tischer BK, Osterrieder N (2010) Herpesvirus - a zoonotic threat? *Veterinary Microbiology* 140:266–270. doi: 10.1016/j.vetmic.2009.06.020.
- Todesco M, Pascual MA, Owens GL, et al (2016) Hybridization and extinction. *Evolutionary Applications* 9:892–908. doi: 10.1111/eva.12367
- Tovar-Sanchez, E., & Oyama, K. (2006). Effect of hybridization of the *Quercus crassifolia* × *Quercus crassipes* complex on the community structure of endophagous insects. *Oecologia*, 147(4), 702–713.
- Traversa D, Cesare A Di, Milillo P, et al (2009) Infection by *Eucoleus aerophilus* in dogs and cats: Is another extra-intestinal parasitic nematode of pets emerging in Italy? *Research in Veterinary Science* 87:270–272. doi: 10.1016/j.rvsc.2009.02.006
- Trigo TC, Schneider A, De Oliveira TG, et al (2013) Molecular data reveal complex hybridization and a cryptic species of Neotropical wild cat. *Current Biology* 23:2528–2533. doi: 10.1016/j.cub.2013.10.046
- Troyer JL, VandeWoude S, Pecon-Slattery J, et al (2008) FIV cross-species transmission: An evolutionary prospective. *Veterinary Immunology and Immunopathology* 123:159–166. doi: 10.1016/j.vetimm.2008.01.023
- Twyford AD, Ennos RA (2012) Next-generation hybridization and introgression. *Heredity* 108:179–89. doi: 10.1038/hdy.2011.68
- UNAIDS (2016) Global AIDS update.
- Urra F (2003) El gato montés en Navarra: Distribucion, ecologia y conservacion. Universidad Autonoma de Madrid, Spain. PhD Thesis.
- Vähä J-P, 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. doi: 10.1111/j.1365-294X.2005.02773.x
- Van Dijk J, Gustavsen L, Mysterud A, et al (2008) Diet shift of a facultative scavenger, the wolverine, following recolonization of wolves. *Journal of Animal Ecology* 77:1183–1190. doi: 10.1111/j.1365-2656.2008.01445.x
- van Heezik Y, Smyth A, Adams A, Gordon J (2010) Do domestic cats impose an unsustainable harvest on urban bird populations? *Biological Conservation* 143:121–130. doi: 10.1016/j.biocon.2009.09.013
- Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite

References

- data. *Molecular Ecology Notes* 4:535–538. doi: 10.1111/j.1471-8286.2004.00684.x
- Verardi A, Lucchini V, Randi E (2006) Detecting introgressive hybridization between free-ranging domestic dogs and wild wolves (*Canis lupus*) by admixture linkage disequilibrium analysis. *Molecular Ecology* 15:2845–2855. doi: 10.1111/j.1365-294X.2006.02995.x
- Vigne J-D, Guilaine J, Debue K, et al (2004) Early taming of the cat in Cyprus. *Science* 304:259. doi: 10.1126/science.1095335
- Vila C (1997) Multiple and ancient origins of the domestic dog. *Science* 276:1687–1689. doi: 10.1126/science.276.5319.1687
- 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. doi: 10.1101/gr.116301.110
- VonHoldt, B. M., Kays, R., Pollinger, J. P., & Wayne, R. K. (2016). Admixture mapping identifies introgressed genomic regions in North American canids. *Molecular Ecology*, 25(11), 2443-2453.
- Vos P, Hogers R, Bleeker M, et al (1995) AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Reseach* 23:4407–4414. doi: 10.1093/nar/23.21.4407
- Vuillaume B, Valette V, Lepais O, et al (2015) Genetic evidence of hybridization between the endangered native species *Iguana delicatissima* and the invasive *Iguana iguana* (*Reptilia, Iguanidae*) in the Lesser Antilles: Management implications. *PLoS One* 10:e0127575. doi: 10.1371/journal.pone.0127575
- Wasieri J, Schmiedeknecht G, Förster C, et al (2009) Parvovirus Infection in a Eurasian Lynx (*Lynx lynx*) and in a European Wildcat (*Felis silvestris silvestris*). *Journal of Comparative Pathology* 140:203–207. doi: 10.1016/j.jcpa.2008.11.003
- Wayne RK, Shaffer HB (2016) Hybridization and endangered species protection in the molecular era. *Molecular Ecology* 25:2680–2689. doi: 10.1111/mec.13642
- Weber D, Stoeckle T, Roth T (2008) Entwicklung und Anwendung einer neuen wilkatzen Nachweismethode.
- Wedrowicz F, Karsa M, Mosse J, Hogan FE (2013) Reliable genotyping of the koala (*Phascolarctos cinereus*) using DNA isolated from a single faecal pellet.

References

- Molecular Ecology Resources* 13:634–41. doi: 10.1111/1755-0998.12101
- Weir BS, Cockerham CC (1984) Estimating F -statistics for the analysis of population structure. *Evolution* (N Y) 38:1358–1370.
- White LA, Forester JD, Craft ME (2015) Using contact networks to explore mechanisms of parasite transmission in wildlife. *Biological Reviews* 92(1):389–409. doi: 10.1111/brv.12236
- Whiteley, A. R., Fitzpatrick, S. W., Funk, W. C., & Tallmon, D. A. (2015). Genetic rescue to the rescue. *Trends in Ecology & Evolution*, 30(1), 42–49.
- WHO World Health Organization. www.who.int. Accessed 1 Jan 2017
- Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* 163:1177–1191.
- Wirtz P (1999) Mother species-father species: unidirectional hybridization in animals with female choice. *Animal Behaviour* 58:1–12. doi: 10.1006/anbe.1999.1144
- Wolf DE, Takebayashi N, Rieseberg LH (2001) Predicting the risk of extinction through hybridization *Conservation Biology* 15:1039–1053. doi: 10.1046/j.1523-1739.2001.0150041039.x
- Wolfe ND, Dunavan CP, Diamond J (2007) Origins of major human infectious diseases. *Nature* 447:279–283. doi: 10.1038/nature05775
- Woodroffe R, Donnelly CA, Ham C, et al (2016) Badgers prefer cattle pasture but avoid cattle: implications for bovine tuberculosis control. *Ecology Letters* 19:1201–1208. doi: 10.1111/ele.12654
- Woolhouse MEJ (2001) Population biology of multihost pathogens. *Science* 292:1109–1112. doi: 10.1126/science.1059026
- Woolhouse MEJ, Gowtage-Sequeria S (2005) Host range and emerging and reemerging pathogens. *Emerging Infectious Diseases* 11:1842–1847. doi: 10.3201/eid1112.050997
- Woolhouse MEJ, Haydon DT, Antia R (2005) Emerging pathogens: The epidemiology and evolution of species jumps. *Trends in Ecology and Evolution* 20:238–244. doi: 10.1016/j.tree.2005.02.009
- Wright S (1943) Isolation by Distance. *Genetics* 28:114–138.
- Wright S (1951) The genetic structure of populations *Annals of eugenics* 15-323-354.

References

- Würstlin S, Segelbacher G, Streif S, Kohlen A (2016) Crossing the Rhine: a potential barrier to wildcat (*Felis silvestris silvestris*) movement? *Conservation Genetics* 17:1435–1444. doi: 10.1007/s10592-016-0874-x
- Yamaguchi N, Driscoll CA, Kitchener AC, et al (2004) 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, Nussberger B (2015) *Felis silvestris*. In: IUCN Red List Threat. Species. <http://dx.doi.org/10.2305/iUCN.UK.2015-2.RLTS.T60354712A50652361.en>. Accessed 6 Oct 2017
- Zhuo X, Feschotte C (2015) Cross-species transmission and differential fate of an endogenous retrovirus in three mammal lineages. *PLoS Pathogens* doi: 10.1371/journal.ppat.1005279

APPENDICES

Appendix 1

References for the figure 1.2

- Ackermann RR, Bishop JM (2010) Morphological and molecular evidence reveals recent hybridization between *gorilla* taxa. *Evolution* 64:271–290. doi: 10.1111/j.1558-5646.2009.00858.x
- Alda F, Doadrio I (2014) Spatial genetic structure across a hybrid zone between European rabbit subspecies. *PeerJ* 2:e582. doi: 10.7717/peerj.582
- Arnold ML, Meyer A (2006) Natural hybridization in primates: one evolutionary mechanism. *Zoology* 109:261–76. doi: 10.1016/j.zool.2006.03.006
- Arteaga MC, McCormack JE, Eguiarte LE, Medellín RA (2011) Genetic admixture in multidimensional environmental space: Asymmetrical niche similarity promotes gene flow in armadillos (*Dasybus novemcinctus*). *Evolution* 65:2470–2480. doi: 10.1111/j.1558-5646.2011.01329.x
- Berthier P, Excoffier L, Ruedi M (2006) Recurrent replacement of mtDNA and cryptic hybridization between two sibling bat species *Myotis myotis* and *Myotis blythii*. *Proc R Soc B Biol Sci* 273:3101–3123. doi: 10.1098/rspb.2006.3680
- Bogdanov AS, Bannikova AA, Pirusskii YM, Formozov NA (2009) The First Genetic Evidence of Hybridization between West European and Northern White-breasted Hedgehogs (*Erinaceus europaeus* and *E. roumanicus*) in Moscow Region. *ISSN Biol Bull Formozov Izv Akad Nauk Seriya Biol* 36:1062–3590. doi: 10.1134/S106235900906017X
- Cabria MT, Michaux JR, Gómez-Moliner BJ, et al (2011) Bayesian analysis of hybridization and introgression between the endangered european mink (*Mustela lutreola*) and the polecat (*Mustela putorius*). *Mol Ecol* 20:1176–1190. doi: 10.1111/j.1365-294X.2010.04988.x
- Dalton DL, Zimmermann D, Mnisi C, et al (2017) Hiding in Plain Sight: Evidence of Hybridization between Cape Mountain Zebra (*Equus zebra zebra*) and Plains Zebra (*Equus quagga burchelli*). *African J Wildl Res* 47:59–64.
- de Jong YA, Butynski TB (2010) Three Sykes's Monkey *Cercopithecus mitis* x Vervet Monkey *Chlorocebus pygerythrus* Hybrids in Kenya. *Primate Conserv* 25:43–56. doi: 10.1896/052.025.0109
- Goedbloed DJ, van Hooft P, Megens H-J, et al (2013) Reintroductions and genetic introgression from domestic pigs have shaped the genetic population structure of Northwest European wild boar. *BMC Genet* 14:43. doi:

10.1186/1471-2156-14-43

- Hamada Y, San AM, Malaivijitnond S (2016) Assessment of the hybridization between rhesus (*Macaca mulatta*) and long-tailed macaques (*M. fascicularis*) based on morphological characters. *Am J Phys Anthropol* 159:189–198. doi: 10.1002/ajpa.22862
- Hapke A, Gligor M, Rakotondranary SJ, et al (2011) Hybridization of mouse lemurs: different patterns under different ecological conditions. *BMC Evol Biol* 11:297. doi: 10.1186/1471-2148-11-297
- Jones EP, Van Der Kooij J, Solheim R, Searle JB (2010) Norwegian house mice (*Mus musculus musculus/domesticus*): Distributions, routes of colonization and patterns of hybridization. *Mol Ecol* 19:5252–5264. doi: 10.1111/j.1365-294X.2010.04874.x
- Koju NP, He K, Chalise MK, et al (2017) Multilocus approaches reveal underestimated species diversity and inter-specific gene flow in pikas (*Ochotona*) from southwestern China. *Mol Phylogenet Evol* 107:239–245. doi: 10.1016/j.ympev.2016.11.005
- Malukiewicz J, Boere V, Fuzessy LF, et al (2015) Natural and Anthropogenic Hybridization in Two Species of Eastern Brazilian Marmosets (*Callithrix jacchus* and *C. penicillata*). *PLoS One* 10:e0127268. doi: 10.1371/journal.pone.0127268
- McKenzie L, Cooper D (1997) Hybridization between tammar wallaby (*Macropus eugenii*) populations from Western and South Australia. *J Hered* 88:398.
- Nijman IJ, Otsen M, Verkaar ELC, et al (2003) Hybridization of banteng (*Bos javanicus*) and zebu (*Bos indicus*) revealed by mitochondrial DNA, satellite DNA, AFLP and microsatellites. *Heredity* 90:10–16. doi: 10.1038/sj.hdy.6800174
- Oliveira R, Godinho R, Randi E, et al (2007) Molecular analysis of hybridisation between wild and domestic cats (*Felis silvestris*) in Portugal: implications for conservation. *Conserv Genet* 9:1–11. doi: 10.1007/s10592-007-9297-z
- Polyakov A V., White TA, Jones RM, et al (2011) Natural hybridization between extremely divergent chromosomal races of the common shrew (*Sorex araneus*, *Soricidae*, *Soricomorpha*): Hybrid zone in Siberia. *J Evol Biol* 24:1393–1402. doi: 10.1111/j.1420-9101.2011.02266.x
- Randi E, Hulva P, Fabbri E, et al (2014) Multilocus detection of wolf x dog hybridization in Italy, and guidelines for marker selection. *PLoS One* 9:e86409. doi: 10.1371/journal.pone.0086409
- Sankararaman S, Mallick S, Dannemann M, et al (2014) The genomic landscape of

- Neanderthal ancestry in present-day humans. *Nature* 507:354–357. doi: 10.1038/nature12961
- Schröder O, Lieckfeldt D, Lutz W, et al (2016) Limited hybridization between domestic sheep and the European mouflon in Western Germany. *Eur J Wildl Res* 62:307–314. doi: 10.1007/s10344-016-1003-3
- Thompson CW, Stangl FB, Bradley RD (2015) Ancient hybridization and subsequent mitochondrial capture in ground squirrels (*Genus Ictidomys*). *Occasional papers of the Museum of Texas Tech University*, 331 :1-19.
- Zhang P, Han J, Lu Z, Chen R (2014) Molecular evidence of a captive-born intergeneric hybridization between bottlenose and Risso's dolphins: *Tursiops truncatus* x *Grampus griseus*. *Aquat Mamm* 40:5–8. doi: 10.1578/AM.40.1.2014.5
- Zinner D, Groeneveld LF, Keller C, Roos C (2009) Mitochondrial phylogeography of baboons (*Papio spp.*) – Indication for introgressive hybridization? *BMC Evol Biol* 9:83. doi: 10.1186/1471-2148-9-83

Appendix 2

Article describing a panel of 12 microsatellite markers developed by Antagene, for the study of the mountain hare and the European hare. This article was written as part of Antagene's activities.

A single multiplex of twelve microsatellite markers for the simultaneous study of the brown hare (*Lepus europaeus*) and the mountain hare (*Lepus timidus*)

Marie-Pauline Beugin^{1,2}  | Jérôme Letty³ | Cécile Kaerle¹ | Jean-Sébastien Guitton³ | Lina Muselet¹ | Guillaume Queney¹ | Dominique Pontier²

¹ANTAGENE, Animal Genomics Laboratory, La Tour de Salvagny (Lyon), France

²Univ Lyon, Université Lyon 1, CNRS, Laboratoire de Biométrie et Biologie Evolutive UMR5558, F-69622 Villeurbanne, France

³Research Department, National Hunting and Wildlife Agency (ONCFS), Juvignac, Nantes, France

Correspondence

Marie-Pauline Beugin, ANTAGENE, Animal Genomics Laboratory, La Tour de Salvagny (Lyon), France.
Email: mpbeugin@antagene.com

Abstract

The management of hunted species is challenging, as it must conciliate the conservation of species and their sustainable exploitation. Nongenetic tools are widely used in this context but they may present limitations notably when species can hybridize or when large-scale spatial monitoring is required to establish optimal management actions. This is why genetic tools have been more and more integrated in wildlife management practices. However, the markers proposed are often amplified in small multiplexes when larger ones could allow to better cope with the small quantities of DNA obtained with noninvasive sampling methods. Here, we propose a unique multiplex of 12 autosomal microsatellite markers for the study of two hare species that exist in sympatry in some areas in Europe and are hunted notably in France: the brown hare *Lepus europaeus* and the mountain hare *L. timidus*. We tested 17 markers previously used in these two species or other lagomorph species, from which 12 were included in this single multiplex. Diversity was between 4 and 30 alleles per locus totaling 126 alleles, and we showed that these markers possess appropriate genetic resolution for individual and species identification for the populations under study. This multiplex panel represents the largest number of microsatellites amplified in one reaction proposed for these two hare species and provides a cost-effective and valuable tool for further hybridization studies and the management of hares.

KEYWORDS

hares, hybridization, individual identification, microsatellites, species identification

1 | INTRODUCTION

The two most widespread species of hares in Europe are the brown hare (*Lepus europaeus*), present in most parts of Europe, and the mountain hare (*L. timidus*) that can be found in northern Europe and in the Alps. Both species have been classified by the International Union for the Conservation of Nature as "Least Concern" (Smith & Johnston, 2008). In France, the brown hare is one of the most hunted species on

the territory (the national bag is of about 600,000 hares; estimations between 587,080 and 667,207 individuals considering a 95% confidence interval for the hunting season 2013–2014; Aubry et al., 2016), whereas mountain hare hunting is limited to the Alps (the national bag is of several hundred; ONCFS 2011). This hunting dimension adds sustainable exploitation issues to the classic conservation questions and makes management plans even more essential to ensure the viability of hare populations in France. Sustainable management tools, such as

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2017 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

wildlife reserve and limitation of hunting bag or of hunting season, have already been developed for the brown hare and have led to an adequate hunting management of this species (Péroux, Bray, Mauvy, Lartiges, & Marboutin, 2005; Smith, Jennings, & Harris, 2005). In contrast, the impact of hunting on mountain hare populations is uncertain and the direct assessment of the mountain hare demography through in situ observations and using the same tools encounters practical problems due to the harshness of its habitat. Furthermore, morphological differentiation between the two species is error prone due to extensive resemblance and to the possibility of hybridization within contact zones (Thulin, Stone, Tegelström, & Walker, 2006). New tools, based on genetic data, are thus required to improve the specific management of these hare species and to address large-scale issues like the spatial range and structure of populations, their genetic diversity, or the monitoring of individuals over time. These are urgent questions for the mountain hare as climate change may threaten its habitat range, particularly in the Alps (Acevedo, Jiménez-Valverde, Melo-Ferreira, Real, & Alves, 2012; Bisi, Wauters, Preatoni, & Martinoli, 2015; Gobiet et al., 2014).

The genetics of hares species has first been studied using mitochondrial DNA notably in order to unravel the phylogenetic relationships between these species (Alves, Ferrand, Suchentrunk, & Harris, 2003; Stamatis et al., 2009; Wu et al., 2005). Results showed that the history of hares, both past and present, has been influenced by different processes (genetic drift, anthropogenic introductions, and translocations, post-glacial recolonizations from refuges—Kasapidis, Suchentrunk, Magoulas, & Kotoulas, 2005; Suchentrunk et al., 2006; Stamatis et al., 2009) which have promoted encounters between species and the occurrence of hybridizing events (Melo-Ferreira et al., 2007, 2014; Thulin, Stone, et al., 2006; Zachos, Ben Slimen, Hackländer, Giacometti, & Suchentrunk, 2010). As a result, the mitochondrial genome of hares is characterized by multiple introgressions, which led many authors to turn to nuclear markers to identify species or assess levels of ongoing hybridization (Alves et al., 2003, 2006; Estonba et al., 2006). Several types of nuclear markers can be applied and microsatellites, notably thanks to their high polymorphism, allow to reach a high resolution to answer wildlife genetics issues (Morin et al., 2012).

No microsatellite has been specifically designed either for the brown hare or for the mountain hare. However, such markers have already been applied to both species, thanks to the possibility for cross-amplification of microsatellites (Barbará et al., 2007). The microsatellites used in both species were developed for the European rabbit (*Oryctolagus cuniculus*) and the scrub hare (*L. saxatilis*) (Andersson, Thulin, & Tegelström, 1999; Antoniou, Magoulas, Platis, & Kotoulas, 2013; Campos, De Bellocq, Schaschl, & Suchentrunk, 2011; Canu et al., 2013; Djan, Popović, & Veličković, 2014; Estonba et al., 2006; Fickel et al., 2005; de Freitas, 2006; Fulgione, Maselli, Pavarese, Rippa, & Rastogi, 2009; Hamill, Doyle, & Duke, 2006; Melo-Ferreira et al., 2014; Mengoni, Mucci, & Randi, 2015; SurrIDGE, Bell, Rico, & Hewitt, 1997; Thulin, Fang, & Averianov, 2006; Thulin, Malmsten, & Laurila, 2012; Thulin, Stone, et al., 2006; Zachos et al., 2010). None of these studies used more than eight markers in the same multiplex. However,

given the increasing use of noninvasive genetics and the subsequent search for an optimized use of available DNA, larger multiplexes are expected as they use of a lower quantity of DNA (Beja-Pereira, Oliveira, Alves, Schwartz, & Luikart, 2009).

In this study, we developed a 12-microsatellites multiplex and assessed its utility in French populations of brown and mountain hares (1) to validate the specific identification performed based on morphologic data, (2) to assess the power of discrimination between individuals of the same species, and (3) to detect putative clues of an ongoing hybridization between the two hare species within the contact zone.

2 | MATERIALS AND METHODS

2.1 | Samples collecting and DNA extraction

Tissue samples ($N = 532$ for brown hares, $N = 141$ for mountain hares) from hunted hares were collected between 2003 and 2009 in south-eastern France (Figure 1). The sampling area included regions where only the brown hare was present, consisting mostly of plains below 1,500 m altitude (see supplementary, Figure S1). The sampling area also encompassed regions above 2,000 m altitude where mountain hares were predominantly observed. Both species were observed between 1,500 and 2,000 m. Hare species identification was made in the field by hunters. DNA was extracted from ear tissues using purification column kits (Nucleospin 96 Tissue kit, Macherey-Nagel) following the manufacturer procedure.

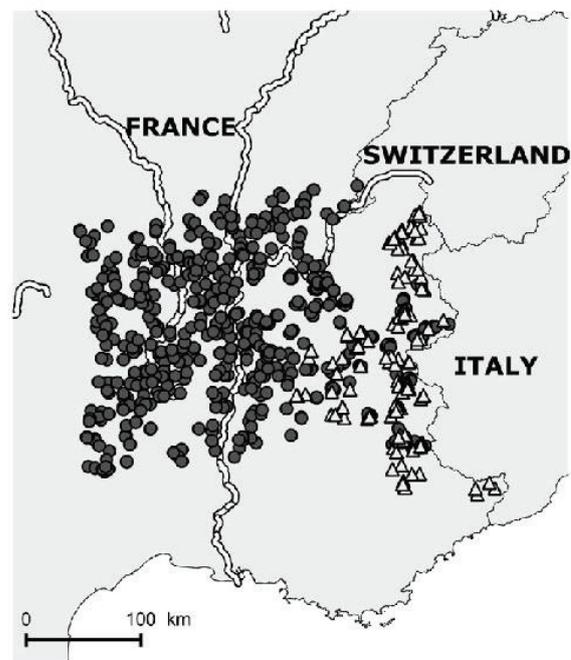


FIGURE 1 Geographical locations of sampling areas. Gray circles and white triangles represent brown and mountain hares, respectively

2.2 | Microsatellite markers selection

We selected 17 microsatellite markers described in the literature: SAT2, SAT3, SAT5, SAT08, SAT12, SAT13 (Mougel, Mounolou, & Monnerot, 1997), SOL30, SOL33, SOL8 (Rico et al., 1994; SurrIDGE et al., 1997), OCELAMB, OCELS1B (van Haeringen, den Bieman, van Zutphen, & van Lith, 1996), D0UTR4 (Korstanje et al., 2001), D7UTR4 (Korstanje et al., 2003) developed for the European rabbit and LSA2, LSA3, LSA6, LSA8 (Kryger, Robinson, & Bloomer, 2002) developed for the scrub hare. Sixteen of them have already been used in both the brown and/or the mountain hare. As for D7UTR4, it has been developed for the European rabbit and have not been applied to any species of hare as far as we know.

We found 23 studies using microsatellites for the study of the brown and/or the mountain hare (see supplementary, Table S1). The 17 markers we tested over the 31 available were ranked among the most used and all had been reported as polymorphic. Overall, no more than eight microsatellites were amplified in a single multiplex (mean = 5.5, SD = 2.23) over the 23 studies.

2.3 | Primers design and multiplex design

A first simplex step was performed in order to check whether all loci were successfully amplified. Only loci correctly amplified were tested for the multiplex construction. A pigtail (Brownstein, Carpten, & Smith, 1996) was added to all reverse primers in order to avoid the addition of adenine nucleotides by the polymerase at the end of the amplification. This first step also allowed us to estimate the allele size range of each locus as well as to assess their polymorphism. The primers of loci with overlapping alleles were dyed with different color fluorescent dyes. At least 10 base pairs (bp; 15–20 bp for loci with high polymorphism) separated loci labeled with the same fluorescent dye. Several adjustment steps followed in order to reduce the number of artifacts on electrophoregrams and to maximize the results readability.

2.4 | Multiplex genotyping

PCR reactions were performed step-by-step following a unidirectional workflow starting in a clear room with positive air pressure to prepare sensitive reagents (enzymes, primers) and then in a pre-PCR room to assemble DNA and reagents. The PCR reaction occurred a primer solution containing 5 μ l of mastermix Taq polymerase (Type-it, QIAGEN), 1.35 μ l of 12 primers pairs at a final concentration of 0.08–0.6 μ mol/L and 30 ng of DNA to be amplified. Each pair of primers was coupled with a fluorescent dye label (details in Table 1). PCR amplifications happened in 96-well microplates in a post-PCR area with negative air pressure. The samples were first denatured at 95°C during 5 min. Then, thirty cycles followed (denaturation step: 95°C, 30 s; annealing step: 55.9°C, 90 s; elongation step: 72°C, 30 s) and a final elongation step at 60°C during 30 min. PCR products were resolved on a capillary sequencer ABI PRISM 3130XL (Applied Biosystem) with formamide (denaturing conditions) and an internal size marker (600 liz; Applied Biosystem) in one migration. The electrophoregrams were analyzed

using GENEMAPPER 4.1 (Applied Biosystem/Life Technologies) independently twice by different operators. Results were then compared, and ambiguous loci were set to missing data.

2.5 | Statistical analysis

2.5.1 | Data selection

The power of resolution for the individual identification was assessed by calculating the probability of identity P(ID) and the probability of identity between siblings P(ID)sib using GenAIEx 6.501 (Peakall & Smouse, 2012). These probabilities also allowed us to determine the minimum number of amplified loci required to have no ambiguity for individual identification. According to Waits, Luikart, and Taberlet (2001), we used 0.01 as threshold for the P(ID). We removed from our dataset any individual typed under the minimum number of loci determined.

Species identification for each individual was determined using the Bayesian software STRUCTURE 2.3.4 (Falush, Stephens, & Pritchard, 2003; Pritchard, Stephens, & Donnelly, 2000). We used an admixture model with correlated allele frequencies between populations and the option POPINFO = 0. The program was run with a Monte Carlo Markov chain length of 1,000,000 after a burn-in of 100,000 iterations, and from one to four genetic clusters ($K = 1-4$). Ten independent runs were carried out for each value of K , and the information from the outputs of the 10 runs for each K was compiled using CLUMPP (Jakobsson & Rosenberg, 2007). We determined the number of clusters (K) that best describe the data following the method of Evanno, Regnaut, and Goudet (2005) implemented in STRUCTURE HARVESTER online web 0.6.94 (Earl & VonHoldt, 2011). We then chose to discard any individual for which the genetic assignment did not match the morphologic identification (individuals had to show a probability of assignment higher than 0.9 to the opposite cluster to which they had been morphologically assigned).

Once the species of each individual was determined, null alleles for each locus in each species were looked for with the software MicroChecker 2.2.3 (Van Oosterhout, Hutchinson, Wills, & Shipley, 2003). The significance of null allele frequencies was then assessed using a binomial exact test following de Meeüs, Béati, Delaye, Aeschlimann, and Renaud (2002). Dropout probability per locus was assessed according to De Meeüs, Humair, Grunau, Delaye, and Renaud (2004). All subsequent analyses were performed with and without the loci for which we detected significant signs of null allele frequencies or dropout. Finally, we tested for deviation from Hardy-Weinberg equilibrium and linkage disequilibrium for each pair of loci separately for brown hares and mountain hares with FSTAT 2.9.3.2 (Goudet, 1995). Significance was assessed applying Bonferroni's correction.

2.5.2 | Detection and characterization of hybrids

In each cluster, we selected the eighty individuals best assigned. All selected individuals reached a posterior probability larger than 0.99. They were then used to simulate parental and hybrid (F1, F2,

TABLE 1 Description of the microsatellite markers of the 12-plex. The accession number in NCBI is given for each locus if the information is available. The pigtailed are reported in bold in the reverse sequence of the primers. The given concentration corresponds to the final concentration of each primer in the primer solution

LOCUS	Accession number	Amorces 5'-3'	Fragment length	Dye	Concentration (μmol/L)	Reference
LSA2	AF491763	F:GGTACTCTATTAGGGAACCCG R: GTGCTT GGCTAGTTGCCATTAGCTCCC	236-254	FAM	0.4	Kryger et al. (2002)
LSA3	AF491764	F:GGATATCAAGAACATGCC R: GTGCTT TGGCTTAAAGGCTGATCTG	199-211	FAM	0.3	Kryger et al. (2002)
LSA6	AF491766	F:CCTAAGATGAATGGATAAGTT R: GTGCTT CTCTCTGTTTTCTGGAGCA	163-171	FAM	0.2	Kryger et al. (2002)
LSA8	AF491768	F:AAGGTATAAATGGGCACT R: GTGCTT GACTGAAATGATGTGCTACC	171-189	VIC	0.6	Kryger et al. (2002)
OCELAMB		F:AGTCACATTTGGCAITTCGTGA R: GTGCTT TTCCTTTTGAATTTAGGATCCACAGC	107-119	NED	0.12	van Haeringen et al. (1996)
OCELS1B		F:TCAGGTATTTGGAAAGTGAATCCC R: GTGCTT TACTGTATATCAAAAGGCATGACCC	143-167	PET	0.3	van Haeringen et al. (1996)
SAT8	X99889	F:CAGACCCGGCAGTTGCAGAG R: GTGCTT GGGAGAGAGGGATGGAGGTATG	88-98	VIC	0.08	Mougel et al. (1997)
SAT12	X99891	F:CTTGAGTTTTAAATTCGGGC R: GTGCTT GTTTGGATGCTATCTCAGTCC	108-136	FAM	0.08	Mougel et al. (1997)
SAT13	X99892	F:GCCTTACCTTTGTGGGG R: GTGCTT TCAGTTTTGAAGGACACCTGC	109-115	VIC	0.2	Mougel et al. (1997)
SOL30	X79215	F:TGCAGCACTTAGTCTCAGGTC R: GTGCTT TCCCGAGCCCAAGATATTGTTACCA	155-201	NED	0.3	Rico et al., 1994; Surridge et al. (1997)
SOL33	X94683	F:GGCCAAATAGTACTGATCCATGT R: GTGCTT GAAGGCTCTGAGATCTAGAT	197-215	PET	0.4	Rico et al., 1994; Surridge et al. (1997)
SOL8	X79217	F:ATCGCACCCATATCTGAGAGAATC R: GTGCTT GGATTGGGCCCTTTGCTCACACTTG	106-122	PET	0.2	Rico et al., 1994; Surridge et al. (1997)

Backcrosses) genotypes with the function *hybridize* implemented in the R package *adegenet* 2.0.1 (Jombart, 2008). A total of 1,800 simulated genotypes were generated and analyzed using NewHybrids 1.1 (Anderson & Thompson, 2002) in order to estimate the resolution power of the 12-plex for the detection of hybridization. We considered that an individual was assigned to a genotypic class when the probability for this class exceeded 0.90 (Godinho et al., 2011).

2.5.3 | Species diversity and genetic structure

Observed and expected heterozygosity, mean number of alleles per locus, number of private alleles were assessed using GenAlEx. Estimations of F_{st} and F_{is} per locus, between and among populations were calculated using Weir and Cockerham's estimators (1984) implemented in GENETIX 4.05.2 (Belkhir, Borsa, Chikhi, Raufaste, & Bonhomme, 1996).

3 | RESULTS

3.1 | Characteristics of the multiplex

The final marker set consisted of 12 multiplexed microsatellite markers (LSA2, LSA3, LSA6, LSA8, OCELAMB, OCELS1B, SAT8, SAT12, SAT13, SOL30, SOL33, SOL8). Of the 17 markers originally considered, two (SAT2 and SAT5) were excluded because their amplification was not satisfactory. SAT3 was also eliminated because of its low polymorphism. Finally, D7UTR1 and D0UTR4 proved impossible to multiplex with the other markers and were thus discarded. Two to four markers were allocated to the same fluorescent dye (see Table 1) with an average spacing of 28 bp between two markers of the same dye.

3.2 | Data selection

According to the values of P(ID) ($2.9 \cdot 10^{-10}$ for the brown hare, $1.5 \cdot 10^{-07}$ for the mountain hare, see supplementary material, Table S2) and P(ID)sib (0.00015 and 0.0013, respectively), individuals could be reliably identified with 12 microsatellite loci. In fact, nine markers were enough to reach the threshold of 0.01 (P(ID)sib for eight

markers equal to 0.0004 and 0.01 for the brown and the mountain hare, respectively).

The most likely number of clusters was $K = 2$. More than 97% of the individuals were assigned to one of the clusters with a posterior probability higher than 0.9; 81% were assigned with probabilities higher than 0.99 (Figure 2). F_{st} estimate indicated significant differentiation between the two clusters ($F_{st} = 0.19$; CI 95%: 0.13–0.25). According to the microsatellite markers, errors of morphologic identification of the species by hunters were detected for 11 individuals. These individuals as well as all individuals genotyped with less than nine markers were removed from the analysis. The final dataset, after PI calculations and STRUCTURE analysis, resulted in 521 brown hares and 122 mountain hares.

Deviation from Hardy–Weinberg equilibrium varied among loci within both species. Some loci that were out of HW equilibrium: SAT12 and LSA8 in the mountain hare, SOL33 in the brown hare, LSA3, SAT13, and SOL30 in both species. There was no linkage disequilibrium between any pair of loci after the Bonferroni correction. Only the locus LSA3 showed significant signs of null alleles. We also found signs of allelic dropout for the loci LSA6 and SAT12. In following analyses, we compared the results using 12 markers and nine markers (LSA3, LSA6, and SAT12 removed). The same 11 errors of species assignation also appeared with the set of nine markers. The posterior probabilities of assignation were similar to the ones obtained with 12 markers (see Figure 2). The F_{st} value calculated over nine markers was slightly higher than what was obtained with the whole panel ($F_{st} = 0.21$; CI 95%: 0.14–0.28). The confidence interval overlapped widely the one obtained with the 12 markers.

3.3 | Detection and characterization of hybrids

The profiles obtained with 12 or nine microsatellites were very similar (see Figure 3). Overall, each hybrid category was mostly assigned to its real class by the computer program. Using 12 microsatellites, the mean assignment of F1 and F2 to parental classes was null. On average, 85.6% of the F1 and 64.6% of the F2 were assigned as such, respectively. Thus, these hybrid classes were reliably detected using the microsatellite panel. On the contrary, the distinction between

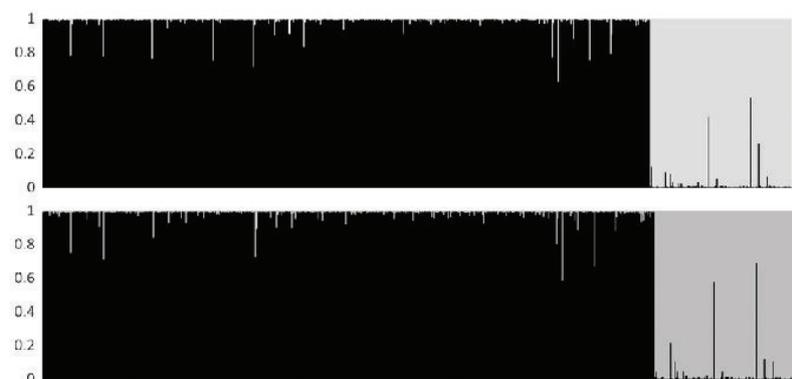


FIGURE 2 Posterior assignment probabilities obtained within the STRUCTURE analysis for the sampled hares with 12 (top plot) and nine (bottom plot) markers and $K = 2$. Brown hare membership is represented in black while mountain hare assignment is in gray. The 11 dubious individuals are not represented

backcross and parental classes was not always clear. The mean assignment of the backcross hybrids to the parental classes was equal to 7.4% on average and, inversely, the mean assignment of parental genotypes to backcross classes reached 5.3% on average. When only nine markers were used, F2 hybrids and backcrosses tended to be considered as parental individuals more often (<1% of F2 hybrids and 12%–16% of the backcrosses). This loss of resolution was expected as we reduced the number of loci.

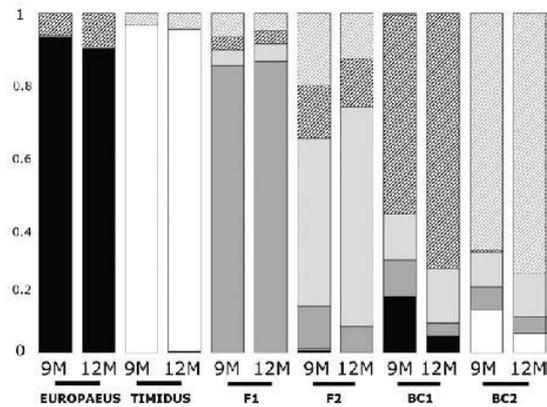


FIGURE 3 Mean probability of assignment of each simulated categories to the six parental and hybrid classes tested with NEWHYBRIDS. Both results obtained with the whole microsatellite panel (12M) and the reduced panel of 9 markers (9M) are represented. Black bars represent brown hares, white bars mountain hares, dark gray stands for F1, light gray for F2, black stripes for F1*brown hare hybrids (BC1), and gray stripes for F1*mountain hare hybrids (BC2)

3.4 | Species diversity and genetic structure

All loci were polymorphic with a mean number of alleles per locus of 10.5 (standard error = 2.19) in brown hare and 6.08 (SE = 1.06) in mountain hare (see Table 2). When considering only alleles present at frequency higher than 5%, the mean numbers were much closer (brown hare: mean = 3.83, SE = 0.53; mountain hare: mean = 2.91, SE = 0.39) as were the numbers of private alleles in the two species when considering frequencies higher than 5% (four for the brown hare, five for the mountain hare). The brown hare showed a lower F_{is} ($F_{is} = 0.11$; CI95%:0.092–0.312) than the mountain hare ($F_{is} = 0.20$; CI95%: 0.15–0.24). Using the panel of nine loci, F_{is} estimates were lower ($F_{is} = 0.078$; CI95%: 0.057–0.097 for the brown hare and $F_{is} = 0.13$; CI95%:0.083–0.18 for the mountain hare). However, the order of magnitude between the two values remained comparable.

4 | DISCUSSION

Our goal was to generate a set of microsatellite markers in brown and mountain hares that could be used cost-effectively to distinguish species and hybrids, and to survey populations and individuals. In this study, we found that 12 markers could be amplified reliably in a single PCR reaction in the two target species.

4.1 | Characterization of the French hare populations and individuals

All 12 loci were polymorphic although two of them (LSA6 and SAT08) showed only one allele present at a frequency higher than 5%. The

TABLE 2 Estimates of the genetic diversity estimators for the brown hare and the mountain hare. N stands for the number of alleles; H_o and H_e are the observed and expected heterozygosity, respectively. The expected frequencies of null alleles are given in the column “Null”. Loci in Hardy-Weinberg disequilibrium are indicated with a “*” for the brown hare and a “#” for the mountain hare. Statistically significant null alleles frequencies are indicated with “***”. The last column indicates dropout p -value for both species

Locus	Lepus europaeus					Lepus timidus					Dropout p -value
	N	F_{is}	H_o	H_e	Null	N	F_{is}	H_o	H_e	Null	
LSA2	18	0.025	0.69	0.71	0.012	13	0.051	0.7	0.73	0.016	.99
LSA3*#	6	0.47	0.35	0.66	0.22**	7	0.54	0.33	0.71	0.26**	.52
LSA6	4	0.074	0.21	0.22	0.031	2	0.66	0.008	0.024	0.082	.028
LSA8#	6	0.073	0.56	0.60	0.029	4	0.55	0.17	0.39	0.22	.67
OCELAMB*	8	0.10	0.43	0.48	0.044	4	0.048	0.28	0.29	0.038	.36
OCELS1B	14	0.0095	0.85	0.86	0.004	12	0.064	0.67	0.72	0.038	.11
SAT08	4	0.040	0.37	0.38	0.015	2	0	0.008	0.008	-0.0041	.073
SAT12#	13	0.022	0.81	0.83	0.0089	7	0.27	0.52	0.7	0.14	.038
SAT13*#	9	0.059	0.74	0.79	0.028	4	0.19	0.5	0.61	0.087	.076
SOL30*#	30	0.095	0.62	0.68	0.047	9	0.18	0.44	0.54	0.10	.32
SOL33*	4	0.37	0.29	0.46	0.17	3	0.13	0.35	0.4	0.057	.092
SOL08	10	0.036	0.77	0.80	0.016	6	-0.015	0.61	0.59	-0.012	.29
Mean	10.5	0.11	0.56	0.62	-	6.08	0.2	0.38	0.47	-	
Variance	2.19	0.0018	0.064	0.057	-	1.08	0.0043	0.068	0.07	-	

substantial null allele frequency that we have found in both species for LSA3 was observed by Mengoni et al. (2015). Thus, in the future studies, it may be relevant to conduct analyses including or excluding this marker even if its impact on the results appears limited especially in our hybridization study. For both species, the estimated mean number of alleles, including rare alleles, fell within the range of values found in other European populations of brown hares (11 in Fickel et al., 2005 and Mengoni et al., 2015 vs. 10.6 in our study) and mountain hares (between 5 and 9 in Hamill et al., 2006; Zachos et al., 2010 and Rehnus & Bollmann, 2016 vs. 6.1 in our study). Such polymorphism made the discrimination of individuals possible based on nine markers over the 12 available in the multiplex. This set of optimized microsatellite markers represents a powerful tool for future noninvasive genetic approaches (e.g., based on feces), often associated with a higher rate of amplification failure. The analysis of a complementary dataset (unpublished data) made of noninvasive samples (feces) of mountain hares ($n = 38$) showed a microsatellite amplification success reaching 97.71%, quite similar to the amplification success obtained in Rehnus and Bollmann (2016) on noninvasive samples (feces). The estimates of allelic dropout (0.17 across loci) and false allele (0.014) were consistent with the ranges obtained in previous noninvasive genetic studies (Frantz et al., 2003; Steyer, Simon, Kraus, Haase, & Nowak, 2012). In particular, this panel may be used to estimate other population parameters such as population size and survival rates when more common tools (such as direct counts or telemetry for example) reveal impractical (Kohn et al., 1999; Lukacs & Burnham, 2005). More widely, this panel of microsatellites should allow investigating the population genetic structures of the two hare species.

4.2 | Specific identification of hares and hybrids detection

The punctual inconsistencies between the morphological and the genetic identification (1.4% and 6.8% of individuals wrongly assigned to the brown hare and the mountain hare, respectively, when considering exclusively individuals sampled in regions where both species are observed) show that morphologically similar individuals can belong to distinct species. The panel of microsatellites might then come to validate morphological identifications, in particular in mountain areas, where the accurate estimation of the species distribution over time is crucial in the context of climate change.

Moreover, this panel allowed to detect reliably first and second generation hybrids, as well as a majority of backcrossed individuals. Hybridization has been reported in several occasions (Thulin, Fang, et al., 2006; Thulin, Stone, et al., 2006) and is expected to happen more and more notably due to the fragmentation of the landscape and the change in populations ranges caused by climate change (Parmesan & Yohe, 2003). Thus, it might represent a growing issue for the conservation of the mountain hare. The present data show signs of hybridization as one individual appears clearly as a hybrid (assigned to several of the hybrid classes of Newhybrids but never to the parental classes, using nine or 12 markers). This individual was located in the contact zone (see supplementary, Figure S2). Four other individuals showed signs of hybridization and were mostly assigned to the backcrossed

categories. However, given the resolution of the microsatellite panel for this class of hybrids, we cannot know for sure whether they are hybrids. The presence of such suspicious individuals reinforces the need for a genetic tool to further investigate to what extent hybridization might represent a threat for the conservation of this species in the Alps. Thus, the 12-plex hereby proposed represents an interesting tool for the quantification of interspecific reproduction events although it proved efficient mainly for the detection of recent events of hybridization. Its resolution power remains limited for older events, and more markers are required to reliably detect backcrosses.

4.3 | Using multiplexes in wildlife management

The multiplexing of several markers is always a challenge because of the interactions between primers that may prevent correct amplification of some loci and potentially lead to genotyping errors such as allelic dropout. However, a careful primer selection and multiple adjustments allow for the coamplification of a substantial number of microsatellites (Guichoux et al., 2011; Hill, Butler, & Vallone, 2009). The 12-plex described in this study is the largest, to our knowledge, to have been proposed for the simultaneous study of the brown hare and the mountain hare. The multiplexing approach is particularly appropriate considering the development of noninvasive genetic studies for which DNA extracted from samples is invariably in extremely small quantities (Beja-Pereira et al., 2009). However, multiplex genotyping panels rarely consist of more than eight markers (Guichoux et al., 2011). Larger multiplexes have been proposed, for example, in humans (26 microsatellites, Hill et al., 2009) or in apricots (20 microsatellites, Campoy, Martínez-Gómez, Ruiz, Rees, & Celton, 2010), but they remain few. The use of next-generation sequencing methodologies should promote their development as the technology allows for the simultaneous detection of many markers (Gardner, Fitch, Bertozzi, & Lowe, 2011), increasing the probability to find compatible microsatellite markers (i.e., markers that do not overlap when attributed to the same dye).

ACKNOWLEDGMENTS

The molecular analyses were funded by the National Hunting and Wildlife Agency (ONCFS): funding contracts 2004/08, 2005/19/6171 and 2009/22/6171 and the CNRS. We are very grateful to hunters and to technical staff of ONCFS and of Departmental Hunting Federations of south-eastern France (SD and FDC of Ain, Alpes-de-Haute-Provence, Hautes-Alpes, Alpes-Maritimes, Ardèche, Drôme, Isère, Loire, Haute-Loire, Lozère, Puy-de-Dôme, Rhône, Savoie et Haute-Savoie) who collected genetic samples for this study. Particularly, we wish to thank P. Auliac, C. Blachier, R. Corti, G. Coursat, P. Crégu, B. Degrange, P. Douvre, F. Étienne, P.-F. Galvin, J. Hureau, N. Jean, F. Normand, C. Randon, C. Rieutort, P. Roche, P. Sibut, D. Siméon, P. Vika, F. Vilain, and S. Zimmermann. We thank Antonio Canu and an anonymous referee for helpful comments.

CONFLICT OF INTEREST

None declared.

DATA ACCESSIBILITY

The data is available from the dryad repository: <http://dx.doi.org/10.5061/dryad.d4368>.

AUTHOR CONTRIBUTION

J-SG and JL designed the study and coordinated field data sampling. LM, CK, and GQ developed the panel of microsatellite markers. MPB conducted the analysis and prepared the initial draft of the manuscript. GQ and DP supervised the analysis and writing of the manuscript. JL and JSJ contributed to the writing of and critically revised the manuscript. All authors approved the final version of the manuscript.

REFERENCES

- Acevedo, P., Jiménez-Valverde, A., Melo-Ferreira, J., Real, R., & Alves, P. C. (2012). Parapatric species and the implications for climate change studies: A case study on hares in Europe. *Global Change Biology*, *18*, 1509–1519.
- Alves, P. C., Ferrand, N., Suchentrunk, F., & Harris, D. J. (2003). Ancient introgression of *Lepus timidus* mtDNA into *L. granatensis* and *L. europaeus* in the Iberian Peninsula. *Molecular Phylogenetics and Evolution*, *27*, 70–80.
- Alves, P. C., Harris, D. J., 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. *Molecular Phylogenetics and Evolution*, *40*, 640–641.
- Anderson, E. C., & Thompson, E. A. (2002). A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*, *160*, 1217–1229.
- Andersson, A. C., Thulin, C. G., & Tegelström, H. (1999). Applicability of rabbit microsatellite primers for studies of hybridisation between an introduced and a native hare species. *Hereditas*, *130*, 309–315.
- Antoniou, A., Magoulas, A., Platis, P., & Kotoulas, G. (2013). Assessing the genetic landscape of a contact zone: The case of European hare in northeastern Greece. *Genetica*, *141*, 23–40.
- Aubry, P., Anstett, L., Ferrand, Y., Reitz, F., Klein, F., Ruetter, S., ... Migot, P. (2016). Enquête nationale sur les tableaux de chasse à tir. Saison 2013–2014. Résultats nationaux. *Faune Sauvage*, *310*(Suppl), 8.
- Barbará, T., Palma-Silva, C., Paggi, G. M., Bered, F., Fay, M. F., & Lexer, C. (2007). Cross-species transfer of nuclear microsatellite markers: Potential and limitations. *Molecular Ecology*, *16*, 3759–3767.
- Beja-Pereira, A., Oliveira, R., Alves, P. C., Schwartz, M. K., & Luikart, G. (2009). Advancing ecological understandings through technological transformations in noninvasive genetics. *Molecular Ecology Resources*, *9*, 1279–1301.
- Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N., & Bonhomme, F. (1996). GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. *Laboratoire génome, Populations, Interactions, CNRS UMR, 5000*, 1996–2004.
- Bisi, F., Wauters, L. A., Preatoni, D. G., & Martinoli, A. (2015). Interspecific competition mediated by climate change: Which interaction between brown and mountain hare in the Alps? *Mammalian Biology*, *80*, 424–430.
- Brownstein, M. J., Carpten, J. D., & Smith, J. R. (1996). Modulation of non-templated nucleotide addition by Taq DNA polymerase: Primer modifications that facilitate genotyping. *BioTechniques*, *20*, 1004–1006, 1008–1010.
- Campos, J. L., De Bellocc, J. G., Schaschl, H., & Suchentrunk, F. (2011). MHC class II DQA gene variation across cohorts of brown hares (*Lepus europaeus*) from eastern Austria: Testing for different selection hypotheses. *Mammalian Biology*, *76*, 251–257.
- Campoy, J. A., Martínez-Gómez, P., Ruiz, D., Rees, J., & Celton, J. M. (2010). Developing microsatellite multiplex and megaplex PCR systems for high-throughput characterization of breeding progenies and linkage maps spanning the apricot (*Prunus armeniaca* L.) genome. *Plant Molecular Biology Reporter*, *28*, 560–568.
- Canu, A., Scandura, M., Luchetti, S., Cossu, A., Iacolina, L., Bazzanti, M., & Appollonio, M. (2013). Influence of management regime and population history on genetic diversity and population structure of brown hares (*Lepus europaeus*) in an Italian province. *European Journal of Wildlife Research*, *59*, 783–793.
- De Meeûs, T., Humair, P. F., Grunau, C., Delaye, C., & Renaud, F. (2004). Non-Mendelian transmission of alleles at microsatellite loci: An example in *Ixodes ricinus*, the vector of Lyme disease. *International Journal for Parasitology*, *34*, 943–950.
- Djan, M., Popović, D., & Veličković, N. (2014). Comparison of allozyme and microsatellite variability in brown hare populations (*Lepus europaeus Pallas*) from Vojvodina (Serbia). *Biologia Serbica*, *35*, 24–30.
- Earl, D. A., & VonHoldt, B. M. (2011). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*, 359–361.
- Estonba, A., Solís, A., Iriondo, M., Sanz-Martin, M. J., Perez-Suarez, G., Markov, G., & Palacios, F. (2006). The genetic distinctiveness of the three Iberian hare species: *Lepus europaeus*, *L. granatensis*, and *L. castroviejoi*. *Mammalian Biology - Zeitschrift für Säugetierkunde*, *71*, 52–59.
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, *14*, 2611–2620.
- Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, *164*, 1567–1587.
- Fickel, J., Schmidt, A., Putze, M., Spittler, H., Ludwig, A., Streich, W. J., & Pitra, C. (2005). Genetic structure of populations of European brown hare: Implications for management. *Journal of Wildlife Management*, *69*, 760–770.
- Frantz, A. C., Pope, L. C., Carpenter, P. J., Roper, T. J., Wilson, G. J., Dalahay, R. J., & Burke, T. (2003). Reliable microsatellite genotyping of the Eurasian badger (*Meles meles*) using faecal DNA. *Molecular Ecology*, *12*, 1649–1661.
- de Freitas, H. M. (2006). *Natural hybridization between the Iberian hare (Lepus granatensis) and the brown hare (L. europaeus) in northern Iberian Peninsula*. PhD thesis. Universidade do Porto.
- Fulgione, D., Maselli, V., Pavarese, G., Ripa, D., & Rastogi, R. K. (2009). Landscape fragmentation and habitat suitability in endangered Italian hare (*Lepus corsicanus*) and European hare (*Lepus europaeus*) populations. *European Journal of Wildlife Research*, *55*, 385–396.
- Gardner, M. G., Fitch, A. J., Bertozzi, T., & Lowe, A. J. (2011). Rise of the machines - recommendations for ecologists when using next generation sequencing for microsatellite development. *Molecular Ecology Resources*, *11*, 1093–1101.
- Gobiet, A., Kotlarski, S., Beniston, M., Heinrich, G., Rajczak, J., & Stoffel, M. (2014). 21st century climate change in the European Alps-A review. *Science of the Total Environment*, *493*, 1138–1151.
- Godinho, R., Llaneza, L., Blanco, J. C., Lopes, S., Alvares, F., Garcia, E. 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.
- Goudet, J. (1995). FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity*, *86*, 485–486.
- Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Léger, P., Lepais, O., ... Petit, R. J. (2011). Current trends in microsatellite genotyping. *Molecular Ecology Resources*, *11*, 591–611.
- van Haeringen, W. A., den Bieman, M., van Zutphen, L. F., & van Lith, H. A. (1996). Polymorphic microsatellite DNA markers in the rabbit (*Oryctolagus cuniculus*). *Journal of Experimental Animal Science*, *38*, 49–57.
- Hamill, R. M., Doyle, D., & Duke, E. J. (2006). Spatial patterns of genetic diversity across European subspecies of the mountain hare, *Lepus timidus*. *Heredity*, *97*, 355–365.
- Hill, C. R., Butler, J. M., & Vallone, P. M. (2009). A 26plex autosomal STR assay to aid Human identity testing. *Journal of Forensic Sciences*, *54*, 1008–1015.

- Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics (Oxford, England)*, 23, 1801–1806.
- Jombart, T. (2008). adegenet: An R package for the multivariate analysis of genetic markers. *Bioinformatics (Oxford, England)*, 24, 1403–1405.
- Kasapidis, P., Suchentrunk, F., Magoulas, A., & Kotoulas, G. (2005). The shaping of mitochondrial DNA phylogeographic patterns of the brown hare (*Lepus europaeus*) under the combined influence of Late Pleistocene climatic fluctuations and anthropogenic translocations. *Molecular Phylogenetics and Evolution*, 34, 55–66.
- Kohn, M. H., York, E. C., Kamradt, D. A., Haught, G., Sauvajot, R. M., & Wayne, R. K. (1999). Estimating population size by genotyping faeces. *Proceedings. Biological sciences/The Royal Society*, 266, 657–663.
- Korstanje, R., Gillissen, G. F., Kodde, L. P., Den Bieman, M., Lankhorst, A., Van Zutphen, L. F. M., & Van Lith, H. A. (2001). Mapping of microsatellite loci and association of aorta atherosclerosis with LG VI markers in the rabbit. *Physiological Genomics*, 6, 11–18.
- Korstanje, R., Gillissen, G. F., Versteeg, S. A., Van Oost, B. A., Bosma, A. A., Rogel-Gaillard, C., ... Van Lith, H. A. (2003). Mapping of rabbit microsatellite markers using chromosome-specific libraries. *Journal of Heredity*, 94, 161–169.
- Kryger, U., Robinson, T. J., & Bloomer, P. (2002). Isolation and characterization of six polymorphic microsatellite loci in South African hares (*Lepus saxatilis* F. Cuvier, 1823 and *Lepus capensis* Linnaeus, 1758). *Molecular Ecology Notes*, 2, 422–424.
- Lukacs, P. M., & Burnham, K. P. (2005). Review of capture-recapture methods applicable to noninvasive genetic sampling. *Molecular Ecology*, 14, 3909–3919.
- de Meeüs, T., Béati, L., Delaye, C., Aeschlimann, A., & Renaud, F. (2002). Sex-biased genetic structure in the vector of Lyme disease, *Ixodes ricinus*. *Evolution*, 56, 1802–1807.
- Melo-Ferreira, J., Boursot, P., Randi, E., Kryukov, A., Suchentrunk, F., Ferrand, N., & Alves, P. C. (2007). The rise and fall of the mountain hare (*Lepus timidus*) during Pleistocene glaciations: Expansion and retreat with hybridization in the Iberian Peninsula. *Molecular Ecology*, 16, 605–618.
- Melo-Ferreira, J., Farello, L., Freitas, H., Suchentrunk, F., Boursot, P., & Alves, P. C. (2014). Home-loving boreal hare mitochondria survived several invasions in Iberia: The relative roles of recurrent hybridisation and allele surfing. *Heredity*, 112, 265–273.
- Mengoni, C., Mucci, N., & Randi, E. (2015). Genetic diversity and no evidences of recent hybridization in the endemic Italian hare (*Lepus corsicanus*). *Conservation Genetics*, 16, 477–489.
- Morin, P. A., Archer, F. I., Pease, V. L., Hancock-Hanser, B. L., Robertson, K. M., Huebinger, R. M., ... Taylor, B. L. (2012). An empirical comparison of SNPs and microsatellites for population structure, assignment, and demographic analyses of bowhead whale populations. *Endangered Species Research*, 19, 1–27.
- Mougel, F., Mounolou, J., & Monnerot, M. (1997). Nine polymorphic microsatellite loci in the rabbit, *Oryctolagus cuniculus*. *Animal Genetics*, 28, 58–71.
- ONCFS (2011). Evolution de la répartition communale du petit gibier de montagne en France. décennie 2000-2009. *Faune Sauvage*, 290 (Suppl.), 16.
- Parmesan, C., & Yohe, G. (2003). A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, 421, 37.
- Peakall, R., & Smouse, P. E. (2012). GenALEX 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics*, 28, 2537–2539.
- Péroux, R., Bray, Y., Mauvy, B., Lartiges, A., & Marboutin, E. (2005). La démographie du Lièvre d'Europe et sa sensibilité à la chasse: évaluation par modélisation. *Faune Sauvage*, 269, 24–32.
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945–959.
- Rehner, M., & Bollmann, K. (2016). Non-invasive genetic population density estimation of mountain hares (*Lepus timidus*) in the Alps: Systematic or opportunistic sampling? *European Journal of Wildlife Research*, 62, 737–747.
- Rico, C., Rico, I., Webb, N., Smith, S., Bell, D., & Hewitt, G. (1994). Four polymorphic microsatellite loci for the European wild rabbit, *Oryctolagus cuniculus*. *Animal Genetics*, 25, 367.
- Smith, R. K., Jennings, N. V., & Harris, S. (2005). A quantitative analysis of the abundance and demography of European hares *Lepus europaeus* in relation to habitat type, intensity of agriculture and climate. *Mammal Review*, 35, 1–24.
- Smith, A. T., & Johnston, C. H. (2008). *Lepus timidus*. *The IUCN Red List of Threatened Species*. Versions 2015.1.
- Stamatis, C., Suchentrunk, F., Moutou, K. A., Giacometti, M., Haerer, G., Djan, M., ... Mamuris, Z. (2009). Phylogeography of the brown hare (*Lepus europaeus*) in Europe: A legacy of south-eastern Mediterranean refugia? *Journal of Biogeography*, 36, 515–528.
- Steyer, K., Simon, O., Kraus, R. H. S., Haase, P., & Nowak, C. (2012). Hair trapping with valerian-treated lure sticks as a tool for genetic wildcat monitoring in low-density habitats. *European Journal of Wildlife Research*, 59, 39–46.
- Suchentrunk, F., Ben Slimen, H., Stamatis, C., Sert, H., Scandura, M., Appollonio, M., & Mamuris, Z. (2006). Molecular approaches revealing prehistoric, historic, or recent translocations and introductions of hares (genus *Lepus*) by humans. *Human Evolution*, 21, 151–165.
- Surridge, A. K., Bell, D. J., Rico, C., & Hewitt, G. M. (1997). Polymorphic microsatellite loci in the European rabbit (*Oryctolagus cuniculus*) are also amplified in other lagomorph species. *Animal Genetics*, 28, 302–305.
- Thulin, C.-G., Fang, M. Y., & Averianov, A. O. (2006). Introgression from *Lepus europaeus* to *L. timidus* in Russia revealed by mitochondrial single nucleotide polymorphisms and nuclear microsatellites. *Heredity*, 143, 68–76.
- Thulin, C.-G., Malmsten, J., & Laurila, A. (2012). Differences in body mass, health status and genetic variation between insular and mainland brown hares (*Lepus europaeus*) in Sweden. *European Journal of Wildlife Research*, 58, 897–907.
- Thulin, C.-G., Stone, J., Tegelström, H., & Walker, C. W. (2006). Species assignment and hybrid identification among Scandinavian hares *Lepus europaeus* and *L. timidus*. *Wildlife Biology*, 12, 29–38.
- Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M., & Shipley, P. (2003). *Micro-Checker*, version 2.2.3. Hull, UK: Department of Biological Sciences and Department of Computer Science, University of Hull.
- Waits, L. P., Luikart, G., & Taberlet, P. (2001). Estimating the probability of identity among genotypes in natural populations: Cautions and guidelines. *Molecular Ecology*, 10, 249–256.
- Weir, B. S., & Cockerham, C. C. (1984). Estimating *F*-statistics for the analysis of population structure. *Evolution*, 38, 1358–1370.
- Wu, C., Wu, J., Bunch, T. D., Li, Q., Wang, Y., & Zhang, Y. (2005). Molecular phylogenetics and biogeography of *Lepus* in Eastern Asia based on mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*, 37, 45–61.
- Zachos, F. E., Ben Slimen, H., Hackländer, K., Giacometti, M., & Suchentrunk, F. (2010). Regional genetic in situ differentiation despite phylogenetic heterogeneity in Alpine mountain hares. *Journal of Zoology*, 282, 47–53.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Beugin M-P, Letty J, Kaerle C, et al.

A single multiplex of twelve microsatellite markers for the simultaneous study of the brown hare (*Lepus europaeus*) and the mountain hare (*Lepus timidus*). *Ecol Evol*. 2017;00:1–9.

<https://doi.org/10.1002/ece3.2943>

Appendix 3

Article presenting a panel of 20 single nucleotide polymorphism markers developed by Antagene. The markers were specifically selected to allow for the detection of hybridization between the domestic pig and the European wild boar. This is an example of highly differentiated marker panels. This article was written as part of Antagene's activities.

A set of 20 multiplexed single nucleotide polymorphism (SNP) markers specifically selected for the identification of the wild boar (*Sus scrofa scrofa*) and the domestic pig (*Sus scrofa domesticus*)

Marie-Pauline Beugin^{1,2} · Eric Baubet³ · Caroline Dufaure De Citres¹ · Cécile Kaerle¹ · Lina Muselet¹ · François Klein³ · Guillaume Queney¹

Received: 5 April 2017 / Accepted: 10 April 2017
© Springer Science+Business Media Dordrecht 2017

Abstract In the last few decades, populations of wild boars have been spreading throughout Europe. This expansion is problematic because of conflicts with human activities and/or sanitary reasons. Thus, the need to develop tools to understand wild boar populations' increase as well as to manage efficiently their populations appears crucial but challenging. As ancestor of the domestic pig, one of the main issues associated with this population spread is the possible hybridization between the two subspecies which complicates the identification of pure individuals. In this study, we propose a set of 20 single nucleotide polymorphism (SNP) markers, allocated in two 10-plexes, selected in order to identify the two subspecies. The multiplexes were validated on reference populations of wild boars and domestic pigs in order to assess their polymorphism and their discriminative power. The selected markers allowed for an exacerbated differentiation between the wild boars and the domestic pigs compared to other studies and thus allows for an efficient identification of species. The low size of this panel of 20 SNP markers makes it a tool compatible

with economic constraints associated with wildlife management compared to tools based on chips.

Keywords Low-size panel · Marker selection · Candidate-gene approach

Introduction

In the last few decades, populations of wild boars have been spreading throughout Europe (Vernesi et al. 2003; Servanty et al. 2010; Massei et al. 2015). In France, the populations have been multiplied by five in 20 years (Saint-Andrieux and Barboiron 2015). This recent population explosion has made of the management of wild boars populations a topic of crucial interest (Scandura et al. 2011). Furthermore, as it may increase the frequency of interactions between wild boars and their close relatives the domestic pigs in farms, it may promote hybridizing events (Vernesi et al. 2003; Goedbloed et al. 2013a). The presence of such hybrids on the field makes difficult the detection of "pure" individuals because hybrids are cryptic and can display a morphology very close to pure morphotypes. Karyotypes (Popescu et al. 1980) have been used to complement morphological information but the possible presence of old generations of hybrids and the events of hybridization between European and Asian domestic pigs to build commercial breeds (Giuffra et al. 2000) considerably lower the resolution power of this genetic tool. Thus, lately, the need to increase the toolbox available to identify and monitor individuals has been pointed out (Scandura et al. 2011).

The wild boar genetics have benefited from the agromonomical value of the domestic pig, thus, a large set of genetic markers has already been developed and applied to wild boars (Giuffra et al. 2000; Vernesi et al. 2003;

Electronic supplementary material The online version of this article (doi:10.1007/s12686-017-0738-9) contains supplementary material, which is available to authorized users.

✉ Marie-Pauline Beugin
mpbeugin@antagene.com

¹ ANTAGENE, Animal Genomics Laboratory, 6 allée du Levant, 69890 La Tour de Salvagny (Lyon), France

² Univ Lyon, Université Lyon 1, CNRS, Laboratoire de Biométrie et Biologie Evolutive UMR5558, F-69622 Villeurbanne, France

³ Office National de la Chasse et de la Faune Sauvage, Unité Cervidés Sangliers, Birieux; Bar le Duc, France

Scandura et al. 2011; Kolodziej et al. 2012; Goedbloed et al. 2013b; Vilaça et al. 2014; Iacolina et al. 2015). Recently, studies took advantage of sequencing technologies to develop single nucleotide polymorphism (SNP) markers. However, the tools currently available, based on these markers, are mostly chips which provide substantial but untargeted information only accessible using costly devices. Here, we developed two SNP markers 10-plexes specifically designed for the identification of the wild boar and the domestic pigs. Such panels are expected to increase the genetic toolbox available for regular monitoring by wildlife managers in order to help them cope with wild boars management issues.

Materials and methods

Sampling and DNA extraction

Two reference populations were sampled in order to build a panel of markers based on true wild boars on one hand, true domestic pigs on the other hand. Wild boar samples were collected on hunted animals in the Parc Naturel des Vosges du Nord, in La Petite Pierre, in France. This population has been monitored since the 1980s by the ONCFS, the French hunting and wildlife agency. Given the isolation of this population and the absence of releases in past decades, the wild boars from La Petite Pierre are considered as “pure”. The pig reference population was obtained from a national campaign of sampling conducted in French butchereries to collect meat pieces from commercial breeds (Landrace, Large White, Piétrain, Duroc). We focused the sampling on commercial breeds (in opposition to rustic breeds) because they have been highly selected to reach agronomical requirements (Wilkinson et al. 2013), especially regarding traits such as prolificacy or tameness, which could continue unbalancing wild boar populations dynamics if introgressed (Rhymer and Simberloff 1996; Lindberg et al. 2007). Overall, 38 wild boars in La Petite Pierre and 43 domestic pigs were sampled. All samples were stored in 90° alcohol at -20 °C then DNA was extracted using column kits (Nucleospin 96 Tissue kit, Macherey–Nagel) and following the manufacturer procedure.

SNP detection and selection

Firstly, we searched for SNP markers using a candidate gene approach. We searched for genes involved in processes regulating traits of agronomical value such as reproduction, prolificacy, body size... because we expect such genes to have been particularly selected to generate domestic pigs breeds and on which we expect

a pig-specific diversification of the DNA across breeds (Wilkinson et al. 2013). One hundred and 22 genes were found in the literature. We selected 36 of them based on the polymorphism of the SNP they were associated to. More precisely, known polymorphic sites in the wild boar were excluded and sites with a clear predominant or a fixed allele in the domestic pig were preferred. These 36 genes contained 42 SNPs (see supplementary, Table S1). For each of them, a fragment between 500 and 750 bp, centred on the known SNP, was amplified over four randomly chosen “individual DNA” (two wild boars and two domestic pigs) and two “DNA pools” (mixed DNA of 10 randomly chosen individuals for each subspecies). All PCR reactions were performed step by step following a unidirectional workflow. The PCR was conducted in a total volume of 10 µl (Gold Taq buffer 1X, MgCl₂ 2.5mM, dNTP 0.05mM, Gold Taq 0.03U/µl, primers 0.4 µM, DNA 3 ng/µl). PCR products successfully amplified were then sequenced using ABI PRISM 3130 XL (Applied Biosystems). Sequences were aligned using Sequencher v4.6. (Gene Codes Corporation) and polymorphisms were detected by eye. A total of 65 informative SNPs, i.e. polymorphic in the domestic pig and/or the wild boar, were detected. In a second step, more SNP markers were selected from sequences obtained in Goedbloed et al. (2013a, see supplementary Table S2) for 88 wild boars and 120 domestic pigs using the Illumina Porcine SNP60 genotyping beadchip (Ramos et al. 2009).

A total of 20 SNP markers was retained applying the following decision rule: one allele had to be fixed in the wild boar (both for individuals and pool) and the other allele had to be predominant or fixed in the domestic pig if possible (again both in individuals and pool). This decision rule led to select the most differentiated markers. The distribution range of the markers was also considered for the selection in order to maximize the chances to successfully build the multiplexes.

SNP genotyping

Primers used for the SNP genotyping are provided in supplementary material (see Supplementary, Tables S3 and S4). The SNP were genotyped by primer extension using the ABI PRISM® SNaPshot™ Multiplex System (Applied Biosystems). We used the QIAGEN Multiplex PCR Taq in a total volume of 10 µl (1nX Mastermix, primer mix—primers concentrations ranging between 0.1 and 0.8 µM in the PCR reaction-, 4 ng/µl of DNA). PCR products were then purified using ExoSap-IT (Affymetrix) and used for the SNaPshot reaction. The SNaPshot products were furthermore analysed using ABI PRISM 3130XL (Applied Biosystems) in one migration with an injection of 1.2 kV/23 s and an internal size marker

Appendices

(GS-120Liz, Applied Biosystems) in formamide (denaturing conditions). All PCR and SNaPshot reactions included positive and negative controls in order to check for the reaction good development. The results were then analysed on Genemapper v4.1 (Applied Biosystem/Life Technologies).

SNP validation and statistical analysis

The validation of the informative SNP markers occurred on the set of 43 domestic pigs and 38 wild boars. The individuals used for the SNP selection are included in this validation set and were used as controls. Deviations from Hardy Weinberg equilibrium and linkage equilibrium were tested using the software Genepop 4.3 (Raymond and Rousset 1995). Allelic frequencies and heterozygosity estimates were computed with Genalex 6.501 (Peakall and Smouse 2012) and Weir and Cockerham's *Fst* (1984) was assessed on Genetix 4.05.2 (Belkhir et al. 1996). A discriminant analysis of principal components (Jombart et al. 2010) was also carried out in order to check for the proper distinction between the wild boar and domestic pig groups using the dedicated function in the package adegenet (Jombart 2008).

Results

We found no significant deviation from Hardy–Weinberg or linkage equilibrium. Overall, the 20 markers reached an *Fst* value of 0.67 (90% confident interval: 0.55–0.76, see Table 1). This high differentiation was obtained both with markers presenting the same predominating allele in both species and two different predominating alleles. Twelve out of the 20 markers presented divergent predominating alleles according to the subspecies. Domestic pigs showed higher levels of heterozygosity. This is consistent with the fact that we preferred markers monomorphic in wild boars populations while they could be polymorphic in the domestic pigs. Overall, the panel allowed for a sharp distinction between wild boars and domestic pigs using the discriminant analysis (Fig. 1).

Discussion

The 20 SNP markers of the panel were highly differentiated compared to other SNP panels used in studies on wild boars (0.15–0.25 in Goedbloed et al. 2013a; 0.15–0.38 in; Pérez-Enciso et al. 2015). The sharp distinction allowed

Table 1 Characterisation of the 20 SNPs used in the two 10-plexes in the wild boar and in the domestic pig

SNP	Fst	Wild boar			Domestic pig		
		MA	He	Ho	MA	He	Ho
SNP2	0.19	A (1.00)	0.00	0.00	A (0.80)	0.32	0.36
SNP4	0.44	A (1.00)	0.00	0.00	A (0.55)	0.50	0.6
SNP12	0.54	G (1.00)	0.00	0.00	C (0.56)	0.49	0.45
SNP15	0.56	T (0.96)	0.08	0.08	A (0.64)	0.46	0.54
SNP23	0.40	G (1.00)	0.00	0.00	G (0.58)	0.49	0.56
SNP27	0.33	G (1.00)	0.00	0.00	G (0.65)	0.45	0.37
SNP28	0.16	C (1.00)	0.00	0.00	C (0.82)	0.29	0.35
SNP33	0.13	T (0.91)	0.17	0.18	T (0.69)	0.43	0.49
SNP39	0.21	G (1.00)	0.00	0.00	G (0.77)	0.36	0.28
SNP41	0.25	A (1.00)	0.00	0.00	A (0.73)	0.39	0.44
SNP43	0.98	G (1.00)	0.00	0.00	A (0.98)	0.05	0.05
SNP46	0.91	C (0.96)	0.08	0.08	T (0.95)	0.09	0.05
SNP48	0.75	C (0.90)	0.19	0.16	A (0.88)	0.21	0.23
SNP50	0.61	C (0.70)	0.42	0.40	T (0.94)	0.11	0.12
SNP52	0.94	C (0.97)	0.05	0.05	T (0.97)	0.07	0.07
SNP53	0.83	A (0.95)	0.10	0.11	G (0.90)	0.19	0.21
SNP54	0.74	A (0.96)	0.08	0.08	G (0.80)	0.32	0.35
SNP55	0.42	G (0.78)	0.35	0.29	A (0.74)	0.38	0.33
SNP56	0.90	C (0.95)	0.10	0.11	T (0.95)	0.09	0.09
SNP57	0.74	T (0.93)	0.13	0.14	C (0.84)	0.27	0.33

MA corresponds to the predominating allele for each locus and sub-species. The frequency of this allele is given between parentheses. Finally, He and Ho correspond to the expected and observed heterozygosity respectively

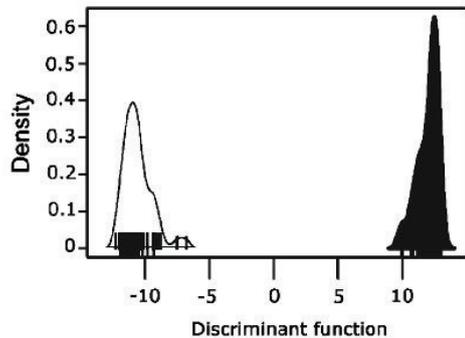


Fig. 1 Representation of the two groups identified with the discriminant analysis of discriminant components. Domestic pigs are represented by the *white distribution* and wild boars by the *black distribution*

by the panel makes it a valuable tool for the management of the two sub-species. Apart from identification issues, it may be useful for the detection of hybrids (Godinho et al. 2014) even though this application was not tested here. On the contrary, it should not be applied to answer general population genetics questions because the selection process used to build the panel may generate biases in statistical analyses.

The primary goal of this panel was to differentiate “pure” wild boars from commercial breeds of domestic pigs. The will to work on “pure” wild boars necessarily limited the number of individuals on which we could develop and test the panel because most populations of wild boars show signs of introgression and only a few are free of restocking (Scandura et al. 2011; Goedbloed et al. 2013b). In future studies, the panel of SNP markers should be applied to more populations of “pure” or “presumed pure” wild boars in order to discard any specific population effect on the high level of differentiation we observed. Also, it should be used on populations which have experienced hybridizing events in order to assess how introgression may have reduced the differentiation between sub-species and to test if this introgression may affect somehow the resolution power of the panel.

Finally, at a time when whole-genome sequencing and wide SNP-chips are increasingly used, we propose a low-size panel of markers. Such approach might puzzle given the actual trend which consists in producing “big data” through the use of new-generation sequencing (NGS) technologies. However, as described in Shafer et al. (2014), the aims in genomics and in applied conservation genetics differ and thus, the tool required in those two fields may not be the same. Plus, questions currently addressed in conservation genetics do not necessarily require the use of all the information produced by NGS (McMahon et al.

2014). Wildlife monitoring is indeed deeply anchored in a socio-economic environment driven by political policies that ask for targeted answers, often in a limited time with limited financial resources. Thus, the huge amount of data produced in genomics is hardly adapted and transposable as it is to applied genetics. Tools must then be built to take advantage of the extraordinary amount of data available thanks to genomics advances and adapt it to conservation genetics’ issues (Monzón et al. 2014). A low-sized panel of markers, chosen and selected in order to answer a targeted question might be an example of bridging tools between genomics and conservation genetics.

Acknowledgements We thank Sébastien Devillard for helpful comments on the manuscript.

References

- Belkhir K, Borsa P, Chikhi L, Goudet J, Bonhomme F (2004) GENETIX 4.05, Windows TM software for population genetics. Laboratoire génome, populations, interactions, CNRS UMR, 5000
- Giuffra E, Kijas JMH, Amarger V et al (2000) The origin of the domestic pig: independent domestication and subsequent introgression. *Genetics* 154:1785–1791
- Godinho R, López-Bao JV, Castro D et al (2014) Real-time assessment of hybridization between wolves and dogs: combining non-invasive samples with ancestry informative markers. *Mol Ecol Resour* 5:2–5. doi:10.1111/1755-0998.12313
- Goedbloed DJ, Megens HJ, Van Hooft P et al (2013a) Genome-wide single nucleotide polymorphism analysis reveals recent genetic introgression from domestic pigs into Northwest European wild boar populations. *Mol Ecol* 22:856–866. doi:10.1111/j.1365-294X.2012.05670.x
- Goedbloed DJ, van Hooft P, Megens H-J et al (2013b) Reintroductions and genetic introgression from domestic pigs have shaped the genetic population structure of Northwest European wild boar. *BMC Genet* 14:43. doi:10.1186/1471-2156-14-43
- Iacolina L, Scandura M, Goedbloed DJ et al (2015) Genomic diversity and differentiation of a managed island wild boar population. *Heredity (Edinb)*. doi:10.1038/hdy.2015.70
- Jombart T (2008) Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24:1403–1405. doi:10.1093/bioinformatics/btn129
- Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet* 11:94. doi:10.1186/1471-2156-11-94
- Kolodziej K, Theissinger K, Brün J et al (2012) Determination of the minimum number of microsatellite markers for individual genotyping in wild boar (*Sus scrofa*) using a test with close relatives. *Eur J Wildl Res* 58:621–628. doi:10.1007/s10344-011-0588-9
- Lindberg J, Björnerfeldt S, Bakken M et al (2007) Selection for tameness modulates the expression of heme related genes in silver foxes. *Behav Brain Funct* 3:18. doi:10.1186/1744-9081-3-18
- Massei G, Kindberg J, Licoppe A et al (2015) Wild boar populations up, numbers of hunters down? A review of trends and implications for Europe. *Pest Manag Sci* 71:492–500. doi:10.1002/ps.3965

Appendices

Conservation Genet Resour

- McMahon BJ, Teeling EC, Höglund J (2014) How and why should we implement genomics into conservation? *Evol Appl* 1–9. doi:10.1111/eva.12193
- Monzón J, Kays R, Dykhuizen DE (2014) Assessment of coyote-wolf-dog admixture using ancestry-informative diagnostic SNPs. *Mol Ecol* 23:182–197. doi:10.1111/mec.12570
- Peakall R, Smouse PE (2012) GenALEX 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* 28:2537–2539
- Pérez-Enciso M, Burgos-Paz W, Ramos-Onsins SE (2015) On genetic differentiation between domestic pigs and Tibetan wild boars. *Nat Genet* 47:190–192. doi:10.1038/ng.3191
- Popescu C, Quéré J, Franceschi P, Boscher J (1980) Observations chromosomiques chez le sanglier français (*Sus scrofa scrofa*). *Genet Sel Evol* 12:395–400. doi:10.1186/1297-9686-12-4-395
- Ramos AM, Crooijmans RPM a, Affara N a. et al (2009) Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. *PLoS ONE*. doi:10.1371/journal.pone.0006524
- Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J Hered* 86:248–249. doi:10.1111/j.0021-8790.2004.00839.x
- Rhymer JM, Simberloff D (1996) Extinction by hybridization and introgression. *Annu Rev Ecol Syst* 27:83–109. doi:10.1146/annurev.ecolsys.27.1.83
- Saint-Andrieux C, Barboiron A (2015) Tableaux de chasse ongulés sauvages Saison 2014–2015. *Faune Sauvag* 308:1–8
- Scandura M, Iacolina L, Apollonio M (2011) Genetic diversity in the European wild boar *Sus scrofa*: Phylogeography, population structure and wild × domestic hybridization. *Mamm Rev* 41:125–137. doi:10.1111/j.1365-2907.2010.00182.x
- Servanty S, Besnard S, Michau F et al (2010) La modélisation démographique: un outil d'aide à la gestion du sanglier? *Faune Sauvag* 288:37–43
- Shafer ABA, Wolf JBW, Alves PC et al (2014) Genomics and the challenging translation into conservation practice. *Trends Ecol Evol* 30:78–87. doi:10.1016/j.tree.2014.11.009
- Vernesì C, Crestanello B, Pecchioli E et al (2003) The genetic impact of demographic decline and reintroduction in the wild boar (*Sus scrofa*): a microsatellite analysis. *Mol Ecol* 12:585–595. doi:10.1046/j.1365-294X.2003.01763.x
- Vilaça ST, Biosà D, Zachos F et al (2014) Mitochondrial phylogeography of the European wild boar: the effect of climate on genetic diversity and spatial lineage sorting across Europe. *J Biogeogr* 41:987–998. doi:10.1111/jbi.12268
- Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution* 38:1358–1370
- Wilkinson S, Lu ZH, Megens HJ et al (2013) Signatures of diversifying selection in European pig breeds. *PLoS Genet*. doi:10.1371/journal.pgen.1003453

