Modification of the structure of wild boar populations by hunting and influence on reproductive processes

Thibault Gayet

N° d'ordre NNT : 2018LYSE1125

THESE de DOCTORAT DE L'UNIVERSITE DE LYON
opérée au sein de
l'Université Claude Bernard Lyon 1

Ecole Doctorale N°341
Evolution, Ecosystèmes, Microbiologie, Modélisation

Spécialité de doctorat : Ecologie
Discipline : Biologie évolutive

Soutenue publiquement le 13/07/2018, par :

Thibault GAYET

Modification of the structure of wild boar populations by hunting and influence on reproductive processes

Devant le jury composé de :

Albano BEJA-PEREIRA, Chercheur, Université de Porto	Rapporteur
Dustin J. PENN, Professeur adjoint, Université de Vienne	Rapporteur
Emmanuelle GILOT-FROMONT, Professeure, LBBE, Université Lyon 1	Examinatrice
Nirmala SEON-MASSIN, Directrice adjointe DRE, ONCFS	Examinatrice
Ludovic SAY, Professeur, LBBE, Université Lyon 1	Directeur de thèse
Eric BAUBET, Ingénieur chef de projet, ONCFS	Co-directeur de thèse
Sébastien DEVILLARD, MCU, LBBE, Université Lyon 1	Invité

UNIVERSITE CLAUDE BERNARD - LYON 1

Président de l'Université	**M. le Professeur Frédéric FLEURY**
Président du Conseil Académique	M. le Professeur Hamda BEN HADID
Vice-président du Conseil d'Administration	M. le Professeur Didier REVEL
Vice-président du Conseil Formation et Vie Universitaire	M. le Professeur Philippe CHEVALIER
Vice-président de la Commission Recherche	M. Fabrice VALLÉE
Directrice Générale des Services	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

Modification of the structure of wild boar populations by hunting and influence on reproductive processes

Thibault GAYET

A thesis presented for the degree of

Doctor of Philosophy

LBBE
BIOMETRIE ET BIOLOGIE EVOLUTIVE

Office National de la Chasse et de la Faune Sauvage

A mon oncle Aimé,
amateur de la bête noire

Foreword

My PhD grant was fund by the *French Hunting and Wildlife Agency* (Office National de la Chasse et de la Faune Sauvage, ONCFS), which is a public institution. Expenses related to my PhD were financially supported by both the *Laboratory of Biometry and Evolutive Biology* (Laboratoire de Biométrie et Biologie Evolutive, LBBE), which is an academic research laboratory, and the ONCFS. The ONCFS is divided in several departments with specific missions. I belonged to the *Expertise and Research Department*, whose missions are to acquire knowledge on wildlife species by conducting field studies and to provide expert counseling and technical support for field managers. Therefore, you will find two popularization articles (in French) in the last two appendices aiming to give feedbacks to partners involved and information to managers. Focusing a little closer, I was part of the *Wild ungulates* team, in the group *Expertise and Management*. My PhD was part of the project aiming to get information on wild boar biology.

My original project was to investigate some of the mechanisms already proposed to explain the demography of wild boar currently observed. It included the disruption of the mating system, that I developed broadly in the thesis, and hybridization with the domestic pig. Initially, we wanted to study the impact of hybridization on the fitness of hybrids. Indeed, hybridization may give fitness advantages to hybrids and their offspring by introducing artificially selected genes of the pig in wild boar. It is suggested that hybrids and their descendant have increased the litter size and/or the individual growth rate for example. In contrary, artificially selected traits (by human) in the pig to increase productivity may be too maladaptive to be maintained in the wild, where resources are limited. Initially, we aimed to investigate if there is a link between the level of hybridization of an individual and its fitness. Sadly, detangling which scenario actually takes place was quickly abandoned because preliminary study of the diagnostic power of genetic tools available gave rather disappointing results. Due to the low level of genetic differentiation between wild boar and domestic pig, identifying hybrids beyond first generation hybrids (direct offspring of a wild boar and a pig) was not possible. However, this study lead to a publication available in first appendix. Also, the analytical skills obtained allowed me to take part of another project aiming to develop a new analytical technic to study hybridization which is in second appendix.

Abstract

The wild boar (*Sus scrofa scrofa*) is a peculiar species. It is an appreciated game species for hunters, a nightmare for farmers and a subject of debate for the society in general. The tenfold increase of the population over the last decades in France and all over Europe, despite increased hunting pressure, generated great human-wildlife conflict. The wild boar is responsible for great economic losses due to damaged crops, vehicle collision, diseases transmission and ecosystem disturbances. Improving management strategies becomes a prime interest to avoid such conflicts, or at least keep them under control. Obtaining information on the species is a first step toward good management strategies. The objective of my work is, in a first part, to characterize the mating system of the wild boar using genetic tools (microsatellite markers) and to identify some parameters influencing the reproductive processes, focusing especially on hunting. The second part focus on the investigation of the influence of the mating system on wild boar life history traits. My researches are based on the study of several populations contrasting in their hunting practices and on longitudinal data of a highly monitored population. The study is based on data collected on wild boars killed by hunting. Genotypes were obtained for pregnant females and their litter and paternity analyses were realized to measure the number of fathers in a litter and estimate multiple paternity rates (proportion of litter sired by more than one father). I was able to show that the mating system is mainly promiscuous (several males mate with several females) contrasting with the polygyny (a dominant male monopolizing a group of females) usually described in this species. Moreover, reproductive processes, estimated by the number of mates of a female and the multiple paternity rates, are influenced by hunting variations in a population. I also showed that number of fathers has positive effect on female fecundity. High rates of multiple paternity together with high genetic diversity were found in a heavily hunted population, suggesting multiple paternity may buffer yearly bottlenecks. However, the increase of number of fathers is not associated with increase of within-litter variation.

Keywords: Harvesting, Mating system, Paternity analysis, Population genetic, Ungulate.

Résumé

Le sanglier (*Sus scrofa scrofa*) est une espèce à part entière. C'est une espèce de gibier particulièrement appréciée des chasseurs, un cauchemar pour les agriculteurs et un sujet de débat pour la société en général. La multiplication par dix des populations au cours des dernières décennies en France et dans toute l'Europe, malgré une pression de chasse accrue, a engendré de nombreux conflits entre les humains et la faune sauvage. Le sanglier est responsable de grandes pertes économiques dues aux dommages aux cultures, aux collisions avec les véhicules, à la transmission de maladies et aux dégradations des écosystèmes. L'amélioration des stratégies de gestion devient un intérêt majeur pour éviter, ou contrôler, de tels conflits. La récolte d'informations sur l'espèce problématique est un premier pas vers de bonnes stratégies de gestion. L'objectif de mon travail est, dans un premier temps, de caractériser le système d'appariement du sanglier à l'aide d'outils génétiques (marqueurs microsatellites) et d'identifier certains paramètres influençant les processus de reproduction, notamment la chasse. Dans un deuxième temps, mon travail se concentre sur l'étude de l'influence du système d'appariement sur les traits d'histoire de vie du sanglier. Mes recherches sont basées sur l'étude de plusieurs populations contrastées dans leurs pratiques de chasse et sur des données longitudinales d'une population intensivement suivie. L'étude est basée sur des données recueillies sur des sangliers tués à la chasse. Les génotypes ont été obtenus pour les femelles gestantes et leur portée et des analyses de paternité ont été réalisées pour mesurer le nombre de pères dans une portée et estimer les taux de paternité multiples (proportion de portées engendrées par plus d'un père). J'ai été en mesure de montrer que le système d'appariement est principalement de la promiscuité (plusieurs mâles s'accouplent avec plusieurs femelles) contrastant avec la polygynie (un mâle dominant monopolisant un groupe de femelles) habituellement décrite chez cette espèce. De plus, les processus de reproduction, estimés par le nombre de partenaires d'une femelle et les taux de paternité multiples, sont influencés par les variations de chasse dans une population. J'ai aussi montré que le nombre de pères avait un effet positif sur la fécondité des femelles. Des taux élevés de paternité multiple et une grande diversité génétique ont été constatés ensemble dans une population fortement chassée, ce qui suggère que la paternité multiple peut tamponner les goulots d'étranglement annuels. Cependant, l'augmentation du nombre de père n'est pas associée à une augmentation de la variation intra-portée.

Mots-clés : Analyse de paternité, Génétique des populations, Ongulés, Prélèvement, Système de reproduction.

Acknowledgments

First, I want to warmly thank the members of my PhD defense jury for evaluating my work to make it (and me) better. Many thanks to Albano Beja-Perreira and Dustin Penn for agreeing to review my manuscript and to Emmanuelle Gilot-Fromont and Nirmala Seon-Massin for accepting to be present for the defense.

Je suis évidemment très reconnaissant à mes chefs Éric, Ludovic et Sébastien pour m'avoir encadré pendant près de cinq ans maintenant, et avoir permis à cette thèse de prendre forme. Entre mon manque de confiance et ma timidité hiérarchique, je ne crois pas vous avoir facilité la tâche... Éric, je te remercie pour m'avoir régulièrement sorti de derrière l'écran pour mettre la main à la patte -sans mauvais jeu de mot- et de m'avoir permis de finir la thèse dans les meilleures conditions. Ludovic, merci pour avoir toujours répondu présent et pour avoir géré l'ensemble des manip' labo (et leurs aléas) car tout ça m'a évité pas mal de stress. Sébastien, merci pour l'intégration au projet *snapclust* avec Monsieur adegenet, T.Jombart, j'ai vraiment apprécié participer au développement d'une technique avec une telle équipe. Merci à tous les trois pour votre confiance.

Je tiens également à remercier Marlène pour avoir chaperonné une bonne partie de mes travaux. Merci pour m'avoir permis de travailler avec Christophe Pélabon et de découvrir l'environnement de travail à la NTNU, je garde un très bon souvenir de l'intermède Norvégien, aussi bref qu'intense !

Quelque peu formel, mais non moins sincère, je remercie l'ONCFS pour avoir assuré le fonctionnement de cette thèse de bout en bout, que ce soit mon financement ou celui des analyses, mais aussi pour m'avoir permis d'intégrer l'Unité Ongulés sauvages (CNERA CS à l'époque !). Je remercie tou·te·s les membres de cette équipe pour leur bonne humeur lors des réunions où les soirées tenaient plus du dîner entre copains que de la réunion de collègues. Je remercie François Klein et Eric Marboutin pour leur investissement de même qu'Ophélie et Sandrine, mais également William, Sonia, Maryline, Christine pour les bons moments passés. Bien-sûr, je n'ai pas oublié l'équipe de Châteauvillain. Merci à Serge, qui répondait patiemment à toutes mes questions durant la tournée des pièges, le sanglier de A à Z, je suis très content d'avoir commencé ma thèse quand tu faisais encore partie du groupe, j'espère que tu profites de ton repos dans le Jura. Merci aussi à Henry et Cyril pour les bons moments passés aux Essarts.

Côté LBBE, je tiens également à remercier Pascale et Sabrina, tou·te·s les doctorant·e·s n'ont pas la chance d'avoir des gens investis dans le bon déroulement d'une vie d'équipe et chaque présentation était l'occasion de conseils très constructifs. Je remercie le pôle administratif et informatique pour le soutien technique durant ces quelques années.

Il me serait difficile de nommer toutes les personnes, que ce soit du milieu professionnel ou de ma vie personnelle, qui ont participer, chacune à leur façon, à la concrétisation de cette thèse, mais puissent ces quelques mots suffire à leur exprimer ma gratitude. Je suis convaincu que ce sont les petites choses, qui au fond, m'ont permis de tenir jusqu'à la fin.

Je remercie bien évidement tous ces gens qui se sont frottés à la Recherche (on ne sait pas toujours de quoi d'ailleurs) entre les murs du LBBE.

Il y a celles/ceux rencontré·e·s avant même le début de la thèse dans un bureau surpeuplé (2014, never forget) et qui sont restés bon gré mal gré à mes côtés jusqu'à la fin. Merci à Nico (*King* Kiki, coucou Marion au passage) pour tes plantes et autres lance-bonhommes, tous ces bons moments passés et votre soutien (le manuscrit qui passe à 4 pages… Tachycardie, Hyperventilation). Je remercie également les Masternelles, un grand merci à Pauline, Thérèse, Marine, Agathe et Mélissa pour les séances de décompressions collectives au-delà des frontières, j'ai trop gole-ri.

Merci à tou·te·s celles/ceux qui m'ont (sup)porté. Célia, un grand merci pour toutes tes petites attentions psychopo et ces moments à apprécier l'instant ou les longues discussions à refaire le monde, tout ça a beaucoup compté pour moi. It maters ! Merci Pierre, pour les zitch et tout ce qui allait autour. Merci au duo de choc de la pause-café : Elodie, pour savoir me mettre des coups de pieds aux fesses, la tête sur les épaules et d'avoir partagé le blues des analyses génétiques, et Louise, pour les bons plans de Loulou et pour me rappeler que "c'est comme ça et c'est pas grave". Merci Mickaël pour les bons moments autour d'une mousse, les parties de fléchettes et les guppys !

Je remercie mes co-bureaux, qui ont subi mes sautes d'humeur (soupirs -surtout dernièrement- et noms d'oiseaux lancés à R), d'hier Marion R., Nicolas L., Jérémie, Coraline et d'aujourd'hui M-P (tes dessins et tes chasses à la mouche), Séb (ta bonne humeur) et Elise (tes gifs). Merci pour les discussions au-dessus de l'écran et les batailles à lancer de gommes et autres.

Bien entendu, je n'ai pas oublié tou·te·s les autres, merci à toutes et à tous pour les discussions plus ou moins éclairées pendant les pauses repas, les sorties "on boit juste une verre", le terrain Chizé (Chizé... x( ), et tant d'autres bons moments improvisés, on aura bien ri ~~(rire tue)~~ : la matriarche Morgane, Vérane, Jenn', Tim', Eliane, Sylvie, Victor, Valentine, Salomé, Pierre M., Laura, Kamal, Elisa, et les permanents Etienne, Laurent et les allumées du terrain Jeanne et Sylvia et les petits stagiaires qui ont récidivé (Rebecca, Florentin).

Je remercie à celles et ceux d'une époque plus lointaine : les quelques Lyonnais·es temporaires qui m'ont assurées quelques sorties loin des fous cités plus haut : Carine, Célia (bisous à Ezra), Constance, Johan, Soso, Calypso et les autres d'ici et d'ailleurs : Elodie G., Clara, Elodie P., Clément, Laurène, Kéké.

Merci Mélanie, mon amie de toujours. S'il y a une vingtaine d'année, dans la classe unique de Boissey et ses 20 élèves, on nous avait dit qu'on en arriverait là...

Je remercie bien sûr la famille, Nelly, Guillaume, (& Bogey) pour leur soutien moral et mes parents et grands-parents qui ont toujours veillés à ce que je ne manque de rien, surtout de nourriture, et qui m'ont soutenu sans forcément comprendre en quoi consistait ce que je fais "Ils vont toujours bien tes sangliers ?". MERCI !

Pour finir en beauté, merci à Mathieu pour ta patience et ton soutien et pour avoir affronté la tempête thèse avec philosophie.

Contents

List of figures

List of tables

Chapter I

Introduction

In this chapter, I introduce a general context of my thesis. I start by presenting problems arising from overabundant species. Then, I present theoretical context of causes explaining multiple male mating behavior of females. I further describe known consequences of multiple paternity ensuing from such behavior by focusing on studies about mammals and birds. The aim of my thesis was to investigate modifications of wild boar reproductive processes in relation to hunting and understanding the evolutionary consequences on the species, focusing especially on multiple paternity.

General context

Biotic homogenization refers to the actual trends of the decline of multitude of species in favor of great increase in distribution range and density of few species sharing special traits like high fecundity and variability, broad diet and adaptability to human disturbances (McKinney and Lockwood 1999; Jeschke and Strayer 2006). It is often associated to invasiveness of non-indigenous species introduced by humans in new biota that outcompete or directly prey on local species. However, environmental disturbances can also alter ecological equilibrium and promotes local species able to face, or even benefiting from, the changes (Garrott et al. 1993). They are the so-called "Winners" of Baskin (1998), impacting negatively other species. Their reproduction and/or expansion are increased due to the changes and they become overabundant, meaning they affect human well-being, reduce density of other species and/or cause ecosystems dysfunctions (Caughley 1981). Overabundance often results of anthropogenic modifications of the environment, as the focus species take advantage over less adaptable ones in the modified landscape (Garrott et al. 1993). To avoid detrimental effects of such species considered as pests, management strategies aiming at controlling population growth are implemented. Variety of techniques were applied ranging from culling with traps or poison (Bosch et al. 2000), increased hunting when the species is a game species (Vercauteren et al. 2011; Koons et al. 2014), to fertility control considered more ethical in order to respond to social pressure (Adderton Herbert 2004) or combination of several technics (Cooper and Herbert 2001). Contrasted results were obtained due to high adaptability of the species or compensatory mechanisms (Bosch et al. 2000; Cooper and Herbert 2001; Simard et al. 2013). However modeling approaches showed that management strategies can be improved by adapting culling to respond to population fluctuations but it requires knowledge about the focused population of the species (Chee and Wintle 2010).

The success of overabundant species is associated to life history traits that promote their demography. They show high fecundity due to large litter size and frequent reproduction events (Capellini et al. 2015), broad diet (Jeschke and Strayer 2006), good dispersal abilities (Hulme et al. 2008) and high plasticity allowing them to tolerate environmental variations and establish in new habitats (Rosecchi et al. 2001). These traits are the most commonly identified ones that help species to reach locally abundant densities, but also to increase their distribution range. Moreover, during the invasion process, in addition to phenotypic plasticity, species may display significant evolutionary changes that favor their adaptation to newly colonized habitats (Shine 2012). Other species-specific factors can influence both life history traits and evolutionary

changes, particularly the mating system. Indeed, mating system can affect genetic diversity, which influences the rate of occurrence of evolutionary changes, and fecundity. Many studies addressed these topic in plants (Sun and Ritland 1998; Rambuda and Johnson 2004; Barrett et al. 2008), probably due to the wide range of mating systems existing in plants. However, studies in animals remain scarce, especially mammals, although mating system is known to influence their demography (Holman and Kokko 2013).

Monogamy, polygamy, multiple paternity

Across species, a wide range of behaviors exists to find a mating partner and access to reproduction. These behavioral strategies displayed by individuals of a species to obtain mates define its mating system (*sensu* Emlen and Oring 1977). It includes the number of mates of an individual, the way it acquires mating, the social interactions between the pair of individuals and the parental cares. Several classifications were proposed over time depending on the ability to monopolize mates directly or indirectly by protecting ressource (Emlen and Oring 1977), on the social bonds between partners (Davies 1991), on the number of mates (Shuster and Wade 2003), or a combination of several of these parameters (Clutton-Brock 1989; Shuster and Wade 2003). Shuster and Wade (2003) identified 12 major groups themselves divided in a total of 42 subcategories, in order to describe precisely each mating system. If this allows a fine description of the mating system of a species based on its social and behavioral characteristics, it does not influence the outcomes of the reproduction. Hereafter, for a matter of clarity, we will only consider mating systems based on the number of mates (Box 1). Indeed, whether a male mates with several females because he was able to monopolize a group of females, a territory including several female territories or because he was the most successful male in a lek, he still reproduces with several females and the variance in number of mating partners is higher between males than if each male reproduce with one female (variance of 0) or if all males reproduce with all females (Figure I 1).

It is interesting to highlight that these classifications are mostly based on behavioral observations. With the rise of genetic studies, due to lower costs of molecular techniques, this classification became subject of debate. Initially, more than 90% of bird species were considered to be monogamous (strictly, *i.e.* lifelong paring of two individuals, or sequentially) as couple were observed to provide biparental cares to their offspring (Lack 1968). However, in the end of the last century, evidences of extra-pair copulations started to undermine this conclusion (Birkhead 1987). In their review, Griffith et al. (2002) showed that monogamy

occurs in 14% of the species and only 25% of species previously considered monogamous. They obtain the diametrically opposed conclusion of Lack (1968) that monogamy is more the exception than the rule. Similarly, mammal species were described as mainly polygynous but genetic analyses tend to show that the reproductive processes are more complicated (Clutton-Brock 1989; Soulsbury 2010). Low ranking males also access to reproduction and observed mating behavior can be a poor predictor of male reproductive success (Coltman et al. 1999; Rus Hoelzel et al. 1999). This lead to distinguish between social mating system (the mating system identified by behavioral studies) and genetic mating system which corresponds to the measured number of partners based on successful mating producing offspring. Each mating system, social or genetic, depends on ecological parameters depending on species-specific traits and on the environment.

Initially, it was accepted that reproductive success increases more for males than for females with the number of mating partners (Bateman 1948). From males point of view, monogamy was puzzling and questioned male advantages to mate with only one female. Latter, Emlen and Oring (1977) stated that monogamy occurs if mating partners are not able to monopolize any other individuals of the opposite sex. This is the case when (i) individuals are very dispersed in space due to large territories and/or resource scattered in the environment (no ecological potential for polygamy) or (ii) because life history traits of the species prevent the acquisition of additional mating despite environmental potential (biological constraints for polygyny). The later depends on traits such as the synchrony of estrus in females or biparental care requirement for the offspring. Indeed, many birds, reproducing in large colonies (ecological potential for polygamy or promiscuity) and producing only one offspring (monotocous) like greater flamingo (*Phoenicopterus roseus*) or northern garnet (*Morus bassanus*), display social monogamy. Often, if not always, long biparental cares are required to raise the brood (species-specific constraints) and the desertion of a parent would reduce the probability of successful reproduction (Bart and Tornes 1989). Thus, losing the brood due to desertion of a parent would be costly considering the time and energy invested to find a mate and reproduce, except if the defecting parent did not sire the brood (Hamilton 1964; Trivers 1972). Consequently, biparental cares are maintained in species where the risk of cuckoldry is low (Griffin et al. 2013). Overall, fitness of parents is increased if the couple stays together to raise their brood.

Box 1: Classification of mating systems based on number of mates, synthetized from Davies (1991) and Shuster and Wade (2003).

Monogamy: Each sex mates with a single partner for life. When partner changes between reproductive events, this is called serial monogamy or sequential polygamy.

Polygamy: Males and/or females mate with varying number of partners divided in three main types:

- **Polygyny**: Females mate permanently with a single male, males mate with varying number of females.

- **Polyandry**: Males mate permanently with a single female, females mate with varying number of males.

- **Polygynandry***: Males mate exclusively with several females and vice versa.

Promiscuity*: Males and females mate with any females and any males respectively without any pair bond (random mating).

***Both polygynandry and promiscuity define multiple-partner mating in both sexes, however the notion of social bond is implied in polygynandry (higher probability of mating with some individuals than others). It is not the case for promiscuity where mating is all individuals have the same probability to mate with any individuals of the other sex).

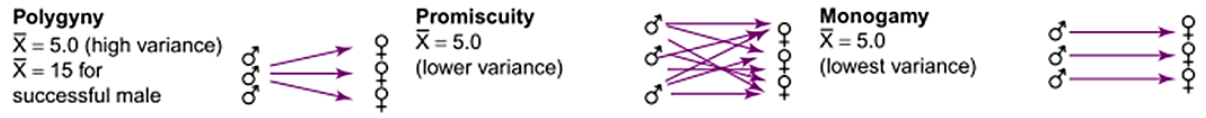


Figure I 1 Means and variances of fitness of polygamous, promiscuous and monogamous males assuming that the mean litter size is five. Mean fitness is five for females in all cases (from Wolff and MacDonald (2004)).

Among other traits, mammals are characterized by mammary glands, developed in females, producing milk to feed their young. The high proportion of mammal species displaying polygynous mating system (around 90%, see Clutton-Brock 1989) was explained by the fact that parental cares can be assumed by females alone. As male are relieved from paternal cares, their reproductive success will mainly depend on the number of females they will reproduce with. This lead to high competition between males to access reproduction and thus increased of sexual selection (Wade 1979). This is especially true if operational sex-ratio (relative proportion of males to females available for mating during a breeding event (Emlen and Oring 1977)) is balanced or biased toward females. Males that are able to monopolize females and reproduce obtain high reproductive success (Figure I 1), while other males who do not reproduce at all obtain a null reproductive success (Wade and Shuster 2004). This is why social polygyny is

associated with a high variance in male reproductive success and a lower one for female, as all females can be fertilized by few males. Indeed virtually, only one male produces enough spermatozoids to fertilize all available females. However, as observed by Clutton-Brock (1989) mating systems are not as fixed as their definitions suggest and several strategies tend to co-exist, which was later confirmed by genetic analyses. Reproductive processes are influenced by individual decisions that may change with time (Gowaty and Hubbell 2009). That is why, despite their attempt to monopolize females, dominant males do not necessarily sire all offspring of all females he mated with because females may mate with several males (Coltman et al. 1999; Heckel et al. 1999; Rus Hoelzel et al. 1999).

Independently of the mating system, when females of polytocous species (species producing more than one offspring at a reproductive event) mate with several males, they can produce litters/broods sired by more than one male. This is true only if spermatozoids from several males lead to successful fecundation. The occurrence of such litters is multiple paternity, and by definition, it does not exist in monotocous species as one offspring can only have one father. We will not consider sequential polygamy of monotocous species as multiple paternity. Progenies produced over the lifetime of monotocous females include half-siblings (sibling with the same mother and different fathers) and, sometimes, some full-siblings (same mother and same father) but there is only one father per reproductive event. Also, by definition, multiple paternity should not occur in monogamous and polygynous species as females mate with only one male. On the other hand, females from polytocous species displaying social polyandry mating systems should mainly produce multiple sired litters/broods (considering all mating produce at least one offspring). It is now clear that mating system are more complex than their definitions. When multiple paternity occurs in monogamous species, it is often referred as extra-pair paternity (but for monogamous monotocous species, it also includes offspring sired by a male which is not the social mate). It is noteworthy that when considering social polygyny strictly, in theory all offspring produced by all females monopolized by a single male are half-siblings (same father, different mothers). Offspring produced by females monopolized by other males are not siblings. When multiple paternity occurs or in polyandrous and promiscuous species, the half-sibling relationships get more complicated as offspring from different females can share the same father and offspring from a male can be found in different females while some offspring are full-siblings. As one male can sire the whole offspring of a female for a breeding event, multiple paternity was subject of numerous studies to understand why this observation is so widespread in animals (Griffith et al. 2002; Uller and Olsson 2008;

Eccard and Wolf 2009). The following parts will focus on explaining why and what are the consequences of multiple paternity in mammals.

Proximal causes of multiple paternity

The trivial answer to the question 'Why does multiple paternity exist?' would be 'because some females mate with several males'. While this answers the question, it does not give any insight to understand this observation. Thus, investigating the causes of multiple paternity is equivalent to understanding why females engage in multi-male mating. Indeed, this behavior is thought to be costly for females due to energetic losses and increased risks of contracting sexually and non-sexually transmitted diseases due to increased contact with conspecifics (Parker and Birkhead 2013). However, on a meta-analysis including 48 species, Lemaître and Gaillard (2013) showed that multi-male mating has no cost on female mortality, acknowledging the fact that such study does not allow to detect if costs in some species are offset by benefits in others.

Over the years, several causes for multi-male mating were identified and review studies attached to detail them (Wolff and MacDonald 2004; Parker and Birkhead 2013). Darwin (1871) already noted that males are more eager to mate than females and, especially in mammals, they pursue them to do so. This behavior was later described as sexual coercion and mainly occurs when females are not guarded by males. Females engage in mating with several males because resisting sexual harassment is energetically costly. This strategy is referred as convenience polyandry. Several studies showed that females are lead to exhaustion by males chasing them or are directly injured by males attempting to mate (Garshelis et al. 1984; Réale et al. 1996; Endo and Doi 2002). When female mammals do not protect themselves, they may still engage in multi-male mating to protect their offspring (Klemme and Ylönen 2010). Indeed, infanticide is widespread in mammals (Hrdy 1979). There are several advantages for males to kill unrelated offspring, including, but non-exhaustingly, that cannibalism can provide food resources for the perpetrator, decreases competition for its own offspring (reduction of maternal allocation to offspring from another male or their absence also mean more resources for its offspring when weaned), and also because lactation is often associated with anestrus (Trivers 1972; Hrdy 1979). Thus, the death of its offspring may lead the female to be receptive again for mating. As killing its own offspring induces great costs for a male (waste of time and energy allocated to reproduction for example), female may solicit several males to mate with them in order to create an uncertainty of paternity. In several species, females were observed to mate

while already pregnant confirming that the behavior is adaptive to prevent infanticide, but obviously, this will not induce multiple paternity (Van Noordwijk and van Schaik 2000). However, multiple paternity may appear when this behavior is expressed during estrus. Female may also mate with multiple males in order to secure their own reproductive success. Indeed, males may be sterile due to genetic defect (Wu et al. 1996), but they may also suffer from sperm depletion if they mate with a lot of females in a short period of time, frequent when estrus are synchronized, or if the sex-ratio is strongly biased (Preston et al. 2001; Milner-Gulland et al. 2003). It is noteworthy that, for domestic animal husbandry, guidelines are set for the different species concerning semen collection for artificial insemination to maintain high fecundation rates (Schilling and Vengust 1987; Leboeuf et al. 2000). The time recommended between two sampling often exceed 24h proving that sperm quality is not optimal after few mating (e.g. maximum of 3 samplings a week for the pig, Frangež et al. 2005). Moreover, it was shown that males can modulate the quantity of sperm provided in their ejaculate, suggesting they could preserve themselves with a female in favor of another one (Wedell et al. 2002). By mating with several males, females increase their chance to mate with at least one male able to guaranty the fecundation of all ovules. Marginally, multi-male mating can favor the offspring survival if it increases the number of males providing material benefits or paternal care, but both traits are rather rare among mammals (Wolff and MacDonald 2004). Moreover, and, as shown in birds, this may induce more costs due to desertion of the social father than benefits provided by other males. Finally, it was suggested that females mate with several males by chance, when they meet several males during their receptive period, if no cost is associated with multiple mating, resulting in random mating (Sutherland 1985; Hubbell and Johnson 1987). Kokko and Mappes (2013) also suggested it can be advantageous for females to mate with every males they meet if the probability of finding a mate is low. Indeed, in such ecological context, it would be very costly to refuse a mating opportunity and die virgin, so multi-male mating could be favored.

Evolutionary consequences of multiple paternity

Independently of its causes, multiple paternity is not neutral and may be the source of evolutionary changes in populations. By mating with several males and reducing the variance in their reproductive success, female allows greater contribution of males to the next generation. Multiple paternity allows transmission of more genetic diversity from a generation to another than single-male mating (Sugg and Chesser 1994; Pearse and Anderson 2009). This is true at the population level, but it is also verified at individual scale. It is suggested that female engage

in multi-male mating to increase the genetic diversity of their offspring, as it is often associated with fitness benefits (Thonhauser et al. 2016). As female tend to reduce the variation of their reproductive success by increasing the variability of the offspring to maximize their fitness, this process is often referred as 'genetic bet-hedging' (Fox and Rauter 2003; Holman 2016). Multiple paternity may also increase female fitness in species where sperm competition occurs. Male-male competition for reproduction is not over when they mate with a female. It goes on in female genital tract to fecund the ovule. The male with the sperm of better quality (concentration and velocity of spermatozoa) outcompetes males with lower sperm quality and has a higher reproductive success (Preston et al. 2003; Malo et al. 2005). In species where multiple paternity is frequent, relative size of the testes is higher (Soulsbury 2010), and increase of testes size is associated with higher production of sperm in order to maximize fecundation and the paternity share (the proportion of offspring produced by a male in a litter). The male winning the sperm competition should sire a larger proportion of the litter. Females can also influence the competition by removing sperm of less desirable male after mating to maximize the chance of being impregnated by favored males (female-female mating behavior are reported in the pig to expell ejaculate, see Aguilera-Reyes et al. 2006). As females choose after copulation to bias paternity in favor of some preferred males, this strategy is often associated as 'good gene' effect on offspring fitness. Multiple paternity was also associated with increased litter size for females. Indeed, studies found that multiple sired litters tend to be larger than single-sired litters (Hoogland 1998; Stockley 2003). This can be linked to multiple mating of female trying to overcome sperm depletion. Thus, multiple-male mating can influence life history traits of females and could be a factor enhancing population dynamic, in particular in invasive species.

Context of the study

The wild boar (*Sus scrofa*) raises serious concerns due to its proliferation in all countries where it is present. Also, it is an important game species which was initially hunted for meat consumption and it still represents an important food resource for some human communities (Sales and Kotrba 2013). In Europe, even if wild boar meat is still consumed, it is mainly a sport hunting species where big males are appreciated for their tusks used as trophy. Hunting bags from all European countries show clear positive evolution of the number of wild boars killed since 1980 (Figure I 2), with five years mean population growth rates above 1.40 over the study period (Massei et al. 2015). Only in France, from 35,893 wild boars shot in 1973,

when the French Hunting and Wildlife Agency started recording wild boar hunting bags, the number increased progressively to the highest value ever recorded of 693,613 in 2016.

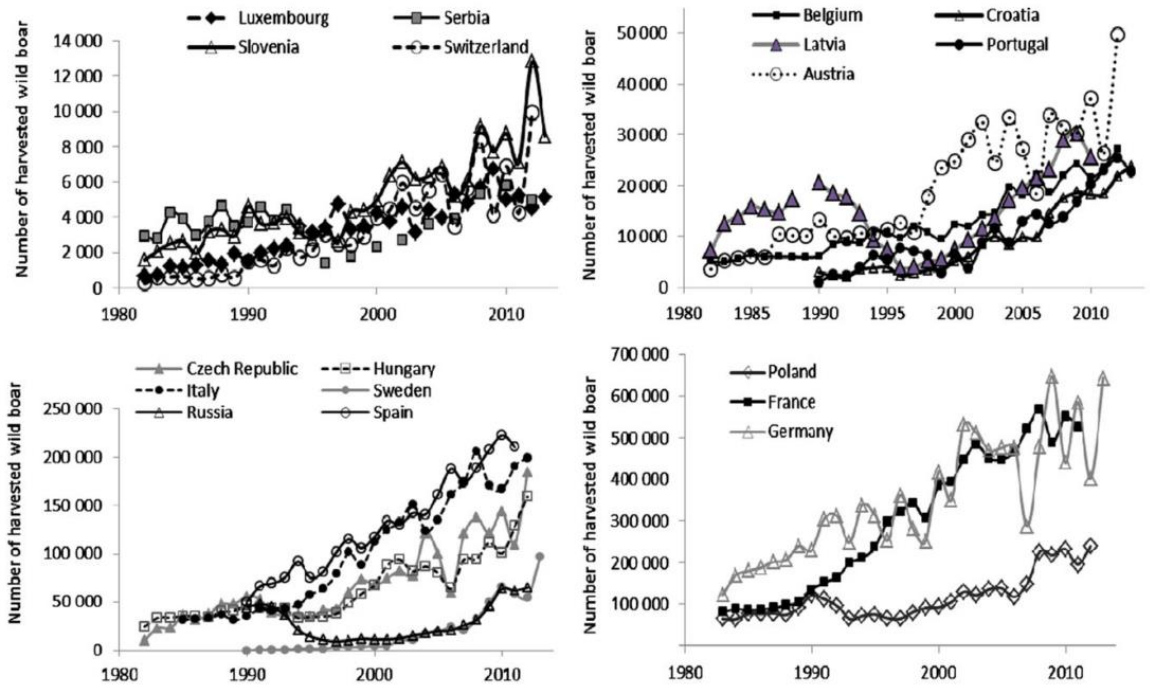


Figure I 2 Evolution of wild boar hunting bags from selected European countries (from Massei et al. 2015).

Such demographic increase leads to important human-wildlife conflicts. The species can cause severe damages in a variety of agricultural crops such as maize, wheat, grapes and potatoes (Schley and Roper 2003). Damage occur more frequently in cultivated fields located near forest (Calenge et al. 2004) and may lead to important economic losses for farmers (Pimentel et al. 2005; Linkie et al. 2007; Schley et al. 2008). Costs of compensation associated to damaged crops range from more than €500.000 for small countries such as Luxembourg or Slovenia to €32 million in France (Massei et al. 2015). Grassland are not spared. If rooting behavior displayed by wild boar foraging for underground food resources can have a positive effect on plant richness and diversity when moderate, it strongly reduced plant cover and alter soil properties when too intensive (Massei and Genov 2004; Barrios-Garcia and Ballari 2012; Bueno et al. 2013). Increase in wild boar number also raises health concerns as high densities are associated with higher contact rates and disease transmission (Rossi et al. 2005; Acevedo et al. 2007). Moreover, the species carries numerous diseases, as reservoir or as host, susceptible to be transmitted to other animal species, especially to pigs (leading once again to great economical losses for farmers) and then indirectly, or directly, to humans (Ruiz-Fons et al. 2008). Indeed, wild boar-human contacts greatly increased over the last decades. Sightings of

wild boars roaming in urban areas become more and more common, including in big cities like Berlin or Barcelona (Cahill et al. 2012; Stillfried et al. 2017). As hunting is forbidden close to habitations, wild boars identify these protected areas and tend to concentrate locally (Tolon et al. 2009). These concentrations of individuals often lead to conflicts when wild boars degrade private gardens and public parks while foraging for food or when they are responsible of collisions with vehicles, which may lead to people injury or death. Costs associated to these accidents are very high and were estimated to be on average more than €45 million per year in Spain and over €100 million in France (Vignon and Barbarreau 2008; Sáenz-de-santa-maría and Tellería 2015).

The ecological and economic impact of the high number of wild boar does not need to be demonstrated any further. It is noteworthy that all these disturbances lead to societal conflicts, mostly involving hunters. Farmers suffering from great economic losses blame hunters for inadequate hunting practices that aim to maintain high densities of games species. Hunters argue in response that damage would not occur if farmers manage their lands better by fencing their crops for example (Storie and Bell 2017). Debates also occur within society in a broader scale, especially in recently recolonized areas. They are nourished by news articles that often focus on the negative effects of wild boar, especially the danger it represents to humans when attacked (Goulding and Roper 2002; Van Herzele et al. 2015). Rare accidents may happen raising public concerns. Generally, hunters are accused of releasing farm animals not afraid of human and facilitating overpopulation to kill for pleasure, while conservationists denounce senseless behavior of people and proliferation of game species due to the lack of predators (Van Herzele et al. 2015). Despite the critics, recreational hunting have been proven to, if not decrease, at least regulate wild boar population growth locally (Geisser and Reyer 2004; Quirós-Fernández et al. 2017), especially when harvesting is wisely planned, and slight increase of hunting effort targeting the most influential individuals can have significant demographic consequences (Gamelon et al. 2012). However, even high hunting mortality is not always enough and heavily hunted populations can still display positive growth rates (Toïgo et al. 2008). Moreover, the number of hunters is decreasing and they are aging in most European countries and even with increased of wild boar harvested *per capita*, the wild boar population keeps growing (Massei et al. 2015). Altogether, the disturbances raise the wild boar as an ecological, economic and societal problematic species.

Reasons for wild boar proliferation have already been investigated and some were identified explaining this demography. First of all, wild boar is peculiar among ungulates. It has

the highest reproductive potential of them all, together with a rather low natural adult mortality (Massei et al. 1996; Bieber and Ruf 2005; Toïgo et al. 2008; Keuling et al. 2013). Gamelon et al. (2011) also showed that wild boar adapts to high hunting pressure by advancing their reproduction allowing progeny to reproduce at one year of age. Moreover, mild winters allow to maintain better body condition leading to increased winter survival in adults. With climate change, mild winters tend to be more frequent making the environment more favorable for the species (Jedrzejewski et al. 1992; Vetter et al. 2015). Survival and reproduction are also favored by change in agricultural practices, improving food resources availability, and rural desertion, reducing human disturbances (Sáez-Royuela and Tellería 1986; Schley and Roper 2003; Massei and Genov 2004). Also, female fertility is increased for hybrids where hybridization with the domestic pig occurs, as pig selected traits genes are introduced in wild boar (Fulgione et al. 2016). To face such an adaptable species, improving management strategies is of great importance. As explained before, the first step to good management strategies is good knowledge of the focus species.

Aim of the thesis

The European wild boar is hunted all over its distribution range but its demography does not seem influenced despite locally high hunting efforts. Hunting is known to induce great changes on structure and genetic characteristics of populations (Harris et al. 2002; Allendorf and Hard 2009). Understanding how wild boar reacts and adapts to the high hunting pressure is of prime interest. In this species, the hunting targets mostly big males for their trophies. Males grow overdeveloped canines all along their life, so oldest males have the longest teeth and are the most favored individuals of hunters (Kierdorf et al. 2004). In heavily hunted populations, such males can be rare because, firstly, intensive hunting leads to reduced survival and very few males reach old age, and, secondly the few individuals reaching old age are easily recognized and preferentially harvested. During my thesis, I investigated if the removal of such males by hunting disrupts the polygynous mating system. As big males are removed from the populations, female monopolization should decrease and their number of mating partners increase, reducing variance in male reproductive success as the mating systems tend toward promiscuity (Figure I 1).

To measure how removal of males from the population influences reproductive processes and to understand how reproduction is shared among remaining males, the study was realized at two different scales. Firstly, a large scale where I compared five populations of wild boars

contrasting mostly in their hunting pressure. The contribution of males to reproduction was estimated from proportions of litters sired by more than one male (multiple paternity rates) and from number of fathers within litters (Chapter IIIA). Secondly, at a narrower scale, as longitudinal study is not possible in all populations, I focused on one of them which was already monitored for several years. This allowed me to evaluate between year variations of the contribution of males to reproduction and to include in the analysis other factors, especially yearly variations of resources availability (Chapter IIIB). Proportions of big males in hunting bag were also registered as it is inversely linked to the number of big males remaining in the population. I expected that mating system disruption should increase with the intensity of hunting and the proportion of big males killed at hunting. Thus, I predicted higher multiple paternity rates and number of fathers within a litter in the most heavily hunted populations and/or years. Concerning the influence of resources on the mating systems, I expected that years of high resource availability would favor female monopolization by males as female groups are more concentrated. As part of my work, I also tried to identify the consequences of the increase of number of sires on life history traits of the species. As explained before, multiple paternity is known to be a female strategy to increase their fecundity and to produce more diverse genetically, and thus, phenotypically offspring (genetic bet-hedging). These hypotheses were tested by measuring the relation between the number of fathers and the number of fetuses within a litter (Chapter IVA) and their phenotypic variations (Chapter IVB). Finally, as fathers of the fetuses are known from previous analyses, I also investigated if their genetic characteristics and body mass were linked to their mating success and reproductive success (number of partners and number of offspring respectively, Chapter V). Biggest males are the most competitive and should be the most able to monopolize females and father high number of offspring. Also, males displaying higher genetic diversity and higher differentiation from the mother, allow females to diversify their litters increasing the fitness of their offspring. They should be favored in detriment of males with lower genetic diversity. Thus, both the number of partners and offspring should increase with male body weight and individual heterozygosity.

Chapter II

Material and methods

In this chapter, I will present information that are common to the following chapters. I start by the general presentation of the wild boar species, including its classification, distribution and biology. Then, I give a description of the study sites with some details about the monitoring realized as part of the data collection for my thesis in the different wild boar populations. I finish with information about the genetic markers used in this study and the number of genotyped individuals. The genetic and statistical analyses description will be provided in each corresponding chapters.

Wild boar

Classification

The wild boar (*Sus scrofa*) belong to the *Suidae* family which is part of the order *Cetartiodactyla* (Figure II 1). The species includes 16 wild subspecies recognized by the world conservation union IUCN, distinguished by geographical and morphological characteristics (Oliver and Leus 2008; Keuling et al. 2017). They are divided in four main groups (Table II 1). The domestic pig can also be added because it is the domestic subspecies (*Sus scrofa domesticus*). Surprisingly, the number of chromosomes varies in the species, from 2n=36 in Europe to 2n=38 in central Europe, Asia and in the pig (Fang et al. 2006). Despite the variation in chromosome number, hybridization between the different subspecies is possible and offspring are viable and fertile.

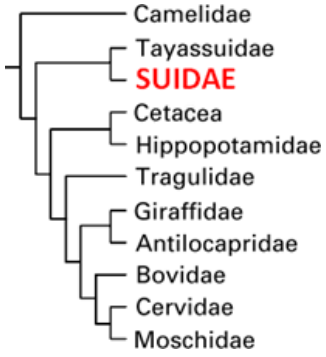


Figure II 1 Supertree representing the phylogenic classification of cetartiodactyl from Price et al. (2005)

Table II 1 Repartition of the 16 subspecies in 4 groups of *Sus scrofa* according to IUCN (Oliver and Leus 2008)

	Group			
	Western	**Indian**	**Eastern**	**Indonesian**
Subspecies	*S. s. scrofa*	*S. s. davidi*	*S. s. sibiricus*	*S. s. vittatus*
	S. s. meridionalis	*S. s. cristatus*	*S. s. ussuricus*	
	S. s. algira	*S. s. affinis*	*S. s. leucomystax*	
	S. s. attila		*S. s. riukiuanus*	
	S. s. lybicus		*S. s. taivanus*	
	S. s. nigripes.		*S. s. moupinensis*	

Distribution

The wild boar is one of the most widely distributed mammals (Massei and Genov 2004). Initially, its native range spreads through Eurasia and Middle-East. Wild boar was lead to extinction in British Isles and Scandinavia over the 17th century but it was then reintroduced

and is now recovering (Booth 1981; Rosvold and Andersen 2008). Similarly, over the last centuries, it colonized all other continents due to intentional introduction by human and/or escape from farms (Figure II 2), whether it was with the classic wild form (*S. scrofa scrofa*), the domestic pigs (*S. s. domesticus*) that form feral population or hybrid swarm of those two, like in Australia and USA (Gabor et al. 1999; Dexter 2003). It was able to adapt to a wide variety of habitats and climates starting from temperate forests, to latter colonized areas ranging from semi-arid taiga to tropical forests (O'Brien et al. 2003; Barrios-Garcia and Ballari 2012; Bengsen et al. 2014). This shows the great plasticity of the species.

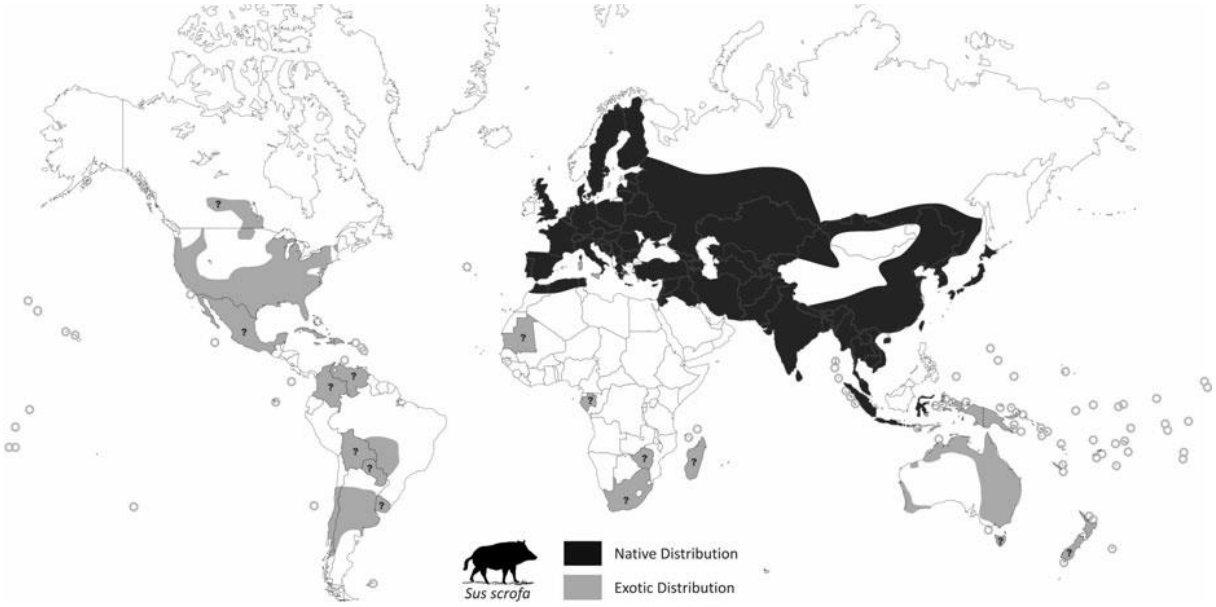


Figure II 2 Worldwide distribution of *Sus scrofa* (wild boar and feral pigs). The species native range demarked in black and introduced range in gray. Gray circles indicate the islands where *S. scrofa* have been introduced. (?) denotes occurrence but unknown distribution (from Barrios-Garcia and Ballari (2012)).

Biology

The wild boar is a medium size and sexually dimorphic ungulate. The body weight ranges from 35 to 350 kg for biggest subspecies and a height varying between 55 et 110 cm, but high variations exist depending of the environment (Spitz et al. 1998; Powell 2004). In the population of Châteauvillain, wild boar mean adult weight is 72 ± 11 kg for females and 102 ±16 kg for males (Toïgo et al. 2008). It can live up to 10 years but its life expectancy is often greatly reduced by hunting (Jezierski 1977; Toïgo et al. 2008). The structures of wild boar populations reported in the literature often show high proportions of young individuals as hunting occurs

everywhere, with few or complete absence of old individuals (Fernández-llario and Mateos-quesada 2003; Herrero et al. 2008). The mortality during the first months of life can be high leading to relatively low juvenile survival (Náhlik and Sandor 2003; Bieber and Ruf 2005). Except hunting, the main causes of wild boar mortality are vehicle collisions, disease, starvation, especially where snow cover limits foraging in winter, and finally predation by wolves where they co-occur (Jedrzejewski et al. 1992; Okarma et al. 1995).

The wild boar is not a territorial species, but females are philopatric. They live in small matriarchal groups of close relatives with their piglets and/or yearlings (Dardaillon 1988; Kaminski et al. 2005; Podgórski et al. 2014). The composition of the group can change throughout the year especially during the breeding season, when male wild boars usually solitary join female groups for mating (Mauget 1980). The size of home range of an individual or group varies depending of the resource in the forest, the time of the year and hunting. Its size ranges from few dozen hectares to few thousands (Maillard and Fournier 1995; Massei et al. 1997; Keuling et al. 2008a). The home range is wider during hunting period which also often corresponds to the mating period of the species. Females tend to remain in, or close to, the group where they are born while males have a higher probability of dispersing. Moreover, when they disperse, males disperse on average further than females (mean of 3.8km *versus* 1.6km, Keuling et al. 2010). It is mainly a nocturnal species, with highest activity around sunset, but as for its home range, its behavior may change depending on factors such as the time of year and the hunting pressure (Russo et al. 1997; Powell 2004; Keuling et al. 2008b).

Diet

The wild boar is omnivore. It is an opportunistic feeder which includes a wide variety of food in its diet depending of available resources in its environment and in time (Figure II 3). This eases its establishment and colonization of new habitats. Vegetable matters are the most important part of its diet, with great proportion of acorn and beechnut during years of high production (Schley and Roper 2003). Maize can also represent a significant part of the alimentation as it is often used for supplementary feeding to maintain wild boars in the forest when forest fruit production is low, thus protecting crops (Calenge et al. 2004). Food from animal origin is also consumed such as earthworms, eggs, small animals (rodents, birds, amphibian, etc…). Wild boar can also scavenge on bigger carcasses, including conspecific carcasses, and also, predatory behaviors on fawns of wild ungulates and livestock have been reported (Ballari and Barrios-García 2014). Around cities, it was also observed to forage on

garbage and food for pets (Cahill et al. 2012). The wild boar shows great plasticity concerning its diet which allowed it to settle in very different environments as explained above. From continental forests with high food availability, it also adapts to environments like Mediterranean forests where oaks are scarce, leading to very low forest fruits availability (Massei et al. 1996). It is noteworthy that the main influential food items remains forest fruits in forest habitat (Geisser and Reyer 2005) and pulsed productions of acorn and beechnut influence reproductive outputs of females (Gamelon et al. 2017). Also, breeding strategy changes between the two populations studied and depends on the kind of forest fruits produced, highlighting again the great plasticity of the species.

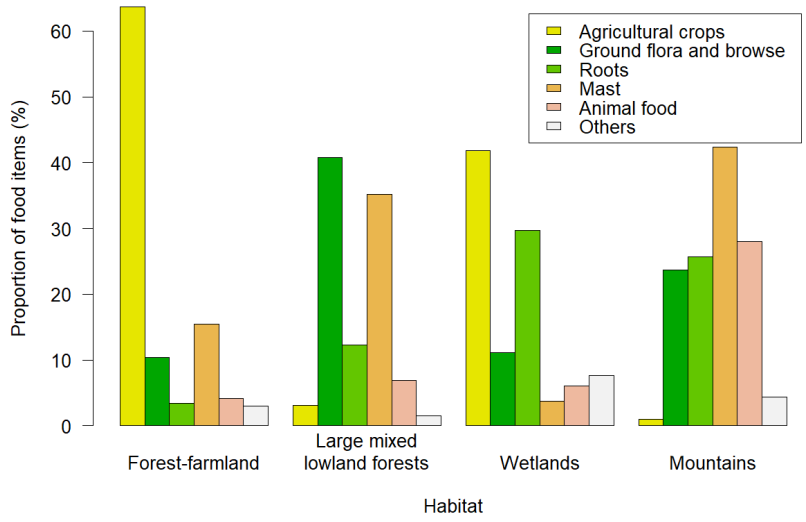


Figure II 3 Proportions of food items in the autumn-winter (rutting period) wild boar diet from various habitats (adapted from Keuling et al. 2017).

Reproduction

As suggested by the sexual dimorphism, the mating system of the wild boar is polygynous, with males competing to monopolize a group of females showing synchronized estrous cycle (Graves 1984; Delcroix et al. 1990). Males are sexually mature and start producing sperm at 7 months but usually do not access reproduction before 3 years while females can start reproduce around one year of age when they reach about 37% of their adult body mass (Mauget and Boissin 1987; Servanty et al. 2009). Also, the species adapts to the hunting with females starting to reproduce earlier in life when the hunting pressure is high (Herrero et al. 2008; Gamelon et al. 2011). The rutting period spreads from October to January with a pic around mid-December (Kozdrowski and Dubiel 2004). Farrowing occurs 115 days after mating (3 months, 3 weeks, 3 days), in a nest constructed by the female (Baubet et al. 2009), so a birth

pic generally occurs around mid-April. The number of piglets produced per female averages 5 but it also depends on the mother age, its body weight and the population (Bieber and Ruf 2005; Servanty et al. 2007; Bywater et al. 2010; Gamelon et al. 2013b). Neonatal phase is poorly known (survival, stillbirth) because disturbances can induce a desertion of the litter by the mother (Baubet et al. 2009). Between 2 weeks and 6 months piglets grow fast (around 100g/day), without difference of growth rate between males and females (Gaillard et al. 1992). They are weaned at 4 months old but females remain in the group while males leave at 14 months of age (Jensen and Recén 1989; Kaminski et al. 2005).

Data collection

Châteauvillain-Arc-en-Barrois

The main wild boar population of the study is from the 11,000ha forest of Châteauvillain-Arc-en-Barrois (48°02′N; 4°55′E) in North-East of France. This population is monitored by the French Hunting and Wildlife Agency (Office National de la Chasse et de la Faune Sauvage (ONCFS)) for over almost four decades, particularly for capture-mark-recapture and spatial uses studies. The forest clump is divided in two areas: an 8,500ha surface of national forest from where samplings come from (red surrounded area, Figure II 4) and a 2,500ha part of private and communal forest. In this population, around 600 wild boars are killed annually (Figure II 5). The climate is intermediate between continental and oceanic, and characterized by mild winters and cool summers. The mean monthly temperature ranged from 1.9 ± 2.1°C in January to 18.9 ± 1.8°C in July (Météo France) while the average monthly precipitation was 75.1 ± 37.5mm over the study period (2007-2016). The forest is mainly composed of oak (*Quercus petraea*, 41%) and beech (*Fagus sylvatica*, 30%) and surrounded by agricultural fields. Acorn and beechnut production fluctuates greatly among years, with a year of very high production followed by several years of low production (Liebhold et al. 2004). Wild boars favored these food resources when available, but they can also cause severe crop damages, especially years of low production. To maintain wild boar in the forest and avoid agricultural damages, hunters spread maize in the forest as supplementary food resource. Also, most part of the forest is surrounded by electric fences to prevent wild boars from going in surrounding crop fields. Fences are also used around the corn crop close to the forest.

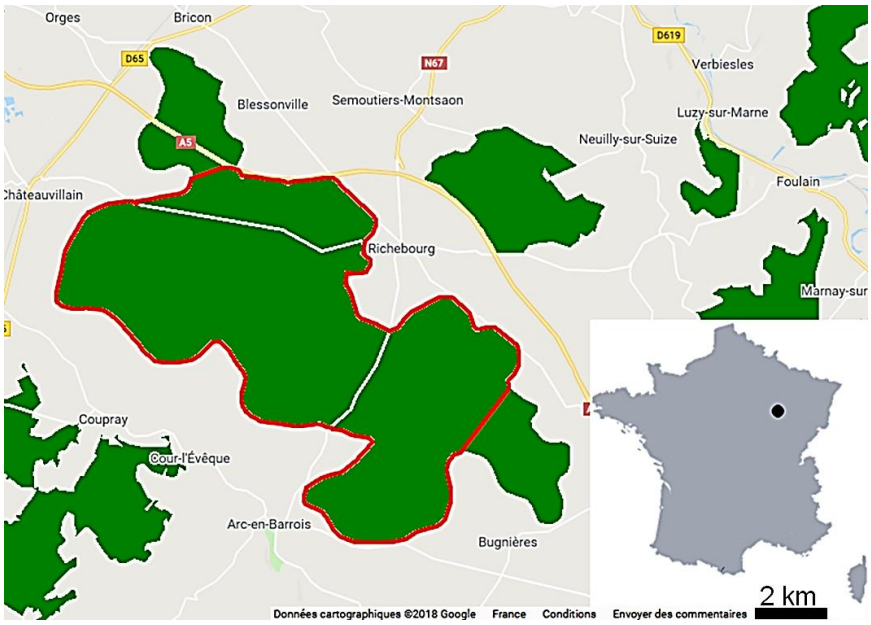


Figure II 4 Location and representation of the forest of Châteauvillain-Arc-en-Barrois. The red line represents the national forest where hunting takes place.

La Petite Pierre, Chambord, Chizé, Belval

Four other sites were also included in the study. The forest of the National Reserve of La Petite Pierre is a 2,800ha open forest located in North-East of France (48°5′N, 7°E, Figure II 6a). Like the previous study site, the climate is continental, with oceanic influences but the forest clump composition is different. The main tree species are silver fir (*Abies alba*), douglas-fir (*Pseudotsuga douglasii*), Norway spruce (*Picea abies*) and European beech (*Fagus sylvatica*). Nearly 150 wild boars are killed each year (Figure II 5). The forest of the Domaine National de Chambord is a 5,440ha forest located in central France (47°36′N, 1°31′E, Figure II 6b). It is enclosed in a 32-km-long stone wall. The climate is mild humid temperate characterized by moderately warm summers and no dry season. The forest is mainly composed of oaks (*Quercus* spp) and pines (*Pinus* spp). Chambord shows the highest number of wild boar killed each year with around 950 individuals shot (Figure II 5). Considering its size and that the population is closed, it is the population with the highest density of wild boars. The Réserve Biologique Intégrale of Chizé is a 2,614ha fenced forest located in Western France (46°50′N, 0°25′W, Figure II 6c). The climate is oceanic with Mediterranean influences, characterized by mild winters and hot summers with frequent summer droughts (average temperatures of 6°C in January and 20°C in August). Oak and beech are the two main tree items in the forest. Like in La Petite Pierre, around 150 individuals are killed each year (Figure II 5). Finally, the forest of Belval is located in North-East of France (49°3'N, 5°E, Figure II 6d) is a 650ha enclosed private

forest. Oak and beech are the main tree items in the forest clump. The climate is between continental and oceanic like Châteauvillain and La Petite Pierre. The number of wild boar killed reaches 90 individuals each year.

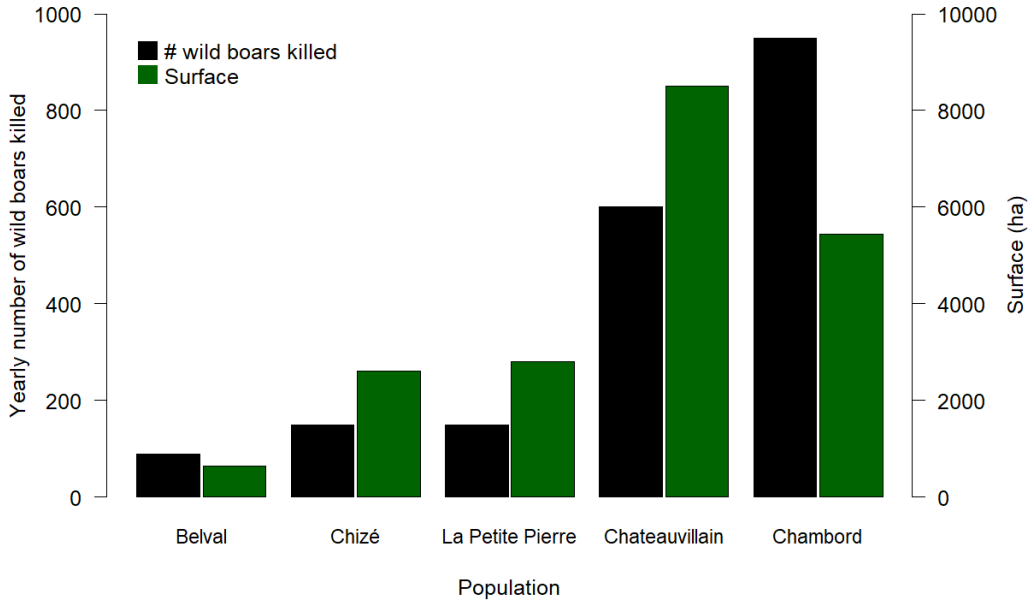


Figure II 5 Number of wild boars killed annually and surface of the study areas for each populations of wild boars.

Obtaining precise estimations of wild boar density in a population remains a challenge (Engeman et al. 2013) and no data was available for most of the populations. The main criteria used to get a proxy of population density remains hunting bag (Maillard et al. 2010). Considering the number of wild boars killed per unit of space, the highest densities are found in Chambord and Belval (Figure II 5) which are both closed populations. They are the only two populations where number of wild boars killed by 100ha exceed 10. On the other hand, La Petite Pierre and Chizé are the populations with the lowest values. High number of wild boars are killed in Châteauvillain, but it is also the largest study area which makes this population intermediate in term of density of wild boars.

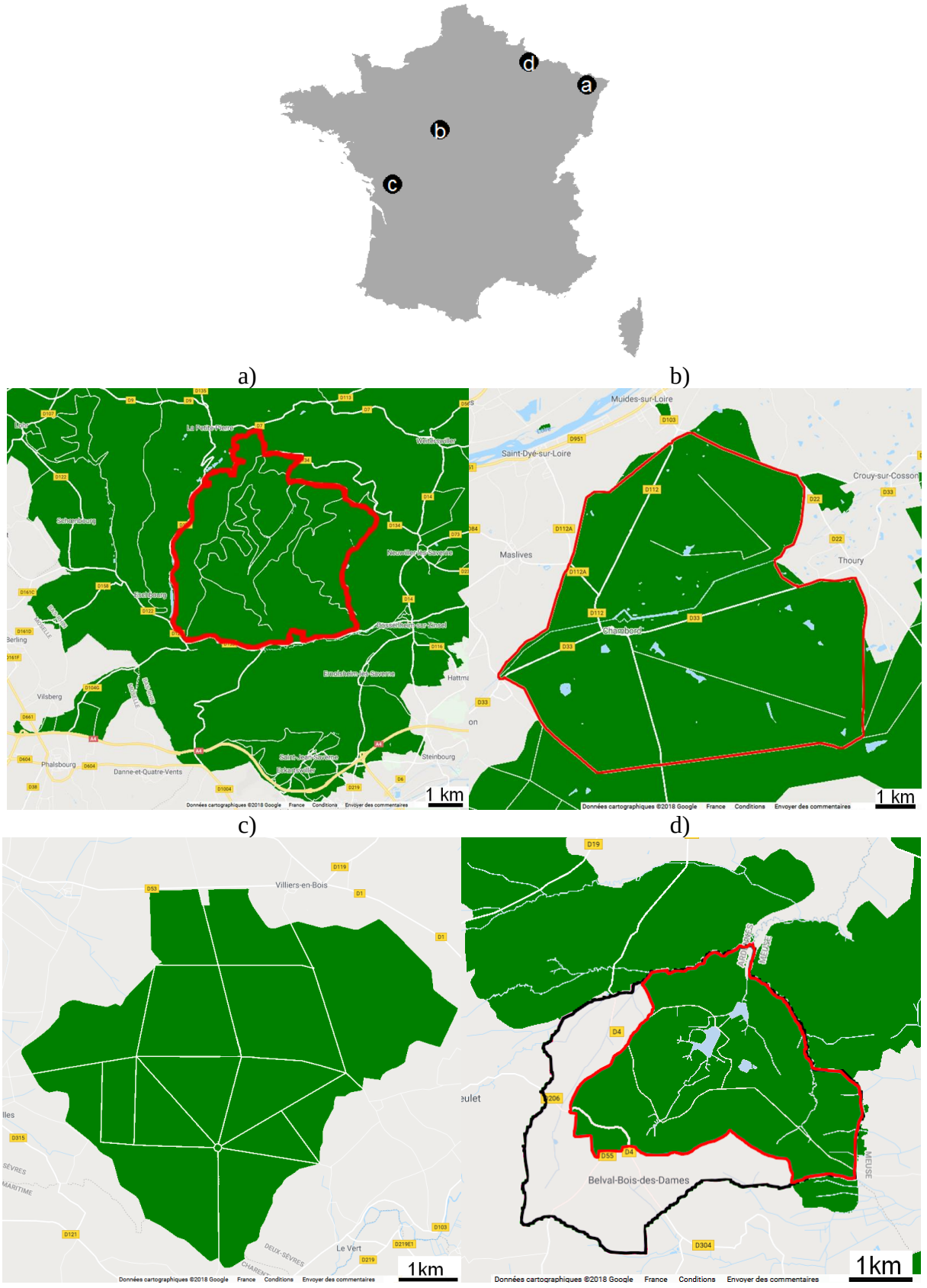


Figure II 6 Location and representation of the study areas: a) National Reserve de La Petite Pierre, b) Domaine National de Chambord, c) Réserve Biologique Intégrale de Chizé and d) Domaine de Belval.

Sampling

Wild boars are hunted in each of these study sites between October and February with variations according to the year and the study sites. In Châteauvillain, hunting events are very frequent as they take place every week-end during this period. In Chambord, Chizé and La Petite Pierre, hunting events are more occasional, only once (sometimes twice) a week but not every week. In Chizé and La Petite Pierre, they occur in a shorter period of time as they start later and finish sooner. Finally, they are only occasional in Belval, spread over few days in the winter period. Hereafter, as a hunting season overlaps two years, it will be named after the year when hunting started (i.e. 2007 for the 2007-2008 hunting season). The sampling started in the 2007 hunting season in Châteauvillain until 2015. For other populations, the sampling started later. For La Petite Pierre, it took place from 2009 to 2013, from 2011 to 2015 in Chizé and Chambord and finally from 2012 to 2015 in Belval (Table II 2).

Data are collected on individuals killed at hunting. For each individuals, sex, dressed weight (*i.e.* without the digestive system, heart, lungs, liver, reproductive tract and blood) and age based on teeth eruption and replacement patterns (Baubet et al. 1994) are recorded. Stomach contents are sampled for diet analysis (see Baubet et al. 2004 for details). For each female, genital tract and ovaries are observed to assess her reproductive status. When she is pregnant, fetuses are removed from the uterus to be weighed and measured. A piece of the ear of the mother and of each fetus (or the whole fetus, depending on its size) of the litter are sampled and stored in alcohol in an individual tube. Some males were also sampled for paternity analysis (usually heavier than 50kg, and mostly over 70kg in Chambord). Some non-reproductive adult females were also sampled for another study (see Table II 5). They were included in analysis for population estimation of genetic parameters (Table II 2, Table II 5). Overall, around 6000 individuals were sampled all populations and all years combined (Table II 2).

Table II 2 Yearly number of individuals sampled for the five populations of wild boar.

Sampling year	Châteauvillain	La Petite Pierre	Chambord	Chizé	Belval
2007	254				
2008	90				
2009	421	83			
2010	292	65			
2011	596	220	407	118	
2012	285	120	136	23	40
2013	201	140		1	78
2014	279	233	584	131	70
2015	261		499		30
2016	378				

Presentation of the set of microsatellites

An initial set of 13 microsatellites markers was used in this study but the marker S0386 was removed due to a lot of individuals displaying amplification failure (Table II 3). Ten out of these twelve markers were chosen from a larger set of 27 markers initially designed for pig (*Sus scrofa domesticus*), which is the domestic sub-species of the wild boar. This initial set was developed by a working group of the Food and Agriculture Organization of the United Nations (ISAG/FAO Standing Committee 2004). The aim was to develop species specific markers allowing to measure the genetic diversity within domestic animal. The markers were chosen from previous studies with strict characteristics. They needed to be identified in mapping studies to avoid linked markers (at least separate by more than 35cM when not possible), exhibit Mendelian inheritance, and be high quality marker (low allelic dropout, low mismatch rate). The last two markers (SW2021 and SW2496) were also selected from pig diversity analysis studies (Vernesi et al. 2003). Finally, an additional marker, AMEL, was used for sex determination (Fontanesi et al. 2008), especially important for small fetuses where genital organs are not visible to the naked eye. More details about all markers are available on the following website that records genetic map and markers for the pig http://www.thearkdb.org/arkdb/.

Table II 3 Microsatellite markers information used for paternity study of the five wild boar populations including the name of the locus, its chromosome location, the sequences for the primers used for amplification and expected and observed size in our sample.

Locus	Chromosome	Primer sequence	Expected size	Observed size
CGA	1	ATAGACATTATGTAAGTTGCTGAT GAACTTTCACATCCCTAAGGTCGT	250-320	217-299
SW240	2	AGAAATTAGTGCCTCAAATTGG AAACCATTAAGTCCCTAGCAAA	90-150	161-185
SW2021	3	GCGACACATGAGATAAAACTGC AATCCACAGGCTTACTCAGATG	100-130	99-143
SO005	5	TCTTCCCTCCTGGTAACTA GCACTTCCTGATTCTGGGTA	200-280	208-272
SO228	6	GGCATAGGCTGGCAGCAACA AGCCCACCTCATCTTATCTACACT	220-250	213-247
SW122	6	TTGTCTTTTTATTTTGCTTTTGG CAAAAAAGGCAAAAGATTGACA	110-120	111-129
SO068	13	AGTGGTCTCTCTCCCTCTTGCT CCTTCAACCTTTGAGCAAGAAC	210-260	209-261
SO215	13	TAGGCTCAGACCCTGCTGCAT TGGGAGGCTGAAGGATTGGGT	135-169	133-171
SW2496	14	TATAGCATTTGGATGTTCCACG GCCCAAATAAAGTGGTCTATGC	180-230	185-234
SO355	15	TCTGGCTCCTACACTCCTTCTTGATG TTGGGTGGGTGCTGAAAAATAGGA	240-280	280-269
SW936	15	TCTGGAGCTAGCATAAGTGCC GTGCAAGTACACATGCAGGG	80-120	91-114
SW24	17	CTTTGGGTGGAGTGTGTGC ATCCAAATGCTGCAAGCG	96-120	95-120
AMEL	X/Y	GTTTAAGCCCTGATGGGTCA CCGGGATAGAACTCTGGTCA	♂: 171,181 ♀: 181,181	♂: 171,181 ♀: 181,181

The set of markers shows high variation of allele number (min_A=6.17 ± 4.11 in Belval, max_A = 12.08 ± 8.07 in Châteauvillain) and genetic diversity (ranging from 0.58 in Chambord and La Petite Pierre to 0.61 in Châteauvillain) in all studied populations (Table II 4). Variation in allele number is important for paternity analysis. For example, some rare alleles can improve assignment of an offspring to its father if this allele is present in both individuals and high genetic diversity allows to distinguish individual from one another. In highly inbred populations, a lot of individuals may share the same genotype (combination of allele for loci studied). No population showed deviation from Hardy-Weinberg equilibrium (tested on adults for each hunting seasons and each populations) allowing to perform analysis with most genetic software (lack of deviation from Hardy-Weinberg equilibrium is a hypothesis to perform the

statistical analysis). Estimated using GenALEx (Peakall and Smouse 2006, 2012), probabilities of identity (P_{ID}, the probability that two random individuals have exactly the same genotype) were very low in all populations (the highest value being $8{,}1\times10^{-10}$ in Chambord) suggesting analyzing the genotype allows to discriminate one individuals from one the other beyond any doubt. Also, probabilities of identity for siblings (P_{IDsib}, the probability that two sibling individuals have exactly the same genotype) were low. These values fit recommendations from literature that suggest value of below 0.0001 for wildlife forensic cases (Waits et al. 2001). Also, they are similar to values of another study of parentage in wild boar (Costa et al. 2012). Altogether, results for this set of markers allow to be confident for the paternity analysis. Moreover, as fetuses were sampled from pregnant mothers, mother-offspring relationships are known (except when sampling mistake occurs but they are easily detected). Including this information with certainty allows to greatly improve paternity analysis (Jones et al. 2010). Also, the comparison of genotype of mothers against their whole litter showed that genotyping errors are rare according to the marker selection (except when sampling mistake occurred highlighted by genotype incompatibility between the mother and all fetuses).

Table II 4 Number of individuals included (N), mean ± sd of number of alleles (A), allelic richness (Ar, calculated on 1000 subsampling of 96 individuals, based on minimum number of individuals in a population, found for Belval), observed (Ho) and expected heterozygosity (He) and differentiation index (F_{is}), difference from Hardy-Weinberg equilibrium (HW, NS for non-significant) tested with *Fstat*, probability of identity of random individuals (P_{ID}) and identity between full sibling (P_{IDsib}) from Genalex, across the 12 microsatellite loci, calculated using adult individuals, for the five populations of wild boar.

	Châteauvillain	Chambord	Chizé	La Petite Pierre	Belval
N	1385	458	115	315	96
A	12.08 ± 8.07	8.5 ± 6.2	7.5 ± 5.37	8.08 ± 5.82	6.17 ± 4.11
Ar	8.6 ± 5.33	6.76 ± 4.52	7.28 ± 5.11	7.18 ± 5.02	6.17 ± 4.11
H_o	0.59 ± 0.24	0.56 ± 0.23	0.63 ± 0.22	0.58 ± 0.29	0.61 ± 0.3
H_e	0.61 ± 0.25	0.58 ± 0.24	0.64 ± 0.23	0.58 ± 0.29	0.6 ± 0.27
F_{is}	0.03 ± 0.04	0.03 ± 0.06	0.02 ± 0.06	-0.01 ± 0.06	0.01 ± 0.1
HW	NS	NS	NS	NS	NS
PID	6.9×10^{-11}	8.1×10^{-10}	6.8×10^{-11}	2.1×10^{-10}	1.6×10^{-10}
PID_{sib}	1.5×10^{-04}	2.5×10^{-04}	9.6×10^{-05}	2.1×10^{-04}	1.6×10^{-04}

Table II 5 Repartition of sampled and genotyped individuals for the five wild boar populations. 'Other' represent non-reproductive females included in allele frequency analyses.

	Class	Châteauvillain	La Petite Pierre	Chambord	Chizé	Belval
2007	Mother	38				
	Fetus	182				
	Male	34				
	Other					
2008	Mother	7				
	Fetus	34				
	Male	49				
	Other					
2009	Mother	52	11			
	Fetus	280	64			
	Male	89	2			
	Other		5			
2010	Mother	30	5			
	Fetus	142	30			
	Male	118	23			
	Other		6			
2011	Mother	36	23	43	11	
	Fetus	192	124	223	76	
	Male	213	38	86	26	
	Other	154	34	49	4	
2012	Mother	10	3			2
	Fetus	39	10			7
	Male	133	33		15	21
	Other	102	74		8	10
2013	Mother	17	20			4
	Fetus	105	111			33
	Male	77	4		1	22
	Other	1	2			19
2014	Mother	32	29	85	13	8
	Fetus	178	200	454	81	57
	Male	63	3		24	4
	Other	2			13	
2015	Mother	32		56		3
	Fetus	150		236		24
	Male	60		137		3
	Other			2		
2016	Mother					
	Fetus					
	Male	36				
	Other					

Chapter IIIA
Investigating factors influencing multiple paternity rates

Characterization of mating system in five wild boar populations subject to varying hunting pressure

Abstract: Intraspecific variations in mating system are reported in numerous species, especially when they live in broad ecological contexts. This induces between population variability in proportion of females engaging in multiple male mating, which depends on the number of male available. For hunted ungulates species, hunting is known to influence population structure, especially when males are preferentially targeted for trophy hunting. Here we investigated how variations in hunting pressure and yearly proportion of big males' removal impact probability of multiple paternity within a litter and the number of mating partners of females in five wild boar (*Sus scrofa scrofa*) populations. We found high rates of multiple paternity in all studied populations confirming the promiscuous mating system recently reported of wild boar with high within population variation. However, variations in hunting pressure and removal of big males did not influence the probability of multiple paternity neither the number of mating partners of females, once the population with the highest sample size was removed. The large magnitude of within population variations in mating systems of wild boar show the great plasticity already reported in the species.

Keywords: polyandry; multiple paternity; harvesting; reproduction monopolization; multiple male mating

Introduction

Multiple sired-litters are common in mammals. Females often mate with several males at a reproductive event (Stockley 2003). The proportion of females engaging multi-male mating varies among species, from close to 0% for monogamous species to 100% for highly promiscuous species (Taylor et al. 2014), leading to inter-species variations of multiple paternity rates (Eccard and Wolf 2009). The number of males that females mate with also changes depending on species-specific characteristics. For example, mate guarding behaviors by male decrease the probability of multi-male mating of female (Kokko and Morrell 2005). However within-species variations also exist depending on the environmental context individuals live in. For domestic cat (*Felis catus*), proportion of litters sired by several males is lower in a low density population than in areas of higher density of cats (Say et al. 1999, 2002). Indeed, in low density, males can more easily defend females against competitors and monopolize paternity of their whole litters. Also, by definition, in such ecological contexts, females have a lower probability of meeting several males than in high population densities leading to low multiple-male mating (Kokko and Mappes 2013). In addition to population density, another important parameter influencing the encounter rate of potentials mating partners is the operational sex ratio (Emlen and Oring 1977), consists of sex-ratio only considering reproducing males and females. Indeed, when the operational sex ratio is highly females skewed, only few males are available for reproduction and thus, multiple-male mating will be rare.

The age structure of animal populations changes over time. The range of variations is especially high for hunted populations as harvesting artificially reduces survival differentially in age classes (Langvatn and Loison 1999; Solberg et al. 2000; Frank et al. 2017). Moreover, reducing female survival can influence their reproductive strategy. Low survival changes the trade-off between reproduction and survival (Gamelon et al. 2011). In populations where hunting is intensive, females of iteroparous species (reproduction in several reproductive events) may not get more than one breeding occasion and should favor any strategy maximizing their reproductive output early in life (Proaktor et al. 2007). Multi-male mating can be favored in such context. It reduces the probability of reproductive failure compared to single-male mating and also increases the litter size (Stockley 2003). This is possible only if female have the opportunity to meet and mate with several males (Martin et al. 2014). In polygynous species, females are monopolized by dominant males which try to maximize their reproductive success by preventing competing males to mate (Emlen and Oring 1977). Such species are characterized

by strong sexual dimorphism with males displaying secondary sexual characters. Older males are often preferentially targeted by hunters because they have the biggest trophies. Their removal allows females to engage in multi-male mating as they are no longer maintained in harem.

Trophy hunting impacts the structure of the population leading to a decrease of their number (Loveridge et al. 2007; Douhard et al. 2016). As few males are enough to fertilize many females, the consequences of their disappearance were neglected for a long time in population dynamic models. Rankin and Kokko (2007) showed that depending on populations characteristics, changes in sex ratio or absolute number of males can have important consequences. They also highlighted that selective harvesting can have severe consequences depending on the mating system of the population. Moreover, the change in population structure in trophy hunted populations leads to changes in its operational sex-ratio (Milner et al. 2007). Thus, through the reduction of the number of males in the population, trophy hunting can lead to change in its mating systems. Milner-Gulland et al. (2003) reported a switch of mating system in an intensively hunted population of saiga antelope (*Saiga tatarica tatarica*) where mainly males are killed for their horns. Dominant females where anecdotally observed surrounding available males while normally males would defend a harem of up to 30 females. While this case is extreme, changes in operational sex ratio can release sexual competition between males and change reproductive patterns (Singer and Zeigenfuss 2002).

The wild boar (*Sus scrofa scrofa*) is hunted all over its distribution range (Massei et al. 2015). The mating system of the species is polygyny (Mauget 1980; Dardaillon 1984). Males feature tusks growing during their entire life, they used to fight other males to monopolize group of females. These tusks are appreciated trophies collected by hunters (Kierdorf et al. 2004). Thus, old males are preferentially killed and the number of adult males in the population can be low in heavily hunted populations due to low survival and higher removal rates (Fernández-llario and Mateos-quesada 2003; Toïgo et al. 2008). Multiple paternity was reported based on genetic studies in hunted populations of wild boars, highlighting multi-male mating behavior of females with variation in number of mates (Delgado et al. 2008; Pérez-González et al. 2014; Gayet et al. 2016). However, whether the proportion of female engaging in multi-male mating and the number of partners they mate with are linked to hunting pressure or not remains to be studied.

We characterized the mating system of wild boar in five different populations varying mainly in hunting pressure. We investigated how the removal of big males influenced multiple

paternity rates in the population. We then measured the impacts of hunting and of the proportion of big males in the hunting bag on (i) the probability for a litter to be multiply-sired and (ii) the number of sires within a litter. We expected higher rates of multiple paternity in heavily hunted populations due to increased disruption of the population structure by harvesting. Finally, we expected increased probabilities for litters to be multiply sired together with increased number of fathers in a litter when the proportion of big males was high in the hunting bag. Indeed, female monopolization should decrease when big competitive males are removed from the population and allows medium males, still present, to get access to females.

Material and Methods

Samplings were realized in the five wild boar populations of Belval, Chambord, Châteauvillain, Chizé and La Petite Pierre (described in Chapter II). The number of animals killed in a given year was recorded and reported to the hunting area to estimate a number of individuals killed per unit of space (Figure III 1), which is a proxy of hunting pressure. Belval and Chambord have the highest number of individuals killed per 100ha, while Chizé and La Petite Pierre have the lowest values. Châteauvillain is between the two groups (Figure III 1). The mean litter size ranged from 4.92 in Chambord to 7.12 in Belval with intermediate values for the three other populations (Table III 1).

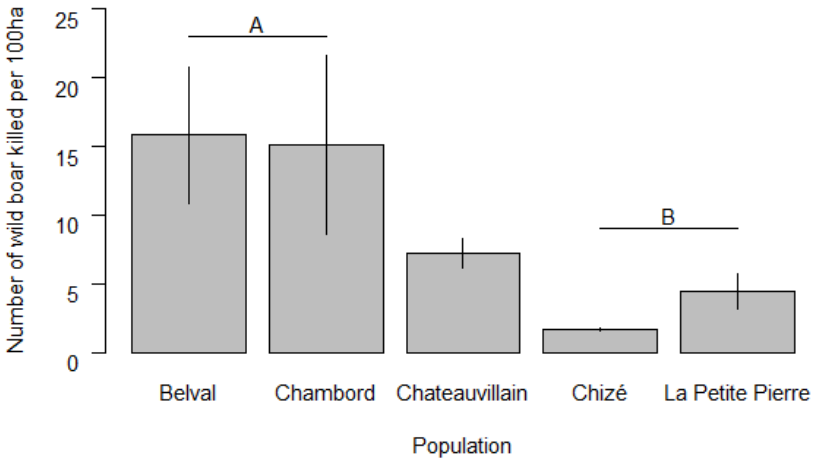


Figure III 1 Number of wild boars killed per 100ha (median ± SD) for five wild boar populations ($n_{year\text{-}Belval}$=4, $n_{year\text{-}Chambord}$=3, $n_{year\text{-}Châteauvillain}$=9, $n_{year\text{-}Chizé}$=2, $n_{year\text{-}La\ Petite\ Pierre}$=5). A and B symbolize significant difference between groups (Kruskal-Wallis test: $\chi^2 = 19.057$, df = 4, p-value < 0.001).

Molecular and paternity analysis

All samples were genotyped for 12 microsatellite loci (see Chapter IVA for more details). Individuals whose genotyping completely failed, were excluded from analysis and when the genotype of the mother or a fetus was not obtained, we excluded the whole mother-litter couples. This reduced the dataset to 511 litters across all populations (details in Table III 1). All genotypes obtained were analyzed using the software COLONY (Jones and Wang 2010). As mothers were already known, it was used to identify fathers among putative males from the hunting bag or to assign a genotype if the father was not sampled. The analyses were performed for each population and each year. The population of Châteauvillain is well studied and we know that the probability of being killed each year for males is very high in this population (estimated survival of 0.23 [0.17; 0.30], Toïgo et al. 2008). Accordingly, all sampled males from year n, subadult and adult males killed year n+1 and adult males from year n+2 were included as possible fathers in COLONY analyses (Table III 1). Since the other populations are not as well studied, all males killed during the focused year and the following years were included as possible fathers in the analyses (Table III 1). We considered that 50% of reproductive males were sampled each year. We acknowledge this is speculative considering the variation of number of males sampled between years in different population. Nevertheless, our results were only slightly influenced by the proportion of males sampled when it varied between 20% and 70% (Figure III 2) and results of COLONY are known to be consistent across different values of this parameter (Harrison et al. 2013). For all analyses, we considered both sex polygynous, did not set population allele frequency, and used the full likelihood analysis method with a medium precision. The markers were chosen codominant, with an allelic dropout of 0.001 and a typing error of 0.01 for each locus. The default values were chosen for all other parameters. The results allowed us to estimate the number of fathers in each litter and then to know if the litter was sired by one or multiple males. Multiple paternity in a litter was coded as 0 (no multiple paternity) when only one male sired the litter and 1 (multiple paternity) when several males contributed to the litter.

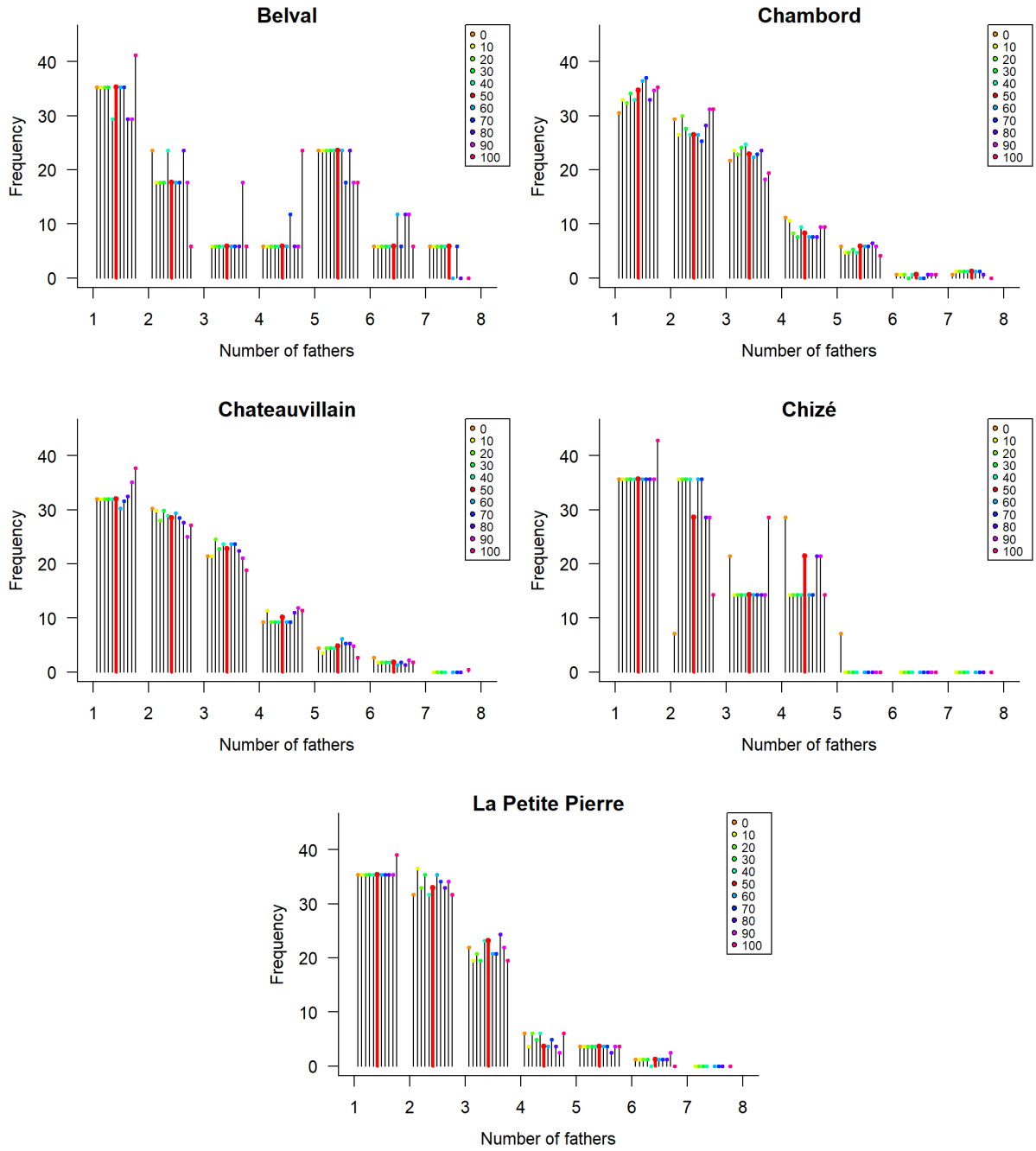


Figure III 2 Frequency of the number of fathers (sampled or inferred by COLONY) found in a litter for the five populations of wild boars analyzed with COLONY using different values for the proportion of fathers sampled varying from 0 to 100, increasing from left to right (also indicated by the color). The red bar highlights the value selected for analysis (50%).

Table III 1 Number of litters (mothers), offspring and males included in paternity analyses with COLONY for the five populations of wild boars and the different years used for the analyses. Mean litter size (± SD) are also reported.

		Populations				
Year	Class	Châteauvillain	La Petite Pierre	Chizé	Chambord	Belval
	Litters	32				
2007	Offspring	154				
	♂	141				
	Litters	7				
2008	Offspring	34				
	♂	233				
	Litters	47	10			
2009	Offspring	263	56			
	♂	307	103			
	Litters	28	4			
2010	Offspring	134	23			
	♂	355	101			
	Litters	34	21	10	42	
2011	Offspring	184	115	71	218	
	♂	417	78	66	223	
	Litters	8	1			2
2012	Offspring	34	3			7
	♂	273	40			50
	Litters	16	17			4
2013	Offspring	99	98			33
	♂	161	7			29
	Litters	27	29	4	77	8
2014	Offspring	144	192	20	408	57
	♂	120	3	24	137	7
	Litters	29			51	3
2015	Offspring	133			210	24
	♂	89			137	3
Overall	Litter size	5.17 ± 1.67	5.94 ± 1.53	6.5 ± 1.56	4.92 ± 1.63	7.12 ± 2

Statistical analysis

Multiple paternity rates were measured in all five populations and were compared using Kruskal-Wallis test. In order to get information on the proportion of big males removed from each population and each year, we recorded the weight of all males killed a given year (except Belval and the year 2009 of La Petite Pierre for which information were not available). To focus only on sexually matured males, we removed males with a dressed body weight inferior to 30kg (which correspond to full body mass of 38kg (Mauget and Boissin 1987; Baubet 1998)), as the proportion of males producing sperm above this threshold weight is low (Mauget and Boissin 1987). For each population and each year, the weight of the biggest males in the hunting bag

was used as reference. The yearly proportions of big males (*Prop*) were estimated as the proportion of males with a weight higher than the 80% percentile of the biggest male killed this year in all four populations (Figure III 3). To quantify the variance of the weight of males killed a given year, we also calculated the difference between the 80% percentile weight value and the median of the weight of sexually mature males (*Diff*). To measure the influence of these two parameters on mating system we performed two analyses. The first measured the probability of a litter to be multiply-sired using the litter status (*pMP*) as a dependent variable in a binomial regression model, while the number of fathers (*Nf*) of a litter was used as a dependent variable in a Poisson regression model. In both cases, the mother body mass (*BMm*), the proportion of big males (*Prop*) and the difference between the 80% percentile weight of the biggest males and the median value of mature males (*Diff*) and the number of wild boars killed per unit of space (*Nha*), were included as explanatory parameters. We also included the litter size (*LS*) for the second model as we know there is a positive relation between *LS* and *Nf* (Chapter IVA). Finally, we included the population as a random factor in all models. We acknowledge that we should include operational sex ratio. However, estimating population size of wild boar is difficult without heavy capture-mark-recapture protocol and the data were not available for all the focused populations (Sweitzer et al. 2000). Analyses were performed including and excluding Châteauvillain population, as the big sample size of this population may weight on the outcome on the results. Correlation between parameters was verified using the Variance Inflation Factor (VIF). Parameters with highest VIF were gradually removed until all VIF were below 3 (Zuur et al. 2009). The best model was selected based on AICc criterion and when several obtained $\Delta AICc < 2$, we used model averaging to get the parameters estimates (Burnham and Anderson 2004). All statistical analyses were performed using R software version 3.3.3 (R Core Team 2017).

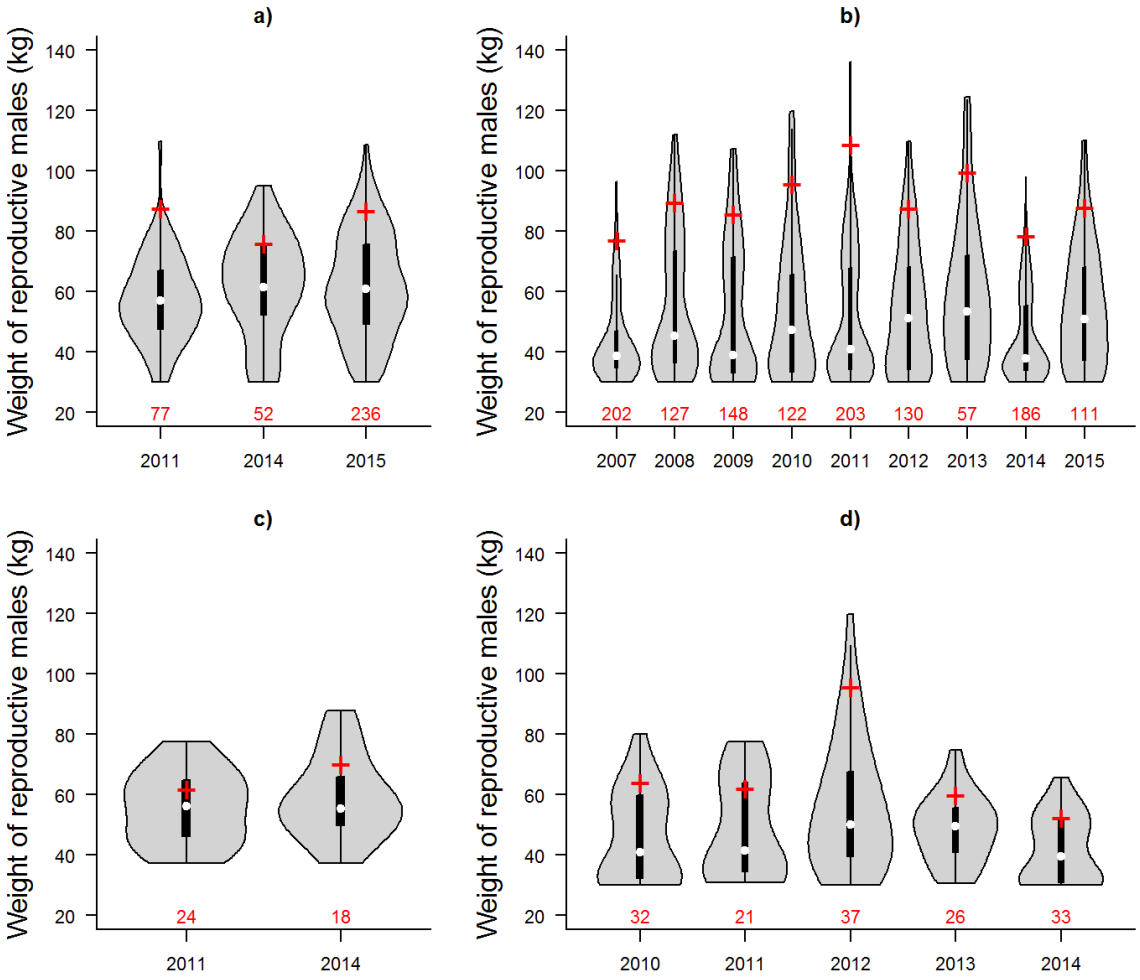


Figure III 3 Weight of reproductive males killed at hunting each year from the four wild boar populations of a) Chambord, b) Châteauvillain, c) Chizé and d) La Petite Pierre. Sample size are given by the red values below each plot. The width shows the density of points. The white points show the median, the thick black vertical lines the central quartiles and the fine vertical lines 1.5 the inter-quartile space. Red crosses show the threshold of 80% of the weight of the heaviest male killed a given year.

Results

Between populations variations of multiple paternity rates

All seasons combined, the number of fathers per litter ranged from one to four and up to seven depending of the population (Figure III 4a and Figure III 2). The highest numbers of fathers were found in Belval (median = 2 ± 2.08 SD) and Chambord (median = 2 ± 1.32 SD). The maximum number of fathers reached six for both Châteauvillain (median = 2 ± 1.25 SD) and La Petite Pierre (median = 2 ± 1.12 SD). The smallest maximal number of fathers per litters was found in Chizé (median = 2 ± 1.19 SD) reaching only four. In all populations, the majority of litters (more than 50%) were sired by one to three males (Figure III 2). Multiple paternity

rates across seasons were high and do not varies between population (Figure III 4b, Kruskal-Wallis test: $\chi^2 = 1.043$, df = 4, p-value = 0.90). High inter-annual variations were observed especially in population with low sample sizes (Belval and Chizé).

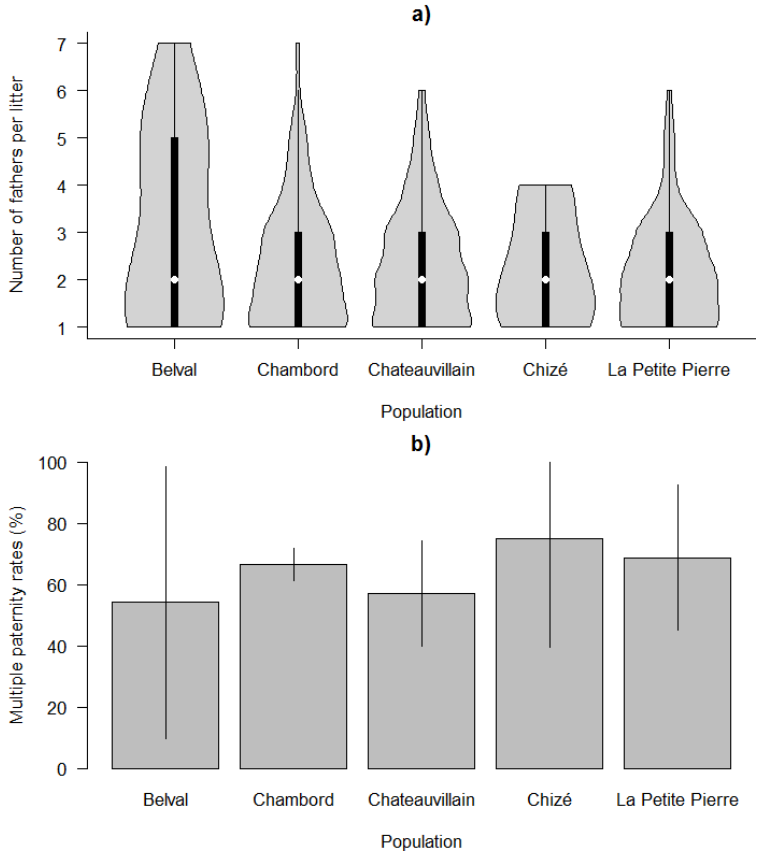


Figure III 4 a) Number of fathers per litter and b) multiple paternity rates (median across seasons ± SD) obtained from the analysis with COLONY in five populations of wild boars ($n_{litter\text{-}Belval}$=17, $n_{litter\text{-}Chambord}$=170, $n_{litter\text{-}Châteauvillain}$ =228, $n_{litter\text{-}Chizé}$=14, $n_{litter\text{-}La\ Petite\ Pierre}$=82). In the first figure, the width shows the density of points. The white points show the median, the thick black vertical lines the central quartiles and the fine vertical lines 1.5 the inter-quartile space.

Investigation of parameters influencing multiple paternity

Three models supported the data to explain *pMP* in a litter (ΔAICc < 2, Supplementary material Table III S3a). The averaged models included the proportion of big males in hunting bag (*Prop*), the difference of the median weight males killed and the threshold for big males (*Diff*), the mother body mass (*BMm*) and the number of wild boars killed per unit of space (*Nha*) when all population were included in analyses (Table III 2a). Only *Prop* showed a significant positive effect on the *pMP* (β = 0.374 ± 0.177, p = 0.035, Table III 2a), however when the population of Châteauvillain was removed, effects no longer exist and the best model was the null model (Table III 2b). When focusing on the number of fathers per litter, *BMm*, *Diff*, *Nha*

and *LS* were included in the models best supported by the data whether Châteauvillain was included or not (Table III 2c and Supplementary material Table III S3b, Table III 2d). *Prop* was also in the averaged model including all four populations, however, the only significant parameter in both cases was the positive effect of the litter size ($\beta = 0.113 \pm 0.032$, $p < 0.001$ with Châteauvillain, Table III 2c and $\beta = 0.113 \pm 0.045$, $p = 0.013$ when the population of Châteauvillain was removed, Table III 2d).

Table III 2 Estimates, standard errors, z statistics, and P values of parameters linked with the probability of occurrence of multiple paternity (a.) and the number of fathers (c.) in a litter for four populations and without Châteauvillain (b) and (d) from full averaged model. Values for the proportion of males killed with a dressed weigh above the annual threshold (*Prop*) and the difference between the median of dressed weigh of males killed and the annual threshold (*Diff*), the number of wild boars killed per surface unit (*Nha*), the mother body mass (BM_m) and the litter size (*LS*) were obtained from averaged model strongly supported by the data. Châteauvillain was removed from (b.) and (d.). Significant parameters are in bold ($n_{litter\text{-}Chambord}$=166, $n_{litter\text{-}Châteauvillain}$ =226, $n_{litter\text{-}Chizé}$=14, $n_{litter\text{-}La\ Petite\ Pierre}$=65).

a.

Parameter	Estimate	SE	z-test statistic	P-value
Intercept	0.728	0.145	-	-
Prop	**0.374**	**0.177**	**2.108**	**0.035**
Diff	0.556	0.344	1.611	0.107
BMm	0.033	0.076	0.431	0.666
Nha	0.020	0.065	0.305	0.760

b.

Parameter	Estimate	SE	z-test statistic	P-value
Intercept	0.651	0.135	-	-

c.

Parameter	Estimate	SE	z-test statistic	P-value
Intercept	0.826	0.033	-	-
LS	**0.113**	**0.032**	**3.493**	**<0.001**
Prop	0.040	0.058	0.691	0.490
Diff	0.058	0.071	0.812	0.417
Nha	0.008	0.024	0.345	0.730
BMm	-0.003	0.016	0.174	0.862

d.

Parameter	Estimate	SE	z-test statistic	P-value
Intercept	0.819	0.043	-	-
LS	**0.113**	**0.045**	**2.493**	**0.013**
Diff	0.005	0.022	0.226	0.821
BMm	0.003	0.021	0.154	0.877
Nha	0.003	0.020	0.130	0.897

Discussion

Overall, multiple paternity rates estimated with COLONY are high in all five populations of wild boars. However, despite variations in hunting intensities between populations, results did not show influence of the number of individuals killed per 100ha or the proportion of big males in the hunting bag on probability of multiple paternity or on the number of fathers within a litter once the population of Châteauvillain was removed from analysis.

High multiple paternity rates were observed in the five populations but there was no difference between populations. These rates are higher than most of those already reported in other populations of wild boars based on genetic studies (see Chapter IVA, Table IV 2). The genetic mating system of the species can be confidently defined as promiscuous, at least in hunted populations of dense deciduous forests. The disruption of population age structure by hunting is well documented (Langvatn and Loison 1999; Milner et al. 2007). Moreover hunting may change individuals repartition in space and influence mating opportunity (Milner-Gulland et al. 2003; Keuling et al. 2008b; Davidson et al. 2011). Thus, even low hunting pressures may induce changes in the mating system. The lack of variation between populations does not fit with results obtain from some other polytocous species where multiple paternity rates change between populations (Wakabayashi et al. 2017) and other ungulate species where the mating systems varies according to the environmental context (Gosling 1991). Besides, except for hunting, amplitude of variation of ecological context is rather low between our focus populations. Indeed, all populations are located in Northern France and the forest clump composition is rather favorable for wild boars. Investigating multiple paternity rates in populations with greater ranges of ecological contexts would be interesting to investigate precisely mating system of wild boar.

Variations of hunting were observed between populations considering the number of individuals killed by unit of space (*Nha)*. This parameter allows to get information readily comparable between populations, but it is greatly sensitive to the density of populations. Two of the three enclosed populations showed very high values of *Nha*. Both are populations were wild boars are fed with maize (pers. com.) allowing high densities and bigger hunting bags. Usually, in open populations, supplementary feeding is used by hunters to maintain wild boars in forest and protect crop fields (Calenge et al. 2004) but in that case, maize is used to maintain individuals in good body conditions. High values of *Nha* can be explained by high population densities and be poor predictor of hunting intensity. However independently of the density, *Nha* remains meaningful to quantify disturbance induced by hunting in populations. However, obtaining good estimates of the number of individuals in each population, and proportions of reproductive wild boars removed, would greatly improve the confidence in the observed patterns. Such information is difficult to obtain especially for large study areas such as Châteauvillain or high-density population as Chambord. Despite great variations between populations, in this study, *Nha* did not influence neither *pMP* nor the number of fathers in a litter. This hunting parameter influences other population characteristics, but not those

investigated in this study. Another possible explanation is that since we used population as a random factor, its effect may have been encompassed in the population effect. Once population effect removed, variations in *Nha* were not sufficient to influence significantly the explained variables. Investing influence of yearly variations of *Nha* within each population would be interesting. We were not able to perform these analyses here since, excepted for Châteauvillain, sampling did not cover enough years (Chizé, Chambord) or included enough litters per year (La Petite Pierre) to have good statistical power.

Probability of multiple paternity (*pMP*) within a litter was not significantly influenced by any parameter included in our analysis except when the Châteauvillain population, having the highest sample size, was part of the analysis. This showed two things. First, *pMP* does not change neither due to yearly variations of the removal of big males in the population of Chambord, Chizé and La Petite Pierre, nor due to the yearly changes of the variance of weight of the males killed at hunting. We did not include any time parameters in the models. Hunting takes place during the rut of wild boar, and males are preferentially targeted by hunters (Gamelon et al. 2012). The probability of multiple-male mating may change during the hunting season along with the decrease of the proportion of males in the population over the hunting period. Secondly, some variable in our models influenced *pMP* in the population of Châteauvillain enough to influence the results for the analysis including all populations. Especially, *Prop* had a significant positive effect in average models where all populations were analyzed. The increase of the removal of big males induces a decrease in the number of males with the capacity to monopolize females. Females are more available for other males to mate with, and, without big males, the competition between males decreases (Singer and Zeigenfuss 2002; Kokko and Rankin 2006). However, the number of sire per litter was no significantly influenced by *Prop* whether Châteauvillain was part of the analysis or not. This suggested that more litters were sired by several males (increase of *pMP*) but litters were not sired by more males (no effect of the number of fathers) when big males disappeared from the population. Indeed, high multiple paternity rates can be observed in populations where females reproduce with a maximum of two males but the number of fathers would only slightly change. The litter size showed a positive effect on the number of fathers. Indeed, the probability of detecting several fathers in large litters is higher than in small litters as the number of fathers ranges from one to a maximum corresponding to the number of fetuses in the litter. However, the increase of the number of mating partners can also induce an increase of female fertility by decreasing the number of unfertilized eggs (Stockley 2003, also see Chapter IVA).

In conclusion, we did not find any influence of hunting on multiple paternity rates, or the number of fathers in our study populations when analyzed all together. However, the population of Châteauvillain seems apart from the others, as whether it was included or not in the analyses greatly influenced results. Investigating more precisely the mechanisms in this population could shed light in parameters influencing mating patterns in wild boar.

Acknowledgment

We are grateful to all people who helped collect harvested wild boars in the different study sites over the year. We thank the ONF, and F. Jehlé who allowed us to work on Châteauvillain population. We also thank the Domaine National de Chambord which allowed us to work on Chambord population, the François Sommer foundation and E. Richard who supervised data collection and supported data analysis in Belval and Chambord. We are thankful to all members of the ONCFS involved in data collection the different study sites and to D. Pierrard for genetic sampling in Belval. This work was supported by the French National Hunting and Wildlife Agency (ONCFS) and by the University of Lyon. ONCFS contributes also by supporting TG, providing his Ph.D. fellowship. COLONY analyses were performed, using the computing facilities of the CC LBBE/PRABI.

Supplementary Material

1. Model selection

Table III S3 Model selection to test the effect on (a.) the probability of occurrence of multiple paternity in a litter and (b.) the number of fathers in a litter of the proportion of males killed with a dressed weigh above the annual threshold (*Prop*), the difference between the median of dressed weigh of males killed and the annual threshold (*Diff*), the mother body mass (BM_m), the litter size (*LS*) and the number of wild boars killed per surface unit (*Nha*) in four populations of wild boars in France. The model retained is in bold (ΔAICc<2). 'X' denotes that the explanatory variable was included in the model ($n_{litter\text{-}Chambord}$=166, $n_{litter\text{-}Châteauvillain}$ =226, $n_{litter\text{-}Chizé}$=14, $n_{litter\text{-}La\ Petite\ Pierre}$=65).

a.

Intercept	BM_m	Prop	Nha	Diff	df	logLik	AICc	ΔAICc	AICc weight
X		**X**		**X**	**4**	**-294.842**	**597.771**	**0.000**	**0.331**
X	**X**	**X**		**X**	**5**	**-294.285**	**598.700**	**0.929**	**0.208**
X		**X**	**X**	**X**	**5**	**-294.574**	**599.277**	**1.506**	**0.156**
X	X	X	X	X	6	-293.958	600.098	2.327	0.103
X				X	3	-297.700	601.451	3.680	0.052
X	X			X	4	-296.730	601.546	3.775	0.050
X			X	X	4	-297.543	603.173	5.402	0.022
X	X		X	X	5	-296.666	603.462	5.691	0.019
X					2	-299.798	603.622	5.851	0.018
X	X				3	-299.309	604.670	6.899	0.010
X			X		3	-299.411	604.873	7.102	0.009
X		X			3	-299.788	605.628	7.857	0.007
X	X		X		4	-298.971	606.029	8.258	0.005
X	X	X			4	-299.306	606.698	8.927	0.004
X		X	X		4	-299.410	606.906	9.135	0.003
X	X	X	X		5	-298.944	608.016	10.245	0.002

b.

Intercept	LS	BM_m	Prop	Nha	Diff	# parameter	Log Likelihood	AICc	ΔAICc	AICc weight
X	**X**					**3**	**-764.831**	**1535.713**	**0.000**	**0.149**
X	**X**		**X**		**X**	**5**	**-762.802**	**1535.732**	**0.020**	**0.148**
X	**X**				**X**	**4**	**-763.981**	**1536.049**	**0.336**	**0.126**
X	**X**		**X**	**X**	**X**	**6**	**-762.029**	**1536.239**	**0.527**	**0.114**
X	**X**	**X**				**4**	**-764.716**	**1537.517**	**1.805**	**0.060**
X	**X**	**X**	**X**		**X**	**6**	**-762.719**	**1537.618**	**1.906**	**0.057**
X	X		X			4	-764.825	1537.735	2.023	0.054
X	X			X		4	-764.829	1537.744	2.031	0.054
X	X			X	X	5	-763.921	1537.971	2.258	0.048
X	X	X			X	5	-763.936	1538.002	2.289	0.047
X	X	X	X	X	X	7	-761.910	1538.061	2.348	0.046
X	X	X	X			5	-764.700	1539.529	3.816	0.022
X	X	X		X		5	-764.712	1539.553	3.840	0.022
X	X		X	X		5	-764.822	1539.773	4.060	0.020
X	X	X		X	X	6	-763.872	1539.924	4.212	0.018
X	X	X	X	X		6	-764.692	1541.566	5.853	0.008
X			X		X	4	-768.720	1545.525	9.813	0.001
X		X	X		X	5	-767.785	1545.699	9.987	0.001
X		X				3	-770.620	1547.292	11.579	0.000
X			X	X	X	5	-768.590	1547.309	11.597	0.000
X		X	X	X	X	6	-767.626	1547.432	11.720	0.000
X						2	-771.722	1547.469	11.757	0.000
X		X			X	4	-769.786	1547.657	11.944	0.000
X					X	3	-771.193	1548.438	12.725	0.000
X				X		3	-771.332	1548.716	13.003	0.000
X		X		X		4	-770.318	1548.722	13.009	0.000
X			X			3	-771.402	1548.856	13.143	0.000
X		X	X			4	-770.504	1549.093	13.381	0.000
X		X		X	X	5	-769.654	1549.436	13.724	0.000
X				X	X	4	-770.956	1549.998	14.285	0.000
X			X	X		4	-771.130	1550.347	14.634	0.000
X		X	X	X		5	-770.261	1550.650	14.938	0.000

Chapter IIIB
Investigating factors influencing multiple paternity rates

Hunting variations shape reproductive processes in a wild boar population

Abstract: Identifying origins of within population variations in reproductive strategies is an increasingly studied subject of research. Several mechanisms have already been described including female choice, male strategy, or ecological factors to explain variations of proportion of females engaging in multiple male mating. However, these mechanisms remain to be studied for many species. The wild boar (*Sus scrofa scrofa*) shows high rates of multiple paternity with within population yearly variations. In this study, we investigated how availability of food resources (proportion of forest fruits in stomach) and hunting processes (yearly number of animals killed, proportion of big males killed) influence probability of multiple paternity and number of partners of female wild boars. Forest fruits did not influence mating patterns however proportion of big males killed and median weight of males showed significant effect. This suggest that availability and quality of males in the population influence reproductive processes in wild boar and that selective hunting can impact these parameters.

Keywords*:* polyandry; multiple paternity; multiple male mating; mating system variation; harvesting; food resource

Introduction

The mating system of animal species or population is defined by ecological factors described and summarized by Emlen and Oring (1977). While it is convenient to categorize the whole species/population, all individuals of a population do not exhibit the same mating strategy. Some females from monogamous species often produce multiple sired broods (Ophir et al. 2008; Arct et al. 2015) leading to the distinction between social and genetic monogamy. Also, some females from polygynandrous species produce single-sired broods, creating variations in multiple paternity rates (proportion of broods sired by more than one male) within and between species (Trexler et al. 1997; Lank et al. 2002; McEachern et al. 2009). However, despite increasing interest in the subject, detangling if these variations of strategies are due to female choice, male strategy, ecological factors or a combination of one or more factors remain to be explored in many species (but, for a review in mammals, see Wolff and MacDonald 2004). A recent study by Wells et al. (2017), in golden-mantled ground squirrels (*Callospermophilus lateralis*), tested three hypotheses to explain multiple paternity rate variations over a 18 years period: the encounter rate (i) where multiple paternity increases with the probability that females meet males (Kokko and Rankin 2006), male monopolization (ii) in which multiple paternity rates decrease with capacity of male to monopolize females (Emlen and Oring 1977; Shuster and Wade 2003), female choice (iii) suggesting multiple paternity increase with female physical condition (Cotton et al. 2006). Yearly variations of multiple paternity rates were best explained by male monopolization hypothesis as it was mostly influenced by female aggregations and the number of competitors a male had to fight to monopolize a group of females.

Mating system of mammals is known to be influenced by population density, predation and food availability (Say et al. 1999, 2002; Kamler et al. 2004; Martin and Martin 2007). Also, food distribution in space and time is of great importance, especially in ungulates, as it shapes aggregation patterns of females and, in return, influences the capacity of males to defend group of females for reproduction (Brashares and Arcese 2002; Pérez-González and Carranza 2011). When resources distribution is sparse, females are expected to be scattered in the environment and increase their displacements when foraging (Brashares and Arcese 2002), decreasing the capacity of males to maintain them in group (Clutton-Brock and Harvey 1978). Thus, social monogamy is often observed in such situations with one male siring the litter, as most ungulates are monotocous or only slightly polytocous (litter size ranging mainly between one and three, Gaillard et al. 2000a). But increased females movement can also increase their probability to

meet different males, leading to serial monogamy (multiple male mating) (Kokko and Mappes 2013) and increase of multiple paternity rates. However, when resource are abundant and/or clumped, females are expected to remain in groups that a single male can more easily defend against competitors forming a harem (Emlen and Oring 1977; Clutton-Brock and Harvey 1978). The ensuing social mating system is polygyny with low multiple paternity rates as only one male should reproduce with the females of the group.

Unlike other ungulates, the wild boar (*Sus scrofa scrofa*) is a highly polytocous species with a litter size averaging five (Servanty et al. 2007), offering great opportunity to detect multiple male mating. The reproduction of this species is influenced by forest fruits availability (acorn and beechnut mostly), which mainly shape reproductive outputs of females (Servanty et al. 2009; Gamelon et al. 2017). Despite high adaptability for its diet, acorns and beechnuts remain the main food items of wild boar (Schley and Roper 2003). Forest fruits production is highly variable and unpredictable due to masting reproductive strategy of oaks (*Quercus sp*.) and beech (*Fagus sp*.). This strategy consists of massive production of fruits a given year followed by several years of low production, with high synchronization of trees of a same forest (Liebhold et al. 2004). Thus, mast years represent years of abundant food resources distributed in space, in contrary of years without mast production where resources are scattered. As for other ungulates species, resources distribution in space and time can influence capacity of males to monopolize groups of females and in return multiple paternity rates. However, to our knowledge, no study investigated the link between resource and mating system in the wild boar so far. Moreover, as a game species, the wild boar is subject to intensive hunting with big males especially targeted for their tusk used as trophy (Kierdorf et al. 2004). Survival of male can be heavily impacted in population with intensive harvesting (Toïgo et al. 2008). Disruption of the population structure ensuing from hunting was shown to influence reproductive processes differentially depending on the population (Chapter IIIA).

The wild boar shows between year variations in multiple paternity rates that exceed between population variations (Chapter IIIA, Figure III 4). However, factors influencing the proportion of females that mate with several males and their number of partners remain poorly investigated. We used long term monitoring of the wild boar population of Châteauvillain to investigate the influence of food resource availability and the effect of the population structure disruption on the mating system in this population. We expect high probability of multiple paternity (*pMP*) and high number of fathers (*Nf*) in a litter (promiscuous mating system) when both food resources are scarce and the number of big males in the population is low. When food

resources are scarce, groups of females would be scattered in the environment and harder to monopolize by a single male leading to increased probabilities of multiple paternity. In mast production years, when resource are abundant, the movement of females groups should be narrower leading to low level of multiple paternity. Moreover, when big males are removed from the population, females are more available for other males creating opportunity for multiple paternity.

Material and Methods

Study site and sample collection

The wild boar population is located in the 11,000 ha Châteauvillain-Arc-en-Barrois forest (48°02′N; 4°55′E, France) and is described in detail in Gayet et al. (2016). During nine hunting seasons (2007-2015), tissue samples were collected from 210 hunted pregnant females and their full litters with fetuses big enough to be measured (1092 fetuses, mean litter size = 5.2 ± 1.66 SD), 305 non-breeding females, females with putative missing fetuses due to bullet wound in the uterus or female with fetuses too small to be measured (those females were only included in the genetic diversity analyses), and from 895 putative reproductive males (also sampled in 2016) with a dressed body mass (*i.e.* without the digestive system, heart, lungs, liver, reproductive tract and blood) higher than 30kg. Crown-rump length of fetuses was measured (in millimeters) to calculate the gestation stage in days, from the average length of fetus within the litter, using relation from Henry (1968). The Julian mating date (using the 1st of July of each year as reference) was calculated by subtracting the gestation stage in days to the date of kill. Stomach contents analysis were realized during the hunting period to identify wild boar diet as described in Baubet et al. (2004). The proportion of three major items, acorn, beechnut and corn was measured and pooled for each hunting month. Finally, for each year, the number of wild boars killed for 100ha and the weight of all sexually mature males killed a given year were recorded (above 30kg of dressed body mass corresponding to a full body mass of 38kg (Mauget and Boissin 1987; Baubet 1998)).

Molecular and paternity analysis

All samples were genotyped for 12 microsatellite loci (see Chapter IVA for more details). Individuals, whose genotyping failed, were excluded from analysis, including the whole mother-litter couples if the genotype of the mother or a fetus was not obtained. This reduced the dataset to 871 putative males, 202 litters and 1049 fetuses. All genotypes obtained were

analyzed using the software COLONY (Jones and Wang 2010). As mothers were already known, it was used to identify fathers among putative males from the hunting bag or to assign a genotype if the father was not sampled. The analyses were realized for each year, including as possible fathers all sampled males from year n, subadult and adult males killed year n+1 and adult males from year n+2 considering the probability of being killed each year for males is very high in this population (Toïgo et al. 2008). This lead to include 141, 233, 307, 355, 417, 273, 161, 120 and 89 sampled males in the analysis, from 2007 to 2015 respectively. We considered 50% of reproductive males were sampled each year. We performed analysis for other proportions of sampled males and the number of fathers was not influenced by the proportion between 20% and 70%. For all analyses, we considered both sex polygynous, did not set population allele frequency, and used the full likelihood analysis method with a medium precision. The markers were chosen codominant, with an allelic dropout of 0.001 and a typing error of 0.01 for each locus. The default values were chosen for all other parameters. The results allowed us to estimate the number of fathers in each litter and then to know if the litter was sired by one or multiple males.

Statistical analysis

To estimate the proportion of big males in the hunting bag each year, we choose to take advantage of the long-term monitoring realized by the French Hunting and Wildlife Agency on this population. As the weight of all adult males killed by hunting since 1982 was available, we determined the 80% percentile value of weight. This allowed us to estimate a historical threshold of the weight of the biggest adult males in the population. We then measured the proportion of sexually mature males killed each year with a weight higher than this threshold (*Prop*). This parameter quantifies the relative number of big males killed each year. As the threshold is fixed through years, we directly recorded the median dressed weight of sexually mature males killed at hunting (*Medw*) to measure how the weight of an average reproductive male varies between years.

We performed two types of generalized linear mixed-effect models to investigate parameters influencing the probability of a litter to be multiple sired and the number of males that sire a litter. Indeed, the proportion of litter displaying multiple paternity can vary greatly but the number of fathers may only slightly change. For the first, we used a binomial regression model using the litter status (*MP*) as response variable (0 for single sired litters, 1 for litters with multiple paternity). For the second, we used a Poisson regression model using the number of fathers (*Nf*) as response variable. We included the mother body mass (*BMm*), the Julian date

of mating (*D*), and the proportion of each of the three items in stomach content of the month when mating occurred: acorn (*A*), beechnut, (*B*) and corn (*C*). This reduced the sample size to 149 litters as some litters where conceived out of the hunting period when information for stomach content is not available. We also added in the model the number of wild boars killed by surface unit (*Nha*) and the parameters *Prop* and *Medw* described above as biological effects. The year was included as a random factor. The litter size (*LS*) was included as a confounding variable in the model investigating for the number of fathers as we know *Nf* and *LS* are linked (Chapter IVA). All numeric variable were scaled before analysis. The collinearity between variables in the models was verified using the Variation Inflation Factor (VIF) with a threshold of three (Zuur et al. 2009). As stomach proportion of acorn and beechnut were highly correlated, we summed them to create a parameter forest fruits (*FF*). Parameters with a VIF value above 3 were removed from full models starting from the proportion of corn (*C*) as this parameter was included only as a co-factor, followed by parameters showing the highest VIF value. The model selection started with the full additive model and the best models explaining the data were selected based on AICc (ΔAICc < 2) and averaged (Burnham and Anderson 2004). All statistical analyses were performed using R software version 3.3.3 (R Core Team 2017).

Results

The proportions of forest fruits and corn in stomach content varied during the study period. When proportions of forest fruit where high, the proportions of corn was low and vice versa.

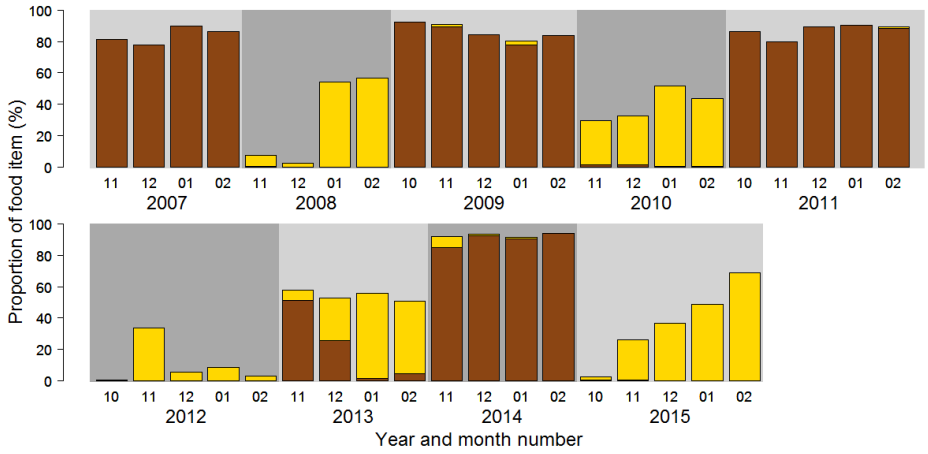


Figure III 5 Evolution of the proportion of forest fruits (*FF*, brown) and corn (*C*, yellow) in wild boar stomach content killed by hunting over the study period. Months are represented by their number (10 for October, 11 November). The grey background separates hunting seasons.

The total number of wild boars killed per unit of space each year ranged from 6.1 to 9.34 individuals per 100ha (median= 7.22 ± 1.08 SD, Figure III 6a). The proportion of males heavier than the threshold weight corresponding to the 80% weight value obtained by adult males killed since 1982 showed a median of 3.08% ± 2.59 SD (min= 0.54%, max= 7.38%, Figure III 6c). The median of the weight of reproductive males killed at hunting ranged from 37.85 to 53.3 kg (median = 45.2.56 ± 6.08 SD, Figure III 6d).

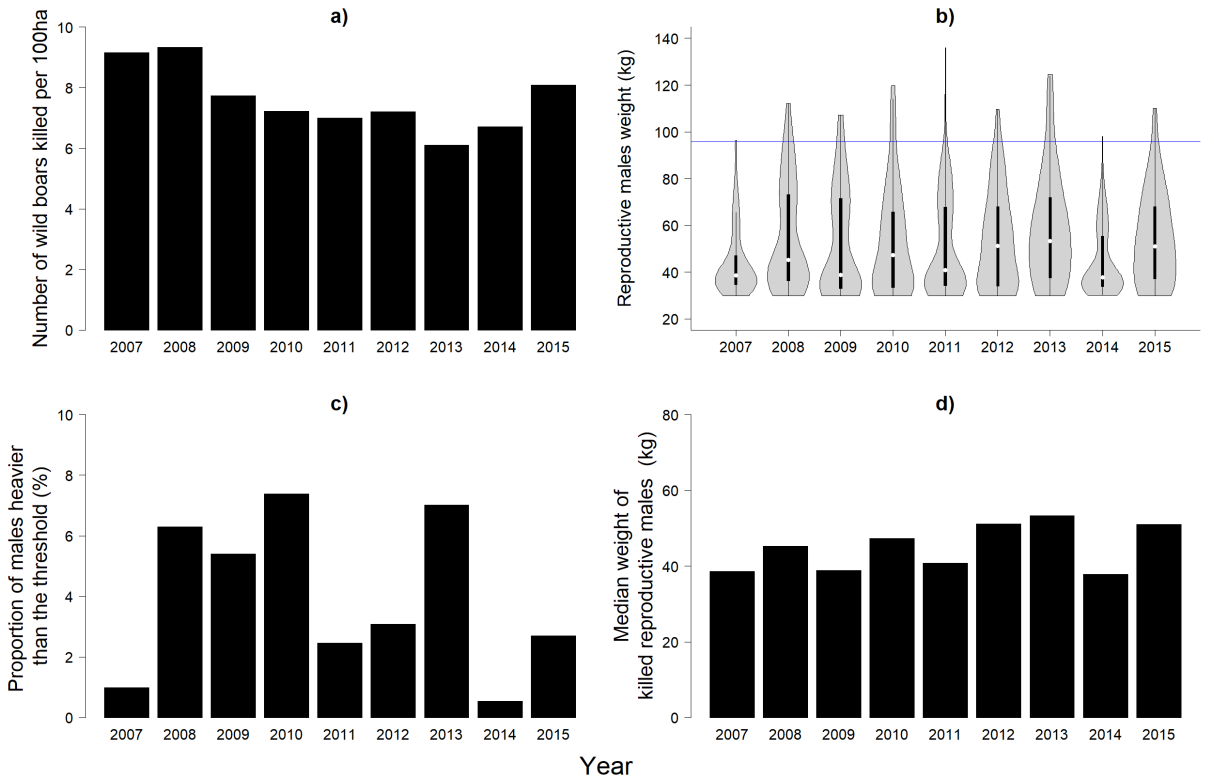


Figure III 6 a) Yearly number of wild boars killed by hunting, b) Dressed body weight of reproductive males wild boars killed by hunting according to the year. The horizontal blue line represent the 80% threshold mass obtained from historical analysis of adult wild boar males since 1982. The width shows the density of points. The white points show the median, the thick black vertical lines the central quartiles and the fine vertical lines 1.5 the inter-quartile space. c) Yearly proportion of males killed by hunting with a dressed body mass above the historical 80% threshold of the weight of adult males since 1982 (*Prop*). d) Yearly median weight of reproductive males killed by hunting (*Medw*).

Both the proportion of corn (*C*) and the proportion of forest fruits (*FF*) showed high co-linearity with other parameters included in the models (VIF > 3), so they were excluded from the full models. Two models were best supported by the data (ΔAICc < 2) for the *pMP* and were averaged (supplementary material Table III S5a). They included all the parameters of the full model as it was the second best model. The Julian date of mating (*D*, β = 0.624 ± 0.223, p = 0.006, Table III 4a) and the yearly proportion of big males killed at hunting (*Prop*, β = 0.838 ± 0.246, p = 0.001, Table III 4a) showed significant positive effects on the *pMP* within a litter. The yearly number of wild boars killed (*Nha*, β = -0.457 ± 0.207, p = 0.028, Table III 4a) and the median weight of reproductive males (*Medw*, β = -0.476 ± 0.227, p = 0.038, Table III 4a) had significantly negative effects. The mother body mass did not influence significantly (BM_m, β = 0.116 ± 0.195, p = 0.553, Table III 4a) the *pMP* in a litter. Concerning the number of fathers in a litter, six models were best supported by the data and were averaged (supplementary

material Table III S5b). *Nha* was excluded from the best models. Only the yearly proportion of big males killed showed a significantly positive effect (*Prop*, $\beta = 0.374 \pm 0.177$, $p = 0.035$, Table III 4b). The other parameters did not show significant effect on the number of fathers within a litter (Table III 4b).

Table III 4 Estimates, standard errors, z statistics, and P values of parameters linked with the probability of occurrence of multiple paternity (a.) and the number of fathers (b.) in a litter. Values for the Julian date of mating (*D*), the proportion of males killed with a dressed weigh above the historical threshold (*Prop*) and median dressed weight of reproductive males (*Medw*), the number of wild boar killed per surface unit (*Nha*), the mother body mass (BM_m) and the litter size (*LS*) were obtained from averaged models strongly supported by the data. Significant parameters are in bold ($n_{litters} = 149$).

a.

Parameter	Estimate	SE	z-test statistic	P-value
Intercept	0.860	0.205	-	-
D	**0.624**	**0.223**	**2.774**	**0.006**
Prop	**0.838**	**0.246**	**3.377**	**0.001**
Nha	**-0.457**	**0.207**	**2.192**	**0.028**
Medw	**-0.476**	**0.227**	**2.076**	**0.038**
BM_m	0.116	0.195	0.594	0.553

b.

Parameter	Estimate	SE	z-test statistic	P-value
Intercept	0.769	0.057	-	-
D	0.027	0.053	0.517	0.605
Prop	**0.196**	**0.059**	**3.295**	**0.001**
Medw	-0.027	0.055	0.498	0.619
BM_m	-0.003	0.023	0.144	0.886
LS	0.093	0.064	1.440	0.150

Discussion

In the population of wild boars of Châteauvillain, we observed that Julian date of mating (*D*) and the yearly proportion of big males killed (*Prop*) showed a positive effect on the probability of multiple paternity (*pMP*) within a litter. The hunting intensity measured as the number of wild boars killed a given year (*Nha*) and the median weight of reproductive males (*Medw*) showed negative effects on multiple paternity probability. Only *Prop* had also a positive effect on number of fathers in a litter. Parameters included as confounding factors (BM_m and *LS*) showed no significant effect on any parameters.

The resources proportion measured as the proportion of corn (*C*) and forest fruits (*FF*) in stomach contents were highly correlated with other parameters, especially the yearly median

weight of reproductive males (*Medw*), so they were excluded from the models presented here. This did not allowed to test the influence of resource availability on the mating system. To perform this analysis, we removed *Medw* from the full models and we obtained VIF values below three for *FF* (*C* was still removed based on VIF values). In this case, *FF* was maintained in averaged models but it did not influenced significantly the *pMP* within a litter, and neither the number of sires in a litter (supplementary material Table III S6). Despite studies demonstrating the link between resources and multiple paternity in mammals (Asher et al. 2008; Cameron et al. 2011), we did not find such effect in this wild boar population. This does not fit our prediction as these results show that variations in forest fruits availability does not influence the mating system. In this population, the range of variations of food availability may be buffered by supplementary feeding. In years of poor mast production, the proportion of corn increased in stomach contents and we know that hunters provide more corn in the forest to avoid crop damages. Thus, female wild boar may aggregate around feeding places the same way they aggregate to feed on forest fruits. Also, the wild boar is a highly opportunistic feeder contrary to other ungulates, so distribution of food resources may be a poor estimator of mating system variations.

Both the *pMP* within a litter and the number of fathers in a litter were positively associated with the yearly proportion of big males in the hunting bag *Prop*. This result fits our prediction. As the proportion of big males killed increases, their number in the population decreases. Moreover, big males are often killed at the beginning of the hunting season which start before of the rutting period of wild boar (Mauget and Boissin 1987). The positive effect of the mating date (*D*) on *pMP* supports this hypothesis. Indeed, it suggests that females are less monopolized as the hunting season progresses. As big males are the most suitable to monopolize females and defend them against competitors, their removal by hunting allows other males, that would not reproduce if big males were present, to have access to females (Hogg and Forbes 1997). The increase of *pMP* and *Nf* observed fit the theory of increased of number of possible mating partners over the hunting season. Positive effect of the males density has already been reported to influence multiple paternity rate (Martin et al. 2014; Wells et al. 2017). However, the lack of effect of the mating date on the number of fathers suggests that *Nf* does not increase with the opportunity of multiple male mating. This can be explained by the increasing removal of males during hunting. Females are less monopolized by big males but, in the same time, less males are available in the population for mating. However, it is noteworthy that high values of *Prop* could also reflect higher proportion of males in the population and hunting bag would be a

proxy of the structure of the global population but *Prop* and the median weight of reproductive males (*Medw*) were not highly correlated based in VIF analysis. Thus, the proportion of big males killed is not linked to the yearly median weight of males, the positive effects of *Prop* on *pMP* and *Nf* is trustworthy.

We observed a negative effect of *Medw* on *pMP* but not on *Nf*. High median weight of males killed can be link to two different scenarios. First, high *Medw* values imply that heavy males are removed of the population and the proportion of small and/or light males remaining increases. This should release between male competition and be associated with higher multiple paternity rates (Zedrosser et al. 2007) contrary to what is observed here. Second, high *Medw* values can also be associated with high weight of males in the population a given year. In this scenario, the hunting bag reflect the population trend. This second possibility is supported by correlation between *Medw* and *FF*. Increase of median body mass of males means higher number of males of good quality. Good quality males are able to defend females against competitors, decreasing multiple male mating opportunity, leading to a decreased *pMP* (Singer and Zeigenfuss 2002; Zedrosser et al. 2007). Moreover, *Nha* showed a similar pattern to *Medw* with a negative effect on *pMP* but no effect on *Nf*. This can be explain because years of mast production allow better body condition leading to increased survival of wild boar, especially for young males (Focardi et al. 2008), and then increased hunting bags. However, the lack of effect of both *Medw* and *Nha* on *Nf* shows that number of fathers does not varies with average males quality and number of animal killed in the population, which temper this previous deduction as we could expect a decrease of the number of father when *pMP* decreases. To verify more precisely mechanism influencing male access to females and their probability to reproduce, identifying all fathers (reproductive males achieving reproduction) and obtaining their morphologic characteristics at the time of mating would be required. However, our data do not allow that (but see Chapter V for exploratory analysis) and such study suggests intensive and costly sampling and analysis procedures in such a large population.

In this wild boar population, variations of availability of food resource did not directly influence mating system but may indirectly modulate male quality and the intra-sexual competition. Also, the yearly proportion of big males killed, which are the most suitable to monopolize females, lead to increase of both probability of multiple paternity and number of mating partners of females. Altogether, these results suggest that mating system in wild boar is mainly influenced by the capacity of males to defend females against competitors. This capacity depends on hunting which change intra-sexual competition between males. Investigating in

detail which characteristics of male wild boars influence their reproductive success should highlight interesting results for both evolutionary biology and management researchers.

Acknowledgment

We thank all those who helped collect harvested wild boars, the ONF (Office National des Forêts), and F. Jehlé who allowed us to work in the study area. This work was supported by the French National Hunting and Wildlife Agency (ONCFS) and by the University of Lyon. ONCFS contributes also by supporting TG, providing his Ph.D. fellowship. COLONY analyses were performed, using the computing facilities of the CC LBBE/PRABI.

Supplementary Material

1. Model selection

Table III S5 Model selection to test the effect on (a.) the probability of occurrence of multiple paternity in a litter and (b.) the number of fathers in a litter of the Julian date of mating (*D*), the proportion of males killed with a dressed weigh above the historical threshold (*Prop*) and the median dressed weight of reproductive males (*Medw*), the mother body mass (BM_m), the litter size (*LS*), and the number of wild boar killed per surface unit (*Nha*) in the population of Châteauvillain, France. The model retained is in bold (ΔAICc<2). 'X' denotes that the explanatory variable was included in the model ($n_{litters}$ = 149).

a.

Intercept	*D*	BM_m	*Prop*	*Nha*	*Medw*	# parameter	Log Likelihood	AICc	ΔAICc	AICc weight
X	**X**		**X**	**X**	**X**	**6**	**-80.96**	**174.51**	**0.00**	**0.23**
X	**X**	**X**	**X**	**X**	**X**	**7**	**-80.15**	**175.08**	**0.58**	**0.17**
X	X		X	X		5	-83.14	176.69	2.19	0.08
X	X	X	X	X		6	-82.08	176.75	2.24	0.07
X	X		X			4	-84.26	176.81	2.30	0.07
X	X		X		X	5	-83.29	177.00	2.49	0.06
X	X	X	X			5	-83.32	177.06	2.55	0.06
X	X	X	X		X	6	-82.59	177.77	3.26	0.04
X	X	X		X		5	-84.27	178.96	4.45	0.02
X			X	X		4	-85.37	179.03	4.52	0.02
X			X			3	-86.51	179.19	4.68	0.02
X	X	X				4	-85.52	179.32	4.82	0.02
X	X			X		4	-85.77	179.81	5.31	0.02
X	X					3	-86.88	179.93	5.42	0.01
X			X	X	X	5	-84.91	180.24	5.73	0.01
X		X	X	X		5	-85.12	180.67	6.16	0.01
X		X	X			4	-86.22	180.71	6.21	0.01
X			X		X	4	-86.28	180.84	6.33	0.01
X	X	X		X	X	6	-84.21	181.00	6.50	0.01
X	X	X			X	5	-85.50	181.43	6.92	0.01
X	X			X	X	5	-85.72	181.86	7.35	0.01
X	X				X	4	-86.86	182.01	7.50	0.01
X		X	X	X	X	6	-84.77	182.12	7.61	0.00
X		X	X		X	5	-86.07	182.55	8.04	0.00
X						2	-89.40	182.89	8.38	0.00
X				X		3	-88.48	183.13	8.62	0.00
X		X				3	-88.71	183.59	9.08	0.00
X		X		X		4	-87.77	183.82	9.31	0.00
X					X	3	-89.27	184.70	10.19	0.00
X				X	X	4	-88.46	185.21	10.70	0.00
X		X			X	4	-88.56	185.39	10.89	0.00
X		X		X	X	5	-87.75	185.92	11.41	0.00

b.

Intercept	*D*	*LS*	BM_m	*Prop*	*Nha*	*Medw*	# parameter	Log Likelihood	AICc	ΔAICc	AICc weight
X		**X**		**X**			**4**	**-227.33**	**462.94**	**0.00**	**0.12**
X	**X**	**X**		**X**		**X**	**6**	**-225.57**	**463.73**	**0.78**	**0.08**
X	**X**	**X**		**X**			**5**	**-226.76**	**463.93**	**0.99**	**0.07**
X		**X**		**X**		**X**	**5**	**-226.83**	**464.08**	**1.14**	**0.07**
X				**X**			**3**	**-229.05**	**464.26**	**1.31**	**0.06**
X		**X**	**X**	**X**			**5**	**-227.21**	**464.84**	**1.89**	**0.04**
X		X		X	X		5	-227.28	464.97	2.03	0.04
X				X		X	4	-228.44	465.15	2.21	0.04
X	X	X		X	X	X	7	-225.19	465.17	2.23	0.04
X		X	X	X		X	6	-226.59	465.76	2.82	0.03
X	X	X	X	X		X	7	-225.51	465.82	2.87	0.03
X	X			X		X	5	-227.72	465.87	2.92	0.03
X	X	X		X	X		6	-226.66	465.90	2.96	0.03
X	X			X			4	-228.82	465.92	2.97	0.03
X		X		X	X	X	6	-226.67	465.93	2.98	0.03
X	X	X	X	X			6	-226.74	466.06	3.12	0.02
X			X	X			4	-228.91	466.10	3.16	0.02
X				X	X		4	-228.92	466.13	3.18	0.02
X				X	X	X	5	-228.16	466.74	3.80	0.02
X		X	X	X	X		6	-227.16	466.92	3.97	0.02
X	X			X	X	X	6	-227.23	467.05	4.11	0.01
X			X	X		X	5	-228.39	467.20	4.26	0.01
X	X	X	X	X	X	X	8	-225.16	467.34	4.40	0.01
X	X		X	X			5	-228.50	467.43	4.48	0.01
X	X		X	X		X	6	-227.48	467.55	4.60	0.01
X		X	X	X	X	X	7	-226.44	467.67	4.73	0.01
X	X			X	X		5	-228.65	467.73	4.78	0.01
X		X					3	-230.85	467.87	4.93	0.01
X			X	X	X		5	-228.78	467.99	5.04	0.01
X	X	X	X	X	X		7	-226.64	468.08	5.14	0.01
X	X	X					4	-230.06	468.40	5.45	0.01
X	X		X	X	X	X	7	-226.94	468.67	5.73	0.01
X			X	X	X	X	6	-228.12	468.83	5.88	0.01
X	X		X	X	X		6	-228.30	469.20	6.25	0.01
X		X			X		4	-230.47	469.22	6.28	0.00
X	X	X			X		5	-229.60	469.62	6.67	0.00
X		X	X				4	-230.82	469.92	6.98	0.00
X		X				X	4	-230.82	469.93	6.98	0.00
X							2	-233.18	470.43	7.49	0.00
X	X	X				X	5	-230.06	470.54	7.59	0.00
X	X	X	X				5	-230.06	470.54	7.59	0.00
X		X	X		X		5	-230.45	471.32	8.38	0.00
X		X			X	X	5	-230.47	471.37	8.42	0.00
X					X		3	-232.65	471.46	8.52	0.00
X	X	X			X	X	6	-229.52	471.64	8.70	0.00
X	X						3	-232.78	471.72	8.78	0.00
X			X				3	-232.79	471.74	8.80	0.00
X	X	X	X		X		6	-229.60	471.79	8.84	0.00
X		X	X			X	5	-230.80	472.01	9.07	0.00
X						X	3	-233.13	472.43	9.49	0.00
X	X		X				4	-232.13	472.55	9.60	0.00
X	X				X		4	-232.18	472.64	9.70	0.00
X	X	X	X			X	6	-230.06	472.71	9.76	0.00
X			X		X		4	-232.25	472.77	9.82	0.00
X	X		X		X		5	-231.48	473.37	10.43	0.00
X		X	X		X	X	6	-230.45	473.50	10.55	0.00
X					X	X	4	-232.65	473.57	10.63	0.00
X			X			X	4	-232.73	473.74	10.80	0.00

X	X				X	4	-232.77	473.82	10.88	0.00
X	X	X	X	X	X	7	-229.52	473.84	10.89	0.00
X	X		X		X	5	-232.13	474.68	11.73	0.00
X	X			X	X	5	-232.15	474.71	11.77	0.00
X			X	X	X	5	-232.25	474.91	11.97	0.00
X	X		X	X	X	6	-231.43	475.45	12.51	0.00

2. Model results including food resources

Table III S6 Estimates, standard errors, z statistics, and P values of parameters linked with the probability of occurrence of multiple paternity (a.) and the number of fathers (b.) in a litter. Values for the Julian date of mating (*D*), the proportion of males killed with a dressed weigh above the historical threshold (*Prop*) and the proportion of forest fruits in stomach content (*FF*), the number of wild boar killed per surface unit (*Nha*), the mother body mass (BM_m) and the litter size (*LS*) were obtained from averaged models strongly supported by the data. Significant parameters are in bold ($n_{litters}$ = 149).

a.

Parameter	Estimate	SE	z-test statistic	P-value
Intercept	0.831	0.238	-	-
D	**0.561**	**0.242**	**2.297**	**0.022**
Prop	**0.829**	**0.294**	**2.802**	**0.005**
Nha	-0.196	0.240	0.810	0.418
FF	0.255	0.300	0.845	0.398
BMm	0.118	0.197	0.596	0.551

b.

Parameter	Estimate	SE	z-test statistic	P-value
Intercept	0.770	0.057	-	-
Prop	**0.194**	**0.060**	**3.230**	**0.001**
D	0.027	0.055	0.491	0.624
FF	0.017	0.049	0.355	0.722
BMm	-0.004	0.024	0.152	0.879
LS	0.091	0.065	1.397	0.162

Chapter IVA
Consequences of multiple paternity

On the evolutionary consequences of increasing litter size with multiple paternity in wild boar (*Sus scrofa scrofa*)

Abstract: Understanding how some species may be able to evolve quickly enough to deal with anthropogenic pressure is of prime interest in evolutionary biology, conservation and management. Wild boar (*Sus scrofa scrofa*) populations keep growing all over Europe despite increasing hunting pressure. In wild boar populations subject to male-selective harvesting, the initially described polygynous mating system may switch to a promiscuous/polyandrous one. Such a change in the mating system, where potentially more males sire a litter at one reproductive event, may be associated with the retention of high genetic diversity and an increase of litter size. We tested these hypotheses by estimating the number of sires per litter based on a 6-year long monitoring of a wild boar population subject to particularly high harvesting pressure. Our results show a high and stable genetic diversity and high rates of multiple paternity compared to other populations, thus depicting a promiscuous/polyandrous mating system in this population. We also show that litter size is positively linked to the number of sires, suggesting that multiple paternity increases fecundity. We finally discuss that multiple paternity may be one of the factors allowing rapid evolution of this population by maintaining both genetic and phenotypic diversity.

Keywords*:* harvesting; polyandry; mating system; selective hunting; fecundity

Gayet T., S. Devillard, M. Gamelon, S. Brandt, L. Say, and E. Baubet. 2016. On the evolutionary consequences of increasing litter size with multiple paternity in wild boar (*Sus scrofa scrofa*). *Evolution* (N. Y). 70:1386–1397.

Introduction

Human exploitation, through hunting or fishing, affects the size of free-ranging populations. High harvesting pressures lead to the removal of a large proportion of individuals, inducing a strong yearly decline of population size. Since genetic diversity is linked to population size (Frankham 1996), intensively harvested populations undergo great genetic loss every year (Harris et al. 2002). This may affect their adaptive potential (Amos and Balmford 2001; Barrett and Schluter 2008) leading to a demographic decline and, in the worst case scenario, to extinction (Gilpin and Soulé 1986; Rosser and Mainka 2002; see Spielman et al. 2004 for a meta-analysis on 170 taxa) if the populations are unable to respond to the new selective pressures. For example, overexploitation is the main factor that induced the collapse of several fisheries in the past century (Hutchings and Myers 1995; Jackson et al. 2001). In addition, for populations facing anthropogenic pressures through exploitation, some modifications have also been observed in phenotypic traits (Coltman et al. 2003; Douhard et al. 2016), demography (Milner et al. 2007; Servanty et al. 2011) or genetic characteristics (Harris et al. 2002; Allendorf et al. 2008).

Selective harvesting, the intensification of harvest efforts geared toward individuals showing phenotypic traits favored by hunters (Milner et al. 2007), may affect the structure of populations. For instance, in populations subject to size-selective harvesting, where the largest adults are preferentially removed, the age and sex structure is biased toward the young and females (Milner et al. 2007). Such a change in age and sex distribution may have strong consequences on the mating system (Kokko and Rankin 2006; Milner et al. 2007). In ungulate populations, the mating system is known to be influenced by density of males and females (Isvaran 2005). For species showing a polygynous mating system with female monopolization by males, intra-sexual competition is diminished when larger males are removed and younger males are more likely to obtain paternities by harassment of females (Isvaran 2005). The mating system switches to a promiscuous/polyandrous one and the multiple paternity rate (defined as the proportion of litters showing more than one father) increases. As the number of males accessing reproduction grows, the variance of male reproductive success decreases, allowing greater genetic diversity to pass on from one year to the other compared to the polygynous mating system (Nunney 1993; Sugg and Chesser 1994; Pearse and Anderson 2009). We hypothesize that these processes (promiscuous mating system and multiple paternity) may have the potential to buffer the loss of genetic diversity due to intensive harvesting. However, multiple paternity effects have been the topic of a debate. On the one hand, in their review,

Jennions and Petrie (2000) described most of the genetic benefits for females, including the increase of genetic diversity within a litter, which tends to increase the genetic diversity at the population scale (Pearse and Anderson 2009). On the other hand, Lotterhos (2011) tempered this result by showing that the positive link between multiple paternity and effective population size, which reflects population genetic diversity, is not always true but depends on the litter size, the number of reproductive events and the female's number of mates over her lifetime.

Surprisingly, wild boars (*Sus scrofa scrofa*) do not fit into the classical frame of reduced population size due to intensive harvesting. Their populations are growing all over Europe despite the continuous increase of hunting pressure (Massei et al. 2015). The mating system of this species has been originally described as polygynous (Mauget 1980; Dardaillon 1984) which is consistent with the sexual dimorphism displayed by the species (Ralls 1977). Interestingly, with the rise of molecular genetic techniques, a growing literature has shown that multiple paternity occurs (Delgado et al. 2008; Poteaux et al. 2009) and may be common in some populations of wild pigs (Delgado-Acevedo et al. 2010; Costa et al. 2012). A positive effect of multiple paternity on litter size has even been highlighted in the domestic pig (*Sus scrofa domesticus*) which is the domestic counterpart of the wild boar. Due to its economic importance (Orr and Shen 2006), the pig is the subject of many studies that aim at understanding the mechanisms underlying reproduction to improve production. The number of artificial insemination events is known to have a positive effect on litter size (Kemp and Soede 1996; Corrêa et al. 2002). Moreover, several sires are commonly used in pig husbandry to increase litter size, hence productivity (Badinel 2010). However, until now, only a few studies have focused on the link between litter size and number of fathers in the wild (DiBattista et al. 2008; Thonhauser et al. 2014, but see Waller and Bilkei (2002) for evidence of larger litter sizes with increasing number of sires per litter in free-ranging pigs).

Recent studies conducted on the wild boar population of Châteauvillain-Arc-en-Barrois which suffers from a particularly high and male-selective harvesting pressure (Toïgo et al. 2008; Servanty et al. 2011), have shown that selection for both earlier birth date and earlier sexual maturity in females could occur over just a few generations to adapt to the harvesting regime (Gamelon et al. 2011; Servanty et al. 2009). But whether multiple paternity may (i) have a positive effect on wild boar fecundity through larger litter size as shown in pigs, and (ii) maintain a high genetic diversity through time, which is then transmitted to each generation, thereby buffering yearly genetic loss and allowing a high ability to respond to new selective pressure, remain understudied questions. Hence, using six years of data sampling, we addressed

these questions. We expected a high level of genetic diversity together with a high rate of multiple paternity in this population, compared to other populations of the same species, triggered by a likely disruption of the polygynous mating system. Moreover, we investigated how the litter size is related to the number of sires within litters and predicted a larger litter size when the number of sires increases.

Material and Methods

Study site and sample collection

The wild boar population is located in the 11,000 ha Châteauvillain-Arc-en-Barrois open forest (48°02′N; 4°55′E, France) surrounded by agricultural fields, thus immigration rate is low (unpublished data). The number of individuals was estimated to be between 1,200 and 1,500 (Gamelon et al. 2011). In this heavily hunted population, wild boars have a 40% probability of being shot every year, rising to 70% for adult males (Toïgo et al. 2008). Moreover, the population exhibits a particularly short generation time for an ungulate which was previously estimated to be 2.27 years (Servanty et al. 2011). During six hunting seasons (2007-2012), the number of wild boar killed annually ranged from 567 to 794 (mean 635). Tissue samples were collected from 165 hunted pregnant females and their full litters (845 fetuses, mean litter size = 5.1 ± 1.63 SD), 264 non-breeding females (included only in the genetic diversity analyses), and from 627 putative reproductive males with a dressed body mass (*i.e.* without the digestive system, heart, lungs, liver, reproductive tract and blood) higher than 30kg (Gamelon et al. 2012). Body mass has been shown to be a structuring factor, more appropriate than age for this species (Gamelon et al. 2012), so weight was recorded for individuals.

Molecular analysis

All tissue samples were stored in alcohol in an individual hermetic straight container of 25ml and then genotyped for 12 microsatellite loci (Supplementary Material 1, Table IV S3). For each sample, 20-80 ng/µl of total genomic DNA was extracted using a buffer lyse. A few milligrams of tissue were pounded and then incubated first at 56°C for 2-3 hours and then at 72°C for 20 minutes in 200µl volumes containing 4µl of Tris HCL 1M, 0.3µl of $MgCl_2$ 1M, 5µl of KCl 1M, 1µl of Tween 20 and 1µl of K proteinase (20mg/ml, EUROBIO). Selective amplification was carried out for 12 microsatellite loci divided into 2 PCR multiplexes by polymerase chain reaction (PCR). PCR were conducted in 96-well microtitre plates in final

volumes of 20µl containing 10µl of PCR Multiplex Master Mix (2x, QIAGEN), 0.6µl of each primer (10mM) and 2µl of the extraction product. PCR was conducted using a BIOBLOCK PTC 100 thermal cycler with the following program: 95°C/15 min, 30 cycles with 94°C/30 s, 57°C/1.30 min and 72°C/1 min denaturing, annealing and extension temperatures respectively, and finally 60°C/30 min. The sizes of PCR amplified products were resolved by GENOSCREEN (http://www.genoscreen.fr) using an APPLIED BIOSYSTEMS 3730xl DNA Sequencing Analyzer (Supplementary Material 1, Table IV S3). Three mother-litter pairs and two males were removed from the analysis due to the high number of missing genotypes.

Genetic and paternity analysis

CERVUS 3.0.7 (Kalinowski et al. 2007) was used to compare observed (*Ho)* and expected (*He)* heterozygosity (using only adult genotypes to avoid biases from family genetic links, Supplementary Material 1, Table IV S3), to identify mother-offspring mismatches, to estimate the null allele rate and to conduct maximum likelihood paternity analyses. CERVUS compares potential sires using mismatches in the fetus-mother-male trio and likelihood ratio scores. Males were considered to be a fetus's genetic sire when there was no father-offspring mismatch, when the fetus–mother–male trio had a positive likelihood of detection (LOD) and when a ΔLOD higher than the 80% critical likelihood ratio (determined by simulations (Supplementary Material 2)). Putative sires for fetuses sampled in a given hunting season *i* were composed of all males sampled during the hunting season *i*, all yearlings and adult males sampled in the hunting season *i+1*, and adult males sampled in the hunting season *i+2*. Overall, the numbers of putative sires were 141, 233, 307, 346, 332 and 122 for the six hunting seasons (2007-2012), respectively. The number of sires per litter N_C, estimated with CERVUS, was recorded as the number of identified fathers.

To circumvent CERVUS failures to identify sires for all fetuses within a litter, GERUD 2.0 (Jones 2005) was also used to provide a second measure of the minimum number of males contributing to each litter N_G. Using known maternal genotypes, GERUD calculates the minimum number of fathers contributing to a given litter by subtracting the known maternal alleles from fetus genotypes, simulating all possible paternal genotypes, and determining the combinations of the remaining alleles that yield the fewest possible sires (Jones 2005). The error rate was estimated by simulations using GERUDsim (Supplementary Material 2). Due to computational limitations, we used for each mother-litter array only the five most polymorphic loci showing no missing data, when possible. In 87 out of 160 mother-litter arrays, the five loci

were the five most variable ones overall. Among the 73 remaining mother-litter arrays, 45 were analyzed with one other locus, 19 using two other loci and nine with all loci showing no missing data. Within these constraints, we performed an exhaustive search for the number of possible combinations of fathers that could explain each progeny array and recorded the minimum number of sires for each litter N_G. Since GERUD uses exclusion to estimate the number of male genotypes contributing to a given progeny array, estimates using this program are considered very conservative and should never overestimate the number of sires N_G for a litter (Jones 2005).

To circumvent CERVUS failures to identify sires of all fetuses within a litter and GERUD conservatism, we complemented our analysis with a less conservative analytical approach based on the maximal number of paternal alleles. Each fetus inherits one allele from its father and one allele from its mother so that the maximum number of alleles in a monopaternal litter is four if both parents are heterozygous and a maximum of two paternal alleles would be identified if single paternity occurs. For each litter and each locus, we calculated the number of alleles, from which we retrieved the known number of maternal alleles to estimate the number of paternal alleles. The maximal number of paternal alleles over the 12 loci was retained to obtain N_{PA}. We acknowledge that this number is also conservative, as only one allele in a litter at a given locus does not preclude multiple paternities.

Three different estimates of the multiple paternity rate (*MPR*) could thus be obtained, allowing us to evaluate consistency: the proportion of litters having more than one putative sire (CERVUS, MPR_C), the proportion of litters having a minimum number of sires N_G higher than one (GERUD, MPR_G) and the proportion of litters having more than two paternal alleles for at least one locus (maximal number of paternal alleles approach, MPR_{PA}).

Statistical analysis

To identify the key-variables driving the variability of the litter size *LS* across the litters, we performed a Poisson regression model where the response variable was *LS* and in which the locus-specific *Ho* and the observed number of alleles *A* for the locus showing the maximal number of paternal alleles in each litter were included as confounding variables. The mother dressed body mass BM_m and, finally, the maximal number of paternal alleles N_{PA}, as a proxy of the number of fathers, were included as main biological effects. We started model selection from the full additive model and then we selected the model with the lowest AICc in order to get estimates and standard error for each predictor variable.

To ensure that the pattern of relationship revealed between *LS* and N_{PA} was not an artefact due to the positive structural relationship between N_{PA} and *LS* (*i.e.* the impossibility to observe a N_{PA} higher than *LS*), we used a permutation test. We performed 10,000 random permutations of N_{PA}, *Ho* and *A* values kept together as a triplet against the pair of *LS* and BM_m values in the range of possible values (permutations where N_{PA} was higher than *LS* were not allowed). Each permutated dataset was analyzed using the model selected from the analysis of the observed dataset (model with the lowest AICc) to obtain the averaged coefficient associated to N_{PA}. The effect of N_{PA}, obtained with our observed dataset, was tested by calculating the exact p-value, against the distribution of permuted values, using the method described by Phipson and Smyth (2010).

The same analyses were performed using N_G along with BM_m as biological effect in the model to explain *LS* variability. The same permutation approach was also carried out by permuting N_G values against *LS* and BM_m values in the range of possible values. The analysis was not performed with CERVUS due to the few litters with all fathers identified. Moreover, litters with few fathers are more likely to be fully resolved than litters with a higher number of fathers. All analyses were performed in R 3.1.3 software (R Core Team 2017).

Results

Paternity analyses and multiple paternity rate

Overall, mean allelic diversity was *A* = 11.25 alleles per locus, ranging from two to 25 and mean expected heterozygosity was *He* = 0.602 ranging from 0.125 to 0.891 (Table IV 2, see Supplementary Material 1, Table IV S3 for details). Ten out of 12 loci showed very small deviations from expected heterozygosity and 11 out of 12 a low frequency of null alleles (<0.05 per locus) (Supplementary Material 1, Table IV S3).

Two mothers showed loci mismatches with all their presumed offspring, leading us to consider that a sampling mistake occurred at the collecting site. Therefore, they were removed from further analyses. Overall, the sire of 44.77% of the fetuses was identified among the set of candidate fathers by CERVUS and 10% of the litters (i.e. 16 out of 160) were fully resolved. Albeit, CERVUS failed to identify any father for 23.75% of the litters (i.e. 38 out of 160). The number of sires N_C ranged from one to five for the litters with at least one father identified (mean N_C = 1.78 sires per litter ± 0.86 SD, n_{litter} = 122, Figure IV 1a). The results obtained with GERUD were very similar. The minimum number of sires N_G ranged from one to four (mean

N_G = 1.69 sires per litter ± 0.63 SD, n_{litter} = 160, Figure IV 1b). Using the maximal number of paternal alleles' approach, N_{PA} ranged from one to four (mean N_{PA} = 2.34 sires per litter ± 0.72 SD, n_{litter} = 160, Figure IV 1c). The multiple paternity rate obtained with GERUD (MPR_G = 0.606, n = 160) was higher than with CERVUS (MPR_C = 0.438, n = 16 fully resolved litters). With the last approach, the multiple paternity rate was the lowest (MPR_{NPA} = 0.338, n = 160).

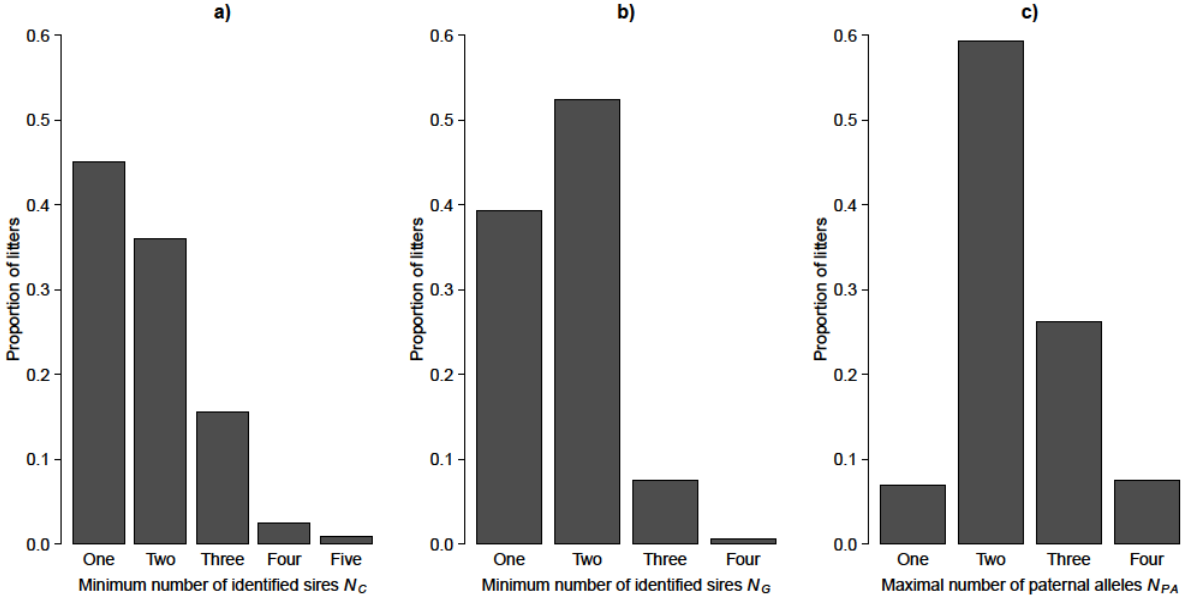


Figure IV 1 Distribution of the estimation of the minimum number of sires per litter using a) CERVUS (N_C, $n_{litters}$ = 122); and b) GERUD (N_G, $n_{litters}$ = 160) and c) the maximal number of paternal alleles per litter (NPA, $n_{litters}$ = 160).

Factors explaining the variability of N_{PA} and N_G

One model including the mother body mass BM_m, the observed number of alleles A for the locus showing the maximal number of paternal alleles and the maximal number of paternal allele N_{PA} was supported by the data (ΔAICc < 2, Supplementary Material 3, Table IV S4). LS increased significantly with BM_m (β = 0.010 ± 0.002, p < 0.001, Table 2a) and A (β = 0.017 ± 0.007, p = 0.021, Table IV 1a). Once the effect of BM_m and A was removed, LS was positively linked to N_{PA} (β = 0.112 ± 0.047, p = 0.018, Figure IV 2a). The probability for the random effect of N_{PA} on LS to be greater than the observed effect of N_{PA} was small ($p(N_{PA\ permuted} > N_{PA\ observed})$ = 0.002 with 10 000 values of beta from the permutated data set). The positive influence of increasing value of N_{PA} on LS was significantly higher than the basal link (Figure IV 2c).

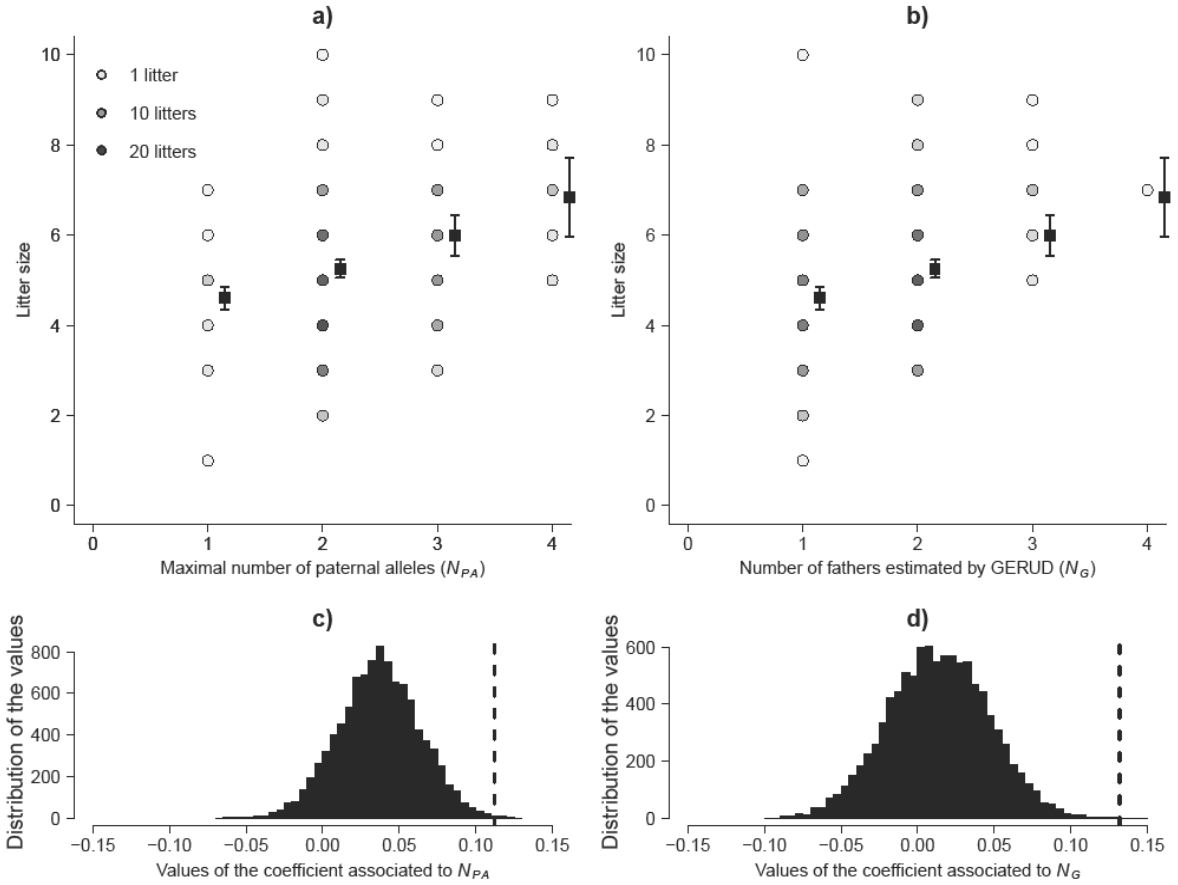


Figure IV 2 Effect of the number of sires per litter estimated by a) the maximal number of paternal alleles (N_{PA}) and b) GERUD (N_G), on the litter size ($n_{litters}$ = 160). Circles, whose colour indicates the number of litters, correspond to observations, and squares (± standard error) correspond to predicted litter sizes. Note that predicted values were obtained assuming a mother body mass (BM_m), and a number of alleles *A* equal to the mean observed values (i.e. 50.24kg, and 21.35, respectively).Distribution of the values of the coefficient associated to **c)** N_{PA} and **d)** N_G, obtained from 10,000 random permutations of the dataset. The dashed lines correspond to the observed averaged values of the coefficients (N_{PA} = 0.112 and N_G = 0.132). Note that $p(N_{PA\ permuted} > N_{PA\ observed}) = 0.002$ with 10,000 values of $N_{PA\ permuted}$ and $p(N_{G\ permuted} > N_{G\ observed}) < 0.001$ with 10,000 values of $N_{G\ permuted}$.

Only the full model was supported by the data with the GERUD approach (Supplementary Material 4, Table IV S5). BM_m was positively linked with *LS* ($\beta = 0.011 \pm 0.002$, $p < 0.001$, Table 2b). Again, *LS* increased significantly with the number of fathers ($\beta = 0.132 \pm 0.054$, $p = 0.015$, Table IV 1b, Figure IV 2b) and the probability for the random effect of N_G to be greater than the observed effect of N_G on *LS* was small ($p(N_{G\ permuted} > N_{G\ observed}) < 0.001$, with 10 000 values of beta, Figure IV 2d). Thus, the positive effect of N_G was significantly higher than the basal link between *LS* and the number of sires. Overall, the higher the number of sires in a litter, the larger the litter size.

Table IV 1 Estimates, standard errors, z statistics and p values of parameters linked with litter size (*LS*). a) Values for the number of alleles *A*, the mother dressed body mass BM_m and the maximal number of paternal alleles N_{PA} were obtained from the model strongly supported by the data (Table S3). b) Values for the mother dressed body mass BM_m and the number of father estimated by GERUD (N_G) were obtained from the model strongly supported by the data (Table S4). Significant parameters are in bold ($n_{litters}$ = 160).

a)

Parameter	Estimate	Standard error	z-test statistic	p value
Intercept	0.475	0.222	2.14	-
A	**0.017**	**0.007**	**2.30**	**0.02**
$\boldsymbol{BM_m}$	**0.010**	**0.002**	**4.34**	**<0.001**
$\boldsymbol{N_{PA}}$	**0.112**	**0.047**	**2.36**	**0.02**

b)

Parameter	Estimate	Standard error	z-test statistic	p value
Intercept	0.852	0.153	5.57	-
$\boldsymbol{BM_m}$	**0.011**	**0.002**	**4.63**	**<0.001**
$\boldsymbol{N_G}$	**0.132**	**0.054**	**2.44**	**0.015**

Discussion

Our results show that the average number of alleles per locus and heterozygosity were high and moderate, respectively, in this wild boar population, despite intensive hunting. The rates of multiple paternity varied between 0.338 and 0.606 depending on the approach used for estimating the number of sires. Regardless of the method used, we found larger litter sizes with increasing number of sires.

Table IV 2 List of the eight papers including the present study dealing with multiple paternity in *Sus scrofa* populations (on 25 September 2015). The subspecies (*S.s.d.* for *S.s.domesticus* corresponds to feral pig populations, *S.s.s* for *S.s.scrofa*), the sample size, the number of microsatellite loci (N_{loci}), the mean number of alleles (*A*), the mean allelic richness (*Ar*, obtained from 1,000 random subsamplings of our dataset corresponding to the sample size of the study cited), the mean observed heterozygosity (*Ho*), the mean expected heterozygosity (*He*), the mean number of litters (N_{litter}), the mean litter size (*LS*), the multiple paternity rate (*MPR*) with the number of litters showing multiple paternity in brackets, information about harvesting pressure found in the study and the reference of the study (with information about population´s location when required) are provided. Values in brackets show the range of the values when available. Bold values for *A*, *Ar*, *Ho* and *He* show values higher than the ones obtained in the present study. Bold values for *MPR* show values higher than the mean value of our three methods (i.e. 46%).

Subspecies	Sample size	N_{loci}	*A*	*Ar*	*Ho*	*He*	$N_{litters}$	*LS*	*MPR*	Harvesting pressure information in the study	Study (Population)
S.s.d	354	14	8.14 [4-17]	9.96 [2-25]	0.575 [0.367-0.756]	**0.68** [0.504-0.833]	11	5.64* [3-10]	0% (0)	*A population 'sampling rate' of 70% was assumed on the basis of published estimates of feral pig capture rates from studies that used very similar trapping methods to those used in this study.*	Hampton et al. (2004)
S.s.d	55	13	4.62 [2-6]	7.55 [1-20]	**0.663** [0.333-0.889]	**0.641** [0.5-0.822]	21	No data	**48%** (10)	*Prior to our study, this region was subjected to at least two years of intensive feral pig control, mainly through aerial baiting with 1080 (sodium monofluoroacetate) baits.*	Spencer et al. (2005)
S.s.d	409	12	8.3 [4-12]	10.12 [2-25]	**0.595** [0.294-0.78]	**0.68** [0.39-0.838]	2	5.4** [2-11]	**50%** (1)	*Feral pig were trapped, harvested or removed by aerial and ground shooting as part of damage-control management activities and sport hunting.*	Delgado-Acevedo et al. (2010) (Brooks)
							5		40% (2)		Delgado-Acevedo et al. (2010) (Cameron)
							2		0% (0)		Delgado-Acevedo et al. (2010) (Coryell)
							8		13% (1)		Delgado-Acevedo et al. (2010) (Dimmit)
							1		0% (0)		Delgado-Acevedo et al. (2010) (Hidalgo)
							1		**100%** (1)		Delgado-Acevedo et al. (2010) (Kerr)
							11		27% (3)		Delgado-Acevedo et al. (2010) (Kleberg)
							16		38% (6)		Delgado-Acevedo et al. (2010) (McMullen)
							18		39% (7)		Delgado-Acevedo et al. (2010) (San Patricio)

(Continued)

Subspecies	Sample size	N_{loci}	*A*	*Ar*	*Ho*	*He*	$N_{litters}$	*LS*	*MPR*	Harvesting pressure information in the study	Study (Population)
S.s.s	9	7	4.14 [3-6]	4.63 [1-13]	**0.603** [0.339-0.731]	0.576 [0.443-0.707]	9	5.56*** [5-7]	11% (1)	*The hunting pressure is high and so is the number of wild boars taken per 100ha of shooting area (Fernandez-Llario et al. 2003)* From this study: 9.49 ± 8.65/100 ha, n=17 hunt, mean ± SD)	Delgado et al. (2008)
S.s.s	488	12	6.58 [2-16]	10.32 [2-25]	0.518 [0-0.83]	0.552 [0.21-0.87]	21	4.05 [2-6]	10% (2)	40% probability of being shot up to 70% for males from Toigo et al. 2008 5.21/100 ha (±2.66, range: 0.64 -9.27) from Servanty et al. 2009	Poteaux et al. (2009) (Population of the study)
S.s.s	49	14	6.21 [3-14]	7.42 [1-20]	**0.718** [0.444-1]	**0.698** [0.396-0.901]	5	5.8	40% (2)	No information	Costa et al. (2012) (Hungary)
	72		5.07 [3-10]	7.95 [2-21]	0.542 [0.268-0.861]	0.552 [0.259-0.776]	5	6.2	20% (1)	No information	Costa et al. (2012) (Portugal)
	46		5.5 [4-12]	7.3 [1-20]	**0.647** [0.364-0.909]	**0.646** [0.411-0.906]	5	4.8	40% (2)	No information	Costa et al. (2012) (Spain)
S.s.s	181	14	9.64 [5-24]	10.76 [2-25]	0.548 [0.451-0.833]**	0.553 [0.412-0.837]**	27	4.3	Different estimates not directly comparable with ours (MPR defined as number of sires per sow)	Past experience with the study areas suggests a sampling intensity between 10–20%	Pérez-González et al. (2014) (Spain Western Iberian Peninsula)
	188				**0.658** [0.451-0.833]**	**0.632** [0.412-0.837]**	35	3.9			Pérez-González et al. (2014) (Spain Azagala)
	74				**0.634** [0.451-0.833]**	**0.632** [0.412-0.837]**	13	3.5			Pérez-González et al. (2014) (Spain Santa Amalia)
	260				**0.683** [0.451-0.833]**	**0.692** [0.412-0.837]**	35	5.9			Pérez-González et al. (2014) (Hungary)
S.s.s	1054	12	11.25 [2-25]	-	0.590 [0.107-0.845]	0.602 [0.125-0.891]	160	5.1 [1-10]	MPR_G = 61% MPR_C = 44% MPR_{NPA} = 34%	40% probability of being shot up to 70% for males from Toigo et al. 2008 5.21/100 ha (±2.66, range: 0.64 -9.27) from Servanty et al. 2009	Present study

*only litters with three or more piglets/fetuses were analyzed

** no data per population available

***only litters with five or more fetuses were selected

The average number of alleles and the allelic richness we reported are the highest among all the studies dealing with multiple paternity in wild boar (Table IV 2). It is noteworthy that the average number of alleles is sensitive to the sample size (our sample size is twice as big as the largest dataset), and both the average number of alleles and the allelic richness may vary with the loci analyzed (Table IV 2). Regarding the average heterozygosity, we found a moderate value (*He* = 0.602) compared to other studies (Table IV 2). Remarkably, this value is closer to the average heterozygosity reported for 14 non-threatened taxa (*He* = 0.699) than to the one reported for their 14 taxonomically-related threatened taxa (*He* = 0.407) provided in Frankham et al. (2002). Therefore, despite the strong hunting pressure, the genetic characteristics of this population are similar to those of other wild boar populations (Table IV 2) characterized, for some of them, with a weaker hunting pressure. Thus, our studied population definitely does not show any characteristics of endangered taxa. However, we acknowledge that despite the fact that comparing heterozygosities is less sensitive to sample size than comparing allelic richness, the comparison may still be sensitive to the loci used. Interestingly, four microsatellite loci used in our study have also been used by Poteaux et al. (2009) on data collected between 1999 and 2001 in the same population. The expected heterozygosity remains constant through time according to the four common loci ($He_{1999\text{-}2001} = 0.518$ *versus* $He_{2007\text{-}2012} = 0.548$) while the average allelic richness is higher ($A_{1999\text{-}2001} = 7.75$ *versus* $Ar_{2007\text{-}2012} = 9.64$ from 1,000 subsamplings of 488 individuals). Thus, both allelic number and expected heterozygosity showed no decrease over time despite the fact they are separated by at least twice the length of the generation time of the population and six hunting seasons. Such findings highlight that this heavily hunted population does not display any evidence of genetic loss over time on the studied loci.

Around 60.6% of the litters showed multiple paternity with GERUD, 43.7% with CERVUS, and this rate was only 33.8% with the maximal number of paternal alleles' approach. This might suggest that the multiple paternity rate is underestimated with this last method for which at least three paternal alleles have to be identified within a litter to classify it as a litter with multiple paternity. However, in our dataset, most of the litters display two paternal alleles, which could be obtained with one or more fathers. To our knowledge, these rates of multiple paternity obtained from the first long-term study at the population level, are among the highest ever reported in the species (Table IV 2). Multiple paternity rates are high (Table IV 2) but they likely underestimate the proportion of females that mate with more than one male. Indeed, multiple paternity rates only measure the number of successful matings that lead to multiple

sired litters, and do not correspond to the proportion of females that mate with more than one male. Such a proportion may potentially be higher than the reported multiple paternity rates, suggesting that the mating system in this population is predominantly promiscuous/polyandrous.

In this population where intensive hunting pressure, especially targeting males, occurs for a long time relative to the short generation time (2.27 years, Servanty et al. 2011), we observed both promiscuous/polyandrous mating system and stable genetic characteristics. We raise the hypothesis that such a mating system might have appeared as an evolutionary response to high hunting pressure due to the lack of dominant males, and be preserved since it has the ability to maintain high genetic variability within a litter (Pérez-González et al. 2014). It is also possible that this mating system appeared due to a tendency for females to mate promiscuously in the absence of dominant males, and it is preserved by the continual removal of large dominant males from the population. Equations from Nunney (1993) showed that multiple paternity (likened through random union of gametes) can increase effective population size by 10% and up to 50% when compared to harem polygyny depending on the proportion of males in the population. This was measured considering a constant population size, a generation time of 2.5 years, harem sizes of 1 (monogamy) and 5 females, thus consistent with group size of female wild boars (Dardaillon 1988; Podgórski et al. 2014). Thus, we hypothesize that the high rates of multiple paternity measured in our study favor the retention of a high genetic diversity through year. Unfortunately, without genetic monitoring of our population before the beginning of intensive hunting, it is impossible to quantify the change of the mating system and its influence on genetic diversity. Moreover, the population studied here is non-fenced and thus open to emigration and immigration. Even if the immigration rate is known to be low in our population (unpublished data), it is difficult to unravel the relative contribution of mating system and migration to the genetic diversity. We strongly encourage further studies to investigate the link between hunting pressure and mating systems. One exciting perspective could be to analyze multiple paternity rates among populations with different hunting pressures, as well as in non-hunted ones, to strengthen the link between hunting intensity and mating system.

After some evidence in domestic pigs (e.g. Waller and Bilkei 2002), we provide here the first empirical evidence of a positive link between multiple paternity and litter size in a free-ranging population. This finding was supported by the permutation tests, which showed that the observed relation is stronger than any structural relationship between litter size and the number

of sires. The pig illustrates the capacity of this species to cope with strong directional selective pressures (Gepts and Papa 2002). It is now well known that the time lapse for optimal fertilization is very short in domestic sows (Soede et al. 1995; Nissen et al. 1997). Therefore, the probability of presence of healthy sperm at ovulation time in the female genital tract increases with the number of artificial insemination events, thereby maximizing the number of fertilized ovules (Kemp and Soede 1996; Corrêa et al. 2002). However, the sperm quality of the boar strongly decreases after one ejaculation for, at least, the next two days (Frangež et al. 2005). Increasing the number of sires for a female is thus used in pig husbandry to obtain optimal fertilization and maximal litter sizes with natural reproduction (Badinel 2010). The underlying behavioral and physiological mechanisms involved in free-ranging wild boars remain to be studied.

In conclusion, high rates of multiple paternity are measured under intensive harvesting regime where rapid evolutionary changes were previously observed (Servanty et al. 2009; Gamelon et al. 2011). This lead us to hypothesize that multiple paternity might be a key basis for exploited populations of wild boar to display evolution over just a few generations (Gamelon et al. 2011; Servanty et al. 2011). It allows the population to withstand the harvesting pressure (Gamelon et al. 2012) through an increase in the number of reproductive males, an unusual pattern in ungulates (Ginsberg and Milner-Gulland 1994), which, we show here, induces larger litter sizes. It is noteworthy that litter size is a key life history trait of fecundity, a major component in demography. Therefore, multiple paternity could be one of the factors contributing to the actual increase of wild boar abundance (Massei et al. 2015). However, the access to reproduction for younger and/or weaker males that would normally not garner matings may have long term negative consequences. Indeed, these males may carry and transmit poor quality genes that could be deleterious on the long run for the population. This study raises question about other ungulate species' strategies to buffer negative consequences of size and sex selective harvesting (Hard et al. 2006), since they are generally unable to modulate the number of offspring produced per reproductive event. Some changes of mating systems can be expected with a decrease of harem size allowing more males to reproduce each year when hunting is intensive. In contrast, many game species from birds to small mammals produce several offspring per reproductive event; investigating to what extent the pattern observed here apply to these species in intensive harvesting context is an interesting challenge.

Acknowledgment

We are grateful to all those who helped collect harvested wild boars and to the ONF, and to F. Jehlé who allowed us to work in the study area. We thank Diane Gonzalez for the preliminary analysis of the first years' data. We thank J-M. Gaillard and J. O'Brien for their helpful comments which markedly improved our paper and C. Carter for the proofreading. This work was supported by the French National Hunting and Wildlife Agency (ONCFS), by the University of Lyon and by the Research Council of Norway through its Centres of Excellence funding scheme, project number 223257.

Supplementary Material

1. Characteristics of the genetic markers

Table IV S3 Number of alleles A, observed Ho and expected He heterozygosity, Hardy-Weinberg equilibrium (NS: equilibrium; *: non-equilibrium) and estimated frequency of null alleles for each locus as provided by CERVUS for adult wild boars (n = 1054 genotypes). Fluorescent dyes used for the resolution of the PCR amplified products and their expected size have been added.

Locus	A	Ho	He	HW	F(Null)	Dye	Expected size
CGA	25	0.845	0.891	*	0.026	Fam	250-320
SO005	23	0.832	0.891	NS	0.034	Ned	200-280
SW2496	16	0.836	0.826	NS	-0.007	Ned	180-230
SO068	15	0.562	0.573	NS	0.010	Hex	210-260
SW2021	11	0.748	0.744	NS	-0.004	Hex	100-130
SW240	8	0.662	0.681	NS	0.015	Ned	90-150
SO228	8	0.638	0.653	NS	0.012	Hex	220-250
SW122	8	0.432	0.427	NS	-0.007	Fam	110-120
SW24	7	0.708	0.703	NS	-0.003	Hex	96-120
SW936	6	0.529	0.516	NS	-0.011	Fam	80-120
SO355	6	0.107	0.125	*	0.068	Fam	240-280
SO215	2	0.186	0.193	NS	0.019	Fam	135-169

2. Estimation of confidence for paternity analysis with CERVUS and GERUD

a) CERVUS analysis

A simulation of parentage analysis determines a critical likelihood score for several levels of confidence at the population level. Genotypes were simulated for 10 000 offspring, with 50% of candidate fathers sampled based on an estimated population size of 1200-1500 wild boars (Gamelon et al. 2011) and the stable-age structure in this population (Servanty et al. 2011), and ≥5 loci used. A conservative overall genotyping error rate of 0.01 was assumed. Strict and relaxed confidence intervals of 95% and 80% were specified for population-level assignment probabilities.

b) GERUD analysis

To evaluate the power of our microsatellite loci to detect multiple paternity, we used the program GERUDsim 2.0 (Jones 2005). Using the observed allele frequencies within the population, GERUDsim simulates sets of offspring genotypes based on user-specified litter sizes (in our case 5), draws a sample of offspring, and then estimates the number of sires present in each litter. We ran 1000 iterations of the simulation, each using a single multilocus maternal genotype and up to four randomly generated paternal multilocus genotypes (two fathers siring 4/1 and 3/2 fetuses, three fathers sharing paternity of 3/1/1 and 2/2/1 fetuses, four fathers siring 2/1/1 fetuses) to evaluate the probability of correctly determining the number of sires within litters. Simulations were conducted using the five most polymorphic loci. The reconstructed number of sires equaled the number assigned by the program in 94.4% and 97.6% of the two-father (sharing 4/1 and 3/2 offspring, respectively) simulations, in only 35.7% and 0% of the three-father (sharing 2/2/1 and 3/1/1 offspring respectively), and 0% of the four-father simulations performed in GERUDsim. Albeit in no circumstances was the number of real fathers overestimated by the heuristic search algorithm employed by GERUDsim. Given our GERUDsim simulations, our estimates of the minimum number of males contributing to litters in GERUD are very conservative.

3. Model selection for the analysis of the effect of number of sires on litter size using maximal number of paternal alleles as a proxy

Table IV S4 Model selection to test the effect of the locus-specific observed heterozygosity *Ho* and observed number of alleles *A*, the mother dressed body mass BM_m and the maximal number of paternal alleles in a litter N_{PA} on the litter size (*LS*) in the wild boar (*Sus scrofa*) population of Châteauvillain, France. The model retained is in bold (ΔAICc<2). 'X' denotes that the explanatory variable was included in the model ($n_{litters}$ = 160).

Intercept	*Ho*	*A*	BM_m	N_{PA}	# parameter	Log Likelihood	AICc	ΔAICc	AICc weight
X		**X**	**X**	**X**	**4**	**-301.90**	**612.05**	**0**	**0.48**
X	X	X	X	X	5	-301.90	614.18	2.13	0.17
X	X		X	X	4	-303.36	614.98	2.93	0.11
X		X	X		3	-304.64	615.43	3.38	0.09
X			X	X	3	-304.65	615.45	3.40	0.09
X	X	X	X		4	-304.64	617.54	5.48	0.03
X	X		X		3	-306.09	618.33	6.28	0.02
X			X		2	-307.58	619.23	7.18	0.01
X		X		X	3	-311.15	628.46	16.41	0.00
X	X	X		X	4	-311.05	630.36	18.31	0.00
X	X			X	3	-312.14	630.44	18.39	0.00
X				X	2	-314.33	632.74	20.69	0.00
X		X			2	-315.61	635.29	23.24	0.00
X	X	X			3	-315.42	636.99	24.94	0.00
X	X				2	-316.47	637.02	24.97	0.00
X					1	-319.15	640.32	28.27	0.00

4. Model selection for the effect of number of sires estimated by GERUD on the litter size

Table IV S5 Model selection to test the effects of the mother dressed body mass BM_m and the number of fathers estimated by GERUD N_G on the litter size (*LS*) in the wild boar population of Châteauvillain, France. The model retained is in bold (ΔAICc<2). 'X' denotes that the explanatory variable was included in the model ($n_{litters}$ = 160).

Intercept	BM_m	N_G	# parameter	Log Likelihood	AICc	ΔAICc	AICc weight
X	**X**	**X**	**3**	**-304.64**	**615.43**	**0**	**0.87**
X	X		2	-307.58	619.23	3.80	0.13
X		X	2	-315.14	634.36	18.94	0.00
X			1	-319.15	640.32	24.90	0.00

Chapter IVB
Consequences of multiple paternity

Does multiple paternity explain phenotypic variation among offspring in wild boar?

Abstract: Wild boar (*Sus scrofa*) females produce large litters with diversified offspring in terms of body mass. Additionally, multiple paternity within a litter has been observed in this promiscuous species. One can hypothesize that multiple paternity represents the mechanism by which females increase within-litter phenotypic variation. Combining long-term monitoring data with paternity analyses in a wild boar population, we tested whether the increase in the number of fathers within a litter explained the increase in within-litter variation in offspring mass observed in large litters. We showed that heavy females mated earlier during the rut, produced larger litters with a higher number of fathers and more variable fetus mass than lighter females. Within-litter diversification of offspring mass increased with gestation stage and litter size, suggesting differential allocation of maternal resource among offspring in utero. However, we found only a weak paternal effect on offspring mass and no direct effect of the number of fathers on the within-litter variation in offspring mass. These results indicate that within-litter diversification of offspring mass is unlikely related to multiple paternity in this species.

Keywords*:* fetus mass, paternity analysis, phenotypic polymorphism, siblings

Gamelon M., T. Gayet, E. Baubet, S. Devillard, L. Say, S. Brandt, C. Pélabon, and B-E. Sæther. Does multiple paternity explain phenotypic variation among offspring in wild boar? *Submitted version*

Introduction

Natural selection on body size is generally positive (Kingsolver and Diamond 2011), also during early life. For example, in mammals and birds, offspring with high body mass often exhibit high survival (see Ronget et al. 2018 for meta-analyses). However, because of a trade-off between size and number of offspring (Smith and Fretwell 1974; Lloyd 1987; Winkler and Wallin 1987), producing many large offspring is not a sustainable reproductive tactic for polytocous species. Thus, maternal resources are either equi-allocated among offspring (Smith and Fretwell 1974), or differentially allocated among them (see e.g. Kühl et al. 2007 in saiga antelope *Saiga tatarica*) leading to within-litter/clutch variation in offspring mass. In variable and unpredictable environments, such a diversification of offspring phenotypes may contribute to minimizing variance in reproductive success among years (Philippi and Seger 1989; Starrfelt and Kokko 2012; Sæther and Engen 2015) and thus maximizing fitness (Kaplan and Cooper 1984; Gamelon et al. 2013b). Interestingly, within-litter/clutch variation in offspring mass can result from multiple paternity if offspring from different fathers genetically differ in their ability to acquire and/or use maternal resources (Watson 1991; Yasui 1998, 2001; Fox and Rauter 2003).

In wild boar (*Sus scrofa*), litter size increases with mother body mass. Heavy females produce large litters with a mixture of heavy and light offspring, whereas lighter females produce litters with similar-sized offspring (Gamelon et al. 2013b). In this polytocous species, contrary to other large mammalian species of herbivores (Gaillard et al. 2000b), piglet body mass has little influence on survival (Baubet et al. 1995) allowing females to produce a large range of offspring phenotypes. Furthermore, by producing diversified offspring phenotypes, heavy females may match the mass of their offspring with teat productivity, thus decreasing within-litter competition to get access to maternal milk, and thereby increasing the chance of rearing many offspring at a given breeding event (Gamelon et al. 2013a). The species has been classically described as polygynous with female monopolization by males. However, multiple paternity has been reported suggesting a promiscuous mating system, and it has been shown that the number of fathers increases with litter size (Gayet et al. 2016). One can thus hypothesize that multiple paternity is the mechanism by which wild boar females increase within-litter variation in offspring mass. If mating with multiple males is the pathway by which females increase the phenotypic polymorphism of their offspring, differences in piglet mass should be partly determined by paternally derived alleles, and we expect a paternal genetic effect on offspring mass as well as more variable offspring in litters sired by many fathers.

Taking advantage of a unique long-term monitoring of a wild boar population, we tested the hypothesis that multiple paternity mediates within-litter diversification of offspring phenotypes. We extended previous works linking female body mass with diversification of offspring phenotypes (Gamelon et al. 2013b) by including paternity analyses. We identified fathers of fetuses from females killed during hunting and tested for a paternal effect on fetus mass. Moreover, we explored the pathways through which female body mass influences the diversification of offspring phenotypes by testing specifically a direct effect of the number of fathers per litter on phenotypic variation among offspring.

Material and methods

Study site and data collection

The study was conducted in northeastern France in the 11,000ha forest of Châteauvillain-Arc-en-Barrois. In this area, wild boars are heavily hunted each year between October and February and the annual survival of adult females is 0.48 [95% CI: 0.44; 0.51] and 0.23 [0.17; 0.30] for adult males (Toïgo et al. 2008). Between 2007 and 2014, we recorded the dressed body mass (*BM*: body mass without digestive tract, heart, lungs, liver, reproductive tract and blood) of 136 pregnant females shot with the sampling date. For each female, we also recorded the litter size (*LS*) and each fetus (n=711) was weighed, measured (crown-rump length, in millimeters) and sexed. From the average fetus length within a litter (*Length*), we estimated gestation stage in days by applying the model of Henry (Henry 1968): gestation stage (in days) = 23.43 + 0.32* *Length* (in mm) (Gamelon et al. 2013b). From this estimated gestation stage and the sampling date, we back-calculated the timing of mating. In order to account for yearly variation in the timing of the mating season, we expressed the timing of mating as the number of days elapsed since the first female has mated in each particular season (*Timing*). Thus, a *Timing* of zero characterizes the most precocious female in each given year. The average fetus length at sampling depends on both the timing of mating and the sampling date. Indeed, for a given timing of mating, a female will have short fetuses when killed soon after mating and longer fetuses when killed later. However, because both the mating season (ranging between July and January, see Results) and the sampling period (from October to February) are wide, there is no correlation between the timing of mating and the average fetus length within a litter when sampled.

Paternity assessment

Tissue samples were collected from all mothers, fetuses and from 762 putative fathers (i.e. putative reproductive males shot). Among the 136 litters, data of the full litters were available for 116 of them (617 fetuses, mean ± SD litter size = 5.32 ± 1.61). All tissue samples were genotyped at 12 microsatellite loci (see Gayet et al. 2016 and Supplementary Material 1 for details). The genotypes of mothers, offspring and putative reproductive males, as well as the known mother-offspring relationships were analyzed using COLONY 2.0.6.1 (Jones and Wang 2010) and for each hunting season *t*, we identified the father (whether the male was sampled or not) of each fetus. Parentage among individuals was inferred by maximal likelihood. We analyzed all the litters considering as putative fathers all males sampled at season *t*, subadult (*i.e.*, between one and two years of age) and adult (*i.e.*, two years of age or older) males sampled at season *t+1* and adult males sampled at season *t+2*. For the analysis in COLONY, the parameters were set as follow: both sex polygamous (because of the promiscuous mating system characterizing the studied population), unknown population allele frequency, full likelihood analysis method, medium precision, codominant markers, proportion of males sampled of 50% (this value is approximated but it only slightly influences results; (Harrison et al. 2013), an allelic dropout of 0.001 and a typing error of 0.01 for each locus. Other parameters were set to their default values. This software allowed identifying fathers (n=235) of each fetus, whether the father was sampled or not. Thus, some of the fathers were identified from tissue samples while others, not sampled, were assigned by COLONY. The probability of identity that is the probability that two fathers drawn in the population have the same genotype at multiple loci, was estimated to be $7.9.10^{-11}$. For the offspring, the probability of identity was estimated to be $1.5.10^{-4}$. These values indicate high confidence for individual identification (Waits et al. 2001). The probability of inferred father, which is the probability to infer correctly a sampled male as the father of a given offspring, was high: 0.974 ± 0.094.

Effect of father identity on fetus mass

For multiple paternity to translate into an increase in within-litter variation in offspring mass, father identity should affect offspring mass in utero. We estimated this effect for fathers that have produced more than one offspring (n = 148 fathers, 624 fetuses in total) using a linear mixed-effect model with individual fetus mass as response variable, fetus sex and mother identity as fixed factors and father identity as a random factor and assumed a Gaussian distribution. Including maternal identity and sex as fixed effects allowed us correcting offspring mass for factors (female body mass and condition, gestation stage, year and litter size) inducing

among-litter variation in body mass as well as the sex effect on offspring mass. The remaining part of the variance in offspring mass thus only results from paternal effects and residual variation. We calculated the paternal effect as the ratio of the variance in offspring mass due to father identity, divided by the total variance: $\frac{\sigma^2_{\text{Father}}}{\sigma^2_{\text{Father}}+\sigma^2_{\text{Residuals}}}$, where σ^2_{Father} is the random variance associated with the father identity, and $\sigma^2_{\text{Residuals}}$ is the residual variance. Half-sibs in different litters, i.e. from the same father but different mothers, may have different body mass simply because they were sampled at different gestation stages. Using mother identity as fixed factor does not entirely account for this effect because mass does not increase linearly during gestation. Neglecting such non-linear growth may artificially increase the residual variance and thus decrease the estimate of the paternal effect. Therefore, the response variable fetus mass was log-transformed to perform the analysis on a proportional scale. We ran 260,000 Markov Chain Monte Carlo iterations, with a burn-in of 10,000 iterations thinning every 250th observation, and non-informative priors were used (for the variance structures (R and G), we used an expected variance of 1 and 0.002 degree of belief parameter for the inverse-Wishart). We computed the posterior modes and the 95% credible intervals of this ratio, of fetus sex and of the variance associated with the father identity with the "HPDinterval" function of the package MCMCglmm (Hadfield 2010) in R 3.4.0 (R Core Team 2017). We assessed convergence with the functions "heidel.diag" (Heidelberger and Welch´s convergence diagnostic) and "geweke.diag" (Geweke´s convergence diagnostic) in R and from visual inspection. We checked normality and homoscedasticity of the residuals.

Effect of the number of fathers on within-litter variation in fetus mass

To assess whether multiple paternity mediates the increase of within-litter variation, we estimated the within-litter variation in mass by calculating the coefficient of variation (*CV* = SD/mean) of fetus mass (on the natural scale) for each full litter (n=116), corrected for small samples as suggested by Haldane (1955). Using confirmatory path analyses (Shipley 2009, 2013), we then determined the causal pathways from mother body mass to within-litter variation, through the number of fathers (*F*) within a litter and/or litter size (*LS*). We included a correlation between *F* and *LS* (Gayet et al. 2016). Because mating ranged between July and January (see Results) and because females were killed from October to February, we observed litters at different periods of the year and at different gestation stages. Therefore, we included both the timing of mating (*Timing*) and the average fetus length (*Length*) (as a measure of gestation stage) in our models. We used the Akaike Information Criterion corrected for small

sample size (AICc) for model selection among the ones presented inTable IV S6 and Supplementary Material Figure IV S5. We recovered the standardized regression coefficients and their associated SE. The analyses were implemented using the package piecewiseSEM (Lefcheck 2016) in R version 3.4.0 (R Core Team 2017).

Results

Paternity assessment

Among the 116 litters analyzed, 15 had all fathers known (i.e. identified from sampled males), 30 had some fathers known while the others were assigned by COLONY, and 71 litters had all fathers assigned by COLONY. Fathers sired on average 3.03 ±2.54 (mean ± SD) offspring (Figure IV 3), with one (75.7% of the cases), two (18.3%), three (5.1%) or four (0.9%) partners. The average number of fathers within a litter was 2.28 ±1.28 (mean ± SD) (Figure IV 3) and multiple paternity was observed in 63.8% of the litters.

Effect of father identity on fetus mass

The linear mixed-effect model evaluating paternal effect on fetus mass *in utero* showed no lack of convergence (Supplementary Material S3). After accounting for maternal effects, fetus mass was 5% [95% CRI: 0.04; 0.07] lighter in female offspring than in males, in accordance with previous studies (Servanty et al. 2007). The variance associated with paternal identity, σ^2_{Father} , was low 0.0005 [95% CRI: 0.0002; 0.002]. The ratio $\frac{\sigma^2_{\text{Father}}}{\sigma^2_{\text{Father}}+\sigma^2_{\text{Residuals}}}$ was 0.09 [95% CRI: 0.03; 0.21] indicating that paternal identity explained 9% of the within-litter variance, which is the variance remaining when sex and all maternal effects were accounted for.

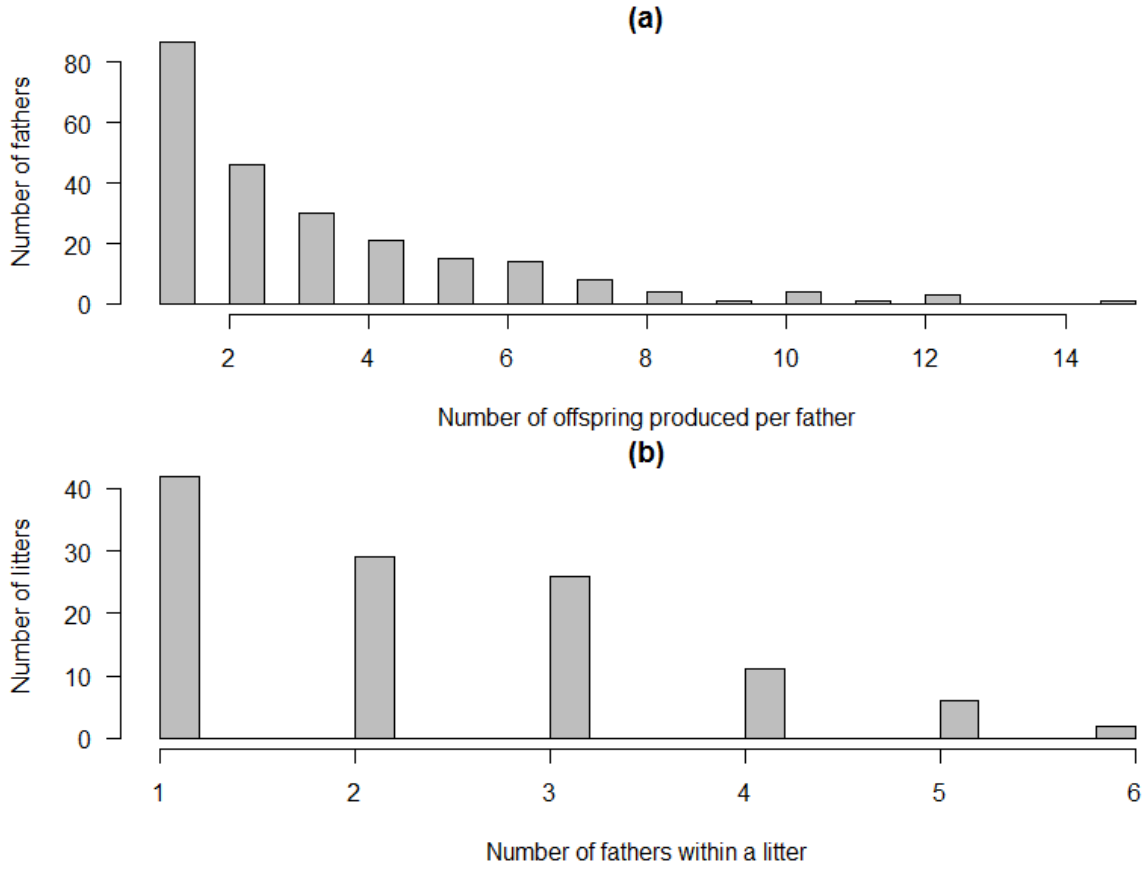


Figure IV 3 (a) Number of sampled offspring per father for the 235 identified fathers in the wild boar population of Châteauvillain-Arc-en-Barrois, France; (b) Number of litters with 1 to 6 fathers observed, for the 116 litters included in the study. Note that fathers may be sampled in the population and identified from tissue samples or not sampled and assessed by COLONY.

Effect of the number of fathers on within-litter variation in fetus mass

Although heavy females produced large litters sired by many fathers with diversified offspring mass, our path analysis did not indicate any direct link between the number of fathers per litter and the within-litter variation in fetus mass. Indeed, the best path models (Table IV S6), close in terms of AICc values, never included direct effect of the number of fathers per litter on CV of fetus mass. However, some of these models included indirect positive effects of mother body mass and number of fathers per litter on the within-litter variation through an increase of litter size (Figure IV 4). Looking more specifically at the relationship between the number of fathers per litter and CV of fetus mass, the full model we tested (model 8, seeTable IV S6 and Supplementary Material Figure IV S5) confirmed no effect of multiple paternity on within-litter variation (effect size ± SE = 0.003 ±0.10).

Table IV S6 Model fit of the 13 competing path models exploring the relationship between female body mass (*BM*), number of fathers within the litter (*F*), litter size (*LS*), timing of mating (*Timing*), mean fetus length (*Length*) and within-litter variation in fetus mass (*CV*) for each litter (n=116). Displayed are the number of parameters (N), the AICc of the tested models, and the difference between each model and the best one (ΔAICc).

Model notation	N	AICc	ΔAICc
1. F~BM / LS~ BM / Timing~BM / Length~Timing / CV~LS+Length	16	59.03	0
2. F~BM / LS~ BM / Timing~BM / Length~Timing / CV~Length	15	60.67	1.65
3. F~BM / LS~ BM / Timing~BM / Length~Timing / CV~ LS+Length+F	17	61.76	2.73
4. F~BM / LS~ BM / Timing~BM / Length~Timing / CV~LS	15	62.05	3.03
5. F~BM / LS~ BM / Timing~BM / Length~Timing / CV~Length+F	16	62.84	3.81
6. F~BM / LS~ BM / Timing~BM / Length~Timing / CV~Length	16	63.18	4.15
7. F~BM+Timing / LS~ BM / Timing~BM / Length~Timing / CV~LS+Length	16	64.20	5.17
8. F~BM+Timing / LS~ BM / Timing~BM / Length~Timing / CV~LS+Length+F	18	64.47	5.45
9. F~BM+Timing / LS~ BM / Timing~BM / Length~Timing / CV~LS	16	64.68	5.65
10. F~BM+Timing / LS~ BM / Timing~BM / Length~Timing / CV~Length+F	17	65.51	6.48
11. F~BM+Timing / LS~ BM / Timing~BM / Length~Timing / CV~LS+F	17	66.85	7.82
12. F~BM / LS~ BM / Timing~BM / Length~Timing / CV~F	15	67.50	8.48
13. F~BM+Timing / LS~ BM / Timing~BM / Length~Timing / CV~F	16	70.11	11.08

The earliest mating reported in our study occurred in mid-July (for 2014) and the latest in mid-January (for 2011) suggesting a particularly long mating season. We provided evidence that female body mass was negatively associated with *Timing*, a metric indicating how precocious was the mating for a female in a given season (Figure IV 4). Therefore, heavy females reproduced earlier than lighter ones during the mating season. Moreover, within-litter variation increased with gestation stage (defined as the average fetus length *Length*) (Figure IV 4). Because within-litter variation in offspring mass was estimated using the coefficient of variation (CV), this effect indicates that variation in offspring mass increases more during gestation than the expected proportional increase of the standard deviation with the mean. Although based on cross-sectional data, this result suggests that offspring differ in their growth rate.

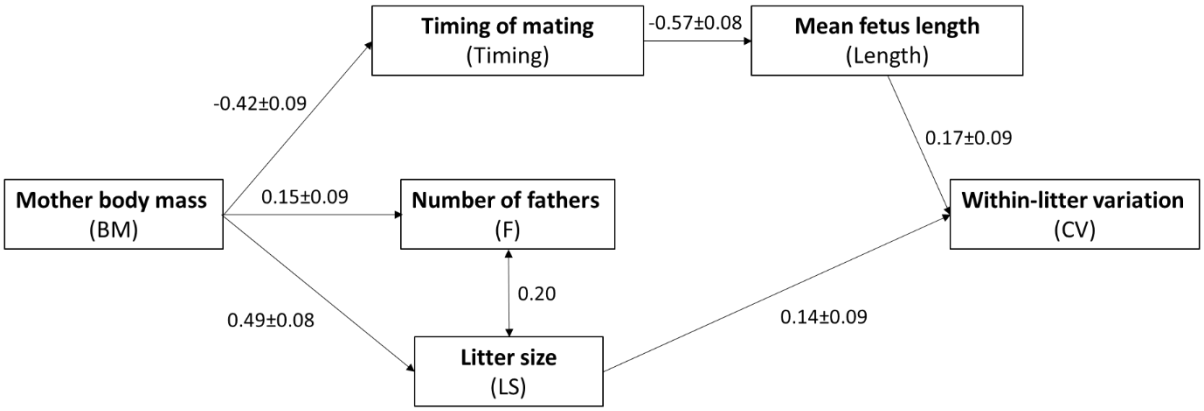


Figure IV 4 Path model with the best fit (see Table IV S6) showing how mother body mass (*BM*) and number of fathers per litter (*F*) influence the within-litter variation in fetus mass (*CV*) through litter size (*LS*), timing of mating (*Timing*) and mean fetus length (*Length*). Numbers indicate standardized regression coefficients and their associated SE.

Discussion

Our findings showed that, contrary to expectations, the diversification of offspring phenotypes within a litter did not directly result from multiple paternity and the genetic diversification of the offspring. Indeed, although larger litters were sired by more fathers as previously observed (Gayet et al. 2016) and contained fetuses of more variable mass than smaller litters, this within-litter variation in fetus mass did not directly result from the number of fathers siring the litter. This result is further supported by the lack of paternal effect on fetus mass *in utero*, as indicated by the small proportion of the within-litter variance explained by paternal identity. Although expected for early-life stages (Wilson et al. 2005), this weak

paternal effect on offspring mass strongly limits the possibly for the females to diversify the mass of their offspring by mating with several, genetically distinct, fathers. It is noteworthy that, due to increasing genetic diversity among offspring, other types of genetic effects such as dominance or epistatic interactions may also affect within-litter variance in offspring mass (Neff and Pitcher 2005). Exploring such effects would require repeated measurements of offspring mass produced by a given pair of mother and father, which is unfortunately impossible in our study system.

Our path analysis identifies the most likely pathways through which female body mass affects within-litter variation in fetus mass. Depending on their body mass, females mate at different periods during the rut. Heavy/old females mate earlier during the rut and have larger litters sired by a high number of fathers than lighter/younger ones. These findings suggest inter-individual heterogeneity among females, with earlier mating and thus parturition dates in old and heavy females compared to young and light ones (see Feder et al. 2008 for evidence on bighorn sheep *Ovis canadensis*). Because wild boar females having reached 33–41% of their full body mass are able to reproduce (Servanty et al. 2009), it is likely that light/young females are primiparous, born in spring and reaching this threshold body mass to reproduce only later during the mating season. In turn, large litters produced by heavy females and gestation stage tended to directly influence within-litter variation in offspring mass.

The increase in CV of fetus mass during gestation indicates that initial differences in body mass among offspring are magnified during gestation most likely due to different growth rates among offspring. This differential growth is not affected by the fathers' genotype and the number of fathers in the litter. Indeed, if multiple paternity was involved in within-litter variation in offspring mass, through different abilities among half-sibs to acquire and/or use maternal resources, we would have detected a direct effect of the number of fathers on within-litter diversification. This is not supported by our observations and we regard multiple paternity as an unlikely mechanism to explain diversification of offspring mass in large litters. Differential maternal allocation among offspring might explain differences in offspring mass. Indeed, mothers of long-lived iteroparous organisms may change the phenotype of their offspring by allocating resources differentially among them within a reproductive attempt (Kühl et al. 2007). Several mechanisms, not mutually exclusive, could explain differential maternal allocation among offspring such as developmental constraints (e.g., position of the offspring in the uterus, Bautista et al. 2015), or sibling rivalry *in utero* to get access to maternal resources (Mock and Parker 1997; Hudson and Trillmich 2007). Whether differential maternal

allocation among offspring is adaptive or results from developmental constraints remains to be carefully explored and offers promising avenues of research.

Supplementary Material

1. Microsatellite information.

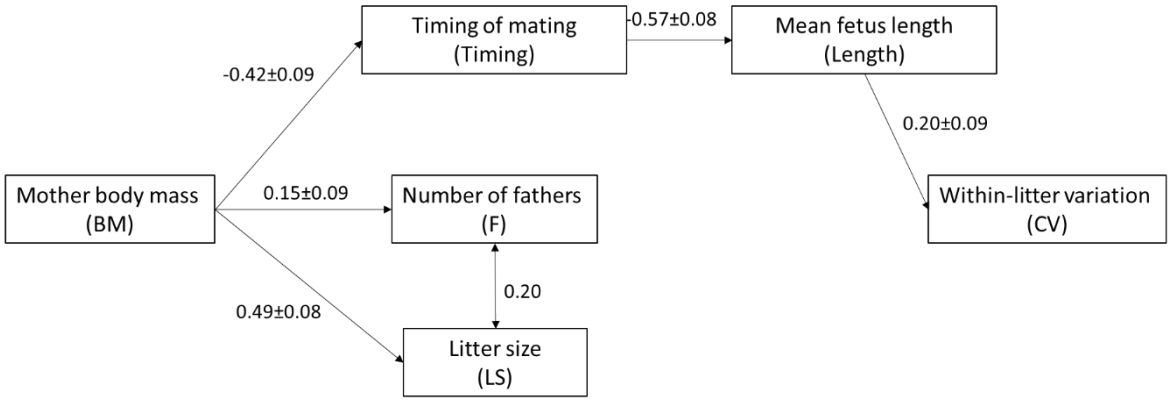
Table IV S7 Number of alleles *A*, observed *Ho* and expected *He* heterozygosity, estimated frequency of null alleles for each locus *F_Null* provided by packages adegenet and PopGenReport, linkage disequilibrium *LD* obtained with Fstat software for each season with all other loci (NS for non-significant), P-values from 1920 randomizations of alleles among individuals within seasons to measure deviation from Hardy-Weinberg equilibrium *HW* for each season provided by Fstat (adjusted P-value at 5% was 0.00052), fluorescent dyes *Dye* used for the resolution of the PCR amplified products and their expected size *Expected size* for adult wild boars (n = 1244 genotypes). The set of markers, not linked to the sex, is known for a good quality of the loci and a low typing error rate. For more details on the microsatellites used, see Appendix 7 in FAO (FAO 2011) and Groenen (2003). The probabilities of identity between two random individuals and two siblings were estimated to be 7.9×10^{-11} and 1.5×10^{-4} respectively using Genalex.

Locus	A	Ho	He	F_Null	LD	HW S1	HW S2	HW S3	HW S4	HW S5	HW S6	HW S7	HW S8	Dye	Expected size
CGA	26	0.841	0.889	0.027	NS	0.013	0.001	0.001	0.171	0.054	0.001	0.027	0.168	Fam	250-320
SO005	23	0.842	0.89	0.027	NS	0.014	0.011	0.001	0.001	0.001	0.021	0.868	0.929	Ned	200-280
SO068	15	0.548	0.565	0.015	NS	0.041	0.013	0.007	0.710	0.698	0.489	0.135	0.066	Hex	210-260
SO215	2	0.187	0.194	0.019	NS	0.804	0.731	0.138	0.091	0.636	0.405	0.499	0.485	Fam	135-169
SO228	8	0.636	0.652	0.013	NS	0.039	0.066	0.027	0.133	0.682	0.484	0.101	0.388	Hex	220-250
SO355	7	0.112	0.126	0.057	NS	0.001	0.001	0.001	0.014	0.002	0.101	0.452	1.000	Fam	240-280
SW122	8	0.443	0.437	-0.007	NS	0.642	0.452	0.897	0.966	0.909	0.592	0.741	0.250	Fam	110-120
SW2021	11	0.75	0.749	-0.001	NS	0.770	0.534	0.296	0.317	0.535	0.628	0.426	0.034	Hex	100-130
SW24	7	0.706	0.7	-0.005	NS	0.488	0.122	0.420	0.432	0.570	0.845	0.811	0.567	Hex	96-120
SW240	8	0.66	0.68	0.014	NS	0.217	0.049	0.119	0.025	0.347	0.120	0.181	0.191	Ned	90-150
SW2496	17	0.836	0.857	0.012	NS	0.778	0.326	0.406	0.576	0.109	0.047	0.931	0.894	Ned	180-230
SW936	6	0.541	0.521	-0.019	NS	0.970	0.472	0.674	0.659	0.788	0.985	0.980	0.916	Fam	80-120
Overall						0.0042	0.0005	0.0005	0.0021	0.0286	0.0391	0.3089	0.1026		

2. Path models.

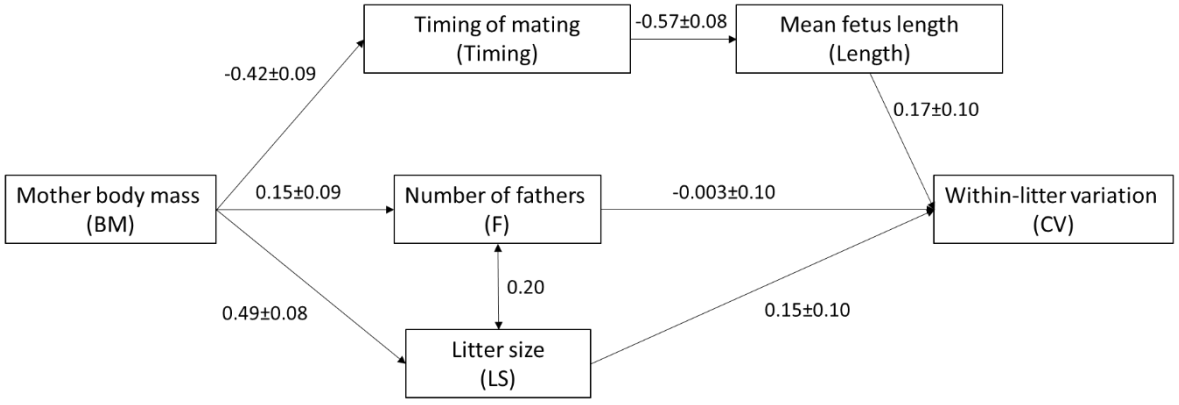
Figure IV S5 Thirteen different competing path models (see table 1) have been tested. The selected path model (model 1, table 1) is shown in figure 2. The other 12 path models, exploring how mother body mass (*BM*) and number of fathers per litter (*F*) influence the within-litter variation in fetus mass (*CV*) through litter size (*LS*), timing of mating (*Timing*) and mean fetus length (*Length*) in the wild boar population of Châteauvillain-Arc-en-Barrois, France, are shown below.

Model 2.



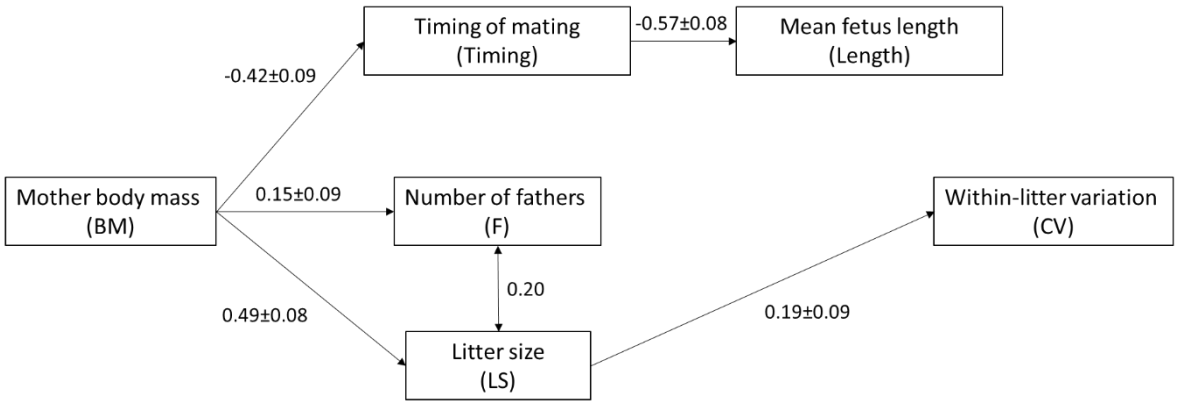
AICc=60.67

Model 3.



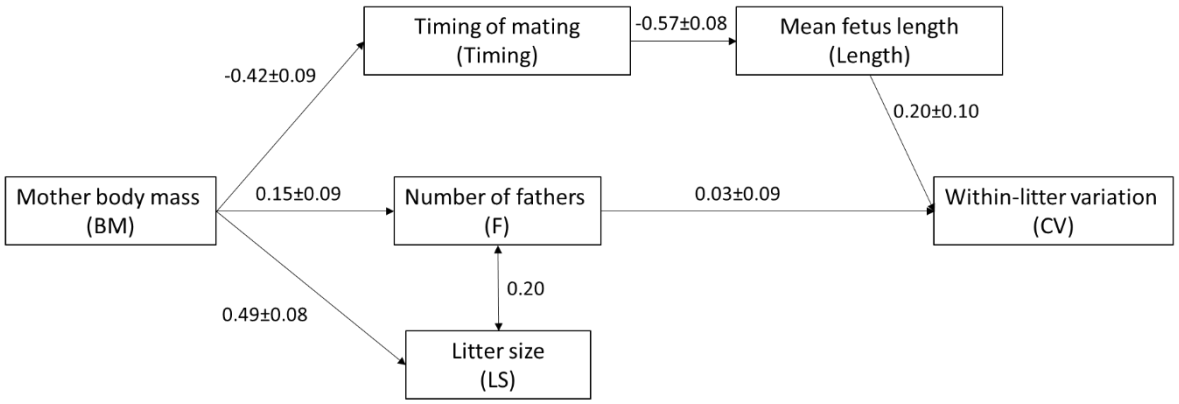
AICc=61.76

Model 4.



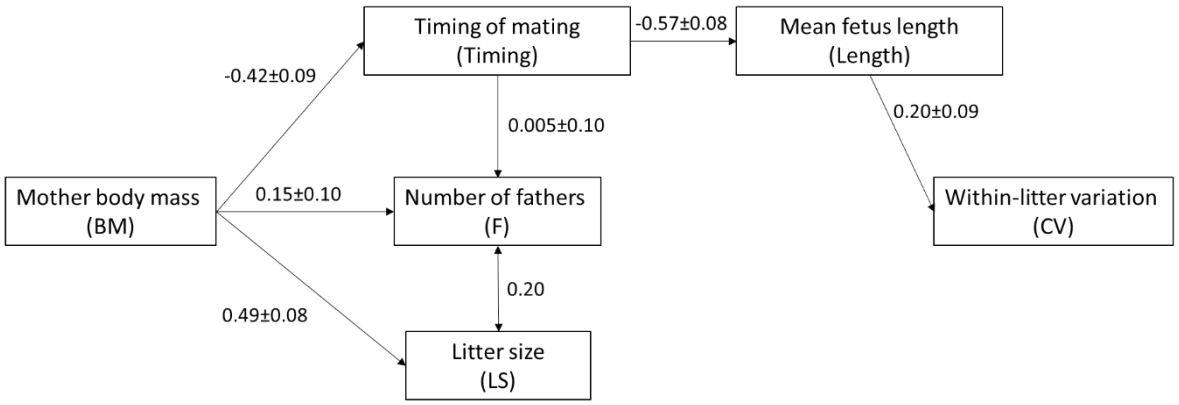
AICc=62.05

Model 5.



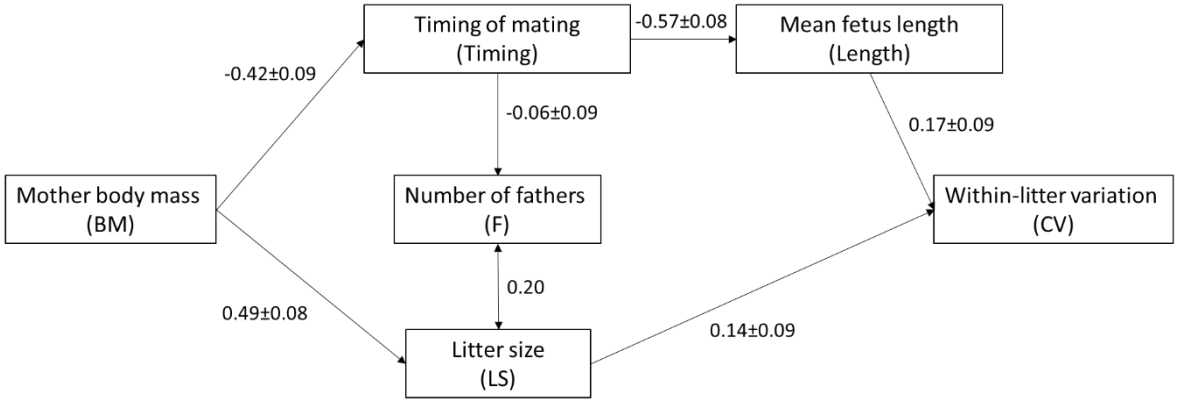
AICc=62.84

Model 6.



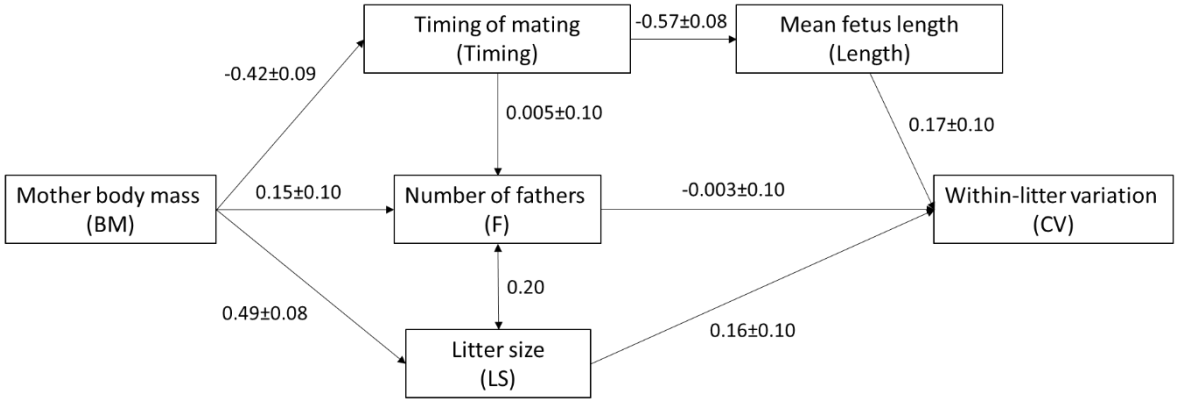
AICc=63.18

Model 7.



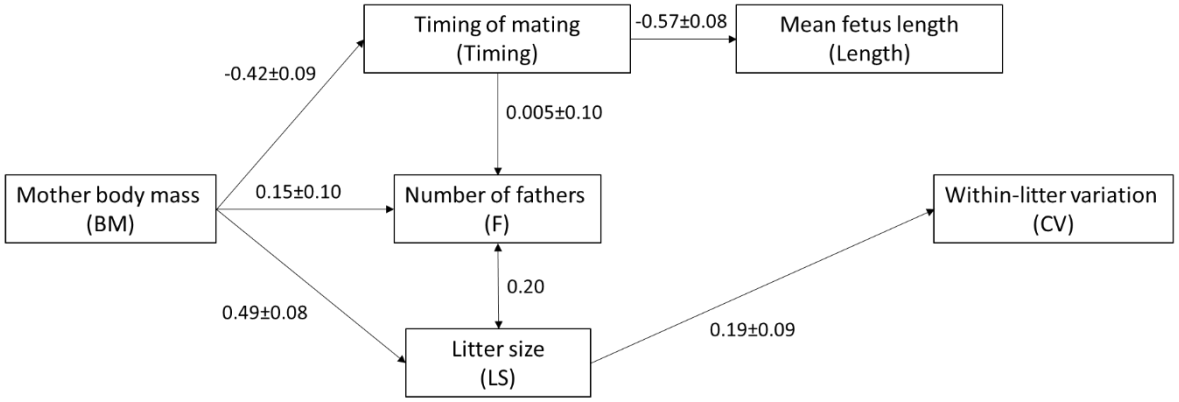
AICc=64.20

Model 8.



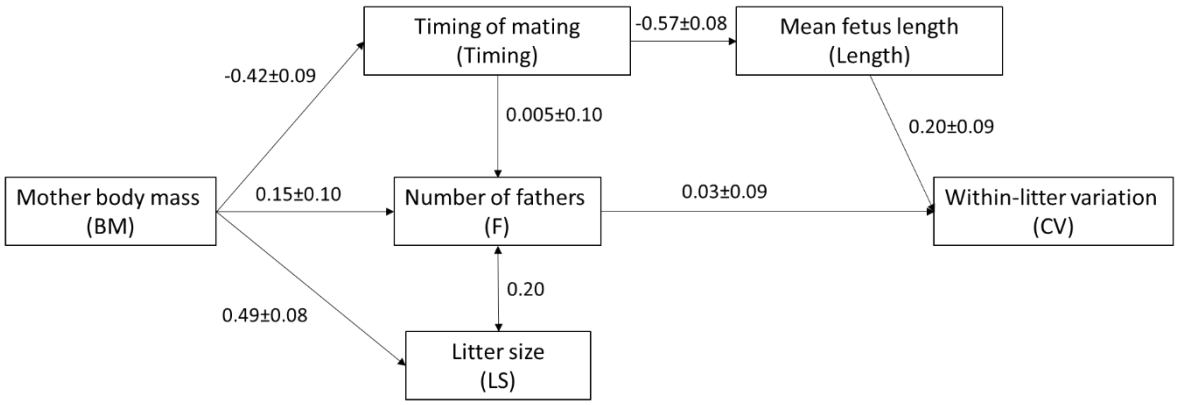
AICc=64.47

Model 9.



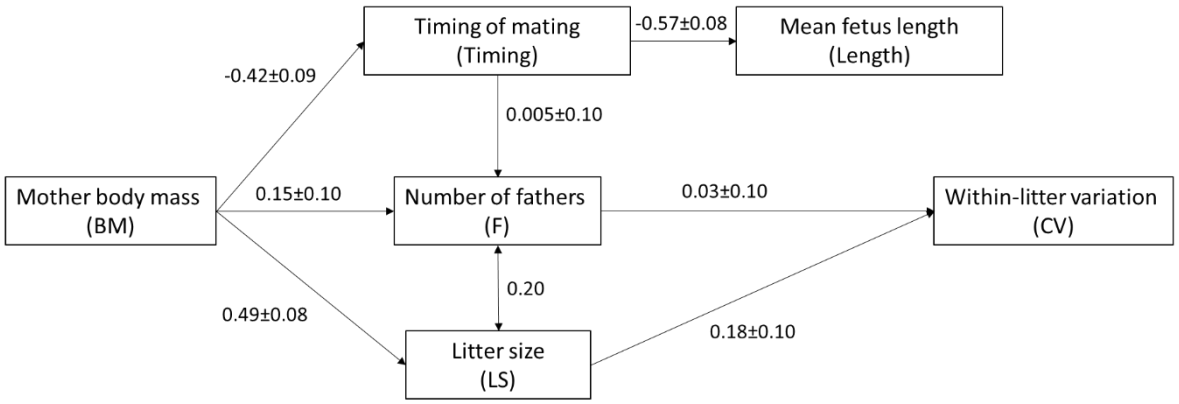
AICc=64.68

Model 10.



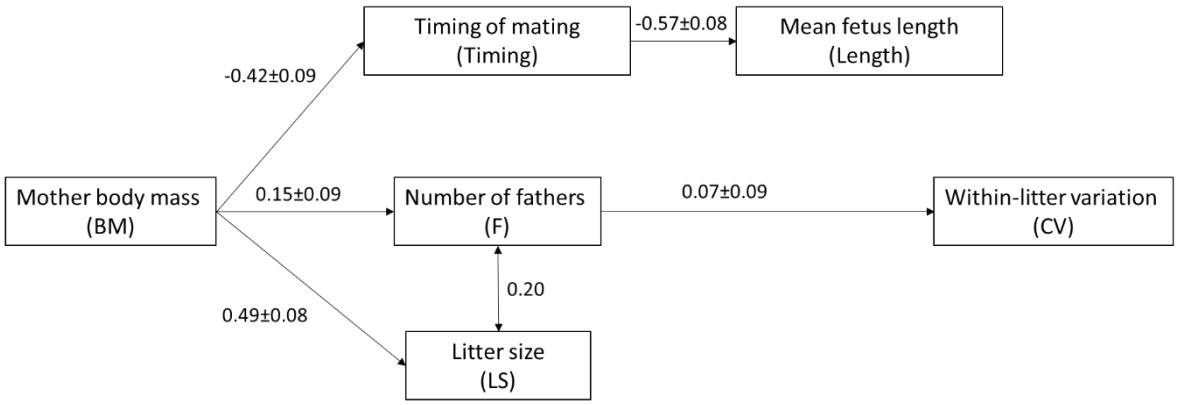
AICc=65.51

Model 11.



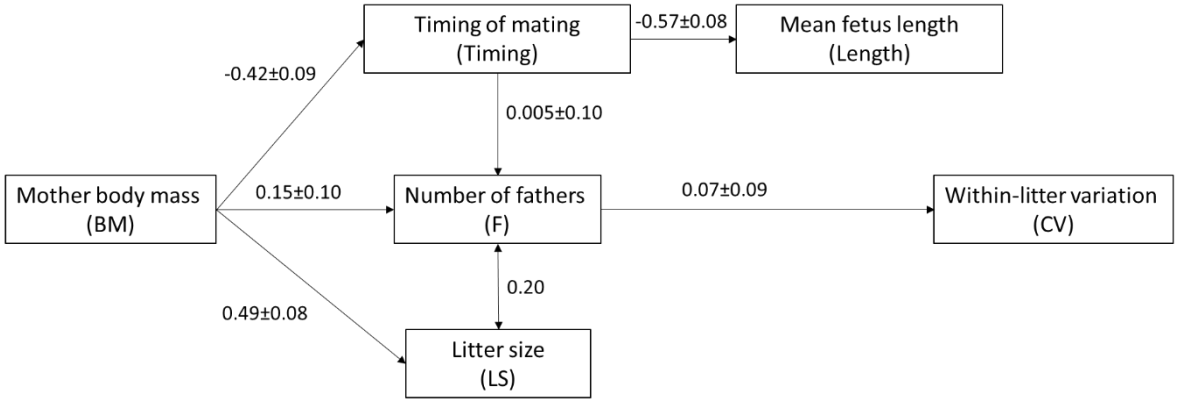
AICc=66.85

Model 12.



AICc=67.50

Model 13.



AICc=70.11

3. Model convergence checks

Figure IV S6 Assessment of the convergence of the linear mixed-effect model linking individual fetus mass (log-transformed) as response variable, to fetus sex and mother identity as explanatory variables and father identity as a random effect through (a) visual inspection. Note that because many mother identities were included in the analysis, only the convergence diagnostics for the intercept, sex and father identity are shown.

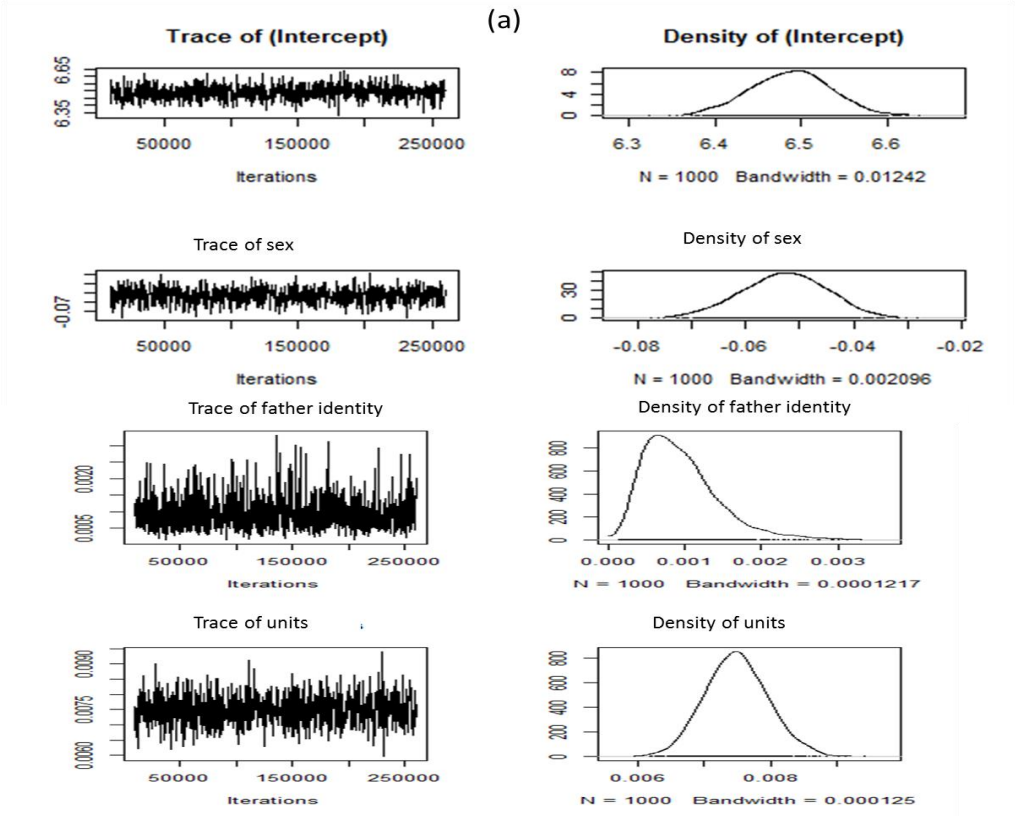


Table IV S8 Assessment of the convergence of the linear mixed-effect model linking individual fetus mass (log-transformed) as response variable, to fetus sex and mother identity as explanatory variables and father identity as a random effect through (a) Heidelberger and Welch´s convergence diagnostic and (b) Geweke´s convergence diagnostic. Note that because many mother identities were included in the analysis, only the convergence diagnostics for the intercept, sex and father identity are shown.

(a)	**Stationarity test**	**P-value**	**Halfwidth test**	**Mean**	**Halfwidth**
Intercept	Passed	0.108	Passed	6.49	0.003
Sex	Passed	0.813	Passed	-0.053	0.0006
Father ID	Passed	0.318	Passed	0.001	$3.07.10^{-5}$
Units	Passed	0.807	Passed	0.007	$2.92.10^{-5}$
(b)	**Z-score**				
Intercept	-1.702				
Sex	1.154				
Father ID	0.301				
Units	0.144				

Chapter V
Characteristics of reproductive males

In this chapter, I investigated if characteristics of male wild boars could influence their mating success and reproductive success. Despite intensive sampling, rather low proportion of fathers were identified among fetuses. The study is far from exhaustive and has several pitfalls, considering that sample males may sire offspring which are non-sampled, male characteristics when they are killed does not necessarily reflect their characteristics when they mate. However, these results allow to get some highlights about which males reproduce in the population of Châteauvillain.

Introduction

Sexual dimorphism is often associated with intense intra-sexual competition to access reproduction for the sex displaying the most exaggerated characteristics (Darwin 1871). In ungulates species where sexual dimorphism occurs, it is displayed by males. It takes a wide range of forms, from antlers in Cervidae, horns in Bovidae, elongated teeth in Suidae and few Cervidae species, to increased body size compared to females (Jarman 1983; Emlen 2008). All these characteristics are used by males to access reproduction by outcompeting other males. The sexual dimorphism is especially important in species with polygynous mating system (only one male reproduce with several females). As few males reproduce with many females, the variance in male mating success (number of mating) is high and competition is particularly intense (Wolff and MacDonald 2004). Only males able to monopolize group of females access copulation and increase their reproductive success (number of offspring sired, in other words, successful mating). However, sneaker reproductive strategies are described in male ungulates, with subordinate males waiting for a mating opportunity with a female instead of trying to monopolize a whole group (Isvaran 2005; Simmons and Fitzpatrick 2012). This explains why, in some species, genetic studies showed that observed number of mating can be a poor predictor of male reproductive success (Amos et al. 1993; Coltman et al. 1999; Rus Hoelzel et al. 1999).

Genetic studies changed conclusions based on behavioral observations regarding mating processes for many species. This is especially true for species where females engage in multiple male mating. Genetic analyses of their offspring highlighted behavior that were never observed before, due to short male-female interaction, nocturnal mating or inaccessible mating place. From there, studies compared male mating success from paternity analyses and from traditional methods including harem size and number of observed copulations with contrasting results. For example, for fallow deer (*Dama dama*) high number of observed mating lead to high number of fawns fathered (Say et al. 2003), while elephant seal (*Mirounga leonina*) showed rather low reproductive success compared to number of mating for some dominant males (Rus Hoelzel et al. 1999). Paternity assignment also allowed to investigate characteristics of fathers. For example, Sorin (2004) showed that in fallow deer older males have higher probability of reproducing but yearlings also father some fawns. Amos et al. (2001) showed in a multiple species analysis that relatedness between parents has a negative impact on their reproductive success.

From the long-term monitoring of the wild boar population of Châteauvillain, we performed paternity analyses on litters of pregnant females killed at hunting to quantify number

of mating partners and offspring for males. Paternity studies require intense sampling of females, but also of males to be able to identify which males access reproduction and which ones do not. Working on pregnant females from hunting bag allows to have good quality data regarding litter size and mother offspring relationships, however this does not give access to litters of female remaining alive in the population. This suggests that we do not know if males, that do not sire any offspring in our litter of sampled females, did not sire any offspring at all or if they only sire some in other females. However, we can have information regarding sampled males that sired offspring in sampled females. We evaluated if genetic characteristics or body mass of males influence their mating success and reproductive success. Despite the promiscuous mating system of wild boars, we expect a higher number of mating partners and offspring sired for big males. Also, we expect higher proportions of litters sired by males with higher genetic diversity.

Material and methods

We used results of COLONY analyses (Jones and Wang 2010) from the Chapter IIIA. In order to ensure sufficient samples numbers, only the population of Châteauvillain has been investigated in the present study. COLONY assigned a father to each fetus among sampled males. If no sampled male's genotype explained the genotype of a fetus, a potential father identification was created for the reconstructed genotype. This analysis resulted in 347 fetuses assigned among 85 sampled males (832 fetuses were assigned among 280 different fathers inferred from COLONY).

Based on the 365 fathers (sampled and inferred from COLONY) and the 226 sampled mothers, we have calculated the mating success (number of partners) and the reproductive success (number of fetuses sired) of all these individuals (thus only wild boars that produced at least one progeny were included). Following Jones (2009), we calculated the opportunity for sexual selection for each sex (I_s) and the between sex difference (ΔI), describing the difference in strength of sexual selection (Shuster and Wade 2003). The Bateman's gradients (β_{ss}) for males and for females was estimated by fitting linear models with the mating success as an explanatory variable and the reproductive success as an explained variable.

Focusing on sampled fathers (85 sampled males that sired at least one progeny for which the dressed body mass was known), we recorded the number of partners the males reproduced with and offspring produced per males and per year. Indeed, males that reproduced during several years are more likely to mate with more females and produce more offspring. We also

estimated individual heterozygosity for each male using the GENHET function (Coulon 2010) in R. The number of mating partner per year and of fetuses produced per year were log transformed and were used as explained variable in two different Gaussian regression models. The effect of individual heterozygosity (*Ho*) and of the body mass of males (*BM*) were included in both models

The pairwise relatedness was measured by the index of Queller and Goodnight (1989). They were calculated between mothers and fathers ($r_{Q\&G\text{-}pair}$) using SPAGeDI version 1.5 (Hardy and Vekemans 2015), when the father was identified among sampled males. We also calculated the relatedness between the mothers and all possible mating partners (males sampled included in COLONY analysis *i.e.* all sampled males killed the same year n as the mother, all adults and subadults of the year n-1 and only adults of the year n-2, Table III 1) to obtain the relatedness for random mating ($r_{Q\&G\text{-}possible}$). For each female, the values of relatedness with the fathers of its fetuses were compared to the values of relatedness with all other possible partners (random mating). The mating was considered to be with a male more related than random when the relatedness of the couple was higher than 97.5% of the relatedness values obtained with all possible matings ($r_{Q\&G\text{-}pair} > r_{Q\&G\text{-}possible\text{-}97.5}$) and with a male less related than random when the relatedness of the couple was below 2.5% of the relatedness values obtained with all possible mating ($r_{Q\&G\text{-}pair} < r_{Q\&G\text{-}possible\text{-}2.5}$).

Results

Males mated with one to five females killed at hunting (median = 1 ± 0.8 SD whether males were sampled or not, Figure V 1a) while females mated with one to six males (median = 2 ± 1.25 SD, Figure V 1b). Males produced 1 to 23 offspring (median = 3 ± 3.41 SD for sampled males and median = 2 ± 2.94 SD for inferred males, Figure V 1c) while females had 2 to 10 fetuses (median = 5 ± 1.67 SD for inferred males, Figure V 1d). Most males (67%) produced offspring the same year as they were killed (Figure V S5). Only three males produced offspring over more than one year (maximum of two years, Table V S1).

The opportunity for sexual selection was similar for females ($I_{females} = 0.29$) than for males ($I_{males} = 0.29$). The Bateman's gradient was significantly positive for both females ($y = 0.29x + 4.50$, $r^2 = 0.04$, $p = 0.001$, Figure V 2) and males ($y = 2.35x - 0.18$, $r^2 = 0.37$, $p < 0.001$, Figure V 2). However, the number of females they mated with and the number of offspring sired were not influenced by male individual heterozygosity or body mass (Figure V 3, see also Table V S2).

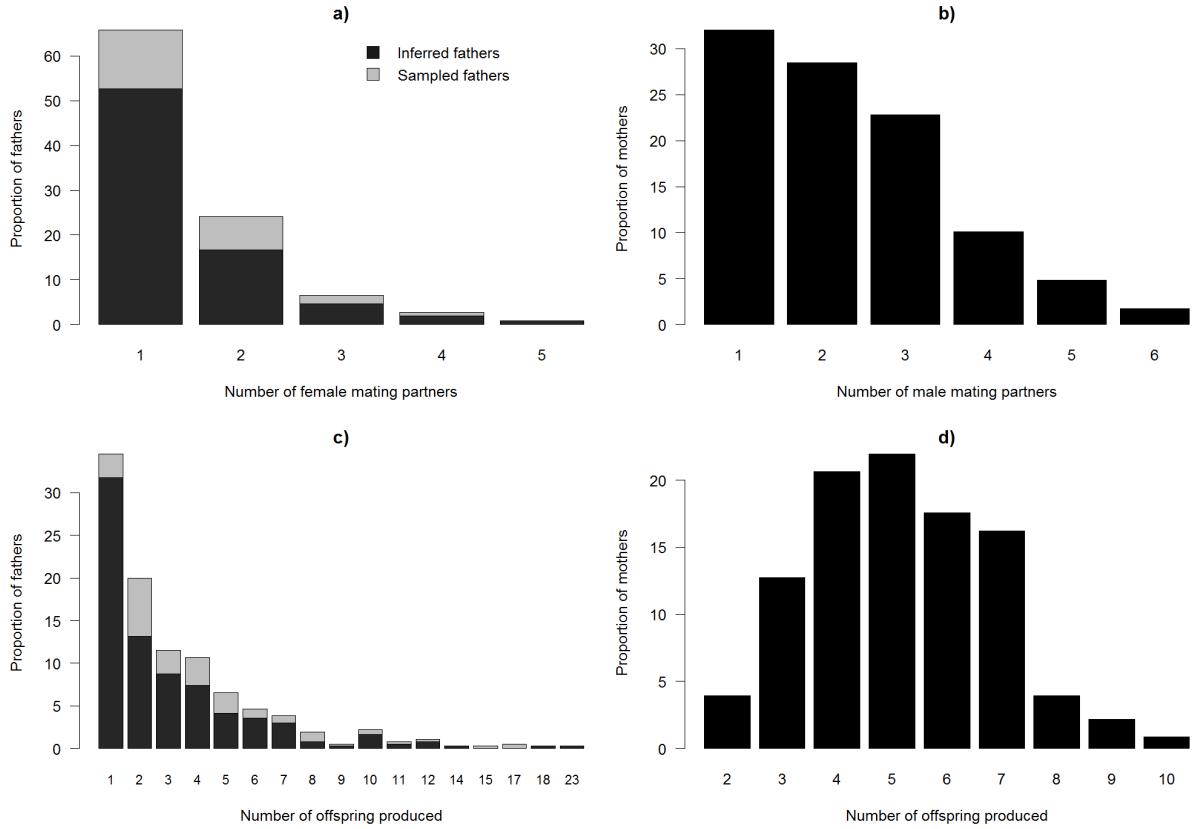


Figure V 1 Distribution of the estimated number of mating partners for a) males (n_{males} = 365) and b) females ($n_{females}$ = 226) obtained from COLONY and number of offspring produced for c) males (estimated with COLONY) and d) females (n_{fetus} = 1179). For males, the light grey parts represent the proportion of fathers from sampled males and dark grey parts the proportions of males inferred by COLONY.

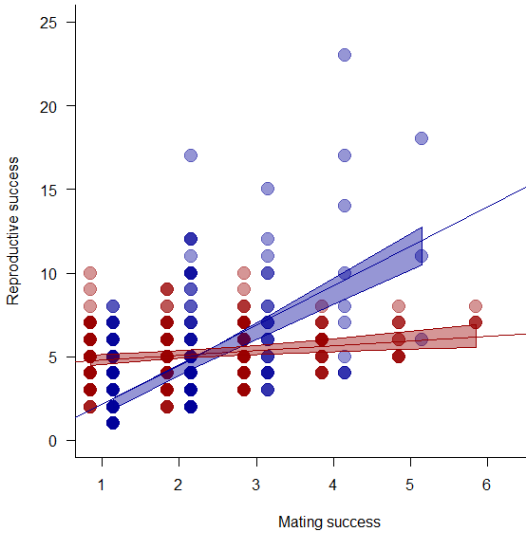


Figure V 2 Bateman's gradients for males (blue, n_{males} = 365) and females (red, $n_{females}$ = 226) wild boar of the population of Châteauvillain (2007–2015). The intensity of the color of the points reflect the number of observations (n_{fetus} = 1179 for both sex, see also Figure V 1c and d). Shading represents 95% confidence intervals.

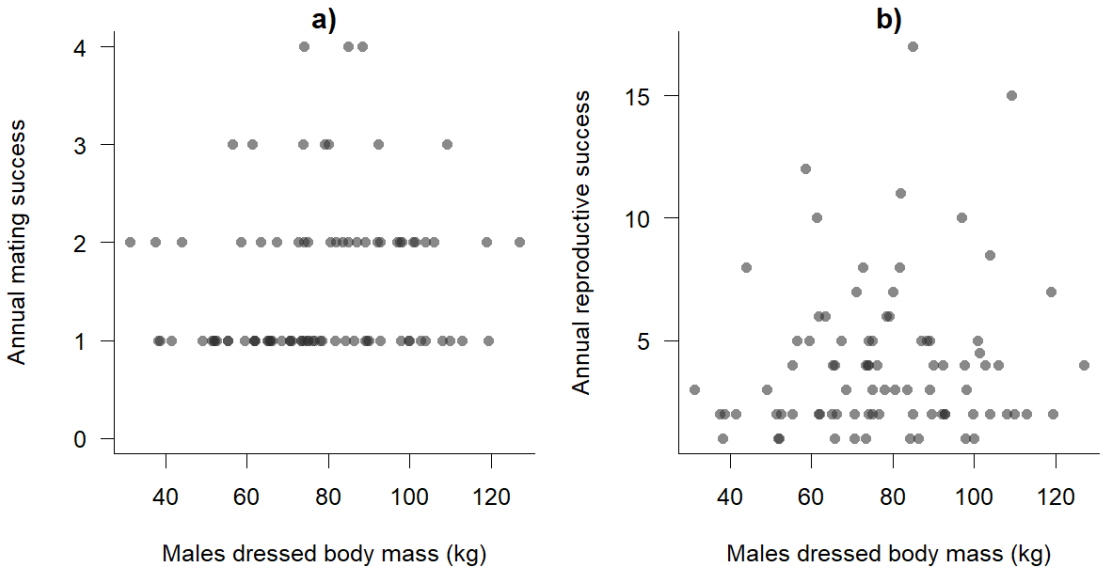


Figure V 3 Annual mating success and reproductive success according to the dressed body weight of wild boar males of the population of Châteauvillain that sired at least one offspring in litters of females sampled at hunting ($n_{fathers}$ = 85). The intensity of the color of the points shows the density of observation.

Values of relatedness between mothers and fathers were not different from random mating in 121 out of 131 mothers-father pairs ($r_{Q\&G\text{-}possible\text{-}2.5} < r_{Q\&G\text{-}pair} < r_{Q\&G\text{-}possible\text{-}97.5}$, Figure V 4). Two mating pairs were less related than random and seven were more related than random (Figure V 4).

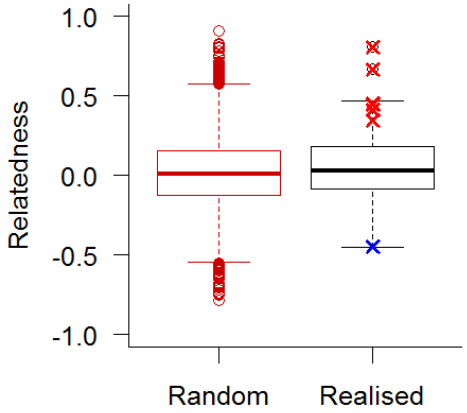


Figure V 4 Comparison of the relatedness from Queller and Goodnight expected under random mating choice (red boxplots, n= 26994) and from realized mating (with sampled fathers only, black boxplots, n=131) of female wild boars from the population of Châteauvillain. Comparisons were realized for each mating pairs to possible matings calculating 97.5 and 2.5 percentile. Red crosses symbolize mating pairs obtaining pairwise relatedness values above the $r_{Q\&G\text{-}possible\text{-}97.5}$ value. Blue crosses symbolize mating pairs obtaining pairwise relatedness values below the $r_{Q\&G\text{-}possible\text{-}2.5}$ value. Note that the value above $r_{Q\&G\text{-}possible\text{-}97.5}$ or below $r_{Q\&G\text{-}possible\text{-}2.5}$ are not necessarily extreme values as the comparison depends of relatedness of the mating pair relative to all possible mating pairs and all are pooled in the graph.

Discussion

Comparison of males and females mating success and reproductive success showed that medians of both were higher for females, but variance of reproductive success was higher for males, with 1 to 23 fetuses sired by a male. Males gain higher reproductive success by increasing their number of mating partners compared to females, but the opportunity for sexual selection was not different between sexes. Male characteristics did not influence their mating success neither their reproductive success. Altogether, results highlight that mating seems to be random for wild boar of the population of Châteauvillain.

We observed a positive relationship between the number of mating partners and the number of offspring produced for both males and females. Such pattern was already reported in Eastern chipmunk (*Tamias striatus*), although with a less steeper slope for males (Bergeron et al. 2012). This suggests that both males and females, but especially males, obtain fitness benefits from additional mates.

Contrary to what is reported in other ungulates, we did not find any influence of male body mass on their mating success or reproductive success (McElligott et al. 2001; Preston et al. 2003). However, wild boar can show great annual variation of body mass, especially rutting males. The body mass of the day they are killed may be a poor proxy of their competitive ability during the rut, especially if the time elapsed is long. Some males sired offspring more than one year before being killed at hunting. However, most males were sampled the same year as the mating was recorded suggesting that males have few mating opportunities before being killed at hunting. In this heavily hunted population, sneaker males may obtain many mating opportunities due to the disturbances caused by hunting (females often scattered by hunting dogs) and the removal of big males, homogenizing the reproductive success. The null value of ΔI shows that sexual selection is low in this population and not higher in males than in females which supports the promiscuous mating system previously described for this species (Bergeron et al. 2012). Considering these results, no mate choice exists in this population. Investigating if this is true for other populations remains to be tested.

The randomness of the mating system is supported by genetic results. Analyses of relatedness and the results about individual heterozygosity also showed that mating occur randomly and there is no preference toward genetically similar or dissimilar mates. These results are contradictory to the conclusions of the study of Pérez-González et al. (2017) who found evidence of outbreeding avoidance in five populations of wild boars from Spain, Portugal and Hungary. They hypothesized that inbred mating may be associated with fitness benefits,

but they were not able to test it. In our focus population, such benefits may not be strong enough to favor outbreeding avoidance. Moreover, high hunting pressure greatly reduce survival. Mating opportunistically may be a best strategy than risking a reproductive failure by trying to select partner based on genetic characteristics (Kokko and Mappes 2013). Also, the results may be different because their methodology to assess relatedness between mating partners was different. Indeed, inferring parents' relatedness from fetuses allowed them to get a higher sample size but it did not allow to measure relatedness of mating partners directly, which can influence the outcome of the result (Van De Casteele et al. 2001)

We acknowledge this study has some limits. Most litters produced a given year cannot be sampled as they are produced by females that are not killed, especially big females which are protected. Thus, it is difficult to quantify the proportion of undetected fetuses produced by sampled males. We cannot know if we sampled the whole progeny of a male that may lead to an underestimation of both their mating and reproductive success. However, the results highlight random mating that would suggest that samples are representative of the overall trend in the population. The probability of having or missing fetuses sired by a male should be the same for all other males. Another limit is the sampling protocol of males. Due to budget constraints, sampling focused mainly on big males as they were supposed to be the most successful. This may also explain the lack of effect of body mass on mating success.

Supplementary Material

1. Dressed body mass of sampled males

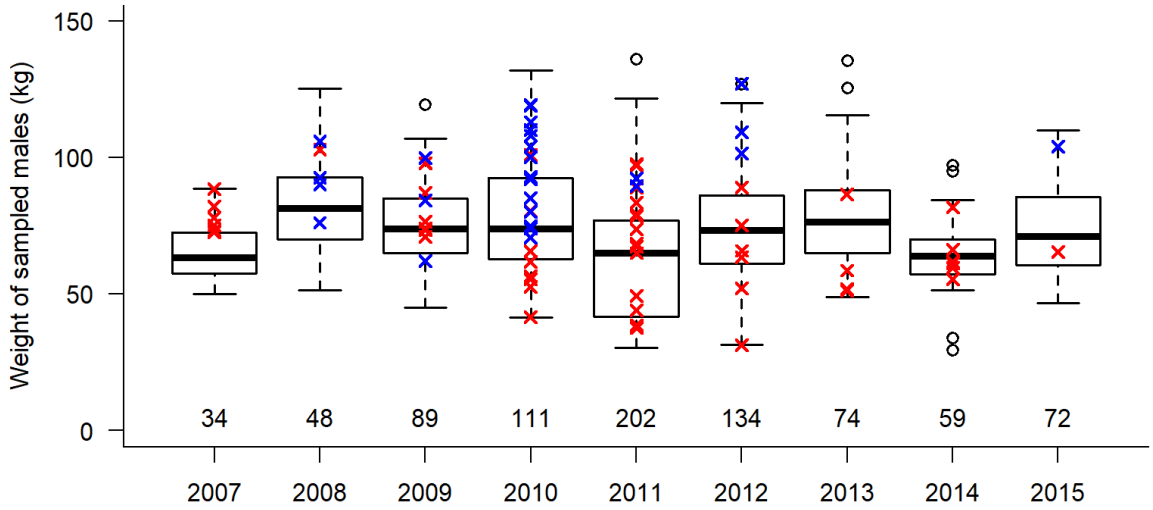


Figure V S5 Dressed body mass of males sampled for paternity analysis from the population of Châteauvillain. Sample size are given by the values below each plot (note that for b. they slightly differ from Table II 5 as weight was not available for all sampled males). For the second figure, the horizontal thick bars in the middle show the median, box upper and lower ends show quartiles, and upper and lower bars show extreme values, except when outliers are present (white dots), they show 1.5 the inter-quartile space. Red crosses indicate the dressed weight of identified fathers sampled the same year as their offspring (n_{males} = 57). Blue crosses show dressed weights of males that sired at least one offspring a year previous their sampling year (n_{males} = 28).

2. Mating and sampling year for identified fathers

Table V S1 Distribution of identified fathers ($n_{fathers}$ = 85) from the wild boar population of Châteauvillain separated by year of sampling and year when they produced offspring (mating year). When two years of mating are reported, male produced offspring during both years.

		Sampling year								
		2007	2008	2009	2010	2011	2012	2013	2014	2015
Mating year	2007	6	4							
	2008		2	3						
	2009			8	14					
	2009-2010				1					
	2010				9	2	1			
	2011					15	1			
	2011-2012						1			
	2012						6			
	2013							4		
	2013-2014									1
	2014								6	
	2015									1

3. Model selection for mating success and reproductive success

Table V S2 Model selection to test the effect on (a.) the yearly number of offspring produced and (b.) the number of female mating partners of the individual heterozygosity index (*Ho*) and the dressed body mass (*BM*) for the male wild boars from the population of Châteauvillain, France (n_{males} = 85).

a.

Intercept	*Ho*	*BM*	df	logLik	AICc	delta	weight
1.162			**2**	**-87.59**	**179.32**	**0**	**0.434**
0.837		0.004	3	-86.92	180.14	0.82	0.289
1.312	-0.153		3	-87.46	181.21	1.89	0.169
0.982	-0.145	0.004	4	-86.80	182.11	2.78	0.108

b.

Intercept	*Ho*	*BM*	df	logLik	AICc	delta	weight
0.092		**0.003**	**3**	**-48.68**	**103.67**	**0**	**0.368**
0.356			2	-49.76	103.68	0.01	0.367
0.289	0.068		3	-49.70	105.70	2.03	0.133
0.018	0.074	0.003	4	-48.61	105.72	2.05	0.132

Chapter VI

Discussion and perspectives

Overview

On the first part (Chapter IIIA), the investigation on the mating system between five populations of wild boars showed high levels of multiple paternity in all populations, clearly demonstrating the promiscuous mating system of the species in hunted populations. Longitudinal data analyzed in the population of Châteauvillain (Chapter IIIB) showed that probability of female to engage in multiple male mating was influenced by the proportion of big males in hunting bags and body mass of dead males, but not by food resources availability. This suggests that the mating system, in this population at least, is mostly influenced by the quantity and quality of available males in the population.

Investigating the relation between the number of fathers estimated with different methods and the number of offspring of females highlighted the positive effect of the number of males contributing to a litter on the fertility of females (Chapter IVA). Increased number of males contributing to the next generation allows to maintain high levels of genetic diversity to withstand yearly bottleneck ensuing from harvesting, compared to a polygynous mating system. No effect of the father identity was found on fetus mass (Chapter IVB) and the number of fathers did not influence phenotypic diversity within litter previously reported in wild boar. This confirms that variation of mass of fetuses in a litter is due to different allocation of female resources rather than a paternal effect.

Finally, reproductive success of both males and females increased with the number of mating partners (Chapter V). However, neither male body mass or genetic characteristics influenced their reproductive success showing that mating is random and bigger males do not sire more offspring.

Parentage studies

Since the development of molecular makers, they were widely used to study very different topics in ecology and evolution. These topics include, but not only, speciation, hybridization, migration and sexual selection. At the beginning, only low resolution markers were known, mostly protein based marker like allozyme (Hubby and Lewontin 1966). Parentage analyses were developed later because they require a higher discriminant power. The discovery of PCR in 1983 (Mullis et al. 1986) and the characterization of microsatellites in 1984 (Jeffreys et al. 1985) made possible reliable genetic identification. From there and up to now, microsatellites became markers of choice in forensic science and paternity studies (Jones and Ardren 2003).

Moreover, nowadays a lot of software are available for parentage analysis depending on the question investigated and data sampling. Most of them are described in Jones et al. (2010).

For my PhD, I used four different technics of paternity analysis each having its advantages and disadvantages. Here are some feedbacks of my experience about the different methods, they may not be adapted for analysis with significantly higher sample size or number of loci (Table VI 1). The simplest was to calculate the number of paternal allele (NPA). I performed the analysis using R software (R Core Team 2017), which is very quickly executed. New samples can be analyzed very easily by slightly adapting the code. However, this method only allows to have a proxy of the number of fathers in a litter. GERUD (Jones 2005) estimates the minimum number of genotypes (in our cases fathers, as mother-offspring relations were known) necessary to explain the genotypes of a litter. Simulations confirmed that it tends to underestimate the real number of fathers. However, technical limits worth to be mentioned for this software. Indeed, the analysis is time consuming for several reasons. The whole dataset needs to be partitioned in a lot of smaller ones. To avoid computational problems, we only included five loci (the most polymorphic) in the analysis out of twelve. Then, the software takes as an input only one mother-litter array and each of them must be analyzed separately (with repeated selection of parameters), giving an individual output. Also, as allele frequencies are required, analyses need to be performed again when new genotypes of adults are included, in order to be the most accurate and up to date possible. Thus, all mother-litter arrays must be analyzed individually again with the new allele frequencies to remains consistent over the different sampling years. The software itself is not hard to use but the analysis can be laborious. CERVUS (Kalinowski et al. 2007) allows to assign parents (in my case fathers) from sampled individuals (males) to offspring. It is often the favored method for parentage identification (Jones et al. 2010). However, it is especially efficient when most of possible breeders are sampled as it assigns a father among sampled males only. Thus, no father is assigned if no sampled males fit its criteria leading to litters with missing data (incomplete resolution of the fathers of the litter) which do not allows to find the number of mating partners of females. As for GERUD, the use of the software requires few interventions of the user but to a lesser extent, as all fathers and offspring can be put in two arrays (one for each) and it makes fewer outputs. Finally, COLONY (Jones and Wang 2010) was used because like CERVUS, it can assign fathers from sampled males and like GERUD, it makes genotypes reconstitution when no sampled males is identified as father. This means we can estimate the number of father for every litters included in the analysis. However, it was not used for Chapter IA because permutation analysis was not

possible to test the positive link between litter size and number of father later reported in Chapter V (no configuration showed a number of offspring higher or equal than the number of fathers). The computational time can be long however computing facilities in the lab allowed to perform a lot of analyses simultaneously. This allowed me to perform analyses with different parameters, especially changing the proportion of sampled males which is a determinant parameters in paternity analysis (Jones et al. 2010).

Table VI 1 Description of positive and negative sides of four parentage methods used for paternity analysis in wild boars concerning their use and results obtained.

	NPA[μ]	GERUD	CERVUS	COLONY
Estimation of the number of fathers	Proxy	Yes	Yes	Yes
Father identification (from sampled males)	No	No	Yes	Yes
Number of loci restriction	No	Yes	No	No
Father genotype reconstruction (non-sampled males)	No	Yes	No	Yes
Analysis speed	Fast	Fast	Fast	Slow (but parallelization possible)*
Automating analysis possible	Yes (R code easily adapted)	No (lot of user's manipulations required)	Yes (Run from command line)	Yes (Run from command line)
Input format	Flexible (User's choice)	Not easy	Easy	Adaptable template with the software
Output format	Flexible (User's choice)	One array per litter	One array per analysis	Many arrays per analysis
Linux version available	Not required*	No	Not required*	Yes*

[μ]Number of paternal alleles

*only from my own judgement considering the data I had to work with. For COLONY, Linux version allowed quick parallelized calculations on the computing facilities of the CC LBBE/PRABI.

Multi-sites studies

Working with the French Hunting and Wildlife Agency (ONCFS) allowed me to have access to exceptional data from five different populations of wild boars. Studies based on multiple sites are essential to infer general conclusions on species characteristics. This is especially true when investigating parameters depending on social factors, such as reproduction mechanisms. Population characteristics often have high impact on these processes for variety of taxa (Schoen and Brown 1991; Griffith et al. 2002; Maher and Burger 2011), although I did not record any variation between populations in our study. It is noteworthy my results are obtained from descriptive analyses, as I was not able to include fine scale information for all populations. More in depth investigations should be realized by including parameters similar to the study in Châteauvillain (Chapter IIIB: resource availability; mating date). Also, conclusions would be greatly improved by obtaining information of the structure of each population. Except in Châteauvillain where Capture-Mark-Recapture protocol is performed for many years, other populations are not as intensively monitored. Such intensive population monitoring requires a lot of field work and to invest time and money to capture and tag wild boars. This explains why it was not possible to collect so much information in every populations. However, obtaining information of the population structure before rutting (and hunting) would be very interesting. In particular, estimating the number of reproductive males (separating big males from medium size males) and females before hunting and linking them with the hunting bag records would allow to get temporal estimations of which individuals remain in the populations. Thus, parameters such as the operational sex-ratio and the proportion of big males could be inferred to measure the intra-sexual competition between males. In large and open population like Châteauvillain, such intense monitoring may be difficult. Indeed, it is already heavily monitored and increasing catch rate would be problematic. Data from such long-term monitoring are priceless (around forty years now!) and such program should not be forsaken. However, if achievable, leading intense monitoring in closed population like Belval with smaller number of individuals could be very interesting. Indeed, small closed populations have already proven helpful to shade light on interesting mechanisms in ungulates, including reproductive processes. For instance, Soay sheeps (*Ovis aries*) from the Hirta island, with a similar surface to Belval, are subject to many studies. In this population, where promiscuous mating system was observed, males mating success and reproductive success were investigated regarding morphological measures such as body mass, horn length and testes size and evidence of sperm depletion in male were reported (Preston et al. 2001, 2003). It is noteworthy that this population

is intact from human intervention for many year, contrary to Belval where wild boars are supplementary fed. Similarly, the island of Vega hosts a population of around 25 moose individuals (*Alces alces*) allowing very precise studies as 100% of animal are radio-collared. Moreover, experimental tests were realized by manipulating population structure by removing older males to mimic selective hunting, for example (Sæther et al. 2004). This is only possible in small and isolated populations. Of course, observations in closed populations like Belval should be confirmed with observations in free-ranging ones, as they do not necessarily reflect natural processes.

Life history traits: where does wild boar fit?

Belonging to ungulates, wild boar is often compared with other species such as red deer (*Cervus elaphus*), as they co-occurs in the same areas, or bighorn sheep (*Ovis canadensis*), widely studied for hunting consequences, in studies of behavior and life history traits (Mysterud 2000; Scillitani et al. 2010; Frantz et al. 2012; Prévot and Licoppe 2013; Gamelon et al. 2014). Also, due to their overabundance and agroforest damages ensuing, it can also be compared to white-tailed deer (*Odocoileus virginianus*), which is a problematic species in North-America (Russell et al. 2001; Rutberg et al. 2004; Gortázar et al. 2006; Barrios-Garcia and Ballari 2012). These comparisons make sense regarding several traits and phylogeny (Figure II 1, Price et al. 2005). However, other traits highly deviate from other ungulates including diet (Schley and Roper 2003), litter size (Servanty et al. 2007), juvenile survival (Jezierski 1977), and senescence (Gamelon et al. 2014). These traits are closer to far different species. In late 70's, Jezierski (1977) already pointed out striking similarities between wild boar and European hare (*Lepus europaeus*) concerning juvenile survival patterns and reproduction of young individuals. However, he noticed that female wild boars reproduce when they reach two years of age while female hares can reproduce in the end of their first year. It is noteworthy that reproduction in wild boar females in their first year of age occurs, associated with heavy hunting mortality, which gives support to the comparison with hare (Gamelon et al. 2011). Focardi et al. (2008) confirmed, once again, Jezierski's comparison by reporting low ratio of weight at independence to adult weight, contrary to most ungulates, and closer to species in the fast end of the life-history continuum. They also point out that mammals with comparable litter size and body mass are mostly found in carnivores, for example gray wolf (*Canis lupus*), arguing that such reproductive outputs are made possible by their diet. Indeed, except suids, no ungulate produce litter larger than three offspring but in the same time, most ungulates are strictly herbivore.

Finally, wild boar exhibit traits of species from the slow part of the slow-fast continuum (Gaillard et al. 1989, 2000b).

Regarding its mating system, due to the big litter size, the wild boar is hardly comparable to other ungulates in term of multiple-male mating. However, promiscuous mating system has already been reported in Soay sheep (*Ovis aries*), with reports of females mating with up to ten different males during 47 mating event in one day based on behavioral observations (Grubb and Jewell 1973, cited in Coltman et al. 1999). It is noteworthy that such observations are possible in this population because it lives in a deserterd island. Habitat and nocturnal behavior of most of other ungulate species does not allow to record this kind of behavior. Moreover, for ungulates, number of offspring per litter does not allow to measure such intense multiple-male mating of females. In my study (Chapter III), the highest number of sire reported was seven which probably underestimates the number of mating partners of the females. However, estimating number of mating is not possible as most of them occur at night and in the forest. Thus, the estimation from the number of fathers contributing to a litter makes inevitably the wild boar close to other mammals with similare litter sizes, mostly rodents like squirrels or marmots (Martin et al. 2014; Wells et al. 2017). To confirm the mating strategy, identifying (with genetic or behavioral study) which males reproduce in an hunted population would be interesting. Just like with Soay sheep, different males strategies would probably co-occurs with big males defending groups of females and smaller males using sneaker strategy by remaining near the females and trying to get access whenever it is possible for them.

Demography of wild boar: Do males matter?

This question was already raised in a paper from Rankin and Kokko (2007) in a general context approach. They explain how effect of the change of the proportion of males depends on species-specific characteristics. Also, using simple models, they show how female fertilization probability can be reduced when proportions of males are very low. At the very least, removal of males in population subject to density-dependence can increase resource availability for females by reducing intraspecific competition (Clutton-Brock et al. 2002). This may lead to direct increase of density as reproduction of females is improved. Subsequently, increase of proportion of males can decrease fitness of females. Less directly, high proportion of males can impair female fitness with sexual aggression and harassment during rutting period (Réale et al. 1996). On the other hand, decreasing proportion of males in the population can impact density.

In population with high rates of removal of males, in most extreme cases, sperm limitation can occur leading to decreased fertility of females (Milner-Gulland et al. 2003). Also, in some species high turn-over of males induce increased infanticide as male would kill female offspring to induce a new estrus (Whitman et al. 1987). Altogether and as concluded by Rankin and Kokko (2007), the effect of males on population demography is not as trivial as it seemed (Caswell 2001) and demographic models including males should be considered.

My results (Chapter IV) support this statement. The number of partners of a female has a positive impact on its fertility. This suggest that the number of males available for mating (present in the population and able to access females without competitor) can influence wild boar demography. However, further studies need to be realized to measure the real effect of males. Gamelon et al. (2012) developed a body weight-structured model to estimate growth rate of wild boar populations. They showed that body weight was a better structuring factor for this species, and an easier information to collect than age which is usually used in demographic models (Caswell 2001). Indeed, this model mainly includes sex and weight of wild boar shot a given year. Implementing this model with a parameter allowing the fecundity of females to vary in response to the relative proportion of small and big males (the ratio $\frac{Number\ of\ small\ males}{Total\ number\ of\ reproductive\ males}$ for example so that fertility increases when the opportunity for females monopolization decreases) could allow to measure how multiple paternity can impact wild boar demography. Such results would greatly complement studies focusing on quantifying how males can impact demography. However, it is worth mentioning that such studies are interesting from a fundamental research point of view to shade light on processes on population demography, but management applications are rather limited especially in a species like the wild boar.

First of all, the increased fecundity of females observed with their number of mating partners is significant but would hardly justify management strategies. Indeed, protecting big males would reduce multiple-male mating opportunity (and multiple paternity rates), but small males will probably still use sneaker strategy. Encouraging small males culling could be a solution but the hunting effort required would be very important to observe a reduction of female fertility. Moreover, the reduction of fertility observed would be disappointing regarding invested efforts. Gamelon et al. (2012) showed that slight increase of hunting effort on big females could have a great impact on wild boar demography as they are the one with the highest reproductive outputs. The cost benefits ratio of this strategy is thus better due to its simplicity and efficiency.

The second pitfalls for management applications concerning wild boars, independently of the method proposed, is the reluctancy of hunters to change their habits. In the seventies, big females where protected to increase wild boar populations because in that time, wild boars were rarer (Servanty 2007). This way of thinking remains up to now even of populations are high and still increasing. For example, in the population of Châteauvillain, big females (> 50kg) are still protected and forfeits must be paid by the transgressor, and the amount is proportional to the weight of the female killed. When big females are not protected, it is not rare to hear hunters complain when one of them killed pregnant females, near the end of the hunting season when fetuses are easy to observe: "Thanks to you, we just lost six wild boars for the next hunting season!" if the dead females was pregnant with six fetuses (conversation I heard during field work). If hunters are aware of the problems associated with the demographic explosion of the wild boar, they are also worried about the sustainability of their practice and want to guarantee the results of the next hunting season. Indeed, Keuling et al. (2016) elegantly showed the gap between their perception of wild boar management at local scale and at large scale. In their survey, most hunters agreed that populations of wild boars should be reduced, however a smaller proportion agreed measures should take place in their own hunting ground. Authors' conclusion was hunters consider that regulation of wild boar population is "somebody else's problem". Such studies highlight the importance of the social part of management strategies, especially for species like wild boar which inevitably forces people from very different contexts to interact, as already mentioned in the introduction. The problematic part is that each stakeholder has its own aims, making discussion complicated and often with pointless conclusions. Indeed, managers want to reduce populations to respond to both agricultural and societal concerns about the demographic explosion. Hunters want abundant game species to get a return on investment on their hunting license and they feel pressured to increase hunting pressure by manager. Due to the will to practice their activity in their own terms, the society often gets a bad image of hunters, especially with over-mediated cases of deviant behavior from hunters such as the killing of Cecil the lion (Nelson et al. 2016).

Conclusion

In five hunted populations of wild boars, we found high levels of multiple paternity highlighting a promiscuous mating system of the species in such context. Moreover, from longitudinal analyses, we observed variations of mating system in response to hunting pressure showing the high plasticity of the wild boar to face environmental changes. Promiscuity is

known to promote the transmission of genetic variability from one generation to the other by increasing the number of males that contribute to reproduction. These traits allow this already very plastic species to maintain a high adaptive potential. Moreover, wild boar females have high reproductive outputs, and we recorded an increase of fecundity with the number of mating partners. To my knowledge, it is the first time that the mating system is shown to increase the population demography in this species. The wild boar displays life history traits associated to invasive species including high fecundity, broad diet and high adaptability explaining the increasing population trends currently observed. The high population dynamic of the species can also be influenced by other factors that can impact the fecundity of the species. For example, changes in agricultural practices provide high quantity of food that can influence reproduction. Also, hybridization is a major evolutive force because it can induce great evolutionary changes in a very short time scale in a species. For the wild boar, hybridization with domestic pig was already suggested as a factor explaining its demography. Indeed, domestic pig is the product of intense artificial selection to improve growth rate and female fecundity to increase productivity. Hybridization between the two subspecies can introduce pig selected genes in wild boar populations and increase fecundity of hybrids and their descendants when introgression occurs (invasion of the genome of a species by another one's genome). Investigating if the individual level of introgression and fecundity are linked remains to be studied. However, finding discriminant genetic markers is difficult due to the low differentiation between pig and wild boar. New genetic markers and/or methods need to be developed to investigate such topic (see Appendices A and B).

Bibliography

Acevedo, P., Vicente, J., Höfle, U., Cassinello, J., Ruiz-Fons, F. and Gortazar, C. **2007**. Estimation of European wild boar relative abundance and aggregation: a novel method in epidemiological risk assessment. *Epidemiol. Infect.* 135:519–527.

Adderton Herbert, C. **2004**. Long-acting contraceptives: A new tool to manage overabundant kangaroo populations in nature reserves and urban areas. *Aust. Mammal.* 26:67–74.

Aguilera-Reyes, U., Zavala-Paramo, G., Valdez-Alarcon, J. J., Cano-Camacho, H., Garcia-Lopez, G. I. and Pescador-Salas, N. **2006**. Multiple mating and paternity determinations in domestic swine (*Sus scrofa*). *Anim. Res.* 55:409–417.

Allendorf, F. W., England, P. R., Luikart, G., Ritchie, P. A. and Ryman, N. **2008**. Genetic effects of harvest on wild animal populations. *Trends Ecol. Evol.* 23:327–337.

Allendorf, F. W. and Hard, J. J. **2009**. Human-induced evolution caused by unnatural selection through harvest of wild animals. *Proc. Natl. Acad. Sci.* 106:9987–9994.

Amos, W. and Balmford, A. **2001**. When does conservation genetics matter? *Heredity (Edinb).* 87:257–265.

Amos, W., Twiss, S., Pomeroy, P. P. and Anderson, S. S. **1993**. Male mating success and paternity in the grey seal, *Halichoerus grypus*: a study using DNA fingerprinting. *Proc. R. Soc. B Biol. Sci.* 199–207.

Amos W, Wilmer JW, Fullard K, Burg TM, Croxall JP, Bloch D and Coulson, T. **2001**. The influence of parental relatedness on reproductive success. *Proc. R. Soc. B-Biological Sci.* 268:2021–2027.

Arct, A., Drobniak, S. M. and Cichoń, M. **2015**. Genetic similarity between mates predicts extrapair paternity-a meta-analysis of bird studies. *Behav. Ecol.* 26:959–968.

Asher, M., Lippmann, T., Epplen, J. T., Kraus, C., Trillmich, F. and Sachser, N. **2008**. Large males dominate: Ecology, social organization, and mating system of wild cavies, the ancestors of the guinea pig. *Behav. Ecol. Sociobiol.* 62:1509–1521.

Badinel, J.-P. **2010**. Biology and genetics of reproduction. Pp. 218–241 *in* Rotschild, M. F. and Ruvinky, A., eds. *The Genetics of the Pig*. CABI Cambridge, MA, USA.

Ballari, S. A. and Barrios-García, M. N. **2014**. A review of wild boar *Sus scrofa* diet and factors affecting food selection in native and introduced ranges. *Mamm. Rev.* 44:124–134.

Barrett, R. D. H. and Schluter, D. **2008**. Adaptation from standing genetic variation. *Trends Ecol. Evol.* 23:38–44.

Barrett, S. C. H., Colautti, R. I. and Eckert, C. G. **2008**. Plant reproductive systems and evolution during biological invasion. *Mol. Ecol.* 17:373–383.

Barrios-Garcia, M. N. and Ballari, S. A. **2012**. Impact of wild boar (*Sus scrofa*) in its introduced and native range: a review. *Biol. Invasions* 14:2283–2300.

Bart, J. and Tornes, A. **1989**. Importance of monogamous male birds in determining reproductive success - Evidence for house wrens and a review of male-removal studies. *Behav. Ecol. Sociobiol.* 24:109–116.

Baskin, Y. **1998**. Winners and losers in a changing world. *Bioscience* 48:788–792.

Bateman, A. J. **1948**. Intra-sexual selection in Drosophila. *Heredity (Edinb).* 2:349–368.

Baubet, E. **1998**. Biologie du sanglier en montagne : biodémographie, occupation de l'espace et régime alimentaire. PhD Thesis, Université Claude Bernard Lyon1, Lyon, France.

Baubet, E., Bonenfant, C. and Brandt, S. **2004**. Diet of the wild boar in the french Alps. *Galemys* 16:101–113.

Baubet, E., Brandt, S., Jullien, J.-M. and Vassant, J. **1994**. Valeur de l'examen de la denture pour la détermination de l'âge chez le sanglier (*Sus scrofa*). *Gibier Faune Sauvag.* 11:119–132.

Baubet, E., Servanty, S. and Brandt, S. **2009**. Tagging piglets at the farrowing nest in the wild : some preliminary guidelines. *Acta Silv. Lignaria Hungarica* 5:159–166.

Baubet, E., Van Laere, G. and Gaillard, J. M. **1995**. Growth and survival in piglets. *IBEX J. Mt. Ecol.* 3:71.

Bautista, A., Rödel, H. G., Monclús, R., Juárez-Romero, M., Cruz-Sánchez, E., Martínez-Gómez, M. and Hudson, R. **2015**. Intrauterine position as a predictor of postnatal growth and survival in the rabbit. *Physiol. Behav.* 138:101–106.

Bengsen, A. J., Gentle, M. N., Mitchell, J. L., Pearson, H. E. and Saunders, G. R. **2014**. Impacts and management of wild pigs *Sus scrofa* in australia. *Mamm. Rev.* 44:135–147.

Bergeron, P., Montiglio, P.-O., Réale, D., Humphries, M. M. and Garant, D. **2012**. Bateman gradients in a promiscuous mating system. *Behav. Ecol. Sociobiol.* 66:1125–1130.

Bieber, C. and Ruf, T. **2005**. Population dynamics in wild boar *Sus scrofa* : ecology, elasticity of growth rate and implications for the management of pulsed resource consumers. *J. Appl. Ecol.* 42:1203–1213.

Birkhead, T. R. **1987**. Sperm competition in birds. *Trends Ecol. Evol.* 2:268–272.

Booth, W. D. **1981**. Wild boar farming in the united kingdom. *IBEX J. Mt. Ecol.* 3:245–248.

Bosch, M., Oro, D., Cantos, F. J. and Zabala, M. **2000**. Short-term efects of culling on the ecology and population dynamics of the yellow-legged gull. *J. Appl. Ecol.* 37:369–385.

Brashares, J. S. and Arcese, P. **2002**. Role of forage, habitat and predation in the behavioural plasticity of a small African antelope. *J. Anim. Ecol.* 71:626–638.

Bueno, C. G., Azorín, J., Gómez-García, D., Alados, C. L. and Badía, D. **2013**. Occurrence and intensity of wild boar disturbances, effects on the physical and chemical soil properties of alpine grasslands. *Plant Soil* 373:243–256.

Burnham, K. P. and Anderson, D. R. **2004**. Multimodel inference: understanding AIC and BIC in model selection. *Sociol. Methods Res.* 33:261–304.

Bywater, K. A., Apollonio, M., Cappai, N. and Stephens, P. A. **2010**. Litter size and latitude in a large mammal: The wild boar *Sus scrofa*. *Mamm. Rev.* 40:212–220.

Cahill, S., Llimona, F., Cabañeros, L. and Calomardo, F. **2012**. Characteristics of wild boar (*Sus scrofa*) habituation to urban areas in the Collserola Natural Park (Barcelona) and comparison with other locations. *Anim. Biodivers. Conserv.* 2:221–233.

Calenge, C., Maillard, D., Fournier, P. and Fouque, C. **2004**. Efficiency of spreading maize in the garrigues to reduce wild boar (*Sus scrofa*) damage to Mediterranean vineyards. *Eur. J. Wildl. Res.* 50:112–120.

Cameron, C., Berteaux, D. and Dufresne, F. **2011**. Spatial variation in food availability predicts extrapair paternity in the arctic fox. *Behav. Ecol.* 22:1364–1373.

Capellini, I., Baker, J., Allen, W. L., Street, S. E. and Venditti, C. **2015**. The role of life history traits in mammalian invasion success. *Ecol. Lett.* 18:1099–1107.

Caswell, H. **2001**. *Matrix population models*. 2nd edn. Sinauer, Sunderland, Massachusetts,USA.

Caughley, G. **1981**. Overpopulation. Pp. 7–19 *in* Jewell, P., Holt, S. and Hart, D., eds. *Problems in Management of Locally Abundant Wild Mammals*.

Chee, Y. E. and Wintle, B. A. **2010**. Linking modelling, monitoring and management: An integrated approach to controlling overabundant wildlife. *J. Appl. Ecol.* 47:1169–1178.

Clutton-Brock, T. H. **1989**. Mammalian Mating Systems. *Proc. R. Soc. B Biol. Sci.* 236:339–372.

Clutton-Brock, T. H., Coulson, T. N., Milner-Guiland, E. J., Thomson, D. and Armstrong, H. M. **2002**. Sex differences in emigration and mortality affect optimal management of deer populations. *Nature* 415:633–637.

Clutton-Brock, T. H. and Harvey, P. H. **1978**. Mammals, resources and reproductive strategies. *Nature* 273:191–195.

Coltman, D. W., Bancroft, D. R., Robertson, A., Smith, J. A., Clutton-Brock, T. H. and Pemberton, J. M. **1999**. Male reproductive success in a promiscuous mammal: behavioural estimates compared with genetic paternity. *Mol. Ecol.* 8:1199–1209.

Coltman, D. W., O'Donoghue, P., Jorgenson, J. T., Hogg, J. T., Strobeck, C. and Festa-Bianchet, M. **2003**. Undesirable evolutionary consequences of trophy hunting. *Nature* 426:655–658.

Cooper, D. and Herbert, C. **2001**. Genetics, biotechnology and population management of over-abundant mammalian wildlife in Australasia. *Reprod. Fertil. Dev.* 13:451–458.

Corrêa, M. N., Lucia Jr., T., Afonso, J. A. B. and Deschamps, J. C. **2002**. Reproductive performance of early-weaned female swine according to their estrus profile and frequency of artificial insemination. *Theriogenology* 58:103–112.

Costa, V., Pérez-González, J., Santos, P., Fernández-Llario, P., Carranza, J., Zsolnai, A., Anton, I., Buzgó, J., Varga, G., Monteiro, N. and Beja-Pereira, A. **2012**. Microsatellite markers for identification and parentage analysis in the European wild boar (*Sus scrofa*). *BMC Res. Notes* 5:479–481.

Cotton, S., Small, J. and Pomiankowski, A. **2006**. Sexual selection and condition-dependent mate preferences. *Curr. Biol.* 16:755–765.

Coulon, A. **2010**. Genhet: An easy-to-use R function to estimate individual heterozygosity. *Mol. Ecol. Resour.* 10:167–169.

Dardaillon, M. **1984**. Le sanglier et le milieu camarguais : dynamique coadaptative. PhD Thesis, Université Paul Sabatier, Toulouse, France.

Dardaillon, M. **1988**. Wild boar social groupings and seasonal changes in the Camargue, southern France. *Z. Saugetierkd.* 53:22–30.

Darwin, C. **1871**. *The descent of man and selection in relation to sex*.

Davidson, Z., Valeix, M., Loveridge, A. J., Madzikanda, H. and Macdonald, D. W. **2011**. Socio-spatial behaviour of an African lion population following perturbation by sport hunting. *Biol. Conserv.* 144:114–121. Elsevier Ltd.

Davies, N. B. **1991**. Mating systems. Pp. 263–294 *in* Krebs, J. R. and Davies, N. B., eds. *Behavioural Ecology: An Evolutionary Approach*. Oxford: Blackwell Scientific Publications.

Delcroix, I., Mauget, R. and Signoret, J. P. **1990**. Existence of synchronization of reproduction at the level of the social group of the European wild boar (*Sus scrofa*). *J. Reprod. Fertil.* 89:613–617.

Delgado-Acevedo, J., Zamorano, A., DeYoung, R. W., Campbell, T. A., Hewitt, D. G. and Long, D. B. **2010**. Promiscuous mating in feral pigs (*Sus scrofa*) from Texas, USA. *Wildl. Res.* 37:539–546.

Delgado, R., Fernández-Llario, P., Azevedo, M., Beja-Pereira, A. and Santos, P. **2008**. Paternity assessment in free-ranging wild boar (*Sus scrofa*) – Are littermates full-sibs? *Mamm. Biol. - Zeitschrift für Säugetierkd.* 73:169–176.

Dexter, N. **2003**. The influence of pasture distribution, and temperature on adult body wieght of feral pigs in a semi-arid environment. *Wildl. Res.* 30:75–79.

DiBattista, J. D., Feldheim, K. a., Gruber, S. H. and Hendry, A. P. **2008**. Are indirect genetic benefits associated with polyandry? Testing predictions in a natural population of lemon sharks. *Mol. Ecol.* 17:783–795.

Douhard, M., Festa-Bianchet, M., Pelletier, F., Gaillard, J. M. and Bonenfant, C. **2016**. Changes in horn size of Stone's sheep over four decades correlate with trophy hunting pressure. *Ecol. Appl.* 26:309–321.

Eccard, J. A. and Wolf, J. B. W. **2009**. Effects of brood size on multiple-paternity rates: a case for 'paternity share' as an offspring-based estimate. *Anim. Behav.* 78:563–571. Elsevier Ltd.

Emlen, D. **2008**. The evolution of animal weapons. *Annu. Rev. Ecol. Evol. Syst.* 39:387–413.

Emlen, S. T. and Oring, L. W. **1977**. Ecology, sexual selection, and the evolution of mating systems.

Endo, A. and Doi, T. **2002**. Multiple copulations and post-copulatory guarding in a free-living population of sika deer (*Cervus nippon*). *Ethology* 108:739–747.

Engeman, R. M., Massei, G., Sage, M. and Gentle, M. N. **2013**. Monitoring wild pig populations: A review of methods. *Environ. Sci. Pollut. Res.* 20:8077–8091.

Fang, M., Berg, F., Ducos, A. and Andersson, L. **2006**. Mitochondrial haplotypes of European wild boars with 2n = 36 are closely related to those of European domestic pigs with 2n = 38. *Anim. Genet.* 37:459–464.

FAO. **2011**. *Molecular genetic characterization of animal genetic resources. FAO Animal Production and Health Guidelines.* Rome.

Feder, C., Martin, J. G. A., Festa-Bianchet, M., Bérubé, C. and Jorgenson, J. **2008**. Never too late? Consequences of late birthdate for mass and survival of bighorn lambs. *Oecologia* 156:773–781.

Fernández-llario, P. and Mateos-quesada, P. **2003**. Population structure of the wild boar (*Sus scrofa*) in two Mediterranean habitats in the western Iberian Peninsula. *Folia Zool* 52:143–148.

Focardi, S., Gaillard, J.-M., Ronchi, F. and Rossi, S. **2008**. Survival of wild boars in a variable environment: unexpected life-history variation in an unusual ungulate. *J. Mammal.*

89:1113–1123.

Fontanesi, L., Scotti, E. and Russo, V. **2008**. Differences of the porcine amelogenin X and Y chromosome genes (AMELX and AMELY) and their application for sex determination in pigs. *Mol. Reprod. Dev.* 75:1662–1668.

Fox, C. W. and Rauter, C. M. **2003**. Bet-hedging and the evolution of multiple mating. *Evol. Ecol. Res.* 5:273–286.

Frangež, R., Gider, T. and Kosec, M. **2005**. Frequency of boar ejaculate collection and its influence on semen quality, pregnancy rate and litter size. *Acta Vet. Brno* 74:265–273.

Frank, S. C., Ordiz, A., Gosselin, J., Hertel, A., Kindberg, J., Leclerc, M., Pelletier, F., Steyaert, S. M. J. G., Støen, O.-G., Van de Walle, J., Zedrosser, A. and Swenson, J. E. **2017**. Indirect effects of bear hunting: a review from Scandinavia. *Ursus* 28:150–164.

Frankham, R. **1996**. Relationship of genetic variation to population size in wildlife. *Conserv Biol* 10:1500–1508.

Frankham, R., Briscoe, D. A. and Ballou, J. D. **2002**. *Introduction to conservation genetics*. Cambridge University Press, Cambridge.

Frantz, A. C., Bertouille, S., Eloy, M. C., Licoppe, A., Chaumont, F. and Flamand, M. C. **2012**. Comparative landscape genetic analyses show a Belgian motorway to be a gene flow barrier for red deer (*Cervus elaphus*), but not wild boars (*Sus scrofa*). *Mol. Ecol.* 21:3445–57.

Fulgione, D., Rippa, D., Buglione, M., Trapanese, M., Petrelli, S. and Maselli, V. **2016**. Unexpected but welcome. Artificially selected traits may increase fitness in wild boar. *Evol. Appl.* 9:769–776.

Gabor, T. M., Hellgren, E. C., Bussche, R. a. and Silvy, N. J. **1999**. Demography, sociospatial behaviour and genetics of feral pigs (*Sus scrofa*) in a semi-arid environment. *J. Zool.* 247:311–322.

Gaillard, J., Festa-Bianchet, M., Yoccoz, N. G., Loison, A. and Toïgo, C. **2000**a. Temporal variation in fitness components and population dynamics of large herbivores. *Annu. Rev. Ecol. Syst.* 31:367–93.

Gaillard, J. M., Festa-Bianchet, M., Delorme, D. and Jorgenson, J. **2000**b. Body mass and individual fitness in female ungulates: bigger is not always better. *Proc. R. Soc. B Biol. Sci.* 267:471–477.

Gaillard, J. M., Pontier, D., Brandt, S., Jullien, J. M. and Allainé, D. **1992**. Sex differentiation in postnatal growth rate: a test in a wild boar population. *Oecologia* 90:167–171.

Gaillard, J. M., Pontier, D., Lebreton, J. D., Trouvilliez, J. and Clobert, J. **1989**. An analysis of demographic tactics in birds and mammals. *Oikos* 56:59–76.

Gamelon, M., Besnard, A., Gaillard, J. M., Servanty, S., Baubet, E., Brandt, S. and Gimenez, O. **2011**. High hunting pressure selects for earlier birth date : wild boar as a case study. *Evolution (N. Y).* 65:3100–3112.

Gamelon, M., Douhard, M., Baubet, E., Gimenez, O., Brandt, S. and Gaillard, J.-M. **2013**a. Fluctuating food resources influence developmental plasticity in wild boar. *Biol. Lett.* 9:20130419.

Gamelon, M., Focardi, S., Baubet, E., Brandt, S., Ronchi, F., Venner, S., Sæther, B. and Gaillard, M. **2017**. Reproductive allocation in pulsed-resource environments : a

comparative study in two populations of wild boar. *Oecologia* 183:1065–1076.

Gamelon, M., Focardi, S., Gaillard, J., Gimenez, O., Bonenfant, C., Franzetti, B., Choquet, R., Ronchi, F., Baubet, E. and Lemaître, J.-F. **2014**. Do age-specific survival patterns of wild boar fit current evolutionary theories of senescence ? *Evolution (N. Y).* 68:3636–3643.

Gamelon, M., Gaillard, J.-M., Baubet, E., Devillard, S., Say, L., Brandt, S. and Gimenez, O. **2013**b. The relationship between phenotypic variation among offspring and mother body mass in wild boar: evidence of coin-flipping? *J. Anim. Ecol.* 82:937–945.

Gamelon, M., Gaillard, J.-M., Servanty, S., Gimenez, O., Toïgo, C., Baubet, E., Klein, F. and Lebreton, J.-D. **2012**. Making use of harvest information to examine alternative management scenarios: a body weight-structured model for wild boar. *J. Appl. Ecol.* 49:833–841.

Garrott, R. A., White, P. J. and Vanderbilt White, C. A. **1993**. Overabundance: an issue for conservation biologists?

Garshelis, D. L., Johnson, A. M. and Garshelis, J. A. **1984**. Social organization of sea otters in Prince William Sound, Alaska. *Can. J. Zool.* 62:2648–2658.

Gayet, T., Devillard, S., Gamelon, M., Brandt, S., Say, L. and Baubet, E. **2016**. On the evolutionary consequences of increasing litter size with multiple paternity in wild boar (*Sus scrofa scrofa*). *Evolution (N. Y).* 70:1386–1397.

Geisser, H. and Reyer, H.-U. **2004**. Efficacy of hunting, feeding, and fencing to reduce crop damage by wild boars. *J. Wildl. Manage.* 68:939–946.

Geisser, H. and Reyer, H.-U. **2005**. The influence of food and temperature on population density of wild boar *Sus scrofa* in the Thurgau (Switzerland). *J. Zool.* 267:89–96.

Gepts, P. and Papa, R. **2002**. Evolution during domestication. P. *in Encyclopedia of Life Sciences*. John Wiley & Sons Ltd, Chichester.

Gilpin, M. E. and Soulé, M. **1986**. Minimum viable population: processes of species extinction. Pp. 19–34 *in Conservation Biology: The Science of Scarcity and Diversity*. Sinauer Associates, Sunderland, Massachusett.

Ginsberg, J. R. and Milner-Gulland, E. J. **1994**. Sex-biased harvesting and population dynamics in ungulates: implications for conservation and sustainable use. *Conserv. Biol.* 8:157–166.

Gortázar, C., Acevedo, P., Ruiz-Fons, F. and Vicente, J. **2006**. Disease risks and overabundance of game species. *Eur. J. Wildl. Res.* 52:81–87.

Gosling, L. M. **1991**. The alternative mating strategies of male topi, *Damaliscus lunatus*. *Appl. Anim. Behav. Sci.* 29:107–119.

Goulding, M. J. and Roper, T. J. **2002**. Press responses to the presence of free-living Wild Boar (*Sus scrofa*) in southern England. *Mamm. Rev.* 32:272–282.

Gowaty, P. A. and Hubbell, S. P. **2009**. Reproductive decisions under ecological constraints: It's about time. *Proc. Natl. Acad. Sci.* 106:10017–10024.

Graves, H. B. **1984**. Behavior and ecology of wild and feral swine (*Sus Scrofa*). *J. Anim. Sci.* 58:482–492.

Griffin, A. S., Alonzo, S. H. and Cornwallis, C. K. **2013**. Why do cuckolded males provide paternal care? *PLoS Biol.* 11:e1001520.

Griffith, S. C., Owens, I. P. F. and Thuman, K. A. **2002**. Extra pair paternity in birds: A review

of interspecific variation and adaptive function. *Mol. Ecol.* 11:2195–2212.

Groenen, M. a. M., Joosten, R., Bosher, M. Y., Amiges, Y., Rattink, A. P., Harlizius, B., Poel, J. J. van der and Crooijmans, R. P. M. A. **2003**. The use of microsatellite genotyping for population studies in the pig with individual and pooled DNA samples. *Arch. Zootec.* 52:145–155.

Grubb, P. and Jewell, P. A. **1973**. The rut and the occurence of oestrus in the Soay sheep on St Kilda. *J. Reprod. Fertil.* 19:491–502. cited in Coltman et al. 1999. Male reproductive success in a promiscuous mammal: behavioural estimates compared with genetic paternity. *Mol. Ecol.* 8:1199–1209.

Hadfield, J. D. **2010**. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *J. Stat. Softw.* 33.

Haldane, J. B. S. **1955**. The measurement of variation. *Evolution (N. Y).* 9:484–484.

Hamilton, W. D. **1964**. The genetical evolution of social behaviour. *J. Theor. Biol.* 7:17–52.

Hampton, J., Pluske, J. R. and Spencer, P. B. S. **2004**. A preliminary genetic study of the social biology of feral pigs in south-western Australia and the implications for management. *Wildl. Res.* 31:375–381.

Hard, J. J., Mills, L. S. and Peek, J. M. **2006**. Genetic implications of reduced survival of male red deer *Cervus elaphus* under harvest. *Wildlife Biol.* 12:427–441.

Hardy, O. J. and Vekemans, X. **2015**. SPAGeDi 1.5: a program for Spatial Pattern Analysis of Genetic Diversity.

Harris, R. B., Wall, W. A. and Allendorf, F. W. **2002**. Genetic consequences of hunting: what do we know and what should we do? *Wildl. Soc. Bull.* 30:634–643.

Harrison, H. B., Saenz-Agudelo, P., Planes, S., Jones, G. P. and Berumen, M. L. **2013**. Relative accuracy of three common methods of parentage analysis in natural populations. *Mol. Ecol.* 22:1158–1170.

Heckel, G., Voigt, C. C., Mayer, F. and Von Helversen, O. **1999**. Extra-harem paternity in the white-lined bat *Saccopteryx bilineata* (Emballonuridae). *Behaviour* 136:1173–1185.

Henry, V. G. **1968**. Fetal Development in European Wild Hogs. *J. Wildl. Manage.* 32:966.

Herrero, J., García-Serrano, A. and García-González, R. **2008**. Reproductive and demographic parameters in two Iberian wild boar *Sus scrofa* populations. *Acta Theriol. (Warsz).* 53:355–364.

Hogg, J. T. and Forbes, S. H. **1997**. Mating in bighorn sheep: Frequent male reproduction via a high-risk "unconventional" tactic. *Behav. Ecol. Sociobiol.* 41:33–48.

Holman, L. **2016**. Bet hedging via multiple mating: A meta-analysis. *Evolution (N. Y).* 70:62–71.

Holman, L. and Kokko, H. **2013**. The consequences of polyandry for population viability, extinction risk and conservation. *Philos. Trans. R. Soc. B Biol. Sci.* 368:20120053.

Hoogland, J. **1998**. Why do female Gunnison's prairie dogs copulate with more than one male? *Anim. Behav.* 55:351–9.

Hrdy, S. B. **1979**. Infanticide among animals: a review, classification, and examination of the implications for the reproductive strategies of females. *Ethol. Sociobiol.* 1:13–40.

Hubbell, S. and Johnson, L. **1987**. Environmental variance in lifetime mating success, mate

choice, and sexual selection. *Am. Nat.* 130:91–112.

Hubby, J. L. and Lewontin, R. C. **1966**. A molecular approach to the study of genic heterozygosity in natural populations. I. The number of alleles at different loci in *Drosophila pseudoobscura*. *Genetics* 54:577–594.

Hudson, R. and Trillmich, F. **2007**. Sibling competition and cooperation in mammals: challenges, developments and prospects. *Behav. Ecol. Sociobiol.* 62:299–307.

Hulme, P. E., Bacher, S., Kenis, M., Klotz, S., Kühn, I., Minchin, D., Nentwig, W., Olenin, S., Panov, V., Pergl, J., Pyšek, P., Roques, A., Sol, D., Solarz, W. and Vilà, M. **2008**. Grasping at the routes of biological invasions: A framework for integrating pathways into policy. *J. Appl. Ecol.* 45:403–414.

Hutchings, J. A. and Myers, R. A. **1995**. The biological collapse of Atlantic cod off Newfoundland and Labrador: an exploration of historical changes in exploitation, harvesting technology, and management. Pp. 37–94 *in The North Atlantic fisheries: successes, failures, and challenges*. Island Institute Studies, Charlottetown, Prince Edward Island. Canada.

ISAG/FAO Standing Committee. **2004**. *Secondary Guidelines for Development of National Farm Animal Genetic Resources Management Plans. Measurement of Domestic Animal Diversity (MoDAD): Recommended Microsatellite Markers*.

Isvaran, K. **2005**. Variation in male mating behaviour within ungulate populations: Patterns and processes. *Curr. Sci.* 89:1192–1199.

Jackson, J. B., Kirby, M. X., Berger, W. H., Bjorndal, K. A., Botsford, L. W., Bourque, B. J., Bradbury, R. H., Cooke, R., Erlandson, J., Estes, J. A., Hughes, T. P., Kidwell, S., Lange, C. B., Lenihan, H. S., Pandolfi, J. M., Peterson, C. H., Steneck, R. S., Tegner, M. J. and Warner, R. R. **2001**. Historical overfishing and the recent collapse of coastal ecosystems. *Science* 293:629–637.

Jarman, P. **1983**. Mating system and sexual dimorphism in large, terrestrial, mammalian herbivores. *Biol. Rev.* 58:485–520.

Jedrzejewski, W., Jedrzewska, B., Okarma, H. and Ruprecht, A. L. **1992**. Wolf predation and snow cover as mortality factors in the ungulate community of the Białowieża National Park, Poland. *Oecologia* 90:27–36.

Jeffreys, A. J., Wilson, V. and Thein, S. L. **1985**. Hypervariable minisatelite regions in human DNA. *Nature* 314:67–73.

Jennions, M. D. and Petrie, M. **2000**. Why do females mate multiply? A review of the genetic benefits. *Biol. Rev. Camb. Philos. Soc.* 75:21–64.

Jensen, P. and Recén, B. **1989**. When to wean - Observations from free-ranging domestic pigs. *Appl. Anim. Behav. Sci.* 23:49–60.

Jeschke, J. M. and Strayer, D. L. **2006**. Determinants of vertebrate invasion success in Europe and North America. *Glob. Chang. Biol.* 12:1608–1619.

Jezierski, W. **1977**. Longevity and mortality rate in a population of wild boar. *Acta Theriol. (Warsz).* 22:337–348.

Jones, A. G. **2005**. GERUD 2.0: a computer program for the reconstruction of parental genotypes from half-sib progeny arrays with known or unknown parents. *Mol. Ecol. Notes* 5:708–711.

Jones, A. G. **2009**. On the opportunity for sexual selection, the bateman gradient and the maximum intensity of sexual selection. *Evolution (N. Y).* 63:1673–1684.

Jones, A. G. and Ardren, W. R. **2003**. Methods of parentage analysis in natural populations. *Mol. Ecol.* 12:2511–2523.

Jones, A. G., Small, C. M., Paczolt, K. A. and Ratterman, N. L. **2010**. A practical guide to methods of parentage analysis. *Mol. Ecol. Resour.* 10:6–30.

Jones, O. R. and Wang, J. **2010**. COLONY: A program for parentage and sibship inference from multilocus genotype data. *Mol. Ecol. Resour.* 10:551–555.

Kalinowski, S. T., Taper, M. L. and Marshall, T. C. **2007**. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol. Ecol.* 16:1099–106.

Kaminski, G., Brandt, S., Baubet, E. and Baudoin, C. **2005**. Life-history patterns in female wild boars (*Sus scrofa*): mother–daughter postweaning associations. *Can. J. Zool.* 83:474–480.

Kamler, J. F., Ballard, W. B., Lemons, P. R. and Mote, K. **2004**. Variation in mating system and group structure in two populations of swift foxes, *Vulpes velox*. *Anim. Behav.* 68:83–88.

Kaplan, R. H. and Cooper, W. S. **1984**. The evolution of developmental plasticity in reproductive characteristics: an application of the "adaptive coin-flipping" principle. *Am. Nat.* 123:393–410.

Kemp, B. and Soede, N. M. **1996**. Relationship of weaning-to-estrus interval to timing of ovulation and fertilization in sows. *J. Anim. Sci.* 74:944–949.

Keuling, O., Baubet, E., Duscher, A., Ebert, C., Fischer, C., Monaco, A., Podgórski, T., Prevot, C., Ronnenberg, K., Sodeikat, G., Stier, N. and Thurfjell, H. **2013**. Mortality rates of wild boar *Sus scrofa* L. in central Europe. *Eur. J. Wildl. Res.* 59:805–814.

Keuling, O., Lauterbach, K., Stier, N. and Roth, M. **2010**. Hunter feedback of individually marked wild boar *Sus scrofa* L.: Dispersal and efficiency of hunting in northeastern Germany. *Eur. J. Wildl. Res.* 56:159–167.

Keuling, O., Podgórski, T., Monaco, A., Melletti, M., Merta, D., Albrycht, M., Genov, P. V, Gethöffer, F., Vetter, S. G., Jori, F., Scalera, R. and Gongora, J. **2017**. Eurasian wild boar *Sus scrofa* (Linnaeus, 1758). Pp. 202–233 *in* Press, C. U., ed. *Ecology, Conservation and Management of Wild Pigs and Peccaries*.

Keuling, O., Stier, N. and Roth, M. **2008**a. Annual and seasonal space use of different age classes of female wild boar *Sus scrofa* L. *Eur. J. Wildl. Res.* 54:403–412.

Keuling, O., Stier, N. and Roth, M. **2008**b. How does hunting influence activity and spatial usage in wild boar *Sus scrofa* L.? *Eur. J. Wildl. Res.* 54:729–737.

Keuling, O., Strauß, E. and Siebert, U. **2016**. Regulating wild boar populations is "somebody else's problem"! - Human dimension in wild boar management. *Sci. Total Environ.* 554–555:311–319. Elsevier B.V.

Kierdorf, U., Konjević, D., Janicki, Z., Slavica, A., Keros, T. and Čurlik, J. **2004**. Tusk abnormalities in wild boar (*Sus scrofa* L .). *Eur. J. Wildl. Res.* 50:48–52.

Kingsolver, J. G. and Diamond, S. E. **2011**. Phenotypic selection in natural populations: what limits directional selection? *Am. Nat.* 177:346–357.

Klemme, I. and Ylönen, H. **2010**. Polyandry enhances offspring survival in an infanticidal

species. *Biol. Lett.* 6:24–26.

Kokko, H. and Mappes, J. **2013**. Multiple mating by females is a natural outcome of a null model of mate encounters. *Entomol. Exp. Appl.* 146:26–37.

Kokko, H. and Morrell, L. J. **2005**. Mate guarding, male attractiveness, and paternity under social monogamy. *Behav. Ecol.* 16:724–731.

Kokko, H. and Rankin, D. J. **2006**. Lonely hearts or sex in the city? Density-dependent effects in mating systems. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 361:319–334.

Koons, D. N., Rockwell, R. F. and Aubry, L. M. **2014**. Effects of exploitation on an overabundant species: The lesser snow goose predicament. *J. Anim. Ecol.* 83:365–374.

Kozdrowski, R. and Dubiel, A. **2004**. The effect of season on the properties of wild boar (*Sus scrofa* L.) semen. *Anim. Reprod. Sci.* 80:281–289.

Kühl, A., Mysterud, A., Erdnenov, G. I., Lushchekina, A. A., Grachev, I. A., Bekenov, A. B. and Milner-Gulland, E. J. **2007**. The 'big spenders' of the steppe: sex-specific maternal allocation and twinning in the saiga antelope. *Proc. R. Soc. London B Biol. Sci.* 274:1293–1299.

Lack, D. **1968**. *Ecological adaptations for breeding in birds*. Methuen, London.

Langvatn, R. and Loison, A. **1999**. Consequences of harvesting on age structure, sex ratio and population dynamics of red deer *Cervus elaphus* in central Norway.

Lank, D. B., Smith, C. M., Hanotte, O., Ohtonen, A., Bailey, S. and Burke, T. **2002**. High frequency of polyandry in a lek mating system. *Behav. Ecol.* 13:209–215.

Leboeuf, B., Restall, B. and Salamon, S. **2000**. Production and storage of goat semen for artificial insemination. *Anim. Reprod. Sci.* 62:113–141.

Lefcheck, J. S. **2016**. piecewiseSEM: Piecewise structural equation modelling in r for ecology, evolution, and systematics. *Methods Ecol. Evol.* 7:573–579.

Lemaître, J. F. and Gaillard, J. M. **2013**. Polyandry has no detectable mortality cost in female mammals. *PLoS One* 8:1–8.

Liebhold, A., Sork, V., Peltonen, M., Koenig, W., Bjørnstad, O. N., Westfall, R., Elkinton, J. and Knops, J. M. H. **2004**. Within-population spatial synchrony in mast seeding of North American oaks. *Oikos* 104:156–164.

Linkie, M., Dinata, Y., Nofrianto, A. and Leader-Williams, N. **2007**. Patterns and perceptions of wildlife crop raiding in and around Kerinci Seblat National Park, Sumatra. *Anim. Conserv.* 10:127–135.

Lloyd, D. G. **1987**. Selection of offspring size at independence and other size-versus-number strategies. *Am. Nat.* 129:800–817.

Lotterhos, K. E. **2011**. The context-dependent effect of multiple paternity on effective population size. *Evolution (N. Y).* 65:1693–706.

Loveridge, A. J., Searle, A. W., Murindagomo, F. and Macdonald, D. W. **2007**. The impact of sport-hunting on the population dynamics of an African lion population in a protected area. *Biol. Conserv.* 134:548–558.

Maher, C. R. and Burger, J. R. **2011**. Intraspecific variation in space use, group size, and mating systems of caviomorph rodents. *J. Mammal.* 92:54–64.

Maillard, D. and Fournier, P. **1995**. Effects of shooting with hounds on size of resting range of

wild boar (*Sus scrofa* L.) groups in mediterranean habitat. *Ibex J. Mt. Ecol.* 3:102–107.

Maillard, D., Gaillard, J.-M., Hewison, M., Ballon, P., Duncan, P., Loison, A., Baubet, C. T. E., Bonenfant, C., Garel, M. and Saint-Andrieux, C. **2010**. Ungulates and their management in France. Pp. 441–474 *in European ungulates and their management in the 21st century*.

Malo, A. F., Roldan, E. R. S., Garde, J., Soler, A. J. and Gomendio, M. **2005**. Antlers honestly advertise sperm production and quality. *Proc. R. Soc. B Biol. Sci.* 272:149–157.

Martin, J. G. A., Petelle, M. B. and Blumstein, D. T. **2014**. Environmental, social, morphological, and behavioral constraints on opportunistic multiple paternity. *Behav. Ecol. Sociobiol.* 68:1531–1538.

Martin, J. K. and Martin, A. A. **2007**. Resource distribution influences mating system in the bobuck (*Trichosurus cunninghami*: Marsupialia). *Oecologia* 154:227–236.

Massei, G. and Genov, P. V. **2004**. The environmental impact of wild boar. *Galemys* 16:135–145.

Massei, G., Genov, P. V and Staines, B. W. **1996**. Diet, food availability and reproduction of wild boar in a Mediterranean coastal area. *Acta Theriol. (Warsz).* 41:307–320.

Massei, G., Genov, P. V, Staines, B. W. and Gorman, M. L. **1997**. Factors influencing home range and activity of wild boar (*Sus scrofa*) in a Mediterranean coastal area. *J. Zool.* 242:411–423.

Massei, G., Kindberg, J., Licoppe, A., Gačić, D., Šprem, N., Kamler, J., Baubet, E., Hohmann, U., Monaco, A., Ozoliņš, J., Cellina, S., Podgórski, T., Fonseca, C., Markov, N., Pokorny, B., Rosell, C. and Náhlik, A. **2015**. Wild boar populations up, numbers of hunters down? A review of trends and implications for Europe. *Pest Manag. Sci.* 12:492–500.

Mauget, R. **1980**. Régulation écologiques, comportementales et physiologiques (fonction de reproduction) de l'adaptation du sanglier, *Sus scrofa* L., au milieu. Ph D Thesis, Université François Rabelais, Tours, France.

Mauget, R. and Boissin, J. **1987**. Seasonal changes in testis weight and testosterone concentration in the european wild boar (*Sus scrofa* L.). *Anim. Reprod. Sci.* 13:67–74.

McEachern, M. B., McElreath, R. L., Van Vuren, D. H. and Eadie, J. M. **2009**. Another genetically promiscuous "polygynous" mammal: mating system variation in <i>Neotoma fuscipes<i/>. *Anim. Behav.* 77:449–455. The Association for the Study of Animal Behaviour.

McElligott, A. G., Gammell, M. P., Harty, H. C., Paini, D. R., Murphy, D. T., Walsh, J. T. and Hayden, T. J. **2001**. Sexual size dimorphism in fallow deer (*Dama dama*): Do larger, heavier males gain greater mating success? *Behav. Ecol. Sociobiol.* 49:266–272.

McKinney, M. L. and Lockwood, J. L. **1999**. Biotic homogenization: a few winners replacing many loosers in the next mass extinction. *Trends Ecol. Evol.* 14:450–453.

Milner-Gulland, E. J., Bukreeva, O. M., Coulson, T., Lushchekina, A. A., Kholodova, M. V, Bekenov, A. and Grachev, I. A. **2003**. Reproductive collapse in saiga antelope harems. *Nature* 422:135–135.

Milner, J. M., Nilsen, E. B. and Andreassen, H. P. **2007**. Demographic side effects of selective hunting in ungulates and carnivores. *Conserv. Biol.* 21:36–47.

Mock, D. W. and Parker, G. A. **1997**. *The Evolution of Sibling Rivalry*. Oxford University Press.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. and Erlich, H. **1986**. Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction.

Mysterud, A. **2000**. The relationship between ecological segregation and sexual body size dimorphism in large herbivores. *Oecologia* 124:40–54.

Náhlik, A. and Sandor, G. **2003**. Birth rate and offspring survival in a free-ranging wild boar *Sus scrofa* population. *Wildlife Biol.* 9:37–42.

Neff, B. D. and Pitcher, T. E. **2005**. Genetic quality and sexual selection: an integrated framework for good genes and compatible genes. *Mol. Ecol.* 14:19–38.

Nelson, M. P., Bruskotter, J. T., Vucetich, J. A. and Chapron, G. **2016**. Emotions and the ethics of consequence in conservation decisions: lessons from Cecil the lion. *Conserv. Lett.* 9:302–306.

Nissen, A. K., Soede, N. M., Hyttel, P., Schmidt, M. and D'Hoore, L. **1997**. The influence of time of insemination relative to time of ovulation on farrowing frequency and litter size in sows, as investigated by ultrasonography. *Theriogenology* 47:1571–1582.

Nunney, L. **1993**. The influence of mating system and overlapping generations on effective population size. *Evolution (N. Y).* 47:1329–1341.

O'Brien, T. G., Kinnaird, M. F. and Wibisono, H. T. **2003**. Crouching tigers, hidden prey: Sumatran tiger and prey populations in a tropical forest landscape. *Anim. Conserv.* 6:131–139.

Okarma, H., Jedrzejewska, B., Jedrzejewski, W., Krasinski, Z. A. and Milkowski, L. **1995**. The roles of predation, snow cover, acorn crop, and man- related factors on ungulate mortality in Białowieża Primeval Forest, Poland. *Acta Theriol. (Warsz).* 40:197–217.

Oliver, W. and Leus, K. **2008**. *Sus scrofa*. *IUCN Red List Threat. Species 2008* e.T41775A1:http://dx.doi.org/10.2305/IUCN.UK.2008.RLTS.T41775.

Ophir, A. G., Phelps, S. M., Sorin, A. B. and Wolff, J. O. **2008**. Social but not genetic monogamy is associated with greater breeding success in prairie voles. *Anim. Behav.* 75:1143–1154.

Orr, D. E. and Shen, Y. **2006**. World pig production, opportunity or threat? Pp. 3–8 *in Proceedings of the 2006 Midwest Swine Nutrition Conference, Indianapolis, Indiana, USA.*

Parker, G. A. and Birkhead, T. R. **2013**. Polyandry: the history of a revolution. *Philos. Trans. R. Soc. B Biol. Sci.* 368:20120335.

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

Peakall, R. and Smouse, P. E. **2012**. GenALEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* 28:2537–2539.

Pearse, D. E. and Anderson, E. C. **2009**. Multiple paternity increases effective population size. *Mol. Ecol.* 18:3124–3127.

Pérez-González, J. and Carranza, J. **2011**. Female aggregation interacts with population structure to influence the degree of polygyny in red deer. *Anim. Behav.* 82:957–970.

Pérez-González, J., Costa, V., Santos, P., Carranza, J., Zsolnai, A., Fernández-Llario, P., Monteiro, N. M., Anton, I. and Beja-Pereira, A. **2017**. Heterozygosity decrease in wild boar mating system - a case of outbreeding avoidance? *J. Zool.* 302:40–48.

Pérez-González, J., Costa, V., Santos, P., Slate, J., Carranza, J., Fernández-Llario, P., Zsolnai, A., Monteiro, N. M., Anton, I., Buzgó, J., Varga, G. and Beja-Pereira, A. **2014**. Males and females contribute unequally to offspring genetic diversity in the polygynandrous mating system of wild boar. *PLoS One* 9:1–22.

Philippi, T. and Seger, J. **1989**. Hedging one's evolutionary bets, revisited. *Trends Ecol. Evol.* 4:41–44.

Phipson, B. and Smyth, G. K. **2010**. Permutation P-values should never be zero: calculating exact P-values when permutations are randomly drawn. *Stat. Appl. Genet. Mol. Biol.* 9:1–16.

Pimentel, D., Zuniga, R. and Morrison, D. **2005**. Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecol. Econ.* 52:273–288.

Podgórski, T., Scandura, M. and Jędrzejewska, B. **2014**. Next of kin next door - philopatry and socio-genetic population structure in wild boar. *J. Zool.* 294:190–197.

Poteaux, C., Baubet, E., Kaminski, G., Brandt, S., Dobson, F. S. and Baudoin, C. **2009**. Socio-genetic structure and mating system of a wild boar population. *J. Zool.* 278:116–125.

Powell, D. M. **2004**. Pigs (Suidae). Pp. 275–290 *in* Kleiman, D.G., Geist, V. & McDade, M., ed. *Grzimek's Animal Life Encyclopedia, second edition, volume 15, Mammals IV*.

Preston, B. T., Stevenson, I. R., Pemberton, J. M., Coltman, D. W. and Wilson, K. **2003**. Overt and covert competition in a promiscuous mammal: the importance of weaponry and testes size to male reproductive success. *Proc. R. Soc. B Biol. Sci.* 270:633–640.

Preston, B. T., Stevenson, I. R., Pemberton, J. M. and Wilson, K. **2001**. Dominant rams lose out by sperm depletion. *Nature* 409:681–682.

Prévot, C. and Licoppe, A. **2013**. Comparing red deer (*Cervus elaphus* L .) and wild boar (*Sus scrofa* L .) dispersal patterns in southern Belgium. *Eur. J. Wildl. Res.* 59:795–803.

Price, S. A., Bininda-emonds, O. R. P. and Gittleman, J. L. **2005**. A complete phylogeny of the whales, dolphins and even-toed hoofed mammals (Cetartiodactyla). *Biol. Rev.* 80:445–473.

Proaktor, G., Coulson, T. and Milner-Gulland, E. J. **2007**. Evolutionary responses to harvesting in ungulates. *J. Anim. Ecol.* 76:669–678.

Queller, D. C. and Goodnight, K. F. **1989**. Estimating relatedness using genetic markers. *Evolution (N. Y).* 43:258–275.

Quirós-Fernández, F., Marcos, J., Acevedo, P. and Gortázar, C. **2017**. Hunters serving the ecosystem: the contribution of recreational hunting to wild boar population control. *Eur. J. Wildl. Res.* 63:4–9. European Journal of Wildlife Research.

R Core Team. **2017**. R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria.

Ralls, K. **1977**. Sexual dimorphism in mammals: avian models and unanswered questions. *Am. Nat.* 111:917–938.

Rambuda, T. D. and Johnson, S. D. **2004**. Breeding systems of invasive alien plants in South Africa: Does Baker's rule apply? *Divers. Distrib.* 10:409–416.

Rankin, D. J. and Kokko, H. **2007**. Do males matter? The role of males in population dynamics. *Oikos* 116:335–348.

Réale, D., Boussès, P. and Chapuis, J.-L. **1996**. Female-biased mortality induced by male sexual harassment in a feral sheep population. *Can. J. Zool.* 74:1812–1818.

Ronget, V., Gaillard, J. M., Coulson, T., Garratt, M., Gueyffier, F., Lega, J. C. and Lemaître, J. F. **2018**. Causes and consequences of variation in offspring body mass: meta-analyses in birds and mammals. *Biol. Rev.* 93:1–27.

Rosecchi, E., Thomas, F. and Crivelli, A. J. **2001**. Can life-history traits predict the fate of introduced species? A case study on two cyprinid fish in southern France. *Freshw. Biol.* 46:845–853.

Rosser, A. M. and Mainka, S. A. **2002**. Overexploitation and species extinctions. *Conserv. Biol.* 16:584–586.

Rossi, S., Fromont, E., Pontier, D., Crucière, C., Hars, J., Barrat, J., Pacholek, X. and Artois, M. **2005**. Incidence and persistence of classical swine fever in free-ranging wild boar (*Sus scrofa*). *Epidemiol. Infect.* 133:559–568.

Rosvold, J. and Andersen, R. **2008**. Wild boar in Norway-is climate a limiting factor? *Nat. Hist.* 28.

Ruiz-Fons, F., Segalés, J. and Gortázar, C. **2008**. A review of viral diseases of the European wild boar: Effects of population dynamics and reservoir rôle. *Vet. J.* 176:158–169.

Rus Hoelzel, A., Le Boeuf, B. J., Reiter, J. and Campagna, C. **1999**. Alpha-male paternity in elephant seals. *Behav. Ecol. Sociobiol.* 46:298–306.

Russell, F. L., Zippin, D. B. and Fowler, N. L. **2001**. Effects of white-tailed deer (*Odocoileus virginianus*) on plants, plant populations and communities: a review. *Am. Midl. Nat.* 146:1–26.

Russo, L., Massei, G. and Genov, P. V. **1997**. Daily home range and activity of wild boar in a mediterranean area free from hunting. *Ethol. Ecol. Evol.* 9:287–294.

Rutberg, A. T., Naugle, R. E., Thiele, L. A. and Liu, I. K. M. **2004**. Effects of immunocontraception on a suburban population of white-tailed deer Odocoileus virginianus. *Biol. Conserv.* 116:243–250.

Sáenz-de-santa-maría, A. and Tellería, J. L. **2015**. Wildlife-vehicle collisions in Spain. *Eur. J. Wildl. Res.* 61:399–406.

Sæther, B.-E. and Engen, S. **2015**. The concept of fitness in fluctuating environments. *Trends Ecol. Evol.* 30:273–281.

Sæther, B., Solberg, E. J., Heim, M., Stacy, J. E., Jakobsen, K. S. and Olstad, R. **2004**. Offspring sex ratio in moose *Alces alces* in relation to paternal age: an experiment. *Wildlife Biol.* 10:51–57.

Sáez-Royuela, C. and Tellería, J. L. **1986**. The increased population of the wild boar (*Sus scrofa* L .) in Europe. *Mamm. Rev.* 16:97–101.

Sales, J. and Kotrba, R. **2013**. Meat from wild boar (*Sus scrofa* L.): A review. *Meat Sci.* 94:187–201. Elsevier Ltd.

Say, L., Devillard, S., Natoli, E. and Pontier, D. **2002**. The mating system of feral cats (*Felis catus L.*) in a sub-Antarctic environment. *Polar Biol.* 25:838–842.

Say, L., Naulty, F. and Hayden, T. J. **2003**. Genetic and behavioural estimates of reproductive skew in male fallow deer. *Mol. Ecol.* 12:2793–2800.

Say, L., Pontier, D. and Natoli, E. **1999**. High variation in multiple paternity of domestic cats (*Felis catus* L.) in relation to environmental conditions. *Proc. Biol. Sci.* 266:2071–4.

Schilling, E. and Vengust, M. **1987**. Frequency of semen collection in boars and quality of ejaculates as evaluated by the osmotic resistance of acrosomal membranes.

Schley, L., Dufrêne, M., Krier, A. and Frantz, A. C. **2008**. Patterns of crop damage by wild boar (*Sus scrofa*) in Luxembourg over a 10-year period. *Eur. J. Wildl. Res.* 54:589–599.

Schley, L. and Roper, T. J. **2003**. Diet of wild boar *Sus scrofa* in Western Europe, with particular reference to consumption of agricultural crops. *Mamm. Rev.* 33:43–56.

Schoen, D. J. and Brown, A. H. **1991**. Intraspecific variation in population gene diversity and effective population size correlates with the mating system in plants. *Proc. Natl. Acad. Sci.* 88:4494–4497.

Scillitani, L., Monaco, A. and Toso, S. **2010**. Do intensive drive hunts affect wild boar (*Sus scrofa*) spatial behaviour in Italy? Some evidences and management implications. *Eur. J. Wildl. Res.* 56:307–318.

Servanty, S. **2007**. Dynamique d'une population chassée de sangliers (Sus scrofa scrofa) en milieu forestier. PhD Thesis, Université Claude Bernard Lyon1, Lyon, France.

Servanty, S., Gaillard, J.-M., Allaine, D., Brandt, S. and Baubet, E. **2007**. Litter size and fetal sex ratio adjustment in a highly polytocous species: the wild boar. *Behav. Ecol.* 18:427–432.

Servanty, S., Gaillard, J.-M., Toïgo, C., Brandt, S. and Baubet, E. **2009**. Pulsed resources and climate-induced variation in the reproductive traits of wild boar under high hunting pressure. *J. Anim. Ecol.* 78:1278–1290.

Servanty, S., Gaillard, J. M., Ronchi, F., Focardi, S., Baubet, É. and Gimenez, O. **2011**. Influence of harvesting pressure on demographic tactics: implications for wildlife management. *J. Appl. Ecol.* 48:835–843.

Shine, R. **2012**. Invasive species as drivers of evolutionary change: Cane toads in tropical Australia. *Evol. Appl.* 5:107–116.

Shipley, B. **2009**. Confirmatory path analysis in a generalized multilevel context. *Ecology* 90:363–368.

Shipley, B. **2013**. The AIC model selection method applied to path analytic models compared using a d-separation test. *Ecology* 94:560–564.

Shuster, S. M. and Wade, M. J. **2003**. Mating systems and strategies. P. *in Mating systems and strategies*.

Simard, M. A., Dussault, C., Huot, J. and Côté, S. D. **2013**. Is hunting an effective tool to control overabundant deer? A test using an experimental approach. *J. Wildl. Manage.* 77:254–269.

Simmons, L. W. and Fitzpatrick, J. L. **2012**. Sperm wars and the evolution of male fertility. *Reproduction* 144:519–534.

Singer, F. J. and Zeigenfuss, L. C. **2002**. Influence of trophy hunting and horn size on mating behavior and survivorship of mountain sheep. *J. Mammal.* 83:682–698.

Smith, C. C. and Fretwell, S. D. **1974**. The optimal balance between size and number of offspring. *Am. Nat.* 108:499–506.

Soede, N. M., Wetzels, C. C., Zondag, W., de Koning, M. A. and Kemp, B. **1995**. Effects of time of insemination relative to ovulation, as determined by ultrasonography, on fertilization rate and accessory sperm count in sows. *J. Reprod. Fertil.* 104:99–106.

Solberg, E. J., Loison, A., Sether, B. E. and Strand, O. **2000**. Age-specific harvest mortality in a Norwegian moose *Alces alces* population. *Wildlife Biol.* 6:41–52.

Sorin, A. **2004**. Paternity assignment for white-tailed deer (*Odocoileus virginianus*): mating across age classes and multiple paternity. *J. Mammal.* 85:356–362.

Soulsbury, C. D. **2010**. Genetic patterns of paternity and testes size in mammals. *PLoS One* 5:e9581.

Spencer, P. B. S., Lapidge, S. J., Hampton, J. O. and Pluske, J. R. **2005**. The sociogenetic structure of a controlled feral pig population. *Wildl. Res.* 32:297–304.

Spielman, D., Brook, B. W. and Frankham, R. **2004**. Most species are not driven to extinction before genetic factors impact them. *Proc. Natl. Acad. Sci. U. S. A.* 101:15261–15264.

Spitz, F., Valet, G. and Lehr Brisbin, I. **1998**. Variation in body mass of wild boars from southern France. *J. Mammal.* 79:251–259.

Starrfelt, J. and Kokko, H. **2012**. Bet-hedging-a triple trade-off between means, variances and correlations. *Biol. Rev. Camb. Philos. Soc.* 87:742–55.

Stillfried, M., Fickel, J., Börner, K., Wittstatt, U., Heddergott, M., Ortmann, S., Kramer-Schadt, S. and Frantz, A. C. **2017**. Do cities represent sources, sinks or isolated islands for urban wild boar population structure? *J. Appl. Ecol.* 54:272–281.

Stockley, P. **2003**. Female multiple mating behaviour, early reproductive failure and litter size variation in mammals. *Proc. R. Soc. London. Ser. B Biol. Sci.* 270:271–278.

Storie, J. T. and Bell, S. **2017**. Wildlife management conflicts in rural communities: a case-study of wild boar (*Sus scrofa*) management in Ērgļu Novads, Latvia. *Sociol. Ruralis* 57:64–86.

Sugg, D. W. and Chesser, R. K. **1994**. Effective population sizes with multiple paternity. *Genetics* 137:1147–1155.

Sun, M. and Ritland, K. **1998**. Mating system of yellow starthistle (*Centaurea solstitialis*), a successful colonizer in North America. *Heredity (Edinb).* 80:225–232.

Sutherland, W. J. **1985**. Chance can produce a sex difference in variance in mating success and explain Bateman's data. *Anim. Behav.* 33:1349–1352.

Sweitzer, R. A., Vuren, D. Van, Gardner, I. A., Boyce, W. M. and Waithman, J. D. **2000**. Estimating sizes of wild pig populations in the north and central coast regions of California. *J. Wildl. Manage.* 64:531–543.

Taylor, M. L., Price, T. A. R. and Wedell, N. **2014**. Polyandry in nature: A global analysis. *Trends Ecol. Evol.* 29:376–383. Elsevier Ltd.

Thonhauser, K. E., Raveh, S., Thoß, M. and Penn, D. J. **2016**. Does multiple paternity influence offspring disease resistance? *J. Evol. Biol.* 29:1142–1150.

Thonhauser, K. E., Thoß, M., Musolf, K., Klaus, T. and Penn, D. J. **2014**. Multiple paternity in wild house mice (*Mus musculus musculus*): Effects on offspring genetic diversity and body mass. *Ecol. Evol.* 4:200–209.

Toïgo, C., Servanty, S., Gaillard, J. M., Brandt, S. and Baubet, E. **2008**. Disentangling natural

from hunting mortality in an intensively hunted wild boar population. *J. Wildl. Manage.* 72:1532–1539.

Tolon, V., Dray, S., Loison, A., Zeileis, A., Fischer, C. and Baubet, E. **2009**. Responding to spatial and temporal variations in predation risk : space use of a game species in a changing landscape of fear. *Can. J. Zool.* 1137:1129–1137.

Trexler, J. C., Travis, J. and Dinep, A. **1997**. Variation among populations of the sailfin molly in the rate of concurrent multiple paternity and its implications for mating-system evolution. *Behav. Ecol. Sociobiol.* 40:297–305.

Trivers, R. L. **1972**. Parental investment and sexual selection. Pp. 136–179 *in In: Sexual Selection and the Descent of Man (ed. Campbell B), London: Heinemann*.

Uller, T. and Olsson, M. **2008**. Multiple paternity in reptiles: patterns and processes. *Mol. Ecol.* 17:2566–2580.

Van De Casteele, T., Galbusera, P. and Matthysen, E. **2001**. A comparison of microsatellite-based pairwise relatedness estimators. *Mol. Ecol.* 10:1539–1549.

Van Herzele, A., Aarts, N. and Casaer, J. **2015**. Wildlife comeback in Flanders: tracing the fault lines and dynamics of public debate. *Eur. J. Wildl. Res.* 61:539–555.

Van Noordwijk, M. A. and van Schaik, C. P. **2000**. Reproductive patterns in eutherian mammals: adaptations against infanticide? Pp. 322–360 *in Infanticide by males and its implications*. Cambridge University Press.

Vercauteren, K. C., Anderson, C. W., Van Deelen, T. R., Drake, D., Walter, W. D., Vantassel, S. M. and Hygnstrom, S. E. **2011**. Regulated commercial harvest to manage overabundant white-tailed deer: An idea to consider? *Wildl. Soc. Bull.* 35:185–194.

Vernesi, C., Crestanello, B., Pecchioli, E., Tartari, D., Caramelli, D., Hauffe, H. and Bertorelle, G. **2003**. The genetic impact of demographic decline and reintroduction in the wild boar (*Sus scrofa*) : A microsatellite analysis. *Mol. Ecol.* 12:585–595.

Vetter, S. G., Ruf, T., Bieber, C. and Arnold, W. **2015**. What is a mild winter? Regional differences in within-species responses to climate change. *PLoS One* 10:1–17.

Vignon, V. and Barbarreau, H. **2008**. Collisions entre véhicules et ongulés sauvages : quel coût économique ? Une tentative d'évaluation. *Faune Sauvag.* 27:31–35.

Wade, M. J. **1979**. Sexual selection and variance in reproductive success. *Am. Nat.* 114:742–747.

Wade, M. J. and Shuster, S. M. **2004**. Sexual selection: harem size and the variance in male reproductive success. *Am. Nat.* 164:E83-9.

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

Wakabayashi, H., Noda, S. and Saitoh, T. **2017**. Intraspecific variation in the frequency of multiple paternity in the Japanese intraspecific variation in the frequency of multiple paternity in the japanese wood mouse (Apodemus speciosus). *Mammal Study* 42:81–88.

Waller, C. M. and Bilkei, G. **2002**. Effect of mating frequency on sow reproductive performance under field conditions and natural mating in large outdoor production units. *Reprod. Domest. Anim.* 37:116–118.

Watson, P. J. **1991**. Multiple paternity as genetic bet-hedging in female sierra dome spiders, *Linyphia litigiosa* (Linyphiidae). *Anim. Behav.* 41:343–360.

Wedell, N., Gage, M. J. G. and Parker, G. A. **2002**. Sperm competition, male prudence and sperm-limited females. *Trends Ecol. Evol.* 17:313–320.

Wells, C. P., Tomalty, K. M., Floyd, C. H., McElreath, M. B., May, B. P. and Van Vuren, D. H. **2017**. Determinants of multiple paternity in a fluctuating population of ground squirrels. *Behav. Ecol. Sociobiol.* 71. Behavioral Ecology and Sociobiology.

Whitman, K., Starfield, A. M., Quadling, H. S. and Packer, C. **1987**. Sustainable trophy hunting of African lions. 6–11.

Wilson, A. J., Coltman, D. W., Pemberton, J. M., Overall, A. D. J., Byrne, K. A. and Kruuk, L. E. B. **2005**. Maternal genetic effects set the potential for evolution in a free-living vertebrate population. *J. Evol. Biol.* 18:405–414.

Winkler, D. W. and Wallin, K. **1987**. Offspring size and number: a life history model linking effort per offspring and total effort. *Am. Nat.* 129:708–720.

Wolff, J. O. and MacDonald, D. W. **2004**. Promiscuous females protect their offspring. *Trends Ecol. Evol.* 19:127–134.

Wu, C., Johnson, N. A. and Palopoli, M. F. **1996**. Haldane's rule and its legacy: why are there so many sterile males? *Trends Ecol. Evol.* 11:281–284.

Yasui, Y. **2001**. Female multiple mating as a genetic bet-hedging strategy when mate choice criteria are unreliable. *Ecol. Res.* 16:605–616.

Yasui, Y. **1998**. The `genetic benefits' of female multiple mating reconsidered. *Trends Ecol. Evol.* 13:246–250.

Zedrosser, A., Bellemain, E., Taberlet, P. and Swenson, J. E. **2007**. Genetic estimates of annual reproductive success in male brown bears: The effects of body size, age, internal relatedness and population density. *J. Anim. Ecol.* 76:368–375.

Zuur, A. F., Ieno, E. N., J.Walker, N., Saveliev, A. A. and Smith, G. M. **2009**. *Statistics for Biology and Health*. Springer New York.

Appendices

Appendix A: SNPs or Microsatellites? Assessing the reliability of different molecular markers to study hybridization between wild boar and domestic pig (*Sus scrofa*).

Abstract: Hybridization between wild and domestic species or subspecies is widespread in vertebrates, but may be difficult to detect, especially when subspecies remain genetically close despite strong anthropic selection in the domestic counterpart. Developing molecular tools to enable efficient identification of hybrids is crucial for a better understanding of both evolutionary and conservation consequences of hybridization in the wild. Here we compared the efficiency of a set of 20 SNPs (Single Nucleotide Polymorphism) with a set of 12 microsatellites to detect hybridization between wild boars (*Sus scrofa scrofa*) and domestic pigs (*S. s. domesticus*). The accuracy of the two sets of molecular markers in detecting hybrid individuals was investigated with two different standard Bayesian analyses on simulated genotypes. Parental and hybrid individual detection was also performed on the real genotypes of 270 wild boars, 57 pigs and 139 phenotypically anomalous wild boars (PAWs). Both simulation and real genotype analyses showed similar capacity of both sets of markers to detect hybridization. Overall, first generation hybrids were the only hybrids detected well. In the PAWs sample, various proportions of molecular hybrids (up to 100%) were detected, depending on the marker and the method used. In addition, the probability of being detected as a hybrid increased with the number of domestic pig traits borne by PAWs. We concluded that despite the different characteristics of the sets of markers, they performed equally well in detecting hybridization. Choice of the molecular marker should thus be based on the economic costs of each type of marker.

***Keywords**:* Bayesian analysis, Hybridization, Introgression; Microsatellite; Single nucleotide polymorphism; Sus scrofa.

Gayet T., L. Say, E. Baubet, S. Brandt and S. Devillard. SNPs or Microsatellites? Assessing the reliability of different molecular markers to study hybridization between wild boar and domestic pig (*Sus scrofa*). *In prep*

Introduction

Hybridization is defined as interbreeding between individuals from two genetically different populations regardless of their phylogeny (sensu Rhymer and Simberloff, 1996). These populations are often considered as evolutionary significant units (ESUs, Ryder, 1986). Hybridization has the potential to induce great changes on fitness in ESUs in short evolutionary timescales (Barrett and Schluter, 2008; Feulner *et al.*, 2013). The impact of hybridization is particularly strong when it is linked to human activities. Indeed, ESUs separated by geographical barriers and exposed to different selective pressures may be brought in contact by anthropogenic factors such as habitat fragmentation or loss, or exotic species introduction. Moreover, high hybridization rates may be maintained by significant gene flux through perpetual releases of domestic or captive-raised individuals (Huxel, 1999; Laikre *et al.*, 2010). As these anthropogenic driven hybridizations do not occur naturally, they are usually considered a threat to the genetic integrity of the ESUs. They can lead to loss of genetic diversity, genetic adaptations through introgression and, in more extreme cases, to a complete change in the genetic structure of the ESU (Rhymer and Simberloff, 1996; Allendorf *et al.*, 2001; Laikre *et al.*, 2010).

Understanding the consequences of hybridization on fitness and on the genetic characteristics of the ESU is therefore of prime interest in population ecology and population genetic studies, and also in population management. Game species such as the wild boar (*Sus scrofa scrofa*) have often been the topic of such discussions due to the problematic demographic increase shown by the species in the past decades (Massei *et al.*, 2015). In addition, wild boar managers have formulated hypotheses on the putative role of hybridization in this demographic increase (Fulgione *et al.*, 2016). Indeed, wild boar can produce fertile offspring with the pig (*Sus scrofa domesticus*), its domestic counterpart. The pig was, and still is, heavily selected to improve production characteristics such as growth rate and fertility (Rauw *et al.*, 1998; Skewes *et al.*, 2008) putting wild boar at risk of being introgressed by domestic genes enhancing productivity (Fulgione *et al.*, 2016). Moreover, the wild boar was subject to restocking practices that used, in part, individuals raised in farms; it is now well-known that breeders often crossed captive wild boars with pigs to increase litter size and growth rate (Goulding, 2001; Canu *et al.*, 2014). Hybridization between the wild boar and the domestic pig also raises methodological questions on the detection of hybrids for both legislation purposes (for example, hybrids are forbidden in wild boar farms in France) and evolutionary biology studies. Due to historical and recent gene flux between both sub-species (Scandura *et al.*, 2011; Goedbloed *et al.*, 2013),

molecular diagnostic tools to reliably identify hybrids are sorely lacking. Thus, measuring individual introgression rate, defined as the invasion of the genetic material of an ESU (here an individual) by another ESU (Mallet, 2005), remains difficult.

Initially, the only method for detecting hybridization was based on individual karyotype, as the pig has 38 chromosomes while wild boar has only 36 in Europe (Scandura *et al.*, 2011). However this method is not very reliable, as 50% of F2 hybrids may display 36 chromosomes (McFee *et al.*, 1966) despite having half of the genome from the pig. Later, numerous sets of molecular markers were developed to study hybridization in wild boar. They have been proven efficient in detecting some past hybridization events (Scandura *et al.*, 2008; Goedbloed *et al.*, 2013; Canu *et al.*, 2014) but not in inferring hybridization rates in wild boar populations or individual rates of admixture. For example, Asian pig mitochondrial DNA (mtDNA) haplotype appeared in European pig breeds when species were crossed to improve European pig breeds (Giuffra *et al.*, 2000). The presence of Asian pig mtDNA in European wild boar clearly proves past hybridization events (Fang *et al.*, 2006; Scandura *et al.*, 2008) but it only shows hybridization with domestic pigs that bear the Asian mtDNA, and it is not transmitted by male hybrids. To circumvent this pitfall, male specific markers were also investigated on the Y-chromosome (Iacolina *et al.*, 2016). A non sex-specific gene coding for a protein involved in coat color, melanocortin-1 receptor (MC1R), was also used to detect hybridization in wild boar, as some allelic forms are specific to the domestic pig (Canu *et al.*, 2016). Finally, more neutral markers such as SNPs (Single Nucleotide Polymorphism) and microsatellites have also been used, as in numerous taxa, without giving hybrid identification beyond any doubt for wild boar (Scandura *et al.*, 2008, 2011; Frantz *et al.*, 2012; Goedbloed *et al.*, 2013; Canu *et al.*, 2014).

Therefore, the identification of hybrids, estimation of the population hybridization rate and the level of individual introgressions are still a major concern in both wild boar management and evolutionary biology. Comparing different sets of molecular markers in terms of their reliability in detecting hybrids and their levels of introgression is a first step toward a better understanding of the wild boar/domestic pig hybridization complex and its consequences in evolutionary biology. Such a comparison must rely on a simulated dataset but would also benefit from the inclusion of real hybrid individuals in data analysis. Despite increasing interest in the wild boar-domestic pig hybridization complex, to our knowledge no study including individuals of known hybrid origin has been carried out. Obtaining such individuals would require experimental crosses, which are difficult to carry out due to legislation and costly husbandry. However, individuals with a hybrid phenotype (i.e. displaying phenotypic

characters of both wild boar and domestic pig) can be found in natural populations. Including such individuals in analysis is important to quantify the ability of markers to detect hybridization and validate genetic tools only if they are able to detect it. This kind of approach has already proved efficient for other hybridization complexes (see Godinho *et al.*, 2011 for wolf×dog; Nussberger *et al.*, 2013 for wildcat×dometic cat).

In this study, we compared the efficiency of two sets of molecular markers, 20 SNPs and 12 microsatellites, to detect hybridization and quantify the level of introgression of the pig genome in wild boars. The 20 SNPs were specifically developed for the study of wild boar hybridization (Beugin *et al.*, 2017) while the microsatellites were commonly used in genetic population studies including paternity analysis (Gayet *et al.*, 2016). The genotypes of 270 wild boars and 57 pigs from commercial breeds defined the parental populations for both sets of markers. We conducted simulation analysis to assess the efficiency of both types of molecular markers to detect hybrids of different generations including F1 hybrids and backcrosses. We also added 139 individuals collected in wild populations of wild boars showing phenotypic evidence of pig introgression in order to verify the ability of both sets of markers to detect hybridization and quantify the level of introgression on real genotypes. This allowed us to be sure that hybrids detected in the wild boar population were not artefacts of the genetic analysis.

Material and Methods

Sample collection

Two hundred and seventy wild boars, i.e. with a typical wild boar phenotype, were sampled from the population located in the Châteauvillain-Arc-en-Barrois forest (48°02′N; 4°55′E, France) for five hunting seasons (2007-2011). In addition, 57 domestic pigs were sampled from commercial butchery meat as described in Beugin et al. (2017). A national sampling campaign was carried out by the ONCFS (French hunting and wildlife agency) for two years: hunters were asked to collect samples from wild boars showing evidence of hybridization with domestic pig (called thereafter phenotypically anomalous wild boars, abbreviated PAWs). We asked hunters to report criteria justifying sampling of a PAW. The criteria were part of a list of phenotypic evidence of hybridization with domestic pig described by Pinet (2005, see Supplementary Material, Table S1). A total of 139 samples of PAWs were obtained from all parts of France (Supplementary Material, Figure S1). PAWs were grouped according to the number of phenotypic characteristics from the pig that they displayed and that

hunters reported: no information (we assumed missing information was due to a careless mistake, n_{PAW0}=43), one criterion (n_{PAW1}=83), two criteria (n_{PAW2}=8), three criteria (n_{PAW3}=4) and four criteria (n_{PAW4}=1). Hereafter, PAWx is used to identify individuals with 'x' pig phenotypic reported by hunters (PAW+ indicates PAWs with one or more criteria).

Molecular analysis

All tissue samples were stored in alcohol in an individual hermetic straight container of 25ml and then were genotyped for 12 microsatellite and 20 SNP loci. The protocol of extraction, amplification and genotyping for SNPs is detailed elsewhere (Beugin *et al.*, 2017). For microsatellite loci, the procedure has already been described in Gayet *et al.* (2016).

Genetic analysis

All analyses were carried out only with individuals that had more than 50% of the genotypes obtained for both microsatellites and SNPs. A total of 447 individuals were included in analyses: 261 wild boars, 50 domestic pigs and 136 PAWs. A qualitative analysis of markers was first conducted with R 3.3.3 software (R Core Team, 2017) using *adegenet* 2.0.1 (Jombart, 2008) and *diveRsity* 1.9.90 (Keenan *et al.*, 2013). For each SNP and microsatellite locus, the differentiation index between pigs and wild boars (F_{ST}), informativeness of ancestry index as described in Rosenberg (2003), expected heterozygosity (He) and allelic frequency of the most common allele (F) for pigs and wild boars were calculated. For microsatellite loci, the total number of allele (N), the allelic richness (Ar) and the number of private alleles were estimated in each population. Values obtained for pigs and wild boars were compared using Wilcoxon tests.

Two Bayesian statistics-based software programs were used to investigate hybridization patterns between wild boars and pigs with both sets of markers. The first was the clustering method STRUCTURE 2.3.4 (Pritchard *et al.*, 2000). It was used to infer the most probable number of parental groups (K), and also to assign each individual to one (or several) of the K parental groups with defined probabilities (q_{ik}). All analyses were carried out with 20,000 iterations of burn-in followed by 500,000 iterations of MCMC (Markov chain Monte Carlo). According to the documentation, the parameter λ (index of correlation of allele frequencies) was first estimated for K=1 and set to 0.6877 for microsatellites and to 0.7823 for SNPs in further analyses. Other parameters were set as follows: admixture model, correlated allele frequencies among populations without including population information and other parameters were set as their default value. Using these parameters for SNPs and microsatellites, 20 independent runs

were realized from K=1 to K=9. The program STRUCTURE HARVESTER (Earl and VonHoldt, 2011) was then used to estimate the most likely number of clusters K that fit the data with the method described by Evanno et al. (2005). The results of the 20 runs were averaged using CLUMPP (Jakobsson and Rosenberg, 2007) to get averaged q_{ik} values for each individual.

The second method used NEWHYBRIDS software (Anderson and Thompson, 2002) with the version available on https://github.com/eriqande/newhybrids. It infers an individual posterior probability of assignment to each of 10 genotype frequency classes: pig, wild boar, F1, F2, both first generation backcrosses (F1 × (pig or wild boar)), and double-backcrosses (backcross[F1 × pig] × pig, backcross[F1 × wild boar] × wild boar, and backcross[F1 × (pig or wild boar)] × F1). The analysis was carried out with 20 independent runs of 20,000 iterations of burn-in and followed by 500,000 iterations of MCMC. As the origin of the pigs is known, we used the parameter z to specify they belong to the genotype class of pig but all other parameters were set as default. The 20 runs were then averaged.

Simulation analysis

These Bayesian approaches do not permit statistical assessment of the efficiency of loci to detect hybridization. Simulations have been made to overcome this problem (Vähä and Primmer, 2006). For both sets of markers, all pigs showing a $q_{i\text{-}Pig}$ (individual probability to belong to pig group) equal to or higher than 0.99 with STRUCTURE (34 with SNPs and 36 with microsatellites) were used to create the reference population of pure pig. Subsequently, the same number of wild boars was kept to create the reference population of pure wild boars (they all had a $q_{i\text{-}WB}$ higher than 0.996 whether it was with SNPs or microsatellites). We used the function *hybridize* from *adegenet* to create 100 individuals of each of the 10 genotype classes described above. A total of 20 datasets of 1000 individuals each containing the 10 different hybrid genotypic classes were simulated for SNPs and for microsatellites. They were analyzed with STRUCTURE and NEWHYBRIDS with the same parameters used for real datasets. These simulations allowed us to obtain threshold values of assignment to the different hybrid classes for individuals whether datasets were real or simulated.

For each marker in STRUCTURE, thresholds were determined based on simulated datasets results to create a decision rule to assign a real individual to one of the parental or hybrid groups. For clarity, we only considered probabilities of assignation to the wild boar group (i.e. $q_{i\text{-}WB}$, acknowledging for K=2, $q_{i\text{-}Pig} = 1 - q_{i\text{-}WB}$). The threshold of assignment to the wild boar group ($q_{WB\text{-}WB\text{-}95\%}$) was calculated as the $q_{i\text{-}WB}$ value obtained by 95% of simulated wild boars. Thus,

a real individual with a $q_{i\text{-}WB}$ higher than $q_{WB\text{-}WB\text{-}95\%}$ would be considered a true wild boar. Symmetrically, a threshold of assignment for the pig group ($q_{Pig\text{-}WB\text{-}95\%}$) was also calculated using $q_{i\text{-}WB}$ values obtained for simulated pigs. Real individuals obtaining $q_{i\text{-}WB}$ below $q_{Pig\text{-}WB\text{-}95\%}$ would be assigned to pigs but those with $q_{i\text{-}WB}$ between $q_{WB\text{-}WB\text{-}95\%}$ and $q_{Pig\text{-}WB\text{-}95\%}$ would be considered hybrid. To go further and estimate individual introgression level, threshold values were also calculated for assignment to the eight hybrid genotypic groups. However STRUCTURE only gives assignment to parental populations, so we calculated an upper and a lower threshold for each genotypic ($q_{H\text{-}WB\text{-}97.5\%}$ and $q_{H\text{-}WB\text{-}2.5\%}$ respectively, H indicating the hybrid genotypic class considered), containing 95% of simulated individuals' $q_{i\text{-}WB}$ values of the given genotypic group. The genotypic class of a real individual can be assessed using these thresholds. Firstly, we compare the $q_{i\text{-}WB}$ obtained by a given real individual to $q_{WB\text{-}WB\text{-}95\%}$. If $q_{i\text{-}WB} > q_{WB\text{-}WB\text{-}95\%}$, the individual is considered a wild boar, otherwise a hybrid (or assigned to pig if $q_{i\text{-}WB} < q_{Pig\text{-}WB\text{-}95\%}$). To identify the hybrid class, the $q_{i\text{-}WB}$ must be between the $q_{H\text{-}WB\text{-}97.5\%}$ and $q_{H\text{-}WB\text{-}2.5\%}$ of the H hybrid genotypic class, and out of all the others. Then the individual can be considered a hybrid of the H class. For each marker, the medians of obtained $q_{i\text{-}WB}$ values per genotypic class were compared to expected $q_{i\text{-}WB}$ values (considering $q_{Pig\text{-}WB}=0$, $q_{WB\text{-}WB}=1$, $q_{F1\text{-}WB}=0.5$, $q_{F2\text{-}WB}=0.5$, $q_{Backcross.Pig\text{-}WB}=0.25$, $q_{Backcross.WB\text{-}WB}=0.75$, $q_{Backcross.WB\times WB\text{-}WB}=0.875$…etc.) using chi-squared tests.

For each marker in NEWHYBRIDS, 10 thresholds (one for each of the 10 genotype classes described above) were calculated to get a decision rule to assign an individual to one of the 10 genotype classes. Each threshold was calculated, based on the results obtained for each of the 10 simulated genotype classes, considering the probability of assignment to the right group obtained by 95% of simulated individuals of the group (for example, the probability of assignment to the F1 group obtained by simulated F1). Thus, a real individual was considered to belong to a genotypic group if its individual probability of assignment to this group was higher than the threshold of this group and below all other nine thresholds. For example, an individual was considered to be F1 if the assignment to the F1 group was higher than the threshold value of F1 group and assignments to other genotypic groups were below corresponding threshold values.

The results of the simulation allowed us to evaluate the accuracy (the ratio of the number of individuals correctly assigned to the group and the number of all individuals assigned to the group) per genotypic group, for both sets of markers and for each method, as described in Vähä and Primmer (2006) as the efficiency (proportion of individuals of a genotypic group correctly

assigned to the genotypic group) is already set at 95%. The accuracy of the sets of markers for each analysis were compared using chi-squared tests. Thresholds obtained made it possible to classify PAWs in parental or hybrid groups with both methods and to verify the ability of both sets to detect them as hybrids.

Results

Markers description

SNP data analysis showed that an allele is present in four loci in the wild boar population (frequency of 1) but they were also present in more than 50% of the pigs (Table 1). The genetic diversity measured as the expected heterozygosity was significantly higher in the domestic population than in the wild one (Wilcoxon test: W=77.5, $p < 0.01$, Table 1). With microsatellites, both allelic richness and expected heterozygosity were the same between pigs and wild boars (Wilcoxon test: W = 57, p = 0.41 and W = 42, p = 0.08 respectively; Table 2). Some alleles were specific to each population (present in one but not in the other) but none of these specific alleles were fixed in any population. Differentiation indexes were higher with SNPs than with microsatellites. The F_{ST} ranged from 0.19 to 0.96 with a median of 0.57 for SNPs (Table 1) *versus* a median value of 0.13 for microsatellites ranging from 0.05 to 0.50 (Table 2) (Wilcoxon test: W = 12, $p < 0.01$). However, the informativeness indexes were not significantly different (Wilcoxon test: W = 155, p = 0.18), with SNPs showing a median of 0.19 *versus* 0.31 for microsatellites (Table 1 and 2).

Table 1: Information for SNP loci for wild boar (Wb) and pig hybridization analysis. The name, the F_{ST} between wild boar and pig and the informativeness index (I_n) are given for each locus. For each population, the number of individuals genotyped (n), the frequency of the major allele (F) and the expected heterozygosity at the locus (He) are provided. PAWs were excluded from the analysis.

Locus	F_{ST}	I_n	n_Wb	F Wb	He_Wb	n_Pig	F Pig	He_Pig
SNP56	0.96	0.55	261	0.99	0.03	50	0.95	0.10
SNP46	0.93	0.54	261	0.97	0.06	49	0.96	0.08
SNP43	0.92	0.55	261	0.96	0.08	50	0.98	0.04
SNP53	0.83	0.39	261	0.94	0.12	50	0.88	0.21
SNP15	0.81	0.27	261	0.99	0.02	50	0.63	0.47
SNP48	0.76	0.34	261	0.90	0.18	50	0.87	0.23
SNP4	0.72	0.21	205	1	0	47	0.51	0.50
SNP12	0.70	0.20	261	0.98	0.05	50	0.56	0.49
SNP52	0.64	0.34	261	0.77	0.35	50	0.97	0.06
SNP23	0.58	0.13	259	0.98	0.03	50	0.60	0.48
SNP41	0.56	0.12	245	1	0	50	0.70	0.42
SNP50	0.55	0.26	261	0.73	0.39	50	0.93	0.13
SNP27	0.54	0.11	259	0.99	0.02	50	0.67	0.44
SNP55	0.50	0.17	261	0.82	0.30	50	0.75	0.38
SNP57	0.47	0.18	248	0.74	0.39	50	0.84	0.27
SNP2	0.45	0.08	237	1	0	50	0.78	0.34
SNP54	0.44	0.15	261	0.77	0.36	50	0.77	0.35
SNP39	0.44	0.08	253	1	0	50	0.79	0.33
SNP28	0.24	0.04	249	0.98	0.03	50	0.83	0.28
SNP33	0.19	0.04	255	0.92	0.15	50	0.70	0.42

Table 2: Information for microsatellite loci for wild boar (Wb) and pig hybridization analysis. The marker name, the F_{ST} between wild boar and pig, the informativeness index (I_n), and the total number of alleles (N) are given for each locus. For each population, the number of individuals genotyped (n), the allelic richness (Ar), the number of private alleles (Np), the frequency of the most frequent allele (F) and the expected heterozygosity at the locus (He) are provided. PAWs were excluded from the analysis.

Locus	F_{ST}	I_n	N	N Wb	Ar Wb	Np Wb	F Wb	He Wb	N Pig	Ar Pig	Np Pig	F Pig	He Pig
SO355	0.50	0.22	8	250	2.83	0	0.95	0.1	50	7.12	4	0.39	0.73
SO068	0.38	0.48	14	243	8.92	4	0.63	0.55	48	8.87	2	0.53	0.67
SW936	0.31	0.30	9	261	4.84	0	0.69	0.48	50	8.37	3	0.38	0.76
SW122	0.26	0.21	7	259	4.61	0	0.73	0.45	49	6.26	2	0.35	0.76
SW24	0.17	0.40	9	255	5.28	2	0.42	0.7	43	6.88	3	0.31	0.81
SW240	0.17	0.36	10	259	6.39	3	0.49	0.65	50	6.99	3	0.24	0.81
SW2496	0.09	0.32	16	250	9.99	7	0.24	0.83	44	8.65	2	0.28	0.83
SO228	0.07	0.16	10	259	3.78	2	0.43	0.65	50	7.51	4	0.39	0.77
CGA	0.07	0.34	19	245	14.25	8	0.16	0.9	44	10.13	1	0.22	0.86
SO005	0.07	0.32	24	233	14.33	10	0.18	0.88	48	12.57	2	0.18	0.88
SW2021	0.05	0.25	16	257	9.01	4	0.38	0.75	50	10.99	6	0.28	0.85
SO215	0.05	0.08	5	260	2	1	0.91	0.17	50	3.61	3	0.86	0.25

STRUCTURE results with real and simulated data

Using the method described by Evanno et al. (2005) based on STRUCTURE analyses, the most likely number of genetic groups was K = 2 for both sets of markers. Wild boars are separated from domestic pigs (Figure 1, a and c).

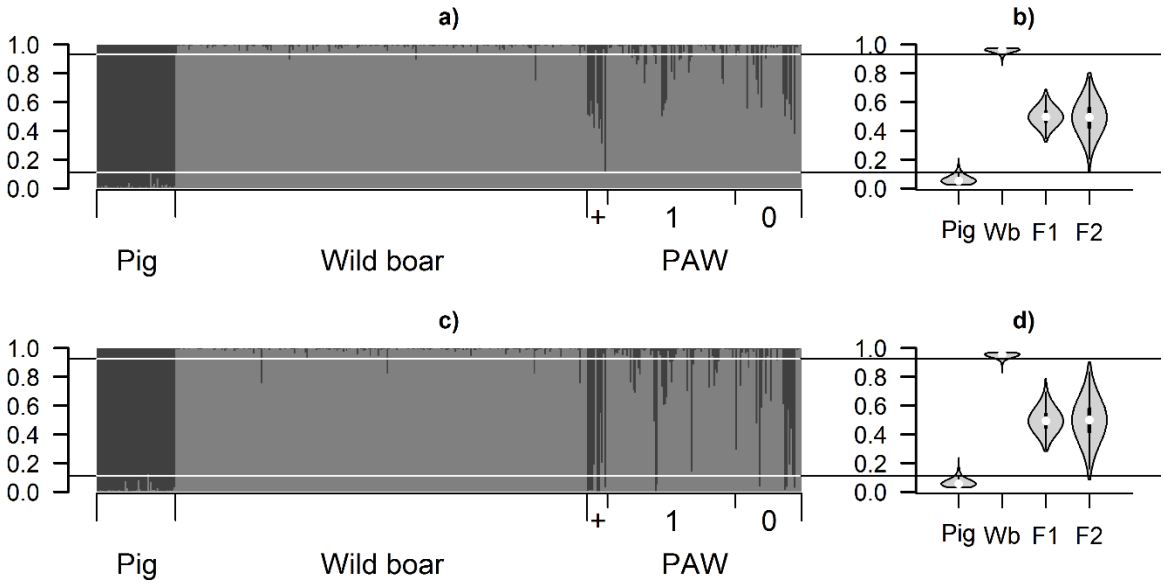


Figure 1: Assignment probabilities to the two genetic clusters 'pig' (dark gray) and 'wild boar' (light gray) inferred by STRUCTURE analysis performed on domestic pigs (n = 50), wild boars (n = 261) and PAWs from the national sampling campaign (separated by the number of phenotypic traits associated with hybridization recorded by the hunters: more than one n_{+}=13, one trait n_1=81 and no trait reported n_0=42) a) with SNPs and c) with microsatellites. Each individual is represented by a vertical bar.

Assignment probabilities to the wild boar group obtained by 4 genotype classes (n=2000 for each class) simulated with b) SNPs and d) microsatellites. The width shows the density of points. The white points show the median, the thick black vertical lines the central quartiles and the fine vertical lines 1.5 the inter-quartile space.

The horizontal lines represent the threshold values determined as the value 95th percentile of assignment values to the wild boar group obtained by simulated wild boar and pig ($q_{WB\text{-}WB\text{-}95\%}$ = 0.931 and $q_{Pig\text{-}WB\text{-}95\%}$ = 0.111 with SNPs and $q_{WB\text{-}WB\text{-}95\%}$ = 0.924 and $q_{Pig\text{-}WB\text{-}95\%}$ = 0.111 with microsatellites for wild boar and pig respectively).

Simulated datasets analysis gave information about the range of $q_{i\text{-}WB}$ values that may be obtained for the different genotypic classes (Supplementary Material, Figure S2). Results were similar between SNPs and microsatellites with lower ranges of $q_{i\text{-}WB}$ values for parental populations (pigs and wild boars) than for other genotypic groups. In both cases, the median of $q_{i\text{-}WB}$ values obtained for each genotypic group was similar to expected values (χ^2 = 3.23, df = 9, p-value = 0.95 for SNPs and χ^2 = 3.26, df = 9, p-value = 0.95 for microsatellites). The ranges obtained by wild boar and pig were separate from each other for both sets of markers. The $q_{i\text{-}}$

$_{WB}$ values ranged from 0.855 to 0.972 for wild boars and from 0.023 to 0.207 for pigs with SNPs. For microsatellites, $q_{i\text{-}WB}$ values ranged from 0.829 to 0.969 for wild boars and from 0.029 to 0.237 for pigs with SNPs. These ranges were also separate from F1's (from 0.282 to 0.790 with microsatellites and from 0.320 to 0.687 with SNPs, Figure 1, b and d). Assignment threshold values were calculated to evaluate the significance of q_{ik} values obtained on real data. Overall accuracies were similar between SNPs and microsatellites ($\chi^2 = 0.588$, df = 9, p-value = 0.999, Supplementary Material, Table S2). The threshold values of assignment to the wild boar population calculated using simulated wild boar results were estimated at $q_{WB\text{-}WB\text{-}95\%} = 0.931$ with SNPs and 0.924 with microsatellites. Symmetrically, thresholds were estimated for assignment to the pig group $q_{Pig\text{-}WB\text{-}95\%} = 0.110$ and 0.111 with SNPs and microsatellites respectively (Figure 1). Other hybrid genotypic groups displayed ranges of $q_{i\text{-}WB}$ values that strongly overlapped each other (Supplementary Material, Figure S2). These hybrid genotypic groups were poorly differentiated one from another because many individuals were assigned to several groups. Thus accuracy was very low for hybrid and introgressed groups with a maximum of 42.5% with SNPs and 34.7% with microsatellites (Supplementary Material, Table S2). This means less than 50% of individuals assigned to a group are really from this group. So threshold values ($q_{H\text{-}WB\text{-}97.5\%}$ and $q_{H\text{-}WB\text{-}2.5\%}$) were not reported for these groups. Moreover, some backcrosses with wild boar obtained $q_{i\text{-}WB}$ values higher than the thresholds $q_{WB\text{-}WB\text{-}95\%}$ for both sets of markers. This means they would be wrongly assigned to the wild boar groups (explaining the accuracy of 89.9% with SNPs and 83.4% for microsatellites) and that wild boar detected by this method are in fact a mixture of pure individuals and backcrosses of different generations. However, backcrosses with pig were always separated from the wild boar genetic cluster. To summarize, individuals with $q_{i\text{-}WB}$ values above $q_{WB\text{-}WB\text{-}95\%}$ will be assigned to the wild boar parental population. Real individuals with $q_{i\text{-}WB}$ values below $q_{WB\text{-}WB\text{-}95\%}$ values cannot be considered as individuals from the wild boar group. They will be identified as hybrids. However their hybrid class cannot be categorized due to overlapping ranges of values of different hybrid genotypic groups and the low accuracy of detection. Individuals with a high proportion of the genome from one of the parental populations (especially Bc.WB×WB and Bc.P×P) may not be identified as hybrids and would be assigned to this parental population.

On real genotypes, 1.92% of wild boars obtained $q_{i\text{-}WB}$ values below the $q_{WB\text{-}WB\text{-}95\%}$ threshold value with both SNPs and microsatellites (Table 3). PAWs from the sampling campaign also showed $q_{i\text{-}WB}$ values below the $q_{WB\text{-}WB\text{-}95\%}$ threshold. This was the case for 31%

and 42.9% out of 42 PAW0, and for 35.1% and 37.2% out of the 94 PAW+, for SNPs and microsatellites respectively. The proportion of individuals showing $q_{i\text{-}WB}$ below $q_{WB\text{-}WB\text{-}95\%}$ increased with the number of criteria for both sets of markers because 27.2% and 30.9% out of the 87 PAW1 and 75% and 62.5% out of the 8 PAW2 were detected as hybrids, with SNPs and microsatellites respectively. All PAW3 and PAW4 obtained $q_{i\text{-}WB}$ values below $q_{WB\text{-}WB\text{-}95\%}$. Overall, 23.4% of PAW+ were identified as hybrids with both sets of markers (Table 3).

NEWHYBRIDS results with real and simulated data

With both sets of markers, the pigs were well separated from the wild boars using NEWHYBRIDS (Figure 2 a and c).

With both SNPs and microsatellites, simulation analysis results showed pigs, wild boars and F1 were globally well identified (Figure 2 b and d) as they were assigned to their right group with high probabilities, despite some individuals obtaining low probabilities of right assignment (down to 0 for F1). Thus, proportions of correct assignment were high especially compared to other genotypic classes. Threshold values were calculated using probabilities of assignment above which 95% of simulated individuals were assigned to their right genotypic group. No difference in overall accuracy was found between SNPs and microsatellites (χ^2 = 1.98, df = 9, p-value = 0.992). Accuracies for pigs, wild boars and F1 were the highest for both set of markers (Supplementary Material, Table S3). Exact assignment occurred mainly for wild boar and F1 with both SNPs (more than 91% of accuracy) and microsatellites (more than 73% of accuracy). Further introgression levels were assigned to their right genotypic group with lower values of probabilities, thus with less accuracy. Indeed, some genotypic groups obtained a high proportion of individuals with a probability of 0 to be assigned to the right group (Figure 2 b and d for the F2 hybrid class, Supplementary Material, Figure S4) giving 95% threshold values close to 0. Only threshold values of assignment for pigs, wild boars and F1 were considered for further analysis for both sets of markers due to low accuracies for other genotypic groups. These thresholds were always higher with SNPs than with microsatellites. Probabilities of assignment above which 95% of simulated wild boars were assigned were 0.773 and 0.707 for SNPs and microsatellites respectively. These 95% threshold values of assignment to pig were 0.786 with SNPs and 0.731 with microsatellites. Threshold of assignment to F1 were 0.578 with SNPs and 0.289 with microsatellites.

On real genotypes, proportions of wild boars that displayed probabilities of assignment below the wild boar threshold was 6.13% with SNPs and 5.75% with microsatellites (Figure 2a

and 2c). Considering the 42 PAW0, 38.1% and 50% obtained assignment to wild boar below the threshold, with SNPs and microsatellites respectively. Focusing on PAW+, 43.6% of the 94 showed no significant assignment to wild boar group with both SNPs and microsatellites. These proportions increased with the number of criteria for both sets of markers as 35.8% and 38.7% of PAW1, 87.5% and 62.5% of PAW2 and 100% of PAW3 and PAW4 were not assigned to the wild boar group with SNPs and microsatellites respectively (Supplementary Material, Figure S5). Overall, 25 PAW+ were identified as hybrids with both sets of markers (Table 3). No individual was assigned to the F1 genotypic group (Supplementary Material, Figure S3) significantly suggesting that individuals detected as hybrids have deeper introgression levels.

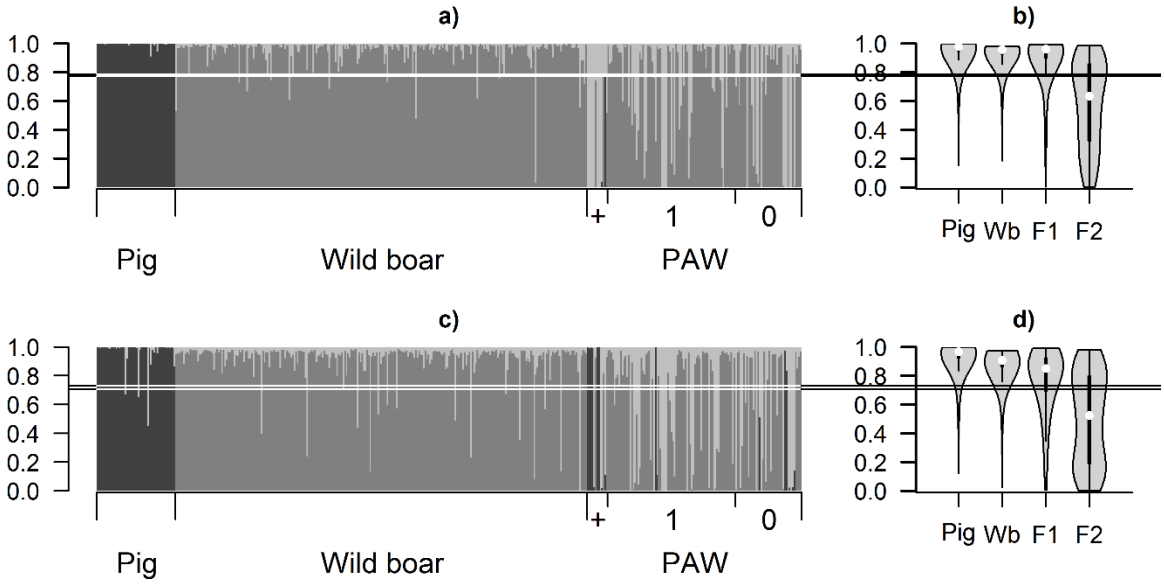


Figure 2 Assignment probabilities to three genotype frequency classes 'pig' (dark gray) and 'wild boar' (gray) and 'hybrid' (light grey) that pool all the 8 hybrid genotype classes inferred by NEWHYBRIDS analysis performed on domestic pigs (n = 50), wild boars (n = 261) and PAWs from the national sampling campaign (separated by the number of phenotypic traits associated with hybridization recorded by the hunters: more than one n_+=13, one trait n_1=81 and no trait reported n_0=42) with a) SNPs and c) microsatellites. Each individual is represented by a vertical bar.

Assignment probabilities to the pig, wild boar (Wb), F1 and F2 genotypic groups obtained by pig, wild boar, F1 and F2 individuals (n = 2000 for each class) simulated with b) SNPs and d) microsatellites. The width shows the density of points. The white points show the median, the thick black vertical lines the central quartiles and the fine vertical lines 1.5 the inter-quartile space.

The horizontal lines represent the threshold values determined as the value 95th percentile of values obtained by simulated wild boar and pigs (0.773 and 0.786 with SNPs and 0.707 and 0.731 with microsatellites for wild boar and pigs respectively).

(Colored figures are available in the Supplementary Material, Figure S3)

Table 3: Number of common individuals between two analyses (STRUCTURE and NEWHYBRIDS) and/or two markers (SNPs and microsatellites) with assignment probabilities to the wild boar group lower than the threshold values for wild boars (bold values, n = 261) and PAW+ (i.e. phenotypic anomalous wild boars with at least one trait associated with hybridization recorded by the hunters, normal values, n = 94). The diagonal shows the number of individuals with assignment probabilities to the wild boar group lower than the threshold values in each analysis and for each marker.

		Structure		Newhybrids	
		SNP	Microsatellite	SNP	Microsatellite
Structure	SNP	**5** - 33	**1** - 22	**5** - 33	**1** - 23
	Microsatellite	NA	**5** - 35	**2** - 24	**5** - 35
Newhybrids	SNP	NA	NA	**16** - 41	**2** - 25
	Microsatellite	NA	NA	NA	**15** - 41
				All combined	**1** - 22

Table 4: Proportions of real wild boars (n=261) and phenotypically anomalous wild boars (PAWx, x indicating the number of criteria reported by hunters, and n_0=42, n_1=81, n_2=8, n_3=4 and n_4=1) identified as hybrids 0 to 4 times with STRUCTURE and NEWHYBRIDS analysis with a) SNPs, b) microsatellites, c) all analyses combined

a)

Real individuals group	Number of identification as hybrid	0	1	2
	Wild boar	93.87	4.21	1.92
	PAW0	61.9	7.14	30.95
	PAW1	64.2	8.64	27.16
	PAW2	12.5	12.5	75
	PAW3	0	0	100
	PAW4	0	0	100

b)

Real individuals group	Number of identification as hybrid	0	1	2
	Wild boar	94.25	3.83	1.92
	PAW0	50	7.14	42.86
	PAW1	61.73	7.41	30.86
	PAW2	37.5	0	62.5
	PAW3	0	0	100
	PAW4	0	0	100

c)

Real individuals group	Number of identification as hybrid	0	1	2	3	4
	Wild boar	88.89	7.66	2.68	0.38	0.38
	PAW0	40.48	9.52	21.43	4.76	23.81
	PAW1	44.44	12.35	24.69	3.7	14.81
	PAW2	12.5	12.5	12.5	0	62.5
	PAW3	0	0	0	0	100
	PAW4	0	0	0	0	100

Discussion

Marker description and informativeness

The comparison between SNP and microsatellite sets showed higher differentiation indexes for SNPs. This is consistent with the design of the set of SNPs since they were especially developed for the study of hybridization (Beugin *et al.*, 2017). In addition, the SNPs

were located in coding regions which may have been under selection in pigs (Beugin *et al.*, 2017) increasing differentiation indexes between pigs and wild boars. Private alleles occurred in domestic pigs (although less than in Beugin *et al.* (Beugin *et al.*, 2017)) but they were not fixed. For microsatellites, some alleles were also private to the pig parental group. Finding these private alleles in wild boar may reveal hybridization as finding Asian mtDNA in a wild boar proves that hybridization occurred in the wild boar lineage (Fang *et al.*, 2006; Scandura *et al.*, 2008). However, these pig private alleles cannot be considered diagnostic since they are not fixed in the pig parental group. Even if F_{ST} indexes differed between sets of markers, the informativeness indexes did not, highlighting that both sets of markers may perform equally well in detecting hybridization. This last point was confirmed by the simulated data analysis results and fits previous studies that showed that fewer microsatellite loci are required than SNPs to obtain a given amount of information (Rosenberg *et al.*, 2003; Liu *et al.*, 2005).

Simulated parental individuals identified and separated from F1 hybrids

Wild boars and pigs used to create simulated datasets had high assignment value to their respective groups. However, individuals of parental populations simulated using these individuals were assigned to their respective group with lower probabilities with both sets of markers with STRUCTURE. As for NEWHYBRIDS, some simulated parental individuals obtained very low assignment probabilities to their groups. These results can be explained by the sensitivity of both STRUCTURE and NEWHYBRIDS to the proportion of hybrids in the datasets (Vähä and Primmer, 2006), as proportions of hybrids in the simulated datasets are higher than those in the real ones. Moreover, the population sampling size can also influence the results (Puechmaille, 2016) and simulated data sets did not have the same characteristics as the real ones (more individuals, higher proportions of hybrids compared to parental individuals). Real individuals from parental populations obtained initially high assignment for their genotypic group while simulated individuals from parental population got lower values when analyzed in the simulated datasets. So the threshold values tended to be lower than expected and, for a real individual to be excluded from the wild boar genotypic group, the $q_{i\text{-}WB}$ needs to be low. With STRUCTURE, we considered $q_{WB\text{-}WB\text{-}95\%}$ and $q_{Pig\text{-}WB\text{-}95\%}$ thresholds were conservative for the identification of hybrids (hybrid origin is sure for an individual assigned to hybrids, i.e. $q_{i\text{-}WB} < q_{WB\text{-}WB\text{-}95\%}$) and relaxed for the parental population (some hybrids may be assigned to wild boar, i.e. $q_{i\text{-}WB} > q_{WB\text{-}WB\text{-}95\%}$). Similarly, with NEWHYBRIDS, as some simulated individuals from parental population obtained low assignment probabilities, thresholds are low compared to what could be expected. Thresholds are once again relaxed for identification of individuals

from parental populations. However, accuracy for parental group assignments was high with both SNPs and microsatellites and with both methods. This proves that despite the low threshold values compared to what could be expected, these thresholds are reliable and meaningful. In the STRUCTURE analysis, ranges of $q_{i\text{-}WB}$ values for F1 hybrids did not overlap parental $q_{i\text{-}WB}$ ranges. However, ranges of $q_{i\text{-}WB}$ values for F1 hybrids strongly overlapped all other genotypic class $q_{i\text{-}WB}$ ranges with both sets of markers. Thus, STRUCTURE can identify hybrids but not estimate the introgression level, and both types of markers performed equally well. Vähä and Primmer (2006) showed that 24 microsatellite loci for a F_{ST} of 0.12 are required to distinguish F1 from parental population with STRUCTURE. With half this number of loci and a similar median value of F_{ST}, our set of microsatellite markers allowed F1 differentiation from parental populations, suggesting the efficiency might also depend on loci used. NEWHYBRIDS performed slightly better because assignment probabilities for the F1 group were better with both SNP and microsatellites. Moreover, accuracy for F1 was high with both sets of markers. This suggests that F1 could be identified with confidence using NEWHYBRIDS. Overall, assignments to parental groups are reliable with both SNPs and microsatellites and the two methods, and so are assignments to F1 with NEWHYBRIDS with both sets of markers.

On the difficulty of quantifying individual introgression level beyond F1

Simulation results showed that genotypic groups other than parental and F1 groups were identified with low accuracies, independently of the marker and the method. Indeed, more than 50% of individuals assigned to any of these groups were wrongly assigned. With NEWHYBRIDS, assignment probabilities to the right group were low for individuals from these groups with both sets of markers, and high proportions of individuals obtained a probability of 0 to be assigned correctly. This led to very low threshold values, which were not reliable. With STRUCTURE, the genotypic group was impossible to infer since ranges of $q_{i\text{-}WB}$ values obtained for simulated hybrid groups (including F1) strongly overlapped each other. This led threshold values to be close to one another. Even if by default 95% of individuals were assigned to their right genotypic class, high proportions were also significantly assigned to other genotypic groups. As genotypic groups could not be distinguished, thresholds were not reliable. With both sets of markers and both methods, some backcrosses were even assigned to the closest parental population. This can be explained firstly by the high proportion of the parental population genome in the most introgressed genotypic groups (in theory, 87.5% of pig and wild boar genome for, respectively, Bc.P×P and Bc.Wb×Wb). Secondly, the relaxed thresholds for the parental population, allowing more assignment mistakes in parental populations than in hybrids,

may cause some hybrids to be wrongly assigned to parental groups. It is noteworthy that backcrosses with pig (Bc.P, Bc.P×P or Bc.P×F1) never overlapped the range of values of wild boar even though the wild boar identification is relaxed. However, these kinds of individuals should be rare and may only occur where pigs are kept in outdoor enclosures, which is a rare practice except in a few regions (Canu *et al.*, 2014). In the wild, the most likely crosses are backcrosses with wild boars (Bc.Wb, Bc.Wb×Wb or Bc.Wb×F1). Thus these genotypic groups are the most important to detect, but simulations showed they may not be identified. To distinguish the purebred population from F1 and backcrosses, Vähä and Primmer (2006) suggested the use of 48 microsatellite loci with a F_{ST} value of 0.21. Our microsatellite set is thus not powerful enough to do that. Our simulation analyses results can be compared to the study realized by Godinho et al. (2011) on wolf and dog hybridization, since they used same Bayesian approaches. Their STRUCTURE study showed high identification efficiency for the parental population, distinction between parental from F1 and from F2 and a decrease of identification for more introgressed individuals. With NEWHYBRIDS, overall they obtained better identification rates of parental groups up to the first generation backcrosses. This is mainly due to the fact that they investigated for simple backcrosses with parental population. As assignment probabilities sum to 1 for each individual, decreasing the number of possible assigned clusters tends to increase assignment probabilities in each cluster. However, this strategy does not enable the detection of more introgressed individuals, nor does it estimate how more introgressed levels are assigned in this framework.

On the detection of hybrids in real populations

Using thresholds obtained by simulation analyses, between 30% and 50% of PAW0 were excluded from the wild boar group. This suggests that they were not sampled by mistake by hunters as high proportions proved to be hybrid individuals (similar proportions to PAW1). Depending on the marker and the analysis, between 35.1% and 43.6% of PAW+ (i.e. 33 and 41 out of 94), obtained assignment below threshold for wild boar. The proportion of PAW+ identified as hybrid was clearly dependent on the number of criteria indicating hybridization. For both sets of markers, the proportions increased from PAW1 to PAW2, to finally reach 100% of hybrids detected for PWA3 and PAW4 (Table 4). Indeed, the probability of expressing phenotypic traits should increase with the proportion of pig genome in the individuals, which decrease with the time elapsed since the hybridization event. Both types of markers performed again equally well in identifying PAWs as hybrids (Table 4) even if they did not recognize some individuals. This can be explained by different selective pressures on microsatellites and

SNPs. Indeed, SNPs were chosen in coding genome region (Beugin *et al.*, 2017) increasing the probability to be expressed and counter-selected in the wild (Frantz *et al.*, 2013; Battocchio *et al.*, 2017), while microsatellites are considered more neutral. Overall, 23.4% of PAW+ were assigned to hybrids in all analyses proving the ability of our sets of markers to detect hybridization (Table 3). As no individual was significantly assigned to the F1 group with NEWHYBRIDS, their levels of introgression were not assessed.

So far, with simulation analysis we showed both sets of markers were able to identify parental individuals with high accuracy. Thus hybrid individuals are recognized as such when excluded from parental groups, but individual introgression rates cannot be assessed. Analysis of genotypes of PAWs showed high proportions of individuals assigned to hybrid groups with both SNPs and microsatellites, proving them reliable in identifying hybrids in real datasets. As no set outperformed the other, we used both to estimate hybridization rates for a wild boar population. Between 1.9% and 7.3% of 261 wild boars were identified as hybrids, depending on the marker and the method. As none was assigned to the F1 group with NEWHYBRIDS, the introgression level was not considered. Only one real wild boar was identified by all four analyses giving confidence about the fact that hybridization events happened in the past in this population (Table 3). The proportion of hybrids in our population is consistent with other studies on free-ranging populations of wild boars in Europe, which ranged from around 2% to 10% (Scandura *et al.*, 2008, 2011; Koutsogiannouli *et al.*, 2010; Frantz *et al.*, 2013; Goedbloed *et al.*, 2013; Canu *et al.*, 2014, 2016). Some individuals, PAWs or wild boars, were identified as hybrids by some analyses but not all of them. Apart from the different information provided by markers, this can be explained, on the one hand because they may have been wrongly identified as wild boar in one or several cases since we were conservative in identifying hybrids. On the other hand, we chose the threshold as the 95^{th} percentile of simulated value, thus some wild boar may also be falsely assigned as hybrids for one or several analyses, but not for the others.

The detection of PAWs as hybrids, associated with simulation analyses, made it possible to prove both the efficiency and the reliability of our markers to detect hybridization. Both sets of markers were equally efficient in detecting hybridization because none outperformed the other whether with simulated or real data. However none of them is able to identify individual levels of introgression after F1 hybrids. More studies are thus required to estimate individual introgression rates. Next generation sequencing methods such as RADSeq may be promising tools to investigate the whole genome scale for new sets of markers. Powerful diagnostic

markers could help study fine-scale hybridization to understand the impact of hybridization on wild boar life history traits like Fulgione *et al.* (2016), to improve management practices when choosing pure population sources for restocking and to improve legislation to detect fraud. Nevertheless both sets can measure introgression rates at the population scale. For future studies, it may be worthwhile to consider the cost benefits ratio of different sets depending on the question investigated. SNP genotyping is cheaper than microsatellite genotyping, but microsatellites enable investigation of other population genetic questions. We hope these results provide researchers and managers with guidelines in the use of genetic tools to investigate hybridization issues and new insights in the design of sets of markers for genetic studies.

Acknowledgments

We are grateful to all those who helped collect harvested wild boars, especially P. Van den Bulck, to the ONF, and to F. Jehlé, who allowed us to work in the study area. We thank Réseau Ongulés Sauvages for the organization of the national sampling campaign of phenotypically anomalous wild boars. This work was supported by the French national hunting and wildlife agency (ONCFS) and by the University of Lyon. ONCFS contributes also by supporting T. Gayet with a PhD fellowship.

References

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

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

Barrett RDH, Schluter D (2008). Adaptation from standing genetic variation. *Trends Ecol Evol* **23**: 38–44.

Battocchio D, Iacolina L, Canu A, Mori E (2017). How much does it cost to look like a pig in a wild boar group? *Behav Processes* **138**: 123–126.

Beugin M-P, Baubet E, Dufaure De Citres C, Kaerle C, Muselet L, Klein F, *et al.* (2017). 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*). *Conserv Genet Resour*: 1–5.

Canu A, Costa S, Iacolina L, Piatti P, Apollonio M, Scandura M (2014). Are captive wild boar more introgressed than free-ranging wild boar? Two case studies in Italy. *Eur J Wildl Res* **60**: 459–467.

Canu A, Vilaça ST, Iacolina L, Apollonio M, Bertorelle G, Scandura M (2016). Lack of polymorphism at the MC1R wild-type allele and evidence of domestic allele introgression

across European wild boar populations. *Mamm Biol* **81**: 477–479.

Earl DA, VonHoldt BM (2011). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* **4**: 359–361.

Evanno G, Regnaut S, Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* **14**: 2611–2620.

Fang M, Berg F, Ducos A, Andersson L (2006). Mitochondrial haplotypes of European wild boars with 2n = 36 are closely related to those of European domestic pigs with 2n = 38. *Anim Genet* **37**: 459–464.

Feulner PGD, Gratten J, Kijas JW, Visscher PM, Pemberton JM, Slate J (2013). Introgression and the fate of domesticated genes in a wild mammal population. *Mol Ecol* **22**: 4210–4221.

Frantz AC, Massei G, Burke T (2012). Genetic evidence for past hybridisation between domestic pigs and English wild boars. *Conserv Genet* **13**: 1355–1364.

Frantz AC, Zachos FE, Kirschning J, Cellina S, Bertouille S, Mamuris Z, *et al.* (2013). Genetic evidence for introgression between domestic pigs and wild boars (Sus scrofa) in Belgium and Luxembourg: a comparative approach with multiple marker systems. *Biol J Linn Soc* **110**: 104–115.

Fulgione D, Rippa D, Buglione M, Trapanese M, Petrelli S, Maselli V (2016). Unexpected but welcome. Artificially selected traits may increase fitness in wild boar. *Evol Appl* **9**: 769–776.

Gayet T, Devillard S, Gamelon M, Brandt S, Say L, Baubet E (2016). On the evolutionary consequences of increasing litter size with multiple paternity in wild boar (*Sus scrofa scrofa*). *Evolution (N Y)* **70**: 1386–1397.

Giuffra E, Kijas JMH, Amarger V, Carlborg Ö, Jeon J-T, Andersson L (2000). The origin of the domestic pig: independent domestication and subsequent introgression. *Genetics* **154**: 1785–1791.

Godinho R, Llaneza L, Blanco JC, Lopes S, Álvares F, García EJ, *et al.* (2011). Genetic evidence for multiple events of hybridization between wolves and domestic dogs in the Iberian Peninsula. *Mol Ecol* **20**: 5154–5166.

Goedbloed DJ, van Hooft P, Megens H-J, Langenbeck K, Lutz W, Crooijmans RPMA, *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.

Goulding MJ (2001). Possible genetic sources of free-living wild boar (*Sus scrofa*) in southern England. *Mamm Rev* **31**: 245–248.

Huxel GR (1999). Rapid displacement of native species by invasive species: effects of hybridization. *Biol Conserv* **89**: 143–152.

Iacolina L, Brajković V, Canu A, Šprem N, Cubric-Curik V, Fontanesi L, *et al.* (2016). Novel Y-chromosome short tandem repeats in *Sus scrofa* and their variation in European wild boar and domestic pig populations. *Anim Genet* **47**: 682–690.

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

Jombart T (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**: 1403–1405.

Keenan K, McGinnity P, Cross TF, Crozier WW, Prodöhl PA (2013). diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods Ecol Evol* **4**: 782–788.

Koutsogiannouli EA, Moutou KA, Sarafidou T, Stamatis C, Mamuris Z (2010). Detection of hybrids between wild boars (*Sus scrofa scrofa*) and domestic pigs (*Sus scrofa f. domestica*) in Greece, using the PCR-RFLP method on melanocortin-1 receptor (MC1R) mutations. *Mamm Biol für Säugetierkd* **75**: 69–73.

Laikre L, Schwartz MK, Waples RS, Ryman N (2010). Compromising genetic diversity in the wild: unmonitored large-scale release of plants and animals. *Trends Ecol Evol* **25**: 520–529.

Liu N, Chen L, Wang S, Oh C, Zhao H (2005). Comparison of single-nucleotide polymorphisms and microsatellites in inference of population structure. *BMC Genet* **6**: 1–5.

Mallet J (2005). Hybridization as an invasion of the genome. *Trends Ecol Evol* **20**: 229–237.

Massei G, Kindberg J, Licoppe A, Gačić D, Šprem N, Kamler J, *et al.* (2014). Wild boar populations up, numbers of hunters down? A review of trends and implications for Europe. *Pest Manag Sci* **12**: 492–500.

McFee AF, Banner MW, Rary JM (1966). Variation in chromosome number among European wild pigs. *Cytogenet Genome Res* **5**: 75–81.

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

Pinet J-M (2005). *L'élevage du sanglier de race pure* (Adeprina, Ed.).

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. *Mol Ecol Resour* **16**: 608–627.

R Core Team (2017). R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria.

Rauw WM, Kanis E, Noordhuizen-Stassen EN, Grommers FJ (1998). Undesirable side effects of selection for high production efficiency in farm animals: a review. *Livest Prod Sci* **56**: 15–33.

Rhymer JM, Simberloff D (1996). Extinction by hybridization and introgression. *Annu Rev Ecol Syst* **27**: 83–109.

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

Ryder OA (1986). Species conservation and systematics: the dilemma of subspecies. *Trends Ecol Evol* **1**: 9–10.

Scandura M, Iacolina L, Apollonio M (2011). Genetic diversity in the European wild boar *Sus scrofa*: phylogeography, population structure and wild x domestic hybridization. *Mamm*

Rev **41**: 125–137.

Scandura M, Iacolina L, Crestanello B, Pecchioli E, Di Benedetto MF, Russo V, *et al.* (2008). Ancient vs. recent processes as factors shaping the genetic variation of the European wild boar: are the effects of the last glaciation still detectable? *Mol Ecol* **17**: 1745–1762.

Skewes O, Morales R, Mendoza N, Smulders FJM, Paulsen P (2008). Carcass and meat quality traits of wild boar (*Sus scrofa s.* L.) with 2n = 36 karyotype compared to those of phenotypically similar crossbreeds (2n = 37 and 2n = 38) raised under the same farming conditions 1. Carcass quantity and meat dressing. *Meat Sci* **80**: 1200–1204.

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. *Mol Ecol* **15**: 63–72.

Supplementary Material (SM)

SM1 Figure S1 Origin of sampled phenotypic anomalous wild boars (PAWs) included in the analysis.

SM2 Table S1 Classification based on morphological characteristics from Pinet (2002) to select reproductive wild boars in farm, used in this study to count the number of phenotypic characteristics from the pig displayed by PAWs

SM3 Figure S2 Assignment probabilities to the wild boar group inferred by STRUCTURE analysis obtained by the 10 genotype classes (n=2000 for each class) performed on simulated datasets a) with SNPs and c) with microsatellites.

SM4 Table S2 Significant assignments of individuals from each of the 10 simulated genotypic groups (2000 individuals per group) given by STRUCTURE with a) SNPs and b) microsatellites.

SM5 Figure S3 Assignment probabilities to 10 genotype frequency classes inferred by NEWHYBRIDS analysis performed on domestic pigs, wild boars and PAWs from the national sampling with a) SNPs and b) microsatellites.

SM6 Figure S4 Assignment probabilities of the 10 genotypic groups to their own group (n=2000 for each class) simulated with a) SNPs and b) microsatellites.

SM7 Figure S5 Assignment probabilities of sampled domestic pigs, wild boars and PAWs to the wild boar genetic group obtained with STRUCTURE a) with SNPs and b) with microsatellites and obtained with NEWHYBRIDS with c) SNPs and d) microsatellites.

SM8 Table S3 Significant assignments of individuals from each of the 10 simulated genotypic groups (2000 individuals per group) given by newhybrids with a) SNPs and b) microsatellites.

Supplementary Material 1

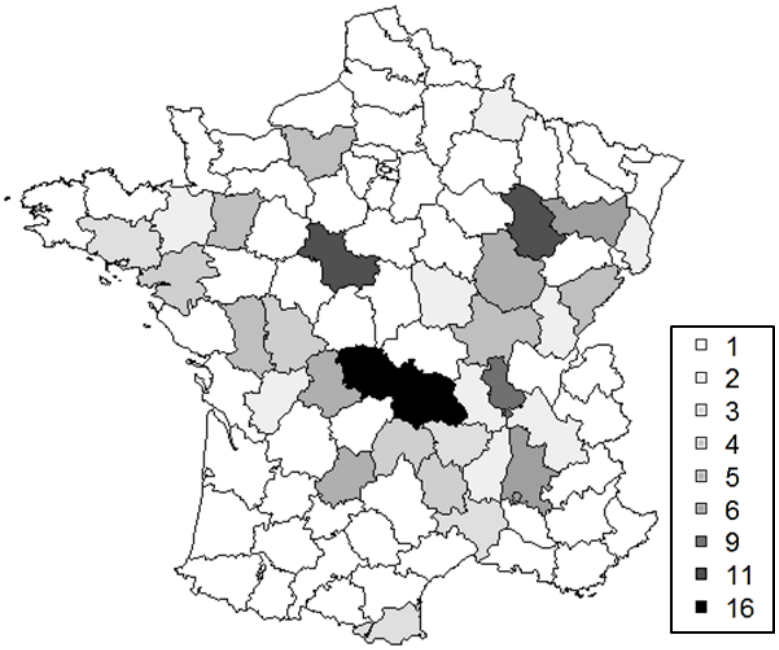


Figure S1 Origin of sampled phenotypic anomalous wild boars (PAWs) included in the analysis. The intensity of the color of the area is proportional to the number of sampled individuals.

Supplementary Material 2

Table S1: Classification based on morphological characteristics from Pinet (2002) to select reproductive wild boars in farm, used in this study to count the number of phenotypic characteristics from the pig displayed by PAWs

Wild boar phenotype	Phenotype to classify as PAW (hybrid)
Narrow head, straight profile	Large head, concave profile
Narrow, straight, long snout	Short snout
Black snout	Spotted snout
Pointed and erected ears	Large, slightly drooping ears
Straight tail	More or less curled tail
No white hair except near jowls	White spotted coat and fair toe horn
Narrow rump and three angled hump	Straight back
Stripped piglet	Piglet without strip

Supplementary Material 3

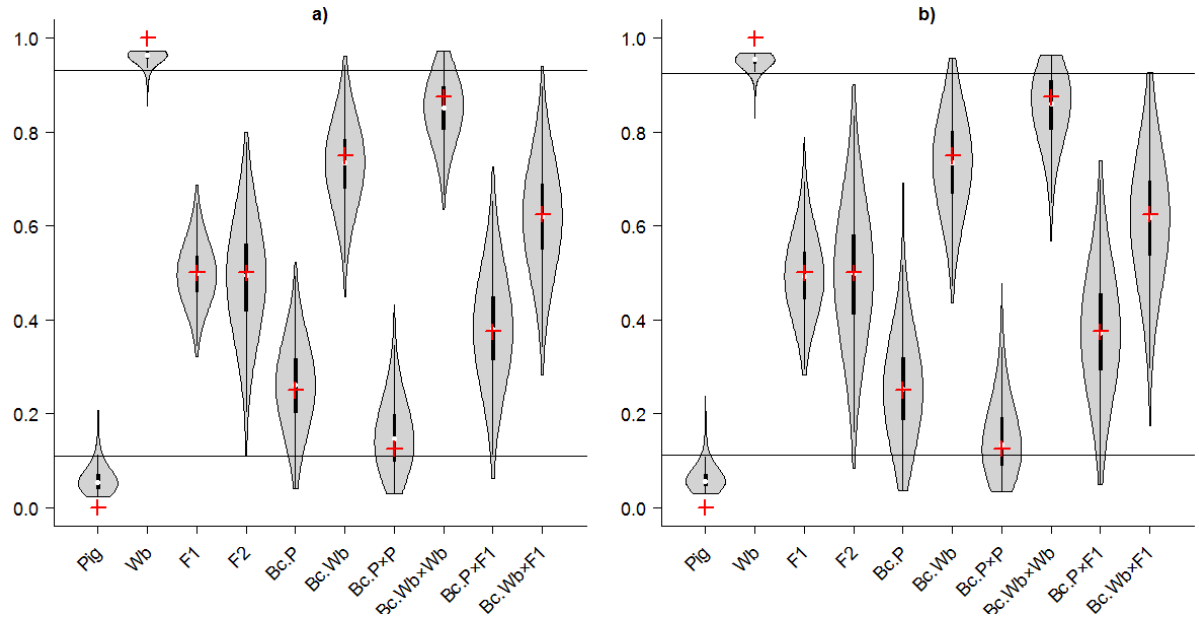


Figure S2 Assignment probabilities to the wild boar group inferred by STRUCTURE analysis obtained by the 10 genotype classes (n=2000 for each class) performed on simulated datasets a) with SNPs and c) with microsatellites. Wb stands for wild boar, P for pig, Bc for backcross (e.g. Bc.P are the backcrosses between F1 and Pig), and, × shows a cross between flanking genotypic groups. The width shows the density of points. The white points show the median, the thick black vertical lines the central quartiles and the fine vertical lines 1.5 the inter-quartile space. The red + indicates the expected $q_{i\text{-}WB}$ value for each genotypic class. The horizontal lines represent the threshold values determined as the 95th percentile of values obtained by simulated wild boar and pig ($q_{WB\text{-}WB\text{-}95\%}$ = 0.9322 and $q_{Pig\text{-}WB\text{-}95\%}$ = 0.113 with SNPs and $q_{WB\text{-}WB\text{-}95\%}$ = 0.928 and $q_{Pig\text{-}WB\text{-}95\%}$ = 0.107 with microsatellites for wild boar and pig respectively, see main text).

Supplementary Material 4

Table S2 Significant assignments of individuals from each of the 10 simulated genotypic groups (2000 individuals per group) given by STRUCTURE with a) SNPs and b) microsatellites. Wb stands for wild boar, P for pig, Bc for backcross and × shows a cross between flanking genotypic groups. Accuracies of the identification of each genotypic groups calculated with the method described by Vähä and Primmer (2006) are also given. Note that the sums in row are not equal to 2000 as individuals might have assignment probabilities to the different genotypic groups higher than corresponding thresholds for several genotypic classes leading thus for these individuals to be assigned in several groups.

a)

		Genotypic group assigned									
		Pig	Wb	F1	F2	Bc.P	Bc.Wb	Bc.P ×P	Bc.Wb ×Wb	Bc.P ×F1	Bc.Wb ×F1
Genotypic group simulated	Pig	1899	0	0	0	137	0	1305	0	4	0
	Wb	0	1900	0	0	0	3	0	937	0	0
	F1	0	0	1900	2000	147	164	0	0	1859	1862
	F2	0	0	1423	1899	502	400	93	29	1620	1507
	Bc.P	70	0	106	758	1900	0	1433	0	1609	57
	Bc.Wb	0	8	140	552	0	1900	0	1232	54	1649
	Bc.P×P	611	0	2	86	1460	0	1899	0	577	1
	Bc.Wb×Wb	0	205	0	27	0	1280	0	1900	0	502
	Bc.P×F1	7	0	872	1655	1287	52	462	2	1897	718
	Bc.Wb×F1	0	1	905	1485	54	1320	3	369	705	1900
	Accuracy (%)	73.4	89.9	35.5	22.4	34.6	37.1	36.6	42.5	22.8	23.2

b)

		Genotypic group assigned									
		Pig	Wb	F1	F2	Bc.P	Bc.Wb	Bc.P ×P	Bc.Wb ×Wb	Bc.P ×F1	Bc.Wb ×F1
Genotypic group simulated	Pig	1900	0	0	0	323	0	1529	0	40	0
	Wb	0	1900	0	0	0	39	0	1363	0	2
	F1	0	0	1900	1999	580	528	19	19	1875	1892
	F2	4	0	1480	1900	715	758	187	150	1642	1614
	Bc.P	138	0	287	1091	1900	10	1524	1	1753	207
	Bc.Wb	0	29	386	1151	4	1900	0	1431	224	1732
	Bc.P×P	760	0	21	231	1592	0	1900	0	962	12
	Bc.Wb×Wb	0	347	23	207	0	1493	0	1896	7	801
	Bc.P×F1	23	0	1069	1739	1473	160	662	13	1900	949
	Bc.Wb×F1	0	1	1180	1757	187	1481	14	591	975	1899
	Accuracy (%)	67.3	83.4	29.9	18.9	28.1	29.8	32.6	34.7	20.3	20.9

Supplementary Material 5

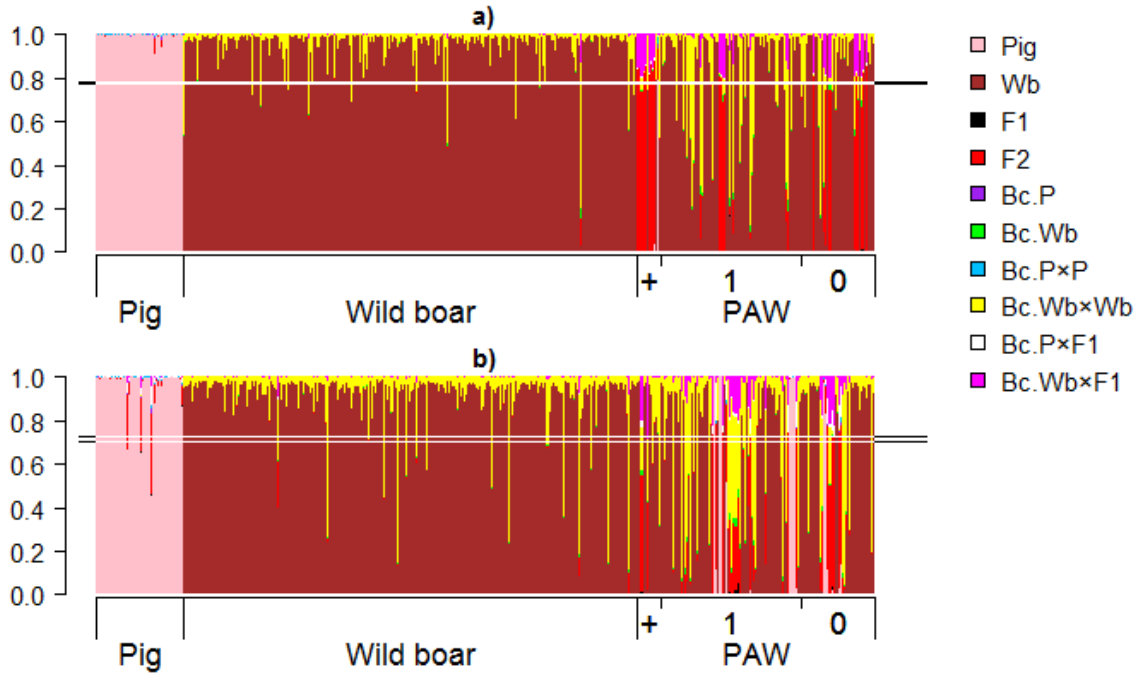


Figure S3 Assignment probabilities to 10 genotype frequency classes inferred by NEWHYBRIDS analysis performed on domestic pigs (n = 50), wild boars (n = 261) and PAWs from the national sampling campaign (separated by the number of phenotypic traits associated with hybridization recorded by the hunters: more than one n_{+}=13, one trait n_{1}=81 and no trait reported n_{0}=42) with a) SNPs and b) microsatellites. Each individual is represented by a vertical bar divided in colors whose length is proportional to the assignment probability to each class.

Supplementary Material 6

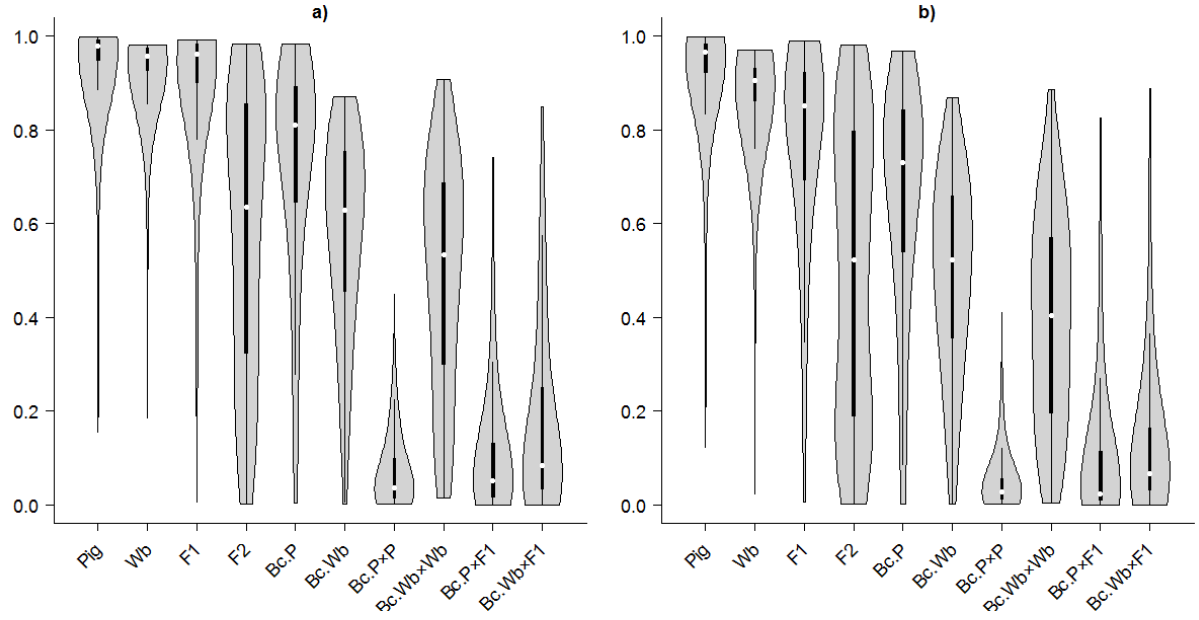


Figure S4 Assignment probabilities of the 10 genotypic groups to their own group (n=2000 for each class) simulated with a) SNPs and b) microsatellites. Wb stands for wild boar, P for pig, Bc for backcross and × shows a cross between flanking genotypic groups. The width shows the density of points. The white points show the median, the thick black vertical lines the central quartiles and the fine vertical lines 1.5 the inter-quartile space.

Supplementary Material 7

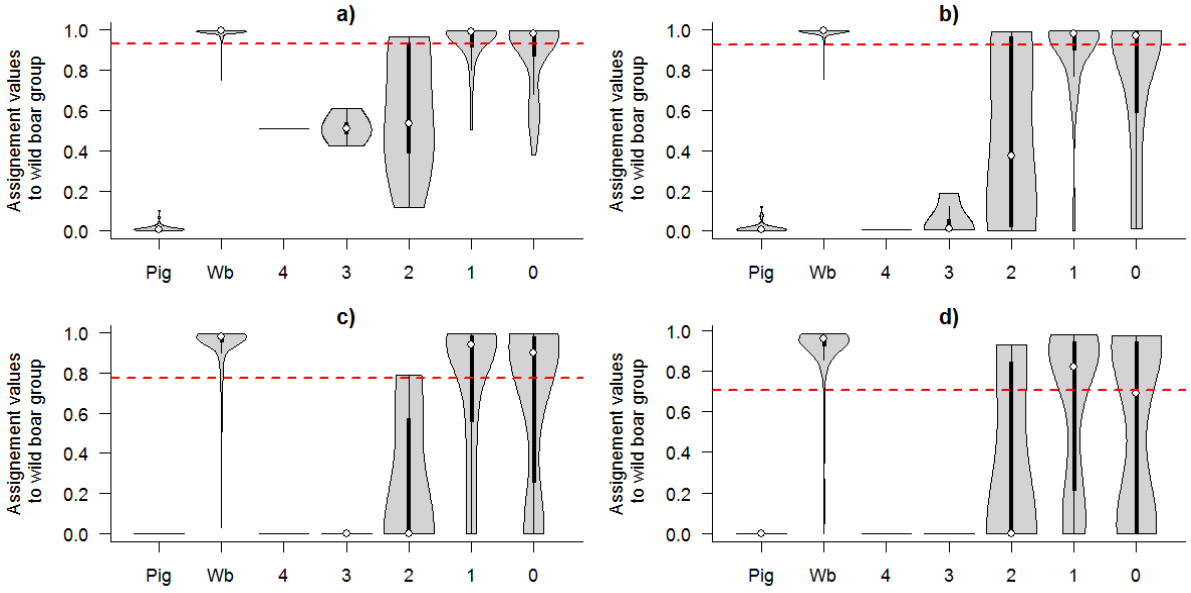


Figure S5 Assignment probabilities of sampled domestic pigs (n = 50), wild boars (n = 261) and PAWs (separated by number of phenotypic traits indicating hybridization n_4=1, n_3=4, n_2=8, n_1=81 and n_0=42) to the wild boar genetic group obtained with STRUCTURE a) with SNPs and b) with microsatellites and obtained with newhybrids with c) SNPs and d) microsatellites. The red horizontal dashed lines represent the threshold values estimated for each analysis and each marker with simulation under which an individuals can no longer be considered a wild boar.

Supplementary Material 8

Table S3 Significant assignments of individuals from each of the 10 simulated genotypic groups (2000 individuals per group) given by newhybrids with a) SNPs and b) microsatellites. Wb stands for wild boar, P for pig, Bc for backcross and × shows a cross between flanking genotypic groups. Accuracies of the identification of each genotypic groups calculated with the method described by Vähä and Primmer (2006) are also given. Note that the sums in row are not equal to 2000 as individuals might have assignment probabilities to the different genotypic groups higher than corresponding thresholds for several genotypic classes leading thus for these individuals to be assigned in several groups.

a)

		Genotypic group assigned									
		Pig	Wb	F1	F2	Bc.P	Bc.Wb	Bc.P ×P	Bc.Wb ×Wb	Bc.P ×F1	Bc.Wb ×F1
Genotypic group simulated	Pig	1900	0	0	0	68	0	1495	0	93	0
	Wb	0	1900	0	0	0	5	0	714	0	2
	F1	0	0	1900	290	33	70	2	0	459	409
	F2	1	0	49	1900	437	176	211	38	1763	1587
	Bc.P	58	0	12	669	1900	0	1623	0	1719	130
	Bc.Wb	0	6	39	706	0	1900	0	1379	213	1693
	Bc.P×P	576	0	0	124	1342	0	1900	0	1063	10
	Bc.Wb×Wb	0	174	0	71	0	1311	0	1900	4	834
	Bc.P×F1	6	0	23	1685	1187	30	747	2	1900	923
	Bc.Wb×F1	0	1	42	1689	64	812	17	419	1123	1899
	Accuracy (%)	74.77	91.3	92.01	26.63	37.77	44.14	31.69	42.68	22.79	25.36

b)

		Genotypic group assigned									
		Pig	Wb	F1	F2	Bc.P	Bc.Wb	Bc.P ×P	Bc.Wb ×Wb	Bc.P ×F1	Bc.Wb ×F1
Genotypic group simulated	Pig	1900	0	0	7	203	0	1606	0	175	0
	Wb	0	1900	0	0	0	33	0	1548	0	29
	F1	0	0	1900	1106	258	286	34	17	944	873
	F2	2	0	146	1900	631	443	276	148	1718	1685
	Bc.P	88	0	68	1233	1900	5	1574	0	1783	303
	Bc.Wb	0	25	161	1279	5	1900	0	1445	470	1792
	Bc.P×P	594	0	3	430	1540	0	1900	0	1197	49
	Bc.Wb×Wb	0	300	9	382	0	1550	0	1900	76	1128
	Bc.P×F1	19	0	123	1775	1321	93	797	16	1900	1117
	Bc.Wb×F1	0	0	159	1792	170	1050	35	557	1262	1899
	Accuracy (%)	72.99	85.39	73.96	19.18	31.52	35.45	30.54	33.74	19.95	21.4

Received: 6 October 2017 | Accepted: 10 December 2017

DOI: 10.1111/2041-210X.12968

RESEARCH ARTICLE

Methods in Ecology and Evolution BRITISH ECOLOGICAL SOCIETY

A fast likelihood solution to the genetic clustering problem

Marie-Pauline Beugin[1,2] | Thibault Gayet[1,3] | Dominique Pontier[1] | Sébastien Devillard[1] | Thibaut Jombart[4]

[1]Univ Lyon, Laboratoire de Biométrie et Biologie Evolutive, CNRS, Université Claude Bernard Lyon 1, Villeurbanne, France

[2]ANTAGENE, Animal Genomics Laboratory, La Tour de Salvagny, France

[3]Office National de la Chasse et de la Faune Sauvage, Unité Cervidés Sangliers, Montfort, Birieux, France

[4]Department of Infectious Disease Epidemiology, School of Public Health, MRC Centre for Outbreak Analysis and Modelling, Imperial College London, London, UK

Correspondence
Thibaut Jombart
Email: thibautjombart@gmail.com
and
Marie-Pauline Beugin
Email: marie.pauline.beugin@gmail.com

Funding information
Medical Research Council Centre for Outbreak Analysis and Modelling; National Institute for Health Research—Health Protection Research Unit for Modelling Methodology; LabEx ECOFECT, Grant/Award Number: ANR-11-LABX-0048

Handling Editor: Thomas Hansen

Abstract

1. 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.
2. 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 optimisation, using the Expectation-Maximisation (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.
3. 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.
4. *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 co-dominant 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
EM algorithm, genetic assignment, genetic clustering, hybridisation, microsatellites, population membership, relative performances, SNP

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 (Corander, Waldmann, & Sillanpää, 2003; Falush, Stephens, & Pritchard, 2003; Pritchard, Stephens, & Donnelly, 2000), the classification of isolates into distinct lineages in microbiology (Feil, Li, Aanensen, Hanage, & Spratt, 2004; Maiden et al., 1998), the investigation of social or ecological units in molecular ecology (Jombart, Devillard, & Balloux, 2010; Sugg, Chesser, Stephen Dobson, & Hoogland, 1996), and the identification of various types of hybrids in conservation genetics (Allendorf, Leary, Spruell, & Wenburg, 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 (Anderson & Thompson, 2002; Corander et al., 2003; Falush et al., 2003; Jombart et al., 2010; Pritchard et al., 2000).

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 (Anderson & Thompson, 2002; Corander et al., 2003; Dupanloup, Schneider, & Excoffier, 2002; Falush et al., 2003; Pritchard et al., 2000), 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 et al., 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 et al., 2002) or Markov Chain Monte Carlo (MCMC) techniques (Corander et al., 2003; Falush et al., 2003; Pritchard et al., 2000; Vähä & Primmer, 2006). While more efficient implementations have been developed (Alexander, Novembre, & Lange, 2009; Raj, Stephens, & Pritchard, 2014; Tang, Peng, Wang, & Risch, 2005), 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 et al., 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 et al., 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 et al., 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 the 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 et al., 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 et al., 2010). Our method can also be used for identifying various types of hybrids between two parental populations. Besides, being an 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 et al., 2000; Falush et al., 2003), BAPS, ADEGENET's *find.cluster* (Jombart et al., 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. Here, *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; Jombart et al., 2017; Kamvar, Tabima, & Grünwald, 2014; Paradis, 2010; Popescu, Huber, & Paradis, 2012).

2 | MATERIALS AND METHODS

2.1 | Rationale of *snapclust*

2.1.1 | Model likelihood

We consider a dataset of allelic profiles $x = \{x_{i,j}\}$ where i indexes individuals ($i = 1, ..., N$) and j indexes loci ($j = 1, ..., J$), so that $\boldsymbol{x}_{i,j}$ 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 $\boldsymbol{g} = \{g(i)\}$, where $g(i)$ defines the group of individual i, with groups indexed by $k = 1, ..., K$. Under the HW model, this likelihood is defined as:

$$p(\boldsymbol{x}_{i,j}|\boldsymbol{f}_{g(i),j}, \Pi) = M(\boldsymbol{x}_{i,j}, \boldsymbol{f}_{g(i),j}, \pi)$$

where M is the probability mass function of the multinomial distribution, $\boldsymbol{f}_{g(i),j}$ is the vector of allele frequencies in group $g(i)$ at locus j, and π 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 the following:

$$p(\boldsymbol{x}_i|\boldsymbol{f}_{g(i)}, \pi) = \Pi_j p(\boldsymbol{x}_{i,j}|\boldsymbol{f}_{g(i),j}, \pi)$$

where $\boldsymbol{f}_{g(i)} = \{f_{g(i),1}, ..., f_{g(i),J}\}$ and $\boldsymbol{x}_i = \{x_{i,1}, ..., x_{i,J}\}$. If we further assume independence of individuals conditional on their group memberships, the general likelihood is given by the following:

$$p(\boldsymbol{x}|\boldsymbol{f}, \boldsymbol{g}, \pi) = \Pi_i p(\boldsymbol{x}_i|\boldsymbol{f}_{g(i)}, \pi)$$

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

$$LL(\boldsymbol{g}) = \sum_i \sum_j M(\boldsymbol{x}_{i,j}, \boldsymbol{f}_{g(i),j}, \pi)$$

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 π 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:

$$p(g(i) = k) = p(\boldsymbol{x}_i|\boldsymbol{g}(i) = k, \boldsymbol{f}_k, \pi) / \sum_q p(\boldsymbol{x}_i|g(i) = q, \boldsymbol{f}_q, \pi)$$

2.1.2 | 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 that correspond to $w = 0.25$ and $w = 0.125$ respectively. The likelihood of a hybrid is defined as before, but using the allele frequencies mixture as follows:

$$p(\boldsymbol{x}_i|\boldsymbol{g}(i) = H, \boldsymbol{f}_A, \boldsymbol{f}_B, w, \pi) = \Pi_j p(\boldsymbol{x}_{i,j}|\boldsymbol{f}_{H,j}, \pi)$$

2.1.3 | Optimisation procedure

Here, *snapclust* achieves fast likelihood maximisation using the EM algorithm (Dempster et al., 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 *snapclust* maximises a log-likelihood rather than between-group distances (Jombart et al., 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, that is, the mean frequencies of alleles in individuals of a given group. The algorithm, adapted from the use of EM for maximising 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, that is, is less than an arbitrary threshold (set to 10^{-10} by default).

2.1.4 | 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 suboptimal 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 et al., 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 most datasets. The three methods are available in the implementation of the algorithm, as well as any other user-defined initial clusters.

2.1.5 | 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, 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:

$$\mathrm{AIC} = -2L' + 2(K(P-J))$$

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):

$$\mathrm{AIC}_c = -2L' + 2(K(P-J)N)/(N-KP+KJ-1)))$$

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

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

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):

$$\mathrm{KIC} = -2L' + 3(K(P-J)+1)$$

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 et al., 2010).

2.1.6 | 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.

2.2 | Simulations

2.2.1 | Simulated datasets without hybrids

The datasets were simulated using QUANTINEMO (Neuenschwander, Hospital, Guillaume, & Goudet, 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

TABLE 1 Parameters used in the simulations using the computer program QUANTINEMO

Parameter	Value	Parameter	Value
Generations	10,000	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

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. To avoid prohibitive computational times, we allowed the number of populations and the dispersal model to vary randomly across replicate, rather than adding systematically new combinations of parameters to the pool of data to simulate. The number of individuals per population was fixed to 100.

2.2.2 | 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 × P2), first-generation backcrosses (BC1: F1 × P1 and F1 × P2), and second-generation backcrosses (BC2: (F1 × P1) × P1 and (F1 × P2) × 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.

2.2.3 | 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 (Falush et al., 2003; Hubisz, Falush, Stephens, & Pritchard, 2009; Pritchard et al., 2000), BAPS 5.4 (Cheng, Connor, Sirén, Aanensen, & Corander, 2013; Corander et al., 2003; Tang, Hanage, Fraser, & Corander, 2009), and ADEGENET's `find.clusters` (Jombart et al., 2010). The *snapclust* analysis was carried out using default parameters (group assignment initialisation 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 where 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 preferred a linear regression on these rates to a binomial generalized linear model (GLM) for practical reasons. The calculation of *TPR* and *TNR* being based on all pairwise comparisons of individuals within simulated datasets, a binomial GLM would have required millions of observations to be included, leading to prohibitive computations. Instead, 360 TPR and TNR values were analysed, that is, one per simulation.

We tested for the effects of the number of loci, the dispersal model, the overall F_{st} 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 1), as well as between the number of loci and the F_{st}. 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 and thus across all conditions of differentiation, number of loci and populations and dispersal model, we compared *TPR* and *TNR* predicted by the respective models by transforming predicted logit rates back to their original scale.

2.2.4 | Analyses of simulated datasets with hybrids

The clustering of individuals in the 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, F_{st}, 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 F_{st}. 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%.

2.3 | 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, Jombart, Dufour, & Moazami-Goudarzi, 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 hybridisation 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 et al., 2010), employing cross-validation to determine the optimal number of principal components to retain. The R script required to reproduce the simulated data and run the analyses are provided in Supplementary Material.

3 | RESULTS

3.1 | 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 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 S1).

Multivariate linear regression captured a large fraction of the variation in logit(*TPR*) values (Adjusted R^2: 61%, Table S2). Predicted *TPR* and *TNR* we pooled by methods across all simulations to compare overall performances of the different approaches. The results showed marginal variations in performances of the methods, with mean predicted *TPR* varying from 96.7% (IQR: 95.4%–99.1%) for *find.clusters* to 99.0% (IQR: 98.5%–99.7%) for BAPS, with 98.0% (IQR: 97.1%–99.5%) for STRUCTURE and 98.1% (IQR:97.3%–99.6%) for *snapclust*. Similar results were observed for *TNR* values, although the model explained a smaller fraction of the variance (Adjusted R^2: 44%, Table S3). Predicted *TNR* values were very high for all four methods: 96.7% (IQR: 95.4%–99.1%) for *find.*clusters, 98.9% (IQR: 98.6%–99.7%) for BAPS, 98.1% (IQR: 97.1%–99.5%) for STRUCTURE, and 98.1% for *snapclust* (IQR: 97.3%–99.5%).

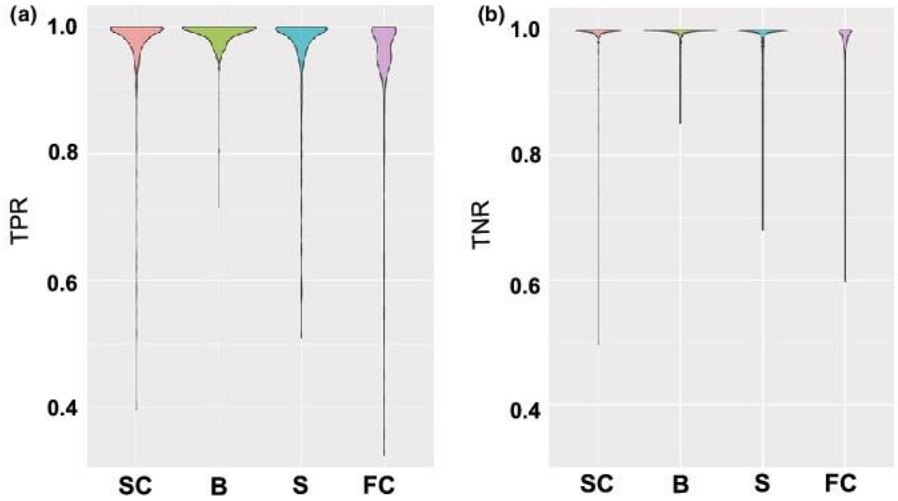


FIGURE 1 Comparison of the various methods on simulated genetic clusters. *Notes*: This figure shows the distribution of (a) the true positive rates (*TPR*) and (b) true negative rates (*TNR*) obtained over all the 360 simulations for the four different methods: *snapclust* (SC), BAPS (B), STRUCTURE (S) and *find.clusters* (FC) for the clustering of individuals in absence of hybrids. This width of the enveloppes reflects the density of points

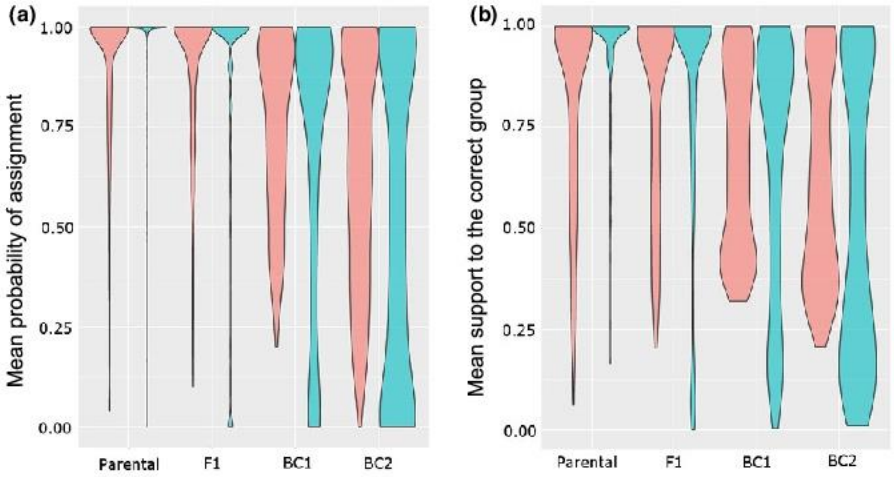


FIGURE 2 Comparison of *snapclust* (red) and NEWHYBRIDS (blue) for the identification of hybrids using simulated data. *Notes*: This figure shows the distributions of (a) the mean proportion of correct group assignment and (b) the support (i.e. group membership probability) for the true class across 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 enveloppes reflects the density of points

Other parameters impacted the performances of the different methods in terms of *TPR* and *TNR* (Tables S2 and S3). In both cases, increased number of loci and larger F_{st} generally improved *TPR* and *TNR* values, although a saturation effect was observed so that the effects of large numbers of loci and stronger F_{st} effectively cancelled out. For instance, for F_{st} of approximately 0.1, the increase in the number of loci from 50 to 500 allowed to increase the *TPR* from 0.88 to 0.97 across all methods, while the same increase in the number of loci for F_{st} of approximately 0.6 yielded *TPR* values ranging between 0.99 and 1. In addition, increasing the number of populations led to improved *TNR* (Table S3).

3.2 | 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 2, Tables S4–S5). The final model of the proportion of correct assignment explained most of the variation in the results (adjusted R^2 = 63.78%). Increased number of loci (t = 23.32; $p = 2.74 \times 10^{-110}$) and stronger F_{st} (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 hybridisation (Figure 2, Table S4–S5). 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 2, Table S5), 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 0.54 s.

3.3 | 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 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 (Figure S1). 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 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 S2), which showed that individuals were structured along a cline of genetic differentiation between the two parental populations, with considerable overlap between "neighbouring" groups (Figure 3c).

4 | DISCUSSION

We have introduced "*snapclust*," a new genetic clustering method which achieves fast maximum likelihood identification of the optimal

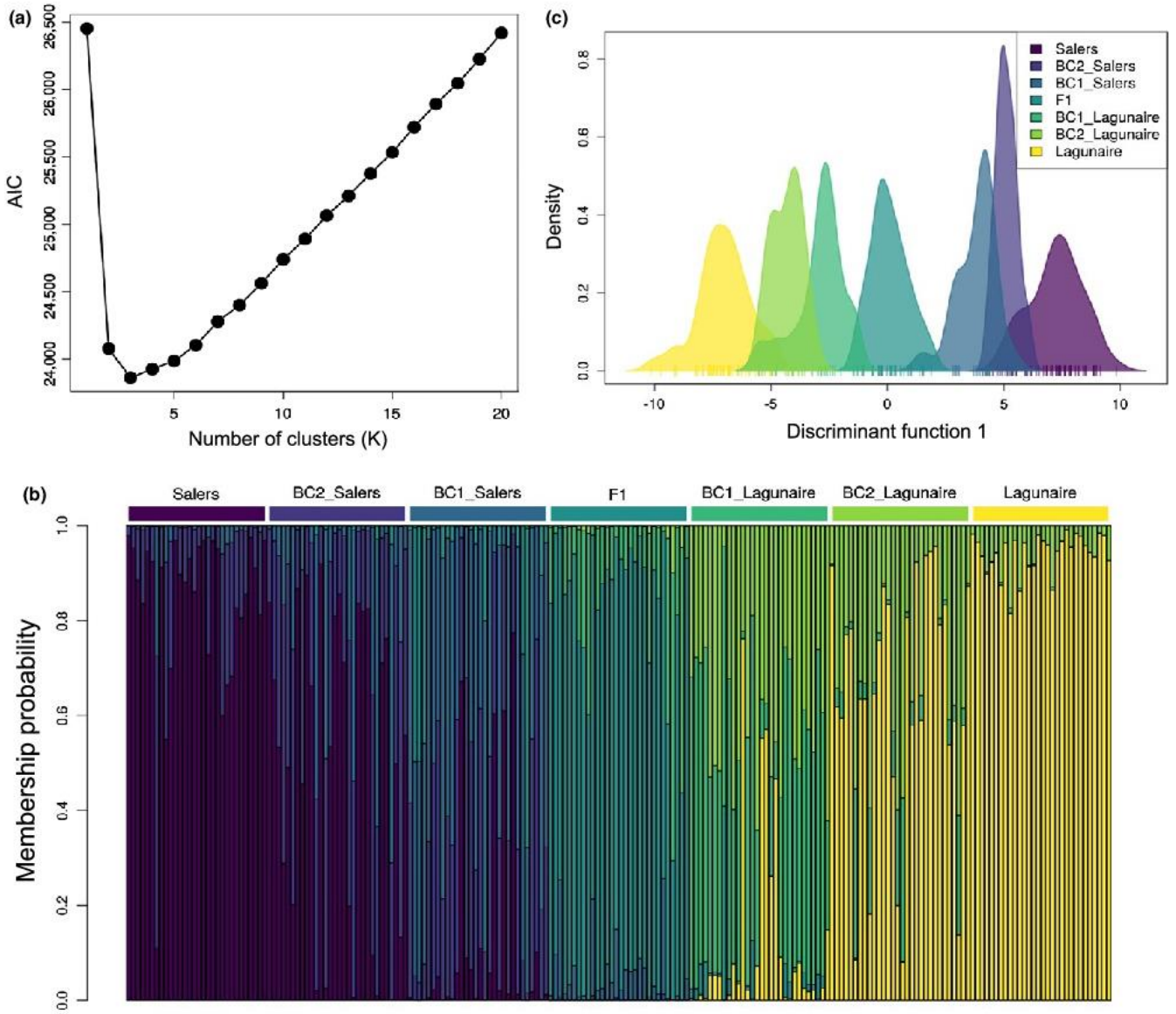


FIGURE 3 Illustration of *snapclust* using simulated hybrids from cattle breed microsatellite data. *Notes*: (a) Representation of the Akaike Criterion value according to the number of populations (*K*) considered. (b) Representation of the individual probability of assignment obtained with the function *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

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 small 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 backcrosses, while being again tremendously more computer efficient. The combination of likelihood estimation and EM algorithm for cluster detection is not new (Fraley & 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 (Fraley & Raftery, 1998). As such, we believe the kind of approach introduced here offers exciting prospects for extending previous efforts to make model-based genetic clustering methods more computer-efficient (Alexander et al., 2009; Raj et al., 2014; Tang et al., 2005).

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 to run a large number of times (e.g. 100) on datasets obtained by random resampling (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 et al., 2014). It would be relatively easy to apply in the case of hybridisation 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, optimisation 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 this 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 (Vähä & 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 co-dominant 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 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, Devillard, Dufour, & Pontier, 2008), or investigating genetic clusters (Jombart et al., 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 (Archer, Adams, & Schneiders, 2017; Goudet, 2005; Kamvar et al., 2014; Paradis, 2010) and phylogenetics (Bortolussi, Durand, Blum, & François, 2006; Jombart et al., 2017; Popescu et al., 2012; Revell, 2012; Schliep, 2011) will enhance its usefulness for the community.

ACKNOWLEDGEMENTS

T.J. 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. M.-P.B. and D.P. are supported by the LabEx ECOFECT (ANR-11-LABX-0048). We thank Jérome 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.

DATA ACCESSIBILITY

This work does not involve empirical datasets. The dataset microbov, used in our illustration, is distributed with the R package ADEGENET: https://cran.r-project.org/web/packages/adegenet/index.html

ORCID


Thibaut Jombart http://orcid.org/0000-0003-2226-8692

REFERENCES

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In E. Parzen, K. Tanabe, & G. Kitagawa (Eds.), *Selected papers of Hirotugu Akaike* (pp. 199–213). Springer Series in Statistics. New York, NY: Springer. https://doi.org/10.1007/978-1-4612-1694-0

Akogul, S., & Erisoglu, M. (2016). A comparison of information criteria in clustering based on mixture of multivariate normal distributions. *Mathematical & Computational Applications*, *21*, 34. https://doi.org/10.3390/mca21030034

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*, 1655–1664. https://doi.org/10.1101/gr.094052.109

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

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

Archer, F. I., Adams, P. E., & Schneiders, B. B. (2017). STRATAG: An r package for manipulating, summarizing and analysing population genetic data. *Molecular Ecology Resources*, *17*, 5–11. https://doi.org/10.1111/1755-0998.12559

Bortolussi, N., Durand, E., Blum, M., & François, O. (2006). apTreeshape: Statistical analysis of phylogenetic tree shape. *Bioinformatics*, *22*, 363–364. https://doi.org/10.1093/bioinformatics/bti798

Cavanaugh, J. E. (1999). A large-sample model selection criterion based on Kullback's symmetric divergence. *Statistics & Probability Letters*, *42*, 333–343. https://doi.org/10.1016/S0167-7152(98)00200-4

Cheng, L., Connor, T. R., Sirén, J., Aanensen, D. M., & Corander, J. (2013). Hierarchical and spatially explicit clustering of DNA sequences with BAPS software. *Molecular Biology and Evolution*, *30*, 1224–1228. https://doi.org/10.1093/molbev/mst028

Corander, J., Waldmann, P., & Sillanpää, M. J. (2003). Bayesian analysis of genetic differentiation between populations. *Genetics*, *163*, 367–374.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, *39*, 1–38.

Dupanloup, I., Schneider, S., & Excoffier, L. (2002). A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology*, *11*, 2571–2581. https://doi.org/10.1046/j.1365-294X.2002.01650.x

Excoffier, L., & Heckel, G. (2006). Computer programs for population genetics data analysis: A survival guide. *Nature Reviews Genetics*, *7*, 745–758. https://doi.org/10.1038/nrg1904

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.

Feil, E. J., Li, B. C., Aanensen, D. M., Hanage, W. P., & Spratt, B. G. (2004). eBURST: Inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *Journal of Bacteriology*, *186*, 1518–1530. https://doi.org/10.1128/JB.186.5.1518-1530.2004

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution*, *17*, 368–376. https://doi.org/10.1007/BF01734359

Felsenstein, J. (2004). *Inferring phylogenies*. Sunderland, MA: Sinauer Associates.

Fraley, C., & Raftery, A. E. (1998). How many clusters? Which clustering method? Answers via model-based cluster analysis. *Computer Journal*, *41*, 578–588. https://doi.org/10.1093/comjnl/41.8.578

Fraley, C., & Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, *97*, 611–631. https://doi.org/10.1198/016214502760047131

Goudet, J. (2005). HIERFSTAT: A package for r to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, *5*, 184–186. https://doi.org/10.1111/j.1471-8286.2004.00828.x

Hubisz, M. J., Falush, D., Stephens, M., & Pritchard, J. K. (2009). Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, *9*, 1322–1332. https://doi.org/10.1111/j.1755-0998.2009.02591.x

Jombart, T. (2008). ADEGENET: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*, 1403–1405. https://doi.org/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*, 3070–3071. https://doi.org/10.1093/bioinformatics/btr521

Jombart, T., Archer, F., Schliep, K., Kamvar, Z., Harris, R., Paradis, E., … Lapp, H. (2017). APEX: Phylogenetics with multiple genes. *Molecular Ecology Resources*, *17*, 19–26. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/10.1038/hdy.2008.130

Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). POPPR: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, *2*, e281. https://doi.org/10.7717/peerj.281

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: GSE*, *39*, 545–567. https://doi.org/10.1186/1297-9686-39-5-545

Legendre, P., & Legendre, L. F. J. (2012). *Numerical ecology*. Amsterdam, the Netherlands: Elsevier.

Maiden, M. C., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., Urwin, R., … Spratt, B. G. (1998). Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences of the United States of America*, *95*, 3140–3145. https://doi.org/10.1073/pnas.95.6.3140

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. https://doi.org/10.1093/bioinformatics/btn219

Paradis, E. (2010). PEGAS: An R package for population genetics with an integrated-modular approach. *Bioinformatics*, *26*, 419–420. https://doi.org/10.1093/bioinformatics/btp696

Popescu, A.-A., Huber, K. T., & Paradis, E. (2012). APE 3.0: New tools for distance-based phylogenetics and evolutionary analysis in R. *Bioinformatics*, *28*, 1536–1537. https://doi.org/10.1093/bioinformatics/bts184

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

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Raj, A., Stephens, M., & Pritchard, J. K. (2014). FASTSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics*, *197*, 573–589. https://doi.org/10.1534/genetics.114.164350

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, *66*, 846–850. https://doi.org/10.1080/01621459.1971.10482356

Revell, L. J. (2012). PHYTOOLS: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution/British Ecological Society*, *3*, 217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x

Schliep, K. P. (2011). PHANGORN: Phylogenetic analysis in R. *Bioinformatics*, *27*, 592–593. https://doi.org/10.1093/bioinformatics/btq706

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464. https://doi.org/10.1214/aos/1176344136

Sugg, D. W., Chesser, R. K., Stephen Dobson, F., & Hoogland, J. L. (1996). Population genetics meets behavioral ecology. *Trends in Ecology & Evolution*, *11*, 338–342. https://doi.org/10.1016/0169-5347(96)20050-3

Tang, J., Hanage, W. P., Fraser, C., & Corander, J. (2009). Identifying currents in the gene pool for bacterial populations using an integrative approach. *PLoS Computational Biology*, *5*, e1000455.

Tang, H., Peng, J., Wang, P., & Risch, N. J. (2005). Estimation of individual admixture: Analytical and study design considerations. *Genetic Epidemiology*, *28*, 289–301. https://doi.org/10.1002/(ISSN)1098-2272

Vähä, J.-P., & Primmer, C. R. (2006). Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Molecular Ecology*, *15*, 63–72.

Wright, S. (1951). The genetical structure of populations. *Annals of Eugenics*, *15*, 323–354.

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, Gayet T, Pontier D, Devillard S, Jombart T. A fast likelihood solution to the genetic clustering problem. *Methods Ecol Evol.* 2018;00:1–11. https://doi.org/10.1111/2041-210X.12968

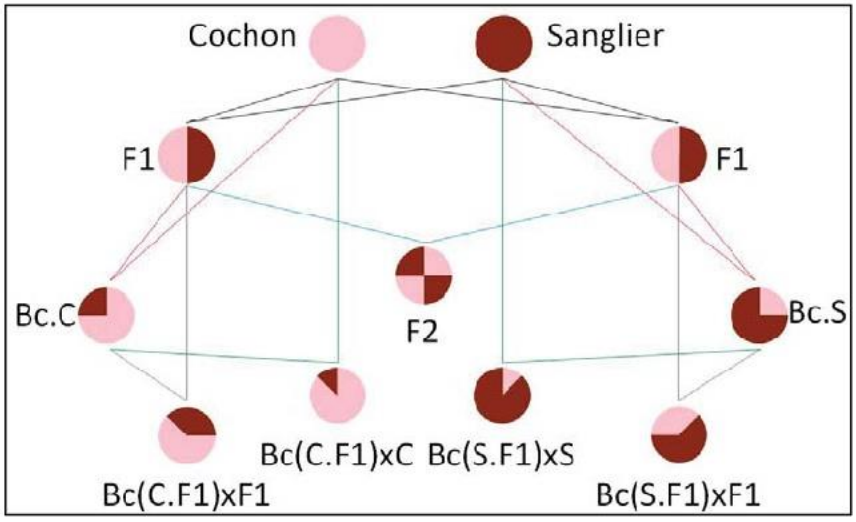
Avancées sur la mise au point d'un outil pour identifier les animaux issus de croisement entre sanglier et cochon

Eric BAUBET [1], Thibault GAYET [123], Ludovic SAY [34], Sébastien DEVILLARD [23], Guillaume QUENEY [5]

([1]) ONCFS– Unité Cervidés-sanglier
([2]) CNRS UMR 5558
([3]) Laboratoire Biométrie et Biologie Evolutive
([4]) Université Claude Bernard Lyon I
([5]) ANTAGENE, Laboratoire de recherche et d'analyses en génomique animale

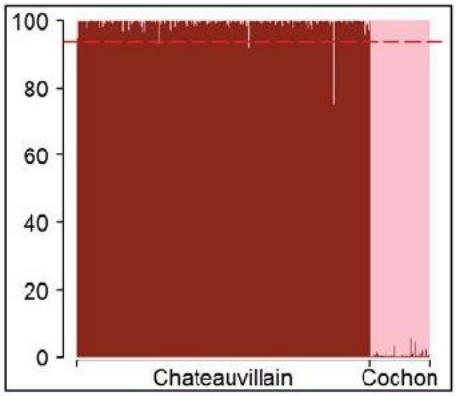
L'objectif du travail engagé par la Direction Recherche et Expertise de l'ONCFS, en partenariat avec le laboratoire ANTAGENE et l'université Claude Bernard Lyon I est d'obtenir un outil d'identification au niveau individuel du degré de mélange d'un individu sanglier lambda pour savoir où il se situerait sur un gradient allant du sanglier « pur » à celui de cochon « pur ». Cette approche est basée sur l'outil SNP (Single Nucleotide Polymorphism) qui est une variation ponctuelle de l'ADN avec 2 formes (allèles) différentes.

L'idée est de sélectionner les allèles qui diffèrent entre les deux populations étudiées, de façon à ce qu'une forme de l'allèle soit fixée (c'est-à-dire toujours la même) chez le porc par exemple, et l'autre forme soit fixée chez le sanglier (ou encore le SNP présente un allèle uniquement chez le sanglier, ou très rare chez le cochon). A l'heure actuelle ce travail repose sur l'analyse de 20 marqueurs SNP (10 spécifiques du sanglier et 10 spécifiques du cochon). Différentes analyses statistiques basées sur des simulations ont montré qu'à partir de ces 20 marqueurs, il est possible de relativement bien distinguer et classer les individus issus des populations parentales « cochons » et « sangliers » ainsi que ce qui pourrait être des individus « hybrides ». Néanmoins, l'outil disponible actuellement n'est pas assez puissant pour déterminer de quels types de croisement résultent ces individus identifiés comme « hybrides ». En effet, il reste une forte imprécision pour identifier si l'hybride est issu d'un croisement de première génération (F1), de seconde génération (F2) ou d'un rétrocroisement (figure ci-contre).



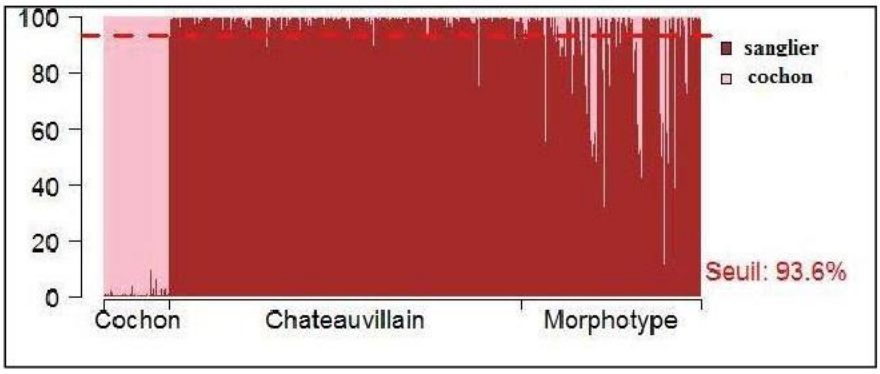
Schématisation théorique des proportions de sanglier « pur » ou cochon attendues dans les génomes au fil des générations et des croisements entre différents types d'individus. F1 = individu de 1er génération issue d'un croisement entre deux souches parentales pures (cochon et sanglier). F2 individu hybride de 2eme génération issu du croisement de 2 individus étant eux même des F1. Bc.C (ou .S) = individu de deuxième génération issu d'un animal F1 et d'une souche parentale (Bc.C = rétrocroisement cochon ou alors Bc.S = rétrocroisement sanglier) etc...

Lors de l'application de cet outil sur des données réelles (échantillons réalisés sur 270 sangliers de la population de Châteauvillain), il a été possible d'identifier un taux de mélange populationnel de 1,15%, (figure ci-contre) ce qui est compatible avec d'autres résultats publiés à l'échelle européenne.



Analyses génotypiques des marqueurs SNP pour les individus issus de la population de Châteauvillain (sanglier) et le lot échantillon de cochon de référence pour cette étude. La ligne pointillée rouge indique le seuil de 93,6% estimé dans l'analyse, pour permettre une discrimination entre un individu considéré comme sanglier « pur » ou un considéré comme hybride. La couleur marron figure la part du génome issue de la composante « sanglier » et celle rose la part issue de celle cochon. 3 individus de la population de sanglier testée (barre rose dépassant vers le bas la ligne de référence) sont ainsi classés en hybride ce qui représente un taux de 1,15%. N sanglier = 261 ; N cochon = 52.

Pour apprécier la qualité de cet outil, il a été décidé d'utiliser des échantillons d'individus avec un phénotype, si possible, montrant des caractéristiques intermédiaires entre porcs et sangliers. Ainsi, ont été collectés par l'intermédiaire du réseau ongulés sauvages, des individus sur deux exercices cynégétiques, pour lesquels des caractéristiques tels qu'un pelage tacheté, des oreilles tombantes, un groin court, un nombre d'allaites et/ou de fœtus anormalement élevés etc... devaient être constatés (un support photographique étant souhaité). Les premières analyses indiquent que plus d'un tiers (35,1%) des individus collectés lors de cette enquête ressortent effectivement avec un classement hybride (figure ci-contre).



Analyses génotypiques des marqueurs SNP pour les individus issus de trois lots (les individus de la population sanglier de Châteauvillain, le lot cochon, et le lot dit « Morphotype », individus collectés via le réseau ongulés sauvages. La ligne pointillée rouge indique le seuil 93,6% estimé dans l'analyse, pour permettre une discrimination entre individu considéré comme sanglier « pur » ou ceux considérés comme hybride. La couleur marron figure la part du génome issue de la composante « sanglier » et en rose la part issue la composante « cochon ». Nsanglier = 256 ; Ncochon = 47, Nmorphotype = 131.

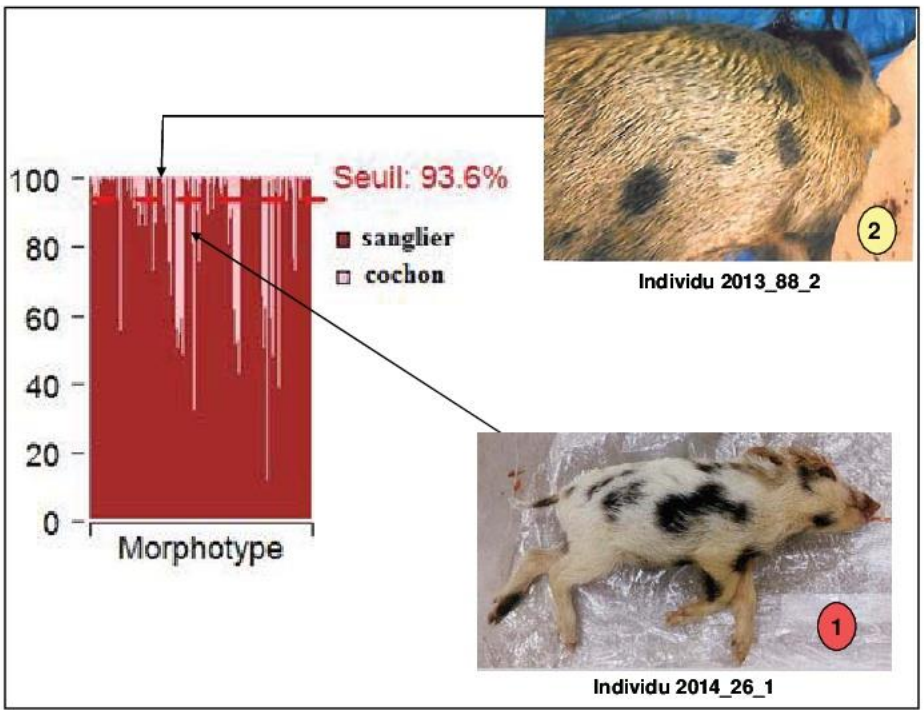
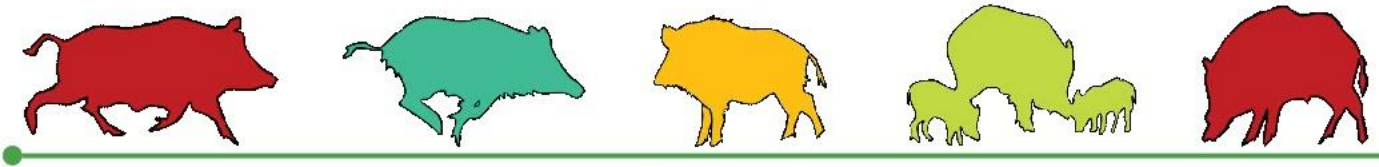


Illustration de deux profils génétiques obtenus en utilisant les 20 marqueurs SNP. En 4-1 : L'individu 2014_26_1 est clairement identifié comme hybride la proportion de génome cochon dépasse largement le seuil de 93,6% retenu pour discriminer entre sanglier et autres (hybrides ou cochon). En revanche, en 4-2 : l'individu 2013_88_2, n'a qu'une très faible proportion de cochon dans son génome et se retrouve donc classé sanglier alors que son pelage présente des taches noires.

Ce premier résultat est encourageant, et insiste à poursuivre l'investigation. En effet, il nous reste encore à croiser l'information disponible issue des clichés photographiques de qualité variable avec le statut génotypique des individus quand cela est possible. En effet, tous les individus n'ont pas forcément fait l'objet d'une prise d'information photographique et/ou dans certains cas celle-ci ne permet pas d'identifier précisément le caractère ayant conduit à l'intégration de l'individu en question dans l'échantillonnage. D'après les premiers éléments en notre possession, le caractère basé uniquement sur l'aspect du pelage ne semble pas suffisant. En effet si certains individus montrent un pelage tacheté et présentent un génotype renvoyant à un individu hybride (figure ci-contre, cas 1), d'autres présentent des caractéristiques de colorations non uniformes alors que l'analyse révèle un génotype identifiant l'animal comme un sanglier (figure ci-contre cas 2).

Nous tenons à remercier chaleureusement les responsables du réseau Ongulés sauvages, Christine Saint-Andrieux et Aurélie Barboiron pour leur participation à cette étude, ainsi que tous les interlocuteurs techniques qui ont dégagé du temps pour collecter et envoyer des échantillons. Un grand merci notamment aux personnels ayant agi avec efficacité sur le terrain. Nous vous informons aussi qu'une restitution individualisée aura lieu pour chacun des départements ayant fourni des données. Le document final est en cours de rédaction.



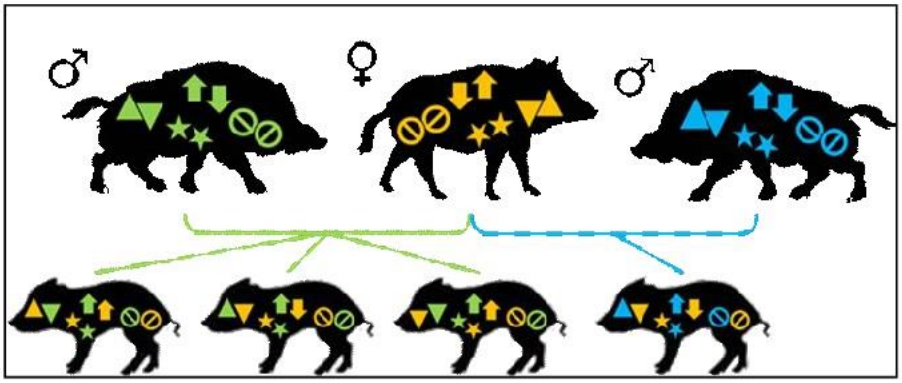
Quel système de reproduction chez le sanglier ?

Thibault GAYET [1,2], Sébastien DEVILLARD [2], Marlène GAMELON [3], Serge BRANDT [1], Ludovic SAY [2], Eric BAUBET [1]

[1] ONCFS Unité Cervidés Sanglier Birieux, [2] Université de Lyon, CNRS, Laboratoire de Biométrie et Biologie Evolutive, [3] Department of Biology, Centre for Biodiversity Dynamics, Norwegian University of Science and Technology.

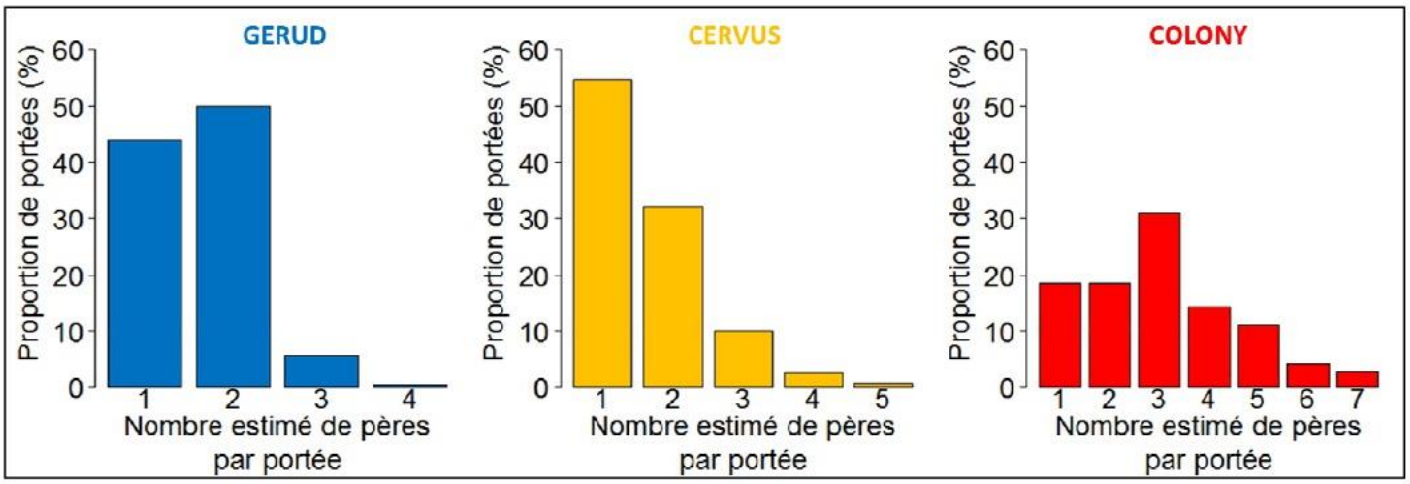
La reproduction du sanglier est généralement décrite comme étant assurée par un seul mâle dominant, qui empêche les autres mâles d'accéder aux femelles d'un même groupe. On parle alors de système de reproduction polygyne. Il est communément admis que ce sont les vieux mâles, les plus gros et les mieux armés qui ont la capacité de contrôler l'accès à un groupe de femelles et de s'accoupler avec elles. Ces conclusions sont principalement basées sur des observations comportementales réalisées par Mauget dans les années 80 dans la population de Chizé (79).
Cependant, dans certaines populations de sangliers très fortement chassées, les vieux mâles sont rares, voire absents. Nous avons étudié comment fonctionnait le système de reproduction dans ce contexte.

Sur le site d'étude de Chateauvillain-Arc-en-Barrois (52) la probabilité d'être abattu chaque année pour un mâle atteint 70% et, par conséquent, très peu d'individus atteignent plus de 3 ans. Pour voir si dans ces conditions le système de reproduction était toujours polygyne, nous avons réalisé des prélèvements sur 217 femelles gestantes tuées à la chasse et sur l'ensemble de leurs fœtus. Les analyses de fragments d'ADN (appelés « microsatellites ») des petits des différentes portées ainsi que de leur mère permettent alors d'estimer le nombre de mâles qui ont engendré chaque portée (cf figure ci-contre).



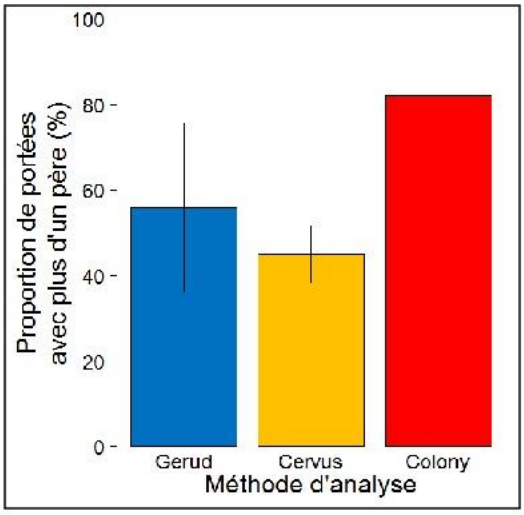
Représentation schématique simplifiée de la transmission de l'information génétique (ADN) des parents vers la descendance. Chaque motif représente une zone de l'ADN (locus), le sens du motif représente différentes formes que peut prendre un locus (allèle). Tous les loci réunis composent le génome d'un individu. Chaque progéniture reçoit la moitié du génome de sa mère et la moitié du génome de son père. Ceci permet d'identifier le nombre de pères ayant participé à une portée sachant que le génome de la mère est connu.

L'estimation du nombre de pères par portée a été réalisée à l'aide de plusieurs logiciels afin de vérifier la cohérence des résultats. En effet, certaines méthodes tendent à sous-estimer le nombre de pères par portée (méthodes GERUD et CERVUS). Il a été possible d'observer entre 1 et 4 pères par portée avec toutes les méthodes, et même jusqu'à 7 avec COLONY. Une moyenne de 1,6 pères par portée est obtenue avec à la fois GERUD et CERVUS, alors qu'elle s'élève à 3,1 avec COLONY, la 3ème méthode utilisée. Contrairement à ce qui est connu des observations comportementales, toutes les femelles ne se reproduisent donc pas avec un seul et unique mâle.



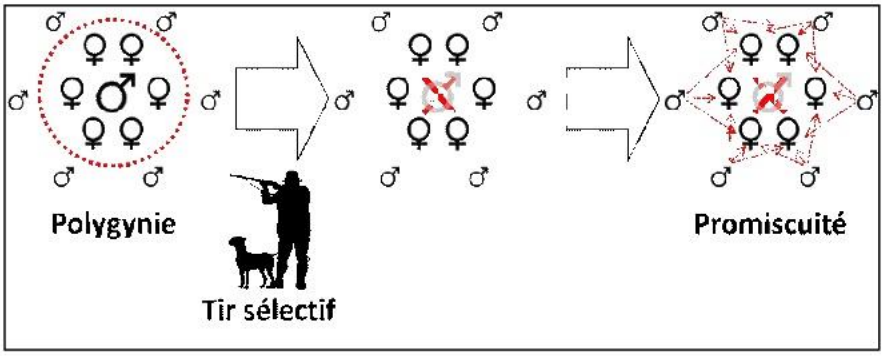
Distribution de l'estimation du nombre de pères utilisant 3 logiciels d'analyses génétiques GERUD *($n_{portées}$= 212),* CERVUS *($n_{portées}$= 159), et* COLONY *($n_{portées}$= 217).*

A partir du nombre de pères par portée, il est possible de déduire la proportion de portées engendrées par plus d'un mâle. Encore une fois, les résultats varient selon les analyses mais dans tous les cas, plus de 45% des portées sont issues de plusieurs pères avec un maximum de 82% observé avec COLONY. En considérant qu'avec GERUD et CERVUS, le nombre de pères a tendance à être sous-estimé, il est probable que la proportion de portées engendrées par plusieurs pères est plus importante en réalité. De fait, plus de la moitié des femelles analysées se sont accouplées avec plusieurs mâles.



Proportion de portées engendrées par plus d'un mâle en fonction du logiciel d'analyses génétiques GERUD *($n_{portées}$= 212),* CERVUS *($n_{portées}$= 159), et* COLONY *($n_{portées}$= 217).*

Les mâles s'accouplant avec plusieurs femelles et les femelles s'accouplant avec plusieurs mâles, le système de reproduction des sangliers de la population de Châteauvillain-Arc-en-Barrois est donc davantage promiscuite que polygyne. Etant donnée la très forte pression de chasse sur les mâles, la promiscuité peut être apparue après la diminution du nombre de vieux mâles dans cette population. Les femelles peuvent alors s'accoupler avec plusieurs mâles, ce qui peut être avantageux d'un point de vue évolutif.



Représentation schématique de la modification du système de reproduction du sanglier d'un système type polygynie (un mâle dominant se reproduit avec plusieurs femelles) vers un système de type promiscuité (un ou plusieurs mâles se reproduisent avec une ou plusieurs femelles) dû à l'absence des vieux mâles prélevés préférentiellement à la chasse.

Cet avantage évolutif peut se traduire notamment par une augmentation de la fertilité des laies. En effet, dans cette population, il a été constaté un effet positif du nombre de pères sur la taille de la portée. On peut supposer qu'il y ait augmentation du nombre d'ovules fécondés, comme c'est le cas chez le porc domestique.