UNIVERSITY OF SASSARI

DISSERTATION FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN
ENVIROMENTAL BIOLOGY
UNIVERSITY OF SASSARI, 2008
XXI CYCLE

The use of genetic markers in the study of social structure in mammals: wolf and wild boar as case studies

Candidate:
Laura Iacolina

Tutor and Supervisor:
Prof. Marco Apollonio

Co-*Supervisor:*
PhD Massimo Scandura

UNIVERSITÁ DI SASSARI

TESI DI DOTTORATO IN
BIOLOGIA AMBIENTALE
UNIVERSITÁ DI SASSARI, 2008
XXI CICLO

L'uso di marcatori molecolari nelle studio di strutture sociali nei mammiferi: il lupo ed il cinghiale come casi studio

Candidata:
Laura Iacolina

Tutor e Supervisore:
Prof. Marco Apollonio

Co-*Supervisore:*
PhD Massimo Scandura

Abstract

This dissertation shows the usefulness of molecular markers in highlighting some crucial aspects of the population and social structure of mammals. Specifically, my study is focused on two species: wild boar (*Sus scrofa*) and wolf (*Canis lupus*). Because of their keystone role in the natural ecosystems, the knowledge of their biology and population dynamics is of paramount importance for the management and conservation of the large mammals' communities in Italy.

First I give a picture of the genetic status of the wild boar in Europe. Then the genetic structure is analysed at a finer scale, with particular attention to relationships: within wild boar social groups and within litters. Within groups the predicted matrilineal social structure was not confirmed, as a low degree of genetic relatedness was observed within groups. While within litters multiple paternity, previously detected in European wild boars, is confirmed revealing that polyandry can become common in some populations.

Then, after the definition of an epirical method in assesing non-invasive samples quality, the genetic structure of a wolf population in a portion of the Italian Apennines was investigated through a six years period. Autosomal and Y-linked microsatellites were employed, thus allowing non-biased reconstruction of genetic patterns within and among wolf packs identifying low levels of gene-flow among adiacent packs, cryptic genetic structure and an higher variability in the male lineage than in females.

Riassunto

Questa tesi mostra l'utilità dei marcatori molecolari nello studio di alcuni aspetti cruciali della popolazione e della struttura sociale dei mammiferi. In particolare sono stati investigati il cinghiale (*Sus scrofa*) ed il lupo (*Canis lupus*).

In primo luogo si è definito il quadro dello stato genetico del cinghiale in Europa. In seguito la struttura genetica è stata analizzata ad una scala più fine, concentrandosi in particola modo sulle relazioni di parentela all'interno dei gruppi sociali e delle cucciolate. Lo studio della struttura sociale del cinghiale è stato condotto integrando I dati genetici con i dati spaziali della popolazione toscana oggetto di studio. La previsione di una struttura sociale matriarcale non è stata confermata dai dati, infatti, si è riscontrato un basso grado di parentela fra gli individui appartenenti alla stessa unità sociale. L'alto ricambio osservato all'interno della popolazione, principalmente dovuto ad un'alta mortalità dovuta a caccia e bracconaggio, sembra essere la causa principale dello scostamento dall'atteso.

Nella stessa popolazione è stata testata la presenza di multipaternità all'interno delle cucciolate. Sono state analizzate dodici famiglie, costituite da scrofe incinte, abbattute durante la stagione venatoria, e dai rispettivi feti. La multipaternità, precedentemente riscontrata in popolazioni di cinghiale europeo, è stata rilevata anche nella popolazione in esame, rivelando come la poliandria possa diventare comune in alcune popolazioni.

Dopodichè, una volta definito un protocollo d'analisi che permettesse l'utilizzo affidabile di campioni noninvasivi, si è proceduto a ricostruire la struttura genetica di una popolazione di lupo appenninico in un arco di tempo di sei anni. Tale analisi è stata condotta mediante l'utilizzo di microsatelliti autosomici e localizzati sul cromosoma sessuale Y, in modo da ottenere un quadro storico delle relazioni tra branchi, o all'interno di un branco, che non fosse influenzato dai meccanismi di trasmissione parentale e da possibili diversi pattern di dispersione fra sessi. Si è dunque osservato un basso flusso genico tra branchi adiacenti con conseguente strutturazione genetica ed una variabilità all'interno della linea maschile che si discosta dal monomorfismo riportato per la linea materna nella specie.

Introduction

This thesis consists of four main parts dealing with the importance of the use of molecular markers in highlighting some crucial aspects of the population and social structure in two mammalian species: the wild boar (*Sus scrofa*) and the wolf (*Canis lupus*). In the *First part* I draw a picture of the genetic status of the wild boar in Europe. In the *Second part* I analyze the genetic structure at a finer scale, with particular attention to relationships within wild boar social groups and within litters. In the *Third part* I point out the reliability of noninvasive DNA wolf samples, which I used in the *Fourth part* to track individuals in a population and to reconstruct fine-scale genetic population structure using differently inherited molecular markers.

The first half of this dissertation is centerend aroud *Sus scrofa* species and the factors influencing its gene dynamics at a wide geographic scale.

The wild boar is a typical r-strategist showing high ecological adaptability, opportunistic feeding and very high reproductive potential (Boitani et al. 1995; Fernandez-Llario and Mateos-Quesada 1998; Geisser and Reyer 2005). We can therefore suppose that after the last glaciation the wild boar easily recolonized Central and Northern European forests, thus reaching an almost continuous and stable distribution modified only by seasonal variations (Jedrzejewska et al. 1997; Bieber and Ruf 2005). However, the population went through a demographic decline in the fist part of the XX century (Apollonio et al. 1988; Oliver et al. 1993), followed by a complete recovery and increase in density throughout Europe after the Second World War (Sáez-Royuela and Tellería 1986; Feichtner 1998; Danilkin 2001).

Togheter with the characteristics the species itself (e.g. dispersal rates) and the landscape (e.g. geographical barriers) (Avise 2004), this expansion-contraction pattern is likely to have strongly affected levels and patterns of genetic variation across the species range. The aim of the *First Part* (**Chapter I**) of this study was to investigate which forces shaped the genetic structure in the European wild boar and their total effect on nowadays populations. With this purpose I investigted both the effects of natural and anthropogenic factors, taking into account the extensive human manipulation that game and domestic species are subject to. I used a large set of molecular data in order to distinguish and estimate the effects of different processes.

As regards natural events I investigated if post-glacial dispersal patterns are still detectable in the present European wild boars and the possible effect of more recent

demographic events (i.e. the contraction-expansion of the population in the last century) on its genetic diversity. Thus, first I addressed the human influence investigating if divergent pig genomes, which were subjected to strong selective and likely drift effects during and after domestication, could have introgressed into the wild boar genome due to local hybridization between the two forms. The fenomenon can occur in areas where pigs are reared in semi-wild conditions (e.g. in Bulgaria and Sardinia, Apollonio et al. 1988; Genov et al. 1991), as well as in other regions (e.g. Central Italy) where pigs were occasionally crossed with wild boars in captivity, and hybrids were released for hunting purposes (Randi et al. 1989). In addition, I checked if the genetic variation in some wild boar populations has been affected also by artificial long-distance migrations associated to uncontrolled and rarely documented restocking plans over the last 50 years. Restocking could have modified the genetic variation both by contributing to the recent demographic expansion and by mixing genetic pools belonging to different subspecies or diverging populations. For example, wild boars from Central Europe are reported to have been repeatedly introduced into Italy (Apollonio et al. 1988; Olivier et al. 1993). The genetic impact of such events in some areas is controversial, with authors suggesting either massive (Randi 2005) or limited (Vernesi et al. 2003) introgression.

All the above mentioned processes have potentially left a signature in the present-day wild boar genetic variation. I used mitochondrial and nuclear markers (microsatellites) to test whether they did or not, and which was their respective contribution.

After having investigated the driving forces sharing genetic variation at a wide scale, in the *Second Part* I focused on a wild boar population in Central Italy (Tuscany), studying the effects of spatial and social organization.

The social organization of wild boar is centered around philopatric adult females, which are facultative cooperative breeders. The basic social unit was reported to be a matrilineal group, with one or more related sows, and one or more cohorts of offspring (Briedermann 1986). After weaning, most females stay with their mothers, and only about 20% of yearling females leave the natal group and disperse while the others are likely to reproduce while still joining their mother social group (Kaminski et al. 2005). However, genealogical relationships in female groups have been poorly investigated thus far and deviations from this commonly accepted scheme have been rarely documented (see Gabor et al. 1999).

Female wild boars typically maintain long-term fidelity to relatively small home ranges (Spitz and Janeau 1990), and a high percentage of adjacent females exhibit overlapping home ranges (Boitani et al. 1994). Accordingly, one would expect overlapping home ranges to reflect

a common female lineage, and genetic relatedness to be inversely correlated with the spatial distance between individuals.

The reproductive features of the species, as well as its breeding system, are likely to affect the genetic structure in a population, influencing both the size and the composition of social groups and the duration of inter-individual associations. In addition, the organization of wild boar groups can vary temporally, with splitting into subgroups and merging of subgroups occurring frequently in a population (Gabor et al. 1999; Kaminski et al. 2005). Importantly, however, the role of demographic and extrinsic factors (e.g. hunting), possibly affecting both the composition and the stability of social groups, has not been systematically investigated.

The primary objective of **Chapter II** was to evaluate the nature of wild boar associations in relation to the genetic relatedness among individuals. I first verified the correlation between geographic and genetic distance among individuals in a population, and then considered the spatial behavior of social units in relation to their composition and the intra-group degree of relatedness. Specifically, I addressed the following questions: Is the geographic distance between individuals inversely correlated to their genetic relatedness? Is genetic relatedness higher for individuals belonging to the same social unit than for individuals belonging to different social units? And, are all adult females in a social group close relatives (mother-offspring or full-sisters)?

While **Chapter III** aims at verifying the occurrence of multiple paternity in the Central Italian population under study. Multiple paternity is known to occur in the domestic pig (Aguillera-Reyes et al 2006), where it usually leads to an increased litter size (Xue et al. 1998), and it has been recently discovered also in a Portuguese wild boar population (Delgado et al. 2008). In this chapter I investigated the occurrence and frequency of multiple paternity. Then I compared results obtained for the Italian population with those of Delgado et al. (2008) with the purpose of understanding if differences are present.

The second half of this dissertation deals with the study of population structure and social biology in the wolf (*Canis lupus*).

The conservation of wolf natural populations represents a priority in several European countries, where the species is endangered or was, in the recent past, severely threatened (Promberger and Schröder 1993). The Italian wolf population suffered a strong persecution until 1971, when wolf hunting was forbidden and poison baits banned. In 1976 the species obtained the fully protected status and from then on the population recovered.

In Italy, killing wolves is illegal, so the only tissue samples available are those belonging to wolf carcasses accidentaly found in the field (death caused by vehicle collision or illegal culling). It is very difficult to obtain samples from live wolves, because of their elusiveness that make uneasy to capture wild individuals. For this reason the main source of information on the population is noninvasive sampling and monitoring. Samples like faeces, hairs, urine, haematic residuals, etc. were proved to represent a valuable source of DNA for genetic analyses (for a review Morin and Woodruff 1996; Piggott and Taylor 2003). Those kind of source samples can be used, with the employment of species-specific markers that minimize the risk of mistyping due to contaminations, for a wide range of analyses according to the purpose of the investigation. Noninvasive collected samples can be used in the investigation of the matrilineal component of a populationby the analysis of mitochondrial DNA, that is present in both sexes, shared by more than one individual and inherited only from the mother, (e.g. Valière et al. 2003). The same samples can be used for the investigation of the patriline in the population by the analysis of chromosome Y-linked markers (e.g. Eriksson et al. 2006), present only in the male component of the population, inherited from father to son and shared by more individuals. The use of noninvasive sampling also allows to identify single individuals with the use of biparentally inherited autosomal markers (i.e. microsatellites, that can provide a unique genotype for each individual in the population (Adams and Waits 2007).

However some pitfalls on noninvasive samples were pointed out (Taberlet et al. 1996; Kohn and Wayne 1997). The use of very small or degraded DNA samples for PCR-based genetic analyses, apart from being extremely sensitive to any source of contamination, might be accompanied by the occurrence of stochastically-generated errors, thus leading to inconsistent results (Goossens et al. 1998; Taberlet et al. 1999). To overcome this limitation, some methods were proposed, aimed at reaching a fixed reliability threshold by replicating amplifications several times. The required number of replications was defined either from a priori generalised error probabilities (Navidi et al. 1992; Taberlet et al. 1996) or from maximum-likelihood estimates of genotype reliability, computed from individual dropout rate and allele frequencies (Miller et al. 2002).

In the *Third part* (**Chapter IV**)of this dissertation I contributed to develop an approach for achieving a reliable data set of wolf genotypes from multiple sources of DNA, including both tissue from carcasses and noninvasively collected samples. This method relies on the relationship between sample quality and amplification outcome, which is ultimately related to the occurrence of typing errors (allelic dropout, false alleles). After DNA extraction, templates are amplified once at each locus and a conservative rating system (Q-score) is adopted to define the quality of single-locus amplifications. A significant relationship was found between

quality scores and error rate., thus allowing to predict the chance a genotype has of being affected by errors on the basis of its Q-score. Genotypes not reaching a satisfactory confidence level can either be replicated to become reliable or excluded from the data set. Despite the possible decrease in overall yield, this method could provide a good compromise between accuracy in genotyping and effectiveness in screening large data sets for long-term or large-scale population surveys. However, to achieve complete and reliable data sets, replicated amplifications are necessary for those samples and loci providing poor results.

In the *Fourth part* of this dissertation, after the definition of the methodological approach, I was able to investigate the genetic structure of a wolf population in a portion of the northern Apennines, where wolves and wild ungulates reached high densities and no natural or artificial barrier limited wolf dispersal.

The use of highly polymorphic microsatellite loci and the development of individual-based statistical methods (Cornuet et al. 1999; Pritchard et al. 2000; Beerli and Felsenstein 2001; Wilson and Rannala 2003; Paetkau et al. 2004) have contributed to improve the resolution power of genetic analysis, thus allowing the uneven partitioning of genetic diversity to be detected at a finer spatial scale (see Pearse and Crandall 2004). As a consequence, several genetic investigations have revealed cryptic barriers to gene flow among subpopulations or social units, highlighting previously unsuspected levels of local genetic differentiation (Taylor et al. 1997; Coltman et al. 2003; Fredsted et al. 2005; Swanson et al. 2005). In referring several cases in which the local divergence exceeded the regional divergence in mammals, Storz (1999) observed that the genetic differentiation among adjacent social groups or spatially defined breeding units may represent an important microevolutionary force in mammalian populations.

As discussed above genetic diversity in a population is strongly influenced by the breeding system and the dispersal patterns, which, in turn, depend on age and sex of dispersers (Chesser 1991a; 1991b). Most social mammals show polygynous mating and female philopatry (Clutton-Brock 1989) and such philopatry generates a sex-biased spatial genetic structure (Purdue et al. 2000; Spong et al. 2002; Coltman et al. 2003; Nussey et al. 2005). In addition, dispersal is often age-biased. However, only the proportion of successful dispersal mediates gene flow in a population. Therefore, all factors affecting rate and pattern of dispersal, on the one hand, and sex and age of dispersing individuals, on the other, can potentially modify the extent of spatial genetic differentiation within the population.

Wolf social structure relies on social units (packs), which are basically composed of a breeding pair and their offspring, but can be joined by members of previous litters and,

occasionally, by unrelated individuals (see Mech and Boitani 2003). Packs originate from the encounter and mating of two dispersing individuals, which are usually unrelated (Smith et al. 1997). The new pair establishes its own territory in an area undefended by other wolves. The chance a pair has to start a family is inversely proportional to the degree of saturation of suitable areas (Packard 2003). Once a pack has settled in its own territory, the stability of the family unit is largely dependent on the fate of the breeding pair (Brainerd et al. 2008).

In **Chapter V** I focused on the fine genetic structure within the free-living wolf population inhabiting the province of Arezzo. Specifically, implementing a noninvasive genotyping strategy, I investigated whether any substantial geographic structuring was present, observing the relatedness within and among neighbouring packs at increasing geographic distances. Finally, as the noninvasive sampling method allowed to obtain multiple locations for one individual, I investigated which proportion of wolves in the population had frequented more than one area, where a pack lives, within the Province territory.

However the species behaviour is not the only force shaping genetic variability within a population. The Italian wolf population suffered a strong persecution that, during the period 1950-1970, led to a very low number of individuals throughout the peninsula (Cagnolaro et al. 1974) and their presence was extremely restricted to a few areas of the Apennine mountains. From then on, the wolf population recovered in Italy and is still growing in size and recolonizing its historical range (Boitani 2003).

As consequence of its history, the Italian wolf shows the effect of a severe bottleneck which, in association with its prolonged isolation, led to genetic erosion (Lucchini et al. 2004). In fact, a single mitochondrial haplotype was found in the whole Italian wolf population (Wayne et al. 1992, Vilà et al. 1997, Vila et al. 1999, Randi et al. 2000). On the contrary, nuclear markers revealed a less severe reduction in genetic diversity, with a remarkable reduction in heterozygosity only in the recently re-established alpine wolf population, while in the Apennines the population showed only a slightly lower variability with respect to other European populations (Scandura et al. 2001, Lucchini et al. 2004, Fabbri et al. 2007).

In **Chapter VI** I aimed at integrating the knowledge acquired in the former chapter by implementing biparental data with Y-linked markers (microsatellites). Specifically, I investigated the spatial and temporal distribution of Y haplotypes in the male lineage of the Arezzo population, checking if male gene flow is somehow constrained within the study area. Furthermore, I reconstructed the patriline in an intensively studied wolf pack over a four-year period. This study revealed for the first time that, unlike maternally-inherited mtDNA

markers, male-specific Y-chromosome markers show a relevant level of variation even at a limited geographic scale.

References

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

Aguillera-Reyes U., Zavala-Páramo G., Valdez-Alarcón J.J., Cano-Camacho H., García-López G.I., Pescador-Salas N. (2006) Multiple mating and paternity determinations in domestic swine (*Sus scrofa*). Animal Research 55: 409-417.

Apollonio M., Randi E., Toso S. (1988) The systematics of the wild boar (*Sus scrofa* L.) in Italy. Bollettino Di Zoologia 3: 213–221.

Avise J.C. (2004) Molecular markers natural history, and evolution. 2nd edn. Sinauer & Associates, Sunderland, Massachusetts.

Beerli P., Felsenstein J. (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. Proceedings of the National Academy of Science USA 98: 4563-4568

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

Boitani L. (2003) Wolf conservation and recovery. In: Mech L.D., Boitani L. (eds) Wolves: Behavior, Ecology, and Conservation, The University of Chicago Press, Chicago, pp. 1-34.

Boitani L., Mattei L., Nonis D., Corsi F. (1994) Spatial and activity patterns of wild boars in Tuscany, Italy. Journal of Mammalogy 75: 600-612.

Boitani L., Trapanese P., Mattei L., Nonis D. (1995) Demography of a wild boar (*Sus scrofa*, L.) population, Tuscany, Italy. Gibier Faune Sauvage 12: 109–132.

Brainerd S.M., Andrén H., Bangs E.E., Bradley E.H., Fontaine J.A., Hall W., Iliopoulos Y., Jimenez M.D., Jozwiak E.A., Liberg O., Mack C.M., Meier T.J., Niemeyer C.C., Pedersen H.C., Sand H., Schultz R.N., Smith D.W., Wabakken P., Wydeven A.P. (2008) The effects of breeder loss on wolves. Journal of Wildlife Management 72 (1): 89-98.

Briedermann L. (1986) Schwartzwild. Neumann-Neudamm ed. Berlin, Germany.

Cagnolaro L., Rosso D., Spagnesi M., Venturi B. (1974) Investigation on the wolf (*Canis lupus*) distribution in Italy, in Canton Ticino and Canton Grigioni (Switzerland). Ricerche di Biologia della Selvaggina, 59, 1-75 [in Italian].

Chesser R.K. (1991a) Influence of gene flow and breeding tactics on gene diversity within populations. Genetics 129: 573-583.

Chesser R.K. (1991b) Gene diversity and female phylopatry. Genetics 127: 437-447

Clutton-Brock T.H. (1989) Female transfer and inbreeding avoidance in social mammals. Nature 337: 70-72.

Coltman D.W., Pilkington J.G., Pemberton J.M. (2003) Fine-scale genetic structure in a free-living ungulate population. Molecular Ecology 12: 733-742.

Cornuet J.M., Piry S., Luikart G., Estoup A., Solignac M. (1999) New methods employing multilocus genotypes to select or exclude populations as origin of individuals. Genetics 153: 1989-2000.

Danilkin A.A. (2001) The wild boar: an unprecedented spread or restoration of the species range? Doklady Biological Sciences 380: 457–460.

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

Eriksson J., Siedel H., Lukas D., Kayser M., Erler A., Hashimoto C., Hohmann G., Boesch C., Vigilant L. (2006) Y-chromosome analysis confirms highly sex-biased dispersal and suggests a low male effective population size in bonobos (*Pan paniscus*). Molecular Ecology 15: 939-949.

Fabbri E., Miquel C., Lucchini V., Santini A., Caniglia R., Duchamp C., Weber J.-M., Lequette B., Marucco F., Boitani L., Fumagalli L., Taberlet P., Randi E. (2007) From the Apennines to the Alps: colonization genetics of the naturally expanding Italian wolf (*Canis lupus*) population. Molecular Ecology 16(8): 1661-1671.

Feichtner B. (1998) Ursachen der Streckenschwankungen beim Schwarzwild im Saarland. Zeitschrift für Jagdwissenschaft 44: 140–150.

Fernandez-Llario P., Mateos-Quesada P. (1998) Body size and reproductive parameters in the wild boar *Sus scrofa*. Acta Theriologica 43: 439–444.

Fredsted T., Pertoldi C., Schierup M.H., Kappeler P.M. (2005) Microsatellite analyses reveal fine-scale genetic structure in grey mouse lemurs (*Microcebus murinus*). Molecular Ecology 14: 2363-2372.

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

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

Genov P., Nikolov H., Massei G., Gerasimov S. (1991) Craniometrical analysis of Bulgarian wild boar (*Sus scrofa*) populations. Journal of Zoology London 225: 309–325.

Goossens B., Waits L.P., Taberlet P. (1998) Plucked hair samples as a source of DNA: Reliability of dinucleotide microsatellite genotyping. Molecular Ecology 7: 1237–1241.

Jedrzejewska B., Jedrzejewski W., Bunevich A.N., Milkowski L., Krasinski A. (1997) Factors shaping population densities and increased rates of ungulates in Bialowieza Primeval Forest (Poland and Belarus) in the 19th and 20th centuries. Acta Theriologica 42: 399–451.

Kaminski G., Brandt S., Baubet E., Baudoin C. (2005) Life-history patterns in female wild boars (*Sus scrofa*): mother-daughter postweaning associations. Canadian Journal of Zoology 83: 474-480.

Kohn M.H., Wayne R.K. (1997) Facts from feces revisited. Trends in Ecology and Evolution 12: 223–227.

Lucchini V., Galov A., Randi E. (2004) Evidence of genetic distinction and long-term population decline in wolves (*Canis lupus*) in the Italian Apennines. Molecular Ecology 13 (3): 523-536.

Mech L.D., Boitani L. (2003) Wolf social ecology. In: Mech L.D., Boitani L. (eds.) Wolves: Behavior, Ecology, and Conservation. The University of Chicago Press, Chicago, USA, pp. 1-3.

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

Morin P.A., Woodruff D.S. (1996) Noninvasive genotyping for vertebrate conservation. In: Molecular Genetic Approaches in Conservation (eds. Smith T.B., Wayne R.K.), pp. 298–313. Oxford University Press, New York.

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

Nussey D.H., Coltman D.W., Coulson T., Kruuk L.E.B., Donald A., Morris S.J., Clutton-Brock T.H., Pemberton J. (2005) Rapidly declining fine-scale spatial genetic structure in female red deer. Molecular Ecology 14: 3395-3405.

Oliver W.L.R., Brisbin I.L., Takahashi S. (1993) The Eurasian wild pig, *Sus scrofa*. In Pigs, Peccaries and Hippos - Status Survey and Action Plan. (ed. Oliver W.L.R.), pp. 112–121. IUCN/SSC, Gland, Switzerland.

Packard J.M. (2003) Wolf behavior: reproductive, social, and intelligent. In: Mech L.D., Boitani L. (eds) Wolves: Behavior, Ecology, and Conservation, The University of Chicago Press, Chicago, p. 38.

Paetkau D., Slade R., Burden M., Estoup A. (2004) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. Molecular Ecology 13: 55-65.

Pearse D.E., Crandall K.A. (2004) Beyond FST: Analysis of population genetic data for conservation. Conservation Genetics 5: 585–602.

Piggott M.P., Taylor A.C. (2003) Remote collection of animal DNA and its applications in conservation management and understanding the population biology of rare and cryptic species. Wildlife Research 30: 1–13.

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

Promberger C, Schröder W (Eds.) (1993) Wolves in Europe: Status and perspectives. Munich Wildlife Society, Ettal, Germany.

Purdue J.R., Smith M.H., Patton J.C. (2000) Female philopatry and extreme spatial genetic heterogeneity in white-tailed deer. Journal of Mammalogy 81: 179-185.

Randi E., Apollonio M., Toso S. (1989) The systematics of some Italian populations of wild boar (*Sus scrofa* L.): a craniometric and electrophoretic analysis. Zeitschrift für Säugetierkunde 54: 40–56.

Randi E., Lucchini V., Christensen M.F., Mucci N., Funk S.M., Dolf G., Loeschcke V. (2000) Mitochondrial DNA variability in Italian and East European wolves: detecting the consequences of small population size and hybridization. Conservation Biology 14: 464-473.

Randi E. (2005) Management of wild ungulate populations in Italy: captive-breeding, hybridisation and genetic consequences of translocations. Veterinary Research Communications 29: 71–75.

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

Scandura M., Apollonio M., Mattioli L. (2001) Recent recovery of the Italian wolf population: a genetic investigation using microsatellites. Mammalian Biology 66: 321-331.

Smith D., Meier T., Geffen E., Mech L.D., Burch J.W., Adams L.G., Wayne R.K. (1997) Is incest common in gray wolf packs? Behavioural Ecology 8: 384-391.

Spitz F., Janeau G. (1990) Spatial strategies: an attempt to classify daily movements of wild boar. Acta Theriologica 35 (1-2): 129.

Spong G., Stone J., Creel S., Björklund M. (2002) Genetic structure of lions (*Panthera leo* L.) in the Selous Game Reserve: implications for the evolution of sociality. Journal of Evolutionary Biology 15: 945–953.

Storz J.F. (1999) Genetic consequences of mammalian social structure. Journal of Mammalogy 80: 553-569.

Swanson B., Fuhrmann R., Crabtree R. (2005) Elevational isolation of red fox populations in the Greater Yellowstone Ecosystem. Conservation Genetics 6: 123-131.

Taberlet P., Griffin S., Goossens B., Questiau S., Manceau V., Escaravage N., Waits L.P., Bouvet J. (1996) Reliable genotyping of samples with very low DNA quantities using PCR. Nucleic Acids Research 24: 3189–3194.

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

Taylor A.C., Horsup A., Johnson C.N., Sunnucks P., Sherwin B. (1997) Relatedness structure detected by microsatellite analysis and attempted pedigree reconstruction in an endangered marsupial, the northern hairy-nosed wombat *Lasiorhinus krefftii*. Molecular Ecology 6: 9-19.

Valière N., Fumagalli L., Gielly L., Lequette, B., Poulle M.-L., Weber J.-M., Arlettaz R., Taberlet P. (2003) Long-distance wolf recolonization of France and Switzerland inferred from non-invasive genetic sampling over a period of 10 years. Animal Conservation 6: 83-92.

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

Vilà C., Savolainen P., Maldanado J.E., Amorim I.R., Rice J.E., Honeycutt R.L., Crandall K.A., Lundeberg J., Wayne R.K. (1997) Multiple and ancient origins of the domestic dog. Science 276: 1687-1689.

Vilà C., Amorim I.R., Leonard J.A., Posada D., Castroviejo J., Petrucci-Fonseca F., Crandall K.A., Ellegren H., Wayne R.K. (1999) Mitochondrial DNA phylogeography and population history of the gray wolf *Canis lupus*. Molecular Ecology 8: 2089-2103.

Wayne R.K., Lehman N., Allard M.W., Honeycutt R.L. (1992) Mitochondrial DNA variability of the gray wolf: genetic consequences of population decline and habitat fragmentation. Conservation Biology 6: 559-569.

Wilson G.A., Rannala B. (2003) Bayesian inference of recent migration rates using multilocus genotypes. Genetics 163: 1177-1191.

Xue J.L., Lucia T., Koketsu Y., Dial G.D., Marsh W.E. (1998) Effect of mating frequency and weaning-to-mating interval on sow reproductive performance. Swine Health and Production 6 (4): 157-162.

FIRST PART

Wild boar population structure throughout Europe

CHAPTER I

Ancient vs. recent processes as factors shaping the genetic variation of the European wild boar: are the effects of the last glaciation still detectable?

Ancient vs. recent processes as factors shaping the genetic variation of the European wild boar: are the effects of the last glaciation still detectable?

Massimo SCANDURA, Laura IACOLINA, Barbara CRESTANELLO, Elisa PECCHIOLI, Maria Francesca DI BENEDETTO, Vincenzo RUSSO, Roberta DAVOLI, Marco APOLLONIO and Giorgio BERTORELLE

Molecular Ecology (2008) 17, 1745–1762

Molecular Ecology (2008) **17**, 1745–1762 doi: 10.1111/j.1365-294X.2008.03703.x

Ancient vs. recent processes as factors shaping the genetic variation of the European wild boar: are the effects of the last glaciation still detectable?

M. SCANDURA,* L. IACOLINA,* B. CRESTANELLO,†‡ E. PECCHIOLI,‡§ M. F. DI BENEDETTO,* V. RUSSO,¶ R. DAVOLI,¶ M. APOLLONIO* and G. BERTORELLE†

Dipartimento di Zoologia e Genetica Evoluzionistica, Università di Sassari, via Muroni 25, 07100 Sassari, Italy, †Dipartimento di Biologia, Università di Ferrara, via Borsari 46, 44100 Ferrara, Italy, ‡Centro di Ecologia Alpina, Viote del Monte Bondone, 38100 Trento, Italy, §Biosfera, Biologia della Conservazione, Ricerca e Didattica, 50122 Firenze, Italy, ¶DIPROVAL, Università di Bologna, Via Rosselli, 42100 Reggio Emilia, Italy

Abstract

The European wild boar is an important game species, subjected to local extinctions and translocations in the past, and currently enormously and worryingly expanding in some areas where management is urgently required. Understanding the relative roles of ancient and recent events in shaping the genetic structure of this species is therefore not only an interesting scientific issue, but it represents also the basis for addressing future management strategies. In addition, several pig breeds descend from the European wild boar, but the geographical location of the domestication area(s) and the possible introgression of pig genomes into wild populations are still open questions. Here, we analysed the genetic variation in different wild boar populations in Europe. Ten polymorphic microsatellites were typed in 252 wild boars and the mtDNA control region was sequenced in a subset of 145 individuals. Some samples from different pig breeds were also analysed. Our results, which were obtained considering also 612 published mtDNA sequences, suggest that (i) most populations are similarly differentiated, but the major discontinuity is found along the Alps; (ii) except for the Italian populations, European wild boars show the signature of a postglacial demographic expansion; (iii) Italian populations seem to preserve a high proportion of preglaciation diversity; (iv) the demographic decline which occurred in some areas in the last few centuries did not produce a noticeable reduction of genetic variation; (v) signs of human-mediated gene flow among populations are weak, although in some regions the effects of translocations are detectable and a low degree of pig introgression can be identified; (vi) the hypothesis of an independent domestication centre in Italy is not supported by our data, which in turn confirm that Central European wild boar might have represented an important source for domestic breeds. We can therefore conclude that recent human activities had a limited effect on the wild boar genetic structure. It follows that areas with high variation and differentiation represent natural reservoirs of genetic diversity to be protected avoiding translocations. In this context controlling some populations by hunting is not expected to affect significantly genetic variation in this species.

Keywords: genetic diversity, microsatellites, mtDNA, phylogeography, population expansion, *Sus scrofa*

Received 17 September 2007; revision received 27 November 2007; accepted 11 January 2008

Introduction

Present-day levels and patterns of genetic variation are strongly affected by the characteristics of a species (e.g. dispersal rates), the landscape (e.g. geographical barriers),

Correspondence: Massimo Scandura, Fax: +39 079 228665; E-mail: scandura@uniss.it

but also by different specific events that occurred in the past (Avise 2004). For example, the genetic impact of the last glaciation was relevant to several species (Hewitt 2000), but many other processes with a potential effect on the genetic variation certainly occurred since then. This is especially true for game and domestic species because of extensive human manipulation. In this study, we focused on several European populations of wild boar (*Sus scrofa*), whose genetic variation patterns have been possibly affected by both domestication and hunting/management activities. We used a large set of molecular data in order to distinguish and estimate the effects of different processes that occurred at different timescales. In addition, this data set enabled us to test a specific hypothesis regarding the domestication of the species in Europe.

The Eurasian wild boar is one of the most widely distributed terrestrial mammals. Its geographical range, excluding recent introductions, extends from Western Europe and Northern Africa to Japan. Possibly originated in Southeastern Asia, where the highest numbers of wild pig taxa are observed (Lucchini *et al*. 2005), *S. scrofa* can be now subdivided into at least two major genetic clades roughly corresponding in the domestic form to European and Asian pigs (Giuffra *et al*. 2000; Okumura *et al*. 2001; Alves *et al*. 2003; Larson *et al*. 2005). The separation of these two clades can be dated back to between 0.5 and 0.9 million years ago, suggesting at least two independent domestication events (Giuffra *et al*. 2000; Kijas & Anderson 2001; Alves *et al*. 2003). Subspecies are usually classified into four major groups, with the European wild boar corresponding to the so-called 'Western races' group (Oliver *et al*. 1993).

The quaternary climatic oscillations, and in particular the last glaciation and the subsequent warm period, produced remarkable consequences on the levels and patterns of genetic variation in several species (Taberlet *et al*. 1998; Hewitt 2000; Petit *et al*. 2003; Hofreiter *et al*. 2004). As regards the European wild boar, however, at least four additional processes related to human activities have occurred ever since: the domestication in the Neolithic; a severe bottleneck in different areas in the last few centuries; a demographic expansion in the last 50 years; several more or less uncontrolled introductions of individuals, which also occurred in the last decades, to restock areas where wild boar was extinct or present at low density.

Postglaciation dispersal

The model initially proposed for the population dynamics of several species during the last glaciation (i.e. one or more southern refugia and postglacial re-expansion towards northern areas) is probably too simplistic for some taxa (see for example Magri *et al*. 2006). However, the current distribution of the wild boar and its dispersal and reproductive potential suggest that the genetic variation in this species should be initially investigated with this model in mind. In fact, as the wild boar is being only sporadically observed in northern areas like central and northern taiga (Briedermann 1990; Danilkin 2001), its presence in Central and Northern Europe during the last Ice Age, when the permafrost almost isolated Iberia, Italy, and the Balkans (Hewitt 2000), seems unlikely. In addition, unlike for example other ungulate species, the wild boar shows the typical attributes of r-strategists: high ecological adaptability, opportunistic feeding, and very high reproductive potential (Boitani *et al*. 1995; Fernandez-Llario & Mateos-Quesada 1998; Geisser & Reyer 2005). We can therefore suppose that after the last glaciation, the wild boar easily recolonized Central and Northern European forests, thus reaching an almost continuous and stable distribution modified only by seasonal variations (Jedrzejewska *et al*. 1997; Bieber & Ruf 2005).

Domestication and hybridization with pigs

Direct consequences on the wild populations during the domestication process are expected to be limited, but recent effects related to the co-existence of domestic and wild forms should be considered. In some areas, pigs are reared in semi-wild conditions (e.g. in Bulgaria and Sardinia, Apollonio *et al*. 1988; Genov *et al*. 1991) and crossbreeding with the wild form is possible. Furthermore, in other regions (e.g. Central Italy), pigs were occasionally crossed with wild boars in captivity, and hybrids were released for hunting purposes (Randi *et al*. 1989; I. Boschi, unpublished report). Therefore, divergent pig genomes, which were subjected to strong selective and likely drift effects during and after the domestication, could have introgressed into the wild boar genetic variation.

Overhunting and demographic decline

In the last few centuries, loss of habitat and overhunting drove the wild boar to extinction in some European regions such as the British Isles, Scandinavia, and several Italian and Western Russian areas (Apollonio *et al*. 1988; Oliver *et al*. 1993). A demographic decline was documented in many other countries, and yet the genetic effects of this event, when analysed in a few geographically restricted areas, were surprisingly not evident (Vernesi *et al*. 2003).

Recent expansion

After the Second World War, the density and geographical distribution of the wild boar have increased almost everywhere in Europe (Sáez-Royuela & Tellería 1986; Feichtner 1998; Danilkin 2001) as a consequence of several factors whose relative weight is uncertain. These factors are: global warming, changes in agricultural practices,

setting up of artificial feeding sites, reduced numbers of predators, increase of mast seeding of beech, and restocking (Bieber & Ruf 2005; Geisser & Reyer 2005). Growth rates of wild boar populations have been so high in some areas that damages to agricultural cultivations and natural ecosystems are frequently reported (Singer *et al*. 1984; Welander 2000; Schley & Roper 2003; Geisser & Reyer 2004).

Translocations

The genetic variation in some wild boar populations has possibly been affected also by artificial long-distance migrations associated to uncontrolled and rarely documented restocking plans over the last 50 years. Restocking could have modified the genetic variation both by contributing to the recent demographic expansion and by mixing genetic pools belonging to different subspecies or differentiated populations. For example, wild boars from Central Europe were repeatedly introduced into Italy (Apollonio *et al*. 1988). The genetic impact of such events in some areas is controversial, with authors suggesting either massive (Randi 2005) or limited (Vernesi *et al*. 2003) introgression. It is noteworthy that two opposite wild boar management policies, none of which considering the conservation of genetic biodiversity as a priority, are often suggested by local authorities: eradication, which is meant to remove the problems from cultivated areas, and restocking, in order to preserve traditional forms of 'social' hunting practices.

All the five processes above have potentially left a signature in the present-day wild boar genetic variation. We used mitochondrial and nuclear markers to test whether they did or not, and which was the possible role of each of them. Considering the wild boar as a model, the results we obtained are also valuable for a better reconstruction of the historical events that affected other species. Therefore, our results have important and more general implications for the development of management and conservation plans of game species.

Finally, we also addressed a topic which is not directly related to the wild boar genetic structure, and yet it is relevant to the understanding of the origin and the number of independent domestication events in this species. Archaeological evidences suggest that, like many other domestic animals, European pigs were domesticated in the Near East and selected breeds were subsequently introduced into Europe by Neolithic farmers (Epstein & Bichard 1984). However, two *in loco* domestication processes, one in Central Europe and the other in Italy, were recently hypothesized on the basis of the analysis of mitochondrial DNA sequences (Larson *et al*. 2005). This hypothesis, which is relevant also for the management of local breeds and the conservation of pig diversity, was tested using the same large data set.

Materials and methods

Sampling and DNA isolation

Hair, skin, skeletal muscle, or ear tissue from 252 wild boars were collected in 15 different sampling areas across Europe (Fig. 1) and stored in 95% ethanol at –20 °C. Sample abbreviations used later in the text are specified in Fig. 1. Additionally, 67 Italian domestic pigs from five historical breeds were sampled for comparison: Cinta Senese (coded Pig 1), Sarda (Pig 2), Calabrese (Pig 3), Casertana (Pig 4) and Nera Siciliana (Pig 5). Five commercial pigs (Large White, Pig 6) were also sampled. All the wild boar sampling areas in Italy, possibly with the exclusion of the Castel Porziano Presidential Reserve, and certainly at different degrees, had been subjected to occasional restocking with unknown genetic effects. The samples from four localities (IFlo, IMrp, ICpr and HDif) had already been used in a previous study (Vernesi *et al*. 2003). Total genomic DNA was extracted by using commercial kits (QIAGEN) or the standard phenol–chloroform method (Sambrook *et al*. 1989), followed by concentration in Microcon-30 columns (Amicon), and kept at –20 °C.

Mitochondrial DNA sequencing

Almost the entire control region (CR) was amplified by polymerase chain reaction (PCR) using two primers developed by Alves *et al*. (2003) (Ss.L-Dloop 5'-CGCCATCA-GCACCCAAAGCT3' and Ss.Hext-Dloop 5'-ATTTTGGGA-GGTTATTGTGTTGTA3') anchoring at positions 16569 and 1128 of the complete pig mitochondrial DNA (mtDNA) genome (GenBank Accession no. AF034253; Lin *et al*. 1999). Reactions were performed in an Applied Biosystems 2420 thermal cycler, with amplification conditions set at 35 cycles of 92 °C for 1 min, 62 °C for 1 min and 72 °C for 1 min, followed by a final extension step at 72 °C for 10 min. PCR products were purified by Exo/SAP digestion and a 411-bp fragment, including the hypervariable extended termination associated sequences (ETAS) domain (Sbisà *et al*. 1997), was directly sequenced using the forward primer SS.L-Dloop and the BigDye Terminator kit version 3.1 (Applied Biosystems). This region was also selected to maximize the size of possible alignments including already published GenBank sequences. Fragments were finally purified in columns loaded with Sephadex G-50 and run in an ABI PRISM 3100 Avant automatic sequencer (Applied Biosystems). Ambiguous positions were verified by re-sequencing the target region with the internal reverse primer Ss.Hint-Dloop (5'-TGGGCGATTTTAGGTGAGATGGT3'), mapping at position 465 of the pig mtDNA. Sequences were obtained for a subsample of 145 wild boars (between 8 and 12 per sampling location) and 47 domestic pigs from the Italian historical breeds (Pig 1 to Pig 5). Commercial

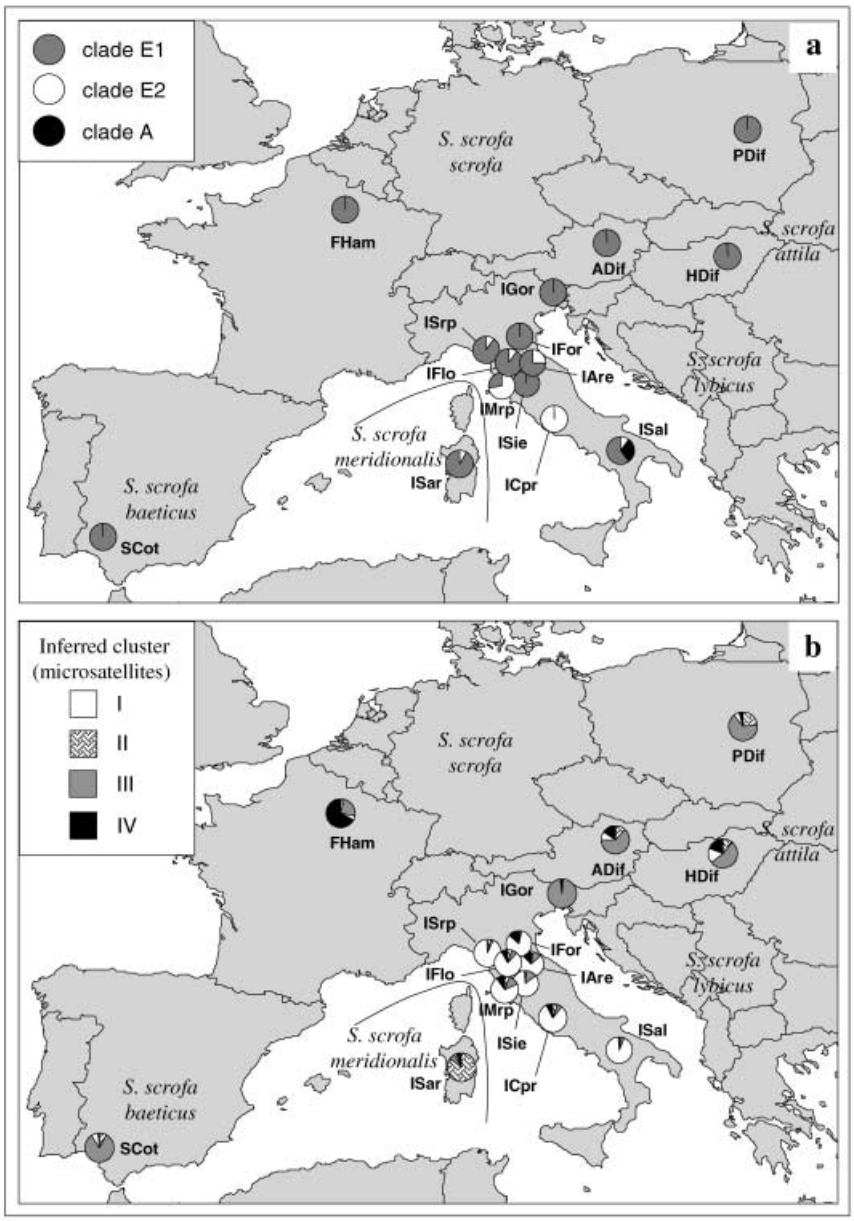


Fig. 1 Geographical locations of wild boar samples. In (a), pie charts represent proportions of each of the three main mtDNA clades (E1, E2 and A) in each sampling area. In (b), pie charts indicate proportions of membership of each sampled population to the four clusters inferred by structure analysis 2 (see text for details). The following abbreviations are used: IAre (Arezzo, Italy), IFor (Forì, Italy), ISie (Siena, Italy), ISal (Salerno, Italy), IFlo (Florence, Italy), IMrp (Maremma Regional Park, Italy), ISrp (San Rossore Regional Park, Italy), ICpr (Castel Porziano Presidential Reserve, Italy), ISar (Sardinia, Italy), IGor (Gorizia, Italy), SCot (El Coto, Spain), FHam (Haute Marne, France), ADif (Austria, different areas), PDif (Poland, different areas), HDif (Hungary, different areas). Subspecies indications on the map are after Groves (1981) and Apollonio *et al*. (1988).

pigs (Pig 6) were not analysed at this marker because of the large amount of data already available in GenBank.

Microsatellite genotyping

A panel of 10 polymorphic microsatellites was selected for the analysis: S026, S215, S355, SW72, SW461, SW857, SW1492, SW2021, SW2496, and SW2532 (details at www.thearkdb.org). Five of them (S026, S215, S355, SW72, and SW857) are in the set recommended by the Food and Agriculture Organization (FAO) to analyse pig diversity (Barker *et al*. 1998), and the rest had been successfully used to study genetic relationships among some European wild boar populations (Vernesi *et al*. 2003). This set was used to genotype all the sampled wild boars and 40 pigs which were raised in wild boar sampling areas (Tuscany and Sardinia). Each PCR was performed in a 10-μL reaction volume, containing 3 μL of DNA solution, 0.5 U of *Taq* DNA polymerase (Euroclone), 1× PCR buffer (Euroclone), 2.5 mM $MgCl_2$, 100 μM of each dNTP and 2 pmol of each primer. The forward primer of each pair was labelled with an ABI fluorescent dye (6-FAM, HEX or TET). The amplification profile was set up with an initial step of denaturation at 95 °C for 3 min, followed by 35 cycles of 92 °C for 45 s, T_a (54–60 °C) for 45 s, and 72 °C for 30 s. A further extension step of 72 °C for 10 min concluded the reaction.

PCR-amplified microsatellite alleles were sized using capillary electrophoresis ABI PRISM automatic sequencers

and internal ROX-500 size standard (Applied Biosystems). The GENEMAPPER software (Applied Biosystems) was used to analyse electrophoretic data.

Genotypes were obtained for all the 252 sampled wild boars, for a subsample of two historical Italian breeds (Pig 1 and Pig 2), and for the Italian commercial pigs (Pig 6).

Mitochondrial DNA data analysis

A total of 192 novel mtDNA CR sequences (411 bp) were obtained (GenBank Accession nos: EU362409–EU362600) and aligned with 612 sequences available in GenBank (Ursing & Arnason 1998; Lin *et al*. 1999; Giuffra *et al*. 2000; Kijas & Andersson 2001; Okumura *et al*. 2001; Kim *et al*. 2002; Randi *et al*. 2002; Alves *et al*. 2003; Yang *et al*. 2003; Gongora *et al*. 2004; Larson *et al*. 2005; Fang & Andersson 2006; Fang *et al*. 2006) using the function CLUSTAL W (Thompson *et al*. 1994) implemented in MEGA 3.0 (Nei & Kumar 2000) and adjusted by eye. Overall, the aligned sequences corresponded to 1065 individual pigs from all over the world (including 231 European wild boars). Haplotypes were collapsed from the entire data set using COLLAPSE version 1.2 (D. Posada, available at http://darwin.uvigo.es), setting deletions as fifth state.

Number of different haplotypes, haplotype (h) and nucleotide (π) diversity, and the mean number of pairwise nucleotide differences between haplotypes (k) were computed using the software ARLEQUIN version 3.01 (Excoffier *et al*. 2005). Allelic richness [$AR_{(r)}$] for each sampled population was calculated from haplotype frequencies using the rarefaction method proposed by El Mousadik & Petit (1996) with the software RAREFAC (R. Petit, www.pierroton.inra.fr/genetics/labo/Software/Rarefac/index.html). The rarefaction size r was set to the smallest sample size among the groups included in the analysis.

A median-joining (MJ) network of haplotypes (Bandelt *et al*. 1999) was created with the software NETWORK 4.1.0.9 (Fluxus Technology), using equal weights for all the mutations and setting the parameter ε to zero, in order to restrict the choice of feasible links in the final network. This approach is especially useful in reconstructing genealogies among closely related taxa, for example, for interpopulation analysis (Bandelt *et al*. 1999). Distributions of pairwise nucleotide differences between haplotypes (mismatch distributions), which are informative on the recent demographic history of a population (Slatkin & Hudson 1991; Rogers & Harpending 1992), were analysed in wild boar and pig populations according to the sudden expansion model as implemented by ARLEQUIN. The age of the expansion was estimated using a generalized nonlinear least-square method, which is based on the minimization of the sum of squared deviations between the observed and the expected mismatch distributions (Schneider & Excoffier 1999). Confidence intervals are obtained using a parametric bootstrap approach based on 1000 simulated samples (Schneider & Excoffier 1999). Finally, Tajima's D (Tajima 1989) and Fu's F_S (Fu 1997) were used to test the null hypothesis of demographic stability, under the common assumption (Avise 1995) that the mtDNA control region can be used as a marker of demographic processes, even though natural selection (mainly at linked regions) cannot be excluded (Ballard & Kreitman 1995). The significance of these statistics was evaluated with 1000 simulated samples as implemented in ARLEQUIN.

Microsatellite data analysis

In order to evaluate the levels of genetic variability in the sampled populations, observed and expected (unbiased gene diversity, Nei 1987) heterozygosities were computed with GENETIX version 4.05 (Belkhir *et al*. 2001). In addition, allelic richness and private allelic richness were calculated per population using HP-RARE (Kalinowski 2005), setting rarefaction according to the smallest sample size in each computation. The possible effect of demographic bottlenecks on the microsatellite variation was tested using the method implemented in the software BOTTLENECK version 1.2 (www.montpellier.inra.fr/URLB/bottleneck/bottleneck.html) (Cornuet & Luikart 1996). The test was performed under three alternative models of microsatellite evolution: the infinite allele model (IAM), the stepwise-mutation model (SMM) and the two-phased model (TPM, set with 10% IAM and 90% SMM). Following the authors' suggestions, the Wilcoxon test was used to test the overall differences between the expected heterozygosity and the heterozygosity predicted from the number of alleles.

Deviations from Hardy–Weinberg equilibrium (HWE) were tested for each population and each locus using the Markov chain method proposed by Guo & Thompson (1992), implemented in the software GENEPOP version 3.4 (Raymond & Rousset 1995). Parameters of the Markov chain expressed as dememorizations/batches/iterations were 10 000/100/5000. The significance level was modified for multiple testing across populations and across loci using the sequential Bonferroni correction (Holm 1979). Deviation from linkage equilibrium (LE) was tested for each pair of loci in each population (810 tests in total) using the log-likelihood ratio approach as implemented in the software FSTAT (Goudet 2001) and a sequential Bonferroni correction (Holm 1979). The minimum significance level for the sequential Bonferroni correction should be set to 0.05/810, when the tests are independent, and to higher values when they are not. To avoid excessive rejections of the linkage disequilibrium hypotheses due to nonindependent LE tests, we set the threshold at 0.05/180, that is, we considered the number of loci times the number of populations, as the effective number of independent tests. This choice did not affect our conclusions, since LE was the rule for our microsatellite

markers. We used MICRO-CHECKER version 2.2.3 (Van Oosterhout *et al*. 2004) to detect signs of the possible occurrence of null alleles, that is, homozygote excess evenly distributed among homozygote size classes at one locus.

In order to evaluate levels of genetic heterogeneity among sampling areas, Weir and Cockerham's estimator of F_{ST} (Weir & Cockerham 1984) was computed using the program GENETIX. Significant deviations from zero were tested over 1000 permutations. Molecular distances between alleles and corresponding indices such as R_{ST} (Slatkin 1995) were not used for the analysis of microsatellites to avoid unpredictable results due to probable multiple-step mutations. F_{ST} values were also used to test the relationship between genetic and geographical distances, using the Mantel test (Mantel 1967) as implemented in ARLEQUIN version 3.01 (Excoffier *et al*. 2005).

Pairwise genetic distances between sampling areas were calculated by the program POPULATIONS version 1.2.28 (Langella 2002). A neighbour-joining (NJ) tree of populations, based on the D_A distance (Nei *et al*. 1983), was used to represent the relationships among groups. A consensus tree was obtained by bootstrapping (1000 replicates) distance values over loci.

A Bayesian cluster analysis was carried out using the method implemented in STRUCTURE version 2.1 (Pritchard *et al*. 2000). We first explored which value of K (number of clusters) maximized the likelihood of the data $[P(D \mid K)]$. Simulations were performed by replicating 10 runs for each value of K comprised between 1 and 20, with the following settings: admixture model (initial $\alpha = 1.0$), no population information, correlated allele frequencies, burn-in length: 20 000, Markov chain Monte Carlo (MCMC) length: 1 000 000. Selected burn-in and MCMC lengths allowed the convergence of the chain. All other parameters were set at their default values. The results were then used to evaluate the most likely partition of our data set, adopting the method proposed by Evanno *et al*. (2005), which relies on the second order rate of change of the likelihood function with respect to K. Once defined the most reliable value of K, the genetic contribution of each inferred cluster to the predefined populations as well as to each individual was investigated.

The Bayesian analysis was also used to study the behaviour of each predefined population when the data set was split into a variable number of clusters, starting from $K = 2$ up to the most reliable value of K. For each K, the run providing the highest value of ln $[P(D)]$ was used. In this descriptive analysis, original groups were assigned to different clades when their composition can be unquestionably assigned to one of the inferred groups. Clearly, this analysis as well as all the tree-based representations implies that early splits (i.e. splits obtained at small K values) can be used to identify the most relevant partitions.

Following the approach introduced by Sacks *et al*. (2004), we also considered the partition which better subdivided the wild boar populations by virtue of their geographical locations. The value of the 'geographical index' (i.e. the average geographical distance between individual locations within clusters divided by the average pairwise distance irrespective of clusters, Sacks *et al*. 2004), which can be computed for each K, is expected to be close to 1 when genetic clusters do not correspond to geographical groups, and lower when each of the K inferred groups includes adjacent populations. When the geographical index stops to decrease for increasing values of K, the geographically meaningful number of groups is reached (Sacks *et al*. 2004).

Results

Mitochondrial DNA variation

A total of 192 mitochondrial CR sequences (411 nucleotides) were analysed: 98 from Italian wild boars, 47 from wild boars sampled in five other European countries, from Spain to Poland, and 47 from five Italian pig breeds. In total, 26 haplotypes (14 of which had never been detected before) and 31 segregating sites (28 substitutions and 3 indels) were identified (see Table S1, Supplementary materials). Different estimates of mitochondrial variability in each population are shown in Table 1. Sample sizes are quite small, and a large variation is expected, and actually observed, across populations at a single locus. However, once we pool the samples into three major groups — wild boars from Italy, wild boars from Europe excluding Italy, and pigs — a clear pattern emerges: genetic diversity in Italy is similar to, or larger than variation observed when several European countries are jointly considered, while pigs show only slightly lower levels of diversity (see Table 1). Sixteen out of 20 haplotypes observed in wild boars are detected in Italy and only seven in the rest of Europe. Of the 28 segregating sites observed in the wild populations, just one is monomorphic in Italy, and 22 in Europe when Italy is not considered. The expectations of these two measures of variation are affected by the sample size, which is clearly larger in the Italian group. But the pattern does not change much when allelic richness, haplotype and nucleotide diversity (i.e. statistics whose expectations are not affected by sample size) are considered (see Table 1).

The high diversity observed in Italian wild boars is due to private mutations and haplotypes, which are mainly related to the exclusive presence in Italy of the three major *Sus scrofa* mtDNA lineages: E1, E2, and A (see Fig. 1a and 2). The codes for these clades were introduced by Giuffra *et al*. (2000) and correspond to clades D1, D4 and D2, respectively, in Larson *et al*. (2005). Clades E1 and A are widely distributed, respectively, in Europe and eastern Asia, while E2 is a second European mtDNA lineage, separated from E1 by five fixed transitions. E2 haplotypes are not observed outside Italy.

Table 1 Genetic variability observed in mtDNA control region sequences (411 bp) in European wild boar populations and in five Italian pig breeds. Values in round brackets correspond to analyses performed excluding the Asian H15 haplotypes (see text). Rarefaction sizes used to compute the allelic richness for each wild boar population and for major groupings are reported in square brackets next to the estimated value

Population	No. of sequences	No. of different haplotypes	Allelic richness	No. of polymorphic sites	Haplotype diversity (*h*)	Nucleotide diversity (π)
IAre	8	4	3.00 [8]	12	0.821	0.013
IFor	10	3	1.98 [8]	5	0.711	0.007
ISie	8	1	0.00 [8]	0	0.000	0.000
ISal	10	6 (5)	4.36 [8]	20 (14)	0.889	0.021
			(4.00[7])		(0.905)	(0.013)
IFlo	10	4	2.78 [8]	11	0.778	0.008
IMrp	11	2	0.99 [8]	9	0.436	0.010
ISrp	10	2	0.80 [8]	9	0.200	0.004
ICpr	10	3	1.80 [8]	2	0.600	0.002
ISar	12	7	4.31 [8]	15	0.864	0.007
IGor	9	1	0.00 [8]	0	0.000	0.000
SCot	9	1	0.00 [8]	0	0.000	0.000
FHam	10	2	0.80 [8]	1	0.200	0.000
ADif	10	4	2.60 [8]	5	0.711	0.006
PDif	8	2	1.00 [8]	2	0.571	0.003
HDif	10	2	0.98 [8]	2	0.356	0.002
WB Italy	98	16 (15)	10.94 [47]	27 (21)	0.874	0.013
			(10.21 [7])		(0.866)	(0.012)
WB Europe (non-Italy)	47	7	6.00 [47]	6	0.829	0.005
WB overall	145	20 (19)	13.08 [47]	28 (22)	0.902	0.011
			(12.47 [47])		(0.898)	(0.010)
Pig 1	10	2	0.78 [5]	2	0.356	0.002
Pig 2	5	3	2.00 [5]	4	0.800	0.005
Pig 3	6	2	0.83 [5]	2	0.333	0.002
Pig 4	11	2	0.88 [5]	1	0.436	0.001
Pig 5	15	5	1.90 [5]	5	0.705	0.005
DP overall	47	10	9.00 [47]	8	0.827	0.005

WB, wild boar; DP, domestic pig.

The sequences in our data set were then aligned with 612 *S. scrofa* sequences available in GenBank and corresponding to European wild boars and pig breeds. In the overall data set, composed of 804 sequences, a total of 114 different haplotypes (coded as H1–H114) were identified. A complete list of sequences and corresponding haplotypes used in the present study is reported in Table S2, Supplementary material. Of the 70 haplotypes observed in European *S. scrofa* (n = 646 individuals), those with the highest frequency and the widest geographical distribution are H22 (137 individuals), H23 (64) and H29 (42), all of which belong to clade E1 and are shared by wild and domestic individuals (see the network in Fig. 2, singletons are not represented). Of the haplotypes classified in the clade E2, H11 (18) and H12 (13) are the most frequent and are both found in four sampling localities in continental Italy. The two most frequent haplotypes of clades E1 and E2 had been already found in a few Italian museum specimens (dating late 18th–early 19th century) analysed by Larson *et al*. (2005). The Asian clade (clade A) is present in European pig breeds and also in a few wild individuals. In Italy, this clade is observed only in three wild boars from Southern Italy (ISal), all sharing the same haplotype H15. This sequence is not uncommon in European pig breeds (see Fig. 2), and was probably transferred to the wild through hybridization events deliberately induced between domestic breeds and Italian wild boar stocks reared in captivity. The haplotype H15 was excluded from further analyses on account of its belonging to a divergent and exotic clade. The genetic variation observed in ISal and overall in Italy when H15 is excluded is also reported in Table 1: the previously described general pattern does not change. None of the pig samples, including those from the five Italian breeds analysed in this study, shows E2 haplotypes (see Table S1).

The mismatch analysis was initially performed on the two groups of samples which were most clearly differentiated at mtDNA sequences according to the presence/absence of clade E2: Italy and Europe excluding Italy. The distributions

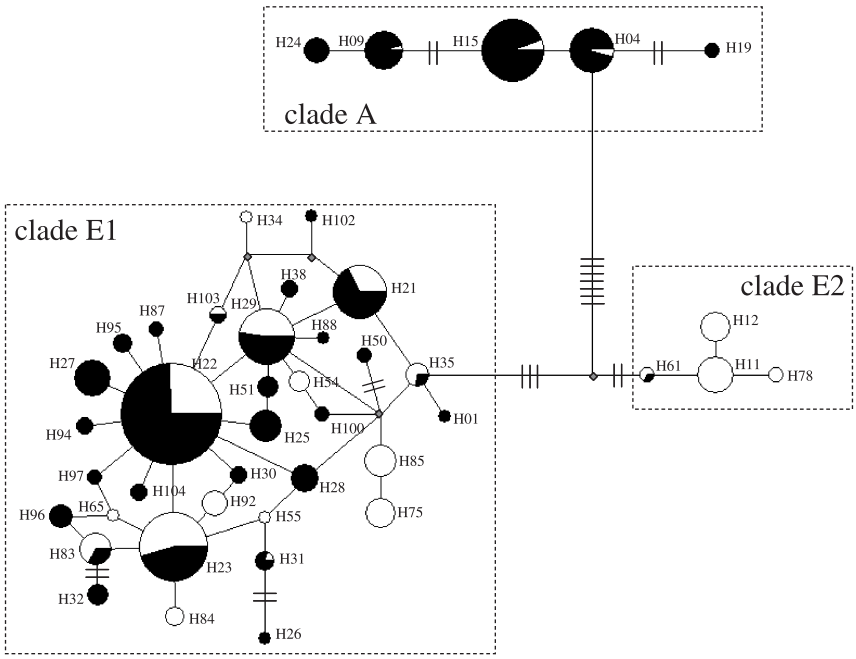


Fig. 2 Median-joining network (Bandelt *et al*. 1999) based on the joint mtDNA data set (original and published sequences). Only haplotypes with a frequency ≥ 2 in the data set were included. Circles indicate sequences observed in wild boar (white) or domestic pig (black) individuals. Size of circles is proportional to the haplotype frequency. Gray diamonds are median vectors, that is haplotypes not observed in the data. Solid branches connecting circles represent single nucleotide changes; lines fragmented by *i* by dashes indicate *i* inferred changes.

obtained either including or leaving out the GenBank sequences (which were possibly affected by non-random sampling) are almost identical. Therefore, only the results obtained with the larger data set are presented. The mismatch distribution in European wild boars ($n = 109$) is smooth and unimodal (Fig. 3a), as expected in the case of a past demographic expansion. The Fu's neutrality test supports the expansion hypothesis ($F_S = –8.25$, $P < 0.01$), whereas the Tajima's D is negative but not significant ($D = –0.84$, $P = 0.22$). The estimated expansion age is 1.36 mutational units (95% CI = 0.58–1.91). On the contrary, the shape of the distribution is clearly ragged and multimodal when the Italian samples are analysed, and the neutrality tests are not significant (Fig. 3b; $n = 116$; $D = 0.95$, $P = 0.86$; $F_S = –0.004$, $P = 0.56$).

The possibility that the clades E1 and E2 observed in Italy simply correspond to two different populations which recently mixed was considered by performing separate mismatch analyses on these clades. According to the results (Fig. 3c, d), the pattern of E1 variation in Italy is not the same as in Europe, since the mismatch distributions have different shapes and the neutrality tests provide different results ($n = 78$; $D = –0.73$, $P = 0.23$; $F_S = –2.30$, $P = 0.17$). The mismatch distribution of the Italian clade E2 is unimodal (Fig. 3d), but Tajima's D and Fu's F_S are far from significance ($n = 38$; $D = –0.11$, $P = 0.48$; $F_S = –0.88$, $P = 0.26$).

Finally, the mismatch distribution was computed for two groups of pig samples, the first including all the European breeds (still excluding the Asian clade) and the second including the Italian breeds only. As regards the former group (European breeds, Fig. 3e), the shape is very similar, although centred at slightly different pairwise difference values, to the distribution observed in the European wild boar (Fig. 3a). Again, neutrality tests point at a demographic expansion ($n = 305$; $D = –1.54$, $P < 0.05$; $F_S = –26.2$, $P < 0.001$), and the estimated expansion age is 0.85 mutational units (95% CI: 0.38–1.12). As regards the latter group (Italian breeds, Fig. 3f), the shape of the distribution is very different from that observed in the Italian wild boar. It partially resembles the European pig breeds distribution (although unusually flat in the central classes), whereas neutrality tests point at a different pattern of genetic variation and do not support the expansion hypothesis ($n = 47$; $D = 0.06$, $P = 0.57$; $F_S = –2.40$, $P = 0.13$).

Genetic variation at microsatellite loci

Between 7 (S026 and S215) and 22 (SW2496) alleles per locus are found across the 292 genotyped individuals. Mean levels of heterozygosity (Table 2) are moderate, averaging 0.57 (SD = 0.05) in the 15 wild boar populations and 0.62 (SD = 0.11) in three pig breeds. Average levels of both heterozygosity and allelic richness are relatively homogeneous across wild boar populations, ranging between 0.47 and 0.62 and between 2.6 and 3.5, respectively. Genetic variation is similar or slightly higher in the joined Italian samples than in the pooled European group (Table 2). An influence on this pattern of individuals from ISal introgressed with Asian pig genomes can be probably excluded. In fact, most populations show higher levels of genetic variation than ISal (Table 2). As regards single populations, Sardinian

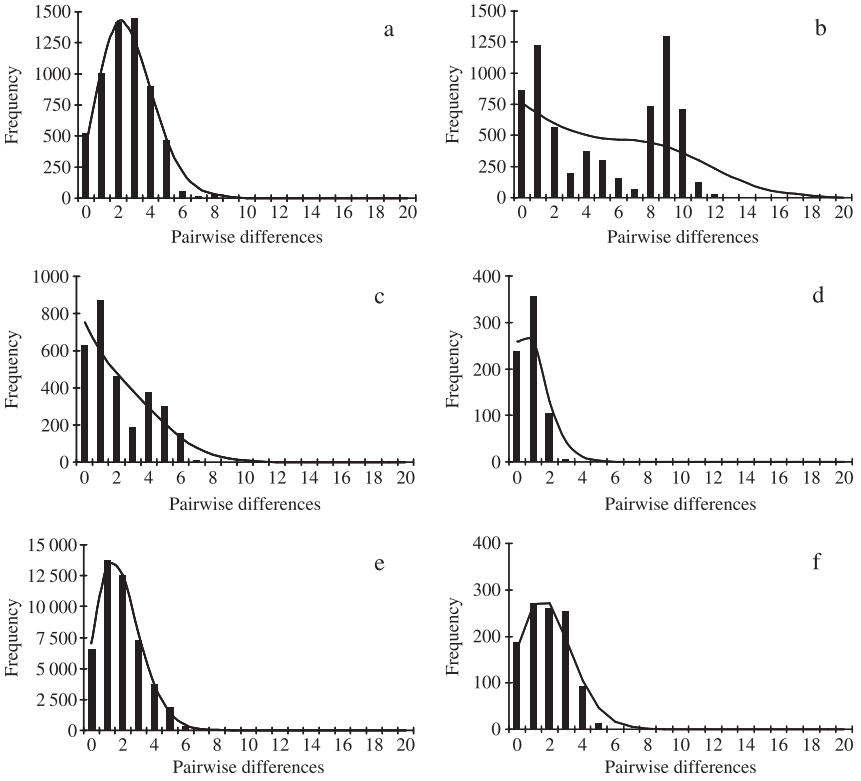


Fig. 3 Mismatch distributions based on pairwise site differences between sequences. (a) European wild boars (excluding Italy); (b) Italian wild boars, all haplotypes; (c) Italian wild boars, clade E1 only; (d) Italian wild boars, clade E2 only; (e) European domestic pigs, clade E1 only; (f) Italian domestic pigs. The expected curve (solid line) was obtained from simulated values computed from the data under the model of demographic expansion. Haplotypes of the Asian clade (A) were excluded from the analysis.

Table 2 Genetic variability detected at microsatellite loci in European wild boar and pig populations. Expected heterozygosity was calculated as gene diversity (Nei 1987). Allelic richness and private allelic richness were calculated with rarefaction set at 14 genes for each population and at 76 genes for groups analyses. Pig 6 was excluded from this analysis for its low sample size

Population	*N*	Average expected heterozygosity	Average observed heterozygosity	Average allelic richness	Average private allelic richness
IAre	10	0.53	0.51	3.84	0.01
IFor	10	0.61	0.59	4.18	0.09
ISie	10	0.62	0.61	3.95	0.10
ISal	10	0.55	0.47	3.56	0.02
IFlo	20	0.62	0.57	4.19	0.13
IMrp	11	0.49	0.45	3.63	0.04
ISrp	15	0.51	0.47	3.04	0.00
ICpr	19	0.54	0.52	3.67	0.01
ISar	41	0.61	0.50	4.27	0.54
IGor	19	0.47	0.47	3.10	0.08
SCot	15	0.60	0.59	3.82	0.26
FHam	20	0.58	0.57	3.88	0.17
ADif	13	0.62	0.55	4.27	0.05
PDif	19	0.61	0.59	3.84	0.07
HDif	20	0.61	0.59	4.45	0.14
WB Italy	165	0.65	0.51	7.96	1.43
WB Europe (non-Italy)	87	0.66	0.58	7.54	1.02
WB overall	252	0.66	0.53	8.44	3.18
Pig 1	22	0.51	0.47	3.31	0.20
Pig 2	13	0.61	0.57	4.26	0.44
Pig 6	5	0.73	0.61	—	—
DP overall	40	0.62	0.52	6.85	1.21

WB, wild boar; DP, domestic pig.

wild boars (ISar) and Sardinian pigs (Pig2) show higher levels of allelic diversity and the highest proportions of private alleles. The lowest values of diversity are found in Northeastern Italy (IGor) and in the enclosed population in San Rossore Regional Park (ISrp).

Averaging across loci, observed heterozygosity is slightly smaller than expected under HWE in all populations. In the locus-by-locus analysis, the number of significant tests after Bonferroni correction is six (out of 180): one in a pig sample (where substructuring, inbreeding, or selection are likely), one in Florence (where translocations are documented, historically and genetically, Vernesi *et al*. 2003), and two each in the probably heterogeneous samples from Poland and Sardinia. Deviations from linkage equilibrium are significant only in 26 pairwise tests (out of 810), concentrated in the pig and the Polish samples. The possible presence of null alleles, tested separately for each locus and each population with MICRO-CHECKER, is limited (13 significant results), randomly distributed across loci, and mainly concentrated again in the Sardinian and the pig samples.

The bottleneck test fails to identify the genetic signature of a demographic decline, no matter the model of microsatellite evolution selected for this analysis. The relative excess of heterozygosity expected in bottlenecked populations (Cornuet & Luikart 1996) is not observed either in the single-locus analysis or in the Wilcoxon test combining the results for the different loci. In contrast, a general deficiency of heterozygosity is found in some populations, but this result is not always consistent across the different mutation models.

Genetic differentiation among populations

The genetic divergence between populations was analysed using the multilocus microsatellite data set. The overall F_{ST} values are relatively high and significant (0.14 and 0.15, excluding and including domestic pigs, respectively; $P < 0.001$ in both cases). Genetic differentiation due to differences between wild and domestic groups is significant but very limited ($F_{CT} = 0.030$, $P = 0.01$). In the wild boar, pairwise F_{ST} values range between 0.00 and 0.31 across Europe (including Italy) and are comparable to those observed between wild boar and domestic pig samples (range 0.10–0.25). Most pairwise values are similar to each other and similar to the global F_{ST}, which is in agreement with a moderately structured tree of populations with short (and moderately supported) internal branches (Fig. 4). No significant correlation was detected between genetic and geographical distances (Mantel test: $r = 0.153$; $P = 0.19$).

When the population structure is analysed with the Bayesian method implemented in the program STRUCTURE, some levels of partitioning in geographically meaningful groups clearly emerge. In the following, we will consider two different partitions which are identified by means of

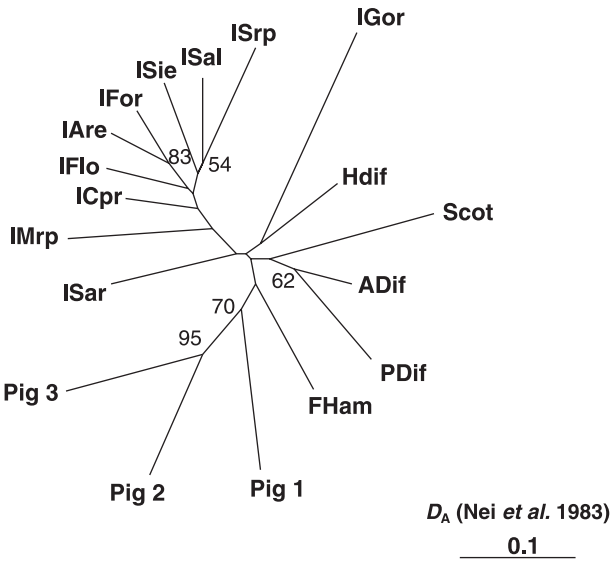


Fig. 4 Neighbour-joining tree based on the Nei's *et al*. (1983) distance (D_A) between populations computed on microsatellite data. Bootstrap support at internodes is shown if > 50%.

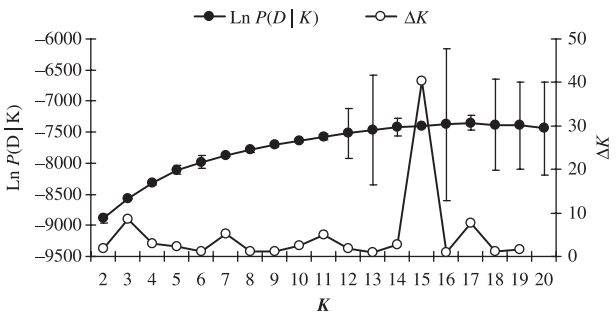


Fig. 5 Posterior probability of the data {ln [$P(D \mid K)$]} and values of ΔK (Evanno *et al*. 2005) as a function of K (number of clusters), as resulting from the simulations in STRUCTURE (structure analysis 1).

the likelihood (structure analysis 1) and the 'geographical index' (structure analysis 2) approaches. The more rigorous likelihood approach identifies 15 clusters (Fig. 5). Accordingly, a detailed analysis of the relationship between inferred clusters and original populations is provided. The 'geographical index' approach identifies four clusters and could thus be regarded as a synthetic representation of the major geographical groups.

Structure analysis 1: likelihood-based partition (Table 3 and Fig. 6)

The large number of clusters and the contribution of each population to each inferred cluster confirm the relatively large genetic divergence between most of the samples. All the populations from Central-Southern Italy, with the exception of IMrp, have major components in the same

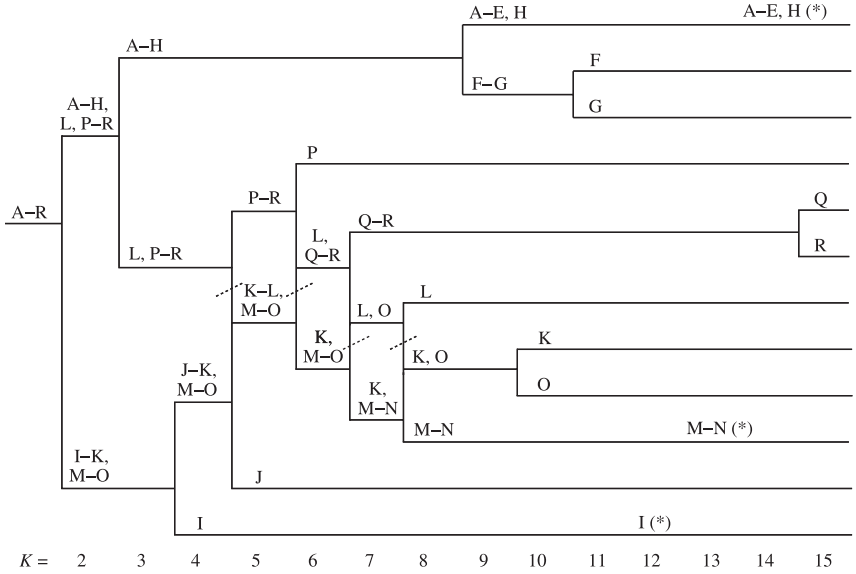


Fig. 6 Diagram showing how populations split into the clusters inferred by the program STRUCTURE as the value of *K* increases. Populations are referred to as: A, IAre; B, IFor; C, ISie; D, ISal; E, IFlo; F, IMrp; G, ISrp; H, ICpr; I, ISar; J, IGor; K, SCot; L, FHam; M, ADif; N, PDif; O, HDif; P, Pig 1; Q, Pig 2; R, Pig 3. Splitting of one population or a group of populations into multiple exclusive clusters is indicated by an asterisk. Reticulations indicate that the attribution of some original groups is not consistent across all *K* values; dashed lines indicate likely solutions, based on clustering at lower *K* values.

Table 3 Partition of the 18 sampled populations into the 15 clusters inferred by the program STRUCTURE (structure analysis 1). This was the most supported value of *K*, obtained by simulated data assuming the admixture model and ignoring population information (10 replicated runs, each with 1 000 000 iterations of data collection after a burn-in of 20 000 iterations). Proportions higher than 0.1 are in bold. Letter codes used in Fig. 6 are reported

			Inferred clusters														
Code	Pop	*N*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A	IAre	10	0.057	0.020	0.016	0.018	0.058	0.015	0.050	0.019	0.006	**0.370**	0.019	0.020	0.016	0.029	**0.286**
B	IFor	10	0.040	0.009	0.039	0.019	0.028	0.011	0.015	0.075	0.005	**0.328**	0.008	0.019	0.096	0.044	**0.265**
C	ISie	10	0.020	0.007	0.010	0.036	0.050	0.025	0.010	0.082	0.013	**0.331**	0.013	0.007	0.030	0.014	**0.353**
D	ISal	10	0.024	0.013	0.029	0.009	0.012	0.017	0.010	**0.257**	0.019	**0.245**	0.011	0.010	0.060	0.007	**0.276**
E	IFlo	20	0.036	0.012	0.017	0.051	0.010	0.016	0.014	0.048	0.013	**0.333**	0.038	0.056	0.041	0.021	**0.295**
F	IMrp	11	0.016	0.009	0.015	0.007	0.014	0.036	0.014	0.064	0.038	0.057	0.007	0.011	**0.652**	0.009	0.052
G	ISrp	15	0.009	0.006	0.009	0.004	0.007	0.009	0.008	**0.862**	0.006	0.025	0.005	0.007	0.010	0.007	0.027
H	ICpr	19	0.048	0.015	0.012	0.013	0.015	0.080	0.033	0.083	0.030	**0.251**	0.009	0.020	0.043	0.041	**0.308**
I	ISar	41	0.020	**0.529**	0.013	0.010	0.015	**0.232**	0.016	0.021	0.031	0.011	0.034	0.027	0.013	0.018	0.011
J	IGor	19	0.011	0.007	0.005	0.006	0.019	0.009	0.010	0.005	**0.893**	0.005	0.006	0.005	0.006	0.006	0.006
K	SCot	15	0.015	0.014	0.013	0.015	**0.802**	0.008	0.007	0.054	0.008	0.012	0.012	0.010	0.012	0.005	0.013
L	FHam	20	0.020	0.010	**0.702**	0.030	0.029	0.015	0.016	0.010	0.009	0.009	0.011	0.055	0.036	0.037	0.010
M	ADif	13	**0.181**	0.020	0.055	**0.178**	0.040	0.012	0.032	0.016	0.034	0.019	**0.196**	0.035	**0.133**	0.029	0.022
N	PDif	19	**0.115**	0.012	0.020	**0.363**	0.061	0.012	0.008	0.022	0.031	0.010	**0.294**	0.007	0.015	0.019	0.011
O	HDif	20	**0.549**	0.016	0.044	0.010	0.026	0.017	0.030	0.037	0.093	0.031	0.033	0.019	0.029	0.034	0.031
P	Pig 1	22	0.013	0.011	0.011	0.011	0.007	0.006	0.015	0.011	0.012	0.014	0.014	0.063	0.009	**0.790**	0.014
Q	Pig 2	13	0.013	0.009	0.007	0.006	0.005	0.009	**0.597**	0.007	0.009	0.009	0.006	**0.289**	0.007	0.017	0.009
R	Pig 3	5	0.013	0.009	0.021	0.036	0.038	0.012	**0.129**	0.010	0.007	0.019	0.023	**0.643**	0.009	0.013	0.018

three inferred clusters (8, 10 and 15, see Table 3). IMrp has one major component in cluster 13, poorly shared by all the other samples. Sardinian and northeastern Italian samples have almost 'private' components, namely clusters 2 and 6, and cluster 9, respectively. Of the European populations, the Spanish and the French samples correspond to private clusters 5 and 3, respectively, while Hungarian, Polish and Austrian wild boars share three cluster components (1, 4, and 11). No one of the wild boar populations shows a relevant proportion of any of the three clusters associated to domestic pigs, although the analysis conducted on an individual basis reveals some levels of admixture. Provided that the three pig samples have a cumulative proportion of clusters 7, 12, and 14, comprised between 78% and 91%, and that the homologous proportion in wild boar populations averages 6%, we arbitrarily classified single wild boars having a cumulative proportion of these clusters > 18% (three times the 'background' proportion observed in

wild boars across Europe) as hybrids. Indeed, the higher similarity between these individuals and the pigs could be possibly explained by introgression and not by common ancestry. Using this threshold, 7% and 9% of the wild boars in Italy and in Europe excluding Italy, respectively, seem to have genomes affected by pig introgression. In Sardinia, only two specimens (5%) fell into this category, but their pig contribution is very high (> 80%). According to the same approach, about 7% of wild boar individuals collected in Italy seems to be affected by introgression of Eastern European wild boar genomes.

The results of the Bayesian analysis obtained with different numbers of inferred groups are summarized hierarchically in Fig. 5. By progressively increasing the number K of components, from $K = 1$ up to $K = 15$ (the X-axis in the diagram), we analysed the partition of each original group into the K-inferred clusters, thus identifying decreasingly important subgroups of samples. Each branch in the figure represents a group inferred at the corresponding K value, while nodes represent the progressive splits of inferred groups when K increases. Original groups were assigned to the branches of the diagram (i.e. to the inferred groups) when their composition was mostly associated to a single inferred group. The most basal split, for $K = 2$, does not separate the domestic and the wild forms as expected: one group actually includes wild boars from Central-Southern Italy and France as well as domestic pigs, whereas the wild boar populations from Sardinia and Northeastern Italy group with the rest of the European populations. For $K = 3$, Central-Southern Italian wild boar populations are separated from pigs, unlike the French one which still groups with the domestic breeds. The Sardinian population is the first original population to emerge separately with a nonshared inferred group, at $K = 4$. The following four steps are somewhat confused because of the French and the Hungarian populations' ups and downs in the diagram branches. This erratic pattern produces the reticulations between $K = 5$ and $K = 8$, which can be, however, easily resolved by considering the earlier assignments of these two populations (see dashed lines in the diagram). Like in previous analyses, the next splits support the divergence of the Northeastern Italian sample (IGor = 5), the clustering of all Central-Southern Italian populations (subdivided into three components only for $K > 10$), and the grouping of most of the other European populations (which starts to split for $K > 7$). Interestingly, the French population is the last group which separates from the domestic pig (only for $K > 7$).

Structure analysis 2: geographically based partitions (Fig. 1b)

This analysis allows for a map-based representation of a reduced number of genetically homogeneous, geographically related groups (see Fig. 1b). It can also be regarded as a zoom of the diagram in Fig. 6 for $K = 4$. All populations from Central-Southern Italy have major components in cluster I, with a cumulative contribution by other clusters in all cases < 30%. Similarly, the Sardinian samples are assigned to the almost 'private' cluster II. Cluster III corresponds to the major genetic component in the rest of the European populations, with the exception of the French sample where the cluster IV prevails. Cluster IV represents the domestic pigs (90% of individual pigs have an assignment probability to this cluster of ≥ 80%). This result is not to be taken as an evidence of pig introgression in France, since in the more accurate structure analysis 1 (see above), none of the French individuals was identified as hybrid. Instead, it supports the hypothesis of a higher-than-average genetic relationship between domestic pigs and French wild boars (see also Figs 4 and 6). When compared with the distribution of the mtDNA clades (see Fig. 1a), the results of this analysis confirm the remarkable divergence between European and Central-Southern Italian wild boars. They additionally suggest that, notwithstanding the frequency of the typical Italian mtDNA clade (E2), large contributions of exotic gene pools can be excluded in all Italian populations.

Discussion

By the analysis of one mitochondrial and 10 independent nuclear markers in wild boars collected across Europe, we evaluated the possible contribution of different natural and human-related processes in shaping the present genetic diversity of the species in the old continent. The role of past wide-scale events, like range and size fluctuations occurring during the last glacial and postglacial periods, are pointed to as the main force leading to the observed levels of differentiation in Europe. In addition, the comparison of the genetic diversity of wild and domestic pigs is compatible with the idea of a domestication centre in Central Europe, but it does not provide support to the hypothesis of an independent domestication in the Italian peninsula.

Postglacial dispersal

Our comparative analysis of wild boar control region sequences shows the signature of a past demographic expansion, which could well have followed the range contraction occurring in Europe during the last glaciation. Assuming a sudden expansion model, this event can be dated back to approximately 1.36 mutational units in the past. This figure is obviously imprecise, because of stochastic variations in the coalescent and the sampling processes, and its conversion into years appears even more difficult. At least a fivefold range of variation can be found in the per-site/per-year mutation rates which have been used in several studies on the hypervariable mtDNA control region

in mammals. For the analysis of ungulates in particular, Randi *et al*. (1998) suggested a range between 4% and 8% sequence divergence per site per million years, while Birungi & Arctander (2000) suggested an interval between 10% and 20%. Accordingly, if we consider only the errors associated to the calibration, the expansion age of the wild boar should be dated back to between 16 000 BP and 80 000 BP. Several authors argued for a mutation rate closer to the upper limit of these intervals (e.g. Bradley *et al*. 1996 and Kim *et al*. 2002 opted for a 15% rate), and we therefore believe that the molecular data we analysed support the hypothesis of a demographic expansion of the wild boar following the last glaciation. Interestingly, the age of the expansion estimated from the similarly unimodal mismatch distribution in the European domestic pig data set (Fig. 3e) is about 40% younger, that is, 10 000 years ago using a 20% rate. Considering the errors associated to these estimates, also for the reason that the method assumes a single demographic event, these data appear consistent with a pig Neolithic expansion associated to the domestication process, or they might actually correspond to a sort of average that keeps also the signs of the postglaciation expansion of their wild ancestors.

Two additional results support the hypothesis that the pattern of genetic diversity in the European wild boar was shaped by the last postglacial colonization event from one or more southern refugia. First, the divergence between European samples outside Italy is similar in different comparisons, and no correlation is observed between genetic and geographical distances. This finding is not consistent with a stable equilibrium between drift and migration (i.e. under isolation by distance), while it is expected in case of a rapid colonization and a simultaneous demographic expansion from a common source population. Second, the mitochondrial variation in Italy is not compatible with a demographic expansion, and both mtDNA and nuclear markers show a similar or even higher level of variation in Italy than in the rest of the European regions jointly analysed. This result is remarkable at the mtDNA: two major clades, E1 and E2, separated by at least 50 000 years, are observed in Italy, whereas only E1 is found elsewhere. The simplest explanation for this pattern seems to be the process of contraction into southern refugia and the following re-expansion towards northern areas. The recent finding (Larson *et al*. 2007) that E2 haplotypes were present in the present-day Croatia about 11 000 years ago is not unexpected under this view, considering that Northern Adriatic was not submerged during the last glaciation (Van Andel 1989). According to this scenario, Italian diversity represents a large fraction of the wild boar preglaciation diversity, which was then preserved without any major impact by subsequent demographic processes. Different hypotheses about the location of source refugia appear equally likely: it could be Italy itself, with the loss of rare E2 haplotypes during the colonization, or some other southern areas in Europe, or a combination of these two. Actually, the Italian peninsula seems to have played for several species a minor role in the recolonization processes than the Iberian and Balkan peninsulas, probably on account of the Alps being a greater physical barrier to the dispersal of individuals (Hewitt 2000). A wider sampling in Iberia and in the Balkans will be necessary to identify which refugium area contributed most to the present gene pool of the European wild boar. Similarly, more locations throughout Europe should be screened for the presence of E2 haplotypes, as to exclude that they occur outside Italy at low frequency.

Domestication and hybridization with domestic pigs

Wild boars and domestic pigs share the most common mtDNA haplotypes and microsatellite alleles, and the population tree (Fig. 4) only weakly resolves the two groups. Also the Bayesian clustering analysis confirms that wild and domestic forms are not more divergent than other pairs of wild boar populations. It seems therefore that the differentiation from the wild boar during the domestication, remarkable at morphological traits and thus probably also at their genetic determinants, was not accompanied by a strong founder effect. As already suggested in relation to other domestication events which implied a long co-existence between the domestic and the wild forms (e.g. dog and cattle), occasional and/or deliberate hybridizations could have played a role in reducing their genetic divergence (Vilà *et al*. 2005; Beja-Pereira *et al*. 2006). A fraction between 5% and 10% of the wild boar individuals we analysed shows the effects of pig introgression, but the global contribution of pig genomes in the wild populations is clearly lower (as supported by the population analysis). Remarkably, we even found wild boar individuals in Southern Italy with Asian pig mtDNA, usually observed in some ameliorated European breeds crossbred with Asian pigs (Giuffra *et al*. 2000). This evidence is consistent with the observation by Fang *et al*. (2006), who assayed an approximately 10-fold lower frequency of Asian haplotypes in wild boar populations than in domestic breeds in Europe (3% vs. 30%). This result supports the view that some levels of hybridization between wild boars and domestic pigs occurred in the past and possibly still occur today.

Domestication: the origin of European breeds

The similarity between European pig breeds and wild boars, when considered in connection with the finding that Middle Eastern mtDNA lineages are not observed in European pig breeds (Larson *et al*. 2005), unequivocally suggests that modern breeds in Europe descend from local wild populations. Additionally, the population tree and the Bayesian inference of population clustering based on

nuclear markers indicate that the wild boar population to be most closely related to the pig is located in northern France. This result is compatible with the hypothesis of Larson *et al*. (2005, 2007) based on modern and ancient mtDNA sequences, whereby Central Europe was an important domestication centre. A second domestication centre for the European breeds was actually proposed by Larson *et al*. (2005) to be located in Southern Europe, and more specifically in mainland Italy. This hypothesis, although intriguing, was supported only by two museum Sardinian specimens with possible feral origin showing E2 mtDNA sequences. Similarly, Larson *et al*. (2007) found two Bronze Age Sardinian individuals morphologically classified as 'domestic/feral' showing E2 sequences, although all ancient Italian samples firmly attributed to domestic pigs had European E1 haplotypes. In our analysis at both mtDNA and nuclear markers, we do not find any specific relationship between Italian pigs and Italian wild boars, and no E2 haplotype is found in the 47 pigs belonging to the five local breeds we analysed. Although the pattern of genetic variation in the Italian breeds suggests a different demographic history from that of other European breeds (no signs of demographic expansion are detected), our data seem to exclude an independent domestication event in Italy, or, at least, that pigs possibly domesticated in Italy have left descendants in modern breeds.

Overhunting and demographic decline

The genetic impact of the demographic decline that affected several wild boar populations in Europe during the last two centuries seems to be extremely limited, if any.

In general, the level of genetic variations at the mtDNA sequences in the whole data set is within the range observed in other ungulate populations (Vernesi *et al*. 2002; Feulner *et al*. 2004; Randi *et al*. 2004). The average microsatellite heterozygosity (0.66) is only slightly lower than the average value observed in 14 non-endangered species (0.70) and much higher than the average value of 0.41 found across 14 threatened taxa (Frankham *et al*. 2004). The genetic variation, as expected in nonpanmictic species, is lower within single localities, with large differences between samples especially at the mtDNA region. This is probably a consequence of the increase of drift effects due to the smaller effective population sizes at this marker, and possibly this is also related to our small sample sizes. We note however that the levels of variation are relatively high almost everywhere, also within samples, with only two restricted groups in as many regional parks (ISrp and IMrp) and one sample with an acknowledged history of recent recolonization (IGor), having very low variation both at mitochondrial and microsatellite loci.

More specifically, the mismatch analysis and two neutrality tests applied to mtDNA sequences are all indicative of either stability or expansion, and a specific method developed to identify a bottleneck using microsatellite markers does not reveal the deviation expected in case of demographic decline. In other words, both the levels and the patterns of diversity observed at mtDNA sequences and microsatellites are consistent with the idea that the population size and distribution range contractions did not affect the overall genetic variability.

Recent expansions and translocations

The possibility to detect the genetic effects of the rapid growth occurred in the last 50 years in several regions is questionable. Provided that a few generations of expansion are not sufficient to accumulate enough mutations in the gene genealogy, to capture statistically the reduced drift effects which are typical of this process is very difficult. However, it is interesting to note that a first attempt to identify these effects using an approximate Bayesian computation (ABC) approach (Beaumont *et al*. 2002), which can be used to model complex demographic processes, suggests that a large growth rate occurred in the last 10–20 generations and was larger in Central Italy than anywhere else in Europe (results not shown). Therefore a tentative conclusion would be that Italy still preserves the preglaciation diversity, which was not severely affected by the decline occurred from the Middle Ages until the end of World War II, and was recently frozen by the modern expansion.

As for the translocation events, it was suggested that poorly documented and usually uncontrolled restocking plans drastically affected genetic variation in Central and Southern Italy, and also speeded up the recent expansion process (Randi 2005; Apollonio *et al*. in press). Since translocations were carried out using also animals from Central Europe, this might have artificially contributed to the higher genetic variation observed in Italy. However, in agreement with Vernesi *et al*. (2003), we can exclude a major impact of such human-mediated migrations. If this phenomenon were extensive, we would have observed (i) the clustering of some Central and Southern Italian samples with some Central European groups; (ii) the presence of inferred clusters (in the Bayesian analysis) shared by some Central and Southern Italian and Central European groups; (iii) a large proportion of individuals from Central and Southern Italy assigned to other European populations. None of these predictions is met by our genetic analyses, and only a limited fraction (7%) of individuals sampled in Italy have significant proportions of their genomes that can be related to wild boar populations abroad. In addition, the plausible hypothesis that the native boars in Italy had E2 haplotypes, while E1 haplotypes were introduced by recent translocations of wild boars from Central Europe is contradicted by the fact that the variation patterns of E1 sequences in Italy and in the rest of Europe are different. We therefore conclude

that restocking from Central Europe had a limited genetic impact and an accordingly a marginal role in the recovery of the Italian population.

Conclusions and management implications

In conclusion, the most important event in shaping the observed pattern of diversity seems to have been the last glaciation, which was followed by a sudden demographic and spatial expansion from one or more southern refugia. The genetic signature of more recent processes, which were mostly related to human activities, can be detected but it appears marginal. Clearly, a wider sampling of European populations, including several locations in the Iberian Peninsula and in the Balkans, is necessary to better clarify postglacial dynamics. However, our data point to a single area of discontinuity which corresponds to the Alps. Wild boars sampled south of this chain show, indeed, higher levels of overall genetic variation, a private mtDNA haplogroup and endemic diversity at microsatellite loci.

The wild boar is a rather invasive species, with a relevant impact on biodiversity, agriculture and livestock, and it is also of large interest for hunters. Therefore, it is important to consider the implications of our results for the management of this species. We believe that (i) Italian populations represent a reservoir of genetic diversity in Europe which should be preserved; for example, following the arguments in Petit *et al*. (1998), in the eventual case of extinction of the wild boar in Italy, the reduction of nuclear allelic richness would be about twice as much as in the case of the simultaneous extinction in five other European regions (11% compared to 6%), and a high divergent mtDNA clade would be lost; however, since almost all the Italian groups are genetically very similar, and even documented demographic reductions did not affect significantly the genetic variation, hunting is still recommended to reduce the population size in some areas; (ii) animals should not be translocated from one European region to another, especially across the Alps; even though the level of differentiation is probably not enough to maintain the current subdivision into subspecies, some level of local adaptation is expected and should not be compromised by hybridization; accidental escapes from wild boar farms should also be prevented for the same reason; (iii) artificial crossbreeding with domestic pigs should be avoided and genetic controls in wild boar farms should be enforced. A different matter are those situations where the genetic exchange between wild and domestic pigs has a historical background, due to the long-lasting practice to rear pigs in a natural state (e.g. in Sardinia). In these cases, the present genetic identity of the wild population is influenced by the prolonged gene flow between the two forms, and thus, in absence of significant introgressions from other areas, they can be referred to as a joint evolutionary unit.

Acknowledgements

We are deeply grateful to all those who collaborated with us in collecting samples: Henrik Okarma (Poland), Laszlo Sugar (Hungary), Juan Carranza (Spain), Franz Suchentrunk (Austria), Eric Baubet (France), Roberto Mazzoni Della Stella (Siena), Stefano Antonacci (Gorizia), Paolo Varuzza (Salerno), Luca Tonini (Maremma Regional Park), Riccardo Bozzi and several local hunters and pig breeders. The Istituto Nazionale per la Fauna Selvatica (INFS) kindly allowed us to collect samples from the Castel Porziano Presidential Reserve. We also thank Lidia Migliori, Giovanni Manca and David Caramelli for technical assistance, and Umberto Albarella for useful suggestions to an early draft of the manuscript. Three anonymous referees made constructive and valuable comments on this paper. The study was possible thanks to the financial contribution of the Sardinian Regional Government, the University of Ferrara, The Italian Ministry for University Research (MIUR PRIN 2005) the Centre for Alpine Ecology, and the Autonomous Province of Trento. L. Iacolina benefited by a fellowship sponsored by the Sardegna Resorts srl.

References

Alves A, Ovilo C, Rodriguez MC, Siliò L (2003) Mitochondrial DNA sequence variation and phylogenetic relationships among Iberian pigs and others domestic and wild pig populations. *Animal Genetics*, **34**, 319–324.

Apollonio M, Randi E, Toso S (1988) The systematics of the wild boar (*Sus scrofa* L.) in Italy. *Bollettino Di Zoologia*, **3**, 213–221.

Apollonio M, Ciuti S, Pedrotti L, Banti P (in press) Ungulate management in Italy in the XXI Century. In: *Ungulate Management in Europe in the XXI Century* (eds Apollonio M, Andersen R, Putman R). Cambridge University Press, Cambridge, UK.

Avise JC (1995) Mitochondrial DNA polymorphism and a connection between genetics and demography of relevance to conservation. *Conservation Biology*, **9**, 686–690.

Avise JC (2004) *Molecular Markers Natural History, and Evolution*, 2nd edn. Sinauer & Associates, Sunderland, Massachusetts.

Ballard JW, Kreitman M (1995) Is mitochondrial DNA a strictly neutral marker? *Trends in Ecology & Evolution*, **10**, 485–488.

Bandelt H-J, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37–48.

Barker JSF, Hill WG, Bradley D, Nei M, Fries R, Wayne RK (1998) *Measurement of Domestic Animal Diversity (MoDAD): Original Working Group Report*. FAO, Rome.

Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian Computation in population genetics. *Genetics*, **162**, 2025–2035.

Beja-Pereira A, Caramelli D, Lalueza-Fox C *et al*. (2006) The origin of European cattle: evidence from modern and ancient DNA. *Proceedings of the National Academy of Sciences, USA*, **103**, 8113–8118.

Belkhir K, Borsa P, Chikhi L, Raufauste N, Bonhomme F (2001) *GENETIX 4.02, logiciel sous WindowsTM pour la génétique des populations*. Laboratoire Génome, Populations, Interactions. CNRS UMR 5000, Université de Montpellier II, Montpellier (France). Available from URL: http://www.genetix.univ-montpz.fr/genetix/genetix.htm

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

Birungi J, Arctander P (2000) Large sequence divergence of mitochondrial DNA genotypes of the control region within populations of the African antelope, kob (*Kobus kob*). *Molecular Ecology*, **9**, 1997–2008.

Boitani L, Trapanese P, Mattei L, Nonis D (1995) Demography of a wild boar (*Sus scrofa*, L.) population, Tuscany, Italy. *Gibier Faune Sauvage*, **12**, 109–132.

Bradley DG, MacHugh DE, Cunningham P, Loftus RT (1996) Mitochondrial diversity and the origins of African and European cattle. *Proceedings of the National Academy of Sciences, USA*, **93**, 5131–5135.

Briedermann L (1990) *Schwarzwild*. VEB Deutscher Landwirtschaftsverlag, Berlin, Germany.

Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, **144**, 2001–2014.

Danilkin AA (2001) The wild boar: an unprecedented spread or restoration of the species range? *Doklady Biological Sciences*, **380**, 457–460.

El Mousadik A, Petit RJ (1996) High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. *Theoretical and Applied Genetics*, **92**, 832–839.

Epstein J, Bichard M (1984) Pig. In: *Evolution of Domesticated Animals* (ed. Mason IL), pp. 145–162. Longman, New York.

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

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

Fang M, Andersson L (2006) Mitochondrial diversity in European and Chinese pigs is consistent with population expansions that occurred prior to domestication. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 1803–1810.

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 $2n = 38$. *Animal Genetics*, **37**, 459–464.

Feichtner B (1998) Ursachen der Streckenschwankungen beim Schwarzwild im Saarland. *Zeitschrift für Jagdwissenschaft*, **44**, 140–150.

Fernandez-Llario P, Mateos-Quesada P (1998) Body size and reproductive parameters in the wild boar *Sus scrofa*. *Acta Theriologica*, **43**, 439–444.

Feulner PGC, Bielfeldt W, Zachos FE *et al*. (2004) Mitochondrial DNA and microsatellite analyses of the genetic status of the presumed subspecies *Cervus elaphus montanus* (Carpathian red deer). *Heredity*, **93**, 299–306.

Frankham R, Ballou JD, Briscoe DA (2004) *A Primer of Conservation Genetics*. Cambridge University Press, Cambridge, UK.

Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 915–925.

Geisser H, Reyer H-U (2004) Efficacy of hunting, feeding, and fencing to reduce crop damage by wild boars. *Journal of Wildlife Management*, **68**, 939–946.

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

Genov P, Nikolov H, Massei G, Gerasimov S (1991) Craniometrical analysis of Bulgarian wild boar (*Sus scrofa*) populations. *Journal of Zoology, London*, **225**, 309–325.

Giuffra E, Kijas JMH, Amarger V *et al*. (2000) The origin of the domestic pig: independent domestication and subsequent introgression. *Genetics*, **154**, 1785–1791.

Gongora J, Fleming P, Spencer PBS *et al*. (2004) Phylogenetic relationships of Australian and New Zealand feral pigs assessed by mitochondrial control region sequence and nuclear GPIP genotype. *Molecular Phylogenetics and Evolution*, **33**, 339–348.

Goudet J (2001) FSTAT, *a Program to Estimate and Test Gene Diversities and Fixation Indices Version 2.9.3*. Available from URL: http://www.unil.ch/izea/softwares/fstat.html.

Groves C (1981) *Ancestors for the Pigs: Taxonomy and Phylogeny of the Genus Sus*. Research School of Pacific Studies, Australian National University, Canberra, Australia.

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

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

Hofreiter M, Serre D, Rohland N *et al*. (2004) Lack of phylogeography in European mammals before the last glaciation. *Proceedings of the National Academy of Sciences, USA*, **101**, 12963–12968.

Holm S (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, **6**, 54–70.

Jedrzejewska B, Jedrzejewski W, Bunevich AN, Milkowski L, Krasinski A (1997) Factors shaping population densities and increased rates of ungulates in Bialowieza Primeval Forest (Poland and Belarus) in the 19th and 20th centuries. *Acta Theriologica*, **42**, 399–451.

Kalinowski ST (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes*, **5**, 187–189.

Kijas JMH, Andersson L (2001) A phylogenetic study of the origin of the domestic pig estimated from the near-complete mtDNA genome. *Journal of Molecular Evolution*, **52**, 302–308.

Kim K-I, Lee J-H, Li K *et al*. (2002) Phylogenetic relationship of Asian and European pig breeds determined by mitochondrial DNA D-loop sequence polymorphism. *Animal Genetics*, **33**, 19–25.

Langella O (2002) POPULATIONS 1.2.28. *Population Genetic Software (Individuals or Populations Distances, Phylogenetic Trees)*. Available from URL: http://www.bioinformatics.org/project/?group_id=84.

Larson G, Dobney K, Albarella U *et al*. (2005) Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science*, **307**, 1618–1621.

Larson G, Albarella U, Dobney K *et al*. (2007) Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *Proceedings of the National Academy of Sciences, USA*, **104**, 15276–15281.

Lin C-S, Sun Y-L, Liu C-Y *et al*. (1999) Complete nucleotide sequence of pig (*Sus scrofa*) mitochondrial genome and dating evolutionary divergence within Artiodactyla. *Gene*, **236**, 107–114.

Lucchini V, Meijaard E, Diong CH, Groves CP, Randi E (2005) New phylogenetic perspectives among species of South-east Asian wild pig (*Sus* sp.) based on mtDNA sequences and morphometric data. *Journal of Zoology*, **266**, 25–35.

Magri D, Vendramin GG, Comps B *et al*. (2006) A new scenario for the Quaternary history of European beech populations: palaeobotanical evidence and genetic consequences. *New Phytologist*, **171**, 199–221.

Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209–220.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nei M, Kumar S (2000) *Molecular Evolution and Phylogenetics*. Oxford University Press, Oxford, UK.

Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. *Journal of Molecular Evolution*, **19**, 153–170.

Okumura N, Kurosawa Y, Kobayashi E *et al*. (2001) Genetic relationship amongst the major non-coding regions of mitochondrial DNAs in wild boars and several breeds of domesticated pigs. *Animal Genetics*, **32**, 139–147.

Oliver WLR, Brisbin IL, Takahashi S (1993) The Eurasian wild pig, *Sus scrofa*. In *Pigs, Peccaries and Hippos – Status Survey and Action Plan*. (ed. Oliver WLR), pp. 112–121. IUCN/SSC, Gland, Switzerland.

Petit RJ, El Mousadik A, Pons O (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation Biology*, **12**, 844–855.

Petit RJ, Aguinagalde I, de Beaulieu J-L *et al*. (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. *Science*, **300**, 1563–1565.

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

Randi E (2005) Management of wild ungulate populations in Italy: captive-breeding, hybridisation and genetic consequences of translocations. *Veterinary Research Communications*, **29**, 71–75.

Randi E, Apollonio M, Toso S (1989) The systematics of some Italian populations of wild boar (*Sus scrofa* L.): a craniometric and electrophoretic analysis. *Zeitschrift für Säugetierkunde*, **54**, 40–56.

Randi E, Mucci N, Pierpaoli M, Douzery E (1998) New phylogenetic perspectives on the Cervidae (Artiodactyla) are provided by the mitochondrial cytochrome *b* gene. *Proceedings of the Royal Society B: Biological Sciences*, **265**, 793–801.

Randi E, D'Huart JP, Lucchini V, Aman R (2002) Evidence of two genetically deeply divergent species of waterhog, *Phacochoerus africanus* and *P. aethiopicus* (Artiodactyla: Suiformes) in East Africa. *Mammalian Biology*, **67**, 91–96.

Randi E, Alves PC, Carranza J, Milošević-Zlantanović S, Sfougaris A (2004) Phylogeography of roe deer (*Capreolus capreolus*) populations: the effects of historical genetic subdivisions and recent nonequilibrium dynamics. *Molecular Ecology*, **13**, 3071–3083.

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

Rogers A, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, **9**, 552–569.

Sacks BN, Brown SK, Ernest HB (2004) Population structure of California coyotes corresponds to habitat-specific breaks and illuminates species history. *Molecular Ecology*, **13**, 1265–1275.

Sáez-Royuela C, Tellería JL (1986) The increased population of the wild boar (*Sus scrofa* L.) in Europe. *Mammal Review*, **16**, 97–101.

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

Sbisà E, Tanzariello F, Reyes A, Pesole G, Saccone C (1997) Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. *Gene*, **205**, 125–140.

Schley L, Roper TJ (2003) Diet of wild boar *Sus scrofa*. Western Europe, with particular reference to consumption of agricultural crops. *Mammal Review*, **33**, 43–56.

Schneider S, Excoffier L (1999) Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics*, **152**, 1079–1089.

Singer FJ, Swank WT, Clebsch EEC (1984) Effects of wild pig rooting in a deciduous forest. *Journal of Wildlife Management*, **48**, 464–473.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457–462.

Slatkin M, Hudson RR (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics*, **129**, 555–562.

Taberlet P, Fumagalli L, Wust-Saucy A-G, Cosson J-F (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453–464.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**, 4673–4680.

Ursing BM, Arnason U (1998) The complete mitochondrial DNA sequence of the pig (*Sus scrofa*). *Journal of Molecular Evolution*, **47**, 302–306.

Van Andel TH (1989) Late Quaternary sea-level changes and archaeology. *Antiquity*, **63**, 733–745.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) Micro-checker: a software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.

Vernesi C, Pecchioli E, Caramelli D *et al*. (2002) The genetic structure of natural and reintroduced roe deer (*Capreolus capreolus*) populations in the Alps and central Italy, with reference to the mitochondrial DNA phylogeography of Europe. *Molecular Ecology*, **11**, 1285–1297.

Vernesi C, Crestanello B, Pecchioli E *et al*. (2003) The genetic impact of demographic decline and reintroduction in the wild boar (*Sus scrofa*): a microsatellite analysis. *Molecular Ecology*, **12**, 585–595.

Vilà C, Seddon J, Ellegren H (2005) Genes of domestic mammals augmented by backcrossing with wild ancestors. *Trends in Genetics*, **21**, 214–218.

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

Welander J (2000) Spatial and temporal dynamics of wild boar (*Sus scrofa*) rooting in a mosaic landscape. *Journal of Zoology*, **252**, 263–271.

Yang J, Wang J, Kijas J, Liu B, Han HYuM, Yang H, Zhao S, Li K (2003) Genetic diversity present within the near-complete mtDNA genome of 17 breeds of indigenous Chinese pigs. *Journal of Heredity*, **94**, 381–385.

This research is a result of the collaboration among the Department of Zoology and Evolutionary Genetics of the University of Sassari, the Department of Biology and Evolution of the University of Ferrara, the Centre of Alpine Ecology in Trento, and the Biosfera Association in Florence. Massimo Scandura is a researcher and Laura Iacolina and M. Francesca Di Benedetto are PhD students, members of the team led by Marco Apollonio, interested in the behavioural ecology and population genetics of wild mammals. Elena Pecchioli and Barbara Crestanello are researchers interested in conservation genetics of wild vertebrates in mountain ecosystems. Vincenzo Russo and Roberta Davoli work on gene expression in cattle and pig breeds at the Zootechnical Center of the University of Bologna. Giorgio Bertorelle is interested in the reconstruction of demographic and selective processes using population genetics data in different species.

Supplementary materials

The following supplementary material is available for this article:

Table S1 Segregating sites and different haplotypes observed at the mtDNA control region (411 bp) of 145 European wild boar and 47 pig samples sequenced in this study. Haplotype frequencies observed in wild boar sampling areas and pig breeds are shown. The following abbreviations are used for wild boar sampling areas: IAre (Arezzo, Italy), IFor (Forlì, Italy), ISie (Siena, Italy), ISal (Salerno, Italy), IFlo (Florence, Italy), IMrp (Maremma Regional Park, Italy), ISrp (San Rossore Regional Park, Italy), ICpr (Castel Porziano Presidential Reserve, Italy), ISar (Sardinia, Italy), IGor (Gorizia, Italy), SCot (El Coto, Spain), FHam (Haute Marne, France), ADif (Austria, different areas), PDif (Poland, different areas), HDif (Hungary, different areas). The following abbreviations are used for domestic pig breeds: Pig 1 (Cinta Senese), Pig 2 (Sarda), Pig 3 (Calabrese), Pig 4 (Casertana) and Pig 5 (Nera Siciliana).

Table S2 List of the 804 *Sus scrofa* mitochondrial control region sequences downloaded from GenBank or obtained by the authors which were used in the study.

This material is available as part of the online article from:
http://www.blackwell-synergy.com/doi/abs/
10.1111/j.1365-294X.2008.03703.x
(This link will take you to the article abstract).

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

SECOND PART

Social structure in a wild boar population

CHAPTER II

Non-kin association in wild boar social units

Non-kin association in wild boar social units

Laura IACOLINA, Massimo SCANDURA, Paolo BONGI and Marco APOLLONIO

Dipartimento di Zoologia e Genetica Evoluzionistica, Università di Sassari, Via Muroni 25, I-07100 Sassari, Italy.

Submitted to Journal of Mammalogy

ABSTRACT

We investigated the social organization of wild boar (*Sus scrofa*) using genetic and spatial data from a study population in Tuscany, Italy. In total 120 wild boars of different sexes and age classes were captured and monitored from 2002 to 2006. All of them were genetically analysed by using 10 polymorphic microsatellites (H_e = 0.693, k = 6.6) and a matrix of pairwise relatedness was calculated. In addition, a reference sample of fully related individuals was created by genotyping 11 sows and their foetuses (n = 56). Spatial data were gathered for 65 animals which had been fitted with either radiocollars or ear transmitters. Sixteen social units were identified by capture data and confirmed by observations and telemetry. A correlation between inter-individual spatial distance and relatedness was observed only in summer-early autumn and seemed to be associated to the presence of piglets. The prediction of matrilinearity in wild boar social units was not confirmed, as a high proportion of unrelated boars was observed within groups. Aggregations of unrelated adult sows (with their litter) and unrelated yearlings were detected in the study population. The high turn-over rate of the population due to human-caused mortality seems to be the main factor responsible for this altered social structure. Accordingly, the observed social organization would turn out from grouping of unrelated survivors, that is promoted by the need to cope with predation risk by wolves.

Keywords: genetic relatedness, microsatellites, social structure, *Sus scrofa*, wild boar.

INTRODUCTION

Gene dynamics within a population are strongly influenced by breeding system, social structure and dispersal patterns (Apollonio and Hartl 1993, Chepko-Sade and Tang Halpin 1987, Storz 1999). Polygynous breeding and female philopatry are the rule in mammals, and a huge variation can be observed in their social structures, ranging from primarily solitary to highly social ones (Eisenberg and Kleiman 1983). Social systems may also differ from population to population, as a response to different ecological constraints and management practices (Lott 1991, Pope 1998). In any case, this variation has a profound impact on the genetic features of populations (Dobson 1998, Storz 1999). Conversely, the study of the genetic structure of a population at a fine scale can prove helpful in describing its social organization (Sugg et al. 1996). In particular, the knowledge of the genetic relationships between individuals in a population can disclose hidden social interactions which are important to fully understand the behavioural ecology of the species (e.g. non-kin based cooperation, Blundell et al. 2004).

The application of molecular techniques provides a tool to establish kin relationships within a population, thus enabling one to test multiple hypotheses in relation to the spatial and social organization of the species under study. Molecular data have revealed close spatial associations among kin in several mammal species, including Florida black bear (*Ursus americanus floridanus*, Moyer et al. 2006), raccoon (*Procyon lotor*, Ratnayeke et al. 2002), grey mouse lemur (*Microcebus murinus*, Wimmer et al. 2002) and African lion (*Panthera leo*, Spong and Creel 2004). However, the hypothesis that relatedness influences the spatial organization was not always confirmed, like in the studies on snowshoe hare (*Lepus americanus*, Burton and Krebs 2003), white-tailed deer (*Odocoileus virginianus*, Comer et al. 2005), and bobcat (*Lynx rufus*, Janečka et al. 2006).

The wild boar, *Sus scrofa*, is an important wildlife species, in both economical and ecological terms. Its widespread recovery across Europe during the last 50 years has raised concerns about the management of this species, which is considered a pest by some and a resource by others. Effective management strategies should take into account several aspects of this species' biology, its social behaviour being one of the most important.

The social organization of wild boar is centered around philopatric adult females, which are facultative cooperative breeders. According to Briedermann (1986), the basic social unit is a matrilineal group, with one or more related sows, and one or more cohorts of offspring. After weaning, most females stay with their mothers, and only about 20% of yearling females leave

the natal group and disperse (Kaminski et al. 2005). Once yearling females have achieved the appropriate growth condition, they are likely to reproduce while still in the social group with their mother (Kaminski et al. 2005). However, genealogical relationships in female groups have been poorly investigated thus far and deviations from this commonly accepted scheme have been rarely documented (see Gabor et al. 1999).

Female wild boars typically maintain long-term fidelity to relatively small home ranges (Spitz and Janeau 1990), and a high percentage of adjacent females exhibit overlapping home ranges (Boitani et al. 1994). Accordingly, one would expect overlapping home ranges to reflect a common female lineage, and genetic relatedness should be inversely correlated with the spatial distance between individuals.

In comparison to other ungulates, wild boar are characterized by several peculiarities such as very high reproductive output (3-6 piglets per litter), early reproduction in females, and a weak mother-offspring bond (Carranza 1996, Cousse et al. 1994, Kaminski et al. 2005). These features obviously affect their social structure, influencing both the size and the composition of social groups, and the duration of inter-individual associations. In addition, the organization of wild boar groups can vary temporally, with fissions into subgroups and fusions of subgroups occurring frequently in a population (Gabor et al. 1999, Kaminski et al. 2005). Importantly, however, the role of demographic and extrinsic factors (e.g. hunting) possibly affecting both the composition and the stability of social groups has not been systematically investigated.

The primary objective of this study was to evaluate the nature of wild boar associations in relation to the genetic relatedness among individuals. We first verified the correlation between geographic and genetic distance among individuals in a population, and then considered the spatial behavior of social units in relation to their composition and the intra-group degree of relatedness. Specifically, we addressed the following questions: Is the geographic distance between individuals inversely correlated to their genetic relatedness? Is genetic relatedness higher for individuals belonging to the same social unit than for individuals belonging to different social units? And, are all individuals in a social group relatives?

MATERIALS AND METHODS

Study area. – The study was carried out in the Alpe di Catenaia, a 12,000-ha mountainous area along the Apennines in Tuscany, Italy (43°48' N, 11°49' E). The area includes a natural reserve (2730 ha) and nearby zones which are open to hunting (Fig. 1).

Elevation ranges from 490 to 1414 m a.s.l. in the Alpe di Catenaia, and the climate is temperate, with hot and dry summers, and cold, rainy winters. Most of the study area (85%) is covered by forests, while the remaining 15% consists of scrubland, cultivated areas, orchard, vineyard, olive grove and human settlements. The only other wild ungulate species in the area was roe deer (*Capreolus capreolus*), and wolves (*Canis lupus*) were the only predators. Also, wild boar represented the staple prey item for wolves in this area (Mattioli et al. 1995, Mattioli et al. 2004), and a wolf pack of 5–6 individuals established its territory and maintained a presence in the area throughout the study period.

The wild boar is a game species which is intensively hunted in Tuscany. Outside the protected area of the natural reserve, wild boars are managed by local hunters. Drive hunts with dogs are conducted from September to January, when 300-900 animals are legally killed in the area each year.

Animal captures, radiotracking and group definition. – This study was carried out from spring 2002 to winter 2005/2006. Wild boars were captured by cage traps baited with maize , except in February-March, when they were captured using a vertical drop net. Cage traps allowed for the simultaneous capture of up to 9 individuals per capture event. Animals were sexed and classified into one of 3 age classes: piglets (from birth to about 12 months; hereinafter referred to as PGL), yearlings (12 to 24 months old; YRL) or adults (>24 months; AD). Upon capture individuals were blindfolded, fitted with ear tags (Allflex, Northfield, Minnesota), weighed, measured, and aged by teeth eruption and wear patterns (Bridermann 1986). Zoletil® (10 ml / 10 kg) was used to immobilize relatively large animals (≥35 kg). Hair samples for genetic analyses were collected and stored in plastic envelopes at –20°C. Sixty-five wild boars were radiocollared. Thirty-one animals (≥ 30 kg) were fitted with TXV-10 radiocollars (Televilt, Lindesberg, Sweden), twenty-four animals (< 30 kg) were fitted with TXP-R ear transmitters (Televilt). Both types of transmitters ranged on 151 MHz. The procedures we used in this work conform to all relevant Italian wildlife and animal welfare legislation.

We calculated the locations of radiocollared animals by triangulation from three different reference points (White and Garrott 1990). We collected a minimum of 8 locations/animal/month (range 8 to 14). Locations were distributed uniformly over the day (discontinuous telemetry, Swihart and Slade 1985), with consecutive positions separated by ≥ 12 hours. We plotted all locations onto a 1:10,000 digital map of the study area. We estimated the accuracy of locations by locating test transmitters which had been placed in different habitats within the study area (Harris et al. 1990). Error for positions was in the range of ± 100 m for fair signals within the study area.

Genetic analysis. – One hundred and twenty captured wild boars were sampled for hair (n = 99) or tissue (n = 21) for genetic analyses. The 21 tissue samples were obtained from marked animals killed by hunters, or animals found dead in the study area. We also collected tissue reference samples from 11 pregnant sows killed during the hunting season and their fetuses (from 4 to 6 per female). Total genomic DNA was extracted using GenElute Mammalian Genomic DNA Miniprep kit (Sigma-Aldrich, St.Louis, Missouri) for tissue samples and Instagene Matrix (Bio-Rad, Hercules, California) for hair samples and kept at –20°C.

All individual animals were typed by a panel of 10 polymorphic microsatellites selected for the analysis: s090, s155, sw24, sw122, sw461, sw2021, sw2492, sw2496, sw2532 and IGF1 (details at http://www.thearkdb.org). Each polymerase chain reaction (PCR) was performed in a 10-μL reaction volume, containing 3 μL of DNA solution, 0.5 U of *Taq* DNA polymerase (Euroclone), 1 x PCR buffer (Euroclone), 2.5 mM $MgCl_2$, 100 μM of each dNTP and 2 pmol of each primer. The forward primer of each pair was labelled with an ABI fluorescent dye (6-FAM, HEX or TET). The amplification profile was set up with an initial step of denaturation at 95 °C for 3 min, followed by 35 cycles of 92 °C for 45 s, T_a (52-65°C) for 45 s, and 72 °C for 30 s. A further extension step of 72 °C for 10 min concluded the reaction. PCR-amplified microsatellite alleles were sized using capillary electrophoresis in an ABI PRISM 3100-Avant automatic sequencer (Applied Biosystems, Foster City, California). The GENEMAPPER software (Applied Biosystems) was used to analyse electrophoretic data.

Data analysis. – Software program RANGES 6 (Kenward et al. 2003) was used to estimate monthly home ranges based on the 95% minimum convex polygon (MCP) method (Southwood 1966). MCP home range was preferred over the Kernel method because of the limited number of fixes available for animals each month (Kernohan et al. 2001). However, in calculating home range centroids from fix locations, the Kernel method was preferred over alternative methods (harmonic or arithmetic mean), because it relies on locations density.

Individual wild boar were partitioned into social units according to capture data, which were subsequently confirmed by observations and telemetry. We assumed that individuals captured together in the same trap or moving together when caught in the nets were part of a social unit. We took into account only those associations which were confirmed by visual observations or telemetry data during the first month following the animals' capture. In order to confirm subsequent groupings, we evaluated the concurrence and the distribution of locations of each pair of individuals over one month. Accordingly, we assumed that 2 individual wild boar were associated in a social unit during a specific month when >50% of

their locations during a 2 hr time period were closer than 500 m. The occurrence and composition of each social unit was checked each month from July 2002 to February 2006. However, only data referred to the period April 2003 – March 2005 were used in our calculations, since this was the time span during which we had a fairly large sample. Fix distributions and home range overlaps were visualized in ARCVIEW GIS 3.2 (ESRI, Redlands, California).

We evaluated the level of genetic variability of the population based on observed heterozygosity (H_o) and expected heterozygosities (H_e), which were estimated by GENALEX 6 (Peakall and Smouse, 2005). GENEPOP 3.2 (Raymond and Rousset 1995) was used to estimate F_{IS} (inbreeding coefficient, ranging between -1 and 1) and to test loci for Hardy-Weinberg (HWE) and linkage equilibrium (LE). The sequential Bonferroni correction was applied to correct significance thresholds in case of multiple tests (Rice 1989).

We used GENALEX to calculate a matrix of pairwise relatedness for all the sampled individuals in the population. As coefficient of relatedness, we chose the unbiased r_{xy} statistics introduced by Queller and Goodnight (1989). Pairwise r_{xy} values range from –1 to +1, with zero indicating the relatedness in a random draw of alleles from the population. Theoretically, in a randomly mating population a relatedness value of 0.5 is expected for parent-offspring and full siblings. Actually, deviations from such expectation are common and this value may vary considerably (Queller and Goodnight 1989). The relatedness matrix was calculated for the data set including all available genotypes (n = 120). In addition, we obtained an empirical data set of fully related individuals (parent-offspring and full-sibs), by genotyping 11 sows killed during the hunting season, together with their foetuses (4-6 per sow, n = 56). The distribution of r_{xy} values of these 'true' family groups (n = 163 comparisons) was used as reference.

Our a priori prediction was that the geographic distance between individuals in the population would be inversely proportional to their genetic distance, based on the idea that closely related animals would either belong to the same social unit or occupy home ranges in relatively closer proximity than more distantly related animals. We tested this hypothesis by estimating the correlation between pairwise relatedness values and spatial distances between the home range centroids in our sample of radiotagged animals (n = 65). In order to account for the lack of independence among pairwise values, we performed a Mantel test for matrix correspondence in GENALEX, testing significance of the correlation coefficient by 9,999 random permutations (Smouse et al. 1986), and applied Bonferroni correction. Because piglets typically move in close association with their mothers during their first year of life

(Bridermann 1986), the inclusion of piglets in this analysis may obscure the effect of relatedness in the post-weaning establishment of individual home ranges. Therefore, the Mantel test was repeated after removing from the matrix all the comparisons which included piglets.

Moreover, we assessed the degree of relatedness between all the members of each social unit identified by field data. Accordingly, we classified each pairwise interaction over a monitored period into either "group", when the two individuals joined the same social unit for at least one month, or "non-group", when the two individuals were never detected in the same social unit. We compared the two corresponding relatedness distributions between each other and to the reference sample, in order to evaluate the deviation from a state of full-relatedness (i.e. from a theoretic r_{xy} value of 0.5). Finally, we explored levels of relatedness within groups including adult females, by comparing PGL-FAD, YRL-FAD, and FAD-FAD associations with those within reference families (sows + foetuses). Likewise, we evaluated the possible composition of groups of subadults by comparing the relatedness in YRL-YRL associations with those obtained for PGL-PGL, and for the reference true siblings (fetuses from the same sow). Two-sample randomization tests (10,000 iterations) were used in POPTOOLS 2.7.5 (Hood 2006) to test for differences between means. Descriptive statistics and graphs were performed using SPSS v. 13.0 (SPSS Inc.).

We finally used the program KINGROUP 2 (Konovalov et al. 2004) to test different hypotheses regarding the relationship between pairs of adult females. The program uses a simulation routine to calculate a ratio between the likelihoods associated to two specific alternative hypothesis (e.g. full-sibs versus unrelated). Using allele frequencies in the real population, KINGROUP generates simulated distributions of r_{xy} for each of the kinship categories corresponding to the null hypothesis (e.g. full-sibs) and the primary hypothesis (e.g. unrelated). From these distributions, it calculates the confidence threshold of the likelihood ratio (i.e. the values needed to reject the null hypothesis). This method allowed to confidently assess kinship between sows in a group.

RESULTS

Sample composition and spatial data. – We captured a total of 120 wild boars, 65 of which were fitted with radiotransmitters and radiotracked between 2002 and 2006. A total of 16 social units were identified in our sample on the basis of capture and spatial associations. Mortality from hunting and poaching was high and caused 86% of the deaths of our study animals. Annual mortality amounted to 47% for adults, 75% for yearlings and 48% for piglets

(mortality in the first two months of life was not considered because very young piglets could not be radiotagged). Consequently, each single wild boar was monitored for an average of 8.6 months.

In the two-year period (April 2003 – March 2005), the average monthly locations of each radiocollared individual were 8.3 (± 2.3 *SD*) and a total of 4,546 radiolocations were obtained.

The composition of our sample in the first year (2003) differed markedly from that in the second year (2004), especially in relation to the proportion of juveniles, which dropped from an average of 50% of the sample in 2003 to only 8% in 2004. Because of this difference we treated the 2 years separately in the statistical analysis. Monthly home ranges differed between years as well, averaging 187.1 ha (209.6 *SD*, standard deviation) in 2003 and 50.7 ha (65.1 *SD*) in 2004. Similarly, the mean overlap between home ranges was twice as high in 2003 compared to 2004 (30.0% vs. 15.6%). Descriptive data of the temporal (monthly) variation in the sample composition and spatial behaviour are shown in Figure 2.

Genetic variation. – A total of 66 different alleles were found at the analysed loci (min 3, max 12 per locus, k = 6.6). Average H_o and H_e were similar, amounting respectively to 0.688 and 0.693. The overall F_{IS} in the population was very close to zero (0.006). The population did not show any significant deviation from HWE, both at single loci and overall (Fisher's method, P = 0.377), while linkage disequilibrium resulted only for three (out of 45) loci combinations (sw2532-sw2496, s090-sw2496, sw122-sw2532). However, each of these markers was mapped in a different chromosome, so that physical linkage could be excluded. Accordingly, in the statistical analyses, we assumed that alleles at different loci were independent.

The coefficient of relatedness in our sample of 120 wild boars averaged -0.010 (± 0.209 *SD*). The reference sample represented by 11 sows and their litters (foetuses) provided a mean relatedness of 0.599 (± 0.130 *SD*), slightly higher than the value of 0.5 which is theoretically expected for comparisons between first-degree relatives (parent-offspring and full-sibs).

Relatedness, spatial patterns and social units. – The correlation between spatial distance and genetic relatedness, as resulting from the Mantel test, fluctuated during the study period (Table 1), proving significantly negative only in summer and autumn (July-October) 2003. This could be related to the piglets' presence, given that no correlation was observed in 2004 when the sample composition was biased towards subadults and adults (Fig. 2). Moreover, the repetition

of this analysis without piglets resulted in a complete lack of significance during the 2-year study period.

The hypothesis that individuals belonging to the same social unit were more related than non-associated individuals was confirmed. The randomization test showed a significantly higher relatedness among individuals of the same social unit ($P<0.001$), although the width of the range suggests that unrelated individuals can group together (Fig. 3). In fact, intra-group comparisons differed from the reference sample ($P<0.001$), thus suggesting a deviation from the full-relatedness hypothesis.

As regards the type of association within putative matrilineal social units, YRL-FAD associations showed the highest levels of relatedness, followed by PGL-FAD and FAD-FAD (Fig. 4). The relatedness between adult females in a group was significantly lower than the relatedness observed in YRL-FAD ($P=0.017$) and PGL-FAD ($P=0.026$) associations, but each of them significantly differs from the distribution observed in the reference families (all $P<0.001$). Similarly, pairs of yearlings (YRL-YRL) showed a low average level of relatedness when compared to PGL-PGL associations and to control sibling pairs (Fig. 5), thus deviating from the expectation of sibship. The average relatedness of FAD-FAD dyads was equal to 0.082 ± 0.155 (mean $\pm$ *SD*). The likelihood analysis with KINGROUP allowed to confidently exclude that 5 out of 9 pairs were represented by close relatives (r_{xy} ranging between -0.212 and 0.140). Indeed, for all of them, the null hypotheses of full-sibs, half-sibs and parent-offspring could be rejected at a 95% confidence. Three pairs were found in association for less than 3 months (usually because of the death or signal loss of one female in the pair), while the other two dyads had resulted together for as long as 6 consecutive months.

DISCUSSION

Here we combined genetic and radiotelemetry data with the aim to test different hypotheses in relation to the wild boar social structure.

Even in species with low behavioural plasticity, the social organization tends to vary under different ecological conditions (Lott 1991). Based on what was previously known for the behaviors of wild boar, we would expect social units in a population to be composed of relatives, and social units would have a higher chance of being surrounded by related than unrelated individuals. Because wild boar are social and only adult males are solitary, the overall spatial segregation among individuals can be predicted to correlate with genetic relatedness.

We observed a negative correlation between geographic distance (i.e. distance between home-range centroids) and genetic relatedness only in summer and early autumn, i.e. during the period between parturition and weaning of juvenile wild boar. Based on previous studies

(Kaminski et al. 2005), this is also the period when the social affinity in groups is stronger, a common pattern for suids in general (Byers 1983, Somers et al. 1995). A remarkable difference was observed between years. In 2003 relatedness correlated significantly with distance in the period July – October only, whereas no correlation was observed during the following year. The most obvious explanation for the observed correlation in 2003 is that this pattern resulted from the higher number of piglets in social units at that time of the year, when they were strongly associated with their mothers. Such a difference could thus be related to the very small number of piglets in our samples during 2004 (Fig. 2), due to a low capture success. Yet, besides the role of piglets, we cannot exclude that the observed difference was caused by other ecological factors (e.g. food availability, climate, etc.).

The overall weak and temporarily limited correlation between genetic and spatial distance does not fit a model of social structure where relatives tend to stay close even though they occasionally belong to different groups. A similar pattern was found in white-tailed deer (Comer et al. 2005), where the observed weak correlation between genetic relatedness and spatial association in does contradicted the 'rose-petal' hypothesis of social organization in this species (Porter et al. 1991). Comer at al. (2005) considered this apparent contradiction as a possible effect of heavy harvesting, as suggested by the altered age structure in the female population, that could have limited the occurrence of persistent and cohesive social groups. Hunting could also help explain the pattern observed in our study population, as suggested by the relatedness analyses within social units.

Individuals in a group resulted to be more related than individuals that were never found in association. This result agrees with the expectation of matrilineal social units, although the divergence from the reference families suggests that low-related or unrelated individuals could also be found in association. When we evaluated the intra-group relatedness with respect to the age class of individuals, we obtained unexpected results. In particular, in 5 out of 9 cases, adult sows joining the same social unit were neither sisters, half-sisters nor mother-daughter. Furthermore, for each age class combination that we took into consideration, the range of relatedness values suggested the simultaneous presence of unrelated and fully related individuals. Thus, contrary to common expectation, associations of both unrelated adult sows and unrelated yearlings were frequent in our study population. The higher level of relatedness shown by the YRL-FAD with respect to FAD-FAD dyads (Fig. 4) suggested that the individuals remaining in association with adult females after their first year of age usually were the offspring of adult females in the group. The high number of piglets in a group implied that multiple litters of different sows were often associated. This could easily explain the low relatedness of PGL-FAD dyads when compared to the control mother-offspring groups.

The occurrence of non-kin associations within groups was reported for feral pigs in Texas (Emlen 1997, Gabor et al. 1999) but has never been demonstrated for European free-living wild boars. Temporal associations of unrelated individuals can be accounted for by the possible benefits deriving from group living, beyond the fitness consequences of a kin-based cooperative behaviour (Griffin and West 2002). For instance, group living represents an effective anti-predator strategy (Hamilton 1971). Wolves are the most important natural predators of wild boar in Europe, where they basically select for young boars (Jędrzejewski et al. 2000, Mattioli et al. 1995). Adult wild boar are less vulnerable to wolf predation and benefit from group living less than young individuals (Mattioli et al. 1995). This could be a key factor explaining the solitary life of adult males. On the contrary, adult females also have to warrant protection for their litter. Therefore they might be urged by the presence of predators to join a group of other females, regardless of kinship. This could well be the case in our study area, where a wolf pack was stably present during data collection and it mainly relied on wild boar as a prey species (Apollonio and Mattioli 2006).

In ungulates, grouping as a response to predation risk is quite common, especially when predators use cooperative hunting techniques (Creel and Winnie 2005, FitzGibbon 1990, Lingle 2001). When prey groups mob predators as an anti-predator strategy, individuals have a higher chance to avoid predation when they form larger groups. This is for instance the behaviour adopted by mule deer against coyote attacks (Lingle 2001). Wild boar groups include adult sows, which, in presence of piglets, can react aggressively against predators (Heck and Raschke 1980).

Also human activities are likely to influence the social structure of wild boars. Hunting was shown to affect both the social and spatial behaviour of wild boar, increasing social affinity (Kaminski et al. 2005) and inducing a variation of home range size, as well as temporary departures from traditional resting sites (Baubet et al. 1998, Maillard and Fournier 1995, Sodeikat and Pohlmeyer 2002). These effects are more evident when drive hunting is practiced, as occurs in the surroundings of the Alpe di Catenaia protected area, as revealed by a parallel study on roe deer spatial patterns in the same area (Bongi et al. 2007).

The study wild boar population was indeed characterized by an overall high mortality rate, mostly due to hunting and poaching (86% of deaths in our sample). Such losses modified the size and the composition of social units and, as a consequence, non-kin associations might have been formed so as to replace dead individuals and maintain the advantages of group living. This explanation would entail a high turn-over and a dynamic composition of social units, where the loss of one or more relatives is compensated by the acceptance of unrelated

individuals to the group. This process may be enhanced by the high proportion of sows killed during the hunting season in the study area (in a bag of 2648 kills, 31% were represented by subadult or adult sows, unpubl. data). In fact, unlike in other European countries (e.g. Germany), the Italian legislation allows hunters to kill adult sows.

In addition to the above mentioned explanations, an average similarity which is lower than expected within a social unit can also be a consequence of multiple paternity. Indeed, when piglets in a litter have different fathers (i.e. they are half-sibs), the overall intra-litter relatedness will be lower than when they are full-sibs (i.e. a single father). The effect can be more pronounced when it involves different cohorts within the same social unit. However, observations of sows breeding with several males are poorly documented (Barrett 1978). In a recent study, Delgado et al. (2008) confirmed that this phenomenon may occur at low frequencies, detecting limited signs of multiple paternity in a wild boar population in Portugal. A preliminary study suggests that this is the case also in our study population (Iacolina et al. in prep.).

Our results suggest that interactions between kin do not play an exclusive role in wild boar sociality. The matrilineal structure of social units in this species may thus exhibit exceptions under certain conditions. In our study area, the weak correlation between genetic relatedness and spatial distance, and the occurrence of unrelated adult females within a group suggest a frequent deviation from matrilinearity. We believe that the rearrangements of wild boar social groups were likely due to the combination of high human-caused mortality, and constant exposure to predation risk.

Further studies are warranted to investigate the temporal and spatial dynamics of non-kin associations, their occurrence under different conditions (e.g. hunting vs. non-hunting areas) and the role of predators as a driving force in promoting the group formation in wild boar. Finally, the presence of unrelated individual wild boar in a social unit opens a series of questions regarding the possible fitness benefits associated to cooperative breeding in this species.

ACKNOWLEDGMENTS

We want to thank E. Donaggio, N. Cappai, A. Viviani, R. Tognini, M. Derudas, V. Mariano, P. Lamberti for their effort in field activities, and many local hunters, especially the URCA association, students and volunteers for their assistance in capturing and monitoring wild boars. We are also thankful to S. Ciuti and S. Luccarini for advices and suggestions during the study, and to P. Banti, G. Chianucci, P. Pedone and L. Mattioli for their logistical support. We are finally grateful to A. Binelli, who kindly revised linguistic style, and to R.A. Sweitzer and to an anonymous reviewer for their contribution in improving a preliminary draft of the manuscript. The study was financed by Regione Toscana, Provincia d'Arezzo, and by the Ministero dell'Università e della Ricerca (MIUR PRIN 2005). L. Iacolina benefited from a fellowship sponsored by Sardegna Resorts s.r.l.

TABLE 1. – Temporal variation (per month, April 2003 – March 2005) in the correlation between geographic distance and relatedness (Mantel test) in wild boars in the Alpe di Catenaia. Significant correlations, evaluated over 9,999 random permutations, are in bold ($P < 0.00213$; Bonferroni correction). a) Correlations computed including piglets, b) correlations computed excluding piglets.

a)

		WINTER			SPRING			SUMMER			AUTUMN			
		J	F	M	A	M	J	J	A	S	O	N	D	
Mantel test	*R*				-0.070	0.009	-0.155	-0.235	-0.219	-0.260	-0.329	-0.167	-0.114	**2003**
	p				0.252	0.513	0.058	**0.001**	**0.001**	**0.000**	**0.000**	0.014	0.110	
	R	-0.075	-0.052	-0.011	-0.050	-0.011	-0.166	-0.134	-0.076	-0.043	-0.005	-0.177	0.019	**2004**
	p	0.208	0.261	0.436	0.222	0.420	0.009	0.028	0.141	0.274	0.409	0.080	0.518	
	R	-0.026	-0.154	0.403										**2005**
	p	0.376	0.299	0.128										

b)

		WINTER			SPRING			SUMMER			AUTUMN			
		J	F	M	A	M	J	J	A	S	O	N	D	
Mantel test	*R*				0.241	0.306	0.367	0.395	0.292	0.352	0.508	-0.010	-0.113	**2003**
	p				0.068	0.085	0.037	0.023	0.075	0.047	0.014	0.522	0.436	
	R	-0.150	-0.016	-0.031	0.025	-0.004	0.172	0.017	-0.009	-0.005	-0.117	-0.093	-0.180	**2004**
	p	0.436	0.536	0.437	0.360	0.543	0.039	0.355	0.532	0.548	0.190	0.284	0.190	
	R	0.145	0.225	0.225										**2005**
	p	0.249	0.221	0.215										

FIG. 1. – Study area in the Alpe di Catenaia, Arezzo, Italy. Borders of the natural reserve (hatched line) and location of cage traps used to capture wild boars (asterisks) are shown.

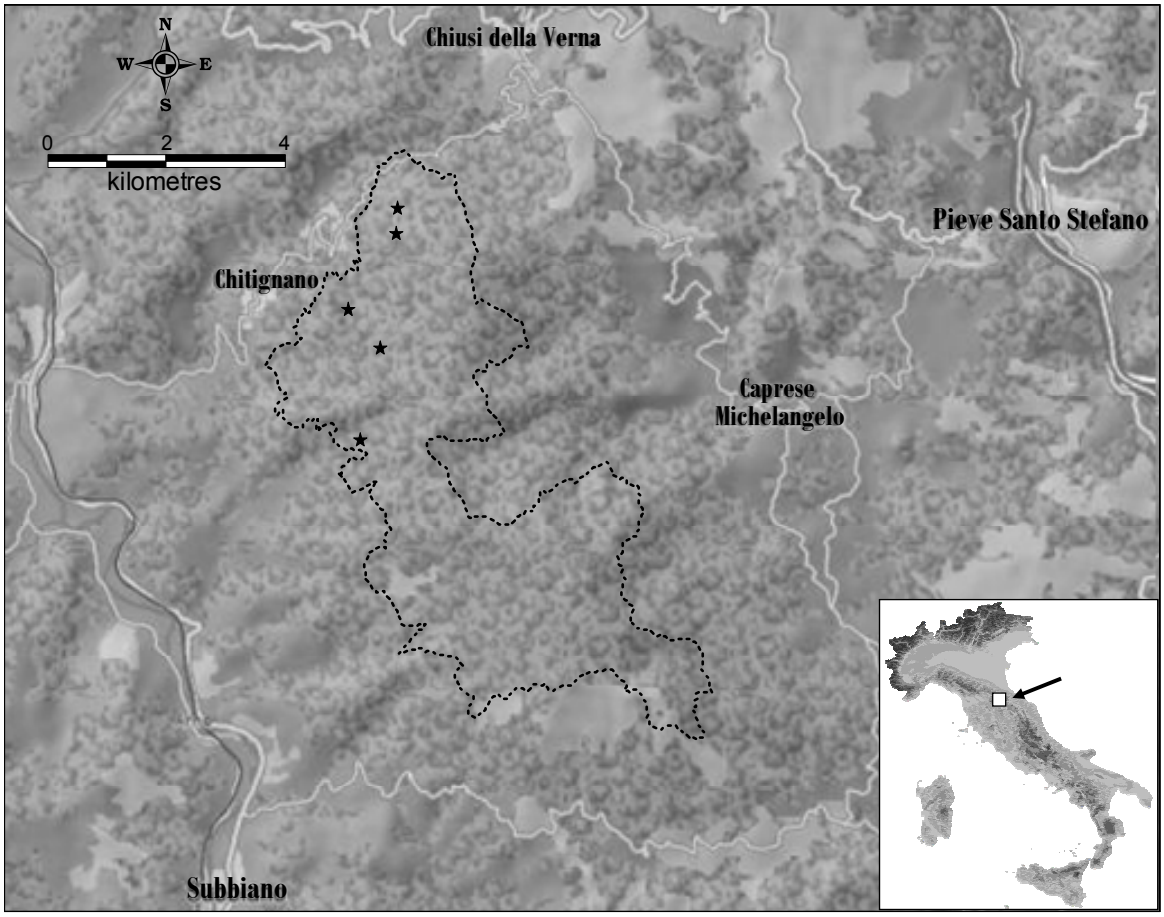


FIG. 2. – Sample variation and its spatial behaviour during the study period (April 2003 – March 2005) in the Alpe di Catenaia wild boar population: a) sample size; b) sample composition (AD - adults, YRL - yearlings, PGL - piglets); c) home range size; d) home range overlap. Home range size is shown as the mean (± *SD*) of all individual home ranges calculated by the MCP method using 95% of fix locations. Home range overlap refers to the average percentage of all pairwise overlaps between monthly individual home ranges.

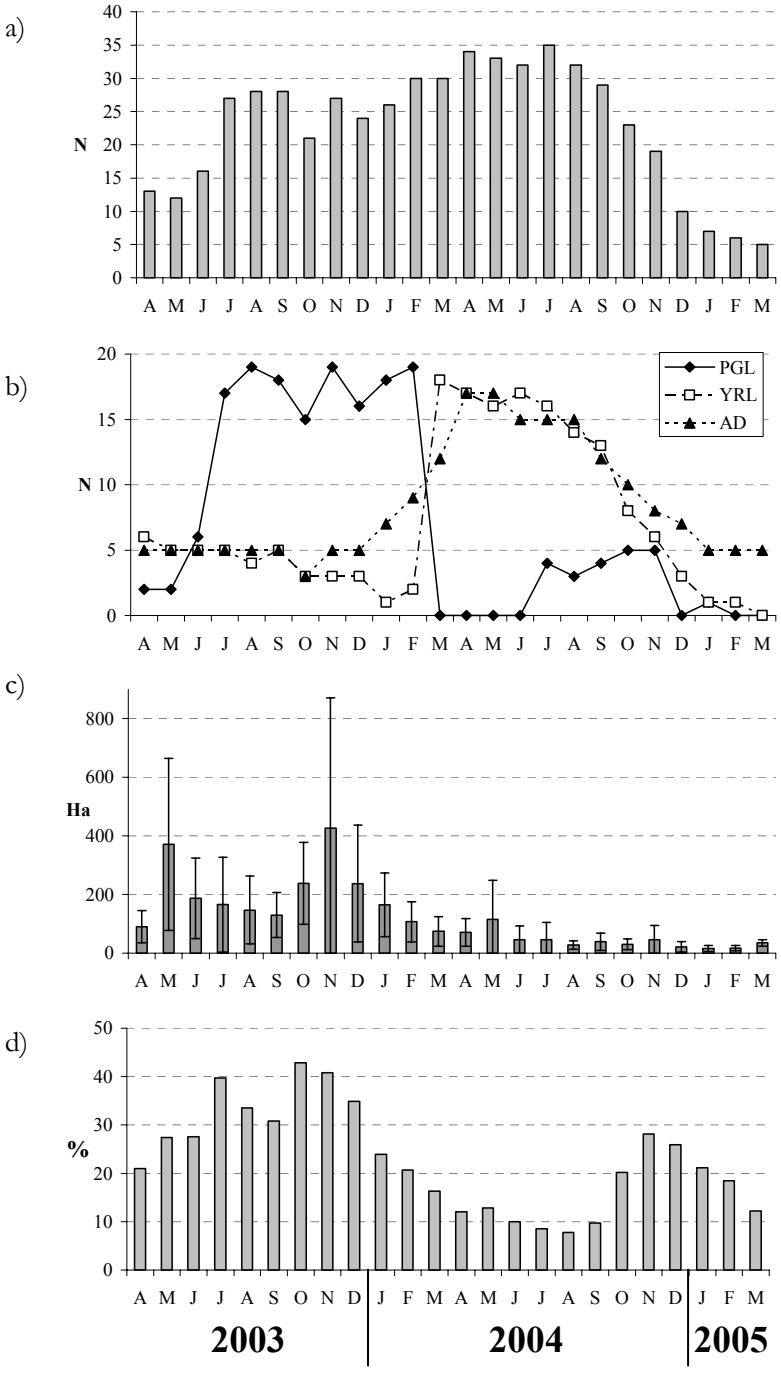
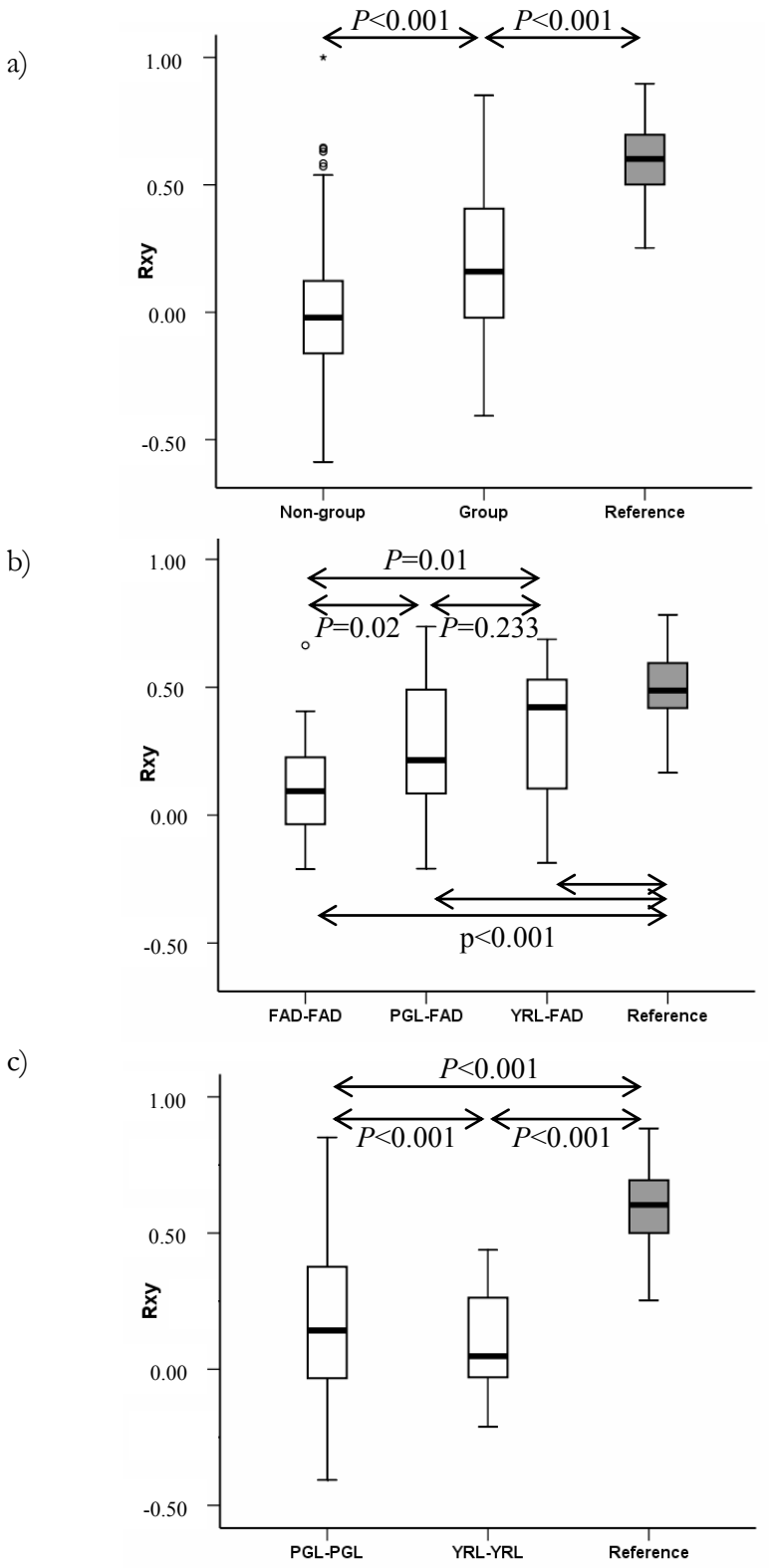


FIG. 3. – Relatedness estimated distributions a) between wild boars joining the same social unit ('Group'; *n* = 215 comparisons) and moving separately in the study area ('Non-group'; *n* = 1598); b) within wild boar social units for different age class associations. PGL – piglet, YRL – yearling, FAD – adult female. As refence, relatedness observed in 11 sow-foetuses families (*n* = 163) is reported. c) intra-class within wild boar social units (PGL – piglet, YRL – yearling). As reference, relatedness observed between littermates (foetuses, *n* = 118 comparisosn) is reported.P-values refer to two-sample randomization tests for differences between means (*n* = 10,000 iterations).



LITERATURE CITED

APOLLONIO, M., AND G. B. HARTL 1993. Are biochemical-genetic variation and mating system related in large mammals? Acta Theriologica Suppl.2:175-185.

APOLLONIO, M., AND L. MATTIOLI. 2006. The wolf in the Arezzo province, Italy. Editrice Le Balze, Montepulciano, Italy.

BARRETT, R. H. 1978. The feral hog at Dye Creek Ranch, California. Hilgardia 46:283-355.

BAUBET, E., S. BRANDT, AND C. TOUZEAU. 1998. Effet de la chasse sur les stratégies d'occupation de l'espace des sangliers (*Sus scrofa*). Analyses préliminaires. Gibier Faune Sauvage 15:655-658.

BLUNDELL, G.M., M. BEN-DAVID, P. GROVES, R. T. BOWYER, AND E. GEFFEN. 2004 Kinship and sociality in coastal river otters: are they related? Behavioural Ecology 15(5):705-714.

BOITANI, L., L. MATTEI, D. NONIS, AND F. CORSI. 1994. Spatial and activity patterns of wild boars in Tuscany, Italy. Journal of Mammalogy 75:600-612.

BONGI, P., S. GRIGNOLIO, S. CIUTI, AND M. APOLLONIO. 2007. Influence of hunting with dogs on roe deer spatial behaviour: differential responses according to sex and age classes. Hystrix, Italian Journal of Mammalogy (N.S.) 1, Suppl. 2007:238.

BRIEDERMANN, L. 1986. Schwartzwild. Neumann-Neudamm ed. Berlin, Germany.

BURTON, C., AND C. J. KREBS. 2003. Influence of relatedness on snowshoe hare spacing behavior. Journal of Mammalogy 84:1100-1111.

BYERS, J. 1983. Social interactions of juvenile collared peccaries, *Tayassu tajacu* (Mammalia: Artiodactyla). Journal of Zoology, London 201:83-96.

CARRANZA, J. 1996. Sexual selection for male body mass and the evolution of litter size in mammals. American Naturalist 148:81-100.

CHEPKO-SADE, B. D., AND Z. TANG HALPIN (EDS.). 1987. Mammalian dispersal patterns: the effects of social structure on population genetics. The University of Chicago Press, Chicago, Illinois, USA.

COMER, C. E., J. C. KILGO, G. J. D'ANGELO, T.C. GLENN, AND K. V. MILLER. 2005. Fine-scale genetic structure and social organization in female white-tailed deer. Journal of Wildlife Management 69:332-344.

COUSSE, S., F. SPITZ, M. HEWISON, AND G. JANEAU. 1994. Use of space by juveniles in relation to their postnatal range, mother, and siblings: an example in the wild boar, *Sus scrofa* L. Canadian Journal of Zoology 72:1691-1694.

CREEL, S., AND J. A. WINNIE. 2005. Responses of elk herd size to fine-scale spatial and temporal variation in the risk of predation by wolves. Animal Behaviour 69:1181-1189.

DELGADO, R., P. FERNÁNDEZ-LLARIO, M. AZEVEDO, A. BEJA-PEREIRA, AND P. SANTOS. In press. Paternity assessment in free-ranging wild boar (*Sus scrofa*) – Are littermates full-sibs? Mammalian Biology.

DOBSON, F.S. 1998. Social structure and gene dynamics in mammals. Journal of Mammalogy 79:667-670.

EISENBERG, J. F., AND D. G. KLEIMAN. 1983. Advances in the study of mammalian behavior. Special Publication, The American Society of Mammalogists, 7:1-753.

EMLEN, S.T. 1997. Predicting family dynamics in social vertebrates. In Behavioural Ecology (Krebs, J.R. and Davies, N.B., eds), pp. 228–253, Blackwell Scientific.

FITZGIBBON, C. D. 1990. Mixed-species grouping in Thomson's and Grant's gazelles: the antipredator benefits. Animal Behaviour 39:1116-1126.

GABOR, T. M., E. C. HELLGREN, R. A. VAN DEN BUSSCHE, AND N. J. SILVY. 1999. Demography, sociospatial behaviour and genetics of feral pigs (*Sus scrofa*) in a semi-arid environment. Journal of Zoology 247:311-322.

GRIFFIN, A. S., AND S. A. WEST. 2002. Kin selection: fact and fiction. Trends in Ecology and Evolution 17:15-21.

HAMILTON, W. D. 1971. Geometry of the selfish herd. Journal of Theoretical Biology 31:295-311.

HARRIS, S., W. J. CRESSWELL, P. G. FORDE, W. J. TREWHELLA, T. WOOLLARD, AND S. WRAY. 1990. Home-range analysis using radio-tracking data - a review of problems and techniques particularly as applied to the study of mammals. Mammal Review 20:97-123.

HECK, L., AND G. RASCHKE. 1980. Die Wildsauen: Naturgeschichte, Ökologie, Hege und Jagd. Verlag Paul Parey, Hamburg, Germany.

HOOD, G. M. 2006. PopTools version 2.7.5. Available on the internet. URL http://www.cse.csiro.au/poptools

JANEČKA, J. E., T. L. BLANKENSHIP, D. H. HIRTH, M. E. TEWES, C. W. KILPATRICK, AND L. I. GRASSMAN JR. 2006. Kinship and social structure of bobcats (*Lynx rufus*) inferred from microsatellite and radio-telemetry data. Journal of Zoology 269(4):494-501.

JĘDRZEJEWSKI, W., B. JĘDRZEJEWSKA, H. OKARMA, K. SCHMIDT, K. ZUB, AND M. MUSIANI. 2000. Prey selection and predation by wolves in Białowieża primeval forest, Poland. Journal of Mammalogy 81(1): 197-212.

KAMINSKI, G., S. BRANDT, E. BAUBET, AND C. BAUDOIN. 2005. Life-history patterns in female wild boars (*Sus scrofa*): mother-daughter postweaning associations. Canadian Journal of Zoology 83:474-480.

KENWARD, S. E., A. B. SOUTH, AND S. S. WALLS. 2003. Ranges6 v. 1.2: for the analysis of tracking and location data. Anatrack Ltd. Wareham, UK.

KERNOHAN, B. J., R. A. GITZEN, AND J. J. MILLSPAUGH. 2001. Analysis of animal space use and movements. Pp. 126-168 in Radio tracking and animal populations (J. J. Millspaugh and J. M. Marzluff, eds.). Academic Press, California, USA.

KONOVALOV, D.A., C. MANNING, AND M.T., HENSHAW. 2004. KINGROUP: a program for kinship reconstruction and kin group assignments using genetic markers. Molecular Ecology Notes 4:779-782.

LINGLE, S. 2001. Anti-predator strategies and grouping patterns in white-tailed deer and mule deer.

Ethology 107:295-314.

LOTT, D. F. 1991. Intraspecific variation in the social systems of vertebrates. Cambridge University Press, Cambridge.

MAILLARD, D. AND P. FOURNIER. 1995. Effects of shooting with hounds on size of resting range of wild boar /*Sus scrofa* L.) groups in mediterranean habitat, Ibex, Journal of Mountain Ecology 3:102-107.

MATTIOLI, L., M. APOLLONIO, V. MAZZARONE AND E. CENTOFANTI. 1995. Wolf food habits and wild ungulate availability in the Foreste Casentinesi National Park; Italy. Acta Theriologica 40:387–402.

MATTIOLI, L., C. CAPITANI, E. AVANZINELLI, I. BERTELLI, A. GAZZOLA, M. APOLLONIO. 2004. Predation by wolves (*Canis lupus*) on roe deer (*Capreolus capreolus*) in north-eastern Apennine, Italy. Journal of Zoology, London, 2004. 264: p. 249-258.

MOYER, M. A., J. W. MCCOWN, T. H. EASON, AND M. K. OLI. 2006. Does genetic relatedness influence space use pattern? A test on Florida black bears. Journal of Mammalogy 87:255-261.

PEAKALL, R., AND P. E. SMOUSE. 2005. GenAlEx 6: Genetic Analysis in Excel. Population genetic software for teaching and research. Australian National University, Canberra, Australia.

POPE, T. R. 1998. Effects of demographic change on group kin structure and gene dynamics of red howling monkey populations. Journal of Mammalogy 79:692-712.

PORTER, W. F., N. E. MATHEWS, H. B. UNDERWOOD, R. W. SAGE, AND D. F. BEHREND. 1991. Social organization in deer: implications for localized management. Environmental Management 15:809-814.

QUELLER, D. C., AND K. F. GOODNIGHT. 1989. Estimating relatedness using genetic markers. Evolution 43:258-275.

RATNAYEKE, S., G. A. TUSKAN, AND M. R. PELTON. 2002. Genetic relatedness and female spatial organization in a solitary carnivore, the raccoon, *Procyon lotor*. Molecular Ecology 11:1115-1124.

RAYMOND, M., AND F. ROUSSET. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. Journal of Heredity 86:248-249.

RICE, WR. 1989. Analyzing tables of statistical tests. Evolution 43:223-225.

SMOUSE, P. E., J. C. LONG, AND R. R. SOKAL. 1986. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. Systematics Zoology 35:627-632.

SODEIKAT, G., AND K. POHLMAYER. 2002. Temporary home range modifications of wild boar family groups (*Sus scrofa* L.) caused by drive hunts in Lower Saxony (Germany). Zeitschrift für Jagdwissenschaft. 48:161-166.

SOMERS, M. J., A. E. RASA, AND B. L. PENZHORN. 1995. Group structure and social behaviour of warthogs, *Phacochoerus aethiopicus*. Acta Theriologica 40:257-281.

SOUTHWOOD , T. R. E. 1966. Ecological Methods. Methuen Ed. London, UK.

SPONG, G., S. CREEL. 2004. Effects of kinship on territorial conflicts among groups of lions, *Panthera leo*. Behavioural Ecology and Sociobiology 55:325-331.

STORZ, J. F. 1999. Genetic consequences of mammalian social structure. Journal of Mammalogy 80:553-569.

SUGG, D. W., R. K. CHESSER, F. S. DOBSON, AND J. L. HOOGLAN. 1996. Population genetics meets behavioral ecology. Trends in Ecology and Evolution 11:338-342.

SWIHART, R. K., AND N. A. SLADE. 1985. Testing for independence of observation in animal movements. Ecology 66(4):1176-1184.

WIMMER, B., D. TAUZT, AND P. M. KAPPELER. 2002. The genetic population structure of the gray mouse lemur (*Microcebus murinus*), a basal primate from Madagascar. Behavioral Ecology and Sociobiology 52:166-175.

CHAPTER III

High frequency of multiple fathering in Italian wild boar litters.

High frequency of multiple fathering in Italian wild boar litters.

Laura IACOLINA, Massimo SCANDURA, Nadia CAPPAI and Marco APOLLONIO

Dipartimento di Zoologia e Genetica Evoluzionistica, Università di Sassari, Via Muroni 25, I-07100 Sassari, Italy.

Manuscript

ABSTRACT

We documented the occurence of multipe paternity (MP) in a free ranging wild boar (*Sus scrofa*) population under heavy hunting pressure in Tuscany, Italy. Twelve families, constituted by a pregnant sow killed during the hunting season and her fetuses, were genetically analysed using ten polymorphic microsatellites. MP was inferred for six (50%) of the twelve analysed litters, both by studying Mendelian inheritance of alleles and by using a likelihood-based algorithm in the software COLONY. Multiple fathering involved litters of both adult and subadult sows. Our results confirm a previous evidence of MP in European wild boars, but they reveal that polyandry can become common in some populations. These new insights suggest that the emblematic ecological plasticity of the wild boar has a correspondence on its reproductive biology. We discussed mechanisms that can promote multiple mating in wild boar, with special reference to male-male competition and female (precopulatory and postcopulatory) mate choice. Our arguments prompt to study male spatial behavior during the mating season, to disclose mechanisms regulating male access to female groups during the estrus, together with a consideration of the possible role of strong hunting pressure.

Keywords: mating systems**,** microsatellites, multiple paternity, *Sus scrofa*, wild boar.

INTRODUCTION

Different breeding strategies are found in mammals, going from monogamy to extreme polygamy, and most of them involve little paternal care of offspring (Clutton-Brock 1991). Because females are the sex investing most in reproduction, they are more choosy and may decide to mate with more than one male, if this behaviour entails fitness benefits (Trivers 1972). On the other side, males are expected to maximize their matings but also to adopt pre- and post-copulatory strategies to prevent their partner from mating with other males, in order to keep their own reproductive success high. However, Isvaran and Clutton-Brock (2007) observed that in mammals, independently from the adopted breeding system, extra-group paternity is quite common. Even species that exhibit a monogamous behaviour can reveal the occurrence of extra-pair mating and litters can be sired by multiple males (Carmichael et al. 2007; Clutton-Brock and Isvaran 2006; Crawford et al. 2008; Nielsen and Nielsen 2007; Sorin 2004). Recently, several studies disclosed the occurrence of multiple paternity (MP) in mammalian species with different social behaviour, such as beaver (*Castor canadensis*, Crawford et al 2008), racoon (*Procyon lotor*, Nielsen and Nielsen 2007), arctic fox (*Alopex lagopus*, Carmichael et al 2007), white-tailed deer (*Odocoileus virginianus*, DeYoung et al. 2002, Sorin 2004) and pronghorn antelope (*Antilocapra americana*, Carling et al. 2003). This was interpreted by the authors in terms of evolutionary benefits, as species with high levels of polyandry were found to have significantly lower rates of early reproductive failure and a larger litter size (Stockley 2003).

The wild boar (*Sus scrofa*), when compared to other ungulates, is characterized by high reproductive performances. Sows can give birth to up to 10 piglets and can reproduce at the age of 8-9 months (Briedermann 1986). Matriarchal social groups are constituted by one or more adult females followed by one or more cohorts of their offspring (Kaminski et al. 2005). It was reported that sows in a group are facultative cooperative breeders, that can synchronise their estrus and mate with solitary adult males, joining the social groups during the breeding season (Dardaillon 1988). Estrus in a female last for 1.8 days on average (Henry 1968) and matings in a group are usually completed within 8 days (Briedermann 1986). Nevertheless, a temporal shift of the estrus may be observed in different social groups of the same area (Briedermann 1986). In October-November adult males begin to search actively for receptive females, and, after reaching a group of estrus sows, they start a courtship behaviour described in detail by Beuerle (1975). At this occasion more males can be interested in the same group of females, giving rise to competitive interactions. Depending on the level of dominance that is established during these interactions, one single or more than one male can mate with the

receptive females (Briedermann 1986). In the wild boar, each female can mate up to 4-5 times during the 2-days estrus period (Briedermann 1986), less than observed in the domestic pig (Xue et al. 1998). Therefore, even though it is difficult to document it by direct observation, one cannot exclude that different males can mate with the same female during the breeding season. Therefore, this species is a suitable candidate for MP.

In agreement with this expectation, Aguilera-Reyes et al. (2006) showed MP to occur in the domestic form, proving that swine litters can be sired by up to 3 males. They investigated the reproductive success of three male pigs (two different pure-bred and one hybrid) after their successive mating with 18 sows. They found that the hybrid was the most successful male (i.e. sired most of the offspring), as it was preferentially chosen by females. Aguilera-Reyes et al. (2006) speculated that females favoured the male with the better genetic quality and that its high reproductive success could be explained in terms of female cryptic choice. Interestingly, they remarked the possible role of female-female contacts (e.g. pseudo-mating) in determining the final fertilization success. Similar behavioural strategies and cryptic selection are likely to occur in the wild form too, as proved by Delgado et al (2008), who recently found MP to occur in a wild boar population in Portugal. In their study population, MP does not seem to be common as only one litter out of nine showed signs of MP, suggesting that mate-guarding by the dominant male could result effective in preventing MP.

In a hunted wild boar population, in central Italy, we recently found an unexpected low degree of relatedness within social units (Iacolina et al. subm.). As main possible explanation we pointed to the high mortality rate, mostly due to hunting (Iacolina et al. subm.), that does not only affect the demographic structure of the population, but even the social system. Besides the presence of unrelated sows in some social units, we also found low levels of relatedness between piglet and subadults, that had been captured together or found in association during the study. We argued that the occurrence of MP in the population could be explored as a possible factor influencing kinship within groups. If confirmed, this interpretation would lead to new insights on the wild boar reproductive biology.

Our goal in the present study was to prove the occurrence of MP in the wild boar study population of Alpe di Catenaia, Tuscany, Italy. We then discussed the meaning of this behaviour with respect to the demographic features of the population and to the ecological constraints of the area.

MATERIALS AND METHODS

Study area. – Samples were collected in the Alpe di Catenaia, a 12,000-ha mountainous area along the Apennines in Tuscany, Italy (43°48' N, 11°49' E). Only the central part of the area is protected (2730 ha), while the nearby zones are open to hunting. Elevation ranges from 490 to 1414 m a.s.l.. Climate is temperate, with hot and dry summers, and cold rainy winters. Most of the study area (85%) is covered by woods, while the remaining 15% consists of scrubland, cultivated areas, orchard, vineyard, olive grove and human settlements. The only other wild ungulate species in the area is roe deer (*Capreolus capreolus*). The wolf (*Canis lupus*) is the main natural predator of wild boars, that here represent the staple of its diet (Mattioli et al. 1995, Mattioli et al. 2004).

The wild boar is a game species which is intensively hunted in Tuscany. Outside the protected area wild boars are managed by local hunters, who carry out drive hunts with dogs from September to January. Approximately 600-1500 animals are legally killed every year in the area.

Sample collection. – Samples had been collected in the period 2002-2007. Pregnant sows were examined by field necropsy. Each sow was weighted (dressed weight) and her age assessed according to teeth eruption and wear patterns (Bridermann 1986). Uteri were removed and stored until veterinary exam. Then fetuses were carefully extracted from each uterus, washed and a sample of each of them, such as a sample of the uterus itself, was stored in 96% ethanol until DNA extraction.

For the analysis, we randomly selected twelve pregnant sows, that were analysed with their litters (from 4 to 6 fetuses per female, $n = 61$, Table 1). Four females were adults (> 24 months) and eight were yearlings (12-24 months). The mean body weight was 71.7 ± 10.5 Kg for adult females and 55.5 ± 8.0 Kg for yearlings (overall mean 60.9 ± 11.6). The number of fetuses per litter was on average 5.1 ± 0.8 (4.7 ± 1.0 for adult and 5.2 ± 0.7 for yearling females).

Genetic analysis. – Total genomic DNA was extracted from tissue samples by using Sigma commercial kit GenElute Mammalian Genomic DNA miniprep (Sigma-Aldrich, St.Louis, Missouri).

All the individuals were typed by a panel of 10 polymorphic microsatellites: s090, s155, sw24, sw122, sw461, sw2021, sw2492, sw2496, sw2532 and IGF1 (details at http://www.thearkdb.org). Each locus mapped in a different chromosome, so that physical linkage could be excluded. Accordingly, in the statistical analyses, we assumed the

independence of alleles among loci. Each polymerase chain reaction (PCR) was performed in a 10-μL reaction volume, containing 3 μL of DNA solution, 0.5 U of *Taq* DNA polymerase (Euroclone, Siziano, Italy), 1 x PCR buffer (Euroclone), 2.5 mM $MgCl_2$, 100 μM of each dNTP and 2 pmol of each primer. The forward primer of each pair was labelled with an ABI fluorescent dye (6-FAM, HEX or TET). The amplification profile was set up with an initial step of denaturation at 95 °C for 3 min, followed by 35 cycles of 92 °C for 45 s, T_a (52-65°C) for 45 s, and 72 °C for 30 s. A further extension step of 72 °C for 10 min concluded the reaction. PCR-amplified microsatellite alleles were sized using capillary electrophoresis in an ABI PRISM 3100-Avant automatic sequencer (Applied Biosystems, Foster City, California). The GENEMAPPER software (Applied Biosystems) was used to analyse electrophoretic data, that were systematically double checked.

All fetuses were sexed. When visual inspection was insufficient to establish their gender with certainty, molecular sex identification was performed by co-amplifying a SRY gene fragment (used primers: 5'-CATTGTGTGGTCTCGTGATC-3' and 5'-AGTCTCTGTGCCTCCTCGAA-3', Richard et al. 1994) and a portion of the mitochondrial cytochrome B gene (L14841 and H15149, Kocher et al. 1989) as positive PCR control.

Data analysis. – In order to evaluate the level of genetic variability of the population, observed (H_o) and expected heterozigosity (H_e) were estimated by MSTOOLS 3 (Park 2001) analysing samples together with a larger dataset of the same population (120 samples used for a previous study, Iacolina et al. subm.). GENEPOP 4.0 (Raymond and Rousset 1995) was used to estimate F_{IS} and to test loci for Hardy-Weinberg equilibrium (HWE).

The minimum number of fathers actually required to explain genotypes within a litter, whose mother was known, was calculated following two different methods. The first is based on simple Mendelian inheritance of codominant microsatellite alleles, whereby offspring inherit one allele per locus from each parent. In fact, we visually created a pool of non-maternal alleles for each locus and determined how many fathers were required (at minimum) to explain the littermates genotypes. We conservatively interpreted as evidence of MP in a litter the case in which two or more loci contained overnumeral non-maternal alleles. The occurrence of >2 non-maternal alleles at a single locus was instead interpreted as possibly due to a germ-line mutation (if confirmed after a replicated genotyping).

The second approach to detect evidences of MP relies on the use of COLONY 1.2 (Wang 2004), a Fortran program implementing a maximum likelihood method to assign individuals sampled from a single generation of a population into full-sib families nested within half-sib families (colonies) using genetic data (codominant genetic markers). For the

analysis we used as population allele frequencies those obtained for the extended data set (n = 193) and set the rate of both allelic drop out and other typing errors at each locus at the value of 0.02. After the first run the alleles supporting multiple siring or highlighted by the program as potentially mistyped were re-checked for trivial errors (typing, interpretation, etc.) and independently genotyped once again together with control samples. Finally, inconsistencies which could neither be attributed to genotyping errors nor to MP were interpreted as mutations. In order to check consistency across runs, COLONY analyses were repeated 10 times with different seeding numbers, but leaving all other settings unmodified. In addition, we verified the potential effect of error rate by replicating COLONY analyses with different rates of allelic dropout and of other typing errors, according to the range reported in the literature for analysis of high-quality DNA samples (from 0.00 to 0.10, Baker et al. 2004; Hoffman and Amos 2005; Pompanon et al. 2005).

RESULTS

Variability in the selected set of markers was verified in the population. A total of 76 different alleles were found (4-12 alleles per locus, k = 7.1). Average H_o and H_e were similar, amounting to 0.687 and 0.699 respectively. The overall F_{IS} in the population was 0.016 and the population showed an overall slightly significant deviation from HWE (χ^2 = 34.6, df = 20, P = 0.022), apparently due to a single locus (sw1492, P =0.001). Proportion of missing data amounted to 1.4%.

Minimum numbers of fathers required to explain offspring genotypes within a litter are shown in Table 2. Genetic data (number of different non-maternal alleles per locus) could support a number of different possible mating configurations, but the most parsimonious solution, consisting in the smallest number of possible fathers for each litter, was considered. It reveals that at least in 6 out of 12 families one father was not sufficient to explain the observed littermates' genotypes.

Results of COLONY analysis were consistent across replications. In fact, according to all COLONY runs, six litters were represented by a single group of full-sibs (i.e. one father), five litters were composed each by two nested groups of full-sibs within a cluster of half-sibs (i.e. two fathers), and in one case three groups of full-sibs were detected (i.e. three fathers, Table 2). Results of the analysis were consistent even in presence of higher rates of genotyping errors. The outcome was indeed identical up to an error rate of 0.08. When accounting for an

error rate of 0.10 the number of monogamous families increased from six to eight, but MP was still required to explain genotypes of the remaining four litters.

MP cases detected by COLONY matched with those resulting from Mendelian inheritance analysis. Cases of single and multiple siring had the same frequency in adult and yearling females (2 MPs out of 4 in adult and 4 MPs out of 8 in yearling sows). Furthermore, age (in month) and body weight of the mothers showed no appreciable effect on MP, as well as litter size (Mann-Whitney U-tests: age: U = 15.0, P = 0.617; body weight: U = 11.5, P = 0.297; litter size: U = 16.0, P = 0.733).

Considering the whole dataset a single possible mutation case, involving a fetus that did not show any maternal allele at a single locus, has been found, resulting in a mutation rate of 0.0016 gametes per generation. This is obviously a minimum rate, as mutations that do not result in the lack of a maternal allele, as well as mutations in the paternal line, would have not been detected. Finally, no evidence of null alleles was found at any of the ten microsatellites in the analysed families.

DISCUSSION

Multiple paternity was found at high frequency in the Alpe di Catenaia wild boar population in central Italy. Six out of twelve litters (50%) proved to have been sired by more than one boar. Considering that in our study MP was detected by a parsimony approach (e.g. two fathers with similar genotypes might have been interpreted as one single father), the actual degree of polyandry shown by females in the population can be still higher. On the other hand, we considered the possibility to have introduced an opposite bias if, by chance, we had selected a sample of large-sized litters, thus maximizing the chance to detect MP. However, our sample did not deviate from a sample of 96 uteri collected during the hunting season 2007-2008 in the same area (litter size averaging 4.96 ± 1.11, Cappai et al. 2008).

The high rate of MP we observed in the Alpe di Catenaia population is in agreement with the low degree of relatedness that we observed within social groups in the area (Iacolina et al. subm). Actually, if females in a group are fertilized by more than one male, wild boar social units will show a complex pattern of kin relationships and an average relatedness lower than expected, even though maintaining their matrilineal social organization. The combined effect of MP and high hunting mortality can thus easily explain the observed pattern.

MP frequency in our study is higher than that recently reported by Delgado et al. (2008) for a Portuguese wild boar population (Alentejo), where one litter out of nine showed genetic evidences of multiple fathering. Considering the limited sample size, this difference could be due to a different age bias in the two samples. Actually, the Alentejo sample showed a higher proportion of adult females, heavier and with a larger litter size, than those included in the present study (Table 3). Provided that adult sows in good physical conditions are more productive than subadults (Fonseca et al. 2004) and have more experience, we would expect that they are more effective in maximizing benefits of multiple mating. However, the observed difference in MP frequency between the two samples is opposite to this prediction, seemingly ruling out a possible effect of age. In addition, in our study, MP occurred at the same rate in adult and subadult females' litters. This clue suggest that multiple mating by females is not age-dependent, althought our limited sample size prompts for further investigations.

As the likelihood to mate with multiple males depends on the encounter rate, one could expect an effect of the population density. Both the Italian and the Portuguese populations are object of high hunting pressure performed with the same method (drive hunts with dogs). Assuming the hunting bag as proportional to population abundance (being boars hunted ad libitum in both areas), the two populations show similar wild boar densities (Table 3), yet a difference in the environmental productivity (mediterranean vs. temperate habitats). This may suggest that female mating strategy in the population is also not density-dependent. However, a possible source of variability may be due to different spatial behaviour due to enviromental differences (e.g. distribution of food resources) that can affect the mate encounter rate in a population.

MP depends on the breeding strategies adopted by sexes, which are aimed at maximizing their respective fitness. According to Briedermann (1986) the wild boar mating system is fundamentally polygynic, where dominance is established among competing males and matings are monopolized by the dominant male, who guards estrus females, chasing away competitors. In species where such tending behavior is adopted by males during the breeding season, the occurrence of multiple mating depends on the likelihood a male has to sistematically chase away all competitors during the female's fertile period. According to the "displacement hypothesis", invoked by Sorin (2004) to explain MP in white-tailed deer, MP occurs when a low-rank male approaches and mates with a female, and is followed by a high-rank dominant male that chace him and subsequently copulates with the same female. Each of them can father her offspring, thus determining MP. An alternative hypothesis ("sneaking hypothesis") implies that while a tending male is engaged in chasing away a competitor, a third male can profit by his inattention and mate with the female. Mate guarding, rank-based

displacement and sneaking by subordinate males are commonly found in cervids (Endo and Doi 2002, Sorin 2004), and are also reported in suids (Somers et al. 1995). Multiple paternity can, however, be favoured by a different mating strategy. If estrus is asynchronous among female groups and the costs of serching for a mate is not high, some males can maximize their reproductive success by roaming in search of estrus females, renouncing to any active mate defense. This spatial behavior would imply different males to have access to the same females during their estrus, thus allowing MP to occur ("roaming hypothesis").

These hypotheses are not mutually exclusive, as the corresponding behaviors can represent alternative male mating strategies. The chance a male has to guard effectively his mate (or a group of potential mates) depends on his physical status and attitude (e.g. age, body size, body condition, aggressiveness, etc.), but also on the duration of the estrus period. In wild boar, a sow is receptive for approximately 48 hours (Henry 1968), a relatively long time in comparison with other ungulates (e.g. 12-24 hours in cervids, Clutton-Brock et al. 1982, Knox et al. 1988, Wong and Parker 1988). Sows in a group tend to synchronise their estrus, but, altogether, the presence of estrus females in a group can last for as long as one week (Briedermann 1986). Consequently, a male should spend a considerable amount of energy to prevent a previously inseminated female from extra-pair copulations and this effort would be magnified to defend a group of females. It can be argued that only adult boars in good body conditions are able to monopolize access to a group of females. In addition, higher the number of potential mates for each fertilizable female in the population, higher is expected to be the cost associated to mate guarding by males (DeYoung et al. 2002). The commonly practised hunting methods (i.e. drive hunt with dogs) usually do not cause a strong distortion of the population structure, as they lead to random culling (Massolo and Mazzoni della Stella 2006). In Alpe di Catenaia, according to hunting bag data, approximately 20% of the males are adult (> 2 years), with an approximate ratio of 1 adult male : 3.2 females (M. Apollonio, unpubl. data), similar to the ratio observed in Spanish wild boar populations (1 : 2.7-3.2, Fernandez-Llario and Mateos-Quesada 2003). At this level of male-male competition, mate guarding can become a costly strategy. Moreover, as matings occur during the hunting period, a dominant boar can be removed (killed or chased) during his stay with a group of females, leaving the occasion to mate to other males.

Up till now we considered the decisional status of male wild boars, but the role of females should not be overlooked. Multiple paternity is predicted to be beneficial also to females, which are expected to promote polyandry for a variety of reasons (Wolff and Macdonald 2004). First, by making uncertain paternity of their offspring, females can reduce the risk of infanticide. Second, they can simply accept approaching males, in order to avoid

sexual harassment. Third, multiple mating can represent a defence against male infertility and genetic incompatibility, increasing the chance to be fertilized by compatible sperms (Zeh and Zeh 2001). Fourth, females can enhance genetic diversity in the litter, increasing the chance that at least some offspring will survive. Fifth, their goal could be to provide "good genes" to their offspring, through mechanisms of cryptic mate choice or sperm competition (Carling et al. 2003). Wolff and Macdonald (2004), reviewing cases of multi-male mating in mammals, concluded that the first hypothesis (i.e. paternity confusion) seems to be supported in most species. On the other hand, multiple mating was found to be associated with higher intrauterine survival and larger litter size in the domestic pig (Waller and Bilkei 2002), suggesting a direct benefit for females enabling MP. The existence of cryptic mate choice by females was suggested by Aguilera-Reyes et al. (2006) to interpret multiple siring in dometic pig litters. They obtained evidences of female behavioral preference for the male (a hybrid) who turned out to have the highest fertilization success. Similar mechanisms can occur also in the wild form, as suggested by Delgado et al. (2008).

Our findings provide new insights into the mating system of free-ranging wild boar populations. Further studies are warranted to investigate within- and between-sexes interactions during the breeding period, in order to understand the evolutionary advantages of promiscuity and MP in this species. Finally, possible effects of the overlap between hunting and mating periods should be carefully investigated.

ACKNOWLEDGEMENTS

We thank A. Cossu for his help in laboratory analysis; E. Donaggio, E. Bertolotto, P. Bongi and the wild boars teams Bobi 1, Pipistrello, 1° Squadra Catenaia, Calbenzano-Vogognano, Caprese-Sovaggio, Monna-Ponte alla Piera, and O.P.N. for helping us in collecting samples. We are also in debt with S. Ciuti for discussion of results and for his comments to the manuscript. The study was financed by Regione Toscana, Provincia d'Arezzo, which also provided logistical support and by the Ministero dell'Università e della Ricerca (MIUR PRIN 2005). L. Iacolina benefited from a fellowship sponsored by Sardegna Resorts s.r.l.

TABLE 1 - Lists of the wild boar pregnant sows and litters (fetuses) sampled in Alpe di Catenaia, Italy, and analysed for multiple paternity.

Sow	Month of death	Estimated age	Age class	Full weight (Kg)	Litter size	Litter composition
F01	November 2002	19-20 months	Sub-Adult	50	4	1F+3M
F02	January 2004	5-7 years	Adult	66	4	3F+1M
F03	January 2006	3-4 years	Adult	70	6	4F+2M
F04	January 2006	3-4 years	Adult	87	4	3F+1M
F05	January 2006	15-16 months	Sub-Adult	44	6	3F+3M
F06	December 2006	22-24 months	Sub-Adult	48	6	6M
F07	January 2007	22-24 months	Sub-Adult	68	6	1F+5M
F08	December 2006	> 10 years	Adult	64	5	3F+2M
F09	December 2006	19-20 months	Sub-Adult	58	5	2F+3M
F10	December 2006	22-24 months	Sub-Adult	55	5	2F+3M
F11	December 2006	22-24 months	Sub-Adult	58	5	2F+3M
F12	December 2006	22-24 months	Sub-Adult	63	5	2F+3M

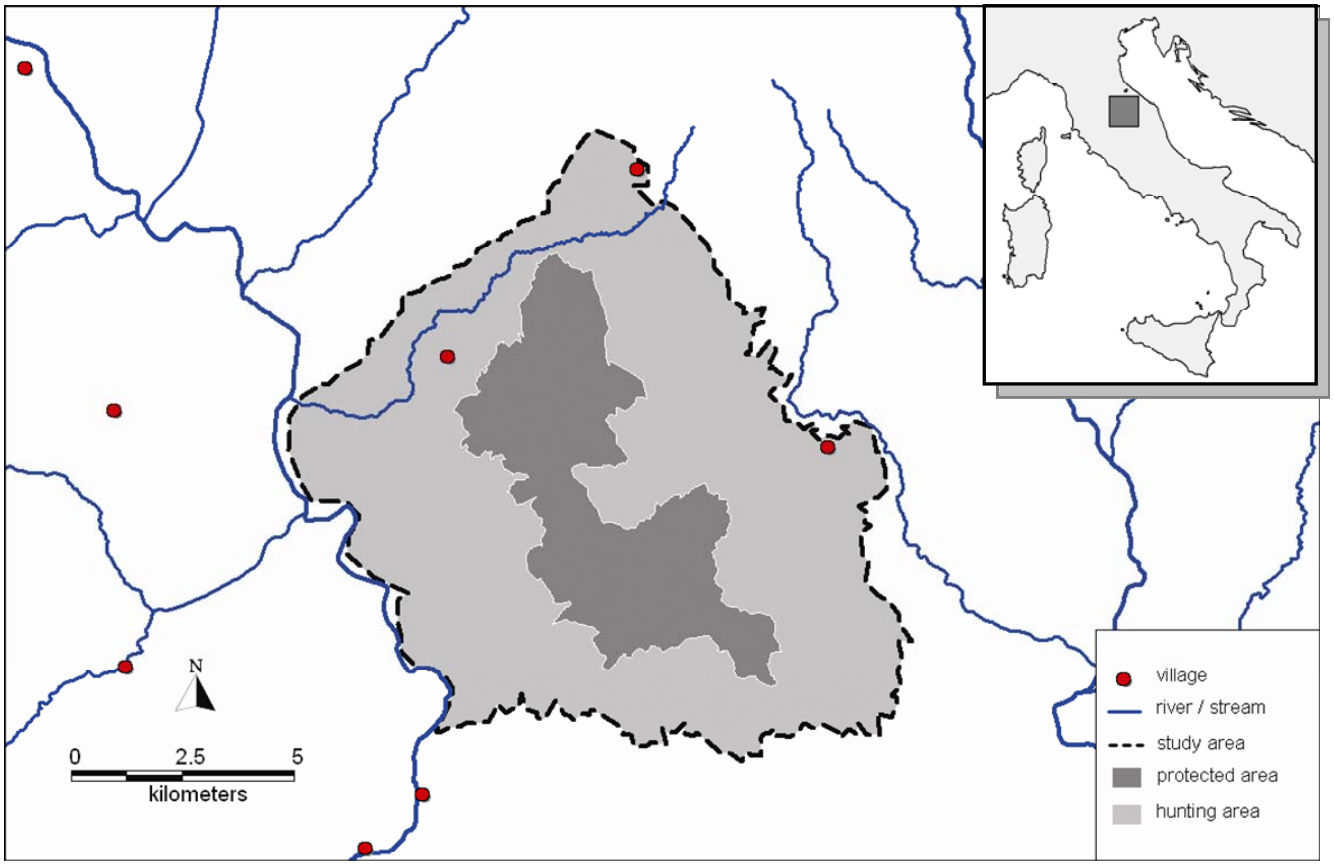
TABLE 2. - Evidence of multiple paternity in wild boar litters in the Alpe di Catenaia population (Tuscany, Italy). Number of non-maternal alleles detected at each microsatellite locus within wild boar litters and the minimum number of fathers required to explain littermates' genotypes are reported. The estimated number of fathers corresponds to the number of full-sibs groups detected by maximum likelihood analysis in COLONY within each litter. Data supporting MP are highlighted in bold.

Litter	**Mother**	**sw1492**	**sw122**	**s090**	**sw2532**	**sw461**	**IGF1**	**sw2021**	**sw2496**	**sw24**	**s155**	**Minimum nr. of fathers**	**Estimated nr. of fathers (COLONY)**
Litter 1	F01	1	1	1	1	1	2	2	2	1	2	1	1
Litter 2	F02	2	1	2	**3**	1	2	**3**	1	2	1	**2**	**2**
Litter 3	F03	1	1	2	1	2	2	2	1	2	2	1	1
Litter 4	F04	1	1	2	1	2	2	1	2	2	2	1	1
Litter 5	F05	2	**3**	2	2	**3**	2	2	**3**	2	1	**2**	**2**
Litter 6	F06	1	2	1	2	**3**	2	**3**	2	2	1	**2**	**2**
Litter 7	F07	1	1	2	1	1	2	2	2	2	2	1	1
Litter 8	F08	**3**	**3**	**3**	2	**3**	2	1	**3**	**3**	1	**2**	**2**
Litter 9	F09	2	2	2	2	1	2	2	2	2	1	1	1
Litter 10	F10	1	2	2	1	1	2	2	2	2	2	1	1
Litter 11	F11	**3**	1	1	2	2	2	2	**4**	**3**	1	**2**	**3**
Litter 12	F12	2	**3**	2	2	1	**3**	2	**3**	2	1	**2**	**2**

TABLE 3. - Comparison between general features and sample composition of the investigated Alpe di Catenaia population (Italy, this study) and the Alentejo population (Portugal, Delgado et al. 2008). Data of average litter size and hunting bag in the Alentejo population are reported respectively in Fonseca et al. 2004 (§) and Fernadez-Llario et al. 2003 (°). Hunting bags are calculated as cumulative average of two consecutive hunting seasons in Alpe di Catenaia, and as average of 17 hunts in the same hunting season in Alentejo.

	Population			
	Litter size ($\bar{x}\pm SD$)	**Hunting bag (boars/100ha, $\bar{x}\pm SD$)**	**Annual rain fall (mm)**	**Bioclimatic zone**
Alpe di Catenaia (Italy)	4.8 ± 1.2	9.6 ± 1.3	800 - 1300	temperate
Alentejo (Portugal)	4.4 ± 1.7 (§)	9.5 ± 8.6 (°)	500 - 700	mediterranean
	Sample			
	Litter size ($\bar{x}\pm SD$)	**% of adult sows**	**Sows' weight ($\bar{x}\pm SD$)**	**MP frequency**
Alpe di Catenaia (Italy)	5.1 ± 0.8	33% (4/12)	60.9 ± 11.6	50% (6/12)
Alentejo (Portugal)	5.6 ± 0.7	89% (8/9)	76.1 ± 16.6	11% (1/9)

FIG. 1- Study area in the Alpe di Catenaia, Arezzo, Italy. The natural reserve (dark grey) and surrounding hunting areas (light grey). Main villages and rivers are shown.



LITERATURE CITED

AGUILLERA-REYES, U., G. ZAVALA-PÁRAMO, J. J. VALDEZ-ALARCÓN, H. CANO-CAMACHO, G. I. GARCÍA-LÓPEZ, AND N. PESCADOR-SALAS. 2006. Multiple mating and paternity determinations in domestic swine (*Sus scrofa*). Animal Research 55:409-417.

BAKER, P. J., S. M. FUNK, M. W. BRUFORD, AND S. HARRIS. 2004. Polygynandry in a red fox population: implications for the evolution of group living canids. Behavioural Ecology 15(5):766-778.

BEUERLE, W. 1975. Freilanduntersuchengen zum Kampf- und Sexualverhalten des europäischen Wildschweines (*Sus scrofa* L.). Zeitschrift für Tierpsychologie 39:211-258.

BRIEDERMANN, L. 1986. Schwartzwild. Neumann-Neudamm ed. Berlin, Germany.

CAPPAI, N., E. BERTOLOTTO, E. DONAGGIO, AND M. APOLLONIO. 2008. Birth distribution and litter size in a wild boar population in Tuscany Apennine. Proceedings of the 7th International Symposium on Wild Boar (*Sus scrofa*) and on Sub-order Suiformes. 28th-30th August 2008, p. 73.

CARLING, M. C., P. A. WISEMAN, AND J. A. BYERS. 2003. Microsatellite analysis reveals multiple paternity in a population of wild pronghorn antelopes (*Antilocapra americana*). Journal of Mammalogy 84(4):1237-1243.

CARMICHAEL, L.E., G. SZOR, D. BERTEAUX, M.A. GIROUX, C. CAMERON, AND C. STROBECK. 2007. Free love in the far north: plural breeding and polyandry of artic foxes (*Alopex lagopus*) on Bylot Island, Nunavut. Canadian Journal of Zoology 85 (3):338-343.

CLUTTON-BROCK, T.H., F. GUINNESS, AND S. D. ALBON. 1982. Red Deer: Behavior and Ecology of Two Sexes. University of Chicago Press, Chicago, IL.

CLUTTON-BROCK, T.H. 1991. The evolution of parental care. Princeton University Press ed. Princeton, USA.

CLUTTON-BROCK, T.H., AND K. ISVARAN. 2006. Paternity loss in contrasting mammalian societies. Biology Letters 2:513-516.

CRAWFORD, J. C., Z. LIU, T. A. NELSON, C. K. NIELSEN, AND C. K. BLOOMQUIST. 2008. Microsatellite analysis of mating and kinship in beavers (*Castor canadensis*). Journal of Mammalogy 89 (3):575-581.

DARDAILLON, M. 1988. Wild boar groupings and their seasonal changes in the Camargue, South France. Zeitschrift für Jagdwissenschaft 53:22-30.

DELGADO, R., P. FERNÁNDEZ-LLARIO, M. AZEVEDO, A. BEJA-PEREIRA, AND P. SANTOS. 2008. Paternity assessment in free-ranging wild boar (*Sus scrofa*) – Are littermates full-sibs? Mammalian Biology 73:169-176.

DEYOUNG, R. W., S. DEMARAIS, R. A. GONZALES, R. L. HONEYCUTT, AND K. L. GEE. 2002. Multiple paternity in white-tailed deer (*Odocoileus virginianus*) revealed by DNA microsatellites. Journal of Mammalogy 83:884-892.

ELLEGREN, H. 2000. Microsatellite mutations in the germline: implications for evolutionary inference. Trends in Genetics 16:551-558.

ENDO, A., AND T. DOI. 2002. Multiple copulations and post-copulatory guarding in a free-living population of Sika deer (*Cervus nippon*). Ethology 108:739-747.

FERNANDEZ-LLARIO, P., AND P. MATEOS-QUESADA. 2003. Population structure of the wild boar (*Sus scrofa*) in two Mediterranean habitats in the western Iberian Peninsula. Folia Zoologica 52:143-148.

FERNANDEZ-LLARIO, P., P. MATEOS-QUESADA, A. SILVÉRIO, AND P. SANTOS. 2003. Habitat effects and shooting techniques on two wild boar (*Sus scrofa*) populations in Spain and Portugal. Zeitschrift Jagdwissenschaft 49:120-129.

FONSECA, C., ET AL. 2004. Reproduction in the wild boar (*Sus scrofa* Linnaeus, 1758) populations of Portugal. Galemys 16(n° especial): 53-65.

HENRY, G.V. 1968. Lenght of estrous cycle and gestation in European wild hogs. Journal of Wildlife Management 32(2):406-408.

HOFFMAN, J. I., AND W. AMOS. 2005. Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. Molecular Ecology 14:599-612.

IACOLINA, L., M. SCANDURA, P. BONGI, AND M. APOLLONIO. (subm.) Non-kin associations in wild boar social units. Journal of Mammalogy

ISVARAN, K., AND T. CLUTTON-BROCK. 2007. Ecological correlates of extra-group paternity in mammals. Proceedings of the Royal Society London B 274:219-224.

KAMINSKI, G., S. BRANDT, E. BAUBET, AND C. BAUDOIN. 2005. Life-history patterns in female wild boars (*Sus scrofa*): mother-daughter postweaning associations. Canadian Journal of Zoology 83:474-480.

KNOX, W. M., K. V. MILLER, AND R. L. MARCHINTON. 1988. Recurrent estrous cycles in white-tailed deer. Journal of Mammalogy 69:384-386.

KOCHER, T. D., ET AL. 1989. Dynamics of mitochondrial DNA evolution in animals: Amplification and sequencing with conserved primers, Proceedings of the National Academy of Sciences USA 86:6196-6200.

MASSOLO, A., R. MAZZONI DELLA STELLA. 2006. Population structure variations of wild boar *Sus scrofa* in central Italy. Italina Journal of Zoology 73(2):137-144.

MATTIOLI, L., M. APOLLONIO, V. MAZZARONE, AND E. CENTOFANTI. 1995. Wolf food habits and wild ungulate availability in the Foreste Casentinesi National Park; Italy. Acta Theriologica 40:387–402.

MATTIOLI, L., C. CAPITANI, E. AVANZINELLI, I. BERTELLI, A. GAZZOLA, AND M. APOLLONIO. 2004. Predation by wolves (*Canis lupus*) on roe deer (*Capreolus capreolus*) in north-eastern Apennine, Italy. Journal of Zoology London 264: 249-258.

NIELSEN, C. L. R., AND C. K. NIELSEN. 2007 Multiple paternity and relatedness in southern Illinois racoons (*Procyon lotor*). Journal of Mammalogy 88(2):441-447.

PARK, S. D. E. 2001. Trypanotolerance in West African cattle and the population genetic effects of selection. Ph.D. dissertation, University of Dublin, Dublin, Ireland.

POMPANON, F., A. BONIN, E. BELLEMAIN, AND P. TABERLET. 2005. Genotyping errors: causes, consequences and solutions. Nature Review Genetics 6:847-846.

RAYMOND, M., AND F. ROUSSET. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. Journal of Heredity 86:248-249.

RICHARD, K.R., S.W. MCCARREY, AND J.M. WRIGHT. 1994. DNA sequence from the SRY gene of the sperm whale (*Physeter macrocephalus*) for use in molecular sexing. Canadian Journal of Zoology 72:873-877.

SOMERS, M. J., O. A. E. RASA, AND B. L. PENZHORN. 1995. Group structure and social behaviour of warthogs Phacochoerus aethiopicus. Acta Theriologica 40:257-281.

SORIN, A. B. 2004. Paternity assignment for white-tailed deer (*Odocoileus virginianus*): mating across age classes and multiple paternity. Journal of Mammalogy 85(2):356-362.

STOCKLEY, P. 2003. Female multiple mating behaviour, early reproductive failure and litter size variations in mammals. Proceedings of the Royal Society London B 270:271-278.

TRIVERS, R. L. 1972. Parental investment and sexual selection. Pp. 136-179 in Sexual selection and the descent of man (B. Campbell, ed.). Aldine, Chicago, IL.

XUE, J. L., T. LUCIA, Y. KOKETSU, G. D. DIAL, AND W. E. MARSH. 1998. Effect of mating frequency and weaning-to-mating interval on sow reproductive performance. Swine Health and Production 6(4):157-162.

WALLER, C. M., AND G., BILKEI. 2002. Effects of mating frequency on sow reproductive performance under field conditions and natural mating in large outdoor production units. Reproduction Domestic Animals 37:116-118.

WANG, J. 2004. Sibship reconstruction from genetic data with typing errors. Genetics 166:1963-1979.

WOLFF, J. O., AND D. W. MACDONALD. 2004. Promiscuous females protect their offspring. Trends in Ecology and Evolution 19:127-134.

WONG, B., K. L. PARKER. 1988. Estrus in black-tailed deer. Journal of Mammalogy 69(1):168-171.

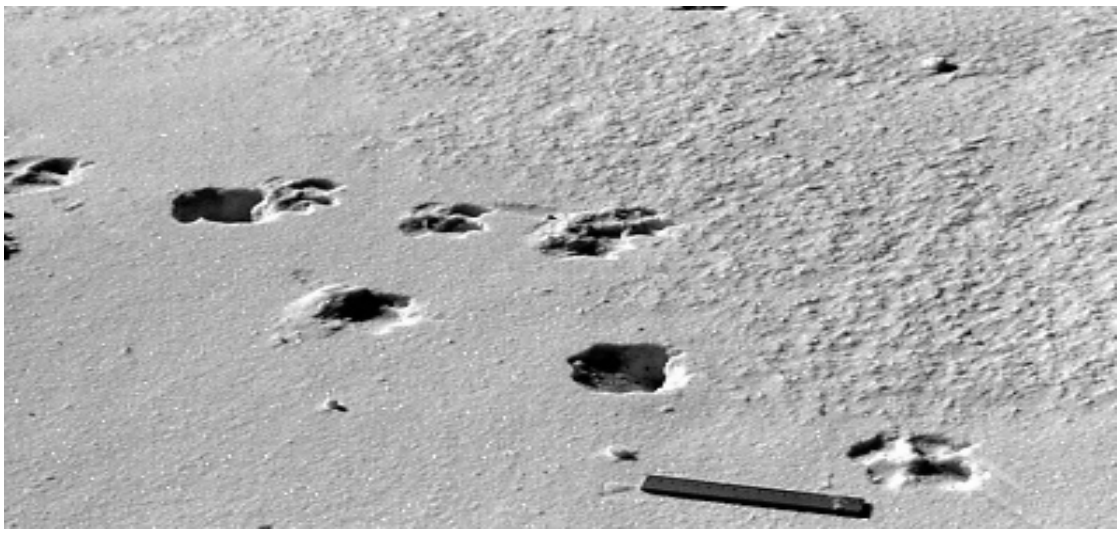
ZEH, J. A., AND D. W. ZEH. 2001. Reproductive mode and the genetic benefits of polyandry. Animal Behaviour 61:1051-1063.

THIRD PART

Non-invasive methodology approach in the study of a wolf population

CHAPTER IV

An empirical approach for reliable microsatellite genotyping of wolf DNA from multiple noninvasive sources



An empirical approach for reliable microsatellite genotyping of wolf DNA from multiple noninvasive sources

Massimo SCANDURA, Claudia CAPITANI, Laura IACOLINA and Marco APOLLONIO

Conservation Genetics (2006) 7:813–823

Conservation Genetics (2006) 7:813–823
DOI 10.1007/s10592-005-9106-5

An empirical approach for reliable microsatellite genotyping of wolf DNA from multiple noninvasive sources

Massimo Scandura[1,2,*], Claudia Capitani[1], Laura Iacolina[1] & Apollonio Marco[1]
[1]*Department of Zoology and Anthropology, University of Sassari, Via Muroni 25, I-07100, Sassari, Italy;* [2]*Lehrstuhl für Verhaltensforschung, Universität Bielefeld, Morgenbreede 45, D-33615, Bielefeld, Germany (*Corresponding author: Phone: +39-079-228667; Fax: +39-079-228665; E-mail: scandura@uniss.it)*

Received 28 July 2005; accepted 15 December 2005

Key words: microsatellites, noninvasive genotyping, quality control, scoring errors, wolf

Abstract

Wildlife management and conservation take advantage of the possibility to study free-living populations by collecting and analysing noninvasive samples. Nevertheless, the commonly adopted approaches, aimed at preventing results being affected by genotyping errors, considerably limit the applicability of noninvasive genotyping. An empirical approach is presented for achieving a reliable data set of wolf (*Canis lupus*) genotypes from multiple sources of DNA collected in a monitored population. This method relies on the relationship between sample quality and amplification outcome, which is ultimately related to the occurrence of typing errors (allelic dropout, false alleles). After DNA extraction, templates are amplified once at each locus and a conservative rating system (*Q*-score) is adopted to define the quality of single-locus amplifications. A significant relationship was found between quality scores and error rate (ER) ($r^2 = 0.982$). Thus it was possible to predict the chance a genotype has of being affected by errors on the basis of its *Q*-score. Genotypes not reaching a satisfactory confidence level can either be replicated to become reliable or excluded from the data set. Accordingly, in the present case study, 48–73% of all single-locus and 51–53% of all multilocus (ML) genotypes reached a sufficient (99 and 95%, respectively) reliability level after a single amplification per locus. Despite the possible decrease in overall yield, this method could provide a good compromise between accuracy in genotyping and effectiveness in screening large data sets for long-term or large-scale population surveys. However, to achieve complete and reliable data sets, replicated amplifications are necessary for those samples and loci providing poor results.

Introduction

The recourse to genetic approaches in wildlife and conservation biology has increased in the last years, especially after that previously disregarded materials – like faeces, hairs, urine, haematic residuals, etc. – were proved to represent a valuable source of DNA for genetic analyses (for a review Morin and Woodruff 1996; Piggott and Taylor 2003). The most important implication to their use is the possibility to collect biological information ''noninvasively'', i.e. without a detectable impact on the population under study. Shortly after the announcement of the availability of this new tool for population studies, besides the praises of noninvasive genotyping being sung, some pitfalls were pointed out (Taberlet et al. 1996; Kohn and Wayne 1997). It was revealed that the use of very small or degraded DNA samples for PCR-based genetic analyses, apart from being extremely sensitive to any source of contamination, might be accompanied by the occurrence of stochastically-generated errors, thus leading to inconsistent results (Goossens et al. 1998; Taberlet

et al. 1999). Scoring errors may arise from the amplification failure of one of the two alleles at a heterozygous locus (''allelic dropout'') or from the amplification of erroneously generated alleles (''false alleles''), either due to stochastic events during PCR or to the amplification of concurrent unspecific DNA templates. Consequently, both false homozygous and false heterozygous genotypes may be generated. To overcome this limitation, some methods were proposed, aimed at reaching a fixed reliability threshold by replicating amplifications several times. The required number of replications was defined either from *a priori* generalised error probabilities (Navidi et al. 1992; Taberlet et al. 1996) or from maximum-likelihood estimates of genotype reliability, computed from individual dropout rate and allele frequencies (Miller et al. 2002). Great variation was observed both in amplification success and in the frequency of genotyping errors, depending on species studied, source of DNA, and laboratory protocols and facilities (Piggott and Taylor 2003; Broquet and Petit 2004). Therefore, it was suggested, when planning a noninvasive genetic survey, to precede it with a pilot study aimed at evaluating the amplification success and at carefully estimating ER for the employed combination of sample types, molecular markers, and DNA protocols (Taberlet and Luikart 1999; Broquet and Petit 2004). Nevertheless, some relevant aspects have been pointed out. First, a huge variability in DNA abundance and quality can exist among samples (Gagneux et al. 1997; Goossens et al. 1998; Morin et al. 2001; Miller et al. 2002), determined by several factors like number of hair follicles used for DNA isolation (Goossens et al. 1998; Paetkau 2003), sampling season (Lucchini et al. 2002; Maudet et al. 2004), as well as time elapsing between sample deposition, collection and analysis (Sloane et al. 2000). This variation is particularly evident when different sources of DNA (faeces, hairs, etc.) are employed. Second, a few ''problematic'' samples may negatively affect the estimation of ER, when calculated across all samples and loci (Sloane et al. 2000; Parsons 2001; Hedmark et al. 2004; Scandura 2005); this effect is stronger if the estimation relies on a limited number of samples. Third, relying on theoretical models based on the worst-case scenario (i.e. 100% allelic dropout – Taberlet et al. 1996), to define the required number of replicates assuring that all genotypes are correct at a 99% confidence level, leads to a magnification of the effort (time, consumables and required DNA) often exceeding the actual need (Sloane et al. 2000; Miller et al. 2002).

Considering that the suitability of noninvasive samples for DNA-based population surveys largely depends on the effort required to achieve confident data (Paetkau 2003), a procedure based on a sample-by-sample evaluation seems to be more effective, when a large variance in sample reliability is expected. Such evaluation can precede microsatellite genotyping and pass through an accurate quantification of each DNA sample (Goossens et al. 1998; Morin et al. 2001), but this step is costly (both in terms of money and of depleted DNA solution) and problematic at limiting conditions (few template molecules). Alternatively, a post-amplification screening based on preliminary amplifications with one (or more) of the most robust microsatellite markers in the set in use (Sloane et al. 2000) or consisting in the scrutiny of single-locus (SL) genotypes can lead to discard low-quality samples and to prevent the occurrence of genotyping errors (Paetkau 2003).

We constructed a large data set, represented by the results of single-tube amplifications of wolf (*Canis lupus*) DNA performed at 10 microsatellite loci during a DNA-based population survey. Samples had been obtained either from wolf carcasses or by noninvasive collection. At this point, according to the literature, we expected a non-negligible proportion of those genotypes was affected by errors (dropouts or false alleles). In that case, such an amount of molecular data would not be suitable to study the genetic structure of the wolf population. Hence, we managed to scrutinise the quality of the collected information, by estimating the proportion of erroneous genotypes in a subset of the available data. In order to avoid any sampling bias, ER was estimated by performing a random selection of sample/marker pairs from the complete data set and replicating their PCRs until a high-confidence genotype was obtained. Moreover, since the yield of a PCR is dependent on the quality and quantity of the provided DNA template (Golenberg et al. 1996), we expected that the outcome of microsatellite amplifications could be used to judge the quality of templates as well as to predict their sensitiveness to genotyping errors.

Therefore, the main aim of the study was to develop a straightforward empirical procedure to

select *a posteriori* reliable genotypes in a data set and discard those reaching an inadequate confidence level. In addition, we evaluated whether genotype reliability could be assessed on the basis of amplification profiles, verifying the relationship between amplification outcome and genotyping ER. Finally, we explored data quality in the original data set, describing the effect of different factors possibly affecting sample reliability.

Materials and methods

Sample collection

Noninvasive samples were collected from 1998 to 2004 during a long-term survey on a wolf population in Italy. As sources of DNA for the analysis, three kinds of samples were used: scats, shed hairs, and blood spots on the snow. Most noninvasive samples were collected in winter while tracking wolves in snow. In this season, samples were supposed to be more suitable for genetic analysis, as regards freshness and DNA preservation. Only fresh faecal samples were collected (usually < 3 days old). Indeed, in winter most samples were collected following wolf tracks in snow within 24–72 h after a snowfall. In summer, scats become rapidly dry, thus wet (fresh) excrements were collected in absence of precipitations. Once a scat was collected, a portion of it was carefully removed and stored in a polypropylene collection tube containing 95% ethanol. Hairs were gathered from the substrate (snow, soil, barbed wire, etc.) using a plastic bag turned inside-out, in order to avoid any manipulation. They were collected individually or pooled together, depending on the grouping pattern observed by the collector. Blood stains were exclusively found on snow; their collection, storage conditions and DNA extraction methods were described in a previous study (Scandura 2005). Subsequently, all samples were transferred to the laboratory where they were stored at −18 °C until analysis. All processed samples, at the moment of their collection, were attributed to wolf on the basis of morphology and association with wolf signs (e.g. tracks).

A total of 464 noninvasively collected samples was analysed, made up of 272 scats, 157 shed hairs and 35 blood spots. Additional samples were represented by plucked hairs or skeletal muscle from 34 wolf carcasses recovered in the study area between 1991 and 2004.

DNA extraction and typing

DNA was extracted from faecal samples using either the GeneClean for Ancient DNA kit (BIO101 Inc.) or the QIAamp DNA stool kit (Qiagen), following the manufacturers' instructions. DNA isolation from hair follicles relied on Chelex-100 resin (Biorad) according to Walsh et al. (1991). A pre-screening of hair samples was conducted, during which hairs were individually scrutinised under a magnifying glass. Up to 10 hairs carrying visible follicles were used per single extraction. Whenever possible, 8–10 hairs were used in a volume of 200 μl. Alternatively, volume was proportionally reduced to a minimum of 60 μl for a single hair. Forceps and tweezers were always alcohol-flamed before handling a new sample. DNA from diluted blood samples was isolated from 200 μl of melted blood–snow mixture using the QIAamp DNA Blood kit (Qiagen) and suspended in a variable volume, depending on the estimated initial dilution (Scandura 2005).

Ten microsatellites including dinucleotides (109, 123, 204, 250 and 377 – Ostrander et al. 1993; 1995), and tetranucleotides (FH2004, FH2054, FH2137, FH2158 and FH2175 – Francisco et al. 1996), derived from the domestic dog, were selected for wolf genotyping. Allele sizes of loci ranged from 110 to 320 bp.

Amplifications were performed in 10 μl reaction volume containing 3 μl of template, 0.5 U of Euro *Taq* DNA polymerase (EuroClone), 10 mM Tris–HCl, 50 mM KCl, 3 mM $MgCl_2$, 0.1 mg/ml BSA, 100 μM of each dNTP and 2 pmol of each primer. The PCR profile was set up with an initial denaturation step at 95 °C for 3 min, followed by 35 cycles of amplification consisting in 40 s at 92 °C, 40 s at the established annealing temperature (55–58 °C), and 30 s at 72 °C. A final 10-min extension step at 72 °C was added. One primer of each pair was end-labelled with fluorescent ABI dyes (6-FAM, TET and HEX). The outcome of each amplification was verified by running 5 μl of each PCR product on 2% agarose gel containing ethidium bromide. A 100 bp DNA ladder (MBI Fermentas) was included in each run, as reference for fragment size and band intensity. If the

expected product was observed in the gel, 1 μl of PCR solution was loaded into a well of a 96-well tray, previously filled with 12 μl formamide and 0.4 μl GS-500 ROX size standard (Applied Biosystems). Thereafter, alleles were sized by capillary electrophoresis on an ABI PRISM 3100-Avant automatic sequencer and allele lengths were determined by the GENEMAPPER software (Applied Biosystems).

In order to avoid cross-contamination among DNA samples during extraction and PCR, some precautions were taken: all reactions were performed in a laminar flow hood, aerosol-resistant pipette tips were used and pre- and post-PCR experiments were carried out in separate rooms. In all cases, the occurrence of contamination events was monitored by including negative controls in each PCR experiment. Finally, any amplification in which contamination was suspected, was discarded.

Quality assessment

The quality control procedure relied on the attribution of a quality score (*Q*-score) to every SL-genotype, referred to the outcome of two intermediate steps of the analysis. Single-locus scores were then averaged over loci to provide a multilocus score, expression of the overall quality of the ML-genotype. The assigned scores referred to: (i) amplification quality (PCR score) and (ii) microsatellite profile quality (SEQ score). After running PCR products on agarose gel, the pattern of each sample run was evaluated on an UV-light transilluminator and a score (PCR score) was assigned, considering band intensity, sharpness, and "cleanness" (absence of nonspecific products) (Figure 1a). The second check concerned the quality of the microsatellite profile, evaluated considering peak shape and height, together with the presence and conformation of shadow peaks (Figure 1b). In this case too, a score (SEQ score) was attributed to the sample. Both PCR and SEQ scores decreased with the better quality of the genotype, ranging between 1 (very good) and 4 (bad). Since score attribution is subjective, in order to make it conservative, the evaluation was always carried out by the same person (M.S.). A specific database was constructed, reporting for each SL-genotype the history of repetitions and the codified quality of the relative amplifications.

Estimate of ER

Two hundred sample/locus combinations were drawn from the data set including all samples typed at a minimum of three of the 10 microsatellites used for individual recognition. Single-locus amplifications were carried out for each selected DNA sample until a "consensus genotype" was obtained. Consensus SL-genotype for a given sample was defined, that supported by at least two replicates for heterozygotes (each allele scored twice) and by three replicates for homozygotes (three identical homozygous profiles). All PCR repetitions carried out for each sample/locus combination concurred in estimating ER. Two classes of genotyping errors were estimated according to Broquet and Petit (2004). The rate of allelic dropout (ADO) was calculated as the number of positive amplifications involving the loss of one allele divided by the overall number of positive amplifications attributed to heterozygous individuals (according to the corresponding consensus genotypes). Similarly, the rate of false alleles (FA) was estimated as the number of PCRs in which false alleles occurred, divided by the overall number of positive amplifications (irrespective of whether they corresponded to homozygous or heterozygous genotypes). Overall ER was estimated as the number of detected SL-genotypes differing from the respective consensus genotype divided by the total number of scored SL-genotypes. The effects of genotype quality and microsatellite length on ER were tested using its arcsine-square-root transformation to meet the assumption of normality (Sokal and Rohlf, 1995). In the former case, the error rate was estimated, as reported above, for each *Q*-score class irrespectively of the nature of samples and markers used, and a model was developed by regression analysis with the aim to predict the reliability of a genotype on the basis of its assigned *Q*-score. Microsatellite size and ER were tested for correlation by using the median allele size of each marker in the data set.

To summarise, the analysis proceeded as follows: (1) all DNA samples underwent PCR amplifications under standard conditions at all loci (if no product was detectable at one locus, the PCR was repeated up to 3 times), (2) PCR products and microsatellite profiles were checked for quality and rated, (3) 200 sample/locus pairs,

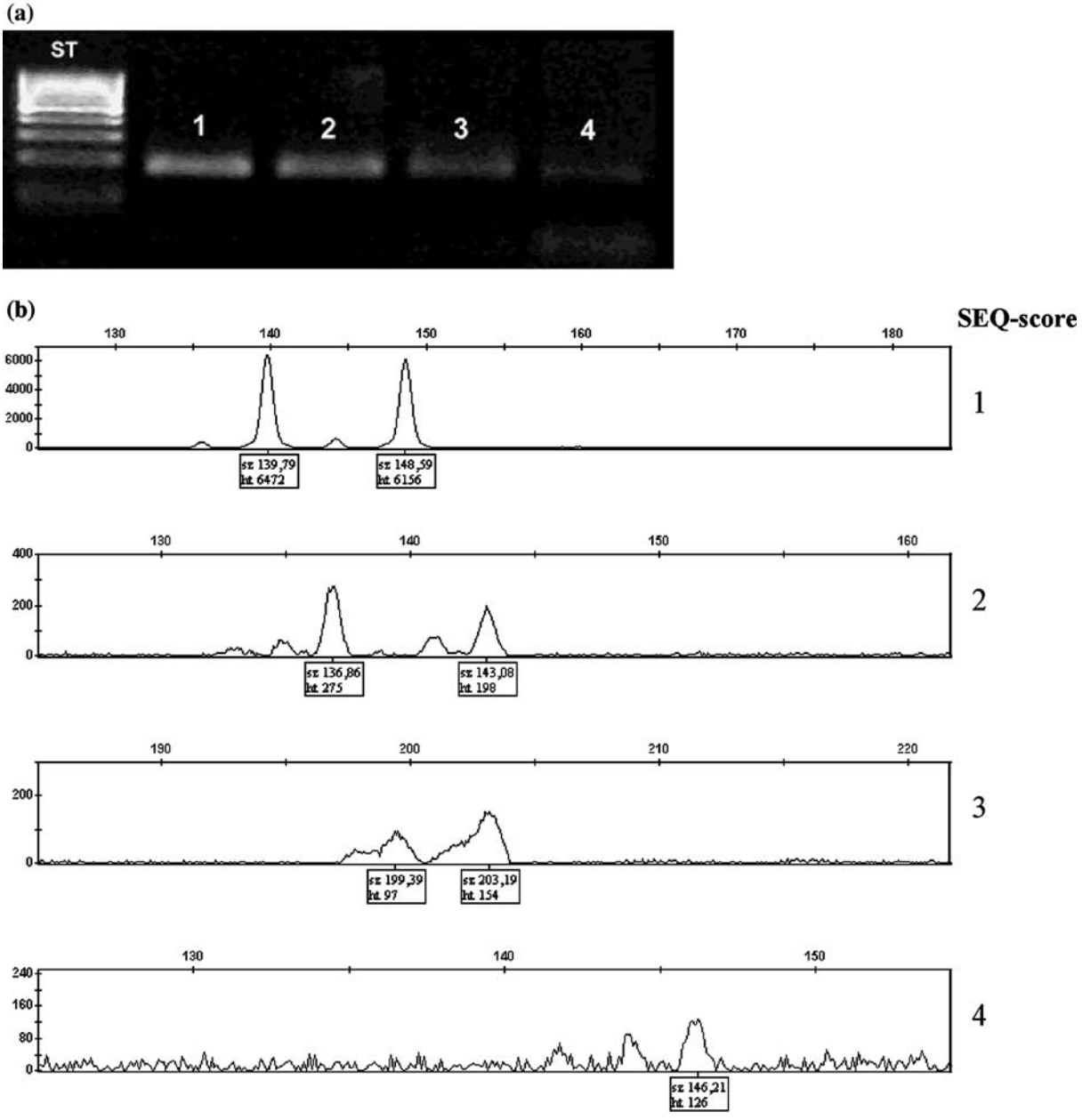


Figure 1. Examples of *Q*-score attribution to (a) microsatellite amplification and (b) peak detection steps in noninvasive genotyping of wild wolves. ST – 100 bp DNA ladder size standard.

representing all loci and sample types, were randomly selected from the data set, (4) for each of them PCRs were replicated until a consensus genotype was yielded, (5) ER was estimated for the replicated data set and referred to sample type, locus and *Q*-score.

Exploration of data quality

Once the dependence of ER on genotype data quality had been verified, the effect of sample and microsatellite features on the latter was investigated, by exploring the complete data set. For

calculations, the following nominal variables were taken into account: source of DNA (tissue, hairs, scat, diluted blood), collection period (cold season = Nov–Apr; warm season = May–Oct), microsatellite type (di- or tetra-nucleotide) and locus length (<190 or >190 bp). Variation in Q-score was previously examined in relation to each single variable using contingency tables and testing departures from a random distribution by χ^2 test for goodness of fit (Sokal and Rohlf 1995). Variables showing a remarkable effect on genotype quality were tested for inclusion in a multi-factorial model by ordinal logistic regression analysis. Similarly, multilocus quality scores per sample were examined, in order to verify their dependence on sample quality and to evaluate their relation to the overall number of scored loci, a relevant variable in noninvasive population surveys. Statistical calculations were performed using SPSS version 13.0 (SPSS Inc.).

Results

Genotype quality and ER

The complete data set was composed by 238 noninvasive samples (141 faecal, 73 shed-hair and 24 diluted blood samples) and 31 genotyped wolf carcasses (17 tissues and 14 plucked hairs), which were successfully typed at, on average, 7.0 microsatellites (SD = 2.2). For 162 replicated sample/locus pairs (93 different samples) among those selected for the analysis, corresponding to a total of 446 amplifications, a consensus genotype was obtained (Table 1). The difference respect to the planned 200 replicated combinations was due to the exhaustion of some DNA samples (mostly hairs) before reaching a consensus genotype. Of the performed PCRs, 313 (70%) concerned heterozygous and 133 (30%) homozygous states. Errors occurred in 3.6% of the amplifications and were either due to amplification failure of one allele at a heterozygous locus (undetectable at homozygous loci) or associated with the appearance of falsely generated alleles. Specifically, rates of allelic dropout and false alleles were estimated at 2.9 and 1.6%, respectively. With respect to sample type, faecal samples showed the highest average frequency of dropout (5.6%), whereas hairs had the highest average FA rate (3.6%), as reported in Table 1.

Quality scores assigned to the outcome of each amplification ranged between 2 and 7. ER increased from 1.0%, for amplifications with a score of 2, up to 12.5% for bad amplifications (Q-score ≥ 6). No error was detected in the few (n = 3) amplifications with score 7. No correlation was found between ER and the median allele length at each locus ($r = 0.155$, n = 10, $P = 0.67$.). As shown in Figure 2, a significant linear relationship was observed between quality score and ER ($r^2 = 0.982$, df = 3, $F = 162.36$, $P = 0.001$). The observed regression represents a model which allows to predict the risk of producing erroneous genotypes on the basis of data quality.

Single-locus genotype quality and reliability

Subsequently, the model was implemented to explore the complete data set, in order to determine overall levels of reliability. Out of 1840 SL-genotypes, 1349 (73%) were expected to be correct with ≥95% confidence, on the basis of their associated Q-score (Figure 3), and 889 (48%) of them with a ≥99% confidence. Of the remaining genotypes, 731 (40%) could reach, on the basis of the prediction, the latter reliability threshold after a single replication, which confirmed the first result.

Table 1. Composition of the replicated data set and estimated error rates

Sample class (number of samples)	Number of selected sample/locus items	Total number of PCRs	Number of detected genotyping errors	Number of detected allelic dropout (%)	Number of detected false alleles (%)
Faeces (50)	61	169	11	7 (5.6)	4 (2.4)
Hairs (15)	23	55	2	0 (0.0)	2 (3.6)
Diluted blood (21)	66	188	2	2 (1.8)	0 (0.0)
Tissue (7)	12	34	1	0 (0.0)	1 (2.9)
Total (93)	162	446	16	9 (2.9)	7 (1.6)

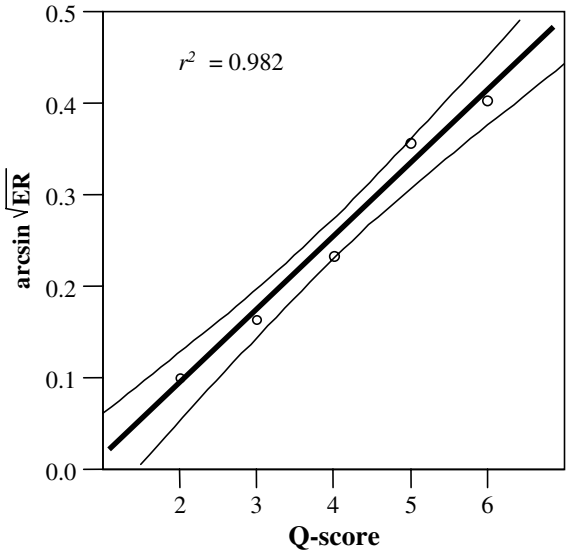


Figure 2. Relationship between assigned *Q*-score and error rate (ER), based on 446 PCRs (95% confidence intervals of the regression line are shown). The arcsine transformation of ER values was used. Both allelic dropouts and false alleles contributed to the calculation of overall ER. The equation for the regression is: arcsin $\sqrt{\text{ER}} = 0.07993(Q\text{-score}) - 0.06846$.

SL-genotype quality varied greatly in relation to sample type and to the size class of the used markers (Table 2). Instead, no effect of sampling season and microsatellite composition was detected. Among samples, tissues collected from wolf carcasses and blood spots occasionally found on snow revealed the best performance in terms of reliability (92% and 88% of genotypes, respectively, having a *Q*-score above the 95% reliability threshold). No model based on logistic regression analysis was able to confidently predict for a given sample the probability of producing a reliable genotype at a specific locus, relying merely on sample features and used marker (data not shown).

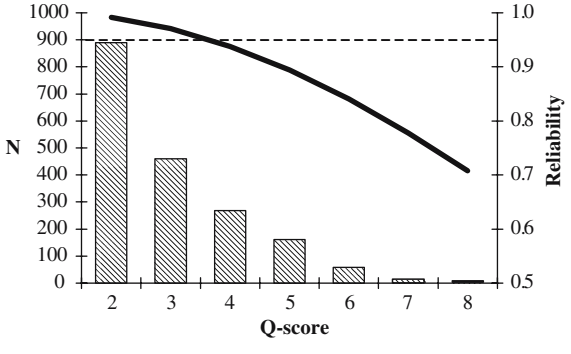


Figure 3. Distribution of *Q*-scores across the data set and corresponding reliability level (solid line), estimated from the model. Hatching indicates the 95% reliability threshold.

Multilocus genotype quality and reliability

With regard to ML-genotypes, mean *Q*-score was 3.1 (SD = 0.8), ranging between 2.0 and 5.8. Once again a relevant variation was observed among sample classes (single classification ANOVA, $F = 12.694$, df = 3, $P < 0.001$), as shown in Figure 4. Blood spots performed significantly better than hairs (*t*-test, $t = -3.716$, df = 62.3, $P < 0.001$) and faecal samples (*t*-test, $t = 7.073$, df = 49.6, $P < 0.001$), while the two latter classes differed from each other (*t*-test, $t = 3.321$, df = 217, $P < 0.01$). On the contrary, the difference in *Q*-scores between collection periods was negligible both referring to all samples (*t*-test, $t = 0.884$, df = 258, $P = 0.38$) and to faecal samples only (*t*-test, $t = -0.001$, df = 135, $P = 0.99$). Furthermore, *Q*-score was highly correlated with the number of scored loci per sample ($r = -0.522$, n = 260, $P < 0.001$). If low-quality SL-genotypes are excluded from the database, out of 269 typed samples, 143 (53%) and 138 (51%) ML-genotypes were confidently typed at five or more loci, referring respectively to the 95 and 99% reliability thresholds. Five was considered the minimum number of markers needed for population inventories or for similar uses at the heterozygosity level observed in the study. The reduction in overall yield following quality check differed among sample classes and was higher for scats (−34%) and plucked-hairs (−25%) and minimum for blood spots (−3%). Altogether, half of the extracted DNA samples failed to amplify at the tested microsatellite loci, and half of those giving positive amplifications failed to produce a reliable ML-genotype. In brief, one processed sample out of four produced a reliable microsatellite genotype by a single-tube approach.

Discussion

Main goal of the present study was to demonstrate that the reliability of a genotype could be predicted from the outcome of the relative amplifications. In other words, the probability a genotype has of being affected by errors could be *a posteriori*

Table 2. Contingency tables showing the effect of sample and microsatellite features on SL-genotype quality (*Q*-score) in the complete data set ($n = 1840$)

	Q-score							Total number of genotypes (%)	χ^2	df	*P*
	2	3	4	5	6	7	8				
Sample type											
Diluted blood	162	42	19	4	3	0	1	231 (12%)	157.1	18	0.000
Faeces	299	223	153	99	40	8	2	824 (45%)			
Hairs	314	164	83	52	13	1	0	627 (34%)			
Tissue	114	31	8	3	1	1	0	158 (9%)			
Sampling period											
May–Oct	106	62	30	25	9	2	0	234 (13%)	3.9	6	0.687
Nov–Apr	783	398	233	133	48	8	3	1606 (87%)			
Microsatellite size											
Short (< 190 bp)	609	328	182	87	35	5	2	1248 (68%)	17.5	6	0.008
Long (> 190 bp)	280	132	81	71	22	5	1	592 (32%)			
Microsatellite type											
Dinucleotide	529	297	172	88	34	5	1	1126 (61%)	8.8	6	0.185
Tetranucleotide	360	163	91	70	23	5	2	714 (39%)			
Total	889	460	263	158	57	10	3	1840 (100%)			

Microsatellite loci were pooled in relation to allele size and repeat motif. Short (110–190 bp) markers included 109, 123, 250, 377, FH2054 and FH2137, whereas long (190–310 bp) markers included 204, FH2004, FH2158 and FH2175. Additional information on the adopted groupings are described in the text.

estimated on the basis of its quality. However, this is only feasible by applying a rigorous and strictly conservative protocol to assess the quality of each single-locus amplification. The empirical approach we used, based on the qualitative evaluation of two steps in the production of a genotype (PCR products and microsatellite profile), allowed us to estimate the probability of errors associated with a specific genotype, by referring to a linear model developed from the data set at hand. In this case, after the first round of positive amplifications, further replications could be avoided for 48–73% of the successfully processed samples, depending on the reliability threshold adopted. The remaining genotypes could either be discarded or confirmed by additional replicates (i.e. multiple-tube approach).

The rationale behind our study is that we find it extremely wasteful to ignore the huge variance in sample quality, treating from the beginning all DNA samples like the worst one, as the multiple-tube approach actually does. Indeed, hairs can provide variable amounts of DNA (Higuchi et al. 1988; Gagneux et al. 1997; Goossens et al. 1998; Morin et al. 2001) and huge differences in DNA quality and quantity can be found in faecal samples, depending on their freshness and composition, as well as on microclimatic conditions (Lucchini et al. 2002; Murphy et al. 2003; Maudet et al. 2004). Hairs used in the present study had different origins, as they might result from follicle atrophy and passive release (moulting), or they may be actively plucked either by the owner itself, by other wolves in the pack or by contact with objects like barbed wire, bushes, tree trunks, etc. In addition to this, differences in follicle size and in the number of follicles included in each extraction contribute to the high variability in sample quality. As a result of this variation, it could not be worth analysing some samples and, whenever possible, it is convenient to concentrate the effort in genotyping only good-quality samples. Evidently, the earlier the step at which poor error-prone samples are discarded, the greater the benefit deriving from their exclusion to the trade-off of the study. However, a qualitative evaluation of collected samples prior to analysing them is problematic. Hairs may be checked for follicle presence and condition (Taberlet et al. 1997; Woods et al. 1999; Triant et al. 2004), and even the number of follicles may be an indicator of DNA quality (Goossens et al. 1998; Paetkau 2003). More complex is the screening of faeces, as time of deposition is difficult to assess and many environmental factors can affect their condition. Once collected, faecal DNA status depends on storing conditions and

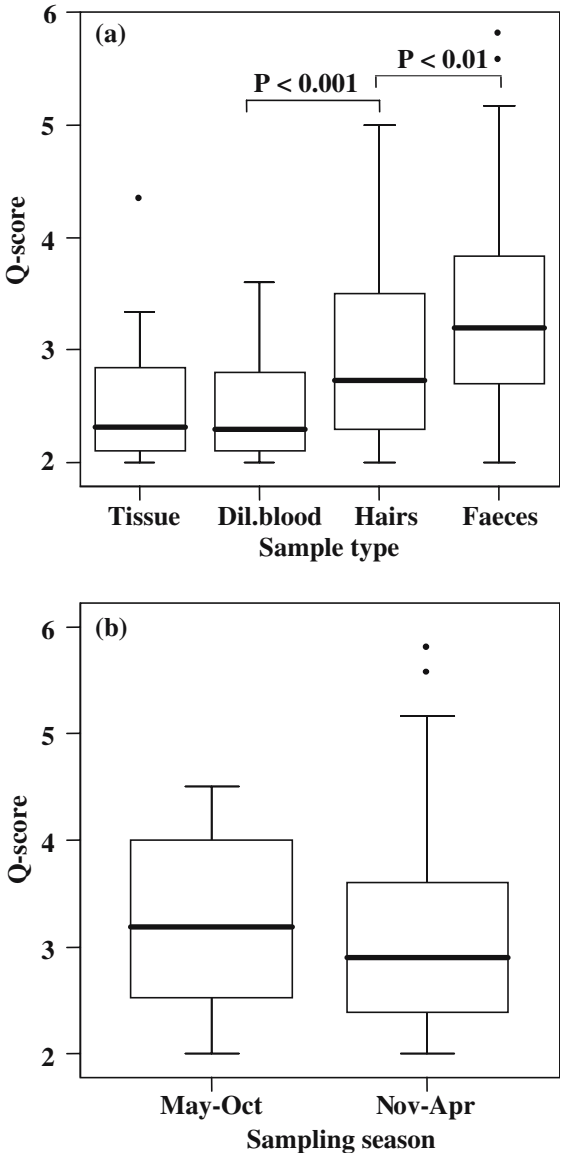


Figure 4. Boxplots of multilocus *Q*-scores attributed to genotypes in relation to (a) sample type and (b) collection period. Significant pairwise differences are reported (*t*-test for independent samples). Outliers are represented by dots.

laboratory procedures, which in this case should be homogenous among samples. Nevertheless, at this step, it is almost impossible for a laboratory technician to assess sample quality. Due to these difficulties the only way to differentiate samples by quality is after DNA isolation, or even after the first amplification. Pre-amplification (Constable et al. 2001; Morin et al. 2001) and post-amplification (Sloane et al. 2000; Paetkau 2003) screening methods proved effective in excluding poor samples from further analyses, affording a considerable saving in the overall effort.

In the present study, we indirectly evaluated sample quality from the outcome of each single amplification, and, more notably, we used the latter as a predictor of typing reliability. Our model was based on a random selection of single-locus amplifications of different samples, in order to avoid the bias originating when using full ML-genotypes of a limited number of samples. Dropout events and the occurrence of false alleles were considered in the calculation of ER. For both sources of error, observed rates were in the range reported for mammal hairs and faeces (see Table 1 in Broquet and Petit 2004), but were relatively low if compared to previous data on noninvasive genotyping of wolf populations (Lucchini et al. 2002; Creel et al. 2003). The marked regression we obtained allowed us to estimate the confidence associated with all other genotypes in the data set, and thus to evaluate for which of them replications were needed. This was done on a locus-by-locus basis, as it was not possible to achieve a multilocus estimation of ER by combining single-locus values, due to their lack of independence. In fact, a sample producing an erroneous genotype at one locus has a higher chance to be incorrectly typed with a second marker (sample quality effect). The effect of the length of amplified regions on genotype reliability revealed somehow contradictory. Allele size seemed to do not affect directly ER, but short (<190 bp) microsatellites produced more high-quality genotypes than larger markers (>190 bp).

We averaged the *Q*-score over loci to obtain an overall estimate of ML-genotype quality per sample, that was less affected, than a single-locus estimate, by the bias due to differences in performance of the markers in use. This index proved largely dependent on the class of noninvasive samples used: blood residuals performed better than hairs, which in turn were more reliable than faeces. On the contrary, no effect of the sampling season was detected. However, a large proportion of samples (87%) was collected in winter and some classes were exclusively (blood spots) or prevalently (hairs) collected on the snow. Therefore, only the performance of faecal DNA was actually comparable between seasons, but even in this case differences in average *Q*-score were not significant,

unlike what was elsewhere observed (Lucchini et al. 2002). Multilocus *Q*-score was highly correlated with the number of typed loci, demonstrating that a high sample quality enhances the chance to perform successful and reliable amplifications. After quality check, the number of samples providing a reliable genotype at 5 or more loci was found to amount to just one forth of all processed samples. This does not mean that the remaining samples should inevitably be culled, but that for them replicated amplifications are necessary for reaching an acceptable confidence level.

Since the method we present is based on subjective criteria, researchers engaged in long-time or large-scale population surveys based on noninvasive genotyping and interested in using this method should perform a pilot study, in order to develop a model that best fit the quality of their data. In these cases, we propose the following procedure: (i) DNA samples are extracted and PCR-amplified at different concentrations with one or more robust microsatellite markers (ii) samples successfully amplifying at these loci are typed once with the complete set of microsatellites; (iii) amplification outcome is checked according to its quality and consequently rated; (iv) a subsample of sample/locus combinations is randomly chosen from the data set and used to accurately estimate the ER by replicating each amplification; (v) the relationship between genotype quality and ER is studied by regression analysis; (vi) the reliability threshold is defined according to the final use of molecular data and employed to explore the complete data set; (vii) PCR replications are finally carried out for samples not reaching the fixed reliability threshold until a consensus genotype, based on a sufficient number of replications (see Miller et al. 2002), is achieved. Exploration of data quality could benefit from the use of specific softwares, like GIMLET (Valière 2002) and MICRO-CHECKER (Van Oosterhout et al. 2004), helping in the identification of scoring errors, or like DNAMIX (Beecham 2004) useful to identify unintentional cross-sample contaminations. Once the reliability of a ML-genotype has been estimated, there exists a further source of information which can increase its confidence level: data redundancy. As noticed by Paetkau (2003), the evidence that several samples in a data set share the same ML-genotype and that the corresponding pairwise match probabilities (Woods et al. 1999) are sufficiently low (i.e. < 0.01) is highly indicative of the reliability of the matching genotype. On the other hand, it could happen that two samples match all alleles but one or two. In this cases, mismatches should be looked on suspiciously, as they could have originated from undetected errors (e.g. dropout); such differences have to be confirmed by replicated PCRs.

In the proposed approach, the first amplification is decisive and therefore the maximum effort should be concentrated at this step. It follows that when limited amounts of template are available, instead of diluting DNA in a greater volume sufficient for multiple PCR repetitions, the template should be concentrated in order to maximize success probabilities and to assure the best performance in the initial round of PCRs. This is particularly important to make samples such as single hairs or diluted blood spots scorable by a sufficient number of microsatellite markers (Paetkau 2003; Scandura 2005). Moreover, even the choice and the number of scored loci is important (Taberlet and Luikart 1999).

In long-term DNA-based population surveys, the effort should be properly balanced in order to assure a sufficient level of data reliability maximising at the same time the number of processed samples (and not *vice versa*). It should be stressed that target level of reliability should be established according to the objectives of the research. We believe that the above-reported approach could help in creating large microsatellite-based data sets for noninvasive surveys of free-living populations.

Acknowledgements

We wish to thank all the people who participated in the wolf monitoring program in the province of Arezzo, among whom we are particularly grateful to Elisa Avanzinelli, Alessia Viviani, Andrea Gazzola, Paolo Lamberti and Luca Mattioli. We also thank Francesca Di Benedetto for her contribution to lab activities and James Burge for linguistic revision. Suggestions by the associate editor and two anonymous reviewers allowed to improve the final draft of the manuscript. Financial support was provided by the Provincial Administration of Arezzo and by the Italian Ministry of University and Research (COFIN 2003, # prot. 2003053710).

References

Beecham GW (2004) DNAMIX v.3 Available at: http://www.statgen.ncsu.edu/~gwbeecha/.

Broquet T, Petit E (2004) Quantifying genotyping errors in noninvasive population genetics. *Mol. Ecol.*, **13**, 3601–3608.

Constable JL, Ashley MV, Goodall J, Pusey AE (2001) Noninvasive paternity assignment in Gombe chimpanzees. *Mol. Ecol.*, **10**, 1279–1300.

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

Francisco LV, Langston AA, Mellersh CS, Neal CL, Ostrander EA (1996) A class of highly polymorphic tetranucleotide repeats for canine genetic mapping. *Mamm. Genome*, **7**, 359–372.

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

Golenberg EM, Bickel A, Weihs P (1996) Effect of highly fragmented DNA on PCR. *Nucleic Acids Res.*, **24**, 5026–5033.

Goossens B, Waits LP, Taberlet P (1998) Plucked hair samples as a source of DNA: Reliability of dinucleotide microsatellite genotyping. *Mol. Ecol.*, **7**, 1237–1241.

Hedmark E, Flagstad Ø, Segerström P, Persson J, Landa A, Ellegren H (2004) DNA-Based individual and sex identification from wolverine (*Gulo gulo*) faeces and urine. *Conserv. Genet.*, **5**, 405–410.

Higuchi R, Von Beroldingen CH, Sensabaugh GH, Erlich HA (1988) DNA typing from single hairs. *Nature*, **332**, 543–546.

Kohn MH, Wayne RK (1997) Facts from feces revisited. *Trends Ecol. Evolut.*, **12**, 223–227.

Lucchini V, Fabbri E, Marucco F, Ricci S, Boitani L, Randi E (2002) Noninvasive molecular tracking of colonizing wolf (*Canis lupus*) packs in the western Italian Alps. *Mol. Ecol.*, **11**, 857–868.

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

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

Morin PA, Woodruff DS (1996) Noninvasive genotyping for vertebrate conservation. In: Molecular Genetic Approaches in Conservation (eds. Smith TB, Wayne RK), pp. 298–313. Oxford University Press, New York.

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

Murphy MA, Waits LP, Kendall KC (2003) The influence of diet on faecal DNA amplification and sex identification in brown bears (*Ursus arctos*). *Mol. Ecol.*, **12**, 2261–2265.

Navidi W, Arnheim N, Waterman MS (1992) A multiple-tubes approach for accurate genotyping of very small DNA samples by using PCR: Statistical considerations. *Am. J. Hum. Genet.*, **50**, 347–359.

Ostrander EA, Sprague GF, Rine J (1993) Identification and characterization of dinucleotide repeat $(CA)_n$ markers for genetic mapping in dog. *Genomics*, **16**, 207–213.

Ostrander EA, Mapa FA, Yee M, Rine J (1995) One hundred and one new simple sequence repeat-based markers for the canine genome. *Mamm. Genome*, **6**, 192–195.

Paetkau D (2003) An empirical exploration of data quality in DNA-based population inventories. *Mol. Ecol.*, **12**, 1375–1387.

Parsons KM (2001) Reliable microsatellite genotyping of dolphin DNA from faeces. *Mol. Ecol. Notes*, **1**, 341–344.

Piggott MP, Taylor AC (2003) Remote collection of animal DNA and its applications in conservation management and understanding the population biology of rare and cryptic species. *Wildl. Res.*, **30**, 1–13.

Scandura M (2005) Individual sexing and genotyping from blood spots on the snow: A reliable source of DNA for noninvasive genetic surveys. *Conserv. Genet.*, DOI 10.1007/s10592-005-9041-5.

Sloane MA, Sunnucks P, Alpers D, Beheregaray LB, Taylor AC (2000) Highly reliable genetic identification of individual northern hairy-nosed wombats from single remotely collected hairs: A feasible censusing method. *Mol. Ecol.*, **9**, 1233–1240.

Sokal RR, Rohlf FJ (1995) *Biometry: The principles and practice of statistics in biological research*, 3rd edn. WH Freeman and Company, New York.

Taberlet P, Griffin S, Goossens B, Questiau S, Manceau V, Escaravage N, Waits LP, Bouvet J (1996) Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Res.*, **24**, 3189–3194.

Taberlet P, Camarra J-J, Griffin S, Uhrès E, Hanotte O, Waits LP, Dubois-Paganon C, Burke T, Bouvet J (1997) Noninvasive genetic tracking of the endangered Pyrenean brown bear population. *Mol. Ecol.*, **6**, 869–876.

Taberlet P, Luikart G (1999) Non-invasive genetic sampling and individual identification. *Biol. J. Linn. Soc.*, **68**, 41–55.

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

Triant DA, Pace RM, Stine M (2004) Abundance, genetic diversity and conservation of Louisiana black bears (*Ursus americanus luteolus*) as detected through noninvasive sampling. *Conserv. Genet.*, **5**, 647–659.

Valière N. (2002) GIMLET: A computer program for analysing genetic individual identification data. *Mol. Ecol.*, **2**, 377–379. Available at: http://www.pbil.univ-lyon1.fr/software/Gimlet/gimlet.htm.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: A software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes*, **4**, 535–538. Available at: http://www.microchecker.hull.ac.uk/index.jsp.

Walsh PS, Metzger DA, Higuchi R (1991) Chelex-100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques*, **10**, 506–513.

Woods JG, Paetkau D, Lewis D, McLellan BN, Proctor M, Strobeck C (1999) Genetic tagging of free-ranging black and brown bears. *Wildl. Soc. Bull.*, **27**, 616–627.

FOURTH PART

Genetic structure in a wolf population

CHAPTER V

A fine-scale genetic survey reveals low levels of interpack gene flow in a wolf population of the Italian Apennines

A fine-scale genetic survey reveals low levels of interpack gene flow in a wolf population of the Italian Apennines

Massimo SCANDURA[1], Laura IACOLINA[1], Claudia CAPITANI[2], Andrea GAZZOLA[1], Alessia VIVIANI[1], Luca MATTIOLI[3] and Marco APOLLONIO[1]

[1]Dipartimento di Zoologia e Genetica Evoluzionistica, Università di Sassari, Via Muroni 25, I-07100 Sassari, Italy.

[2]Joint Research Centre (JRC)- Institute for Environment and Sustainability (IES) - Rural, Water and Ecosystem Resources Unit- via E.Fermi, Ispra (VA)

[3]Servizio Piano Faunistico, Provincia di Arezzo, Piazza della Libertà 3, I-57100, Arezzo, Italy.

Manuscript

ABSTRACT

We investigated local gene flow in a high-density wolf (*Canis lupus*) population of the Italian Apennines, where no effective barrier to wolf dispersal were present. From 1998 to 2004 we examined wolf carcasses and non-invasively collected samples, focusing our effort on three mountain districts, separated by two main rivers, where wolf packs showed a high stability. With the use of nine autosomal microsatellites we successfully genotyped 177 samples providing with the identification of 74 wolves, which were genetically sexed. Resampling rate of non-invasively sampled individuals was low in the population, but three males were present for as long as 5-6 consecutive years. Genetic relatedness steeply decreased when it was calculated for wolves sampled in neighbouring mountain areas, suggesting that interpack migration at a limited distance scale is infrequent in the population. In addition, pedigree reconstruction in an intensively monitored pack over a four year period revealed that no individual born in this pack was sampled in other pack territories in the following years. The detected low gene flow resulted in a remarkable genetic structuring, confirmed by bayesian analysis. We associate these results to mortality patterns and a high long-distance dispersal rates. Overall, the present study suggests the existence of cryptic genetic structure in the Apennine wolf population and highlights the importance of studying social patterns and population dynamics to sustain wolf conservation in Italy.

Keywords: *Canis lupus*, wolf, autosomal microsatellites, genetic structure, migration, gene flow

INTRODUCTION

Due to their generally low Ne/N (i.e. effective population size / census size) ratio, wolf (*Canis lupus*) populations, are sensitive to the detrimental genetic implications that usually follow a bottleneck. Loss of genetic variability and high inbreeding have been observed in isolated bottlenecked populations (Liberg 2005).

However, recent studies proved that wolves prevent the risk of genetic deterioration and inbreeding depression by mechanisms of incest avoidance and social structuring varying in dependence of ecological constraints. On one hand, it was confirmed by several genetic investigations that wolves tend to mate with unrelated partners (Smith et al. 1997, Vonholdt et al. 2008). On the other hand, both dispersal patterns and social mechanisms (adoption of unfamiliar individuals, splitting, merging or budding of packs), can continuously modify the partition of individuals into social units, promoting the encounter of unrelated wolves (Lucchini et al. 2002, Grewal et al. 2004, Jedrzejewski et al. 2005, Vonholdt et al. 2008). In addition, fitness advantages by heterozygote individuals were shown to constrain the inbreeding level of isolated populations (Bensch et al. 2006).

Many recent studies have contradicted the previously accepted idea depicting a wolf pack as a social unit constituted by a breeding pair and their offspring of one or more litters (Mech 1970), revealing more complex structures that may result from a variety of dynamic processes (Mech & Boitani 2003). For instance, the occurrence of non-breeding adults that are not relatives of the breeding pair seems to be not a rare event in wolf packs, especially in exploited populations affected by high mortality rates (Grewal et al. 2004, Jedrzejewski et al. 2005). Indeed, mortality is one of the most important determinants of dynamism in the social structure of a population, especially if it affects breeders (Brainerd et al. 2008).

More controversial is the nature of dispersal in wolf populations, with respect to the distance travelled by wolves and to a possible bias between sexes. Dispersing wolves can cover more than 1000 kilometers (Wabakken et al. 2007), but the proportion of long-range dispersers becoming successful breeders seems to change a lot among areas (Kojola et al. 2006). Whether the attitude toward long-range dispersal is selected in wolves is debatable, as isolated populations with frequent short-range interpack dispersal are able to maintain their levels of heterozygosity and inbreeding over time (Vohnoldt et al. 2008).

No agreement exists on whether dispersal patterns in wolves are sex biased. Field data on radio-collared wolves show slight or no difference in dispersal attitudes between sexes

(Mech 1987; Gese and Mech 1991; Kojola et al. 2006), whereas male-biased dispersal is suggested by genetic investigations carried out in North America and Europe (Lehman et al. 1992, Seddon et al. 2005, Fabbri et al. 2007, Vonholdt et al. 2008).

In a saturated population, the constitution of a new pack is expected to be quite rare, while more frequent should be the recruitment of alien individuals by existing packs. The latter case may take place by the adoption of a non-breeding member or by the replacement of a breeder (after its death or by its usurpation). Accordingly, high mortality rates would promote turnover of individuals within packs and increase the frequency of splitting and merging processes. This scenario implies that the exchange of genes at a local scale would favour a genetic homogenization in the population.

In this study, we investigated the genetic structure of a wolf population in Italy at a local scale. The study population had rapidly recovered after a bottleneck (in the period 1960-1970), and at the beginning of the 2000s appeared to have been saturated by a number of close packs, occupying the main mountain areas (Apollonio et al. 2004; Capitani et al. 2006). No meaningful natural or artificial barrier to wolf dispersal exists in the area. Yet wolves are protected by the national law, high levels of induced mortality, mainly due to poaching and traffic accidents, are reported.

Under these conditions, migration of individuals at a limited geographic scale should be favoured and a high genetic homogeneity is expected to arise in the population. Therefore, we specifically tested the following predictions: *i*) individual movements to close pack territories are frequent as result of short-range dispersal, pack splitting and merging/adoption processes, *ii*) genetic relatedness is high between neighbouring packs and slowly decreases with distance, and *iii*) no substantial genetic structuring can be found in the population.

METHODS

Study area and wolf population - The study was conducted in a mountainous district, covering approximately 3000 km2, in the province of Arezzo, in central Italy. Major mountains, rarely exceeding 1500 m a.s.l., are protected by a national park (Foreste Casentinesi N.P.) and five natural reserves, and are mostly covered by forests. Villages are concentrated in lowlands and along the main valleys formed by the Arno and the Tevere rivers (Fig. 1). The area harbours an abundant and diverse community of wild ungulates, comprising wild boar *Sus scrofa* and roe deer *Capreolus capreolus*, which are ubiquitous, and red deer *Cervus elaphus*, fallow deer *Dama dama*, and mouflon *Ovis orientalis musimon*, whose presence is more limited. Here,

the recovery of the wolf population started in the 1980s. Wolves rapidly spread over the area, occupying the main ridges. In the 1990s, several packs had established their territories in the national park (Apollonio et al. 2004), where they fed mostly on wild ungulates (Mattioli et al. 1995; 2004). Starting from 1998, several packs were ascertained to have occupied all major mountains and some surrounding hilly zones in the province.

The local wolf population was monitored continuously from January 1998 to December 2004. No wolf was captured, tagged or radio-collared. Data on population dynamics and spatial distribution (approximate location of pack territories and homesites) were achieved by direct observations, wolf-howling and snow-tracking (see Gazzola et al. 2002, Apollonio et al. 2004, Capitani 2005). Nine to twelve different packs (on average 10.3 ± 1.2) were annually counted in the province of Arezzo. Pack size was between 2 and 8, averaging 4.0 ± 0.6 (mean ± sd) in winter. An overall winter density of 2.3 wolves per 100 km2 and an average distance of 11.1 ± 1.8 km between neighbouring pack rendezvous were estimated in the area (Capitani et al. 2006). High pack productivity and high fidelity to summer rendezvous sites were ascertained during the study period (Capitani 2005).

Sample collection and DNA extraction - Opportunistic sampling was conducted in the study area during the period 1998-2004. Sample collection was focused in three mountain districts (A to C), where wolf presence was stable at the beginning of the study, and in a few peripheral areas, where more recently wolf packs had settled their own territories (Fig. 1). Tissue, hairs, scats and blood residuals in snow were used as sources of genomic DNA. Muscle, skin or hair samples were taken from wolf carcasses discovered within the study area. Nine of them predated the beginning of the study, and had been collected by the staff of the provincial administration in the period 1991-1998. Fresh (< 1 week) faecal samples were collected in wolf areas, mostly along tracks in snow, and stored in polypropylene tubes filled with 96% ethanol. Shed hairs were stored dry in plastic or paper envelopes. Blood residuals were found along wolf tracks in snow and collected as described elsewhere (Scandura 2005). Whenever possible, all samples were kept cold immediately after their collection and then stored at –18 °C until analysis.

The Chelex-method (Walsh et al. 1991) was employed to isolate DNA from one to ten follicles per hair sample. The QIAamp Tissue and Stool kits (Qiagen) were used to extract DNA from tissues or blood and excremental samples, respectively. Suppliers' protocols were followed in all cases, except for melted blood/snow mixtures (see Scandura 2005). Precautions were taken to minimise the contamination risk of low-copy DNA samples. DNA isolation and amplification were performed in separate rooms. PCRs were prepared in a dedicated laminar

flow hood and the workspace sterilized before each preparation. Blank extractions were conducted to monitor possible contaminations. Negative (no DNA) and positive controls were included in each PCR tray.

Microsatellite genotyping and genotype reliability - Nine canine autosomal microsatellites, including five dinucleotides (cxx.109, cxx.123, cxx.204, cxx.250 from Ostrander et al. 1993, and cxx.377 from Ostrander et al. 1995), and four tetranucleotides (FH2004, FH2054, FH2137 and FH2175 from Francisco et al. 1996), were selected for wolf genotyping (Table 1). One primer in each pair was end-labelled with fluorescent dyes (Applied Biosystems). Details of the amplification protocols are reported in Scandura et al. (2005). The presence of amplification products was detected on 2% agarose gel. Successful PCRs were run on an ABI PRISM 3100-Avant automatic sequencer and allele lengths were subsequently determined using the GENEMAPPER v. 3.2 software (Applied Biosystems).

Genotyping errors are likely to occur in all dataset, especially in those including poor DNA samples, and, depending on their frequency, they can affect the outcome of the performed analyses and the consequent biological conclusions (Pompanon et al. 2005). A preliminary investigation on a partial wolf dataset revealed a limited error rate (up to 5.6% allelic dropout and up to 3.6% false alleles, depending on the sample class, Scandura et al. 2006). With the aim to prevent the occurrence of genotyping errors in the final dataset, a specific multi-step protocol was developed, also considering that both high- and low-quality DNA samples were used.

1) All samples were genotyped once (one positive PCR per locus) following the approach described in Scandura et al. (2006). A quality score (Q-score) was attributed to each single-locus genotype on the basis of the amplification profile. The corresponding expected error rate (ER) was then estimated using the regression function in Scandura et al. (2006) and the expected multilocus reliability was computed as $1-\Pi ERi$. Only genotypes showing an expected reliability higher than 0.5 were retained, while low-quality genotypes were discarded.

2) The retained (uncorrected) genotypes were used to perform the assignment test for species determination (see below). Indeed, as confirmed by simulated data (data not shown), a limited error rate, i.e. < 10%, is likely to do not affect significantly the outcome of this kind of analysis. Samples that were not confidently assigned to wolves were excluded form futher analyses.

3) The wolf (uncorrected) genotypes were compared eachother by the software GIMLET (Valière 2002) in order to identify groups of identical genotypes.

4) At this step, unique genotypes were evaluated for reliability. Each single genotype was considered reliable if at least one of the following cases occurred: i) it was obtained from a tissue sample, ii) two or more samples shared this genotype and the corresponding match probability for siblings (Woods et al. 1999) was <0.01, iii) it was confirmed by at least two independent replicates if heterozygote and four independent replicates if homozygote.

5) Due to the possible correction of some genotype during step 4, all reliable genotypes were run once again in GIMLET checking for correspondence and a new (corrected) set of unique genotypes was obtained.

6) Finally, we used the software MICRO-CHECKER (Van Oosterhout et al. 2004) to detect possible biases in the data, which might be due to trivial typographic errors, scoring errors or null alleles.

Species identification - Two other canids are sympatric with wolves in the study area: the domestic dog and the red fox (*Vulpes vulpes*). Consequently, species identification was necessary for all genotypes deriving from noninvasive samples. In order to discriminate among wolf, dog and fox, uncertain genotypes were compared to those of 30 wolves (carcasses recovered over 10 years within the study area), 20 dogs (including different races and mongrels), and 5 red foxes. Genetic discrimination is posssible due to allele differences at the typed microsatellites (non-overlapping allelic ranges are observed at some loci between wolf/dog and red fox). The statistical approach for the correct attribution of an uncertain genotype relied on a Bayesian clustering method, implemented in the software STRUCTURE version 2.1 (Pritchard et al. 2000). The program was run (100,000 iterations as burn-in period and 100,000 iterations for data collection) assuming three populations ($K = 3$). For each uncertain genotype a probability (q) was estimated to be assigned to each inferred cluster, on the basis of the allelic similarity with the reference samples. Since wolves and dogs can crossbreed in nature (Vilà and Wayne 1999), the admixture model was assumed. Therefore genotypes were either probabilistically assigned to one population or shared by two of them in case of admixed ancestry. In this calculation, genotypes were attributed to the local wolf population if the probability of assignment to the corresponding cluster was $q[\text{wolf}] \geq 0.95$.

Sex determination - Sex determination of non-invasively sampled individuals and non-sexed wolf carcasses was carried out by a canid-specific amplification of the DBX/DBY region, according to Seddon (2005). Conditions of the multiplexed PCR matched those for microsatellites, except for primer concentration (0.15 μM for DBX, 0.10 μM for DBY) and amplification profile (40 cycles of touchdown with annealing at 60-50 °C). PCR products were then electrophoresed on a 2% agarose gel containing ethidium bromide for visual detection.

Each set of reactions included a male control (Y-positive), a female control (Y-negative) and a blank (no template).

Statistical analysis - Genetic diversity in the population was evaluated using all available wolf genotypes and calculating mean number of alleles per locus (*k*), expected unbiased (H_e; Nei 1978) and observed heterozygosity (H_o). GENETIX v. 4.02 (Belkhir et al. 2001) was used to calculate the per-locus and overall FIS (Weir and Cockerham 1984) in the population and to test their significance over 10,000 permutations of alleles throughout the data set. Departures from Hardy-Weinberg equilibrium (HWE) and linkage equilibrium (LE) were tested using GENEPOP 4 (Raymond and Rousset 1995) and applying sequential Bonferroni corrections to account for multiple tests (Sokal and Rohlf 1995).

Due to the fact that: *i*) no pedigree information was available from other sources (e.g. sighting of marked animals), *ii*) sampling in the area was not exaustive (i.e. a low proportion of putative parents was sampled for most wolves) and *iii*) allele frequency estimates relying on dead animals might not be representative of the population (due to a possible bias in the proportion of immigrants or satellites), accurate parentage assignment could not be performed and we could not define the composition of social units yearly present in the area. Therefore, indirect methods were used to test our predictions.

First, we looked at resampling data with the aim to verify the occurrence of the same wolf in more than one pack territory in different years. Resampling rate was calculated and the distribution of distances among different locations of the same individual was evaluated.

Second, we used pairwise genetic relatedness to study how the signal of coancestry can change with distance. As estimator of genetic relatedness the Queller & Goodnight's (1989) pairwise relatedness *R* was calculated from genotype data in GENALEX v. 6 (Paekall and Smouse 2005).

This analysis was restricted to individuals sampled during the period 2000-2003 in three mountain areas separated by two main rivers (A, B and C in Fig. 1, schematic view in Fig. 2). Presence of one or two wolf packs was monitored They represent a good model, since in each area at least two adjacent packs were present during this period.

Field activities conducted every year in these areas allowed to monitor wolf presence and to locate packs' rendezvous sites. During census activity, every detected rendezvous was attributed to a hypothetical pack, whose size and reproductive status (presence/absence of a litter) was assessed from summer howl records (field techniques and methods are described in Capitani et al. 2006). According to pack locations provided by wolf-howling surveys, the

genotyped individuals were partitioned into 'putative' social units. The affiliation of each given individual to a social unit was then obtained by looking for the rendezvous site resulting the closest to the geographic centre of all his sampling locations in a year (up to a maximum of 6 km distance). A 6-km cutoff was adopted according to the average interpack distance observed in the study area (11.1 ± 1.8 km, Capitani et al. 2006). In so doing we assumed that pack territories were stable during the year, as suggested by the general consistency between wolf-howling (summer) and snow-tracking (winter) data.

Once individual affiliations to social units were established, pairwise relatedness values were computed for the following combinations (Fig. 2): 1) within the same unit (WU), 2) between units in the same area (BU-SA), 3) between units in neighbouring areas (BU-NA), 4) between units in distant areas (BU-DA). In order to avoid the confounding effect of temporal stratification in our data, this analysis was performed on an annual basis (i.e. only pairs of wolves that had been sampled in the study area during the same year were considered). A total of 21 packs-year were considered for this computation.

Differences between the means of the distributions obtained for each distance class were tested for significance by a randomization test (Manly 1991), using a Monte Carlo simulation to obtain by permutations a distribution of 1000 values of the difference between means, to which the real difference was compared.

Third, we studied the possibility that individuals had dispersed from a central and stable pack to the surrounding packs in the province. The Alpe di Catenaia (AC) pack (Fig. 1) was intensively monitored during the study and the position of its rendez-vous site was nearly the same in the years 2000-2003. Kinship was inferred, using the program CERVUS v. 3 (Marshall et al. 1998, Kalinowski et al. 2007), among all individuals that were found in this period at a distance < 6 km from the rendez-vous site. Pack composition derived from genetic data was corroborated by the known association of genotypes along winter trails in snow. Once determined the wolves representing the AC breeding pair, we looked among all genotypes sampled in the following years outside the AC territory which was compatible to be an offspring of that pair. This possibility was tested by mismatch analysis in CERVUS and on the basis of the relatedness coefficient *R*.

Fourth, the occurrence of migrants within the study area was studied by an assignment test, that was performed by a bayesian clustering analysis in the software STRUCTURE v. 2.1 (Pritchard et al 2000). The program estimates the log likelihood, Pr ($X|K$), i.e., the probability of the data associated to a certain number of subpopulations (K). Values of K comprised between 2 and 10 were tested, considering that 10 was the average number of packs detected

yearly in the area. Convergence was obtained by 100,000 replications following a burn-in period of 100,000 iterations (admixture model, alpha = 1, correlated allele frequencies). Five runs were performed for each value of *K* to verify consistency of results. The number of consistent genetic groups in our sample (i.e. the optimal value of *K*) was inferred from results of simulations, following the method used by Garner et al. (2004). The contribution of each of these inferred groups to single individuals sampled in one area was used to evaluate their status as "resident" or "immigrant". Resident wolves were animals that shared a high percentage (> 80%) of membership to the genetic cluster being the most frequent in the area where they were sampled. Immigrats were wolves that had > 80% membership to a cluster which was the most common elsewhere (i.e. the sampling area was other than the area of origin). This analysis was aimed at identifying possible migration events at a limited geographic scale, and therefore it was restricted to the three areas (A, B and C in Fig. 1) for which we had more data, using all genotypes collected in the period 2000-2003. Such an approach was reported to be a robust method to detect immigration, when populations not included in the sample could have genetically contributed to the composition of the individuals under study (Manel et al. 2002, Spencer & Hampton 2005).

RESULTS

Wolf genotyping and genetic diversity - Four-hundred-ninety-eight samples were processed, 30 of which corresponded to wolf carcasses recovered in the period 1991-2004. After culling non-invasive samples that failed to amplify or produced low-quality amplifications, a total of 177 multilocus genotypes was obtained. MICRO-CHECKER did not detect any bias in our dataset, that could be attributed to errors in allele scoring, allelic dropout or null alleles. The scored genotypes were attributed to 86 different individuals (Psib = 6.7×10-5-2.1×10-2). Among them, the bayesian cluster analysis revealed 4 foxes and 8 dogs, each with a confidence of assignement >95%. By contrary, all other 44 uncertain genotypes were attributed to wolves, but two of them showed signs of a possible event of hybridization with dogs in their ancestry (P[wolf] < 95%,).

Molecular gender determination allowed to establish that 36 individuals were males and 31 were females (for 7 the sex could not be assessed). Among the recovered carcasses, age was determined for 23 individuals, most of which resulted pups or yearlings (17/23 = 74%).

A total of 55 alleles was observed at the nine loci (3-9 per locus), while average Ho and He across loci equalled 0.647 and 0.660, respectively. No significant difference was detected

between heterozygosity values calculated for carcasses and noninvasive samples over loci (paired t-test, $p > 0.05$), suggesting that allelic dropout was not a concern in our dataset. The population deviated from HWE (Fisher's method, $p = 0.0021$), but this was not due to inbreeding ($FIS = 0.0199$, $p > 0.10$). Six pairs of loci (out of 36) showed a significant association at the genotypic level (linkage disequilibrium), but none of them was located on the same chromosome. Violation of both HWE and LE are likely to be consequence of population structuring, by virtue of the simultaneous occurrence of different rare alleles in different subpopulations (or social units).

Resampling - Each genotype deriving from non-invasive sampling was obtained on average from the analysis of 2.4 samples. However, one each two individuals was sampled just once (Tab. 1 and Fig. 3a). Referring to individuals that were sampled at least twice, approximately 80% of all resampling events occurred at a distance < 6 km from the first sampling site, without differences between sexes (Fig. 3b). With respect to the geographical partition in Fig. 1, just a single wolf (male W58) was sampled in more than one area, since he was found in area B during 2002 and in C during 2003 and 2004.

Only three individuals, all males, were sampled over a period longer than 3 years (W11, W18 and W19). Male W11 was non-invasively monitored over a period of six consecutive years (the entire study period), in the same area frequented for three consecutive years (2001, 2002 and 2003) by female W38 (see. Tab. 1). According to CERVUS analysis and to pairwise relatedness values, W11 and W38 were compatible with representing the parental pair of the AC pack durig this .

Genetic relatedness - As expected, relatedness decreased with distance in the population (Fig. 4). In intra-pack comparisons (WU, $n = 68$), relatedness averaged 0.310 ± 0.042 (mean ± standard error), but it decreased to 0.133 ± 0.034 between packs of the same area (BU-SA, $n = 74$). Pairwise relatedness dropped to average values of 0.011 ± 0.018 when individuals sampled in neighbouring areas were considered and it was negative (-0.124 ± 0.023, BU-NA, $n = 220$) for comparisons between distant areas. The observed differences in mean relatedness between classes were all significant (randomization test, $P < 0.01$ for all comparisons). Hence, the signal of relatedness was lost very early moving away from a pack.

Migration analysis and assignment test - Six individuals in our sample were likely to have been sired by the breeding pair W11-W38 (Fig. 5). Five of them had been sampled, between 2000 and 2003, within a radius of 6 km from the AC pack rendezvous site, while the other one (W48) was a yearling that was found dead in December 2002 at 13 km distance. This wolf had

been shot and was found at the side of a provincial road, thus leading to argue that a post-mortem translocation had occurred. In addition his finding place was not included in any wolf territory. No other wolf sampled after year 2000 in the study area was consistent with having originated from the AC pack.

The bayesian assignment test allowed to evaluate possible migrations of single individuals in the population. The partition with $K = 3$ received the highest support from simulations (Tab. 2), and the contribution of the three inferred groups to the genetic composition of wolves sampled in the three areas is shown in Fig. 5. Individuals in aerea A had on average 88% membership to cluster I, wolves in area B showed 75% membership to cluster III and those in area C had 64% membership to cluster II.

A total of 32 wolves out of 47 (68%) were classified as resident, i.e. they were sampled while staying in the area of their most likely origin. Four wolves turned out to be immigrants and all of them seemingly had dispersed from area A. Two of them were sampled in area B and two in area C. Three immigrants were females. According to the previous result, no wolves from area B was found in the neighbouring areas.

DISCUSSION

In the present study, we considered an area central to the present wolf's range in the Italian peninsula, where the species fully recovered after its historical bottleneck that occurred in the '60-70s (Cagnolaro et al. 1974). By 1998 the local population was monitored and a high pack density was detected in mountain areas, where suitable territories appeared saturated by packs (Apollonio & Mattioli 2006). In fact, no remarkable rearrangement of pack territories was recorded during the monitoring period, inter-pack distances were roughly constant, and a high fidelity to summer homesites was noticed (Capitani 2005, Capitani et al. 2006). The present study provides evidence of limited gene flow in these areas. Accordingly, in absence of radio-tracking data on wolf movements in Italy, genetic data suggest the existence of a cryptic genetic structure at a small geographic scale.

Temporal and spatial patterns, inferred by non-invasive resampling of genotyped individuals in the population, suggest that immigration of individuals into neighbouring packs is infrequent. Once genotyped for the first time, an individual had a high chance (approximately 80%), if resampled, to be found in the same area (i.e. pack territory). But half the wolves have not been found again after their first sampling event. This low recapture rate can be associated to both high mortality and high dispersal rate, provided that most mortality

events are usually undetected and wolves leaving the natal pack may easily emigrate out of the study area.

In agreement with resampling data, no evidence of short-term dispersal from the AC pack was detected during the period 2000-2003. During this period, the rendezvous site of this pack was constantly detected in the same small valley (Capitani et al. 2006) and a pair of individuals (male W11 and female W38) was repeatedly sampled in the pack range. Parentage analysis confirmed them to represent the breeding pair of the pack. Although six wolves in our sample were likely offspring of this pair, each of them was found during the course of one single year and all but one were sampled in the AC range only. The exception is represented by W48, who was killed at the periphery of the AC pack territory, but we suspect that it had been killed in the AC area and translocated after his death. Therefore, despite of its position and apparent stability, a low short-range migration rate can be deduced for the members of this pack.

As expected, genetic relatedness decreased with distance. Although two rivers separated areas A, B and C (Figg. 1 and 2), they cannot represent effective barriers to wolf dispersal. Nevertheless, while wolves of different packs in the same area showed a moderate degree of relatedness, mean *R* dropped immediately to zero when it was calculated across neighbouring areas.

The level of genetic differentiation among the three areas that comes out from bayesian analysis confirms the existence of constraints to local gene flow, with the two main valleys representing areas of genetic discontinuity. Nonetheless four potential cases of recent migration among areas were detected in the population. In all cases area A acted as source area.

On the basis of previous data obtained for North America (Vonholdt et al. 2008), we predicted that a population with the characteristics observed in our study area in the Apennines (i.e. clumped and quite stable pack territories) should show high levels of genetic homogeneity, as consequence of frequent interpack migration. This prediction was not confirmed by our study.

In Yellowstone, a number of different mechanisms like pack fission, pack merging and adoption of dispersers promotes breeding between unrelated adult wolves, allowing to retain genetic variability over time (Vonholdt et al. 2008). However, such social dynamics accounting for inbreeding avoidance were observed in an isolated recovered wolf population (i.e. Yellowstone), where the risks associated with dispersing to reach the closest populations is

very high because of the absence of suitable habitat corridors. In Italian Apennines, on the contrary, wolves can find a continuum of suitable habitats and natural corridors that can render long-range dispersal an effetcive strategy. The most striking evidence in favour of this hypothesis is the natural recolonization of the western Alps (Valière et al. 2003; Fabbri et al. 2007).

The maintenance of a remarkable degree of structuring in the population can be consequence of pack stability. Despite mortality rate is not negligible in the area, it seems to be mostly in charge of young individuals (Capitani 2005). If the loss of a breeder is not frequent in the population, breeding pairs can reproduce for several years and, if undisturbed, they can retain the same homesites and the same territories from year to year, as suggested by the analysis of summer locations of pack choral responses to acoustic stimulations (Capitani et al. 2006). This is what we actually observed in the AC pack.

Fine-scale genetic data agree with evidences obtained at a higher spatial scale in Italy. The limited short-range interpack dispersal observed by us in a mountain region in Tuscany reinforces the idea that an opposite trend occurs in the Italian population. In addition to some occasional reports (e.g. Boitani 1992; Ciucci et al. 2005), indeed, genetic data obtained by Fabbri et al. (2007) for the Italian population revealed that multiple events of long-range dispersal could explain the genetic diversity observed in the recently recolonized Alpine population. Actually, long range dispersal is common in wolves and is particularly frequent among young age classes, which can disperse over hundreds of kilometers (Mech & Boitani 2003). Dispersal distances may be very high particularly in expanding populations, like observed in France (Valière et al. 2003) and in Sweden (Vilà et al. 2002; Wabakken et al. 2007). In Finland, both radiotelemetry and genetic data revealed that long-distance migrants represent a high proportion of dispersing individuals, being the average dispersal distance was close to 100 km (Aspi et al. 2006, Kojola et al. 2006).

With respect to this hypothesis, mountain areas like that monitored by us in the Arezzo province are likely to represent sources supplying with migrants to the recolonization of peripheral sink areas. This interpretation can explain the ongoing expansion pattern of wolves in central Italy, where wolves are colonizing peripheral hilly areas and lowlands, after having fully recolonized mountain areas.

Overall, these results suggest the existence of a cryptic genetic structure in the Apennine wolf population, which can result from a combination of demographic and dispersal patterns, like stability of breeding pairs and the tendency to replace a breeder loss with individuals coming from far away (i.e. long-range dispersers).

Moreover, the present study suggests that the reconstruction of genealogies within a population can provide valuable insight into the dynamics influencing both genetic an social patterns in this species. The knowledge of these patterns, if supported by future studies, is of paramount importance for the conservation of this carnivore in Italy.

ACKNOWLEDGEMENTS

We thank M. Alboni, E. Avanzinelli, F. Benvenuti, M. Geri, D. Giustini, S. Marsili, A. Vanni and all the people who helped us in collecting samples. We also thank S. Cappelli and D. Del Chiaro for their contribution to laboratory activities, and L. Mauri for bioacoustic analysis of recorded howls. We are grateful to Fritz Trillmich for his helpful suggestions. The wolf monitoring project was funded by the Italian Ministry for the University and Research (PRIN 2003), the Regional Government of Tuscany, and the Provincial Administration of Arezzo. We are grateful to the Administration ex-ASFD of Pratovecchio for their logistic support. L. Iacolina benefited by a fellowship provided by the Sardegna Resorts srl.

TABLE 1. – Sampled genotypes through years, identified by individual ID, sex and number of sampling events (N). Shadows define the minimum period of presence of the wolf in the area. Numbers in circles indicate the number of sampling events in a single month. Death of monitored individuals is indicated by "X" and corresponds to the finding of its carcass. For these individuals, additional information concerning cause of death and age are reported. Geographic areas where wolves were sampled refer to those in Fig. 1.

Genotype	Sex	N	1998	1999	2000	2001	2002	2003	2004	Area	Info
W18	M	6		❶		❷❷		❶		A	
W23	M	1			❶X					A	injured by vehicle collision and recovered by CFS
W29	F	1			❶					A	
W20	-	1		❶						A	
W32	M	2				❶		❶		A	
W33	F	8				❶❶	❶❷❷	❶		A	
W34	M	3				❷❶				A	
W35	F	2				❶❶				A	
W40	M	1					❶X			A	cub (3-4 months), driven
W42	M	1					❶			A	
W43	-	1					❶			A	
W44	M	5					❶❸	❶		A	
W10	M	1	❶X							B	2-3 years old, killed by snare
W11	M	10	❶	❷	❷	❷❶❶		❶		B	
W21	F	3		❷	❶					B	
W27	M	1			❶					B	
W26	F	1			❶X					B	> 1 year old, shot
W37	F	1				❶				B	
W38	F	5				❷	❷	❶		B	
W39	F	1				❶				B	
W45	M	15					❺❾❶			B	
W46	M	7					❹❷❶			B	
W50	F	3					❷❶			B	
W48	M	1					❶X			B	yearling (1-2 years), shot
W60	F	1						❶		B	
W66	-	1						❶		B	
W74	F	1							❶X	B	cub (8 months), shot
W12	F	1	❶X							B	3-4 years old, black fur, shot
W25	M	1			❶X					B	> 1 year old, shot
W53	M	1					❶			B	
W54	F	2					❶❶			B	
W55	M	3					❷❶			B	
W65	F	1						❶X		B	cub (6-7 months), driven
W71	-	1						❶		B	
W73	M	1						❶X		B	
W31	M	7			❶		❺❶			C	
W14	M	1	❶							C	
W15	-	1	❶							C	
W19	M	7		❶	❷	❶	❶	❶❶		C	
W47	F	1					❶			C	
W28	F	3			❶	❷				C	
W30	F	2			❶	❶X				C	cub (< 1 year), shot
W36	F	2				❶❶X				C	cub (< 1 year), poisoned
W41	M	1					❶X			C	cub (5-6 months), driven
W49	F	1					❶X			C	cub (7-8 months), unknown cause of death
W51	F	5					❷❶	❷		C	
W52	M	1					❶X			C	cub (7-8 months), unknown cause of death
W56	F	1					❶			C	
W58	M	8					❶	❶❶	❺	C, B	
W57	M	3					❶	❷		C	
W70	M	1						❶		C	
W68	M	1						❶		C	
W69	F	1						❶		C	
W72	M	1							❶X	C	driven?
W13	F	1	❶X								yearling (1-2 years), shot
W16	M	1		❶							
W17	F	1		❶							
W22	M	1			❶X						cub (< 1 year), driven?
W24	-	1			❶						
W67	M	1						❶X			cub (< 1 year), unknown cause of death
W61	M	1						❶X			cub (5-6 months), shot
W59	M	1						❶X			> 1 year old, driven?
W62	F	1						❶			
W63	M	1						❶X			> 5 years old, shot
W64	F	2						❶❶			

TABLE 2. Mean likelihood obtained with the software STRUCTURE (Pritchard et al. 2000) for simulations with different values of *K* (i.e. number of inferred homogeneous clusters) and mean values of $\Delta\ln P(D) = \ln P(D)_{k+1} - \ln P(D)_{k}$.

	lnP(D)		**ΔlnP(D)**	
K	**Mean**	**SD**	**Mean**	**SD**
1	-1091.3	1.6		
2	-1041.8	7.4	49.5	7.4
3	-958.1	10.9	83.7	14.3
4	-902.7	5.5	55.4	8.7
5	-930.2	9.4	-27.4	13.1
6	-947.1	17.0	-17.0	21.2
7	-966.1	50.3	-26.1	64.5
8	-959.0	12.1	14.3	59.9
9	-980.4	23.4	-21.4	23.4
10	-988.3	57.5	-7.9	57.2

FIG. 1. – Simplified map showing the distribution of typed samples (filled squares and stars) in the province of Arezzo (Italy). Circles represent approximate pack locations (6-km buffers around summer pack locations of 2002-2003). Protected areas are in dark grey. At the bottom right: the partition into three areas (A, B, and C) separated by the two main rivers.

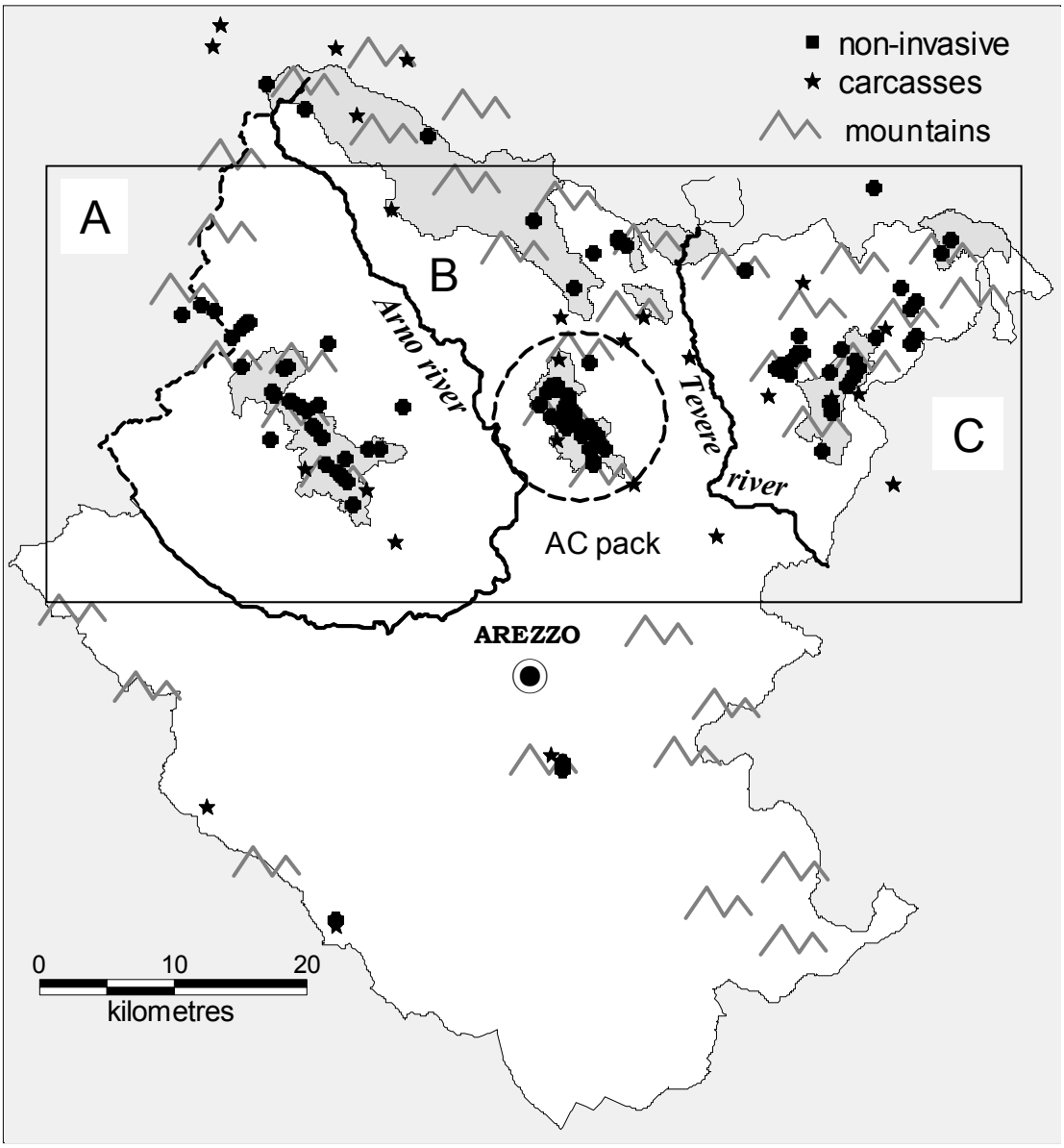


Fig. 2. – Scheme reproducing the situation within the rectangle in Fig. 1, which the relatedness analysis refers to. Four classes of distance used for comparisons among sampled wolves are shown: within social units (WU), between units in the same area (BU-SA), between units in neighbouring areas (BU-NA)and between distant ares (BUDA). Rounded squares represent units, separeted into the tree areas (A, B and C) by the two main rivers (Arno and Tevere).

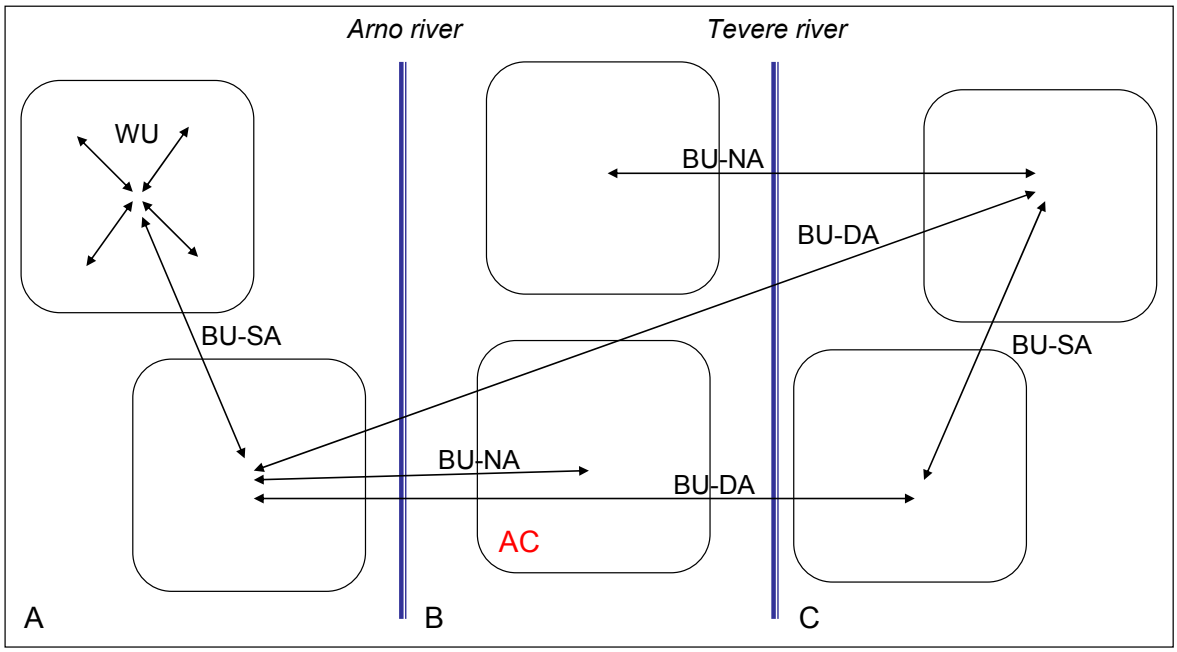
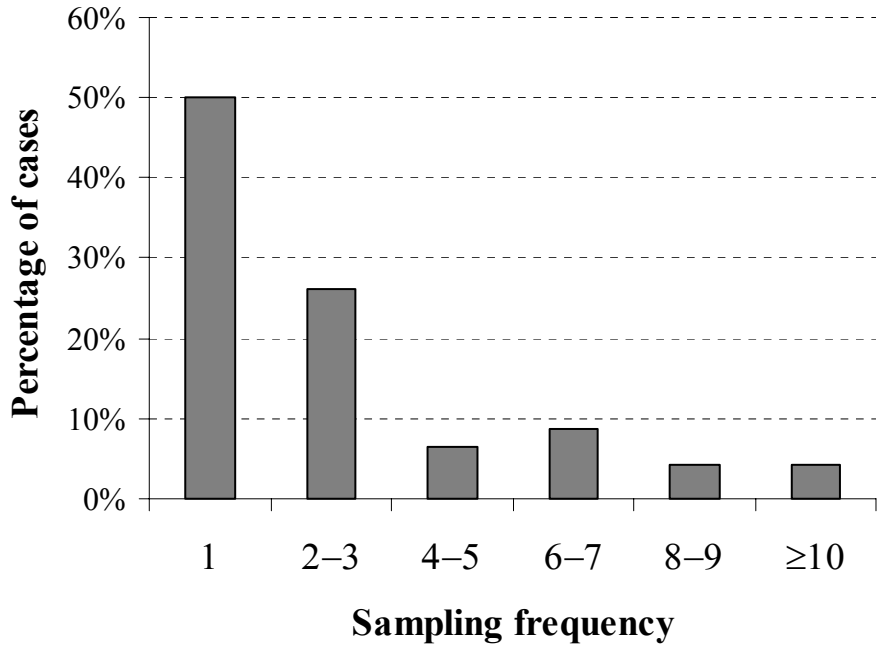


FIG. 3. – Resampling of individual wolves in the study area. *a)* Percentage distribution of the sampling frequency. *b*) Percentage distribution of classes of resampling distance (data divided by sex are shown).



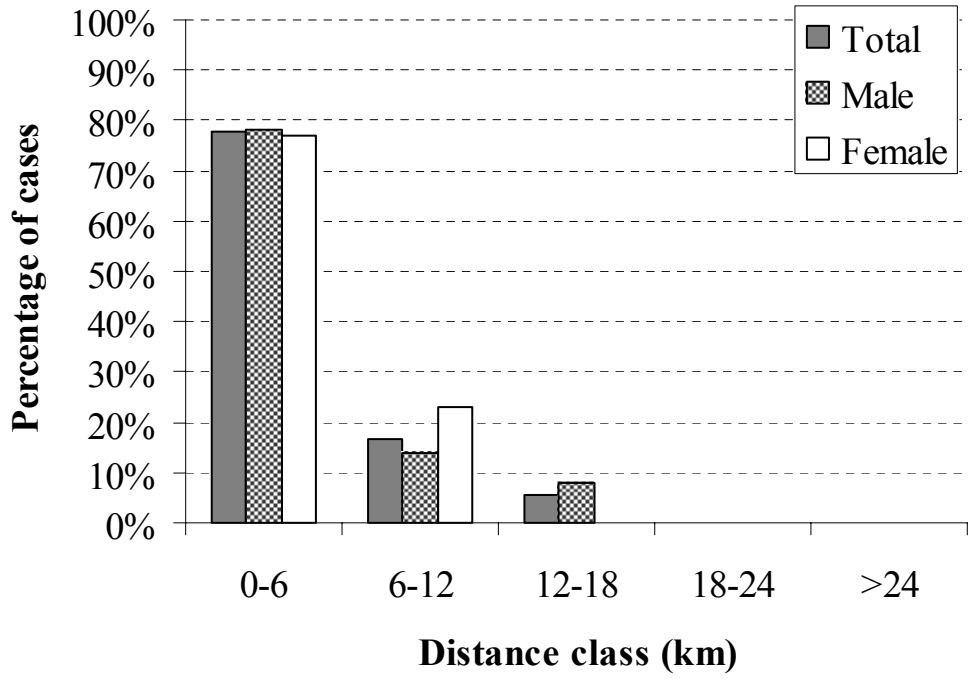


FIG. 4. – Variation in the pairwise genetic relatedness between wolves belonging to different distance classes (see Fig. 2). Significance of the difference between mean values were obtained by randomization tests.

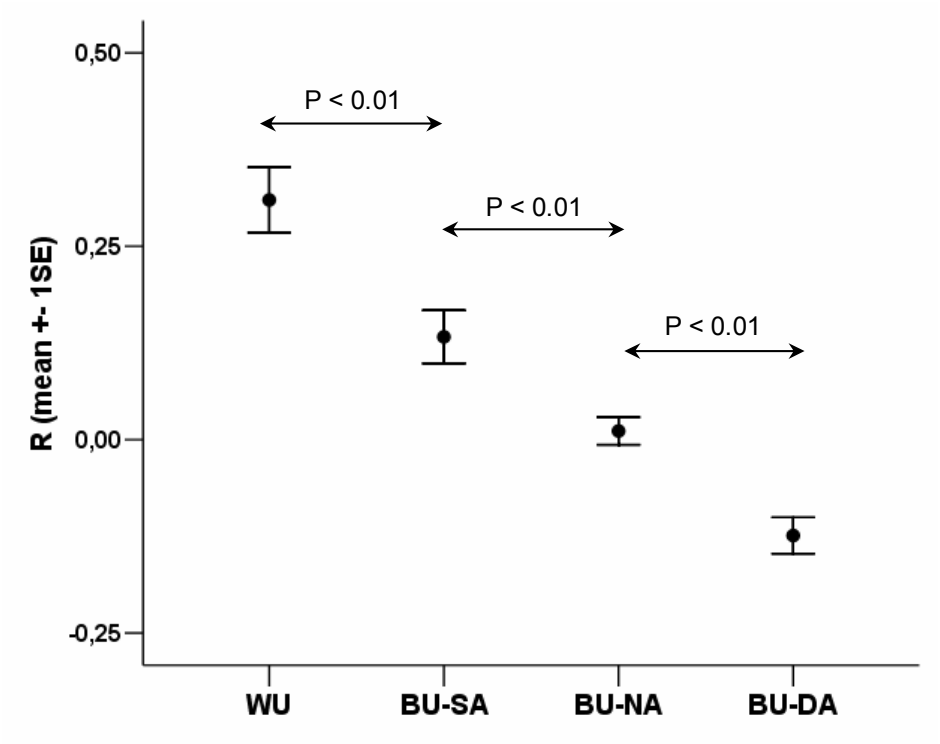
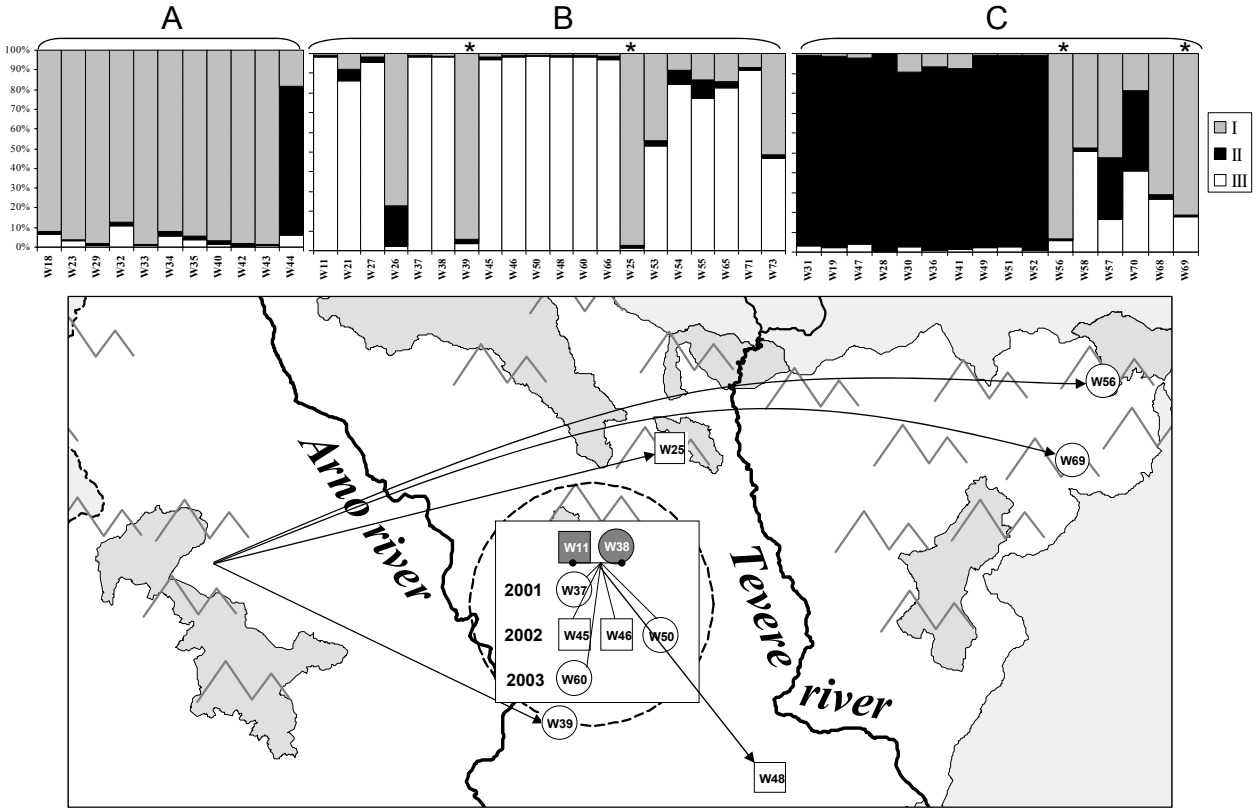


FIG. 5. – Proportions of membership to the three clusters resulting from the cluster analysis in Structure are reported for wolves sampled in areas A, B and C in the period 2000-2003. The asterisk identify the genotypes identified as migrants, and represented in the map below. The box included in the map reports the reconstructed genealogy (only sampled individuals) in the AC pack and the wolf W48, born in this pack and recovered dead at the periphery of the pack territory.



LITTERATURE CITED

Apollonio M, Mattioli L, Scandura M, et al. 2004 Wolves in the Casentinesi Forests: insight for wolf conservation in Italy from a protected area with a rich wild prey community. Biological Conservation 120: 249-260.

Apollonio M, Mattioli L (eds) 2006 Il lupo in provincia di Arezzo. Editrice Le Balze, Montepulciano, Italy. [in Italian]

Aspi J, Roininen E, Ruokonen M, Kojola I, Vila C 2006 Genetic diversity, population structure, effective population size and demographic history of the Finnish wolf population. Molecular Ecology 15: 1561-1576.

Belkhir K, Borsa P, Chikhi L, Raufauste N, Bonhomme F 2001 GENETIX, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions CNRS UMR 5000, Université de Montpellier II, Montpellier, France

Bensch S, Andrén H, Hansson B, et al. 2006 Selection for heterozygosity gives hope to a wild population of inbred wolves. PLoS One 72: 1–7.

Boitani L 1992 Wolf research and conservation in Italy. Biological Conservation 61: 125-132.

Brainerd SM, Andrén H, Bangs EE, et al. 2008 The effects of breeder loss on wolves. Journal of Wildlife Management 72 (1): 89-98.

Cagnolaro L, Rosso D, Spagnesi M, Venturi B 1974 Investigation on the wolf (*Canis lupus*) distribution in Italy, in Canton Ticino and Canton Grigioni (Switzerland). Ricerche di Biologia della Selvaggina 59: 1-75 [in Italian].

Capitani C 2005 Dynamics of a wolf (*Canis lupus*) population in Eastern-Central Apennines. PhD Dissertation, University of Sassari, Italy.

Capitani C, Mattioli L, Avanzinelli E, Gazzola A, Lamberti P, Mauri L, Scandura M, Viviani A Apollonio M 2006 Selection of rendezvous sites and reuse of pup raising areas among wolves *Canis lupus* of north-eastern Apennines, Italy. Acta Theriologica 51: 395-404.

Ciucci P, Maiorano L, Andreani M, Reggioni W, Boitani L 2005 Dispersione a lungo raggio di un lupo dall'Appennino Settentrionale alle Alpi Marittime: movimenti, comportamento spaziale ed eterogeneità ambientale. Hystrix, Italian Journal of Mammalogy (n.s.) supp. 2005, p. 65

Fabbri E, Miquel C, Lucchini V, et al. 2007 From the Apennines to the Alps: colonization genetics of the naturally expanding Italian wolf (*Canis lupus*) population. Molecular Ecology 16: 1661-1671.

Francisco LV, Langston AA, Mellersh CS, Neal CL, Ostrander EA 1996 A class of highly polymorphic tetranucleotide repeats fon canine genetic mapping. Mammalian Genome 7: 359-362.

Garner TWJ, Pearman PB, Angelone S 2004 Genetic diversity across a vertebrate species' range: a test of the central-peripheral hypothesis. Molecular Ecology 13: 1047-1053.

Gazzola A, Avanzinelli E, Mauri L, Scandura M, Apollonio M 2002 Temporal changes of howling in south European wolf packs. Italian Journal of Zoology 69: 157-161.

Gese EM, Mech LD 1991 Dispersal of wolves (*Canis lupus*) in northeastern Minnesota, 1969-1989. Canadian Journal of Zoology 69: 2946-2955.

Grewal SK, Wilson PJ, Kung TK, et al. 2004 A genetic assessment of the eastern wolf (*Canis lycaon*) in Algonquin provincial park. Journal of Mammalogy 85: 625-632.

Jedrzejewski W, Bracnicki W, Veit C*,* et al. 2005 Genetic diversity and relatedness within packs in an intensively hunted population of wolves *Canis lupus*. Acta Theriologica 50: 3-22.

Kalinowski ST, Taper ML, Marshall TC 2007 Revising how the computer program cervus accommodates genotyping error increases success in paternity assignment. Molecular Ecology 16: 1099–1006.

Kojola I, Aspi J, Hakala A, et al. 2006 Dispersal in an expanding wolf population in Finland. Journal of Mammalogy 87: 281-286.

Lehman N, Clarkson P, Mech LD, Meier TJ, Wayne RK 1992 A study of the genetic relationships within and among wolf packs using DNA fingerprinting and mitochondrial DNA. Behavioral Ecology and Sociobiology 30: 83-94.

Liberg O, Andrén H, Pedersen H-C, et al. 2005 Severe inbreeding depression in a wild wolf (*Canis lupus*) population. Biology Letters 1: 17-20.

Lucchini V, Fabbri E, Marucco F*,* et al. 2002 Noninvasive molecular tracking of colonizing wolf (*Canis lupus*) packs in the western Italian Alps. Molecular Ecology 11: 857-868.

Manel S, Berthier P, Luikart G 2002 Detecting wildlife poaching: identifying the origin of individuals with bayesian assignment tests and multilocus genotypes. Conservation Biology 16: 650-659.

Manly BFJ 1991 Randomization and Monte Carlo Methods in biology. Chapman and Hall, London, Example 4.1 (page 51).

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

Mattioli L, Apollonio M, Mazzarone V, Centofanti E 1995 Wolf food habits and wild ungulate availability in the Foreste Casentinesi National Park, Italy. Acta Theriologica 40: 387-402.

Mattioli L, Capitani C, Avanzinelli E, et al. 2004 Predation by wolves (*Canis lupus*) on roe deer (*Capreolus capreolus*) in north-eastern Apennine, Italy. Journal of Zoology, London 264: 249-258.

Mech DL 1970 The Wolf. New York: Natural History Press.

Mech LD 1987 Age, season, distance, direction, and social aspects of wolf dispersal from a Minnesota pack. In: Chepko-Sade BD, Tang Halpin Z (eds.) Mammalian Dispersal Patterns. The University of Chicago Press, Chicago, USA pp. 55-74.

Mech LD, Boitani L 2003 Wolf social ecology. In: Mech LD, Boitani L (eds.) Wolves: Behavior, Ecology, and Conservation. The University of Chicago Press, Chicago, USA, pp. 1-34

Nei M 1978 Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89: 583-590.

Ostrander EA, Sprague GF, Rine J 1993 Identification and characterization of dinucleotide repeat $(CA)_n$ markers for genetic mapping in dog. Genomics 16: 207-213.

Ostrander EA, Mapa FA, Yee M, Rine J 1995 One hundred and one new simple sequence repeat-based markers for the canine genome. Mamm Genome 6: 192-195.

Peakall R, Smouse PE 2005 GenAlEx 6: Genetic Analysis in Excel. Population genetic software for teaching and research. Australian National University, Canberra, Australia.

Pompanon F, Bonin A, Bellemain E, Taberlet P 2005 Genotyping errors: causes, consequences and solutions. Natural Rev Genet 6: 847-846.

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

Queller D C, Goodnight KF 1989 Estimating relatedness using genetic markers. Evolution 43: 258-275.

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

Scandura M 2005 Individual sexing and genotyping from blood spots on the snow: a reliable source of DNA for non-invasive genetic surveys. Conservation Genetics 6(5): 871-874.

Scandura M, Capitani C, Iacolina L, Apollonio M 2006 An empirical approach for reliable microsatellite genotyping of wolf DNA from multiple noninvasive sources. Conservation Genetics 7: 813-823.

Seddon JM 2005 Canid-specific primers for molecular sexing using tissue or non-invasive samples. Conservation Genetics 6: 147-149.

Seddon JM, Parker HG, Ostrander EA, Ellegren H 2005 SNPs in ecological and conservation studies: a test in the Scandinavian wolf population. Molecular Ecology 14: 503-511.

Smith D, Meier T, Geffen E, et al. 1997 Is incest common in gray wolf packs? Behavioral Ecology 8: 384-391.

Sokal R, Rohlf FJ 1995 Biometry: the principles and practise of statistics in biological research. Freeeman, New York.

Spencer PBS, Hampton JO 2005 Illegal translocation and genetic structure of feral pigs in Western Australia. Journal of Wildlife Management 69: 377-384.

Valière N 2002 gimlet: a computer program for analysing genetic individual identification data. Molecular Ecology 2: 377–379.

Valière N, Fumagalli L, Gielly L, et al. 2003 Long-distance wolf recolonization of France and Switzerland inferred from non-invasive genetic sampling over a period of 10 years. Animal Conservation 6: 83-92.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P 2004 Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes 4: 535-538.

Vilà C, Wayne RK 1999 Hybridization between wolves and dogs. Conservation Biology 13: 195-198.

Vilà C, Sundqvist AK, Flagstad Ø, et al. 2002 Rescue of a severely bottlenecked wolf (*Canis lupus*) population by a single immigrant. Proceedings of the Royal Society London B 270: 91-97.

Vonholdt BM, Stahler DR, Smith DW, et al. 2008 The genealogy and genetic viability of reintroduced Yellowstone grey wolves. Molecular Ecology 17: 252-274.

Wabakken P, Sand H, Kojola I, Zimmermann B, Arnemo JM, Pedersen HC, Liberg O 2007 Multistage, long-range natal dispersal by Global Positioning System-collared Scandinavian wolf. Journal of Wildlife Management 71: 1631-1634.

Walsh PS, Metzger DA, Higuchi R 1991 Chelex-100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. BioTechniques 10: 506-513

Weir BS, Cockerham CC 1984 Estimating F-statistic for the analysis of population structure. Evolution 38: 1358-1370.

Woods JG, Paetkau D, Lewis D, McLellan BN, Proctor M, Strobeck C 1999 Genetic tagging of free-ranging black and brown bears. Wildlife Society Bulletin 27: 616-627.

CHAPTER VI

Variability at Y-chromosome microsatellites within a free ranging wolf population in central Italy

Variability at Y-chromosome microsatellites within a free ranging wolf population in central Italy

Massimo SCANDURA[1], Laura IACOLINA[1], Claudia CAPITANI[2], Andrea GAZZOLA[1], Luca MATTIOLI[3] and Marco APOLLONIO[1]

[1]Dipartimento di Zoologia e Genetica Evoluzionistica, Università di Sassari, Via Muroni 25, I-07100 Sassari, Italy.

[2]Joint Research Centre (JRC)- Institute for Environment and Sustainability (IES) - Rural, Water and Ecosystem Resources Unit- via E.Fermi, Ispra (VA)

[3]Servizio Piano Faunistico, Provincia di Arezzo, Piazza della Libertà 3, I-57100, Arezzo, Italy.

Manuscript

ABSTRACT

We investigated the genetic variability within the male lineage in a free ranging wolf (*Canis lupus*) population in central Italy. From 1998 to 2004 we examined wolf carcasses and non-invasively collected samples such as scats, hairs and diluted blood spots. Thirty-one samples, resulting to be male according to autopsy or molecular sexing, were amplified at four Y-chromosome microsatellite loci. Two markers were polymorphic and identified four different haplotypes. The detected variation is remarkable, if we consider the limited scale of the monitored area and especially if compared with the absence of variation in the maternally-inherited markers (i.e. mitochondrial DNA). Spatio-temporal distribution of the four haplotypes in the area suggested an influence of the recolonization pattern, which is likely based on long-range dispersal of individuals. We also specifically investigated the patriline within a pack over a four year periods identifying the replacement of the breeding male by a new unrelated individual, confirming the trend already observed in this species.

Keywords: *Canis lupus*, wolf, Y-chromosome, microsatellites, genetic variability, gene flow.

INTRODUCTION

The conservation of wolf (*Canis lupus*) natural populations represents a priority in several European countries, where the species is endangered or was, in the recent past, severely threatened (Promberger and Schröder 1993). The Italian wolf population suffered a strong persecution until 1971, when wolf hunting was forbidden and poison baits banned. This change in attitude was completed in 1976 when the species obtained the fully protected status. However, during the period 1950-1970 the number of wolves throughout the peninsula was very low (Cagnolaro et al. 1974) and their presence was extremely restricted to a few areas of the Apennine mountains. From then on, the wolf population recovered in Italy and is still growing in size and recolonizing its historical range.

As consequence of its history, the Italian wolf shows in the mitochondrial line the effect of a severe bottleneck, which, in association with its prolonged isolation, led to genetic erosion (Lucchini et al. 2004). A single mitochondrial haplotype was found in the whole Italian wolf population (Wayne et al. 1992, Vilà et al. 1997, Vila et al. 1999, Randi et al. 2000). On the contrary, nuclear markers revealed a less severe reduction in genetic diversity, with remarkable reduction in heterozygosity only in the recently re-established Alpine wolf population, while the Apennines population showed only a slightly lower variability with respect to other European populations (Scandura et al. 2001, Lucchini et al. 2004, Fabbri et al. 2007).

Actual knowledge on the Italian wolf population is based on matrilineal (mitochondrial) and bi-parental (autosomic) markers but nothing is known about paternal inherited markers, possibly biasing the picture of population structure and history. Males and females do not play symmetrical roles in structuring populations, as strong asymmetries between sexes may affect dispersal and breeding strategies (Greenwood 1980). Y chromosome markers have been widely used in humans and other primates (Tosi et al. 2000, Hammer et al. 2003, Eriksson et al. 2006, Douadi et al. 2007, Kawamoto et al. 2008). The use of paternal inherited markers in mammals has increased in the past ten years, and it is mainly related to studies on recolonization, sex-biased dispersal rates and phylogeography (Van Hooft et al. 2002, Meadows et al. 2006, Yannic et al. 2008).

In wolf those kind of markers was recentently used to investigate the recolonization patterns that led to the population recovery in Scandinavia, giving an idea of the contribution of each sex in the recolonization event, as well as of the gene flow among adiacent countries (Sundqvist et al. 2001). Combining the use of differently inherited markers, former studies have proved the presence of sex-bias in gene flow in the North American populations.

Musiani et al. (2007) reconstructed the relationships between packs with different behaviours (migratory vs. non-migratory) and natal habitat (tundra/taiga vs. boreal coniferous forest). Vilà et al. (2003) showed the hybridization patterns between a female wolf and a male dog in Scandinavia. The combination of paternal, maternal and bi-parentaly inherited markers has also been used to reconstruct population structure and kin relationships within and among packs in a North American population (Grewal et al. 2004).

In the present study, we focused on the fine genetic structure within a free-living wolf population in Italy, by integrating previous biparental data with Y-linked markers. Specifically, we investigated the spatial discontinuity in the distribution of genetic variation in the male lineage, checking whether gene flow is somehow limited over short distances and even in absence of physical barriers. Furthermore, we reconstructed the patriline in a wolf pack over a four-year period.

METHODS

Study area and wolf population - The study was conducted in the province of Arezzo (central Italy) in the mountainous district that covers approximately 3000 km^2. Major mountains, rarely exceeding 1500 m a.s.l., are protected by a national park (Foreste Casentinesi N.P.) and five natural reserves. Villages are concentrated in lowlands and along the main valleys formed by the Arno and the Tevere rivers. Within the study area the community of wild ungulates is abundant and diverse, comprising wild boar (*Sus scrofa*), roe deer (*Capreolus capreolus*), red deer (*Cervus elaphus*), fallow deer (*Dama dama*), and mouflon (*Ovis orientalis musimon*). In the province of Arezzo the wolf population started to recover in the 1980s. Wolves rapidly spread over the area, occupying the main ridges. About ten years later, several packs had established their territories in the national park (Apollonio et al. 2004), where they fed mostly on wild ungulates (Mattioli et al. 1995; 2004). At present, wolf presence covers all major mountains and protected areas.

The local wolf population was monitored from 1998 to 2004. Data on population dynamics and spatial distribution (approximate location of pack territories and homesites) were estimated by direct observations, wolf-howling and snow-tracking (for details on the field methodology see Gazzola et al. 2002, Apollonio et al. 2004, Capitani 2005). Packs were annually counted (range 9-12, average 10.3 ± 1.2) within the province of Arezzo territory, with pack size comprised between 2 and 8, averaging 4.0 ± 0.6 (mean ± sd) in winter. An overall winter density of 2.3 wolves per 100 km^2 was estimated in the area.

The average distance between adjacent packs, calculated as the mean of all pairwise distances between rendezvous sites of contiguous packs, amounted to 11.1 ± 1.8 km during the study (see Capitani et al. 2006). High productivity and high fidelity to summer rendezvous sites were ascertained in local packs in the study area (Capitani 2005). In 2003-2004 two additional packs were detected for the first time, expanding the area of wolf presence, and they both produced a litter in summer 2003.

Sample collection and DNA extraction - Opportunistic sampling was conducted in the study area during the period 1998-2004. Sample collection was concentrated in mountain districts where 2 or more packs were present. Non-invasive samples (hairs, scats and blood residuals in snow) were used as sources of genomic DNA, muscle, or skin samples were collected only from wolf carcasses discovered within the study area (for collection methods see Scandura et al. in prep.).

The Chelex-method (Walsh et al. 1991) was employed to isolate DNA from one to ten hair follicles per sample. The QIAamp Tissue and Stool kits (Qiagen, Hamburg, Germany) were used to extract DNA from tissues or blood and excremental samples, respectively. Suppliers' protocols were followed in all cases, except for melted blood/snow mixtures (see Scandura 2005).

Sex determination - According to Seddon (2005) sex determination was carried out by a canid-specific amplification of two intronic regions of the sexual chromosomes of canids, one each in the DBY gene (DBY7) and the DBX gene (DBX6). A multiplexed PCR was used, as specified for microsatellites, except for primer concentration (0.15 μM for DBX6, 0.10 μM for DBY7) and amplification profile (40 cycles of touchdown with annealing at 60-50 °C). PCR products were visualised on a 2% agarose gel containing ethidium bromide for visual detection. Male templates produced two bands (249 bp and 118 bp), whereas in females only the DBX 249-bp fragment was amplified. Each set of reactions included a male control (Y-positive), a female control (Y-negative) and a blank (no template).

Y-linked microsatellite genotyping - Samples which provided reliable genotyping at autosomal loci (see Scandura et al. 2006 for evaluation of reliability and description of autosomal microsatellite amplification) and resulted to be male according to autopsy or molecular sexing (N=31) were amplified at four Y-chromosome microsatellite loci: MS34A, MS34B, MS41A and MS41B (Sundqvist et al. 2001). Amplifications were performed in 10 μl reaction mix containing 3 μl of template, 0.5 U of EuroTaq DNA polymerase (EuroClone, città, stato), 10 mM Tris-HCl, 50 mM KCl, 3 mM MgCl2, 0.1 mg/mL BSA (Promega, città, stato), 100 μM of each dNTP and 2 pmol of each primer (the forward one was labelled with

an ABI dye). The amplification profile was set up with an initial step of denaturation at 95 °C for 3 min, followed by 40 cycles of 92 °C for 40 s, Ta (touchdown 72-65 °C) for 45 s, and 72 °C for 30 s. A further extension step of 72 °C for 10 min concluded the reaction. One PCR blank was included in each set of amplifications, for contamination checking, as well as a positive control, for correct amplification checking. Successful PCRs were run on an ABI PRISM 3130 automatic sequencer and allele lengths were subsequently determined using the PEAK SCANNER v. 1.0 software (Applied Biosystems Foster City, California).

Y-DNA variation - Level of polymorphism at Y-chromosome microsatellites was evaluated with respect to the number of different alleles and to allele frequencies. Furthermore, alleles at multiple loci were combined to construct haplotypes. Temporal and geographical distribution of haplotypes in the territory of the province of Arezzo was evaluated in MapInfo Professional 8.0 (Mapinfo Corporation). As samples included in this study had been previously genotyped with a set of 9 autosomal microsatellites (see Scandura et al. in prep.). We were able to combine data obtained with autosomal and Y-linked markers. On this way, we performed a parentage analysis to assess the kinship relationships between individuals in the Alpe di Catenaia pack, which was intensively monitored during the study. Parentage in this pack was assessed using the software CERVUS 3.0 (Marshall et al. 1998, Kalinowski et al. 2007), accounting, for each individual, for the likelihood of inherintance of each allele in its genotype from a panel of potential parents. In this analysis, temporal constrains (e.g. data of birth or death, if known) were used to restrict the number of potential parents. Finally, the resulting pedigree of all sampled individuals in Alpe di Catenaia was evaluated for consistency with Y microsatellite data.

RESULTS

According to Scandura et al. (in prep.), with the use of 9 polymorphic microsatellite autosomal loci, we successfully genotyped 177 samples that identified 86 different individuals within the study area in the period 1998-2004. Of them, 4 resulted to be foxes and 8 dogs, the remaining 74 being attributed to wolves (assignment probability > 95%). The 31 male samples resulting from the analysis of autosomal loci were further investigated here, those samples resulting to belong to the same individual were considered only once (Y-genotypes considered N=26).

Among Y chromosome loci, locus MS41A was monomorphic for allele 206, while MS41B presented irregular motif repeats and so they were excluded. The two loci remaining

were polymorphic with three alleles each (Fig. 1) and yielded four different haplotypes (Fig. 2). The temporal occurrence of haplotypes is shown in fig. 3. The dominant haplotype (H1) was present throughout the study period in different areas; the second most common, H2, firstly appeared in 2002 and was present in the province until the end of the monitoring; while H3 (two individuals) and H4 (one individual) belonged to wolves found dead that were sampled for the first (and last) time the same year of their death.

The structure of the Alpe di Catenaia pack has been investigated since 2000. Combining field and genetic data, we were able to reconstruct relationships among pack members. Genealogy within the pack is shown in fig. 4. In the period 2000-2003 the breeding pair was stably composed by female W38 and male W11 (which presented haplotype H1); in 2004 an unrelated individual (W58), carrying H2 haplotype, replaced the breeding male, mating with the same female W38, as supported by parentage analysis in CERVUS.

DISCUSSION

In this study, we examined the variability in the male lineage in a Central Italian wolf population. We found a relevant level of polymorphism at two markers linked to the Y chromosome. Each of them showed three different alleles and four different haplotypes were found in our samples. Such amount of variation, detected at a limited geographic scale, is significant if compared to the absolute lack of variation observed in the mitochondrial DNA of the Italian wolf population (Vilà et al. 1999; Randi et al. 2000). This result has two possible explanations: 1) the bottleneck experienced by the Italian wolf population was less dramatic in the male than in the female fraction; 2) the presence of multiple alleles at Y microsatellites is a consequence of genetic introgression from the domestic dog.

Under the former hypothesis, many (at least four) male lineages and only one matriline would have been present in the surviving wolves during the demographic minimum, that occurred between 1960 and 1970 (Cagnolaro et al. 1974).

A similar pattern has been observed by Sundqvist et al. (2001) in Scandinavia, where they found two different Y-haplotypes occurring in a 16-year period after the wolf recolonization, while a single mitochondrial haplotype was present in the same population (Ellegren et al. 1996). The variability found in Scandinavia is lower than that of other Northern countries, like Russia and Baltic states, but consistent with the observed number of alleles in the population at autosomal loci (Ellegren et al. 1996), involving a minimum of three founders (two males and a female) (Sundqvist et al. 2001).

Wolves and domestic dogs can hybridize (Vilà and Wayne 1999) and the origin of multiple Y haplotypes could be associated to introgressive hybridization between the two animals. Evidences of hybridization have been found in Italy at a not negligible frequency (Verardi et al. 2006) and suspect hybrids had been sampled also in the Arezzo province (Scandura et al. in prep.). So this hypothesis cannot be ruled out. However, the presence of signs of hybridization only in male uniparental markers (and not in mtDNA) would contradict the idea that the chance of mating between a male wolf and a bitch is higher than the opposite pairing case (Vilà and Wayne 1999).

We found that the most frequent haplotype (H1) was shared by all the analyzed samples throughout the study area in the first four years of monitoring (1998-2001), and, only afterwards, we recorded the appearance of three new haplotypes. The increased number of haplotypes that we observed is probably caused by the immigration of wolves dispersing from other regions. This is likely related with the ongoing expansion of the wolf population in Italy. During last century wolves were confined into restricted areas of central-southern Italy but they are actually recovering throughout their historic range, reaching saturation levels in some areas like the one under study. The recolonization pattern could have brought in the study area Y-haplotypes previously confined into other parts of the peninsula.

Long-range dispersal is likely to be bidirectional and we should thus assume that some individuals immigrated yearly into the study population and that some of them may have been successfully integrated into local packs (Mech and Boitani 2003). Adopted individuals are favoured in replacing a member of the breeding pair, due to the general attitude towards outbreeding observed in the species (Smith et al. 1997). In fact, these intruders are potential carriers of novel alleles thus they could contribute to constrain inbreeding at low levels in the population. This process can help explain the prolonged maintenance of the genetic differentiation among local packs. In this study we reconstructed the genealogy of an intensively monitored pack. We found that the originary breeding pair (♂ W11 + ♀ W38) remained stable and successfully reproduced since 2000 until 2003. Then, few months after the last sampling of the breeding male (carrying H1 haplotype), we detected the presence of a new unrelated male (W58, carrying H2 haplotype) first recorded in the eastern part of the study area the year before. W58 replaced the breeding male in the pack as a pup was identified as descendant from him and female W38 in 2004. Replacement of the male member of the breeding pair by an unrelated individual is reported to be frequent in wolf populations (Vonholdt et al. 2008).

This study demonstrates that a new class of paternally inherited molecular markers, that has been neglected in previous studies on Italian wolves, can instead be useful to investigate social dynamics and male-specific dispersal patterns, as well as to give support to the reconstruction of genealogies in wild wolf populations.

Our results prompt to check Y-haplotype variability across the whole Italian population, to get an idea of the overall level of diversity and to produce a complete map of the haplotype distribution throughout the peninsula. Particularly interesting would be the analysis of the recolonized alpine population, which will allow to comprehend the relative contribution of the two sexes in promoting the range expansion in the Alps.

In conclusion, we believe that future genetic surveys on the Italian wolf population, even those based on a non-invasive sampling strategy, should largely benefit from the addition of uniparental Y-linked microsatellites to the commonly used sets of autosomal markers.

ACKNOWLEDGEMENTS

We thank C. Capitani, L. Mattioli, M. Geri and all the people who helped us in collecting samples. We also thank P.F. Sechi for his contribution to laboratory activities. The wolf monitoring project was funded by the Provincial Administration of Arezzo, the Regional Government of Tuscany, and the Italian Ministry for the University and Research (PRIN 2003). L. Iacolina benefited by a fellowship provided by the Sardegna Resorts srl.

FIG. 1. – Alleles per locus and frequencies at two microsatellites Y-linked loci.

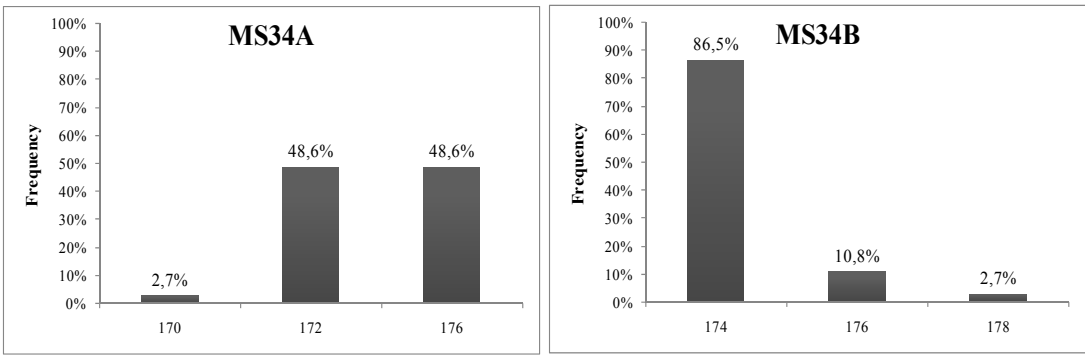
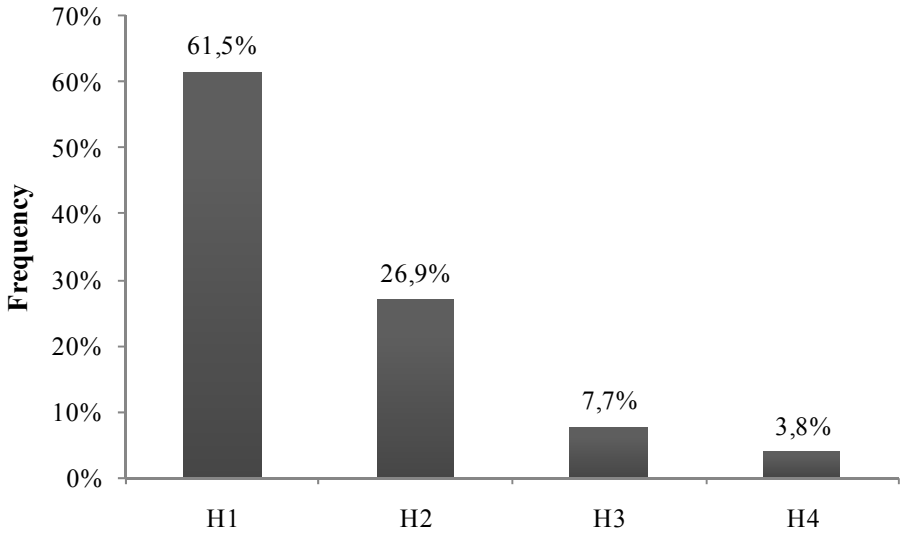


FIG. 2. – Y chromosome haplotypes found in the population and allele sizes.



MS34A	176	172	172	170
MS34B	174	174	176	178

FIG. 3. – Y chromosome haplotypes distribution through years and areas of the Arezzo Province (protected areas are shadowed). Differences in the number of reported haplotypes are due to obsillations in the occurrence of male sampling through years.

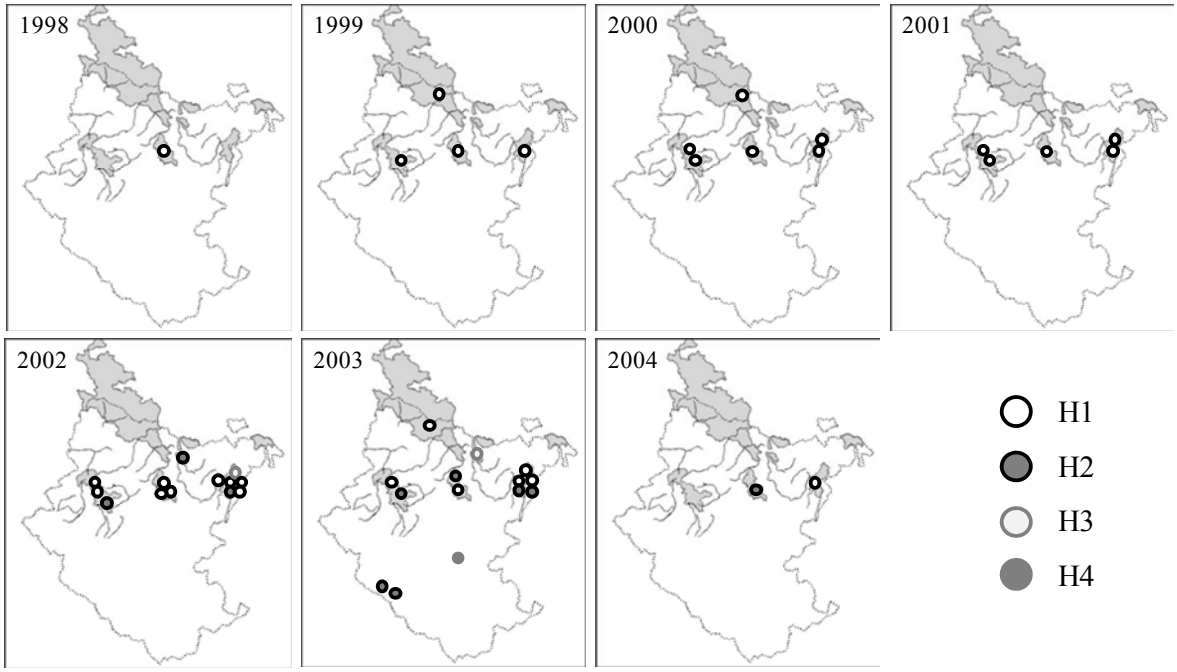
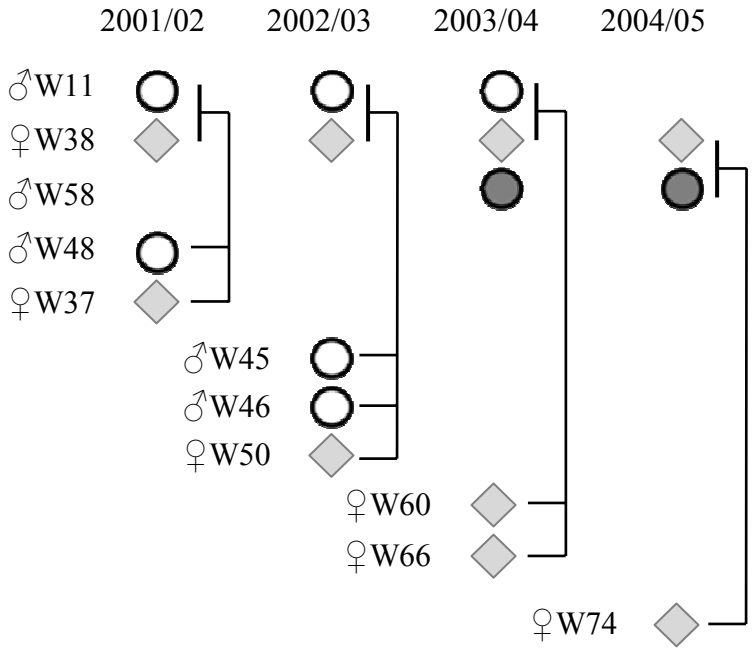


FIG. 4. – Reconstruction of Alpe di Catenaia wolf pack's genealogy in the time span 2000-2004.



LITTERATURE CITED

Apollonio M, Mattioli L, Scandura M, et al. 2004 Wolves in the Casentinesi Forests: insight for wolf conservation in Italy from a protected area with a rich wild prey community. Biological Conservation 120: 249-260

Cagnolaro L, Rosso D, Spagnesi M, Venturi B 1974 Investigation on the wolf (*Canis lupus*) distribution in Italy, in Canton Ticino and Canton Grigioni (Switzerland). Ricerche di Biologia della Selvaggina, 59, 1-75 [in Italian].

Capitani C 2005 Dynamics of a wolf (*Canis lupus*) population in Eastern-Central Apennines. PhD Dissertation, University of Sassari, Italy.

Capitani C, Mattioli L, Avanzinelli E, et al. 2006 Selection of rendezvous sites and reuse of pup raising areas among wolves Canis lupus of north-eastern Apennines, Italy. Acta Theriologica 51: 395-404.

Douadi MI, Gatti S, Levrero F, et al. 2007 Sex-biased dispersal in western lowland gorillas (*Gorilla gorilla gorilla*). Molecular Ecology 16: 2247-2259.

Eriksson J, Siedel H, Lukas D, et al. 2006 Y-chromosome analysis confirms highly sex-biased dispersal and suggests a low male effective population size in bonobos (*Pan paniscus*). Molecular Ecology 15: 939-949.

Ellegren H, Savolainen P, Rosén B 1996 The general history of an isolated population of the endangered grey wolf *Canis lupus*: a study of nuclear and mitochondrial polymorphism. Philosophical Transactions of the Royal Society, Series B 351: 1661-1669.

Fabbri E, Miquel C, Lucchini V, et al. 2007 From the Apennines to the Alps: colonization genetics of the naturally expanding Italian wolf (*Canis lupus*) population. Molecular Ecology 16:1661-1671.

Gazzola A, Avanzinelli E, Mauri L, Scandura M, Apollonio M 2002 Temporal changes of howling in south European wolf packs. Italian Journal of Zoology 69: 157-161.

Greenwood PJ 1980 Mating System, Philopatry and Dispersal in Birds and Mammals. Animal Behaviour 28: 1140-1162.

Grewal SK, Wilson PJ, Kung TK, et al. 2004 A genetic assessment of the eastern wolf (*Canis lycaon*) in Algonquin provincial park. Journal of Mammalogy 85: 625-632.

Hammer MF, Blackmer F, Garigan D, Nachman MW, Wilder JA 2003 Human population structure and its effects on sampling Y chromosome sequence variation. Genetics 164: 1495-1509.

Kalinowski ST, Taper ML, Marshall TC 2007 Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Molecular Ecology 16: 1099-1006.

Kawamoto Y, Tomari KI, Kawai S, Kawamoto S 2008 Genetics of the Shimokita macaque population suggest an ancient bottleneck. Primates 49: 32-40.

Lucchini V, Galov A, Randi E 2004 Evidence of genetic distinction and long-term population decline in wolves (*Canis lupus*) in the Italian Apennines. Molecular Ecology 13(3): 523-536.

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

Mattioli L, Apollonio M, Mazzarone V, Centofanti E 1995 Wolf food habits and wild ungulate availability in the Foreste Casentinesi National Park, Italy. Acta Theriologica 40: 387-402.

Mattioli L, Capitani C, Avaninelli E, Bertelli I, Gazzola A, Apollonio M 2004 Predation by wolves (*Canis lupus*) on roe deer (*Capreolus capreolus*) in north-eastern Apennines, Italy. Journal of Zoology 264: 1–10.

Meadows JRS, Hanotte O, Drögemüller C, et al. 2006 Globally dispersed Y chromosomal haplotypes in wild and domestic sheep. Animal Genetics 37. 444-453.

Mech LD, Boitani L (Eds.) 2003 Wolves: Behavior, Ecology, and Conservation, The University of Chicago Press, Chicago, USA.

Musiani M, Leonard JA, Cluff HD, et al. 2007 Differentiation of tundra/taiga and boreal coniferous forest wolves: genetics, coat colour and association with migratory caribou. Molecular Ecology 16: 4149-4170.

Promberger C, Schröder W (Eds.) 1993 Wolves in Europe: Status and perspectives. Munich Wildlife Society, Ettal, Germany.

Randi E, Lucchini V, Christensen MF, et al. 2000 Mitochondrial DNA variability in Italian and East European wolves: detecting the consequences of small population size and hybridization. Conservation Biology 14: 464-473.

Scandura M, Apollonio M, Mattioli L 2001 Recent recovery of the Italian wolf population: a genetic investigation using microsatellites. Mammalian Biology 66: 321-331.

Scandura M 2005 Individual sexing and genotyping from blood spots on the snow: a reliable source of DNA for non-invasive genetic surveys. Conservation Genetics 6: 871-874.

Scandura M, Capitani C, Iacolina L, Apollonio M 2006 An empirical approach for reliable microsatellite genotyping of wolf DNA from multiple noninvasive sources. Conservation Genetics 7: 813-823.

Seddon JM 2005 Canid-specific primers for molecular sexing using tissue or non-invasive samples. Conservation Genetics 6: 147-149.

Smith D, Meier T, Geffen E, et al. 1997 Is incest common in gray wolf packs? Behavioral Ecology 8: 384-391.

Sundqvist A-K, Ellegren H, Olivier M, Vila C 2001 Y chromosome haplotyping in Scandinavian wolves (*Canis lupus*) based on microsatellite markers. Molecular Ecology 10: 1959-1966.

Tosi AJ, Morales JC, Melnick DJ 2000 Comparison of Y chromosome and mtDNA phylogenies leads to unique inferences of macaque evolutionary history. Molecular Phylogenetics and Evolution 17: 133-144.

Van Hooft WF, Groen F, Prins HHT 2002 Phylogeography of the African buffalo based on mitochondrial and Y-chromosomal loci: Pleistocene origin and population expansion of the Cape buffalo subspecies. Molecular Ecology 11: 267-279.

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

Vilà C, Savolainen P, Maldanado JE, et al. 1997 Multiple and ancient origins of the domestic dog. Science 276: 1687-1689.

Vilà C, Amorim IR, Leonard JA, et al. 1999 Mitochondrial DNA phylogeography and population history of the gray wolf *Canis lupus*. Molecular Ecology 8: 2089-2103.

Vilà C, Wayne RK 1999 Hybridization between wolves and dogs. Conservation Biology 13: 195-198.

Vilà C, Walker C, Sundqvist AK, et al. 2003 Combined use of maternal, paternal and bi-parental genetic markers for the identification of wolf-dog hybrids. Heredity 90: 17-24.

Vonholdt BM, Stahler DR, Smith DW, et al. (2008) The genealogy and genetic viability of reintroduced Yellowstone grey wolves. Molecular Ecology 17, 252-274.

Walsh PS, Metzger DA, Higuchi R 1991 Chelex-100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. BioTechniques 10: 506-513.

Wayne RK, Lehman N, Allard MW, Honeycutt RL 1992 Mitochondrial DNA variability of the gray wolf: genetic consequences of population decline and habitat fragmentation. Conservation Biology, 6, 559-569.

Yannic G, Basset P, Hausser J 2008 Phylogeography and recolonization of the Swiss Alps by the Valais shrew (*Sorex antinorii*), inferred with autosomal and sex-specific markers. Molecular Ecology 17: 4118-4133.

Conclusions

The data collected in this study allowed me to get new insight into the population genetic structure, at different spatial scale, of two wild species, wild boar (*Sus scrofa*) and wolf (*Canis lupus*), using molecular markers.

In the *First part* (**Chapter I**), by the analysis of one mitochondrial and 10 independent nuclear markers in wild boars collected across Europe, I evaluated the possible contribution of different natural and human-related processes in shaping the present genetic diversity of the species in the old continent. The role of past wide-scale events, like range and size fluctuations occurring during the last glacial and postglacial periods, are pointed to as the main force leading to the observed levels of differentiation in Europe.

The most important event in shaping the observed pattern of diversity seems to have been the last glaciation, which was followed by a sudden demographic and spatial expansion from one or more southern refugia. The comparative analysis of wild boar control region sequences shows the occurrence of an exclusive mtDNA lineage in peninsular Italy and the signature of a past demographic expansion, which could well have followed the range contraction occurring in Europe during the last glaciation.

The genetic signature of more recent processes, which were mostly related to human activities, can be detected but it appears marginal. Data point to a single area of discontinuity which corresponds to the Alps. Wild boars sampled south of this chain show, indeed, higher levels of overall genetic variation, a private mtDNA haplogroup and endemic diversity at microsatellite loci.

After having investigated the driving forces shaping genetic variation at a wide scale, I focused on a wild boar population in Central Italy (Tuscany), studying the effects of spatial and social organization (*Second part*).

In **Chapter II**, I explored relatedness within social groups in Alpe di Catenaia. I found a correlation between inter-individual spatial distance and genetic relatedness only in summer-early autumn and it seemed to be associated to the presence of piglets. The prediction of matrilinearity in wild boar social units was not confirmed, as a low degree of relatedness was observed within groups. Aggregations of unrelated adult sows (with their litter) were detected in the study population. The high hunting pressure also help explain the pattern observed in our study. The study wild boar population was indeed characterized by an overall high

mortality rate, mostly due to hunting and poaching (86% of deaths in our sample). The resulting high turn-over could have affected the distribution of genes in the population, accounting for the observed deviation from the expected pattern. Accordingly, I argue that the observed social organization would turn out from grouping of unrelated survivors, that is promoted by the presence of wolves in the area. The advantages of group living in presence of predators can indeed be safeguarded by joining groups of conspecifics, irrespectively from their degree of relatedness.

Another possible explanation to the low degree of relatedness within groups could be multiple paternity, investigated in **Chapter III**. Multiple paternity was found at high frequency in the Alpe di Catenaia wild boar population. Six out of twelve litters (50%) proved to have been sired by more than one boar. Considering that in our study multiple paternity was detected by a parsimony approach (e.g. two fathers with similar genotypes might have been interpreted as one single father), the actual degree of polyandry shown by females in the population can be still higher. On the other hand, we excluded the risk to have introduced an opposite bias, possibly due to a selection of large-sized litters in our sample, thus maximizing the chance to detect multiple paternity.

Multiple paternity frequency in our study is higher than that recently reported by Delgado et al. (2008) for a Portuguese wild boar population (Alentejo). An effect of an age difference in the samples of females used in the two investigations was ruled out, as multiple paternity interested both adult and subadult females' litters without apparent differences. Analogously, no effect of body mass on multiple paternity frequency was observed.

As the likelihood to mate with multiple males depends on the encounter rate, one could expect an effect of the population density. Assuming the hunting bag as proportional to population abundance, the two populations show similar wild boar densities (Fernadez-Llario et al. 2003). This also suggests that female mating strategy in the population is not density-dependent.

I therefore concluded that multiple paternity is likely to be related to other (ecological, demographic, etc.) factors influencing the breeding strategy adopted by sexes, which are aimed at maximizing their respective fitness. Further investigations on within- and between-sexes interactions during the breeding period are warranted, in order to understand the evolutionary advantages of promiscuity and multiple paternity in this species.

The second half of this dissertation deals with the social organization and patterns of population assembly in the wolf.

First I contributed to develop a methodological approach for achieving a reliable data set of genotypes from multiple sources of DNA collected in a monitored population (*Third part*). This method relies on the relationship between sample quality and amplification outcome, which is ultimately related to the occurrence of genotyping errors (allelic dropout, false alleles). Genotypes not reaching a satisfactory confidence level can either be replicated to become reliable or excluded from the data set.

After the definition of the analisys protocol, in the *Fourth part*, I analysed samples collected in an area central to the present Italian wolf's range, investigating the fine-scale genetic structuring within an area showing high pack densities and local saturation (Apollonio & Mattioli 2006). In **Chapter V**, I collected genetic data on the wolf population inhabiting the province of Arezzo, using a combination of biparentaly inherited autosomal markers and a non-invasive sampling strategy. This genetic investigation revealed no remarkable sign of gene flow among neighbouring packs and a consequent appreciable differentiation between close geographic areas, even in absence of physical barrier to wolf dispersal. I thus rejected the hypothesis of frequent interpack migration observed in the Yellowstone wolf population (Vonholdt et al. 2008), while supporting the idea that in the Italian population long-range dispersal is the key factor maintaining the genetic diversity, as alredy suggested to interpret the recent wolf recolonization of the Alps (Fabbri et al. 2007).

This second hypothesis is even supported by the use of paternally inherited markers. In **Chapter VI,** indeed, I used two microsatellites located in the Y chromosome on the same sample that I had previously analyzed with biparental markers. This investigation revealed the presence in the study area of a single Y haplotype (H1) till 2002 and the sudden appearance since then that of three new haplotypes, that in some areas replaced the former one.

Foreign unrelated individuals are favoured in replacing a member of the breeding pair, due to the general attitude towards outbreeding observed in this species (Smith et al. 1997). This process has been observed in the Alpe di Catenaia (AC) pack. During the period 2000-2003, the rendezvous site of this pack was constantly detected in the same small valley (Capitani et al. 2006) and a pair of individuals (male W11 and female W38) was repeatedly sampled in the pack range. Parentage analysis performed in **Chapter V** confirmed them to represent the breeding pair of the pack. Six wolves in our sample were likely offspring of this pair, and each of them was found during the course of one single year in the Alpe di Catenaia range only or in its proximity. Then, in 2003, the presence in the AC pack of a new unrelated male (W58), carrying a different Y-haplotype (H2), was detected and he was likely to have replaced the former breeding male, as a pup born in 2004 was identified as daughter of the

pair W58-W38. The pedigree reconstruction reported in **Chapter V** is in agreement with the transmission of Y-linked markers in the AC pack, as presented in **Chapter VI**.

Despite of its central position and apparent stability, the Alpe di Catenaia pack during the study period did not represent a source of migrants that joined neighbouring packs. Thus, if mortality did not account for this result, the only possible deduction is that animals dispersed out of the study area.

Analogously, the appearence of previously undetected haplotypes in the Arezzo population can be justified by the arrival of long-range dispersers coming from other areas of the wolf range in the peninsula.

Overall, these results suggest the existence of a cryptic genetic structure in the Apennine wolf population, which can result from a combination of demographic and dispersal patterns, like stability of breeding pairs and the tendency to replace a breeder loss with individuals coming from far away (long-range dispersers). This knowledge, if supported by future studies, is of paramount importance for the conservation of this carnivore in Italy.

References

Apollonio M., Mattioli L. (eds.) (2006) Il lupo in provincia di Arezzo. Editrice Le Balze, Montepulciano, Italy.

Capitani C., et al. (2006) Selection of rendezvous sites and reuse of pup raising areas among wolves Canis lupus of north-eastern Apennines, Italy. Acta Theriologica 51: 395-404.

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

Fabbri E., et al. (2007) From the Apennines to the Alps: colonization genetics of the naturally expanding Italian wolf (Canis lupus) population. Molecular Ecology 16: 1661-1671.

Fernandez-Llario P., Mateos-Quesada P., Silvério A., Santos P. (2003) Habitat effects and shooting techniques on two wild boar (*Sus scrofa*) populations in Spain and Portugal. Zeitschrift Jagdwissenschaft 49: 120-129.

Smith D., et al. (1997) Is incest common in gray wolf packs? Behavioral Ecology 8: 384-391.

Vonholdt B.M., et al. (2008) The genealogy and genetic viability of reintroduced Yellowstone grey wolves. Molecular Ecology 17: 252-274.

AKNOWLEDGEMENTS

This work was financially supported by the

Sardinia Resort S.r.l.

who gave me the opportunity to perform this research.

Here I owe my debt of gratitude to my financial supporter.

Thank you

Questo lavoro è stato possibile grazie ai finanziamenti erogati da

Ministero dell'Istruzione, dell'Università e della Ricerca

Provincia d'Arezzo

Regione Toscana

Sardinia Resort S.r.l.

Ringrazio in particolare Dr. P. Banti, Dr. G. Chianucci, Dr. P. Pedone e Dr. L. Mattioli.

Per cominciare una premessa, sapete bene che non sono brava con le parole, quindi sappiate sin d'ora che il mio affetto e la mia gratitudine vanno ben oltre quel che riuscirò a mettere nero su bianco.

Innanzitutto vorrei ringraziare il prof. Apollonio per avermi dato questa possibilità.

Un grazie di cuore va a Massimo, per la sua guida, la sua pazienza, gli insegnamenti e la pazzia!!! Senza il suo aiuto e sostegno tutto questo non sarebbe stato possibile. Ringrazio anche le sue bimbe per l'allegria e gli aneddoti che hanno reso più leggeri questi tre anni.

Altro grazie sentito va a Simone per tutte le volte che mi è venuto in soccorso, i suoi insegnamenti ed i preziosi commenti su tutto il lavoro svolto ma non solo!

E poi come dimenticare tutti gli amici che hanno condiviso con me gioie e dolori di questo dottorato?! Un abbraccio alle mie mitiche Bialowieza girls che sono sempre importanti compagne di viaggio (in senso lato oltre che letterale): Silvia, Sara ed Anna. Voglio ringraziare Francesca D.B., Daria, Tiziana, Antonio, PierFrancesco, Pasquella, Amalia, Nicoletta, Kristina, Giovanni, Mario, Marco e tutto lo "staff"(siete troppi per mettere tutti i nomi, ma voi sapete benissimo chi siete!) del laboratorio di genetica per l'aiuto, la competenza, le chiacchere e l'amicizia. E poi ci sono Francesca, Rossana ed Emiliana che dai luoghi più disparati sono sempre state presenti. Nonché gli amici che ora dividono con me la pressione di questo periodo: Alberto, Paolo M., Paolo B., Valentina e Ivan. Ragazzi, ce la possiamo fare!!!

Ovviamente un ringraziamento va a tutti quelli che mi hanno aiutato (e ospitato) per le raccolte dati ma non solo. Tutti i membri di questo grande gruppo con le sue varie sedi distaccate: Emanuela, Elisa, Barbara (meravigliosa compagna di stanza), la piccola Nadia (che ormai avrà il telefono fuso per tutte le volte che l'ho chiamata in cerca di aiuto), super Andrea e lo Sgrigno (grazie ragazzi per i consigli, il sostegno e la simpatia), Fabiella, Roberta e Stefano, tutti i cacciatori per la loro collaborazione (in modo particolare l'URCA). Ma anche Siriano ed Elena per aver svolto, ognuno a proprio modo, ruoli fondamentali in questo mio percorso.

Non posso dimenticare anche gli amici che nel corso del tempo hanno preso altre strade ma che tanto mi hanno aiutato: Claudia e Fabio. E coloro che pur non facendo parte di questo "universo parallelo" sono da anni amici insostituibili: Martina, Maggie e Gisella, ma anche Edith, Arika, Roser, Paola, Manuela, Daniela, Vania e Mary che in un arco spazio-temporale molto ampio sono stati al mio fianco in mille modi. Grazie a tutti voi!

Ma sopra ogni cosa voglio ringraziare la mia famiglia. I miei genitori, mia sorella e Francesco D., il mio compagno di vita Francesco per tutte le volte che hanno riso con me, raccolto i cocci del mio umore e calmato o sopportato i miei nervi. Lo so, vi ho messo tutti a dura prova, ma voi siete sempre rimasti al mio fianco e questo non ha prezzo!! Grazie! Grazie!! Grazie!!!

Laura Iacolina
Dipartimento di Zoologia e Genetica Evoluzionistica
Via Muroni 25
I-07100 Sassari
Italy
e-mail: liacolina@uniss.it
Tel.: +39(0)79228628
Fax: +39(0)79228665