CONSERVATION GENETICS OF THE GENUS Sus

Randi E.

Istituto Nazionale per la Fauna Selvatica, via Cà Fornacetta 9, 40064 Ozzano dell'Emilia, Bologna, Italy.

Abstract: Recent technical advances in biochemical and molecular biology make it possible to study the population genetics and phylogenetics using biological samples collected undestructively during field researches, or collected from museum specimens. Applications of conservation genetics of the genus *Sus* range from the definition of a molecular phylogeny of the Suiformes to the description of boundaries between species and subspecies and the identification of the origins of wild, feral, domestic or cross-breeded populations. In this paper I review information on the genetic structure of west European wild pig populations, and present recent findings on nucleotide variability of mitochondrial DNA genes. The amplification of DNA through the polymerase chain reaction and following nucleotide sequencing of selected mitochondrial DNA genes, are effective methods which can be applied to describe patterns of within- and among- species genetic variability in the Suidae.

Keywords: *Sus scrofa,* Suidae, Molecular biology, Populations, Hybridization, Mitochondrial DNA.

IBEX J.M.E. 3:6-12

1. Introduction

The Suidae comprise a number of species of great biological interest and economical value. Suids interact with human activities, because they are domesticated, reared, crossed, translocated, hunted, eated, and in certain cases, venerated or persecuted. They have a place in many traditional cultures and in the everyday life of millions of humans. Such notwithstanding, the biology of many species is still poorly known, and their management and conservation cannot take advantage of the necessary basic information.

Conservation genetics, the application of population (and in particular small populations) genetic models and biochemical and molecular technologies to conservation biology, is aimed to preserve biodiversity at the different levels it is organized (genes, populations, species, evolutionary lineages), and can be applied to address a wide range of still open questions about the biology of the Suidae.

First of all we need to develop a molecular phylogeny of the Suidae, which must integrate non-molecular knowledges, and constitute the background within which to describe the time scales and patterns of their evolution. This framework is necessary to delineate species (as well as subspecies and populations) boundaries and relationships. Just to mention a few examples: the wild pig fauna of entire regions (e.g., the Philippines) is poorly known, and species/subspecies boundaries have not been clearly delineated so far (Oliver *et al.*, 1993). The patterns of speciation and partitions of genetic diversity among populations of the African suids await description (Grubb, 1993), as well as it is not known the extent of genetic divergence among the 3 (or more ?) subspecies of the babirusa (Macdonald, 1993). Captive breeding is frequently mentioned as a necessary option to secure the preservation of some endangered suids (first of all, the highly endangered Pygmy hog, *Sus salvanius*; Oliver & Deb Roy, 1993). Captive propagation with the perspective of reintroduction in the wild, needs the preservation of the largest possible fraction of the species' gene pool. Quantification of existing gene diversity within small populations, breeding plans aimed to retain the maximum possible gene diversity and to avoid inbreeding and inbreeding depression are efforts required to reconstitute viable populations in the wild. Suids represent a resource for human populations. Two species have been domesticated (*Sus scrofa* and *S. celebensis*; Groves, 1981), and other are easily bred in captivity (National Research Council, 1983). Most populations of the Eurasian wild pig are intensively exploited and managed, but the genetic, demographic and ecological impacts of intense hunting pressure and animal translocations are still poorly known. In selected cases it will be important to determine the origins of feral pig populations, and evaluate eventual genetic peculiarities, because, besides the obvious problems they pose as pests in exotic habitats, they could have some value for conservation and as possible source of human incomes (Brisbin, 1990), once appropriately managed and controlled.

2. Population genetics of *Sus scrofa*.

Multilocus Protein (enzymes and non-enzymatic proteins) Electrophoresis (MPE) is an inexpensive and simple technique, widely used to study population genetics and phylogenetics of many animal species. Proteins migrate in an electric field because some aminoacids are charged. If a point mutation (or a deletion, an insertion) changes the aminoacid composition of a protein, it will change (in about 30% of the cases) its net electric charge, and its rate of electrophoretic migration will be different and distinguishable from the original molecular form. Appropriate staining recipes will, then, reveal allelic variation at single structural loci. Therefore, it is possible to identificate and count genotypes and alleles, compute within-population estimates of gene diversity, between-population genetic distances, and work up a number of population genetic models like Hardy-Weinberg equilibrium, linkage disequilibrium, gene flow, geographical population structuring, and so forth. Limitations of protein electrophoresis derive from the low mutation rates at structural loci, and from possible selective value of some allelic variants. lt is not expected any aminoacid substitution in populations which were recently separated, and the recovery of heterozygosity is slow after a recent bottleneck (Lande & Barrowclough, 1987). In some cases proteins may be not sufficiently variable to allow discriminating among recently isolated populations.

Some populations of the Eurasian wild pig have been recently studied, using karyotype mapping (Bosma *et al.*, 1984), blood groups (Kurosawa *et al.*, 1979) and MPE (Tanaka *et al.*, 1983; Randi *et al.*, 1989). Results concordantly showed the existence of an east-west clinal variation in allele frequencies, which produced dendrograms separating the eastern from the western phenotypically recognized subspecies (Kurosawa *et al.*, 1979). Studies of restriction fragment length polymorphism of the mitochondrial DNA (Watanabe *et al.*, 1986; Lan & Shi, 1993) supported these conclusions. The observed genetic gap is roughly correspondent to the distributions of the two prevalent karyotypes, 2n = 36 and 2n = 38 (Bosma *et al.*, *op.cit.*). An insufficient sampling of wild pig populations along an east-west transect from central Asia to central Europe prevents us to understand the origins of these concordant patterns of genetic variation: the apparent clinal variation, and the apparent main genetic gap could be the results of recent dispersal from different and anciently isolated centers of origins, and/or of isolation by distance acting on populations with restricted (at least across some areas) gene flow. Multivariate morphometrics of skull measures indicated a more or less linear north-east to south-west dimensional cline, perhaps mainly due to environmental effects and adaptations to food availability on growth rates and adult body size (Randi *et al.*, 1989). Blood groups and MPE showed comparatively greater genetic distances between Asian and European-American domestic pig breeds (Tanaka *et al.*, *op.cit.*), supporting the idea of polyphyletic domestication (Oliver *et al.*, 1993). Some of the modern pig breeds originated from Asian stocks or possibly from crosses between Asian and European pig strains (Ollivier & Sellier, 1982). These informations can be used to investigate the population genetic structure and the putative origins of local wild pig populations, particularly in cases where translocations and/or introductions could have originated populations of uncertain origins or unknown genetic make up. We have used MPE to describe genetic variability in some west European wild pig populations (*S.s. scrofa*), and to assess the genetic structure of the formerly described ltalian subspecies (*S.s. meridionalis*, the Sardinian wild pig, and *S.s. majori*, the Maremma wild pig; Randi *et al.*, 1989). Clustering multilocus pairwise genetic distances indicated small divergence among *S.s. majori* and other west European wild pig populations (Fig.1). These results prompted us to reject the validity of the subspecies *S.s. majori*, which must be considered, at best, an ecotype phenotypically adapted to a Mediterranean type habitat. Within relatively short geographic distances, and in contrast with the presence of significant body size differences, allozyme variability is often not structured enough to evidence significant population gaps. We have recently studied genetic variability in some Bulgarian wild pig populations sampled in localities where were described (Genov *et al.*, 1991) the existence of different morphological phenotypes: a larger one living in the northern plains, and a smaller one occurring in the southern mountains. MPE showed results concordant with morphometrics (Fig. l), but allele divergence among northern and southern populations was small (Randi *et al.*, 1992). Such small distances are not necessary attributable to genetic isolation and absence of gene flow, but most probably to genetic disequilibrium and drift, with consequent ran-

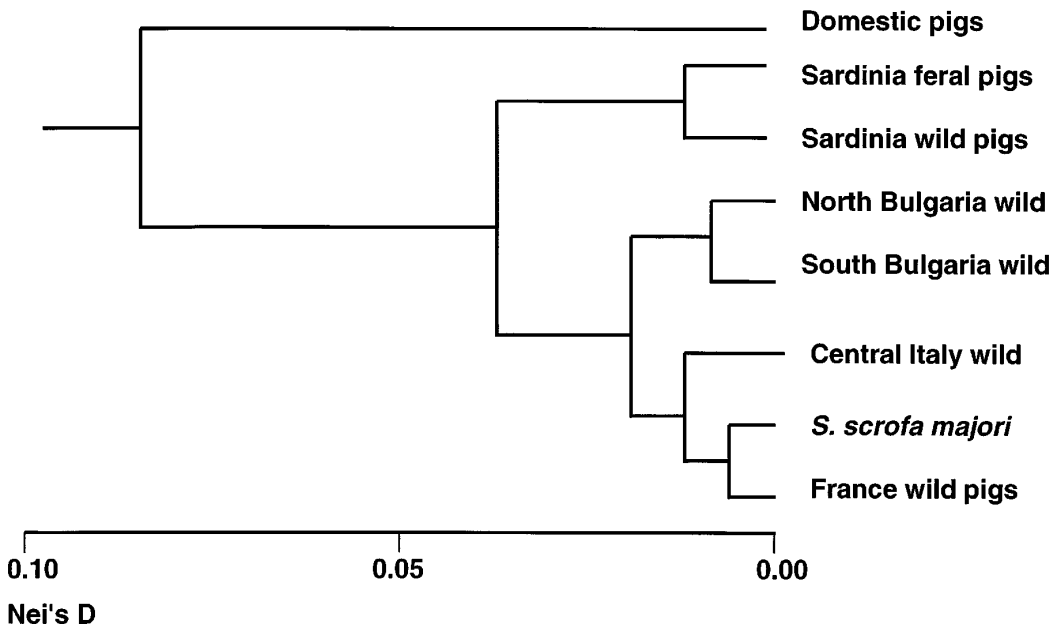

Figure 1 - UPGMA dendrogram obtained by clustering allozyme Nei's genetic distances among some west European wild, feral and domestic pig populations.

dom fluctuations in allele frequencies, due to the genetically small effective size of recently established populations. We observed greater genetic distance values separating wild and feral Sardinian pigs from other west European wild pig populations, thus validating the subspecies *S.s. meridionalis* (Fig. 1). The locus 6PGD is a possible marker of hybridization between wild and domestic pigs. This locus was polymorphic in almost all the studied pig breeds (Franceschi & Ollivier, 1981), while it was monomorphic in all the studied west European wild pig populations (Randi *et al.*, 1989). It was monomorphic in Sardinian ferals, as well as in the Smoky Mountains pigs (USA), which originated, at least in part, from introduced European wild pigs (Smith *et al.*, 1980).

We hypothesized that 6PGD is polymorphic in Asian wild pigs and that this polymorphism was introduced in modern pig breeds through hybridization (Randi *et al.*, 1989). This hypothesis has been supported by Kurosawa and Tanaka's paper (1991) showing extensive 6PGD polymorphism in *S.s. leucomystax* and *S.s. riukiuanus* (with some monomorphic populations, possibly due to recent bottlenecks and isolation), and *S.s. taivanus*. 6PGD and other allozyme polymorphisms (PGM, PGI) are linked to artificially selected meat production characters (e.g., in the linkage group PGI-HAL-6PGD; Rasmusen, 1983), and can show allele frequency divergence among wild and domestic pigs. These enzyme loci could be used to locate the geographic origins of certain wild pig stocks, and to detect hybridization and introgression of domestic genes at the population level.

3. Mitochondrial DNA nucleotide sequencing

Using Restriction Fragment Length Polymorphisms (RFLP) we can estimate rates of nucleotide substitutions through restriction endonuclease digestions of purified target DNA, followed by detection of restriction fragments by hybridization with a cloned labelled probe. The main advantage of this method (as well as of most DNA methods) is given by the possibility to choose the appropriate target DNA sequence to be studied. DNA is not a homogeneous strip of nucleotides, but is a structured macromolecule with single copy structural genes and repetitive DNA, with genes evolving at a low pace, or hot spots of mutation and recombination, *i.e.*, sequences with high rates of molecular evolution. Using cloned hypervariable sequences as probes, it is

possible to identificate the single individual (DNA fingerprinting), while using slowly evolving sequences as probes (*e.g.*, RNA genes), it is possible to study phylogenetic relationships among very divergent taxa. The main drawback of RFLP method is the need of good quality and abundant target DNA, which cannot so easily be obtained during field work or undestructively collected from endangered species. Recently direct nucleotide sequencing has become feasible at the population level, thanks to the discovery of the possibility to amplificate DNA using the polymerase chain reaction (PCR). The target DNA sequence is amplified *in vitro* using two oligonucleotide primers (synthesized *in vitro*). One primer is complementary to the sequence flanking one end of the target DNA, while the other primer is complementary to the other flanking sequence. The presence of free nucleotides and of a thermostable DNA polymerase, results in the extension of both primers, which copy the target DNA sequence. PCR is performed in an automated machine, the thermal cycler, which controls a 3-temperature cycle: DNA denaturation, primer hybridization, DNA extension. This cycle can be repeated 20-50 times, doubling the quantity of target DNA each cycle, and producing, at the end, a million times the amount of target sequence present initially. This pure DNA can be directly sequenced. Nucleotide sequences produce enormous amounts of genetic information: the single nucleotide is the character (with four possible states), and sequences are exactly comparable through different laboratories. Sequences are cumulative information, and DNA databases are exponentially growing.

One of the most informative application of both RFLP and PCR in population and conservation genetics is the study of mitochondrial DNA (mtDNA). MtDNA is present in the mitochondria, it is maternally inherited (through the oocytes), it is haploid and does not recombine. MtDNA shows (at least in mammals) an average rate of molecular evolution 5-10 times faster than average single copy nuclear DNA. This makes mtDNA the molecule of choice to study genetic divergence among conspecific populations, and to determine maternal phylogenies. We are running a project to sequence selected mtDNA genes in the Suidae, with the aim: 1) to identificate reliable slow-evolving regions, which can retain unambiguous phylogenetic signals and which can allow reconstructing the evolutionary patterns of taxa within the Suiformes; 2) to identificate fast-evolving hypervariable regions, which can be used to describe patterns of genetic variability at the interface between conspecific and intraspecific populations. We have designed oligonucleotide PCR primers to amplificate and to sequence the entire mitochondrial cytochrome b (CYB), a protein-coding gene, with intermediate rate of sequence evolution in mammals, and the mitochondrial control region (D-LOOP), a non-coding region involved in the replication of the mtDNA, and usually evolving at high rate in mammals. We have sequenced about 600 bp of the CYB and 500 bp of the D-LOOP in several species of the Suidae and in samples of different populations of west European wild pig. The main results (Fig. 2) indicate that:

1) CYB showed low sequence variation in *S. scrofa*, and divergence among species was due to point mutations only. Preliminar calibration of the CYB molecular clock agrees with presumed paleontological information, and other molecular findings, on times of speciation within the Suiformes, and suggests that CYB is a suitable gene to obtain reliable phylogenetic signals at the ordinal level. 2) D-LOOP showed extensive sequence reorganization among different species, due to both point mutations and duplications of repeated motifs, as well as to insertion or delection of single bases. Therefore, D-LOOP seems to be an unreliable source of phylogenetic information in the Suidae. West European wild pig samples showed a surprisingly low level of sequence variation at the D-LOOP. Most populations shared the same (CYB + D-LOOP) mtDNA haplotype, including domestic pigs, and some Sardinian wild and feral pigs. Sequence divergence within west European *S. scrofa* (wild and domestic) is less than 1%, and we have not found any mtDNA fixed sequence difference between wild and western domestic pigs. About 60% of the studied Sardinian wild pig specimens showed a different mtDNA haplotype (0.3% nucleotide divergence from the common haplotype), which supports the indications of genetic peculiarity of the Sardinian pig population, and further confirms his subspecific status.

4. Conclusions

The analysis of patterns of nucleotide sequence divergence, both among- and within-species, is a preliminar step, necessary to identificate mtDNA genes which could be used as a relia-

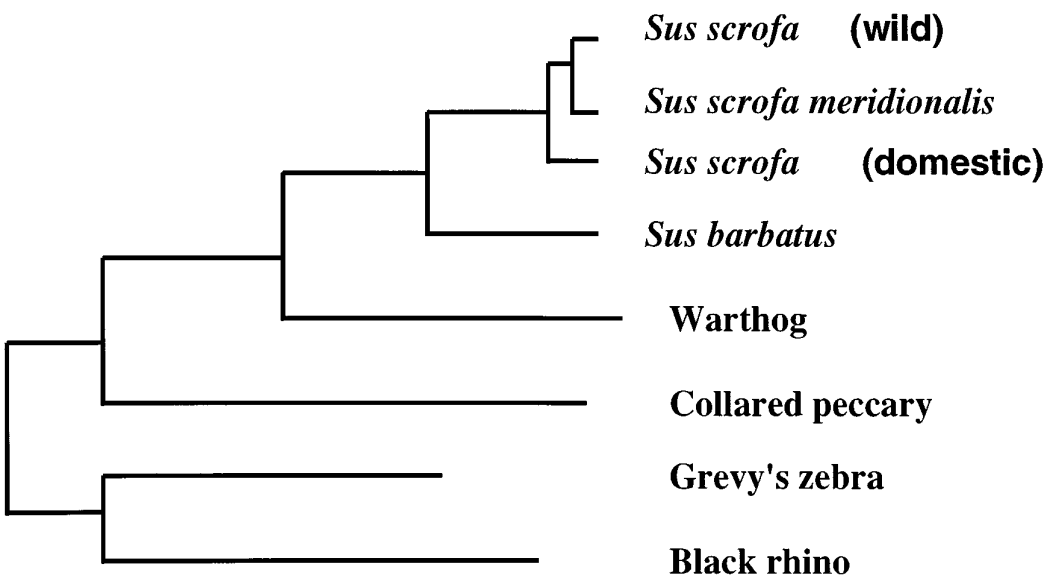

Figure 2 - Neighbor-joining tree showing phylogenetic relationships among Suiformes, obtained using nucleotide sequences of the mitochondrial DNA cytochrome b gene.

ble source of information for intraordinal phylogenetics or population genetics studies in the Suiformes. The mtDNA CYB is a protein-coding gene, and as expected it is structurally conserved across species, as consequence of evolutionary constraints. In the Suiformes CYB seems to evolve at approximately 2% nucleotide substitutions per million year, accordingly with the average rate of mtDNA molecular evolution in mammals. Our findings suggest that CYB can be reliably used to describe times of evolution and pattern of phylogenetic divergence in the Suiformes. CYB will be particularly useful to individuate phylogenetic gaps due to pre- or post-Pleistocene isolation of evolutionary lineages, and could be successfully applied to disentangle species relationships in areas, like the Philippines, of particular biogeographic complexity. The D-LOOP appears to evolve at a faster rate in the Suiformes. lt is mainly a non-transcribed region, with lower constraints to both genome size and structural reorganization. Correct alignment of reorganized D-LOOP regions could be difficult in anciently separated taxa, and consequently the D-LOOP could be reliable to track phylogeny at lower ranks only.

Some portions of the D-LOOP should be hypervariable within species, and therefore should be informative for the description of phylogeographic relationships among conspecific populations or clusters of closely related species. The 500 bp of the D-LOOP we have sequenced in west European wild pig samples, so far, showed a surprisingly low rate of nucleotide variability. Genetic variability among populations was concordantly low at nuclear allozyme loci, mtDNA CYB, as well as at mtDNA D-LOOP, supporting the idea of a recent colonization of western Europe by *S. scrofa* populations which survived a relatively intense bottleneck. These findings furtherly stress the existence of high phenotypical plasticity in *S. scrofa,* which favours great adaptability to very different habitat and food conditions, and suggest caution in using morphometry as a tool for intraspecific taxonomy. Within such a framework, it is noteworthy the presence of a peculiar mtDNA haplotype in about 60% of the studied Sardinian wild pigs. Pigs (as well as most of the present large vertebrates) were most probably introduced by man in Sardinia (Groves, 1989). The observed genetic heterogeneity of the Sardinian pig population could have been originated from multiple introductions of pigs of different geographic origins. The peculiar mtDNA Sardinian haplotype has not been detected in any other

west European wild pig population, so far. lt could represent the original mtDNA haplotype introduced in Sardinia, probably through semi-domesticated pigs from the Middle East. The other mtDNA haplotype detected in the Sardinian pig population is the common west European pig mtDNA haplotype, which could have been introduced in Sardinia later. This hypothesis points out to two problems. First of all we wish to track the geographic origin and distribution of the Sardinian mtDNA haplotype, with a particular attention to the status of the wild pigs of Corsica and Andalusia, which are currently described as *S.s. meridionalis* (Groves, 1981). The second one has more relevant implications for conservation. The Sardinian pig population is a genetically mixed one. lt is probable that the actual Sardinian pig population is a patchwork of ancient feral pigs (probably originated in the Middle East), which have been introgressed with recent western domestic pigs. If so, we need to analyse this patchwork, mapping haplotypes occurence and frequencies, and enforce the preservation of ancient Sardinian pigs (the true *S.s. meridionalis*) in the areas which will be eventually discovered to be unpolluted.

Concordantly with karyotypes, protein markers and mtDNA RFLP, we expect that mtDNA nucleotide sequences will show greater divergence among Asian and European wild pig populations, as well as domestic pig breeds (work in progress). The individuation of fixed mtDNA sequence differences between different wild pig populations, or between wild and domestic pig breeds will greatly aid the analysis of the genetic origins of managed wild pig populations, as well as the individuation of wild x domestic crosses.

MtDNA is maternally inherited and can characterize maternal ancestries only. PCR can be used to perform a very sensitive fingerprinting analysis through the amplification of microsatellites. Microsatellites consist of 10-50 copies of 1 to 6 bp repeats, which are randomly interspersed in all eucariotic DNAs. There is an enormous number of microsatellites and, as many microsatellites have 4 or more alleles of different length (generated probably by asymmetrical crossing-over), they can describe an enormous amount of variability. These repeats are flanked by DNA with unique sequences, which can be used for locus-specific priming for PCR amplification. Microsatellites can be used as fingerprinting, to detect hybrids if parentals are fixed for different alleles, and for population genetics, if population divergence has been very recent. We can take advantage of the great number (probably more than 400) microsatellite loci which have been described in the domestic pig (Rohrer *et. al.*,1994), and that work perfectly on wild pigs' DNA (unpublished results).

One of the most important advantages of PCR is that it makes possible to amplify very small quantities of DNA (in theory starting from a single DNA molecule), also if recovered in very bad conditions. DNA suitable for PCR can be obtained from almost any kind of small samples (one drop of blood, one hair root, a few square millimeters of skin biopsy), also if in very bad state of preservation (old museum skins, old bones, archaeological remains). Fresh and old samples can be preserved at room temperature in absolute ethanol, without necessity to frozen them. It is therefore possible to collect samples for PCR amplification and DNA sequencing as by-product of many field work projects, as well as from museum specimens.

5. Acknowledgements

I wish to thank A. L. Bolelli, P. De Marta and V. Lucchini for assistance during laboratory analyses at Istituto Nazionale per la Fauna Selvatica, and C. H. Diong (Singapore), P. Genov (Sofia), G. Massei (Firenze), G. Tosi (Milano), A. Zilio (Milano) for collaborative projects and sample collection.

REFERENCES

BOSMA A. A., DE HAAN N. A. & MACDONALD A. A., (1984) - Karyotype variability in the Wild boar *(Sus scrofa)*. Symposium International sur le Sanglier. F. Spitz & D. Pépin (eds), Toulouse, *Les Colloques de l'INRA*, n° 22: 53-56.

BRISBIN I. L. JR., (1990) - A consideration of feral swine *(Sus scrofa)* as a component of conservation concerns and research priorities for the Suidae. *Bongo*, 18: 283-293.

FRANCESCHI P. F. & OLLIVIER L., (1981) - Fréquences de quelques gènes importants dans les population porcines. *Z. Tierzuchtg. Zuchtgsbiol.*, 98: 176-186.

GENOV P., NIKOLOV H., MASSEI G. & GERASIMOV S., (1991) - Craniometrical analysis of Bulgarian Wild boar *(Sus scrofa)* populations. *J. Zool., Lond.*, 225: 309-322.

GROVES C. P., (1981) - Ancestors for the pigs: Taxonomy and phylogeny of the genus *Sus*. *Technical Bull. No 3*. Department of Prehistory, Research School of Pacific Studies, Australian National University, Canberra.

GROVES C.P., (1989) - Feral mammals of the Mediterranean islands: documents of early domestication. In: *The walking larder: patterns of domesti-*

cation, pastoralism and predation. J.Clutton-Brock (ed.), Umvin Hayman, London: 46-58.

GRUBB P., (1993) - The Afrotropical suids *(Phaco-choerus, Hylochoerus* and *Potamochoerus)*. Taxonomy and description. In: *Pigs, Peccaries and Hippos*. W. L. R. Oliver (ed), IUCN, Gland, Switzerland: 66-75.

KUROSAWA Y., OISHI T., TANAKA K. & SUZUKI S., (1979) - Immunogenetic studies on wild pigs in Japan. *Anim. Blood Grps. Biochem. Genet.*, 10: 227-233.

KUROSAWA Y. & TANAKA K., (1991) - PGD variants in several wild pig populations of east Asia. *Animal Genetics*, 22: 357-360.

LAN H. & SHI L., (1993) - The origin and genetic differentiation of native breeds of pigs in southwestern China: an approach from mitochondrial DNA polymorphism. *Biochem. Genet.*, 31: 51-60.

LANDE R. & BARROWCLOUGH G. F., (1987) - Effective population size, genetic variation, and their use in population management. In: *Viable populations for conservation*. M. E. Soulè (ed.), Cambridge University Press, Cambridge: 87-123.

MACDONALD A. A., (1993) - The babirusa *(Babyrousa babyrussa)*. In: *Pigs, Peccaries and Hippos*. W. L. R. Oliver (ed.), IUCN, Gland, Switzerland: 161-171.

NATIONAL RESEARCH COUNCIL, (1983) - Little-known Asian animals with a promising economic future. Washington.

OLIVER W. L. R., COX C. R. & GROVES C. P., (1993) - The Philippine warty pigs. *(Sus philippensis* and *S. cebifrons)*. In: *Pigs, Peccaries and Hippos*. W. L. R. Oliver (ed.), IUCN, Gland, Switzerland: 145-155.

OLIVER W. L. R. & DEB ROY S., (1993) - The pygmy hog *(Sus salvanius)*. In: *Pigs, Peccaries and Hippos*. W. L. R. Oliver (ed.), IUCN, Gland, Switzerland: 121-129.

OLIVER W. L. R., GROVES C. P., COX C. R. & BLOUCH R. A., (1993) - Origins and domestication and the pig culture. In: *Pigs, Peccaries and Hippos*. W. L. R. Oliver (ed.), IUCN, Gland, Switzerland: 171-179.

OLLIVIER L. & SELLIER P., (1982) - Pig genetics: a review. *Ann. Genet. Sel. Anim.*, 14: 481-544.

RANDI E., APOLLONIO M. & TOSO S., (1989) - The systematics of some Italian populations of Wild boar *(Sus scrofa)*: a craniometric and electrophoretic analysis. *Z. Säugetierk.*, 54: 40-56.

RANDI E., MASSEI G. & GENOV P., (1992) - Allozyme variability in Bulgarian Wild boar populations. *Acta Theriologica*, 37: 271-278.

RASMUSEN B. A., (1983) - Isozymes in swine breeding. In: *Isozymes*, Vol. 11. Alan R. Liss, Inc., New York: 249-268.

ROHRER, G. A., ALEXANDER L. J., KEELE J. W., SMITH T. P. & BEATTIE C. W., (1994) - A microsatellite linkage map of the porcine genome. *Genetics*, 136: 231-245.

SMITH M. W., SMITH M. H. & BRISBIN I. L. JR., (1980) - Genetic variability and domestication in swine. *J. Mamm.*, 61: 39-45.

TANAKA K., OISHI T., KUROSAWA Y. & SUZUKI S., (1983) - Genetic relationships among several pig populations in East Asia analysed by blood groups and serum protein polymorphisms. *Anim. Blood Grps. Biochem. Genet.*, 14: 191-200.

WATANABE T., HAYASHI Y., KIMURA J., YASUDA Y., SAITOU N., TOMITA T. & OGASAWARA N., (1986) - Pig mitochondrial DNA: polymorphism, restriction map, orientation and sequence data. *Biochem. Genet.*, 24: 385-396.