SOME METHODS FOR ANALYSING GENETIC MARKER DATA IN A BIODIVERSITY SETTING - EXAMPLE OF THE PIGBIODIV DATA

QUELQUES MÉTHODES D'ANALYSE DES DONNÉES DE MARQUEURS GÉNÉTIQUES POUR DES ÉTUDES DE BIODIVERSITÉ - EXEMPLE DES DONNÉES PIGBIODIV

SanCristobal, M.[1], C. Chevalet[1], J.-L. Foulley[2] and L. Ollivier[2]

[1]INRA. Laboratoire de Génétique Cellulaire. 31326 Castanet Tolosan Cedex. France.
[2]INRA. Station de Génétique Quantitative et Appliquée. 78352 Jouy en Josas Cedex. France.

ADDITIONAL KEYWORDS

Biodiversity. Genetic marker. Genetic distance. Pig.

MOTS-CLEFS ADDITIONNELS

Biodiversité. Marqueur génétique. Distance génétique. Porc.

SUMMARY

Biodiversity is an important factor to consider when establishing conservation programmes. The PigBioDiv project aimed to assess the genetic diversity of the European pig by using genetic markers. This chapter presents some methods available for analysing such data. The two marker technologies used in PigBioDiv, namely microsatellites and AFLP (described in chapters 3 and 4 by Groenen *et al.,* 2003 and Plastow *et al.,* 2003), allow a fairly precise evaluation of the within-breed and the between-breed variation. The meanings of the Reynolds and standard Nei genetic distances are given under some population genetics models. Various ways are shown for exploiting the information those distances provide, such as comparisons between distances, tree building and analysis of breed diversity. The need to combine within- and between-breed diversity is emphasised and ways of combining those two components of biodiversity are discussed.

RÉSUMÉ

La biodiversité est un facteur important à considérer quand on établit des programmes de conservation. Le projet PigBioDiv avait pour objectif d'évaluer la diversité génétique du porc européen à l'aide de marqueurs génétiques. Ce chapitre présente des méthodes disponibles pour analyser des données de ce genre. Les deux techniques de marquage utilisées dans Pig BioDiv, microsatellites et AFLP (décrits aux chapitres 3 et 4 pour Groenen *et al.,* 2003 et Plastow *et al.,* 2003), permettent une évaluation assez précise de la variation intra-race et entre races. La signification des distances de Reynolds et Nei standard est donnée pour certains modèles de génétique des populations. On montre diverses façons d'exploiter l'information que ces distances fournissent, telles que des comparaisons entre elles, la construction d'arbres et l'analyse de la diversité des races. La nécessité de combiner les deux composantes intra et entre races de la diversité est soulignée, et diverses façons de les combiner sont discutées.

INTRODUCTION

Biodiversity is a large concept that all countries now have in mind.

Preserving the current diversity of the living material on earth is fundamental for future generations to survive. The general idea is that what is lost cannot be rebuilt. In the case of animals, in the recent past, more and more effective breeding programmes have been implemented, and have led to an emphasis on a few specialised stocks. Consequently, breeds that are less suited to current needs tend to see their numbers decline and to be eventually lost. At the same time, the high selection pressure put on the main breeds leads to a drastic reduction in the number of animals that leave progeny for breeding. Conservation of variation, however, is necessary to meet future agricultural challenges and particularly food needs, as well as to preserve the rich agricultural heritage of the various regions of the world. An overview of the historical background of animal genetic resources and the reasons for concern about their future, both in the developing and developed world, has been presented by Barker (2002). As detailed in these Proceedings (Ollivier, Amigues and Boscher, 2003), the PigBioDiv project aimed to assess the diversity of one particular livestock species (pig) within one particular continent (Europe).

The focus in this chapter will be on the genetic diversity assessed by using genetic markers. Such markers are widely available in the pig as in most farm animal species. Some are (nearly) neutral, some are not. Some are highly polymorphic, others have less variability. In the present project, the neutral genetic background is studied, since it allows the estimation of parameters that are of prime interest for population geneticists. Future evolution of our set of populations can be predicted using these parameters (Nunney, 2000). It must be noted that the variability of neutral markers differs from the variability of markers under selection. Both kinds of diversity, however, are expected to be correlated because of disequilibria generated by random drift or by hitch-hiking effects between neutral and selected linked loci (Bataillon *et al.*, 1996; Slatkin and Wiehe, 1998).

The emphasis in this chapter is on methodology. Tools will be presented for measuring within-breed and between-breed genetic variation and the way to combine both will be discussed. In contrast with previous studies on pigs (e.g. Laval *et al.*, 2000; Martinez *et al.*, 2000; Sun *et al.*, 2002), large-scale strategies are needed in the analyses. An overview of PigBioDiv results can be found in SanCristobal *et al.* (2002). Detailed results will not be given here but full analyses of the project data are in preparation and will be presented elsewhere.

MARKER POLYMORPHISMS

MICROSATELLITES

This is a marker technology widely used in farm animal species. The data on 11 pig populations from the PiGMaP pilot-project (Laval *et al.*, 2000) were combined with those obtained over the 59 breeds of the PigBioDiv project (see chapter 2 of these Proceedings). Meanwhile, the number of loci investigated was increased from 27 in PiGMaP to 50 in PigBioDiv. Microsatellite loci are known to be multi-allelic and co-dominant, and precisely mapped

on the pig genome (see Groenen *et al.*, 2003 in these Proceedings). The technique is based on the evaluation of fragment sizes corresponding to the various alleles, which implies establishing a coding system for loci typed in different laboratories. This was needed for combining PiGMaP and PigBioDiv data for the 27 loci common to both projects. By using a set of 4 control DNA used in both projects a correspondence was established.

Overall, microsatellite polymorphism was quite large over the 70 breeds sampled, since the average number of alleles per locus was 14.5. Within breed, the observed and effective numbers of alleles per locus were 4.5 and 2.7 respectively. The effective number of alleles takes account of uneven allele frequencies: a rare neutral allele will most probably be lost in the next generation, and have a very low weight in the calculation of the effective allele number. This number corresponds to the ideal situation where the allele frequencies are equal, for the same level of heterozygosity. The proportion of monomorphic populations for each locus was low, since only 2 percent of the populations sampled in the present project were monomorphic on average across loci. As a consequence, expected heterozygosity per breed was high, with an overall value of 56 percent.

AFLP

The AFLP technique is described by Plastow *et al.* (2003) in these Proceedings. This technique was only applied in the 59 breeds of the PigBioDiv project, and 148 AFLP loci were typed. Such loci, contrary to microsatellites, are not mapped on the pig genome. They are known to be biallelic and they were here scored dominantly, which meant that only presence or absence of a given band could be scored. The proportion of monomorphic populations for each locus - i.e. when the band was either always present or always absent - was large, since an average over loci of 63 percent of the 59 populations sampled were monomorphic. In contrast with microsatellites, a low expected heterozygosity per breed of 11.6 percent was noted, with 1.4 and 1.2 mean within-breed observed and effective numbers of alleles, respectively.

A question is the comparison of performance of the two types of markers for diversity analyses: numerous but less polymorphic vs polymorphic but less numerous and more expensive. As well, genomic properties are different: microsatellites have no known function, while AFLP may correspond to any region of the genome. Both are a priori assumed neutral.

WITHIN-BREED VARIATION

The analysis of within-breed genetic variation was based on the individual typings reported above. The contrast between the AFLP and microsatellites polymorphisms has already been pointed out. Variation across loci in the percentage of monomorphic populations also differed. It was more pronounced with AFLP (range 3-98 percent) compared to microsatellites (range 0-16 percent). In addition a bimodal

distribution of this percentage over AFLP loci was observed, suggesting differential evolution forces according to the genome region considered.

F_{IS} statistics (Weir and Cockerham, 1984) are commonly used for testing the Hardy-Weinberg equilibrium. The test is based on the comparison between observed and expected heterozygosities (averaged over the loci). Clearly only microsatellite genotypes can be used in such an analysis. An excess of homozygotes ($F_{IS} > 0$) denotes some inbreeding or heterogeneity of the population, i.e. a population composed of several sub-groups (the so-called Wahlund effect). In the reverse, an excess of heterozygotes ($F_{IS} < 0$) can be due to a specific management system aimed at reducing the inbreeding, as sometimes implemented in small local populations.

Considering now a particular (microsatellite) locus, an excess of homozygotes may arise from technical artefacts such as the presence of null alleles. It can easily be shown that, assuming Hardy-Weinberg equilibrium, the heterozygotes deficit (*HD*) is an increasing and rather simple function of the null allele frequency (p_0), since the expected proportion of heterozygotes for the null allele (such genotypes appear as *false* homozygotes) is $2p_0(1-p_0)$ and the proportion of homozygotes (undetected) is p_0^2. Then $HD = 2p_0(1-p_0)/(1-p_0^2) = 2p_0/(1+p_0)$. One microsatellite (S0386) indeed showed a highly positive value of F_{IS}, and a rather consistent within-breed deficit of heterozygotes. A possible explanation would be the presence of a null allele previously detected at this locus in familial studies (Archibald *et al.*, personal communication mentioned page 199 in Laval *et al.*, 2000). The presence of null alleles at one locus out of 50 is expected to have a small influence on averaged F_{IS} statistics, and a negligible effect on further analyses, such as estimation of genetic distances.

Evidence on clustering of individuals within particular breeds may help explain observed F_{IS} statistics. A tree based on pairwise distances between individuals can be built using the UPGMA method. This distance between individuals is an approximate estimate of one minus twice the kinship coefficient (Malécot 1946; Chevalet, 1980). When a population is made out of two sub-groups, such a tree can be drawn to visualise the two groups, and help explain a significantly positive F_{IS} statistics. Individual trees can also detect outliers, which may be due to sampling or label errors.

BETWEEN-BREED VARIATION AND GENETIC DISTANCES

The most useful information for between-breed analysis is derived from allele frequencies. They can be estimated from individual typings. Bulk typing can also be used, and reduce genotyping costs (see Groenen *et al.* (2003) in these Proceedings). Allele frequencies are compared between populations via the computation of a genetic distance. Numerous genetic distances are found in the literature and, going back to the earliest concepts proposed, genetic distances may be seen as serving purposes either of clustering of populations or of study of their evolution (Nei, 1987). They can

be partitioned into groups according to mathematical criteria (e.g. Lefort-Buson and de Vienne, 1985) or population genetics criteria (e.g. Laval *et al.*, 2002).

Models underlying genetic distances

A genetic distance is supposed to tell us something about the genetic differences between two populations. The conservationist approach may see genetic distance as a measure of genetic *differentness* between populations at a given time, without reference to any model supposed to have generated the differences seen. In contrast, the population geneticist approach might well be summarised by saying that *genetic distance is not an abstract, idealized measure of « differentness ». It is an estimate of a parameter of the model which is thought to have generated the differences we see* (Felsenstein, 2000). The model will serve to predict future differences and future diversity, through the understanding of evolutionary forces having led to the current differentiation of the populations under study. In PigBioDiv, two classic measures of genetic distances were retained, namely Reynolds (D_R) and standard Nei (D_S). Each of which is more appropriate for a particular genetic model which will be presented now.

The Reynolds distance (Reynolds *et al.*, 1983) has a mathematically simple expectation under the pure genetic drift model, excluding any admixture as well as mutations. Under such a model D_R increases linearly with time: $E(D_R)= t/2N_e$ (approximately), where t is the number of generations and N_e the effective size of the 2 populations considered. More generally, the harmonic mean of the two populations may be considered. If two populations have effective sizes N_i and N_j respectively, then $E(D_R)= t((1/4N_i) + (1/4N_j))$. In fact, this expectation is the meaningful parameter in a pure drift model: the average variation in the inbreeding coefficient. If the effective sizes of a pair of populations would be available, then the time since their divergence could be derived from an estimate of the Reynolds distance. It is however difficult to estimate accurately effective population sizes, since the distribution of family sizes is needed (Hill, 1972).

The expectation of the standard Nei distance (Nei, 1972) shows a linear increase with time, however under a slightly more complex model, since it assumes that the populations have reached a mutation-drift equilibrium. Then $E(D_S)= 2vt$, where v is the mutation rate per generation and t the number of generations since divergence.

The expectations of the Reynolds distance under mutation-drift equilibrium and of the standard Nei under pure drift are also available (Laval *et al.*, 2002) but these expectations are less simple, and so less attractive. Another important difference between D_R and D_S is that the latter, contrary to the former, is expected to depend on the type of marker considered, particularly the founder heterozygosity H_0. The two distances are approximately related as follows for small distances (see Laval *et al.*, 2002): $E(D_S)= E(D_R)H_0/(1-H_0)$.

Exploitation of genetic distance data

The number of genetic distance estimates increases as the square of

the number *n* of populations (the number of distance figures is *n*(*n*-1)/2). It is obvious therefore that a framework of analysis is needed for such large amounts of data. Three such frameworks may be considered, namely (i) *comparisons* of a given measure of distance across types of markers, or across measures of distances for a given type of marker, with the purpose of throwing some light on possible models of divergence, (ii) drawing *trees*, in view of *phylogeny* inferences (Takezaki and Nei, 1996) or of breed clustering as a visual summary of a large distance matrix, and (iii) analysis of *between-breed diversity*, with an objective of biodiversity management.

Various *comparisons* between distances were briefly discussed by SanCristobal *et al.* (2002) and will be further developed in SanCristobal *et al.* (in preparation) and Plastow *et al.* (in preparation). The incidence of different patterns of marker polymorphism is worth considering. A particular problem arises in dealing with those AFLP loci which are fixed for the same allele in pairs of populations, so-called identically monomorphic (IM) loci. An alternative is either to ignore IM loci, as recommended by Weir (1996) when such loci are rare, or to assume a null distance at IM loci. Because of the high proportion of IM loci in our data, rather different multilocus distances were to be expected according to the option taken. Comparing Reynolds and Nei distances for microsatellites is also of interest, since our data strongly suggested a significant role of mutations in the divergence between the *Meishan* and the European breeds, thus highlighting the different causes of divergence behind those two distances.

Once a pairwise genetic distance matrix is available, it can be summarised and visualised by drawing a *tree*, like a Neighbor-Joining tree (Seitou and Nei, 1987). An example of such a tree derived from PigBioDiv data is given in SanCristobal *et al.* (2002). It is generally admitted that when dealing with breeds of farm animals the interpretation of trees in terms of phylogeny can be misleading (SanCristobal *et al.*, 2002; Weir, 2002). Trees, however, offer useful classification tools, by allowing breeds (or lines) clustering within the whole set analysed. Trees may sometimes reveal unsuspected topologies or unexpected positions of some breeds which may be worth further investigation.

When *analysing between-breed diversity*, individual breed contributions may be derived from any set of distances. In a context of species conservation, Weitzman (1992, 1993) showed how to derive a diversity function *V* from a set of genetic distances in order to evaluate the relative loss of diversity resulting from the extinction of any given species. He also defined a diversity expected after a given period of time, based on the extinction probability of each species. The *marginal diversity* of any species was then obtained as minus the partial derivative of the expected diversity with respect to the extinction probability of the species considered. Weitzman also showed that the algorithm leading to *V* generates a rooted tree which may be interpreted as an evolutionary tree, and whose branch lengths correspond to the diversity lost when the corres-

ponding species goes extinct.

This approach was extended to the situation of livestock breeds diversity by Thaon d'Arnoldi *et al.* (1998), and has served to evaluate the relative merits of endangered breeds when setting conservation programmes. A relative loss of diversity for a breed k may be defined as $V(k)=[V(S)-V(S\backslash k)]/V(S)$, where $V(S)$ is the diversity of the whole set of breeds considered and $V(S\backslash k)$ is the diversity of the set deprived of breed k. The quantity $V(k)$ may also be termed marginal diversity, as it can be shown to be equal to the marginal diversity defined by Weitzman for the particular situation of zero probabilities of extinction. This method has already been used in farm animal species (and particularly in the pig by Laval *et al.*, 2000). In most studies highly unequal contributions of the breeds considered were evidenced. The method, however, becomes computationally demanding when the number of breeds is high. A software has been developed in the framework of PigBio Div in order to obviate this difficulty (see Derban *et al.*, 2002).

Other approaches to diversity analysis have been proposed. Petit *et al.* (1998) presented a method to evaluate the contributions of individual populations to the total diversity on the basis of genetic markers. They showed how to derive the contribution of each population to the total diversity by using the classical gene diversity parameters of Nei (1977). These population contributions may in turn be partitioned into two components, a contribution to the between-population diversity and a contribution to the average within-population diversity, these two components adding up to the total contribution. A similar method, based on the concept of co-ancestry, instead of gene diversity, was proposed by Caballero and Toro (2002). They showed that the ranking of breeds based on total diversity can be quite different from the ranking based on between-breed diversity and warned against conclusions which might be drawn if only the latter were considered. Hence the need to examine how to best combine those two components of diversity.

COMBINING WITHIN- AND BETWEEN-BREED VARIATION

Ranking breeds for conservation purposes is a difficult task. The main issue is how to deal with small populations. When a breed is made of very few breeding animals, it becomes endangered and may eventually become extinct. This situation, however, is not in itself a sufficient reason for giving conservation priority to the most endangered breeds, since this kind of reasoning might lead to give a conservation priority to a (nearly) inbred line, even though it might contain no specific allele richness. In contrast, large populations generally have a higher potential for future genetic improvement due to their higher internal genetic variability.

A proper biodiversity analysis should combine within- and between-breed aspects. A global structural index is provided by the F_{ST} statistics of Wright (Weir and Cockerham, 1984). An analysis of variance, based on genotypes and taking the factor *breed* into account, splits the total variance into a within-breed and a between-breed

variance (Weir and Cockerham, 1984). When individual breed contributions to marker diversity are evaluated using the method of Petit *et al.* (1998), the total breed contributions to gene diversity may be shown to be the result of weighting the between-breed contribution (to the total between-breed diversity) and the within-breed contribution (to the mean within-breed diversity) by F_{ST} and 1-F_{ST} respectively. A similar definition applies to the total diversities, based on the concept of kinship, which have been proposed by Caballero and Toro (2002) and Eding *et al.* (2002).

Different weights, however, may be desirable in some contexts, as suggested for instance by Chaiwong and Kinghorn (1999). Using the data of Laval *et al.* (2000), Ollivier and Foulley (2002) gave an illustration of how the two components of diversity might be combined in a flexible manner. Depending on the context, different weights will apply. As emphasised by Barker (2002), setting priorities should be considered separately according to whether the choice of breeds is for genetic improvement, for comparative evaluation of breeds or for conservation of endangered breeds. On the other hand, and in a long-time perspective, one should consider the economic advantage of being able to cope with changing production-marketing systems (Smith, 1986), which would require maintaining diverse breeding stocks and so giving more emphasis to between-breed variation.

DISCUSSION AND CONCLUSIONS

Two types of markers were considered in PigBioDiv, microsatellites and AFLP, which are rather different in polymorphism as well as in ease and costs of development. Both are assumed to be neutral. Though the project focussed on inferences on populations, based on genetic marker data, attention has also been given here to the dual question of analysing marker behaviour, given the populations at hand. Specific behaviours of some markers could be tested, which could provide signatures for the relative strength of drift, mutation or selection, as possible sources of diversity, or even for technical artefacts such as null alleles. In practice, the data were directly downloaded from the project database (see Russell *et al.* 2003) and for most purposes existing software were used (Becker *et al.*, 1988; Belkhir *et al.*, 1998; Felsenstein, 2000). However, because of the large variations in the number of microsatellites typed in each breed (for reasons given in Groenen *et al.*, 2003), specific programs had to be used in order to allow for missing data.

The emphasis has been on the genetic models underlying the two genetic distances analysed. Such models are useful for trying to understand the evolutionary forces behind the present diversity situation, and for predicting future evolution of diversity (see Nunney, 2000). One should keep in mind that the evolutionary processes are expected to be quite different in species and natural populations compared to breeds within species under domestication pressure, not to mention the time scale dimension of the comparison.

We have not considered here purely descriptive and model-free statistical

tools, which are available in large numbers (e g see Laloë *et al.*, 2002). Graphical data analysis such as Principal Component Analysis is another useful tool, which may be used for instance for visualising the ratio of the between- to the within- breed variations.

Another aspect not addressed in this paper is the assignment of individuals to their breed that can be performed using the genetic marker data generated by the project. It can be noted that the power of an assignment test depends on the differentiation of the breeds and thus again on the balance of within- and between- breed variation.

In conclusion, it is recognised that the analysis of molecular markers for biodiversity purposes is a complex process. The need to combine the within-breed and the between-breed components of diversity has been emphasised. As we have seen, several methods for measuring those components are available (including the analysis of allelic richness not considered here). Interrelations among various measures of genetic diversity also need further study (Barker, 2001), and the PigBioDiv project has created a significant resource to undertake these types of studies.

A marker-based description of genetic diversity of a set of breeds needs a variety of statistical tools, some of which were presented here. This stage may not need any theory on previous evolution processes. In contrast, the prediction of future genetic diversity requires the knowledge and understanding of evolutionary forces acting on the set (or subsets) of breeds, at various time horizons. Conservation decisions must take these predictions into account, as well as other aspects (Ruane, 2000). One may also need to take into account that the future direction of animal breeding may change and evolve, as indeed it did in the recent past (Hervieu and Bonnemaire, 2002).

REFERENCES

Barker, J.S.F. 2001. Conservation and management of genetic diversity: a domestic animal perspective. *Can. J. For. Res.*, 31: 588-595.

Barker, J.S.F. 2002. Relevance of animal genetic resources and differences to the plant sector. In: Groeneveld E. and Glodek P. (Eds), Animal Breeding and Animal Genetic Resources. Federal Agricultural Research Centre (FAL), Mariensee and Institute of Animal Genetics, Göttingen, Braunschweig, Germany, 15-21.

Bataillon, T.M., J.L. David and D.J. Schoen. 1996. Neutral genetic markers and conservation genetics: simulated germplasm collections. *Genetics*, 144: 409-417.

Becker, R.A., J.M. Chambers and A.R. Wilks. 1988. The New S Language. Wadsworth and Brooks/Cole, Pacific Grove, California.

Belkhir, K., P. Borsa, J. Goudet, L. Chikhi et F. Bonhomme. 1998. GENETIX, logiciel sous WindowsTM pour la génétique des populations, Laboratoire génome et populations, CNRS UPR 9060, Université de Montpellier II, Montpellier, France.

Caballero, A. and M.A. Toro. 2002. Analysis of genetic diversity for the management of conserved subdivided populations. *Conserv. Genet.*, 3: 289-299.

Chaiwong, N. and B.P. Kinghorn. 1999. Use of genetic markers to aid conservation decisions for groups of rare domestic breeds. *Proc.*

Assoc. Advmt. Anim. Breed. Genet., 13: 365-368.

Chevalet, C. 1980. Calcul des coefficients d'identité, inégalités et distances génétiques. In: Legay J.M.(Eds), Biométrie et Génétique, Soc. Fr. de Biométrie, INRA, Paris, 42-49.

Derban, S., J.-L. Foulley and L. Ollivier. 2002. WEITZPRO: a software for analysing genetic diversity. INRA, Paris.

Eding, H., R.P.M.A. Crooijmans, M.A.M. Groenen and T.H.E. Meuwissen. 2002. Assessing the contribution of breeds to genetic diversity in conservation schemes. *Genet. Sel. Evol.*, 34: 613-633.

Felsenstein, J. 2000. PHYLIP (Phylogeny Inference Package), Department of Genome Science, University of Washington, Seattle.

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

Hervieu, B. and J. Bonnemaire. 2002. Social demand and concerns about animal breeding: the animal, a social construct. *World Congr. Genet. Appl. Livest. Prod.,* 29: 3-13.

Hill, W.G. 1972. Effective size of populations with overlapping generations. *Theor. Pop. Biol.*, 3: 278-289.

Laloë, D., K. Moazami-Goudarzi and D. Chessel. 2002. Contribution of individual markers to the analysis of the relationships among breeds by correspondence analysis. *World Congr. Genet. Appl. Livest. Prod.*, 33: 493-496.

Laval, G., N. Iannuccelli, C. Legault, D. Milan, M.A.M. Groenen, E. Giuffra, L. Andersson, P.H. Nissen, C.B. Joergensen, P. Beeckmann, H. Geldermann, J.-L. Foulley, C. Chevalet and L. Ollivier. 2000. Genetic diversity of eleven European pig breeds. *Genet. Sel. Evol.*, 32: 187-203.

Laval, G., M. SanCristobal and C. Chevalet. 2002. Measuring genetic distances between breeds: use of some distances in various short term evolution models. *Genet. Sel. Evol.*, 34: 481-507.

Lefort-Buson, M. et D. de Vienne. (Eds). 1985. Les distances génétiques. Estimations et applications. INRA, Paris, France.

Malécot, G. 1946. La consanguinité dans une population limitée. *C.R. Acad. Sci. Paris*, 222: 841-843.

Martínez, A.M., J.V. Delgado, A. Rodero and J.L. Vega-Pla. 2000. Genetic structure of the Iberian pig breed using microsatellites. *Anim. Genet.*, 31: 295-301.

Nei, M. 1972. Genetic distances between populations. *Am. Nat.*, 106: 283-292.

Nei, M. 1977. Statistics and analysis of gene diversity in subdivided populations. *Ann. Hum. Genet.*, 41: 221-233.

Nei, M. 1987. Molecular Evolutionary Genetics. Columbia University Press, New-York.

Nunney, L. 2000. The limits to knowledge in conservation genetics. The value of effective population size. *Evol. Biol.*, 32: 179-194.

Ollivier, L. and J.-L. Foulley. 2002. Some suggestions on how to preserve both within- and between-breed genetic diversity. Book of Abstracts of the 53rd Annual Meeting of the European Association for Animal Production, Cairo, Egypt, 1-4 September 2002, page 2.

Ollivier, L., Y. Amigues and M.Y. Boscher. 2003. An EC-funded project on characterisation of genetic variation in the European pig. Objectives, organisation, breed sampling, DNA preparation and circulation. *Arch. Zootec.,* 52: 137-144.

Petit, R.J., A. El Mousadik and O. Pons. 1998. Identifying populations for conservation on the basis of genetic markers. *Conserv. Biol.*, 12: 844-855.

Plastow, G., K. Siggens, M. Bagga, B. Brugmans, H. Heuven and J. Peleman. 2003. Utilization of AFLP® for genetic distance analysis in pigs. *Arch. Zootec.,* 52: 157-164.

Reynolds, J., B.S. Weir and C.C. Cockerham. 1983. Estimation of the coancestry coefficient: basis for a short-term genetic distance.

Genetics, 105: 767-779.

Ruane, J. 2000. A framework for prioritizing domestic animal breeds for conservation purposes at the national level: a Norwegian case study. *Conserv. Biol.*, 14: 1385-1393.

Russell, G.A., A.L. Archibald, C.S. Haley and A.S. Law. 2003. The pig genetic diversity database and the WWW. *Arch. Zootec.,* 52: 165-172.

SanCristobal, M., C. Chevalet, C.S. Haley, G. Russell, G. Plastow, K. Siggens, M. Bagga, M.A.M. Groenen, Y. Amigues, K. Hammond, G. Laval, M.-Y. Bosher, D. Milan, A. Law, E. Fimland, R. Davoli, V. Russo, G. Gandini, A. Archibald, J.V. Delgado, M. Ramos, C. Désautés, L. Alderson, P. Glodeck, J.NN Meyer, J.-L. Foulley and L. Ollivier. 2002. Genetic diversity in pigs - Preliminary results on 58 European breeds and lines. World Congr. *Genet. Appl. Livest. Prod.*, 33: 523-528.

Seitou, N. and M. Nei. 1987. Reconstructing phylogenetic trees. *Mol. Biol. Evol.*, 4: 406-425.

Slatkin, M. and T. Wiehe. 1998. Genetic hitch-hiking in a subdivided population. *Genet. Res.*, 71: 155-160.

Smith, C. 1986. Variety of breeding stock for the production-marketing range, and for flexibility and uncertainty. *World Congr. Genet. Appl. Livest. Prod.*, 10: 14-18.

Sun, F., Y. Zhang, Z. Wang and S. Yang. 2002. Study on the genetic relationship among 18 Chinese local pig breeds using microsatellite DNA markers. *World Congr. Genet. Appl. Livest. Prod.*, 33: 529-532.

Takezaki, N. and M. Nei. 1996. Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics*, 144: 389-399.

Thaon, C. d'Arnoldi, J.L. Foulley and L. Ollivier. 1998. An overview of the Weitzman approach to diversity. *Genet. Sel. Evol.*, 30: 149-161.

Weir, B.S. 1996. Genetic Data Analysis II. Sinauer Associates, Sunderland, Massachusetts.

Weir, B.S. 2002. Genetic distances and population histories: traps and bootstraps. *World Congr. Genet. Appl. Livest. Prod.*, 33: 467-472.

Weir, B.S. and C.C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. *Evolution*, 38: 1358-1370.

Weitzman, M.L. 1992. On diversity. *Quart. J. Econ.*, 107: 363-405.

Weitzman, M.L. 1993. What to preserve? An application of diversity theory to crane conservation. *Quart. J. Econ.*, 108: 157-183.