

Genetic Diversity Analysis Using Lowly Polymorphic Dominant Markers: The Example of AFLP in Pigs

J.-L. FOULLEY, M. G. M. VAN SCHRIEK, L. ALDERSON, Y. AMIGUES, M. BAGGA, M.-Y. BOSCHER, B. BRUGMANS, R. CARDELLINO, R. DAVOLI, J. V. DELGADO, E. FIMLAND, G. C. GANDINI, P. GLODEK, M. A. M. GROENEN, K. HAMMOND, B. HARLIZIUS, H. HEUVEN, R. JOOSTEN, A. M. MARTINEZ, D. MATASSINO, J.-N. MEYER, J. PELEMAN, A. M. RAMOS, A. P. RATTINK, V. RUSSO, K. W. SIGGENS, J. L. VEGA-PLA, AND L. OLLIVIER

From INRA, Station de Génétique Quantitative et Appliquée, 78352 Jouy-en-Josas cedex, France (Foulley and Ollivier); Keygene N.V., PO Box 216, 6700 AE Wageningen, The Netherlands (van Schriek, Brugmans, Heuven, and Peleman); Rare Breeds Survival Trust, 6 Harnage, Shrewsbury, Shropshire SY5 6EJ, United Kingdom (Alderson); Labogena, Domaine de Vilvert, 78352 Jouy-en-Josas cedex, France (Amigues and Boscher); Sygen International, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, United Kingdom (Bagga and Siggins); the Animal Genetic Resources Group, FAO, Viale delle Terme di Caracalla, Roma 00100, Italy (Cardellino and Hammond); DIPROVAL, Università di Bologna, Via Rosselli 107, 42100 Coviolo-Reggio Emilia, Italy (Davoli and Russo); the Facultad de Veterinaria, Universidad de Córdoba, Avda Medina Azahara 9, 14005 Córdoba, Spain (Delgado and Martinez); Nordic Gene Bank Farm Animals, Pb 5003, 1432 Aas, Norway (Fimland); the Department VSA, Università degli Studi di Milano, Via Celoria 10, 20133 Milano, Italy (Gandini); the Animal Genetics Institute, Albrecht-Thaer-Weg 3, 37075 Göttingen, Germany (Glodek and Meyer); Wageningen University, PO Box 338, 6700 AH Wageningen, The Netherlands (Groenen, Harlizius, Heuven, Joosten, and Rattink); ConSDABI-NFP.I-FAO, Centro di Genomica e di Proteomica per la Qualità e per l'Eccellenza alimentare, Località Piano Cappelle, 82020 Benevento, Italy (Matassino); Universidade de Tras-os-Montes e Alto Douro, 5001 Vila Real Codex, Portugal (Ramos); and Laboratorio de Genética Molecular, FESSCR, Apartado Oficial Sucursal 2, 14071 Córdoba, Spain (Vega-Pla). B. Brugmans is now at EasyGene Nieuw Kanaal 7E, 6709 PA Wageningen, The Netherlands.

Address correspondence to L. Ollivier at the address above, or e-mail: louis.ollivier@jouy.inra.fr.

Abstract

DNA markers are commonly used for large-scale evaluation of genetic diversity in farm animals, as a component of the management of animal genetic resources. AFLP markers are useful for such studies as they can be generated relatively simply; however, challenges in analysis arise from their dominant scoring and the low level of polymorphism of some markers. This paper describes the results obtained with a set of AFLP markers in a study of 59 pig breeds. AFLP fingerprints were generated using four primer combinations (PC), yielding a total of 148 marker loci, and average harmonic mean of breed sample size was 37.3. The average proportion of monomorphic populations was 63% (range across loci: 3%–98%). The moment-based method of Hill and Weir (2004, *Mol Ecol* 13:895–908) was applied to estimate gene frequencies, gene diversity (F_{ST}), and Reynolds genetic distances. A highly significant average F_{ST} of 0.11 was estimated, together with highly significant PC effects on gene diversity. The variance of F_{ST} across loci also significantly exceeded the variance expected under the hypothesis of AFLP neutrality, strongly suggesting the sensitivity of AFLP to selection or other forces. Moment estimates were compared to estimates derived from the square root estimation of gene frequency, as currently applied for dominant markers, and the biases incurred in the latter method were evaluated. The paper discusses the hypotheses underlying the moment estimations and various issues relating to the biallelic, dominant, and lowly polymorphic nature of this set of AFLP markers and to their use as compared to microsatellites for measuring genetic diversity.

The current situation of pig breeds in Europe is marked by the coexistence of many small populations of local breeds and a relatively small number of intensively selected breeds (e.g., Large White, Landrace, and Piétrain). This means that it is of particular interest to assess the level of genetic diversity that is present within Europe. An important aspect is the identification and conservation of breeds that may have special characteristics and gene variants (alleles) of future utility and which may have been lost from the improved breeds developed over the last century.

The development of efficient methods to characterize DNA polymorphism now allows such studies to be realized on a significant scale. This study was undertaken to characterize genetic diversity in a large number of pig populations using a large number of markers as recommended by the guidelines established by the FAO (1998). The general objectives of the study and its organization have been described elsewhere (Ollivier et al. 2003). Two marker systems were utilized, microsatellites and AFLP. Both markers have been quite extensively used on several farm animal species, and an analysis of AFLP markers in pigs was recently described (SanCristobal et al. in press b). The present paper builds on this analysis using the approach recently developed by Hill and Weir (2004) and highlights the specific problems arising in the genetic analysis of this dominant and lowly polymorphic marker as compared to the standard methods applying to codominant and highly polymorphic markers such as microsatellites.

Materials and Methods

Populations

The sampling objective was 50 individuals from each of 58 populations representing the variety of populations present in Europe and 1 Chinese breed. The European populations included local breeds (23), national varieties of international breeds (14), and commercial lines (21). A detailed list of the populations analyzed and numbers typed is given in Table 1. More details of the populations can be found at <http://databases.roslin.ac.uk/pigbiodiv>.

AFLP Markers

The AFLP marker data were generated by the technique of Vos et al. (1995) using four *EcoRI/TaqI* primer combinations (PCs) chosen from preliminary analyses (Plastow et al. 2003). The PCs were *EcoRI* + AAG/*TaqI* + CAA (E33/T47: 28 markers), *EcoRI* + AAG/*TaqI* + CTT (E33/T62: 43 markers), *EcoRI* + ACA/*TaqI* + CAC (E35/T48: 39 markers), and *EcoRI* + ATT/*TaqI* + CAC (E46/T48: 38 markers), totaling 148 markers. The range of the number fingerprinted per line was 16–46 (a total of 60 being obtained by pooling the two samples of the Chinese Meishan). The harmonic mean of number fingerprinted per breed/line varied very little across loci, from 36.9 to 37.5, with an average of 37.3.

Within-Breed Diversity and Breed Differentiation

The usual procedures with dominant markers are based on estimating allele frequencies under the assumption of Hardy-Weinberg equilibrium. At any locus, the frequency of the allele “absence of a band” can thus be calculated as the square root of the frequency of animals not showing that band. Calling \hat{p} and \bar{p} this estimate and its average over all populations, respectively, first-order estimates of expected heterozygosity, H_e , and θ (or F_{ST}) can be obtained simply as $H_e = 2\hat{p}(1 - \hat{p})$ and $\theta = V(\hat{p})/\hat{p}(1 - \hat{p})$, as detailed by Lynch and Milligan (1994). Recently, Hill and Weir (2004) have investigated a more robust moment-based method using the observed mean \bar{P} and variance s^2 of the recessive genotype frequency P (i.e., absence of band) at any given locus. Estimates of p (\hat{p}) and θ ($\hat{\theta}$) are obtained by solving the following equations:

$$E(P) = \bar{P}, \quad (1)$$

$$E(V_B + V_W/n_h) = s^2, \quad (2)$$

where V_B and V_W are the between- and within-population variance components of allele frequency and n_h the harmonic mean of the number of individuals scored in each breed for the locus considered. The left-hand sides of Equations 1 and 2 can be expressed as functions of the unknowns p and θ . The expressions to be used are $E(P) = p^2 + \theta p(1 - p)$ and the V_B and V_W functions of p and θ given by Hill and Weir (2004), depending on the p distribution assumed. Here we assumed a pure drift distribution of the recessive allele frequencies p among populations, and used equations 6 and 7 of Hill and Weir (2004). Equations 1 and 2 define a nonlinear system in \hat{p} and $\hat{\theta}$ which can be solved numerically, e.g., via the Newton-Raphson algorithm (see appendix II, pp. 907–908 of Hill and Weir 2004). This procedure provides moment estimates of \hat{p} and $\hat{\theta}$ as well as estimates of asymptotic sampling variances. As a by-product, an unbiased estimate of the average within-breed expected heterozygosity can be obtained as follows:

$$H_e = E[2p(1 - p)] = 2(\hat{p} - \bar{P}). \quad (3)$$

Genetic Distances

As previously for θ , genetic distances can be computed by using the square root estimation of allele frequency. This procedure has been implemented by SanCristobal et al. (2002) on the genetic distances of Reynolds (Reynolds et al. 1983) and Nei (1978). In the biallelic case, \hat{p}_{il} and \hat{p}_{jl} being the square root allele frequencies of populations i and j at locus l , multi-locus Reynolds estimator may be taken either as the arithmetic mean of single-locus distance estimates D_{Rl} , namely,

$$D_{RA} = \frac{1}{L} \sum_{l=1}^L D_{Rl}, \quad \text{where } D_{Rl} = \frac{(\hat{p}_{il} - \hat{p}_{jl})^2}{\hat{p}_{il} + \hat{p}_{jl} - 2\hat{p}_{il}\hat{p}_{jl}}, \quad (4)$$

or as ratios of means as in the PHYLIP package (Felsenstein 2004)

$$D_{RP} = \frac{\sum_{l=1}^L (\hat{p}_{il} - \hat{p}_{jl})^2}{\sum_{l=1}^L (\hat{p}_{il} + \hat{p}_{jl} - 2\hat{p}_{il}\hat{p}_{jl})}. \quad (5)$$

When the l locus is identically monomorphic (IM) in populations i and j , i.e., when the two populations show the same

Table 1. Breeds sampled in the PigBioDiv program

Code ^a	Country	Name (Company) ^b	Category ^c	Average number typed
CZPR01	Czech Republic	Presticke	L	40.5
DEAS01	Germany	Angler Sattelschwein	L	45.4
DEBB01	Germany	Bunte Bentheimer	L	44.7
DEDU03	Germany	Duroc line (S)	C	44.7
DEHA02	Germany	Hampshire line (BHZIP)	C	45.0
DELR14	Germany	Landrace line (BHZIP)	C	45.8
DELW02	Germany	German Large White	I	44.3
DELW10	Germany	Large White line (BHZIP)	C	45.8
DEMA01	Germany	Mangalica	L	28.7
DEPI03	Germany	German Piétrain	I	45.3
DKLR04	Denmark	Danish Landrace (contemporary)	I	43.3
DKLR05	Denmark	Danish Landrace (1970)	I	27.8
ESMJ01	Spain	Manchado de Jabugo	L	34.5
ESNC01	Spain	Negro Canario	L	16.8
ESNI01	Spain	Negro Iberico	L	45.3
ESRE01	Spain	Retinto	L	45.6
FILR06	Finland	Finnish Landrace	I	42.3
FRCR01	France	Créole (Guadeloupe)	L	38.0
FRDR01	France	DRB synthetic line (SCAPAAG)	C	39.0
FRLA01	France	Laconie synthetic line (PAL)	C	41.0
FRLR01	France	French Landrace	I	43.7
FRLR13	France	Landrace line (FH)	C	44.5
FRLW01	France	French Large White (dam line)	I	39.8
FRLW08	France	Large White line (FH)	C	26.7
FRLW09	France	Large White line (PAL)	C	41.0
FRLW12	France	French Large White (sire line)	I	20.4
FRPI02	France	French Piétrain	I	42.2
FRPI05	France	Piétrain line (FH)	C	20.0
FRTM01	France	Tia Meslan synthetic line (PAL)	C	21.8
GBBK01	United Kingdom	Berkshire	L	46.0
GBBL01	United Kingdom	British Lop	L	31.9
GBBS01	United Kingdom	British Saddleback	L	41.7
GBDU02	United Kingdom	Duroc line (PIC)	C	45.7
GBGO01	United Kingdom	Gloucester Old Spots	L	44.7
GBHA01	United Kingdom	Hampshire line (PIC)	C	45.5
GBLB01	United Kingdom	Large Black	L	46.0
GBLE01	United Kingdom	Leicoma synthetic line (PIC)	C	46.0
GBLR10	United Kingdom	Landrace line (PIC)	C	44.2
GBLR11	United Kingdom	Landrace line (PIC)	C	45.8
GBLR12	United Kingdom	Landrace line (PIC)	C	45.4
GBLW05	United Kingdom	Large White line (PIC)	C	46.0
GBLW06	United Kingdom	Large White line (PIC)	C	43.2
GBLW07	United Kingdom	Large White line (PIC)	C	46.0
GBMW01	United Kingdom	Middle White	L	36.7
GBPI04	United Kingdom	Piétrain line (PIC)	C	45.5
GBT A01	United Kingdom	Tamworth	L	40.7
ISLR09	Iceland	Icelandic Landrace	I	32.8
ITCA01	Italy	Calabrese	L	19.0
ITCS01	Italy	Cinta Senese	L	29.5
ITCT01	Italy	Casertana	L	27.7
ITDU01	Italy	Italian Duroc	I	45.7
ITLR03	Italy	Italian Landrace	I	44.5
ITLW03	Italy	Italian Large White	I	44.3
ITNS01	Italy	Nera Siciliana	L	43.7
NOLR08	Norway	Norwegian Landrace	I	45.4
PLPU01	Poland	Pulawska Spots	L	43.9
PTBI01	Portugal	Bisaro	L	45.7
SELS01	Sweden	Linderödssvin	L	33.7
CNMS01	China	Meishan	Imported	59.6
Total	14 countries	59 breeds and lines	23 L/14 I/21 C 1 imported	

^a The breed code is the concatenation of a two-letter country code, a two-letter breed code, and a two-digit count. For example, ISLR09 is the Icelandic Landrace, the 9th Landrace population of the project. Synthetic lines are designated by DR, LA, LE, and TM. The codes in bold are 30 populations representing 24 L or C populations and one population from each of the six main microsatellite-based clusters DU, HA, LR, LW, PI, and Iberian. These 30 populations are assumed to approximate a starlike pattern of phylogeny.

^b Companies are BundesHybridZuchtProgram (BHZIP), France Hybrides (FH), Pen Ar Lan (PAL), Pig Improvement Company (PIC), Schaumann (S), and Société Coopérative Agricole pour l'Assainissement et l'Amélioration Génétique du Cheptel Porcin (SCAPAAG).

^c Breed categories are local breed (L), national variety of international breed (I), and purebred or synthetic commercial line (C).

phenotype (either total presence or total absence of band), Equation 4 is undetermined because both numerator and denominator equal zero. In such cases we made $D_{Rl} = 0$. Corrections of D_{Rl} for sample size were made as in Weir (1996, p. 195), by subtracting $[\hat{p}_{il} + \hat{p}_{jl} - (P_{il} + P_{jl})]/(2n_{hl} - 1)$ from the locus numerators of Equations 4 and 5, where \hat{p} , P , and n_h are as defined previously. For consistency with the zero distance assumed for IM loci, the negative D_{Rl} were also set to zero. The standard distance of Nei (D_S) is zero for IM loci, and the usual multilocus estimate could then be used. Negative D_S were set to zero, as recommended by Nei (1978).

Hill and Weir (2004) also extended their moment method to estimate the Reynolds genetic distances using the squared difference of genotype frequencies P_i and P_j between populations i and j . Their reasoning relies on the following expectation: $E(P_i - P_j)^2 = 2(V_B + V_W/n_h)$, where V_B and V_W are the components of variance of gene frequency, defined above for Equations 1 and 2 and n_h the harmonic mean of the numbers of individuals scored in each population. Thus, in order to construct an unbiased moment estimator, the difference $(P_i - P_j)^2$ should be corrected for the within-population sampling variance V_W/n_h . We used here the average V_W over all populations expressed as a function of the common p and θ under the same model of pure drift as in Equation 2. Letting $f(\theta) = V_B$, where θ here is the specific value for the pair considered, and $C = 0.5[(P_i - P_j)^2 - V_W/n_h]$, the nonlinear system to solve can be written as $f(\theta) = C$, or, alternatively, $g(\theta, C) = 0$, where $g(\theta, C) = C - f(\theta)$. Solutions were obtained iteratively from step n to step $n + 1$ via the Newton-Raphson algorithm, as $\theta^{(n+1)} = \theta^{(n)} - g(\theta^{(n)}, C)/g'(\theta^{(n)}, C)$, where $g'(\theta, C)$ is the derivative of g with respect to θ . The solution in θ was taken as the moment estimate D_{Rl} of the Reynolds distance between populations i and j for locus l , and the arithmetic mean of D_{Rl} was taken as the multilocus moment estimate D_{RM} , wherefrom a Neighbor-Joining tree (Saitou and Nei 1987) was drawn. The sampling variance of D_{RM} is given by Hill and Weir (2004) as a function of the number of loci L and the distances, allele frequencies, and harmonic means of sample sizes at each locus (see their equation 21).

Similarity Index

Various measures of breed similarity, needing no genetic model, are available. We retained the simple matching index between populations i and j at locus l

$$SM_l = P_{il}P_{jl} + (1 - P_{il})(1 - P_{jl}), \quad (6)$$

where P_{il} and P_{jl} are the frequencies of recessive genotypes at the locus considered, which can be averaged over the L loci.

Results

Polymorphism of AFLP Bands

The amount of polymorphism observed for the AFLP bands varied considerably because the proportion of monomorphic populations for a specific band ranged from 3% to 98% according to the locus (Figure 1a) (SanCristobal et al.

submitted). Overall, the percentage of monomorphic AFLP bands per population was high (mean 63%, range 41%–83%). Monomorphism, however, includes complete absence as well as complete presence of a band, as shown by the average within-breed proportion of recessive homozygotes plotted against the proportion of monomorphic populations in Figure 1a. Figure 1b shows an even and slightly asymmetric U-shaped distribution of the recessive genotype frequency.

Within-Breed Diversity and Breed Differentiation

Estimates of recessive gene frequency (p) and $F_{ST}(\theta)$ were obtained by the iterative method indicated above. All 148 F_{ST} estimates were obtained after a maximum of seven iterations and were in the parameter space 0–1, except at one locus, for which we used the Lynch and Milligan (1994) procedure. Table 2 shows the average within-breed parameters for each PC and the biases involved when square roots of allele frequency were used. A highly significant average F_{ST} of 0.11 and a highly significant heterogeneity of F_{ST} among the four PCs were evidenced. Table 2 also shows that expected heterozygosity varied among PCs, and biases were considerably larger for the PC with the lowest allele frequency. F_{ST} and H_e appeared to vary independently among PCs, as also shown by the low overall correlation (–0.15) between the two parameters.

Square Root–Based Genetic Distances

As shown by SanCristobal et al. (2002), the Reynolds distances differed depending on the multilocus estimator used. The arithmetic mean estimator D_{RA} was markedly lower (average 0.10) than the PHYLIP estimator D_{RP} (average 0.27). This peculiar situation is due to the high proportion of AFLP loci being IM. The standard Nei distance showed a higher correlation with D_{RA} than with D_{RP} ($r = 0.94$ versus 0.84 in Table 3) as expected because the zero distances of IM loci enter both D_{RA} and the multilocus Nei distance.

Moment Estimates of Reynolds Distances and Comparison To Other Genetic Distances and Similarity Between Breeds

D_{RM} estimates were made equal to zero in case of negative value of C , which mainly occurred for IM loci, because then $(P_i - P_j)^2 = 0$, and made equal to one in the few cases where $C \geq p(1 - p)$. The average of the 1,711 D_{RM} was 0.12 (range 0.03–0.32), very close to the average F_{ST} . An approximate standard error of individual distances of 0.016 was derived from Hill and Weir (2004), taking 0.11, 0.47, and 37.3 as the average distance, allele frequency, and harmonic mean of sample size, respectively.

A comparison of D_{RM} to the other three genetic distances and the similarity index is presented in Table 3 (all five matrices are available at <http://databases.roslin.ac.uk/pigbioidiv/publications.html>). A much higher correlation of D_{RM} with D_{RA} and D_S than with D_{RP} (0.94 and 0.91 versus 0.73) was observed, reflecting the exclusion of IM loci in the PHYLIP algorithm. The same reason probably explains the particularly

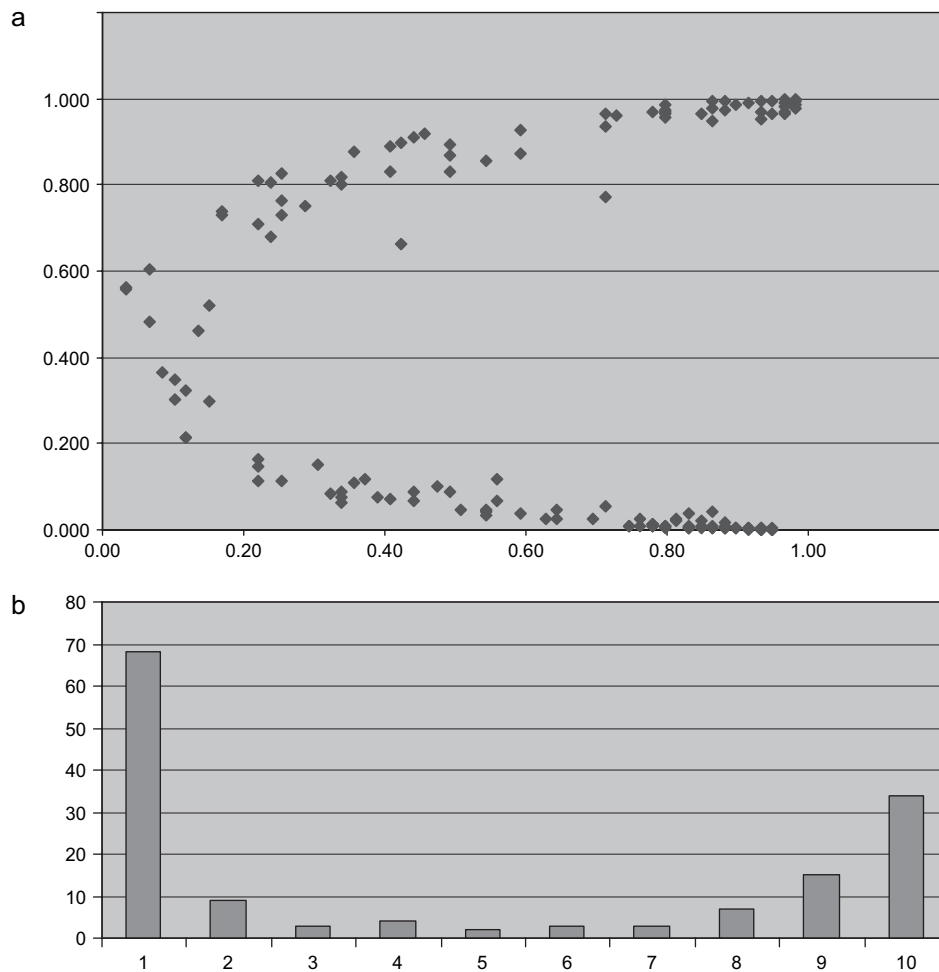


Figure 1. AFLP polymorphism. (a) Average within-breed frequency of the recessive genotype (range 0.001–0.998) versus proportion of monomorphic populations (0.034–0.983) over 148 AFLP loci. (b) Distribution of average within-breed recessive homozygote frequency over 148 AFLP loci. Ordinate, number of loci; abscissa, intervals of 10 percentage points of recessive genotype frequency (1 = 0–10, 2 = 10–20, ..., 10 = 90–100).

Table 2. Moment estimates of the within-breed parameters and F_{ST} values per PC, derived from mean and variance of recessive genotype frequency (P)

PC	Number of loci	p^a	H_e^a	F_{ST}^b	Standard error ^c of mean F_{ST}
E33/T47	28	0.69 (2.5)	0.145 (5.5)	0.091 (13.1)	0.019a
E33/T62	43	0.21 (20.1)	0.098 (16.3)	0.094 (66.9)	0.019a
E35/T48	39	0.49 (3.5)	0.112 (4.5)	0.138 (17.3)	0.020ab
E46/T48	38	0.57 (2.7)	0.141 (2.1)	0.165 (17.4)	0.014b
Total of 59 breeds	148	0.47 (10.7)	0.124 (8.1)	0.114 (31.0)	0.009
Subsample of 30 breeds ^d	137	0.47 (11.6)	0.124 (8.1)	0.078 (42.2)	0.005

p , Within-breed frequency of recessive allele (absence of band); H_e , within-breed expected heterozygosity $2(p - P)$; F_{ST} , weighted average of moment estimate θ (or Wright's F_{ST}).

^a In parentheses negative bias of the square root estimate (%).

^b In parentheses positive bias of the square root estimate (%).

^c Based on average asymptotic sampling variances: values with no letter in common correspond to significantly different F_{ST} at $P < .01$ (t test).

^d Populations shown in bold in Table 1, assumed to exhibit a starlike pattern of phylogeny.

Table 3. Mean values of and correlations between genetic distances and similarity index among the 59 breeds analyzed

	D_{RM}	D_{RA}	D_{RP}	D_S	$1 - SM$
Mean	0.123	0.097	0.273	0.052	0.137
D_{RM}	1	0.935	0.730	0.908	0.887
D_{RA}	—	1	0.774	0.957	0.867
D_{RP}	—	—	1	0.838	0.583
D_S	—	—	—	1	0.891

D_{RM} , moment estimate of Reynolds distance (Hill and Weir 2004); D_{RA} , arithmetic mean of single-locus Reynolds distances (Reynolds et al. 1983) based on square root estimates of allele frequency; D_{RP} , weighted mean of single-locus Reynolds distances based on square root estimates of allele frequency (same weights as in the PHYLIP software); D_S , standard distance of Nei (1978) based on square root estimates of allele frequency; SM, simple matching index.

low correlation of D_{RP} with the simple matching index. The Neighbor-Joining tree based on D_{RM} is given in Figure 2.

Discussion

AFLP Polymorphism

The high average proportion of monomorphic populations and the large variation observed in allelic frequencies are sources of difficulties in the analysis, as already pointed out by Hill and Weir (2004) in their analysis of a subsample of our data. In such a situation one may consider excluding the more monomorphic loci, as suggested by Lynch and Milligan (1994) and done on our data by SanCristobal et al. (in press b). However, the choice of a truncation point is subjective, and loss of information and bias are incurred. Here we have shown the results obtained by using the method of Hill and Weir (2004) without excluding any locus.

Population Diversity

The moment estimations avoid the biases implied in the Lynch-Milligan (LM) procedure. The LM recommendation for minimizing biases is to prune loci for which P is below $3/n$. In our data set, where $n_h = 37.3$, the LM pruning criterion would be $P < 0.08$ and would lead to exclude 43% of the loci. This is well above the level of 10% considered as acceptable by Lynch and Milligan (1994), though below the level of 55% applied in SanCristobal et al. (in press b). Our situation is thus typically one in which the moment-based estimations have most merit against the standard square root-based methods. As shown in Table 2, the negative bias for the recessive allele frequency reached 20% for the PC showing the lowest null phenotype frequency. The positive bias of the square root method for F_{ST} was even higher and also considerably higher for the PC with the lowest allele frequency. The particularly high biases observed for low allelic frequencies confirm the simulations of Hill and Weir (2004: table 4). A scaling effect is, however, to be noted because bias in absolute value tends to decrease with increasing F_{ST} , which leads to overweight large deviations when biases are expressed in relative value. One would expect a positive bias in the Reynolds genetic distances also, contrary to the lower genetic distances derived from square root gene frequency (Table 3). The high correlation between D_{RA} and D_{RM} , however, indicates a rather uniform bias over all population combinations, suggesting that the ranking of distance comparisons between populations would not be much affected by the chosen estimator.

Hill and Weir (2004) have detailed the basic assumptions underlying their method and possible sources of bias remaining, such as, primarily, random mating and homogeneous distances between breeds. The latter implies a starlike pattern of relationship between populations. In our sample of populations, phylogeny has been shown to deviate from a starlike pattern of relationship because lines derived from the same breed tended to cluster significantly based on microsatellite data (SanCristobal et al. 2002, in press a). A star phylogeny can be approximated by keeping only one population from each microsatellite-based cluster. The calculations were redone on a subsample of 30 populations (shown in bold in Table 1) and 137 loci because 11 loci became fully monomorphic after removal of 29 breeds. The results in Table 2 show that average gene frequency was not changed, whereas F_{ST} was lowered and also more accurately estimated. Alternative estimation methods exist, such as maximum likelihood, based on gene frequency distribution, or Bayesian procedures, which have been reviewed by Hill and Weir (2004).

AFLP markers have been used successfully in livestock species, such as goat (Ajmone-Marsan et al. 2001; Marsan et al. 2002), cattle (Buntjer et al. 2002, Negrini et al. 2002), and pigs (Cameron et al. 2003), though AFLP studies appear to be less numerous than microsatellites. The F_{ST} estimate of 0.11 for this sample of pig populations is in keeping with the 0.11 and 0.15 estimates reported in goats (Ajmone-Marsan et al. 2001) and cattle (Negrini et al. 2002), respectively. We have no clear explanation to offer

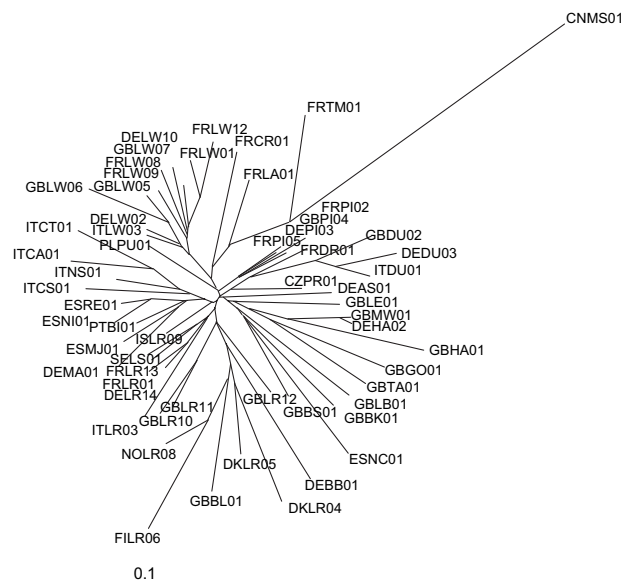


Table 4. Comparisons between Reynolds genetic distances for microsatellites and AFLP

Average distance among (or between)	Reynolds distances		
	Microsatellites (PHYLIP)	AFLP (Equation 4)	AFLP (Equation 5)
All 59 populations	0.21	0.10 (0.16)	0.27 (0.25)
Four Large White populations from France	0.04	0.05 (0.09)	0.14 (0.12)
Meishan breed and European populations	0.35	0.19 (0.29)	0.44 (0.45)

AFLP distances using the 67 most polymorphic loci are given in parentheses.

for the heterogeneity among PCs evidenced in our study, which is contrary to the assumed random amplification of AFLP loci. Most studies simply ignore this possible source of variation (e.g., see for plants the review of Nybom 2004). Campbell et al. (2003), however, noted effects of the PC on assignment success in fish.

More generally, significant PC effects as well as large variation of F_{ST} across loci indicate that some AFLP sequences may depart from neutrality, as previously suggested by SanCristobal et al. (2002). The high variability of F_{ST} across loci may be exemplified by the extreme situation of bands being nearly fixed in some lines and nearly absent in others, leading to F_{ST} values close to one. Variation of F_{ST} across loci, however, depends strongly on the pattern of relationship between populations (Robertson 1975). Interestingly, the coefficient of variation of F_{ST} , expected to be inversely proportional to the square root of the number of populations, was considerably above expectation over the whole sample of 59 breeds as compared to the subsample of 30 populations. A similar pattern was observed on microsatellites when 70 breeds were compared to 37 (Vega-Pla JL, personal communication). This tends to confirm the inflation of variance of F_{ST} which is to be expected in structured phylogenies (Robertson 1975). The Lewontin and Krakauer (1973) test, however, remained highly significant over the reduced sample of 30 populations because the observed variance of F_{ST} was nearly 45 times above the value predicted under the neutrality hypothesis, in keeping with the extreme variability noted above. The sensitivity of AFLP markers to selection effects has also been demonstrated by Cameron et al. (2003). Using the same restriction enzyme combination as in this study, they showed that AFLP markers successfully discriminated between animals from a set of divergently selected lines of pigs that had arisen from the same base population. This suggests that changes of QTL frequencies due to selection were sufficient to generate associated differences in AFLP frequencies, though the role played by genetic drift in selected lines of small size cannot be ignored.

AFLP Polymorphism and Genetic Distances

The low polymorphism of our data has the important consequence that quite frequently the same allele happens to be fixed in pairs of populations. Such loci, termed IM, thus generate a number of zero monolocus distances. The overall proportion of such distances is 47% over the 253, 228 locus-breed pair combinations considered in our study. The molecular nature of such loci merits further consideration to

confirm Mendelian behavior. Though internal evidence from nuclear families is lacking in our data, resemblance between full sibs could be assessed in several breeds (as in the preliminary analysis of Plastow et al. 2003). No deviation from expectation was observed even for extremely monomorphic loci. Of note also is the successful use of AFLP for mapping in several farm animal species including the pig (Plastow et al. 1998).

As noted by Weir (1996, p. 195), IM loci do not carry any information on the duration of the drift process since the populations have diverged. Obviously, the drift model breaks down as loci start to drift all the way to fixation. The PHYLIP software takes the option of computing a “weighted” Reynolds distance which in fact amounts to ignoring IM loci, whereas the IM loci are included in the arithmetic mean estimator of the Reynolds distance as well as in its moment estimation, thus yielding markedly lower distances when the proportion of IM loci is high as in our case. Table 4 shows that the “arithmetic mean” AFLP distances between Meishan and the European breeds were quite below the microsatellite distances. This suggests that by assuming a zero value for IM loci one does not entirely account for the overall divergence process between the two groups. In contrast, the bias in terms of drift appeared to be negligible for breeds having recently diverged because the AFLP distances even slightly exceeded the microsatellite distances among the Large White populations considered (Table 4). The agreement between our F_{ST} estimation of 0.11 and the average Reynolds distances D_{RM} and D_{RA} in Table 3 confirms that Equation 4 more truly reflects the breed diversity situation than Equation 5. The peculiar AFLP polymorphism also explains the stronger correlation between the standard Nei distance and the Reynolds distance obtained with the mean estimator compared to the PHYLIP estimator because the Nei distance does not exclude any locus. A straight application of PHYLIP on AFLP data cannot therefore be generally recommended, unless a check is made on the proportion of IM loci.

Comparison of AFLP and Microsatellite Markers for Genetic Diversity Studies

AFLP and microsatellite loci are numerous and dispersed randomly over the genome, making them both suitable for biodiversity analysis. Evidence against selective neutrality of microsatellite sequences is accumulating (see the review of Zhang and Hewitt 2003). Our data provide similar evidence for AFLP, in line with the discrimination power shown by Cameron et al. (2003).

It is well-known that different marker systems provide diversity estimates often not well correlated (e.g., Mariette et al. 2002). The analyses presented here also show that AFLP and microsatellites may carry different diversity information. The overall expected heterozygosity, H_e , of 0.12 (Table 2) is far below 0.56 for microsatellites (SanCristobal et al. in press a). In addition, the individual breed H_e s obtained for AFLP (using the square root estimate of gene frequency in each breed) were rather weakly correlated ($r = 0.52$) with the corresponding 59 H_e s for microsatellites (Ollivier 2002). As to breed differentiation, the microsatellite F_{ST} of 0.23 (Ollivier 2002), on 70 breeds of which 59 are common to the present study, was about twice as high as the 0.114 value obtained here with AFLP. On the other hand, in a study of breed contributions to diversity based on the same data, using both AFLP and microsatellite genetic distances, correlations weaker than expected were found between the two marker systems, though the partitioning of diversity between the three breed categories showed similar patterns (Ollivier et al. 2005). In an extensive compilation of wild plant diversity studies, Nybom (2004) compared microsatellites and AFLP and pointed out that within-population variation for microsatellites is usually considerably higher than that for dominant markers, as observed in our case. In contrast, estimates of among-population variation with the two types of markers usually agree in wild plants, though large discrepancies have occasionally been reported (e.g., Maguire et al. 2002). We should, however, keep in mind the rather different plant situation because fewer microsatellite studies are available in plants and plants include selfing as well as outcrossing species. More studies on livestock species would therefore be needed to confirm the differences we observe in our pig study.

The possibility of combining diversity information arising from different types of markers may be considered, taking due account of their statistical reliability, though true differences may need to be kept in mind. Based on the standard error of our moment-based estimates of distance for 148 biallelic dominant AFLP loci (0.016) and that for 50 “4.5-allelic” microsatellite loci (0.021 in SanCristobal et al. in press a), two biallelic dominant loci $((175/148)(0.021/0.016)^2 = 2.0)$ may be considered as equivalent to one biallelic codominant locus in our polymorphism situation.

In conclusion, compared to microsatellites, difficulties of analysis arise from the biallelic and dominant status of AFLP. As we have shown, the moment-based estimations of Hill and Weir (2004) can be recommended, instead of using the square root estimator of allele frequency, which leads to considerable biases in allele frequency and genetic diversity estimates. A potential advantage of AFLP over microsatellites is the identification of private alleles in assignment analyses (see Campbell et al. 2003), as also shown by Alves et al. (2003) in testing breed identity in pigs. Because AFLP markers are very easily generated, a large inventory can be made, and informative bands can be selected. We have also shown in this study that PC effects are not to be excluded. Appropriate PCs may thus be selected to obtain “high-quality loci,” as also suggested by Cameron et al. (2003) for discrim-

inating between pigs of particular genotypes and by Campbell et al. (2003) for assignment studies.

Acknowledgments

AFLP is a registered trademark of Keygene N.V. (Wageningen, The Netherlands). The PigBioDiv project was funded by the European Union (contract BIO4-CT98-0188). We are indebted to Prof. W. G. Hill (Edinburgh, UK) for providing useful comments and suggestions while preparing this paper. Contributions by G. Plastow (Abingdon, UK) and C. S. Haley (Roslin, UK) are also gratefully acknowledged.

References

- Ajmone-Marsan P, Negrini R, Crepaldi P, Milanese E, Gorni C, Valentini A, and Cicogna M, 2001. Assessing genetic diversity in Italian goat populations using AFLP markers. *Anim Genet* 32:281–288.
- Alves E, Castellanos C, Ovilo C, Silvio L, and Rodriguez C, 2003. Differentiation of the raw material of the Iberian pig meat industry based on the use of amplified fragment length polymorphism. *Meat Sci* 61:157–162.
- Buntjer JB, Otsen M, Nijman IJ, Kuiper MTR, and Lenstra JA, 2002. Phylogeny of bovine species based on AFLP fingerprinting. *Heredity* 88:46–51.
- Cameron ND, van Eijk MJT, Brugmans B, and Peleman J, 2003. Discrimination between selected lines of pigs using AFLP® markers. *Heredity* 91:494–501.
- Campbell D, Duchesne P, and Bernatchez L, 2003. AFLP utility for population assignment studies: analytical investigation and empirical comparison with microsatellites. *Mol Ecol* 12:1963–1978.
- Felsenstein J, 2004. PHYLIP (phylogeny inference package), version 3.63. Seattle, WA: Department of Genome Science and Department of Biology, University of Washington.
- [FAO] Food and Agriculture Organisation of the United Nations, 1998. Secondary guidelines for development of national farm animal genetic resources management plans. Measurement of domestic animal diversity (MoDAD): Original Working Group Report. Available at: <http://dad.fao.org/en/refer/library/guidelin/workgrp.pdf>. Accessed May 10, 2006.
- Hill WG and Weir BS, 2004. Moment estimation of population diversity and genetic distance from data on recessive markers. *Mol Ecol* 13:895–908.
- Lewontin RC and Krakauer J, 1973. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* 74: 175–195.
- Lynch M and Milligan BG, 1994. Analysis of population genetic structure with RAPD markers. *Mol Ecol* 3:91–99.
- Maguire TL, Peakall R, and Saenger P, 2002. Comparative analysis of genetic diversity in the mangrove species *Avicennia marina* (Forsk.) Vierh. (Avicenniaceae) detected by AFLPs and SSRs. *Theor Appl Genet* 104:388–398.
- Mariette S, Le Corre V, Austerlitz F, and Kremer A, 2002. Sampling within the genome for measuring within-population diversity: trade-offs between markers. *Mol Ecol* 11:1145–1156.
- Marsan PA, Negrini R, Milanese E, Crepaldi P, Cicogna M, Zagdsuren Y, Ertugrul O, Luikart G, and Taberlet P, 2002. Geographic structure in goat diversity as revealed by AFLP molecular markers. 7th World Congr Genet Appl Livestock Prod 33:657–660.
- Negrini R, Milanese E, Perez-Torrecillas C, Moretti M, Bozzi R, Filippini F, Lenstra JA, and Ajmone-Marsan P, 2002. Genetic diversity in central Italian cattle breeds as assessed by biallelic AFLP markers. 7th World Congr Genet Appl Livestock Prod 33:661–664.
- Nei M, 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89:583–590.
- Nybom H, 2004. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Mol Ecol* 13:1143–1155.

- Ollivier L, 2002. BIO4-CT98-0188. Characterisation of genetic variation in the European pig to facilitate the maintenance and exploitation of biodiversity. Scientific final report. Jouy-en-Josas, France: INRA.
- Ollivier L, Alderson L, Gandini GC, Foulley J-L, Haley CS, Joosten R, Rattink AP, Harlizius B, Groenen MAM, Amigues Y, Boscher M-Y, Russell G, Law A, Davoli R, Russo V, Matassino D, Désautés C, Fimland E, Bagga M, Delgado J-V, Vega-Pla JL, Martínez AM, Ramos AM, Glodek P, Meyer J-N, Plastow GS, Siggins KW, Archibald AL, Milan D, SanCristobal M, Laval G, Hammond K, Cardellino R, and Chevalet C, 2005. An assessment of the European pig diversity using molecular markers: partitioning of diversity among breeds. *Conserv Genet* 6:729–741.
- Ollivier L, Amigues Y, and Boscher MY, 2003. An EC-funded project on characterisation of genetic variation in the European pig. Objectives, organisation, breed sampling, DNA preparation and circulation. *Archiv Zootec* 52:137–144.
- Plastow GS, Kuiper M, Wales R, Archibald AL, Haley CS, and Siggins KW, 1998. AFLP for mapping and QTL detection in commercial pigs. 7th World Congr Genet Appl Livestock Prod 26:209–212.
- Plastow G, Siggins K, Bagga M, Brugmanns B, Heuven H, and Peleman J, 2003. Utilization of AFLP for genetic distance analysis in pigs. *Archiv Zootec* 52:157–164.
- Reynolds J, Weir BS, and Cockerham CC, 1983. Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics* 105:767–779.
- Robertson A, 1975. Gene frequency distributions as a test of selective neutrality. *Genetics* 81:775–785.
- Russell GA, Archibald AL, Haley CS, and Law AS, 2003. The pig genetic database and the WWW. *Archiv Zootec* 52:165–172.
- Saitou N and Nei N, 1987. The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425.
- SanCristobal M, Chevalet C, Haley CS, Russell G, Plastow G, Siggins K, Bagga M, Groenen MAM, Amigues Y, Hammond K, Laval G, Boscher M-Y, Milan D, Law A, Fimland E, Davoli R, Russo V, Gandini G, Archibald A, Delgado JV, Ramos M, Désautés C, Alderson L, Glodek P, Meyer J-N, Foulley JL, and Ollivier L, 2002. Genetic diversity in pigs—preliminary results on 58 European breeds and lines. 7th World Congr Genet Appl Livestock Prod 33:525–528.
- SanCristobal M, Chevalet C, Haley CS, Joosten R, Rattink AP, Harlizius B, Groenen MAM, Amigues Y, Boscher M-Y, Russell G, Law A, Davoli R, Russo V, Désautés C, Alderson L, Fimland E, Bagga M, Delgado JV, Vega-Pla JL, Martínez AM, Ramos M, Glodek P, Meyer J-N, Gandini G, Matassino D, Plastow G, Siggins K, Laval G, Archibald A, Milan D, Hammond K, and Cardellino R, in press a. Genetic diversity within and between European pig breeds using microsatellite markers. *Anim Genet*.
- SanCristobal M, Chevalet C, Peleman J, Heuven H, Brugmanns B, van Schriek M, Joosten R, Rattink AP, Harlizius B, Groenen MAM, Amigues Y, Boscher M-Y, Russell G, Law A, Davoli R, Russo V, Désautés C, Alderson L, Fimland E, Bagga M, Delgado JV, Vega-Pla JL, Martínez AM, Ramos M, Glodek P, Meyer J-N, Gandini G, Siggins K, Laval G, Archibald AL, Milan D, Hammond K, Cardellino R, Haley C, and Plastow G, in press b. Genetic diversity in European pigs utilising AFLP markers. *Anim Genet*.
- Vos P, Hogers R, Bleeker M, Reijans M, Van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, and Zabeau M, 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res* 23:4407–4414.
- Weir BS, 1996. *Genetic data analysis II*. Sunderland, MA: Sinauer Associates.
- Zhang DX and Hewitt GM, 2003. Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Mol Ecol* 12:563–584.

Received October 11, 2005

Accepted March 9, 2006

Corresponding Editor: Bruce Weir