UTILIZATION OF AFLP® FOR GENETIC DISTANCE ANALYSIS IN PIGS

UTILIZACIÓN DE LOS AFLP® PARA EL ANÁLISIS DE DISTANCIAS GENÉTICAS EN CERDOS

Plastow, G.[1], K. Siggens[1], M. Bagga[1], B. Brugmans[2], H. Heuven[2] and J. Peleman[2]

[1]PIC Group. Iniversity of Cambridge. CB2 1QP. UK.
[2]Keygene NV. Agrobusiness Park 90. 6708PW Wageningen. The Netherlands.

ADDITIONAL KEYWORDS

Pig. Breed. Genetic diversity. AFLP markers.

PALABRAS CLAVE ADICIONALES

Cerdo. Raza. Diversidad genética. Marcadores AFLP.

SUMMARY

AFLP® was utilized successfully as a tool to generate a large number (more than 100) of DNA markers in a total of 2435 individuals from the EC Pig Biodiversity project. This information was used for an initial genetic distance analysis with the results tending to group the lines and breeds as expected based on the descriptive information.

RESUMEN

Los AFLPs han sido utilizados con éxito como herramienta para generar un gran número (más de cien) de marcadores de ADN en 2435 individuos del proyecto Europeo sobre Biodiversidad Porcina. Esta información fue utilizada para realizar un análisis inicial de distancias genéticas que permitió agrupar líneas y razas según sugería la información descriptiva.

INTRODUCTION

The AFLP® technique is a powerful method for generating a specific and reproducible DNA fingerprint from any DNA sample. These fingerprints may differ between individuals when polymorphisms are present in the DNA fragments being generated. The technique is based on selective PCR amplification of restriction fragments from a total digest of genomic DNA without knowledge of nucleotide sequences of the target organism. The method is therefore ideal for generating large numbers of new polymorphic markers. For this reason it was of interest to determine the utility of AFLP for genetic distance analysis in pigs.

The objective of this Project WorkPackage within the EC Pig Biodiversity project was to use AFLP genotyping to generate new markers for calculating genetic distances between breeds/lines, applying it to individual samples as well as reviewing the possibility for use with bulked samples. Comparison of its utility with

other marker systems will be discussed.

MATERIALS AND METHODS

PREPARATION OF DNA

Blood (20 ml) was collected into potassium EDTA. Red cells were lysed by addition of 30 ml Lysis Buffer (155 mM NH_4Cl, 10 mM $KHCO_3$, 0.1 mM EDTA pH 8.0); samples were mixed and left at room temperature for 10 minutes. Nucleated cells were collected by centrifugation, 3500g x 10 minutes. The cell pellet was re-suspended in 12.5 ml Extraction Buffer (100mM Tris-HCl, 50 mM EDTA, 150 mM NaCl pH 8.0) plus 75 µl proteinase K (20 mg/ml); 2.5 ml of 20 percent SDS was then added and the lysate mixed gently. Lysates were incubated at 50°C for 2 hours.

DNA was extracted twice with phenol (buffered 0.1M Tris-HCl pH 8.0, 0.1 percent 8-hydroxyquinoline) followed by one extraction with chloroform. DNA was precipitated from the aqueous phase by addition of 1/10 volume 3M sodium acetate pH 5.2 and 2 volumes ethanol. Precipitated DNA was hooked out of the tube and washed in 70 percent ethanol. The DNA was drained briefly, dried for 10 minutes at room temperature and re-suspended in 2 ml TE buffer (5 mM Tris-HCl, 1 mM EDTA pH 8.0) overnight at 4°C followed by 2 hours on a rolling platform at RT.

DNA concentration and quality were checked on an agarose gel and each sample was adjusted to 200 µg/ml. In total 2625 samples from 61 lines were generated by the project partners for AFLP analysis.

AFLP FINGERPRINTING

Fingerprints were generated as described by Vos *et al.* (1995). An example is given in **figure 1**. Templates were made starting from 100-500 nanograms of DNA. Once the templates are prepared then a virtually unlimited number of fingerprints can be generated. PCR amplification, using these DNA templates for the enzyme combination *EcoRI/TaqI* were performed on all samples. In the current

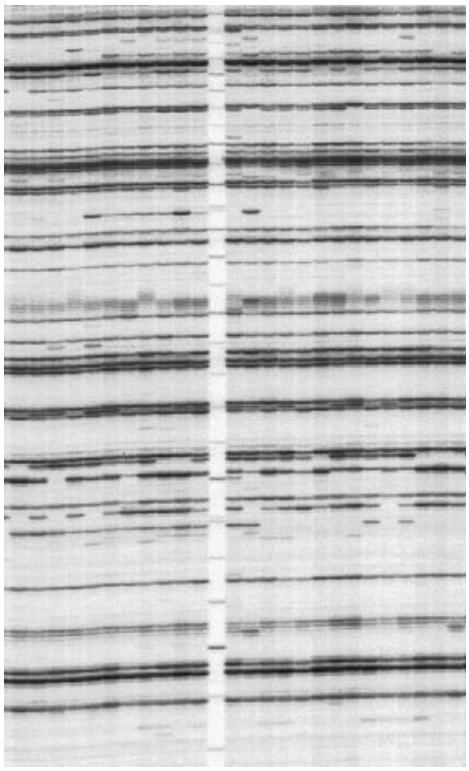

Figure 1. *Detail of an AFLP gel from the project including the 10bp size ladder in the centre.* (Detalle de un gel AFLP del proyecto incluyendo una escala de 10bp de tamaño en el centro).

study four *EcoRI/TaqI* primer combinations (PCs) were chosen based on the results obtained in previous AFLP analyses of pigs (Plastow *et al.*, 1998). The PCs were: *EcoRI*+AAG/*TaqI* +CAA (E33/T47), *EcoRI*+AAG/ *TaqI*+CTT (E33/T62), *EcoRI*+ACA/ *TaqI*+CAC (E35/T48) and *EcoRI* +ATT/*TaqI*+CAC (E46/T48).

AFLP was also established at the Sygen Laboratory using the LiCOR sequencer platform (Myburg *et al.*, 2001). However, the range of fragments visualised for optimal scoring differed with the sequencer platform compared to the Keygene platform. This meant that it would be relatively inefficient to generate markers on two platforms as only a proportion of the fragments would be common between the platforms, resulting in significant loss of information. In addition, whilst it is possible using common samples as standards, to generate comparable information, it is by no means trivial. For this reason it was decided to only generate project information at Keygene.

In order to optimise the 96-lane gel-format a maximum of 46 indivi-duals were fingerprinted per line. However, in some cases the number available was less and in other cases DNA was found to be degraded. Markers between 50 and 600 base pairs were scored. A dominant dataset was created starting from co-dominant scored genotypes using proprietary software and transferred to the project database at the Roslin Institute.

ANALYSIS

The score data was converted for cluster analysis and genetic distance analysis performed as follows: The presence of an AFLP-fragment (D-score) was converted to 1, the absence of an AFLP fragment (B score) to 0 and missing scores were assigned the value 2. Two different similarity matrices were generated consisting of similarity indices for all combinations of lines using NTSYpc software (Rohlf, 1993). Similarity matrices were calculated using the *Simple Matching* (SM=(A+B/n)) and *Jaccard* (J=a/ (a+b+c)° coefficients respectively where *a*, *b*, *c*, *d* and *n* are from a two way frequency table for all pairs of the two objects *i* and *j* **(table I)**.

RESULTS

A pilot analysis was run with a subset of 11 of the lines to confirm and maximise the number of bands that

Table I. a, b, c, d *and* n *are defined as follows for a two-way frequency table for all pairs of two objects* i *and* j. (*a, b, c,* y *n* se definen como sigue en una tabla de frecuencias de doble entrada para todos los pares de dos objetos *i* y *j*).

		J	
		+	-
i	+	a	b
	-	c	d

a + d= number of matched
b + c= number of *unmatched*
n= total sample size

Genstat was used to calculate average similarity coefficient between lines based both on Simple Matching and Jaccard coefficients.

Table II. *Relative statistical efficiencies (relative standard error of the estimate of genetic distance).* (Eficiencias estadísticas relativas (error típico relativo de las estimaciones de distancia genética)).

	4 PCs	3PCs
Full co-dominant scoring	100	88
Partial co-dominant scoring*	90	79
Full dominant scoring	87	76
Pooled samples	54	47

*about 27percent.

would be generated from the AFLP PCs. Test AFLP fingerprints were run on 506 samples (46 per line: 25 males and 21 females) to verify sample quality and PC performance in terms of marker *density*. All of the samples yielded high quality fingerprints (see **figure 1**) and a total of 111 bands could be visualised and scored with the Keygene scoring software. Approximately 27 percent of the markers could be scored co-dominantly. This sample set was made up of 21 fullsibs (i.e. one male and female from each of the unrelated litters sampled) plus the 5 additional males (i.e. all males were analysed) and as expected each individual was most closely related to its litter mate based on simple distance analysis.

In order to determine the strategy for the rest of the genotyping, estimates were made of the relative efficiencies of typing with three PCs (87 markers) or four PCs (111 markers) and being able to score markers co-dominantly, dominantly or a mix thereof; consideration was also given to the implications of trying to use pooled samples (two samples per pool, pooling within line) as a means to reduce genotyping costs. On the assumptions of typing 46 animals with a gene frequency q=0.5 the following relative efficiencies were obtained, see **table II** (Ollivier, personal communication).

Table II shows that having to use dominant scores does not reduce the efficiency dramatically. In comparison it can be seen that using pooled samples is not a real option as the efficiency is significantly reduced. As a comparison with the microsatellite typing in the project it was estimated that four PCs would be approximately 72 percent as efficient as typing 50 microsatellites (assuming an average of 3.5 alleles per MS) (Ollivier, personal communication).

The decision was made to type the full set of animals with four PCs. Differences in DNA quality led to differences in band intensities which made it difficult to separate heterozygotes from the homozygous present class. It was therefore necessary to score all markers in a dominant fashion. Genotyping this wider range of genetic diversity uncovered additional markers not seen in the pilot study and at the same time led to a small number of the pilot markers being rejected as poor

Table III. *The number of markers per PC.* (Número de marcadores por PC).

PC	# of markers
E33/T47	28
E33/T62	43
E35/T48	39
E46/T48	38
Total	148

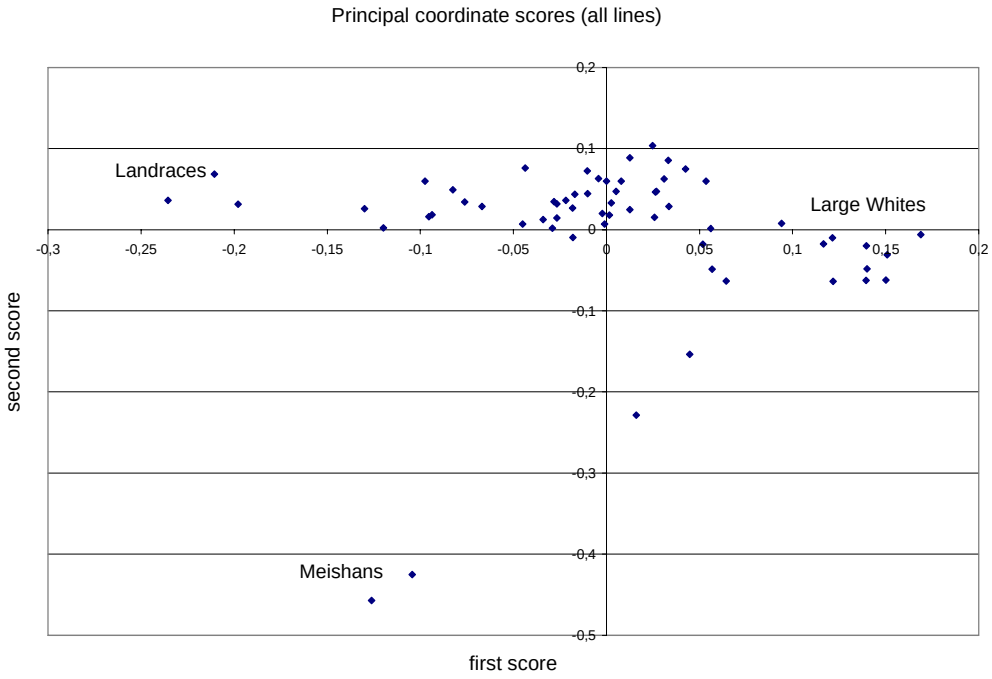

Figure 2. *Principal coordinate scores for all lines.* (Valores de las coordenadas principales para todas las líneas).

quality markers for the larger data set. The number of markers scored per PC is given in **table III**.

A total of 148 markers could be reliably scored for all 2435 individuals that were genotyped. These 148 markers were then used for genetic distance analysis (using the >300,000 data points generated). The similarity coefficients between all individuals were summarized by calculating average similarity coefficients between lines. To visualize these average coefficients a principle co-ordinate analysis was performed and the scores for the first two factors plotted. **Figure 2** is based on analysis of all lines including the Chinese Meishan lines. **Figure 3** presents the results of the analysis without these two *extreme* lines.

DISCUSSION

AFLP is an extremely effective multiplex technology for the generation of random DNA markers. We have demonstrated here that 148 markers can be generated from more than sixty breeds of pigs with four AFLP primer combinations. The initial analysis indicates that such markers have utility in determining genetic distances between breeds. The method is extremely reproducible, particularly

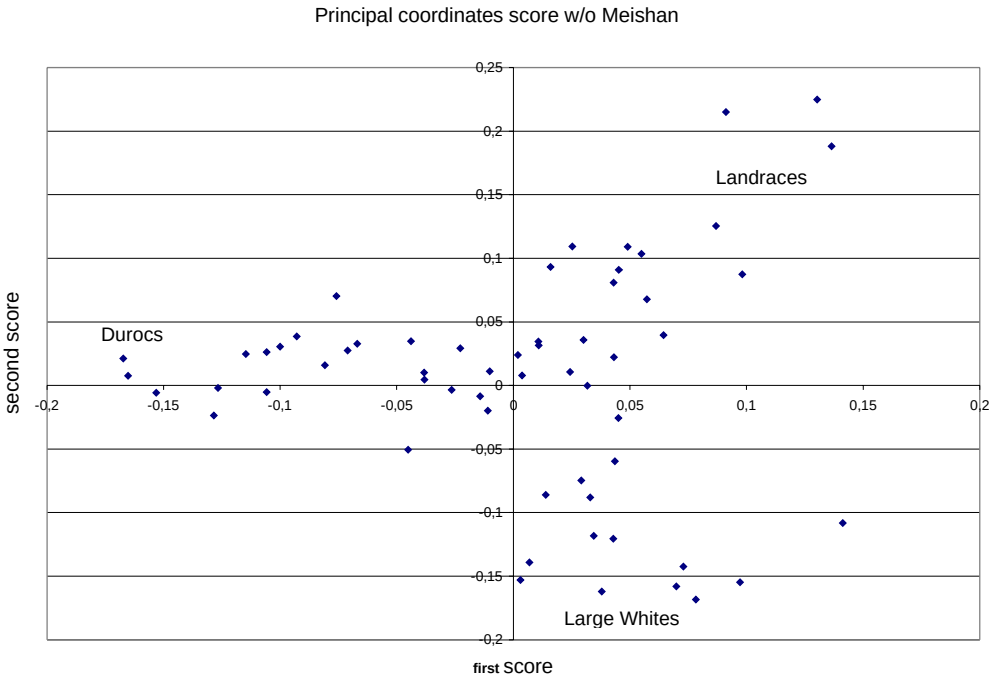

***Figure 3**. Principal coordinate scores of the analysis without MS01 and MS0.* (Puntuaciones de las coordenadas principales del análisis sin MS01 y MS0).

when applied at Keygene. Unfortunately it was not possible to score all the markers co-dominantly (due to differences in DNA quality) which reduced the information content. This could be addressed by adding PCs although this would entail extra cost. Even so, we feel that AFLP will have a possible advantage over microsatellites as less set up and less DNA is required. Thus AFLP could be particularly valuable for animal species where there is a shortage of suitable polymorphic DNA-markers and/or DNA.

Unlike the situation with microsatellites (Groenen *et al.*, 2003) it was not easy to utilize AFLP for pooling strategies. However, if care is taken to normalize DNA concentration then it might be possible to quantify bands to provide some information on allele frequencies within pools. If this was possible then AFLP may be more efficient because of the multiplex nature of AFLP and the relative ease of marker scoring compared to pooled microsatellites due to stutter bands etc. Indeed, initial results at Keygene using dominant AFLP markers in pools have subsequently been encouraging (van Eijk, Pers. Comm.)

DNA quality is a very important criterion for AFLP analysis. Ideally DNA is required to be of high quality in order to reduce background and to generate clean fingerprints over the full fragment range. DNA quality did

vary between breeds and source as judged by the success of AFLP. However, in general there did not seem to be any advantage over microsatellites, approximately the same success rates were observed between the methods (results not shown). It was also the case that the same samples gave problems with both methods. Even so, as noted above better overall DNA quality may have led to co-dominant scoring of some markers thereby increasing the utility of the marker sets for distance analysis.

Genetic diversity studies are an important component of efforts to conserve gene pools and they can also be used to assist in the selection of lines to be used in commercial breeding programs. For example, choosing more distant lines within and between breeds for crossbreeding in order to optimize potential heterosis. The preliminary analysis presented here suggests that AFLP can be used to generate useful genetic distance information. The Chinese Meishan lines are clearly shown to represent a *different gene pool* compared to the European lines, which cluster more closely together **(figure 2)**. Within the European lines the Landrace, Large White and Duroc lines appear in different parts of the cluster analysis with the Landrace and Large White lines tending to be at opposite sides of the plots **(figures 2 and 3)**. This supports their common use in crossbreeding programs for dam lines to generate heterosis for traits such as litter size.

Although, the initial results suggest that AFLP, using four PCs, may not be as effective as utilizing 50 microsatellites (see also San Cristobal *et al.*, 2003, in the present proceeding) it clearly has potential for such studies. Indeed the option exists to increase the number of PCs to reach equivalence; and with recent impro-vements the relative cost may favour the use of AFLP over microsatellites. A more detailed comparison of the AFLP and microsatellite data is now underway (San Cristobal, Chevalet, Heuven, Plastow *et al.* in preparation) and this may provide a clearer view of the relative utility of each of the marker systems for genetic distance analysis. Both systems require a high degree of skill to utilize them effectively, however, AFLP does not require the large amount of development that is required for microsatellite panels and they may also be easier to compare between laboratories (as allele scoring is easier). This makes AFLP particularly interesting for *new* species where large numbers of appropriate microsatellites may not be immediately available.

Even so, a comparison of AFLP and SSRs for determining genetic diversity in mangrove species Avicennia marina (Maguire *et al.*, 2002) showed that either marker system is applicable to expand genetic studies of mangrove. A similar study by David *et al.* (2001) resulted in similar conclusions when both marker systems where used to determine genetic diversity in Carp strains. Similar studies in pigs and cattle have also been performed by Ovilo *et al.* (2000) and Basedow (1998) respectively. AFLP was successfully used by Ajmone Marsan *et al.* (2002, cattle), Chung *et al.* (2001, pigs) and Ajmone-Marsan *et al.* (2001, goats) to study genetic variation within and between breeds/lines.

CONCLUSION

The AFLP-technology allows simple and cost effective determination of genetic diversity of any species without prior knowledge of sequence information.

ACKNOWLEDGEMENT

AFLP® is a registered trademark of of Keygene N.V. covered by patents and patent applications and used by PIC under a registered license.

REFERENCES

Ajmone-Marsan, P., R. Negrini, E. Milanesi, R. Bozzi, I.J. Nijman, J.B. Buntjer, A. Valentini and J.A. Lenstra. 2002. Genetic distances within and across cattle breeds as indicated by bi-allelic AFLP markers. *Animal Genetics*, 33: 280-286.

Ajmone-Marsan, P., R. Negrini, P. Crepaldi, E. Milanesi, C. Gorni, A. Valentini and M. Cicogna. 2001. Assessing genetic diversity in Italian goat populations using AFLP® markers. *Animal Genetics*, 32: 281-288.

Basedow, M. 1998. Use of molecular genetic markers in the study of genetic diversity of German cattle breeds. PhD-thesis of University of Kiel, Germany.

Chung, E.R., W.T. Kim, Y.S. Kim, J.K. Lee and S.K. Han. 2001. Genetic diversity and breed identification of Korean native pig using AFLP markers. *J. An. Sci. and Technology*, 43: 777-788.

David, L., P. Rajasekaran, J. Fang, J. Hillel and U. Lavi. 2001. Polymorphism in ornamental and common Carp strains (*Cyprinus carpio* L.) as revealed by AFLP analysis and a new set of microsatellite markers. *Molecular Genetics and Genomics*, 266: 353-362.

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 microsatellite genotyping for population studies in the pig using individual and pooled DNA samples. *Arch. Zootec.*, 52: 145-155.

Maguire, T.L., R. Peakall and P. Saenger. 2002. Comparative analysis of genetic diversity in mangrove species Avicennia mariana detected by AFLPs and SSRs. *Theoretical and Applied Genetics*, 104: 388-398.

Myburg, A.A., D.L. Remington, D.M. O'Malley, R.R. Sederoff and R.W. Whetten. 2001. High-throughput AFLP analysis using infrared dye-labeled primers and an automated DNA sequencer. *Biotechniques*, 30: 348-357.

Ovilo, C., M.C. Barragan, C. Castellanos, M.C. Rodriguez, L. Silio, M.A. Toro, J.A. Alfonso-de-Almeida and J. Tirapicos-Nunes. 2000. Proc 4th Int. Symposium on Mediterranean Pig. Evora, Portugal. Options Mediterraneennes, Serie A 41: 79-84.

Plastow, G.S., M. Kuiper, R. Wales, A.L. Archibald, C.S. Haley and K.W. Siggens. 1998. AFLP for mapping and QTL detection in commercial pigs. Proceedings of the 67th World Congress on Genetics Applied to Livestock Production 26, Armidale, NSW, Australia 209-212.

Rohlf, F.J. 1993. NTSYS-pc. Numerical taxonomy and multivariate analysis system. Exeter Software NY 11733.

San Cristobal, M., C. Chevalet, J.-L. Foulley and L. Ollivier. 2003. Some methods for analysing genetic marker data in a biodiversity setting - Example of the PigBioDiv data. *Arch. Zootec.,* 52: 173-183.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, Th. Van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper and M. Zabeau. 1995. AFLP: a new technique for DNA fingerprinting. *Nuc. Ac. Res.,* 23: 4407-4414.