Is it possible the breed origin traceability of Iberian pigs?

Alves E., Fernández A.I., García-Cortés L.A., Lopez A., Benítez R., Rodríguez C., Silió L.

in

De Pedro E.J. (ed.), Cabezas A.B. (ed.).
7th International Symposium on the Mediterranean Pig

Zaragoza : CIHEAM
Options Méditerranéennes : Série A. Séminaires Méditerranéens; n. 101

2012
pages 565-571

Is it possible the breed origin traceability of Iberian pigs?

E. Alves, A.I. Fernández, L.A. García-Cortés, Á. López, R. Benítez C. Rodríguez and L. Silió

*Departamento de Mejora Genética Animal, INIA, Ctra. La Coruña, Km 7, 28040 Madrid (Spain)

Abstract. Two different approaches for the verification of *Iberian* breed origin, in both animals alive and meat products, are jointly described in this study. The former is based on two genes that present polymorphisms with alleles that are exclusive (*MC1R*4*) or present high frequencies (*IGF2* g.3072A) in *Duroc* breed and are absent in *Iberian*. The use of these markers allows to discriminate Iberian purebred from crossbred *Duroc* x *Iberian* but it does not fully detect pigs with a lower proportion of *Duroc* genes. The second approach is centered on the use of single nucleotide polymorphisms (SNP) presenting divergent frequencies in both breeds. After hybridization with PorcineSNP60 BeadChip of samples from different *Iberian* (26) and *Duroc* (15) origins, we selected 96 SNPs with differences of allelic frequencies larger than 0.8 and evenly distributed over the 18 autosomes. Simulations were carried out estimating that 48 out of these SNPs would allow the verification of breed origin, with errors ranging from 1% to 4%, both in purebred and in crossbred animals with different *Duroc* proportion. Two diagnostic analyses based respectively on 750 and 230 samples genotyped for the panel of 96 SNPs have been performed with different purposes and satisfactory results.

Keywords. Breed traceability – Iberian – Duroc – SNP.

La traçabilité de l'origine des porcs de race Ibérique est-elle possible ?

Résumé. *Deux approches sont décrites pour vérifier l'origine de la race Ibérique, applicables aux animaux vivants ainsi qu'à leurs produits dérivés, grâce à l'utilisation de marqueurs génétiques. La première est basée sur l'utilisation de deux gènes qui présentent des polymorphismes avec des allèles fixés (MC1R*4) ou qui présentent une fréquence élevée (IGF2 g.3072A) chez la race Duroc et qui sont absents chez la race Ibérique. L'utilisation de ces marqueurs permet de discriminer les génotypes Ibériques purs par rapport aux animaux issus de croisements Duroc x Ibérique, mais ne détecte pas complètement les porcs ayant un fond génétique Duroc moins important. La deuxième approche est basée sur l'utilisation des polymorphismes de type SNP, présentant des fréquences alléliques différentes chez les deux races. Après hybridation des échantillons de différentes origines Ibériques (26) et Duroc (15) sur la puce de génotypage PorcineSNP60 BeadChip, nous avons sélectionné 96 SNPs parmi ceux qui présentaient des différences inter-races de fréquences alléliques supérieures à 0,8 et qui étaient répartis sur les 18 autosomes. Les simulations réalisées estiment que l'utilisation de 48 de ces SNPs permettrait la répartition par race des animaux purs et croisés avec Duroc dans des proportions différentes, en obtenant des taux d'erreur compris entre 1% et 4%. Nous avons réalisé deux analyses diagnostiques basées sur le génotypage de 750 et 230 échantillons, respectivement, sur le panel de 96 SNPs, avec des objectifs différents et des résultats satisfaisants.*

Mots-clés. *Traçabilité – Ibérique – Duroc – SNP.*

I – Introduction

Iberian pigs are the source of highly priced meat and dry-cured products. The optimum quality of meat and hams is associated with both the combination of purebred Iberian genotypes and traditional extensive fattening called *Montanera*. However Iberian pigs are commonly crossbred with Duroc animals, and even other dark coated breeds like *Large Black*, to improve their efficiency for lean growth. The Spanish regulation of 'Iberian' labelling only admits progenies from Iberian sows and boars crosses with males from other breeds than *Duroc*. Moreover cured

products are labelled as '*Ibérico*' or '*Ibérico Puro*' depending if they proceed from *Duroc* x *Iberian* crossbred or from purebred *Iberian* animals. In the case of products to be exported, additional quality controls could be eventually carried out by customer countries. Full traceability is increasingly being demanded by producers and consumers but traditional tagging systems presents several difficulties, mainly in *Iberian* pigs, because of their extensive management system (López-Bote, 1998).

In this context, genetic markers could be a useful tool to check the breed origin of living animals and products as well as to verify the parentage relations registered on the Herd Book. Microsatellite markers have been proposed for estimate the genetic composition of dry-cured *Iberian* hams (García *et al*., 2006). However this approach requires the genotyping of a high number of markers (>25) and preliminary analyses on parental populations are needed to determine their allelic frequencies pattern. In addition, these markers can be difficult to score and are not amenable to automation. Single nucleotide polymorphisms (SNPs) present several advantages over microsatellite markers: higher abundance in the genome, more easily to handle and interpret in laboratory and better compatibility with automation. Although SNP are usually bi-allelic and consequently less informative than microsatellites, this disadvantage can be overcome by genotyping a higher number of SNPs. Our group has been working in the search of *Iberian* or *Duroc* exclusive genetic markers and, despite the close genetic relationship between these breeds, some exclusive alleles were reported on nuclear (Fernández *et al*., 2004) and mitochondrial (Alves *et al*., 2009) genes. Indeed analysis of *MC1R* and *IGF2* alleles are already in use and mitochondrial DNA markers could result particularly useful for *Iberian* maternal origin validation. The problems arise when samples to analyze carry less than 50% of *Duroc* genes. In these cases neither of the available tests does itself results enough to certify genetic origin (Rodriguez-Ramilo *et al*., 2008). Advances in high-throughput DNA sequencing and genotyping have led to the recent commercialization of a high density porcine SNP array. The aim of this work was to use the 60K porcine SNP array to develop a low-density panel of evenly spaced SNPs and check its effectiveness for differentiate purebred *Iberian* from crossbred pigs with a wide range of *Duroc* genes.

II – Materials and methods

1. High-density genotyping

In a first step, 41 samples were analyzed including 15 *Duroc* boars from 12 different genetic origins and 26 *Iberian* pigs (both males and females) from 16 breeding nuclei registered in the Herd Book. Genomic DNA was extracted from blood samples according to the standard phenol-chloroform method or from ear tag biopsies using the PureLink™ Genomic DNA kit (Invitrogen, Spain). All these samples were genotyped for 62,163 SNPs, using the Porcine SNP 60 BeadChip (Illumina, San Diego, CA, USA). Genotyping reactions were performed on an "Infinium DNA Analysis Assay" at the Veterinary Service of Molecular Genetics (Universitàt Autonoma de Barcelona, Spain).

2. SNP selection and evenly spaced low-density genotyping

The porcine SNP60 BeadChip features 62,163 evenly spaced probes with an estimated one marker per 40Kb across the pig genome. Genotyping data were analyzed and those SNPs that failed to produce an amplification product, that have no information about their location or map over sexual chromosomes were eliminated of the study, yielding a total of 45,180 SNPs. A first SNP selection was based on (*i*) their informativity i.e. the SNP that presented divergent frequencies between the two analyzed breeds with between-breed differences of allelic frequencies larger than 0.80 and (*ii*) regular distribution over the 18 autosomes. We also incorporated additional probes in order to check the feasibility of genotyping highly informative SNP for breed origin verification (*IGF2*, *MC1R*, *OCA2*) not included in the Porcine SNP 60

BeadChip. The Illumina Assay Design Tool (ADT) was used for evaluating individual loci and creating the most successful custom genotyping assay. The 96 SNP loci finally selected were simultaneously interrogated with the GoldenGate Genotyping Assay. Two analyses were performed with different goals and based respectively on: *a*) 750 DNA samples of purebred pigs of both breeds and probationary *Iberian* pigs, and *b*) 230 DNA samples from *Iberian* sows and *Duroc* boars and controlled F1 *Duroc* x *Iberian* crossbred pigs.

3. Simulations and statistical analysis

Previous simulations were performed assuming different number of bi-allelic SNPs (24, 48 and 96) with diverse values (from 0.60 to 0.80) for the allelic differences between *Iberian* and *Duroc* breeds. In each simulation, one hundred reproducers of each breed were simulated with the correspondent number of SNP genotypes. From the alleles of reproducers of each breed we sampled two hundreds F1 crossbred pigs, and a similar procedure was used between purebred and F1 individuals to simulate four hundreds pigs of each one of the backcrosses: F1 x *Iberian* and F1 x *Duroc*. According to its profile of allelic frequencies, each one of the simulated pigs was assigned to one of five clusters using an algorithm similar to STRUCTURE (Pritchard *et al*., 2000). One hundred replicates were obtained for each simulated case.

For the analysis of actual genotyping data software *Bayesian Analysis of Population Structure* (BAPS) v5 was used (Corander *et al*. 2003).

III – Results and discussion

1. SNP selection from High-Density Genotyping

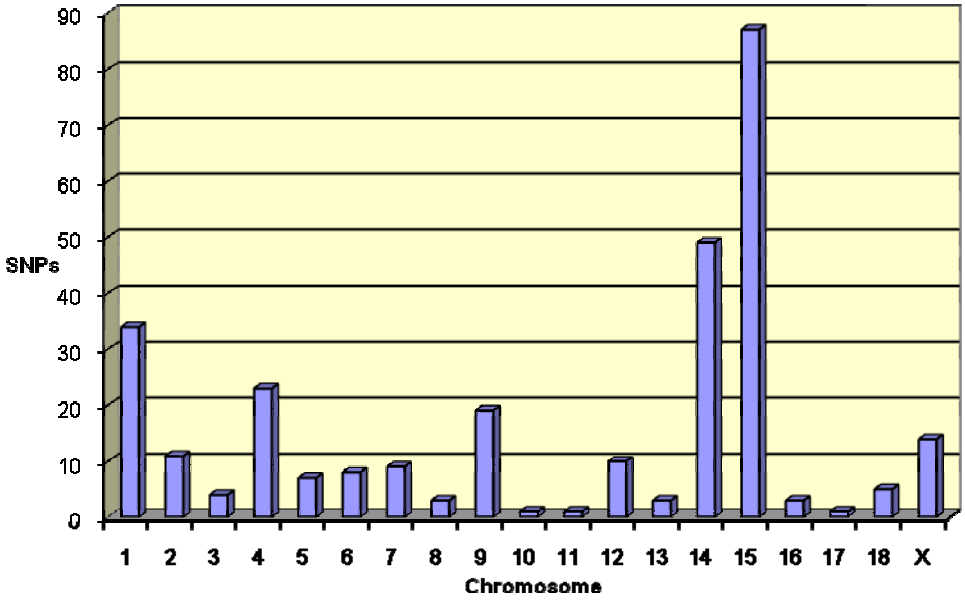

Fig. 1. Distribution over the porcine chromosomes of SNPs showing differences between breeds of allelic frequencies larger than 0.80, obtained from high-density genotyping with SNP60 BeadChip.

The genotyping of 15 *Duroc* and 26 *Iberian* samples with the porcine SNP60 BeadChip revealed a total number of 292 SNPs with allelic frequencies differences larger than 0.80 and mapped on different chromosomes (Fig. 1). The number of SNPs displaying divergent allelic frequencies is proportional to the total number of probes interrogated in each chromosome but it is not related with the size of the chromosome. That is the reason why the highest number of SNPs was observed on SSC14 and SSC15 that are smaller than SSC13. Despite this inconvenient, we tried to carry out the SNP selection taking into account the chromosomes size whenever possible and choice SNPs evenly spaced across the porcine genome. Moreover, as it was said before, we checked the feasibility of genotyping highly informative SNPs for breed origin verification (*IGF2*, *MC1R*, and *OCA2*) not included in the Porcine SNP 60 BeadChip. Assay Design Tool (Illumina) provided satisfactory scores for *OCA2*, *MC1R*2* and *MC1R*4* but not for *IGF2* g.3072A polymorphism. Inclusion of mitochondrial polymorphisms also had to be discarded because GoldenGate Genotyping Assay does not allow the genotyping of this kind of variation. The final set of 96 SNPs included 92 probes selected from the Porcine SNP 60 BeadChip, distributed over the 18 autosomes as follows: 11 on SSC1, eight on SSC2, three on SSC3, six on SSC4, six on SSC5, seven on SSC6, five on SSC7, three on SSC8, seven on SSC9, one on each one of SSC10 and SSC11, six on SSC12, four on SSC13, eight on SSC14 and also on SSC15, two on SSC16, one on SSC17 and five on SSC18. Moreover four additional probes were included, two for *MC1R*2* and one for each one *MC1R*4* and *OCA2*.

2. Simulation results

Results of two of the performed simulations were summarized in Tables 1 and 2. The first case corresponds to the use of a panel of 48 SNPs with remarkable allelic differences between breeds ($|q_{IBERIAN} - q_{DUROC}| > 0.80$), and the second to a larger panel of 96 SNPs with lower divergence of frequencies ($|q_{IBERIAN} - q_{DUROC}| > 0.60$). In both situations, the method performs very well for assigning adequately each individual to its correspondent genetic group. The observed error rates in these simulations ranged from 1% to 4% using 46 very divergent SNPs (Table 1), and from 0.1 % to 2% using 96 moderately divergent SNPs (Table 2).

Table 1. Allocation average proportions (over 100 replicates) to the different groups of simulated purebred and crossbred pigs: *Iberian* (*n* = 100), *Duroc* (*n* = 100), F1 *Duroc* x *Iberian* (*n* = 200), F1 x *Duroc* (*n* = 400) and F1 x *Iberian* (*n* = 400) genotyped for 48 SNPs with allelic differences between breeds greater than 0.80

	Iberian	Duroc	F1 Duroc x Iberian	F1 x Duroc	F1 x Iberian
Iberian	0.992	0.000	0.000	0.000	0.008
Duroc	0.000	0.992	0.000	0.008	0.000
F1 Duroc x Iberian	0.000	0.000	0.960	0.011	0.029
F1 x *Duroc*	0.000	0.002	0.033	0.965	0.000
F1 x Iberian	0.005	0.000	0.034	0.000	0.961

However, some of the assumptions of the simulations favour these positive results: *a)* purebred and crossbred pigs are related, the last ones being progenies of the *Iberian* and *Duroc* reproducers considered in the analyses, *b*) the number of clusters (five) is known and the possible hidden substructure of the purebred populations is ignored, and *c*) all the genotypes are available without missing marker data. The assumptions *a*) and *c*) are clearly unrealistic in practice, and the assumption *b)* is questionable at least for the Iberian breed where a hidden substructure has been previously inferred (Alves *et al*., 2006). The considered proportions of *Duroc* genes could be directly applied to the situation of Iberian products that usually proceed from these crosses. However, a lower proportion of the *Duroc* genome could be present at animals qualified as purebred *Iberian*. These arguments indicate that the ability of these

techniques to solve the practical problems of traceability of *Iberian*-type live pigs and meat products will be lower than the expected according to the simulated results, and the use of panels of at least 96 very divergent SNPs ($|q_{IBERIAN} - q_{DUROC}| > 0.80$) seems advisable.

Table 2. Allocation average proportions (over 100 replicates) to the different groups of simulated purebred and crossbred pigs: Iberian (n = 100), Duroc (n = 100), F1 Duroc x Iberian (n = 200), F1 x Duroc (n = 400) and F1 x Iberian (n = 400) genotyped for 96 SNPs with allelic differences between breeds greater than 0.60

	Iberian	Duroc	F1 Duroc x Iberian	F1 x Duroc	F1 x Iberian
Iberian	0.999	0.000	0.000	0.000	0.001
Duroc	0.000	1.000	0.000	0.000	0.000
F1 Duroc x Iberian	0.000	0.000	0.972	0.009	0.019
F1 x Duroc	0.000	0.004	0.020	0.976	0.000
F1 x Iberian	0.005	0.000	0.015	0.000	0.980

3. Low-density genotyping

The finally selected panel of 96 SNPs was used for genotyping two sets with 750 and 230 DNA samples, respectively, pursuing different objectives. A verification of the purebred *Iberian* origin was performed on a set of analyzed samples proceeding from 82 *Iberian* pigs, 61 *Duroc* and 607 uncertain purebred *Iberian* pigs named here as probationary animals. Fig. 2 represents a graphic summary of the results obtained from one of the performed analysis: admixture based on pre-defined clustering. Purebred pigs from the two breeds were considered as the origin populations and the proportion of alleles proceeding from these populations was estimated for the probationary samples. The presence of *Duroc* genes was observed in different proportions for some of the probationary animals and also for six of the assumed purebred *Iberian*. However, more satisfactory results were obtained from admixture analysis based on the mixture clustering of individuals (results not shown). This analysis assumes the maximum uncertainty about the samples and besides of a cluster grouping the *Duroc* pigs, the clustering of individuals reveals a likely hidden substructure of three different clusters for all the samples (668) of possible *Iberian* origin. The posterior admixture analysis indicates that only a 2.5% of these samples present *Duroc* genes with a probability lower than 0.05.

As it was already mentioned the set of 96 SNPs included *MC1R*4* and *MC1R*2* probes that can supply additional information for problematic breeding nucleus. Besides, an additional genotyping of *IGF2* g.3072A/G was carried out (according to the procedure described by Van Laere *et al*., 2003), for samples proceeding from the herds where *Duroc* genes were detected in putative Iberian animals. Both *MC1R*4* and *IGF2* g.3072A are exclusive alleles of the *Duroc* breed. Their presence in animals from the same breeding nucleus evidence introgression of *Duroc* genes. Moreover, the presence of *MC1R*2* allele was detected in a few animals which indicates introgression of *Large Black* in those herds.

The goal of the second trial was to validate the *Duroc* x *Iberian* origin of commercial crossbred pigs. The 230 analyzed samples corresponded to seven *Duroc* boars, 50 *Iberian* sows and 173 controlled F1 *Duroc* x *Iberian* animals. In this case, admixture analysis based on pre–defined clustering was performed. The *Duroc* boars and the *Iberian* sows were considered as the source of genes of the crossbred animals, and the proportion of genes proceeding from each one of them was estimated for each individual. The obtained results are summarized on Fig. 3. We could infer the presence of about a 10% of *Duroc* genes on four out of the 50 assumed purebred *Iberian* sows, with a probability lower than 0.001. Results also confirm the crossbred origin of all the *Duroc* x *Iberian* pigs. However, for some of the crossbred animals the inferred

percentage of *Duroc* genes was slightly different than the 50% expected. It can be explained by the low size of the origin breed samples that cannot represent all the within population diversity of these breeds for the selected genetic markers.

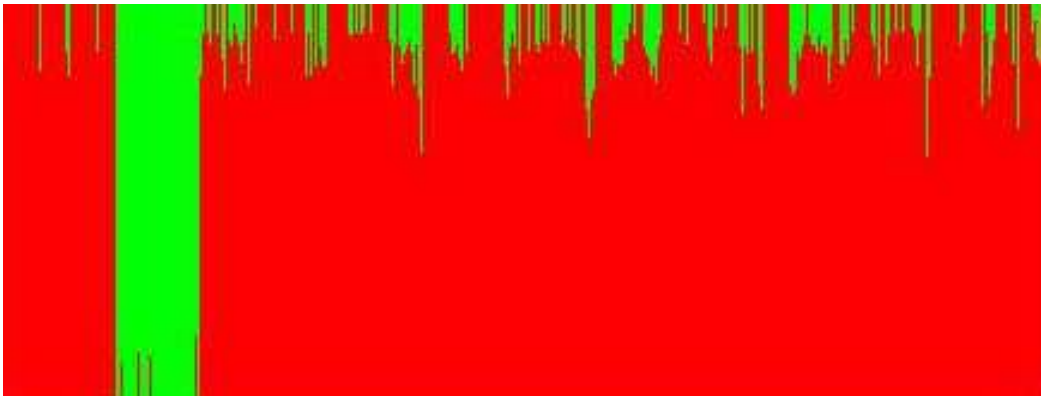

Fig. 2. Proportion of Iberian and Duroc genes on the 750 analyzed samples inferred from admixture based on pre-defined clustering.

Fig. 3. Proportion of Iberian and Duroc genes on the 230 analyzed samples.

It was mentioned before that the selection of SNPs was based on the ADT SNPScore file output that allows include those assays that are predicted to have a high likelihood of success. However this does not guarantee the complete amplification of all the SNPs because low-density genotyping uses VeraCode technology whereas high-density genotyping uses Infinium technology. Hence, we observed that a number of probes, ranging from 10 to 13 failed to produce amplification. Moreover, taken into account the inclusion of two *MC1R**2 probes on the

final set of 96 SNPs the actual number of useful probes varied between 82 and 85 instead the 96 selected SNPs.

IV – Conclusions

This study exemplifies how the recent advances in SNP discovery and high throughput automated genotyping methods can be applied to solve problems of authentication of genetic origin of *Iberian* pigs and their products. Low density genotyping of a moderate number of SNPs (< 90), with divergent frequencies between the *Iberian* and *Duroc* breeds, may be a powerful tool either to infer the purebred *Iberian* origin, to detect animals with low proportion of *Duroc* genes or to validate the *Duroc* x *Iberian* origin of commercial pigs or meat products. However, further research need to be carried out in order to build adequate databases for improving the usefulness of this procedure. We have planned to study a higher number of purebred and crossbred pigs, and to extend the panel of markers with new SNPs discovered using the next generation sequencing technology (Ramos *et al*., 2010).

Acknowledgements

The authors wish to acknowledge Juan García-Casco (AECERIBER) and Mario Gómez (Agroibéricos DeRaza) for providing samples, and technical assistance of Anna Mercadé (UAB). Financial support was received from the INIA CON09-006 and MICINN PET2008-0090 grants.

References

Alves E., Fernández A.I., Fernández-Rodríguez A., Pérez-Montarelo D., Benítez R., Óvilo C., Rodríguez C. and Silió L., 2009. Identification of mitochondrial markers for genetic traceability of European wild boars and Iberian and Duroc pigs. In: *Animal,* 3:9: 1216-1223.

Alves E., Fernández A., Barragán C., Ovilo C., Rodríguez C.and Silió L., 2006. Inference of hidden population substructure of the Iberian pig breed using multilocus microsatellite data. In: Sp. J. Agric. Res., p. 4(1), 37-46.

Corander J., Waldmann P. and Sillanpää M.J., 2003. Bayesian analysis of genetic differentiation between populations. In: *Genetics*, 163: 367-374.

Fernández A., Fabuel E., Alves E., Rodriguez C., Silió L. and Óvilo C., 2004. DNA tests based on coat colour genes for authentication of the raw material of meat products from Iberian pigs. In: J. Sc. Food Agric., 84: 1855–1860.

García D., Martínez A., Dunner S., Vega-Pla J.L., Fernández C., Delgado J.V. and Cañón J., 2006. Estimation of the genetic admixture composition of Iberian dry-cured ham samples using DNA multilocus genotypes. In: *Meat Science,* 72: 560-566.

López-Bote C.J., 1998. Sustained utilization of the Iberian pig breed. In:*Meat Science*, 49(Suppl.). S17–S27.

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

Ramos A.M., Megens H.J., Croijmans R.P.M.A., Schook L.B. and Groenen M.A.M., 2010. The use of next generation sequencing technology in the identification of specific SNPs for breed assignment and traceability of animal products. In: 9th WCGALP, Leipzig August 1-6 2010.

Rodríguez-Ramilo S., Fernández J. and Toro M.A., 2008. Análisis de agrupamiento de cerdos Ibéricos, Duroc y de sus cruces. In: *XV Reunión Nacional de Mejora Genética Animal*, Sevilla 19 a 21 de Junio de 2008.

Van Laere A.S., Nguyen M., Braunschweig M., Nezer C., Collette C., Moreau L., Archibald A.L., Haley C.S., Buys N., Tally M., Andersson G., Georges M. and Andersson L., 2003. A regulatory mutation in *IGF2* causes a major QTL effect on muscle growt in the pig. In: *Nature*, 425: 832-836.