



Worldwide genetic relationships of pigs as inferred from X chromosome SNPs

W. Burgos-Paz*, C. A. Souza^{*,†}, A. Castelló*, A. Mercadé*, N. Okumura[‡], I. N. Sheremet'eva[§], L. S. Huang[¶], I. C. Cho^{**}, S. R. Paiva[†], S. Ramos-Onsins* and M. Pérez-Enciso^{*,††}

*Center for Research in Agricultural Genomics (CRAG), Universitat Autònoma de Barcelona, Bellaterra, 08193, Spain. [†]EMBRAPA, Recursos Genéticos e Biotecnologia (Embrapa Genetic Resources and Biotechnology), Brasília, DF, 02372, Brasil. [‡]STAFF Institute, Society for Techno-Innovation of Agriculture, Forestry and Fisheries, Tsukuba, Ibaraki, 305-0854, Japan. [§]Institute of Biology and Soil Science FEB RAS, 159, Stoletia str, Vladivostok, 690022, Russia. [¶]Key Laboratory for Animal Biotechnology of Jiangxi province, Ministry of Agriculture of China, Jiangxi Agricultural University, Nanchang, 330045, China. ^{**}Subtropical Animal Experiment Station, National Institute of Animal Science, R.D.A., 175-6, O-Deung Dong, Jeju, 690-150, South Korea. ^{††}Institut de Recerca i Estudis Avançats de Catalunya (ICREA), Pg. Lluís Companys 23, Barcelona, Spain

Summary

The phylogeography of the porcine X chromosome has not been studied despite the unique characteristics of this chromosome. Here, we genotyped 59 single nucleotide polymorphisms (SNPs) in 312 pigs from around the world, representing 39 domestic breeds and wild boars in 30 countries. Overall, widespread commercial breeds showed the highest heterozygosity values, followed by African and American populations. Structuring, as inferred from F_{ST} and analysis of molecular variance, was consistently larger in the non-pseudoautosomal (NPAR) than in the pseudoautosomal regions (PAR). Our results show that genetic relationships between populations can vary widely between the NPAR and the PAR, underscoring the fact that their genetic trajectories can be quite different. NPAR showed an increased commercial-like genetic component relative to the PAR, probably because human selection processes to obtain individuals with high productive parameters were mediated by introgressing boars rather than sows.

Keywords animal genetic resources, phylogeography, pig, sex chromosome, single-nucleotide polymorphism, wild boar.

Introduction

Archeological and genetic evidence shows that the pig was repeatedly domesticated from wild boars starting in the Neolithic period, initially in eastern Asia and later in other regions of Asia and of Europe (Larson *et al.* 2005; Fang & Andersson 2006). Our knowledge of the origin of the different pig populations subsequent to domestication is still incomplete. In Asia, differences between northern and southern wild boar populations have been found, and multiple ancestors for domesticated Asian pig populations have been reported (Fang *et al.* 2005; Luetkemeier *et al.* 2010). In Europe, the wild boar was domesticated in the

Neolithic period in different places and eventually formed European local breeds (Giuffra *et al.* 2000; Ramirez *et al.* 2009). An introgression of Asian pigs into European breeds during the 18th and 19th centuries is well documented (Porter 1993; Giuffra *et al.* 2000). This event resulted in today's commercial European breeds, most of which are now widespread internationally. Other populations have been much less widely studied. In the Americas, the species was introduced during Columbus's second trip to the Hispaniola Island (today, Dominican Republic and Haiti) in 1493. The current main genetic component of these animals is supposed to be Iberian, with a possible subsequent introgression of other breeds. As for African populations, Ramirez *et al.* (2009) reported an Asian component in pigs in East African countries that was not observed in the western part of the continent.

Several approaches have been used to study population genetic relationships in pigs, including mtDNA sequences (Giuffra *et al.* 2000; Larson *et al.* 2005; Fang & Andersson 2006), the Y chromosome (Ramirez *et al.* 2009), and

Address for correspondence

M. Pérez-Enciso, Center for Research in Agricultural Genomics (CRAG), Universitat Autònoma de Barcelona, Bellaterra 08193, Spain.
E-mail: miguel.perez@uab.es

Accepted for publication 22 March 2012

autosomal genetic markers (Fan *et al.* 2002; Fang *et al.* 2005; SanCristobal *et al.* 2006). In contrast, X chromosome markers have been neglected so far. Yet, some characteristics make the X chromosome an interesting resource for the study of variability between populations (Schaffner 2004). Despite the structural differences between the X and Y chromosomes, they have a region that allows for segregation in meiosis, called the pseudoautosomal region (PAR). This region, located in the telomeric regions of the X chromosome arms, can recombine with the Y chromosome (Schaffner 2004). The rest of the chromosome constitutes the non-pseudoautosomal region (NPAR). Males carry only one copy of this chromosome, making its effective population size three-quarters that of the autosomes and increasing genetic drift for the NPAR-linked loci. Also, differences between effective population sizes in each gender can influence differences between populations depending on the migration sex ratio (Pool & Nielsen 2009; Casto *et al.* 2010).

The PAR X chromosome structure has been described in bovine (Das *et al.* 2009), horse (Raudsepp & Chowdhary 2008), dog and sheep (Toder *et al.* 1997). In pigs, the first PAR genes (*PRKX*, *KAL1*, and *STS*) were mapped to SSCXp/Yp by Quilter *et al.* (2002). Also, the porcine pseudoautosomal boundary (PAB) was recently mapped by qPCR between *SHROOM2* and *CLCN4* (Raudsepp *et al.* 2012). According to the Scrofa9 assembly (Archibald *et al.* 2010), the *Sus scrofa* X chromosome (SSCX) is 125.8 Mb long and contains 701 coding genes, 32 pseudogenes, and 178 non-coding RNAs (www.ensembl.org).

Here, we report on the analysis of worldwide pig population relationships using SSCX SNP polymorphisms in both the NPAR and the PAR, and we refine the localization of both regions in the porcine genome.

Materials and methods

Samples

A total of 312 DNA samples from 224 males, 87 females, and one unknown-sex individual from 39 pig breeds collected in 30 countries around the world were genotyped in this study (Table S1). These breeds include local breeds and wild boars from Asia and Europe, local creole and feral pigs from the Americas, and a smaller sample from African countries. Information about breed was not available for the latter region. Within commercial breeds, we included the most widely used breeds in the modern pig industry, mostly of European origin, as well as some British local breeds that have been introgressed with Asian germplasm (Table S1).

SNP selection, genotyping, and quality control

We selected 96 candidate SSCX SNPs from sequences described in the study of Wiedmann *et al.* (2008), chosen

according to their position (identified in the PAR) and to fulfill Illumina design requirements. The selected SNPs were genotyped using a Veracode Golden Gate Genotyping Assay Kit and analyzed in a Bead Xpress Reader (Illumina, Inc.). All genotypes were assigned using GENOME STUDIO software (Illumina, Inc.) and subsequently checked manually. Database pruning was conducted with PLINK (Purcell *et al.* 2007), excluding those SNPs with call frequency <90% and minimum allele frequency lower than 0.01 as well as those individuals with >10% missing genotypes. SNPs were annotated using Ensembl's biomart tool in the Scrofa10 assembly (<http://www.biomart.org/>).

Statistical analysis

With the aim of localizing the PARs and NPARs, observed (*Ho*) and expected (*He*) heterozygosities were obtained for each locus within sex using ARLEQUIN 3.5 software (Excoffier & Lischer 2010). Successive analyses were conducted for the PARs and the NPARs independently.

To test the hypothesis of genetic structuring in this sample assay, populations were grouped into seven meta-populations according to genetic, historic, and geographic relationships: Asia, Asian wild boar, local European breeds (Europe), European wild boar (which includes Tunisian wild boar), Africa, and American creole and feral pigs (Ramirez *et al.* 2009). Analysis of molecular variance (AMOVA) implemented in ARLEQUIN software was used to measure the degree of structuring between and within meta-populations. In males, we converted the NPAR haploid genotypes into diploid homozygous genotypes (Casto *et al.* 2010). We also explored alternative approaches such as reconstructing female haplotypes and diploidizing each, but the results did not change; *P*-values were calculated by performing 10 000 permutations.

Two classification approaches, supervised and unsupervised, were performed to characterize genetic relationships between populations. First, clustering of populations was performed with discriminant analysis of principal components (DAPC) implemented in the ADEGENET package (Jombart 2008). Second, probabilistic assignment to *K* groups with a Bayesian method in STRUCTURE software (Pritchard *et al.* 2000) was used. In this analysis, two population structures were evaluated: the first including the seven meta-populations and the second considering the 53 populations (country-breeds combination) independently. For each population structure, 10 simulations with *K* values ranging from 2 to 12, considering an admixture model and allele frequencies correlated, were evaluated. For each simulation, a burn-in period of 50 000 iterations was followed by 500 000 final iterations. To infer the optimal *K* value, STRUCTURE results were analyzed according to the delta *K* method described by Evanno *et al.* (2005).

Results

Quality control

A total of 56 SNPs (Table S2) passed quality control tests and were included in the analysis. The ratio of successfully genotyped SNPs ($56/96 = 0.58$) was much lower than that expected with this technology. The main cause of SNP discard was the high number of missing values per sample, but we also found nine non-segregating SNPs. Data are available in PLINK format upon request.

Localization of the presumable PAR in SSCX

To estimate the localization of the PAR, heterozygosity for each SNP was calculated within sex. The first 33 SNPs, located between the positions of 1.51 and 6.77 Mb, showed heterozygosities higher than zero in males ($H_o = 0.21$) and comparable to those in females ($H_o = 0.20$), whereas the remaining 23 SNPs, localized between 7.86 and 116.43 Mb, showed values equal to zero. In contrast, females showed H_o values higher than zero for those SNPs ($H_o = 0.15$). This suggests that the PAR likely is located between 0 and 7.8 Mb in the SSCX. The PAR and NPAR were analyzed separately.

Allelic frequencies and population structure

The heterozygosities in the PAR and NPAR for each of the seven meta-populations are shown in Table 1. Overall Africa, commercial and American populations showed the highest values of H_o in the PAR and H_e in NPAR. However, these also were the most extensively sampled populations. In contrast, local breeds and wild boars from Europe showed the lowest H_o and H_e . This is in agreement with previous results (e.g., Ramirez *et al.* 2009), which also described low variability levels in European pigs, especially in wild boars. Also, it can be noticed that the NPAR heterozygosities are lower than that of the PAR, except in European local pigs. Nevertheless, the reduction in heterozygosity is lower than expected, that is, 75%.

To gain further insight into population structuring and the effects of each potential structuring level (i.e., breed, country, and chromosome region), we performed several AMOVAS (Table 2). First, we considered all breed–country populations and second, the seven meta-populations described; finally, we analyzed each meta-population individually. Irrespective of classification, the same trends were consistently observed: the F_{ST} values were higher in the NPAR than in the PAR, and the among-population variance component also was higher in the NPAR than in the PAR—even if the within-population component dominates (Table 2). Population structure analyses suggest that 29.7 and 40.5% of the genetic variance was explained by the among-population variance component in the PAR and the NPAR, respectively, when all breed–country groups were considered (Table 2). When pigs were grouped by each of the seven meta-populations, the among-population variance component decreased, explaining only 22.7% and 25.0% of the total variance in the PAR and the NPAR, respectively. This suggests that the meta-population classification was less biologically meaningful than the breed–country arrangement.

Pairwise F_{ST} values among the seven meta-populations showed high differentiation in both the PAR and the NPAR between Asian and other populations (Table 3). Furthermore, pairwise F_{ST} values suggest that each SSCX region tells different, albeit similar, stories about the relationship between populations. In agreement with previous results (Ramirez *et al.* 2009), we found very low differentiation between local European breeds and wild boars; yet again, it was higher in the NPAR than in the PAR, 0.06 vs. 0.05, respectively. In contrast, differentiation between Asian domestic and wild boars was much higher, especially in the NPAR ($F_{ST} = 0.60$). As for derived African and American populations, our analyses indicate that, on average, the closest related breeds were the commercial breeds, followed by the European local breeds. This result agrees with a primary Iberian origin of the American populations followed by an important international breed introgression that has blurred the initial Iberian origin of American creole pigs.

Table 1 Sample size (n), missing values (NA), and observed (H_o) and expected (H_e) heterozygosities for each region of SSCX.

Population	n	%NA	Pseudoautosomal regions		Non-pseudoautosomal region H_e	H_{eNPAR} / H_{ePAR}
			H_o	H_e		
Asia	6	1.48	0.17 ± 0.22	0.27 ± 0.19	0.30 ± 0.21	1.11
Asia WB ¹	37	2.99	0.14 ± 0.16	0.21 ± 0.19	0.15 ± 0.17	0.71
Europe local	19	0.75	0.14 ± 0.12	0.17 ± 0.15	0.16 ± 0.20	0.94
Europe WB	42	1.19	0.14 ± 0.17	0.16 ± 0.19	0.14 ± 0.20	0.87
Commercial	66	0.81	0.28 ± 0.11	0.36 ± 0.11	0.25 ± 0.20	0.69
Africa	28	1.53	0.26 ± 0.15	0.31 ± 0.14	0.25 ± 0.19	0.80
America	114	1.20	0.21 ± 0.13	0.25 ± 0.15	0.20 ± 0.20	0.80
Total	312	1.33	0.21 ± 0.10	0.32 ± 0.13	0.25 ± 0.18	0.78

¹WB, wild boar.

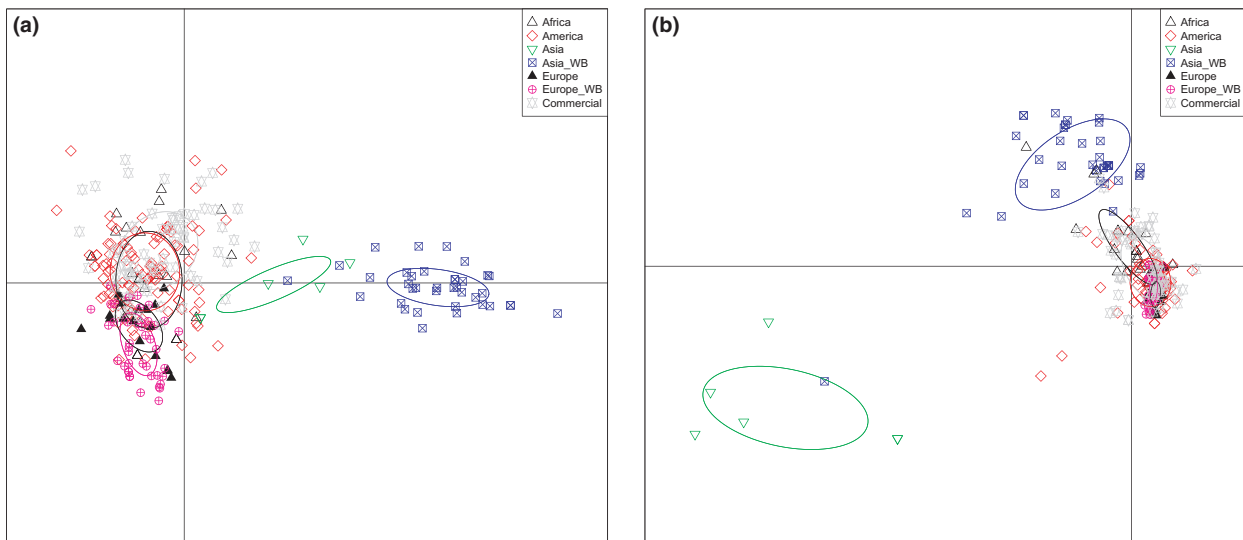


Figure 1 Scatterplot of DAPC analysis among meta-populations in the pseudoautosomal regions (a) and the non-pseudoautosomal region (b).

Probabilistic assignment to genetic groups

Differentiation of Asian populations from other groups was confirmed with DAPC in both SSCX regions (Fig. 1). Principal components retained explained 91% of the genetic variance in both the PAR and NPAR. This analysis (Fig. 1) corroborates the similarity in allelic frequencies between European local pigs and wild boars and also among the non-Asian populations. Asian local populations showed a cline-like relationship between Asian wild boars and commercial populations in the PAR (Fig. 1a). However, the discriminant analysis was more difficult to interpret regarding the NPAR (Fig. 1b). In this case, DAPC showed a higher similarity of European and commercial pigs with Asian wild boars than with Asian domestic pigs. We hypothesize that demographic effects in Asian domestication together with ascertainment bias effect and limited sampling could explain these results.

We also performed an unsupervised Bayesian probabilistic assignment of individuals to genetic groups, as implemented in *STRUCTURE* software. The most likely number of clusters was $K = 2$ when pigs were grouped by meta-population. In agreement with numerous previous results, Asian populations clustered in a separate group from the remaining populations (Fig. 2). Moreover, European domestic and wild boar pigs clustered in a second genetic group with a more than 97% membership proportion. In addition, commercial, African and American populations showed different degrees of admixture of European and Asian genetic components, the latter genetic component being in lower proportion.

To assess structuring in more detail, we also ran *STRUCTURE* with pigs grouped by breed and country. Here, $K = 3$ was the most supported value. Figure 3 shows the average cluster membership of pigs by breed and by chromosome

Table 2 AMOVA results for the different population structures in each SSCX region.

Group	Populations	Pseudoautosomal region			Non-pseudoautosomal region		
		Within populations	Among populations	F_{ST}	Within populations	Among populations	F_{ST}
Breed-country	53	3.26	1.37	0.297	1.50	1.02	0.405
Meta-populations ^{1,2}	7	3.89	1.14	0.227	1.99	0.66	0.250
Asia WB ³	5	2.31	0.54	0.189	1.11	0.95	0.461
Europe	5	2.38	0.24	0.092 ⁴	0.92	0.66	0.417
Europe WB	8	2.33	0.42	0.154	0.99	0.52	0.345
Commercial	16	4.83	0.76	0.135	1.98	1.01	0.338
Africa	5	4.54	0.31	0.065 ⁴	2.35	0.11	0.046
America	13	3.01	0.60	0.167	1.79	0.57	0.242

¹Includes the seven meta-populations Asia, Asia wild boar, Europe, Europe wild boar, Commercial, Africa, and Americas.

²Asia was excluded because all individuals were considered one population.

³WB, wild boar.

⁴No significant statistical differences were found. The rest of the F_{ST} values were highly significant.

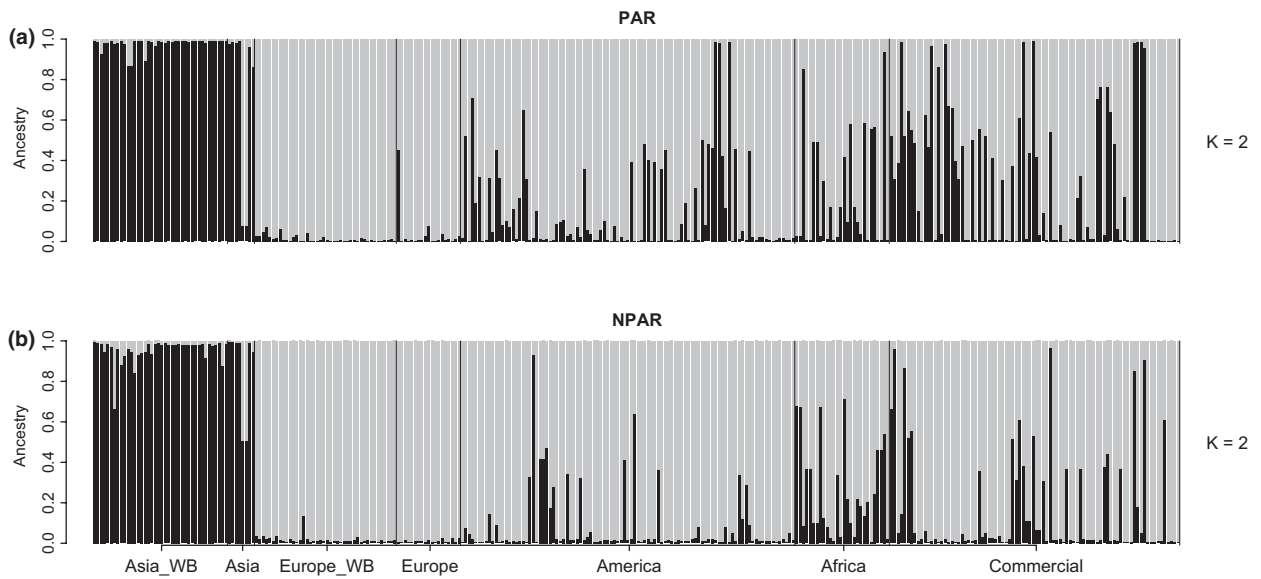


Figure 2 Bayesian probabilistic assignment of individuals considering meta-population structure in the pseudoautosomal regions (a) and the non-pseudoautosomal region (b). Breeds included in each group are described in Table S1.

region. Overall, Asian and European wild boars and Chinese pigs were rather homogeneous, especially for the PAR. Interestingly, Japanese wild boars from Ryukyu and the main islands were slightly different in the NPAR, in parallel with results from mtDNA (Wu *et al.* 2007) which suggest different founder origins. The native Korean pigs showed differences in allelic frequencies to commercial breed introgression signals. In fact, the origin of native Korean pigs is

mixed, with accredited introgression of commercial pigs within the original Korean germplasm. Interestingly, the composition of the PAR and NPAR was quite distinct, with a larger proportion of European clusters in the NPAR, again in agreement with a primary European origin via males.

European populations were made up of two genetic clusters in different proportions depending on the SSCX region. In the PAR, one cluster represented 80.7% of the

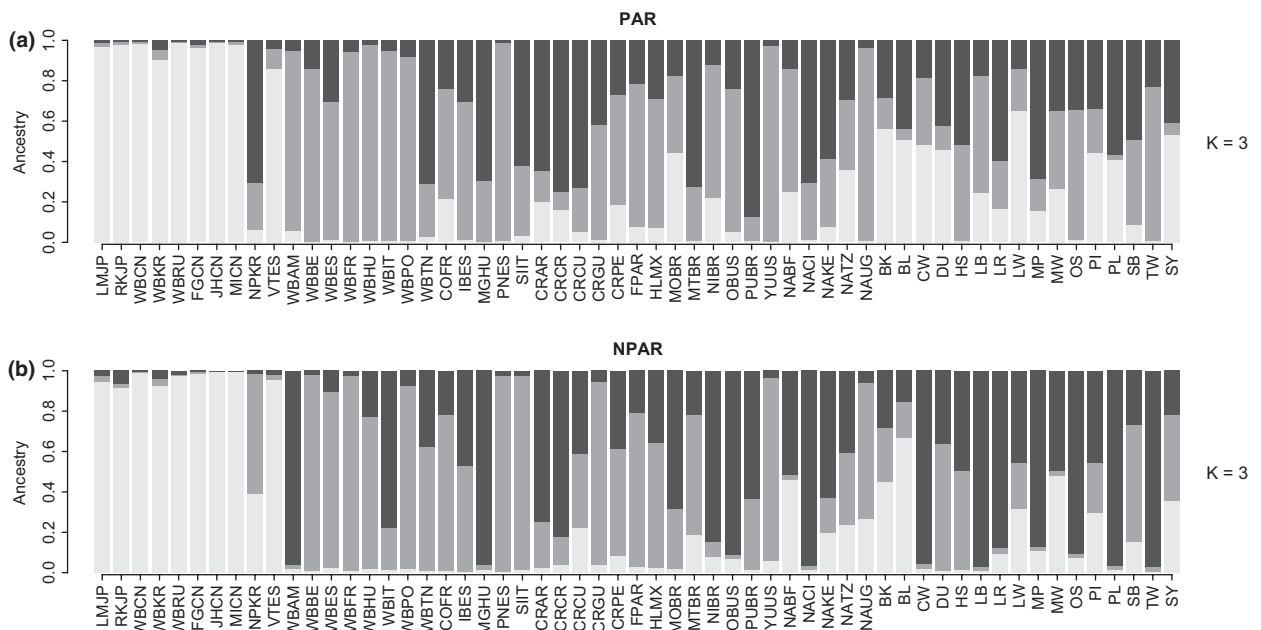


Figure 3 Average Bayesian probabilistic cluster assignment by breed and country in the pseudoautosomal regions (a) and the non-pseudoautosomal region (b). Breed and country codes as in Table S1.

Table 3 Pairwise F_{ST} values in pseudoautosomal regions (below diagonal) and non-pseudoautosomal region (above diagonal).

	Africa	America	Asia	Asia WB ¹	Europe	Europe WB	Commercial
Africa		0.10	0.52	0.27	0.16	0.21	0.02
America	0.03		0.64	0.45	0.05	0.15	0.04
Asia	0.18	0.28		0.60	0.64	0.69	0.57
Asia WB	0.44	0.50	0.22		0.51	0.53	0.34
Europe	0.05	0.03	0.36	0.58		0.06	0.07
Europe WB	0.09	0.08	0.41	0.61	0.05		0.13
Commercial	0.02	0.06	0.11	0.32	0.10	0.15	

¹WB, wild boar.

All F_{ST} values were significant ($P < 0.05$).

genetic component in wild boars; this value decreased to 69.3% in the NPAR, influenced by high differences in allelic frequencies of Italian and Armenian individuals. Mangalitza pigs showed allelic frequencies similar to that of Iberian pigs in the PAR, whereas in the NPAR, they were similar to Armenian pigs. As expected, commercial breeds were mosaics of European and Asian-like clusters in different degrees. However, there were differences between the PAR and NPAR, maybe as a result of differences in the sex ratio of the original breed genetic contribution to each population. For instance, in Duroc and in Pietrain, the Asian-like component decreased in the NPAR. Further, clustering was very different between the PAR and NPAR for breeds like Tamworth, Large Black, and Chester White. As in commercial breeds, three clusters with different proportions for each SSCX region were detected in America populations. Yucatan, Feral Argentinean, and Peruvian pigs were primarily European in origin, whereas an Asian-like component appeared in the NPAR for Brazilian Monteiro and Cuban pigs. In a previous work on the Y chromosome in African populations, we observed a larger Asian component in Eastern than in Western countries; however, this trend is not apparent from our results for SSCX.

Discussion

This study is a first approach to evaluate the usefulness of SNP polymorphisms in the pig X chromosome to detect population structure in a worldwide pig sample. Of 96 candidate SNPs, we retained only 56 polymorphic SNPs. These SNPs were initially obtained from pools of seven pig breeds, namely Duroc, Landrace, Yorkshire, Large White, Hampshire, Berkshire, and Pietrain, so they represent a limited variability in the whole species (Wiedmann *et al.* 2008).

Using indirect evidence of male heterozygosity, we determined that the PAR spans the first 7.8 Mb of the SSCX short arm. Quilter *et al.* (2002) mapped the *STS*, *KALI*, and *PRKX* genes to the PAR using fluorescent in situ hybridization (FISH); in the Sscrofa9 assembly, the first two genes are located at 2.89 and 4.00 Mb, which does suggest

a bigger PAR length in pig than in humans although smaller than in cattle (Ross *et al.* 2005; Das *et al.* 2009). Intriguingly, gene *PRKX* (ENSSSCG00000012833) is located on the Xq tail at 125.63–125.80 Mb. Further, Rohrer *et al.* (1994) reported three markers presumably located in the PAR; one of the markers, namely *SW949*, was aligned against Sscrofa9 assembly and matched two positions: 125.67 (P -value $1.8e-146$) and 125.77 Mb (P -value $6.3e-144$), again on the Xq tail. These data point to possible errors in the X chromosome Sscrofa9 assembly, given that a single PAR is suggested by FISH data.

The *SHROOM2* gene has been associated with PABs in pigs (Raudsepp *et al.* 2012). In the Sscrofa9 assembly, *SHROOM2* (ENSSSCG00000012104) is located at 5.26 Mb and within the bounds of the PAR suggested in this study. The *amelogenin* (*AMEL*) gene has been reported as the ancient Pseudoautosomal Boundary (PAB) in mammals (Iwase *et al.* 2003). This gene is located between 6.89 and 6.90 Mb in the pig genome (ENSSSCG00000012113), and male heterozygosity was zero in this region. Therefore, a comparative map of PAR in humans, cattle, goat, sheep, horse, and dog (Das *et al.* 2009), and the results reported here suggest that the PAR size in pigs is ~7 Mb.

The main feature of the X chromosome in population genetics studies is its low effective size, which theoretically reduces the genetic variability to three-quarters that of autosomes. Our data do show a reduction in the variability in the NPAR vs. PAR (Table 1), but this was much smaller than expected, except in commercial breeds—where the SNPs were ascertained. Although deviations from theoretical variability ratio have been reported (e.g., Gottipati *et al.* 2011), the most likely explanation for our results is SNP ascertainment bias. The analyses from the few published sequence data suggest that, in pigs, the ratio of X/A variability is actually much lower than 0.75 (Amaral *et al.* 2011; Esteve-Codina *et al.* 2011). Moreover, Ma *et al.* (2010) found a large (~31 Mb) recombination cold spot adjacent to the centromere of the pig X chromosome. Here, we found a strong reduction in H_o values in SNPs located between 61.52 and 92.32 Mb, in a position similar to that

in the study of Ma *et al.* Detailed analyses for this region showed a single haplotype in all samples except for two native Korean sows.

Overall, commercial populations showed the highest H_o and H_e values, followed by African and American populations. Given that the populations from these two continents are derived, a lower variability should be expected; these data then would suggest an important introgression of commercial breeds into the 'creole' populations. In general, H_o and H_e values were higher in Asian than in European populations, as reported in other studies using autosomal microsatellite markers (Fan *et al.* 2002; Luetkemeier *et al.* 2010). Moreover, as also reported by Ramirez *et al.* (2009), both domestic and European wild boar populations showed similar heterozygosity values for both SSCX regions and low differentiation, suggesting a possible gene flow between populations. In Asia, H_o and H_e values were higher in domestic populations than in wild boar, in contrast to results from Luetkemeier *et al.* (2010). However, these H_o and H_e values should be interpreted with caution, given the small sample size of some populations and differences in allele frequencies with commercial populations (ascertainment bias effect).

The X chromosome patterns of variability are strongly influenced by the sex ratio: genetic drift is very sensitive to the number of males, whereas the recombination rate depends on that of females. As revealed by STRUCTURE, a variety of different patterns between the PAR and NPAR, especially in commercial breeds, can be seen (Fig. 3). In some breeds, such as Berkshire or British Lop and most wild boars, the PAR and NPAR exhibited similar cluster compositions, but in most breeds, there were large differences between the PAR and NPAR, showing that their genetic trajectories are distinct.

Discriminant analysis of principal components and F_{ST} results showed a low differentiation between Africa, America, and commercial populations. Historical records document a strong international breed introgression in Africa, especially from the maternal side, from Asia and British breeds during the 18th and 19th centuries. Ramirez *et al.* (2009) found a primary European influence in western Africa, whereas eastern Africa exhibited an Asian genetic component. Our results cannot confirm these results; Eastern African populations (Kenya, Tanzania, and Uganda) share a large proportion of a European-like genetic group in both SSCX regions (Fig. 3). American population origins are European, mainly Iberian, with recent commercial breed introgression (Ramirez *et al.* 2009; Souza *et al.* 2009). Our results are consistent with this, showing low differentiation with the commercial population; in addition, all breeds showed a proportion of an Asian-like cluster (Fig. 3). Some populations (Yucatan, Argentinean feral pigs, and Guatemalan pigs) still preserve a high proportion of a European-like genetic cluster proportion in both regions, whereas others (Costa Rican or Argentinean

creoles) show strong influences of a commercial-like genetic component. A particularly interesting case was found in the Brazilian Moura and Monteiro breeds, where the Asian-like component appears in both populations but in different SSCX regions. The origin of the Monteiro breed is possibly European but has been crossbred with Asian breeds to improve reproductive parameters (Grossi *et al.* 2006). Overall, it is difficult to quantify the introgression impact of commercial breeds in American populations.

Conclusions

Thus far, the X chromosome has been poorly studied in the pig species. In this genotyping study, our findings are in agreement with a reduction in the variability in the NPAR, although attenuated because of SNP ascertainment bias. We also find indirect evidence that the PAR comprises approximately ~7 Mb of SSCX. Two main population clusters were detected, corresponding to Asian and European origins. Nevertheless, our results show that genetic relationships between populations can vary greatly between the NPAR and the PAR, underscoring the fact that their genetic trajectories can be quite different. The NPAR showed an increased commercial-like genetic component relative to the PAR, probably because of the fact that human selection processes to obtain individuals with high productive parameters were mediated by introgressing boars rather than sows. Further studies with a much denser SNP panel should allow the detection of selective sweeps in this important chromosome.

Acknowledgements

We thank the many groups and people who helped us to collect samples, in particular the Conbiand network and Embrapa Suínos e Aves, Embrapa Pantanal, Empresa Baiana de Desenvolvimento Agrícola, Universidade Estadual do Sudoeste da Bahia, and Universidade de Brasília. Thanks also to referees for their comments. WBP is funded by COLCIENCIAS (Departamento Administrativo de Ciencia, Tecnología e Innovación, Francisco José de Caldas fellowship 497/2009, Colombia). CAS was funded by a PhD grant from CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brazil) and Universidade Católica de Brasília, Brazil. This work is funded by grants AGL2010-14822 (MICINN, Spain) to MPE, CGL2009-09346 (MICINN, Spain) to SERO, and Consolider project (MICINN, Spain) to Centre of Research in Agricultural Genomics.

References

- Amaral A.J., Ferretti L., Megens H.J., Crooijmans R.P., Nie H., Ramos-Onsins S.E., Perez-Enciso M., Schook L.B. & Groenen M.A. (2011) Genome-wide footprints of pig domestication and selec-

- tion revealed through massive parallel sequencing of pooled DNA. *PLoS ONE* **6**, e14782.
- Archibald A.L., Bolund L., Churcher C. *et al.* & Swine Genome Sequencing Consortium (2010) Pig genome sequence—analysis and publication strategy. *BMC Genomics* **11**, 438.
- Casto A.M., Li J.Z., Absher D., Myers R., Ramachandran S. & Feldman M.W. (2010) Characterization of X-linked SNP genotypic variation in globally distributed human populations. *Genome Biology* **11**, R10.
- Das P.J., Chowdhary B.P. & Raudsepp T. (2009) Characterization of the bovine pseudoautosomal region and comparison with sheep, goat, and other mammalian pseudoautosomal regions. *Cytogenetic and Genome Research* **126**, 139–147.
- Esteve-Codina A., Kofler R., Himmelbauer H., Ferretti L., Vivancos A.P., Groenen M.A., Folch J.M., Rodriguez M.C. & Perez-Enciso M. (2011) Partial short-read sequencing of a highly inbred Iberian pig and genomics inference thereof. *Heredity* **107**, 256–264.
- Evanno G., Regnaut S. & Goudet J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular ecology* **14**, 2611–2620.
- Excoffier L. & Lischer H.E. (2010) ARLEQUIN suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564–567.
- Fan B., Wang Z.G., Li Y.J. *et al.* (2002) Genetic variation analysis within and among Chinese indigenous swine populations using microsatellite markers. *Animal Genetics* **33**, 422–427.
- Fang M. & Andersson L. (2006) Mitochondrial diversity in European and Chinese pigs is consistent with population expansions that occurred prior to domestication. *Proceedings of the Royal Society. B, Biological Sciences* **273**, 1803–1810.
- Fang M., Hu X., Jiang T., Braunschweig M., Hu L., Du Z., Feng J., Zhang Q., Wu C. & Li N. (2005) The phylogeny of Chinese indigenous pig breeds inferred from microsatellite markers. *Animal Genetics* **36**, 7–13.
- Giuffra E., Kijas J.M., Amarger V., Carlborg O., Jeon J.T. & Andersson L. (2000) The origin of the domestic pig: independent domestication and subsequent introgression. *Genetics* **154**, 1785–1791.
- Gottipati S., Arbiza L., Siepel A., Clark A.G. & Keinan A. (2011) Analyses of X-linked and autosomal genetic variation in population-scale whole genome sequencing. *Nature Genetics* **43**, 741–743.
- Grossi S.F., Lui J.F., Garcia J.E. & Meirelles F.V. (2006) Genetic diversity in wild (*Sus scrofa scrofa*) and domestic (*Sus scrofa domestica*) pigs and their hybrids based on polymorphism of a fragment of the D-loop region in the mitochondrial DNA. *Genetics and Molecular Research* **5**, 564–568.
- Iwase M., Satta Y., Hirai Y., Hirai H., Imai H. & Takahata N. (2003) The *amelogenin* loci span an ancient pseudoautosomal boundary in diverse mammalian species. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 5258–5263.
- Jombart T. (2008) ADEGENET: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405.
- Larson G., Dobney K., Albarella U. *et al.* (2005) Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science* **307**, 1618–1621.
- Luetkemeier E.S., Sodhi M., Schook L.B. & Malhi R.S. (2010) Multiple Asian pig origins revealed through genomic analyses. *Molecular Phylogenetics and Evolution* **54**, 680–686.
- Ma J., Iannuccelli N., Duan Y., Huang W., Guo B., Riquet J., Huang L. & Milan D. (2010) Recombinational landscape of porcine X chromosome and individual variation in female meiotic recombination associated with haplotypes of Chinese pigs. *BMC Genomics* **11**, 159.
- Pool J.E. & Nielsen R. (2009) Inference of historical changes in migration rate from the lengths of migrant tracts. *Genetics* **181**, 711–719.
- Porter V. (1993) Pigs. A Handbook to the Breeds of the World. Helm Information
- Pritchard J.K., Stephens M. & Donnelly P. (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959.
- Purcell S., Neale B., Todd-Brown K. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **81**, 559–575.
- Quilter C.R., Blott S.C., Mileham A.J., Affara N.A., Sargent C.A. & Griffin D.K. (2002) A mapping and evolutionary study of porcine sex chromosome genes. *Mammalian Genome* **13**, 588–594.
- Ramirez O., Ojeda A., Tomas A. *et al.* (2009) Integrating Y-chromosome, mitochondrial, and autosomal data to analyze the origin of pig breeds. *Molecular Biology and Evolution* **26**, 2061–2072.
- Raudsepp T. & Chowdhary B.P. (2008) The horse pseudoautosomal region (PAR): characterization and comparison with the human, chimp and mouse PARs. *Cytogenetic and Genome Research* **121**, 102–109.
- Raudsepp T., Das P.J., Avila F. & Chowdhary B.P. (2012) The pseudoautosomal region and sex chromosome aneuploidies in domestic species. *Sexual Development* **6**, 72–83.
- Rohrer G.A., Alexander L.J., Keele J.W., Smith T.P. & Beattie C.W. (1994) A microsatellite linkage map of the porcine genome. *Genetics* **136**, 231–245.
- Ross M.T., Grafham D.V., Coffey A.J. *et al.* (2005) The DNA sequence of the human X chromosome. *Nature* **434**, 325–337.
- SanCristobal M., Chevalet C., Haley C.S. *et al.* (2006) Genetic diversity within and between European pig breeds using microsatellite markers. *Animal Genetics* **37**, 189–198.
- Schaffner S.F. (2004) The X chromosome in population genetics. *Nature Reviews Genetics* **5**, 43–51.
- Souza C.A., Paiva S.R., Pereira R.W., Guimaraes S.E., Dutra W.M. Jr, Murata L.S. & Mariante A.S. (2009) Iberian origin of Brazilian local pig breeds based on *Cytochrome b* (MT-CYB) sequence. *Animal Genetics* **40**, 759–762.
- Toder R., Glaser B., Schiebel K., Wilcox S.A., Rappold G., Graves J. A. & Schempp W. (1997) Genes located in and near the human pseudoautosomal region are located in the X-Y pairing region in dog and sheep. *Chromosome Research* **5**, 301–306.
- Wiedmann R.T., Smith T.P. & Nonneman D.J. (2008) SNP discovery in swine by reduced representation and high throughput pyrosequencing. *BMC Genetics* **9**, 81.
- Wu G.S., Yao Y.G., Qu K.X., Ding Z.L., Li H., Palanichamy M.G., Duan Z.Y., Li N., Chen Y.S. & Zhang Y.P. (2007) Population phylogenomic analysis of mitochondrial DNA in wild boars and domestic pigs revealed multiple domestication events in East Asia. *Genome Biology* **8**, R245.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Breed codes and ISO country codes employed and breed x country sample size.

Table S2 Submission name, dbSNP accession number and position on X chromosome for each SNP.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.