Porcine Colonization of the Americas: a 60k SNP story

C.A. Souza[*], Y. Ramayo, H.J. Megens, M.C. Rodríguez, A. Loarca, E. Caal, H. Soto, M. Melo, M.A. Revidatti, S.A. de la Rosa, I.N. Shemereteva, N. Okumura, I.C. Cho, J.V. Delgado, S.R. Paiva, R.P.M.A. Crooijmans, L.B. Schook, M.A.M. Groenen, S.E. Ramos-Onsins, M. Pérez-Enciso

Introduction

The pig, *Sus scrofa*, is a foreign species in the American continent. It was first introduced by the Spanish and Portuguese colonizers and then successively in multiple, many unrecorded, events. Therefore, although the original introduced pigs should be related to modern Iberian pigs, the phylogeny of current creole pigs that now populate the continent is likely to be very complex. Besides, America harbors a wide diversity of environments ranging from hot tropical and subtropical to altitude (*altiplano*) climates, which has resulted in individuals particularly tolerant or resistant to local diseases and extreme environments distinct from those of temperate Europe. All this has occurred in a time span that is instantaneous when measured on an evolutive scale, providing a unique example of a fast evolutive phenomenon of adaptation. Here, we provide a genome wide overview of 'indigenous' local pig populations sampled across the American continent. In agreement with our initial conjecture, the current genotypic composition of these populations cannot be solely explained by a single Iberian origin. Interestingly, the Spanish America pigs seem to cluster separately from the Portuguese America breeds. Otherwise, there does not exist a clear geographic trend, and the observed pattern cannot be explained by an expansion of an initial colonization event, i.e., a typical stepping stone model.

Material and methods

Samples: The Illumina SNP60k porcine chip was used to genotype 135 pig samples from America (Table 1); almost all animals sampled lived outdoors (village pigs) or were semiferal and did not have an appearance as having been crossed with international white breeds. Yucatan and Ossabaw tissues were bought from Indiana University experimental herds, Moura and Monteiro (Brazil) belonged also to conservation nuclei. We also genotyped European breeds that may have contributed to American pigs: Iberian (n=15), Landrace

---

[*] *Universitat Autònoma de Barcelona* (Spain) / EMBRAPA (Brazil) / *Universidade Católica de Brasília* / *Universidad de Granma* (Cuba) / Wageningen University (Holland) / INIA (Spain) / *Colegio de Medicos Veterinarios y Zootecnistas de Guatemala* / *Universidad de Costa Rica* / *Universidad Nacional del Altiplano* (Peru) / *Universidad Nacional del Nordeste* (Argentina) / *Gobierno de Formosa* (Argentina) / Institute of Biology and Soil Science – Vladivostock (Russia) / STAFF Institute (Japan) / National Institute of Subtropical Agriculture (S. Korea) / *Universidad de Córdoba* (Spain) / University of Illinois (USA) / Centre for Agronomic Research (Spain) / ICREA (Spain)

(n=52), Duroc (n=22) and Large White (n=22). As outgroups, we included European and Tunisian wild boars (n=19) together with Asian wild boars (n=24) from Russia, Korea and Japan, as well as nine Korean native pigs. A total of 298 specimens were used in this study.

Genotypes: Starting with the 62,163 SNPs present in the chip, 45,080 were finally used to carry out the analyses. We discarded 11,591 SNPs assigned to sex chromosomes, without map or ambiguous locations. Further SNP pruning using PLINK (Purcell et al, 2007) criteria excluded SNPs with more than 5% of missing genotypes (3,830), SNPs without ancestral allele state information (1.496) and monomorphic SNPs in all samples (166). The ancestral allele was obtained from S. *verrucosus* (Megens et al, in preparation).

Table 1: Description of American samples

Country	Breed	Region	n
USA	Ossabaw pig	Ossabaw island (Georgia)	8
	Yucatán	Indiana	10
Mexico	Hairless	Nayarit / Veracruz	2
Cuba	Creole	Pinar del Río / S Spiritu / Granma	18
Guatemala	Creole	Baja Veracruz / Salamá	14
Costa Rica	Creole	Guanacaste / Alajuela	12
Peru	Creole	Puno – Titicaca area	16
Argentina	Creole / feral	Misiones / Esteros Iberá / Formosa	24
Brazil	Moura	Santa Catarina	9
	Nilo	Goias	2
	Monteiro	Mato Grosso	10
	Piau	Several states	10

Population Structure Analyses. We performed a Principal Component Analysis (PCA) using Eigenstrat (Prince et al., 2006). To examine relatedness among breeds, we analyzed SNP genotype frequencies with InSTRUCT (Gao et al, 2007) using the admixture model with correlated allele frequencies.

Results and discussion

The overall population structure visualized through the first two PCA components is in Figure 1A. Clearly, the first axis (explaining 34% variance) is primarily a geographic axis that separates the two main clusters, Asia and Europe, where the extremes are represented by the Iberian pig and Asian wild boars. Note that the Korean pig is relatively close to the European area. Within the much larger porcine hapmap study, Megens et al. (in preparation) have shown that the Korean pig is the closest breed to the European international breeds. The second axis (14% of variance) is associated with modern selection and clearly separates the commercial genetic groups (Duroc, Landrace and Large White) but with two strongly divergent clusters: Duroc vs. Large White / Landrace, whereas local pigs (Iberian) have not diverged too much from the European wild boar. All American pigs fit nicely in between and look relatively uniform when compared to species-wide variation; a result that is expected given the fact that they are derived admixed populations. A closer view of American populations only is in Figure 1B. The most remarkable observation is the lack of a general geographical pattern, e.g., distant populations overlap like Argentinean with Guatemala and often there is a large heterogeneity within populations. In some instances, like Peru, the samples are clustered in two groups. This can be explained by the fact that they were sampled in the North and South areas of Titicaca lake but we did not trace all the samples so we cannot confirm this hypothesis. Occasionally, like in Cuba, a sample was an outlier despite the phenotypic similarity to others in the same area. Nevertheless, the PCA axis 1 (7% of variance) indicate a gradient of variation between populations from Brazil and the Spanish colonization region with an overlap of Brazilian Moura breed with Argentina pigs, which are located in the frontier between both countries. The Yucatan and Piau pigs were the most differentiated groups, whereas the pigs from the Georgian island of Ossabaw clustered with the rest of Spanish America pigs, a result in agreement with the historical evidence that these animals descend partly from the Iberian pigs.

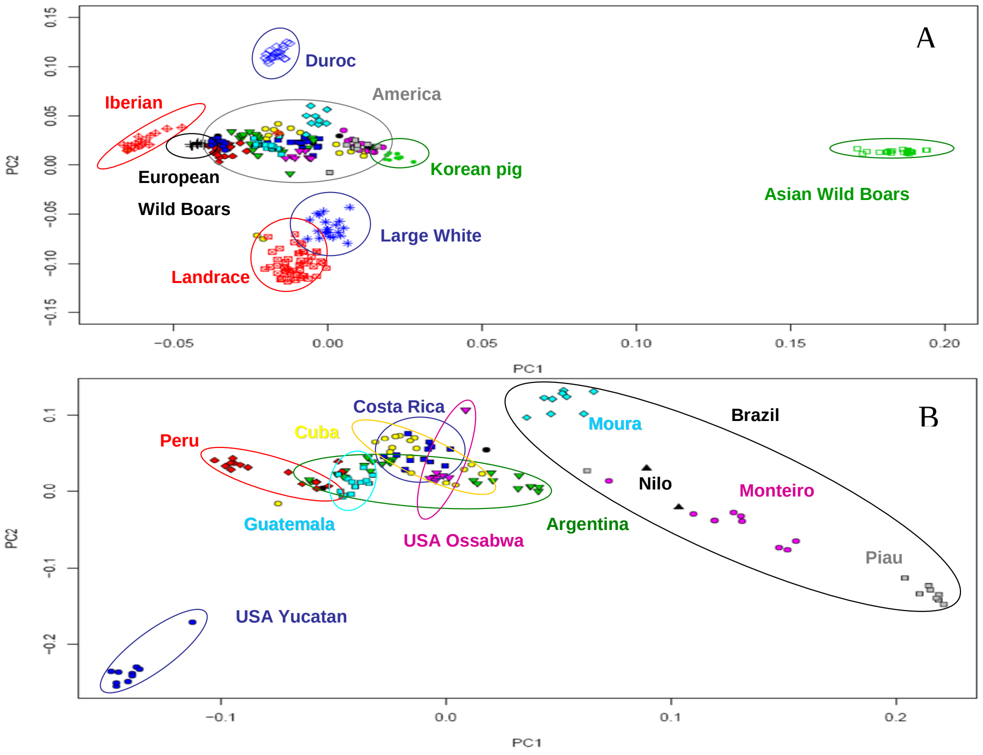

Figure 1: PCA analyses (A) All samples; (B) Analysis only with American populations.

A population clustering assessed using Bayesian inference revealed that the clusters are consistent with the known history of introduction of pigs in the Americas (Figure 2). The first level of clustering (K = 3) reflect a predomination of European origin and showed signatures of admixture with Asia and international breeds Landrace and Large White. The second level (K=5) reveals the contribution of Duroc and Korean pigs to American populations which is corresponded by the Fst index estimates (results not shown). All clustering levels clearly indicate the mixed nature of American populations.

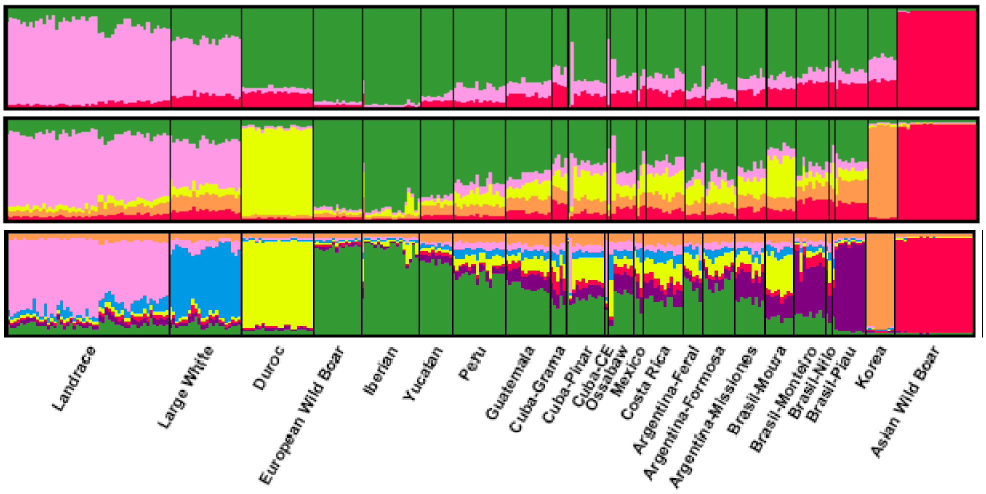

Figure 2: InSTRUCT Bayesian clustering with number of clusters K = 3, 5, and 7.

Conclusion and ongoing developments

As suspected, the history of American porcine populations is far from simple. Here we have confirmed, genomewide, an important but not unique contribution of Iberian pigs. In fact, an Asian component cannot be neglected, particularly in breeds like Piau. It remains to be disentangled whether this component has been introduced directly from Asia or indirectly through introgression via international breeds. A Duroc signal seems relatively minor, with exceptions like Moura (Brazil) or Cuba (Pinar del Río). We are focusing now on (i) identifying local signals of selection to specific climates, e.g., altitude or heat, and (ii) inferring a demographic model that accounts for SNP ascertainment bias; we suspect that its effect on some of the results can be strong.

Acknowlegments

We thank A. Mercadé and A. Castelló for having genotyped part of the the animals and network CONBIAND for facilitating sample collection. C. Souza thanks grants from CAPES and EMBRAPA (Brazil), Y. Ramayo is recipient of a PhD studentship from MICINN (Spain). Work funded by Consolider initiative to CRAG and AGL2007-65563 grants (Spain), EU SABRE project FOOD-CT-2006-01625, and USDA project 2007-04315 (USA).

References

Gao, H., S. Williamson, et al. (2007). *Genetics* **176**: 1635.

Price, A., N. Patterson, et al. (2006). *Nature* **38**: 904.

Purcell, S., B. Neale, et al. (2007). *Am J Human Genet* **81**: 559.