 Genome Data from a 16th Century Pig Illuminates Modern Breed Relationships  

W. Burgos-Paz,¹,² O. Ramírez,³ E. Casas,¹ M. Ballester,¹  
E. Bianco,¹,² I. Olalde,³ V. Novella,⁴ M. Gut,⁵ C. Lalueza-Fox,³ M. Saña,⁴ M. Pérez-Enciso¹,²,⁶  
¹Centre for Research in Agricultural Genomics (CRAG), Bellaterra, Spain, ²Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona, Bellaterra, Spain, ³Institut de Biologia Evolutiva (CSIC-Universitat Pompeu Fabra), PRBB, Barcelona, Spain, ⁴Departament de Prehistòria, Universitat Autònoma de Barcelona, Bellaterra, Spain, ⁵Centro Nacional de Análisis Genómico (CNAG), PCB, Barcelona, Spain, ⁶Institut Català de Recerca i Estudis Avançats (ICREA), Carrer de Lluís Companys 23, Barcelona, Spain. ¹ Equal contribution

ABSTRACT: Ancient DNA (aDNA) provides direct evidence of historical events that have modeled the genome of modern individuals. In livestock, resolving the differences between the effects of initial domestication and of subsequent modern breeding is not straightforward without aDNA data. Here, we have obtained shotgun genomes from a 16th century pig from northeastern Spain, together with three new modern genomes from Iberian pig, Spanish wild boar and a Guatemalan Creole pig. Comparison with genome data shows that the ancient pig is closely related to extant Iberian pigs and to European wild boar. Specific differentiation analyses allowed us to pinpoint genes that have been plausibly affected by initial domestication. Among those, we found genes involved in coat color and in performance, both known functions associated with early domestication process.

Key words: ancient DNA; domestication; next generation sequencing; pig.

INTRODUCTION

Most ancient DNA in livestock studies have focused on very early events, such as domestication, and have provided very limited evidence based primarily on a single locus like mitochondrial DNA (mtDNA). Disentangling the effects of domestication from those of modern breeding on the genome by simply comparing wild and domestic specimens (say wild boar vs. pig) is difficult because domestics carry signatures of modern breeding and selection. To that end, the sequencing of ancient domestic genomes predating the advent of breeds and modern artificial selection era is unavoidable.

To understand better these issues, we present the partial genome sequence of a 16th century pig. In addition, we also sequenced three new genomes to provide evidence on important historical and genetic events like domestication, admixing and the relationship with Creole pigs.

MATERIALS AND METHODS

Archaeological and modern samples. Ancient sample selected for sequencing was a tibia (diaphysis and distal epiphyses) of an adult retrieved from a very well-preserved assemblage (UE 10955) (Font et al. (2008)) at Montsoriu Castle, located in the province of Girona, in the north-east of the Iberian Peninsula. Additionally, three modern samples pertaining to a wild boar from the same Spanish region, an Iberian pig from the highly inbred strain Guadyerbas (Toro et al. (2008)), and an American Creole pig from Guatemala were used for populations relationships comparisons.

DNA extraction. All experimental procedures on ancient samples were performed in a dedicated ancient DNA laboratory (IBE-PRBB, Barcelona) as described elsewhere (Lalueza-Fox et al. (2007)). Three single end lanes of 100 bp length reads were sequenced at Fasteris (www.fasteris.com, Plan-les-Ouates, Switzerland) using HiSeq. Modern samples were sequenced in Centro Nacional de Análisis Genómico (CNAG, www.cnag.cat) also using HiSeq2000 Illumina platform.

Data Alignment and Quality Control. For ancient sample we first removed stretches of N’s and stretches of consecutive bases with 0, 1, or 2 quality scores from the 3’ and 5’ ends of the reads. Reads shorter than 30 nucleotides were discarded for further analyses. After quality control reads were mapped and aligned to the current pig genome assembly (Sscrofa10.2) using BWA (Li and Durbin (2009)). Furthermore, to improve the ancient DNA read mapping we followed recommendations in Schubert et al. (2012). For allele determination we considered only reads with minimum mapping quality of 20, and base quality (Phred score) of at least 30 if there was a single read covering that position or 20 with depths 2-5x. Positions covered with >5x were discarded, as being most likely caused by repetitive or copy number variant regions.

For the modern samples reads were aligned with BWA allowing for 7 mismatches. Genotypes were called using the SAMtools mpileup option (Li et al. (2009)) and filtered with vcftools.pl varFilter, all modern samples were analyzed together setting a minimum depth to 5x and a maximum depth of twice the average sample’s depth plus one, minimum map quality of 20 and minimum base quality of 20. Additionally, aligned bam files were obtained from Groenen et al. (2012) and from Iberian genome (Esteve-Codina et al. (2013)).

Array Genotyping Data. In order to position the 16th century pig among worldwide samples, we combined the genotypes inferred from sequence in the ancient pig with 60k SNP array genotypes from two biodiversity panels (Burgos-Paz et al. (2013); Manunza et al. (2013)). We performed principal component analysis (PCA) and a partially supervised admixture (Alexander et al. (2009)) analysis using K=7 clusters corresponding to origins
Iberian, wild boar, Duroc, Landrace, Large White, Chinese breeds and Hampshire.

**Differentiation analyses.** Given the small time span occurred since divergence between ancient and modern pigs, we searched for regions of extreme differentiation ($F_{ST}$), as this criterion is more sensible to recent events. We sought to investigate specific changes that may have occurred early in domestication by selecting windows and genes showing extreme differentiation in wild boar vs. the ancient and Iberian breeds. We computed $F_{ST}$ among wild boar vs. the Iberian and Ancient samples. Extreme regions in $F_{ST}$ larger than average + 3 SD and with at least five SNPs were analyzed with PANTHER (Protein Analysis Through Evolutionary Relationships) (Mi et al. (2013)).

**RESULTS AND DISCUSSION**

**Genome Quality.** An equivalent to a shotgun efficiency of 0.85% for ancient sample was obtained, similar to those reported in other ancient samples from the Iberian Peninsula (García-Garcerà et al. (2011)). After alignment, 9% of the *Sus scrofa* 10.2 assembly was covered with average depth of 2x (equivalent to a genome wide average depth ~ 0.11×). As for modern sequences, the average number of reads was 355,069,201 resulting in average depths 12-13× after filtering by base and map quality.

**Figure 1.** PCA using autosomal SNPs from the 60k array retrieved in the ancient sample, together with the corresponding genotypes from worldwide sample collection.

**Worldwide Context Inferred from SNP Arrays.** A total of 4,090 autosomal SNPs from the 60k array could be retrieved from the ancient sample. A Principal Component Analysis (PCA, Figure 1) of those SNPs broadly agrees with the original analysis that included the complete SNP dataset from Burgos-Paz et al. (2013) and Manunza et al. (2013). The PC1 axis is primarily geographical, separating Asian from European populations. The Near East (NE) wild boars are closer to European than to Asian pigs, and NE genetic structure grossly coincides with their geographic origin. The ancient sample was located within the modern Iberian pig cluster; and it does not show evidence of Asian admixing either. Unsupervised Admixture analyses suggested $K = 12$ as the optimum number of components. Results with this $K$ value suggest that the Near East component is completely absent from the ancient sample. The admixture analysis, instead, strongly supports a 100% Iberian component to the ancient pig.

**Figure 2.** NJ tree using genetic distances between the ancient and modern samples. WB, wild boar; IB, Iberian; AN, ancient; CR, Creole; HS, Hampshire; DY, Duroc; LW, Large White; LR, Landrace; PI, Pietrain.

**Genomewide Analysis.** We retained 794,514 autosomal SNPs without any missing value across eight modern samples and the ancient pig. We computed autosomal divergence (% of allele differences) between the ancient and the eight modern sequences, which again showed that the Iberian pig is the closest sample to the ancient pig, followed by Spanish wild boar, Hampshire and Creole (Figure 2). The length of ancient homozygous stretches (IBS blocks) shared with the Iberian was also the largest, followed at distance by wild boar (results not shown). All other samples, including Creole pig, were less similar to the ancient pig.

The PCA (Figure 3) has the first axis bounded by the wild boar (WB) and Large White (LW), which is the international breed with the largest Chinese component. The second axis primarily explains divergence with Duroc (DU). The ancient sample is closest to the Iberian pig and wild boar. This is interesting as it demonstrates that the Iberian pigs, at least the traditional strain analyzed here, have not been admixed with Asian pigs.
Differentiation signals. 157 genes were within the regions of $F_{ST}$ larger than genome-average plus 3 SD. Interestingly, after the Bonferroni correction for multiple testing, we observed a significant enrichment in genes related with “carbohydrate metabolic process” ($P = 0.0031$) and “disaccharide metabolic process” ($P = 0.0013$). Three important genes related with galactosidase activity were identified (QOQ237, Glb1l2 and Glb1l3). Among the top most differentiated regions analysis we found the Follicle-stimulating hormone ($FHSB$) and Tyrosinase-related protein 1 ($TYRP1$) genes. Other interesting genes that also appeared in our top 100 most differentiated regions analysis are v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog ($KIT$) and Cryptochrome 2 ($CRY2$), involved in reproduction. Given the small sample size and low coverage of our study, these results should be confirmed with larger sample sizes and higher depth.

CONCLUSION
Our data show that current and ancient Iberian pigs were similar, and different from current international breeds, either because an ancient divergence or caused by Asian introgression. Genome-wide data from an ancient pig, prior to modern intense selection for lean and growth traits, also provides us with an opportunity to understand selection at the gene level and separate them from those brought about by domestication and by Asian introgression. Among the highly differentiated genes between the ancient and Iberian vs. wild boar, we found genes related to coat color ($KIT$, $TYRP1$) and to reproductive performance (Galactose metabolism, $FHSB$, $CRY2$), in all likelihood among the first traits selected during domestication.

ACKNOWLEDGMENTS
Archaeological excavation funded by Direcció General d’Arxius, Biblioteques, Museus i Patrimoni and Museu d’Arqueologia de Catalunya. Work funded by 2010ACOM00030 (AGAUR) to MS, FEDER and BFU2012-34157 grant (Spain) to CLF, and AGL2010-14822 grants (Spain) to MPE. OR is a postdoctoral Researcher from the JAEDOC program cofounded by ESF, EB is recipient of a FPI grant from ministry of Research (Spain), IO of a predoctoral fellowship from the Basque Government (DEUI, Spain), WBP is funded by COLCIENCIAS (Francisco José de Caldas fellowship 497/2009, Colombia).

LITERATURE CITED