Heredity

ORIGINAL ARTICLE
Porcine colonization of the Americas: a 60k SNP story

W Burgos-Paz[1,19], CA Souza[1,2,19], HJ Megens[3], Y Ramayo-Caldas[1,4], M Melo[5], C Lemús-Flores[6], E Caal[7], HW Soto[8], R Martínez[9], LA Álvarez[10], L Aguirre[11], V Iñiguez[12], MA Revidatti[13], OR Martínez-López[14], S Llambi[15], A Esteve-Codina[1], MC Rodríguez[16], RPMA Crooijmans[3], SR Paiva[2], LB Schook[17], MAM Groenen[3] and M Pérez-Enciso[1,18]

The pig, *Sus scrofa*, is a foreign species to the American continent. Although pigs originally introduced in the Americas should be related to those from the Iberian Peninsula and Canary islands, the phylogeny of current creole pigs that now populate the continent is likely to be very complex. Because of the extreme climates that America harbors, these populations also provide a unique example of a fast evolutionary phenomenon of adaptation. Here, we provide a genome wide study of these issues by genotyping, with a 60k SNP chip, 206 village pigs sampled across 14 countries and 183 pigs from outgroup breeds that are potential founders of the American populations, including wild boar, Iberian, international and Chinese breeds. Results show that American village pigs are primarily of European ancestry, although the observed genetic landscape is that of a complex conglomerate. There was no correlation between genetic and geographical distances, neither continent wide nor when analyzing specific areas. Most populations showed a clear admixed structure where the Iberian pig was not necessarily the main component, illustrating how international breeds, but also Chinese pigs, have contributed to extant genetic composition of American village pigs. We also observe that many genes related to the cardiovascular system show an increased differentiation between altiplano and genetically related pigs living near sea level.
Heredity (2013) 110, 321–330; doi:10.1038/hdy.2012.109; published online 19 December 2012

INTRODUCTION

The pig, *Sus scrofa*, originated in Southeast Asia *ca* 5.3 −3.5 MYA (Groenen *et al.*, 2012) the species subsequently colonized the rest of Eurasia and North Africa (Larson *et al.*, 2005) but was absent from America before European colonization. Pigs, together with other livestock species like sheep, cattle or goats, were first introduced by Spaniards and Portuguese from the very beginning of colonization. Actually, the first recorded event of pig import into the new continent dates as early as the second Columbus trip (Crossby, 2003). On the Portuguese side, the first historical evidence of pig introduction dates from 1532 by Martim Afonso de Souza (Mariante and Cavalcante, 2006). According to Crossby (2003), 'the pigs adapted the quickest to the Caribbean environment', and the relevance of the pig as a source of meat from the very early days of conquest is well acknowledged (Elliot, 2007).

Nowadays, the porcine species is made up of a few highly specialized and widespread internationally breeds, well known for their leanness and high fertility. Although these international pig breeds have been replacing or intermixing with local American populations, numerous populations of direct descent from Iberian populations, so called 'creole', still are reported to exist. Currently, village pigs with a putative Iberian ancestry are common among many rural communities in most American countries. These animals are important to local communities not only because they provide food, but also because they are used as savings: they are sold when cash is needed. Normally, village pigs behave as commensal animals and feralization is also common, either because animals escape or because some areas were repopulated on purpose. This has occurred, for example, for hunting purposes in the USA, Argentina or South Brazil (Merino and Carpinetti, 2003). Therefore, although the original pigs introduced in the Americas should have been related to Iberian pigs and in particular to those of the Canary islands, the phylogeny and phylogeography of extant village and creole pigs that now populate the continent is likely to be very complex.

The study of village pigs is not only relevant from a social or historical perspective. America harbors a wide diversity of environments ranging from hot tropical climates to altitude (*altiplano*) dry

climates. Pigs thrive in all these areas, except in the very dry ones, resulting in animals adapted to extreme environments, quite distinct from those of temperate Europe. On a long, evolutionary scale, adaptation is usually characterized by an accelerated rate of non synonymous changes in protein coding regions, or in regulatory regions. Nevertheless, adaptation in the context of domestic species must stand primarily on standing variants, because of the short period of time considered. Pigs were brought into the Americas a few hundred years ago, a very short time on an evolutionary scale. Despite this, dramatic phenotypic changes have occurred. For instance, feral pigs develop much larger resistance to parasites or lack of food than pigs from international highly productive breeds. Some environments like high altitude in the *altiplano* or extreme and continuous heat in Cuba or North East Brazil also poses serious physiological challenges. The fact that adaptation must have occurred in a short time span suggests that rapid changes in allelic frequencies must have occurred, and also that excess of differentiation (for example, F_{ST}) can be a good proxy to detect these events (Akey et al., 2002; Vaysse et al., 2012).

Although some studies of American local pigs (Ramirez et al., 2009; Souza et al., 2009) or in other species like creole cattle (Delgado et al., 2011; Gautier and Naves, 2011) have been reported, they concern a small number of populations and/ or a few markers. In this work, by using a 60k single-nucleotide polymorphism (SNP) chip (Ramos et al., 2009), we provide the first comprehensive genomic analysis of village pigs from a wide sample of American countries, ranging from Cuba to North Argentina. This work was motivated by our interest in answering the following broad questions: (1) What is the origin of American village pig populations and their structure? Although admixing has certainly occurred, it is important to quantify its extent, for example, how much fraction of Iberian germplasm still exists, if any? (2) Is there any relationship between geographic and genetic distance, at least in the most isolated areas where admixing with modern breeds is likely to be rare? (3) And last but not least: is there any signal affecting the distribution of genotypic frequencies as a result of adaptation to extreme environments? All these questions bear relevance to both genetic and historical issues, and answering them will improve our understanding of how organisms adapt rapidly to extreme environments.

MATERIALS AND METHODS
Samples
We focused on sampling village pigs, for example, pigs living in a feral or semi-feral status from rural communities or assigned a 'creole' status, that is, thought to be of Iberian ancestry (Elliot, 2007). Sampling of relatives, for example, sibs, and animals showing evidence of intercrossing with international breeds was avoided. Our results showed that this was not always accomplished, as discussed below. A total of 206 animals from 14 countries were genotyped: the USA, Cuba, Guadeloupe, Mexico, Guatemala, Costa Rica, Colombia, Ecuador, Peru, Bolivia, Paraguay, Uruguay, Argentina and Brazil. These animals showed a wide variety of phenotypes, they lived outdoors, often in extreme climates and environments (Table 1, samples are described with more detail in Supplementary File 1).

Genotypes were also used from a wide hapmap catalog that are either potential founders of the American populations or outgroups (Table 1). These included local Mediterranean pigs from Spain (Iberian and Canary Islands pigs), Portugal (Bisaro) and Sicily (*Nero Siciliano*), international breeds (Duroc, Landrace, Large White, Hampshire) plus four breeds from East China, the most likely origin of pigs exported to other continents: Meishan, Jiangquhai, Jinhua and Xiang pig. Chinese pigs were genotyped because of the accredited partial Asian ancestry of international breeds and to assess whether there is any evidence of direct introgression of Chinese germplasm into the Americas. Finally, we genotyped Western wild boars.

Table 1 Pigs genotyped in this study

Country	Population/ breed	Location	Code	N (n)
Village pigs				
USA	Ossabaw pig	Ossabaw island	USOB	7
	Yucatan	Indiana	USYU	10
	Guinea hog	Several locations	USGH	15
Mexico (MX)	Cuino	Nayarit	MXCU	7
	Hairless	Several locations	MXHL	9
Cuba (CU)	Creole	Pinar del Río (West)	CUWE	5
		Sancti Spiritus (Center)	CUCE	1
		Granma (East)	CUEA	12
Guadeloupe (GP)	Creole	Guadeloupe	GPCR	4
Guatemala (GU)	Creole	Baja Verapaz, Salamá	GUCR	14
Costa Rica (CR)	Creole	Guanacaste, Alajuela	CRCR	12
Colombia (CO)	Zungo	Cereté (Córdoba)	COZU	10
	Creole	Alto Baudó (Chocó)	COCR	11
Ecuador (EC)	Creole	Loja	ECCR	5 (1)
Peru (PE)	Creole	Titicaca area	PECR	16
Brazil (BR)	Moura	Concórdia	BRMO	9
	Monteiro	Poconé	BRMT	10
	Piau	Bahia	BRPU	9
	Nilo	Goias	BRNI	2
Bolivia (BO)	Creole	Oruro	BOCR	6 (3)
Paraguay (PY)	Feral pig	San Pedro	PYFP	3 (3)
Argentina (AR)	Creole	Misiones	ARMS	9
	Feral pig	Esteros del Iberá	ARFP	6
	Semi feral	Formosa	ARFO	10
	Creole	Salta	ARNW	3 (3)
Uruguay (UY)	Cerdo pampa	Rocha	UYCP	1 (1)
Outgroup pigs				
Spain	Iberian	Several locations	ESIB	16
	Canarian	Canary islands	ESCN	4
Portugal	Bisaro	Several locations	PTBI	14
Italy	Black sicilian	Sicily	ITSI	4
Poland, Hungary, Tunisia	Wild boar	Several locations	WB	13
Denmark, Holland, USA	Duroc	Several locations	DU	20
Denmark, Holland, USA	Landrace	Several locations	LR	20
Denmark, Holland, USA	Landrace	Several locations	LW	20
UK, USA	Hampshire	Several locations	HS	14
China	Jiangquhai	Jiangsu	JQ	11
China	Jinhua	Zheijiang	JH	17
China	Xiang pig	Guizhou	XP	13
China	Meishan	Jiangsu	MS	17

$N=$ total sample size; $n=$ number of samples with a high percentage of missing values ($<20\%$) and removed from F_{ST} and ADMIXTURE analyses.

Genotyping and quality control
Samples were genotyped with the Illumina's porcine SNP60 BeadChip (Ramos et al., 2009). Raw data were visualized and analyzed with the Genome Studio software (Illumina, San Diego, CA, USA). Among the 62 163 SNPs initially present on the chip, 46 259 were finally selected using PLINK (Purcell et al., 2007) by pruning monomorphic SNPs or SNPs with an allele frequency below 0.05, SNPs located on the sex chromosomes, SNPs with more than 5% missing genotypes, SNPs not mapped on the Sscrofa10.2 assembly or SNPs for which the ancestral allele could not be identified. The ancestral allele was estimated
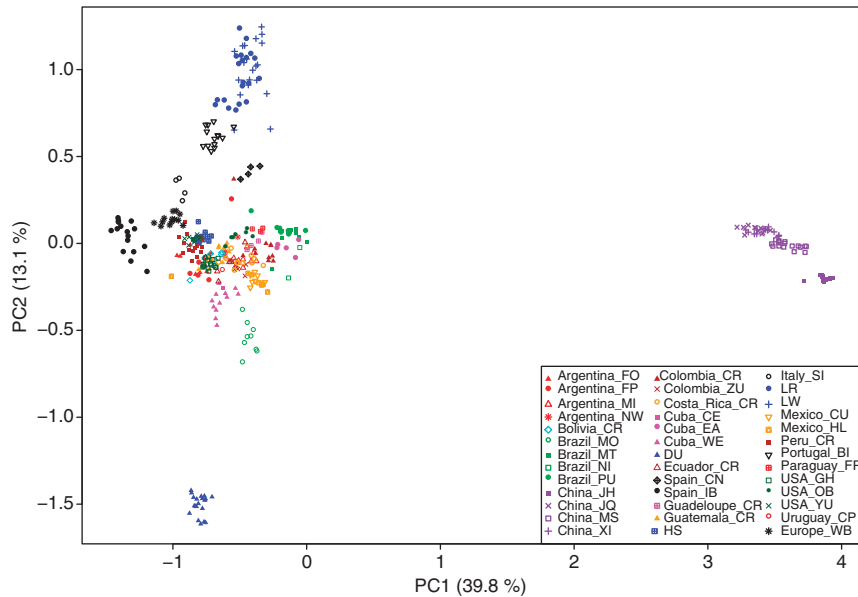
Figure 1 Principal component analysis using all samples.

based on *S. verrucosus* genotypes (Groenen *et al.*, 2012). Raw data had high-genotyping quality (call rate > 0.95) except for a few samples from Paraguay, Bolivia and Uruguay that were retained for their interest but not used in all analyses. Specifically, they were removed from the admixture and F_{ST} analyses.

Analysis

To visualize genetic distances between populations, principal component analyses (PCA) were obtained with smartpca program from EIGENSOFT (Price *et al.*, 2006). A complete relationship between individuals was drawn via a Neighbor Joining algorithm and visualized with DENDROSCOPE v. 2.7.4 software (Huson *et al.*, 2007) using pairwise identity-by-state genetic matrix distance (1-IBS) obtained with PLINK v. 1.07. To examine potential origins of each population, the Maximum Likelihood approach implemented in ADMIXTURE v 1.20 (Alexander *et al.*, 2009) was employed. First, ADMIXTURE was run in an unsupervised manner with a variable number of clusters $K = 2–20$. Lowest 10-fold cross-validation values were used to choose an optimum K-value, as suggested by the authors. Default termination criteria were used. We also considered a partial supervised approach where some samples were assumed to be of known ancestry K. Both PCA and ADMIXTURE were run by pruning markers in high linkage disequilibrium using the option –indep in PLINK. A total of 18 499 markers were selected for these analyses. To determine the relation between genetic and geographical distances of American pig populations, Mantel tests were performed in ADEGENET R package v. 1.3 −1 (Jombart, 2008) using exact sampling-site GPS coordinates and 1-IBS genetic distance matrix. The genetic differentiation between populations was assessed by the F_{ST} fixation index. Following Akey *et al.* (2010), we also considered a standardized F_{ST} measure. For each SNP and population k, we computed

$$d_k = \sum_{j \neq k} \frac{F_{ST}^{kj} - E[F_{ST}^{kj}]}{\text{sd}[F_{ST}^{kj}]}, \quad [1]$$

where $E[F_{ST}^{kj}]$ and $\text{sd}[F_{ST}^{kj}]$ denote the average value and s.d. of F_{ST} between populations k and j, respectively, over all SNPs. Statistics d was obtained either summing across all pairs of populations, that is, a global measure of differentiation, or between population k and their three nearest populations in terms of lowest F_{ST}. This latter statistics is similar to that proposed by Yi *et al.* (2010), and should be more powerful to identify selection than is Akey's statistics (Equation (1)) as it provides a direction to the allele frequency trajectory and reduces noise relative to the global test, where all population pairs are averaged. All populations with $N > 4$ were analyzed individually. Finally, some groups of populations, namely American populations—

excluding Brazil—vs European and international populations were also evaluated. In this case, we used Equation (1) as

$$d = \sum_k \sum_j \frac{F_{ST}^{kj} - E[F_{ST}^{kj}]}{\text{sd}[F_{ST}^{kj}]},$$

where subscripts k and j refer to populations in groups 1 (for example, America) and 2 (for example, Europe). The average d statistics over SNPs in non-overlapping windows of 1 Mb were plotted. Windows with an average d value above 2.0 s.d. (empirical distribution corresponding to the 1% extreme windows) in each population containing at least five SNPs were considered as candidate regions for selection. To complement the differentiation analyses, we also applied a selection test based on homozygosity extent (iHS). In this case, haplotypes were inferred with fastPHASE v. 1.4.0 (Scheet and Stephens, 2006) using subpopulation label information. Haplotype frequencies were then used to evaluate the presence of selective patterns for each SNP across the pig genome as described (Voight *et al.* (2006), and inferred using the rehh R-package v. 1.0 (Gautier and Vitalis, 2012). The 1 Mb windows with extreme average |iHS| scores across SNPs in that window were retained for further analysis.

Gene annotations within candidate regions were obtained by using the preliminary annotation of assembly 10.2 provided by ensembl (Groenen *et al.*, 2012). Overrepresentation of GO categories was determined with the DAVID database (Huang *et al.*, 2009), and pathway analyses were carried out with IPA, the ingenuity system (www.ingenuity.com).

Simulations

Given the difficulty of interpreting some of the results because of SNP ascertainment bias in the chip, we used coalescence simulation under a simplified model. We assumed four populations (Asia, International, Iberian and Creole, Supplementary File 2). Asian pigs diverged from European pigs 1 MYA (assuming one generation every two years), European pigs split into International and Iberian pigs ∼ 500 years ago. Both Iberian and International pigs contributed to creole pigs in approximately equal proportions, international pigs were introgressed with Chinese pigs (10%), whereas Iberian remained isolated. We studied variable Chinese contribution to creole pigs: 0, 1 and 10%. We ran coalescence simulations with mlcoalsim v. 1.9 (Ramos-Onsins and Mitchell-Olds, 2007). Out of the 10 000 independent loci simulated, we randomly selected 1000 such that the frequency spectrum in the International population was approximately flat, as observed in our data, in order to mimic ascertainment bias. Unsupervised and partially supervised
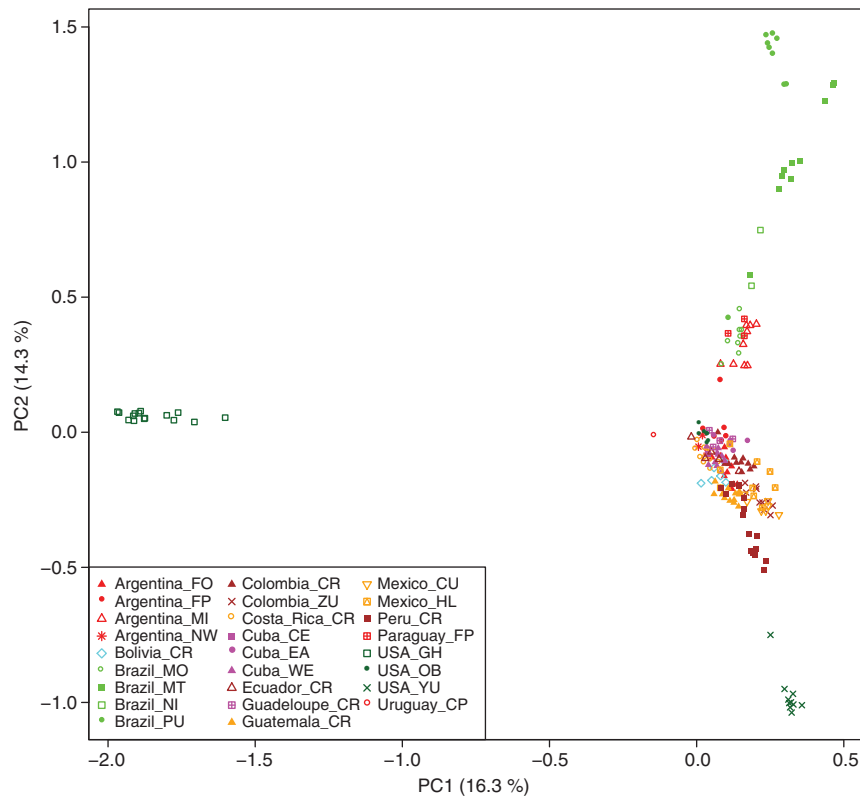
Figure 2 Principal component analysis using American samples only.

ADMIXTURE ($K = 3$) was applied to the simulated data, and we evaluated the bias in estimating the Chinese contribution.

RESULTS

A wide continent with shrunken genetic variation

We know from historical and genetic evidence that American pigs descend primarily from European pigs (Ramirez *et al.*, 2009; Souza *et al.*, 2009). The original flow began with pigs from the Iberian Peninsula and the Canary Islands, followed by a more recent intercrossing with international breeds. The PC analysis (Figure 1) partially agrees with this initial hypothesis. The first axis explains ∼40% of total variance and is predominantly geographical: It reflects the dramatic genetic distance between Asian and European populations. Chinese breeds and the Mediterranean Iberian breed represent both extremes on this axis. Large White and Landrace, international breeds known to have been introgressed with Chinese pigs, lie closer to Asia than do the Iberian pigs or European wild boars, which have remained isolated and unmixed with Asian germplasm. Nevertheless, these international breeds fall clearly within the 'European' neighborhood. Some Iberian pigs seem to be outliers. Although there is good evidence of sub-structuring among Iberian pigs (Alves *et al.*, 2006), we show later that this is caused by introgression from Duroc. Quite interestingly, the second axis, explaining a much lower fraction of variance (13%), primarily reflects the effects of artificial selection, with Landrace/ Large White vs Duroc breed representing the two extremes of the axes. The Iberian pig, an unimproved breed, lies broadly at the same level as wild boar on the second axis. This great distance between Duroc and other international or Mediterranean breeds is somewhat unexpected, as the original Duroc-Jersey breed was created in the USA with pigs of

several ancestries, including Iberian and African animals (Porter, 1993).

As for the American populations, these lie in a relatively wide area in between Iberian, Bisaro, Canary, Landrace and Large White breeds, a symptom of their predominant European descent. American pigs, nonetheless, do form a complex conglomerate of their own that is both explained by both PCA axes, the likely contribution of Iberian)) but also of Duroc, Landrace and Large White (the second axis). Therefore, American populations are clearly admixed. Interestingly, some American populations, like Brazilian Piau or Monteiro or East Cuban pigs, are closer to the Chinese cluster than other American populations. Similarly, Brazilian Moura is closer to Duroc than the rest of the American populations (See also Supplementary Files 3 and 4). An interesting observation is that Portuguese Bisaro and Canarian pigs cluster distantly from Spanish Iberian pigs, despite being from the same geographical or national origin. The traditional view (Porter, 1993) of porcine phylogeography is the presence of two main clades among European pigs: the Mediterranean clade represented, for example, by Iberian pigs, and the Celtic clade from Northern areas, represented by Landrace or Bisaro. Nevertheless, original Canarian pigs should not cluster with these Celtic breeds because they are supposed to represent primigenious pigs, maybe with African ancestry. We hypothesize that the modern Canary pigs we genotyped here are actually introgressed with international and/or Asian breeds. This interpretation agrees with historical records (García-Dory *et al.*, 1990) as well as with the report of Asian lineages in the mitochondrial DNA of Canary pigs (Clop *et al.*, 2004). It is plausible that Asian germplasm was introduced into Canary pigs by the British, who were influential in Canarian agriculture development during late nineteenth century (García and Capote, 1982).

Next, to gain in refinement and to focus on the main goal of this work, the PC analysis was run with American village pigs only (Figure 2). From a strict American point of view, the extreme breeds are Guinea Hog, Yucatan and Brazilian Piau. Our data support a distinct origin of Guinea Hogs from the rest of village pigs in the Americas and from either Yucatan or Ossabaw pigs. In terms of F_{ST}, the closest populations to Guinea Hog were Costa Rican and Formosa (Argentina) pigs, although both relatively high: 0.13 and 0.14, respectively. A point worth mentioning is that Ossabaw and Yucatan pigs were clearly differentiated (average $F_{ST} = 0.16$), despite an assumed shared Iberian ancestry. Yucatan was the closest breed to Spanish Iberian, whereas Ossabaw clustered among other American village pigs, and was in the same clade as Guadelupe pigs in the dendrogramme (Supplementary File 3). Our genotypic data support a clear separation between these breeds.

But perhaps the most noticeable observation from Figure 2 is that the second axis separates Brazilian from the rest of American pigs, with the exception of Moura. Although this partitioning is also seen in Figure 1, it is not so evident when all breeds are analyzed jointly. There also exists variability within Brazilian populations though. Piau was the most distantly related population to the rest of American village pigs, whereas Moura was the closest to, for example, Paraguayan feral pigs or Argentinean Misiones. Although it is tempting to interpret this as two separate routes of colonization, the Portuguese and the Spanish routes, this is not the sole explanation. We shall return to this point later.

A complementary view to that of the PCA is the dendrogramme pictured in Supplementary File 3. Although most pigs from the same population or breed tend to cluster together, exceptions are an Ossabaw pig within the Duroc clade or a Costa Rican pig mixed among Large Whites, both of these are probably recent admixtures with these international breeds. These animals, together with two outlier Iberian pigs, were removed to compute F_{ST} analyses. An interesting outlier is MXHL0140. This is a hairless Mexican pig from Veracruz province that clusters with Yucatan pigs, instead of with the rest of hairless pigs, which are positioned near the Duroc clade. Given the Mexican origins of Yucatan pigs, a plausible explanation is that this pig is actually a survivor of the ancient Mexican pigs currently perpetuated by US Yucatans, whereas extant Mexican 'traditional' breeds have been crossed with Duroc or other alien breeds. The results shown in Table 2, discussed below, suggest that the main source of introgression in Mexican pigs has been the Duroc breed.

Geography and genetic structuring
Neither PC analysis nor dendrogrammes (Figures 1 and 2, Supplementary File 3) reveal any broad clustering by geographic origin. For instance, Peruvian populations were positioned between Yucatan and Guatemalan pigs. Northeast Argentinean and Cuban pigs were scattered among other geographically distant pigs. In some cases, though, geography and genetics correlated: Paraguay feral pigs clustered with nearby Misiones pigs and Bolivian pigs were close to Peruvian ones. The two Colombian populations belonged to the same clade (Supplementary File 3), yet their F_{ST} was 0.19. In general, we did not observe that genetic distance or average F_{ST} was a proxy for geographic distance. To test the relation between geographic and genetic distances, a Mantel test was performed. Figure 3a shows the results for all samples. Except for pigs sampled in the same location, geographic distance explains very little of the variation in genetic distance. The coefficients of determination (r^2) were 0.09 and 0.04, respectively, when pigs from the same location were considered or not. Notice that a reduced genetic distance among pigs in the same

Table 2 Predicted cluster composition using partly supervised ADMIXTURE ($K = 6$)

Population code	IB	LR	LW	DU	HS	CN
USGH	0.00	0.00	0.00	0.00	1.00	0.00
USOB	0.37	0.20	0.11	0.12	0.13	0.07
USYU	0.99	0.00	0.00	0.00	0.01	0.01
CUCE	0.32	0.09	0.07	0.25	0.22	0.05
CUEA	0.41	0.13	0.09	0.13	0.08	0.16
CUWE	0.36	0.13	0.03	0.29	0.16	0.04
GPCR	0.36	0.09	0.23	0.16	0.07	0.09
MXHL	0.49	0.06	0.07	0.20	0.07	0.11
MXCU	0.52	0.00	0.12	0.18	0.06	0.11
GUCR	0.60	0.10	0.03	0.11	0.10	0.05
CRCR	0.36	0.12	0.13	0.19	0.14	0.07
COCR	0.50	0.13	0.09	0.11	0.06	0.12
COZU	0.72	0.02	0.02	0.10	0.04	0.11
ECCR	0.44	0.11	0.09	0.19	0.12	0.04
PECR	0.67	0.06	0.11	0.11	0.04	0.02
BOCR	0.50	0.07	0.10	0.18	0.14	0.01
ARFP	0.47	0.18	0.07	0.20	0.07	0.02
ARFO	0.56	0.11	0.07	0.16	0.07	0.02
ARMS	0.34	0.37	0.02	0.16	0.05	0.06
BRMT	0.18	0.69	0.00	0.03	0.00	0.09
BRMO	0.13	0.30	0.02	0.45	0.04	0.05
BRNI	0.16	0.55	0.00	0.17	0.01	0.12
BRPU	0.02	0.93	0.01	0.02	0.00	0.02
Average	0.41	0.19	0.06	0.15	0.12	0.06

Abbreviations: CN, China; DU, Duroc; HS, Hampshire; IB, Iberian; LR, Landrace; LW, Large White.
Population codes as in Table 1.

site can be due simply to sampling close relatives in the same or nearby villages.

Given the historical complexity of American colonization and because a shorter geographical distance does not necessarily imply a more active trade route, we circumscribed the analyses to a narrower, hopefully simpler space. Two regions were reanalysed separately. First, the North Argentinean pigs (Misiones, Corrientes, Formosa and Salta provinces) together with nearby Paraguay feral pigs; and second, Central America (Mexico, Guatemala and Costa Rica). It can be seen, again, that correlation vanishes and even becomes slightly negative when pigs from the same spot are removed (Figures 3b and c). In Argentina, the r^2 was 0.16, but vanished ($r^2 < 10^{-3}$) when pair of pigs with a geographic distance of zero were removed. Similarly, the r^2 in Central America were 0.23 and 0.15, respectively, in each of the two analyses. This suggests that pigs from nearby locations are genetically related, maybe because local communities exchange animals, but also that pigs can be imported from different remote or foreign locations. Overall, a classical stepping-stone model is not applicable to this human-mediated livestock colonization, where geographic distance explains only a tiny fraction of total genetic variability. Note that this pattern could also reflect an incipient pattern of breed formation. In fact, except for Brazil, legislation on local breeds or populations is very recent in Latin American countries and in general not strictly enforced.

Next, ADMIXTURE was used to characterize genetic structure across American village pigs and their putative ancestral breeds. The unsupervised method detects $K = 14$ clusters as an optimum partition number (Figure 4a). This suggests an underlying highly complex genetic structure, despite the apparent uniformity within American village pigs portrayed by PC (Figures 1 and 2). In Figure 4a, a number
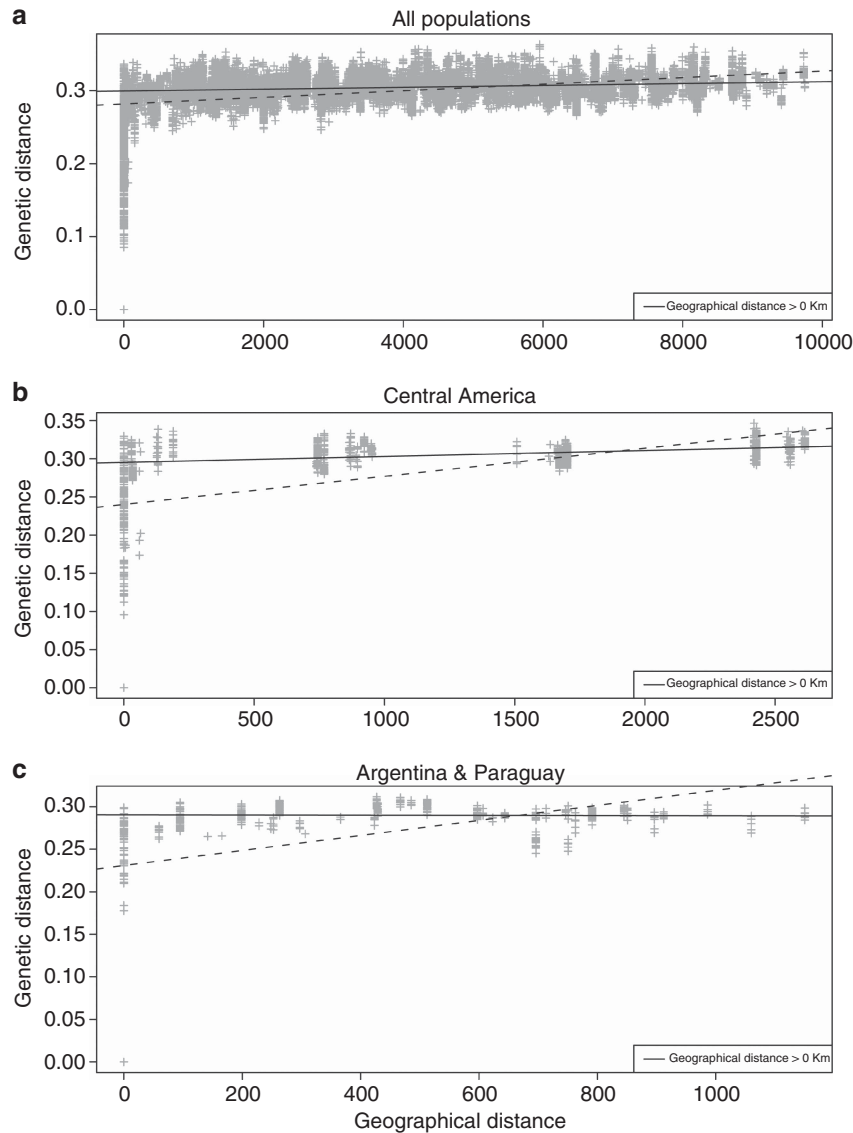
Figure 3 Correlation between geographic and genetic distances: (**a**) all American samples; (**b**) Central American samples; (**c**) North East Argentina and Paraguay. Continuous (dashed) line is regression not including (including) samples from the same location (geographic distance 0).

of populations are identified as homogeneous, that is, Iberian, Duroc, Hampshire, Guinea Hog, Yucatan, Cuino, Piau and Chinese breeds, and, to a lesser extent, Landrace, Large White and Colombian Zungo. Other populations, primarily American, but also Bisaro and Canary, are admixed. In agreement with previous results, the method does not detect a strong structuring between Iberian and European wild boar (Ramirez *et al.*, 2009; van Asch *et al.*, 2012). If we take a uniform cluster assignment as a signature of recent isolation, Figure 4 suggests that Guinea Hog, Yucatan, Cuino, Colombian Zungo and Piau would be the American populations that show less or no degree of recent introgression. Except for Cuino, for which there are no official records, this agrees with the fact that these are established breeds with their own breeding programmes.

It is also illuminating to consider a partially supervised analysis. In this case, some pigs were assigned a predefined cluster. We ran cases $K = 13$ and $K = 6$. With $K = 13$, a predefined cluster was assigned to those pigs from uniform breeds as suggested by the unsupervised analysis (Figure 4a). A value $K = 13$ was used instead of $K = 14$

because no population was assigned fully to a fourteenth cluster. This analysis (Figure 4b) suggests a putative Brazilian Piau cluster to be predominant among Brazilian breeds, primarily in Monteiro, and where Moura is largely introgressed with Duroc. Similarly, a hypothetical Colombian Zungo cluster would be present among many American village populations. A problem with this supervised analysis is that the large number of clusters assumed, without considering historical processes, makes interpretation difficult. To simplify matters, we considered a smaller number of clusters ($K = 6$) that represent all known major origins of American village pigs: Iberian, Landrace, Large White, Duroc, Hampshire and Chinese pigs. Therefore, we make the simplifying, but reasonable, assumption that the genetic make-up of American pigs can be largely explained in terms of these six origins. The analysis (Figure 4c) still shows that American populations are clearly admixed but to different degrees; heterogeneity within populations is also evident. Assuming the hypothesis of these six clusters representing the main ancestral populations of American village pigs, the Iberian pig represents an
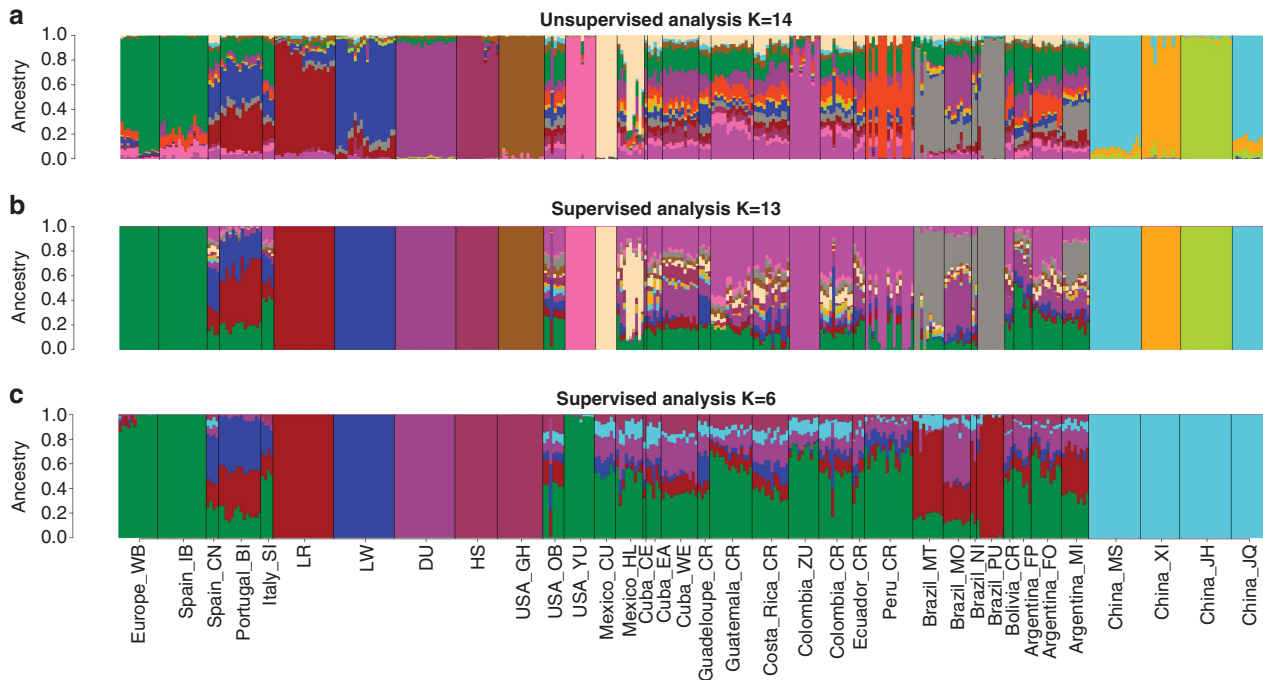
Figure 4 ADMIXTURE analyses: (**a**) unsupervised, $K = 14$; (**b**) supervised, $K = 13$; (**c**) supervised, $K = 6$.

important component, especially in Yucatan, Peruvian and Colombian Zungo pigs.

Nevertheless, this Iberian component varies largely in importance across populations. In fact, PCA analyses (Figure 1) suggests that a pure Iberian ancestry is unlikely. A more specific analysis with ADMIXTURE (Figure 4 and Table 2) confirms that American pigs are partly of Iberian origin, but that this origin is not necessarily predominant, except Yucatan or perhaps Peru and Colombian Zungo. The inferred average Iberian contribution to American village pigs is ∼40%, ranging from Yucatan (∼99%) to Brazilian Moura or Piau (∼0%). Supplementary File 4 shows the F_{ST} between the putative main founders (Iberian and international breeds) and the genotyped American populations. Except for a few populations studied, namely Yucatan, Peruvian altiplano, feral Argentinean pigs and Colombian Zungo, Iberian was not the closest breed. Overall, American populations were equidistant between Landrace, Large White and Iberian breeds, whereas Duroc is the most distant one.

ADMIXTURE also suggests that an Asian component cannot be ruled out for several populations (Table 2 for supervised $K = 6$ and Supplementary File 5, unsupervised $K = 14$). European wild boar is our negative control, and ADMIXTURE does report <1% of Chinese assignment, as in Iberian and Sicilian pigs. Also in agreement with records, Large White and Landrace have variable levels of introgression from Chinese breeds. Bisaro and Canary pigs are likely to be admixed recently with international breeds, the latter displaying a considerable influence of Chinese pigs, in agreement with previous mitochondrial DNA results (Clop *et al.*, 2004). Within the Americas, the breeds with little or no inferred Chinese introgression are Yucatan, Ossabaw, Mexican hairless, Bolivian, Peruvian and some Argentinean pigs. In contrast, Eastern Cuba, Pacific Colombian creole and some Brazilian pigs (Nilo predominantly) may have a non negligible percentage of Chinese germplasm. The closest Chinese breed, in terms of F_{ST}, was consistently the Jiangqhuai breed (Supplementary File 4). This breed is originally from the Taihu lake area, the origin of

the most prolific Chinese pigs, and is also renowned for its good meat quality. In agreement with reports (Porter, 1993), this supports the belief that Chinese pigs were imported to improve upon the characteristics of local European pigs.

Three levels of Chinese migration into the Americas were compared via simulation. Chinese contribution was overestimated with the unsupervised ADMIXTURE, whereas values are better estimated with partially supervised ADMIXTURE, unless migration is very small (1%, Supplementary File 6). For instance, in the unsupervised analysis, the Chinese contributions were estimated to be 8.7, 10.5 and 18.2% when true migration rates were 0%, 1% and 10%, respectively. The equivalent supervised estimates were 4.1%, 4.5% and 11.6%, respectively. In contrast, the contributions of Iberian and International pigs were reasonably well estimated. The simulated site frequency spectra, together with the observed spectra from some populations in our data is in Supplementary File 7, and shows that the simulated model reproduces, approximately, the observed data.

Signals of adaptation: size and altitude

First, we investigated whether there is evidence for any common selective signature between American village pigs, excluding Brazilian samples and minipigs (Yucatan, Cuino and Guinea Hogs) and their European and international ancestors. Supplementary File 8 shows over-represented GO categories ($P < 0.01$) within genes in 1 Mb windows with average d statistics greater than 2 s.d. over the mean, that is, ∼1% extreme windows. Despite the apparent heterogeneity among breeds and populations, it is noteworthy that a few GO categories were highly over-represented. These ontologies are related to development (specifically limb morphogenesis), vitamin A metabolism and behavior. Therefore, this may suggest that a common response among American populations has involved modifying their pattern of development and, perhaps, also by how they respond to external stimuli.
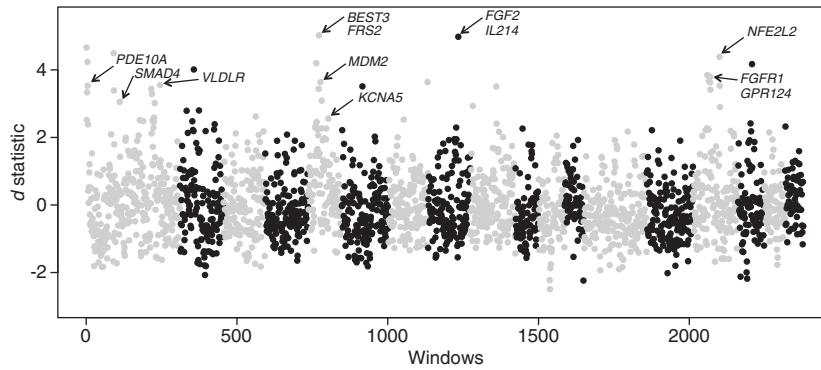
Figure 5 *d* profile in the Peruvian population showing the position of some relevant genes. Each dot represents a 1 Mb window containing at least five SNPs.

Adaptation to altitude was specifically explored. Among the environmental challenges posed by the American continent to livestock life, the Andean altiplano is probably one of the harshest. Figure 5 contains the profile of the *d* statistics; there were 87 extreme windows ($d > 2$ s.d. over the mean) that contained 301 annotated genes. The most significant enriched category was the peptidyl-citrulline biosynthetic process (Supplementary File 8); interestingly, citrulline has been reported to relax blood vessels and may improve adaptation of blood circulation to altitude. It is also remarkable that, among the genes in extreme F_{ST} windows, we found several genes known to be involved in response to hypoxia (*SMAD4, MDM2, VLDLR, KCNA5*) although their corresponding GO categories were not significantly enriched. A detailed inspection showed that a total of 54 out of the 301 annotated genes are also involved in the cardiovascular system phenotype and physiological characteristics of the mammalian heart and blood vessels (Supplementary File 9), and IPA analyses showed that over 70 of the 301 genes were involved in cardiovascular or hematological diseases (Supplementary File 10).

The alternative statistics iHS resulted in far fewer outlier windows, may be because of detection of homozygosity requires denser SNP spacing than that employed here. Only three windows (Supplementary File 11) were over 1.4 s.d. and only the most significant window, that on SSC2, overlapped with the differentiation analysis (Figure 5). There are no reported genes in current porcine assembly for this window. Yet, analysis using our own unpublished RNAseq data allowed us to identify several unannotated genes. A subsequent annotation with blast2go, Gotz *et al.* (2008) identified gene *EMR1*, which is involved in respiratory diseases. The second most extreme window (SSC9) contained three genes, *TBP12, GNG11* and *GNGT1*, which are involved in blood coagulation.

DISCUSSION
We present the most extensive genomic analysis of American creole livestock species to date. The samples genotyped represent a comprehensive overview of the extant genetic variability in American village pigs; these pigs are, importantly, adapted to a wide array of climates and environmental conditions, for example, heat, altitude or diseases. With data at hand, most American populations showed a high degree of admixture, greater than their parental populations, that is, Iberian, Large White, Landrace or Duroc, together with a putative direct Chinese influence. The genetic landscape that we observe is that of a complex conglomerate, in contrast to similar analyses in other species with a much more marked structure, such as dogs. Nevertheless, the analyses of village dogs have also proven to be much more

complex than those of well-established breeds (Boyko *et al.*, 2009). In particular, we did not observe that genetic distance or average F_{ST} was a proxy for geographic distance, likely because livestock populations have a great mobility and corresponding complex genetic histories.

There are two potential problems regarding the interpretation of results. First, the limited number of individuals sampled and second, SNP ascertainment bias. While small samples may not be so relevant when the number of markers is high (Willing *et al.*, 2012), the consequences of SNP ascertainment bias are, however, much more difficult to assess. Theoretical and simulation work have shown that 'PCA projections from genotype data will be similar to PCA projections from resequencing data, but will typically be larger in magnitude' (McVean, 2009), that is, distances will be biased, although the topology will be conserved. To explore, even if tentatively, this issue we ran coalescence simulations. Although our goal was not to comprehensively analyze all potential models, the simulations suggest: (i) that a partially supervised approach is more reliable than an unsupervised method, and (ii) that the estimate of Chinese influence can be biased upwards when the true migration is zero or very small ($\sim 1\%$) but are more accurate as migration rate increases. The supervised ADMIXTURE estimates of Chinese influence with are reasonably large in some populations, notably in Eastern Cuban, Guadeloupe, Mexico, Pacific Colombian and Brazil's Nilo. Therefore, a Chinese contribution in these cases would not be an artefact. Although there is evidence of direct introgression from Asia into the Americas (Ramirez *et al.*, 2009; Lemus and Ly, 2010), this Asian influence might also be indirect, mediated by international breeds. Complete resequencing and comprehensive simulations will help to elucidate this issue.

The term 'creole' (Spanish *criollo*, Portuguese *crioulo*) is used to refer to descendants from the Iberian Peninsula (Elliot, 2007). As with humans, the traditional view is that 'creole' pigs are descendants of pigs imported from the Iberian Peninsula. However, the actual ancestry of the many breeds termed 'creole' throughout the Americas is unknown. Our data suggest that this contribution has been dramatically attenuated in current village pigs. If the contribution of the Spanish Iberian pig to American creoles is smaller than anticipated, we can speculate whether creole pigs have undergone a dramatic introgression with international-breed pigs or whether extant Iberian pigs are different from those of several centuries ago. We favor the first hypothesis: (i) there is little structuring between European wild boar and Iberian pigs (Ramirez *et al.*, 2009; van Asch *et al.*, 2012), (ii) a greater Iberian contribution is ascribed by a supervised analysis to the most preserved or isolated populations

(Yucatan, Peru) than to other populations (Table 2), and (iii) the introduction of international breeds all over the world replacing local livestock is well known. As a result, village pig populations are far from being static genetic pools. In fact, the presence of outliers in some of these populations, rather than being simply 'noise' or errors in sampling, illustrates that village pigs are dynamic populations whose genetic structure can change quickly and deserve conservation. In fact, the ancestral Mexican population of Yucatan pigs is now almost extinct, so current Yucatan mini-pigs should actually more faithfully reflect the ancestral genetic variability of Mexican pigs than do modern *cuino* or *pelón* pigs. In all likelihood, international breeds will continue to be introgressed into American village pig populations, whereas the flow of Iberian pigs was interrupted long ago.

Historical records, mitochondrial DNA data (Souza *et al.*, 2009) and our data support that Brazilian pigs are mostly related to European local pigs, as are the rest of American village pigs. Nevertheless, Brazilian pigs clustered separately at the continent level (Figure 3). Although this result should be considered cautiously, given that the American principal components explain a small fraction of worldwide variance where the Asia—Europe axis is predominant (Figure 1), it seems to be a general trend that Brazilian pigs are closely related among each other (See also the dendrogramme in Supplementary File 3). Can this be explained by different histories from the early days of colonization or is it due to more recent events? Certainly, Portuguese and Castilians divided their area of influence in America from the very beginning due to the Treaty of Tordesillas, in 1493. Empirical support for this hypothesis is also provided by the fact that Bisaro pigs, a Portuguese breed, are genetically closer in terms of F_{ST} to Brazilian populations than are Iberian pigs. Yet, it is worth noting as well that Portugal was ruled by the Spanish Hapsburg dynasty during a large initial period of the colony (1581–1640), therefore increasing trade between and within Iberian kingdoms and their colonies in the Americas. There were also intermittent periods of Dutch rule in NE Brazil, for example, 1624–1654 in Pernambuco. F_{ST}'s also show that Bisaro pigs are nearer to many American populations than are Iberian pigs, which would suggest a predominant Portuguese 'pig colonization' America-wide. Similarly, Canary pigs are also close to American pigs. However, as Figure 4 suggests, there is evidence that both ancient Bisaro and Canary pigs have been intermixed with modern breeds. What is the cause, therefore, of a specific Brazilian signature? First, note that Moura is somewhat separate from the rest of Brazilian pigs, and they exhibit an increased Duroc component. Mariante and Cavalcante (2006) do report that local Brazilian pigs were crossed to Duroc-Jersey to make up Moura. As for the rest of Brazilian breeds, the explanation is not so clear. A Chinese contribution cannot be ruled out, at least in Nilo and in Monteiro. Further, classical studies (Vianna, 1956) mention that Portuguese imported pigs from their colony Macau in China. Interestingly, some pigs in Misiones, Argentina are still called Macau. The ADMIXTURE supervised analysis suggests a strong Landrace component in Piau with $K = 6$ (Table 2 and Figure 4c), whereas larger K suggests a cluster of its own and shared with other Brazilian populations (Figures 4a and b). The Piau breed originated in the states of Goias, São Paulo and Minas Gerais, likely a result of crosses between local and other breeds like Poland China or Duroc, among others (Mariante and Cavalcante, 2006). All in all, it can be hypothesized that the difference between Brazil and Spanish America that we see today is caused by distinct introgression patterns, rather than by distinct initial colonization processes.

A major task in order to understand adaptation at the molecular level is to characterize the genes that have responded to selection, either artificial selection or natural selection as a result of adapting to extreme environments. Our results bear special relevance regarding the adaptation to altitude. Our study identified ∼300 highly differentiated genes. Remarkably, about 54 has a role in blood circulation and four of them (*SMAD4, MDM2, VLDLR, KCNA5*) were *a priori* functional candidates in human studies (Simonson *et al.*, 2010). Among those, a few merit special attention. *FGF2* and *FGFR1* are involved in phenotypic modulation of vascular smooth-muscle cells (Chen *et al.*, 2009). *NFE2L2* has a role in the coordinated upregulation of genes in response to oxidative stress, whereas *GPR124* regulates angiogenesis in the central nervous system (Kuhnert *et al.*, 2010). Additional genes include *BEST3, PDE10A, PDE11A* and *IL21*. *BEST3* is expressed in smooth-muscle cells and is important for regulation affecting vasomotion. *PDE10A* and *PDE11A* are expressed in components of the trigeminovascular pain signaling system (Kruse *et al.*, 2009). *PDE10A* is also involved in progressive pulmonary vascular remodeling, increasing its expression in some pulmonary diseases (Tian *et al.*, 2011). Finally, interleukin 21 signaling has a critical role in promoting the lung inflammatory response to acute pneumovirus infection (Spolski *et al.*, 2012). Adaptation to altitude has received attention in humans (see Cheviron and Brumfield, 2012 for a review), and physiological differences caused by altitude have been studied in cattle (Wuletaw *et al.*, 2011). However, to our knowledge, this is the first report of indirect evidence of genetic adaptation to altitude in livestock. It should be noted that, given the relatively low density of markers and the large window used (1 Mb), the selective footprints described are probably among the most extreme ones and other indirect evidence of selective events are waiting to be identified with more data and with more refined tools.

CONCLUSION

To conclude with a paraphrase of Novembre *et al.* (2008): creole porcine genes in the Americas do not mirror geography. They look rather like a blur of history. Genetic evidence supports the belief that creole pig populations are relatively homogeneous within a short geographic radius, a shared ancestry likely due to the exchange of pigs between nearby communities. Aside from that, geographic distance explains just a tiny fraction of variation in coancestry. Across the Americas, the genomic patterns observed are not compatible with a classical stepping-stone colonization model, reminding us that livestock is highly mobile, especially in the case of pigs. Modern village pigs in the Americas are the result of many independent colonization and introgression events, including may be a direct Chinese introgression. Importantly, these data also confirm our initial hypothesis regarding adaptation: extreme climates have posed important challenges to pigs.

DATA ARCHIVING

Data have been deposited at Dryad: doi:10.5061/dryad.t1r3d.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

Akey JM, Ruhe AL, Akey DT, Wong AK, Connelly CF, Madeoy J et al. (2010). Tracking footprints of artificial selection in the dog genome. *Proc Natl Acad Sci USA* **107**: 1160–1165.

Akey JM, Zhang G, Zhang K, Jin L, Shriver MD (2002). Interrogating a high-density snp map for signatures of natural selection. *Genome Res* **12**: 1805–1814.

Alexander DH, Novembre J, Lange K (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **19**: 1655–1664.

Alves E, Fernández AI, Barragán C, Ovilo C, Rodríguez C, Silió L (2006). Inference of hidden population substructure of the Iberian pig breed using multilocus microsatellite data. *Span J Agric Res* **41**: 37–46.

Boyko AR, Boyko RH, Boyko CM, Parker HG, Castelhano M, Corey L et al. (2009). Complex population structure in African village dogs and its implications for inferring dog domestication history. *Proc Natl Acad Sci USA* **106**: 13903–13908.

Chen P-Y, Simons M, Friesel R (2009). FRS2 via fibroblast growth factor receptor 1 is required for platelet-derived growth factor receptor i2-mediated regulation of vascular smooth muscle marker gene expression. *J Biol Chem* **284**: 15980–15992.

Crossby A (2003). *The Columbian Exchange*. Praeger Publishers: Connecticut.

Delgado JV, Martinez AM, Acosta A, Alvarez LA, Armstrong E, Camacho E et al. (2011). Genetic characterization of Latin-American Creole cattle using microsatellite markers. *Anim Genet* **43**: 2–10.

Cheviron ZA, Brumfield RT (2012). Genomic insights into adaptation to high-altitude environments. *Heredity (Edinb)* **108**: 354–361.

Clop A, Amills M, Noguera JL, Fernandez A, Capote J, Ramon MM et al. (2004). Estimating the frequency of Asian cytochrome B haplotypes in standard European and local Spanish pig breeds. *Genet Sel Evol* **36**: 97–104.

Elliot JH (2007). *Empires of the Atlantic World: Britain and Spain in America 1492-1830*. Yale University Press.

García M, Capote J (1982). *El Cerdo Negro Canario*. Cabildo Insular de La Palma: Excmo.

Gautier M, Naves M (2011). Footprints of selection in the ancestral admixture of a New World Creole cattle breed. *Mol Ecol* **20**: 3128–3143.

García-Dory MA, Martínez-Vicente S, Orozco-Piñán F (1990). *Guía De Campo De Las Razas Autóctonas Españolas*. Alianza Editorial: Madrid.

Gautier M, Vitalis R (2012). Rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics* **28**: 1176–1177.

Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ et al. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res* **36**: 3420–3435.

Groenen M, Archibald A, Uenishi H, Tuggle C, Takeuchi Y, Rothschild M et al. (2012). Pig genomes provide insight into porcine demography and evolution. *Nature* **491**: 393–398.

Huang D, Sherman BT, Lempicki RA (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**: 44–57.

Huson DH, Richter DC, Rausch C, Dezulian T, Franz M, Rupp R (2007). Dendroscope: An interactive viewer for large phylogenetic trees. *BMC Bioinformatics* **8**: 460.

Jombart T (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**: 1403–1405.

Kruse LS, Muller M, Tibak M, Gammeltoft S, Olesen J, Kruuse C (2009). PDE9A, PDE10A, and PDE11A expression in rat trigeminovascular pain signalling system. *Brain Res* **1281**: 25–34.

Kuhnert F, Mancuso MR, Shamloo A, Wang H-T, Choksi V, Florek M et al. (2010). Essential regulation of cns angiogenesis by the orphan G protein-coupled receptor GPR124. *Science* **330**: 985–989.

Larson G, Dobney K, Albarella U, Fang M, Matisoo-Smith E, Robins J et al. (2005). Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science* **307**: 1618–1621.

Lemus C, Ly J (2010). Estudios de sostenibilidad de cerdos mexicanos pelones y cuinos. La iniciativa nayarita. *Revista Computadorizada de Producción Porcina* **17**: 89–98.

Mariante AS, Cavalcante N (2006). *Animals of the Discovery*. Embrapa: Brasilia.

McVean G (2009). A genealogical interpretation of principal components analysis. *PLoS Genet* **5**: e1000686.

Merino M, Carpinetti B (2003). Feral pig Sus scrofa population estimates in bahía Samboronbón conservation area, Buenso Aires Province, Argentina. *J Neotrop Mammal* **10**: 269–275.

Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A et al. (2008). Genes mirror geography within Europe. *Nature* **456**: 98–101.

Porter V (1993). *Pigs: A Handbook to the Breeds of the World*. Comstock Publishing.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**: 904–909.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**: 559–575.

Ramirez O, Ojeda A, Tomas A, Gallardo D, Huang LS, Folch JM et al. (2009). Integrating Y-chromosome, mitochondrial, and autosomal data to analyze the origin of pig breeds. *Mol Biol Evol* **26**: 2061–2072.

Ramos AM, Crooijmans RP, Affara NA, Amaral AJ, Archibald AL, Beever JE et al. (2009). Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. *PLoS One* **4**: e6524.

Scheet P, Stephens M (2006). A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am J Hum Genet* **78**: 629–644.

Ramos-Onsins SE, Mitchell-Olds T (2007). Mlcoalsim: multilocus coalescent simulations. *Evol Bioinform Online* **3**: 41–44.

Simonson TS, Yang Y, Huff CD, Yun H, Qin G, Witherspoon DJ et al. (2010). Genetic evidence for high-altitude adaptation in Tibet. *Science* **329**: 72–75.

Souza CA, Paiva SR, Pereira RW, Guimaraes SE, Dutra Jr WM, Murata LS et al. (2009). Iberian origin of Brazilian local pig breeds based on Cytochrome b (MT-CYB) sequence. *Anim Genet* **40**: 759–762.

Spolski R, Wang L, Wan C-K, Bonville CA, Domachowske JB, Kim H-P et al. (2012). IL-21 promotes the pathologic immune response to pneumovirus infection. *The Journal of Immunology* **188**: 1924–1932.

Tian X, Vroom C, Ghofrani HA, Weissmann N, Bieniek E, Grimminger F et al. (2011). Phosphodiesterase 10a upregulation contributes to pulmonary vascular remodeling. *PLoS ONE* **6**: e18136.

van Asch B, Pereira F, Santos LS, Carneiro J, Santos N, Amorim A (2012). Mitochondrial lineages reveal gene flow between Iberian wild boars and South Iberian pig breeds. *Anim Genet* **43**: 35–41.

Vaysse A, Ratnakumar A, Derrien T, Axelsson E, Rosengren Pielberg G, Sigurdsson S et al. (2012). Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping. *PLoS Genet* **7**: e1002316.

Vianna AT (1956). *Os Suínos: Criação prática e econômica*, 2nd edn. Serviço de Informação Agrícola, Ministério da Agricultura: Rio de Janeiro.

Voight BF, Kudaravalli S, Wen X, Pritchard JK (2006). A map of recent positive selection in the human genome. *PLoS Biol* **4**: e72.

Willing ME, Dreyer C, van Oosterhout C (2012). Estimates of genetic differentiation measured by F(ST) do not necessarily require large sample sizes when using many SNP markers. *PLoS One* **7**: e42649.

Wuletaw Z, Wurzinger M, Hol tT, Dessie T, Sölkner J (2011). Assessment of physiological adaptation of indigenous and crossbred cattle to hypoxic environment in Ethiopia. *Livest Sci* **138**: 96–104.

Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZX, Pool JE et al. (2010). Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* **329**: 75–78.

Supplementary Information accompanies this paper on Heredity website (http://www.nature.com/hdy)