Journal of
Animal Breeding and Genetics

J. Anim. Breed. Genet. ISSN 0931-2668

ORIGINAL ARTICLE

Genetic relationships among American donkey populations: insights into the process of colonization

J. Jordana[1,*], A. Ferrando[1,*], J. Miró[2], F. Goyache[3], A. Loarca[4], O.R. Martínez López[5], J.L. Canelón[6], A. Stemmer[7], L. Aguirre[8], M.A.C. Lara[9], L.A. Álvarez[10], S. Llambí[11], N. Gómez[12], L.T. Gama[13], M.F. Nóvoa[14], R.D. Martínez[15], E. Pérez[16], A. Sierra[17], M.A. Contreras[18], A.M. Guastella[19], D. Marletta[19], G. Arsenos[20], I. Curik[21], V. Landi[22], A. Martínez[22] & J.V. Delgado[22]

1 Departament de Ciència Animal i dels Aliments, Facultat de Veterinària, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain
2 Departament de Medicina i Cirurgia Animals, Facultat de Veterinària, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain
3 Área de Genética y Reproducción Animal, SERIDA-Deva, Gijón, Spain
4 Ministerio de Agricultura, Ganadería y Alimentación, Quetzaltenango, Guatemala
5 Dirección General de Investigación Científica y Tecnológica, Centro Multidisciplinario de Investigaciones Tecnológicas, Universidad Nacional de Asunción, Asunción, Paraguay
6 Departamento de Producción e Industria Animal, Decanato de Ciencias Veterinarias, Universidad Centroccidental Lisandro Alvarado, Barquisimeto, Venezuela
7 Facultad de Ciencias Agrícolas y Pecuarias, Universidad Mayor de San Simón, Cochabamba, Bolivia
8 Centro Biotecnología Reproductiva Animal, Universidad Nacional de Loja, Loja, Ecuador
9 Instituto de Zootecnia, Nova Odessa, SP, Brazil
10 Sede Palmira, Universidad Nacional de Colombia, Palmira, Valle del Cauca, Colombia
11 Instituto de Producción Animal, Área Genética, Facultad de Veterinaria, Universidad de la República (UdelaR), Montevideo, Uruguay
12 Facultad de Medicina Veterinaria y Zootecnia, Universidad Nacional Micaela Bastidas de Apurímac, Abancay, Peru
13 Faculdade de Medicina Veterinária, Universidade Técnica de Lisboa, Lisboa, Portugal
14 AEPGA Associação para o Estudo e Protecção do Gado Asinino, Atenor, Portugal
15 Mejora y Conservación de Recursos Genéticos, Facultad de Ciencias Agrarias, Universidad Nacional de Lomas de Zamora, Llavallol, Argentina
16 Departamento de Salud y Explotación Animal, Universidad de Granma, Bayamo, Cuba
17 Posgrado e Investigación, Instituto Tecnológico de Conkal, Yucatán, Mexico
18 Escuela de Medicina Veterinaria, UST Sede Santiago, Santiago, Chile
19 Sezione di Scienze delle Produzioni Animali, Dipartimento di Agricoltura, Alimentazione e Ambiente, Università degli studi di Catania, Catania, Italy
20 School of Veterinary Medicine, Aristotle University of Thessaloniki, Thessaloniki, Greece
21 Department of Animal Science, Faculty of Agriculture, University of Zagreb, Zagreb, Croatia
22 Departamento de Genética, Facultad de Veterinaria, Universidad de Córdoba, Córdoba, Spain

Keywords

Bayesian inference; Creole livestock; endangered breeds; genetic resources; microsatellites.

Correspondence

J. Jordana, Departament de Ciència Animal i dels Aliments, Facultat de Veterinària, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain.
Tel: +34 935811217;
Fax: +34 935812106;
E-mail: jordi.jordana@uab.cat

Received: 13 February 2015;
accepted: 6 July 2015

Summary

This study presents the first insights into the genetic diversity and structure of the American donkey metapopulation. The primary objectives were to detect the main structural features underlying variability among American donkey populations, identify boundaries between differentiated gene pools, and draw the main colonization pathways since the introduction of donkeys into America in the 15th century. A panel of 14 microsatellite markers was applied for genotyping 350 American donkeys from 13 countries. The genetic structure of this metapopulation was analysed using descriptive statistics and Bayesian model-based methods. These populations were then compared to a database containing information on 476 individuals from 11 European breeds to identify the most likely ancestral donor populations. Results showed the presence of two distinct genetic pools, with confluence of the two in Colombia. The southern pool showed a unique genetic signature subsequent to an older founder event, but lacked any significant influence of modern gene flow from Europe.

  doi:10.1111/jbg.12180

*Both authors contributed equally to this work

The northern pool, conversely, may have retained more ancestral polymorphisms and/or have experienced modern gene flow from Spanish breeds. The Andalusian and, to a lesser extent, the Catalan breeds have left a more pronounced footprint in some of the American donkey populations analysed.

Introduction

Domestic donkeys (*Equus africanus asinus*) descend from African wild ass (*Equus africanus*) populations. The domestication process started about 5000 years ago and had two major separate events: one arose in north-east Africa from Nubian wild ass ancestors (*Equus africanus africanus*) and a second event involved a genetically distinct wild ass population not yet identified and probably extinct at the present time (Beja-Pereira *et al.* 2004; Kimura *et al.* 2011). Donkeys were brought to Europe soon after domestication and were already widely distributed throughout the continent in the classical antiquity (Yanes 2005).

The domestic donkey reached the American continent for the first time at the very end of the 15th century, during the period of Spanish colonization (Laguna 1991; Rodero *et al.* 1992), together with other livestock species (Brookshier 1974; Laguna 1991; Delgado *et al.* 2010). These donkeys mostly originated from southern Spain, but they were also loaded onto ships sailing to America during layover stops in the Canary Islands, where donkeys had been introduced from Northern Africa in the middle of the 15th century (Yanes 2005). The first foundational nucleus was created in the Hispaniola Island (today Dominican Republic and Haiti) with the aim of reproducing and adapting imported domestic species to the new territory prior to extending them to other Antilles islands and to the mainland. Two decades after the first arrivals, the donkey census increased successfully, and even some feral populations developed (Laguna 1991; Yanes 2005).

Shipping routes connected the Caribbean islands with the mainland through two principal routes. The first one reached Mexico and connected with different routes to Florida, New Mexico, California and Texas. The second route arrived at Panama harbours to connect with the routes to Central and South America, easterly through Colombia, Venezuela and northern Brazil, or southerly through Ecuador and Peru (Sponenberg 1992; Delgado *et al.* 2010; see Figure S1). Spanish livestock was also introduced into America from the Caribbean islands, through the coasts of Uruguay and Argentina, and entering inland by going upstream of the River Plate and its tributaries. The southern cone of Argentina, Chile, Uruguay, Paraguay and southern Brazil were under the influence of this route (Delgado *et al.* 2010). Livestock were also brought to America by the Portuguese route, arriving either through the north-east coast or the south-east coast of Brazil. From here, livestock were spread inland towards southern Brazil and Paraguay (Primo 2004).

In the 16th and 17th centuries, Peru became an important centre of mule production, and started its nucleus with donkeys originated from a feral population imported from Jamaica. There was an active trade between Peru and Argentinean Pampas, with commercial routes connecting the Pampas region to important economical communities, north-west to the Peruvian plateau and north-easterly to Bahia lands (Laguna 1991; Santos *et al.* 1992). Another important nucleus of mule production emerged in Mexico (Laguna 1991; Yanes 2005).

In the 18th and 19th centuries, several European donkey breeds were imported to North America, especially the Catalan, much appreciated due to its large size. The Catalan breed was extensively imported throughout the 20th century to Canada, United States, Mexico, Argentina and Brazil (Romagosa 1959), as the production of mules remained an important activity until the first half of the 20th century (Laguna 1991; Yanes 2005).

American donkeys, like most of the other American livestock species (Delgado *et al.* 2010), have not yet been characterized and typified as distinct breeds. Nevertheless, the genetic characterization of American donkeys is essential to investigate the impact of founder events, isolation, genetic drift and episodic waves of unidirectional gene flow from the Old World on the present genetic composition of Creole populations. These results, together with morphological and phenotyping studies aimed at their characterization, are the first step to identify these populations as distinct breeds.

This study has been carried out through the joint efforts of various research institutions associated with the CYTED-XIIH and CONBIAND networks, within the framework of a collaborative project. The main objectives were to: (i) investigate the genetic diversity and structure of American donkey populations;

(ii) detect the existence of geographical patterns of genetic dispersion; and (iii) compare those populations with European donkey breeds that are the current representatives of the most likely source populations from which American donkeys have originated during the colonization process, for an understanding of the shaping process of their genetic pool.

Material and methods

Sample collection

A total 826 individual donkeys were included in this study: 350 American donkeys and 476 European donkeys. The American samples were taken from a wide area (latitude: 38° S–23° N; longitude: 38° W–96° W) and included individuals from 13 countries (Figure S2a). American discrete populations were *a priori* defined according to the country the individuals belonged to. Individuals were sampled from several farms per country, except those from Chile, which originated from a single feral population. All Brazilian individuals sampled were from farms located in the eastern side of the country (Ceará). Genealogical information was lacking, and the degree of familial relationships was unknown for almost all individuals. All American samples consisted of hairs.

European samples belonged to 11 breeds or discrete populations from four Mediterranean countries (Figure S2b): Italy (three breeds), Greece (undetermined breed), Portugal (one breed) and Spain (six breeds). Genotypes for Italian and Spanish breeds were obtained from previous studies (Ferrando *et al.* 2008; Bordonaro *et al.* 2012). Portuguese and Greek samples were genotyped for the present study from hair samples.

DNA extraction and genotype scoring

Genomic DNA was isolated from hair with the DNeasy Blood & Tissue Kit (Qiagen Iberia SL, Barcelona, Spain), following the protocol for animal tissue. Each sample included around 10–15 hair root ends.

A panel of 14 markers was used: *AHT04*, *AHT05*, *ASB23*, *HMS02*, *HMS03*, *HMS05*, *HMS06*, *HMS07*, *HTG04*, *HTG06*, *HTG07*, *HTG10*, *HTG15* and *VHL20* (see Table S1). The 5′ side of each forward primer was fluorescently labelled. All microsatellite loci were amplified as described elsewhere (Aranguren-Méndez *et al.* 2001), with minor modifications: DNA was amplified by means of four multiplex PCR, as detailed in Table S1. Bovine serum albumin was added to each PCR in a final concentration of 0.8 μg/μl, to minimize the effects of inhibitors co-extracted with the DNA. Approximately 10–60 ng of DNA was used per PCR. Diluted PCR products were run on an automated sequencer ABI PRISM 3730 (Life Technologies, Carlsbad, CA, USA) with a size standard. Analysis for sizing was performed with GENEMAPPER v3.7 (Life Technologies).

All but Italian breeds were analysed and coded in the same laboratory. The allele correspondence between databases was made by means of reference samples. Due to technical reasons, the marker *HTG07* was not used for intercontinental comparisons (i.e. 13 markers were used).

Genetic diversity analyses

Allelic count, observed (H_O) and unbiased expected (H_E) heterozygosities and their standard deviations were obtained with GENETIX v.4.05.2 (Belkhir *et al.* 1996–2004). Allelic richness (AR) and *F* fixation indexes were calculated with FSTAT v.2.9.3.2 (Goudet 1995, 2001). Standard errors (SE) of *F*-statistics were obtained by jackknifing over loci. Linkage disequilibrium (LD) between pairs of loci among American populations was also computed, and the critical level of significance of LD tests for a 5% nominal level was adjusted after applying the Bonferroni correction for multiple independent tests (23 660 permutations).

Genetic differentiation

The F_{ST} matrix of genetic distances among American populations was obtained with FSTAT. The p-value of the estimated F_{ST} distances was obtained after 78 000 permutations, and the critical level of significance was adjusted for multiple independent tests. A principal coordinates analysis (PCoA) via covariance matrix with data standardization was performed on the F_{ST} matrix with GENALEX 6.501 (Peakall & Smouse 2006, 2012). The genetic distances D_A (Nei *et al.* 1983) and weighted estimator of D_R (Reynolds *et al.* 1983) matrices among American populations were obtained with POPULATIONS v.1.2.30 (Langella 2002). This programme was also used to obtain the genetic distances tree among American populations and European breeds, using the weighted estimator of D_R and the neighbour-joining algorithm with 5000 bootstraps among loci. The unrooted distance tree was then visualized with TREEVIEW v. 1.6.6 (Page 1996).

Bayesian inference

The structure of the American metapopulation was investigated by means of a Bayesian model-based

clustering method using STRUCTURE v.2.3.4 (Pritchard *et al.* 2000; Falush *et al.* 2003; Hubisz *et al.* 2009). The model-based algorithms implemented in this software are used for inferring the most likely number of K differentiated genetic clusters at Hardy–Weinberg equilibrium (HWE), underlying the genetic variability found in a group of populations. They compute the proportional membership of a genome in each inferred cluster, at population (Q) and individual (q) levels. We assumed that individuals could have arisen from more than one ancestral genetic pool (admixture model) and that allele frequencies were correlated. The programme was run from $K = 1$ to $K = 13$, with 20 independent runs per K. Each run included a burn-in period of 800 000 Markov Chain Monte Carlo (MCMC) steps, followed by 1 000 000 additional iteration steps. The most likely value of K was determined from the ΔK parameter, following the method described by Evanno *et al.* (2005) implemented in STRUCTURE HARVESTER (Earl & vonHoldt 2012). However, we also considered other values of K where the mean likelihood $L(K)$ reached a plateau. We defined a mean value of $Q > 0.80$, averaged among runs, as the threshold to assign predefined populations to a unique cluster. Otherwise, a mixed ancestry was assumed. The analysis was repeated within main clusters detected in the previous analysis. This time, two different models were used in the analyses: the first one was the same as described previously, while the second included the sample location as prior information (i.e. all individuals from the same country shared the same location). This model allows improving the detection of population structure when the data set has a low information content and the location is informative, but it also ignores this prior when there is no correlation between sampling location and population structure (Hubisz *et al.* 2009). Ten independent runs per K, from $K = 1$ to $K = N + 1$ (N, number of populations analysed) were performed with both models. Finally, we carried out 50 independent runs with the global American and European data set under the same parameters as for the American data set, assuming a number of ancestral populations ranging from $K = 2$ to $K = 27$, without sample location as prior, to explore the genetic relationship among American populations and present-day European breeds.

The software CLUMPP (Jakobsson & Rosenberg 2007) was applied for permutation and matching replicated runs of STRUCTURE and obtaining averaged Q and q values among all runs. Plots were visualized and modified with the GENESIS v.0.2.5 software (Buchmann & Hazelhurst 2014).

Results

Genetic diversity of American populations

A total of 98 alleles were observed for the 14 markers analysed across all American individuals (Tables 1 and S1). The number of alleles per locus ranged from 4 (*HMS06*, *HTG06*) to 13 (*AHT05*), and H_E from 0.139 (*HTG04*) to 0.845 (*HTG10*). The overall genetic diversity ($H_E = 0.569 \pm 0.245$) of the American metapopulation was in the range of most European breeds (see Table S2). No significant LD was observed between any pair of loci within populations after applying the Bonferroni correction for multiple independent tests. However, as the classical Bonferroni correction may be too restrictive, p-values were evaluated without applying the correction. Still, for each analysed pair of loci, none of the significant p-values ($p < 0.05$) was consistent for more than three populations at a time.

At the population level, H_E was above the overall mean for the populations sampled in Brazil, Guatemala and Mexico, and close to the mean in those from Cuba and Colombia. Allelic richness was higher in these countries, as well. Brazil also showed the highest number of private alleles (PA = 5), while all other populations except Cuba (PA = 2) had one or none. The lowest genetic diversity was detected in donkeys from Uruguay.

American metapopulation structure

A significant departure from the HWE was detected when all American samples were considered, due to a deficit of heterozygotes ($F = 0.101$). The hierarchical analysis of F-statistics showed that this deficit partly originated from genetic differentiation among populations from different countries ($\theta = 0.061$, SE: 0.006), but was also caused by a within-country deficit ($f = 0.048$, SE: 0.008). When analysed separately, most countries showed none or slight departures from HWE, but this departure was highly significant only for populations from Uruguay and Peru (Table 1).

The F_{ST} pairwise distances varied from 0.0011 to 0.1269 for the various pairs, and were significant ($p < 0.01$) for almost all pairs, except Mexico-Guatemala, and between Peru and populations from most south-western American countries (Argentina, Chile, Paraguay and Bolivia; see Table 2). However, the latter four populations were significantly differentiated from each other. Uruguay and Venezuela were the most distant from each other. Similar results were observed with the D_R and D_A distance matrices (Tables 2 and S3).

Table 1 Main diversity parameters from each American population included in this study for a panel of 14 microsatellite markers: observed heterozygosity (H_O), unbiased expected heterozygosity (H_E) and their standard deviations, within-country excess or deficit of heterozygotes (F_{IS}), mean number of alleles (MNA), rarefaction of allelic richness to 12 individuals (AR), number of private alleles (PA), and genome distribution to cluster A and B for $K = 2$ inferred clusters (values of $Q > 0.80$ are indicated in bold)

Country	Code	Number of samples	H_O	H_E	F_{IS}	MNA	AR	PA	Cluster A	Cluster B
Argentina	ARG	25	0.493 ± 0.220	0.523 ± 0.239	0.059*	4.86	4.13	0	**0.806**	0.194
Bolivia	BOL	30	0.509 ± 0.230	0.549 ± 0.240	0.074*	4.07	3.78	0	**0.849**	0.151
Brazil	BRA	25	0.601 ± 0.217	0.592 ± 0.190	−0.017	5.14	4.35	5	0.288	0.712
Chile	CHI	20	0.501 ± 0.304	0.501 ± 0.281	−0.002	3.79	3.59	0	**0.955**	0.045
Colombia	COL	30	0.524 ± 0.214	0.563 ± 0.237	0.071*	5.00	4.27	1	0.340	0.660
Cuba	CUB	70	0.563 ± 0.236	0.566 ± 0.237	0.004	5.71	4.40	2	0.051	**0.949**
Ecuador	ECU	21	0.507 ± 0.280	0.511 ± 0.260	0.009	3.93	3.63	1	**0.903**	0.097
Guatemala	GUA	15	0.532 ± 0.227	0.573 ± 0.207	0.074*	4.64	4.46	1	0.121	**0.879**
Mexico	MEX	14	0.597 ± 0.268	0.597 ± 0.193	0.000	4.43	4.32	0	0.103	**0.897**
Paraguay	PAR	29	0.467 ± 0.270	0.497 ± 0.261	0.061*	3.86	3.52	0	**0.868**	0.132
Peru	PER	20	0.482 ± 0.247	0.539 ± 0.263	0.109**	4.21	3.94	1	**0.887**	0.113
Uruguay	URU	24	0.365 ± 0.234	0.452 ± 0.277	0.196***	3.57	3.27	1	**0.934**	0.066
Venezuela	VEN	27	0.466 ± 0.270	0.505 ± 0.273	0.079*	4.00	3.50	0	0.059	**0.941**
Overall		350	0.512 ± 0.225	0.569 ± 0.245	0.101***	7.00	–	–	–	–

*p < 0.05, **p < 0.01; ***p < 0.001.

Table 2 F_{ST} values among American populations and significance (above diagonal) and weighted D_R distances (below diagonal). All p-values of the F_{ST} matrix were obtained after 78 000 permutations. The indicative adjusted nominal level (5%) for multiple comparisons was 0.000641

	MEX	GUA	CUB	VEN	COL	BRA	ECU	PER	BOL	PAR	CHI	ARG	URU
MEX	–	0.0055 ns	0.0478**	0.0831**	0.0717**	0.0375**	0.0889**	0.0478**	0.0515**	0.0822**	0.0963**	0.0815**	0.1137**
GUA	0.0077	–	0.0393**	0.0778**	0.0639**	0.0443**	0.0914**	0.0441**	0.0481**	0.0758**	0.1110**	0.0771**	0.1154**
CUB	0.0493	0.0408	–	0.0321**	0.0327**	0.0515**	0.0949**	0.0461**	0.0610**	0.0563**	0.0946**	0.0728**	0.1035**
VEN	0.0885	0.0833	0.0329	–	0.0502**	0.0830**	0.1179**	0.0704**	0.0714**	0.0825**	0.1269**	0.0887**	0.1269**
COL	0.0761	0.0685	0.0336	0.0529	–	0.0393**	0.1040**	0.0308**	0.0384**	0.0582**	0.0670**	0.0487**	0.0936**
BRA	0.0383	0.0463	0.0529	0.0873	0.0408	–	0.0727**	0.0248**	0.0409**	0.0513**	0.0750**	0.0567**	0.0812**
ECU	0.0939	0.0977	0.0999	0.1266	0.1110	0.0755	–	0.0326**	0.0502**	0.0463**	0.0752**	0.0540**	0.0862**
PER	0.0513	0.0482	0.0477	0.0751	0.0332	0.0260	0.0348	–	0.0011 ns	0.0079 ns	0.0151 ns	0.0120 ns	0.0391**
BOL	0.0546	0.0514	0.0632	0.0755	0.0404	0.0424	0.0526	0.0031	–	0.0318**	0.0255**	0.0188**	0.0593**
PAR	0.0871	0.0808	0.0582	0.0874	0.0612	0.0532	0.0484	0.0098	0.0336	–	0.0578**	0.0295**	0.0386**
CHI	0.1017	0.1191	0.0997	0.1371	0.0707	0.0779	0.0787	0.0171	0.0272	0.0605	–	0.0399**	0.0723**
ARG	0.0863	0.0824	0.0760	0.0944	0.0513	0.0589	0.0565	0.0141	0.0204	0.0312	0.0419	–	0.0298**
URU	0.1242	0.1267	0.1099	0.1384	0.1006	0.0862	0.0925	0.0436	0.0636	0.0418	0.0778	0.0330	–

ns, not significant.
**p < 0.01 after applying Bonferroni correction for multiple independent tests.

The PCoA of the F_{ST} matrix summarizes graphically these results (Figure 1). The first axis (accounting for 36.51% of variation) separated two groups. The first one encompassed donkeys from Argentina, Bolivia, Chile, Ecuador, Paraguay, Peru and Uruguay. They followed a scattered cline-pattern among the first axis. The second axis (18.69%) tended to separate moderately Ecuador from the other countries. In this group, populations from Uruguay and Ecuador were the more distant from each other ($F_{ST} = 0.0862$). On the other hand, Mexico, Guatemala, Cuba and Venezuela fell very close to each other in the first axis, whereas the second axis created a clear split in this second group. Brazil and Colombia fell in an intermediate position between both groups.

Bayesian clustering

The most likely value of K for American populations detected with STRUCTURE after applying the Evanno method was $K = 2$ (Figure S3). The existence of two major clusters was consistent with the PCoA, such that the first inferred one (cluster A) gathered animals from Argentina, Bolivia, Chile, Ecuador, Paraguay,

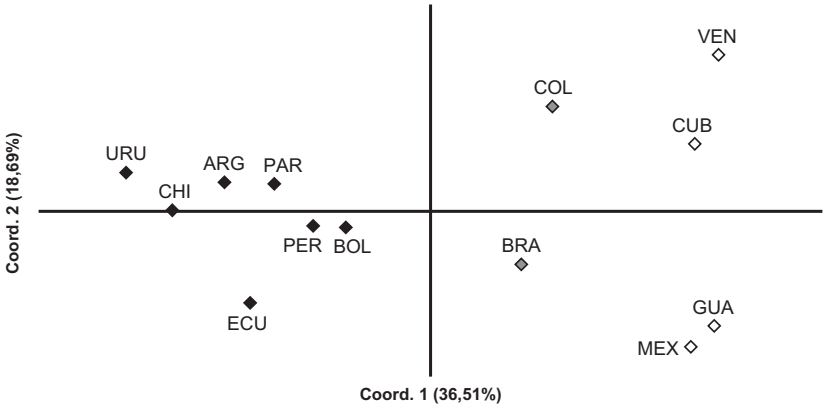

Figure 1 Principal coordinates analysis of American populations constructed with the F_{ST} matrix and applying the covariance with standardization option. The colour of the diamonds corresponds to the principal clusters described in the main text: (i) (black), (ii) (white) and admixed (grey).

Peru and Uruguay. The second one (cluster B) included animals from Cuba, Guatemala, Mexico and Venezuela. Donkeys sampled in Brazil and Colombia had mixed contributions from both clusters ($Q < 0.8$; Table 1, Figure 2). The highest mean Q values for clusters A and B were found in Chile ($Q = 0.955$) and Cuba ($Q = 0.949$), respectively. At the individual level, most donkeys from Colombia tended to group either into one or the other cluster, indicating substructure of the population, whereas Brazilian individuals were more admixed, but overall closer to cluster B ($Q = 0.712$). Nonetheless, $L(K)$ increased until $K = 4$ before reaching a plateau, and secondary structures appeared within both clusters (Figure S3).

Considering that $K = 2$ is probably the uppermost hierarchical division of a more complex structure, we explored the results within each cluster without and with the sample location as a prior. We detected differences between both models, such that cluster A (Figure S4) showed no clear internal structure with the first model, as slight variations of $L(K)$ were observed from $K = 1$ to $K = 5$. However, most individuals from Uruguay consistently clustered separately from all other individuals, whereas the other populations were fairly admixed among all tested K. Conversely, the second model resolved much better. At $K = 2$, most Uruguayan donkeys fell apart all other populations but Argentina, which was admixed. At $K = 3$, Ecuador fell in its own subcluster, whereas the other populations were mostly assigned to a third subcluster, except Argentina whose admixture level increased together with the number of K. Higher values of K suggested a division within the third subcluster, but with moderate Q values. Regarding cluster B (Figure S5), the first model defined two distinct groups that separated Venezuela (subcluster B1) from Mexico and Guatemala (subcluster B2), while the Cuban population was split, with most individuals from the south-eastern area of the island belonging to B1, while most of those from north-western areas joined B2. The second model showed similar results at $K = 2$. However, $L(K)$ still increased until $K = 4$ and ΔK was ambiguous. The existence of three groups made sense, as Cuba and Venezuela were successfully assigned to distinct groups while Mexico and Guatemala formed a single group. At $K = 4$, either there was a split within Mexico (with a small number of samples originating from the same area), or Cuba was split, similarly as observed for $K = 2$. The plot shown in Figure S5 for $K = 4$ is a mean of both clustering solutions, but both divisions are visible in the plot drawn for $K = 5$.

Intercontinental comparisons

The diversity parameters of American and European populations for a common panel of 13 markers are

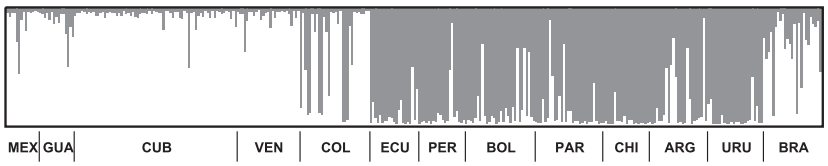

Figure 2 Genome distribution of individuals to each inferred cluster ($K = 2$) by the Bayesian model-based clustering method. Each individual is represented by a vertical single line. The length of coloured lines, vertical axis, is proportional to Q values to inferred clusters. The code of populations is the same as in Table 1.

summarized in Table S2. In Europe, genetic diversity was higher for mainland populations (H_E: 0.555–0.611; AR: 4.07–4.71), and lower for the islander breeds Pantesco and Majorero (H_E: 0.448–0.520: AR: 3.09–3.42). American populations from cluster B showed H_E values close to European breeds (H_E: 0.548–0.562; AR: 4.10–4.31), while cluster A was clearly below (H_E: 0.431–0.533; AR: 3.16–3.99).

The neighbour-joining tree constructed with the D_R distance generally showed bootstrap values near or below 50%, but the bootstraps increased to 88% for the clade including samples from Cuba and Venezuela, and 90% for the clade from Mexico and Guatemala (Figure 3). Populations from cluster A grouped together, and clearly separated from Iberian breeds. Andalusian donkeys were placed close to those from Mexico and Guatemala, while Catalan and Brazilian donkeys fell close to each other.

The analysis with STRUCTURE of the combined American and European data set (Figure S6) suggested the presence of five main clusters following the Evanno method. We did not observe a first split between American and European groups. At $K = 5$, American populations from cluster A formed a consistent group separated from all other populations (Q: 0.61–0.81). Animals from Mexico and Guatemala ($Q = 0.60$), and to a lesser extend Brazil ($Q = 0.49$), shared moderate ancestry to a common cluster together with Andalusian and Catalan donkeys ($Q = 0.79$ and 0.52, respectively). On the other hand, the subcluster B1 did not group consistently with any European breed.

Discussion

Main pathways of the colonization process

American donkey populations are broadly divided into two main clusters, which suggests that the colonization process and expansion of donkeys across America followed at least two main pathways. Cluster A, which includes south-western countries, is presumably the result of an ancient founder effect that took place at the early stages of colonization in the 16th and 17th centuries. Historical records document the creation of a reproductive nucleus of donkeys in the Peruvian plateau and surrounding territories from feral Jamaican donkeys (Laguna 1991; Yanes 2005), and the existence of ancient commercial routes may have contributed in maintaining a gene flow for centuries within this cluster in a

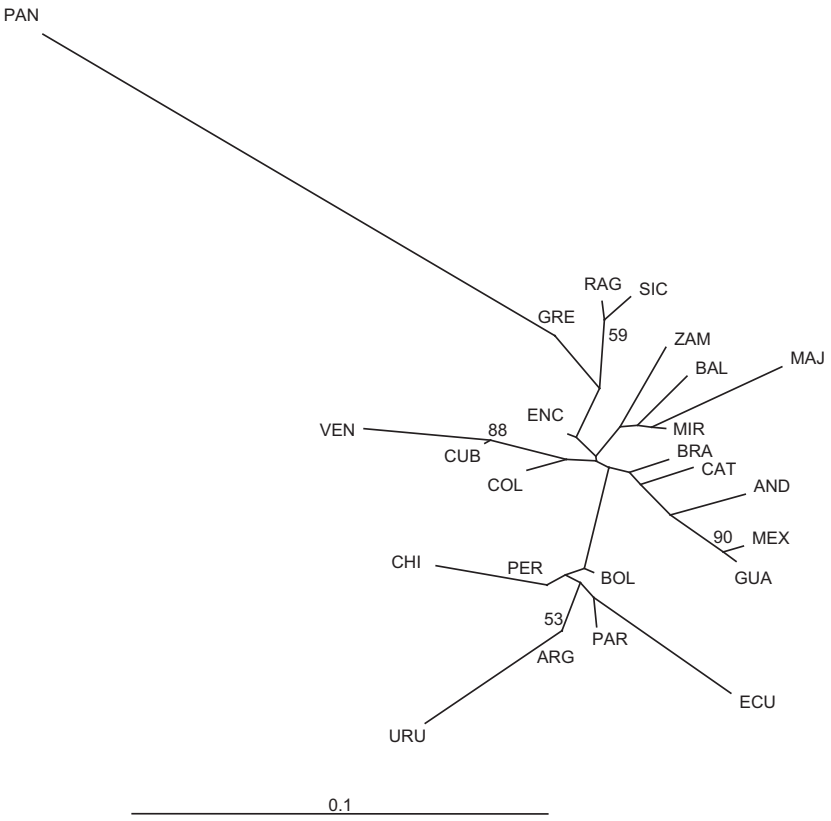

Figure 3 Unrooted tree of American and European populations obtained with the weighted D_R distance of Reynolds *et al.* (1983) and the neighbour-joining algorithm, after 5000 bootstraps over loci. Bootstrap values over 50% are indicated. The genetic distance scale is indicated below the tree. Names of American and European populations or breeds are represented by the first three letters.

north–south direction. This is consistent with the scattered cline-pattern suggested by the PCoA, the lower genetic diversity found in Uruguay and the higher genetic distance between donkeys from this country and from Ecuador. The strong deviation from HWE detected in Uruguay may be the result of an artefact caused by genetic fragmentation among sampled herds or by consanguinity among individuals from the same farm. However, we cannot discard the presence of cryptic substructures in southern parts of South America as a result of ancestral gene flow from a southern route upstream of the River Plate (Delgado *et al.* 2010) or by the arrival of animals from southern Brazil following the Portuguese route (Primo 2004). The central position of Peru in the phylogenetic tree and the genetic diversity levels that it has relative to other countries in this cluster is consistent with its strategic position on the north–south migration route and the important role it played as a centre of mule production during the 16th and 17th centuries. These results suggest that populations included in cluster A seem to have long-evolved independently from northern reproductive nuclei, and have experienced a genetic drift process subsequent to a marked founder event that has created a singular genetic signature.

Cluster B gathers populations closer to the Caribbean area of influence, and animals from Cuba and Venezuela had the highest assignment values to this cluster ($Q > 0.940$). The genetic similarity between Venezuelan and southern Cuba donkeys suggests that both countries were originally populated with individuals from the same reproductive nucleus. The development of the Venezuelan donkey population was most likely the result of a unique founder event with no relevant ensuing gene flow. As regards Cuba, the AR and private alleles detected suggest that high levels of ancestral polymorphisms have been retained compared to other areas. The genetic division among north-western and south-eastern individuals of the island can be the result of an isolation-by-distance pattern throughout the island. However, we cannot discard that Cuba itself was under the influence of two migratory waves from distinct islander sources that possibly underwent a genetic drift process since the establishment of the first donkey nucleus, prior to reaching Cuba and the mainland.

Mexican and Guatemalan populations could have descended from individuals imported from the north-western area of Cuba, which are genetically closer, but other sources such as Jamaica or former Hispaniola cannot be discarded. However, we have not found any details of animal movements during ancient maritime routes between Caribbean islands and the mainland to support this hypothesis.

Colombia appeared to be the contact zone between both clusters, as individuals were fully assigned either to cluster A or B, or were admixed. The role of Brazil is less evident, because even though it is more related to cluster B, it seems to be genetically influenced by both. The high number of private alleles detected in Brazil compared to other American populations also suggests that this country has received a more diverse genetic influence, either in the past through the Portuguese route, or as a result of modern gene flow.

The lasting European genetic signature

The first groups of donkeys introduced into the New World originated from Andalusia (southern Spain) and were of two types: either small-sized (similar to North African populations), or large frame animals (ancestors of present Andalusian breed; see Laguna 1991). Thus, it could be anticipated that a close genetic relationship between American and Andalusian donkeys would be detectable. However, the Spanish genetic signature is no longer recognizable in populations from cluster A, and there are two plausible hypotheses for this result. First, the ancestral genetic pool of this group could have been the African-type small-sized donkeys instead of the large-sized donkeys which now prevail in southern Spain. The second and more plausible hypothesis points to a rapid genetic drift resulting from the joint influence of a founder effect, isolation and selection pressure to a particular phenotype. The influence of a recent gene flow from European breeds, such as Catalan (Romagosa 1959) among Argentinean donkeys was not observed in the D_R tree. However, animals from this country showed the lowest Q value to their own cluster ($Q = 0.806$). On the other hand, the feral Chilean population may have kept a more genuine ancestral pool ($Q = 0.955$) as a result of its isolation from commercial trades due to its feral status.

Conversely to cluster A, the D_R distance tree and the Bayesian model-based clustering analysis suggest that Andalusian donkeys have more likely influenced populations from Mexico and Guatemala. Indeed, a previous study on the mitochondrial genetic diversity of Mexican donkeys showed that the most spread haplotypes among several Spanish breeds were also found in Mexican donkeys and, more particularly, the Andalusian breed shared a private haplotype with Mexican donkeys (López *et al.* 2005). This supports the stronger Andalusian ancestral influence on these populations. Successive migration waves of

importation of Spanish breeds, particularly Catalan and Andalusian, even centuries after the colonization (Laguna 1991; Yanes 2005), may have also contributed to maintain levels of genetic diversity and closer relationships among Brazilian, Mexican and Guatemalan populations and Spanish breeds. The importations of Catalan breed to countries, such as Mexico and Brazil during the 20th century (Romagosa 1959), may explain the central position of the Brazilian population in the D_R phylogenetic tree and its closer relationship to this breed. However, as the Catalan breed did not form a well-defined cluster among European breeds, its influence on other breeds is not easy to track.

The genetic source of Cuban and Venezuelan populations could not be identified among Spanish breeds. This may be due to the fact that both donkeys from Cuba and Venezuela and Spanish breeds diverged from the common ancestral donor population by genetic drift for centuries and have evolved separately with no modern gene flow since their establishment. Alternatively, other ancestral genetic pools should be considered, such as the African-like small-shaped donkeys imported from the Iberian Peninsula or the donkeys from Canary Islands, which were stopovers of caravels during the travel from Spain to the New World. Nonetheless, if Canarian donkeys did contribute to the genetic pool of American donkeys, this is not noticeable nowadays.

To conclude, these results bring an insight into the genetic diversity and structure of the American donkey metapopulation. American donkeys from the cluster A have shaped a genuine genetic signature, clearly differentiated from European breeds. This group has evolved separately from northern countries with a clear contact zone located in Colombia. Differentiated genetic pools are also identified in Cuba and Venezuela, while other countries appeared to have maintained a closer relationship to Spanish breeds. The addition of more American populations to fill the gaps between sampled areas and a more extensive panel of markers should help in depicting a more accurate genetic map of these Creole populations.

Acknowledgements

This study was partially funded by the Ministry of Agriculture, Livestock, Fisheries, Food and the Natural Environment of the Government of Catalonia. We also thank the farmers and the various breeders' associations for their help and collaboration throughout the sampling period.

References

Aranguren-Méndez J., Jordana J., Gómez M. (2001) Genetic diversity donkey breeds using microsatellite DNA markers. *Genet. Sel. Evol.*, **33**, 433–442.

Beja-Pereira A., England P.R., Ferrand N., Jordan S., Bakhiet A.O., Abdalla M.A., Mashkour M., Jordana J., Taberlet P., Luikart G. (2004) African origins of the domestic donkey. *Science*, **304**, 1781.

Belkhir K., Borsa P., Chikhi L., Raufaste N., Bonhomme F. (1996–2004) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5171, Université de Montpellier II, Montpellier, France.

Bordonaro S., Guastella A.M., Criscione A., Zuccaro A., Marletta D. (2012) Genetic diversity and variability in endangered Pantesco and two other Sicilian donkey breeds assessed by microsatellite markers. *ScientificWorldJournal*, **2012**, 648427.

Brookshier F. (1974) Following the conquistadors. In: Publishing Division of the University (ed.), The Burro. University of Oklahoma Press, Norman, pp. 213–238.

Buchmann R., Hazelhurst S. (2014). Genesis Manual, University of the Witwatersrand, Johannesburg. (available at: http://www.bioinf.wits.ac.za/software/genesis/Genesis.pdf; last accessed 25 July 2015)

Delgado J.V., León J.M., Gómez M., Nogales S., Camacho M.E. (2010) Las razas ovinas ibéricas y su participación en la colonización de Iberoamérica. In: Servicio de Publicaciones de la Universidad de Córdoba (ed.), Biodiversidad Ovina Iberoamericana -Caracterización y Uso Sustentable-. Universidad de Córdoba, Córdoba, pp. 17–30.

Earl D.A., vonHoldt B.M. (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.*, **4**, 359–361.

Evanno G., Regnaut S., Goudet J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.*, **14**, 2611–2620.

Falush D., Stephens M., Pritchard J.K. (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

Ferrando A., Goyache F., Jordana J. (2008) Cuantificación de la diversidad genética en las razas asnales españolas para estrategias de conservación. In: IX Simposio Iberoamericano sobre Conservación y Utilización de Recursos Zoogenéticos. Mar de Plata (Argentina), 10–12 December 2008. Memorias, Tomo I, pp. 271–274.

Goudet J. (1995) FSTAT (vers. 1.2): a computer program to calculate F-statistics. *J. Hered.*, **86**, 485–486.

Goudet J. (2001) FSTAT, a Program to Estimate and Test Gene Diversities and Fixation Indices (version 2.9.3). (available at: http://www.unil.ch/izea/softwares/fstat.html. Updated from Goudet (1995).

Hubisz M.J., Falush D., Stephens M., Pritchard J.K. (2009) Inferring weak population structure with the assistance of sample group information. *Mol. Ecol. Resour.*, **9**, 1322–1332.

Jakobsson M., Rosenberg N.A. (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.

Kimura B., Marshall F.B., Chen S., Rosenbom S., Moehlman P.D., Tuross N., Sabin R.C., Peters J., Barich B., Yohannes H., Kebede F., Teclai R., Beja-Pereira A., Mulligan C.J. (2011) Ancient DNA from Nubian and Somali wild ass provides insights into donkey ancestry and domestication. *Proc. R. Soc. Lond. B*, **278**, 50–57.

Laguna E. (1991) El Ganado Español, un descubrimiento para América. Ministerio de Agricultura, Pesca y Alimentación, Madrid, Spain.

Langella O. (2002) Populations 1.2.30. Copyright © 1999, Olivier Langella, CNRS UPR9034. (available at: http://bioinformatics.org/project/?group_id=84; last accessed 27 July 2015)

López C., Alonso R., de Aluja A.S. (2005) Study of genetic origin of the Mexican Creole donkey (*Equus asinus*) by means of the analysis of the D-loop region of mitochondrial DNA. *Trop. Anim. Health Prod.*, **37**(Suppl. 1), 173–188.

Nei M., Tajima R., Tateno Y. (1983) Accuracy of estimated phylogenetic threes from molecular data. *J. Mol. Evol.*, **19**, 153–170.

Page R.D.M. (1996) TREEVIEW: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.*, **12**, 357–358.

Peakall R., Smouse P.E. (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes*, **6**, 288–295.

Peakall R., Smouse P.E. (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. *Bioinformatics*, **28**, 2537–2539.

Primo A.T. (2004) América: conquista e coloniçao. A fantástica história dos conquistadores ibéricos e seus animais na era dos descobrimentos, Ed. Movimento, Porto Alegre, Brazil.

Pritchard J.K., Stephens M., Donnelly P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Reynolds J., Weir B.S., Cockerham C.C. (1983) Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics*, **105**, 767–779.

Rodero A., Delgado J.V., Rodero E. (1992) Primitive Andalusian Livestock and their implication in the Discovery of America. *Arch. Zootec.*, **41**, 383–400.

Romagosa J.A. (1959). El Garañón Catalán. PhD thesis, Universidad de Madrid, Madrid, Spain, pp. 326.

Santos S.A., Sereno J.R.B., Mazza M.C.M., Mazza C.A. (1992) Origin of the Pantaneiro horse in Brazil. *Arch. Zootec.*, **41**, 371–381.

Sponenberg D.P. (1992) Colonial Spanish sheep, goats, hogs, and asses in the United States. *Arch. Zootec.*, **41**, 415–419.

Yanes J.E. (2005) Antecedentes históricos. In: Ministerio de Agricultura, Pesca y Alimentación (ed.), Razas asnales autóctonas españolas, Ministerio de Agricultura, Pesca y Alimentación, Madrid, pp. 39–97.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1 Main shipping routes of donkeys from Europe to America and subsequent colonization pathways throughout the continent (blue arrows) in the 16th and 17th centuries.

Figure S2 Geographical location of sampled animals.

Figure S3 (a) Genome distribution of American donkeys to clusters inferred with STRUCTURE. Plots for the most relevant values of K are shown. (b) Mean likelihood $L(K)$ and standard deviation (SD) over 20 runs from $K = 1$ to $K = 13$. (c), ΔK for each value of K.

Figure S4 Results obtained with STRUCTURE for the analysis within American cluster A averaged over 10 runs, from $K = 1$ to $K = 8$: (a), without sampling location as prior, and (b), country of origin is used as prior.

Figure S5 Results obtained with STRUCTURE for the analysis within American cluster B averaged over 10 runs, from $K = 1$ to $K = 5$: (a), without sampling location as prior, and (b), country of origin is used as prior.

Figure S6 Results obtained with STRUCTURE for the global clustering of American and European populations averaged over 50 runs, from $K = 2$ to $K = 27$.

Table S1 Panel of microsatellite primers used for genotyping domestic donkey *Equus asinus* and diversity parameters of the American metapopulation: NA, number of alleles, H_O, observed heterozygosity, H_E, unbiased expected heterozygosity, F_{IS}, fixation index.

Table S2 Main diversity parameters from each American and European population included in this study for a panel of 13 microsatellite markers observed heterozygosity (H_O), unbiased expected heterozygosity (H_E), mean number of alleles (MNA), rarefaction of allelic richness to 12 individuals (AR), and number and frequency of private alleles (PA).

Table S3 D_A distance matrix among American populations.

