Ancient DNA reveals traces of Iberian Neolithic and Bronze Age lineages in modern Iberian horses

JAIME LIRA,*† ANNA LINDERHOLM,‡ CARMEN OLARIA,§ MIKAEL BRANDSTRÖM DURLING,¶ M. THOMAS P. GILBERT,** HANS ELLEGREN,†† ESKE WILLERSLEV,** KERSTIN LIDÉN,‡ JUAN LUIS ARSUAGA*† and ANDERS GÖTHERSTRÖM*††

*Centro UCM-ISCIII de Investigación sobre Evolución y Comportamiento Humanos, 28029-Madrid, Spain, †Departamento de Paleontología, Universidad Complutense de Madrid, 28040 Madrid, Spain, ‡Archaeological Research Laboratory, Stockholm University, SE-10691 Stockholm, Sweden, §Laboratori d'Arqueologia Prehistòrica. Universitat Jaume I, 12071 Castelló de la Plana, Spain, ¶Department of Forest Mycology & Pathology, Swedish University of Agricultural Sciences, S750 07 Uppsala, Sweden, **Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen, Denmark, ††Department of Evolutionary Biology, Uppsala University, S752 36 Uppsala, Sweden

Abstract

Multiple geographical regions have been proposed for the domestication of *Equus caballus*. It has been suggested, based on zooarchaeological and genetic analyses that wild horses from the Iberian Peninsula were involved in the process, and the overrepresentation of mitochondrial D1 cluster in modern Iberian horses supports this suggestion. To test this hypothesis, we analysed mitochondrial DNA from 22 ancient Iberian horse remains belonging to the Neolithic, the Bronze Age and the Middle Ages, against previously published sequences. Only the medieval Iberian sequence appeared in the D1 group. Neolithic and Bronze Age sequences grouped in other clusters, one of which (Lusitano group C) is exclusively represented by modern horses of Iberian origin. Moreover, Bronze Age Iberian sequences displayed the lowest nucleotide diversity values when compared with modern horses, ancient wild horses and other ancient domesticates using nonparametric bootstrapping analyses. We conclude that the excessive clustering of Bronze Age horses in the Lusitano group C, the observed nucleotide diversity and the local continuity from wild Neolithic Iberian to modern Iberian horses, could be explained by the use of local wild mares during an early Iberian domestication or restocking event, whereas the D1 group probably was introduced into Iberia in later historical times.

Keywords: Bronze Age, domestication, *Equus caballus*, Iberian Peninsula, Middle Ages, mitochondrial DNA, Neolithic, restocking

Received 4 October 2008; revision received 5 October 2009; accepted 8 October 2009

Introduction

Of all the main domestic livestock species (cattle, sheep, goats, pigs and horses), the domestic horse (*Equus cabal-*

Correspondence: Jaime Lira, Fax: 00 34 91 387 75 48;
E-mail: jlira@isciii.es
All dates in the text are reported in conventional radiocarbon age. When different, calibrated years (Cal) BC or AD dates are detailed.
Data deposition footnote: The Portalón and Cova Fosca *Equus caballus* sequences have been deposited in the GenBank with the following accession numbers: DQ683525-DQ683544, GU066502-GU066503.

lus) probably underwent the most complex domestication event, as evidenced through a combination of high mitochondrial diversity with weak phylogeographical structure (MacHugh & Bradley 2001). It has been proposed that the earliest domestication of horses took place in the Eurasian steppes, between the fifth and fourth millennium BC (Anthony & Brown 2003). Horse remains from the sites of Dereivka (Ukraine) and Botai (Kazakhstan) have been widely studied (Levine 2005), providing different versions on the process of this domestication (Anthony & Brown 2003; Levine 2005). A recent study based on metrical, biting damage and organic residue analyses of horse and pottery remains

from the Botai Culture (mid-fourth millennium BC) suggested that these horses were similar to Bronze Age domestic horses, probably ridden and used as a source of milk (Outram *et al.* 2009). In another study, genetic analyses into ancient horse coat colour genetics from samples obtained throughout Eurasia, and spanning in age from the Late Pleistocene to medieval times, similarly confirmed the origin of domestic horses in the Eurasian steppes around 5000 BP (Ludwig *et al.* 2009). Moreover, early remains of harness equipment, presumably related to horse control, have been associated with the Early Bronze Age (Harding 2000). Finally, artistic and textual records place the undisputable evidence of domestic horses in the human societies at least at the end of the third millennium BC (Levine 2005).

High matrilineal diversity has been found in modern horses (Lister *et al.* 1998; Vilà *et al.* 2001; Jansen *et al.* 2002). In contrast, no variation was found when 14.3 kb of Y chromosomal DNA was screened in domestic stallions (Lindgren *et al.* 2004). This pattern could be explained by the polygamous behaviour of this species, but also hints at a domestication process, where only a restricted number of domestic stallions were involved. Once the domestic stocks were fully established, they could have been occasionally restocked with wild mares from different local populations. Alternatively, such genetic signals may have been the reflection of a stock-breeding tradition, where only very few stallions would have been used for a large number of mares. In either case, the observed mtDNA variation is best explained by an involvement of local mares during the domestication process.

A study on mtDNA from 652 horses worldwide has proposed the existence of up to 17 clusters, representing 17 distinct genetic lineages (Jansen *et al.* 2002). Some of these clusters seem to exhibit a strong geographical signature (Jansen *et al.* 2002). One such case is the D1 mtDNA group, which appears to be associated mainly with breeds originating in the Iberian Peninsula and Northern Africa (Jansen *et al.* 2002). From Iberia, 50% of the Andalusian horses and 56% of the Lusitano horses analysed in Jansen *et al.* (2002) belonged to this cluster. Furthermore, 54% of the North African Barb horses also belonged to D1 group. Other horses with a probable Iberian origin also clustered within it (e.g. 31% of the Mustang horses analysed) (Jansen *et al.* 2002). However, all these breeds were also present in other clusters. For example, Andalusian horses also clustered in at least four other groups and the Lusitano horses in at least one group other than D1. An additional mtDNA study that focused on modern Lusitano horses found their haplotypes clustering in four main groups (A, B, C and D). Cluster A appeared overrepresented in the Lusitano sample, as this cluster likely

corresponds to the D1 grouping previously described (Lopes *et al.* 2005). Finally, a study on horses from a largest sample of Iberian breeds (Royo *et al.* 2005), found five major clusters (I-V), corresponding Ia to D1, D2 and D3 clusters defined by Jansen *et al.* (2002).

Iberian horses display a heterogeneous mtDNA pattern, grouping with other modern breeds (Jansen *et al.* 2002). For reasons of the clustering of a majority of horses of Iberian origin within D1 group, an Iberian origin for this group was proposed (Jansen *et al.* 2002; Royo *et al.* 2005). Even though the D1 group is currently the most common group for Iberian horses, nothing is known about its presence in the Iberian Peninsula during prehistoric times. Previous analysis on six ancient Iberian mtDNA samples yielded no association with the D1 group (Seco-Morais *et al.* 2007).

It has been proposed that wild stocks from different localities were used in the horse domestication process and wild Iberian horses were involved (Zeuner 1963). This hypothesis has been further supported by metrical analysis of Iberian horse remains from Neolithic-Bronze Age periods (Uerpmann 1990) and a local domestication event was suggested for some Calcolithic-Bronze Age horse populations.

An alternative hypothesis is that modern Iberian breeds had an origin outside the Iberian Peninsula, but subsequently developed locally into their current form (Lopes *et al.* 2005; Royo *et al.* 2005). One such hypothesis is that the northern breeds (i.e. Garrano, Caballo de Corro, Asturcón, Pottoka and Losino) originated from Celtic migrations, while the southern breeds (i.e. Andalusian, Carthusian, Marismeño, Lusitano or Sorraia) might have originated in North Africa (Lopes *et al.* 2005; Royo *et al.* 2005). In addition, previous studies of the morphological traits of the Sorraia horses have established this breed as the most likely ancestral type of some of the modern Iberian breeds (D'Andrade 1945).

To investigate the antiquity of the D1 group in the Iberian Peninsula and to explore other possible maternal lineages structures, we have analysed mtDNA from 22 horse remains from the Neolithic, Bronze Age and Middle Ages levels from the Spanish archaeological sites of El Portalón de Cueva Mayor (Sierra de Atapuerca, Burgos) and Cova Fosca (Ares del Maestre, Castellón).

Material and methods

Archaeological information

The Portalón site constitutes one of the present day entrances to the Cueva Mayor karst complex in the Sierra de Atapuerca (Burgos), located on the Northern

Iberian Plateau (Fig. 1). The stratigraphic sequence from Portalón site has been extensively radiocarbon-dated, yielding materials that range from the Upper Pleistocene to medieval times (Carretero *et al.* 2008). The sequence contains a Bronze Age layers with an extensive *Equus caballus* record (Castaños 2005). Despite the recovery of a large number of horse remains from the Portalón Bronze Age layers, the establishment of their domestication status has been difficult (Castaños 2005).

We collected 22 horse remains from Portalón (Table S1). Nine were excavated from stratigraphic layers associated with the Initial and Middle Bronze Age, whereas the other 13 were from disturbed layers coming from the clandestine excavation area (Carretero *et al.* 2008) (Supporting Text 1). Thirteen out of the 22 remains (Table 1, Table S1) were radiocarbon-dated. Teeth identification was performed according to the guidelines of the 'New York International Hipparion Conference, 1981' (Eisenmann *et al.* 1988).

The site of Cova Fosca is located in Ares del Maestre (Province of Castellón) in the East of the Iberian Peninsula (Fig. 1). Cova Fosca site displays a stratigraphic sequence with Epipaleolithic to Neolithic levels (Olària 1988). Two horse samples were collected from Neolithic levels and were directly radiocarbon-dated (Table 1, Supporting Text 1). The Cova Fosca samples have been considered as samples from wild horses based on their Neolithic chronology.

Ancient DNA extraction

Ancient DNA was extracted in a dedicated ancient DNA laboratory in Madrid (Centro Mixto UCM-ISCIII, Spain). Two different extraction techniques were used, one based on a phosphate buffer for extraction followed by hybridization and magnetic separation techniques (for biotinylated primers details see Table S2) (Anderung *et al.* 2005), the other based on silica spin columns (Yang *et al.* 1998). The mitochondrial control region was amplified in three overlapping fragments spanning nucleotide positions (np) 15444-15842 (Xu & Árnason 1994). Overlapping fragments were obtained by H1F-H1R, H2F-H2R & H3F-H3R primer pairs (Vilà *et al.* 2001) and by JLH2F-JLH2R primer pairs designed for this study (Table S3). Each primer pair amplified approximately 150 bp. PCR was performed on 9 μl of the beads obtained from the magnetic separation and 5 μl of those from the silica purification. All PCR reactions were undertaken using $1\times$ PCR Buffer (Qiagen), 2.5 mM $MgCl_2$ (Qiagen), 200 μM dNTPs (Sigma), 0.2 μM of each primer and three Units of HotStarTaq DNA Polymerase (Qiagen) in a final volume of 25 μl. The PCR conditions were 10 min at 95 °C, followed by 54 cycles of 30 s at 94 °C, 30 s at 50 °C, 30 s at 72 °C, with a final extension of 7 min at 72 °C.

Amplicons were purified using the ExoSAP-IT purification kit (Amersham Biosciences) following the manufacturer's instructions. Sequencing reactions were performed using the DYEnamic Terminator Cycle sequence kit, and sequences were visualized on a MegaBACE 96 capillary system (GE Healthcare).

To obtain a more detailed insight into the composition of the DNA extracts, we deep-sequenced three of the PCR products from each of five of the samples (ATA03, ATA05, ATA11, ATA12 and ATA18) using Roche GS FLX sequencing chemistry (Roche). Specifically, following PCR purification each of the three PCR products were pooled at an equimolar ratio into five composite pools (one per sample). The DNA in these pools was subsequently converted into FLX libraries using the manufacturer's MID tagging protocol. Each library included a unique oligonucleotide barcode, that post sequencing was used to allocate the sequences to the original sample source. Once built, the five libraries were quantified against a known standard using quantitative PCR (Meyer *et al.* 2008), pooled at equimolar ratio into a final single pool, then emPCR'd and sequenced on 1/8 of a sequencing plate (following the manufacturer's guidelines) (Table S12).

DNA authentication

Standard ancient DNA approaches were taken with regards to preventing the possibility of contamination (Gilbert *et al.* 2005; Willerslev & Cooper 2005). In addition to the standard laboratory procedures, the following are of particular note. Firstly, negative extraction controls, as well as non target-species controls (*Bos taurus* and *Cervus* sp./*Capra* sp. samples for the Portalón

Fig. 1 Geographical location of the Portalón de Cueva Mayor and Cova Fosca sites within the Iberian Peninsula.

Table 1 Stratigraphic position and radiocarbon dates of the Portalón (ATA) and Cova Fosca samples (CFo)

Sample	Stratigraphic position	Conventional radiocarbon age	2 sigma calibrated result
ATA01	Disturbed layers	(Beta-200991) 3640 ± 40 BP	Cal BC 2130 – 1900
ATA02	Disturbed layers	(Beta-253407) 3370 ± 40 BP	Cal BC 1750 – 1590 & Cal BC 1590 – 1530
ATA03	Disturbed layers	(Beta-200992) 3500 ± 40 BP	Cal BC 1920 – 1720
ATA04	Disturbed layers	(Beta-253409) 3690 ± 40 BP	Cal BC 2200 – 1960
ATA05	Disturbed layers	(Beta-253410) 3700 ± 40 BP	Cal BC 2200 – 1970
ATA06	Disturbed layers	(Beta-253411) 3770 ± 40 BP	Cal BC 2300 – 2120 & Cal BC 2090 – 2040
ATA07	Disturbed layers	(Beta-204455) 1010 ± 40 BP	Cal AD 980 – 1050 & Cal AD 1100 – 1140
ATA08	Disturbed layers	(Beta-253412) 3690 ± 40 BP	Cal BC 2200 – 1960
ATA09	Disturbed layers	(Beta-200993) 3600 ± 40 BP	Cal BC 2040 – 1880
ATA10	Disturbed layers	(Beta-253413) 3460 ± 40 BP	Cal BC 1890 – 1680
ATA11	*In situ*		
ATA12	*In situ*	(Beta-204454) 3760 ± 50 BP	Cal BC 2310 – 2030
ATA13	*In situ*		
ATA14	*In situ*	(Beta-200994) 3630 ± 40 BP	Cal BC 2130 – 2080 & Cal BC 2060 – 1890
ATA15	*In situ*		
ATA16	*In situ*		
ATA17	*In situ*		
ATA18	*In situ*		
ATA19	*In situ*		
ATA20	Disturbed layers	(Beta-253408) 3910 ± 40 BP	Cal BC 2480 – 2290
CFo 01	*In situ*	(Beta-222754) 6230 ± 50 BP	Cal BC 5310 – 5040
CFo 02	*In situ*	(Beta-227420) 6200 ± 50 BP	Cal BC 5300 – 5010

and the Cova Fosca samples respectively) were incorporated. Neither yielded any evidence of horse DNA, indicating that sample cross-contamination was unlikely. Secondly, DNA preservation seems plausible. In particular, DNA from the *Bos taurus* samples was extracted and analysed. Cattle yielded *Bos taurus* mtDNA, but no horse sequences were obtained in any of the samples. Moreover, a total of 15 horses were analysed for biochemical preservation through collagen extraction and quantification (Table S4). As a further control, DNA sequences were replicated from six of the horses in an independent laboratory in Uppsala (Sweden), yielding partial sequences with concordant results (Table S4). Lastly, to ensure that post-mortem damage derived miscoding lesions did not affect the sequence accuracy, several extracts were performed per specimen, and several PCR amplifications were sequenced per extract (Hofreiter *et al.* 2001). Depending on each fragment and sequence, a minimum of two and a maximum of nine sequences were performed (Table S5).

DNA analyses

The sequences from Portalón and Cova Fosca horses were aligned with those from a wide variety of modern and ancient horses (Table S6). These included modern Iberian horse breeds, extant horses from breeds of Iberian origin and extant horses from other non-Iberian

breeds, Pleistocene horses from Alaska dated 28 000– 12 000 BP (Vilà *et al.* 2001), Ireland dated around 27 000 BP (McGahern *et al.* 2006), Pleistocene-Holocene *Equus* sp. from America and Eurasia (Weinstock *et al.* 2005), horses from the Scythian period (3rd century BC) (Keyser-Tracqui *et al.* 2005) and the Roman period (1st century AD) (Di Bernardo *et al.* 2004), ancient Swedish horses from historical times and Viking specimens (Vilà *et al.* 2001), horses from the Kwakji archaeological site in Cheju, Korea (AD 700–800) (Jung *et al.* 2002) and from Yakutia (17th or 18th century AD) (Keyser-Tracqui *et al.* 2005), ancient Irish and British horses (McGahern *et al.* 2006), as well as six ancient Iberian sequences related to Upper Pleistocene, Iron Age and from the Battle of Aljubarrota (AD 1385) (Seco-Morais *et al.* 2007). A total dataset of 1140 D-loop sequences obtained from GenBank were used (Accession Numbers in Table S6). The ancient Scythian, Roman, North European and Holocene Irish and British horses, as well as horses from Cheju and Yakutia were considered as ancient domesticates. Roman samples were classified as horses/mules, and Weinstock *et al.* (2005) samples were classified as *Equus* sp. (as their mtDNA are similar to *Equus caballus*, they were included in the analyses). Furthermore, we considered all the Weinstock *et al.* (2005) samples as samples from wild horses.

Sequences were aligned using BioEdit (Hall 1999) and Reduced Median Networks (RMNs) were constructed for the region 15487–15740 np using Network 4.5.1.0

(Bandelt *et al.* 1995) (available at http://www.fluxus-engineering.com).

Two networks were produced. One network was generated with the modifications suggested by Jansen *et al.* (2002): nucleotide positions 15585, 15597 and 15650 were considered mutational hotspots and eliminated from the analysis. Moreover, positions 15659 and 15737 were down-weighted (1.0) (Fig. 2A and B, Fig. S2). The second network was generated without modifications (Fig. S1). Both RMNs were constructed with default weight chosen at $r = 1$ (Supporting Text 1). As the ancient Iberian sequences obtained by Seco-Morais *et al.* (2007) did not align completely with the remaining sequences, we could not include them in the RMNs.

Nucleotide diversity analysis

Bronze Age nucleotide diversity, as measured by the average pairwise difference (π), was calculated for Bronze Age samples from Portalón. This statistic was also calculated for 28 Pleistocene–Holocene wild ancient horses (26 *Equus* sp. with Cova Fosca samples CFo 01 & CFo 02 included), 18 ancient domesticates (that included sequence ATA07 and that excluded the Yakut sequences from 17th-18th century AD), 399 modern Iberian (excluding the Pomba sequence from Luís *et al.* 2006a), and 689 modern horses from other breeds (Table S8). Nonparametric bootstrapping with 10 000 replicates was used to estimate the empirical 95% Confidence Interval (CI) for the nucleotide diversity. Population differentiation was also tested using nonparametric bootstrapping approaches, where π was compared between groups for each bootstrap replicate to calculate exact bootstrap P-values. The two approaches used were pairwise comparisons, as well as simultaneous bootstrapping, where all the groups were compared simultaneously for each replicate (Fig. 3). This approach is useful to test if a single group of sequences has a lower or higher π than the remaining groups. It is, however, a very conservative test.

Neutrality test analysis

Previous studies on modern and ancient domestic cattle have considered observed departures from a null hypothesis of constant population size under the neutral model as population growth as a sign of artificial selection pressure (Bollongino *et al.* 2006). We evaluated the probability of a population growth in horse samples using Tajima's D, Fu's F_s and R_2 statistics (Tajima 1989; Fu 1997; Ramos-Onsins & Rozas 2002). Both Fu's F_s and R_2 statistics are considered as powerful tests for detect-ing population growth. Moreover, the R_2 statistic offers good resolution for detecting population expansion when working with small sample size (Ramos-Onsins & Rozas 2002). Tajima's D and Fu's F_s tests were performed on the Bronze Age Portalón sequences (Table S9), and other subsamples of modern horses from the nonmodified RMN (Table S10). Some of these modern subsamples were distributed in a star-like pattern. The R_2 statistic was performed only on the Bronze Age Portalón sample. Tajima's D and Fu's F_s were carried out with Arlequin Ver. 3.11 (Excoffier *et al.* 2005). Statistical significance of the Tajima's D and Fu's F_s tests was adjusted for multiple comparisons by using sequential Bonferroni corrections (Rice 1989). The R_2 statistic was computed using DnaSP v5.0 (Librado & Rozas 2009). The significance of the R_2 value was evaluated with coalescent simulations fixing the number of segregating sites and comparing the observed value with a null distribution generated by 10 000 replicates with a significance level $\alpha = 0.05$.

Population structure

We investigated population differentiation by grouping the Bronze Age Portalón horses with modern horses, wild Pleistocene-Holocene and other ancient domestic horses (Table S11), and calculating Φ_{ST} under the Analysis of Molecular Variance (AMOVA) (Excoffier *et al.* 1992). Statistical significance was also adjusted for multiple comparisons by sequential Bonferroni corrections (Rice 1989).

Phylogenetic analysis

To test the consistency of the haplotype relationships obtained in the RMNs, phylogenetic relationships were estimated using Bayesian analysis and Maximum Likelihood (ML). The phylogenetic trees were constructed over the haplotypes presented in the modified (202 haplotypes, Fig. 4) and nonmodified RMNs (246 haplotypes, Fig. S3). Subsequently, we included the sequences produced in this study.

The model for DNA substitutions that best fitted the data was selected with ModelTest3.7 (Posada & Crandall 1998) under Hierarchical Likelihood Ratio Test (hLRTs) approach. The model chosen was Tamura & Nei with a gamma shape parameter (α) 0.4534 for the modified RMN dataset and (α) 0.3487 for the nonmodified RMN dataset.

Bayesian analysis was performed with MrBayes 3.1 (Ronquist & Huelsenbeck 2003) using the General Time Reversible (GTR) model of evolution, with gamma-distributed rate variation across sites, setting the prior gamma shape parameter for both datasets
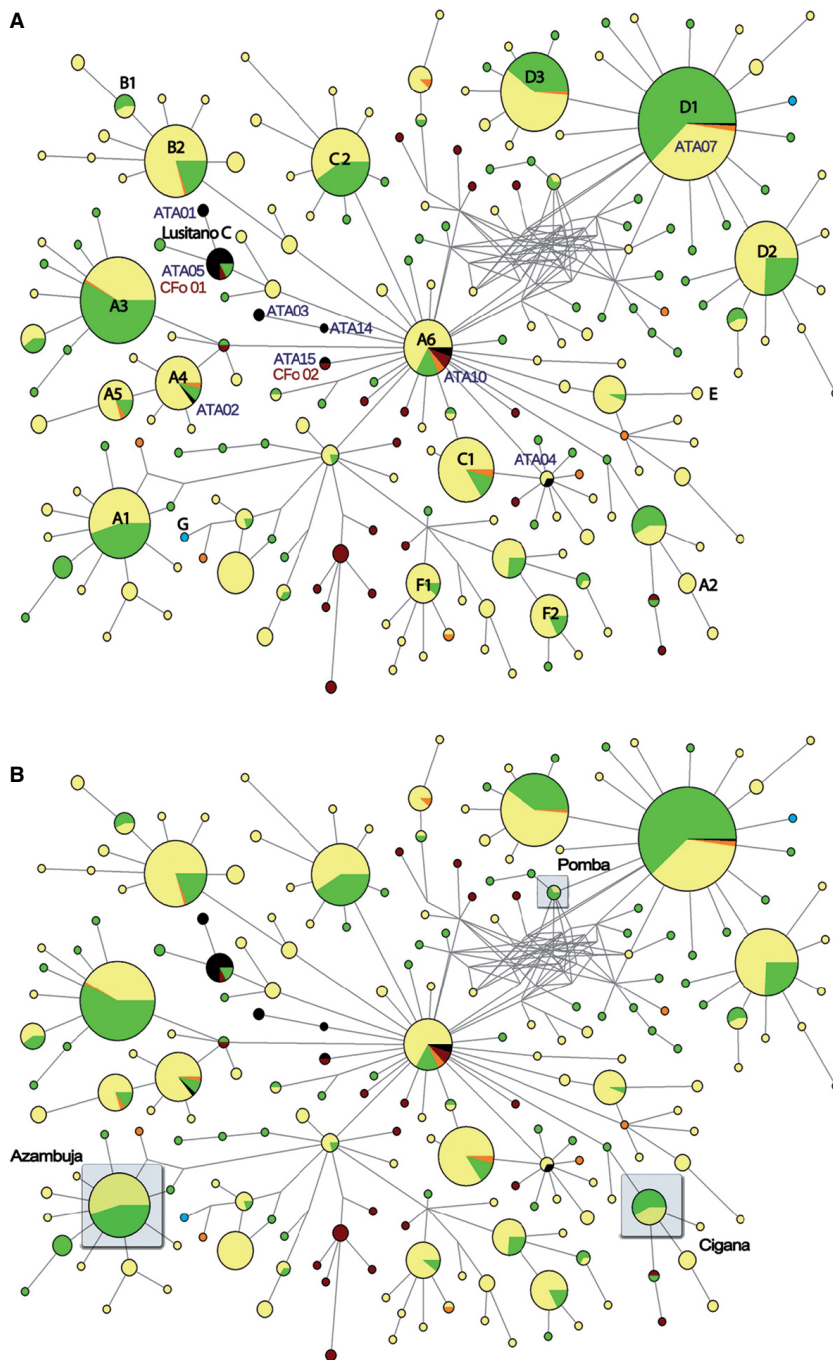
Fig. 2 (A) Network distribution of the Portalón and Cova Fosca samples. Modified RMN of 1156 mtDNA sequences (np 15487 to 15740). Colours code: *black* for ancient Iberian samples; *green* for modern Iberian horses and horses belonging to breeds of an Iberian origin; *yellow* for modern non-Iberian horses, *orange* other ancient samples through Eurasia, *brown* for wild ancient horses. Lusitano group C includes the following sequences and haplotypes: sequences ATA05, ATA09, ATA12, ATA13, ATA16, ATA17, ATA18, ATA19, ATA20 and CFo 01 sharing a haplotype; ATA01 and ATA06 share a haplotype. In the A6 cluster, ATA10 and ATA11 share a haplotype. ATA15 & CFo 02 are one mutation away from A6, and share a haplotype. ATA08 and ATA03 share a haplotype, and are close to the ATA14 haplotype. Both haplotypes are three and two mutations away to A6 respectively. ATA02 is located in A4 group, and ATA04 is located one mutation away to the main C1 haplotype. ATA07 is located in D1 group. Branches are not proportional to their lengths (for identification of rest of the haplotypes and mutations, see Fig. S2 and Supporting Text 3). (B) Maternal lineages in the modern Sorraia horses.
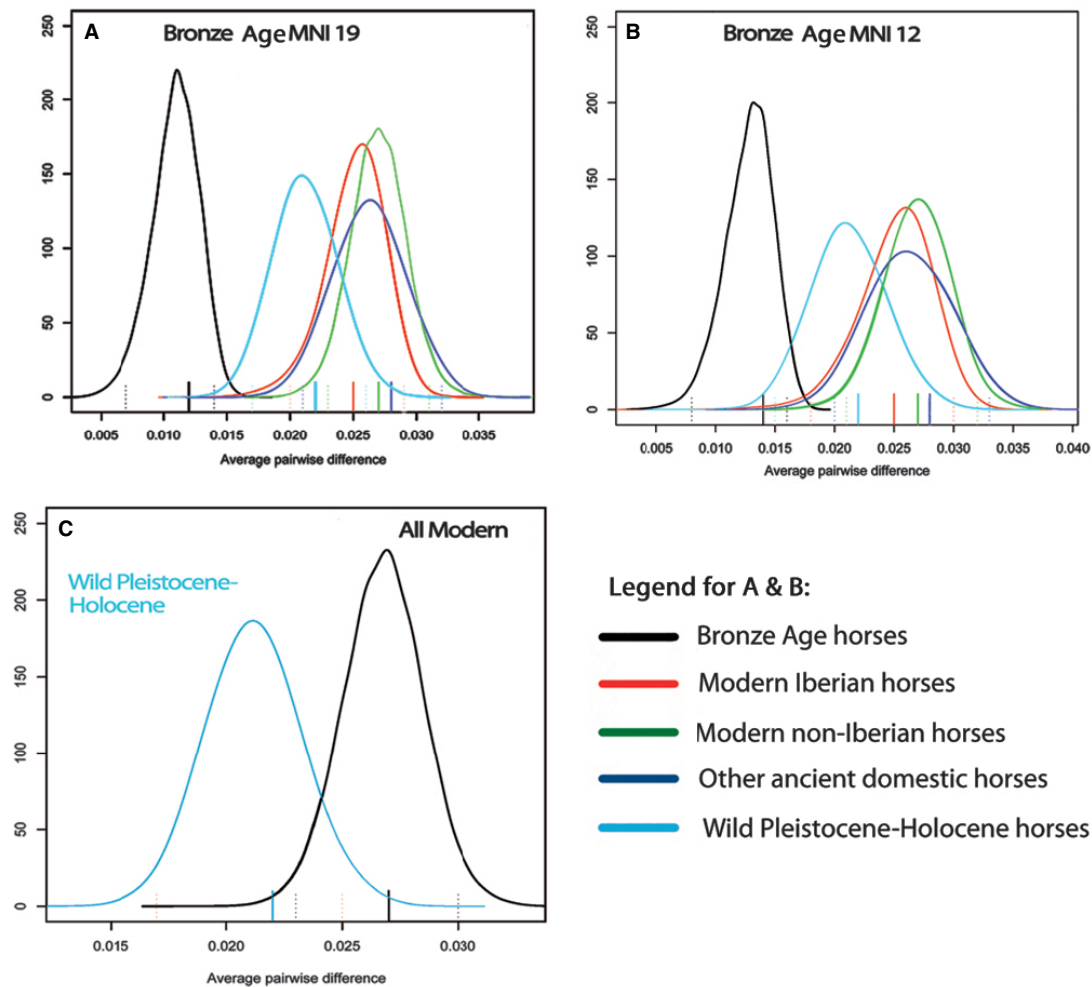
Fig. 3 Distribution of mean number of pairwise differences after nonparametric bootstrapping (10 000 replicates with size correction) among (A) Portalón Bronze Age horses (MNI = 19) with modern Iberian horses, modern non-Iberian horses, other ancient domesticates and wild ancient horses. (B) Portalón Bronze Age horses (MNI = 12) with modern Iberian horses, modern non-Iberian horses, other ancient domesticates and wild ancient horses (C) 28 wild ancient horses with 1088 modern horses.

(trees from Fig. 4.1 and Fig. S3.1). The GTR model was selected as the closest model to Tamura & Nei implemented by MrBayes, following Driscoll *et al.* (2007). The tree-space was explored by four chains over 10^7 generations sampling every 1000 generations. We obtained 10 000 trees, fixing the burn-in value at 25%.

Maximum Likelihood trees were generated using a heuristic search. An initial tree was constructed using a fast distance-based algorithm, and subsequently improved by a subtree pruning regrafting (SPR) algorithm as the branch swapping option using PhyML3.0 web server (Guindon & Gascuel 2003, http://www.atgc-montpellier.fr/phyml/). Reliability of the branches was examined generating 100 bootstrap pseudo-replicates.

Moreover, we tested the phylogenetic relationships of our ancient Iberian sequences with those published by Seco-Morais *et al.* (2007) with a Bayesian and a ML approach for the region 15444-15842 np. We also included in the analysis Lusitano and Pura Raza Española sequences (AY293975-AY293991, AY525091-AY525096, AY805641-AY805664) as well as sequences from the Cigana (AJ413909), Azambuja (AJ413918) and Pomba (DQ339575) as Sorraia maternal lineages. Inferences were performed under the Tamura & Nei model of nucleotide substitution with a proportion of invariable sites (0.5827) and a Gamma distribution shape parameter (0.6267) selected by ModelTest3.7 under hLRTs approach. The Bayesian tree was generated with MrBayes 3.1 under GTR model, setting the proportion of invariable sites as well as the Gamma

4.1. Bayesian inference
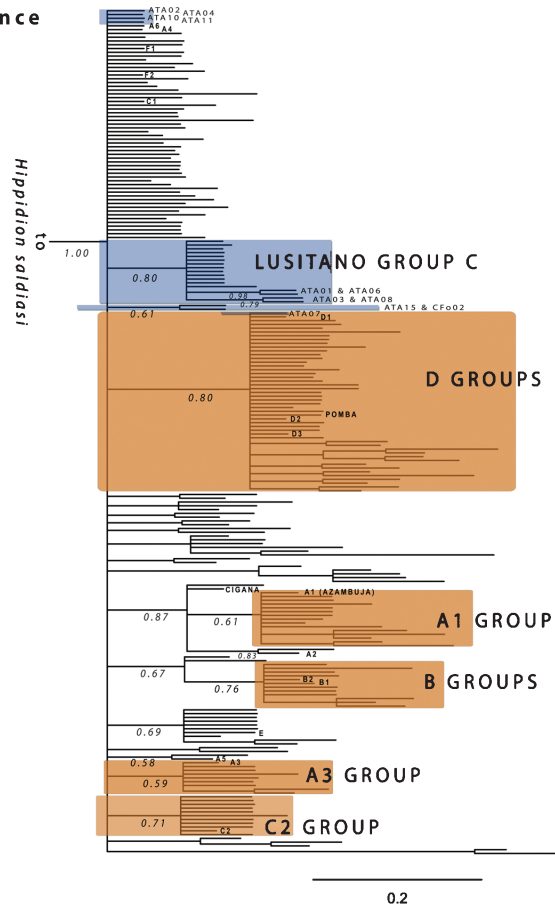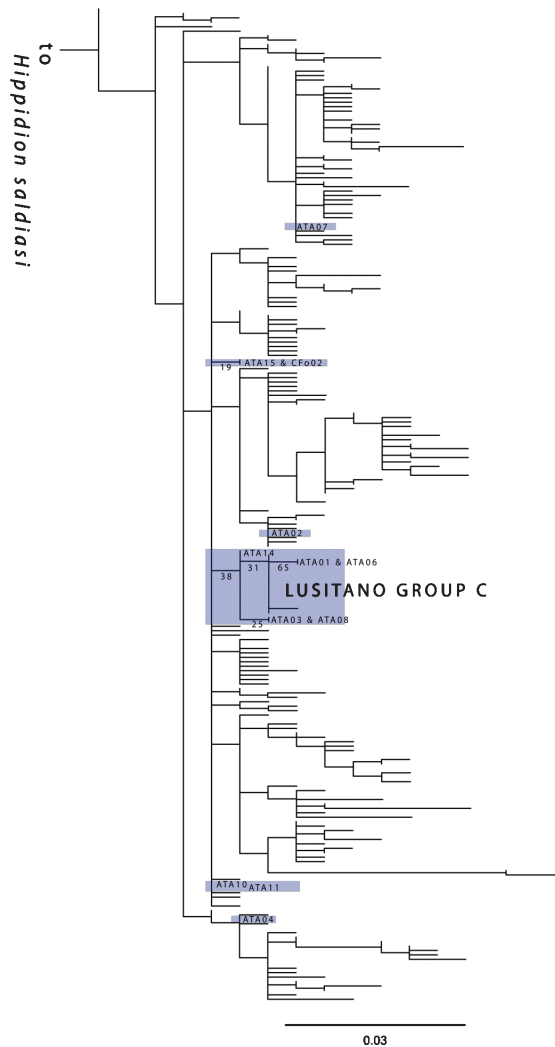
4.2. ML inference

Fig. 4 Phylogenetic trees from the modified RMN haplotypes (Fig. 2). **Tree 4.1. Bayesian inference**. Samples from the Portalón and Cova Fosca sites are highlighted in blue. Motifs from the haplotypes defining the Jansen groups and Sorraia maternal lineages are also added. Groups whose haplotypes cluster together (as in the modified RMN) are highlighted in orange. Posterior probability values are placed below branches. **Tree 4.2. ML inference**. Samples from Portalón and Cova Fosca sites are highlighted in blue. Bootstrap values are placed below branches. Lusitano group C contains the following ancient Iberian samples in both trees: ATA01, ATA03, ATA05, ATA06, ATA08, ATA09, ATA12, ATA13, ATA14, ATA16, ATA17, ATA18, ATA19, ATA20 and CFo01.

shape parameter. Analysis was performed with the same number of chains, generations, sampling frequency and burn-in value as above. The ML tree was gener-ated in the same way as above. Reliability of the branches was assessed by 100 bootstrap resamplings (Fig. 5).

The ML and Bayesian trees were rooted with three *Hippidion saldiasi* sequences (AY152861, AY152862, AY152863) as outgroups (from Orlando *et al.* 2003). FigTree v1.2.2 was used to draw the trees (http://tree.bio.ed.ac.uk/software/figtree/).

Results

Chronologies

Two directly dated horse remains from Cova Fosca date to the Early Iberian Neolithic (CFo 01: 6230 ± 50 BP and CFo 02: 6200 ± 50 BP). Twelve of 13 Portalón samples span a timeframe around 3900–3370 BP. They have been associated to the transition Calcolithic – Bronze Age layer and with Initial-Middle Bronze Age layers (Table 1) (Carretero *et al.* 2008). The last Portalón sample dated back to the Middle Ages. All the Portalón samples recovered from stratigraphic layers have previously been associated to the Initial - Middle Bronze Age (see L Juez, University of Burgos, Burgos; Carretero *et al.* 2008 and Mínguez 2005 for association between faunal remains catalogue label and stratigraphic layers).

Haplotype distribution

Twenty out of 22 faunal remains from Portalón, and the two samples from Cova Fosca yielded the three overlapping mtDNA regions (Table S7). The modified RMN indicates that the ancient Iberian sequences distribute into nine haplotypes that occur at different frequencies (Fig. 2A and B). The medieval Portalón sequence (ATA07) clustered with the D1 group (Jansen *et al.* 2002) (Table S7). ATA04 clustered near the C1 group and ATA02 in the A4 group (both groups defined by Jansen *et al.* 2002) (Table S7). Two additional sequences (ATA10 & ATA11) shared a haplotype and clustered in the previously defined A6 group (Jansen *et al.* 2002) (Table S7), and one sequence (ATA15) was one mutation away from ATA10 and ATA11. Eleven sequences (ATA01, ATA05, ATA06, ATA09, ATA12, ATA13, ATA16, ATA17, ATA18, ATA19 & ATA 20) clustered in the Lusitano group C (Lopes *et al.* 2005) (Table S7). These are distributed in two haplotypes: ATA01 & ATA06 in one haplotype, and the remaining nine sequences in the other. Three sequences (ATA03 & ATA08, ATA14) clustered in different haplotypes, three and two mutations away respectively, from the A6 haplotype. Regarding the Cova Fosca samples, one sequence (CFo 01) clustered in the Lusitano group C, and the other (CFo 02) displayed an identical haplotype to ATA15 (Table S7). The nonmodified RMN displayed a similar distribution, but grouped ATA03, ATA08 and

ATA14 in the Lusitano group C and left ATA04 without a defined group (Fig. S1).

In addition to the D1 group, groups A6, A4 and C1 have previously been observed to contain some horses of Iberian origin (Jansen *et al.* 2002). The Lusitano group C was defined with a restricted number of individuals (Lopes *et al.* 2005) from the Lusitano breed (Fig. 2A, Supporting Text 2 and 3). In the networks, horses from modern American breeds also clustered within this group [an Argentinean Creole horse (Mirol *et al.* 2002) and a Puerto Rican Paso Fino horse (Luís *et al.* 2006b)]. This supports previous studies on the genetic links existing between Iberian and New World horses (Mirol *et al.* 2002; Luís *et al.* 2006b). Interestingly, for all the clusters with ancient Iberian sequences defined in the network, Lusitano group C is the only one with ancient Iberian sequences that incorporates modern individuals of exclusively Iberian origin.

That no samples from Cova Fosca, and just the medieval one from Portalón cluster in the D1 group is striking. Sample ATA07 carried the CCG motif from np 15494 to 15496 that is present in the horses from this group (Royo *et al.* 2005). The ancient Iberian sequences grouped in the Lusitano group C carried a deletion in np 15533, also common in the modern horses from this group. This deletion is also present in ATA03, ATA08 and ATA14.

Moreover, we found no relationship among the ancient Iberian samples and the Sorraia horses. The Sorraia breed constitutes three maternal lineages. None of the subfossil samples clustered with the Azambuja and Cigana lineages (Luís *et al.* 2002) (Fig. 2B). Recently, the new maternal Pomba lineage has been found to carry a haplotype close to the D groups (Luís *et al.* 2006a), and sample ATA07 is just one mutation away from the Pomba haplotype (Fig. 2B).

Interestingly, the 26 wild ancient samples downloaded from the GenBank represent 20 haplotypes that are widely distributed in the modified RMN. Some of these, from Weinstock *et al.* (2005), were placed close to the ancient Iberian sequences. Indeed sequences PLE14, PLE15 and PLE11 (from Siberia and Germany respectively) shared the same haplotype as ATA10 & ATA11. Moreover, both PLE10 and PLE18 (from Yukon and Siberia) were only one mutation away from the A6 group, whereas the Siberian sequence PLE20 was placed two mutations away. Further, PLE21 (from Germany) was one mutation away to ATA02. In the nonmodified RMN, the wild ancient sequences downloaded from the GenBank appeared in 22 haplotypes. PLE14 was near to Lusitano group C, and PLE11 shared haplotype with ATA10 & ATA11.
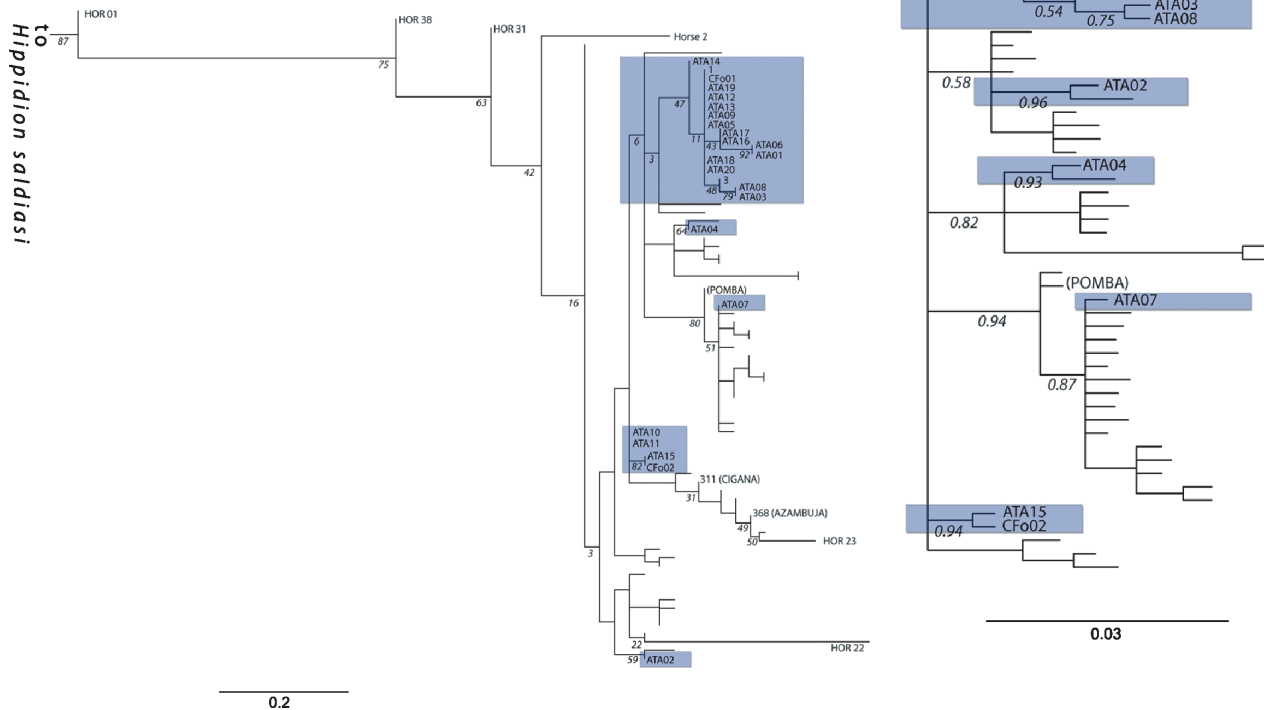
Fig. 5 Phylogenetic analysis comparing Cova Fosca and Portalón horses with other ancient and modern Iberian horse sequences. The Portalón and Cova Fosca branches are marked in blue. **(5.1) Bayesian inference**, with posterior probabilities indicated below branches. **(5.2) ML inference**, with bootstrapping values indicated below branches.

Nucleotide diversity in ancient and modern horses

The modified RMN described eight haplotypes from the Iberian Bronze Age (Fig. 2A). Nucleotide diversity was estimated taking into account sequences and individuals. To help account for the potential problem of analysing multiple remains from single individuals, we performed an estimation of the Minimum Number of Individuals (MNI). A MNI of 12 horse individuals from the Bronze Age Portalón were obtained by pool-

ing morphological, haplotype diversity and chronology data (ATA01, ATA02, ATA03, ATA04, ATA06, ATA09, ATA10, ATA12, ATA14, ATA15, ATA16 and ATA20).

The nucleotide diversity analysis based on these 12 individuals (Table S8), yielded $\pi = 0.014$ (95% CI: 0.008 – 0.016). Furthermore, to assess for a possible overestimation of the genetic diversity using these 12 individuals (as one of the criteria was haplotype differentiation), we performed the same genetic analysis including all

the 19 Bronze Age sequences. The 19 Bronze Age horses yielded a value of $\pi = 0.012$ (95% CI: 0.007–0.014). The two-fold estimations of nucleotide diversity among the Bronze Age horses (MNI = 12 and MNI = 19) was performed comparing each case to 399 modern Iberian, 689 modern non-Iberian, 18 ancient domesticated and 28 wild ancient horses, using nonparametric bootstrapping with size corrected resampling for 12 and 19 respectively, through 10 000 replicates (Fig. 3A and B, Table S8). Both sample sizes (12 and 19) yielded a significantly lower π value than any modern Iberian, modern non-Iberian, other ancient domesticates and wild ancient horses ($P = 0.0122$ and $P = 0.0002$ MNI = 12 & 19, Fig. 3A and B). Interestingly, when we compared π for the 28 wild ancient horses, with 1088 modern domestic horses, a significantly lower value was obtained for the former ($P = 0.0192$, nonparametric bootstrapping with size corrected resampling, sample size 28, 10 000 replicates) (Fig. 3C).

Examining changes in population size

To investigate whether the Bronze Age Portalón horses might have belonged to an expanding population, we calculated Tajima's D and Fu's F_s statistics (Table S9), as well as the R_2 statistic. The Tajima's D results were weakly consistent with a population expansion in Bronze Age Portalón, although no statistical significance was obtained. Similar results were obtained in modern Iberian ($n = 399$) and modern non-Iberian ($n = 689$) horses, as well as in the wild ancient horses ($n = 28$) and ancient domesticates ($n = 18$) (Table S9). We similarly tested subsets of modern domestic *Equus caballus* populations [as performed for modern domestic and wild *Sus* sp. samples by Larson *et al.* (2005)], to investigate whether they underwent population expansion. We obtained results indicating population expansions in modern Iberian and modern non-Iberian horses from D groups and also in other groups (Table S10).

In contrast to the Tajima's D results, Fu's F_s statistics indicated population expansion when performed on both Bronze Age Portalón dataset (MNI = 19 or 12), as well as in the whole modern Iberian, modern non-Iberian and ancient (wild and domestic) horses. Further, Fu's F_s statistics showed results of population growth in some subset of samples from the network, including modern Iberian and modern non-Iberian horses from D groups (Tables S9 and S10).

Lastly, when the R_2 statistic was used to test the null hypothesis of a Portalón constant population size, no signs of population growth were detected in any of the Bronze Age Portalón sample sizes (R_2 value for MNI 12 = 0.1363, $P = 0.3137$ and R_2 value for MNI 19 = 0.1227, $P = 0.3942$).

Population structure

We use AMOVA to investigate the variability at the group level. When testing Portalón samples (MNI = 12 or 19) against modern Iberian, modern non-Iberian and ancient (wild and domestic) samples, the highest differences among populations were found between any Portalón sample set and modern Iberian horses (for MNI = 19, Φ_{ST} 0.22972; $P < 0.00001$ and for MNI = 12, Φ_{ST} 0.2001; $P < 0.00001$, Table S11). These results remain significant after applying sequential Bonferroni corrections.

Phylogenetic analysis

The Bayesian inference provided relationships that ranged from 'supported' to 'unresolved'. The relationships of the haplotypes that constitute the Lusitano group C were moderately supported, with posterior probabilities of 0.80 and 0.93 for the trees indicated in Fig. 4.1 and Fig. S3.1 respectively. Moreover, both trees included ATA03, ATA08 and ATA14 within the Lusitano group C, as also indicated by the nonmodified RMN. In contrast ATA02, ATA04, ATA10, ATA11, ATA15 and CFo02 were not supported as members of the A4, C1 and A6 groups, appearing in branches of the main polytomy generated in both topologies.

Maximum Likelihood phylogenetic inferences indicated different topologies to those of the Bayesian analyses, with the ancient Iberian horses distributed in different clusters through the topologies. In both ML trees, the relationship of the Lusitano group C was supported with low bootstrap values: 38% and 39% bootstrap values for modified and nonmodified RMNs trees respectively (Fig. 4.2 and Fig. S3.2). Moreover, both trees indicated that the sequences ATA03, ATA08 and ATA14 clustered within the Lusitano group C.

Neither the ML nor Bayesian analyses indicated a relationship between the ancient Iberian sequences presented in this study and those published by Seco-Morais *et al.* (2007). Nevertheless, one sample from the Aljubarrota Battle (HOR 23) clustered in the same clade as a member of the Sorraia breed Azambuja lineage (Fig. 5). Although the ML and Bayesian analyses produced different topologies, in both cases the Lusitano group C clustered together, with a bootstrap value of 47 and posterior probability of 0.79 respectively.

Discussion

Previous studies have demonstrated the difficulty of assigning specific modern horse breeds to exclusively concrete haplotypes, and these to geographical areas (Jansen *et al.* 2002). Furthermore, the high matrilineal

diversity already presented in ancient horses from Kazakhstan and China (Keyser-Tracqui et al. 2005; Cai et al. 2009), makes it difficult to determine a clear geographical association (Keyser-Tracqui et al. 2005). The 28 wild ancient horses appear mostly dispersed, clustered with modern horses in four haplotypes, for both RMNs (being in common haplotypes with ATA05 and ATA10).

Given the wide geographical distribution of the ancient wild horses that have been used, one might expect that a large portion of the true variation is represented. If true, and if the domestic horses descended from just one domestication event, then it is reasonable to expect an observation of lower nucleotide diversity among the domestic horses, than in the wild horses, exactly opposite to our findings (Fig. 3C). However, if domestic horse populations were continuously restocked by wild mares taken from a wide geographical area, the domestic population would absorb a large portion of the diversity from the wild populations, and thus a pattern such as that described in Fig. 3C would be expected.

The diversity estimates and nonparametric bootstrap results from the Bronze Age Iberian horses suggest low values of nucleotide diversity (Fig. 3A and B). This could be indicative of a homogeneous population. As the samples span a short timeframe, around 500 years (see Table 1 for chronologies), this may indicate geographical isolation, or a population passing through a bottleneck event.

Independently of the network analysed (Fig. 2A and Fig. S1), 400 modern horses of Iberian origin are distributed with different frequencies among almost 100 haplotypes. Only a limited number of these haplotypes contain exclusively modern horses of Iberian origin. While the Lusitano group C is constituted only by horses of Iberian origin, this is not the case for the A6, A4 and C1 groups. Furthermore, although the Bronze Age Iberian sequences appear in different groups throughout the networks, the majority cluster in the Lusitano group C. In the Bayesian phylogeny, Lusitano group C is a moderately supported cluster, although in the ML phylogeny only low bootstrap support is given. Similar discrepancy has been observed previously (Alfaro et al. 2003; Simmons et al. 2004; Erixon et al. 2003). While nonparametric bootstrapping has been advocated as too conservative in comparison to posterior probabilities (Wilcox et al. 2002), it has also been suggested that posterior probabilities tend to inflate the support values, and they should not be interpreted as support for correctly resolved clades (Simmons et al. 2004).

The signals for population growth in the Bronze Age Portalón horses are not conclusive. While Fu's F_s values reject the null hypothesis of constant population size, neither Tajima's D nor R_2 statistic support this alternative hypothesis. Lack of an expansion signal has also been obtained from some modern domestic horses, where their domestic status is clear (Table S10). We noted that Tajima's D only offers results of population expansion when there is a well-defined star-like pattern. The Tajima's D and R_2 values obtained from the Bronze Age Portalón samples may reflect a relatively constant population size. Furthermore, it should be noted that the Tajima's D results as well as Fu's F_s statistics were performed on subsamples from the network, and should therefore be treated with caution.

The AMOVA analyses show significant differentiation among the Bronze Age and modern horses, as well as with the wild ancient and the ancient domestic horses (Table S11). Although variation due to temporal structure must be considered, there is the same amount of among sample variation between Iberian Bronze Age and wild ancient horses, as well as between Iberian Bronze Age compared with other ancient domesticates from different localities. This is also true for variation within the Iberian Bronze Age horses and modern horses of Iberian origin, where 20–23% is represented by differences between the two groups. Although these two populations may have had Iberian origins, the differences could reflect the selective breeding or a restocking effect occurred during the last 4000 years. Unsurprisingly, when ancient Portalón samples are compared with modern non-Iberian horses, this value is much lower, and most variation is seen within populations. This could be as a result of the diverse origin, and the variation it brings with it, in the non-Iberian sample.

No sequence from the Neolithic or the Bronze Age periods is associated to the D1 group. Even though this is one of the most important groups in modern horses that have an Iberian origin, its star-like structure is consistent with an expansion episode of recent historical times. Indeed, this hypothesis is supported by the fact that the D1 group is absent in the Bronze Age material, first appearing in Iberia in the medieval period (sample ATA07). An independent domestication event in Iberian Peninsula for D1 group can not be ruled out, but the alternative hypothesis of a migration from North Africa also remains open. It could be possible that Lusitano group C has been largely replaced by the D1 group from the Middle Ages onwards. Indeed, a similar pattern for maternal lineage replacement has been reported in European Neolithic pigs (Larson et al. 2007).

No relationships can be inferred between Portalón – Cova Fosca horses and any of the ancient Iberian sequences published by Seco-Morais et al. (2007). The latter sequences therefore may represent more maternal lineages of Iberian origin. However, as previously stated, a migratory origin for the Seco-Morais et al. (2007) Iron Age and medieval sequences cannot be ruled out. Alter-

natively, one can not rule out the possibility of sample misidentification in that dataset, for example direct BLAST comparison of one of the sequences, medieval sample HOR22, yields a best match on *Equus asinus* sequences. This same sequence also is placed basal among horses in Bayesian phylogenetic analyses, whereas ML analyses cluster it among the rest of the horses analysed, although associated with a very long branch.

Lastly, our data do not support the Sorraia horses as a primitive predomestic lineage, in contrast to a previous claim (D'Andrade 1945). The only unique ancient Iberian haplotypes that are closely related to Sorraia sequences are not extremely old, but rather date to the Middle Ages (ATA07 near to Pomba, and HOR 23 near to Azambuja) (Fig. 2B and Fig. 5). However, it is worth noting that today, as a result of the bottleneck they underwent during the 20th century (Luís *et al.* 2002), the only maternal lineages of this breed are constituted by Azambuja, Cigana and Pomba. Thus, the possibility that some of the Portalón haplotypes could represent extinct maternal lineages from Sorraia cannot be completely excluded.

Conclusions

The DNA sequences recovered from Iberian Neolithic and Bronze Age horses indicate that ancient horses from Iberia were incorporated into the domestic stocks. This is supported by the sharing of haplotypes by the Neolithic Cova Fosca, Bronze Age Portalón and modern horses of Iberian origin. The low nucleotide diversity present in Bronze Age Portalón horses could reflect the presence of horses from a geographically restricted population. These results suggest the possibility of an independent domestication episode or the use of Iberian maternal lineages in a restocking process from the wild. Whether such a process qualifies as a conventional domestication event, or simply reflects a recurrent phenomenon of restocking, in which mares were incorporated into domestic herds brought from elsewhere, will require nuclear DNA analysis and a more detailed knowledge of the genetic structure from the earliest Iberian domestic herds. Lastly, relationships have neither been found among D1 group and Neolithic-Bronze Age Iberian horses, nor with the Neolithic-Bronze Age Iberian horses and the Sorraia breed.

References

Alfaro ME, Zoller S, Lutzoni F (2003) Bayes or Bootstrap? A simulation study comparing the performance of Bayesian Markov Chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence *Molecular Biology and Evolution*, **20**, 255–266.

Anderung C, Bouwman A, Persson P *et al.* (2005) Prehistoric contacts over the Straits of Gibraltar indicated by genetic analysis of Iberian Bronze Age cattle. *Proceedings of the National Academy of Sciences of the USA*, **102**, 8431–8435.

Anthony DW, Brown DR (2003) Eneolithic horse rituals and riding in the steppes: new evidence. In: *Prehistoric Steppe Adaptation and the Horse* (ed Levine M, Renfrew C, Boyle K). pp. 55–68, McDonald Institute for Archaeological Research, Cambridge.

Bandelt H-J, Forster P, Sykes BC, Richards MB (1995) Mitochondrial portraits of human populations using Median Networks. *Genetics*, **141**, 743–753.

Bollongino R, Edwards CJ, Alt KW, Burger J, Bradley DG (2006) Early history of European domestic cattle as revealed by ancient DNA. *Biology Letters*, **2**, 155–159.

Cai D, Tang Z, Han L *et al.* (2009) Ancient DNA provides new insights into the origin of the Chinese domestic horse. *Journal of Archaeological Science*, **36**, 835–842.

Carretero JM, Ortega AI, Juez L *et al.* (2008) A Late Pleistocene-Early Holocene archaeological sequence of Portalón de Cueva Mayor (Sierra de Atapuerca, Burgos, Spain). *Munibe (Antropologia-Arkeologia)*, **59**, 67–80.

Castaños PM (2005) Estudio de la fauna de Cueva Mayor de Atapuerca. In: *Estudios Sobre Atapuerca (Burgos): III. Los Materiales del Bronce Final de ''El Portalón'' de Cueva Mayor* (ed. Mínguez MT), PhD thesis, pp. 247–257. Universidad de Deusto and Excma. Diputación Provincial de Burgos, Bilbao and Burgos.

D'Andrade R (1945) O cavalo do Sorraia. *Boletim Pecuário*, **3**, 105–117.

Di Bernardo G, Galderisi U, Del Gaudio S *et al.* (2004) Genetic characterization of Pompeii and Herculaneum Equidae buried by Vesuvius in 79 AD. *Journal of Cellular Physiology*, **199**, 200–205.

Driscoll CA, Menotti-Raymond M, Roca AL *et al.* (2007) The Near Eastern origin of cat domestication. *Science*, **317**, 519–523.

Eisenmann V, Alberdi MT, De Giuli C, Staesche U (1988) *Studying Fossil Horses. Vol I: Methodology*. EJ Brill, Leiden.

Erixon P, Svennblad B, Britton T, Oxelman B (2003) Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. *Systematic Biology*, **52**, 665–673.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of Molecular Variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Excoffier L, Laval G, Schneider S (2005) ARLEQUIN (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.

Fu Y-X (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 915–925.

Gilbert MTP, Bandelt H-J, Hofreiter M, Barnes I (2005) Assessing ancient DNA studies. *Trends in Ecology and Evolution*, **20**, 541–544.

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, **52**, 696–704.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95–98.

Harding AF (2000) *European Societies in the Bronze Age*. Cambridge University Press, Cambridge.

Hofreiter M, Serre D, Poinar HN, Kuch M, Pääbo S (2001) Ancient DNA. *Nature Reviews Genetics*, **2**, 353–359.

Jansen T, Forster P, Levine MA *et al.* (2002) Mitochondrial DNA and the origins of the domestic horse. *Proceedings of the National Academy of Sciences of the USA*, **99**, 10905–10910.

Jung Y-H, Han S-H, Shin T, Oh M-Y (2002) Genetic characterization of horse bone excavated from the Kwakji archaeological site, Jeju, Korea. *Molecules and Cells*, **14**, 224–230.

Keyser-Tracqui C, Blandin-Frappin P, Francfort H-P *et al.* (2005) Mitochondrial DNA analysis of horses recovered from a frozen tomb (Berel site, Kazakhstan, 3rd Century BC). *Animal Genetics*, **36**, 203–209.

Larson G, Dobney K, Albarella U *et al.* (2005) Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science*, **307**, 1618–1621.

Larson G, Albarella U, Dobney K *et al.* (2007) Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *Proceedings of the National Academy of Sciences of the USA*, **104**, 15276–15281.

Levine MA (2005) Domestication and early history of the horse. In: *The Domestic Horse: The Origins, Development and Management of its Behaviour* (ed. Mills DS, McDonnell SM). pp. 5–22, Cambridge University Press, Cambridge.

Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Lindgren G, Backström N, Swinburne J *et al.* (2004) Limited number of patrilines in horse domestication. *Nature Genetics*, **36**, 335–336.

Lister AM, Kadwell M, Kaagan LM *et al.* (1998) Ancient and modern DNA in a study of horse domestication. *Ancient Biomolecules*, **2**, 267–280.

Lopes MS, Mendonça D, Cymbron T *et al.* (2005) The Lusitano horse maternal lineage based on mitochondrial D-loop sequence variation. *Animal Genetics*, **36**, 196–202.

Ludwig A, Pruvost M, Reissmann M *et al.* (2009) Coat color variation at the beginning of horse domestication. *Science*, **324**, 485.

Luís C, Bastos-Silveira C, Cothran EG, Oom MM (2002) Variation in the mitochondrial control region sequence between the two maternal lines of the Sorraia horse breed. *Genetics and Molecular Biology*, **25**, 309–311.

Luís C, Bastos-Silveira C, Costa-Ferreira J, Cothran EG, Oom MM (2006a) A lost Sorraia maternal lineage found in the Lusitano horse breed. *Journal of Animal Breeding and Genetics*, **123**, 399–402.

Luís C, Bastos-Silveira C, Cothran EG, Oom MM (2006b) Iberian origins of New World horse breeds. *Journal of Heredity*, **97**, 107–113.

MacHugh DE, Bradley DG (2001) Livestock genetic origins: Goats buck the trend. *Proceedings of the National Academy of Sciences of the USA*, **98**, 5382–5384.

McGahern AM, Edwards CJ, Bower MA *et al.* (2006) Mitochondrial DNA sequence diversity in extant Irish horse populations and in ancient horses. *Animal Genetics*, **37**, 498–502.

Meyer M, Briggs AW, Maricic T *et al.* (2008) From micrograms to picograms: quantitative PCR reduces the material demands of high-throughput sequencing. *Nucleic Acids Research*, **36**, e5.

Mínguez MT (2005) *Estudios Sobre Atapuerca (Burgos): III. Los Materiales del Bronce Final de ''El Portalón'' de Cueva Mayor*. PhD thesis, Universidad de Deusto and Excma. Diputación Provincial de Burgos, Bilbao and Burgos.

Mirol PM, Peral García P, Vega-Pla JL, Dulout FN (2002) Phylogenetic relationships of Argentinean Creole horses and other South American and Spanish breeds inferred from mitochondrial DNA sequences. *Animal Genetics*, **33**, 356–363.

Olària C (1988) *Cova Fosca. Un Asentamiento Meso-Neolítico de Cazadores y Pastores en la Serranía del Alto Maestrazgo*. Servicio de Publicaciones, Diputación de Castellón, Castellón.

Orlando L, Eisenmann V, Reynier F, Sondaar P, Hänni C (2003) Morphological convergence in *Hippidion* and *Equus* (*Amerhippus*) South American equids elucidated by ancient DNA analysis. *Journal of Molecular Evolution*, **57**, S29–S40.

Outram AK, Stear NA, Bendrey R *et al.* (2009) The earliest horse harnessing and milking. *Science*, **323**, 1332–1335.

Posada D, Crandall KA (1998) MODELTEST: Testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818.

Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, **19**, 2092–2100.

Rice WR (1989) Analyzing tables of statistical tests. *Evolution*, **43**, 223–225.

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.

Royo LJ, Álvarez I, Beja-Pereira A *et al.* (2005) The origins of Iberian horses assessed via mitochondrial DNA. *Journal of Heredity*, **96**, 663–669.

Seco-Morais J, Oom MM, Quesada F, Matheson CD (2007) Ancient Iberian horses: a method to recover DNA from archaeological samples buried under sub-optimal conditions for preservation. *Journal of Archaeological Science*, **34**, 1713–1719.

Simmons MP, Pickett KM, Miya M (2004) How meaningful are Bayesian support values? *Molecular Biology and Evolution*, **21**, 188–199.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

Uerpmann H-P (1990) Die domestikation des pferdes im Chalkolithikum West- und Mitteleuropas. *Madrider Mitteilungen*, **31**, 109–153.

Vilà C, Leonard JA, Götherström A *et al.* (2001) Widespread origins of domestic horse lineages. *Science*, **291**, 474–477.

Weinstock J, Willerslev E, Sher A *et al.* (2005) Evolution, systematics, and phylogeography of Pleistocene horses in the New World: A molecular perspective. *PLoS Biology*, **3**, e241.

Wilcox TP, Zwickl DJ, Heath TA, Hillis DM (2002) Phylogenetic relationships of the dwarf boas and comparison of Bayesian and bootstrap measures of phylogenetic support. *Molecular Phylogenetics and Evolution*, **25**, 361–371.

Willerslev E, Cooper A (2005) Ancient DNA. *Proceedings of the Royal Society B*, **272**, 3–16.

Xu X, Árnason U (1994) The complete mitochondrial DNA sequence of the horse, *Equus caballus*: extensive heteroplasmy of the control region. *Gene*, **148**, 357–362.

Yang DY, Eng B, Waye JS, Dudar JC, Saunders SR (1998) Improved DNA extraction from ancient bones using silica-based spin columns. *American Journal of Physical Anthropology*, **105**, 539–543.

Zeuner FE (1963) *A History of Domesticated Animals*. Hutchinson and Co., London.

J.L. is a PhD student at the Centro Mixto UCM-ISCIII, Madrid. His main research interests concern the horse domestication and the evolutionary biology of different members of the genus *Equus* using ancient DNA techniques. A.L. has a PhD in Archaeometry at the Archaeological Research Laboratory, Stockholm University, and her main interests are ancient DNA and stable isotope analysis on prehistoric populations. C.O. is Full Professor of Prehistory and Director of the Laboratorio de Arqueologia Prehistòrica, Universitat Jaume I, Castelló. Her main interests are the Tardiglacial cultures, the first Neolithic economies, the Late Upper Paleolithic portable art, and the open air art developed through Epipaleolithic to Neolithic. M.B. is a Postdoc at the Department of Forest Mycology and Pathology at the Swedish University of Agricultural Sciences. He works mainly on fungal evolutionary genomics. M.T.P.G. is a molecular biologist with wide ranging interests within ecology, evolutionary biology, forensic genetics, and archaeology. H.E. holds a Chair in Evolutionary Biology at the Evolutionary Biology Centre at Uppsala University. His research focuses on various aspects of evolutionary genomics and on molecular evolution. E.W. is an evolutionary biologist whose research specialises in the study of ancient DNA extracted from substrates as diverse as bone, faeces, soil and ice in order to answer questions of evolutionary, anthropological and archaeological interest. K.L. is Director of the Archaeological Research Laboratory, Stockholm University. Her main interests deal with the Neolithization process using bone chemistry analysis. J.L.A. is Full Professor at the Departamento de Paleontología of the Universidad Complutense de Madrid, and Director of the Centro Mixto UCM-ISCIII, Madrid. His work is focused on paleoanthropology, evolutionary biology and molecular ecology. A.G. is associate professor in Evolutionary Genetics at Uppsala University, and is mainly interested in animal domestication and the Neolithization process.

Supporting information

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

Table S1 Morphological description of the Portalón and Cova Fosca samples

Table S2 Genetic structure of the biotinylated primers

Table S3 Amplification primers

Table S4 Samples analysed for collagen preservation and fragments independently replicated

Table S5 Number of amplification copies of the each fragment used to design the sequences consensus

Table S6 Horse sequences from the GenBank, breed origin and publication

Table S7 Polymorphic sites between the mtDNA D-loop region np. 15487–15740 of Portalón sequences, Cova Fosca sequences, Jansen motifs groups and the Lusitano group C

Table S8 Molecular diversity indices

Table S9 Tajima's D and Fu's F_s tests for the groups used in the study

Table S10 Tajima's D and Fu's F_s tests for different groups of domestic horses from the non-modified RMN

Table S11 AMOVA analysis of Bronze Age Portalón horses (MNI = 19 & 12) against modern Iberian, modern nonIberian, wild Pleistocene-Holocene horses and other ancient domesticates

Table S12 Products from the libraries generated by Roche GS FLX sequencing chemistry

Fig. S1 Non-modified reduced median network.

Fig. S2 Modified reduced median network.

Fig. S3 Phylogenetic trees from the non-modified RMN haplotypes.

Supporting Text 1 Faunal remains collection.

Supporting Text 2 Horses distribution in haplotypes with frequencies >1, from the non-modified RMN.

Supporting Text 3 Horses distribution in haplotypes with frequencies >1, from the modified RMN.