Introduction

The question about the time and the place of horse domestication, a process, which had huge impact on the progress of the civilizations, is disputable. According to the most widely accepted hypothesis, the earliest domestication of the horse has happened in the western parts of the Eurasian steppes, between the Northern Black Sea region and present-day Kazakhstan and Turkmenistan. It seems that it occurred not earlier than the first half and most probably during the middle (even the last third) of the fourth millennium BC (from ~ 5.5 kya; Warmuth et al., 2012; Petersen et al. 2013). The next steps of...
large-scale horse breeding occurred almost simultaneously in Eurasia and North Africa due to the development of the social structure of human society (Vila et al., 2001; Jansen et al., 2002, Lei et al., 2009, Cieslak et al., 2010, Lira et al., 2010). On the other hand, the morphological differences between wild and domestic animals are rather vague and the genetic introgression between them is speculative.

Mitochondrial DNA (mtDNA) studies reveal several possible Eurasian domestication centres, most of them likely being secondary domestication centres. These regions have been proposed mainly because of the high frequencies of specific haplogroups. One of the centres for Western Europe is the Iberian Peninsula, where the dominant haplogroup is L (D; Xi1,2; see Jansen et al., 2002; Cai et al., 2009; Cieslak et al., 2010; Achilli et al., 2012, respectively). For Northern and Central Europe, the the M (C1) haplogroup predominates along with N, B and D haplogroups (Jansen et al., 2002, Cai et al., 2009; Achilli et al., 2012).

Another widely explored domestication centre is in the Turkmen and the Kazakh steppes (Akhal Teke horses), where G, J, Q and A haplogroups prevail (Achilli et al., 2012). An investigation of ancient DNA from Chinese and Mongolian horses has revealed the dominance of A (A) and Q (F) haplogroups (Cai et al., 2009; Achilli et al., 2012). Domestication in the Middle East is characterised by both a high incidence of haplogroups I and O’P, as well as a mixed genetic profile containing all known haplogroups (except for the F-haplogroup, which is characteristic of the Przivalski’s horse only, a wild horse which most probably have not participated in the domestication process: Achilli et al., 2012).

The first study on domestic breeds for the territory of the Balkan Peninsula and for Bulgaria in particular was conducted by Hristov et al. (2017a). It includes results for the mitochondrial profile in three local populations of mountain horses from Stara Planina Mts., the Rhodopes and the Rila-Pirin massif. Although these horses are known trivially under the common name of Karakachan breed, they have been differentiated on the basis of phenotypic traits as three different breeds (Petrov, 1941, Barzev et al., 2005). Petrov considers as typical representatives of modern Bulgarian primitive horses three mountain breeds, Rilo-Pirinski, Staroplaninski and Karakachanski (Rhodopes), as well as two planar breeds, Deliromanski (Ludogorski, East Bulgarian) and Kamchiiski horse.

Molecular analysis of mtDNA of Bulgarian autochthonous mountain horses shows a varied genetic profile covering almost all known haplogroups (Hristov et al., 2017a). The haplotypes of European origin (J, M, N and L) are predominant. A similarity was found in the Staroplaninski and Karakachanski horses due to the dominance of European haplogroups, although with a specific mitochondrial profile for Central and Western European populations (a significantly higher incidence of L-20 (20 %), J, M and N). This type of profile reveals the ancient and native background for both breeds. Unusual results are found with regard to the genetic profile of the Rilo-Pirinski horse compared to the other two populations. A high incidence of A, G, Q and C haplogroups has been established in this local domestic horse, which could be an evidence of genetic introgression of local populations with East Asian populations of the domestic horse (Akhal Teke, Kazakhstan, etc.).

One of the earliest centres of horse domestication could be the North-Eastern (Pontic) part of the Balkan Peninsula, and most likely the eastern Black Sea coast, according to the data from the sunken Early Bronze Age village of Urdoviza (Spassov & Iliev, 1997, see Spassov and others in prep.). This has provoked interest in exploring the DNA data from wild horses in the Balkan region.

This paper aims to study the genetic diversity of wild horses (before their domestication) and to make a comparative analysis with ancient and modern populations.

**Materials and methods**

1. **Archeological samples**

Four samples (bones and dental material) of wild horses from three archaeological sites: Ohoden/Valoga (Early Neolithic), Gradeshnitsa/Malo Pole (Early Neolithic) and Devetak/Karnobat (Early Bronze Age; Figs. 1 and 2) from the collection of NMNH-BAS were studied. The morphological evaluation of the material (coll. of the NMNHS-BAS) of the bone remains was performed (N.S., N.I. and L.H.) on the basis of morphological and morphometric criteria. The remains from Ohoden were insufficient for taxonomic interpretation. The metacarpal fragment of a wild horse from the early Neolithic of Gradeshnitsa -Malo Pole was determined as E. germanicus (a Late Pleistocene horse, survived in Bulgaria till the Holocene) by comparative analysis in previous works (Spassov & Iliev 1997; 1998). The metatarsus of a most probably wild horse from the Early Bronze Age of Devetak had the following metric data: total length – 273.8 mm; width of the diaphysis – 31.24; distal width – 46.47 and with an index
Comparative genetic analysis of subfossil wild horses (from the Neolithic age and Early Bronze age) and present-day...

of Brauner – 11.41. It was determined by comparison (this study) with E. refus using the morphometric criteria (Brauner, 1916; Gromova, 1949).

Ancient DNA isolation
The methodology includes: isolation of ancient DNA (aDNA) from bone remains, amplification of mitochondrial DNA fragments, sequencing and genetic analysis. Ancient DNA was isolated from bone and dental remains from 4 samples dating from the Early Neolithic (6th millennium BC) to the early Bronze Age (3rd millennium BC; Table 1).

The genetic material was isolated according to the protocol of Yang et al. (1998) with minor modifications (Hristov et al., 2017b). To prevent contamination of the bone surface by foreign DNA, the samples were treated sequentially with sodium hypochlo-
rite (40%), 2% hydrochloric acid and washed with ultra-clear H2O several times. After drying for 24-48 h in UV-irradiation medium and constant air filtration, the surface layer was removed, and then a bone powder was obtained, which was further homogenized in metal mortars. Treatment of bone material for isolation of the aDNA was performed from 400-500 mg of bone powder dissolved in 5 ml of lysis buffer (0.5M EDTA, 2% Sodium dodecyl sulfate, 0.1M Tris pH8, 10 μl/ml Mercaptoethanol, 20 μl/ml Proteinase K). The samples were incubated in a hybridizer (Hybridiser HB-2D, Techne, UK) at constant rotation at 55°C for 36-48 h. The samples were centrifuged at 5000 rpm for 1 h, after which the supernatant was filtered with 0.45 μm filters and transferred to 50 ml tubes. DNA isolation was performed using silicone membrane technology, including the use of DNA isolation columns (GeneMatrix, E3520, EURx, Poland) and 5M GuSCN (Sigma-Aldrich) binding reagent V/V. The aDNA bound to the silica columns was purified twice with 70% ethanol wash solution and dissolved in ultrapure water. Isolated aDNA was stored at -20°C.

2. PCR amplification and sequencing
Five different overlapping regions were used to amplify the D-loop region (mitochondrial DNA, HVRI). Primers for amplification are listed in Table 2.

All PCR reactions were performed with 10ng/μl DNA in a final volume of 50 μl (NZYTaq Colourless Master Mix, Cat No – MB040, NZYTech, Portugal). They were performed under the following conditions: Initial denaturation at 94°C for 5 min.; 40 cycles of denaturation at 94°C for 30 seconds, hybridization – 50°C for 30 seconds, elongation at 72°C for 1 minute; end final elongation at 72°C for 10 min. The amplified fragments were separated and visualized on 2% agarose gel electrophoresis.

The successfully amplified products were purified with a PCR purification kit (Gene Matrix, PCR clean-up kit, EURx, Poland) and sequenced in both directions using a PlateSeq kit (Eurofins Genomics Ebersberg, Germany).

3. Statistical processing and analysis of sequence results
All obtained DNA sequences were manually edited and aligned with the MEGA7 program (Kumar et al., 2016) using the horse reference DNA sequence X79547 (Xiufeng & Arnason, 1994). The obtained sequences (about 650 bp) were deposited in the GenBank database National Biotechnology Information Center (NCBI) under accession numbers MG420991 – MG420994.
The sequences were analyzed using their polymorphic positions and the haplogroups were determined according to the nomenclature of Achilli et al. (2012). The obtained sequences were compared to other populations: Carpathian pony (eU093063 – eU093045), Akhal-teke (DQ327950 – DQ327967) and Bulgarian primitive mountain horses (KU601744 – KU601624) (see McGahern et al., 2006; Priskin et al., 2017a, respectively) and mitochondrion – PopSet Acc. No. 74725290 (Achilli et al., 2012) for comparing the genetic profile with the known border populations. Unpublished data were also used in interpreting the results – 57 sequences from the Danube horse breed (MG420898 – MG420955).

This modern breed is representative of lowland horses in Bulgaria created based on mares from the Eastern Bulgarian breed.

**4. Protocol for working with aDNA**

All experiments were performed according to standard precautions in specialized and territorially distinct laboratories for aDNA work: bone material processing, DNA isolation and PCR amplification (Paabo et al., 2004; Willerslev & Cooper, 2005). Briefly, this includes establishing independent laboratories (premises and buildings) for working with aDNA, treating surfaces and solutions with UV radiation (45 W, 72 h), heat treatment (over 180°C, 12 h), acid treatment (2.5M HCl, 48 h) and/or sodium hypochlorite (40%, 48 h), washing with ultrapure water and air filtration in the premises. The dis-

### Table 1. Studied archaeological samples.

<table>
<thead>
<tr>
<th>Periodization</th>
<th>Settlement</th>
<th>Abbreviation</th>
<th>Skeletal element</th>
<th>Product size, bp</th>
<th>Haplogroup</th>
<th>GenBank Acc. Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Early Neolithic</td>
<td>Ohoden/Valoga</td>
<td>Och/Va1</td>
<td>phalanx II</td>
<td>641</td>
<td>G1 MG420992</td>
</tr>
<tr>
<td>2</td>
<td>Early Neolithic</td>
<td>Ohoden/Valoga</td>
<td>Och/Va2</td>
<td>dens</td>
<td>658</td>
<td>G1 MG420993</td>
</tr>
<tr>
<td>3</td>
<td>Early Neolithic</td>
<td>Gradeshnitsa-Malo Pole</td>
<td>G/MP1</td>
<td>metacarpus</td>
<td>257</td>
<td>Q2’3? MG420994</td>
</tr>
<tr>
<td>4</td>
<td>Early Bronze</td>
<td>Devetak/Karnobat</td>
<td>Dev1</td>
<td>metatarsus</td>
<td>651</td>
<td>G1 MG420991</td>
</tr>
</tbody>
</table>

### Table 2. Primers used for amplification of the D-loop region of mtDNA. The position of the primers is relative to the reference sequence NC_001640 (Xiufeng & Arnason, 1994).

<table>
<thead>
<tr>
<th>Region</th>
<th>Primer sense/antisense</th>
<th>Sequence 5’-3’</th>
<th>Reference</th>
<th>Product size, bp</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.</td>
<td>f-15450</td>
<td>cacccaaagtgaaattctac</td>
<td>Hristov et al., 2017a</td>
<td>114</td>
</tr>
<tr>
<td></td>
<td>Ec1_r_15564</td>
<td>gacctagttagccattataaga</td>
<td>Elsner et al., 2016</td>
<td></td>
</tr>
<tr>
<td>II.</td>
<td>f-15450</td>
<td>cacccaaagtgaaattctac</td>
<td>Hristov et al., 2017a</td>
<td>220</td>
</tr>
<tr>
<td></td>
<td>Ec4_r_15670</td>
<td>gaccttaggaggttagcacc</td>
<td>Elsner et al., 2016</td>
<td></td>
</tr>
<tr>
<td>III.</td>
<td>f-15450</td>
<td>cacccaaagtgaaattctac</td>
<td>Hristov et al., 2017a</td>
<td>210</td>
</tr>
<tr>
<td></td>
<td>Ec2_r_15660</td>
<td>gatggggtatgcacgatcaaat</td>
<td>Elsner et al., 2016</td>
<td></td>
</tr>
<tr>
<td>IV.</td>
<td>Ec4_r_15590</td>
<td>gaatggcctatgtcgtg</td>
<td>Elsner et al., 2016</td>
<td>190</td>
</tr>
<tr>
<td></td>
<td>Ec5_r_15760</td>
<td>tagttggggagtgtgctg</td>
<td>Elsner et al., 2016</td>
<td></td>
</tr>
<tr>
<td>V.</td>
<td>Ec4_r_15590</td>
<td>gaatggcctatgtcgtg</td>
<td>Elsner et al., 2016</td>
<td>488</td>
</tr>
<tr>
<td></td>
<td>r-16078</td>
<td>ataacacctagtggctg</td>
<td>Hristov et al., 2017a</td>
<td></td>
</tr>
</tbody>
</table>

This modern breed is representative of lowland horses in Bulgaria created based on mares from the Eastern Bulgarian breed.
Comparative genetic analysis of subfossil wild horses (from the Neolithic age and Early Bronze age) and present-day... 7

Results and discussion

Ancient DNA was successfully amplified from all the tested samples. In the interpretation of the results after analysis of the sequenced fragments/regions for assay, combined overlapping sequences of about 650 bp were used. The established polymorphic positions are presented in Table 3 by comparison with reference sequences X79547 (XUFENG & ARNASON, 1994) and JN398377 (ACHILLI et al., 2012). Mitochondrial haplogroups were determined according to the latest classification of ACHILLI et al. (2012).

Three haplotypes belonging to two haplogroups, Q2’3 and G1, were identified. The dominant haplogroup was G1 (3/4, 75%) presented with two different haplotypes: Och / Va1-2 and Dev1 (transversion at polymorphic position 16072C / A; Table 3).

Haplogroup G is found in both modern and ancient horses. The results of the studies showed that the prevalence of haplogroup G was the highest in the Middle Asian horse breeds – over 16%.

Table 3. Haplogroups and polymorphic sites of the studied samples. (Reference sequence JN398377; Achilli et al, 2012)

<table>
<thead>
<tr>
<th>No</th>
<th>Abbreviation</th>
<th>Haplogroup</th>
<th>Polymorphic positions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Och/ Va1</td>
<td>G1</td>
<td>15492,15539,15582,15594,15599,15632,15647,15663,15700,15717,15867,16028</td>
</tr>
<tr>
<td>2</td>
<td>Och/ Va2</td>
<td>G1</td>
<td>15492,15539,15582,15594,15599,15632,15647,15663,15700,15717,15867,16028</td>
</tr>
<tr>
<td>3</td>
<td>G/MP1</td>
<td>Q2’3</td>
<td>15492, 15582, 16028</td>
</tr>
<tr>
<td>4</td>
<td>Dev1</td>
<td>G1</td>
<td>15492,15539,15582,15594,15599,15632,15647,15663,15700,15717,15867,16028, 16072</td>
</tr>
</tbody>
</table>
Gyulnas Dzhebir, Georgi Yordanov, Iskra Yankova, Daniela Sirakova, Maria Petrova, Boyko Neov, Peter Hristov, Georgi Radoslavov, Latinka Hristova, Nikolai Spassov

(Kazakhstan and Turkmenistan; Fig. 3). For Central and Western Europe this frequency was about two times lower. The subhaplogroup G1 is believed to be typical of the Middle Asian breed Akhal Teke from Kazakhstan (Jansen et al., 2002; Cai et al., 2009; Cieslak et al., 2010; Achilli et al., 2012). This is logical because this region is considered to be the main domestication centre and is a representative of the steppe regions. These studies, though large-scale, have a drawback in terms of “white spots” – the genetic diversity of key geographic regions of Europe, such as the Balkan Peninsula and the Northern Black Sea coast.

The first survey of populations of Bulgarian horses was only published in 2017. This study characterized the mountain type of local horses from Staro Planina Mts., the Rhodopes and the Rila-Pirin massif or over 121 samples in total (Fig. 2). The results showed a typical European mitochondrial profile in the Staroplaninski and Rhodope horses with unusually high frequencies of the European-specific groups M and D (2 to 3 times). The mitochondrial profile of the Rila-Pirin horses differed significantly from the aforementioned with high frequencies of typical Central and Eastern Asian groups, i.e. haplogroups A, G, Q, C. In this study the authors commented on this unusual genetic profile as a massive mix of Asian local horses (genetic drift or introgression).

A similar genetic drift of populations of Central Asian to Central European horses was also found for the Carpathian Mountain Horse, Hucul, and the planar Hungarian and Polish primitive horses (Czernekova, 2013; Cieslak et al., 2010; Priskin et al., 2017). This migration is considered as a result of the “migration of the peoples” and especially the Hungarian settlements during the first millennium of the new era.

Summarizing the data on modern local Bulgarian horses, several conclusions can be made: 1) they show a heterogeneous genetic profile following their geographical localization; 2) although insufficient, the data show a significant difference between the planar and the mountain types of horses (Fig. 2); 3) the so-called Asian G1 group has a high frequency and is even a dominant haplogroup in the mountain (Rilo-Pirinski) and the planar Danube horses.

The ancient DNA data received from the subfossil bones of wild horses from Bulgaria represent a special interest: they are the first such data for the extinct wild horses of the subgenus Equus (Equus) which has been spread in the forest-steppe and steppe regions from Eastern Europe most probably till Kazakhstan in the Early and Middle Holocene. Two species of wild horses, rather different after dentition and limb proportions, existed in the Late Pleistocene of Europe and survived after paleozoological data in Eastern Europe (including Bulgarian): the massive Equus germanicus (= E. latipes) and the slender and more adapted to steppe conditions E. ferus (= E. gmelini) (the so-called Tarpan). This statement, supported by a large number of data was expressed in the papers of Spassov and Iliev (1997; 1998), remained insufficiently known to the science community. The presence of two different wild species in the probable area of the earliest domestication make the question for the wild progenitor of the do-
domestic horse (*Equus caballus*) more complicate. The existence of two different wild horse species in the Holocene of Eastern Europe, which both have been used in the process of the domestication, is supported by the presence of the haplogroup Q in the massive wild horse from the Neolithic of Gradeshnitsa-Malo Pole, determined as *E. germanicus* (Spassov & Iliev 1997), after morphological criteria, and the presence of another haplogroup (G) in the slender wild horse from the Early Bronze Age of Devetak.

In Central Asian (Kazakhstan, Turkmenistan) horses, which are ancient by origin, there is a predominance of the G and Q haplogroups, which we find in prehistoric wild horses from Bulgaria. As these Middle Asian breeds have been grown since ancient times in the region, according to today’s research, especially close to the most probable primary horses’ domestication centre, it could be assumed again that these haplogroups were typical of the wild horses of the Eurasian steppes and forest steppes from the Eastern Black Sea coast to the Kazakh and Iranian open spaces (here we exclude the more eastern species – the Mongolian wild horse, which did not participate in terms of genetic data in the origin of the domestic horse).

Judging by the presence of the haplogroup Q in the massive wild horse *E. germanicus*, from the Neolithic of Gradeshnica, we can assume that it has been involved in the domestication of ancient domestic horses. Haplogroup G could be the typical haplogroup for the slender wild horse of the steppes (*E. ferus*) determined in the Early Bronze Age of Devetak and well presented in the old domestic breeds in the Middle East. Further investigations and a larger sample (number of samples) are needed in order to confirm the hypotheses very hesitantly suggested here. It is not excluded that a part of the territory of Bulgaria (the Eastern Black Sea coast, judging by genetic data and the presence of the domestic horse from the Early Bronze Age of Urdoviza), represents (with the North Black Sea coast) a part of the primary centre of the domestication of the horse.

**Conclusion**

The current study presents the first data on the genetic diversity of the Holocene wild horses of the subgenus *Equus* (*Equus*) from the early Neolithic and the Early Bronze Age in Bulgaria, and with this the first data on the genetic diversity of the extinct Holocene wild horses of this subgenus in Europe. The results show the presence of the Q (in *Equus germanicus* from the Early Neolithic of NW Bulgaria) and G (G1) (in *E. ferus* from the E. Neolithic and the E. Bronze Age of Bulgaria) haplogroups. Both G and Q haplogroups have so far been spread with high frequency in the Middle Asian horse domestic breeds. This preliminary result gives ground to support the statement of the survival of two different Late Pleistocene wild horses in the Holocene of Eastern Europe and with this to support the polyphyletic hypothesis for the origin of the domestic horse in the western part of the Eurasian steppes. Our results give some clarity about the further dispersal of the domestic horse and about its strong genetic polymorphism. The data, though preliminary, can provide interesting information about the wild ancestors and the origin of domestic horses.

Our findings would contribute also to elucidating the origin and migration processes in the formation of local horse breeds. This information is directly related to the understanding of migration and the cultural-historical processes in our region.

The established genetic profile of subfossil wild horse (*E. ferus* – the Tarpan) is the closest to the profile of the Danube horse, where the frequency of the haplogroup G is about 50%.

**Acknowledgements:** This study was financially supported by the project “Thracians – Genesis, Ethnic Development and Cultural Identities” of the Bulgarian Academy of Sciences, and was carried out in the frames of the subproject of the NMHS: “The wild and domestic animals from Prehistoric time till the Thracian epoch.” The team is grateful for the recent local horses’ samples provided by the Executive Director of the National Horse Breeding Association Georgi Yordanov.

**References**


Gyulnas Dzhebir, Georgi Yordanov, Iskra Yankova, Daniela Sirakova, Maria Petrova, Boyko Nen, Peter Hristov, Georgi Radoslavov, Latinka Hristova, Nikolai Spassov

chaeological Science, 36(3), 835-842.


