SHORT COMMUNICATION doi:10.1111/j.1365-2052.2006.01495.x

Evidence for biogeographic patterning of mitochondrial DNA sequences in Eastern horse populations

A. McGahern*, M. A. M. Bower†, C. J. Edwards‡, P. O. Brophy*, G. Sulimova§, I. Zakharov§, M. Vizuete-Forster†, M. Levine†, S. Li¶, D. E. MacHugh* and E. W. Hill*

*Animal Genomics Laboratory, School of Agriculture, Food Science and Veterinary Medicine, College of Life Sciences, University College Dublin, Belfield, Dublin 4, Ireland. †McDonald Institute for Archaeological Research, Downing Street, Cambridge, CB2 3ER, UK. ‡Smurfit Institute of Genetics, Trinity College, Dublin 2, Ireland. §Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow V-333, 117809, Russia. ¶Department of Archaeology, Peking University, Beijing, 100871, China

Summary

Equine mitochondrial DNA (mtDNA) phylogeny reconstruction reveals a complex pattern of variation unlike that seen in other large domesticates. It is likely that this pattern reflects a process of multiple and repeated, although not necessarily independent, domestication events. Until now, no clear geographic affiliation of clades has been apparent. In this study, AMOVA analyses have revealed a significant non-random distribution of the diversity among equine populations when seven newly sequenced Eurasian populations were examined in the context of previously published sequences. The association of Eastern mtDNA types in haplogroup F was highly significant using Fisher's exact test of independence ($P = 0.00000$). For the first time, clear biogeographic partitioning has been detected in equine mtDNA sequence.

Keywords genetic diversity, horse domestication, mitochondrial DNA, phylogeography, polymorphism.

There is widespread debate surrounding the 'when and where' of horse domestication. However, it is widely believed that the Eurasian Steppe, with its vast overland communication routes, was an important centre of domestication (Levine 1999; Anthony & Brown 2003; Benecke & von den Driesch 2003; Kuz'mina 2003; Olsen 2003). A recent report suggests that the Iberian Peninsula may have served as an important glacial refugium for the horse and that this region may also have been the site of a localized domestication (Lopes *et al.* 2005).

Mitochondrial DNA (mtDNA) diversity analyses of most domestic animals reveals a common motif: two major clusters of sequences geographically distributed between east and west (MacHugh & Bradley 2001). This pattern is indicative of independent domestication events mainly centred in Middle/Near East and East Asia, although Africa may also have been an important region for cattle domestication (MacHugh *et al.* 1997; Troy *et al.* 2001; Hanotte *et al.* 2002; Bruford *et al.* 2003). Equine mtDNA phylogeny reconstruction provides an interesting exception to the rule with a distinctive star-like branching structure (McGahern *et al.* 2006; Jansen *et al.* 2002). Existing genetic data are consistent with the capture and exploitation of wild mares having taken place in multiple locations, possibly over a broad time span (Lister 2001; Vila *et al.* 2001; Hill *et al.* 2002; Jansen *et al.* 2002). Consequently horse genetic diversity is complex and currently lacks the biogeographical patterning associated with other large domesticated animals. The high mobility of horses and their use by humans throughout prehistory up to the recent past for spanning great distances, has led to an obscuring of the genetic structure within the species through post-domestication movements. This is particularly apparent among European horses, where modern breeding practices, for example, have led to considerable mixing of populations. However, horses from more isolated populations may show a clearer biogeographical signature.

The analysis of 118 additional sequences from seven previously unsampled horse populations from central and north-eastern Eurasia and China [Akhal Teke ($n = 18$; DQ327950–DQ327967); Guan Mountain ($n = 10$; DQ327839–DQ327847); Mesenskaya ($n = 18$; DQ327968–DQ327985); Mongolian ($n = 16$; DQ327986–DQ328001); Orlov ($n = 18$; DQ328002–DQ328019);

Address for correspondence

E. W. Hill, Animal Genomics Laboratory, School of Agriculture, Food Science and Veterinary Medicine, College of Life Sciences, University College Dublin, Belfield, Dublin 4, Ireland.
E-mail: emmeline.hill@ucd.ie

Accepted for publication 25 May 2006

Vyatskaya ($n = 18$;DQ328020–DQ328037); Yakut ($n = 20$; DQ328038–DQ328057)] has for the first time revealed a significant biogeographical association in equine mtDNA sequences. A 247-bp equine mtDNA control region sequence between positions 15 494–15 740 of reference sequence X79547 (Xu & Arnason 1994) was examined in 962 horses (118 new sequences reported here, 98 sequences reported in McGahern *et al.* (2006) and 746 existing GenBank accessions that are listed in Table S1) sampled from 72 populations from three broad geographic regions (Europe, EUR; Middle East and Africa, MEA; Far East, FE). This collection represents the largest and most comprehensive study of horse matriarchal phylogeography to date. mtDNA haplotypes were grouped into one of seven major haplogroups as described by Jansen *et al.* (2002). The frequency of each haplogroup in the two main geographic regions (EUR and MEA/FE) and the geographic sampling locations and haplogroup assignments of individuals from the seven newly sequenced populations are illustrated in Fig. 1.

A skeleton median-joining network was drawn to determine the correct relationships of the seven main haplogroups, A–G (Fig. 2a), and this highlighted two previously unrecorded haplogroups A7 and F3. In addition, networks were constructed from MEA/FE sequences, including the newly sequenced populations (Fig. 2b), and for EUR sequences (Fig. 2c). In each case the majority of sequences were found in haplogroups A and D. In the European network, sequences were distributed throughout with no apparent affiliation to any one group, although haplogroup D was the most common (35%). Interestingly, haplogroup F was found more frequently in Eastern populations (16.8% FE; 16.5% MEA) than in European populations (9.7%), while a relatively small proportion of Eastern sequences (15.5% FE; 26.9% MEA) were found in haplogroup D. Furthermore, a number of sequences within haplogroup F have been found for the first time in these newly sequenced Eastern populations. The higher incidence of haplogroup F and the relatively fewer occurrences of haplogroup D sequences in Eastern populations suggest a degree of genetic isolation and differentiation from European horse populations.

An analysis of molecular variance (AMOVA) was performed to investigate the partitioning of mtDNA molecular variance among these geographic groups using the Arlequin program (Schneider *et al.* 2000). Populations were grouped into an Eastern cohort (Arabia, China, Egypt, Iran, Japan, Korea and Russia) and a Western cohort (Austria, Belgium, Denmark, Germany, Iceland, Ireland, Netherlands, Norway and UK), revealing a degree of geographic organization of haplotypes. Although only explaining a low level of genetic variation (2.71%), this geographic organization was highly significant ($P = 0.00782$), suggesting that Eastern horse populations demonstrate some geographic clustering to the exclusion of Western horse populations. The majority of sequence diversity was found within populations (90.18%, $P = 0.00000$), as shown within the AMOVA analysis.

To test the concordance of haplogroup F with Eastern populations, a Fisher's exact test of independence (Sokal & Rohlf 1995) was performed in a 2 × 2 contingency table featuring MEA/FE populations and EUR populations with counts of haplogroup F vs. non-haplogroup F sequences. Haplogroup F concordance with MEA/FE population origin was highly significant ($P = 0.00000$). Additionally, a test of independence for haplogroup D sequences (the most common globally) in MEA/FE and EUR populations was also highly significant ($P = 0.000013$).

The results from this survey of equine mtDNA phylogeographic diversity show for the first time a clear biogeo-

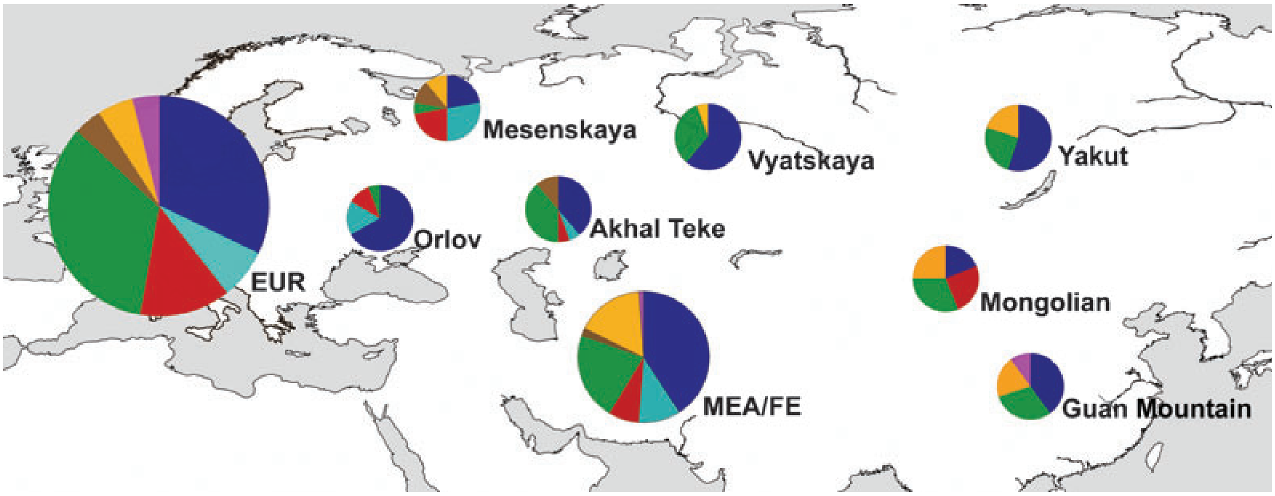

Figure 1 Mitochondrial DNA haplogroup distributions in Eurasian horse populations (EUR, $n = 619$; MEA/FE, $n = 343$). Geographical locations for newly sampled populations are denoted by the position of the individual pie charts. The size of each circle is proportional to the number of sequences sampled from that region. Haplogroups are coloured as follows: dark blue = A, light blue = B, red = C, green = D, brown = E, orange = F and pink = G.

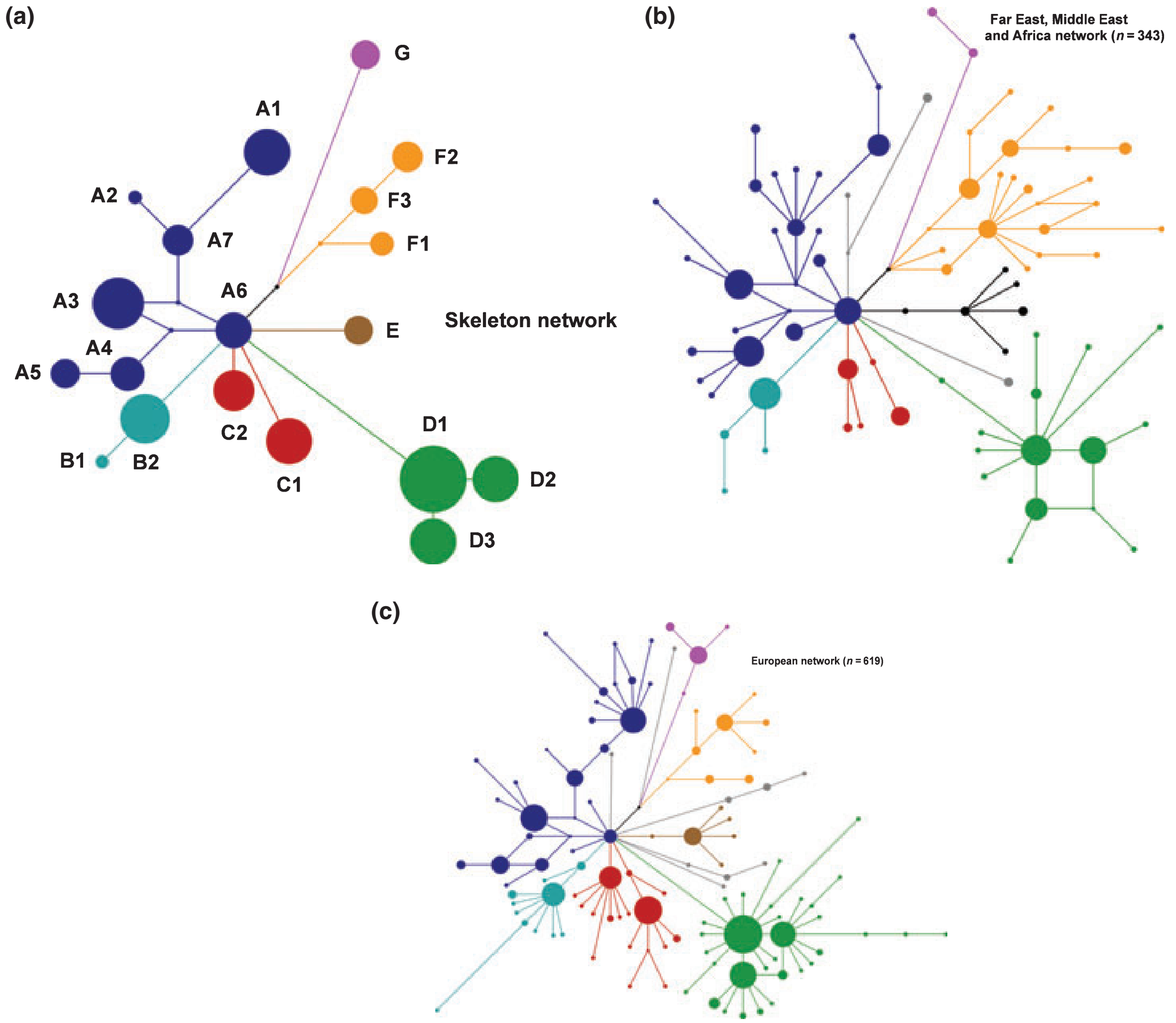

Figure 2 Reduced median phylogenetic networks constructed for horse mitochondrial DNA sequences using 247 bp of the control region. (a) Relationships of the seven main haplogroups, A–G, showing the new, previously unrecorded haplogroups A7 and F3. Two regional haplotype groups are represented including (b) the Middle East Africa and the Far East (MEA/FE $n = 343$) and (c) Europe (EUR, $n = 619$). The spatial arrangement of the skeleton network and the colour codes are preserved in the full data networks. The colour indicates which of the seven haplogroups each sample roots to with dark blue = A, light blue = B, red = C, green = D, brown = E, orange = F and pink = G. Circles represent sequence haplotypes, the area being proportional to the frequency of the haplotype. Points are theoretical intermediate nodes introduced by the median algorithm and branches between haplotypes represent mutations.

graphical pattern within maternal horse lineages and indicate that comprehensive population sampling focused on Eastern horse populations could uncover genetic legacies of the domestication process that have not been obscured by later mobility. Consequently, multidisciplinary investigations encompassing genetics, archaeogenetics and archaeology should enable discrimination between the genetic effects of diffuse domestication and the lack of population structure caused by subsequent mobility, thereby facilitating meaningful archaeological and anthropological inference.

Acknowledgements

The authors thank the Irish Department of Agriculture and Food. We are grateful to Prof. Jiao Nanfeng and Mr Tian Yaqi (Shaanxi Institute of Archaeology), as well as Maša Amatt (Cambridge University), for help sampling the Chinese horse population and to Prof. D. G. Bradley for helpful comments. C.J.E. is supported by Irish Research Council for Science Engineering and Technology Basic Research Grant Scheme (project numbers SC/1999/409 and SC/2002/510). M.A.B, M.L. and M.V.F. are supported by AHRC, the McDonald Institute for Archaeological Research and the Isaac Newton Trust. E.W.H. is supported by a Science Foundation Ireland, President of Ireland Young Researcher Award.

References

Anthony D.W. & Brown D.R. (2003) Eneolithic horse rituals and riding in the steppes: new evidence. In: *Prehistoric Steppe*

Adaptation and the Horse (Ed. by M. Levine, C. Renfrew & K. Boyle), pp. 55–68. McDonald Institute for Archaeological Research, Cambridge, UK.

Benecke N. & von den Driesch A. (2003) Horse exploitation in the Kazakh steppes during the Eneolithic and Bronze Age. In: *Prehistoric Steppe Adaptation and the Horse* (Ed. by M. Levine, C. Renfrew & K. Boyle), pp. 69–82. McDonald Institute for Archaeological Research, Cambridge, UK.

Bruford M.W., Bradley D.G. & Luikart G. (2003) DNA markers reveal the complexity of livestock domestication. *Nature Reviews Genetics* **4,** 900–10.

Hanotte O., Bradley, D.G., Ochieng J.W., Verjee Y., Hill E.W. & Rege J.E. (2002) African pastoralism: genetic imprints of origins and migrations. *Science* **296,** 336–9.

Hill E.W., Bradley D.G., Al-Barody M., Ertugrul O., Splan R.K., Zakharov I. & Cunningham E.P. (2002) History and integrity of thoroughbred dam lines revealed in equine mtDNA variation. *Animal Genetics* **33,** 287–94.

Jansen T., Forster P., Levine M.A., Oelke H., Hurles M., Renfrew C., Weber J. & Olek K. (2002) Mitochondrial DNA and the origins of the domestic horse. *Proceedings of the National Academy of Sciences of the United States of America* **99,** 10905–10.

Kuz'mina E.E. (2003) Origins of pastoralism in the Eurasian steppes. In: *Prehistoric Steppe Adaptation and the Horse* (Ed. by M. Levine, C. Renfrew & K. Boyle), pp. 203–32. McDonald Institute for Archaeological Research, Cambridge, UK.

Levine M.A. (1999) The origins of horse husbandry on the Eurasian Steppe. In: *Late Prehistoric Exploitation of the Eurasian Steppe* (Ed. by M. Levine, Y. Rassamakin, A. Kisleno & N. Tatarintseva), pp. 5–59. McDonald Institute, Cambridge, UK.

Lister A.M. (2001) Tales from the DNA of domestic horses. *Science* **292,** 218–9.

Lopes M.S., Mendonca D., Cymbron T., Valera M., da Costa-Ferreira J. & Machado Ada C. (2005) The Lusitano horse maternal lineage based on mitochondrial D-loop sequence variation. *Animal Genetics* **36,** 196–202.

MacHugh D.E. & Bradley D.G. (2001) Livestock genetic origins: goats buck the trend. *Proceedings of the National Academy of Sciences of the United States of America* **98,** 5382–4.

MacHugh D.E., Shriver M.D., Loftus R.T., Cunningham P. & Bradley D.G. (1997) Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (*Bos taurus* and *Bos indicus*). *Genetics* **146,** 1071–86.

McGahern A.M., Edwards C.J., Bower M.A., Heffernan A., Park S.D.E., Brophy P.O., Bradley D.G., MacHugh D.E. & Hill E.W. (2006) Mitochondrial DNA sequence diversity in extant Irish horse populations and in ancient horses. *Animal Genetics*, doi: 10.1111/j.1365-2052.2006.01506.x

Olsen S.L. (2003) The exploitation of horses at Botai Kazakhstan. In: *Prehistoric Steppe Adaptation and the Horse* (Ed. by M. Levine, C. Renfrew & K. Boyle), pp. 83–103. McDonald Institute for Archaeological Research, Cambridge, UK.

Schneider S., Roessli D. & Excoffier L. (2000) *Arlequin: A Software for Population Genetics Data Analysis.* Genetics and Biometry Lab, Department of Anthropology, University of Geneva, Geneva, Switzerland.

Sokal R.R. & Rohlf F.J. (1995) *Biometry, The Principles and Practice of Statistics in Biological Research.* Freeman, New York.

Troy C.S., MacHugh D.E., Bailey J.F., Magee D.A., Loftus R.T., Cunningham P., Chamberlain A.T., Sykes B.C. & Bradley D.G. (2001) Genetic evidence for Near-Eastern origins of European cattle. *Nature* **410,** 1088–91.

Vila C., Leonard J.A., Gotherstrom A., Marklund S., Sandberg K., Liden K., Wayne R.K. & Ellegren H. (2001) Widespread origins of domestic horse lineages. *Science* **291,** 474–7.

Xu X. & Arnason U. (1994). The complete mitochondrial DNA sequence of the horse, *Equus caballus*: extensive heteroplasmy of the control region. *Gene* **148,** 357–62.

Supplementary Material

The following supplementary material is available for this article online from http://www.blackwell-synergy.com/doi/full/10.1111/j.1365–2052.2006.01495.x:

Table S1 GenBank accession numbers of horse mitochondrial DNA sequences used in the study.