Genetic Characterization of Cleveland Bay Horse Breed

Anas M. Khanshour [1], Eleanore K. Hempsey [2], Rytis Juras [1] and E. Gus Cothran [1,*]

[1]   Department of Veterinary Integrative Biosciences, College of Veterinary Medicine and Biomedical Science, Texas A&M University College Station, TX 77843, USA; Anas.khanshour@gmail.com (A.M.K.); RJuras@cvm.tamu.edu (R.J.)

[2]   Faculty in Biology, Northland Pioneer College, Winslow, AZ 86047, USA; eleanore.hempsey@npc.edu

[*]   Correspondence: gcothran@cvm.tamu.edu

Abstract: The Cleveland Bay (CB) is the United Kingdom's oldest established horse breed. In this study we analyzed the genetic variability in CB horses and investigated its genetic relationships with other horse breeds. We examined the genetic variability among 90 CB horses sampled in the USA compared to a total of 3447 horses from 59 other breeds. Analysis of the genetic diversity and population structure was carried out using 15 microsatellite loci. We found that genetic diversity in CB horses was less than that for the majority of other tested breeds. The genetic similarity measures showed no direct relationship between the CB and Thoroughbred but suggested the Turkman horses (likely in the lineage of ancestors of the Thoroughbred) as a possible ancestor. Our findings reveal the genetic uniqueness of the CB breed and indicate its need to be preserved as a genetic resource.

Keywords: genetic variability; inbred horse; microsatellite; warmblood horse

1. Introduction

The Cleveland Bay (CB) is the only native warmblood horse in the United Kingdom [1] and is said to be the UK's oldest established horse breed with a history of more than 600 years. In the United States, the CB became a popular breed and, by the 1900s, there were over 2000 registered horses. However, this number decreased to fewer than 200 registered horses by 2009 and consequently it was listed by the Livestock Conservancy as a critically endangered breed. Likewise, in the United Kingdom, the Rare Breeds Survival Trust listed the CB as a breed of critical concern as there were fewer than 300 breeding females registered worldwide. Today the CB is used for recreational riding, driving, and equestrian competition. Most are used for breeding purposes but many are just kept as companion animals. The majority of purebreds reside in Great Britain with North America (primarily the USA) having about 25% of the total population. A handful of purebreds are in continental Europe, Australasia, Pakistan, and Japan (The Cleveland Bay Horse Society of North America CBHSNA, personal communication). The CB breed may have a mixed ancestry due to potential crosses with Barb, Arabian and Thoroughbred stallions [2] despite claims of the lack of Thoroughbred blood made in the introduction to the 1884 stud book [3]. It also has a shared history with other horses of the British Isles, including the Shire and Hackney Horse [4]. The CB has contributed to other modern horse breeds such as the Oldenburg, Holstein, and Hanoverian [5].

Genetic studies have unveiled the ancestry and diversity of many horse breeds, the impact of losing a breed to the species' overall diversity, and the effect of inbreeding on animals' health, especially in small populations [6–13]. In this study, we examined the genetic variation and the genetic relationships of the CB horses and compare those to 59 other breeds from Europe and the New World. Comprehensive information about genetic diversity is important for establishing appropriate

conservation and breeding strategies [14]. The information obtained can provide a baseline for conservation efforts which are underway in both the USA and the UK (personal communication from the CBHSNA).

2. Materials and Methods

Hair samples from 90 CB horses, collected from 2006 until 2008, were submitted to the Texas A&M Animal Genetics Laboratory by the horse owners from the USA. The hairs were used for parentage testing using 15 microsatellite markers. Microsatellites genotyping and DNA extraction were done as described before [15]. We also used genotypes of 3447 horses collected from different countries worldwide which represented 59 horse breeds genotyped with the same microsatellite markers. These genotypes came from a database of more than 100 recognized breeds maintained by one of us (EGC) at Texas A&M University (unpublished data). The breeds and their sample sizes are listed in Supplementary Table S1. We calculated standard gene diversity indices using GENEALEX 6 [16]. Majority-rule consensus trees were created with PHYLIP 3.69 [17] and visualized with MEGA4 [18]. The Przewalski horse was used as the out-group as per McCue et al. [19]. A factorial correspondence analysis was carried out using GENETIX v. 4.05 [20]. We used the STRUCTURE 2.3.3 software [21] to study the admixture pattern. Burn-in value was set to 20,000 and MCMC repetitions were 100,000 as recommended by the software developer. Runs for each selected value of k (k = 2 to k = 12) were repeated ten times. The software CLUMP [22] was used to align multiple replicates for each k in order to facilitate the interpretation of clustering results. The DISTRUCT application [23] was used to graphically display the results. The best number of clusters was determined based on Δk value [24] using Structure Harvester [25].

3. Results and discussion

Genetic diversity indices for the CB and 59 other horse breeds are shown in Supplementary Table S1. In the CB, the expected and observed heterozygosity (He = 0.619 and Ho = 0.629) were 1.7 and 1.2 standard deviations (SD) below the average for all breeds, respectively. Likewise, the mean number of alleles in the CB (MNA = 4.86) was found to be 1.7 SD below the average. Based on standard measures of variability (heterozygosities and measures of allelic diversity), the CB was well below the variation levels for almost all other breeds and ranked as the third lowest after the Friesian and Clydesdale horses, both of which are known to be highly inbred breeds. Reported inbreeding levels are 0.26 for the Clydesdale [26] and 0.16 in the Friesian [27]. The current mean inbreeding estimate for the CB is 0.19 (value provided by M. Elis, Registrar for the International Cleveland Bay Registry (ICBR); estimate calculated from pedigree using PedScope v 2.4.01xs from Tenset Technologies Ltd., Cambridge, UK). Because purebred horses generally have lower genetic diversity than non-breed horses [28,29], the low variation seen in the CB is a red flag but, given the breed's recent history, it is not surprising. However, the hope is that current breeding practices will preserve what diversity is still available and potentially increase variability levels through the importation of animals from other countries where the CB still exist. Current breeding practices in the USA primarily involve natural breeding, however, there is a growing use of in vitro fertilization and use of frozen semen to allow breeding to stallions that are not geographically close to the mare being bred to (G. Gordon, ICBR; personal communication). Use of frozen semen would facilitate gene exchange among horses from different countries.

To better understand the relationship between the CB and other horse breeds that may have contributed to its history, a majority-rule consensus tree was generated, Figure 1A. The major clades of the tree were as expected, with the draft and pony breeds falling into several clades together on a single major branch, a clade of breeds that have significant Thoroughbred input, horses of Iberian origin, and most other warmbloods forming other clades. The CB paired with the Connemara Pony in a cluster that included the Welsh Pony and Exmoor Pony. All the breeds in this clade were British Isle horses and this branch fell between the clades of the draft horses and ponies and all other breeds shown in

Supplementary Table S1. The Exmoor is a rather primitive breed [2] with low variability, Supplementary Table S1. The Connemara and Welsh Pony have relatively high variability and both have some crossing to the Thoroughbred (Hendricks 1995). The similarity of the CB and other breeds was also examined using the factorial correspondence analysis FCA for 13 breeds including all breeds from the CB clade and at least one breed representing each of the other sub-clusters in the tree. The CB did not show a strong relationship with any other breeds including the ones in the same clade (Figure 1B). This might be partially due to the low genetic variability in the CB. However, the CB clustered with the other ten breeds (Turkoman, Irish Draught, Garrano, Highland Pony, Welsh Pony, Percheron, Shetland Pony, Miniature Horse, Connemara Pony, and Norwegian), with only the Thoroughbred and Exmoor Pony falling outside of this group. Interestingly, the Turkoman was found to be the closest to the CB. The Turkoman is a possible founder of oriental breeds such as Barb (Edwards 1994) that is believed to have a part in the origin of the Thoroughbred and the CB. The CB and the 10 breeds sharing the same cluster from the FCA plot were then tested using the Bayesian-based analysis, Figure 1C. At a small k value (k = 3), the CB showed similarity with the Turkoman and the Irish Draught, supporting the results found in the FCA. At the best fit of k = 4, based on Delta_K plot (Supplementary Figure S1), the CB separated from the other two clusters and the Przewalski that was used as an out-group. There was no clear admixture pattern among the CB and all other breeds. However, the Turkoman and the Connemara Pony showed very little admixture with the CB. The results from the FCA and the admixture pattern perhaps are the best argument for the genetic uniqueness of the CB, and indicate further the need for this breed to be preserved as a genetic resource.
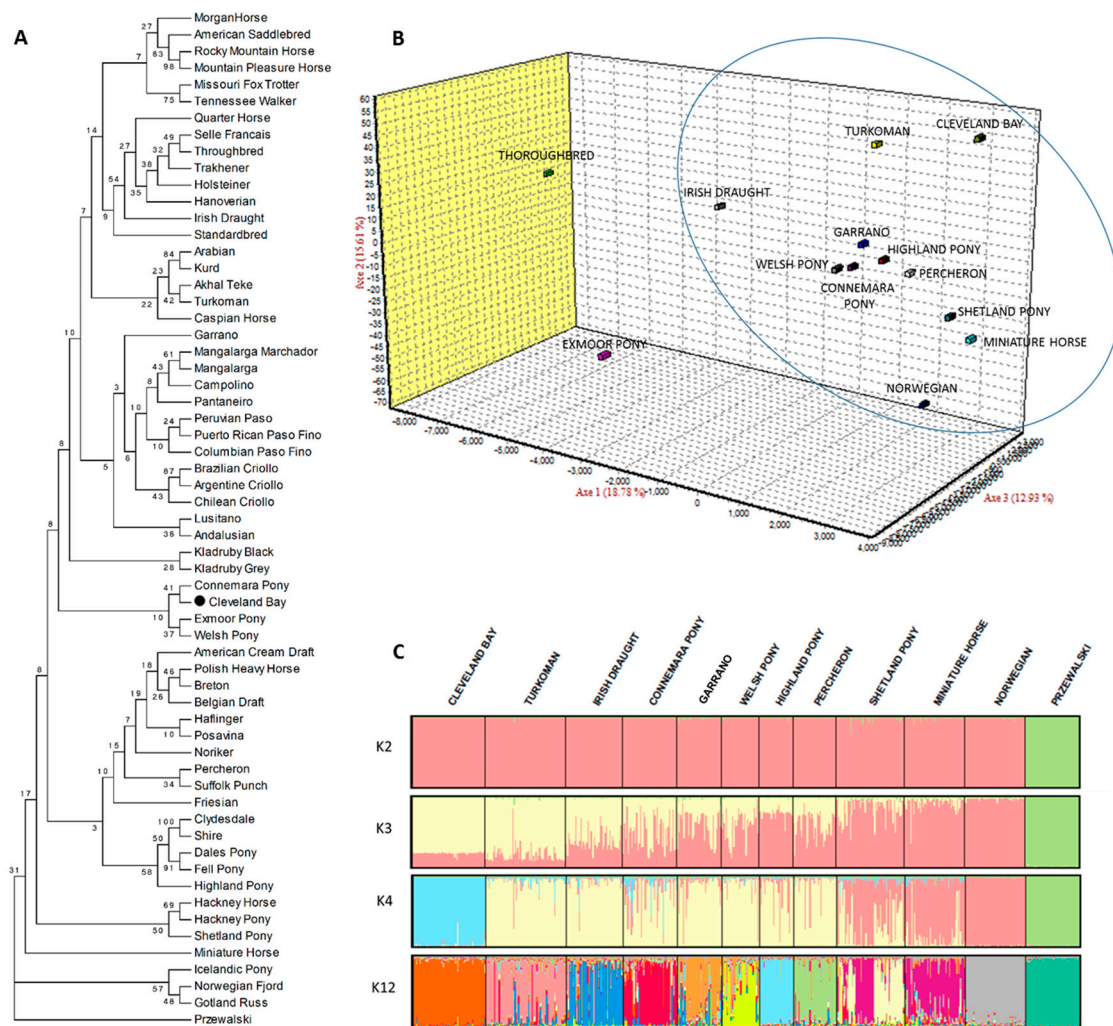
Figure 1. The genetic tree and population structure of the CB breed. (**A**) Dendogram of the genetic relationship among 60 horse breeds and Przewalski as an out-group based on majority-rule consensus RML tree; (**B**) the factorial correspondence analysis of 13 breeds plotted in 3D plot where Axis 1 accounts for 18.78% of the variation. (**C**) Clustering assignment of 11 horse breeds and Przewalski using the Bayesian method. Each individual is represented by a single column which is divided into segments whose size and color represent the relative proportion of the animal genome corresponding to a particular cluster at different values of k.

4. Conclusions

Genetic variation in the CB is quite low, with levels among the lowest found for domestic horse breeds. This low variation is an indication that active conservation measures for the breed should be undertaken. Many horse breeds have disappeared completely or been drastically reduced in number within the past one hundred years, and such low variation indicates that the CB could be at risk. Interest in reviving the CB began in the late 19th century when the studbook was first introduced. Since that time there has been a continued interest to keep the breed from extinction, although a serious threat to the breed in modern times has been the crossing of the CB with other breeds to produce sport horses. Genetic testing and detailed pedigree analysis provide a mechanism that can suggest the best matings to make in order to maintain the present levels of diversity within the breed. Considering the breed's own mixed history, as well as its influence on other breeds across Europe and America, an in depth genetic evaluation of the breed (possibly genome sequencing) seems appropriate. The Cleveland Bay is unique among world horse breeds and conservation is important.

References

1. Austen, C.; Corrie, S.; Swinney, N.J. *The Complete Illustrated Encyclopedia of Horses*; Metro Books: New York, NY, USA, 2008.
2. Hendricks, B.L. *International Encyclopedia of Horse Breeds*; University of Oklahoma Press: Norman, OK, USA, 1995.
3. Dixon, W.S. *The Cleveland Bay Stud Book*; Cleveland Bay Horse Society: Middlesbrough, UK, 1884.
4. Sidney, S.; Sinclair, J.; Charles, W.; Blew, A. *The Book of the Horse*; Cassell & Company: Detroit, MI, USA, 1893.
5. Edwards, E. *The Encyclopedia of the Horse*; Kindersley: New York, NY, USA, 1994.
6. Cothran, E.; Santos, S.A.; Mazza, M.C.M.; Lear, T.L.; Sereno, J. Genetics of the Pantaneiro Horse of the Pantanal Region of Brazil. *Genet. Mol. Biol.* **1998**, *21*. [CrossRef]
7. Cothran, E.G.; van Dyk, E. Genetic Analysis of Three South African Horse Breeds. *J. S. Afr. Vet. Assoc.* **1998**, *69*, 120–125. [CrossRef] [PubMed]
8. Juras, R.; Cothran, E.G.; Klimas, R. Genetic Analysis of Three Lithuanian Native Horse Breeds. *Acta Agric. Scand. Sect. A Anim. Sci.* **2003**, *53*, 180–185. [CrossRef]
9. Achmann, R.; Curik, I.; Dovc, P.; Kavar, T.; Bodo, I.; Habe, F.; Marti, E.; Solkner, J.; Brem, G. Microsatellite Diversity, Population Subdivision and Gene Flow in the Lipizzan Horse. *Anim. Genet.* **2004**, *35*, 285–292. [CrossRef]
10. Luis, C.; Bastos-Silveira, C.; Cothran, E.G.; Mdo, M.O. Iberian Origins of New World Horse Breeds. *J. Hered.* **2006**, *97*, 107–113. [CrossRef] [PubMed]
11. Luis, C.; Juras, R.; Oom, M.M.; Cothran, E.G. Genetic Diversity and Relationships of Portuguese and Other Horse Breeds Based on Protein and Microsatellite Loci Variation. *Anim. Genet.* **2007**, *38*, 20–27. [CrossRef]
12. Plante, Y.; Vega-Pla, J.L.; Lucas, Z.; Colling, D.; de March, B.; Buchanan, F. Genetic Diversity in a Feral Horse Population from Sable Island, Canada. *J. Hered.* **2007**, *98*, 594–602. [CrossRef]
13. Thirstrup, J.P.; Bach, L.A.; Loeschcke, V.; Pertoldi, C. Population Viability Analysis on Domestic Horse Breeds (Equus Caballus). *J. Anim. Sci.* **2009**, *87*, 3525–3535. [CrossRef]
14. Notter, D.R. The Importance of Genetic Diversity in Livestock Populations of the Future. *J. Anim. Sci.* **1999**, *77*, 61–69. [CrossRef]
15. Khanshour, A.; Conant, E.; Juras, R.; Cothran, E.G. Microsatellite Analysis of Genetic Diversity and Population Structure of Arabian Horse Populations. *J. Hered.* **2013**, *104*, 386–398. [CrossRef]
16. Paetkau, D.; Waits, L.P.; Clarkson, P.L.; Craighead, L.; Strobeck, C. An Empirical Evaluation of Genetic Distance Statistics Using Microsatellite Data from Bear (Ursidae) Populations. *Genetics* **1997**, *147*, 1943. [PubMed]
17. Felsenstein, J. *Phylip (Phylogeny Inference Package)*; Version 3.6. Distributed by the Author; University of Washington: Seattle, WA, USA, 2004; Available online: http://evolution.genetics.washington.edu/phylip.html (accessed on 13 August 2017).
18. Tamura, K.; Dudley, J.; Nei, M.; Kumar, S. Mega4: Molecular Evolutionary Genetics Analysis (Mega) Software Version 4.0. *Mol. Biol. Evol.* **2007**, *24*, 1596–1599. [CrossRef] [PubMed]
19. McCue, M.E.; Bannasch, D.L.; Petersen, J.L.; Gurr, J.; Bailey, E.; Binns, M.M.; Distl, O.; Guerin, G.; Hasegawa, T.; Hill, E.W.; et al. A High Density Snp Array for the Domestic Horse and Extant Perissodactyla: Utility for

Association Mapping, Genetic Diversity, and Phylogeny Studies. *PLoS Genet.* **2012**, *8*, e1002451. [CrossRef] [PubMed]

20. Belkhir, K.; Borsa, P.; Chikhi, L.; Raufaste, F.; Bonhomme, F. Genetix 4.05, Windows Tm Software for Population Genetics. In *Laboratorie Genome, Populations, Interactions*; Universite Montpellier II: Montpellier, France, 1996–2004.

21. Pritchard, J.K.; Stephens, M.; Donnelly, P. Inference of Population Structure Using Multilocus Genotype Data. *Genetics* **2000**, *155*, 945–959. [PubMed]

22. Jakobsson, M.; Rosenberg, N.A. Clumpp: A Cluster Matching and Permutation Program for Dealing with Label Switching and Multimodality in Analysis of Population Structure. *Bioinformatics* **2007**, *23*, 1801–1806. [CrossRef] [PubMed]

23. Rosenberg, N.A. Distruct: A Program for the Graphical Display of Population Structure. *Mol. Ecol. Notes* **2004**, *4*, 137–138. [CrossRef]

24. Evanno, G.; Regnaut, S.; Goudet, J. Detecting the Number of Clusters of Individuals Using the Software Structure: A Simulation Study. *Mol. Ecol.* **2005**, *14*, 2611–2620. [CrossRef] [PubMed]

25. Earl, D.A.; von Holdt, B.M. Structure Harvester: A Website and Program for Visualizing Structure Output and Implementing the Evanno Method. *Conserv. Genet. Resour.* **2012**, *4*, 359–361. [CrossRef]

26. Petersen, J.L.; Mickelson, J.R.; Cothran, E.G.; Andersson, L.S.; Axelsson, J.; Bailey, E.; Bannasch, D.; Binns, M.M.; Borges, A.S.; Brama, P.; et al. Genetic Diversity in the Modern Horse Illustrated from Genome-Wide Snp Data. *PLoS ONE* **2013**, *8*, e54997. [CrossRef]

27. Ducro, B.J.; Windig, J.J.; Hellinga, I.; Bovenhuis, H. Genetic Diversity and Measures to Reduce Inbreeding in Friesian Horses. In Proceedings of the 10th World Congress of Genetics Applied to Livestock Production, ASAS, Vancouver, BC, Canada, 17–22 August 2014.

28. Metzger, J.; Tonda, R.; Beltran, S.; Agueda, L.; Gut, M.; Distl, O. Next Generation Sequencing Gives an Insight into the Characteristics of Highly Selected Breeds Versus Non-Breed Horses in the Course of Domestication. *BMC Genom.* **2014**, *15*, 562. [CrossRef]

29. Schubert, M.; Jonsson, H.; Chang, D.; der Sarkissian, C.; Ermini, L.; Ginolhac, A.; Albrechtsen, A.; Dupanloup, I.; Foucal, A.; Petersen, B.; et al. Prehistoric Genomes Reveal the Genetic Foundation and Cost of Horse Domestication. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E5661–E5669. [CrossRef] [PubMed]