Appendix S1. Additional material

Materials and Methods

Horse Blood Samples

Blood samples from 353 horses of 11 breeds and two populations of unknown breed horses were obtained from the Equine Genetics Center herd at Cornell University, farms in the US, and Thoroughbred studs in Newmarket, UK. These comprised 97 Thoroughbreds, 129 Standardbreds, 32 Arabians, 11 Quarter horses, 7 warmbloods, 33 unknown breed ponies, 38 unknown breed horses, and 1 from each of the following breeds: Appaloosa, Friesian, Gypsy Vanner, Hanoverian, Missouri Fox Trotter, and Oldenburg. All samples were obtained following guidelines outlined in Cornell's Institutional Animal Care and Use Committee (IACUC) protocol #1986-0216.

Equine Leukocyte Antigen (ELA) Typing

Historically clusters of pregnancy-induced antisera with similar reaction patterns were used to define the ELA system, which is the species specific name of the MHC of the horse (Bright *et al.* 1978; Bailey *et al.* 1979). International workshops convened in the 1980s identified 19 specificities as products of the ELA system (Bull 1983; Bailey *et al.* 1984; Antczak *et al.* 1986; Bernoco *et al.* 1987; Lazary *et al.* 1988). This serological assay was applied to 218 horses (56 Thoroughbreds, 125 Standardbreds, 4 Arabians, 3 Quarter horses, 6 warmbloods, 1 Missouri Fox Trotter, 1 Oldenburg, 1 Appaloosa, 14 unknown breed ponies, and 7 unknown breed horses) based on tests at Cornell University or the University of Kentucky. Alloantigens referred to in this study included ELA-A1, A2, A3, A4, A5, A6, A7, A8, A9, A10, A15 and A19. These alloantigens were inherited as MHC class I products from a single locus. However, it is also likely that the reagents (complex alloantisera) recognized multiple specificities on multiple antigens and, therefore, defined a haplotype rather than single antigenic molecule. In any case, no horses were ever found possessing more than two of these alloantigens and they were assigned genetically to a putative ELA-A class I MHC locus (Lazary *et al.*1988).

Peripheral Blood Lymphocyte Isolation and Purification of Genomic DNA

Peripheral blood lymphocytes were isolated from heparinized venous blood samples, snap frozen in liquid nitrogen, and stored at -80°C. Genomic DNA (gDNA) was isolated from frozen lymphocytes using the DNeasy Blood & Tissue Kit, following the manufacturer's protocol (Qiagen Inc., Valencia, CA).

Selection of Intra-MHC Microsatellite Markers

The equine MHC region in the horse whole genome sequence (http://genome.ucsc.edu) was analyzed for microsatellite repeat regions using the Sputnik software (http://abajian.net/sputnik/index.html). Pigtail polymerase chain reaction (PCR) primers for repeats with a score above 22 were designed and tested on horses homozygous for five ELA haplotypes: ELA-A2, -A3, -A5, -A9, and -A10. These five horses were later included in the cohort for the microsatellite haplotype analysis. Loci with low polymorphism or ambiguous results were not analyzed further. Primers for three newly designated (COR112, COR113, COR114) and two previously discovered loci, UMN-JH34-2 (GenBank: EF531702) and UM011 (Meyer *et al.*1997; Swinburne *et al.* 2000) were selected for this study based on their polymorphism, stability, and overall performance and were directly labeled with fluorescent tags and synthesized using ABI custom tailed primer pair technology (Applied Biosystems, Foster City, CA). The positions of these markers relative to the equine MHC genes are shown in **Fig. S1**.

Microsatellite PCR Amplification and Fragment Analysis

To amplify microsatellite alleles, 50 ng of gDNA was added to a 10-µl reaction containing 1x PCR Buffer, 1.5 mM $MgCl_2$, 0.2 mM dATP, dCTP, dGTP, dTTP, 1 U *Taq* polymerase, and 0.25 µM forward fluorescent and reverse primers (**Table S4**). In some experiments, two multiplexes were used to amplify COR114 and COR112, and UM011 and UMN-JH34-2 as pairs, respectively. Each multiplex reaction contained identical concentrations of reagents as in a single reaction, except for the addition of two separate sets of primers each at 0.25 µM and 75 ng of gDNA, for a final volume of 10 µl. Amplification was carried out under the following conditions: 1 cycle

of 95°C for 5 minutes, followed by 33 cycles of 95°C for 1 minute, 58°C for 1 minute, and 72°C for 30 seconds, and 1 cycle of 72°C for 20 minutes for the final extension.
For fragment analysis, 2 µl PCR product was premixed with 14 µl Hi-Di™ Formamide and 1 µl GeneScan™-500 LIZ® Size Standard (Applied Biosystems) in a final volume of 17 µl. These samples were sized using the Applied Biosystems 3730x1 DNA Analyzer at the Life Science Core Laboratories Center at Cornell University, Ithaca, NY. Electropherograms were analyzed using Applied Biosystems PeakScanner Software v1.0.
Calculation of Correlation Coefficient (r)

A cohort of 218 horses (breed distribution as previously mentioned under "ELA Typing") that had been serotyped and microsatellite typed were selected for correlation analysis (Klein 1986).
Haplotype Frequency Determination

Haplotype frequencies of MHC haplotypes defined by microsatellites were determined by direct haplotype counting for 63 Thoroughbred horses, 123 Standardbred horses and 26 Arabian horses. These horses included sets of half-siblings but not full-siblings. The Standardbred horses were divided into Pacers and Trotters since they appear to represent different breeding populations within the breed (Bailey, 1983).

Supporting Text
Supporting Evidence for MHC Microsatellite Testing
Microsatellite loci have been useful proxies for detecting MHC polymorphisms in several species (Wagner *et al.* 1996; Santucci *et al.* 2007; Bastos-Silveira *et al.* 2009; Shichi *et al.* 2009), as these markers are valuable in studying haplotype lineages ad divergence. Comparison between different methodologies for characterizing the MHC either by a serological or a molecular approach has been performed in a few studies. In most instances, microsatellite haplotypes were documented as surrogate markers with high fidelity to their serological counterparts (Meagher and Potts 1997; Nunez *et al.* 2004; Penedo *et al.* 2005; Fulton *et al.* 2006).
Wade et al. (2009) compared linkage disequilibrium between horse breeds and found strong conservation, perhaps reflecting the recent divergence of horse breeds. Indeed, many modern horse breeds share founders within the last 10 to 25 generations. Therefore, microsatellite haplotypes spanning the equine MHC were investigated to determine if they could be used as proxies for serologically defined markers among and between horse breeds.
Definition of MHC Microsatellite Haplotypes

Microsatellite testing was performed on DNA samples from 353 horses using the microsatellite markers described above. Variable numbers of alleles for each microsatellite locus were detected: UMN-JH34-2 (12 alleles); COR112 (13 alleles); COR113 (8 alleles); UM011 (13 alleles); COR114 (10 alleles) (**Table S1**). Some of the alleles at each locus were rare and were observed only in a few animals without major contribution to haplotype identification. **Table 1** shows the haplotype constitution, the correlation coefficient, the breeds possessing each haplotype, and the number of individuals found with those haplotypes. Of the 353 horses in this study with 706 total haplotypes, 600 haplotypes were associated with nomenclature from **Table 1** while 106 haplotypes were not associated with a named haplotype according to the rules of nomenclature (approximately 15% of haplotypes).
Correspondence of Microsatellite Haplotypes with Serotypes

Strong correlations were found between MHC haplotypes determined serologically and those determined with microsatellite markers, with *r* values ranging from 0.75 for ELA-A19 and ELA-A15 to 0.95 for ELA-A8 (**Table S2**). In addition, multiple microsatellite haplotypes were associated with 5 of the 12 ELA serotypes (A3, A4, A5, A9, and A10). For example, 5 subtypes were defined for ELA-A3. Of those, 4 shared the 207 allele of UMN-JH34-2 (class I MHC region) but showed differences in the class II region microsatellite markers. Likewise, there were 4 subtypes of ELA-A10, all of which shared the 221 allele of UMN-JH34-2 but varied in class II region microsatellites.
Distribution of Microsatellite Haplotypes between Breeds

The majority of ELA serotypes in this study were determined for Thoroughbred and Standardbred horses (**Table S2**). Some of the serotypes were restricted almost entirely to one or the other of these breeds. For example, ELA-A2 and ELA-A9 were found among Thoroughbred but not Standardbred horses. Conversely, ELA-A1, ELA-A4, and ELA-A8 were found in Standardbreds but not in Thoroughbreds. To a lesser extent, this was also true for the microsatellite haplotypes associated with those serotypes. While serotypes for the ELA antigens ELA-A3, A5, and A10 were found in both breeds, microsatellite haplotypes associated with those serotypes were breed specific. For example, for ELA-A3, subhaplotypes A3a, A3b, A3c, and A3d were associated with Thoroughbred but not Standardbred horses while, conversely, A3e was detected exclusively in Standardbred horses.

The Thoroughbred breed originated over 300 years ago in England from a small group of horses (approximately 100-200) commonly thought to be based on Arabian horses crossed with native English horses (Willet 1970). The American Standardbred breed originated in the United States approximately 150 years ago

and included Thoroughbred horses among its founders. Regardless, the MHC microsatellite haplotypes divided along breed lines.

None of the microsatellite haplotypes of Thoroughbreds and Standardbreds were associated with Arabian horses, even though Arabians were shown to carry some of the same ELA serotypes in previous studies (Antczak *et al.* 1986). Since only few of the Arabian horses tested in this study had been serotyped, it is possible that the microsatellite commonly detected in Arabians (COR6, COR7, COR8; **Table S3**) represent subtypes of known ELA serotypes such as ELA-A1, which was reported at high frequency in Arabians (Antczak *et al.* 1986).

Supporting References

University of California Santa Cruz Genome Bioinformatics [Internet]; c2009.
Available from: http://genome.ucsc.edu/.

Abajian, C. University of Washington [Internet]; c2010.
Available from: http://abajian.net/sputnik/index.html.

Antczak D.F., Bailey E., Barger B., Guerin G., Lazary S., McClure J., Mottironi V.D., Symons R., Templeton J. & Varewyck H. (1986) Joint report of the Third International Workshop on Lymphocyte Alloantigens of the Horse, Kennett Square, Pennsylvania, 25-27 April 1984. *Animal Genetics* **17**, 363-73.

Bailey E., Stormont C., Suzuki Y. & Trommershausen Smith A. (1979) Linkage of loci controlling alloantigens on red blood cells and lymphocytes in the horse. *Science (New York, N.Y.)* **204**, 1317-9.

Bailey E. (1983) Population studies on the ELA system in American standardbred and thoroughbred mares. *Animal Blood Groups and Biochemical Genetics* **14**, 201-11.

Bailey E., Antczak D.F., Bernoco D., Bull R.W., Fister R., Guerin G., Lazary S., Matthews S., McClure J. & Meyer J. (1984) Joint report of the Second International Workshop on Lymphocyte Alloantigens of the Horse, held 3-8 October 1982. *Animal Blood Groups and Biochemical Genetics* **15**, 123-32.

Bastos-Silveira C., Luis C., Ginja C., Gama L.T. & Oom M.M. (2009) Genetic variation in BoLA microsatellite loci in Portuguese cattle breeds. *Anim Genet* **40**, 101-5.

Bernoco D., Antczak D.F., Bailey E., Bell K., Bull R.W., Byrns G., Guerin G., Lazary S., McClure J., Templeton J. & Varewyck H. (1987) Joint report of the Fourth International Workshop on Lymphocyte Alloantigens of the Horse, Lexington, Kentucky, 12-22 October 1985. *Anim Genet* **18**, 81-93.

Bright S., Antczak D.F. & Ricketts S. (1978) Studies on equine leukocyte antigens. Proc. Fourth International Congress on Equine Infectious Diseases, 1976 (Bryans, J. and Gerber, H., eds.), Suppl. No. 1. *Journal of Equine Medicine and Surgery*, 229-36.

Bull R.W. (1983) Joint report of the First International Workshop on Lymphocyte Alloantigens of the Horse held 24-29 October 1981. *Animal Blood Groups and Biochemical Genetics* **14**, 119-37.

Fulton J.E., Juul-Madsen H.R., Ashwell C.M., McCarron A.M., Arthur J.A., O'Sullivan N.P. & L. T.R., Jr. (2006) Molecular genotype identification of the Gallus gallus major histocompatibility complex. *Immunogenetics* **58**, 407-21.

Gustafson A.L., Tallmadge R.L., Ramlachan N., Miller D., Bird H., Antczak D.F., Raudsepp T., Chowdhary B.P. & Skow L.C. (2003) An ordered BAC contig map of the equine major histocompatibility complex. *Cytogenetic and genome research* **102**, 189-95.

Klein J. (1986) The Antibody. In: *Natural History of the Major Histocompatibility Complex.* (ed. by Anonymous), pp. 233-90. John Wiley & Sons, Inc., New York, NY.

Lazary S., Antczak D.F., Bailey E., Bell T.K., Bernoco D., Byrns G. & McClure J.J. (1988) Joint Report of the Fifth International Workshop on Lymphocyte Alloantigens of the Horse, Baton Rouge, Louisiana, 31 October-1 November 1987. *Animal Genetics* **19**, 447-56

Meagher S. & Potts W.K. (1997) A microsatellite-based MHC genotyping system for house mice (Mus domesticus). *Hereditas* **127**, 75-82.

Meyer A.H., Valberg S.J., Hillers K.R., Schweitzer J.K. & Mickelson J.R. (1997) Sixteen new polymorphic equine microsatellites. *Animal Genetics* **28**, 69-70.

Nunez Y., Ponz F. & Gallego F.J. (2004) Microsatellite-based genotyping of the swine lymphocyte alloantigens (SLA) in miniature pigs. *Research in veterinary science* **77**, 59-62.

Penedo M.C., Bontrop R.E., Heijmans C.M., Otting N., Noort R., Rouweler A.J., de Groot N., de Groot N.G., Ward T. & Doxiadis G.G. (2005) Microsatellite typing of the rhesus macaque MHC region. *Immunogenetics* **57**, 198-209.

Santucci F., Ibrahim K.M., Bruzzone A. & Hewit G.M. (2007) Selection on MHC-linked microsatellite loci in sheep populations. *Heredity* **99**, 340-8.

Shichi D., Ota M., Katsuyama Y., Inoko H., Naruse T.K. & Kimura A. (2009) Complex divergence at a microsatellite marker C1_2_5 in the lineage of HLA-Cw/-B haplotype. *J Hum Genet* **54**, 224-9.
Swinburne J., Gerstenberg C., Breen M., Aldridge V., Lockhart L., Marti E., Antczak D., Eggleston-Stott M., Bailey E., Mickelson J., Roed K., Lindgren G., von Haeringen W., Guerin G., Bjarnason J., Allen T. & Binns M. (2000) First comprehensive low-density horse linkage map based on two 3-generation, full-sibling, cross-bred horse reference families. *Genomics* **66**, 123-34.
Tallmadge R.L., Lear T.L. & Antczak D.F. (2005) Genomic characterization of MHC class I genes of the horse. *Immunogenetics* **57**, 763-74.
Wade C.M., Giulotto E., Sigurdsson S., Zoli M., Gnerre S., Imsland F., Lear T.L., Adelson D.L., Bailey E., Bellone R.R., Blocker H., Distl O., Edgar R.C., Garber M., Leeb T., Mauceli E., MacLeod J.N., Penedo M.C., Raison J.M., Sharpe T., Vogel J., Andersson L., Antczak D.F., Biagi T., Binns M.M., Chowdhary B.P., Coleman S.J., Della Valle G., Fryc S., Guerin G., Hasegawa T., Hill E.W., Jurka J., Kiialainen A., Lindgren G., Liu J., Magnani E., Mickelson J.R., Murray J., Nergadze S.G., Onofrio R., Pedroni S., Piras M.F., Raudsepp T., Rocchi M., Roed K.H., Ryder O.A., Searle S., Skow L., Swinburne J.E., Syvanen A.C., Tozaki T., Valberg S.J., Vaudin M., White J.R., Zody M.C., Lander E.S. & Lindblad-Toh K. (2009) Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science* **326**, 865-7.

Wagner J.L., Burnett R.C., DeRose S.A., Francisco L.V., Storb R. & Ostrander E.A. (1996) Histocompatibility testing of dog families with highly polymorphic microsatellite markers. *Transplantation* **62**, 876-7.

Willet, P. (1970) *The Thoroughbred*. p. 288. G.P.Putnum's Sons, New York.