ANIMAL GENETICS Immunogenetics, Molecular Genetics and Functional Genomics

doi: 10.1111/age.12003

Microsatellite variation in the equine MHC

C. L. Brinkmeyer-Langford*, J. J. Cai*, C. A. Gill† and L. C. Skow*

*Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, TX, 77843-4458, USA. †Department of Animal Science, Texas A&M University, College Station, TX, 77843-2471, USA.

Summary

Genes within the major histocompatibility complex (MHC) encode proteins involved in innate and adaptive immune responses. Genetic variation in this region can influence the immune response of an individual animal to challenges from a variety of pathogens; however, a complete documentation of genetic variation in the MHC is lacking for most domestic animals, including horses. To provide additional genetic markers for study of the horse MHC, or ELA (equine lymphocyte antigen), we identified 37 polymorphic microsatellite repeats in ELA and used these variations separately and together with published SNPs to investigate linkage disequilibrium (LD) and haplotype structure in a sample of Thoroughbred horses. ELA SNPs alone detected little LD, but microsatellites, either separately or combined with SNPs, revealed substantially more LD. A subset of markers in very high LD across the breadth of ELA may be predictive of structural polymorphisms or linked epistases that are important drivers of haplotype structure in Thoroughbreds.

Keywords horse, linkage disequilibrium, major histocompatibility complex, polymorphism

Introduction

The factors determining the susceptibility of an individual to a specific disease are complex, but familial associations suggest that inheritance comprises an important predisposing cause. Identification of loci involved in determining host response to pathogens is critical for identifying individuals at risk and for developing appropriate therapies. One region of the mammalian genome frequently associated with disease susceptibility is the major histocompatibility complex (MHC). Most MHC genes encode proteins of the innate or adaptive immune systems; consequently, genetic variation in the MHC frequently predisposes to autoimmune diseases and susceptibilities to various pathogens (Escayg *et al.* 1997; Thorsby 1997; Sharif *et al.* 1998, 1999; Bailey *et al.* 2000; Kaufman 2000; Park *et al.* 2004; Dukkipati *et al.* 2006).

Although associations of MHC haplotypes with disease susceptibilities are commonly detected, the identification of specific variant alleles causal for disease susceptibility has been difficult, due in part due to the peculiar nature of the MHC. The MHC contains some of the most polymorphic genes of the vertebrate genome, but alleles of some of these genes are ancient and predate speciation of related taxa (Klein 1987). Because trans-species polymorphisms confound analysis of MHC history and haplotype structure by single nucleotide polymorphisms (SNPs), many investigators have turned to microsatellite repeats and microsatellite–SNP combinations as a source of more rapidly diversifying gene markers to provide a better understanding of MHC evolution and function (e.g. Mountain *et al.* 2002).

The identification of haplotypes associated with susceptibility to disease can expedite the identification of animals at risk and assist in identifying causative alleles and genetic mechanisms underlying susceptibility phenotypes. Consequently, identification of linkage disequilibrium (LD) and haplotype structure can assist researchers seeking to identify predisposing loci for diseases and other health-related phenotypes. Variation in ELA (equine lymphocyte antigens) has been described serologically, by restriction fragment length polymorphism (RFLP) and by microsatellite repeats, but these markers collectively have revealed little about LD and the haplotype structure of ELA. McCue *et al.* (2012) recently described the design of an equine 50K SNP chip that included 69 SNPs within ELA, which significantly contributes to the tools available for analysis of ELA. However, the low density of SNP markers across ELA (approximately 1 SNP/72.5 Kb) is inadequate to detect LD and haplotype structure in this important region of the

Address for correspondence

C. L. Brinkmeyer-Langford, Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, TX 77843-4458, USA.

E-mail: cbrinkmeyer@cvm.tamu.edu

Accepted for publication 14 August 2012

horse genome. In this communication, we identify and characterize 37 polymorphic microsatellite repeats that map within the boundaries of the horse ELA to provide additional markers for analysis of ELA. To test the power of the microsatellite markers to inform analysis of ELA, we genotyped the Thoroughbred subset of animals previously genotyped for SNPs in the equine HAPMAP project and compared patterns of LD of ELA as revealed separately by microsatellites, by published SNPs (McCue *et al.* 2012) and by both types of markers in combination.

Materials and methods

DNA sources

DNA from three cohorts of horses was used for microsatellite characterization, genotyping and haplotype analysis. These included samples from 48 horses (12 sires and 36 progeny) of 10 breeds and cross-bred stocks in the International Horse Reference Family Panel (IHRFP; Guérin *et al.* 1999, 2003), 174 pedigreed and registered Thoroughbred horses (archived at the College of Veterinary Medicine at Texas A&M University) and 24 Standardbreds kindly provided by Dr. Ernest Bailey, The Gluck Center, University of Kentucky.

In addition, ELA SNPs from horses used as part of the horse whole-genome SNP project were included in SNP-based haplotype analysis. These included 351 horses from 15 breeds, including one Akhal Teke, 21 Andalusian, 25 Arabian, 24 Belgian, 22 Franches Montagnes, 18 French Trotters, 20 Hannoverian, 19 Icelandic, 46 Thoroughbred (from four different herds), 21 Mongolian, 22 Norwegian Fjord, 49 Quarter Horse, 23 Saddlebred, 21 Norwegian Standardbred and 19 Swiss Warmblood horses.

Microsatellite genotyping

Selection

The SPUTNIK DNA microsatellite repeat search utility (http://espressosoftware.com/sputnik/index.html) was used to probe ELA genomic sequence (version 2.0) for microsatellite sequences with motifs of 2–5 bp in length. $(AC:GT)_n$ microsatellites from all regions of ELA were selected for primer design. To maximize the number of polymorphic markers used in analysis, we selected those with ≥ 6 repeated units for further screening due to the increased likelihood of polymorphisms in longer repeats (Weber 1990).

Primer design

Prior to primer design, the 500 bp of flanking sequence around each microsatellite was repeat masked (http://www.repeatmasker.org/) and microsatellites found to be within other repeat elements were excluded from further investigation. Primer design was performed using PRIMER3 (http://frodo.wi.mit.edu/primer3/); resulting primers were screened using BLAST to ensure uniqueness and specificity. Forward primers were given a 5′ M13 tag consisting of the sequence 5′-TTTCCCAGTCACGACGTTG-3′.

Genotyping procedure

Polymorphism and Mendelian inheritance of microsatellites were confirmed by genotyping stallions and selected offspring from the IHRFP. PCR was performed in 10 μl volumes consisting of 2 μl (50 ng) DNA; 2 μl ddH_2O; 1 μl 10X buffer; 0.1 μl 20 mM dNTPs; 1 μl Master Amp™ PCR Enhancer (Epicentre Biotechnologies); 0.3 μl clear Sigma JumpStart Taq Polymerase (Sigma-Aldrich); 2 μl 5 μM M13 NED, VIC, or 6-FAM fluorescent labeling dye; 0.66 μl 1 μM forward primer; and 1 μl 10 μM reverse primer. All reactions were performed with the following cycling conditions: an initial denaturation at 95 °C for 20 min, followed by 35 cycles each consisting of a 30 s at 94 °C denaturation, 30 s at the appropriate annealing temperature (60 °C) and a 30 s extension at 72 °C; with a final 10 min extension at 72 °C. Amplification was verified by running 4 μl of product on a 2% agarose gel with 0.25 μg/ml ethidium bromide, and the remaining product was used for genotyping on an ABI-3730 automated capillary sequencer (PE Applied Biosystems) normalized with internal size standard (GS 500 LIZ®; Applied Biosystems). GENEMAPPER™ 3.5 software (Applied Biosystems) was used to manually score the output and determine microsatellite lengths. Microsatellites identified as polymorphic in this way were genotyped on the remaining horses from each group.

Sequencing

Microsatellite lengths and polymorphic status were verified by sequencing representative amplicons at each locus (SeqWright). Amplicons were cloned into the TOPO Blunt end vector (Invitrogen) and transfected into chemically competent *E. coli* (DH5α). Cultures (1.5 ml each) were collected by centrifugation and delivered to SeqWright for DNA purification and sequencing. Sequencing chromatograms and sequence files were manually examined to validate sequence quality and confirm microsatellite characteristics.

Haplotype analysis

Haplotypes were reconstructed from microsatellite genotyping data using PHASE (Stephens *et al.* 2001). Animals and markers with at least 80% genotype call rates were included in the microsatellite evaluation. Microsatellite data were also analyzed using MIDAS (Gaunt *et al.* 2006) to establish D'_{ij}, χ^2 and r^2 for allelic combinations.

Published 2012. This article is a U.S. Government work and is in the public domain in the USA., doi: 10.1111/age.12003

ELA haplotype analysis was also performed on data generated by the whole genome SNP project, incorporating SNPs located within ELA (http://ftp.ncbi.nih.gov/snp/). Haplotypes were reconstructed from the SNP data with PHASE (Stephens *et al.* 2001) and FASTPHASE (Scheet & Stephens 2006), using data with 90% or higher call rates for animal and marker. For both microsatellites and SNPs, markers with minimum allele frequencies <0.05 were excluded from analysis. Output was visualized and assessed using the Haploview program (Barrett *et al.* 2005).

Results

Microsatellite genotyping

SPUTNIK interrogation of the genomic sequence of ECA20 from positions 27.4 to 33.9 Mb identified a total of 557 microsatellites with repeated units of 2–5 bp in length. This number included 185 dinucleotide, 124 trinucleotide, 150 tetranucleotide and 98 pentanucleotide microsatellites. Of the dinucleotide repeats, 87 were of the motif $(AC:GT)_n$ and 67 (81%) of these contained at least six repeat units. A total of 73 microsatellite repeats were analyzed in this study, including 13 with compound or imperfect dinucleotide composition. As expected, the frequencies of polymorphic microsatellites detected in this study of the ELA region were highly correlated with the length and composition of each marker. Of the 73 microsatellites surveyed for variation, 28 (38.3%) were monomorphic in the populations tested; three (4%) were located in repetitive regions, four (5.4%) produced multiple amplicons and one (1.3%) did not yield an acceptable primer pair. Of the 37 polymorphic microsatellites, three (2.3%) contained compound microsatellites, mean number of alleles = 5.0 (range 3–6). Among the 34 polymorphic microsatellites with perfect repeats, 14 (41%) had fewer than 10 repeat units, mean number of alleles = 3.9 (range 2–8); nine (26%) had repeat units from 10 to 14, mean number of alleles = 7.0 (range 2–13); and 11 (32%) with 15 or more repeat units had mean number of alleles = 9.0 (range 5–13).

Primer pairs were designed for 63 of the 67 $(AC:GT)_n$ microsatellites with ≥ 6 repeat units. The remaining four microsatellites were either located in regions of repetitive sequence or suitable primers could not be designed; these sequences were not considered for markers (Table S1). To provide more even coverage across ELA, we designed six additional primer pairs for $(AC:GT)_n$ microsatellites with five repeated units for a total of 69 primer pairs. All primers were optimized for PCR and evaluated against sires and selected offspring of the IHRFP to determine polymorphism and confirm Mendelian inheritance. Amplicons from each primer pair were sequenced to confirm allele sizes and the sequences deposited in Genbank with accession numbers JQ769146–JQ769184.

Of the 69 primer pairs tested, 37 (53%) amplicons were polymorphic and 30 (41%) were monomorphic among the IHRFP horses. Four primer pairs produced amplicons with multiple bands and were removed from further study. Among the polymorphic microsatellites were 31 newly described markers and six previously described markers (Meyer *et al.* 1997; Roberts *et al.* 2000; Tozaki *et al.* 2001; Milenkovic *et al.* 2005; Tseng *et al.* 2010). An additional six of the markers characterized in this study were identified but not characterized in a survey of *in silico* predicted markers (Mittmann *et al.* 2010).

The 37 polymorphic microsatellites, spanning 4.96 Mb of ELA, were used to genotype the full complement of IHRFP horses and populations of pedigreed Thoroughbreds and Standardbreds (see Materials and methods) to assess polymorphism and verify inheritance. The 37 remaining markers were successfully scored in the Thoroughbred and Standardbred sample sets, whereas only 35 markers were reliably scored in the IHRFP horses (markers 446–219 and 525–382 were removed due to ambiguities in allele calling). Variable DNA quality among the IHRFP samples was likely responsible for the ambiguous genotyping. On average, 95% of the Thoroughbreds and 98% of the Standardbreds were successfully genotyped for each of the 37 microsatellites, whereas only 83% of IHRFP animals on average were successfully genotyped for each marker. The average number of alleles across the panel of microsatellites is 6.4 (range 2–15). Polymorphic information content (PIC) values differed among breeds but averaged 0.49 across all breeds. The numbers and sizes of alleles, PIC values, observed heterozygosity and genotyping success for each of the 37 microsatellite markers across the three discovery and validation populations are presented in Table S1.

The amplicons of all microsatellite alleles were sequenced and compared to the equine reference genome sequence (Twilight) to verify sequence identity and to exclude homoplasy. In doing so, we detected 45 SNPs within 250 bp of 20 of the 37 markers (Table S1). Since SNPs and microsatellite markers have dramatically different mutation rates (Crow 1993; Ellegren 2000; Weber & Wong 1993), analysis of genomic regions using both types of genetic markers can provide complementary evolutionary information provided the two markers, called SNPSTRs, are in strong LD (Mountain *et al.* 2002). SNPSTRs within 250 bases of each other are almost always in complete LD (Agrafioti & Stumpf 2007), and all the SNPs identified within microsatellite amplicons in this study fall within this spacing interval.

Analyzing the microsatellite variation reported here in the context of flanking SNPs should provide useful information on the ancestry and diversification of these markers among different breeds of horses and by extension will inform studies of the evolution of ELA haplotypes.

SNP genotyping

The horse whole-genome SNP project (Wade *et al.* 2009) identified 4392 SNPs in the ELA region (ECA20:

Published 2012. This article is a U.S. Government work and is in the public domain in the USA., doi: 10.1111/age.12003

27 403 526–33 865 081 bp) as delineated by sequence homologies with the classical and extended regions of HLA (Brinkmeyer-Langford *et al.* 2010). Examination of the numbers and distribution of ELA SNPs incorporated into the design of the Illumina 50K SNP chip (McCue *et al.* 2012) revealed that ELA is relatively poorly marked. Of the 4392 candidate SNPs in the ELA region, only 96 met the requirements for eligibility in the design of the Illumina 50K SNP chip. Twenty-seven SNPs from this set were subsequently eliminated from analysis for lack of validating genotype data including 19 that were not confirmed as polymorphic and eight that were typed on too few horses to be properly evaluated. Analysis of the 69 remaining SNPs revealed that that the markers were generally evenly distributed across ELA, with 36 located within the ELA class I (ECA20: 28 737 669–30 620 000) and extended class I regions (ECA20: 27 403 526–28 737 668), six located within a large segmental duplication (ECA20: 30 620 000–31 321 086; Brinkmeyer-Langford *et al.* 2010) 12 within the class III region (ECA20: 31 321 086 –31 896 104), one within a 650 Kb 'gene desert' (ECA20: 31 896 104–32 572 317) between the class III and class II regions and 15 within the class II (ECA20: 32 621 480–33 555 172) and extended class II regions (ECA20: 33 557 799–33 849 668). The average inter-SNP spacing across ELA is ~93.3 kb, more than twice the genome-wide average of ~43 kb (McCue *et al.* 2012). Furthermore, 13 (five pairs and one trio) of the 69 SNPs are tightly clustered in six small segments, each segment spanning <250 bp. The very close spacing of these SNPS thereby effectively reduces the number of informative SNPs in ELA to 62 and increases the average inter-SNP distance to approximately 100 kb.

Using an r^2 value of ≥ 0.8, we determined the minimum set of ELA SNPs (tagSNPs) that could be reliably used to predict the alleles of all 69 ELA SNPs in each breed. An average of 57 tagSNPs, or 92%, of informative SNP markers were required to predict all 69 SNP variants in ELA for each breed except Standardbreds and Thoroughbreds. Standardbreds required 41 tagSNPs to represent ELA SNPs, the fewest of any breed genotyped in this study, and Thoroughbreds required the second fewest, at 45 tagSNPs. These results agree with Tseng *et al.* (2010), who found that Thoroughbreds and Standardbreds required the fewest number of microsatellite markers to define >90% of ELA haplotypes

Analysis of linkage disequilibrium

Microsatellites

Microsatellite genotypes were determined for the Thoroughbred samples used in the Illumina SNP array experiment (McCue *et al.* 2012). Genotypic data were analyzed for LD by MULTIALLELIC INTERALLELIC DISEQUILIBRIUM ANALYSIS (MIDAS) software to provide estimates of D'_{ij}, χ^2 and r^2 for allelic combinations. Graphical output cannot be exported from MIDAS; however, screenshots are shown in Fig. S1. Furthermore, MIDAS output in tabular form is available as Table S2. Examination of the decay of LD with distance indicates that ELA contains two populations of microsatellite markers, one of which decays rapidly to about $\chi^2 \simeq 0.25$ and then plateaus at $\chi^2 \simeq 0.20$ out to 6 Mb; a second set of markers are in complete LD across the entirely of ELA. When both sets of markers are combined, the χ^2 estimate of LD actually increases with distance (Fig. 1a).

SNPs

PHASE output of the ELA SNP genotypes from the EquineSNP50 SNP Array populations (McCue *et al.* 2012) was visualized in HAPLOVIEW. The HAPLOVIEW graphic for TB is presented in Fig. 2, and HAPLOVIEW outputs for all other

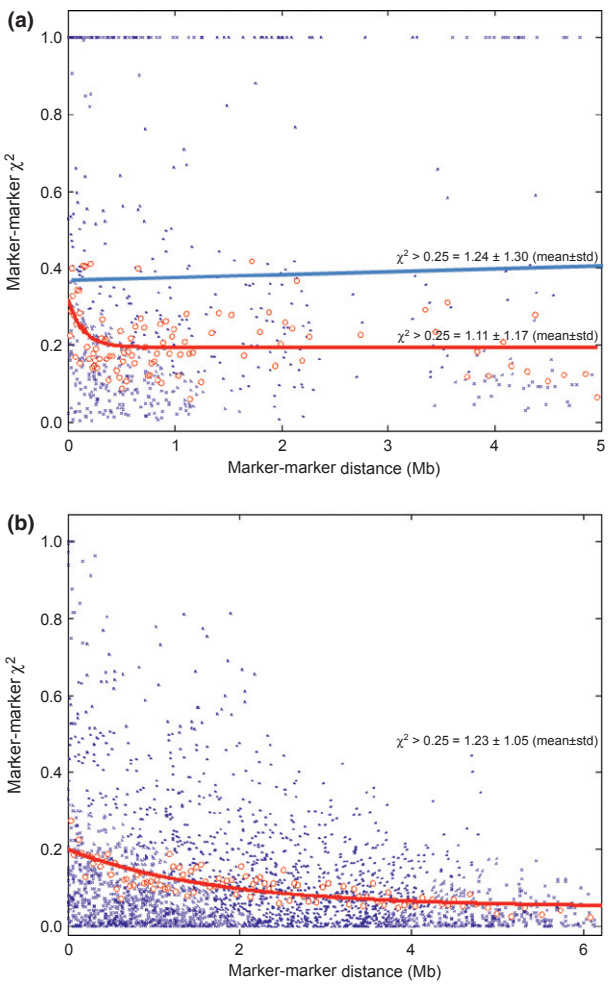

Figure 1 Plots of χ^2 values of 111 pair-wise comparisons of microsatellite (a) and SNP (b) markers, illustrating LD over distance. Regression lines are provided in bold. For the microsatellite plot (a), the upper line (blue) was calculated including the LD = 1 data points and shows LD increasing with distance, whereas the lower line (red) omits the LD = 1 data points and shows LD decreasing with distance.

Published 2012. This article is a U.S. Government work and is in the public domain in the USA., doi: 10.1111/age.12003

breeds analysed are available in Fig. S2. Blocks of LD in SNPs across the ELA were generally <10 kb in size in most breeds (Fig. 1b) and highly correlated with SNP density as expected. Therefore, LD estimated by SNPs across the ELA region appears to be similar to the estimated genome wide LD for most breeds of horses (Wade *et al.* 2009), although the low density and irregular spacing of SNP markers may underestimate LD in the ELA region. Exceptions to this generalization included the Andalusian breed with a single block of LD spanning 683 kb within the class II region and the Thoroughbred breed with six blocks of LD >100 kb, ranging from 131 to 857 kb in size.

Microsatellites and SNPs combined

For integrated marker analysis, SNP genotypes were extracted from the Thoroughbred subset of horses genotyped in the equine whole-genome SNP project (McCue *et al.* 2012) and combined with microsatellite genotypes for analysis of LD in ELA. Normalized χ^2 analysis of LD between multi-allelic markers (Zhao *et al.* 2007) was used to estimate the pair-wise LD between all SNP and microsatellite markers in ELA (Fig. 1a,b). LD, as measured by associations of SNP markers, declined rapidly over distance to about 0.1 (Fig. 1b). This is similar to the results for whole genome LD reported by *McCue et al.* (2012).

A similar examination of the decay of LD with distance, as measured by associations of microsatellite markers (Fig. 1a), indicates that ELA contains two populations of microsatellite markers, one of which decays rapidly as expected and then plateaus at $\chi^2 \simeq 0.20$ over the remainder of ELA 6 Mb. A second set of microsatellite markers are in complete LD across the entirety of ELA. When both sets of markers are combined, the χ^2 estimate of LD actually increases with distance (Fig. 1a) to a value of 0.4 at 5 Mb.

To characterize those microsatellite markers in strong LD (Fig. 1a), we selected all pair-wise comparisons of markers >1 Mb apart with $\chi^2 \geq 0.8$ and represented them as an ARC diagram over the genetic map of ELA (Fig. 3). The size of the dot representing each marker is proportional to the number of links connected with that marker. The data represented in Fig. 3 show a remarkably non-random distribution of ELA markers in strong LD, especially among class I markers, markers around *TNF* in the class III region and markers in the DR/DQ region of class II. The patterns of LD shown in Fig. 3 suggest that certain combinations of alleles are conserved as blocks across the entirety of ELA separated by three regions of low LD (blue bars on Fig. 3). A more detailed examination of the markers used to generate Fig. 3 disclosed that microsatellite markers contributed disproportionately to the observed long-range LD and that the non-random pattern of LD is a consequence of marker distribution For example, of the 87 pair-wise comparisons of markers ≥ 2 Mb apart with LD ≥ 0.8, all such marker pairs were microsatellites. The three regions of ELA without strong LD are regions that contain no or very few microsatellite markers. Although allele specific interactions might explain some of the LD, the patterns of LD in ELA observed in Fig. 3 seem to be largely a function of the type and distribution of genetic markers. In regions exclusively or predominantly marked by SNPs, no long-range LD is detected. This observation is likely due to the age of the

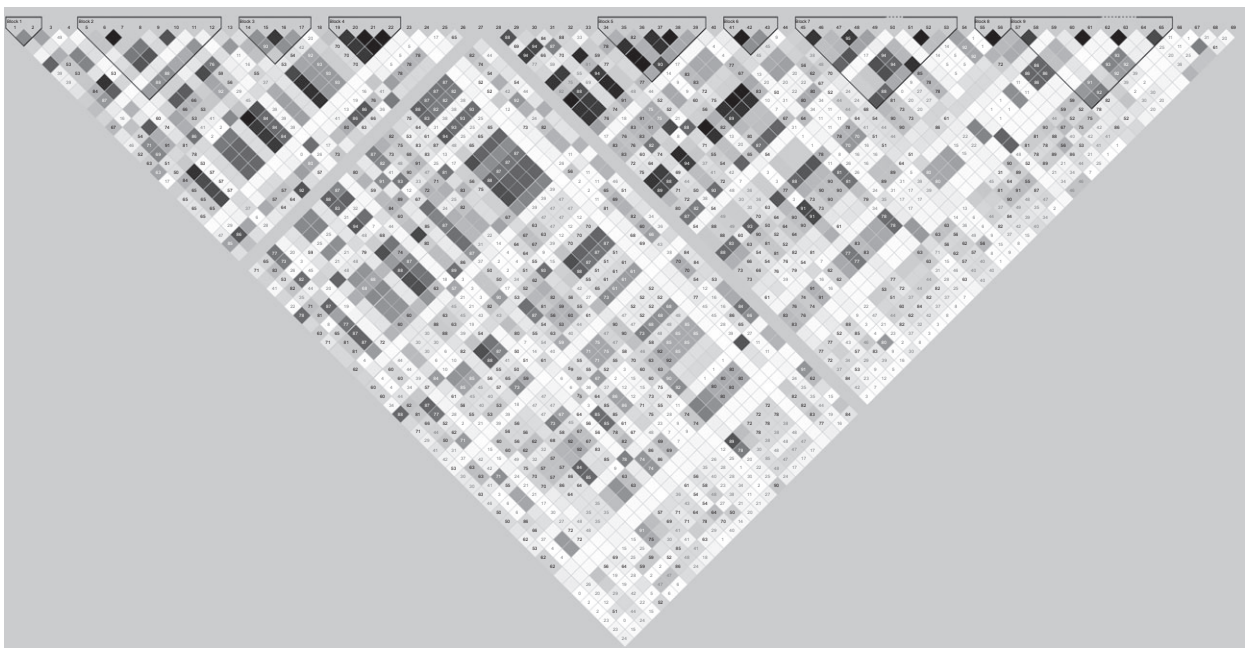

Figure 2 Haploview output showing linkage disequilibrium in the SNP-genotyped Thoroughbred horses. The shaded squares represent pair-wise comparisons between SNPs, with the degree of shading indicating r^2 value (darker shading = higher r^2 value).

Published 2012. This article is a U.S. Government work and is in the public domain in the USA., doi: 10.1111/age.12003

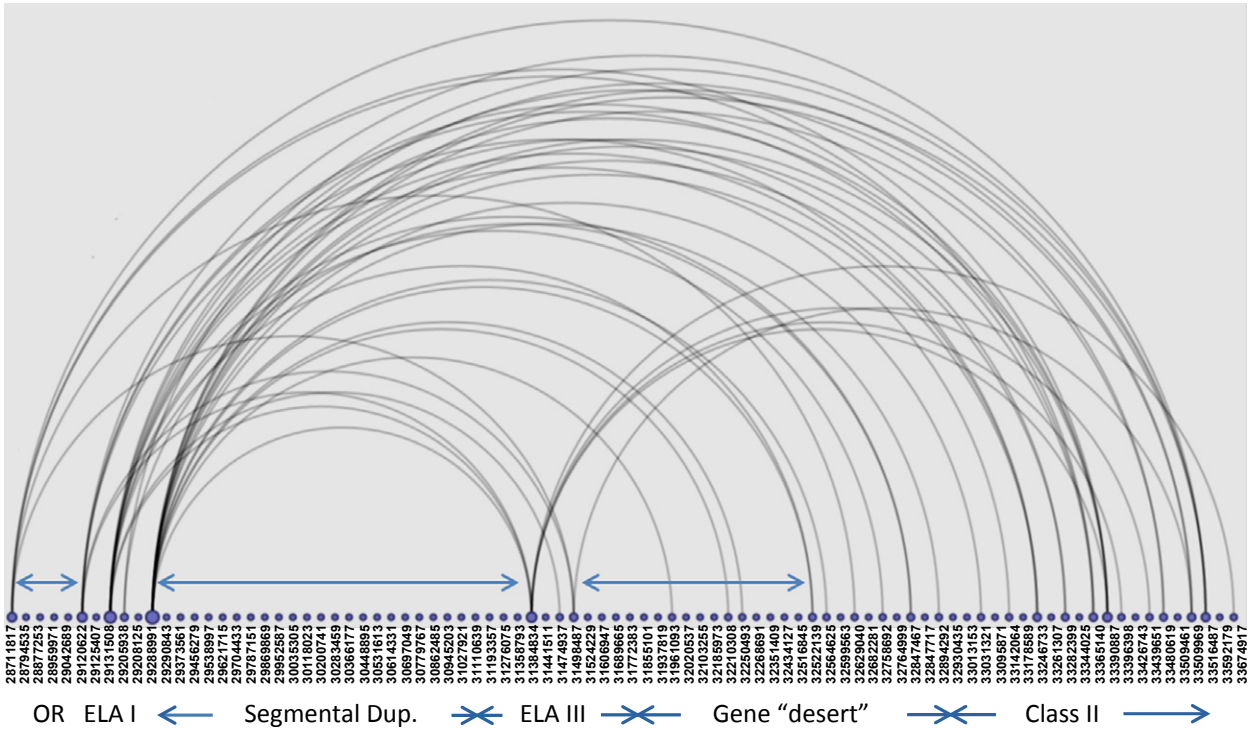

Figure 3 LD over distances of $\geq$1 Mb between SNPs or microsatellite markers. The different classes of MHC genes are found below the graphic, with arrows indicating the part of the MHC occupied by each class. 'OR' indicates olfactory receptor genes; 'Segmental Dup.' indicates a segmental duplication (see text).

SNPs (only 'common' SNPs were selected for the Illumina 50K chip), the low density of SNPs across ELA and the observation of high percentages of SNP homozygotes within breeds (Table S1). Conversely, microsatellite markers are mutationally more dynamic than SNPs and presumably represent more recent divergence. This property, coupled with the multi-allelic nature of microsatellites, identifies LD structure in ELA that is not observable by SNP analysis. This interpretation is consistent with a recent report by Tseng *et al.* 2010, in which a set of only five microsatellite markers distributed across ELA was sufficient to define more than 90% of the serologic haplotypes in a large sample of the Thoroughbred and Standardbred breeds.

Discussion

The domestic horse is presently the only representative of the mammalian order Perissodactyla for which whole genome sequence is available. The analysis of ELA in this study will facilitate comparative studies of the vertebrate MHC and provide for in-depth candidate gene searches for disease susceptibilities in this species. Here we report the characterization of 37 polymorphic microsatellites in the ELA among 10 breeds and cross-bred stocks of horses. Thirty-one of the markers are newly described, and six markers have been previously characterized (Meyer *et al.* 1997; Tozaki *et al.* 2001; Tseng *et al.* 2010). Twenty of the microsatellites also have closely associated SNPs (SNPSTRs; Agrafioti & Stumpf 2007) that may be useful for tracking the origins and dispersion of different ELA haplotypes among equids. In addition, this study examines the haplotype structure of ELA in the Thoroughbred breed as revealed by microsatellites and published SNPs from the horse HAPMAP project. Collectively these analyses define an additional region of the horse genome interrogated for LD.

The selection and distribution of SNPs and microsatellites used in this analysis influenced the results in several ways. First, the design of the Illumina50K SNP chip produced an array of ELA markers with an average distance between SNPs of ~93 Kb. However, five pairs and one trio of SNPs were located within 200 bp of each other in six different regions of ELA, increasing the effective distance between informative SNPs to approximately 100 Kb and illustrating the need for greater SNP density to better define LD across the ELA region and probably the horse genome in general. In addition, the percentage of individuals homozygous at any given SNP varied only slightly across breeds, but was quite high overall, ranging from 56 to 71% (average over 64%), further reducing the informative power of SNPs across ELA.

The 31 microsatellites newly identified in this study, combined with the six microsatellites previously described, will contribute significantly to analysis of ELA structure. Microsatellites are characterized by a relatively high muta-

Published 2012. This article is a U.S. Government work and is in the public domain in the USA., doi: 10.1111/age.12003

tion rate and are preferable to SNPs for analysis of within species/breed diversification. Unfortunately, the distribution of microsatellite markers across ELA is somewhat irregular with no or few microsatellites located in the regions of the segmental duplication and gene desert, two portions of ELA that collectively span almost 1.5 Mb of DNA.

Linkage disequilibrium estimated by SNPs for ELA in this study is less than observed for genome-wide LD in Thoroughbreds (McCue *et al.* 2012) and likely reflects the relative scarcity of SNP markers for ELA on the Illumina SNP chip and the high frequency of homozygotes among the Thoroughbred sample. Conversely, microsatellite markers revealed a surprising amount of LD across the entirety of ELA, much more than the genome average predicted by SNPs. This observation is consistent with studies describing increased LD in the MHCs of other species (Davies *et al.* 1992, 1994a,b; Martin *et al.* 1995; Snoek *et al.* 1998; Carrington 1999; Ahn *et al.* 2002). The patterns of LD revealed by ARC plot representation for Thoroughbreds (Fig. 3) reveal extensive LD for many alleles extending over the entirety of the ELA. We do not consider misassembly of the equine genome sequence as a likely explanation for this observation, as the current assembly has been validated via linkage and radiation hybrid mapping and fluorescent *in situ* hybridization (FISH) mapping (Raudsepp *et al.* 2008). A more plausible explanation for the LD pattern observed in Thoroughbreds may include structural polymorphisms, perhaps one or more inversions, within ELA that would serve to reduce meiotic recombination in this region of the genome. Regions of markedly reduced recombination in the MHCs of other species have been observed (Gyllensten & Erlich 1993; Jeffreys *et al.* 2001; Meyer & Thomson 2001; Cullen *et al.* 2002; Seddon *et al.* 2010), but the mechanisms responsible have yet to be fully identified. The generalized observation of reduced recombination in the mammalian MHC indicates some biological advantage to maintaining certain allelic associations as haplotypes (Raymond *et al.* 2005). Genetic interactions of risk alleles for certain MHC-associated disease have been reported in humans (Lincoln *et al.* 2009; Noble *et al.* 2010). To our knowledge such genetic interactions among 'normal' variants of the MHC have not been reported.

The significant LD observed in Thoroughbreds using microsatellites may provide an avenue to address some of the questions discussed above. Alleles of markers in complete LD across Mb scales of DNA may predictively identify haplotypes with structural polymorphisms or other features responsible for reduced recombination. Consequently these markers may provide the tools with which to identify horses in other populations and from other breeds that likely contain genetic variation needed to search for and characterize structural variation in ELA. In addition, the long range LD observed for certain allelic combinations, identifies specific genomic targets amenable to functional genomic approaches to assess the hypothesis of allelic interactions. Such studies will be informative whether or not structural variation can be demonstrated within the ELA complex.

Conclusions

This work significantly expands the number of genetic markers available for analysis of the equine MHC by contributing an additional 33 microsatellite markers to augment the small set of previously described microsatellite markers and public SNPs. Using SNPs as markers identified little LD in ELA in a survey of breeds included in the equine SNP study (McCue *et al.* 2012). On the other hand, microsatellite markers used separately or in combination with published SNPs revealed significant LD in Thoroughbreds and identified sets of markers that define haplotypes across the entire length of ELA. These markers are in complete LD across Mb scales of DNA and may be predictive of structural rearrangements or functionally important allelic interactions.

Acknowledgements

We wish to acknowledge Alex Holliday for his assistance with genotyping the microsatellites. Ernie Bailey (University of Kentucky) provided DNA for serotyped Thoroughbred and Standardbred horses. We wish to extend our gratitude to members of the equine genomics community who kindly allowed access to the Gentrain SNP-genotyped horse DNA samples, including James Mickelson (University of Minnesota), who provided DNA for the SNP-genotyped horses used in this project. This research was funded by United States Department of Agriculture Cooperative State Research, Education, and Extension Service (USDA-CSREES) postdoctoral grant number 2008-35205-18768 to CB-L and by Morris Animal Foundation grant D07EQ-500, 'Program in Equine Medical Genetics'. The authors declare no conflict of interest.

References

Agrafioti I. & Stumpf M.P. (2007) SNPSTR: a database of compound microsatellite-SNP markers. *Nucleic Acids Research* **35**, D71–5.

Ahn J., Won T.W., Kaplan D.E., Londin E.R., Kuzmic P., Gelernter J. & Gruen J.R. (2002) A detailed physical map of the 6p reading disability locus, including new markers and confirmation of recombination suppression. *Human Genetics* **111**, 339–49.

Bailey E., Marti E., Fraser D., Antczak D. & Lazary S. (2000) Immunogenetics of the Horse. In: *The Genetics of the Horse* (Eds by A. Bowling & A. Ruvinsky), pp. 123–55. CABI Publishing, New York, NY.

Barrett J.C., Fry B., Maller J. & Daly M.J. (2005) HAPLOVIEW: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–5.

Brinkmeyer-Langford C.L., Murphy W.J., Childers C.P. & Skow L.C. (2010) A conserved segmental duplication within ELA. *Animal Genetics* **41**(Suppl 2), 186–95.

Carrington M. (1999) Recombination within the human MHC. *Immunological Reviews* **167**, 245–56.

Crow J.F. (1993) How much do we know about spontaneous human mutation rates? *Environmental and Molecular Mutagenesis* **21**, 122–9.

Cullen M., Perfetto S.P., Klitz W., Nelson G. & Carrington M. (2002) High-resolution patterns of meiotic recombination across the human major histocompatibility complex. *American Journal of Human Genetics* **71**, 759–76.

Davies C.J., Andersson L., Joosten I., Mariani P., Gasbarre L.C. & Hensen E.J. (1992) Characterization of bovine MHC class II polymorphism using three typing methods: serology, RFLP and IEF. *European Journal of Immunogenetics* **19**, 253–62.

Davies C.J., Joosten I., Andersson L. *et al.* (1994a) Polymorphism of bovine MHC class II genes. Joint report of the Fifth International Bovine Lymphocyte Antigen (BoLA) Workshop, Interlaken, Switzerland, 1 August 1992. *European Journal of Immunogenetics* **21**, 259–89.

Davies C.J., Joosten I., Bernoco D. *et al.* (1994b) Polymorphism of bovine MHC class I genes. Joint report of the Fifth International Bovine Lymphocyte Antigen (BoLA) Workshop, Interlaken, Switzerland, 1 August 1992. *European Journal of Immunogenetics* **21**, 239–58.

Dukkipati V.S., Blair H.T., Garrick D.J. & Murray A. (2006) Ovar-MHC–ovine major histocompatibility complex: role in genetic resistance to diseases. *New Zealand Veterinary Journal* **54**, 153–60.

Ellegren H. (2000) Microsatellite mutations in the germline: implications for evolutionary inference. *Trends in Genetics* **16**, 551–8.

Escayg A.P., Hickford J.G. & Bullock D.W. (1997) Association between alleles of the ovine major histocompatibility complex and resistance to footrot. *Research in Veterinary Science* **63**, 283–7.

Gaunt T.R., Rodriguez S., Zapata C. & Day I.N. (2006) MIDAS: software for analysis and visualisation of interallelic disequilibrium between multiallelic markers. *BMC Bioinformatics* **7**, 227.

Guérin G., Bailey E., Bernoco D. *et al.* (1999) Report of the International Equine Gene Mapping Workshop: male linkage map. *Animal Genetics* **30**, 341–54.

Guérin G., Bailey E., Bernoco D. *et al.* (2003) The second generation of the International Equine Gene Mapping Workshop half-sibling linkage map. *Animal Genetics* **34**, 161–8.

Gyllensten U.B. & Erlich H.A. (1993) MHC class II haplotypes and linkage disequilibrium in primates. *Human Immunology* **36**, 1–10.

Jeffreys A.J., Kauppi L. & Neumann R. (2001) Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nature Genetics* **29**, 217–22.

Kaufman J. (2000) The simple chicken major histocompatibility complex: life and death in the face of pathogens and vaccines. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences* **355**, 1077–84.

Klein J. (1987) Origin of major histocompatibility complex polymorphism: the trans-species hypothesis. *Human Immunology* **19**, 155–62.

Lincoln M.R., Ramagopalan S.V., Chao M.J., Herrera B.M., Deluca G.C., Orton S.M., Dyment D.A., Sadovnick A.D. & Ebers G.C. (2009) Epistasis among HLA-DRB1, HLA-DQA1, and HLA-DQB1 loci determines multiple sclerosis susceptibility. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 7542–7.

Martin M., Mann D. & Carrington M. (1995) Recombination rates across the HLA complex: use of microsatellites as a rapid screen for recombinant chromosomes. *Human Molecular Genetics* **4**, 423–8.

McCue M.E., Bannasch D.L., Petersen J.L. *et al.* (2012) A high density SNP array for the domestic horse and extant Perissodactyla: utility for association mapping, genetic diversity, and phylogeny studies. *PLoS Genetics* **8**, e1002451.

Meyer D. & Thomson G. (2001) How selection shapes variation of the human major histocompatibility complex: a review. *Annals of Human Genetics* **65**, 1–26.

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.

Milenkovic D., Mariat D., Swinburne J., Chadi-Taourit S., Binns M. & Guerin G. (2005) Characterization and localization of 17 microsatellites derived from BACs in the horse. *Animal Genetics* **36**, 164–6.

Mittmann E.H., Wrede J., Pook J. & Distl O. (2010) Identification of 21 781 equine microsatellites on the horse genome assembly 2.0. *Animal Genetics* **41**, 222.

Mountain J.L., Knight A., Jobin M., Gignoux C., Miller A., Lin A.A. & Underhill P.A. (2002) SNPSTRs: empirically derived, rapidly typed, autosomal haplotypes for inference of population history and mutational processes. *Genome Research* **12**, 1766–72.

Noble J.A., Valdes A.M., Varney M.D. *et al.* (2010) HLA class I and genetic susceptibility to type 1 diabetes: results from the Type 1 Diabetes Genetics Consortium. *Diabetes* **59**, 2972–9.

Park Y.H., Joo Y.S., Park J.Y., Moon J.S., Kim S.H., Kwon N.H., Ahn J.S., Davis W.C. & Davies C.J. (2004) Characterization of lymphocyte subpopulations and major histocompatibility complex haplotypes of mastitis-resistant and susceptible cows. *Journal of Veterinary Science* **5**, 29–39.

Raudsepp T., Gustafson-Seabury A., Durkin K. *et al.* (2008) A 4,103 marker integrated physical and comparative map of the horse genome. *Cytogenetic and Genome Research* **122**, 28–36.

Raymond C.K., Kas A., Paddock M. *et al.* (2005) Ancient haplotypes of the HLA Class II region. *Genome Research* **15**, 1250–7.

Roberts M.C., Murtaugh J., Valberg S.J., Mickelson J.R. & Alexander L.J. (2000) Twenty-one new equine dinucleotide repeat microsatellites. *Animal Genetics* **31**, 141.

Scheet P. & Stephens M. (2006) A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *American Journal of Human Genetics* **78**, 629–44.

Seddon J.M., Berggren K.T. & Fleeman L.M. (2010) Evolutionary history of DLA class II haplotypes in canine diabetes mellitus through single nucleotide polymorphism genotyping. *Tissue Antigens* **75**, 218–26.

Sharif S., Mallard B.A., Wilkie B.N., Sargeant J.M., Scott H.M., Dekkers J.C. & Leslie K.E. (1998) Associations of the bovine major histocompatibility complex DRB3 (BoLA-DRB3) alleles with occurrence of disease and milk somatic cell score in Canadian dairy cattle. *Animal Genetics* **29**, 185–93.

Sharif S., Mallard B.A., Wilkie B.N., Sargeant J.M., Scott H.M., Dekkers J.C. & Leslie K.E. (1999) Associations of the bovine major histocompatibility complex DRB3 (BoLA-DRB3) with production traits in Canadian dairy cattle. *Animal Genetics* **30**, 157–60.

Snoek M., Teuscher C. & van Vugt H. (1998) Molecular analysis of the major MHC recombinational hot spot located within the G7c

Published 2012. This article is a U.S. Government work and is in the public domain in the USA., doi: 10.1111/age.12003

gene of the murine class III region that is involved in disease susceptibility. *Journal of Immunology* **160**, 266–72.

Stephens M., Smith N.J. & Donnelly P. (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics* **68**, 978–89.

Thorsby E. (1997) Invited anniversary review: HLA associated diseases. *Human Immunology* **53**, 1–11.

Tozaki T., Mashima S., Hirota K., Miura N., Choi-Miura N.H. & Tomita M. (2001) Characterization of equine microsatellites and microsatellite-linked repetitive elements (eMLREs) by efficient cloning and genotyping methods. *DNA Research* **8**, 33–45.

Tseng C.T., Miller D., Cassano J., Bailey E. & Antczak D.F. (2010) Identification of equine major histocompatibility complex haplotypes using polymorphic microsatellites. *Animal Genetics* **41** (Suppl 2), 150–3.

Wade C.M., Giulotto E., Sigurdsson S. *et al.* (2009) Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science* **326**, 865–7.

Weber J.L. (1990) Human DNA polymorphisms and methods of analysis. *Current Opinion in Biotechnology* **1**, 166–71.

Weber J.L. & Wong C. (1993) Mutation of human short tandem repeats. *Human Molecular Genetics* **2**, 1123–8.

Zhao H., Nettleton D. & Dekkers J.C. (2007) Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between single nucleotide polymorphisms. *Genetical Research* **89**, 1–6.

Supporting information

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

Figure S1 Screen shots of MIDAS output.

Figure S2 HAPLOVIEW output showing linkage disequilibrium in the SNP-genotyped horses from 15 breeds (see Materials and methods).

Table S1 Details about microsatellite markers used in this study.

Table S2 Output tables generated by MIDAS analyses described in this article.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

Published 2012. This article is a U.S. Government work and is in the public domain in the USA., doi: 10.1111/age.12003